THÈSE

POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Des Interactions

École doctorale GAIA

Analyse génétique et moléculaire de l'établissement de la symbiose Nod-indépendante chez la légumineuse tropicale *Aeschynomene evenia*

Présentée par Johan QUILBE

Le 26/01/2021

Devant le jury composé de

Guilhem DESBROSSES, Professeur des Universités, Université de Montpellier	Président
Sandra BENSMIHEN, Chargée de Recherche, INRA-LIPM	Rapportrice
Pierre-Marc DELAUX, Directeur de Recherche, CNRS-LRSV	Rapporteur
Fernanda DE CARVALHO-NIEBEL, Directrice de Recherche, INRA-LIPM	Examinatrice
Jean-Michel ANÉ, Professeur des Universités, University of Wisconsin Madison	Examinateur
Jean-François ARRIGHI, Chargé de Recherche, IRD-LSTM	Directeur de thèse

Recherches effectuées au Laboratoire des Symbioses Tropicales & Méditerranéennes
IRD/CIRAD/INRA/SupAgro Montpellier/UM
Campus International de Baillarguet TA A-82/J, 34398 Montpellier Cedex 5, France

Remerciements

Mes premiers mots de remerciements vont aux membres du jury de thèse qui ont accepté d'évaluer ce travail. Les rapporteurs Sandra Bensmihen et Pierre-Marc Delaux ainsi que les examinateurs Fernanda de Carvalho-Niebel, Jean-Michel Ané et Guilhem Desbrosses avec qui j'ai pu, pour certains, partager des moments d'échanges scientifiques aux cours de différents congrès. Je tiens également à remercier les personnes qui ont constitué mon comité de thèse pendant ces trois années. A Benjamin Peret, Pascal Ratet et Clare Gough pour leur bienveillance et les discussions enrichissantes que nous avons partagées.

Je remercie aussi les nombreux collaborateurs avec qui j'ai eu la chance de travailler et qui ont pu contribuer de près ou de loin aux résultats de cette thèse, je m'excuse à l'avance si j'oublie d'en citer certains. Merci à Christophe Klopp et Léo Lamy pour les analyses bio-informatiques ; à Manuel Becana pour les analyses biochimiques ; à Hassen Gherbi, Joël Fardoux et Fabienne Cartieaux pour leurs conseils en biologie moléculaire ; à Marjorie Pervent pour son aide technique sur la QPCR ; à Ronan Rivallan et Pierre Mournet pour la création des banques ADN ; à Laurent Brottier et Amandine Delteil pour leur participation au crible de mutagénèse EMS ; et enfin aux stagiaires Thomas Benichou, Martin Phillipe et Rémi Guyonnet qui ont partagé une partie de l'épopée avec moi.

Je remercie l'ensemble des membres du laboratoire LSTM qui ont participé au quotidien au bon déroulement de cette thèse avec une mention spéciale à Naïma Rezkallah qui est toujours bienveillante et souriante, et à Nathalie Claudé qui est franche et avenante. De nouveau, merci à Marjorie Pervent qui est d'une rigueur et d'une efficacité redoutable, sans elle cette 3ème année de thèse aurait été très compliquée en plus de la crise sanitaire et aussi à Rémi Guyonnet pour l'aide qu'il ma fourni depuis ces 2 dernières années, pour sa bonne humeur et son humour.

Merci aux membres de l'équipe MSLT, Jean-François Arrighi, Eric Giraud, Fabienne Cartieaux, Marjorie Pervent, Nico Nouwen, Joël Fardoux, Djamel Gully, Clémence Chaintreuil, Frédéric Gressent qui ont tous participé à l'aventure *Aeschynomene* et qui m'ont souvent aidé de façon technique ou théorique. Merci à Nico Nouwen avec qui j'ai beaucoup apprécié discuter d'hypothèses scientifiques ou de sujets personnels. Tu es un scientifique brillant.

Je remercie beaucoup Eric Giraud de m'avoir accueilli dans son équipe avec enthousiasme et d'être aussi sympathique avec les non-permanents, comme avec les permanents d'ailleurs. J'ai eu plaisir de voir que les expériences conduites pendant cette thèse t'ont interessé et excité, je suis content d'avoir apporté ma pierre à l'édifice sur l'étude de ce système si particulier *Bradyrhizobium-Aeschynomene* qui te tient à cœur.

Mais mon plus grand remerciement va à Jean-François Arrighi, mon directeur de thèse. Depuis le stage de M2 nous travaillons ensemble, 4 ans que nous nous cotoyons et j'ai pu voir à quel point vous êtes un scientifique brillant et perfectionniste. J'ai beaucoup appris à vos côtés sur le sujet passionant de la symbiose et ce furent de riches années pour moi. Je suis reconnaissant car vous avez su voir en moi un potentiel dont certains doutaient. Vous m'avez grandement appuyé dans tous les moments clés, depuis la soutenance de M2, le concours de l'école doctorale, les comités de thèse et maintenant la rédaction du manuscrit. Pour tout cela je vous dis merci et j'espère le meilleur pour vous dans les années à venir.

Merci à tous les non-permanents que j'ai pu cotoyer pendant ces 3 années et particulièrement à Doriane Dijon, Bryan Vincent, Rémi Guyonnet, Stephan Boivin, Thomas Benichou, Maïwen Lhoir, Alicia Camuel, Bouchra Nasslhassem et Fernando Clavijo. Merci aussi à Albin Teulet et Nicolas Busset qui ont été de très bonne compagnie et avec qui j'ai pu partager sur leur expérience de thèse.

Merci à mes proches et plus particulèrement mes amis d'enfance Ronan Martin et Nicolas Jarlan, ma famille, mes parents qui m'on soutenu dans tous mes choix et ma femme Maëva Grassi qui m'a épaulé pendant tout ce temps. Vous êtes le noyau dur et incassable qui m'a permi d'avancer jusqu'à aujourd'hui.

Une pensée aux personnes qui nous ont quitté entre le début de cette thèse et aujourd'hui, et qui auraient été heureux de pouvoir assister à la fin de cette aventure.

Pour finir merci à vous qui me lisez et que vos chemins soient bercés par la joie et la réussite.

Résumé

Plusieurs espèces appartenant au genre tropical *Aeschynomene* développent une symbiose fixatrice d'azote originale avec des *Bradyrhizobium* photosynthétiques, qui ne passe pas par la reconnaissance des facteurs Nod rhizobiens et la formation de cordons d'infection. Pour mettre à jour les mécanismes moléculaires de cette symbiose dite Nod-indépendante, une approche de génétique directe est mise en œuvre chez *Aeschynomene evenia*. Elle combine le crible d'une population de mutagenèse EMS, ayant permis d'isoler 250 mutants de nodulation, au séquençage du génome d'*A. evenia* pour faciliter l'identification de gènes symbiotiques à partir de ces mutants.

Au cours de ma thèse, j'ai mis à profit le génome nouvellement disponible d'*A. evenia* en menant une analyse comparative de gènes symbiotiques afin d'obtenir une vue d'ensemble des mécanismes potentiellement conservés ou divergents entre *A. evenia* et les légumineuses modèles *M. truncatula* et *L. japonicus*. Ce travail a mis en exergue des différences génétiques importantes, notamment au niveau des étapes de reconnaissance des rhizobiums et d'infection. Ensuite, pour déterminer quels gènes connus chez les légumineuses modèles sont aussi impliqués dans la symbiose Nod-indépendante, j'ai recherché des mutations par Targeted Sequence Capture de ces gènes sur les 250 mutants de nodulation d'*A. evenia*. Cette approche a révélé que nombreux mutants Nod$^-$, qui sont bloqués dans les étapes les plus précoces de l'interaction, sont altérés dans plusieurs gènes de la voie de signalisation symbiotique activée par les rhizobiums : *AePOLLUX, AeCCaMK, AeCYCLOPS, AeNIN* et *AeNSP2*. En revanche, aucune mutation n'a été trouvée dans les gènes codant des récepteurs symbiotiques.

Pour identifier chez *A. evenia* de nouveaux gènes clé de la symbiose Nod-indépendante, j'ai appliqué une approche sans *a priori*, le Mapping-by-Sequencing, sur les mutants Nod$^-$ n'ayant pas de gène connu associé. Cette stratégie a été fructueuse pour identifier deux gènes : *AeRLCK*, qui code un Receptor-Like Cytoplasmic Kinase (RLCK), et *AeCRK*, codant une Cysteine-rich Receptor-like Kinase (CRK). Il est notable que ces deux gènes ne sont pas retrouvés chez les légumineuses modèles et qu'ils appartiennent à deux familles d'acteurs de signalisation précoce pouvant participer à la formation de complexe

récepteurs avec des Receptor-Like Kinases (RLK) pour assurer la transduction d'un signal extracellulaire. Une caractérisation de ces deux nouveaux acteurs symbiotiques a révélé qu'*AeRLCK* est issu d'une duplication en tandem propre aux *Aeschynomene* Nod-indépendantes et qu'*AeCRK* est soumis à une sélection purifiante chez ces mêmes espèces, ce qui suggère qu'ils pourraient fonctionner spécifiquement dans la symbiose Nod-indépendante. Pour préciser leur rôle et les positionner par rapport à la voie de signalisation symbiotique conservée avec les légumineuses modèles, j'ai également initié une analyse comparative fine au niveau phénotypique et moléculaire.

La découverte d'*AeRLCK* et *AeCRK* représente une avancée importante dans notre compréhension des mécanismes moléculaires qui contrôlent l'établissement de la symbiose Nod-indépendante. Elle ouvre maintenant la voie à une analyse fonctionnelle de ces deux acteurs symbiotiques et à une recherche des récepteurs symbiotiques avec qui ils peuvent interagir.

Mots-clés : Symbiose Nod-indépendante, *Aeschynomene*, *Bradyrhizobium*, Génétique, Mutagénèse

Abstract

Many species belonging to the tropical genera *Aeschynomene* develop an original kind of nitrogen-fixing symbiosis with photosynthetic *Bradyrhizobia* that is not based on Nod Factors perception or infection threads formation. To uncover the molecular mechanisms governing this so-called Nod-independent symbiosis, a forward genetic approach has been developed in Aeschynomene evenia. This approach combines the analysis of an EMS mutagenesis population, that allowed the isolation of 250 nodulation mutants, and *A. evenia*'s genome sequencing, that facilitate the identification of symbiotic genes from these mutants.

During this thesis, I utilized the recently sequenced genome of *A. evenia* to conduct a comparative analysis of symbiotic genes, that showed an overview of the conserved or divergent mechanisms comparing to M. truncatula and L. japonicus. This work highlights major genetic differences, in particular during perception and infection steps. After that, to determine which known genes from model legumes are involved in Nod-independent symbiosis, I search mutations with the Targeted-Sequence Capture approach on the 250 *A. evenia* nodulation mutants. This method reveals that many Nod- mutants, that are blocked in the early steps of the interaction, are impaired in genes of the Nod signaling pathway activated by Rhizobia : *AePOLLUX, AeCCaMK, AeCYCLOPS, AeNIN,* and *AeNSP2*. However, no mutations have been found in receptor coding genes.

To identify new key symbiotic genes in *A. evenia*, I performed an approach with no *a priori*, called Mapping-by-Sequencing, on Nod- mutant for which we have no mutations associated. This successful strategy allows the identification of two new genes: *AeRLCK*, coding a Receptor-Like Cytoplasmic Kinase (RLCK) and *AeCRK*, coding a Cystein-rich Receptor-like Kinase (CRK). Notably, these two genes are not present in other model legumes and belong to 2 early signaling actors family wich may participate in the formation of a receptors complex with RLKs to ensure the signal transduction from the membrane. The characterization of these new symbiotic actors revealed that *AeRLCK* originates from a tandem duplication event specific to the Nod-independent *Aeschynomene* and that *AeCRK* is under purifying selection in these species, suggesting that they have specific functions during Nod-independent symbiosis. To precise their role and to situate them in the Nod signaling

pathway conserved with model legumes, I also initiated a phenotypic and molecular comparative analysis.

The *AeRLCK* and *AeCRK* discoveries represent major advances in our understanding of the molecular mechanisms governing the Nod-independent symbiosis. That now paves the way to functional analyses and to the quest of symbiotic receptors with which they can interact.

Key-words : Nod-independent symbiosis, *Aeschynomene*, *Bradyrhizobium*, Genetics, Mutagenesis

Table des matières

Remerciements 1

Résumé 3

Introduction générale 13

1. La symbiose fixatrice d'azote, intérêt écologique et agronomique 14

1.a - L'azote, un enjeu majeur pour la croissance des plantes 14

1.b - Rôle clé des organismes diazotrophes dans la fixation biologique de l'azote 16

1.c - Diversité des symbioses fixatrices d'azote 18

1.d - Le succès des légumineuses grâce à la nodulation fixatrice d'azote 20

2. Mécanismes de la symbiose rhizobienne chez les légumineuses modèles 22

2.a - Les facteurs Nod : molécules signal clés de la symbiose rhizobium-Légumineuse 24

2.b - Reconnaissance des facteurs Nod par la plante-hôte 27

2.c - La voie de signalisation Nod 30

2.d - L'infection symbiotique 36

2.e - Mise en place et fonctionnement du nodule fixateur d'azote 41

3. Unité et diversité des mécanismes de nodulation 48

3.a - Les origines de la nodulation 49

3.b - Diversification des mécanismes de nodulation chez les légumineuses 55

3.c - Les processus d'infection symbiotique alternatifs chez les légumineuses 57

4. Originalités du système symbiotique *Bradyrhizobium-Aeschynomene* 62

4.a - Les *Aeschynomene* : des légumineuses à nodulation de tige 62

4.b - L'existence d'une symbiose indépendante des facteurs Nod 64

4.c - *Aeschynomene evenia* comme espèce modèle de la symbiose FN-indépendante 67

4.d - Approches génétiques et moléculaires pour mettre à jour les mécanismes de la symbiose NF-indépendantes chez *A. evenia* 70

5. Présentation du projet de thèse 73

Chapitre 1 74

1.1 - Contexte de l'étude 75

1.2 - Article 1 : 76

1.3 - Données complémentaires 101

Chapitre 2 135

2.1 - Contexte de l'étude 136

Article 2 : Caractérisation de la voie de signalisation Nod-indépendante chez la légumineuse *Aeschynomene evenia* 136

2.2 - Article 2 : 137
2.3 - ADDENDUM : NSP2, a key symbiotic regulator in the spotlight 159
Chapitre 3 166
3.1 - Contexte de l'étude 167
3.2 - Article 3 : 168
3.3 - ADDENDUM : Identification de trois gènes requis pour l'infection intercellulaire chez *Aeschynomene evenia* 187
Discussion générale 196
1. Les LysM-RLK ne semblent pas impliqués dans la symbiose NF-indépendante..... 198
2. La voie de signalisation Nod est (au moins partiellement) conservée..... 199
3. Un nouveau récepteur-kinase, AeCRK, est requis pour la symbiose Nod-indépendante..... 200
4. Un nouveau RLCK, AeRLCK, est aussi essentiel pour initier la symbiose NF-indépendante 201
5. Mécanismes de l'infection intercellulaire dans un contexte Nod-indépendant..... 202
6. Les *Aeschynomene* : modèles d'étude de la diversité des mécanismes symbiotiques..... 203
Bibliographie.....204

Liste des figures et des tableaux

INTRODUCTION

Figure 1 : Schéma simplifié du cycle de l'azote terrestre

Figure 2 : Représentation schématique du complexe nitrogénase et des cofacteurs contenant du métal associés

Figure 3 : Différents types d'associations endophytes intercellulaires avec des cyanobactéries fixatrices d'azote

Figure 4 : Associations symbiotiques intracellulaires

Figure 5 : Evolution de la symbiose fixatrice d'azote chez les plantes terrestres

Figure 6 : Phylogénie des Leguminosae basée sur la séquence protéique de 81 gènes plastidiques et obtenue par inférence Bayésienne

Figure 7 : Les légumineuses en rotation de culture

Figure 8 : Modèle schématique du développement d'un nodule chez *Medicago truncatula*

Figure 9 : Structure des facteurs Nod et régulation de leur synthèse

Figure 10 : Représentation schématique de 2 groupes de LysM-RLK auxquels appartiennent NFP et LYK3

Figure 11 : Les complexes récepteurs LysM-RLKs sont impliqués dans la perception des micro-organismes

Figure 12 : Gènes et processus impliqués dans les étapes précoces de la nodulation

Figure 13 : Réseau de régulation génétique de l'infection rhizobienne par trois facteurs de transcription NF-YA1, ERN1 et NIN

Figure 14 : L'infection rhizobienne

Figure 15 – Etapes de développement de nodules déterminés et indéterminés

Figure 16 : Evolution de la nodulation chez *Parasponia*, la seule non-legumineuse à interagir symbioiquement avec des rhizobiums

Figure 17 : Profils phylogénétiques des gènes de la symbiose NFN (détailler NFN)

Figure 18 : Représentation de la distribution phylogénétique de plantes nodulantes et non-nodulantes du clade NFN sous l'hypothèse d'un gain unique suivi de multiples pertes de la nodulation

Figure 19 : Duplication ancestrale du génome des Papillonacées datée de 58 millions d'années

Figure 20 : Les deux grands types d'infection rencontrés chez les Légumineuses

Figure 21 : Deux modes d'infection possibles chez la légumineuses *Sesbania rostrata* avec son symbionte *Azorhizobium caulinodans*

Figure 22 : La souche de *Rhizobium leguminosarum* Norway colonise les nodules de *Lotus burtii* en absence de cordon d'infection transcellulaire

Figure 23 : Modèle comparatif des gènes nécessaires pour les programmes d'infection intra- vs inter-cellulaire chez *Lotus japonicus*.

Figure 24 : Les *Aeschynomene*, des légumineuses aux propriétés de nodulation originales.

Figure 25 : Groupes d'inoculations croisées chez les *Aeschynomene.*

Figure 26 : Efficience de la nodulation de tige avec *Bradyrhizobium* ORS278 et ORS285 (et le mutant ORS285Δ*nodB* ne produisant plus de facteurs Nod).

Figure 27 : Arbre phylogénétique (ITS + *trnL*) du genre *Aeschynomene* et d'espèces appartenant à des genres associés.

Figure 28 : Représentation schématique de la cascade de signalisation symbiotique avec les gènes identifiés chez les légumineuses modèles

Figure 29 : Les approches génétiques et les outils disponibles pour étudier les gènes symbiotiques chez les légumineuses.

CHAPITRE 1

Figure 1 : Structure and evolution of the *Aeschynomene evenia* genome

Figure 2 : Comparative analysis of symbiotic receptors, signaling, and infection genes

Figure 3 : Comparative analysis of endoreplication-mediated nodule differentiation

Figure 4 *NCR* genes in the *A. evenia* genome

Figure 5 : Symbiotic and nonsymbiotic globins of *A. evenia*

Figure 6 : Candidate genes of the known symbiotic signaling pathway identified by Targeted Sequence Capture.

Figure 7 : Novel gene involved in the establishment of the NF-independent symbiosis identified by Mapping-by-Sequencing

CHAPITRE 2

Figure 1 : Absence de couronnes de poils axillaires chez les mutants *nsp2* d'*A. evenia*

Figure 2 : Rôle de l'azote et d'*AeNSP2* dans le contrôle de l'expression de gènes réprimés au cours de la nodulation

Figure 3 : Phénotypes de nodulation des mutants *pollux*, *ccamk*, *nsp2* et *nin* d'*A. evenia*

Figure 4 : Phénotypes de nodulation des mutants *cyclops* et *crk* d'*A. evenia*

Figure 5 : Rôle des gènes de signalisation symbiotique dans l'induction de gènes de nodulines précoces chez *A. evenia*

Figure 6 : Phénotypes de mycorhization des mutants des gènes de signalisation d'*A. evenia*

Figure 7 : Analyse de l'expression de gènes marqueurs de la mycorhization chez *A. evenia*

ADDENDUM CHAPITRE 2

Box 1 : Model of NSP2 as a central regulatory hub integrating symbiotic and nutritional cues to mediate symbiotic and non-symbiotic responses

CHAPITRE 3

Figure 1 : Phénotype de nodulation des mutants « One Big Nodule »

Figure 2 : Identification et validation du gène *AeRLCK* comme étant responsable du phénotype OBN

Figure 3 : Infection et organogenèse nodulaire chez les mutants OBN

Figure 4 : Nodulation caulinaire chez les mutants OBN

Figure 5 : Phylogénie d'*AeRLCK*

Figure 6 : Structure du locus Ae01g26600 et reconstitution de l'évènement de duplication en tandem à l'origine d'AeRLCK

Figure 7 : Profils d'expression d'*AeRLCK* et des gènes les plus proches chez *A. afraspera* et *M. truncatula*

Tableau 1 : Données phénotypiques, génétiques et moléculaires des mutants OBN d'*A. evenia*

Tableau 2 : Tests d'allélismes entre les mutants OBN d'*A. evenia*

Tableau 3 : Test de complémentation fonctionnelle du mutant *rlck*

ADDENDUM CHAPITRE 3

Figure 1 : *AePLT*, phénotype des mutants et caractéristiques du gène

Figure 2 : *AeRINRK*, phénotype des mutants et caractéristiques du gène

Figure 3 : *AeRLP*, phénotype des mutants et caractéristiques du gène

Figure 4 : Rôle des gènes de la voie de signalisation Nod-indépendante dans l'induction de l'expression de gènes d'infection au cours de la nodulation

Tableau 1 : Données phénotypiques, génétiques et moléculaires des mutants d'infection d'*A. evenia*

Tableau 2 : Tests d'allélismes des mutants d'*AePLT*

Introduction générale

1. La symbiose fixatrice d'azote, intérêt écologique et agronomique

1.a - L'azote, un enjeu majeur pour la croissance des plantes

Parmi les éléments chimiques nécessaires au développement des êtres vivants, on retrouve tout d'abord les 4 essentiels qui sont le carbone (C), l'hydrogène (H), l'oxygène (O) et l'azote (N). On trouve ensuite ceux généralement sous formes ioniques qui sont nécessaires au fonctionnement de la machinerie cellulaire comme : le phosphore (P), le soufre (S), le sodium (Na), le potassium (K), le magnésium (Mg), le calcium (Ca), le chlore (Cl). Viennent ensuite de nombreux éléments traces comme le fer (Fe), le molybdène (Mo) ou encore le cuivre (Cu), qui sont présents en très faible quantité dans les organismes mais qui sont importantes pour certaines activités biologiques spécifiques. Chez les plantes, la source de carbone organique est générée lors de la réaction chimique de photosynthèse grâce à la lumière, au gaz carbonique présent dans l'air et à l'hydrogène/oxygène constituant les molécules d'eau. Les autres éléments chimiques qu'on appelle alors macro- et/ou micro-éléments sont puisés directement dans le sol grâce au système racinaire de la plante. L'azote, le phosphore et le potassium sont souvent rares et difficilement accessibles dans les sols alors que les plantes sont fortement dépendantes de ces ressources qu'elles ne peuvent trouver que dans leur environnement proche.

L'azote joue un rôle important pour le développement de la vie car il entre notamment dans la composition des acides nucléiques, qui sont la base de l'ADN et de l'ARN, et des acides aminés qui sont les constituants élémentaires des protéines. Dans les sols, l'azote assimilable par les plantes correspond aux formes ioniques non complexées (NO_3^- et NH_4^+) qui sont présentes en très faible quantité, et représentent donc un facteur majeur limitant la croissance des plantes terrestres (Kaye and Hart, 1997). Les formes d'azote assimilables sont principalement générées par des phénomènes biotiques comme le métabolisme des organismes hétérotrophes, qui permet la décomposition de la matière organique contenant de l'azote, ainsi que la fixation biologique de l'azote. Des phénomènes abiotiques comme la foudre, la combustion ou encore le volcanisme interviennent également mais cela ne représente qu'une très faible source d'azote assimilable pour le fonctionnement des écosystèmes (Fig. 1). Depuis que les hommes sont capables de cultiver des plantes, la

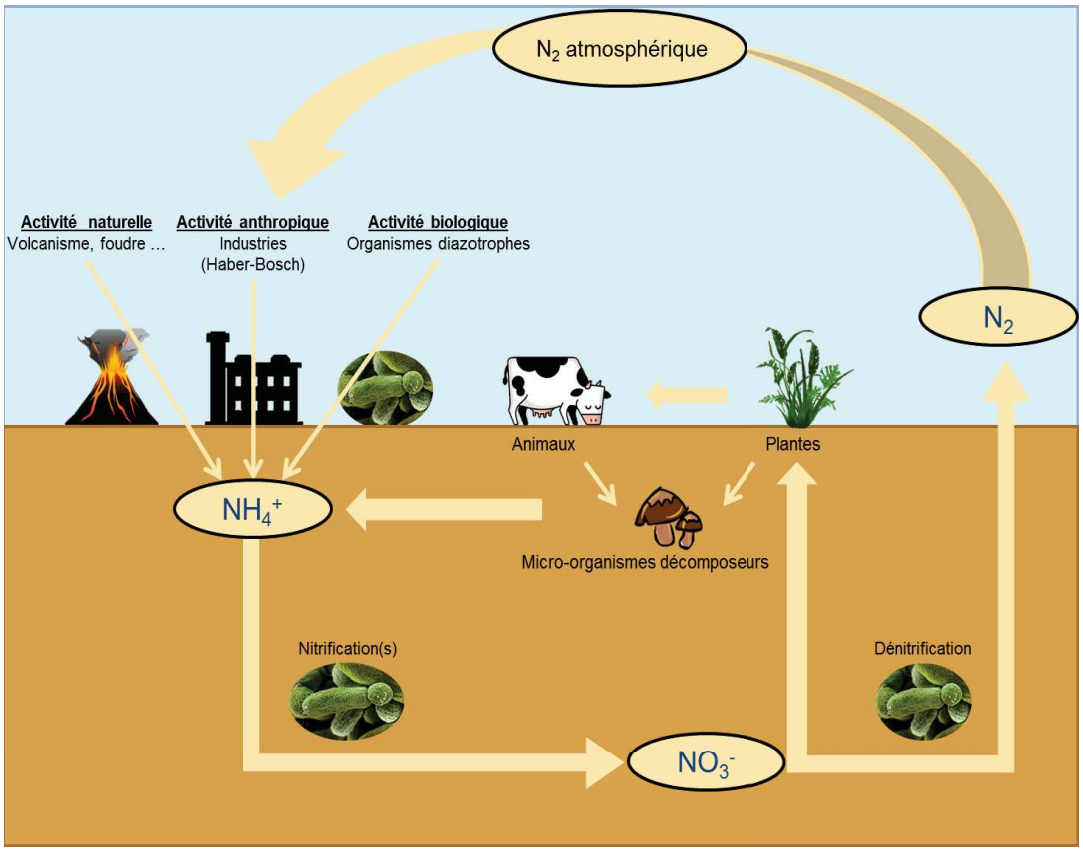


Figure 1 : Schéma simplifié du cycle de l'azote terrestre. Les 3 principales activités génératrices de $NH4^+$ sont listées a gauche. Les flèches représentent les flux des différentes formes d'azote entre les compartiments (atmosphère, géosphère et biosphère).

concentration en azote des sols a posé des problèmes de rendement des cultures. Certains peuples ont développé des procédés pour enrichir les sols comme par exemple l'utilisation des excréments d'animaux riches en éléments azotés ou encore en utilisant des rotations de culture avec des plantes ne dépendant pas de l'azote des sols pour se développer comme les légumineuses (Voisin *et al.*, 2014). Plus récemment, avec l'avènement de l'ère industrielle, un procédé physico-chimique très coûteux en énergie nommé Haber-Bosch a été inventé et permet l'hydrogénation du diazote afin de former l'ion NH3+ qui rentre dans la composition des engrais chimiques (Stranges, 2001). L'utilisation massive de ces engrais chimiques (intrants) en agriculture au cours du dernier siècle a permis d'augmenter considérablement les rendements, c'est ce que l'on a appelé la révolution verte. Le succès des engrais chimiques est tel que sa production représentait en 2018 environ 120 millions de tonnes d'azote par an (Statista 2020). En comparaison, la fixation biologique de l'azote représente environ 110 millions de tonnes/an en milieu terrestre et 140 millions de tonnes/an dans les océans (Gruber and Galloway, 2008). La source anthropique correspond donc à environ 50% de l'apport d'azote dans le cycle biogéochimique naturel. Les mauvaises pratiques culturales comme la monoculture, la dispersion d'engrais (azote), la pulvérisation de pesticides et le labour trop profond des sols ont de nombreuses conséquences négatives. En effet, le système agricultural actuel trop intensif a fini par épuiser les sols et provoque des effets délétères sur l'environnement étant donné que les plantes sont incapables d'assimiler la totalité de l'engrais qui leur est apporté. Le surplus d'engrais azoté pollue les nappes phréatiques en traversant les sols ou sera lessivé par la pluie et rejoindra les cours d'eau. Cette pollution entraîne une élévation du taux de nitrates (NO_3^-) et des autres éléments contenus dans les engrais, cela conduit à une pollution des écosystèmes aquatiques qui a potentiellement un impact sur la santé humaine et sur l'environnement qui peut aller jusqu'à une disparition de la biodiversité (Fields, 2004). D'autre part, les bactéries dénitrifiantes naturellement présentes dans les sols et jouant habituellement un rôle clé dans le cycle de l'azote, augmentent considérablement la production d'oxydes nitreux qui sont néfastes sur l'environnement en contribuant notamment à l'effet de serre et à la dégradation de la couche d'ozone.

C'est pourquoi, la nécessité de développer de nouvelles pratiques agriculturales qui permettent de maintenir de bons rendements, tout en étant durables et respectueuses de

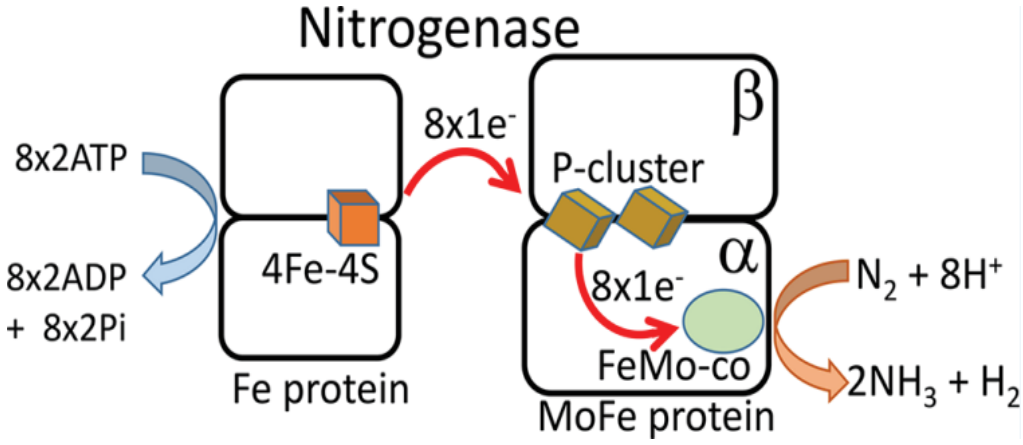


Figure 2 : Représentation schématique du complexe nitrogénase et des cofacteurs contenant du métal associés. La protéine Fe ou dinitrogénase réductase Fe (gauche) utilise de l'énergie sous forme d'ATP pour délivrer les électrons un par un au second composant, la protéine catalytique MoFe ou oxydoréductase Molybdène-Fer (droite). Le transfert d'électron est un processus en plusieurs étapes impliquant trois « metal clusters » et aboutissant à la formation de molécules de NH_3 (Seefeldt *et al.,* 2018)

l'environnement, est urgente. Aujourd'hui, une prise de conscience collective sur les sujets touchant à l'alimentation, l'agriculture et l'environnement donne l'espoir de voir ces pratiques agriculturales se modifier. La communauté scientifique va aider à trouver des solutions pour repenser les modes de production, ce qui pourrait conduire à une seconde « révolution verte » en tenant compte, cette fois des enjeux environnementaux.

1.b – Rôle clé des organismes diazotrophes dans la fixation biologique de l'azote

Bien que l'atmosphère terrestre soit composée à 80% de diazote (N_2) et représente ainsi un immense réservoir, ce diazote n'est pas directement utilisable par la très grande majorité des organismes vivants car il est constitué d'une triple liaison covalente qui le rend extrêmement stable. Seuls certains procaryotes sont capables d'utiliser cette source d'azote minéral (N_2). Ces bactéries dites « diazotrophes » sont les grands acteurs de la fixation biologique de l'azote. Elles regroupent plusieurs clades éloignés par des millions d'années d'évolution, ubiquité qui souligne l'intérêt de ce trait évolutif. Il y a des archées (*Euryarchaeota*), des cyanobactéries (*Nostocaceae, Microcoleaceae…)*, des protéobactéries (*Rhizobiaceae*) ou encore des actinomycètes (*Frankiaceae*). Malgré leurs divergences évolutives, les bactéries diazotrophes ont en commun de posséder tous les gènes *nif* et *fix* nécessaires à la synthèse et au fonctionnement du complexe enzymatique de la nitrogénase qui est capable de rompre la triple liaison du diazote pour former l'ion ammonium NH_4^+ (Sprent, 2007). Ce complexe protéique est formé par deux métalloprotéines, une dinitrogénase (protéine MoFe) et une dinitrogénase réductase (protéine Fe) qui est, elle même un homo-dimère (nifH) avec un site de fixation pour l'ATP et un centre Fer-Soufre (4S-4Fe) (Fig. 2). La dinitrogénase MoFe possède deux centres Fer-Soufre et un co-facteur 7Fe-Mo-9S-C-Homocitrate (FeMo-Co). Malgré la structure sophistiquée de ce complexe protéique, ce système présente une faiblesse à cause de la présence de métalloprotéines et de leurs précurseurs (notamment *nifB*) qui sont très sensibles à l'oxygène, rendant le fonctionnement de la nitrogénase possible uniquement dans un environnement pauvre en O_2. Lors du fonctionnement du complexe, la dinitrogénase réductase oxyde un cofacteur en consommant 2 ATP pour assurer la transmission d'électrons jusqu'au facteur FeMo-Co assurant la réduction du diazote en ammonium selon la stœchiométrie suivante (Fig. 2) :

$N_2 + 8\ H^+ + 8\ e^- + 16\ ATP \rightarrow 2\ NH_3 + H_2 + 16\ ADP + 16\ PI$

Pour transformer une molécule de N_2 en 2 molécules de NH_3, la nitrogénase consomme l'équivalent de 25 ATP (16 + fonctionnement co-facteurs), élevant ce procédé au rang d'une des réactions les plus coûteuses en énergie du monde vivant, ce qui peut expliquer que ce trait est présent chez très peu d'espèces de bactéries (Simpson and Burris, 1984).

Parmi les bactéries diazotrophes, on retrouve différents types d'écologie, d'une part des fixateurs libres produisant leur propre azote qui finit par enrichir l'environnement dans lequel ils se développent, et d'autre part des micro-organismes vivants en association avec des organismes eucaryotes. Cette fixation biologique de l'azote joue un rôle fondamental dans le cycle de l'azote au niveau terrestre et aquatique. Dans l'océan, les cyanobactéries sont responsables de la majorité de l'activité fixatrice d'azote et fournissent un apport de molécules azotées à l'ensemble des communautés d'eucaryotes photosynthétiques comme le phytoplancton, qui alimente ensuite l'ensemble de la chaine trophique, jusqu'aux organismes supérieurs**.** Sur terre, ce sont les α-protéobactéries qui sont responsables de la majorité de la fixation biologique et qui apportent une source de molécules azotées dans les sols et les cours d'eau (Falcón *et al.*, 2004; Sprent and Sprent, 1990). A cause de la sensibilité de la nitrogénase à l'oxygène, les micro-organismes ont développé différents systèmes afin de protéger ce complexe enzymatique clé. Chez certaines espèces de cyanobactéries (*Anabaena flosaquae*), il existe une séparation d'identité cellulaire pour les fonctions de fixation d'azote et de photosynthèse. Ainsi, les hétérocystes sont des cellules spécialisées dans la fixation d'azote qui arborent une paroi très épaisse, réduisant les échanges d'O_2 avec l'extérieur et permettant l'activité nitrogénase (Kangatharalingam *et al.*, 1992). Il y a aussi des espèces de cyanobactéries fixatrices d'azote qui sont incapables de se différenciées en hétérocystes et qui ont une stratégie différente pour protéger la nitrogénase de l'oxygène. En effet, les bactéries du genre *Trichodesmium* sont capables de réaliser une séparation temporelle en réalisant la fixation d'azote la nuit et la photosynthèse le jour (Milligan *et al.*, 2007). En vie libre, les actinomycètes du genre *Frankia* sont également capables de produire des cellules spécialisées dans la fixation d'azote, leur paroi est modifiée par la formation d'une enveloppe lipidique multi-lamellaire (Ghodhbane-Gtari *et al.*, 2014). Cependant, pour réguler la pression partielle en O_2 lors de l'association symbiotique chez les légumineuses et chez la majorité des plantes actinorhiziennes, une stratégie différente s'est mise en place au cours de l'évolution. Sous le contrôle de la plante-hôte, une protéine est produite pour

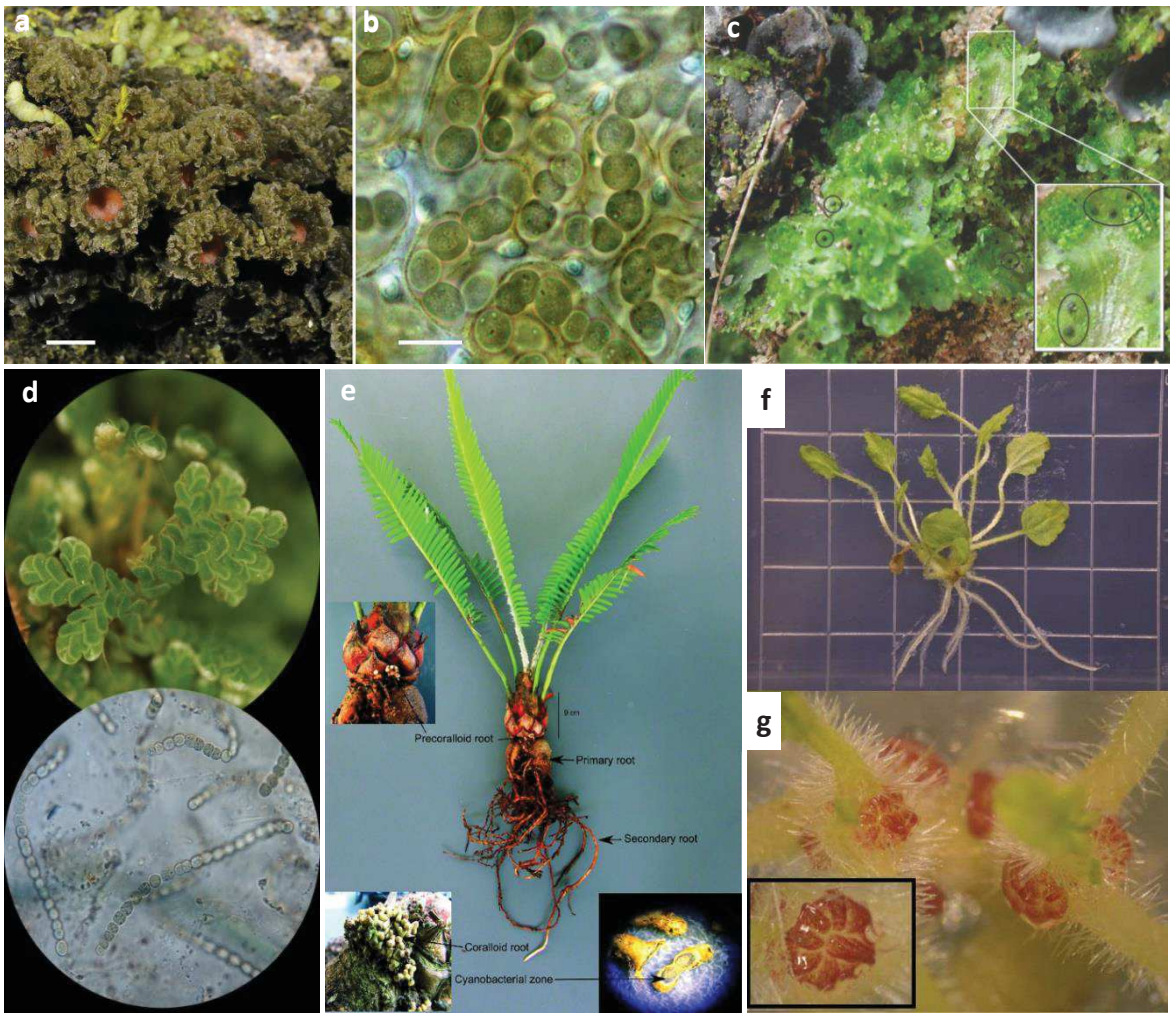


Figure 3 : Différents types d'associations endophytes intercellulaires avec des cyanobactéries fixatrices d'azote. a) Thalles de cyano-lichens (*Leptogium sp./Nostoc* sp.) b) Symbiontes au travers de la partie supérieure hydratée du thalle fongique. c) Hépatiques (*Blasia pusila/Nostoc* sp.) avec zoom sur des colonies intercellulaires (point noir). d) *Azolla* en haut et le symbionte en bas (*Azolla Filiculoides/Anabaena azollae*). e) Cycas et photos de racines coralloïdes (*Dioon edule/Nostoc* sp.). f) Gunnera âgée de 5 semaines (*Gunnera manicata/Nostoc* sp.). g) Glandes rouges se développant sur la tige a la base de chaque pétiole foliaire.

réguler la pression partielle en O_2 lors de l'interaction. Cette protéine que l'on nomme Leghémoglobine (Lhb) appartient à la famille des globines, elle assure le transport des molécules d'O_2 pour l'apporter aux bactéries qui en ont besoin pour leur respiration, tout en diminuant la pression partielle en O_2 afin de ne pas inhiber l'activité de la nitrogénase (Bhattacharya *et al.*, 2013; Downie, 2005).

1.c – Diversité des symbioses fixatrices d'azote

Certains organismes eucaryotes ont su tirer parti de l'incroyable capacité fixatrice d'azote des bactéries diazotrophes en s'associant avec elles pour développer ce que l'on appelle une symbiose mutualiste. Cette association garantit à l'hôte une source d'azote constante tout au long de sa vie sans dépendre de l'environnement dans lequel il grandit tandis que le symbionte profite de l'apport de molécules carbonées/sucres ou autre composés fournis par l'hôte. Dans la diversité des symbioses fixatrices d'azote, on distingue des associations rudimentaires le plus souvent avec des cyanobactéries compatibles et des associations plus sophistiquées souvent avec des rhizobiums et des actinomycètes. Les cyanobactéries sont capables de s'associer avec un large panel d'hôtes comme des plantes, des champignons, des éponges et des protistes (Usher *et al.*, 2007). Par exemple, certaines cyanobactéries fixatrices d'azote s'associent avec des champignons basidiomycètes, le résultat de cette association est communément appelé « cyano-lichen » qui représente environ 15% de la diversité des lichens (les 85% restants s'associent avec des micro-algues vertes). Le champignon fournit l'eau et les minéraux aux cyanobactéries qui en retour fournissent sucres et composés azotés. Cette association mutualiste permet à ces deux types de micro-organismes de coloniser précocement des milieux extrêmes qui sont souvent pauvres en matière organique et en eau, comme la surface des rochers ou les sols nus, et ainsi participer à la succession écologique en tant qu'espèce pionnière (Nash, 1996; Rikkinen, 2015) (Fig. 3a). D'autres cyanobactéries interagissent avec des plantes appartenant à des lignées évolutives diverses comme les Hépatiques (Bryophyte), la fougère aquatique *Azolla* (Pterydophyte), les *Cycas* (Gymnosperme), ou encore les *Gunnera* (Angiosperme). Pour les hépatiques, seuls 4 genres parmi les 340, sont connus pour s'associer avec des cyanobactéries du genre *Nostoc* : *Marchantia* et *Porella* forment une association épiphyte, c'est-à-dire que les bactéries vivent à la surface des thalles, alors que *Blasia* et *Cavicularia* forment une association endophyte. Chez ces deux derniers genres d'hépatiques, les

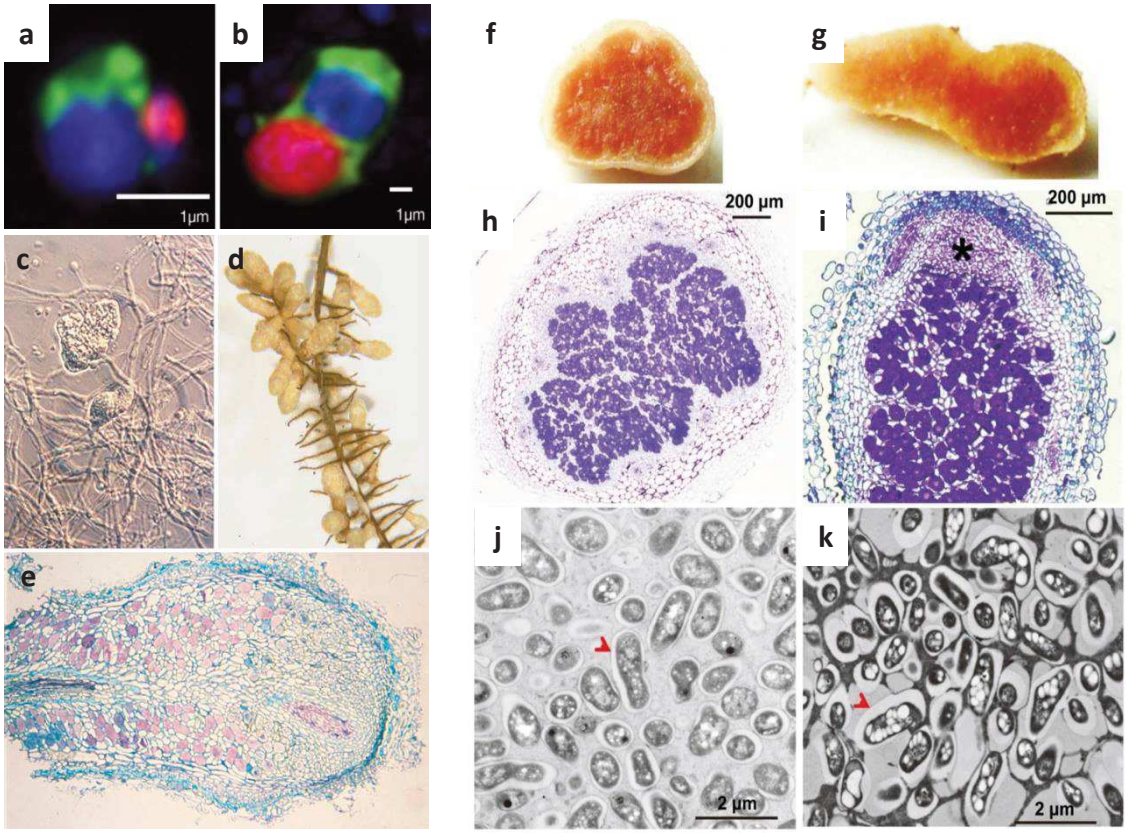


Figure 4 : Associations symbiotiques intracellulaires. a-b) Pico-algue en rouge, cyanobactérie en vert et noyau en bleu (*Prymnesiophyte/Atelocyanobacterium thalassa*). c,d,e) Symbiose actinorhizienne (*Allocasuarina verticillata/Frankia* sp.). f,h,j) Coupe de nodule déterminé de *Vigna unguiculata/Sinorhizobium* sp. NGR234. g,i,k) Coupe de nodule indéterminé de *Leucaena leucocephala/Sinorhizobium* sp. NGR234 (Li *et al*., 2013)

cyanobactéries forment des colonies sombres d'environ 1mm de diamètre dans l'espace intercellulaire de la partie ventrale des thalles (Adams and Duggan, 2008) (Fig. 3c). Ces exemples montrent qu'il existe des associations où les plantes peuvent héberger des cyanobactéries dans leurs tissus de manière intercellulaire mais aucunes structures spécialisées ne sont induites. En revanche, chez *Azolla* qui s'associe avec la cyanobactérie *Anabaena azollae* du genre *Nostoc*, on retrouve de petites cavités spécialisées au-dessous des feuilles présentes constitutivement, où se logent les bactéries formant des hétérocystes qui fixent l'azote pour leur hôte (Peters *et al.*, 1976) (Fig. 3d). Le genre *Cycas* contient des espèces qui s'associent avec des cyanobactéries du genre *Nostoc*. Au cours de leur développement, les racines latérales acquièrent un géotropisme négatif et remontent à la surface du sol puis se divisent pour former des structures boursoufflées et ramifiées qui ressemblent à du corail (Fig. 3e). Le terme de racines coralloïdes a donc été utilisé pour désigner ces structures particulières qui abritent les cyanobactéries majoritairement différenciées en hétérocystes dans la zone intercellulaire du cortex (Grilli Caiola, 1980). *Gunnera* est le seul genre d'angiosperme s'associant avec des cyanobactéries qui sont internalisées dans des cellules hôtes au niveau d'une glande située à l'émergence de chaque pétiole. Ces glandes sont présentes dès le début de la croissance de la plante et ne nécessitent pas la présence des bactéries pour leur développement (Bergman *et al.*, 1992) (Fig. 3f ,g). Ces 3 exemples soulignent qu'il existe des plantes capables de mettre en place des structures spécialisées qui sont présentes au tout début du développement et qui ont pour fonction d'accueillir les cyanobactéries (Kneip *et al.*, 2007). En outre, il existe des exemples de cyanobactéries capables de vivre de manière durable dans le cytoplasme de la cellule hôte végétal comme c'est le cas de la cyanobactérie fixatrice d'azote *Atelocyanobacterium thalassa* (anciennement UCYN-A). Cette bactérie s'associe avec une pico-algue appartenant aux *Prymnesiophyte* et entretient un mode de vie symbiotique intracellulaire obligatoire avec son hôte mettant en lumière un fort degré d'adaptation (Fig. 4a,b). Il a été montré que ce mode de vie d'endosymbiotique obligatoire était dû au petit génome de la cyanobactérie résultant de la perte de nombreux gènes liés au métabolisme et à la photosynthèse (Thompson *et al.*, 2012; Farnelid *et al.*, 2016). Cependant, chez les actinomycètes du genre *Frankia* (Fig. 4c) qui s'associent avec des plantes dites « actinorhiziennes » ainsi que chez les *rhizobiums* qui s'associent avec les légumineuses, l'interaction avec la plante hôte mène à la création d'un organe symbiotique spécifique que

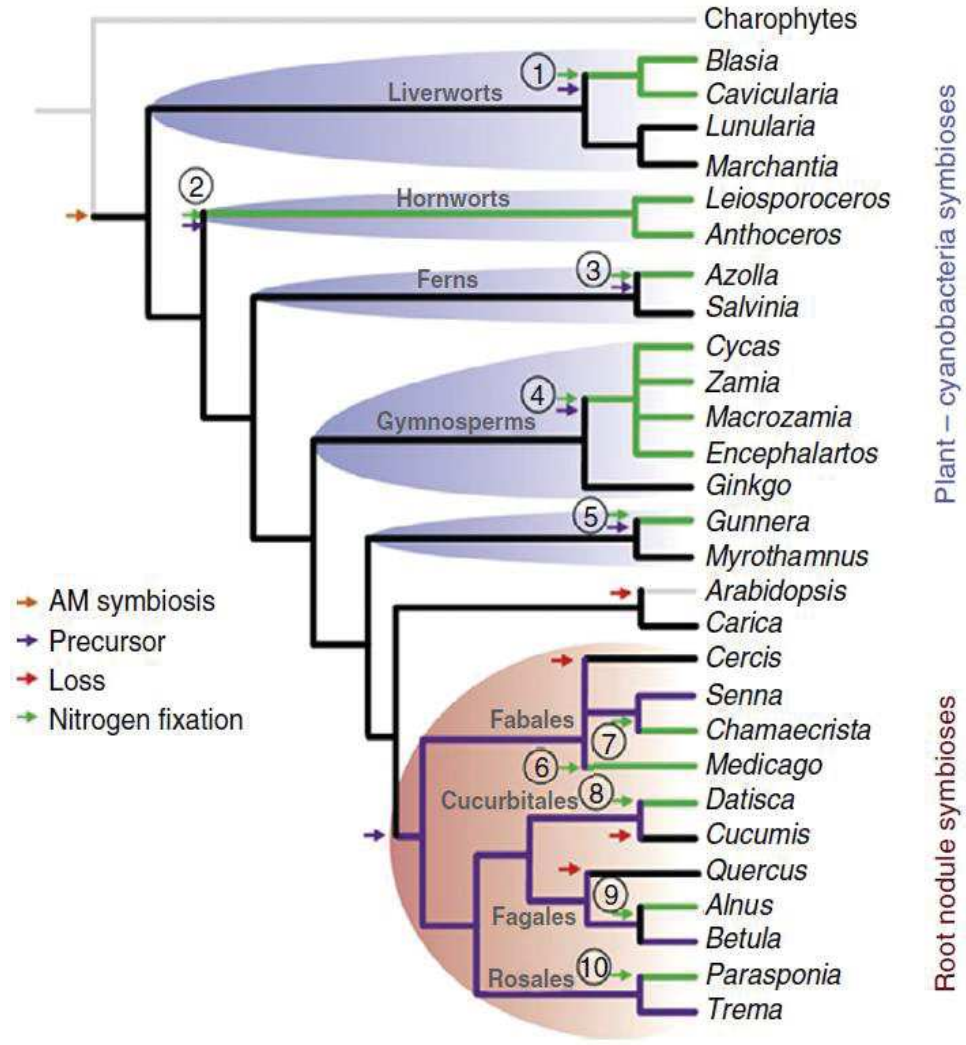


Figure 5 : Evolution de la symbiose fixatrice d'azote chez les plantes terrestres. Les origines proposées de la symbiose fixatrice sont indiquées par des flèches vertes. Les branches violettes représentent les clades prédisposés, les branches noires montrent des espèces capables de former l'association endomycorhizienne (AM) et les branches grises montrent les clades ayant perdu cette association. Les flèches rouges montrent la perte de la prédisposition vers la nodulation ou la symbiose endomycorhizienne. (D'après Delaux et al, 2015)

l'on nomme « nodule » (Fig. 4). Bien que la fonction de fixation d'azote du nodule soit partagée entre ces deux types de symbioses, il y a de grandes différences dans leur développement et leur morphologie finale. Les nodules actinorhiziens semblent correspondre à des racines latérales modifiées avec une vascularisation centrale, de l'endoderme et du cortex hébergeant les Frankia en périphérie. Ces nodules sont composés de plusieurs lobes dont le primordium est initié dans le péricycle (comme la racine latérale), donnant un aspect coralloïde. Chez certaines plantes actinorhiziennes des racines nodulaires prolongent les lobes nodulaires, mais à l'inverse des nodules, elles ne sont pas infectées par la bactérie. Par exemple, dans le genre *Myrica* on retrouve des racines nodulaires dans le prolongement des nodules qui ne sont pas impliquées dans la fixation d'azote alors que le genre *Alnus* ne possède pas ces racines nodulaires (Duhoux *et al.*, 1996).

A l'inverse de ce qui est retrouvé chez les plantes actinorhiziennes, les nodules de légumineuses possèdent tous une vascularisation périphérique et un tissu infecté central correspondant à la zone de fixation de l'azote (Fig. 4). Il existe plusieurs types de nodules chez les légumineuses qui seront décrits dans les parties suivantes. Bien que les nodules actinorhiziens et les nodules de légumineuses se développent tout deux sur la racine, les différences d'organogénèse et de structure suggèrent que leur phénotype dépend de l'hôte.

1.d - Le succès des légumineuses grâce à la nodulation fixatrice d'azote

Toutes les espèces de plantes connues capables de développer des nodules racinaires appartiennent au grand clade des Fabidées, ces plantes sont réparties au sein de 4 ordres qui sont les Fabales, Fagales, Cucurbitales et Rosales. Ensembles, ils forment le clade des NFN (Nitrogen-Fixing Nodulation) (Fig. 5). Parmi les 28 genres appartenant au clade des NFN, seulement 10 d'entre eux, sont capable de réaliser la symbiose fixatrice d'azote et cela ne concerne pas toutes les espèces de ces genres. La capacité à former des nodules fixateurs avec des *Frankia* a été décrite chez 8 lignées appartenant aux Fagales, Cucurbitales et Rosales. Ces plantes actinorhiziennes comme les aulnes (*Alnus*), les myrtes (*Myrica*) ou encore les Céanothes (*Ceanothus*) représentent entre 200 et 400 espèces (Pawlowski and Demchenko, 2012). En revanche, la capacité à développer des nodules avec des rhizobiums est retrouvée dans deux autres lignées de plantes : les légumineuses (*Leguminosae*, Fabales) et le genre *Parasponia* (*Canabaceae*, Rosales, ne contient que 5 espèces capables de

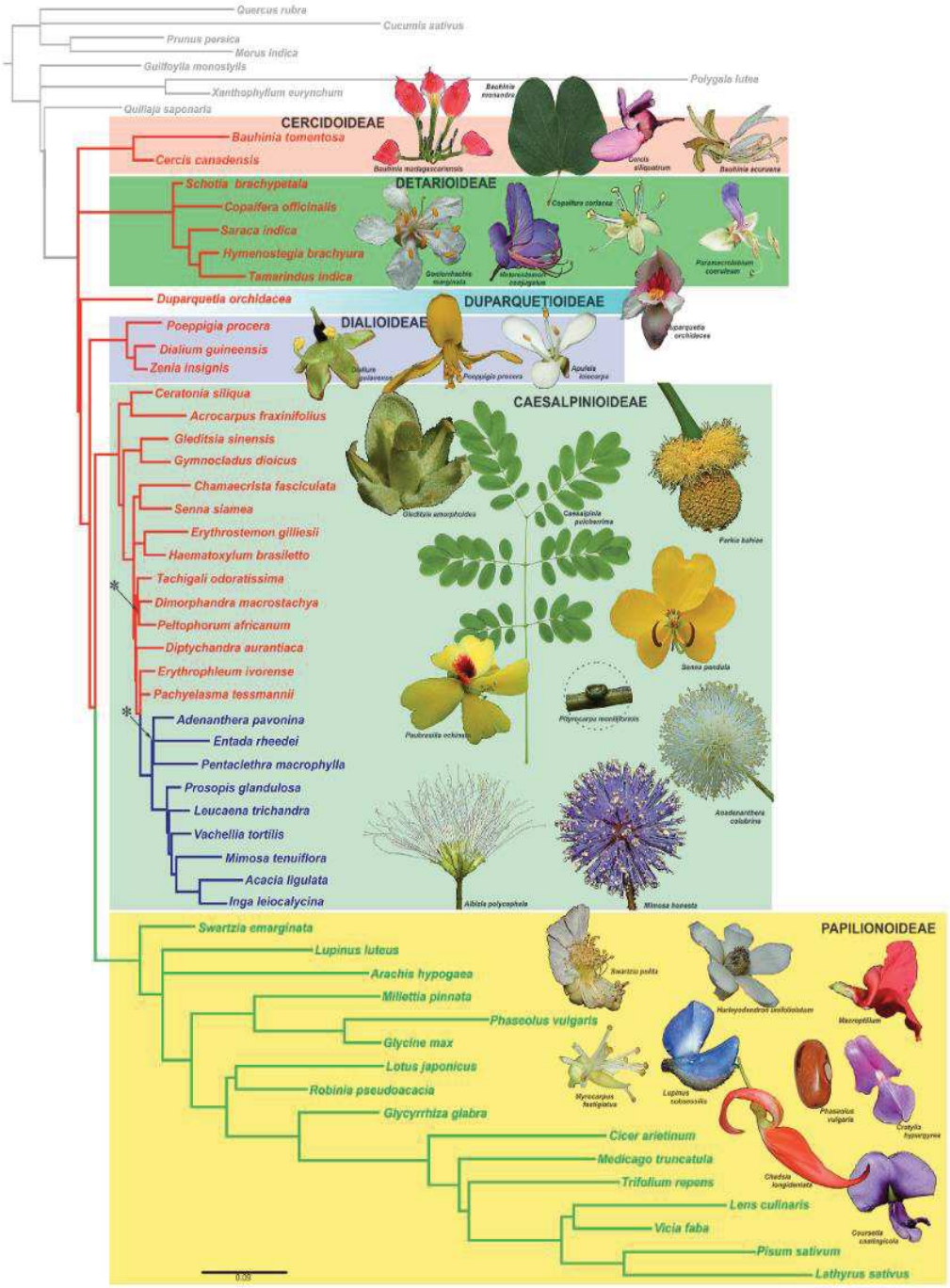


Figure 6 : Phylogénie des Leguminosae basée sur la séquence protéique de 81 gènes plastidiques et obtenue par inférence Bayésienne. En rose la sous-famille des Cercioideae, en vert les Detarioideae, en bleu les Duparquetioideae, en violet les Dialioideae, en vert clair les Caesalpinioideae et en jaune les Papilionoideae. (D'après Azani et al., 2017)

noduler). Parmi les plantes nodulantes, les *Leguminosae* comptent le plus grand nombre de représentants car la majorité des 20.000 espèces sont capables de réaliser la symbiose fixatrice d'azote. Une nouvelle classification des légumineuses qui s'appuie sur des analyses phylogénétiques moléculaires distingue 6 sous-familles qui sont les *Caesalpinioideae* (incluant maintenant les Mimosacées), *Cercidoidea*e, *Detarioideae*, *Dialioideae*, *Duparquetioideae* et *Faboideae* (Fig. 6). La nodulation est proéminente chez les *Faboideae* qui représentent la majorité des espèces de légumineuses (14 000 des 20 000 espèces) (Azani *et al.*, 2017). La radiation évolutive de ce clade, il y a environ 58 millions d'années suit l'extinction massive Crétacé-Paléogène au cours de laquelle environ 30% des plantes terrestres ont disparu, libérant ainsi de nouvelles niches écologiques ce qui a conduit à une grande diversité écologique et morphologique dans **c**ette sous-famille des *Faboideae* (Sprent, 2007; Wilf and Johnson, 2004). Elle est en effet elle-même divisée en quatre principaux clades qui sont les Genistoïdes, les Dalbergioïdes, les Phaseloïdes et les Galegoïdes.

L'homme a su mettre à profit la capacité de nodulation des légumineuses pour répondre à différents besoins alimentaires, et depuis peu de temps, aussi écologiques. La forte teneur en protéine dans les feuilles (utilisées comme fourrage) et dans les grains (nourriture humaine) est corrélée à la capacité fixatrice d'azote dans ce groupe de plantes. Les 4 clades de légumineuses contiennent de nombreuses espèces d'intérêt agronomique comme par exemple chez les Génistoïdes, plusieurs espèces de lupins (*Lupinus albus* et *Lupinus angustifolius*) servant à l'alimentation humaine ou animale et représentant une production mondiale de 1,2 million de tonnes en 2018. Chez les Dalbergioïdes, le représentant le plus important au niveau agronomique est l'arachide (*Arachis hypogeae*) avec laquelle on peut se nourrir directement ou produire de l'huile, et qui a atteint une production mondiale en 2018 de 46 millions de tonnes. Chez les Phaseloïdes regroupant de nombreuses espèces d'intérêt agronomique, le soja (*Glycine max*) est la légumineuse la plus cultivée avec 349 millions de tonnes en 2018. A titre de comparaison la production de la céréale la plus cultivée, le blé, représentait en 734 millions de tonnes en 2018, ce qui placent les légumineuses au deuxième rang de la production végétale mondiale (FAOSTAT, 2018). Enfin, chez les Galegoïdes, on retrouve la lentille et le pois utilisé en alimentation humaine ou encore la luzerne (*Medicago sativa*) qui est une des plantes fourragères les plus utilisée pour

A.



B.

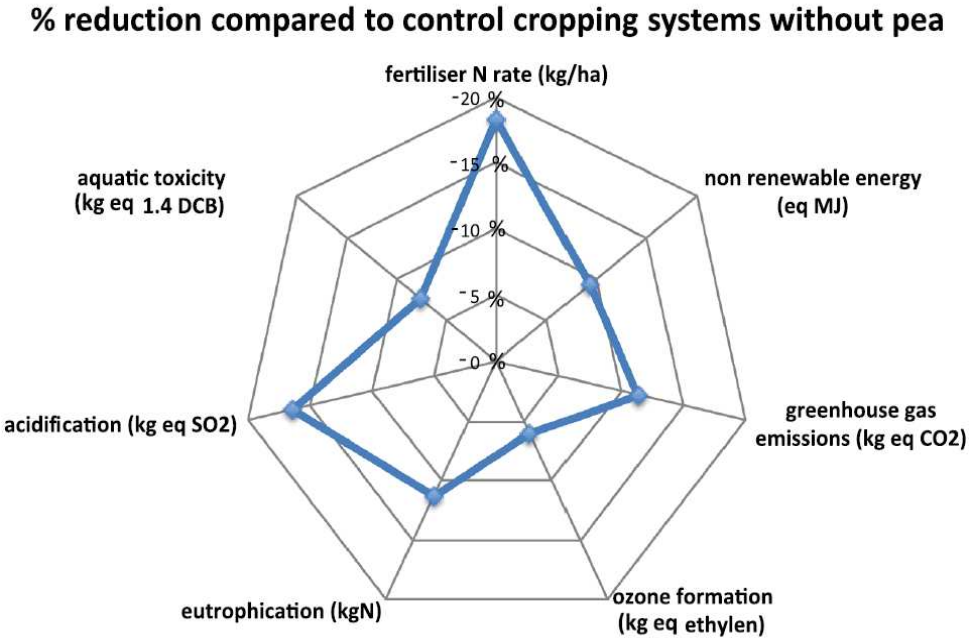


Figure 7 : Les légumineuses en rotation de culture. A) Le pois, *Pisum sativum,* avec ses nodules (gauche), en champ (centre) et les grains (droite). B) Réduction (en %) de l'impact environnemental de champs incluant la culture du pois en comparaison à des champs sans pois. Les diminutions les plus drastiques concernent l'utilisation de fertilisant azoté et l'acidification du sol (D'après Voisin et al.,2014).

l'alimentation du bétail grâce à sa richesse en protéines. D'autre part, les légumineuses sont considérées comme des espèces pionnières car elles peuvent coloniser précocement des terres nues qu'elles vont enrichir en nutriment au cours de plusieurs cycles de vies. Récemment, des projets de végétalisation de sols dénudés ou de re-végétalisation de sols pollués par les activités humaines ont eu lieu et inclus l'utilisation d'espèces locales de légumineuses, par exemple en Nouvelle-Calédonie où *Acacia spirorbis* est utilisé dans cette optique sur des sites miniers (Jaffré *et al.*, 1994). De plus, de nombreuses espèces sont utilisées en rotation de culture de façon empirique depuis des millénaires car cela permet d'enrichir le sol en azote pour la saison suivante. La culture de ces légumineuses rend possible la réduction drastique de l'utilisation de fertilisants azotés sans perte de rendement (Fig. 7). Les légumineuses sont cultivées pour l'alimentation humaine, leur forte teneur en protéine en fait des substituts possibles à la viande qui a un coût économique et écologique élevé. Elles peuvent aussi servir pour l'alimentation animale et participe ainsi à la production globale de nourriture d'origine animale. C'est pourquoi mieux comprendre les mécanismes régissant la symbiose chez les légumineuses est un enjeu important pour inventer l'agriculture de demain. Tous ces éléments convergent pour dire que les légumineuses sont des plantes d'intérêt majeur dû à leur capacité d'interagir avec des bactéries fixatrices d'azote et c'est pourquoi la majorité des recherches scientifiques ont été conduites sur la famille des *Faboideae*. Cependant, les légumineuses ont aussi des aspects plus négatifs comme par exemple le fait qu'elles ont généralement de faibles rendements, que certaines espèces ne fixent pas très efficacement l'azote et que leur culture intensive appauvri les sols.

2. Mécanismes de la symbiose rhizobienne chez les légumineuses modèles

La plupart des légumineuses partagent un processus d'infection et d'organogénèse du nodule similaire qui a été étudié en détail chez plusieurs espèces. Les légumineuses sécrètent un cocktail de molécules dans la rhizosphère, dont des acides aminés qui ont un effet chimiotactique sur certaines bactéries du sol comme les rhizobiums, les attirant jusqu'à la surface de la racine, notamment dans la zone où les poils absorbants sont en croissance et qui sont les seuls infectibles (Compton et al., 2020 DOI : 10.3389/fmicb.2020.581482). Là, les

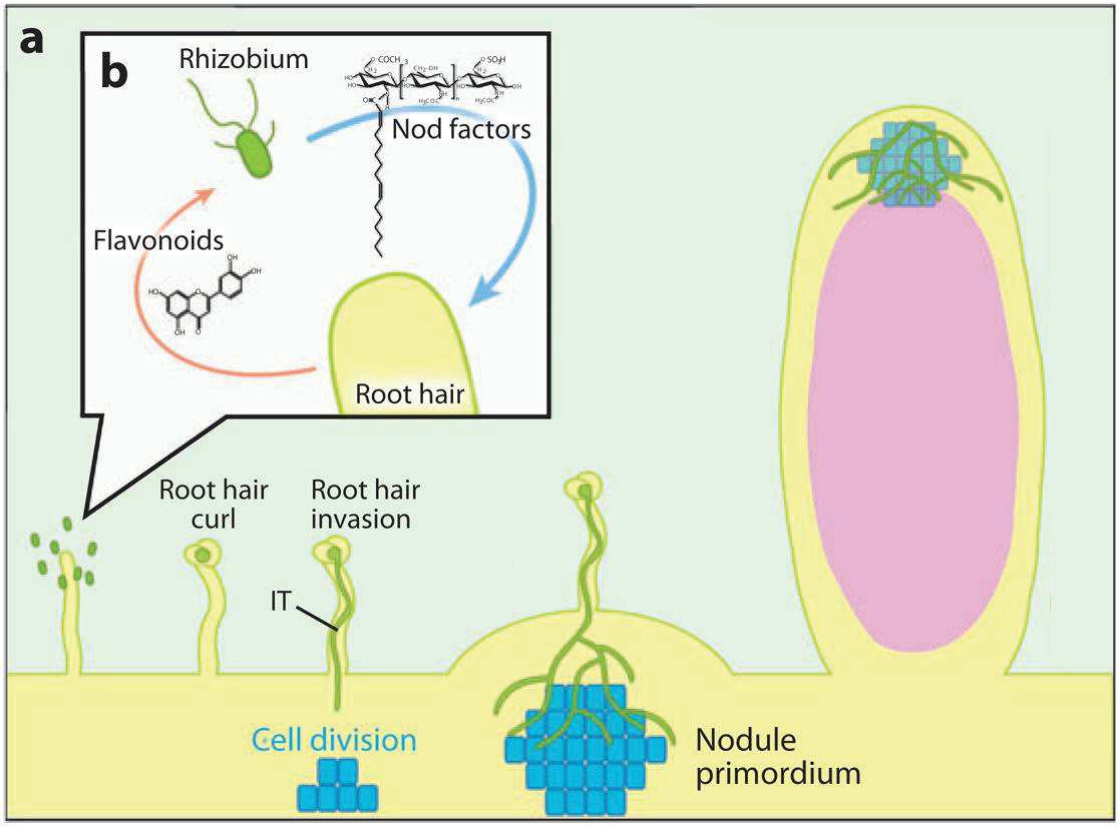


Figure 8 : Modèle schématique du développement d'un nodule chez *Medicago truncatula*. (a-b) Les flavonoïdes exsudés par les racines de la plante hôte dans le sol induisent la production de facteurs Nod chez la bactérie. La perception de ces facteurs Nod par des récepteurs de la plante-hôte induit une déformation du poil absorbant en crosse de berger et des divisions cellulaires dans le cortex sous-jacent au site d'infection. Les cordons d'infection conduisent les bactéries jusqu'aux cellules du primordium nodulaire où elles sont libérées. Ce primordium est à l'origine d'un nodule dont la croissance est assurée par un méristème situé en position apicale et qui présente une zone centrale contenant les bactéries fixatrices d'azote (Gibson *et al.*, 2009)

rhizobiums s'accumulent au niveau de l'apex d'un poil absorbant et induisent une déformation du poil en forme de « crosse-de-berger » (Fig. 8). Cette structure piège les bactéries et crée une chambre d'infection dans laquelle elles se divisent, formant le foyer d'infection. A partir de ce foyer, se développe une structure tubulaire, le cordon d'infection qui a pour vocation de conduire de façon contrôlée par la plante les bactéries dans les assises internes de la racine (Gibson *et al.*, 2008 ; Fournier et al., 2015 ; Turgeon & Bauer, 1985). Cette progression est guidée par le noyau de la cellule qui migre en position apicale du cordon d'infection et par une polarisation du cytosquelette des cellules traversées, formant une structure baptisée « cordon de pré-infection ». Il a été suggéré que l'initiation et le passage du cordon d'infection d'une cellule à l'autre est permise par une dégradation localisée des parois végétales. Le développement des cordons d'infections continue au travers de plusieurs assises corticales jusqu'à atteindre les cellules du primordium nodulaire qui a été induit à distance et en dessous de la cellule épidermique initialement infectée (Timmers *et al.*, 1999). Les bactéries sont alors libérées et endocytées dans certaines cellules du primordium nodulaire, aboutissant au développement d'un nodule dans lequel les cellules du tissu central sont infectées de manière hétérogène (Fig. 8). Les rhizobiums qui se retrouvent au sein de symbiosomes subissent une différenciation en bactéroïde, qui est la forme active pour la fixation de l'azote. Chez certaines légumineuses comme celles du clade des IRLCs (Inverted-Repeat Lacking Clade) cette différenciation est très poussée et elle est irréversible car la bactérie devient incapable de se diviser, elle est donc qualifiée de « terminale » (Oono *et al.*, 2010).

Ce processus symbiotique apparait comme étant sophistiqué car il fait intervenir un programme d'infection contrôlé par la plante-hôte d'un côté et de l'autre un programme d'organogenèse nodulaire pour aboutir à la formation d'un nodule fixateur d'azote. Pour mettre à jour les mécanismes moléculaires régissant cette interaction symbiotique, plusieurs rhizobiums ont été caractérisés au niveau génétique et deux légumineuses ont été choisies comme modèle d'étude, *Medicago truncatula* et *Lotus japonicus*. En raison de leur maniabilité génétique et de leur proximité phylogénétique avec de nombreuses espèces tempérées cultivées, facilitant ainsi le transfert de connaissances (Cook, 1999; Handberg and Stougaard, 1992). Au cours des 20 dernières années, les travaux de recherche menés par la communauté scientifique au niveau international ont permis de mettre à jour de nombreux gènes

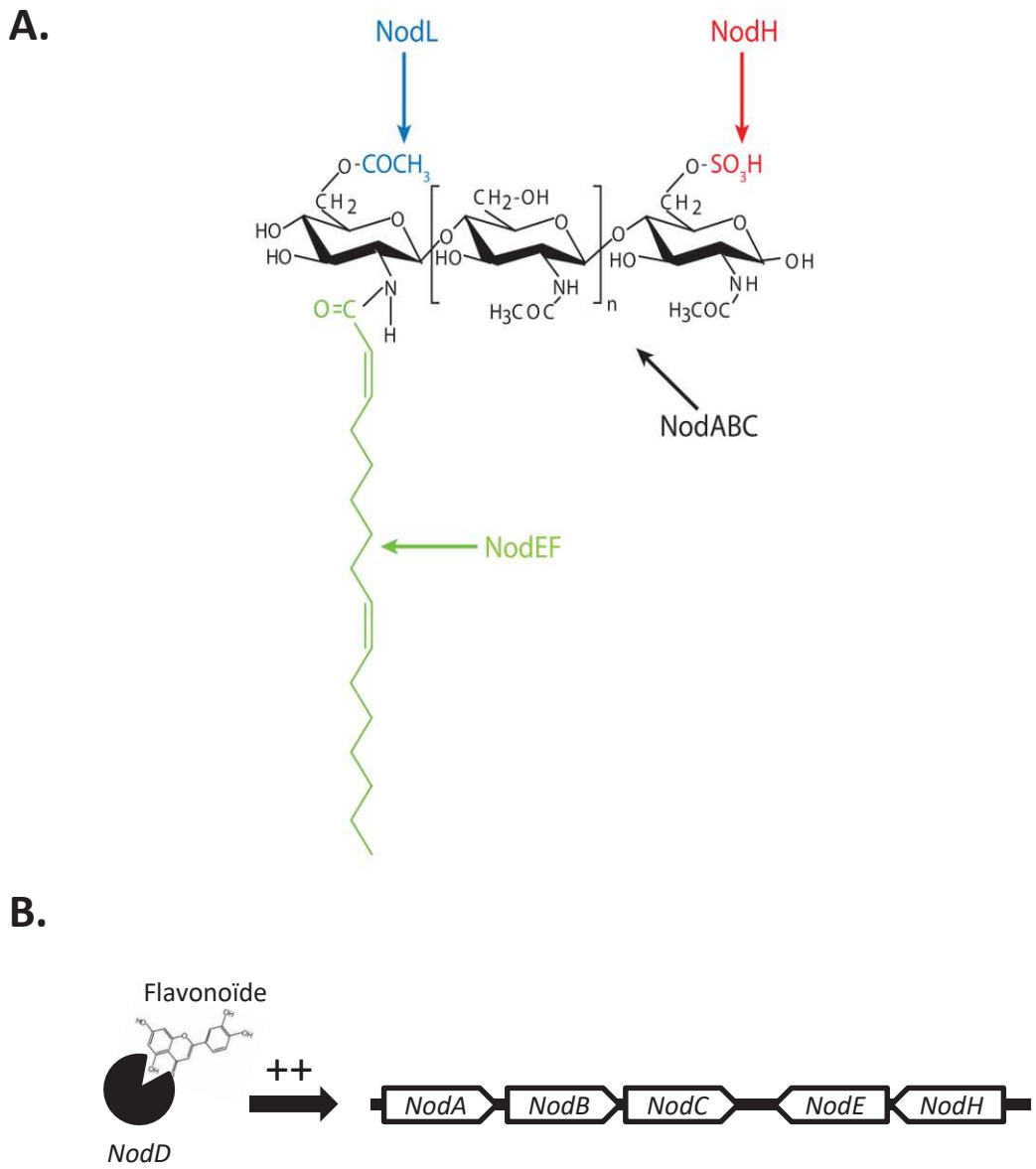


Figure 9 : Structure des facteurs Nod et régulation de leur synthèse. A. Les facteurs Nod produits par la souche *Ensifer meliloti* et les enzymes biosyntéthiques (protéines nod) impliquées dans la voie de biosynthèse. **B.** Mécanisme d'action de la protéine régulatrice NodD sur l'opéron nod.

symbiotiques, apportant un éclairage important sur les mécanismes de la symbiose rhizobium-légumineuse.

2.a - Les facteurs Nod : molécules signal clés de la symbiose rhizobium-Légumineuse

Les légumineuses et plus particulièrement celles qui vivent en milieu tempéré, sont nodulées par un nombre restreint de rhizobiums, cette interaction très spécifique suggère qu'il existe un mécanisme de reconnaissance entre les 2 partenaires. Par exemple malgré leur proximité phylogénétique et le fait qu'elles soient toutes deux isolées de nodules de *P. vulgaris*, la souche *R. etli* possède un spectre d'hôte restreint à une espèce (*P. vulgaris*) alors que *R. tropici* possède un large spectre d'hôte étant capable de noduler plus de 8 genres de légumineuses. La reconnaissance mutuelle est le fruit d'un dialogue moléculaire entre le symbionte et l'hôte qui débute par la production d'un cocktail de flavonoïdes par la plante dans la rhizosphère (Del Pero Martínez and Martínez, 1993). Dans le cocktail de flavonoïdes produits, certains induisent l'expression de gènes *nod* chez la bactérie, qui sont nécessaires à la synthèse et à la modification d'une molécule signal clé de nature Lipo-Chito-Oligosaccharidique (LCO) nommé « Facteur Nod » (Lerouge *et al.*, 1990). Sa structure chimique correspond à une chaîne de plusieurs oligomères de chitine (résidus N-Acetyl Glucosamine ou GlcNAc) sur laquelle se greffent plusieurs décorations moléculaires (Fig. 9A). Ce squelette d'oligomères possède un unique groupement d'acide gras insaturé en partie terminal, il peut être sulfaté à l'autre extrémité et les autres résidus sont N-acétylés. Les différents types de groupements, ajoutés sur le squelette du Chito-Oligosaccharide, jouent un rôle sur la spécificité d'hôte (Dénarié *et al.*, 1996).

La protéine nodD est à la fois senseur de signaux environnementaux et régulateur transcriptionnel (Fig. 9B). Cette protéine est activée après l'interaction avec le flavonoïde compatible et cible des séquences spécifique d'ADN (*nod* box) dans les régions promotrices des gènes de nodulation (gènes *nod*) et active leurs transcription (Fisher and Long, 1993; Peck *et al.*, 2006). Même si nodD est ubiquitaire chez les rhizobiums, ses caractéristiques varient d'une espèce à l'autre et les types de flavonoïdes qu'il reconnaît sont différents. Chez certaines souches comme *Rhizobium leguminosarum bv. trifolii* qui possède un spectre d'hôte restreint, on ne trouve qu'une seule version du gène *nodD* et sa mutation rend la

souche incapable de noduler. A l'inverse chez *Ensifer meliloti* qui possède un plus large spectre d'hôte, il y a plusieurs copies du gène *nodD* dont les mutations de toutes les versions sont requises pour abolir la nodulation (Honma *et al.*, 1990). De nombreux autres gènes sont induits au contact des flavonoïdes dépendamment de *nodD* et interviennent dans la synthèse d'hopanoïdes (Perzl *et al.*, 1997), le catabolisme des rhizopines (Jiang *et al.*, 2001), la fixation d'azote (Dombrecht *et al.*, 2002), la synthèse et/ou modification des polysaccharides de surfaces (Mimmack *et al.*, 1994) et également dans la modulation de l'expression de régulateurs transcriptionnels (Perret *et al.*, 1999). Concernant les gènes *nod* chez *E. meliloti*, ils ont été classés en trois groupes : 1) les gènes canoniques responsables de la synthèse du squelette moléculaire des facteurs Nod (*nodA*, *nodB* et *nodC*) qui sont structurellement et fonctionnellement conservés chez les différentes espèces de *Rhizobium* et *Bradyrhizobium* (Marvel *et al.*, 1987) ; 2) les gènes de nodulation spécifiques à l'hôte (notamment *nodE* et *nodF*) qui jouent un rôle dans la spécificité de l'interaction en branchant des décorations moléculaires sur le squelette du facteur Nod (Demont *et al.*, 1993) ; 3) le gène *nodD* et ses homologues activant la transcription des gènes canoniques et des gènes spécifiques à l'hôte (Mulligan and Long, 1985). Les mutants bactériens des gènes canoniques *nodABC* sont incapables d'établir une interaction symbiotique avec leur hôte suggérant que ces gènes jouent un rôle clé. En revanche les mutants sur les gènes de nodulation spécifiques à l'hôte, montrent une capacité réduite à infecter la totalité de leur partenaires habituels (Debellé *et al.*, 1986). De plus les doubles mutants *nodF/nodL* chez *R. meliloti*, influençant l'état d'acétylation et d'acylation du facteur Nod, ce qui engendrent des problèmes dans le processus d'infection sans altérer l'induction à distance du primordium nodulaire, montrant que les modifications apportées aux facteurs Nod sont essentielles à l'étape d'infection mais ne joue pas de rôle dans l'organogénèse (Ardourel *et al.*, 1994). Pour finir, les facteurs Nod purifiés sont suffisants pour induire des réponses très rapides comme des flux ioniques au niveau de la membrane plasmique et des oscillations calciques au niveau de l'enveloppe nucléaire, ainsi qu'un branchement des poils absorbants et l'induction de divisions de cellules corticales voire l'organogenèse du nodule chez certaines légumineuses. La plupart de ces réponses sont induites à de très faibles concentrations en facteurs Nod suggérant l'existence de récepteurs à haute affinité chez l'hôte (Schultze and Kondorosi, 1998).

De façon surprenante les bactéries du sol capables d'interagir avec les légumineuses que l'on nomme communément *rhizobiums*, appartiennent aux α- et β-protéobactéries (clade paraphylétique). L'arbre phylogénétique généré avec les séquences 16S de plusieurs espèces de rhizobiums possédant les gènes de nodulation présente une topologie différente de l'arbre généré avec la séquence des gènes *nod*. En effet, la similarité des gènes *nod* et des « ilôts symbiotiques » chez les différentes espèces de rhizobiums, suggère que les gènes nécessaires à la nodulation et la fixation d'azote (*nod*, *nif* et *fix*) ont été acquis par transfert horizontal de manière répétée et indépendante. Cela montre qu'en dépit de leur phylogénie les rhizobiums ont convergé vers le même mode de vie symbiotique avec les légumineuses et ce, grâce à l'alternance du mode de vie libre qui permet un échange de matériel génétique avec d'autres bactéries du sol (Masson-Boivin *et al.*, 2009).

Il existe cependant d'autres facteurs qui influencent le succès de l'interaction symbiotique comme la modulation de l'immunité de l'hôte qui est normalement une barrière à la mise en place de la symbiose, car elle perçoit les motifs moléculaires associés aux microbes (MAMP). Etablir une interaction stable nécessite donc une communication tout au long de la vie des partenaires avec une réduction de l'immunité pour ne pas éliminer les symbiontes. Les facteurs Nod participent au processus de réduction des mécanismes de défenses et sont produits tout au long de l'interaction. En effet, il a été montré que la perception des facteurs Nod supprimaient partiellement la réponse MTI (MAMP-Triggered Immunity) chez des légumineuses (*G. max*) mais également chez des non-légumineuses (*A. thaliana*, *S. lycopersicum* et *Z. mays*) (Liang *et al.*, 2013). Même si les facteurs Nod sont les composants bactérien majeur qui permettent la compatibilité de l'interaction et la modulation de l'immunité, d'autres éléments entrent en jeu dans la réduction des réactions de défenses de l'hôte. En effet chez *L. japonicus*, les EPS (exopolysaccharides) de surfaces provenant des *rhizobiums* ont un rôle crucial dans le processus d'accommodation des bactéries pendant le processus d'infection via leur perception par le récepteur LjEPR3, il est suggéré que ce LysM-RLK permet de réduire la réaction immunitaire de l'hôte (Kawaharada *et al.*, 2015)**.** La modulation de l'immunité peut également se faire par l'envoi d'effecteurs protéiques dans le cytoplasme de la cellule eucaryote, comme c'est le cas d'*Ensifer fredii* qui supprime les mécanismes de défenses grâce à son T3SS (Type 3 Secretion System) pour noduler efficacement son hôte le soja (Jiménez-Guerrero *et al.*, 2015).

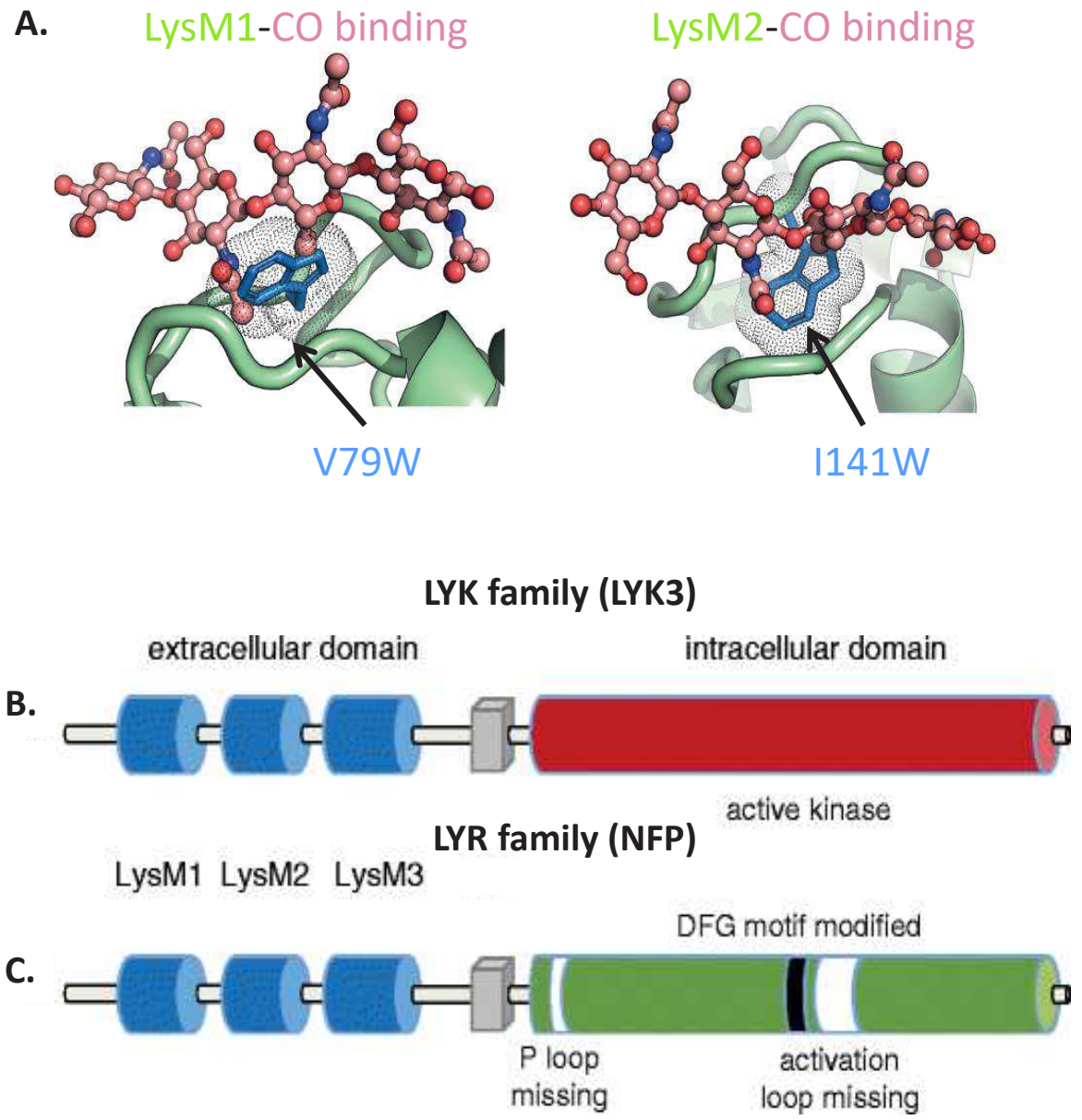


Figure 10 : Représentation schématique de 2 groupes de LysM-RLK auxquels appartiennent NFP et LYK3. A. Analyse des régions LysM1 et LysM2 de NFR1 (récepteur aux facteur Nod) et CERK6 (récepteur à la chitine). Modèle des domaines LysM1 et LysM2 de CERK6 prédit pour lier la chitine. **B.** LysM-RLK avec un domaine extracelluliare contenant 3 domaine LysM, un domaine transmembranaire et un domaine kinase. **C.** un domaine kinase cytoplasmique fonctionnel comme par exemple LYK3 ou (B) un domaine kinase non-fonctionnel comme par exemple NFP. (D'après Fliegman & Bono, 2015 ; Boszoki et al, 2020).

2.b - Reconnaissance des facteurs Nod par la plante-hôte

Afin de comprendre comment la plante hôte perçoit les facteurs Nod et comment ils induisent des réponses symbiotiques conduisant à la formation de nodules fixateurs d'azote, des cribles génétiques ont été réalisés chez les deux légumineuses modèles permettant d'identifier des mutants altérés dans la nodulation. La caractérisation de ceux bloqués dans les étapes les plus précoces de l'interaction symbiotique a montré qu'ils sont altérés dans certaines réponses aux facteurs Nod purifiés, ce qui a conduit à proposer un modèle génétique d'une voie de signalisation Nod activée par les facteurs Nod et conduisant à ces réponses symbiotiques (Catoira *et al.*, 2000). Les gènes situés en haut de cette voie de signalisation et supposés coder les récepteurs putatifs aux facteurs Nod, ont été clonés chez *M. truncatula* (MtNFP/MtLYK3) et *L. japonicus* (LjNFR1/LjNFR5), révélant leur nature de Récepteur Kinase à domaine LysM ou « LysM-RLK » (Arrighi *et al.*, 2006; Limpens, 2003; Madsen *et al.*, 2003; Mulder *et al.*, 2006; Radutoiu *et al.*, 2003). Les mutants de LYK3 chez *M. truncatula* présentent une courbure du poil absorbant mais ils sont incapables d'initier le cordon d'infection et sa croissance alors que les mutants de NFP ne présentent aucune réponse. En revanche, chez L. *japonicus* les mutants de ces 2 gènes ne présentent pas de déformations ni courbures des poils absorbants. A l'époque de leur découverte, des protéines à domaine LysM étaient connues chez les bactéries comme étant capable de lier des molécules contenant des résidus GlcNAc ou N-acetyl Glucosamine (Fig. 10A) comme le peptidoglycane ou la chitine (Buist *et al.*, 2008), renforçant l'idée que MtNFP/MtLYK3 et LjNFR1/LjNFR5 sont les récepteurs aux facteurs Nod. Ce rôle a ensuite été confirmé par des approches de génétique et de biochimie.

Par le fait, ce type d'étude a permis l'identification précise des acides aminés nécessaires pour la liaison au facteur Nod et ils sont situés dans le domaine LysM2 chez MtNFP, LjNFR5 et LjNFR1 (Bensmihen *et al.*, 2011; Broghammer *et al.*, 2012; Sørensen *et al.*, 2014). Lorsque l'on remplace l'ectodomaine complet (ou chacun des domaines LysMs) de MtNFP par celui de son plus proche orthologue chez le pois (PsSYM10), dont les symbiontes ne sont pas compatibles avec *M. truncatula*, la signalisation précoce est restaurée. Cela suggère que NFP n'a pas de rôle dans la spécificité de reconnaissance des décorations du facteur Nod mais plutôt dans la reconnaissance du groupement acyl (Bensmihen *et al.*, 2011).

Étonnamment, le récepteur MtLYK3/LjNFR1 montre une activité d'autophosphorylation alors que le gène MtNFP/LjNFR5 possède un domaine « dead » kinase, c'est-à-dire qu'il est non-fonctionnel pour l'activité de phosphorylation. En conséquence, il a été proposé que la transduction du signal se faisait par des interactions protéine/protéine et plus particulièrement par l'hétérodimérisation de NFP/NFR5 avec LYK3/NFR1 (kinase active) pour former un complexe récepteur (Madsen *et al.*, 2011). Cette interaction a été démontrée par une technique de FRET (Fluorescence Resonance Energy Transfer) montrant que NFP et LYK3 forment un hétérocomplexe à la périphérie des cellules des nodules de *M. truncatula* (Moling *et al.*, 2014).

Les gènes MtLYK3/LjNFR1 et MtNFP/LjNFR5 sont les représentants de deux types de LysM-RLK définies sur la séquence de leur domaine kinase. Le premier type nommé LYK (Fig. 10B), auquel LYK3/NFR1 appartient, possède un domaine kinase RD et montre une activité d'autophosphorylation *in vitro* (Arrighi *et al.*, 2006). Le deuxième type nommé LYR (Fig. 10C), auquel NFP/NFR5 appartient, possède un domaine kinase dépourvu de plusieurs motifs conservés et ne montre aucune activité d'autophosphorylation ou de trans-phosphorylation *in vitro* (Arrighi *et al.*, 2006; Madsen *et al.*, 2011). On peut également trouver des versions tronquées fortement similaires aux LysM-RLK mais qui ne possèdent que la partie extracellulaire contenant les 3 domaines LysM attachée à la membrane par une ancre-GPI, ce sont des LysM-RLPs qui ont été nommés LYMs (Arrighi *et al.*, 2006).

Par ailleurs, dans le génome de *M. truncatula*, les gènes LYK3, LYK4, LYK5, LYK6, LYK7 forment un cluster de gènes de 46Kb et LYK3/LYK4 présentent une forte homologie. Le phénotype du mutant *lyk4* est moins fort que celui de *lyk3* mais il présente tout de même moins de nodules et affecte la morphologie du cordon d'infection (Limpens, 2003). Le gène LYK9, aussi appelé LYS6 chez *Lotus* et CERK1 chez *Arabidopsis* et le riz, est présent chez de nombreuses plantes en dehors du clade des légumineuses. La comparaison des séquences conservées du domaine kinase chez *Arabidopsis* et des analyses phylogénétiques, ont révélées que le gène LYK3 dérive de la duplication ancestrale du gène LYK9/CERK1 et que LYK3 n'est présent que chez les légumineuses (De Mita *et al.*, 2014). Chez *Arabidopsis* CERK1 est impliqué dans l'immunité dirigée contre les champignons pathogènes et chez le riz il a un rôle double entre immunité et symbiose mycorhizienne (Miya *et al.*, 2007; Zhang *et al.*, 2015). Récemment, il a été montré chez *M. truncatula,* que MtLYK9 est impliqué dans

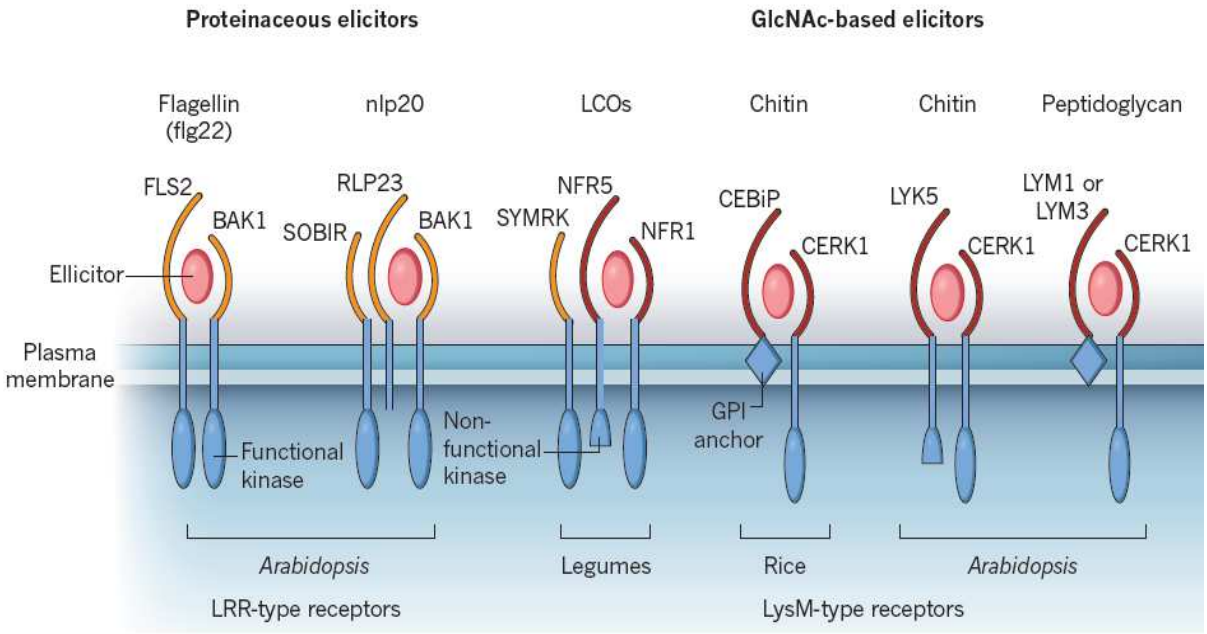


Figure 11 : Les complexes récepteurs LysM-RLKs sont impliqués dans la perception des micro-organismes. Les éliciteurs protéiques sont généralement perçus par des LRR-RLKs (en orange). Les éliciteurs GlcNAc comme les facteurs Nod, la chitine ou le peptidoglycane sont reconnus par des LysM-RLKs motifs LysM (en rouge). Au moins un des récepteurs de ces complexes possèdent un domaine kinase fonctionnel pour la transduction du signal (D'après Zipfel et oldroyd., 2017)

l'immunité car le mutant est plus sensible au pathogène *Aphanomyces euteiches* et qu'il est aussi impliqué dans le processus de mycorhization car le mutant présente une réduction de la colonisation des champignons endomycrohiziens (Gibelin-Viala *et al.*, 2019). De plus, le gène LYK3/NFR1 semble aussi impliqué dans la reconnaissance du partenaire fongique chez les deux légumineuses modèles, car les mutants présentent une légère diminution de la colonisation du champignon, qui est plus marquée chez *L. japonicus* (Zhang *et al.*, 2015). Ces résultats suggèrent que la fonction ancestrale de LYK3 était orientée vers la reconnaissance des LCO-fongiques.

Concernant le second récepteur aux facteurs Nod, le gène NFP est lui aussi positionné dans un cluster de gènes (LYRs) chez *M. truncatula*, soulignant l'existence d'évènements de duplications ancestraux dans cette famille de gènes. En plus d'être impliqué dans la nodulation chez les légumineuses, NFP intervient également dans la résistance aux pathogènes chez *M. truncatula* (Rey *et al.*, 2013). Ce LysM-RLK est présent chez de nombreuses plantes dont les céréales et la tomate, chez cette dernière, la mutation du gène NFP (SlLYK10) entraîne une perturbation de la mycorhization. De manière surprenante, lorsque le gène *SlLYK10* est introduit chez des mutants *nfp* de *M. truncatula*, la nodulation est partiellement restaurée. Le fait que ce gène n'est pas subis de changements de fonction depuis la séparation Rosidae-Asteridae, suggèrent que SlLYK10 fait partie d'un système ancestral de perception des LCO mycorhizien, qui a été directement recruté pendant l'évolution de la symbiose fixatrice d'azote chez les légumineuses (Girardin *et al.*, 2019).

Comme évoqué précédemment et au même titre que les *rhizobiums*, les champignons mycorhiziens produisent des LCOs et des COs à chaîne courte pour établir une endosymbiose racinaire avec leur hôte (Maillet *et al.*, 2011, 2011). Les récepteurs de la plante capables de percevoir ces signaux fongiques appartiennent eux aussi à la famille des LysM-RLK et plus précisément au type LYRIA (SlLYK10) chez les *Solanaceae* (Girardin *et al.*, 2019). Chez le riz, le complexe récepteur aux facteurs Myc a récemment été découvert, il s'agit ici aussi de deux LysM-RLK : OsMYR1 (MYc factor Receptor 1) et OsCERK1. En remplaçant l'ectodomaine de OsMYR1 et de OsCERK1 avec ceux de MtNFP et MtLYK3 respectivement, il est possible d'observer une augmentation des oscillations calciques en réponse à l'application de facteurs Nod chez la lignée de riz transgénique (He *et al.*, 2019).

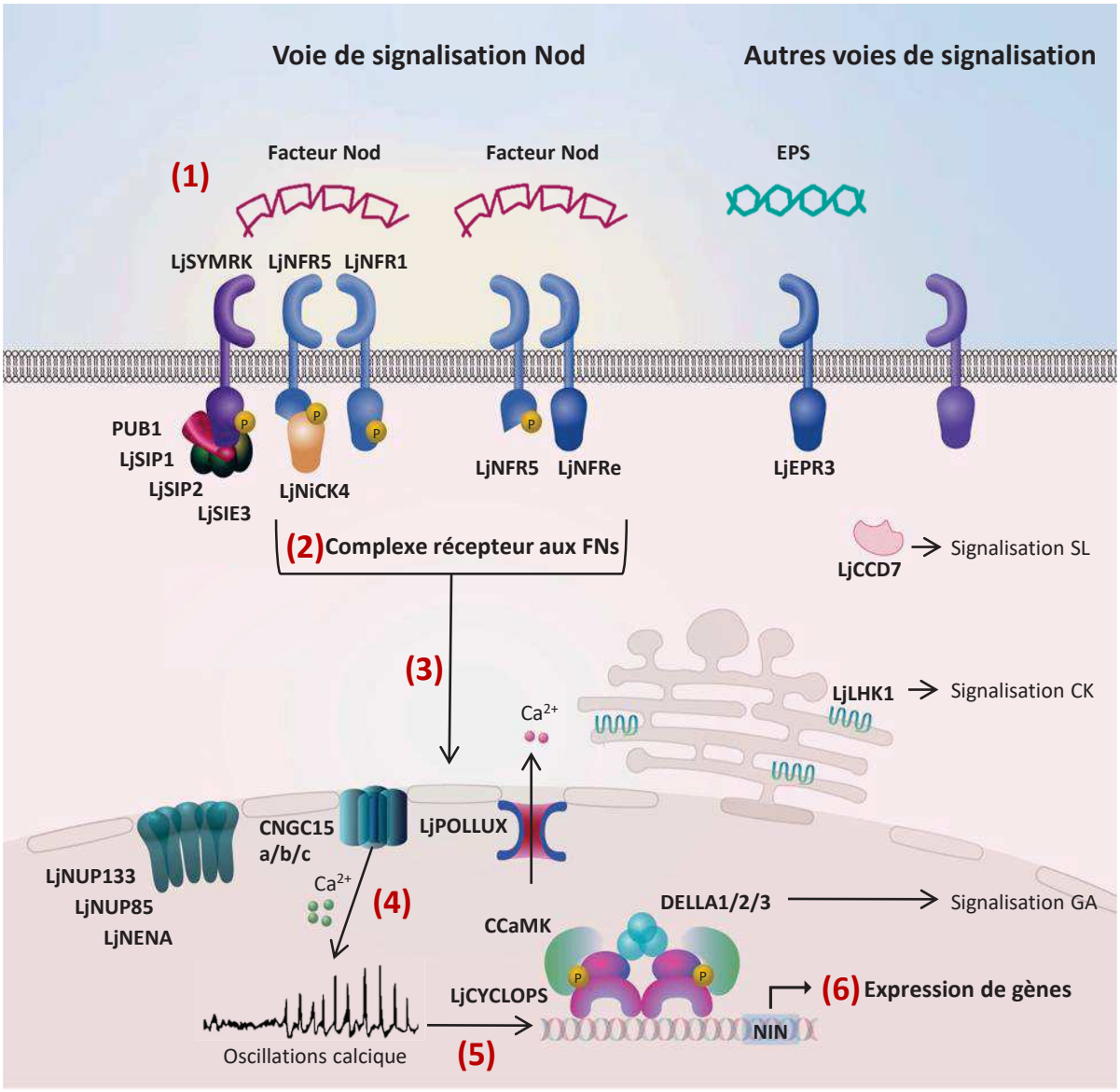


Figure 12 : Gènes et processus impliqués dans les étapes précoces de la nodulation. (1) Production de facteurs Nod (FN) par les bactéries. (2) Perception des FNs par le complexe récepteur. (3) Induction dans le cytosol de réponses biochimiques et physiologiques conduisant à (4) un influx de calcium et des oscillations calciques au niveau de l'enveloppe nucléaire. (5) Décodage dans le noyau des oscillations aboutissant à (6) l'induction de l'expression de nombreux gènes symbiotiques via un réseau de facteurs de transcription. SL= Strigolactones, CK= Cytokinines et GA= Giberellines. (Adapté de Roy et al, 2020)

La famille des LysM-RLK regroupe donc des gènes ayant un rôle dans l'immunité, la nodulation et la mycorhization, on peut retrouver plusieurs gènes impliqués dans la même fonction biologique mais aussi de mêmes gènes fonctionnant dans plusieurs types d'interactions (Fig. 11). La formation d'hétéro-complexes récepteurs est une des clés du fonctionnement des LysM-RLK, qui permet de percevoir différent ligands et d'induire différentes réponses au niveau cellulaire. Chez les légumineuses, il est désormais admis que les récepteurs aux Facteurs Nod dérivent de récepteurs ancestraux orientés vers la perception de LCO/CO fongiques et fonctionnant pendant le processus de mycorhization apparu il y a environ 450 MA (Markmann and Parniske, 2009).

2.c - La voie de signalisation Nod

Des gènes intervenant, en aval de ceux codant les récepteurs des FN, dans « la voie de signalisation Nod » ont été en grande partie clonés à partir des mutants Nod^-. D'autres acteurs ont aussi été identifiés par des approches de génétique inverse et de biochimie, permettant de mieux comprendre les évènements suivant la perception du signal symbiotique au niveau de la membrane comme la transduction du signal dans la cellule et l'activation des programmes génétiques de la symbiose rhizobienne.

Transduction du signal au niveau de la membrane plasmique

Après perception des facteurs Nod, de nombreux évènements moléculaires vont avoir lieu pour transduire le signal jusqu'au noyau, où l'activation de programmes génétiques spécifiques va mener à l'infection et l'organogénèse du nodule (Fig. 12). La protéine NiCK4 (NFR5-interacting Cytoplasmic Kinase 4) a été identifiée lors d'une étude récente de protéomique chez *L. japonicus*, révélant qu'elle était capable d'interagir avec LjNFR5, de phosphoryler LjNFR5 et LjNFR1 *in vitro* et était co-exprimée avec LjNFR5 dans les poils absorbants ainsi que dans les cellules du nodule. L'analyse phénotypique des mutants a montré que NiCK4 participait à l'organogénèse nodulaire et qu'il faisait partie intégrante de la voie de signalisation Nod (Wong *et al.*, 2019).

Au niveau de la membrane plasmique, un récepteur symbiotique nommé SYMRK (SYMbiosis Receptor-like Kinase) joue un rôle majeur dans la transduction du signal lors de l'interaction rhizobienne et endomycorhizienne (Stracke *et al.*, 2002). Ce gène fait partit de la grande famille des LRR-RLK (Leucine Rich-Repeat RLK) qui sont généralement connus pour jouer un

rôle dans le développement ou la réponse aux stress. Une étude chez *L. japonicus* a permis de mieux comprendre la structure et le fonctionnement de ce récepteur. L'ectodomaine de SYMRK est constitué de 3 domaines LRR et d'un domaine MLD (Malectin-Like Domain), ce dernier est clivé en l'absence de stimuli symbiotique au niveau d'un motif « GDPC ». L'analyse des mutants sur le motif GDPC montre que ce motif est nécessaire pour le clivage du MLD et nécessaire pour la fonction symbiotique de SYMRK. Le clivage du MLD induit une dégradation rapide de SYMRK suggérant qu'il s'agit d'un mécanisme de régulation. De plus, il est montré que NFR5 interagis préférentiellement avec la version de SYMRK clivée et que la présence du domaine MLD interfère avec l'interaction NFR5-SYMRK. Il est proposé que le MLD protège SYMRK de la dégradation mais que le clivage expose le domaine LRR, ce qui promeut le renouvellement des protéines membranaires (NFR5). La libération du MLD de SYMRK assure la répartition spatio-temporelle appropriée pour une capacité symbiotique optimale. Les résultats suggèrent que SYMRK agit comme un co-récepteur et initie la signalisation symbiotique en concert avec NFR5 (Antolín-Llovera *et al.*, 2014). Il a également été montré avec une technique de recherche des substrats des protéines kinases (MAKS) chez *M. truncatula* que SYMRK était capable de phosphoryler directement LYK3 *in vivo* et *in vitro* (Jayaraman *et al.*, 2017). A l'heure actuelle, ce que reconnait l'ectodomaine de SYMRK n'est pas connu, en revanche plusieurs protéines ont été identifiées comme interagissant avec le domaine kinase de SYMRK et fonctionnent en aval de celui-ci. Une liste de cibles potentielles de phosphorylation de SYMRK a pu être établie à partir de cet article de Jayaraman et al, 2017. On retrouve par exemple une trehalose-6-phosphate synthase et une peptidylprolyl cis-trans isomérase qui ont un rôle dans le métabolisme des sucres et dans la signalisation dans les réponses aux stress. Une VAMP-associated protein principalement impliquée dans le trafic vésiculaire est également une cible potentielle de SYMRK.

D'autres études ont permis d'identifier des protéines interagissant avec SYMRK, comme par exemple, SIP1 (SYMRK Interacting Protein 1), qui possède un domaine d'interaction avec l'ADN, lui permettant de se lier à des séquences riches en AT, notamment celles présentes dans le promoteur du facteur de transcription NIN (Zhu *et al.*, 2008). La protéine SIP2 (SYMRK Interacting Portein 2) appartient à la famille des MEK Map Kinase. Les cascades de signalisation MAPK aboutissent à des réponses cellulaires comme l'induction de divisions, la différenciation ou des ajustements physiologiques à différents stress. Ici, l'extinction de

l'expression de SIP2 provoque une réduction drastique du nombre de nodules sur les racines transformées (Chen *et al.*, 2012). Une mevalonate synthase (HMGR1) a été plus récemment identifiée chez *M. truncatula* et semble jouer un rôle majeur dans la transduction du signal Nod car l'extinction de l'expression de HMGR1 par RNAi conduit à une forte réduction de la nodulation. Le recrutement de HMGR1 par SYMRK est requis pour la production de composés isoprenoïdes (notamment le mevalonate) qui sont connus pour être des messagers seconds. Ils ont été proposés faire le lien entre la membrane plasmique où sont situés les récepteurs symbiotiques et l'enveloppe du noyau où des flux ioniques sont induits en réponse aux FN (Kevei *et al.*, 2007)**.**

Genèse et décodage d'oscillations calciques au niveau du noyau

Le mévalonate, produit direct de l'activité de HMGR1, est suffisant pour induire des oscillations calciques nucléaires et l'expression de certains gènes symbiotiques chez *M. truncatula* et le mutant *dmi2* (SYMRK) mais pas chez le mutant *dmi1* (POLLUX) (Venkateshwaran *et al.*, 2015). Ce gène, MtDMI1/LjPOLLUX, code un canal cationique localisé au niveau de l'enveloppe nucléaire et ne semble pas être directement responsable des oscillations calciques (Peiter *et al.*, 2007). Chez *L. japonicus*, en plus de LjPOLLUX/MtDMI1, un homologue nommé LjCASTOR est également impliqué dans la génération des oscillations calciques. Une simple mutation induisant un changement d'acide aminé dans la région du filtre sélectif modifie drastiquement le comportement de MtDMI1 par rapport à LjPOLLUX, et montre une meilleure efficacité pour générer les oscillations calciques. Le gène MtDMI1 peut secourir le phénotype de nodulation des mutants *pollux*, *castor* et le double mutant *castor-pollux* alors que LjPOLLUX et LjCASTOR sont tous les deux nécessaires pour complémenter le mutant *dmi1*. Ces données montrent que chez *M. truncatula*, MtDMI1 est capable de remplir sa fonction seul alors que LjPOLLUX et LjCASTOR sont tous deux requis chez *L. japonicus* pour remplir cette même fonction (Venkateshwaran *et al.*, 2012). LjCASTOR/MtDMI1 est un transporteur hautement sélectif de Ca^{2+}, et non pas un transporteur de K+ comme cela avait été proposé en 2008 (Charpentier *et al.*, 2008). La structure de cette protéine a été finement étudiée, plusieurs domaine permettant la liaison au Ca^{2+} ont été découvert et son activation requiert la présence de Ca^{2+} cytosolique ou nucléosolique (Kim *et al.*, 2019). Il existe aussi d'autres canaux calciques qui sont

directement responsables de ces oscillations, une famille de CNGC (Cyclic Nucleotide-Gated Channels) a été étudiée chez *M. truncatula* et a permis d'identifier 3 gènes candidats. L'extinction de l'expression de ces 3 gènes CGNC15a/CGNC15b/CGNC15c, inhibe l'infection rhizobienne et l'induction de gènes précoces en réponse aux facteurs Nod. Il a aussi été montré qu'un complexe composé des canaux CNGC15 et de MtDMI1, est formé sur la membrane nucléaire. Ces données montrent que les protéines CNGC15 sont responsables du transport du calcium durant la signalisation symbiotique, agissant en concert avec DMI1 pour activer et moduler le signal calcique chez *M. truncatula* (Charpentier *et al.*, 2016). La signature typique des oscillations calciques induites par les facteurs Nod est ensuite perçue et décryptée par la protéine nucléaire CCaMK (Calcium and Calcium/Calmodulin-dependent serine/threonine protein kinase) ou MtDMI3 chez *M. truncatula,* dont la mutation inhibe complètement les réponses précoces de nodulation alors que la production d'oscillations calcique est maintenue. Cette protéine possède un domaine « calcium-binding » permettant la liaison du Calcium et un domaine kinase pour la transduction du signal, entre ces deux domaines on retrouve un domaine d'autorégulation permettant de moduler l'activité kinase via le calcium et la calmoduline (Levy, 2004; Gleason *et al.*, 2006). Une fois les oscillations calciques perçues par CCaMK, le signal est transduit par le domaine kinase qui interagit et phosphoryle un activateur transcriptionnel nommé LjCYCLOPS/MtIPD3. Cette protéine possède un domaine de liaison à l'ADN et un domaine kinase, elle est activée par la phosphorylation par CCaMK et, à son tour, active l'expression d'un réseau de facteurs de transcription (Yano *et al.*, 2008 ; Singh *et al.*, 2014).

Facteurs de transcription et induction de gènes symbiotiques

La signalisation à partir de ce point se différencie entre l'infection et l'organogénèse du nodule, plusieurs facteurs de transcription contrôlant l'expression de gènes impliqués dans ces deux processus ont été identifiés. Après phosphorylation de CYCLOPS, il se lie aux promoteurs de deux facteurs de transcription majeurs qui vont être décrits dans les paragraphes suivants et qui sont : ERN1 (Ethylene response factor Required for Nodulation 1) et NIN (Nodule INception), qui sont tous deux impliqués dans l'infection et l'organogénèse du nodule (Yano *et al.*, 2016 ; Singh *et al.*, 2014).

Notons que l'activation des programmes d'infection et d'organogénèse passent aussi par deux facteurs de transcription clé de type GRAS : NSP1 et NSP2, qui interagissent ensemble pour former un hétéro-complexe requis pour la liaison au promoteur et la transcription de gènes codant des nodulines précoces et d'autres facteurs de transcription comme NIN et ERN1 (Hirsch *et al.*, 2009 ; Cerri *et al.*, 2012). Le gène NSP2 intègre également des signalisations hormonales car il est fortement induit dans le cortex racinaire après application de cytokinines, dépendemment de CRE1 (Plet *et al.*, 2011). Le gène LjLHK1/MtCRE1, qui est un récepteur aux cytokinines (CKs), est essentiel pendant la transduction du signal facteur Nod et est impliqué dans la formation du primordium nodulaire. La mutation de ce gène provoque une forte diminution de la nodulation malgré le succès de nombreux évènements d'infection (Murray *et al.*, 2007).

Par ailleurs, les gibbérellines peuvent réguler le processus de nodulation chez les légumineuses, bien que plusieurs rôles ayant un impact négatifs ou positifs sur la nodulation ont été rapportés en fonction de la plante hôte (Ferguson *et al.*, 2005; Maekawa *et al.*, 2009)**.** Les protéines DELLAs sont les régulateurs centraux de la signalisation de l'acide gibbérellique et font également partie intégrante de la voie Nod chez *M. truncatula*. Il est suggéré que l'action des protéines DELLA, dépendamment des gibbérellines, peut former un complexe avec NSP1/NSP2 et NF-YA1, activant ainsi la transcription de ERN1 pour contrôler le processus d'infection (Fonouni-Farde *et al.*, 2016). De plus, les protéines DELLAs peuvent promouvoir la formation du complexe CCaMK/CYCLOPS (DMI3/IPD3) et augmenter le niveau de phosphorylation de CYCLOPS, en plus d'interagir avec NSP1 et NSP2. Il a été démontré que les DELLAs forment un complexe et servent de pont entre CYCLOPS et NSP2 (Jin *et al.*, 2016).

Le gène ERN1 contenant un domaine APETALA2/Ethylene Responsive Factor (AP2/ERF) ainsi que son homologue ERN2, sont exprimés lors de l'infection rhizobienne et dépendent directement de CYCLOPS et NSP2 pour leur activation. Chez *M. truncatula*, les simples mutants *ern1* et *ern2* présentent un blocage de l'infection et un retard de la nodulation alors que les doubles mutants *ern1-ern2* sont incapables d'initier l'infection et d'induire des nodules. Ces résultats suggèrent une fonction redondante de ces gènes dans le processus de nodulation chez *M. truncatula*. En revanche chez *L. japonicus*, où ERN2 est absent et où les mutants *ern1* sont incapables d'initier l'infection et de produire des nodules, ERN1 est

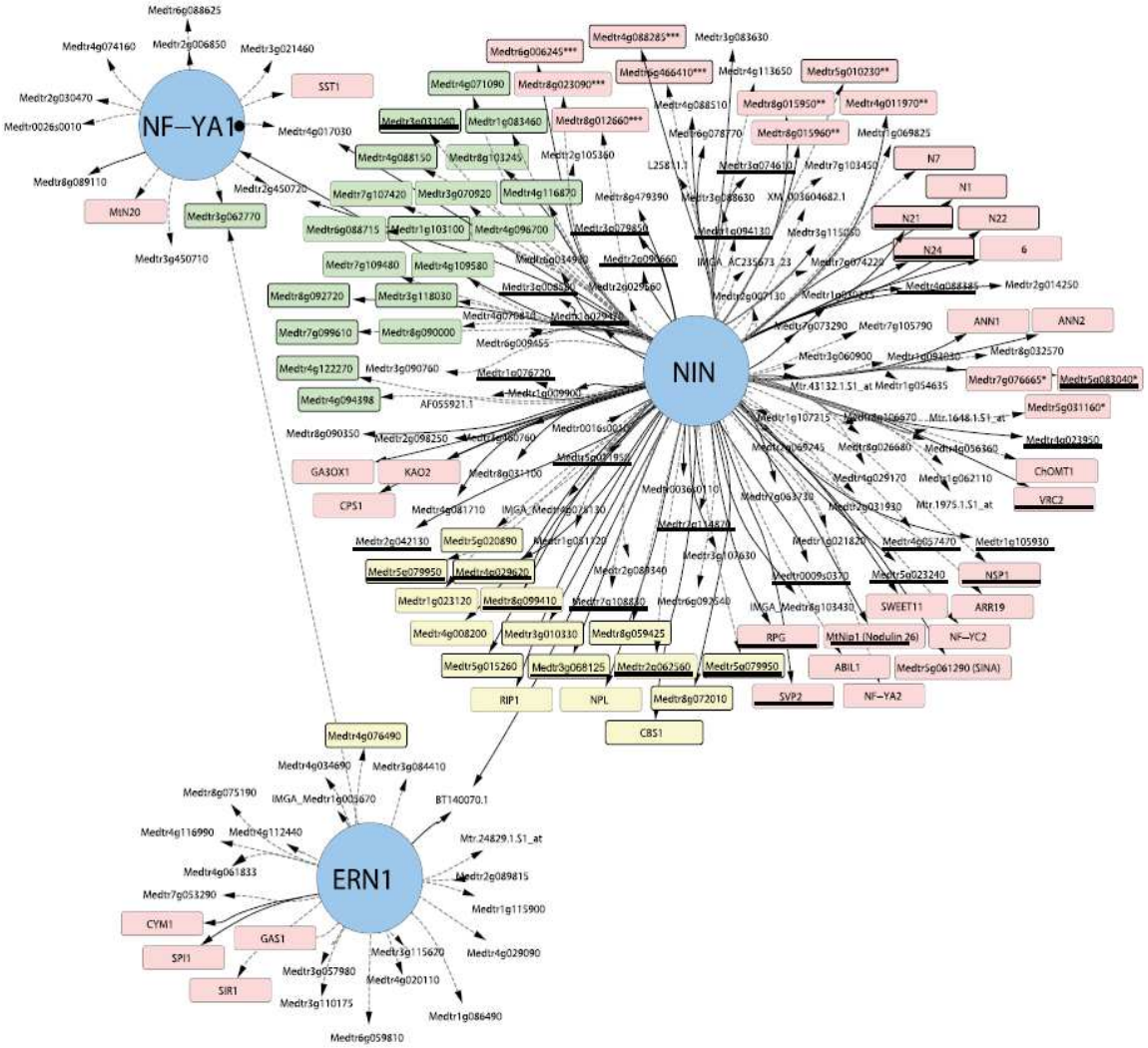


Figure 13 : Réseau de régulation génétique de l'infection rhizobienne par trois facteurs de transcription NF-YA1, ERN1 et NIN. Ce dernier apparait comme un régulateur central du processus de nodulation. Les flèches en ligne pleine correspondent à une régulation positive forte. Les gènes en jaune sont relatifs à la paroi végétale, ceux en vert sont relatifs à la réplication/transcription et ceux en rouge correspondent à des gènes symbiotiques connus. Les gènes surlignés possèdent des homologues identifiés chez *L. japonicus* avec une expérience de ChIP-seq sur NIN. (D'après Liu et al, 2019)

suffisant seul pour réaliser sa fonction dans l'infection et l'organogénèse (Cerri *et al.*, 2016; Kawaharada *et al.*, 2017). Pour finir, le facteur de transcription NIN fait partie de la famille NLP (NIN-Like Protein) chez *Arabidopsis* qui est une famille connue pour être impliquée dans la réponse aux nitrates. Cependant, les gènes LjNLP1 et plusieurs AtNLPs sont activés en réponse au statut azoté de la plante alors que le gène LjNIN a perdu la capacité de s'activer dans ces conditions alors qu'il peut être activé par l'application exogène de CKs (Heckmann *et al.*, 2011; Suzuki *et al.*, 2013). Les mutants pertes de fonction chez *L. japonicus* pour NIN (et aussi pour NF-YA1) ne sont plus capables de former des nodules tandis que la surexpression de NIN induit des structures semblables aux primordia nodulaires dans la zone corticale de la racine en absence de bactéries. L'étude du transcriptome de mutants *nin*, a permis d'identifier de nombreux gènes potentiellement régulés par ce facteur de transcription comme ceux impliqués dans le remodelage de la paroi végétale, le cycle cellulaire, l'assimilation de nutriments et la signalisation phytohormonale (gibbérelline, jasmonate). Le recrutement de ce nombre impressionnant de gènes sous le contrôle de NIN en fait le noyau central de la régulation de l'infection rhizobienne (Liu *et al.*, 2019a) (Fig. 13).

Les analyses génétiques réalisées sur *M. truncatula* et *L. japonicus* ont mis à jour de nombreux acteurs moléculaires jouant un rôle dans la transduction du signal provenant de la membrane jusqu'au noyau et également dans l'induction de l'expression de nombreux gènes sous le contrôle de facteurs de transcriptions spécifiques. La conservation des mécanismes moléculaires entre les deux légumineuses modèles a été démontrée, bien que quelques différences puissent être observées dû à des histoires évolutives différentes. Cette conservation de mécanismes va même au-delà, car chez certains mutants comme *symrk* qui sont Nod^-, les tentatives d'infection de la part d'un champignon endomycrohizien sont toutes infructueuses sur la racine, montrant que *SYMRK* est un composant commun de la signalisation entre la nodulation et la mycorhization (Stracke *et al.*, 2002). Des analyses phénotypiques réalisées sur les mutants des gènes de cette voie de signalisation Nod ont montré que plusieurs gènes avaient un phénotype d'endomycorhization (*SYMRK, POLLUX, CCAMK, IPD3, NSP1, NSP2*) suggérant qu'une partie de cette voie est également partagée avec les plantes mycorhiziennes. Pour NIN, plusieurs études ont obtenu des résultats contradictoires sur la mycorhization, en effet chez *M. truncatula* une première étude montrait une diminution de la colonisation mycorhizienne chez le mutant *nin* (Guillotin *et*

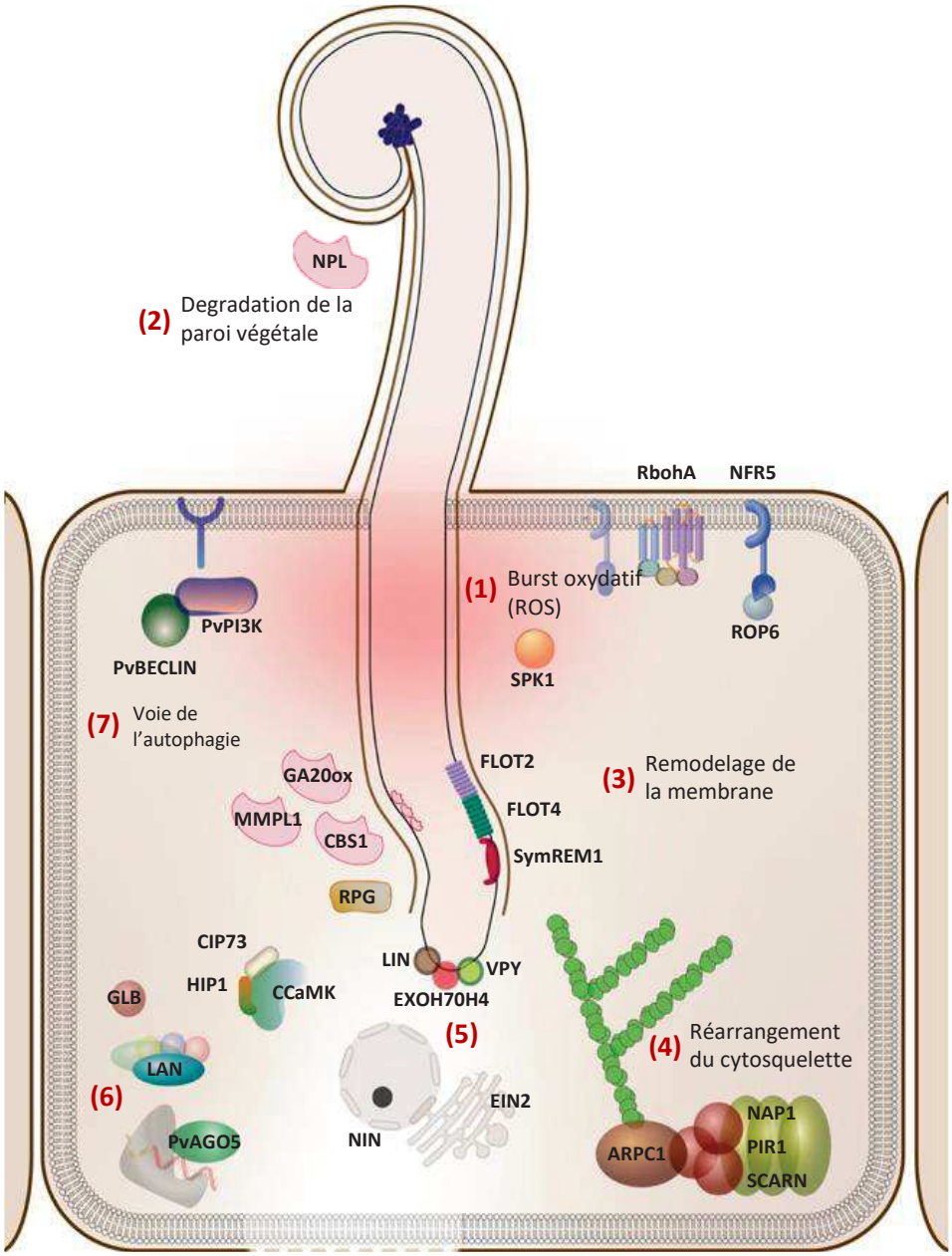


Figure 14 : L'infection rhizobienne. L'entrée des bactéries dans la cellule végétale déclenche (1) un burst oxydatif accompagné par (2) une dégradation locale de la paroi végétale, (3) un remodelage de la membrane et (4) un ré-arrangement du cytosquelette. A l'apex du cordon d'infection on retrouve (5) un complexe protéique nommé infectiosome. En parallèle, (6) plusieurs changements transcriptionnels et post-traductionnels ont lieu et (7) la voie de l'autophagie permet l'accomodation de la bactérie dans la cellule hôte. (Adapté de Roy et al, 2020)

al., 2016) alors qu'une étude plus récente tendait à infirmer ces résultats en montrant que des mutants *nin* sont mycorrhizés de manière semblable aux témoins (Kumar *et al.*, 2020), comme c'est aussi le cas chez *L. japonicus* (Geurts *et al.*, 2016). La principale différence entre les symbioses mycorhizienne et rhizobienne réside dans la formation d'un nouvel organe symbiotique qui n'est induit que dans le cas de l'interaction rhizobium-légumineuse. Il existe donc des gènes nécessaires à l'organogénèse nodulaires comme par exemple *LHK1*/*CRE1* et *NIN*, dont la surexpression induit des nodules spontanés (Tirichine *et al.*, 2007; Soyano *et al.*, 2013) et dont les mutants n'ont pas de phénotype de mycorhization (Laffont *et al.*, 2015; Kumar *et al.*, 2020), montrant qu'ils sont spécifiques au processus de nodulation.

2.d - L'infection symbiotique

Une liste non-exhaustive de gènes impliqués dans l'infection est fournie dans la Figure 14.

Rôle du cytosquelette et des enzymes de dégradation de la paroi lors de l'infection

Le processus d'infection décrit en début de partie 2 est orchestré par de nombreux gènes symbiotiques dont beaucoup ont été identifiés grâce à l'analyse de mutants ou à des approches d'extinctions de leur expression. Les gènes *Nap1*, *Pir1* et *SCARN* qui codent des sous-unités du complexe SCAR/WAVE impliqué dans des réarrangements d'actine s'avèrent être requis pour assurer la polarisation des cellules au cours de l'infection chez *L. japonicus*. Les mutants pour ces gènes sont en effet perturbés dans la formation et la progression du cordon d'infection, conduisant à un nombre élevé de nodule non-infectés. Mais ce ne sont pas des gènes avec une fonction spécifique à la symbiose car les mutants présentent aussi des poils absorbants et des trichomes altérés (Yokota *et al.*, 2009; Miyahara *et al.*, 2010).

Une fois la chambre d'infection formée grâce à une expansion des membranes et le remodelage de la paroi végétale, les bactéries doivent passer la barrière constituée par la paroi végétale du poil absorbant pour former le cordon d'infection. Durant ce processus il semble que le matériel nécessaire au développement de la chambre d'infection et au cordon d'infection comme des enzymes dégradant les composés de la paroi végétale (cellulose, hémicellulose, pectine et protéines) de façon à détendre les mailles de cette structure rigide, soit délivré grâce aux Vesicle-Associated Membrane Proteins ou VAMPs faisant partie de la machinerie d'exocytose de l'hôte (Ivanov, 2012). Parmi les enzymes de dégradation de la paroi, lenzyme NPL (Nodule Pectate Lyase) est capable de dégrader la pectine ainsi que

l'acide polygalacturonique, et son activation dépend de la voie de signalisation Nod (notamment de NIN). Les mutants *npl* produisent des nodules non-infectés et la majorité des évènements d'infection sont stoppés au niveau du foyer d'infection à l'apex du poil absorbant, cela indique que cette enzyme joue un rôle important dans la dégradation de la paroi cellulaire de la plante hôte lors de l'initiation du processus symbiotique (Xie *et al.*, 2012).

Contrôle de l'infection par plusieurs types de récepteurs LysM-RLK

Un gène appartenant à une famille maintenant bien connue, les LysM-RLK, est nécessaire pour initier les cordons d'infections. Le récepteur LjEPR3 (ExoPolysaccharide Receptor 3) n'est pas capable de percevoir les LCOs mais peut lier les exopolysaccharides de surfaces (EPS) présents dans la membrane des bactéries. Les mutants *epr3* montrent des problèmes d'infection et sont incapables de percevoir des EPS purifiés, c'est la première fois qu'un LysM-RLK est décrit comme étant capable de lier d'autres substrats que des dérivés de chitine. Il a été proposé que la reconnaissance des EPS par EPR3 constitue un deuxième niveau de spécificité contrôlant la mise en place de l'interaction symbiotique (Kawaharada *et al.*, 2015; Wong *et al.*, 2020). La perception des facteurs Nod, qui est importante dans les étapes précoces de la symbiose, s'avère également nécessaire pour assurer l'élongation de ce cordon d'infection. En effet, si la majorité des lignées RNAi obtenues pour les gènes *LYK3* et *NFP* chez *M. truncatula* montrent un blocage précoce de l'interaction symbiotique, similaire à celui observé avec les mutants respectifs, certaines lignées RNAi développent des cordons d'infection anormaux et avortés, ce qui montre que ces gènes codent des protéines constituant le complexe récepteur aux facteurs Nod sont impliqués dans ce processus (Arrighi *et al.*, 2006; Limpens, 2003; Smit *et al.*, 2007).

Plusieurs protéines capables d'interagir avec le complexe récepteur aux facteurs Nod et intervenant dans le processus d'infection ont été identifiées. Comme SYMREM1 (SYMbiotic REMorin 1) dont l'extinction de l'expression chez *M. truncatula* conduit à la formation de nodules anormaux et de nombreux cordons d'infection avec une structure en forme de sac, qui avortent tous dans le cortex externe, indiquant que la croissance polaire du cordon a été affectée. La protéine SYMREM1 est localisée à la membrane plasmique au niveau de micro-domaines membranaires et interagit physiquement avec NFP et LYK3, elle permet le

contrôle de la répartition spatiale du complexe récepteur au facteur Nod pendant le développement du nodule (Lefebvre *et al.*, 2010). Les micro-domaines membranaires sont définis par l'accumulation de protéines à un endroit donné sur la surface membranaire, comme c'est le cas par exemple des radeaux lipidiques.

D'autres protéines ont été identifiées comme intervenant dans le même processus, FLOT2 et FLOT4, des flotillin-like localisées à l'extrémité du poil racinaire et dont l'extinction d'expression provoque une augmentation du nombre de cordon d'infection qui avortent. Il est suggéré que FLOT2 et FLOT4 sont impliqués dans la première invagination du cordon dans la cellule du poil racinaire et que FLOT4 est également nécessaire pour la croissance du cordon d'infection. Ces protéines forment des micro-domaines membranaires qui sont dispersés sur le poil absorbant mais la protéine s'accumule à l'apex du poil 1 jour après inoculation de la bactérie. De plus, en absence de bactéries, LYK3 et FLOT4 ne sont pas localisés au même endroit sur le poil absorbant alors qu'ils co-localisent et stabilisent la membrane après l'inoculation (Haney and Long, 2010; Haney *et al.*, 2011). Les micro-domaines membranaires recrutent directement le complexe récepteur aux facteurs Nod et le stabilise, le recrutement de LYK3 sur ces micro-domaines spécifiques est nécessaire au processus d'infection. Le modèle actuel suggère que SYMREM1 sert d'échafaudage secondaire qui requiert FLOT4 pour immobiliser physiquement LYK3 dans les micro-domaines (Liang *et al.*, 2018).

<u>Existence d'un complexe protéique assurant la croissance polaire du cordon d'infection</u>

Plusieurs mutants altérés dans l'infection ont permis d'identifier des gènes dont le rôle précis est resté longtemps obscure. Fort heureusement de récents travaux viennent mettre en lumière les mécanismes de certains de ces gènes symbiotiques et viennent compléter notre savoir sur les premières étapes de la nodulation.

Les mutants du gène LIN (Lumpy INfection) chez *M. truncatula* sont affectés dans la progression du cordon d'infection et le développement du primordium nodulaire. Le gène LIN code une E3 ubiquitin-protein ligase, la dégradation des protéines sous contrôle de l'ubiquitine est un mécanisme universel pour réguler de nombreux processus biologique en éliminant certaines enzymes ou facteur de transcriptions. LIN semble donc être un régulateur de composants de la voie de signalisation Nod et son fonctionnement est

nécessaire pour que les protéines ciblées aient une activité spatio-temporelle correcte (Kiss *et al.*, 2009). Chez les mutants *vpy* le foyer d'infection peut se former à l'apex du poil absorbant mais l'initiation du cordon d'infection est retardée ou absente. La protéine VPY de fonction inconnue, possède un domaine d'interaction protéine-protéine suggérant qu'elle fonctionne en recrutant d'autres partenaires protéiques (Murray *et al.*, 2011). Cette protéine est exprimée pendant la mycorhization et co-localise avec une sous-unité du complexe exocyste nommé EXO70I sur la surface membranaire. Paradoxalement, les mutants *exo70i* sont impactés dans la mycorhization mais pas dans la nodulation. Finalement, de récents travaux ont permis d'identifier un complexe multi-protéique au niveau de la membrane, qui fait intervenir 3 autres protéines pendant la nodulation (LIN, VPY, EXO70H4) et qui est localisé à l'apex du cordon d'infection en croissance. De plus, le mutant *exo70h4* présente un phénotype similaire aux mutants *vpy* et *lin*. En effet, VPY interagit avec LIN grâce à une série de répétition ankyrine située coté C-terminal et il semble que cette interaction stabilise VPY, car son signal fluorescent (VPY-GFP) est réduit chez les mutants *lin* par rapport au sauvage. Il est suggéré que ce complexe moléculaire assure la croissance polaire du cordon d'infection (Liu *et al.*, 2019b).

D'autres gènes sont requis pour l'infection symbiotique

Plusieurs gènes ayant montré un rôle dans le processus d'infection ont été découverts, mais bien souvent leur fonction dans le mécanisme global reste mal comprise.

Parmi les premiers gènes d'infection identifiés, *RPG* code une protéine à motifs coiled-coils induite pendant l'infection mais également impliqué dans l'activation d'autres gènes relatifs à l'infection. Les mutants chez *M. truncatula* arborent des cordons d'infection très épais et qui avancent lentement conduisant à un nombre réduit de nodules fixateurs, ce phénotype suggère qu'il est impliqué dans la croissance polarisée du cordon d'infection. Cependant, sa fonction précise n'a pas encore été révélée malgré qu'il soit considéré comme un gène clé de l'évolution de la nodulation (Arrighi *et al.*, 2008). D'autre part, les mutants de *M. truncatula* sur le gène *CBS1* (Cystathionine-β-Synthase-like) montrent des cordons d'infections défectueux qui sont parfois dépourvus de rhizobiums et qui forment de petits nodules avec peu de symbiosomes. Cette protéine est exprimée uniquement en conditions symbiotiques et est localisée au niveau du cordon d'infection et des symbiosomes. Des

gènes homologues chez *Arabidopsis* sont impliqués dans la maturation de la paroi végétale, suggérant un rôle potentiel de *CBS1* dans la formation de la paroi du cordon d'infection (Sinharoy *et al.*, 2016). Aussi, la protéine RINRK1 (Rhizobial INfection Receptor-like Kinase 1) a récemment été identifiée chez *L. japonicus*, elle appartient à la famille des LRR-RLK et est spécifiquement impliquée dans le processus d'infection. En effet, la plupart des mutants sont incapables d'initier le cordon d'infection et seulement quelques nodules blancs non-infectés peuvent être observés 18 jours après inoculation. Un homologue (RINRK2) a également été identifié et est lui aussi impliqué dans l'infection, bien que la mutation de ce gène ne provoque pas de phénotype, chez le double mutant *rinrk1-rinrk2* il y a un phénotype plus drastique que pour *rinrk1*. Parmi les RLK, RINRK1 est atypique car son domaine kinase n'est pas fonctionnel, ce qui pose la question de savoir avec qui il interagit et comment il assure la transmission du signal au niveau intracellulaire (Li *et al.*, 2020).

Endocytose des bactéries dans des symbiosomes

Une fois que le cordon d'infection atteint le primordium nodulaire, des structures spécialisées au niveau de l'apex du cordon d'infection, dépourvu de paroi et qu'on appelle gouttelettes d'infection, se forment. A partir de ces excroissances les bactéries sont relâchées dans le cytoplasme d'une cellule racinaire. Dans le cytoplasme, les bactéries sont isolées par une membrane péribactéroïdienne d'origine végétale et on appelle l'ensemble le symbiosome où elles fixent l'azote atmosphérique (Brewin, 2004).

L'ensemble des gènes impliqués dans le processus d'infection et les liens qui les unissent ne sont pas encore tous connus. Cela est notamment dû au fait que les efforts de recherche se sont focalisés initialement sur les gènes impliqués dans la perception des rhizobia et la signalisation symbiotique. Néanmoins, la caractérisation de gènes d'infection a déjà permis de mettre en lumière deux traits notables. Tout d'abord, NIN, le facteur de transcription considéré comme central dans la mise en place de la symbiose, est nécessaire pour l'expression de très nombreux gènes symbiotiques intervenant dans le processus d'infection (Liu *et al.*, 2019a). Ensuite, plusieurs gènes relatifs à l'infection comme ceux impliqués dans le réarrangement du cytosquelette ou de l'endocytose sont partagés lors des interactions symbiotiques entre la plante hôte et les rhizobiums/champignons endomycorhiziens, en lien avec la similarité des structures créées dans le cytoplasme de la cellule hôte lors des deux

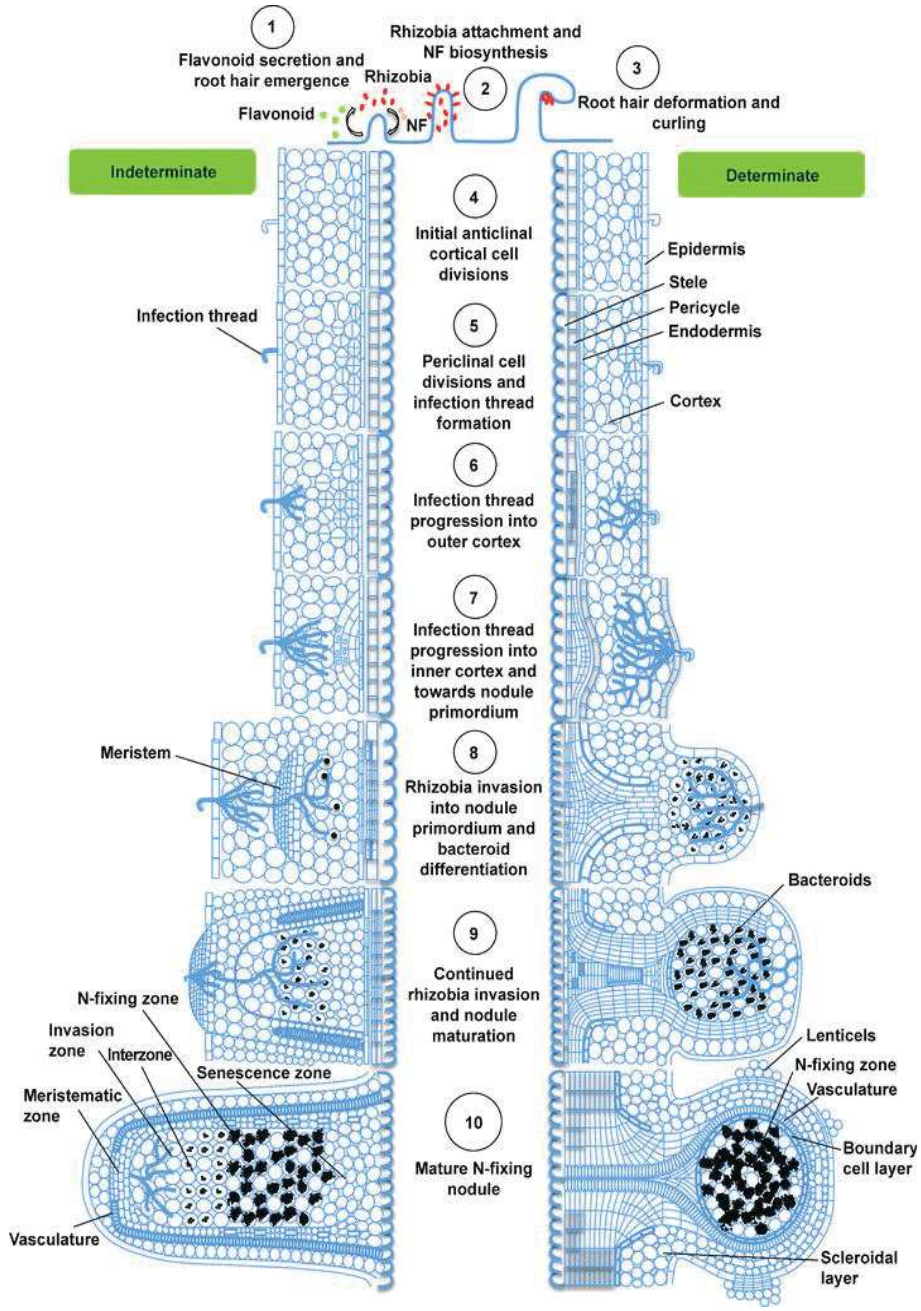


Figure 15 – Etapes de développement de nodules déterminés et indéterminés. Organogénèse de nodule indéterminé de pois (gauche) et déterminé du soja (droite). (1) La production de flavonoïdes attirent les bactérie et (2) induit la synthèse de facteurs Nod. (3) Le poil absorbant se déforme et créé une chambre d'infection contenant les bactéries, à partir de laquelle le cordon est initié. (4-5) Des divisions cellulaires sont induites dans le cortex interne (nodules indéterminés) ou dans la couche de cellule sub-épidermique (nodules déterminés). (6-7) Le cordon d'infection progresse dans les tissus et (8) infecte le primordium nodulaire où les bactéries commencent a se différencier. (9-10) Le nodule devient matûre avec une zonation pour les nodules indéterminés, à l'apex on retrouve un méristème persistant. A l'inverse, pour les nodules déterminés, le méristème est transitoire, les cellules de la zone centrale du nodule sont alors toutes au même stade de développement (Ferguson *et al.*, 2010)

types de symbioses endo-racinaires (cordon d'infection/arbuscule) (Hause and Fester, 2005). Ainsi, comme pour le processus de perception et de signalisation, plusieurs acteurs moléculaires de l'infection rhizobienne sont dérivés de la machinerie permettant la symbiose mycorhizienne.

2.e - Mise en place et fonctionnement du nodule fixateur d'azote

Nodules déterminés et nodules indéterminés

Chez les légumineuses modèles, on distingue deux types de nodules. Pour *M. truncatula* les nodules produits sont de type « indéterminés », de forme allongée avec un méristème persistant donnant lieu à une zonation (Fig. 15). Dans la 1ère zone à l'apex du nodule on retrouve le méristème nodulaire, la 2ème zone correspond à la zone d'infection, la 3ème zone regroupe des cellules symbiotiques matures permettant de fixer l'azote et la 4ème zone observable dans les nodules âgés correspond à la zone de senescence composé de cellules symbiotiques en dégénérescence. Le tissu est infecté de manière hétérogène car les bactéries sont libérées dans certaines cellules issues du méristème alors que des cellules non-infectées subsistent (Van de Velde *et al.*, 2006). En revanche chez *L. japonicus* le méristème est transitoire et donne naissance à un nodule « déterminé » sphérique (Fig. 15). Il n'y a pas de zonation chez cette espèce où la zone de fixation correspond au tissu infecté central du nodule. Dans les deux cas, les nodules se développent à partir du cortex racinaire et, les nodules formés possèdent une vascularisation périphérique facilitant les échanges avec le tissu infecté central (Schubert, 1986; Sprent, 2007). En revanche la racine latérale est issue de la réactivation du cycle cellulaire de cellules fondatrices spécifiques sensibles à l'auxine située dans le péricycle (zone méristématique).

Modalités de l'organogenèse nodulaire

En parallèle de l'infection, la voie de signalisation Nod aboutit à l'activation du programme d'organogenèse du nodule (Madsen *et al.*, 2010). Des formes mutées auto-actives de certaines protéines de la voie de signalisation Nod comme CCaMK et CRE1/LHK1 ou la surexpression de NIN et NF-YA1, sont suffisants pour induire la formation de nodules spontanés sur la racine, traduisant leur rôle clé dans le développement du nodule. Mais, tandis que la perception des facteurs Nod est initiée au niveau des poils absorbants, elle

provoque des réponses au niveau du cortex racinaire où se forme le primordium nodulaire. Ce décalage spatial implique qu'il existe une transduction du signal entre les différentes assises cellulaires. L'inoculation de la bactérie chez *M. truncatula* augmente la connectivité symplastique entre les cellules des tissus externes de la racine. Cette connectivité symplastique est nécessaire pour la colonisation des rhizobiums durant la nodulation et plusieurs molécules transportées ont pu être identifiées comme des facteurs de transcription. Ces travaux suggèrent qu'il existe un signal non-cellulaire autonome permettant d'intégrer les programmes génétiques épidermaux et corticaux mais la nature de ces signaux est encore inconnue (Gaudioso-Pedraza *et al.*, 2018).

La formation du nodule fait intervenir de la signalisation hormonale, notamment les cytokinines (CKs) et l'auxine (IAA). En effet, les lignées RNAi du récepteur aux cytokinines LHK1/CRE1 ne développent plus de nodules (Yang *et al.*, 2015) et un mutant gain de fonction produit des nodules spontanés en absence de bactéries et est connu pour avoir un rôle antagoniste dans le développement des racines latérales, suggérant des points communs dans le contrôle du développement nodulaire et racinaire via les CKs (Tirichine *et al.*, 2007). Chez les légumineuses, l'inoculation ou l'application de facteurs Nod provoque des changements locaux dans l'accumulation d'auxine et l'augmentation de l'expression de transporteurs d'auxine (AUX1), dans les primordia nodulaires et en périphérie des nodules en formation (de Billy *et al.*, 2001; Pacios-Bras *et al.*, 2003). De plus, une étude récente montre que malgré une induction différente, les programmes génétiques de développement des nodules et des racines latérales se chevauchent. En effet, les mutants *lbd16* (LBD16, LoB-Domain protein 16) chez *M. truncatula* sont altérés à la fois dans l'initiation des racines latérales et aussi celle des nodules. Le gène NIN (inductible par les cytokinines) permet l'induction du programme génétique d'organogénèse nodulaire en passant par l'activation de LBD16, qui augmente la biosynthèse d'auxine via l'induction transcriptionnelle de deux gènes : STY (STYlish) et YUC (YUCca). Les cytokinines facilitent donc l'accumulation d'auxine localement grâce à l'induction de LBD16 par NIN, ce qui active un programme génétique s'imbriquant dans celui induit lors de la formation des racines latérales (Schiessl *et al.*, 2019). Un lien fort entre nodule et racine est également révélé par le phénotype original du mutant *noot* chez *M. truncatula* qui est caractérisé par une perte partielle d'identité du nodule qui évolue en racine (noot : nodule-root). Ce phénotype est dû à la perte de fonction du gène

NOOT1 (Nodule rOOT1) appartenant à la famille de co-facteurs de transcription NBCL (NOOT-BOP-COCH-LIKE) (Couzigou *et al.*, 2012). Le paralogue de ce gène, nommé NOOT2 est également exprimé dans le méristème apical nodulaire mais les mutants ne présentent aucun phénotype symbiotique. En revanche, le double mutant *noot1-noot2* montre une perte totale de l'identité nodulaire. Il est suggéré que ces gènes sont responsables de la répression de l'identité racinaire au sein du nouvel organe symbiotique (Magne *et al.*, 2018). Cependant, MtNOOT1 est également nécessaire pour la coordination du développement racinaire. En effet, le mutant montre une élongation racinaire rapide à cause d'un plus gros méristème racinaire et une différenciation des cellules du xylème retardé. MtNOOT1 n'est pas seulement capable de maintenir l'identité du nodule mais coordonne également le développement de la racine primaire de la base à l'apex (Shen *et al.*, 2020).

Différenciation des bactéries et des cellules végétales dans le nodule

Au sein du nodule, les bactéries qui sont libérées dans les symbiosomes se différencient en bactéroïdes pour pouvoir réaliser leur fonction de fixation de l'azote. Suivant la plante-hôte, cette différenciation est plus ou moins poussée, elle peut être associée à une augmentation du nombre de copies du génome, un élargissement des bactéries voire un changement de leur forme. Il existe plusieurs morphotypes de bactéroïdes dépendant de l'hôte, on en trouve des allongés (E : elongated), gonflés (S : swollen), sphériques (SP : spherical) et allongés-branchés (EB : elongated-branched). A l'inverse de ce qui est observé chez *L. japonicus*, certaines légumineuses induisent une différenciation en bactéroïdes dite « terminale » car elle conduit à une perte de la capacité reproductive. Chez *M. truncatula*, cette différenciation terminale est gouvernée par l'expression de peptides de 30 à 50 acides aminés riches en cystéines nommés NCR (Nodule Cystein-Rich peptide). Il a été montré que les NCRs étaient adressés aux bactéries et entraient dans la membrane bactérienne et le cytosol. L'obstruction du transport des NCRs chez un mutant *dnf1* (signal peptidase) est corrélée à l'absence de différenciation terminale chez les bactéries. A l'inverse, l'expression ectopique de NCRs chez des légumineuses qui n'en ont pas ou l'application de NCRs sur des cultures bactériennes provoquent des symptômes de différenciation terminale. Le peptide mature cible les bactéroïdes en interagissant avec la membrane et plusieurs composants cytosoliques pour interférer avec la régulation du cycle cellulaire (Oono *et al.*, 2010;

Mergaert *et al.*, 2006; Van de Velde *et al.*, 2010; Bonaldi *et al.*, 2011). Chez *M. truncatula*, plus de 600 gènes codant des NCRs ont été identifiés dont la majorité sont fortement exprimés lors de la nodulation mais qui peuvent posséder des charges différentes. Les NCRs ont également été trouvés chez d'autres espèces du clade des IRLC auquel appartient *M. truncatula* ; le nombre de NCRs, la composition du cocktail de NCR (cationique/anionique) et le niveau d'expression des peptides cationiques a pu être corrélés aux morphotypes des bactéroïdes contenus dans les cellules végétales (Montiel *et al.*, 2017).

D'autres phénomènes de différenciations ont lieu pendant la formation du nodule, notamment un phénomène d'endo-réduplication des cellules corticales racinaires. Les cellules végétales ne subissant pas l'endo-réduplication restent petites et non infectées, alors que les cellules ayant subi plusieurs cycles d'endo-réduplication peuvent atteindre 80 fois leur taille d'origine. Ce mécanisme pourrait être une voie plus rapide que la prolifération cellulaire pour augmenter la taille de l'organe symbiotique. Le gène *CCS52A* qui est un activateur du « anaphase-promoting complex » (APC), a été identifié comme étant impliqué dans la différenciation des cellules végétales symbiotiques mais pas dans leur prolifération. En effet, la dérégulation de ce gène affecte le développement du nodule avec une réduction de la taille des cellules, un niveau de ploïdie plus faible et des problèmes dans la maturation des cellules symbiotiques aboutissant à la mort des bactéries et des cellules végétales (Cebolla, 1999; Vinardell *et al.*, 2003). Chez *L. japonicus*, deux composants du complexe protéique de la DNA Topoisomerase VI ont été identifiés comme participant au phénomène d'endo-réduplication et au bon fonctionnement du nodule. Il s'agit des gènes VAG1 et SUNERGOS1 dont les mutants présentent un nombre faible de nodules ainsi qu'un retard dans l'apparition de ces derniers (Suzaki *et al.*, 2014; Yoon *et al.*, 2014). Les deux processus de différenciation en bactéroïde et d'endo-reduplication sont liés à l'accommodation des bactéries au mode de vie intracellulaire et permettent d'optimiser l'efficacité de l'association symbiotique.

Fonctionnement du nodule pour la fixation de l'azote

Une fois le nodule formé, de nombreux gènes de la plante-hôte entrent en jeu pour réguler son fonctionnement et garantir l'activité des bactéries au sein des symbiosomes. La fixation de l'azote par les symbiontes est permise par le complexe enzymatique de la nitrogénase

dont la sensibilité à l'oxygène a déjà été évoquée. Afin de pallier au problème de l'oxygène dans les nodules les légumineuses expriment des protéines appartenant à la famille des globines (Glb) nommées Leghemoglobines (Lhb). Les Lhbs ont une affinité plus faible pour l'O2 que les Glbs et sont spécifiquement exprimées dans le nodule, elles ont évolué vers une fonction qui consiste à séquestrer le dioxygène pour l'apporter à la bactérie en dose faible mais continue pour le fonctionnement de son métabolisme, sans inhiber l'activité de la nitrogénase. Ces gènes sont souvent présents en plusieurs copies et font partie des plus exprimés dans le nodule (pour revue (Becana *et al.*, 2020). L'extinction de l'expression par RNAi des 3 gènes codant des Lhbs chez *L. japonicus* cause une élévation de la pression partielle en O_2 dans le nodule et une perte de la capacité à fixer l'azote (Ott *et al.*, 2005).

L'efficacité de la symbiose fixatrice d'azote dépend de la capacité de la plante et de la bactérie à échanger des ressources, les rhizobiums étant totalement dépendants de la plante. Cet échange a essentiellement lieu au niveau de la membrane péribactéroïdienne sur laquelle sont présents de nombreux transporteurs protéiques.

La plante hôte contrôle les flux de carbone et d'azote avec le symbiosome mais également d'autres éléments nécessaires au métabolisme des bactéries comme le phosphore, le souffre, le potassium, le sodium, le calcium, le vanadium, le fer, le molybdène, le nickel et le cobalt (Rosendahl *et al.*, 1992). Les bactéries peuvent métaboliser différentes sources de carbone qui sont principalement apportés par la plante-hôte. Le gène *MtSWEET11* est un transporteur de sucrose exprimé dans le poil racinaire infecté, dans le méristème, la zone d'invasion et de vascularisation des nodules (Kryvoruchko *et al.*, 2016). Le sucrose est un dérivé de la photosynthèse qui est métabolisé en malate, cet acide carboxylique est considéré comme la première source de carbone apportée aux bactéroïdes dans le nodule (Colebatch *et al.*, 2004). Il existe également plusieurs transporteurs de polyols/monosaccharides (PLTs) qui sont spécifiquement et fortement exprimés dans le nodule, et dont l'un d'entre eux est localisé au niveau du symbiosome, suggérant un rôle dans le transport de monosaccharides vers les bactéroïdes (Tian *et al.*, 2017). Comme évoqué précédemment, d'autres éléments chimiques sont nécessaires au fonctionnement des bactéroïdes. Chez *L. japonicus* le gène SST1 est impliqué dans le transport des sulfates et est spécifiquement localisé sur le symbiosome (Krusell *et al.*, 2005). Un gène responsable de l'import de fer dans le symbiosome a également pu être identifié chez *M. truncatula,* il s'agit

de MtFPN2 (Ferroportin2) dont la mutation entraine une réduction sévère de la fixation d'azote (Escudero *et al.*, 2020).

D'autre part, l'azote fixé par les bactéries correspond à la forme d'ammonium (NH_4^+) et se diffuse au travers de la membrane de manière passive au sein du symbiosome, puis deux modes de transports distincts vers le cytoplasme ont pu être identifiés : un canal NH_3 et un canal cationique transportant le K^+, Na^+ et NH_4^+ (Niemietz and Tyerman, 2000). Des canaux « aquaporins-like », qui font partie des protéines les plus abondantes dans le symbiosome sont aussi capables de transporter l'ammonium (Hwang *et al.*, 2010). Confortant l'idée que ces aquaporins-like peuvent jouer un rôle important dans l'export du NH_4^+ depuis les symbiosomes, la noduline 26 (aquaporin) chez le soja, interagit avec la Glutamate Synthase qui est l'enzyme responsable de la production de glutamine à partir de NH_4^+, cation qui représente la forme transportable de l'azote fixé dans le système vasculaire de la plante (Masalkar *et al.*, 2010).

Régulation de la nodulation

Le processus de fixation d'azote est coûteux en énergie et la présence d'un trop grand nombre de nodules sur la racine par rapport aux besoins de la plante hôte en azote aurait un effet délétère sur son développement global. Un système d'auto-régulation de la nodulation (AON) permet à la plante de gérer la densité de nodules présents sur sa racine. A ce jour 3 mécanismes qui passent par des régulations systémiques ont été identifiés.

Plusieurs lignées mutantes *har1* (Hypernodulation Aberrant Root formation 1) de *L. japonicus* présentant un phénotype d'hypernodulation et une insensibilité à l'azote ont été étudiés en détail (Wopereis *et al.*, 2000). Ces mutants *har1* ont permis d'identifier un gène codant une LRR-RLK partageant un fort taux de similarité avec CLAVATA1 (CLV1) d'*A. thaliana* et qui est responsable de la communication longue distance entre la racine et la partie aérienne. L'activation du récepteur MtHAR1/LjSUNN grâce à la perception d'un peptide spécifique CLE, induit lors de la présence d'une forte concentration d'azote dans l'environnement ou après perception des facteurs Nod, va entraîner une régulation négative de la nodulation (Nishimura *et al.*, 2002; Okamoto *et al.*, 2013).

Le second mécanisme de régulation met en jeu le récepteur CRA2 (Compact Root Architecture 2) capable de percevoir des peptides CEP. Ces peptides sont produits dans la racine lors d'une carence en azote et sont perçus par CRA2 qui joue un rôle de régulateur positif de la nodulation et dans le développement des racines latérales par différentes voies, ce mécanisme est complémentaire au système de régulation impliquant SUNN/CLE (Mohd-Radzman *et al.*, 2016).

Le dernier mécanisme de régulation de la nodulation connu à ce jour repose sur la perception d'une phytohormone, l'éthylène. En effet les mutants *sickle* de *M. truncatula* présentent un phénotype d'hyper-infection, un retard dans l'organogénèse ainsi qu'une insensibilité à l'éthylène, cependant ils ne sont pas affectés dans la régulation systémique. Au niveau local, l'éthylène est un signal induisant l'arrêt de la progression de l'infection mais l'expression du gène SICKLE permet de contrer cet effet négatif (Penmetsa, 1997).

Modulation de l'immunité pendant la nodulation

Par ailleurs, l'immunité de la plante joue un rôle très important tout au long de l'interaction avec la bactérie dès son entrée dans le cytoplasme de la cellule hôte, qui devrait en principe induire des mécanismes de défense. Chez *M. truncatula*, une heure après l'inoculation de la bactérie de nombreux gènes de défenses sont induits mais leur expression retrouve un niveau basal environ 6h après inoculation (Lohar *et al.*, 2006). L'expression de gènes de défenses est réduite dans les cellules infectées chez *M. truncatula* par rapport à des cellules non-infectée, suggérant une modification du programme génétique de la cellule pour faciliter l'accommodation des bactéries (Limpens *et al.*, 2013). Plusieurs gènes ont été identifiés comme étant impliqués dans ce processus comme par exemple le gène RSD codant un facteur de transcription de type C2H2 zinc-finger qui pourrait réprimer l'expression du gène codant la protéine associée à la membrane vésiculaire 721a (Vamp721a). Chez les mutants *rsd*, les bactéroïdes entrent en sénescence et des réactions de défenses sont observées dans les nodules (Sinharoy *et al.*, 2013). Des mutants sur deux autres gènes, *DNF2* (Phospholipase-like protein) et *SymCRK* (Cystein-rich Receptor Kinase), montrent un phénotype similaire au mutant *rsd*, c'est-à-dire qu'ils présentent des nodules nécrotiques, expriment des gènes de défenses et accumulent des composés phénoliques. Ces gènes sont

donc également requis pour la persistance des rhizobiums dans le nodule (Bourcy *et al.*, 2013; Berrabah *et al.*, 2014).

Pour conclure, de nombreuses découvertes génétiques ont été réalisées sur les deux légumineuses modèles *M. truncatula* et *L. japonicus* ces 20 dernières années et on continue aujourd'hui encore à découvrir de nouveaux acteurs moléculaires. On se rend maintenant compte de la complexité que cache le mécanisme de nodulation en termes de nombre et de variété de gènes pour que l'interaction soit effective et durable. Un résultat fort de tous ces travaux est que de nombreux gènes impliqués dans la symbiose rhizobienne sont également partagés lors du processus d'endomycorhization avec les Glomeromycètes, qui rappelons-le, concerne plus de 80% des espèces de plantes vasculaires (Smith and Read, 2009). Le niveau de conservation des mécanismes moléculaires entre ces deux types de symbioses nourrit l'espoir que de nombreuses non-légumineuses pourraient posséder un background génétique favorable pour une tentative de transfert de la capacité à noduler. Dans la dynamique actuelle, des projets à gros financements visent à faciliter le transfert de la nodulation chez des non-légumineuses, ce qui permettrait de s'affranchir des engrais chimiques (Bill & Melinda Gates Foundation et Collaborative Crop Resilience Program). Dans cette optique, la communauté scientifique cherche également à comprendre comment a évolué la nodulation et à identifier des mécanismes symbiotiques alternatifs à ceux décrits chez les deux légumineuses modèles.

3. Unité et diversité des mécanismes de nodulation

Grâce à l'avènement des NGS (New Generation Sequencing) et des nouvelles connaissances acquises sur les légumineuses modèles, de nouveaux champs de recherche se sont développés ces dernières années. On cherche désormais à mieux comprendre comment la capacité fixatrice d'azote est apparue et comment elle s'est diversifiée au sein des différents clades capables de produire des nodules.

3.a - Les origines de la nodulation

Une origine unique commune aux légumineuses, *Parasponia* et les plantes actinorhiziennes

Les légumineuses ainsi que 9 autres familles de plantes sont capables de produire des nodules fixateurs d'azote. Pour rappel, ces plantes appartiennent au clade monophylétique des NFN (Nitrogen-Fixing root Nodule) et font partie de 4 ordres distincts (Rosales, Fabales, Fagales, Cucurbitales). Cependant seulement 10 des 28 familles de ce clade contiennent des espèces nodulantes dont certaines sont étudiées de façon approfondie comme *Casuarina glauca* s'associant avec des actinomycètes du genre *Frankia*, *Parasponia andersonii* qui s'associe avec des rhizobiums et évidemment plusieurs espèces de légumineuses.

Afin de mieux comprendre l'histoire évolutive de ces plantes, la communauté scientifique s'efforce de comprendre l'origine de la nodulation. Pour expliquer la disparité du trait « symbiose fixatrice d'azote » dans le clade NFN, deux hypothèses évolutives ont été émises grâce à des analyses phylogénétiques. Tout d'abord l'hypothèse d'une origine unique suivie de pertes massives du trait de nodulation qui ont eu lieu en parallèle dans les différents groupes de plantes. La seconde hypothèse, dite hypothèse de l'origine multiple, soutient que la capacité à noduler serait apparue plusieurs fois au cours de l'évolution de manière indépendante (au moins 5 fois) grâce à une prédisposition génétique ancestrale (Soltis *et al.*, 1995). Cette seconde hypothèse de gains parallèles de la nodulation a largement été acceptée par la communauté scientifique notamment grâce à l'analyse phylogénétique de différentes lignées de plantes actinorhiziennes (Swensen, 1996; Doyle, 1998). Cependant, les données moléculaires et génétiques de l'époque n'étaient pas encore suffisantes pour pouvoir trancher entre les deux hypothèses.

Des études de phylogénomique comparative s'appuyant sur les récentes avancées technologiques en matière de séquençage (NGS) qui ont considérablement réduit les coûts pour le séquençage de transcriptomes et de génomes, ont été récemment mis en œuvre pour dévoiler les mécanismes évolutifs ayant mené à la répartition hétérogène du trait de nodulation chez les différents groupes de plantes. Pour étudier l'apparition et l'histoire d'un trait phénotypique, il est nécessaire de définir un cadre phylogénétique précis pour réaliser des études de phylogénomique comparative. En effet, on a besoin de comparer des espèces

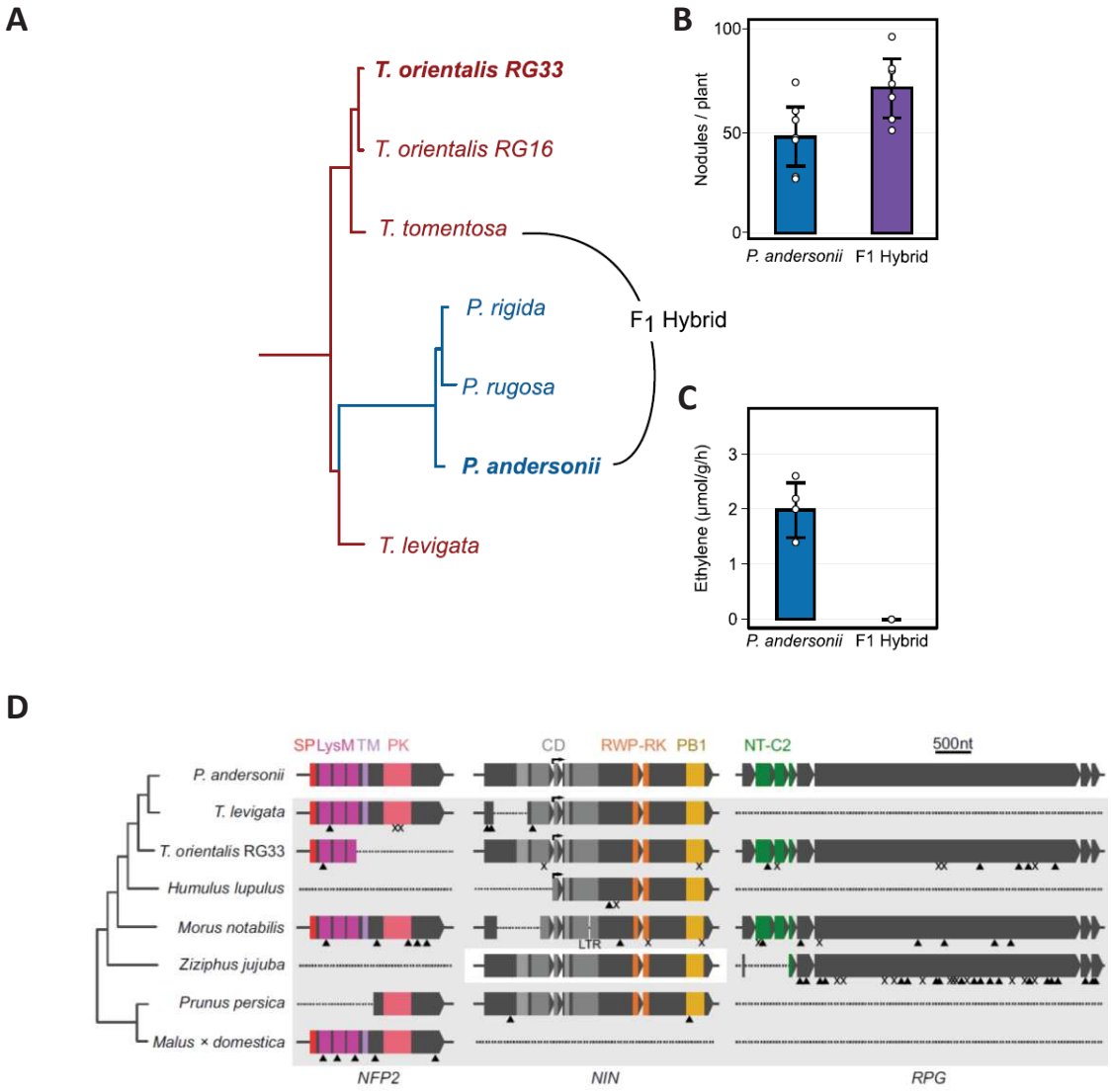


Figure 16 : Evolution de la nodulation chez *Parasponia*, la seule non-legumineuse à interagir symbioiquement avec des rhizobiums. A. Phylogénie des genres *Parasponia* (nodulant) et *Trema* (non nodulant). Les noms en gras correspondent aux espèces séquencées. Des tests d'hybridation interspécifique (*P. andersonii* x *T. tomentosa*) ont été réalisés pour donner des plantes hybrides F1. **B.** Moyenne du nombre de nodules par plante chez *P. andersonii* et les hybrides F1. **C.** Moyenne de l'activité nitrogénase (ARA) chez *P. andersonii* et les hybrides F1. **D.** Pseudogénisation ou perte de NFP2, NIN et RPG dans plusieurs lignées phylognénétiques indépendantes. Chez *H. lupulus* NIN est pseudogénisé alors que NFP2 et RPG ne sont pas retrouvés. Chez *Z. jujuba* NFP2 est perdu et RPG est pseudogénisé, cependant NIN est intact. Les triangles indiquent des changements de phase du cadre de lecture, les croix indiquent des codons stop et LTR indique un rétrotransposon. (D'après Van Velzen et al, 2018)

avec le plus faible degré de divergence possible mais présentant des propriétés contrastées afin d'identitifer les gènes différentiellement présents et potentiellement responsables du trait phénotypique étudié. Cette approche de phylogénomique comparative a d'abord été appliquée à l'étude de l'endomycorhization en comparant les données transcriptomiques d'espèces capables ou non de mycorhizer. Ceci a permis d'identifier 174 gènes symbiotiques impliqués dans l'endomycorhization et qui ont été spécifiquement perdus chez les espèces incapables de réaliser ce type de symbiose (*Brassicaceae*). Cela suggère aussi que les symbiontes ont maintenu une pression de sélection sur le réseau de gènes de leur hôte pendant l'évolution de lignées entières. Avec cette méthode, certains gènes connus chez les légumineuses modèles ont été identifiés et la fonction des nouveaux gènes découverts a été validée par des expériences de génétique inverse (Delaux *et al.*, 2014; Bravo *et al.*, 2016). Des approches similaires ont ensuite été employées pour identifier le gain et la perte de gènes liés à l'évolution de la symbiose fixatrice d'azote dans le clade NFN en comparant des espèces proches, capables ou non de noduler. Elles ont bénéficié du séquençage de génome de nombreuses espèces, ce qui rend ce type d'approche plus puissant car permettant de faire des analyses de synténie pour suivre l'évolution des gènes symbiotiques.

La volonté de comprendre l'origine de la nodulation chez *Parasponia andersonii*, la seule lignée de plante connue capable de noduler avec des rhizobiums en dehors des légumineuses, a conduit à l'étude intensive de ce modèle appartenant à l'ordre des Rosales (Figure 16A). Le genre *Parasponia* contient 5 espèces qui forment un clade niché au sein du genre *Trema* et qui ne contient que des espèces incapables de produire des nodules (Behm *et al.*, 2014). Le séquençage du génome de *Parasponia andersonii* couplé à des données RNAseq ont révélé, d'une part, que *P. andersonii* possède les orthologues putatifs de la plupart des gènes symbiotiques caractérisés comme étant impliqués dans la symbiose rhizobium-légumineuse (96 sur 126 analysés), d'autre part, que *P. andersonii* partage un set de 290 gènes symbiotiques avec *M. truncatula*, dont l'expression est intensifiée dans les nodules de ces deux espèces. Des analyses fonctionnelles par génétique inverse ont également montré que plusieurs gènes impliqués dans l'infection, la signalisation et l'organogénèse chez les légumineuses (CRE1/LHK4, NSP1, NSP2, NIN et NF-YA1) sont nécessaires à la mise en place de la symbiose chez *Parasponia* (van Zeijl *et al.*, 2018; Bu *et al.*, 2020). Les analyses de génomique comparative, entre trois espèces de *Parasponia* dont

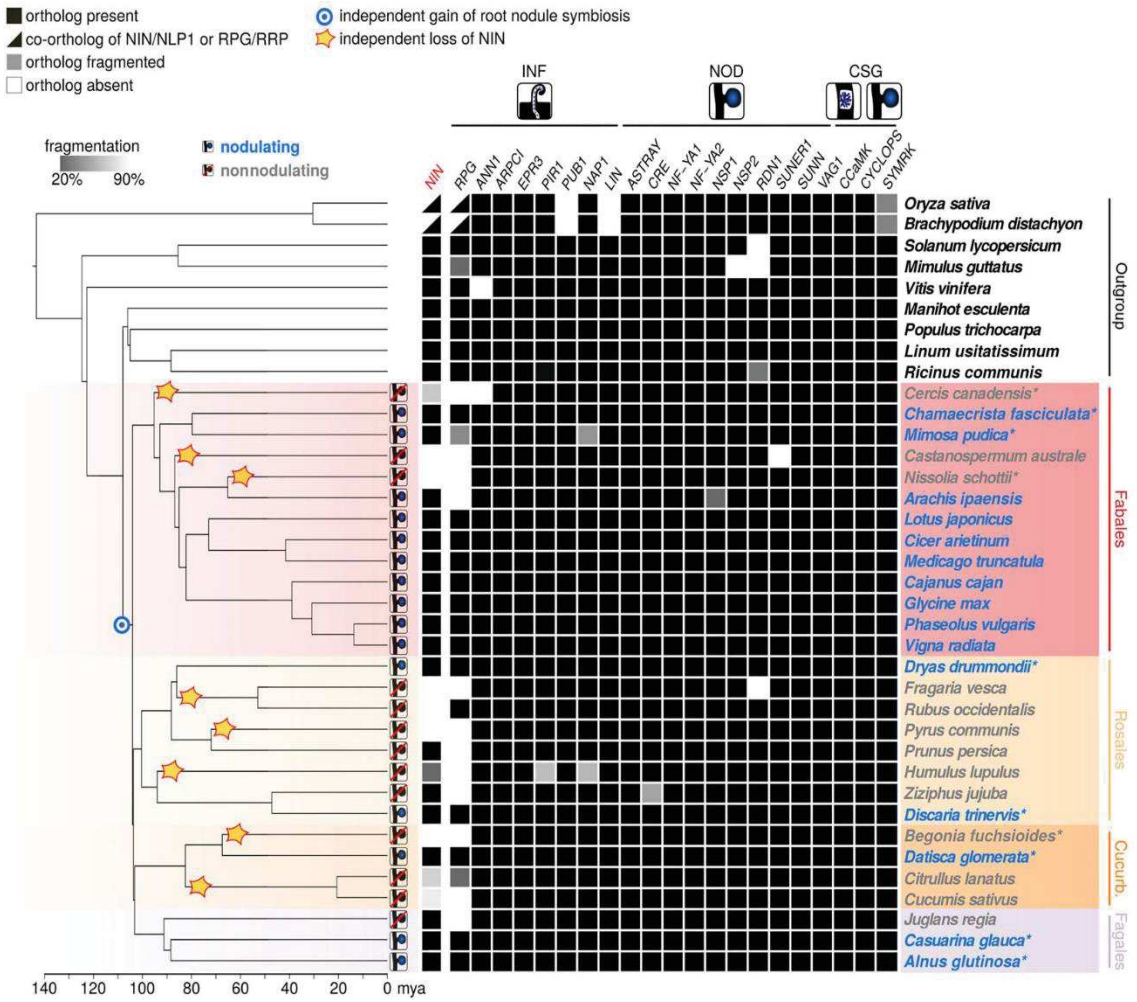
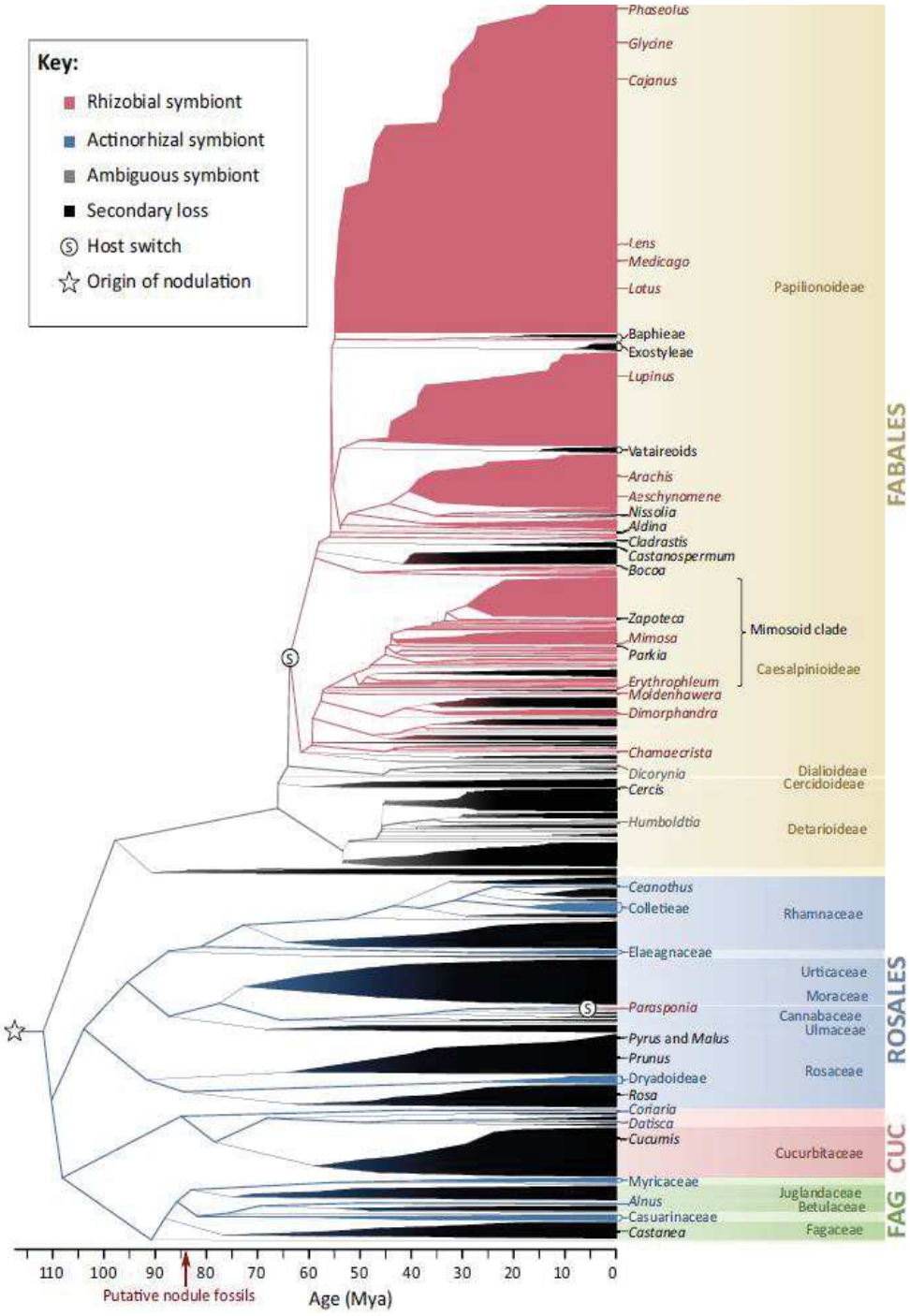


Figure 17 : Profils phylogénétiques des gènes de la symbiose NFN (détailler NFN). Le chronogramme contient des espèces nodulantes (bleus) et non-nodulantes (gris) appartenant aux 4 ordres du clade NFN. Le groupe externe est constitué de 9 espèces hors du clade NFN. La présence et l'absence de 21 gènes est illustrée par les carrés noir, blanc et gris= forme tronquée. Les étoiles indiquent les pertes indépendantes de NIN. INF= gène nécessaires à l'infection, NOD= gènes impliqués dans l'organogénèse nodulaire et CSG= gènes de la voie de signalisation commune à la mycorhization et la nodulation. (D'après Griesmann et al, 2018)

P. andersonii et trois espèces de *Trema* dont *T. tomentosa* qui sont incapables de produire des nodules (Fig. 16A), n'ont pas permis de révéler des variations du nombre de copie de certains gènes qui pourraient expliquer un gain de la nodulation chez *Parasponia*. Au contraire de ce qui était attendu, elles ont révélé des pertes ou pseudogénisations parallèles des orthologues de 7 gènes symbiotiques dont *NIN*, *RPG* et *NFP* chez les Rosales indiquant que c'est l'inactivation spécifique de certains gènes symbiotiques qui a conduit à la perte de la capacité à noduler chez *Trema* (van Velzen *et al.*, 2017) (Fig. 16B).

Ce résultat phare est étayé par un deuxième travail de recherche basé sur une analyse comparative plus large de 37 espèces nodulantes et non-nodulantes réparties dans tous les genres du clade NFN. La recherche de gènes présents chez les espèces nodulantes et absents chez les espèces non-nodulantes, et les analyses de micro-synténie ont révélé la conservation du bloc de synténie contenant *NIN* dans tout le clade NFN et des signatures d'évènements multiples de perte de fonction pour ce régulateur symbiotique clé, ce gène étant absent ou en cours de pseudogénisation chez 10 des 13 espèces non-nodulantes analysées du clade NFN. Ces pertes indépendantes de *NIN* sont observées dans 3 lignées de la famille des légumineuses et 5 chez les non-légumineuses. Elles sont corrélées à l'absence de nodulation, elles semblent donc suffisantes pour expliquer la perte de capacité de la nodulation dans ces lignées. Cependant, le fait que NIN ait été retrouvé chez 3 espèces non-nodulantes suggère que la perte d'autres gènes peut également provoquer la perte de la nodulation. Les analyses comparatives menées sur 21 autres gènes de nodulation, ont montré que de la même manière que *NIN*, c'est le gène *RPG*, impliqué dans la croissance du cordon d'infection, qui a été perdu chez les 3 espèces non-nodulantes qui possèdent encore *NIN*. Ces résultats corroborent et élargissent ceux obtenus avec le système *Parasponia-Trema*, et plaident en faveur d'un gain unique de la capacité à noduler suivi de multiples pertes au sein des 4 ordres du clade NFN (Griesmann *et al.*, 2018) (Fig. 17).

Tous ces résultats appuient donc l'hypothèse que l'ancêtre commun du clade NFN daté d'environ 110 millions d'années était capable de former une symbiose fixatrice d'azote et que de nombreuses pertes de ce trait sont ensuite apparue en parallèle dans différents clades au cours de l'évolution. La perte de ce trait symbiotique peut-être expliquer par le coût énergétique très élevé de la fixation biologique de l'azote dont la plante n'a besoin que



Figur 18 : Représentation de la distribution phylogénétique de plantes nodulantes et non-nodulantes du clade NFN sous l'hypothèse d'un gain unique suivi de multiples pertes de la nodulation. La couleur des branches représente le statut de l'hôte et l'identité du microsymbionte. (D'après van Velzen et al, 2019)

lorsqu'elle est dans un environnement pauvre en azote assimilable (van Velzen *et al.*, 2019) (Fig. 18)**.**

Cette hypothèse est aussi cohérente avec l'observation que les plantes actinorhiziennes, tout comme montré pour *Parasponia* partagent de nombreux gènes de nodulation avec les légumineuses. Tout d'abord, les transcrits de nombreux orthologues des gènes symbiotiques ont été retrouvés dans les transcriptomes de *Casuarina glauca* et *Alnus glutinosa* comme *SYMREM1, RPG, SYMRK, HMGR1, CASTOR, CCaMK, CYCLOPS, NSP1, ERF1, ERN1* et *NIN* qui sont exprimés dans la racine et le nodule (Hocher *et al.*, 2011). De plus, le rôle dans la nodulation de certains gènes comme *SYMRK, CCaMK* et *NIN* a été démontré par des expériences d'extinction d'expression par RNAi (Gherbi *et al.*, 2008a; Svistoonoff *et al.*, 2013; Clavijo *et al.*, 2015), suggérant que la voie de signalisation symbiotique fonctionne de façon comparable pendant la nodulation chez les plantes actinorhiziennes et les légumineuses, cette voie étant aussi en partie commune à la symbiose endomycorhizienne. Même si l'on sait que la CSSP est conservée entre les différents types de nodulation (clade NFN) et la mycorhization, on ne connait pas encore les mécanismes adaptatifs des récepteurs ayant permis le changement de symbionte qui continue malgré tout à activer cette voie de signalisation.

Plusieurs mystères demeurent donc encore autour de l'évolution de la symbiose fixatrice d'azote. Le signal produit par les *Frankia* permettant la reconnaissance mutuelle des partenaires n'a pas encore été découvert mais il semble être d'une nature différente du facteur Nod des rhizobiums. En effet, la découverte d'un facteur de faible masse moléculaire résistant aux chitinases, capable d'induire des oscillations calcique et l'expression du gène CgNIN en réponse au surnageant de *Frankia* permet de dire qu'il existe un facteur bactérien sécrété de nature différente des facteurs Nod et qui induit la nodulation chez *C. glauca* (Chabaud *et al.*, 2016; Cissoko *et al.*, 2018). De plus, l'implication des gènes de la famille des LysM-RLK dans la perception du signal bactérien chez les plantes actinorhizienne n'a pas encore été montrée. Cela suggère que malgré la conservation partielle des mécanismes de signalisation, certaines étapes comme la perception du signal divergent entre ces différents types de symbioses.

Cependant, une étude récente montre que 4 LysM-RLKs contrôlent la nodulation chez *Parasponia* (LYK1, LYK3, NFP1 et NFP2) et que ces gènes ont évolués à partir d'anciens évènement de duplications pré-datant l'origine de la nodulation. La présence de LysM-RLKs orthologues chez les légumineuses et *Parasponia* indique une origine évolutive commune de la nodulation dépendante des facteurs Nod. Aussi, PanLYK1 et PanLYK3 sont essentiels pour la formation d'endomycorhizes chez cette plante (Rutten *et al.*, 2020).

Suite à ces avancées importantes, l'hypothèse a été émise sur le partenaire bactérien ancestral de la symbiose fixatrice d'azote qui aurait d'abord évolué à partir d'espèces de *Frankia* produisant des LCOs. Par la suite, différentes protéobactéries auraient acquis les gènes de biosynthèse des LCOs par transfert horizontal de gènes leur permettant ainsi d'entrer en compétition avec les *Frankia* pour interagir avec les ancêtres des légumineuses et de *Parasponia* (Remigi *et al.*, 2016). C'est le système de protection vis-à-vis de l'oxygène de *Frankia* qui appuie cette hypothèse, car ces actinomycètes possèdent leur propre système de protection vis-à-vis de l'oxygène alors que les rhizobiums dépendent de la leghémoglobine produite par leur hôte, suggérant que les *Frankia* étaient les premiers partenaires des plantes capables de produire des nodules. Puis il y aurait eu plusieurs changements de partenaires bactériens pendant l'évolution du clade NFN : changement de *Frankia* à rhizobium une fois chez l'ancêtre des légumineuses et une fois chez l'ancêtre de *Parasponia* (van Velzen *et al.*, 2018).

Innovations liés à la capacité à développer des nodules fixateurs d'azote

Les processus de signalisation symbiotique et d'infection semblent être en grande partie empruntés à la symbiose endomycorhizienne avec les Gloméromycètes (cf partie II). Cependant l'aptitude à développer des nodules nécessite des gènes spécifiquement impliqués dans le développement de ce nouvel organe qui n'est pas induit lors de la symbiose endomycorhizienne et une adaptation vis-à-vis de l'azote.

Un lien fort est observé entre l'organogénèse du nodule et les cytokinines (CK) qui permettent d'en initier la formation. Une version auto-active du gène capable de percevoir ces CKs (*LjLHK1*) induit une organogénèse nodulaire spontanée, alors qu'a l'inverse le mutant perte de fonction *lhk1* est capable d'initier l'infection mais est fortement impacté dans l'organogénèse (Murray *et al.*, 2007; Tirichine *et al.*, 2007). Une étude comparative a

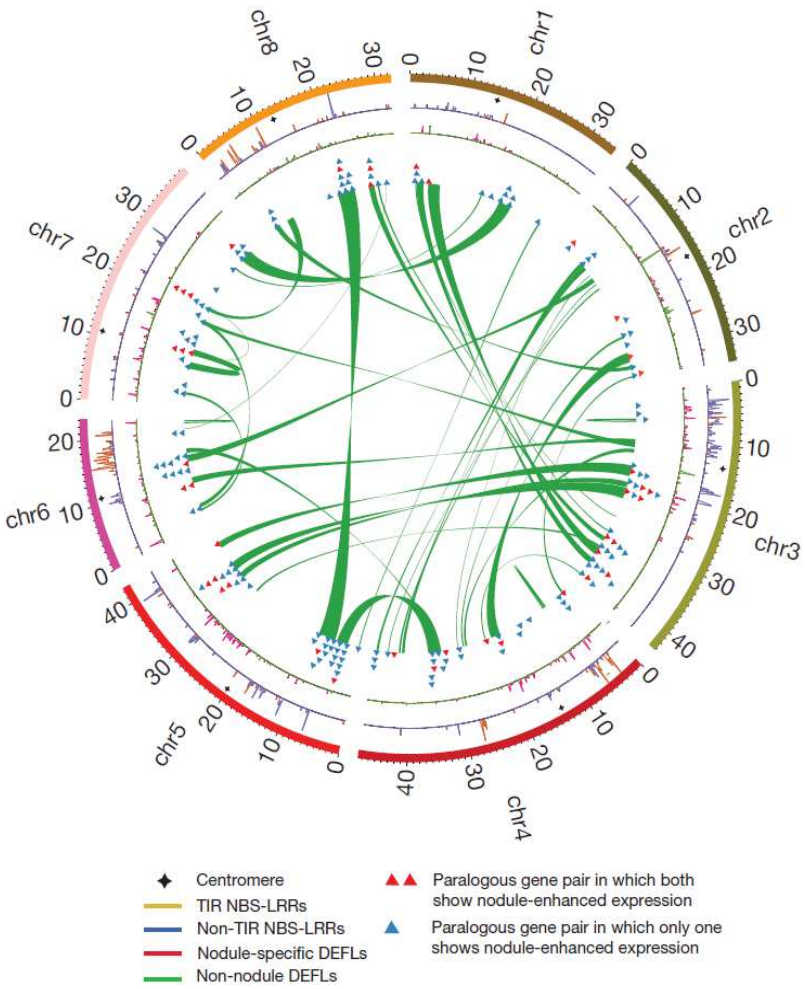
montré que les légumineuses capables de noduler avaient une sensibilité particulière aux CKs au niveau du cortex racinaire qui n'est pas retrouvée chez les légumineuses incapables de noduler et les non-légumineuses (dont des plantes actinorhiziennes). Chez la majorité des plantes l'application de CKs inhibe la croissance racinaire, mais chez les légumineuses nodulantes le développement de pseudonodules indiquent qu'il existe une réponse corticale spécifique et unique en réponse aux CKs. Cela suggère que la capacité à formé des nodules dérivés de cellules corticales a été gagnée chez les légumineuses nodulantes et probablement perdue chez les légumineuses non-nodulantes (Gauthier-Coles *et al.*, 2019).

Les CKs sont également impliqués dans des programmes développementaux comme la formation des racines latérales, la régulation de la taille de la racine et dans la morphogénèse du système vasculaire. Comme décrit en partie 2, les programmes génétiques de développement des nodules et des racines latérales sont intriqués, la formation du nodule met en jeu le gène NIN inductible par les CKs, qui active certains composants du programme de formation des racines latérales pour l'organogénèse nodulaire (Schiessl *et al.*, 2019).

L'évolution de l'organogénèse nodulaire a dû impliquer des programmes de développement déjà existants comme le montre la présence de structures nodulaires ou « nodule-like » chez plusieurs lignées de plantes terrestres, indiquant que les outils génétiques nécessaires à l'organogénèse nodulaire ne sont pas uniquement présents dans le clade NFN (pour revue Markmann and Parniske, 2009). Par ailleurs, les nodules des légumineuses et des plantes actinorhiziennes sont considérés comme fondamentalement différents mais une étude récente sur ces différents types de nodules a permis d'observer que l'ontogénie des nodules de type actinorhizien formés par *P. andersonii* et *Alnus glutinosa* étaient plus proches qu'on ne le pensait de l'ontogénie des nodules de type légumineuse. De plus, la découverte d'une mutation homéotique dans le gène MtNOOT1 qui convertit des nodules de type légumineuse en nodules de type actinorhizien, suggère que ces deux types de nodules partagent une origine évolutive commune et appuie l'hypothèse que la nodulation a évoluée une seule fois chez l'ancêtre commun du clade NFN (Shen *et al.*, 2020).

Plusieurs évènements génétiques majeurs ont participés à l'émergence de la capacité fixatrice d'azote. Le facteur de transcription NIN fait partie de la famille NLP (NIN-Like

A.



B.

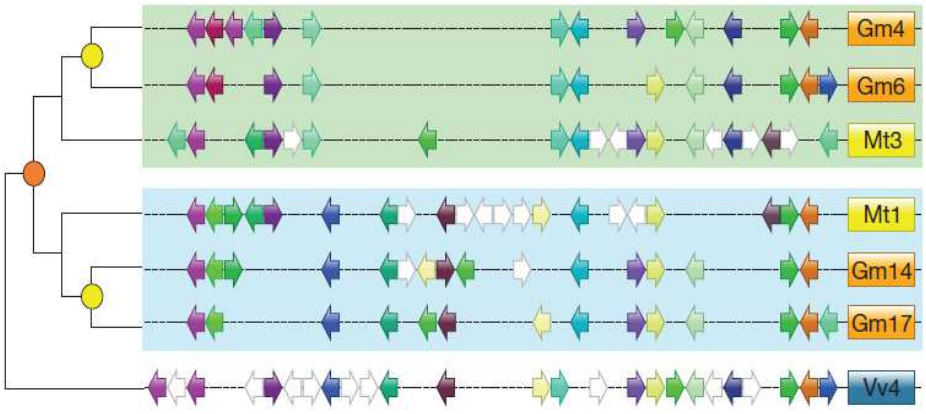


Figure 19 : Duplication ancestrale du génome des Papillonacées datée de 58 millions d'années. A. Représenation Circos du génome de *Medicago* illustrant la "Whole Genome Duplication" (WGD) et les familles de gènes concernés chez *Medicago*. Les flèches en bleu montrent des paires de gènes paralogues dont un seul a une expression forte dans le nodule et les flèches en rouge des paires de gènes paralogues étant tous deux fortement exprimés dans le nodule. **B.** Microsynténie entre les gènes de *M. truncatula* (Mt) et les régions correspondantes chez le soja *G. max* (Gm). Le noeud orange correspond à la WGD de 58 MA. Les noeuds jaunes correspondent à une WGD spécifique à *G. max* datée d'environ 13 MA. (D'après Young et al, 2011)

Protein) qui est une famille connue pour être impliquée dans la réponse aux nitrates chez *Arabidopsis*. Cependant, les gènes *LjNLP1* et plusieurs *AtNLP*s sont activés en réponse au statut azoté de la plante alors que le gène *LjNIN* a perdu la capacité de s'activer dans ces conditions mais il peut être activé par l'application exogène de CKs (Heckmann *et al.*, 2011; Suzuki *et al.*, 2013). La divergence avec les NLPs ancestraux réside dans deux changement majeurs qui sont la régulation de son expression par le partenaire microbien via la voie de signalisation Nod et sa localisation nucléaire constitutive. Cela représenterait la naissance d'une étape clé dans le processus de nodulation qui a été suivi plus tard par d'autres adaptations, comme la régulation de son expression par les CKs au sein de la famille des légumineuses. Il reste encore beaucoup de choses à découvrir sur ce facteur de transcription dont l'expression spatiotemporelle et les fonctionnalités multiples compliquent la compréhension de son fonctionnement globale (Liu and Bisseling, 2020)

3.b - Diversification des mécanismes de nodulation chez les légumineuses

Il est maintenant admis que la nodulation a précédé de plusieurs millions d'années l'apparition des légumineuses et de nombreux éléments indiquent que cette capacité aurait été acquise par l'ancêtre commun le plus récent du clade NFN. Cependant on observe une grande disparité de ce caractère au sein des différentes lignées évolutives, même chez les légumineuses qui regroupent un grand nombre d'espèces par rapport à d'autres familles de plantes, la capacité de noduler n'est pas partagée par toutes les espèces. La sous-famille des *Faboideae* contient beaucoup d'espèces qui sont pour leur majorité capables de produire des nodules fixateurs d'azote et chez qui on observe une grande diversité dans les processus de nodulation. Les chercheurs se sont alors demandé comment pouvait s'expliquer une radiation évolutive aussi importante dans ce clade.

La découverte de traces d'un évènement de Whole-Genome Duplication (WGD) à la racine du clade des *Faboideae* a permis de mieux comprendre la dynamique évolutive de la symbiose rhizobienne. En effet, la comparaison des paternes de duplication chez plusieurs légumineuses montrent qu'une WGD a eu lieu il y a environ 58 MA chez les *Faboideae* (Cannon *et al.*, 2015). La WGD participe à la dynamique des génomes en créant de la redondance génétique qui va potentiellement faciliter l'émergence de nouvelles fonctions de gènes sans compromettre celles qui existent déjà. Les mutations s'accumulent sur une des deux copies du gène, ce qui peut conduire à de la néo/sub-fonctionnalisation ou de la

pseudogénisation, alors que l'autre copie conserve la plupart du temps sa fonction (Freeling, 2006). La contribution de cette WGD sur la nodulation a été étudiée de façon approfondie chez *M. truncatula*. L'analyse génomique a permis de trouver un total de 963 paires de gènes dérivés de la WGD. Parmi les gènes ayant une expression nodulaire, 8 paires ont les deux copies exprimées dans le nodule, alors que 43 paires de gènes n'ont qu'une seule des copies exprimée dans le nodule (Fig. 19). Ces données sont cohérentes avec une origine de la nodulation précédant la WGD, suivie de sub/néo-fonctionnalisation potentielle des gènes dupliqués. L'intense radiation évolutive observée chez les légumineuses est donc en partie attribuée à la WGD de 58 MA (Young *et al.*, 2011).

Il existe plusieurs exemples de gènes symbiotiques aujourd'hui bien connus qui sont concernés par des évènements de duplication et dont la fonction a soit divergée au cours du temps soit été conservée, dans les deux cas les profils d'expression sont souvent contrastés. Chez *M. truncatula* le gène CBS1 et son paralogue CBS2 possèdent une fonction et des profils d'expressions différents. En effet, *MtCBS1* est exprimé principalement dans les nodules et dans le cordon d'infection alors que *MtCBS2* n'est exprimé que dans les racines mycorhizées (Sinharoy *et al.*, 2016). Les gènes *MtNFP* et *MtERN1*, qui sont des paralogues de *MtLYR1* et *MtERN2* exprimés pendant la mycorhization, ont par la suite évolué vers une fonction spécifique lors de la nodulation (Young *et al.*, 2011). Les gènes *MtERN1* et le paralogue *MtERN2* ont une fonction qui semble partiellement redondantes car il est possible de complémenter le phénotype (formation de nodules non-infectés) du mutant *ern1* en exprimant ERN2 sous le contrôle du promoteur d'ERN1 (Cerri *et al.*, 2016). Des LRR-RLKs, RINRK1 et son paralogue RINRK2, sont impliqués dans le processus d'infection rhizobienne mais RINRK2 aurait une fonction mineure et redondante dans l'infection (Li *et al.*, 2020). Pour finir, les facteurs de transcription NOOT1 et le paralogue NOOT2, sont impliqués dans le maintien de l'identité nodulaire. Mais ces gènes possèdent des profils d'expression différents, *MtNOOT1* est exprimé dans la racine primaire et dans le nodule alors que *MtNOOT2* n'est exprimé que lorsque le primordium nodulaire est induit (Couzigou *et al.*, 2012; Magne *et al.*, 2018).

En revanche chez différentes lignées de légumineuses, on observe qu'il existe plusieurs schémas de rétentions et de pertes de paralogues. Par exemple, à l'inverse de ce qui est observé avec *M. truncatula*, chez *L. japonicus* on ne retrouve qu'une seule copie d'ERN1. En effet, chez *L. japonicus* le mutant *ern1* présente le même phénotype que le double

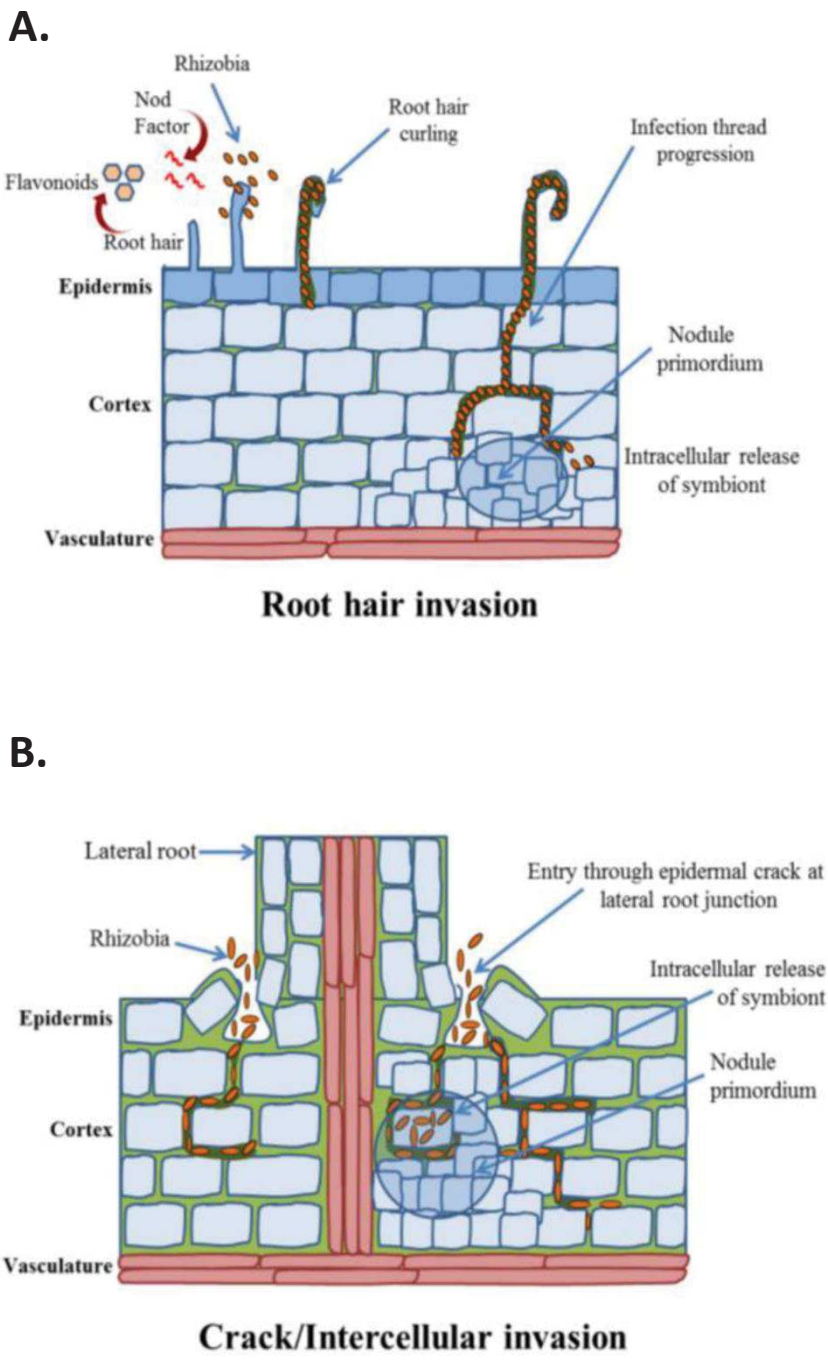


Figure 20 : Les deux grands types d'infection rencontrés chez les Légumineuses. L'infection rhizobienne est initiée au niveau du poil absorbant ou de fissures dans l'épiderme. **A.** L'infection du poil absorbant est initié après adhésion des bactéries à l'apex du poil et perception des facteurs Nod. **B.** L'infection intercellulaire débute par la progression des bactéries dans des brèches de l'épiderme leur permettant un accès direct aux cellules corticales plus profondes. (D'après Sharma et al., 2020)

mutant *ern1/ern2* chez *M. truncatula*, montrant qu'une seule copie est suffisante chez *L. japonicus* pour permettre l'infection. Les analyses moléculaires sur les ratios de mutations non-synonymes/synonymes suggèrent que la séquence de ERN1 est soumise à une pression de sélection forte chez les légumineuses alors que cette pression est beaucoup moins importante pour ERN2 (Yano *et al.*, 2016). D'autre part, le gène *MtSICKLE*, impliqué dans l'autorégulation de la nodulation (AON), est présent en deux copies chez *L. japonicus* et ces 2 gènes ont une fonction redondante (EIN2-1 et EIN2-2). La surexpression de la partie C-terminale de la protéine EIN2 dans un fond mutant *ein2* induit une activation constitutive des réponses à l'éthylène. Les analyses de séquences indiquent que LjEIN2-1/LjEIN2-2 ont divergés depuis longtemps et que l'évolution des récepteurs à l'éthylène chez les légumineuses est plus rapide par rapport aux autres plantes. Il semblerait que l'acquisition de la symbiose rhizobienne est accélérée l'évolution des composants de la signalisation éthylène chez les légumineuses pour leur permettre de contrôler leur nouveau partenaire symbiotique (Miyata *et al.*, 2013). Finalement le LysM-RLK *MtLYR1* dont la duplication a donné naissance à *MtNFP,* a été maintenu chez *M. truncatula* mais aucun rôle symbiotique n'a pu être démontré jusqu'à maintenant, bien que son orthologue *LjLYS11* soit induit lors de la mycorhization chez *L. japonicus*. Chez *M. truncatula* la séquence de *MtLYR1* diverge beaucoup de celle de *MtNFP* et suggère un processus de pseudogénisation (Gough *et al.*, 2018; Rasmussen *et al.*, 2016).

La WGD a la base du clade des Fabaceae a donc permis d'augmenter le contenu en gènes symbiotiques chez cette famille et l'évolution des gènes dupliqués aurait par la suite permis de diversifier les processus symbiotiques déjà existants. Plusieurs nouveaux traits seraient apparus comme le caractère déterminé/indéterminé des nodules, la présence ou l'absence de différenciation terminale en bactéroïdes ou encore des modes d'infection alternatifs. Même si le type de nodule et la différenciation terminale sont illustrées par les deux légumineuses modèles, ces caractères sont assez variables et dispersés dans la phylogénie des légumineuses (Oono *et al.*, 2010).

3.c - Les processus d'infection symbiotique alternatifs chez les légumineuses

Le processus de nodulation chez les 2 légumineuses modèles a été longuement étudié mais elles possèdent le même mode d'infection symbiotique intracellulaire alors qu'il existe une diversité bien plus importante (Fig. 20) qui reste encore peu étudiée dans la famille des Fabaceae (Sprent, 2008).

Existence de processus d'infection intercellulaire « constitutifs »

Comme décrit précédemment, le processus d'infection chez les légumineuses modèles met en jeu la formation d'un cordon d'infection au niveau d'un poil absorbant qui va guider les bactéries jusqu'aux couches de cellules internes de la racine. Cependant, ce processus n'est pas représentatif de l'ensemble des mécanismes d'infection rencontrés dans cette famille car chez de nombreuses légumineuses tropicales et méditerranéennes (environ 25% des genres de légumineuses), comme c'est le cas pour le Lupin (*Lupinus angustifolius*) et l'arachide (*Arachis hypogeae*) qui appartiennent aux clades des Génistoïdes et Dalbergioïdes respectivement, le processus d'infection symbiotique ne fait pas intervenir la formation de cordon d'infection mais se déroule de manière intercellulaire. Même si plusieurs mécanismes d'infection intercellulaire existent, de manière générale le processus se déroule en 3 grandes étapes qui sont i) la traversée de la barrière épidermique racinaire entre deux cellules ou via une fissure ii) la colonisation de la zone corticale en progressant dans l'apoplasme iii) la pénétration des bactéries dans la cellule hôte. Il est probable que différents mécanismes visant à détendre les mailles de la matrice extracellulaire soit mis en place par le symbionte et/ou l'hôte, via notamment lasécrétion d'enzymes, pour faciliter la progression des bactéries comme cela a été montré chez les légumineuses modèles (partie II). Après la colonisation de la surface racinaire et la pénétration des bactéries de façon intercellulaire, elles entrent en contact avec les cellules du cortex et un mécanisme proche de l'endocytose se met en place dans la cellule végétale. Les cellules infectées subissent ensuite des divisions successives à l'origine du nodule (Muñoz *et al.*, 1998; Robledo *et al.*, 2008).

Pour citer des légumineuses utilisant un processus d'infection intercellulaire qui ont été bien décrits dans la littérature, on peut d'abord parler du lupin blanc (*Lupinus albus*) qui utilise un mode d'infection intercellulaire. Lorsque les bactéries arrivent sur la surface de la racine elles s'accumulent au niveau de la jonction entre une cellule de poil absorbant et une cellule épidermique. Aucun cordon d'infection ou courbure de poil absorbant n'a pu être observé, les bactéries progressent simplement entre ces deux cellules pour atteindre la cellule du cortex située juste sous le poil absorbant. L'infection de cette première cellule est concomitante avec l'induction de divisions cellulaires dans le cortex racinaire externe

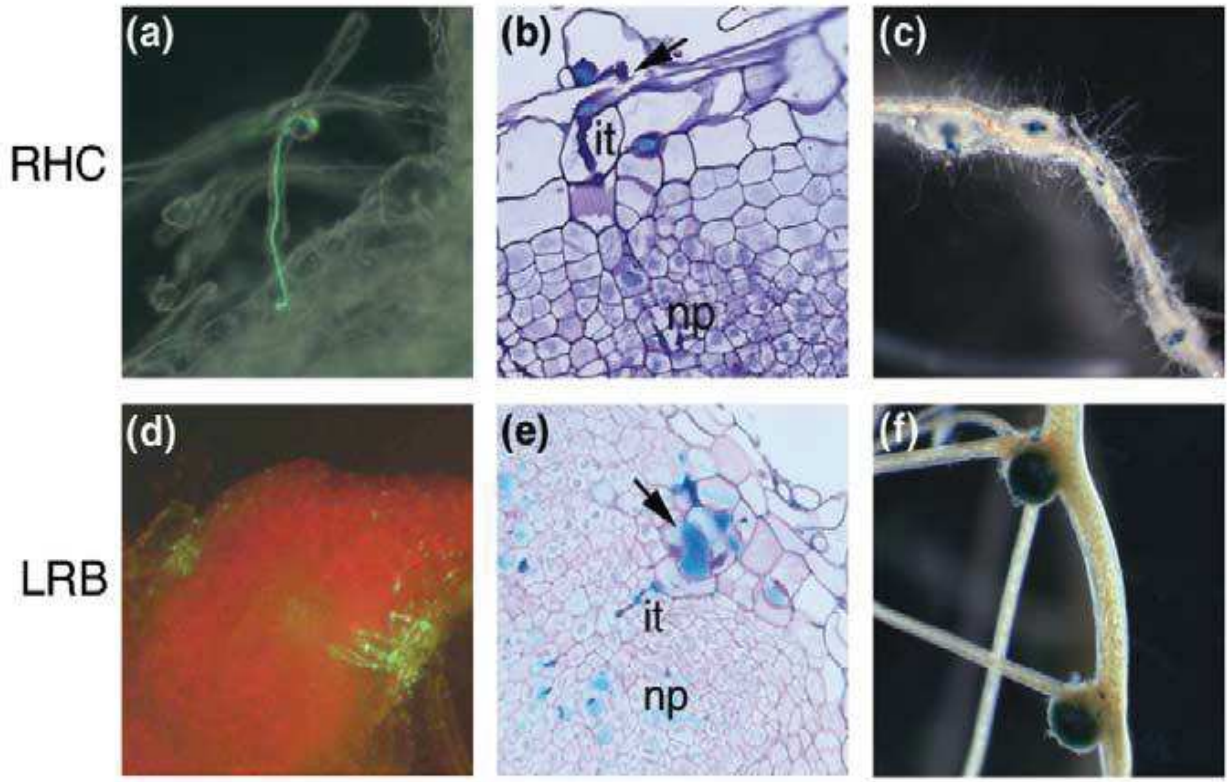


Figure 21 : Deux modes d'infection possibles chez la légumineuses *Sesbania rostrata* avec son symbionte *Azorhizobium caulinodans*. RHC (a-c) : par courbure des poils racinaire et formation de cordons d'infection en condition aéroponique. LRB (d-f) : par infection intercellulaire à la base des racines latérales puis formation de cordons d'infection, en condition hydroponique. Les bactéries sont marquées GFP (a et d) et GUS (b,c,e,f). It : infection thread, np : nodule primordium. (D'après Goormachtig et al, 2004)

sous le poil absorbant. Douze jours après inoculation de la bactérie on peut observer les premiers nodules (Gonzalez-Sama *et al.*, 2004).

Le processus d'infection chez l'arachide est très similaire à celui du lupin à la différence que l'arachide possède une structure de couronne de poils axillaire au niveau de l'émergence des racines latérales. Les *Bradyrhizobium* colonisent la surface racinaire de façon plus intense au niveau des couronnes de poils axillaires d'où ils pénètrent de manière intercellulaire via la lamelle moyenne entre deux cellules de poils axillaires, où se créèent de petites fissures. Par la suite ils sont endocytés dans une cellule corticale dont la paroi est structurellement altérée, cette cellule infectée se divise et donne naissance à un tissu uniformément infecté appelé « nodule aeschynomenoïde » (Boogerd and van Rossum, 1997; Uheda *et al.*, 2001). La structure « couronne de poils axillaires » est présente chez plusieurs Dalbergioïdes comme *Arachis*, *Stylosanthes* et *Aeschynomene*, la sécrétion de mucilage et de différents composés ainsi que les fissures naturelles provoquées par l'émergence de la racine latérale font de cette structure une interface privilégiée pour l'interaction bactérie-plante. La présence des poils axillaires dans deux clades distincts peut suggérer une convergence évolutive, soulignant un trait probablement important pour l'écologie de ces plantes. Le rôle exact des couronnes de poils axillaire n'a pas été démontré à ce jour mais on sait que cette structure est intimement liée à la nodulation. En effet, certaines lignées d'arachide (*A. hypogaea*) obtenues suite à des croisements entre cultivars sont dépourvues de couronnes de poils axillaires et sont également incapables de noduler, et de nombreux gènes étaient dérégulés chez ces lignées (Nambiar *et al.*, 1983; Peng *et al.*, 2017). D'autre part chez *A. hypogeae*, des analyses transcriptomiques comparant des espèces utilisant l'infection intercellulaire à des espèces utilisant l'infection intracellulaire (modèles) ont révélées la présence de la majorité des gènes symbiotiques connus comme les récepteurs aux facteurs Nod et les gènes de la voie de signalisation Nod. Plusieurs gènes de cette voie ont été testés par RNAi comme CCaMK, CYCLOPS et LHK1, permettant de confirmer leur implication dans cette symbiose utilisant un processus d'infection intercellulaire (Sinharoy and DasGupta, 2009; Das *et al.*, 2019; Kundu and DasGupta, 2018). Cependant plusieurs différences ont été observées comme l'absence des transcrits de SymCRK, ROP6, RR9 et SEN1 ainsi que l'absence dans le génome des gènes RPG et DNF2 (Karmakar *et al.*, 2019). Plusieurs modes d'infection co-existent donc à

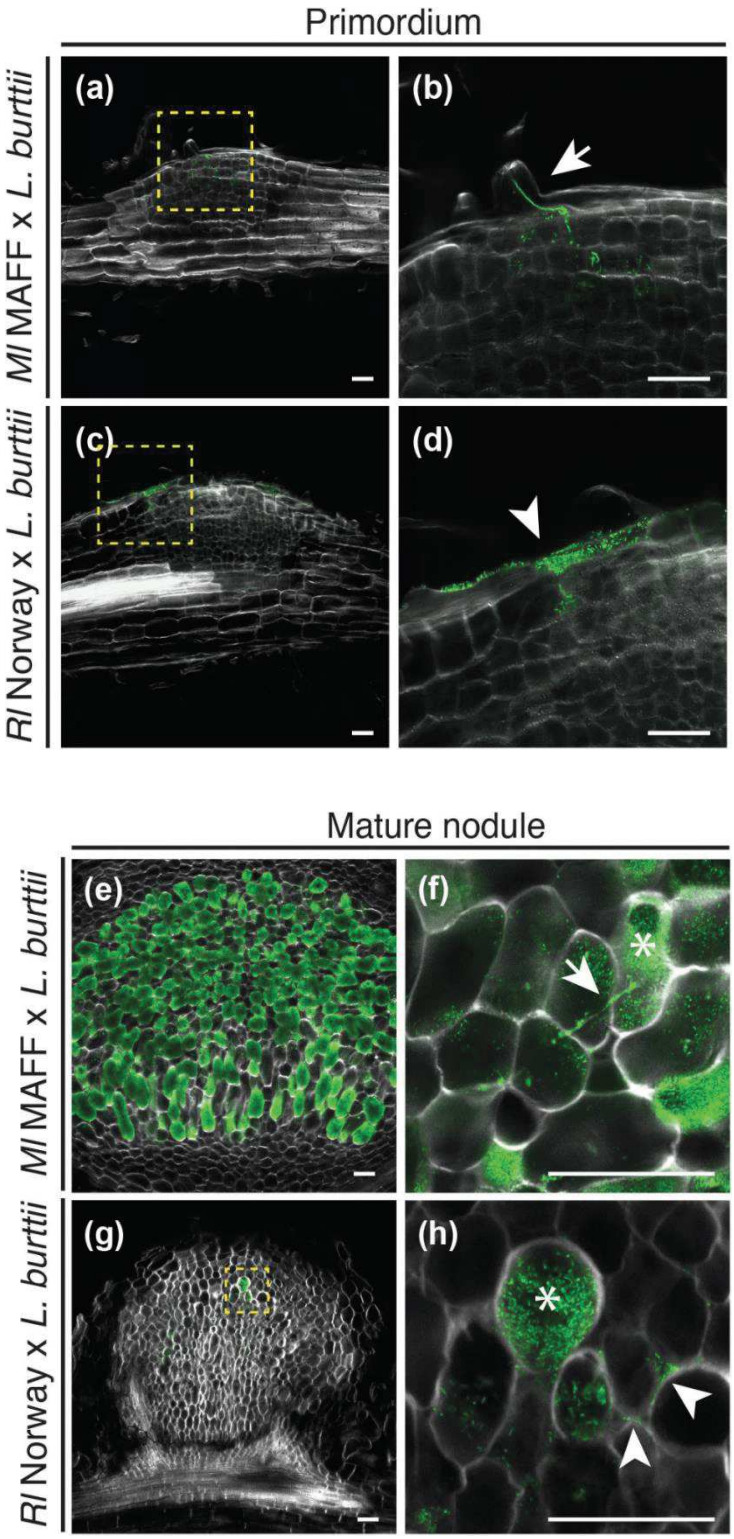


Figure 22 : La souche de *Rhizobium leguminosarum* Norway colonise les nodules de *Lotus burtii* en absence de cordon d'infection transcellulaire. (a-b) La souche *Mesorhizobium loti* infecte le primordium nodulaire 5dpi via des cordons d'infection traversant l'épiderme alors que (c-d) *R. leguminosarum* Norway infecte le primordium nodulaire 11dpi en absence de cordon d'infection. (e-f) *M. loti* infecte totalement le tissu central du nodule, à l'inverse (g-h) *R. leguminosarum* Norway colonise partiellement le nodule de manière intra- et intercellulaire. (D'après Liang et al, 2019)

l'intérieur du clade des légumineuses mais les prérequis génétiques sont encore mal connus.

Existence de processus d'infection intercellulaire « conditionnels »

Le mode d'infection intercellulaire peut être observé aussi dans certaines conditions chez des légumineuses pour lesquelles l'infection via la formation de cordons d'infection est le plus courant. On peut citer l'exemple atypique d'une légumineuse tropicale africaine semi-aquatique, qui a été très étudiée, *Sesbania rostrata*. C'est une des rares Fabaceae possédant un mode d'infection qui diffère suivant l'état émergé ou immergé de la racine. En effet, lorsque la racine n'est pas immergée, la plante utilise le mécanisme classique d'infection intracellulaire avec la courbure des poils absorbants et la formation du cordon d'infection comme chez *L. japonicus* dont elle est proche phylogénétiquement (Fig. 21a,b,c). De plus, comme c'est le cas chez *M. truncatula*, la nodulation dans ces conditions est inhibée par l'éthylène (Goormachtig *et al.*, 2004). En revanche lorsque la racine est immergée, les bactéries présentes au niveau de la fissure de l'épiderme produite par l'émergence de la racine latérale pénètrent de manière intercellulaire au niveau du cortex racinaire (Fig. 21d,e,f) puis induisent une mort cellulaire localisée dépendamment des facteurs Nod, ce qui leur permet de proliférer dans une poche d'infection intercellulaire (Lievens *et al.*, 2005). La formation de cette poche d'infection intercellulaire est aussi dépendante de l'éthylène et c'est à partir de cette structure particulière que des cordons d'infection inter- et intra-cellulaire sont lancés en direction du primordium nodulaire en développement (D'Haeze *et al.*, 2003). Ainsi, *S. rostrata* est capable de noduler de manière semblable aux légumineuses modèles lorsqu'elle est en condition non-immergée, mais lorsque les racines sont immergées, la production d'éthylène inhibe le mécanisme d'infection classique et promeut à la place un mode d'infection alternatif intercellulaire. Ce mécanisme d'infection hybride est donc probablement une adaptation au mode de vie hydrophytique de cette plante (Capoen *et al.*, 2010). On peut également noter que *S. rostrata* possède la structure en couronne de poils axillaires qui n'apparait qu'après l'inoculation de la racine (Boogerd and van Rossum, 1997).

Très récemment, il a été montré (Liang *et al.*, 2019) que certains isolats bactériens naturels de *Lotus* pouvaient infecter de manière intercellulaire la racine sans passer par le poil

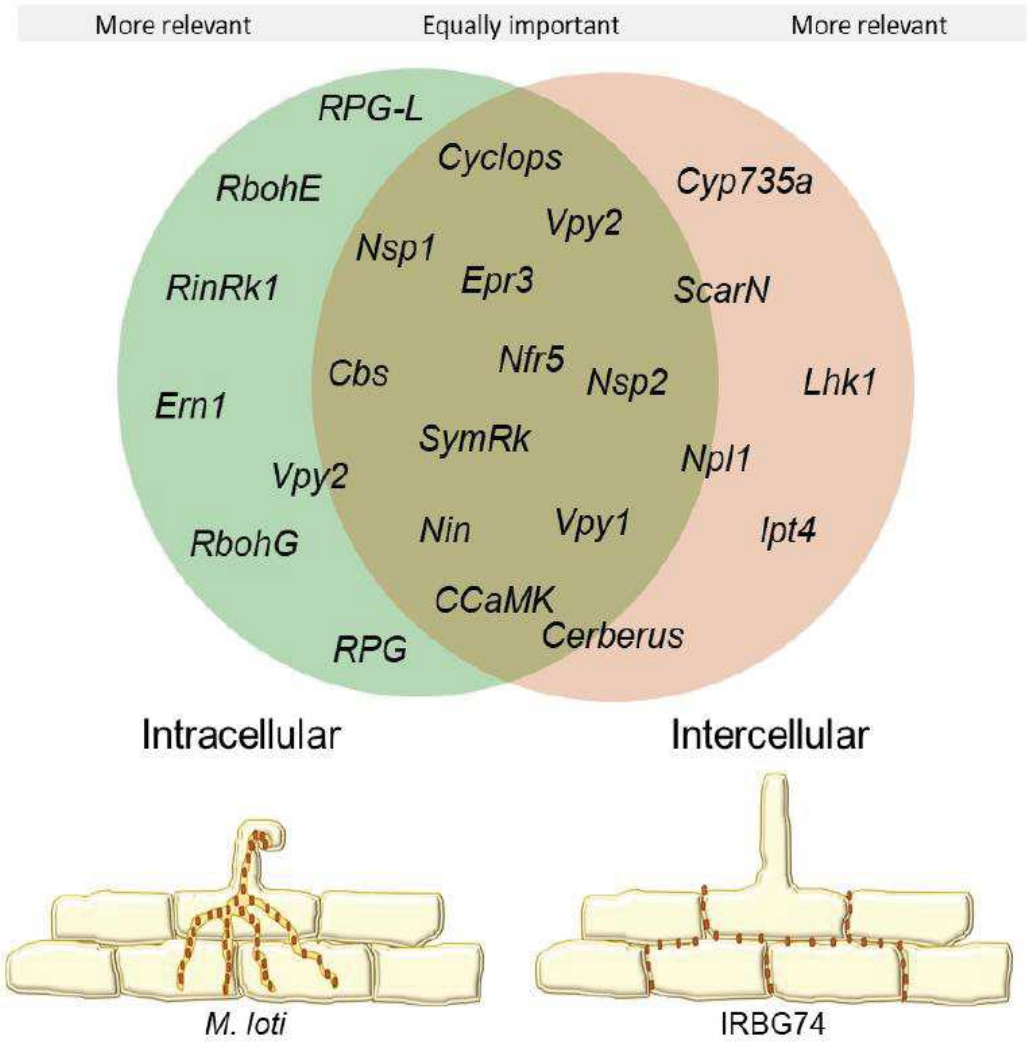


Figure 23 : Modèle comparatif des gènes nécessaires pour les programmes d'infection intra- vs inter-cellulaire chez *Lotus japonicus*. *Lotus* recrute un ensemble de gènes qui sont nécessaires pour la symbiose racinaire quelque soit le processus d'infection de la bactérie. D'autres acteurs symbiotiques sont plus importants pour le processus d'infection intra- ou inter-cellulaire. Il semblerait que l'infection intercellulaire soit sensible à l'absence de certains gènes relatifs aux cytokinines. (D'après Montiel et al, 2020)

absorbant (Fig. 22). Les bactéries s'accumulent en intercellulaire dans le cortex, puis, au moment de l'endocytose, des structures en « chevilles » (peg-like) se forment. La présence de ces structures en cheville est dépendante de la production de facteurs Nod par les bactéries. Ces structures qui ressemblent à des cordons d'infection larges et déformés émergeant des poches de bactéries intercellulaire, ont déjà été décrites chez *Lupinus albus* et *Aeschynomene afraspera* (Gonzalez-Sama *et al.*, 2004; Bonaldi *et al.*, 2011).

Afin d'étudier plus en détail les différents modes d'infection chez cette légumineuse modèle, les phénotypes de nodulation avec la souche de *rhizobium* IRBG74 qui infecte de manière intercellulaire *L. japonicus* et sa souche naturelle *M. loti,* ont été comparés sur différents fonds mutants. L'inoculation de la souche IRBG74 sur plusieurs mutants révèle que RPG, ERN1 et RINRK1 ne sont pas impliqués dans le mécanisme d'infection intercellulaire à l'inverse de NFR5, SYMRK, CCaMK, EPR3, CYCLOPS, NIN, NSP1 et NSP2. Une analyse RNAseq montre également un transcriptome distinct en réponse à l'inoculation d'IRBG74 ou de *M. loti* (Fig. 23). De nombreux gènes relatifs aux cytokinines sont différentiellement exprimés, les mutants *cyp735A* et *ipt4* impliqués dans la biosynthèse de cytokinine sont fortement affectés dans leur nodulation et le mutant *lhk1* ne présente aucuns nodules, suggérant un rôle clé des cytokinines dans le processus d'infection intercellulaire (Montiel *et al.*, 2020). Ces données ont l'intérêt de montrer que les mécanismes d'infection intra- et inter-cellulaire présentent des modalités distinctes, même si dans ce système symbiotique, les cinétiques de nodulation en inter- et intra- sont différentes. Ainsi, explorer les mécanismes de l'infection intercellulaire chez une légumineuse chez qui ce type d'infection est la règle sera aussi important pour mieux comprendre les différences mécanistiques entre ces deux processus d'infection.

Les mécanismes symbiotiques décrits chez les légumineuses modèles sont très sophistiqués alors que nous avons vu qu'il existe des exemples de légumineuses nodulant grâce des mécanismes alternatifs. Etudier l'origine et la diversité des mécanismes de nodulations permet de mieux comprendre les différents types d'interactions dans le but de connaître le set de gènes minimum nécessaire pour aboutir à une symbiose fixatrice d'azote fonctionnelle. A terme, ces connaissances pourraient être utiles pour élaborer des stratégies de transfert de certains gènes clefs sur des non-légumineuses capables de réaliser la symbiose endomycorhizienne afin de tenter d'induire des nodules fixateurs.

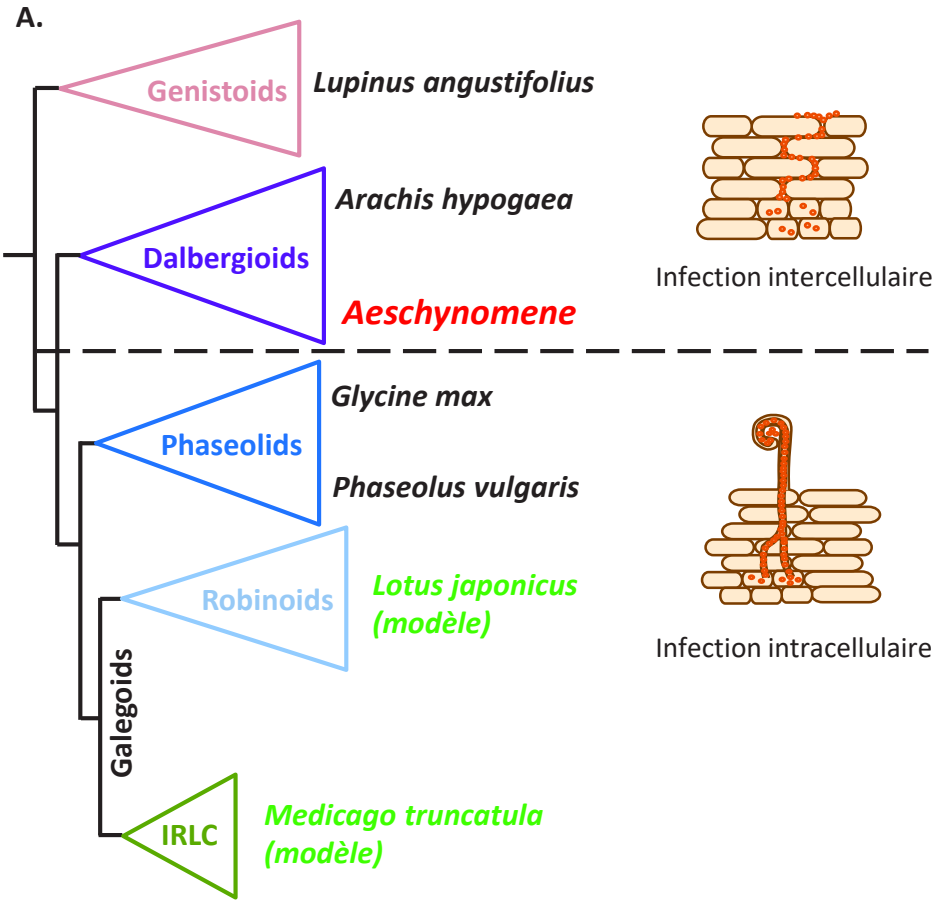
4. Originalités du système symbiotique *Bradyrhizobium-Aeschynomene*

Nous avons vu que l'un des grands enjeux en matière de recherche fondamentale sur la symbiose fixatrice d'azote est d'élucider les adaptations moléculaires responsables de la mise en place d'une symbiose efficace. Les connaissances techniques et scientifiques acquises chez les légumineuses modèles sont une base de travail pour étudier d'autres systèmes symbiotiques afin d'avoir une vue d'ensemble de la diversité des mécanismes de nodulation chez les Fabaceae. Les gènes impliqués dans ces mécanismes sont des outils clés pour tenter de transférer cette capacité à de nouvelles plantes hôtes. Il y a donc un intérêt grandissant à chercher et à étudier des systèmes biologiques possédant des originalités symbiotiques pour pouvoir les comparer aux modèles classiques. A cet égard, des légumineuses tropicales et semi-aquatiques du genre *Aeschynomene* ont attiré l'attention des scientifiques depuis longtemps et ont émergé comme un modèle d'étude de mécanismes symbiotiques alternatifs à ceux rencontrés chez les légumineuses modèles.

4.a - Les *Aeschynomene* : des légumineuses à nodulation de tige

Le genre *Aeschynomene* appartenant au clade des Dalbergioideae, au même titre que l'arachide et partageant le même mode d'infection intercellulaire que cette dernière (Fig. 24A), contient près de 250 espèces dont environ la moitié possède un mode de vie semi-aquatique (mares, sols inondés et marécages) alors que les autres espèces peuvent se développer dans la savane et en forêt sèche (Lavin *et al.*, 2001). Ces différentes espèces sont originaires d'Asie, d'Afrique et d'Amérique; par ailleurs les genres *Soemmeringia*, *Cyclocarpa*, *Kotschya*, *Smithia*, *Geissapsis*, *Bryapsis* et *Humalaria*, sont étroitement liés au genre *Aeschynomen*e (Lavin *et al.*, 2001; Lewis, 2005). Il y a environ une quinzaine d'espèces d'*Aeschynomene* vivant en condition semi-aquatique qui ont développé la capacité de produire des nodules caulinaires en plus des nodules racinaires (Fig. 24B,C).

L'observation des premiers nodules caulinaires a d'ailleurs été faite sur l'espèce *Aeschynomene afrapsera* au début du 20ème siècle par le chercheur Danois Hagerup (Hasgerup, 1924). Un des mécanismes d'adaptation face à l'immersion temporaire ou



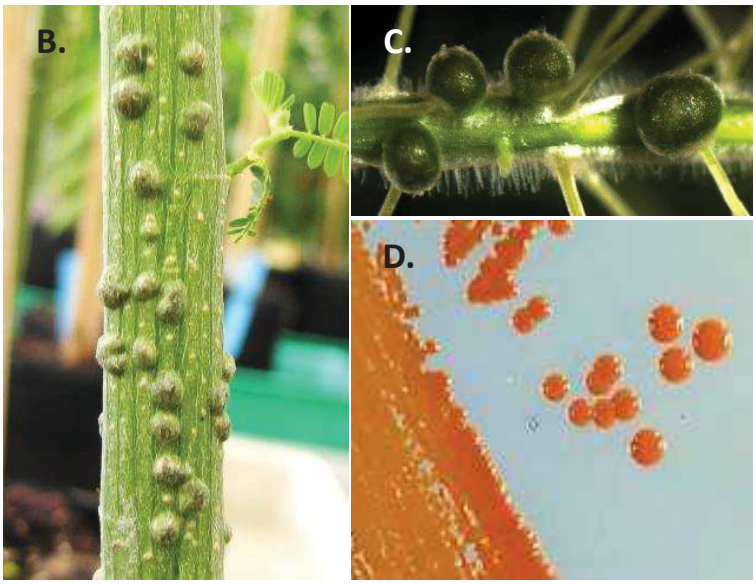


Figure 24 : Les *Aeschynomene*, des légumineuses aux propriétés de nodulation originales.
A. Phylogénie des légumineuses Papilionacées avec les 5 sous-familles les plus représentatives. L'infection intercellulaire a lieu chez les Génistoïdes et les Dalbergioïdes. L'infection intracellulaire a lieu chez les Phaséolides et les Galégoïdes (ce clade contient les 2 espèces modèles). **B.** Nodules caulinaires d'*Aeschynomene* sp. **C.** Nodules racinaires d'*Aeschynomene* sp. **D.** Photo de *Bradyrhizobium* photosynthétiques (pigments caroténoïdes oranges).

permanente, réside dans la présence de primordia racinaires sur la tige dont les méristèmes s'activent au contact de l'eau et qui se développent par la suite en racines adventives. L'émergence de la racine adventive crée des fissures au niveau de l'épiderme de la tige, ce qui fait de ce point le site d'infection privilégié par les bactéries pour former des nodules caulinaires (Alazard and Duhoux, 1988).

Il existe également d'autres genres de légumineuses dans lesquels on retrouve 1 à 2 espèces capables de produire des nodules caulinaires comme *Sesbania* (Galegoideae), *Discolobium* (Dalbergioideae) et *Neptunia* (Mimosoideae) mais cela reste un trait symbiotique très rare chez les légumineuses (Allen and Allen, 1981; Boivin *et al.*, 1997). Cependant, *Sesbania rostrata* est considérée comme le principal modèle d'étude de la nodulation de tige chez qui l'ontologie a été finement étudiée. Ces nodules caulinaires sont de formes ovoïdes, contiennent de la chlorophylle, ont une croissance déterminée et leur zone centrale contient une large masse de cellules infectées (Duhoux, 1984).

En plus de leur particularité par rapport aux nodules de tige, les *Aeschynomene* se distinguent également par l'identité de leur partenaire symbiotique qui appartient exclusivement au genre *Bradyrhizobium*. Parmi les symbiontes isolés de nodules d'*Aeschynomene*, un grand nombre de souches de *Bradyrhizobium* montrent une activité photosynthétique (Fig. 24D). C'est un caractère peu commun chez les rhizobiums qui, en dehors des *Bradyrhizobium* associés aux *Aeschynomene*, n'a été identifié que chez des *Methylobacterium* nodulant des espèces du genre *Lotononis* (Jaftha *et al.*, 2002). Le premier exemple le plus étudié de *Bradyrhizobium* photosynthétique est celui de la souche BTAi1, isolée de nodules caulinaire d'*Aeschynomene indica*, qui synthétise des pigments photosynthétiques de type bactériochlorophylle A et caroténoïde (Eaglesham *et al.*, 1990). Par la suite il a été démontré que les *Bradyrhizobium* photosynthétiques constituaient un sous-clade monophylétique dans l'arbre phylogénétique rDNA 16S de ce genre, bien distinct de *Bradyrhizobium elkanii* et *Bradyrhizobium japonicum*, mais qu'ils ne pouvaient pas être considérés comme une espèce à part entière (Molouba *et al.*, 1999).

D'autre part, il a été montré que l'activité photosynthétique des symbiontes joue un rôle majeur dans l'efficacité de la fixation d'azote des nodules caulinaires (Giraud *et al.*, 2000). Étonnament les gènes relatifs à la photosynthèse sont activés par des longueurs d'ondes

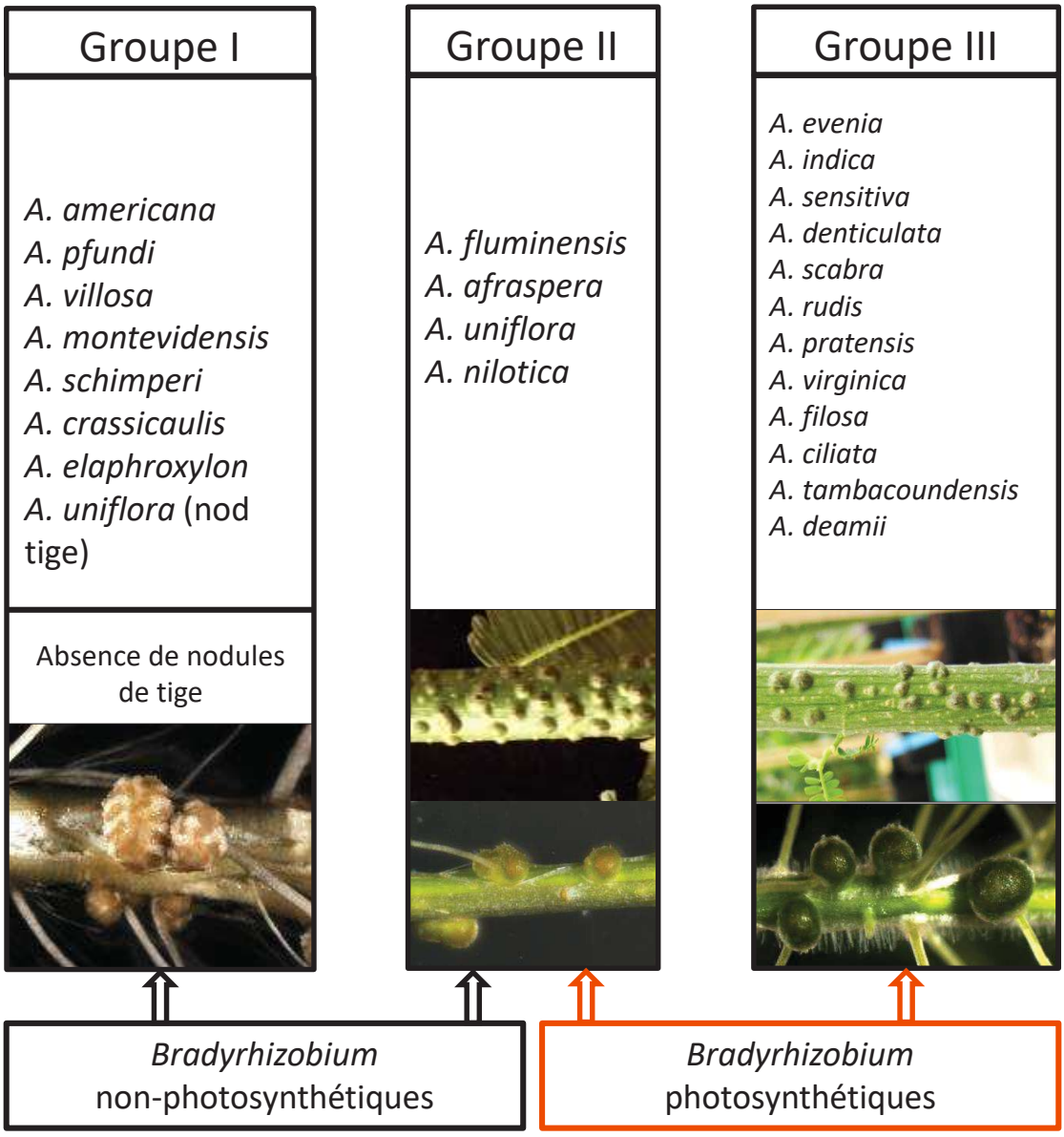


Figure 25 : Groupes d'inoculations croisées chez les *Aeschynomene*. Les espèces du groupe I ne forment pas de nodules de tige (hormis *A. uniflora*). Les *Bradyrhizobium* photosynthétiques nodulent uniquement les groupes II et III, certains d'entre eux sont d'ailleurs spécifiques du groupe III et ne peuvent pas noduler les membres du groupe II. (Adapté d'Alazard, 1985)

situées dans le rouge lointain. Ce mécanisme original de régulation paraît être parfaitement adapté pour promouvoir la photosynthèse dans les nodules caulinaires, puisque c'est la seule longueur d'onde restant après absorption de la lumière par la chlorophylle végétale. Il a aussi été observé que la lumière infra-rouge stimule la fixation d'azote dans les nodules contenant des bactéroïdes photosynthétiques, suggérant que la photosynthèse bactérienne procure de l'énergie pour la fixation d'azote et permet ainsi une croissance plus rapide de la plante en compensant le phénomène de « puit de carbone » ou carbon-sink (Giraud and Fleischman, 2004). De plus, le nombre de nodules caulinaires observés chez certaines espèces comme *A. afraspera* ou *A. nilotica* peut-être extrêmement élevé, permettant à ces légumineuses de fixer l'azote atmosphérique très efficacement. Avec une fixation d'azote estimée à environ 100kg/ha, ces légumineuses ont été proposées comme engrais vert en rizière pour remplacer les engrais azotés chimiques (Alazard and Becker, 1987).

Par ailleurs, des études d'inoculations croisées ont permis de déterminer le spectre d'hôte de différentes souches de *Bradyrhizobium* isolées de nodules caulinaires et racinaires d'*Aeschynomene*, ce qui a mené à la création de 3 groupes d'inoculation croisée (Fig. 25). Le premier groupe (I) correspond à des espèces qui nodulent uniquement sur la partie racinaire et sont associées à des *Bradyrhizobium* non-photosynthétiques (exemple *A. americana*). Les groupes II et III, eux, sont composés des *Aeschynomene* à nodulation caulinaire. Dans le second groupe (II) on retrouve des *Aeschynomene* produisant un nombre impressionnant de nodules caulinaires et qui peuvent s'associer avec des *Bradyrhizobium* photosynthétiques ou non-photosynthétiques (exemple *A. afraspera*). Dans le troisième groupe (III), les *Aeschynomene* produisent des nodules de tige et s'associent uniquement avec des *Bradyrhizobium* photosynthétiques de manière très spécifique (exemple *A. indica*). Dans les groupes II et III, on observe deux niveaux de spécificité des *Bradyrhizobium*, d'abord ceux ayant un spectre d'hôte plus large et pouvant noduler l'ensemble des *Aeschynomene* des groupes II et III, et ensuite ceux ayant un spectre d'hôte plus restreint ne pouvant noduler que les *Aeschynomene* du groupe III (Alazard, 1985).

4.b - L'existence d'une symbiose indépendante des facteurs Nod

La particularité symbiotique la plus notable dans le genre *Aeschynomene*, en plus des traits originaux cités précédemment, est la capacité de s'associer avec des *Bradyrhizobium*





C

Nod protein	ORS278			
	CDS	Similarity (%)	Acc. no.	E value
NodA *	Absent			
NodB	Brado4564	33% to *R. galegae*	P50354	$2e^{-16}$
NodC *	Absent			
NodD	Brado3695	39% to *R.*	Q53061	$1e^{-54}$
NodE *	Absent			
NodG	Brado3311	66% to *M.*	P72332	$4e^{-86}$
NodI	Brado0517	39% to *R. tropici*	Q933C0	$5e^{-37}$
NodJ *	Absent			
NodL	Brado2093	42% to *R. leg.*	P08632	$2e^{-11}$
NodM	Brado3763	78% to *B.*	Q9AQ10	$1e^{-277}$
NodP	Brado5192	75% to *B. elkanii*	BAB55898	$1e^{-133}$
NodQ	Brado5193	66% to *B. elkanii*	BAB55899	$5e^{-236}$

Figure 26 : Efficience de la nodulation de tige avec *Bradyrhizobium* ORS278 et ORS285 (et le mutant ORS285Δ*nodB* ne produisant plus de facteurs Nod). Sur **A.** *A. afraspera* et **B.** *A. sensitiva*. On observe des nodules même en absence de facteurs Nod chez *A. sensitiva*. **C.** Tableau des séquences codantes dans le génome d'ORS278 partageant une similarité avec les gènes *nod* connus. Les étoiles en rouge soulignent l'absence de gène clés pour la synthèse des FNs. (D'après Giraud et al., 2007)

photosynthétiques qui pour certains n'ont pas la capacité de produire des facteurs Nod. En effet en 2007, E. Giraud et son équipe ont découvert en séquençant le génome complet de deux souches de *Bradyrhizobium* (BTAi1 et ORS278) s'associant à *A. sensitiva* et *A. indica*, que les gènes nécessaires à la synthèse des facteurs Nod étaient absents (Fig. 26). Cette découverte majeure a permis de mettre à jour un nouveau paradigme : l'existence d'une symbiose rhizobienne indépendante des facteurs Nod ou « Nod-indépendante ». C'est la première fois que l'on découvre une légumineuse qui interagit avec des rhizobiums sans utiliser les facteurs Nod. Cette découverte soulève de nombreuses questions scientifiques sur la nature du signal bactérien activant la nodulation et sur les étapes de reconnaissances du facteur Nod qui peuvent être contournées (Giraud *et al.*, 2007). Afin de mieux comprendre les mécanismes symbiotiques régissant cette symbiose facteur Nod indépendante côté bactérie, une mutagénèse aléatoire (Tn5) a été entreprise sur la souche ORS278 pour repérer des mutants incapables de produire des nodules avec *A. sensitiva*. Sur plus de 9500 mutants testés, seulement 27 ont eu un impact sévère sur la symbiose avec induction de quelques pseudonodules sur certaines plantes. Cependant aucun mutant strictement Nod^- n'a été observé, cela peut s'expliquer par de la redondance fonctionnelle dans les premières étapes du processus symbiotique, des mutations létales ou une mutagénèse non-saturante. La majorité des mutants présentant un phénotype de nodulation avaient une insertion dans des gènes déjà connus chez les rhizobiums modèles. Plusieurs espèces de *Rhizobium* auxotrophes pour des dérivés de purines et déficients dans la symbiose ont déjà été observés dans d'autres études, ici 14 mutants d'ORS278 affectés dans des gènes relatifs à la biosynthèse des purines ont été obtenus. A l'époque, ces données ont suggéré qu'un dérivé de purine, potentiellement une molécule « cytokinin-like », jouait un rôle majeur dans la formation des nodules chez *A. sensitiva*. Pour vérifier cette hypothèse des mutants de *Bradyrhizobium* ORS285 ont été analysés, ces mutants *miaA* sont affectés dans la production de CKs (2-methylthiol) et montre un retard de nodulation/fixation avec *A. indica* et *A. afraspera.* De plus, *A. indica* produit des nodules plus petits et en plus grand nombre lorsqu'on inocule les mutants *miaAi* par rapport à la souche sauvage. Ces analyses montrent que les CKs produites par la souche ORS285 ne sont pas les molécules signal clés permettant la nodulation Nod-indépendante mais elles contribuent positivement au développement des nodules chez les *Aeschynomene* (Podlešáková *et al.*, 2013).

Des pistes sur l'hypothèse que des molécules de surface pourraient servir de signal ont été explorées, comme l'implication de LPS particuliers produits par les souches ORS278 et BTAi1 lors de l'interaction. En effet un nouveau sucre a été identifié, le bradyrhizose qui forme l'antigène-O de ce LPS. L'analyse de la structure du LPS révèle aussi qu'une molécule hopanoïde est accrochée au lipide A. La structure unique de ce LPS a été étudiée chez la souche de *Bradyrhizobium* ORS285, montrant que l'hopanoïde attaché au lipide A jouait un rôle important dans la vie libre et symbiotique de la bactérie, en renforçant la stabilité de la membrane externe et en conférant un rôle dans le maintien chronique de l'infection intracellulaire (Silipo *et al.*, 2014). En revanche les mutants affectés dans la biosynthèse ou l'accrochage de l'antigène-O n'ont pas présentés de phénotype symbiotique, peut-être à cause de la présence de nombreux autres sucres de surfaces (EPS ou KPS) qui masqueraient l'épitope du LPS (Busset *et al.*, 2016).

Parallèlement à la symbiose entre certaines espèces d'*Aeschynomene* et des *Bradyrhizobium* ne produisant pas de facteurs Nod, il a été montré que des mutants déficients pour la production de facteurs Nod chez *B. elkanii* pouvaient induire des nodules sur *G. max* (soja) muté ou non dans le récepteur aux facteurs Nod, NFR1. Dans ce dernier cas, cette symbiose FN-indépendante n'est rendue possible que par la présence d'un système de sécrétion de type 3 (T3SS) fonctionnel (Okazaki *et al.*, 2013). Ce T3SS permet donc d'activer la nodulation en contournant l'étape de reconnaissance des FN chez le soja.

L'étude de la souche de *Bradyrhizobium* ORS285 a permis de découvrir que le T3SS était également impliqué dans la symbiose avec les *Aeschynomene*. En effet, une large diversité de *Bradyrhizobium* non-photosynthétiques comme *B. elkanii* sont capables d'induire des nodules sur l'espèce Nod-indépendante *A. indica*. C'est un système artificiel qui permet de s'affranchir de la reconnaissance des facteurs Nod et qui montre que le T3SS des *Bradyrhizobium* non-photosynthétiques joue un rôle majeur dans la nodulation à l'inverse de ce que l'on pensait jusqu'à lors. Cependant, la majorité des *Bradyrhizobium* photosynthétiques ne possèdent pas la machinerie du T3SS, à l'exception de la souche ORS285 qui en possède un mais qui n'est pas impliqué dans la mise en place de la symbiose Nod-indépendante, car avec un T3SS muté ou non, la souche induit toujours des nodules (Okazaki *et al.*, 2016).

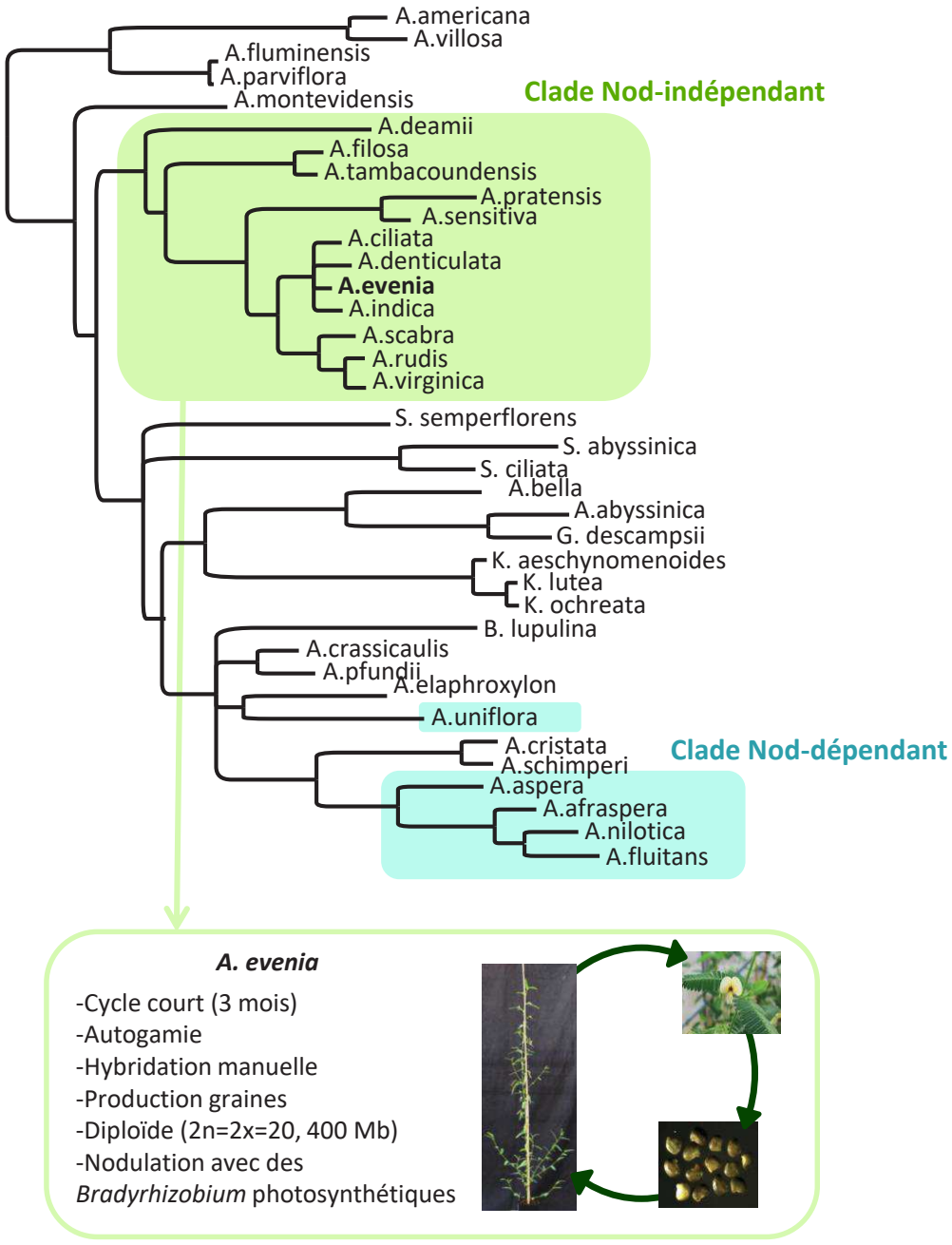


Figure 27 : Arbre phylogénétique (ITS + *trnL*) du genre *Aeschynomene* et d'espèces appartenant à des genres associés. En vert le clade contenant uniquement des espèces nodulant de manière indépendante des facteurs Nod et en bleu le clade des espèces nodulant uniquement avec les facteurs Nod. L'espèce modèle *A. evenia* a été définie en 2012, grâce a ses caractéristiques biologiques, physiologiques et la taille de son génome (Brottier et al., 2018; Arrighi et al, 2012)

Une étude plus récente sur les effecteurs délivrés par la souche de *Bradyrhizobium* non-photosynthétique ORS3257, a permis d'identifier plusieurs d'entre eux comme étant impliqués dans la symbiose avec *A. indica*. Un de ces effecteurs, ernA (pour effector-required for nodulation-A) joue un rôle majeur dans la formation des nodules chez *A. indica*, en effet l'expression d'ernA chez une souche de *Bradyrhizobium* incapable de produire des nodules avec *A. indica* lui confère la capacité de noduler. En plus de la démonstration de son interaction avec les acides nucléiques, il a été montré que l'expression ectopique d'ernA chez *A. indica* provoque un phénotype particulier de « nodule-like » et de « root-like », suggérant un rôle de cet effecteur dans le développement de l'organe symbiotique grâce à un mécanisme moléculaire qui reste encore à clarifier (Teulet *et al.*, 2019).

Ces données indiquent qu'il existe deux processus Nod-indépendant possibles pour interagir avec les *Aeschynomene*, le premier avec des souches non-photosynthétiques qui est dépendant du T3SS et le second avec des souches photosynthétiques dont les mécanismes ne sont pas encore élucidés malgré que de nombreuses analyses fonctionnelles aient été menées sur les bactéries. Le manque de connaissances au niveau génétique et fonctionnelle côté plante hôte, doit être comblé pour ouvrir de nouvelles pistes de recherche afin d'identifier les signaux non-Nod.

4.c - *Aeschynomene evenia* comme espèce modèle de la symbiose Nod-indépendante

Pour comprendre l'évolution des traits originaux que sont la nodulation de tige et la symbiose avec des *Bradyrhizobium* ne produisant pas de facteurs Nod, une phylogénie moléculaire (gènes marqueurs nucléaire et cytoplasmique) du genre *Aeschynomene* et une caractérisation des propriétés de nodulation de différentes espèces ont été réalisées. La présence de primordium de racine adventive sur la tige a été observée dans tous les clades distincts du genre *Aeschynomene*, cette omniprésence dans tous les taxa et les analyses phylogénétiques appuient l'hypothèse d'une prédisposition génétique ancestrale située à la racine du clade des *Aeschynomene* capables de produire ces primordia caulinaires comme adaptation au mode de vie hydrophytique, par la suite des mutations indépendantes se sont produites dans plusieurs clade, permettant la nodulation de tige (Chaintreuil *et al.*, 2013; Brottier *et al.*, 2018). En revanche, le caractère « nodulation facteur Nod-indépendante » est retrouvé dans un unique clade où il n'y a que des espèces Nod-indépendantes, niché au

cœur des *Aeschynomene* Nod-dépendantes (Fig. 27). Cela suggère que la capacité d'interagir avec des *Bradyrhizobium* ne produisant pas de facteurs Nod n'est apparue qu'une seule fois au cours de l'évolution du genre (Chaintreuil *et al.*, 2013; Brottier *et al.*, 2018).

Pour les *Aeschynomene* utilisant un processus symbiotique indépendant des FN, des analyses de caractérisation génétique plus poussées ont été réalisées en déterminant le nombre de chromosome. Au sein de ce clade, 13 espèces sont diploïdes (2n=*2x*=20) et 6 sont polyploïdes (4 espèces 2n=*4x*=40 et 2 espèces 2n=*6x*=60), les analyses montrent que l'ancêtre commun de ce clade était diploïde mais que certaines espèces ont subi des évènements de polyploïdisation récents. Les approches de phylogénie et de cytogénétique ont permis de retracer l'histoire des espèces polyploïdes en montrant que chacune était issue d'une hybridation interspécifique associée à une duplication du génome. Ces travaux ont donc permis de montrer que toutes les *Aeschynomene* Nod-indépendantes partageaient la même origine évolutive (Arrighi *et al.*, 2013; Chaintreuil *et al.*, 2018) (Fig. 27).

Pour choisir parmi ces espèces d'*Aeschynomene* utilisant un processus FN-indépendant, un modèle d'étude, il a fallu déterminer l'espèce avec les meilleures caractéristiques. Pour cela plusieurs paramètres ont été pris en compte comme le niveau de ploïdie du génome, la floraison, la production de graines et la durée du cycle de développement, ce qui a permis d'éliminer plusieurs candidats. Rapidement c'est l'espèce *A. evenia* qui s'est distinguée grâce à un génome diploïde et de petite taille pour une légumineuse (2n=*2x*=20, 400Mb) (, son autogamie (permettant la création de lignées homozygotes), son cycle de vie court (10 semaines), son rendement en graines élevé, sa taille raisonnable et la possibilité de réaliser des hybridations manuelles (Fig. 27). C'est aussi l'espèce Nod-indépendante pour laquelle le plus grand nombre d'accessions est disponible et la diversité génétique est la plus grande (Arrighi *et al.*, 2013; Chaintreuil *et al.*, 2018)**.** D'autre part, *A. evenia* est capable de s'associer avec les souches de *Bradyrhizobium* photosynthétiques ORS278 (sans gènes *nod*) et ORS285 (avec gènes Nod) pour former des nodules fonctionnels. Plusieurs outils ont également été développés comme la transformation par *Agrobacterium rhizogenes* qui permet de transformer uniquement la racine de la plante. Les données montrent que le taux de transformation est supérieur à 50% et que le taux de nodulation est supérieur à 85%. L'ensemble de ces données indiquent que les approches génétiques et fonctionnelles sont

possibles chez *A. evenia*, qui a donc été proposée comme modèle d'étude de la symbiose Nod-indépendante (Arrighi *et al.*, 2012).

Aeschynomene evenia (diploïde) est très proche d'*Aeschynomene indica* (polyploïde) sur laquelle de nombreux travaux scientifiques avaient été réalisés auparavant, facilitant ainsi le transfert de connaissances. En effet, une étude a comparé les mécanismes d'infection d'une espèce d'*Aeschynomene* Nod-dépendante (*A. afraspera*) et d'une *Aeschynomene* Nod-indépendante (*A. indica*). Chez ces deux espèces le processus débute par la colonisation des couronnes de poils axillaires, les bactéries pénètrent ensuite dans le milieu intercellulaire en passant par la jonction entre deux cellules épidermiques. Dans le cas d'*A. afraspera*, cet évènement a lieu au niveau de la jonction entre une cellule de poil axillaire et une cellule épidermique (similaire au processus observé chez l'arachide). En revanche, pour *A. indica* les bactéries passent par la jonction entre deux cellules de poils axillaires. Chez ces deux espèces, on peut observer que les premières cellules infectées au niveau du cortex externe subissent une mort cellulaire, les cellules voisines compressent cette zone et forment une sorte de poche intercellulaire qui va aider les bactéries à se répandre dans d'autres parties de la racine. A l'inverse des cellules du cortex externe, les cellules situées plus en profondeur semblent parfaitement s'accommoder après l'endocytose des bactéries et se divisent pour former le tissu uniformément infecté. La mort cellulaire de certaines cellules corticales ressemble à celles observées chez *Stylosanthes* et les structures de poches d'infection intercellulaire ressemblent à celles observées chez *Sesbania rostrata* (cf partie 2). Chez *A. afraspera* et *A. indica*, le développement du nodule est comparable avec une large zone d'infection centrale entouré par du cortex non-infectés dont la vascularisation est connectée à celle de la racine principale. Cependant, les nodules d'*A. indica* contiennent uniquement des bactéroïdes différenciés de forme ronde alors que les nodules d'*A. afraspera* contiennent des bactéroïdes différenciés en forme de bâtonnets allongés (Bonaldi *et al.*, 2011). Ces différents traits de nodulation font des *Aeschynomene* des modèles atypiques mais idéaux pour étudier des mécanismes de reconnaissance, de signalisation, d'infection et de fonctionnement du nodule.

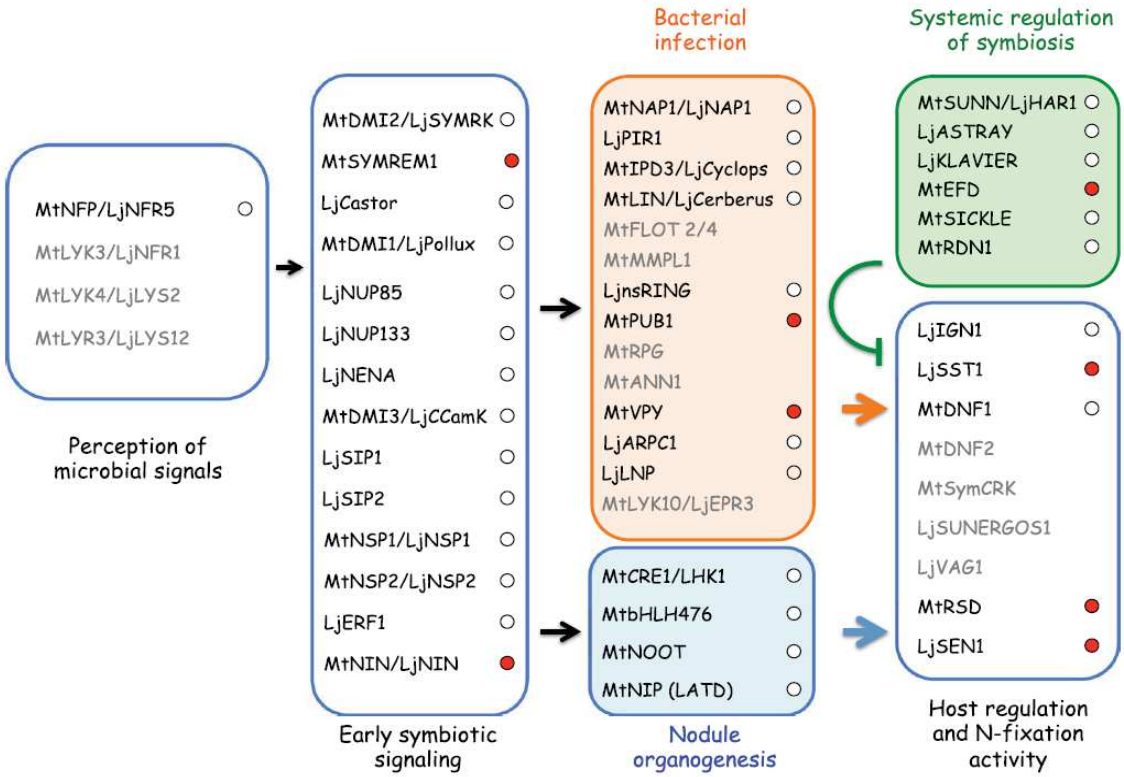
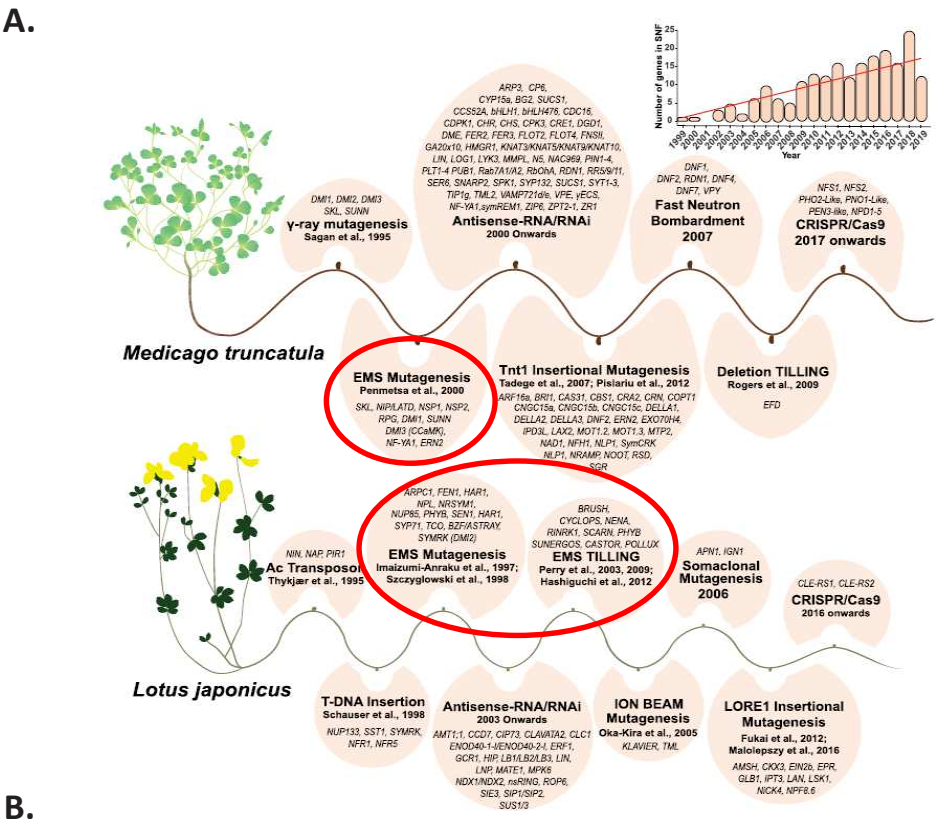


Figure 28 : Représentation schématique de la cascade de signalisation symbiotique avec les gènes identifiés chez les légumineuses modèles. Ces gènes ont été classés selon leur fonction putative dans la nodulation. Les gènes en noir ont été identifiés dans le transcriptome *d'A. evenia* alors que pour les gènes en gris on ne retrouve pas de transcrits. La régulation de ces gènes dans la racine au cours de la nodulation est illustrée par un cercle blanc pour les gènes exprimés constitutivement et par un cercle rouge pour les gènes surexprimés. (D'après Gully et al, 2018)

4.d - Approches génétiques et moléculaires pour mettre à jour les mécanismes de la symbiose Nod-indépendantes chez *A. evenia*

Afin d'avoir un premier aperçu du contenu en gène symbiotiques et de leur expression dans différentes conditions chez le nouveau modèle *A. evenia*, une analyse transcriptomique (RNAseq) a été réalisée sur plusieurs organes. Les orthologues des gènes symbiotiques de *M. truncatula* et *L. japonicus* ont été recherchés et plus de 36 ont été identifiés. Ces analyses ont permis de révéler l'absence d'expression de certains gènes impliqués dans les étapes de reconnaissance de la bactérie comme *LYK3* et *EPR3*, de gènes relatifs à l'infection comme FLOT et RPG, ainsi que des gènes impliqués dans le fonctionnement du nodule comme *DNF2*, *SymCRK* et *VAG1* (Chaintreuil *et al.*, 2016). Une étude transcriptomique réalisée en 2018 a permis d'identifier des gènes connus exprimés différentiellement entre la racine inoculée et prélevée a différent temps (6 hpi, 1, 2, 4 et 6dpi) ou non-inoculée (0 dpi). Dans les réponses précoces (6 hpi) on retrouve des DEGs (Differentially Expressed Gene) impliqués dans le métabolisme, les processus biosynthétiques, les processus cellulaires, la signalisation ou encore la réponse au stress. Concernant les réponses tardives (6 dpi), la majorité des DEGs sont relatifs à des processus cataboliques et métaboliques ainsi que dans la réponse au stress, notamment les NCRs qui sont sollicités lors de la différenciation en bactéroïdes (Gully *et al.*, 2018) (Fig. 28). Ces 2 publications ont conforté l'idée que, malgré un certain niveau de conservation avec les modèles, certains mécanismes symbiotiques présentent des différences majeures avec ceux déjà décrits. Des approches de génétique inverse (RNAi) ont été développées pour estimer le niveau de conservation de la voie symbiotique identifiée chez les légumineuses modèles. L'extinction de l'expression de 4 gènes symbiotiques connus pour leur rôle majeur chez les légumineuses modèles (*SYMRK, CCaMK, LHK1/CRE1* et *DNF1*) révèle une conservation partielle des mécanismes de signalisation chez *A. evenia*. La réduction du niveau de transcrits du gène *AeSYMRK* et *AeLHK1* provoque une baisse de la nodulation se traduisant par un nombre de nodules par plante plus faible que le contrôle. L'extinction d'*AeCCaMK* provoque une baisse sévère du nombre de nodules et l'expression d'une forme auto-active induit la formation de nodules spontanés en absence de bactéries, comme cela a été observé chez le *L. japonicus*. Ces données montrent que ces gènes de signalisation sont requis pour la mise en place d'une symbiose Nod-indépendante chez les *Aeschynomene* (Fabre *et al.*, 2015).



B.

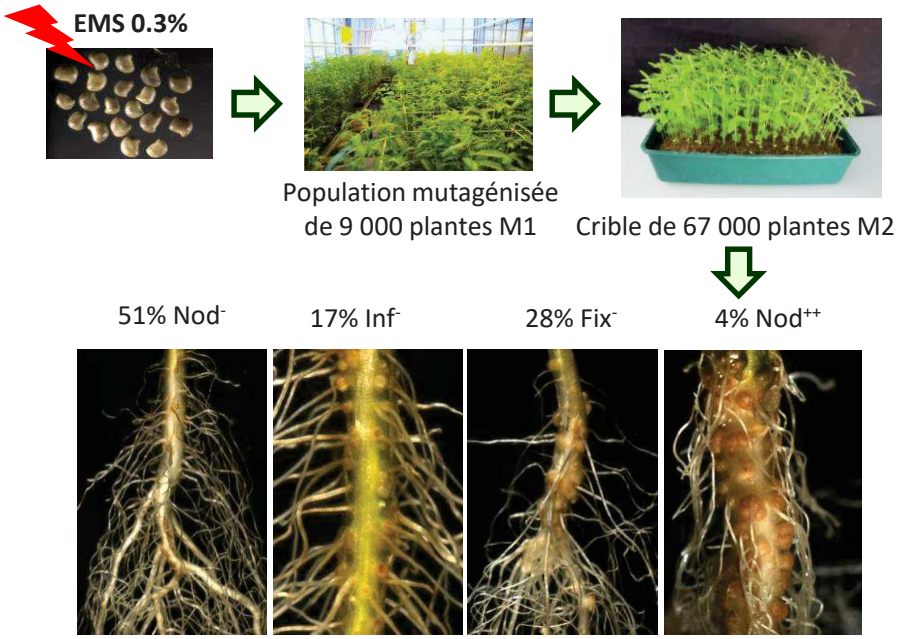


Figure 29 : Les approches génétiques et les outils disponibles pour étudier les gènes symbiotiques chez les légumineuses. A. Représentation des différentes approches de génétique directe, dont des mutagénèses EMS réalisées sur les deux légumineuses modèles (cercles rouges) avec les gènes correspondants trouvés par chacune de ces méthodes durant les 20 dernières années. **B.** La mutagénèse EMS chez *A. evenia* et le crible de la population mutante a abouti à l'identification de 250 mutants altérés dans le processus de nodulation (**A.** : d'après Roy et al., 2020).

De plus, une autre étude a montré que les étapes tardives de la symbiose Nod-indépendante dans le genre *Aeschynomene* mettent en jeu les mêmes mécanismes de différenciation terminale que chez les IRLCs (*M. truncatula*) grâce à l'action de peptide riches en cystéines, les NCRs. Étonnamment des NCRs ont été retrouvés chez les *Aeschynomene* alors qu'on pensait que ces peptides étaient spécifiques au clade des IRLCs, mais la similarité avec les NCRs de *M. truncatula* au niveau du peptide signal et des motifs riche en cystéine, confirme que ce clade distinct possède aussi les gènes nécessaires à leur synthèse (Dalbergioïdaea). Tout comme chez *M. truncatula*, le gène *DNF1* est impliqué dans le transport des NCRs au symbiosome et l'extinction de son expression par RNAi chez *A. evenia* provoque une absence de différenciation des bactéroïdes (Czernic *et al.*, 2015). La similarité de ces mécanismes ouvre la question de savoir si les NCRs des IRLCs partagent une origine commune avec ceux des *Aeschynomene*, dans ce cas ils étaient présents à la racine du clade des légumineuses et plusieurs genres auraient perdu ces gènes au cours de l'évolution de la famille.

Afin d'identifier les nouveaux déterminants génétiques potentiellement spécifique du processus Nod-indépendant, une approche de génétique directe, sans *a priori*, a été développé au laboratoire. La génétique directe de type mutagénèse aléatoire est utilisée depuis de nombreuses années pour identifier des déterminants génétiques impliqués dans certains processus biologiques. Cette approche puissante basée sur l'analyse du phénotype a permis l'identification de nombreux gènes symbiotiques chez les légumineuses modèles. Ces approches sans *a priori* consistent à générer de nombreuses mutations dans le génome de l'espèce ciblée en l'exposant à des agents mutagènes, puis de passer au crible la population de mutants pour isoler ceux qui ont le phénotype recherché. Il existe plusieurs types de mutagénèses aléatoires basées sur différents agents mutagènes qui ont pour but de générer une population mutante comme les rayonnements ɣ, la radiation Fast Neutron, les transposons ou encore l'EMS. Historiquement la première mutagénèse réalisée chez *M. truncatula* a été effectuée par rayonnement gamma (ɣ-ray mutagenesis) et a permis l'isolation d'une vingtaine de mutants affectés dans la nodulation et la mycorhization (Sagan *et al.*, 1995). Par la suite, une mutagénèse EMS à grande échelle a été générée, permettant l'isolation de plusieurs mutants altérés dans différents processus biologique dont la symbiose (Penmetsa and Cook, 2000). Pour finir, une mutagénèse insertionnelle avec le rétro-transposon Tnt1 a permis d'isoler plus de 421 lignées mutants altérés dans la

nodulation (Tadege *et al.*, 2008). En parallèle, des approches similaires (EMS et Tnt) ont été réalisées sur la deuxième légumineuse modèle *Lotus japonicus* ce qui a permis d'isoler de nombreux mutants ayant un phénotype de nodulation (Szczyglowski *et al.*, 1998; Urbański *et al.*, 2012) (Fig. 29A). Ces techniques ont toutes leurs propres avantages et inconvénients, par exemple l'EMS (EthylMethanoSulfonate) est un agent chimique mutagène qui favorise l'apparition de nombreuses mutations ponctuelles (Single Nucleotide Polymorphism) et principalement dans transitions G>C vers A>T. L'intérêt de cette technique est que le génome sature rapidement en mutations, cette approche est gérable au niveau d'une équipe car il n'y a pas besoin de générer une population de mutagénèse trop importante. La recherche des mutations ponctuelles dans le génome, qui pouvait poser problème il y a encore quelques années, est aujourd'hui facilitée par le développement des nouvelles technologies de séquençage.

Dans le but de trouver les déterminants génétiques impliqués dans la symbiose Nod-indépendante, une stratégie de mutagénèse EMS a été initiée chez *Aeschynomene evenia* (Fig. 29B). Des expériences préliminaires ont permis de déterminer les conditions optimales pour que le génome soit saturé en mutations sans que l'effet ne soit trop délétère sur le développement des plantes et la concentration de 0,3% d'EMS a été choisie. Ce traitement a été appliqué sur une population de 8 000 plantes M1 qui ont été autofécondées pour donner naissance à 67 000 plantes M2 criblées sur leur phénotype symbiotique. La présence d'un petit pourcentage de plantes albinos dans cette population confirme que le traitement EMS a fonctionné correctement. Plus de 250 mutants M2 ont montré un phénotype altéré dans la nodulation racinaire dont 51% était strictement incapable de produire des nodules (Nod^-), 17% présentaient une altération du processus d'infection (Inf^-), 28% étaient incapables de fixer l'azote (Fix^-) et 4% étaient hypernodulants (Nod^{++}). Cette collection de mutants de nodulation représente donc une ressource génétique importante pour identifier des gènes symbiotiques. Pour pouvoir remonter du phénotype symbiotique jusqu'au gène impliqué, il est important de posséder une version annotée du génome pour pouvoir comparer la séquence du mutant d'intérêt à une séquence de référence, cela permettra de localiser précisément la mutation et d'identifier le gène en question. Un séquençage complet du génome d'*A. evenia* a donc été entrepris dans le but de pouvoir étudier le contenu en gène symbiotique et servir d'outils pour les analyses futures.

5. Présentation du projet de thèse

C'est dans le contexte très dynamique du développement d'une approche de génétique directe chez *A. evenia*, nouvellement choisie comme modèle d'étude de la symbiose Nod-indépendante, que s'inscrit mon travail de thèse. L'objectif de ce travail a été de progresser dans la compréhension des mécanismes moléculaires qui sous-tendent l'établissement de cette symbiose fixatrice d'azote sans l'intervention de facteurs Nod et de cordons d'infection. Pour cela, j'ai développé une démarche en trois étapes :

• Pour obtenir une vue générale des mécanismes symbiotiques qui sont potentiellement conservés et ceux qui sont divergents chez *A. evenia*, j'ai exploré son génome tout juste annoté et j'ai mené une analyse *in silico* des gènes symbiotiques initialement identifiés chez les légumineuses modèles, *M. truncatula* et *L. japonicus*. Ce travail a conforté l'idée que les mécanismes de reconnaissance du partenaire rhizobien et d'infection sont différents chez *A. evenia*, et il a mis en exergue également des différences notables lors des étapes plus tardives de la nodulation.

• Pour réaliser la dissection génétique des étapes précoces de la symbiose Nod-indépendante, j'ai participé au crible en cours dans l'équipe de la population de mutagenèse EMS développée pour *A. evenia* et isolé des mutants de nodulation. Je me suis consacré ensuite à la caractérisation des mutants altérés dans les étapes les plus précoces de l'interaction pour définir des groupes phénotypiques, établir le déterminisme génétique de ces mutants et les regrouper dans des groupes de complémentation.

• Pour identifier les gènes symbiotiques altérés chez ces mutants de nodulation, j'ai en parallèle mis en œuvre deux stratégies de séquençage. Une approche de Targeted Sequence Capture a été réalisée sur l'ensemble de la collection de 250 mutants pour rechercher de façon ciblée des mutations dans un set de gènes symbiotiques connus chez les légumineuses modèles. Pour les mutants bloqués dans les étapes précoces de l'interaction et qui ne sont pas associés à un de ces gènes, une recherche sans *a priori* de gènes symbiotique a alors été conduite par Mapping-by-Sequencing. Le travail d'analyse *in silico* des gènes symbiotiques, la caractérisation des mutants de nodulation, l'identification et l'analyse de gènes symbiotiques conservés et de gènes nouveaux sont présentés dans les chapitres 1, 2 et 3 qui suivent.

Chapitre 1

1.1 Contexte de l'étude

Plusieurs espèces d'*Aeschynomene* sont uniques parmi les légumineuses en raison de leur capacité à établir une interaction symbiotique qui ne repose pas sur la reconnaissance des facteurs ni sur la formation de cordon d'infection. Ces *Aeschynomene* sont également singulières car elles développent à la fois des nodules de racine et de tige avec des souches de *Bradyrhizobium* photosynthétiques. Malgré l'importance de ces traits symbiotiques pour appréhender la diversité des mécanismes de nodulation, ils restent peu étudiés. Pour combler ces lacunes, *A. evenia* a été choisi comme modèle d'étude génétique et moléculaire de la symbiose dite « Nod-indépendante ». En 2014, le financement du projet ANR AeschyNod a marqué de lancement d'un programme de recherche qui visait à mettre en œuvre le séquençage du génome d'*A. evenia*, donnant ainsi un accès direct aux gènes qu'il contient, et à développer une approche de génétique directe, pour identifier sans *a priori* des gènes importants pour la symbiose Nod-indépendante.

J'ai mis à profit le génome annoté d'*A. evenia* pour mener une analyse *in silico* des gènes symbiotiques initialement identifiés chez les légumineuses modèles, *M. truncatula* et *L. japonicus*. Ce travail a révélé des altérations (perte d'expression ou délétion) pour plusieurs gènes importants pour la reconnaissance du partenaire rhizobien et pour l'infection intracellulaire. De plus, il a mis en évidence des mécanismes différents chez *A. evenia* pour la différenciation et le fonctionnement du nodule, offrant de nouvelles pistes d'étude des étapes tardives de la nodulation. Suite au crible d'une population de mutagenèse EMS chez *A.* evenia, 250 mutants altérés pour la nodulation ont été isolés. Je me suis intéressé aux mutants bloqués dans les étapes les plus précoces de l'interaction car ils représentent les meilleurs candidats pour identifier des gènes de la signalisation Nod-indépendante. Pour cela, j'ai mené une caractérisation phénotypique et génétique de ces mutants et mis en oeuvre deux approches de re-séquençage, le Targeted Sequence Capture et le Mapping-by-Sequencing, qui ont abouti à l'identification de plusieurs gènes symbiotiques. Les résultats obtenus font l'objet d'un premier article en révision.

Article 1 : Genetics of nodulation in *Aeschynomene evenia* uncovers new mechanisms of the rhizobium-legume symbiosis

1.2. Article 1 :

Genetics of nodulation in *Aeschynomene evenia* uncovers new mechanisms of the rhizobium-legume symbiosis

Johan Quilbé[1], Léo Lamy[1,2], Laurent Brottier[1], Philippe Leleux[1,2], Joël Fardoux[1], Ronan Rivallan[3,4], Thomas Benichou[1], Rémi Guyonnet[1], Manuel Becana[5], Irene Villar[5], Olivier Garsmeur[3,4], Bárbara Hufnagel[6], Amandine Delteil[1], Djamel Gully[1], Clémence Chaintreuil[1], Marjorie Pervent[1], Fabienne Cartieaux[1], Mickaël Bourge[7], Nicolas Valentin[7], Guillaume Martin[3,4], Loïc Fontaine[8], Gaëtan Droc[3,4], Alexis Dereeper[9], Andrew Farmer[10], Cyril Libourel[11], Nico Nouwen[1], Frédéric Gressent[1], Pierre Mournet[3,4], Angélique D'Hont[3,4], Eric Giraud[1], Christophe Klopp[2] & Jean-François Arrighi[1*]

[1]IRD, Laboratoire des Symbioses Tropicales et Méditerranéennes (LSTM), UMR IRD/ SupAgro/INRAE/ UM2 /CIRAD, TA-A82/J – Campus de Baillarguet 34398 Montpellier cedex 5, France.

[2]Plateforme bioinformatique Genotoul, BioinfoMics, UR875 Biométrie et Intelligence Artificielle, INRAE, Castanet-Tolosan, France.

[3]CIRAD, UMR AGAP, Montpellier, France.

[4]AGAP, Université Montpellier, CIRAD, INRAE, Montpellier SupAgro, Montpellier, France.

[5]Departamento de Nutrición Vegetal, Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas, Apartado 13034, 50080 Zaragoza, Spain.

[6]BPMP, Univ Montpellier, CNRS, INRAE, SupAgro, Montpellier, France.

[7]Cytometry Facility, Imagerie-Gif, Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France.

[8]BGPI, Université de Montpellier, CIRAD, INRA, Montpellier SupAgro, F-34398 Montpellier, France.

[9]Institut de recherche pour le développement (IRD), University of Montpellier, DIADE, IPME, Montpellier, France.

[10]National Center for Genome Resources, Santa Fe, New Mexico, USA.

[11]LRSV, Université de Toulouse, CNRS, UPS, Castanet-Tolosan, France.

[*] email: jean-francois.arrighi@ird.fr

Abstract

Among legumes (Fabaceae) capable of nitrogen-fixing nodulation, several *Aeschynomene* spp. use a unique symbiotic process that is independent of Nod factors and infection threads. They are also distinctive in developing root and stem nodules with photosynthetic bradyrhizobia. Despite the significance of these symbiotic features, their understanding remains limited. To overcome such limitations, we conducted genetic studies of nodulation in *Aeschynomene evenia*, supported by the development of a genome sequence for *A. evenia* and transcriptomic resources for 10 additional *Aeschynomene* spp. Comparative analysis of symbiotic genes substantiated singular mechanisms in the early and late nodulation steps. A forward genetic screen also showed that *AeCRK*, coding a novel receptor-like kinase, and the symbiotic signaling genes *AePOLLUX*, *AeCCamK*, *AeCYCLOPS*, *AeNSP2* and *AeNIN*, are required to trigger both root and stem nodulation. This work demonstrates the utility of the *A. evenia* model and provides a cornerstone to unravel new mechanisms underlying the rhizobium-legume symbiosis.

Introduction

Legumes (Fabaceae) account for ~27% of the world's primary crop production and are an important protein source for human and animal diets. This agronomic success of legumes relies on the capacity of many species to establish a nitrogen-fixing symbiosis with soil bacteria, collectively known as rhizobia, forming root nodules[1]. Promoting cultivation of legumes and engineering nitrogen fixation in other crops will decrease the input of chemical nitrogen fertilizers and to will help to achieve short and long-term goals aimed at a more sustainable agriculture[2.]

Intensive research mainly performed on two temperate model legumes, *Medicago truncatula* and *Lotus japonicus*, has yielded significant information on the mechanisms controlling the establishment and functioning of the legume-rhizobium symbiosis[1]. In the general scheme, plant recognition of key rhizobial signal molecules, referred to as Nod factors, triggers a symbiotic signaling pathway leading to the development of an infection thread that guides bacteria inside the root and to the distant formation of a nodule meristem where bacteria are delivered and accommodated to fix nitrogen[1].

To further advance our understanding of the rhizobial symbiosis, there is a great interest in tracking the origin of nodulation[3,4] and in uncovering the whole range of symbiotic mechanisms[5,6]. In this quest, some semi-aquatic tropical *Aeschynomene* species constitute a unique symbiotic system because of their ability to be nodulated by photosynthetic bradyrhizobia that lack the canonical *nodABC* genes, necessary for Nod factor synthesis[7,8]. In this case, nodulation is not triggered by a hijacking Type-3 secretion system present in some non-photosynthetic bradyrhizobia[9,10]. Therefore, the interaction between photosynthetic bradyrhizobia and *Aeschynomene* represents a distinct symbiotic process in which nitrogen-fixing nodules are formed without the need of Nod factors. To unravel the molecular mechanisms behind the so-called Nod factor-independent symbiosis, *Aeschynomene evenia* (400 Mb, 2n=2*x*=20) has emerged as a new genetic model[11-13].

A. evenia is also an valuable legume species because: (i) it uses an alternative infection process mediated by intercellular penetration as is the case in 25% of legume species[14,15]; (ii) it is endowed with stem nodulation, a property shared with very few hydrophytic legume species[16,17]; and (iii) it groups with *Arachis* spp., including cultivated peanut (*Arachis hypogaea*) in the Dalbergioid clade, which is distantly related to *L. japonicus* and *M. truncatula*[11]. Previous transcriptomic analysis from root and nodule tissues did not detect expression of several known genes involved in bacterial recognition (e.g. *LYK3* and *EPR3*), infection (e.g. *RPG* and *FLOT*) and nodule functioning (e.g. *SUNERGOS1* and *VAG1)*[12,18]. Such data support the presence of new or divergent symbiotic mechanisms in *A. evenia* in comparison with other well-studied model legumes.

A. evenia is thus a system of prime interest to study the evolution and diversity of the rhizobial symbiosis. To efficiently conduct genetic studies of nodulation in *A. evenia*, we produced a genome sequence for this species along with *de novo* RNA-seq assemblies for 10 additional Nod factor-independent *Aeschynomene* spp. These genomic and transcriptomic data sets were then used for a comparative analysis of known symbiotic genes, leading to the evidence of singular symbiotic mechanisms in *Aeschynomene* spp. Finally, we used the available genome sequence in a forward genetic approach to perform the genetic dissection of nodulation in *A. evenia*. We successfully identified a novel receptor-like kinase, paving the way for unraveling of a new symbiotic pathway.

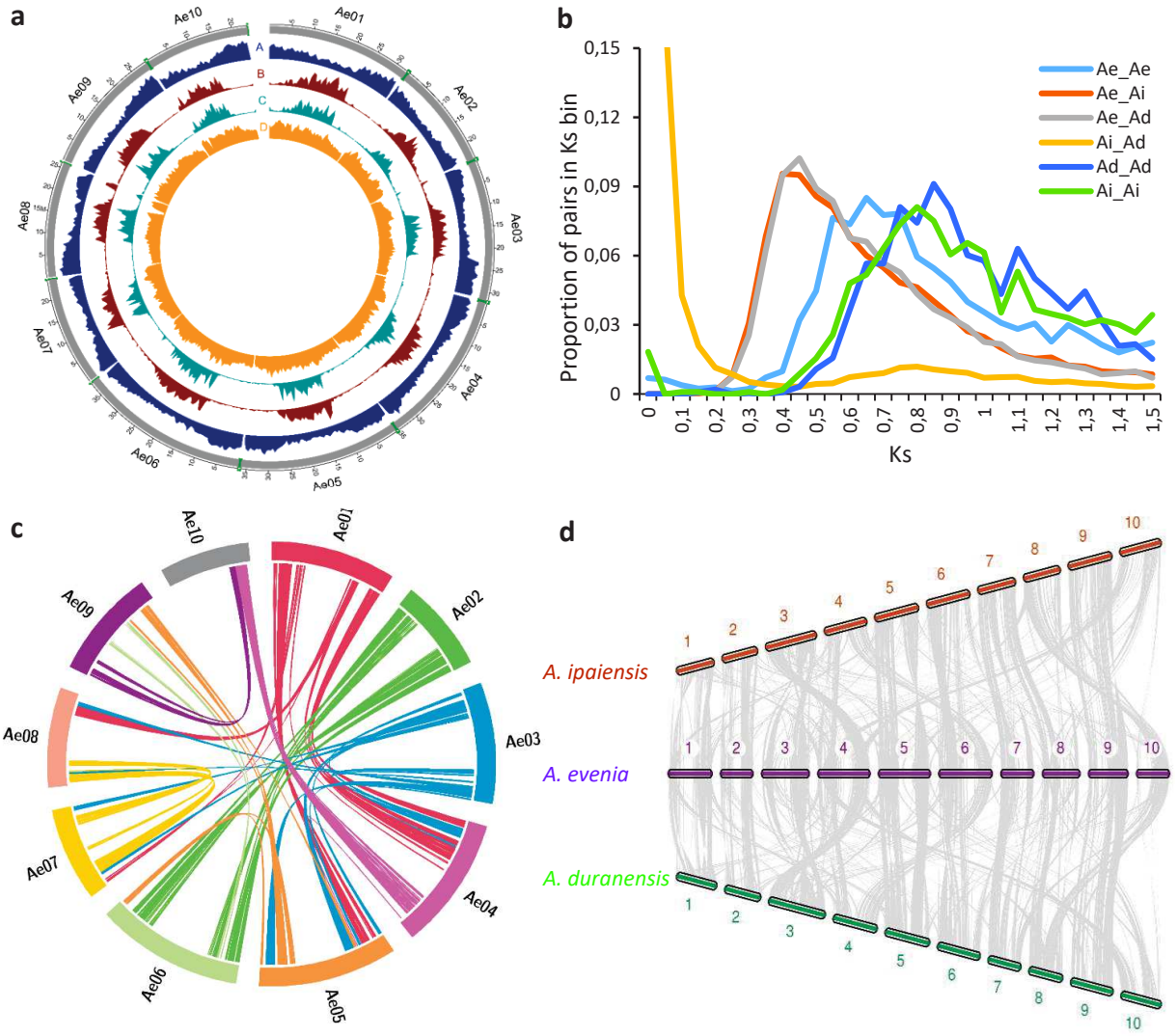


Fig. 1 Structure and evolution of the *Aeschynomene evenia* genome. a Distribution of genomic features along the chromosomes. The outer ring represents the 10 chromosomes with the captured telomeres in green (scale is in Mb). A, Gene density. B, density of transposable elements LTR/Copia. C, density of Gypsy transposable elements. D, Total SNP distribution. Densities are represented in 0.5 Mb bins. **b** Ks analysis of *A. evenia* (Ae) with the *Arachis* species, *A. duranensis* (Ad) and *A. ipaiensis* (Ai). Proportion of gene pairs per Ks range (with bin sizes of 0.05) for indicated species pairing. The shift of the WGD Ks peaks in Ae-Ae vs Ad-Ad and Ai-Ai is notable (0.65 vs 0.85 and 0.8), indicating more rapid accumulation of mutations in *Arachis* species than in *A. evenia*. **c** Syntenic regions in the *A. evenia* genome corresponding to intragenomic duplications. The colored lines are links between colinearity blocks that represent syntenic regions >1 Mb. **d** Syntenic relationships between *A. evenia* (center) and *Arachis sp.*, *A. ipaiensis* (upper) and *A. duranensis* (lower). The syntenic blocks >1 Mb in the *A. evenia* genome are shown. To facilitate comparisons, for *Arachis* species, chromosomes were scaled by factors calculated based on the genome size of *A. evenia*.

Results

A reference genome for the Nod factor-independent legume *Aeschynomene evenia*

As a support to forward genetic and comparative genetic studies of nodulation, a reference genome assembly was produced for *A. evenia* using the inbred CIAT22838 line[12]. To the single-molecule real-time (SMRT) sequencing technology from PacBio RSII platform was used to obtain a 78x genome coverage (Supplementary Tables S1-4). The resulting assembly was 376 Mb, representing 94% to 100% of the *A. evenia* genome, considering the estimated size of 400 Mb obtained by flow cytometry[12,16] or of 372 Mb derived from *k*-mer frequencies (Supplementary Fig. 1). PacBio scaffolds were integrated in the 10 linkage groups of *A. evenia* using an existing genetic map[12], an ultra-dense genetic map generated by Genotyping-by-Sequencing (GBS), and scaffold mapping was subsequently refined on the basis of synteny with *Arachis* spp[19] (Supplementary Fig. 2 and 3). The final 10 chromosomal pseudomolecules anchored 302 Mb (80%) of the genome (Supplementary Fig. 4, Supplementary Table 4). Protein-coding genes were annotated using a combination of *ab initio* prediction and transcript evidence gathered from RNA sequenced from nine tissues/developmental stages of nodulation using both RNA-sequencing (RNA-seq) and PacBio isoform sequencing (Iso-Seq) (Supplementary Tables 5 and 6). The current annotation contains 32,667 gene models (Fig. 1a, Supplementary Table 7). Their expression pattern was also determined by developing a Gene Atlas from the RNA-seq data obtained here (Supplementary Table 8) and from an earlier nodulation kinetics[18]. The identification of 94.4% of the 1,440 genes in the Plantae BUSCO dataset (Supplementary Table 9) confirmed the high quality of the genome assembly and annotation. Approximately 72% of the genes were assigned functional annotations using Swissprot, InterPro, GO, and KEGG (Supplementary Table 10). Additional annotation of the genome included the prediction of 6,558 non-coding RNAs (ncRNAs), the identification of repetitive elements accounting for 53.5% of the assembled genome and mainly represented by LTRs, the effective capture of 16 out of the 20 telomeric arrays, and the distribution of sequence variation along chromosomes based on the resequencing of 12 additional *A. evenia* accessions[20] (Fig. 1a, Supplementary Fig.4 and Supplementary Tables 11-15). Finally, all the resources were incorporated in the AeschynomeneBase (http://aeschynomenebase.fr), which includes a genome browser and user-friendly tools for molecular analyses.

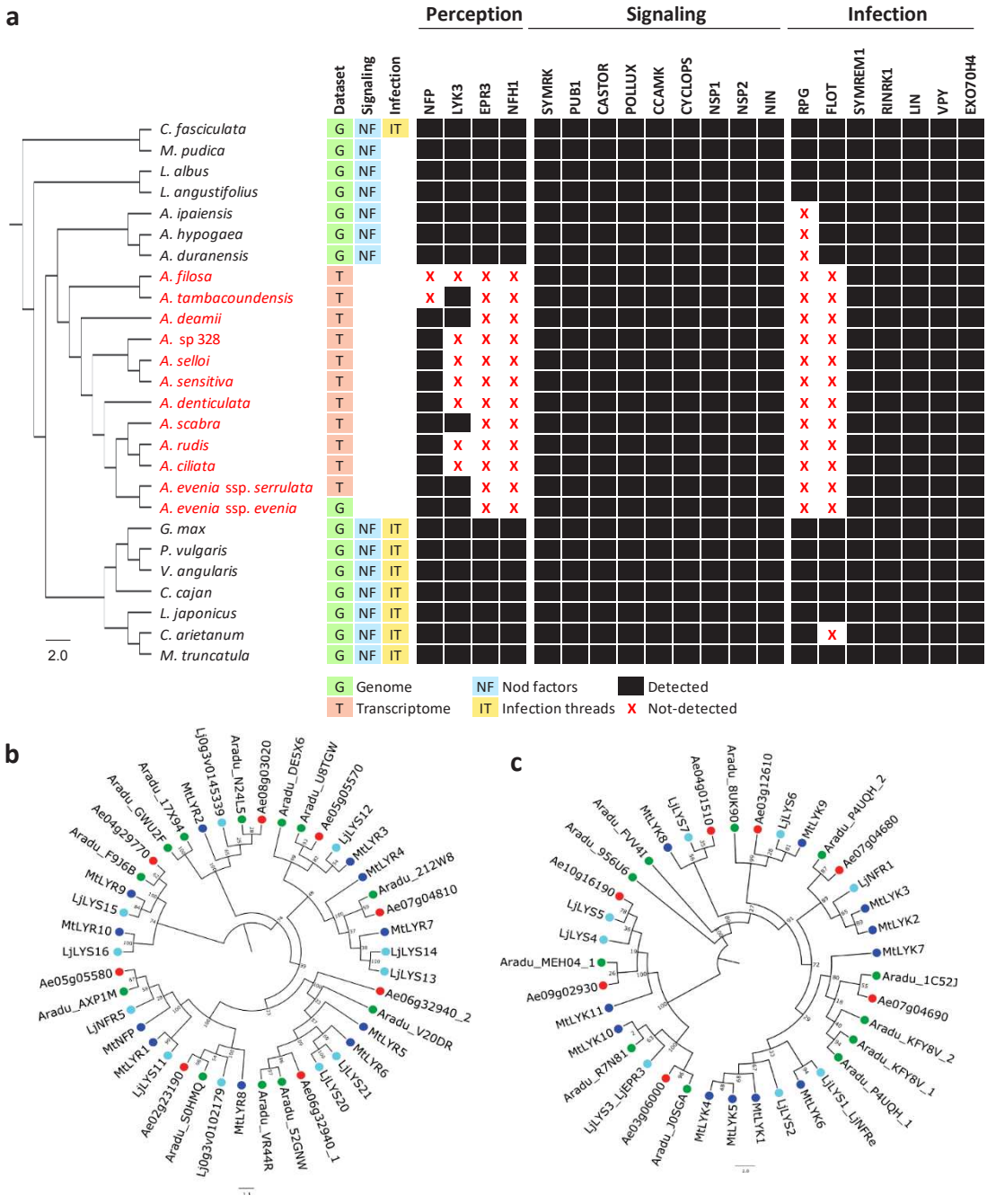


Fig. 2 Comparative analysis of symbiotic receptors, signaling, and infection genes. a Phylogenetic pattern of symbiotic genes involved in rhizobial perception, signaling and infection. The phylogenetic tree containing *Aeschynomene* species (in red), members of the main Papilionoid clades and two non-Papilionoid legume species, was obtained by global orthogroup analysis. All BS values (x1000) were comprised between 92 and 100% and so are not indicated for figure clarity. The presence and absence of genes are indicated in black or with a red cross, respectively. **b** and **c**, Phylogenetic analysis of the LysM-RLK gene family in *A. evenia* (red), *Arachis duranensis* (orange), *M. truncatula* (blue), and *Lotus japonicus* (green). **b** Phylogenetic tree of the *LYR* genes. **c** Phylogenetic tree of the *LYK* genes. Node numbers represent boostrap values (% of 1000 replicates). The scale bar represents substitutions per site.

To trace back the history of the *A. evenia* genome, it was compared to the genomes of *Arachis duranensis* and *Arachis ipaiensis*, which belong to the same Dalbergioid clade. *Aeschynomene* and *Arachis* lineages diverged ~49 MYA but are assumed to share an ancient whole genome duplication (WGD) event that occurred ~58 MYA at the basis of the Papilionoid legume subfamily[19, 21-23]. The shared WGD event, the *Aeschynomene-Arachis* divergence, and the *A. duranensis-A. ipaiensis* speciation were apparent in the synonymous substitutions in coding sequence (Ks) analysis between and within the *A. evenia-A. duranensis-A. ipaiensis* genomes (Fig. 1b). Modal Ks values are ~0.65 for *A. evenia*, i.e. more similar to those reported for *L. japonicus* and *G. max* (both ~0.65) than to those of *A. duranensis* (~0.85) and *A. ipaiensis* (~0.80) that were already reported to have evolved relatively rapidly[19]. In the case of *A. evenia*, it is worth noting that no more recent peak of Ks is visible, indicating it did not undergo any further WGD event. We identified paralogous *A. evenia* genes and orthologous *A. evenia-Arachis* spp. genes using synteny and Ks value criteria (Supplementary Data 1-3). This revealed the blocks of conserved collinear genes resulting from the ~58 MYA WGD event in the *A. evenia* genome (Fig. 1c). A comparison of *A. evenia* with *A. duranensis* and *A. ipaiensis* shows that extensive synteny remains prominent along chromosome arms despite multiple rearrangements (Fig. 1d). To be able to compare *A. evenia* to other *Aeschynomene* spp. that also use a Nod factor-independent process, we performed *de novo* RNA-seq assemblies from root and nodule tissues for 10 additional diploid *Aeschynomene* spp. (Supplementary Tables 16 and 17). Groups of orthologous genes for *A. evenia*, related *Aeschynomene* spp., and several species belonging to different legume clades were then generated using OrthoFinder (Supplementary Table 18). A consensus species tree inferred from single-copy orthogroups perfectly reflected the legume phylogeny and, for the *Aeschynomene* clade, the previously observed speciation with the early diverging species *Aeschynomene filosa*, *Aeschynomene tambacoundensis*, and *Aeschynomene deamii*, and a large group containing *A. evenia*[16,20] (Fig. 2a).

Symbiotic perception, signaling, and infection

In addition to their ability to nodulate in the absence of Nod factors[8,11], *A. evenia* and related *Aeschynomene* spp. use an infection process that is not mediated by the formation of infection threads[14]. This prompted us to perform a phylogenetic analysis of known symbiotic genes[1] based on the orthogroups containing *Aeschynomene* spp. and to exploit the Gene

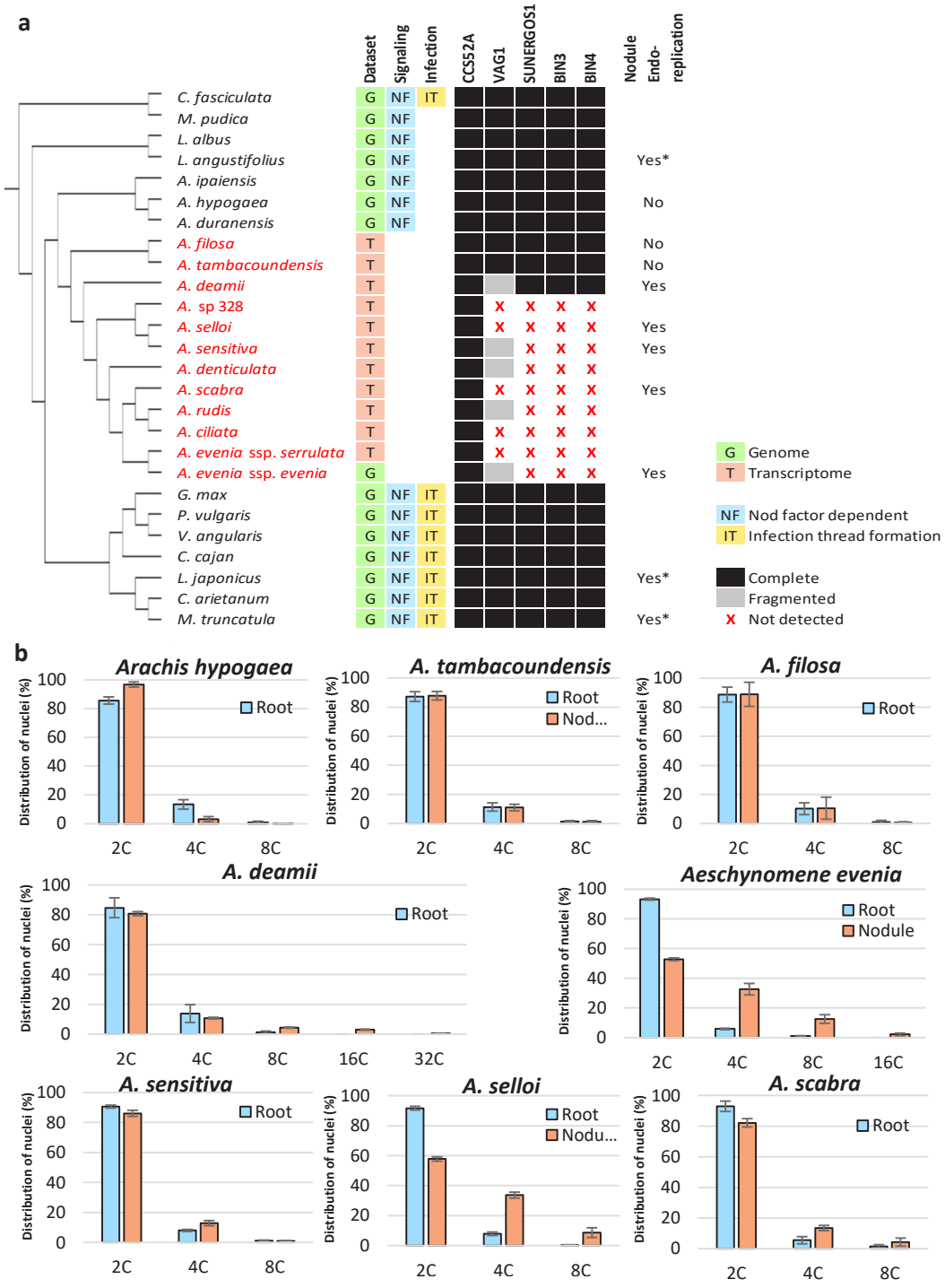
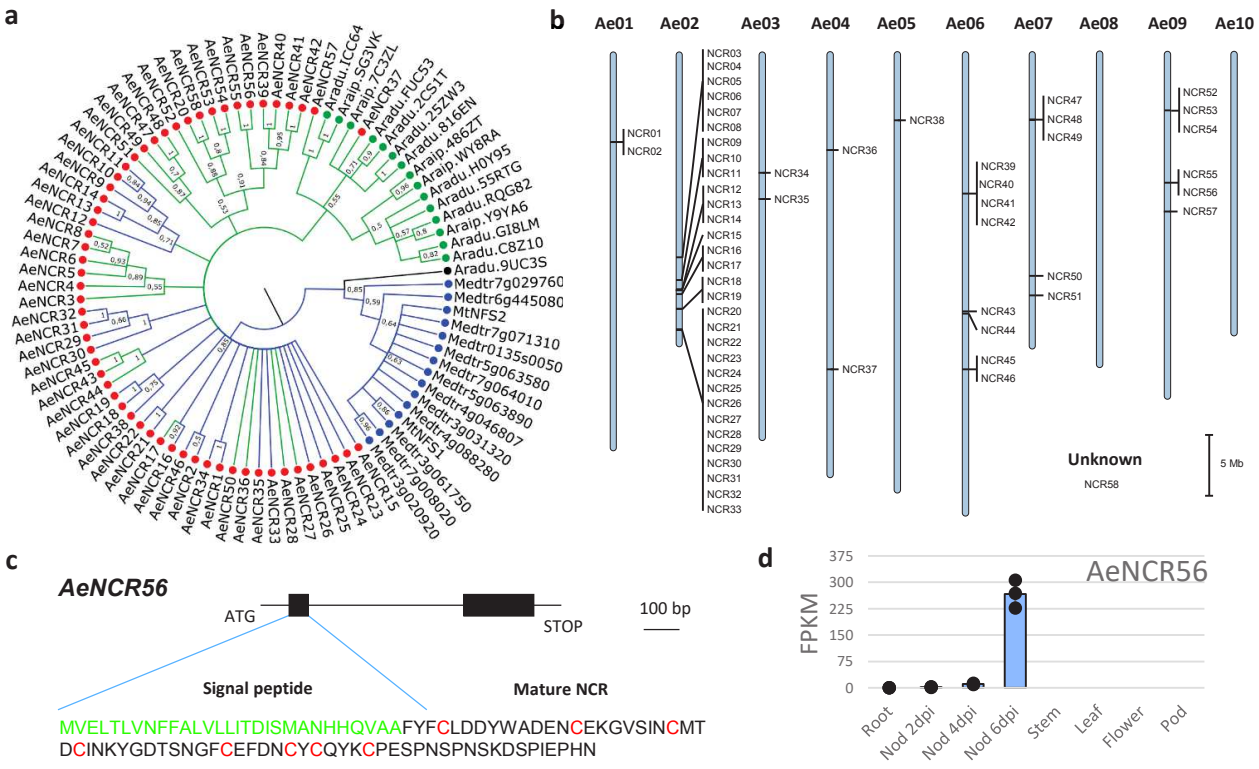


Fig. 3 Comparative analysis of endoreplication-mediated nodule differentiation. a Absence and presence of 5 genes involved in cell endoreplication during nodule differentiation, including the mitotic inhibitor CCS52A and components of the Topoisomerase VI complex, VAG1, SUNERGOS1, BIN3 and BIN4 in legume species. The ML tree containing *Aeschynomene* species (in red), members of the main Papilionoid clades and two non-Papilionoid legume species, was obtained by global orthogroup analysis. All BS values (x1000) were comprised between 92 and 100% and so are not indicated for figure clarity. For the observed occurrence of nodule cell endoreplication, asterisks indicate data from the literature, others come from the present study. **b** Flow cytometric histograms of *Arachis hypogaea* and of several *Aeschynomene* species obtained by measurement of nuclear DNA content in root and nodule cells. Bars indicate standard deviations.

Atlas developed for *A. evenia*. This comparative investigation revealed that the two genes encoding the Nod factor receptors, NFP and LYK3, are present but that *LYK3* is barely expressed in *A. evenia* (Supplementary Data 4). What is more, transcripts of both genes were not detected in the transcriptome of all other *Aeschynomene* species (Fig. 2a). In line with this observation, the gene coding for NFH1 (Nod Factor Hydrolase 1), which mediates Nod factor degradation in *M. truncatula*, was not found in any *Aeschynomene* data (Fig. 2a and Supplementary Data 5). Interestingly, *EPR3*, which inhibits infection of rhizobia with incompatible exopolysaccharides in *L. japonicus*, was not found in the *A. evenia* genome (Fig. 2a). Synteny analysis based on genomic sequence comparison with *A. duranensis* confirmed the complete deletion of *EPR3*, of genes within the *LYK* cluster containing *LYK3* and of the *NHF1* gene in *A. evenia* (Supplementary Fig. 7-9). Extending our analysis to the whole LysM-RLKs/RLPs gene family, to which *NFP*, *LYK3*, and *EPR3* belong, led to the identification of 18 members in *A. evenia* with 7 LYK, 8 LYR, and 3 LYM genes (according to the *M. truncatula* classification[24]) (Fig. 2b,c, Supplementary Fig. 6, Supplementary Data 4). No *Aeschynomene*-specific LysM-RLK genes were found; instead, several members present in other legumes were predicted to be missing in *A. evenia*.

Downstream of the Nod factor recognition step, genes of the symbiotic signaling pathway were identified in *A. evenia* and related *Aeschynomene* spp. (Fig. 2a; Supplementary Data 5). However, variations relative to model legumes were revealed by the detection of orthologs and paralogs probably resulting from the ancestral Papilionoid WGD. Notably, for the genes encoding the LRR-RLK receptor SYMRK and the E3 ubiquitin ligase PUB1, two copies are present, both showing nodulation-linked expression (Fig. 2a, Supplementary Fig. 10 and 11, Supplementary Data 5). It is worth noting that SYMRK and PUB1 are known to interact among with each other and with LYK3 in *M. truncatula*[1]. Considering that *AeLYK3* is probably not involved in Nod factor-independent symbiosis, it remains to be investigated how the presence of two copies of *AeSYMRK* and *AePUB1* in *A. evenia* might contribute to the diversification of the signaling mechanisms[25]. Downstream of SYMRK, the symbiotic signaling pathway leads to the triggering of the plant-mediated rhizobial infection. Determinants such as *VPY*, *LIN*, and *EXO70H4*[26], which are required both for polar growth of infection threads and subsequent intracellular accommodation of symbionts in *M. truncatula*[1], have symbiotic expression in *A. evenia* (Supplementary Data 5). This expression pattern is probably linked to the later symbiotic process since rhizobial



e

Aeschynomene NCR motif 1	**SP-X_n-C-X_n-C-X_3-C-X_n-C-X_n-C-X_1-C-X_n**
Medicago NCR motif a	**SP-X_n-C-X_5-C-X_n-C-X_n-C-X_4-C-X_1-C-X_n**
Aeschynomene NCR motif 2	**SP-X_n-C-X_n-C-X_n-C-X_3-C-X_n-C-X_n-C-X_1-C-X_3-C-X_n**
Defensin-like motif	**SP-X_n-C-X_n-C-X_n-C-X_3-C-X_n-C-X_n-C-X_1-C-X_3-C-X_n**

Figure 4 *NCR* genes in the *A. evenia* genome. a Bayesian phylogenetic reconstruction of relationships between *NCR* genes identified in the genomes of *A. evenia* (red), *A. duranensis* and *A. ipaiensis* (green) and with a few members of *M. truncatula* (blue). Branches in blue correspond to NCRs with the cysteine-rich motif 1 and branches in green correspond to NCRs with the cysteine-rich motif 2. Node numbers indicate posterior probabilities. The scale bar represents substitutions per site. **b** Genome scale organization of *NCR* genes in *A. evenia* visualized with the SpiderMap tool. Vertical bars indicate gene clusters. **c** Typical structure of an NCR gene in *A. evenia* as exemplified with *AeNCR56.* Black boxes represent exons, the first one coding for the signal peptide (green) and the second one for the mature NCR (with conserved cysteines in red). **d** Expression pattern of *AeNCR56* in *A. evenia* aerial organs, in roots and in nodules (Nod) after 2, 4 and 6 days post-inoculation (dpi) with the *Bradyrhizobium* strain ORS278. Expression is given in normalized FPKM read counts. For root and nodule samples, bars are SE of three biological replicates and dots represent individual expression levels. **e** Cysteine-rich motifs 1 and 2, Medicago NCR and defensin structures as previously presented[33]. SP, Signal Peptide, Xn, length of conserved spacings between cysteines. In red, conserved cysteines in the motif 1, in green, additional cysteines found in motif 2 and shared with the defensin signature.

invasion occurs in an intercellular manner in *A. evenia*[14]. In contrast, other key infection genes[1] are expressed at very low levels, as is the case of *NPL* and CBS1, or absent: *RPG* was undetectable in Dalbergioid legume species and *FLOT* genes were completely missing in *Aeschynomene* spp., substantiating mechanistic differences in the infection process (Fig. 2a, Supplementary Fig. 12; Supplementary Data 5).

Nodule development and bacterial accommodation

During nodule development, differentiating plant cells undergo endoreplication leading to an increase in ploidy levels and cell size. The mitotic inhibitor CCS52A, a key mediator of this nodule development process[27,28], is conserved in all *Aeschynomene* spp. (Fig. 3a). However, earlier transcriptomic studies[12,18] failed to detect two genes coding for components of the DNA topoisomerase VI complex, subunit A (SUNERGOS1) and an interactor (VAG1). In *L. japonicus*, these two genes are required for cell endoreplication during nodule formation[29,30]. From previous Arabidopsis studies, the DNA topoisomerase VI is known to contains two other components, the subunit B (BIN3)[31] and a second interactor (BIN4)[32], which were both successfully identified in legumes but not in *A. evenia* (Fig. 3a). Synteny analysis based on genomic sequence comparison with *Arachis* spp. substantiated the specific and complete loss of *SUNERGOS1*, *BIN3*, and *BIN4*, and the partial deletion of *VAG1* in *A. evenia* (Supplementary Fig. 13). A similar pattern could be observed for most *Aeschynomene* spp. However, *SUNERGOS1*, *BIN3*, *BIN4*, and *VAG1* with a distinct truncation were detected in *A. deamii* and the full gene set was present in *A. filosa* and *A. tambacoundensis* as is the case for peanut (*Arachis hypogaea*), indicating that these gene losses are disconnected from the Nod factor-independent character (Fig. 3a, Supplementary Fig. 14). To link these different gene patterns with variations in nodule cell endoreplication, roots and nodules of several species were analyzed by flow cytometry. Contrary to our expectations, whereas no difference in ploidy levels was observed in *A. filosa*, *A. tambacoundensis* or peanut, we discovered higher ploidy levels in nodule cells than in root cells of *A. deamii*, *A. evenia*, *A. scabra*, *A. selloi*, and *A. sensitiva* (Fig. 3b). Taken together, these data reveal a case of gene co-elimination affecting a symbiotic process and it offers the opportunity to investigate how the nodule cell endoreplication process is mediated in the absence of the Topoisomerase VI complex in certain *Aeschynomene* spp.

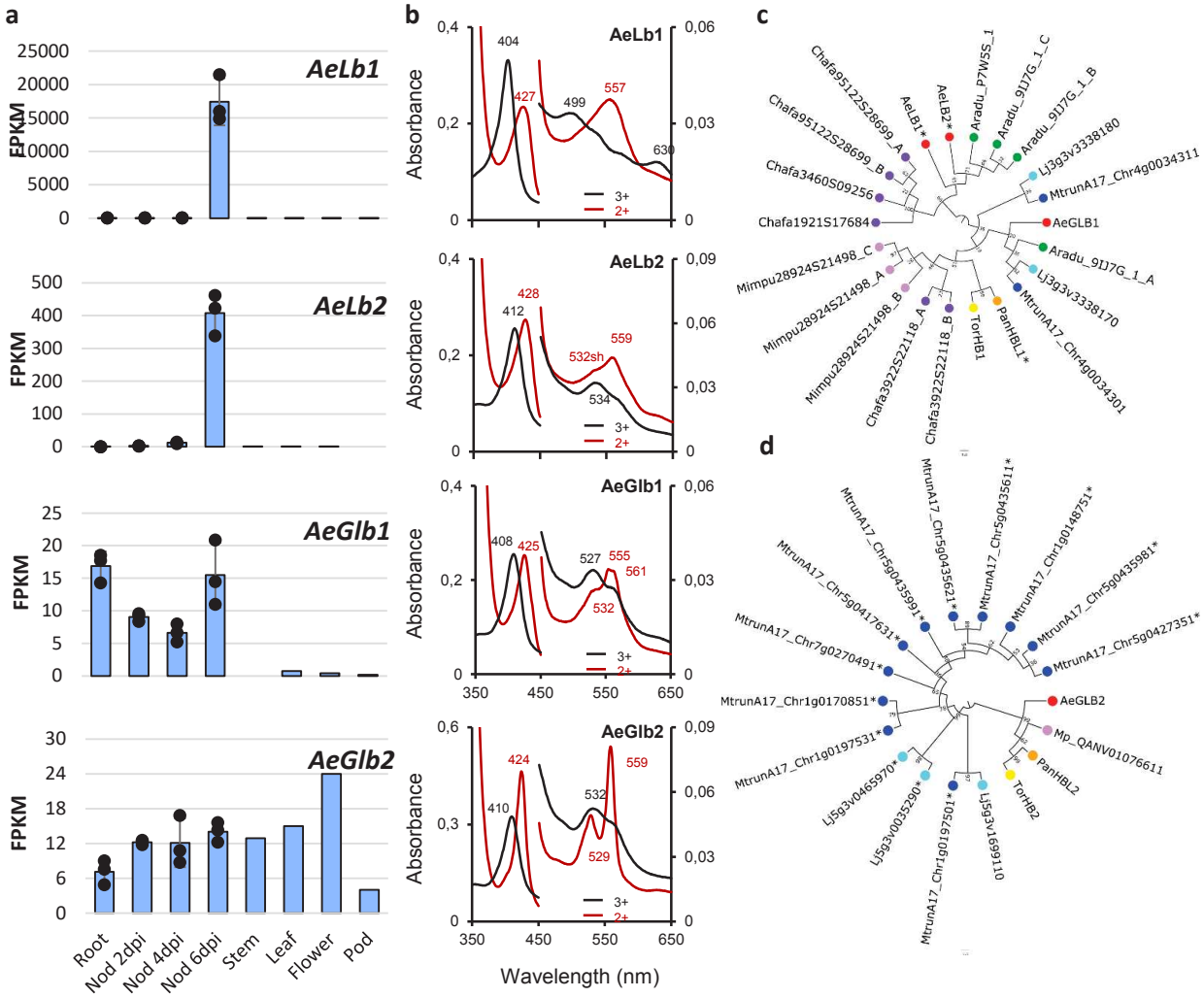


Fig. 5 Symbiotic and nonsymbiotic globins of *A. evenia*. a Expression profiles of *A. evenia* globin genes in aerial organs, roots and nodules (Nod) after 2, 4, and 6 days post-inoculation (dpi) with the *Bradyrhizobium* strain ORS278. Expression is given in normalized FPKM read counts. For root and nodule samples, bars are SE of three biological replicates and dots represent individual expression levels. **b** UV-visible spectra of *A. evenia* globins in the ferric (black) and ferrous (red) form. **c** and **d** Phylogenetic reconstructions of relationships between Lb, class 1 and class 2 globin genes identified in *A. evenia* (Ae), *A. duranensis* (Aradu), *M. truncatula* (Mtrun), *L. japonicus* (Lj), *C. fasciculata* (Chafa), *M. pudica* (Mimpu), and the non-legumes *P. andersonii* (Pan) and *T. orientalis* (Tor). Node numbers represent boostrap values (% of 100 replicates). The scale bar represents substitutions per site. Lbs are marked with an asterisk.

Nodule formation is also accompanied by the differentiation of nodule cell-endocyted rhizobia into nitrogen-fixing bacteroids. In *M. truncatula*, this differentiation is mediated by the expression of a wide set of plant genes coding for nodule specific cysteine-rich peptides (NCRs)[1]. Although NCRs were long thought to be restricted to the IRLC clade to which *M. truncatula* belongs[33], *A. evenia* and other *Aeschynomene* spp. were recently shown to express *NCR*-like genes[34]. We identified 58 such genes in the *A. evenia* genome (Fig. 4a). The *AeNCR* genes are mainly organized in clusters (Fig. 4b) and they are typically composed of two exons encoding the signal peptide and the mature NCR (Fig. 4c). Most *NCR* genes display prominent nodule-induced expression in *A. evenia* that correlates with the onset of bacteroid differentiation (Fig. 4d, Supplementary Data 6). All predicted NCRs contain one of the two previously described cysteine-rich motifs[34,35] (Fig. 4e). Thus, 26 NCRs of *A. evenia* harbor the cysteine-rich motif 1 similar to *M. truncatula* NCRs while 32 NCRs of *A. evenia* have the defensin-like motif 2 (Fig. 4a, Supplementary Fig. 15 and 16). In *A. duranensis* and *A. ipaiensis*, no NCRs with the cysteine-rich motif 1 could be found, whereas 10 and 5 *NCR*-like genes, respectively, with the defensin-like motif 2 were identified (Fig. 4a, Supplementary Fig. 15 and 16) and the expression of most of them is induced in the nodule (LegumeMines database). These features of Dalbergioid NCRs raise questions as to how they emerged and whether they evolved to symbiotic or defense functions.

Nodule functioning involves leghemoglobins derived from class 1 phytoglobins

In nitrogen-fixing nodules, maintaining a low but stable oxygen concentration is crucial to protect the nitrogenase complex. To ensure this function, legumes have recruited leghemoglobins (Lbs) that evolutionary derive from non-symbiotic hemoglobins (now termed phytoglobins; Glbs), and that occur at high concentrations in nodules[36]. We found six globin genes in the *A. evenia* genome (Supplementary Data 7). Two of them are homologous to class 3 Glb genes and were not studied further. Two genes show moderate expression, have homology to class 1 and class 2 Glb genes, and were accordingly designated *AeGlb1* and *AeGlb2* (Fig. 5a). The two other globin genes were found to be highly and almost exclusively expressed in nodules (Fig. 5a). This observation suggested that they encode Lbs and were termed *AeLb1* and *AeLb2*. To unequivocally classify the four proteins, they were purified and characterized for heme coordination (Fig. 5b). Both AeGlb1 and AeGlb2 showed hexacoordination in the ferric and ferrous forms, confirming that they correspond to class 1

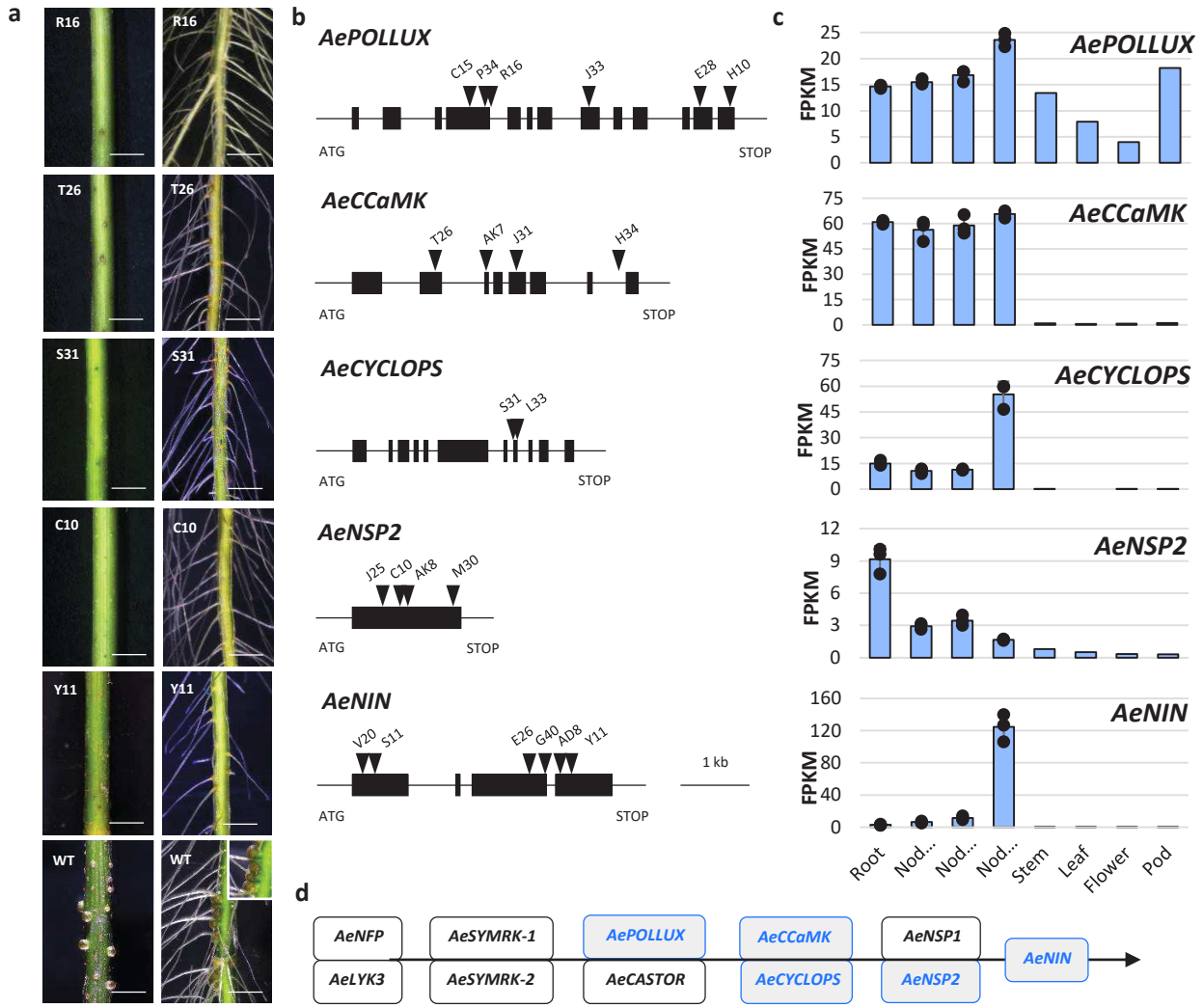


Fig. 6 Candidate genes of the known symbiotic signaling pathway identified by Targeted Sequence Capture. a Nodulation phenotypes observed on stem (left panel) and root (right panel) in EMS mutant plants and the WT line (bottom panels with inset corresponding to a zoom on root nodules). Scale bars: 5 mm. **b** Structure of the different symbiotic genes showing the position of the EMS mutations. Black boxes depict exons, lines represent untranslated regions and introns, and triangles represent mutation sites with the name of the corresponding mutant indicated above. **c** Expression profiles in *A. evenia* aerial organs, in roots and in nodules (Nod) after 2, 4 and 6 days post-inoculation (dpi) with the *Bradyrhizobium* strain ORS278. Expression is given in normalized FPKM read counts. For root and nodule samples, bars are SE of three biological replicates and dots represent individual expression levels. **d** Representation of the Symbiotic Signaling Pathway as inferred from model legumes. Genes in blue are those demonstrated as being involved in the NF-independent signaling in *A. evenia* using the mutant approach.

and class 2 Glbs, respectively. AeLb1 shows complete pentacoordination in both ferric and ferrous form, whereas AeLb2 is hexacoordinate in the ferric form and almost fully pentacoordinate in the ferrous form. AeLb1 is therefore a typical Lb but AeLb2 appears to be an unusual one. All four globins were found to bind the physiologically-relevant ligands, O_2 and nitric oxide (Supplementary Fig. 17). However, the unexpected discovery was that both AeLb1 and AeLb2 cluster with class 1 Glbs and not with class 2 globins, as observed for other legumes[36] (Fig. 5c and 5d). In the globin phylogeny, *AeLb1* and *AeLb2* cluster tightly with certain class 1 Glb genes of *Arachis* and also of the more distantly related legume *Chamaecrista fasciculata*. The *Arachis* genes are also highly expressed in nodules (LegumeMines database) and probably encode Lbs. Among the *C. fasciculata* genes, one was previously evidenced to be highly expressed in root nodules and to code for a putative ancestral Lb named *ppHB*[37] (corresponds to the Chafa1921S17684 gene in Fig. 5c). Sequence and synteny analysis further indicated that *A. evenia* Lbs and class 1 Glb genes are similar and located in a single locus that is conserved in legumes (*SI Appendix,* Fig. S18-S20). This supports the hypothesis that *A.evenia* Lbs arose from class 1 Glbs by local gene duplication and the presence of probably such Lbs in *Arachis* and *Chamaecrista* legumes, further suggests this evolution to be ancient. The finding of Lbs originating from a class 1 Glb challenges our view on the evolution of Lbs in legumes and is only comparable to panHBL1, the symbiotic Glb1 of the non-legume *Parasponia*[3]. However, panHBL1 appears to be different from *A. evenia* Lbs (Fig. 5c). These Lbs thus offer a valuable new case to study the convergent evolution of O_2-transporting Lbs.

Genetic dissection of root and stem nodulation

To uncover genes underpinning the singular symbiotic traits evidenced in *A. evenia*, a large-scale forward genetic screen was undertaken by performing ethyl methane sulfonate (EMS) mutagenesis (Supplementary Table 19). Treating 9,000 seeds with 0.3% EMS allowed us to develop a mutagenized population of 70,000 M_2 plants that were subsequently screened for plants with altered root nodulation (Supplementary Fig. 21). Finally, 250 symbiotic mutants were isolated and sorted into distinct phenotypic categories: [Nod^-] for complete absence of nodulation, [Nod^{-*}] for occasional nodule formation, [Inf^-] for defects in infection, [Fix^-] for defects in nitrogen fixation and [Nod^{++}] for excessive numbers of nodules. The collection of mutants was subjected to Targeted Sequence Capture on a set of selected genes with a

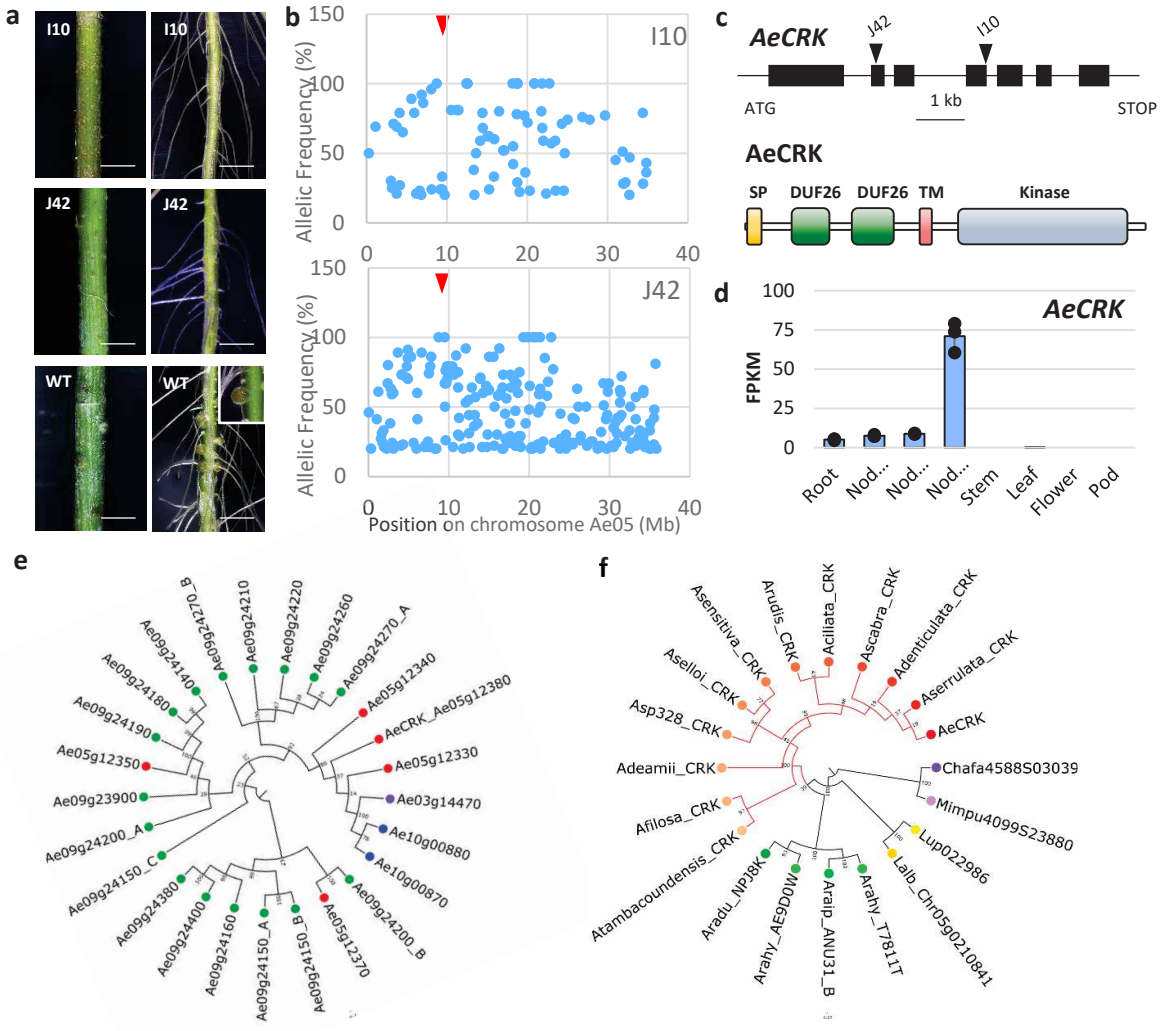


Fig. 7 Novel gene involved in the establishment of the NF-independent symbiosis identified by Mapping-by-Sequencing. a Nodulation phenotypes observed on stem (left) and root (right) in EMS mutant plants and the WT line. Scale bar: 5 mm. **b** Frequency of the EMS-induced mutant alleles in bulks of Nod$^-$ backcrossed F2 plants obtained for the I10 and J42 mutants by Mapping-by-Sequencing. The SNPs representing the putative causal mutations are indicated by the red arrow head. **c** *AeCRK* gene and protein structure. Upper panel, the *AeCRK* gene exons are indicated by black boxes and the position of the EMS mutations indicated by triangles. Lower panel, the predicted AeCRK protein contains a signal peptide (SP), followed by two cysteine-rich domains of unknown function (DUF26), a transmembrane domain (TM) and a kinase domain. **d** *AeCRK* expression pattern in *A. evenia* aerial organs, root and during nodule (Nod) development after inoculation (dpi) with the *Bradyrhizobium* strain ORS278. Expression is given in normalized FPKM read counts. For root and nodule samples, bars are SE of three biological replicates and dots represent individual expression levels. **e** Phylogenetic tree of the *CRK* gene family in *A. evenia*. In total, 25 CRK genes were identified and found to be located either as a singleton on the Ae03 chromosome (purple), in tandem on the Ae10 chromosome (blue) or in clusters on the Ae05 and Ae09 chromosomes (red and green, respectively). **f** Phylogenetic tree of *AeCRK* orthologous genes present in *Aeschynomene* species (*A. ciliata*, *A. deamii*, *A. denticulata*, *A. evenia* var. *evenia* and var. *serrulata*, *A. filosa*, *A. rudis*, *A. scabra*, *A. selloi*, *A. sensitiva*, *A.* sp 328 and *A. tambacoundensis*), *Arachis* species (*A. duranensis*, *A. hypogaea* and *A. ipaiensis*), *Lupinus* species (*L. albus* and *L. angustifolius*), *Chamaecrista fasciculata* and *Mimosa pudica*. The *Aeschynomene* lineage (red) is characterized by a negative selection evidenced in the extracellular domain of *AeCRK*. **e** and **f** Node numbers represent boostrap values (% of 1000 replicates). The scale bar represents substitutions per site.

potential symbiotic role. Analysis of EMS-induced SNPs allowed the filtering of siblings originating from the same screening bulks and led to the identification of candidate mutations.

We decided to focus our genetic work on the Nod$^-$ mutants since they are most probably altered in genes controlling the early steps of nodulation. Moreover, they provide an opportunity to test the role of these genes in stem nodulation whose genetic control is completely unknown so far. For this, Nod$^-$ mutants were backcrossed to the WT line and segregating F_2 progenies were phenotyped for root and stem nodulation after sequential inoculation. These analyses always pointed to a single recessive gene controlling both root and caulinar nodulation (Fig. 6a, Supplementary Table 22). For Nod$^-$ mutants associated with a candidate mutation, the mutations were validated as being causative by genotyping F_2 backcrossed mutant plants and by performing targeted allelism tests. This produced allelic mutant series for five genes of the symbiotic signaling pathway[1]: *AePOLLUX* (6 alleles), *AeCCaMK* (4 alleles), *AeCYCLOPS* (2 alleles), *AeNSP2* (4 alleles), and *AeNIN* (6 alleles) (Fig. 6b, Supplementary Tables 20 and 21). Among these signaling genes, *AePOLLUX* was found to be consistently expressed in all plant organs whereas the other genes are expressed only in symbiotic organs. *AeCCaMK* is constantly expressed in roots and in all stages of nodule development, *AeCYCLOPS* and *AeNIN* are induced during nodulation, and *AeNSP2* is down-regulated during nodulation (Fig. 6c). Thus, mutant analysis revealed that the signaling pathway, described in *M. truncatula* and *L. japonicus*, is partially conserved in *A. evenia* and is necessary for stem nodulation. However, not all known signaling genes were evidenced with the mutant approach (Fig. 6d). In particular, no consistent mutation was found in any member of the LysM-RLK family. Although it cannot be excluded that our mutagenesis was not saturating, this observation again supports the lack of a key role for LysM-RLKs in the early steps of the symbiotic interaction in *A. evenia*. Neither was a causative mutation found for the two paralogs of *SYMRK* in *A. evenia*. In an earlier study, we used RNAi to target *AeSYMRK* (actually *AeSYMRK2*), which reduced the number of nodules[13]. Because *AeSYMRK1* and *AeSYMRK2* are 82% identical in the 296-pb RNAi target region, they were probably both targeted. The functioning of the two receptors during nodulation remains to be investigated.

A novel receptor-like kinase to mediate the symbiotic interaction

Two Nod$^-$ mutants, defective in both root and stem nodulation, were not associated with any known genes and were consequently good candidates to uncover novel symbiotic functions (Fig. 7a). To identify the underlying symbiotic gene, we used a Mapping-by-Sequencing approach on bulks of F_2 mutant backcrossed plants. Linkage mapping for each mutant population identified the same locus on chromosome Ae05, where mutant alleles frequencies reached 100% (Fig. 7b). Analysis of the region containing the symbiotic locus identified mutations in a gene that encodes a Cysteine-rich Receptor-like Kinase (CRK)[38], henceforth named *AeCRK* (Supplementary, Table 22). The predicted 658 aa-long protein harbors a signal peptide, two extracellular DUF26 domains, a transmembrane domain, and an intracellular serine/threonine kinase domain (Fig. 7c, Supplementary Fig. 22). In the mutated forms, the G2228A SNP alters a canonical intron/exon splice boundary probably generating a truncated protein while the G1062A SNP leads to the replacement of G354E in the highly conserved glycine-rich loop of the kinase domain (*SI Appendix,* Table S20). Allelism tests performed with the two Nod$^-$ mutant lines (I10 and J42) indicated that they belong to the same complementation group (Supplementary Table 21). Hairy root transformation of the I10 mutant with the coding sequence of *AeCRK*, fused to its native promoter, resulted in the development of nodules upon inoculation with the *Bradyrhizobium* ORS278 strain, while no nodules were produced in control plants transformed with the empty vector (Supplementary Fig. 23, Supplementary Table 23). The identification of genetic lesions in the two independent *Aecrk* alleles together with the transgenic complementation of the mutant phenotype provide unequivocal evidence that *AeCRK* is required for the establishment of symbiosis *A. evenia*.

AeCRK was found to be expressed in roots with significant up-regulation in nodules, in agreement with its symbiotic function (Fig. 7d). Notably, *AeCRK* is part of a cluster of five *CRK* genes in *A. evenia* but genes of this cluster are interspersed within the CRK phylogeny (Fig. 7e). Although similar *CRK* clusters are located in syntenic regions in other legumes, no putative ortholog to *AeCRK* could be found in *M. truncatula* or *L. japonicus*, and actually in no legume using a root hair- and infection-thread mediated infection process (Fig. 7f, Supplementary Fig. 24, Supplementary Data 8). To gain further insights into the molecular evolution of *AeCRK*, we ran branch model by estimating different synonymous and

nonsynonymous substitution rates ($\omega=\frac{dN}{dS}$) using the phylogenetic tree topology. These analyses, performed on the entire gene sequence and on the four functional domains of *AeCRK* orthologs separately (signal peptide, extracellular, transmembrane and kinase domains), revealed a higher purifying/negative selection acting on the extracellular domain part in the *Aeschynomene* clade (ω_{BG} = 0.480 and ω_{FG} = 0.187, *p-val*=0.017214) (Fig. 7f , Supplementary Table 24). This purifying selection suggests that *AeCRK* could have evolved to adapt nodulation with *nod* gene-lacking photosynthetic bradyrhizobia. These data support that *AeCRK* is a key component of the pathway used by *A. evenia* to trigger symbiosis in the absence of Nod factors and infection threads.

Discussion

A. evenia and a handful of other *Aeschynomene* spp. have gained renown for triggering efficient nodulation without recognition of rhizobial Nod factors nor infection thread formation[8,9,14]. To accelerate the deciphering of this original symbiosis, we conducted in *A. evenia* forward genetics based on an EMS mutagenesis and developed a reference genome sequence to enable re-sequencing strategies of nodulation mutants. This work leads to the demonstration that the triggering of nodulation in *A. evenia* is mediated by several components of the Nod signaling pathway described in model legumes, *AePOLLUX*, *AeCCaMK*, *AeCYCLOPS*, *AeNSP2*, and *AeNIN*, thus significantly extending a previous report of the involvement of *AeSYMRK*, *AeCCaMK*, and *AeLHK1* genes in root nodulation[13]. The present study also reveals that this symbiotic signaling pathway controls not only root but also stem nodulation in *A. evenia*. This dual nodulation is present in few half-aquatic legume species[16,17] such as *Aeschynomene* spp. and *S. rostrata* but the genetics of stem nodulation has remained unknown so far. With the forward genetic screen, not all known genes of the Nod signaling pathway were recovered. Indeed, no causative mutation could be found in *AeCASTOR* or in *AeNSP1*, whereas CASTOR and NSP1 are known to act in concert with POLLUX and NSP2, respectively[1]. In addition, there are no obvious paralogs reported that may function redundantly, as it is probably the case for SYMRK in *A. evenia*. Therefore, either both these genes were unfortunately not targeted by the EMS mutagenesis or a

special evolution of *AePOLLUX* and *AeNSP2* rendered them sufficient for symbiosis as evidenced for DMI1/POLLUX in *M. truncatula*[39]. Also striking is the failure of the mutant approach to demonstrate the involvement of any LysM-RLK member, most notably the Nod factor receptors. In agreement with this observation, *LYK3* is not expressed in *A. evenia*. Conversely, *NFP* remains expressed in *A. evenia*, putatively because it has been co-opted from an ancestral role in the arbuscular mycorrhizal (AM) symbiosis in legumes[40]. Therefore, a comparative genetic analysis of *NFP* and *LYK3* between *A. evenia* and *Aeschynomene patula*, which displays a Nod factor-dependent nodulation and which was recently selected as a suitable complementary model[16], should illuminate their recent evolution and clarify if *NFP* has any role in *A. evenia*.

Finding that the core Nod signaling pathway, but not the upstream Nod factor receptors, is conserved in *A. evenia* suggests that one main difference with other legumes comes from the symbiotic receptor plugged-in the pathway[41]. In line with this idea, a novel receptor-like-kinase belonging to the large CRK family[38] was discovered as being required to trigger nodulation in *A. evenia*. In the legume phylogeny, this gene is present only in Papilionoid lineages using an intercellular infection process and also in the more primitive Caesalpinioid legumes, *Chamaecrista fasciculata* and *Mimosa pudica*, for which infection does not involve root hair invasion either[15]. Such distribution of the *AeCRK* orthologs suggests it could be linked to intercellular infection. Molecular evolutionary analysis further evidenced the extracellular domain of AeCRK orthologs to be under purifying selection in the *Aeschynomene* clade, arguing for a particular evolution with the Nod factor-independent symbiosis. CRKs are repeatedly pointed as important actors of plant early signaling during immunity and abiotic stress[42,43]. They are supposed to be mediators of reactive oxygen species (ROS)/redox sensing through their DUF26 extracellular domains and to transduce the signal intracellularly via their cytoplasmic kinase[43]. Another putative function of their DUF26 domains was recently proposed, based on strong similarity to fungal lectins, as mediating carbohydrate recognition[44]. Therefore, characterization of AeCRK will be crucial to provide information on pending questions: has AeCRK retained the ancestral function or has it been neofunctionalized? Is AeCRK involved in the direct perception of photosynthetic bradyrhizobia or does it mediate ROS/redox sensing during early signaling/infection? Is the Nod factor-independent activation inherently linked to intercellular infection? This could be probably the case since genetic studies in *L. japonicus* evidenced that double mutant lines

were occasionally able to develop nitrogen-fixing nodules in a Nod factor- and infection thread-independent fashion[45]. Additionally, the ability of *L. japonicus* to be infected intracellularly or intercellularly, depending on the rhizobial partner, was recently used to provide insights into the genetic requirements of intercellular infection[46]. It was showed that some determinants required for the infection thread-mediated infection are dispensable for intercellular infection, among which RPG is found. This finding echoes the observed absence of RPG in *A. evenia* and other Dalbergioid legumes for which intercellular infection is the rule. However, other infection determinants (LIN, VPY, EXO70H4 and SYN) that are also involved in intracellular accommodation of symbionts are present in *A. evenia*, suggesting that both the core symbiotic signaling pathway and the machinery mediating intracellular accommodation are conserved, as a general feature of endosymbioses[47]. Continuing the mutant-based gene identification in *A. evenia* is will increase our knowledge on the mechanisms of the as yet under-explored intercellular infection process.

In addition to the intercellular infection process, several symbiotic features present in *A. evenia* are shared with other legumes, including peanut, for which the molecular basis of nodulation is subject of recent investigations[48]. As evidenced previously[34] and in the present work, *Aeschynomene* and *Arachis* spp. express NCR-like genes during bacterial accommodation, in a similar fashion to IRLC legumes, but their symbiotic involvement remains to be clarified. Most remarkable is the discovery that *Aeschynomene* and *Arachis* spp. have recruited some class 1 Glbs as Lbs transporting O_2 in nodule infected cells. Indeed, it is well established that in legumes some class 2 Glbs have evolved to Lbs to ensure such a crucial function[36], but the Dalbergioid lineage appears to be an exception to this pattern of Lb utilization. Comparative genomic analysis in Papilionoid legumes revealed a striking parallel with the presence of two conserved loci where both Glb and Lb genes belonging to class 1 and class 2, respectively, can be found across species. It is therefore tempting to hypothetize that Lbs arose from Glbs by gene tandem duplication and divergent evolution in these two loci, and that they were differentially lost depending on the legume lineages. The presence in the more primitive Caesalpinioid *C. fasciculata* of a hemoglobin that has some characteristics of Lb[37] and is closely related to Dalbergioid Lbs supports that this feature is ancient in legumes. In addition, the presence also in nodulating non-legume species of class 1-derived Lbs (eg. *Parasponia*) or class 2-derived Lbs (eg. *Casuarina*) suggests this dual

evolution to be recurrent[3]. This will be an exciting evolutionary issue to determine how different Glbs adapted to Lbs and if these Lbs have any specific functional specificity.

The discovery of new or alternative mechanisms underpinning the nitrogen-fixing symbiosis strengthens *A. evenia* as a valuable model for the study of nodulation. The successful development of a forward genetic approach supported by a reference genome and companion resources also shows this legume is amenable for genetic research, this research being complementary to the one performed on *M. truncatula* and *L. japonicus*. The acquired knowledge will contribute to characterize the diversity of the symbiotic features occurring in legumes. It is also expected to benefit legume nodulation for agronomic improvement and, ultimately, it could provide leads to engineer nitrogen-fixation in non-legume crops.

Methods

Plant material for genome sequencing. We sequenced an inbred line of *Aeschynomene evenia* C. Wright (evenia jointvetch) obtained by successive selfings from the accession CIAT22838. This accession was originally collected in Zambia and provided by the International Center for Tropical Agriculture (CIAT, Colombia) (http://genebank.ciat.cgiar.org). *A. evenia* was previously shown to be diploid (2n=2x=20) and to have a flow cytometry-estimated genome size of 400 Mb (1C = 0.85 pg)[11,12,16.]

Genome sequencing and assembly into pseudomolecules. High-quality genomic DNA was prepared from root tissue of 15-day-old plants cultured *in vitro* using an improved CTAB method[12], followed by a high-salt phenol-chloroform purification according to the PacBio protocol. DNA was further purified using Ampure beads, quantified using the Thermofisher Scientific Qubit Fluorometry and fragment length was evaluated with the Agilent Tapestation System. A 20-kb insert SMRTbell library was generated using a BluePippin 15 kb lower-end size selection protocol (Sage Science). 55 Single-Molecule Real-Time (SMRT) cells were run on the PacBio RS II system with collections at 4-hourly intervals and the P6-C4 chemistry[49] by the Norwegian Sequencing Center (CEES, Oslo, Norway). A total of 8,432,354

PacBio post-filtered reads was generated, producing 49 Gb of single-molecule sequencing data, which represented a 78x coverage of the *A. evenia* genome. PacBio reads were assembled using HGAP (version included in smrtpipe 2.3.0), the assembly was polished using the Quiver algorithm (SMRT Analysis v2.3.0) and then the SSPACE-LongRead (v1.1) program scaffolded the contigs when links were found (Supplementary*,* Table 1). MiSeq reads were also generated to correct the sequence and estimate the genome size based on *k*-mer analysis (*SI Appendix*, Supplementary Note 1). The *de novo* genome assembly contains 1,848 scaffolds, with a scaffold N50 of ~0.985 Mb and with 90% of the assembled genome being contained in 538 scaffolds Then, we performed the *A. evenia* chromosomal-level assembly using serial analyses (fully described in Supplementary Note 1). The anchored scaffolds were joined with stretches of 100 Ns to generate 10 pseudomolecules named Ae01 to Ae10 according to the linkage group nomenclature for *A. evenia*[12] (Supplementary Tables 3 and 4).

Gene prediction and annotation. First, repeats were called from the assembled genome sequence using RepeatModeler v1.0.11 (https://github.com/rmhubley/RepeatModeler) (*SI Appendix,* Table S12). The genome was then masked using RepeatMasker v4-0-7 (http://www.repeatmasker.org/). Nine tissue specific RNA-Seq libraries (sequenced by the GeT-PlaGe Platform, Toulouse, France) and full-length transcripts generated from Iso-Seq (sequenced by the Cold Spring Harbor Laboratory, New York, USA) (details in *SI Appendix*, Supplementary Methods) were aligned on the unmasked reference with STAR[50]. The resulting bam files were processed with StringTie[51] v1.3.3b to generate gene models in gtf format which were merged with Cuffmerge from Cufflinks[52] v2.2.1 to produce a single gtf file. This gtf was used to extract a corresponding transcript fasta file using the gtf_to_fasta program included in the TopHat[52] v2.0.14 package. The masked genome, the transcript fasta file and the gff files were used to train a novel AUGUSTUS[53] v3.2.3 model. This model was used to call the genes for all chromosomes. The AUGUSTUS prediction and the gtf files were then given to EVM[54] v1.1.1 to refine the model and remove wrongly called genes. This produced a new gff file that was used to extract the corresponding transcripts using gtf_to_fasta. These transcripts were processed with TransDecoder[55] in order to validate the presence of an open reading frame.

To check the completeness of the prediction, a master list of 100 nodulation genes was created and used for some additional manual annotation leading to the current

annotation containing 32,667 gene models (Supplementary Table 7). Alignments of the Illumina RNA-seq clean reads from the nine samples with the STAR software supported 25,301 of the 32,667 predicted genes (Supplementary Table 8). Finally, genome assembly and annotation quality was assessed using the Benchmarking Universal Single Copy Orthologs (BUSCO[56] v3) with the BLAST E-value cutoff set to 10^{-5} (Supplementary Table 9). The BUSCO analysis includes a set of 1,440 genes that are supposed to be highly conserved and single copy genes present in all plants. Gene functions were assigned according to the best match of alignments using BLASP (1e-5) to SwissProt database. The InterPro domains, Gene Ontology (GO) terms and KEGG pathways database associated with each protein were computed using InterProScan with outputs processed using AHRD (Automated Human Readable Descriptions) (https://github.com/groupschoof/AHRD) for selection of the best functional descriptor of each gene product (Supplementary Table 10).

Gene expression analysis. The normalized gene expression counts were computed using Cufflinks package based on the TopHat[51] output results of the RNA-Seq data analysis from the nine samples' (Root N-, Root N+, Nodule 4d, Nodule 7, Nodule 14d, Stems, Leaves, Flowers and Pods) analysis performed for the *A. evenia* genome annotation. Gene expression was calculated by converting the number of aligned reads into FPKM (Fragments per kilobase per million mapped reads) values based on the *A. evenia* gene models. RNA-seq data previously obtained from RNA samples of *A. evenia* IRFL6945[18] were also processed and converted into FPKM.

Orthogroup inference. We inferred orthogroups with OrthoFinder[57] v.0.4.0 to determine the relationships between *A. evenia*, the other diploid *Aeschynomene* taxa and several legume species. In the latter, proteomes were last obtained from the Legume Information System (https://legumeinfo.org/), the National Center for Biotechnology Center (https://www.ncbi.nlm.nih.gov) or from specific legume species websites in March 2020. They included *A. duranensis* (V14167 v1), *A. hypogaea* (Tifrunner v1), *A. ipaiensis* (K30076 v1), *C. cajan* (pigeonpea ICPL87119 v1), *C. fasciculata* (golden cassia v1), *C. arietanum* (chickpea ICC4958 v2), *L. japonicus* (lotus MG-20 v3), *L. albus* (white lupin v1.0), *L. angustifolius* (narrow-leafed lupin Tanjil_v1.0), *G. max* (soybean Wm82.a2.v2), *M. truncatula* (barrel medic MtrunA17r5.0), *M. pudica* (sensitive plant v1), *P. vulgaris* (common bean

G19833 v2) and *V. angularis* (cowpea Gyeongwon v3). Recommended settings were used for all-against-all BLASTP comparisons (Blast+ v2.3.0) and OrthoFinder analyses to generate orthogroups (Supplementary Table 18). Phylogenies were created by aligning the protein sequences using MAFFT[58] v7.205 and genetic relationships were investigated in the trees generated with FastTree[59] v2.1.5 which is included in OrthoFinder. FigTree v1.4.3 (http://tree.bio.ed.ac.uk/) was subsequently used to further process the phylogenetic trees. A consensus species tree was also generated by OrthoFinder, based on alignment of single-copy orthogroups (*i.e*., an orthogroups with exactly one gene for each species).

Symbiotic gene analysis. Nodulation-related genes were collected from recent studies in *M. truncatula* and *L. japonicus*[1,3,24] and the protein sequences were retrieved from orthogroups generated with OrthoFinder for the 12 *Aeschynomene* taxa and the 14 other legume species. Important gene families or processes, such as the LysM-RLK/RLPs[24], components of the Topoisomerase VI complex[29-32], NCRs[33-35], Lbs/Glbs[36-37], and CRK receptors[38] were analyzed in greater detail (*SI Appendix*, Supplementary Methods). For phylogenetic tree reconstructions, protein sequences were aligned with MAFFT v7.407_1 and processed with FastME v2.1.6.1_1 (model of sequence evolution: LG, gamma distribution: 1 and Bootstrap value: x1000) or PhyML v3.1_1 (model of sequence evolution: LG, Gamma model: ML estimate, Bootstrap value: x100) using the NGPhylogeny online tool[60] (https://ngphylogeny.fr/). MrBayes v3.2.2 with two MCMC chains and 10^6 iterations was prefered for NCRs sequences as it gave better results with their short and divergent sequences. Sequence alignments were visualized with Jalview[61] v2.11.0. Microsynteny analysis was performed using the Legume Information System with the Genome Context Viewer (https://legumeinfo.org/lis_context_viewer) and the CoGe Database (https://genomevolution.org/coge/), using the GEvo (genome evolution analysis) tool to visualize the gene collinearity in syntenic regions.

Nodulation mutants. Nodulation mutants were obtained for *A. evenia* and characterized as fully described in the Supplementary Note 3. Briefly, a large scale mutagenesis was performed by treating 9,000 seeds from the CIAT22838 line with 0.30% EMS incubated overnight under gentle agitation. Germinated M_1 seedlings were transferred in pots filled with attapulgite. M_1 plants were allowed to self and 4-6 M_2 pods corresponding to

approximately 40 seeds were collected from individual M_1 plants. Seeds collected from the same tray containing 72 M_1 plants were pooled and defined as one bulk. 116 bulks of M_2 seeds were thus produced to constitute the EMS-mutagenized population. Phenotypic screening for nodulation alterations was conducted on 600 M_2 plants per bulk, 4 weeks after inoculation with the photosynthetic *Bradyrhizobium* strain ORS278. Plants with visible changes in their root nodulation phenotype were retained and allowed to self. The stability and homogeneity of the symbiotic phenotype was analyzed in the M_3 progeny. Whole inoculated roots of confirmed nodulation mutants were examined using a stereomicroscope (Niko AZ100; Campigny-sur-Marne, France) to identify alterations in nodulation and to establish phenotypic groups. The genetic determinism of the nodulation mutants was analyzed by backcrossing them to the CIAT22838 WT parental line according to the established hybridization procedure[11] and by determining the segregation of the nodulation phenotype in the F_2 population, 4 weeks post inoculation with the *Bradyrhizobium* strain ORS278. These F2 plants were also used for additional analyses. Allelism tests were performed between selected nodulation mutants using the same crossing procedure[11] to define complementation groups.

Targeted Sequence Capture. For Targeted Sequence Capture of symbiotic genes in nodulation mutants of *A. evenia*, 404 symbiotic genes known to be involved in the rhizobium-legume symbiosis or identified in expression experiments in *A. evenia*, were selected and their sequence extracted from the *A. evenia* genome to design custom baits with the following parameters: -Bait length 120 nucleotides, Tiling frequency 2x-. These probes were commercially synthetized by Mycroarray® in a custom MYbaits kit (ArborBiosciences, https://arborbiosci.com/). DNA was extracted from roots of M_3 nodulation plants to construct genomic libraries using a preparation protocol developed at the GPTRG Facility of CIRAD (Montpellier, France)(*SI Appendix*, Supplementary Methods). The captured libraries were sequenced on an Illumina HiSeq 3000 sequencer at the GeT-Plage Facility of INRA (Toulouse, France) in 150 bp single read mode. Read alignment and genome indexing were performed in the same way as for PoolSeq. Variations were called with Freebayes v1.1.0 with standard parameters and annotated according to their effect on *A. evenia* genes using SnpEff[62] (v4.3t and "eff -c snpEff.config transcript" parameters). This

file was then manually searched to identify the candidate gene variations able to explain the phenotypes.

Mapping-by-Sequencing. DNA was extracted from pooled roots of 100-120 F_2 backcrossed mutant plants and used to prepare the library for Illumina sequencing on a HiSeq 3000 sequencer at the GeT-Plage Facility of INRA (Toulouse, France) and at the Norwegian Sequencing Center (CEES, Oslo, Norway) as 150 bp-paired end reads. The *A. evenia* genome was indexed with BWA[63] index (v0.7.12-r1039, using standard parameter). Reads were assessed for quality using the FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and aligned on the reference genome with bwa mem using "M" option. The alignment file was compressed, sorted and indexed with Samtools[64] (v1.3.1). Variations were called with Freebayes[65] (v1.1.0, with "-p 100 --use-best-n-alleles 2 –pooled-discrete"). The resulting variation file was annotated using SnpEff[62] (v4.3t and "eff -c snpEff.config transcript" parameters) and SNP indexes corresponding to mutant allele frequencies were calculated. SNP plots with the SNP index and their chromosomal positions were obtained to identify genetic linkages visible as clusters of SNPs with an SNP index of 1. In the genomic regions harboring a genetic linkage, predicted effect of SNPs on genes were analyzed to identify candidate genes.

Data availability

The data reported in this study are tabulated in Datasets S1–S9 and *SI Appendix*. Genome assembly and annotation, accession resequencing and RNA-seq data for *A. evenia* are deposited at NCBI under BioProject ID: PRJNA448804. RNA-seq data for other *Aeschynomene* species are available under the BioProject ID: PRJNA459484. Resequencing data for *A. evenia* nodulation mutants are available under the BioProject ID: PRJNA590707 and PRJNA590847. Accession numbers for all deposited data are given in Supplementary Dataset 9. Genome assembly and annotation for *A. evenia* can also be accessed at the AeschynomeneBase (http://aeschynomenebase.fr) and at the Legume Information System (https://legumeinfo.org). Additional data were obtained from Legume Mines (https://mines.legumeinfo.org) and CoGe (https://genomevolution.org). Biological material and constructs are available for academic research upon reasonable request.

References

1. Roy, S. *et al.* Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *Plant Cell* **32**, 15-41. (2020).
2. Charpentier, M. & Oldroyd, G. How close are we to nitrogen-fixing cereals? *Curr. Opin. in Plant Biol.* **13**, 556-64 (2010).
3. van Velzen, R. *et al.* Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc. Natl. Acad. Sci. U S A* **115**, E4700-E4709 (2018).
4. Griesmann, M. *et al.* Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* **361**, 6398 (2018).
5. Masson-Boivin, C., Giraud, E., Perret, X. & Batut, J. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol*. **17**,458-66 (2009).
6. Sprent, J. I. & James, E. K. Legume-rhizobial symbiosis: an anorexic model? *New Phytol.* **179**, 3-5 (2008).
7. Giraud, E., Hannibal, L., Fardoux, J., Vermeglio, A. & Dreyfus, B. Effect of *Bradyrhizobium* photosynthesis on stem nodulation of *Aeschynomene sensitiva*. *Proc. Natl. Acad. Sci*. *USA* **97**, 14795-14800 (2000).
8. Giraud, E. *et al.* Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science* **316**, 1307-1312 (2007).
9. Okazaki, S. *et al.* Rhizobium-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *ISME J*. **10**, 64-74 (2015).
10. Teulet, A. *et al.* The rhizobial type III effector ErnA confers the ability to form nodules in legumes. *Proc. Natl. Acad. Sci. USA*. **116**, 21758-21768.
11. Arrighi, JF. *et al. Aeschynomene evenia*, a model plant for studying the molecular genetics of the Nod-independent rhizobium-legume symbiosis. *Mol. Plant Microbe Interact.* **25**, 851-861 (2012).
12. Chaintreuil, C. *et al.* A gene-based map of the Nod factor-independent *Aeschynomene evenia* sheds new light on the evolution of nodulation and legume genomes. *DNA Res*. **23**, 365-76 (2016).
13. Fabre, S. *et al.* Nod factor-independent nodulation in *Aeschynomene evenia* required the common plant-microbe symbiotic toolkit. *Plant Physiol*. **169**, 2654-64 (2015).
14. Bonaldi, K. *et al.* 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 (2011).
15. Ibáñez, F., Wall, L. and Fabra, A. Starting points in plant-bacteria nitrogen-fixing symbioses: intercellular invasion of the roots. *J. Exp. Bot.* **68**, 1905-1918 (2017).
16. Brottier, L. *et al.* A phylogenetic framework of the legume genus *Aeschynomene* for comparative genetic analysis of the Nod-dependent and Nod-independent symbioses. *BMC Plant Biol.* **18**, 333 (2018).
17. Boivin, C. *et al.* Stem nodulation in legumes: diversity, mechanisms, and unusual characteristics. *Crit. Rev. Plant Sci.* **16**, 1-30 (1997).
18. Gully, D. *et al.* Transcriptome profiles of Nod factor-independent symbiosis in the tropical legume *Aeschynomene evenia*. *Sci. Rep.* **8**, 10934 (2018).
19. Bertioli, D. J. *et al.* The genome sequences of *Arachis duranensis* and *Arachis ipaiensis* the diploid ancestors of cultivated peanut. *Nat. Genet.* **48**, 438-46 (2016).

20. Chaintreuil, C. *et al.* Naturally occurring variations in the nod-independent model legume *Aeschynomene evenia* and relatives: a resource for nodulation genetics. *BMC Plant Biol*. **18**, 54 (2018).
21. Lavin, M., Herendeen, P. S. & Wojciechowski, M. F. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst. Biol.* **54**, 575-594 (2005).
22. Cannon, S.B. *et al.* Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Mol Biol Evol*. **32**, 193-210 (2015).
23. Zhuang, M. *et al.* The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet*. **51**, 865-876 (2019).
24. Buendia L, Girardin A, Wang T, Cottret L, Lefebvre B. LysM receptor-like kinase and LysM receptor-like protein families: an update on phylogeny and functional characterization. *Front. Plant Sci*. **9**, 1531 (2018).
25. Qiao, Z., Pingault, L., Nourbakhsh-Rey, M. & Libault, M. Comprehensive comparative genomic and transcriptomic analyses of the legume genes controlling the nodulation process. *Front. Plant Sci*. **7**, 34 (2016).
26. Liu, C. W. *et al.* A protein complex required for polar growth of rhizobial infection threads. *Nat. Com*. **10**, 2848 (2019).
27. Cebolla, A. *et al.* The mitotic inhibitor ccs25 is required for endoreplication and ploidy-dependent cell enlargement in plants. *The EMBO J.* **18**, 4476-84 (1999).
28. Gonzalez-Sama, A. *et al.* Nuclear DNA endoreplication and expression of the mitotic inhibitor Ccs52 associated to determinate and lupinoid nodule organogenesis. *Mol. Plant Microbe Interact.* **19**, 173-180 (2006).
29. Yoon, H. J. *et al. Lotus japonicus SUNERGOS1* encodes a predicted subunit A of a DNA topoisomerase VI that is required for nodule differentiation and accommodation of rhizobial infection. *Plant J.* **78**, 811-21 (2014).
30. Suzaki, T. *et al.* Endoreduplication-mediated initiation of symbiotic organ development in Lotus japonicus. *Developmen*t **141**, 2441-5 (2014).
31. Yin, Y. *et al.* A crucial role for the putative Arabidopsis topoisomerase VI in plant growth and development. *Proc. Natl. Acad. Sci. USA*. **99**, 10191-6 (2002).
32. Breuer, C. *et al.* BIN4, a novel component of the plant DNA topoisomerase VI complex, is required for endoreduplication in Arabidopsis. *Plant Cell* **19**, 3655-68 (2007).
33. Montiel, J. *et al.* Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc. Natl. Acad. Sci. USA* **114**, 5041-5046 (2017).
34. Czernic, P. *et al.* convergent evolution of endosymbiont differentiation in dalbergioid and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides. *Plant Physiol*. **169**, 1254-65 (2015).
35. Alunni, B. *et al.* Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in *Medicago truncatula*. *Mol. Plant-Microbe Interact.* **20**, 1138-48 (2007).
36. Becana, M., Yruela, I., Sarath, G., Catalán, P. and Hargrove, M.S. Plant hemoglobins: a journey from unicellular green algae to vascular plants. *New Phytol.* doi/abs/10.1111/nph.16444 (2020).
37. Gopalasubramaniam, S.K. *et al.* Cloning and characterization of a caesalpinoid (Chamaecrista fasciculata) hemoglobin: the structural transition from a nonsymbiotic hemoglobin to a leghemoglobin. *Proteins* **72**, 252-60 (2008).

38. Quezada, E. H., García, G. X., Arthikala, M.K., Melappa, G., Lara, M. & Nanjareddy, K. Cysteine-rich receptor-like kinase gene family identification in the *Phaseolu*s genome and comparative analysis of their expression profiles specific to mycorrhizal and rhizobial symbiosis. *Genes (Basel)* **10**, pii: E59 (2019).
39. Venkateshwaran, M. *et al.* The recent evolution of a symbiotic ion channel in the legume family altered ion conductance and improved functionality in calcium signaling. *The Plant Cell* **24**, 2528-45(2012).
40. Gough, C., Cottret, L., Lefebvre, B. & Bono, J. Evolutionary History of Plant LysM Receptor Proteins Related to Root Endosymbiosis. *Front Plant Sci* **9**, 923 (2018).
41. Geurts, R., Xiao, T. T. & Reinhold-Hurek, B. What Does It Take to Evolve A Nitrogen-Fixing Endosymbiosis? *Trends Plant Sci.* **21**, 199-208 (2016).
42. Berrabah, F. *et al.* A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. *New Phytol*. **203**, 1305-14 (2014).
43. Bourdais, G. *et al.* Large-scale phenomics identifies primary and fine-tuning roles for CRKs in responses related to oxidative stress. *PLoS Genet*. **11**, e1005373 (2015).
44. Vaattovaara, A. *et al.* Mechanistic insights into the evolution of DUF26-containing proteins in land plants. *Commun Biol*. **2**, 56 (2019).
45. Madsen, L. H. *et al.* The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nature Com* **1**, 10 (2010).
46. Montiel, J. *et al.* Distinct signalling routes mediates intercellular and intracellular rhizobial infection 1 in *Lotus japonicus*. bioRxiv preprint doi: https://doi.org/10.1101/2020.05.29.124313 (2020).
47. Radhakrishnan, G.V. *et al.* An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. *Nat Plants* **6**, 280-289 (2020).
48. Sharma, V. *et al.* Molecular basis of root nodule symbiosis between *Bradyrhizobium* and 'Crack-Entry' legume groundnut (*Arachis hypogaea* L.). *Plants* **9**, 276 (2020).
49. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-8 (2009).
50. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
51. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. **33**, 290-5 (2015).
52. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. **7**, 562-78 (2012).
53. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–ii225 (2003).
54. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**, R7 (2008).
55. Haas, B. J. *et al. De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494–1512 (2013).
56. Simão, F. A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-2 (2015).

57. Emms, D. M. & Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. **16**, 157 (2015).
58. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. **30**, 772-80 (2013).
59. Price, M.N., Dehal, P.S. & Arkin A.P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
60. Lemoine, F., *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research* **47**:W260-W265 (2019).
61. Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M., Barton, G.J. Jalview version 2: A Multiple Sequence Alignment and Analysis Workbench. *Bioinformatics* **25**, 1189-1191 (2009).
62. Cingolani, P. *et al.* 1. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* (Austin) **6**, 80-92 (2012).
63. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010).
64. Li, H. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
65. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint* arXiv:1207.3907 [q-bio.GN] (2012).

Acknowledgements

The *A. evenia* genome sequencing and mutagenesis project was supported by a grant from the French National Research Agency (ANR-AeschyNod-14-CE19-0005-01) and by Agropolis Fondation through the «Investissements d'avenir» program (ANR-10-LABX-0001-01) under the reference ID "AeschyMap" AA1202-009. The work on globins and Lbs was supported by the Spanish Ministry of Science and Innovation-European Regional Development Fund (AGL2017-85775-R) obtained by M. B. We are grateful to the different sequencing centers that contributed to this work. The Norwegian Sequencing Centre (NSC) (http://www.sequencing.uio.no) generated both the PacBio DNA sequences and the Illumina sequences. The GeT-PlaGe platform (https://get.genotoul.fr/la-plateforme/get-plage/) was involved in Illumina sequencing. The Cold Spring Harbor Laboratory (https://www.cshl.edu/research/cancer/next-generation-genomics/pacific-biosciences-sequencing/) produced the PacBio Iso-Seq data. Computing was performed thanks to the GenoToul bioinformatics facility (http://bioinfo.genotoul.fr/) and thanks to the CIRAD - UMR AGAP HPC Data Center of the South Green Bioinformatics platform

(https://www.southgreen.fr/). The project also benefited from the expertise and the cytometry facilities of Imagerie-Gif (ttp://www.i2bc.paris-saclay.fr/spip.php?article1139) and of the genotyping facilities of the AGAP laboratory (https://www.gptr-lr-genotypage.com/plateaux-techniques/plateau-de-genotypage-cirad).

Author Contributions

J.F.A. conceived the whole project and supervised data analyses. C. K. determined the sequencing strategies, supervised the genome assembly and annotation, and analyzed sequence data. J.F.A., A.D. and D.G produced DNA and RNA material for *Aeschynomene* spp. P.L. assembled the genome. L.L. refined the genome assembly and annotated the genome. A.F. was involved in functional gene annotation and integration of data into the Legume Information System. L.B. and J.F. generated the *A. evenia* EMS mutagenized population and managed the population screening for nodulation mutants. D.G., C.C., F.C., N.N., F.G., E.G. and L.F. contributed to the mutagenesis project. J.Q., L.B., T.B., R.G. and M.P. undertook the phenotypic and genetic characterization of the nodulation mutants. J.Q., L.B. and C.C. produced plant DNA material and, with R.R and P.M., developed the DNA libraries for GBS and Targeted Sequence Capture. J.Q. produced the mutant DNA material for the Mapping-by-Sequencing and analyzed all the data for mutation identification. J.Q. and J.F.A. conducted analysis of the symbiotic genes. C.L. conducted selective pressure analysis. M.B. and I.V. conducted the analysis of globins. M. Bourge and N.V. performed the flow cytometry analysis. O.G., G.M., A.d'H. and B.H. contributed to the genome analysis and in the production of figures. J.F.A. wrote the manuscript. E.G. and other authors commented on the manuscript.

1.3 . Données complémentaires

Genetics of nodulation in *Aeschynomene evenia* uncover new mechanisms of the rhizobium-legume symbiosis

Johan Quilbé, Léo Lamy, Laurent Brottier, Philippe Leleux, Joël Fardoux, Ronan Rivallan, Thomas Benichou, Rémi Guyonnet, Manuel Becana, Irene Villar, Olivier Garsmeur, Bárbara Hufnagel, Amandine Delteil, Djamel Gully, Clémence Chaintreuil, Marjorie Pervent, Fabienne Cartieaux, Mickaël Bourge, Nicolas Valentin, Guillaume Martin, Loïc Fontaine, Gaëtan Droc, Alexis Dereeper, Andrew Farmer, Cyril Libourel, Nico Nouwen, Frédéric Gressent, Pierre Mournet, Angélique D'Hont, Eric Giraud, Christophe Klopp and Jean-François Arrighi

Supplementary notes, figures and tables

Supplementary Notes

Supplementary Note 1. Genome sequencing, assembly and annotation

MiSeq sequencing for genome sequence polishing and genome size estimation

An Illumina MiSeq paired-end sequencing library was also constructed following the manufacturer's instructions and sequenced at the CIRAD Genotyping Platform (GPTRG) (Montpellier, France). 2 x 4.85 Gb of raw data reads were generated, with read lengths of 300 bp, providing a 24x genome coverage (Supplementary Table 2). The MiSeq reads were trimmed for quality and assembled into contigs using SPADES. These contigs were filtered on coverage (5x < coverage < 17x) to obtain a higher confidence. This Illumina assembly was then compared with the PacBio scaffolds to correct the sequence, leading to the addition of 440,803 nucleotides to the genome sequence. Sequences generated from the MiSeq library were also used to calculate the genome size and estimate the heterozygoty level based on k-mer analysis by making use of GenomeScope v1[1]. The fastq files were processed with Jellyfish[2] version 2.1.1 using k-mer length 21bp in order to produce a count and the histogram file. This file was uploaded in to http://qb.cshl.edu/genomescope/ to generate the k-mer plot as well as the metrics.

Generation of chromosomal pseudomolecules

Two F2 mapping populations derived from a single cross between the reference line CIAT22383 and the mapping parent CIAT3282 were used to build a genetic map. One mapping population comprising 230 F_2 individuals was previously genotyped with a total of 364 SSR and SNP markers, based on *A. evenia* transcript sequences and produced 10 linkage groups. In addition to this conventional molecular marker map, a genotyping-by sequencing (GBS) approach was employed to develop high-density molecular markers as described[3]. For this, individual genomic DNA was extracted and fragmented by the ApeK1 restriction enzyme for another mapping population of 186 F2 plants. After barcoded GBS adapter ligation, sample pooling and PCR amplification, fragments were validated by Agilent Bio-analyzer 2100. The library was sequenced on two Illumina Hiseq2500 lanes producing 471,971,790 single-end 150 bp reads. Bowtie2 software package was able to align 89% of the reads to the

genome assembly scaffolds but only 28% of them presented a single hit. These reads were then processed with Tassel[4] v5.0 that identified tags as defining identical reads if having a minimum of 5 counts. 215,650 tags were used to identify SNPs by local genome alignment and variation calling. SNPs were filtered for quality with a set of vcftools. The resulting genotyping matrix contained 3,189 SNPs, with a minimum 8x coverage and less than 20% of missing data. In turn, it served to pairwise LOD scores and pairwise recombination frequencies matrices with JoinMap. Linkage group data were computed with Scaffhunter[5] (https://github.com/ SouthGreenPlatform/scaffhunter) to order in a single step SNPs and scaffolds.

To anchor the scaffolds onto pseudomolecules, transcripts corresponding to the gene-based markers of the conventional genetic map were searched by BLAST in the genome assembly and they allowed positioning PacBio scaffolds on the 10 *A. evenia* linkage groups. This initial chromosomal placement served as a backbone for adding and ordering scaffolds according to the GBS linkage groups. GBS SNPs consistently covered all the linkage groups with the exception of the upper part of AeLG4. Since a better genome assembly was obtained with HGAP (48x) for this region, scaffolds were replaced to be able to cover the whole AeLG4. Finally, 326 scaffolds were mapped on the linkage groups, totalizing ~360 Mb. Scaffold placement was arbitrary within blocks with the same centiMorgan value. 25 misjoined scaffolds were split at breakpoints identified by flanking GBS SNP locations delineating the switches in genotype calls and often characterized by a stretch of 1000x Ns in the intervening sequence, coming from the SSPACE long read scaffolding. Where allowed by the map data, scaffold position and orientations were adjusted using synteny with the genome of *A. duranensis*. We also included in the pseudomolecules, 64 scaffolds containing 16 Mb of sequence that lacked genetic support, but for which the *Arachis* genomes predicted precise locations based again on conserved synteny. Finally, the anchored scaffolds were joined with stretchs of 100 Ns to generate 10 pseudomolecules. These were numbered according to the *A. evenia* genetic map. Unmapped scaffolds were grouped into an “Unknown” chromosome. Left aside scaffolds were much shorter in size compared to mapped ones (in average 51 kb vs 755 kb) (Supplementary Table 3). Recent long read sequencing and scaffolding technologies should in the future allow integrating them.

As a quality control step, comparisons of genetic and physic (sequence) distances were performed. For this, the gene (transcriptome)-based were mapped onto the 10

pseudomolecules to plots genetic vs. physic distances. As a general trend, markers were mostly located at the chromosome ends showing consistent recombination (rising slope) while they were scarcely present in the pericentromeric parts characterized by markedly diminished recombination (flat central slop). The same approach was applied to the genomic (GBS)-based markers but, to avoid genetic length overestimation due to errors and missing data, the genotyping data was first corrected using GBS_corrector.0.3.py (Guillaume Martin, CIRAD, France). The same recombination patterns were observed but with a mostly regular marker distribution along the pseudomolecules encompassing the pericentromeric regions. Finally, the pseudomolecules structure was compared to the physic and genetic distances of the first and last anchoring marker for each constituent scaffold using SpiderMap (Jean-François Rami, CIRAD, France) and GenoGraph (Olivier Garsmeur, CIRAD, France). This revealed the chromosome arms to be mostly composed of long and orientated scaffolds whereas pericentromeric regions are constituted of numerous short scaffolds, part of them being unoriented and approximately positioned due to the lack of recombination in the central part of the chromosomes.

Chloroplast and mitochondria genomes are usually represented in different proportions when compared to the nucleus genome in the reads. They have also specific gene content which enables their characterization. Contig depth was calculated after realigning reads on contigs and used to filter sequences corresponding to the organelle genomes. The genome assembly contigs were also filtered to remove photosynthetic *Bradyrhizobium* genome sequences that aroused from plant root endosymbiont. It was further checked for microbial contamination by alignment against a bacterial genomes database using Megablast (E-value < 1e-5, > 90% identity, > 200 bp length mapped to scaffold sequences). Scaffolds identified as prokaryotic contamination were put apart. Scaffolds evidenced as being chimeric were split and bacterial sequences removed from the genomic data. Using the expurgated sequences, we checked the genome assembly accuracy by considering the coherence of genomic features such as expressed gene- and repeat-density gradients and by comparing the scaffolds again for synteny with the *Arachis* genomes. Potential scaffold shifts or reorientations were checked against genetic map constraints before being processed to provide an accurate scaffold ordering.

Illumina RNA-Seq and PacBio Iso-Seq

Plants were cultured *in vitro* in liquid BNM (Buffer Nodulation Medium) and inoculated or not with the photosynthetic strain ORS278 or grown in the greenhouse as previously described. Root and nodule material was collected from *in vitro* cultured plants at five different times or growth conditions: 15-days old roots without nitrogen supply (Root N-), 15-days old roots with 0.5 mM KNO3 (Root N+), 4-days old nodules (Nodule 4d), 7-days old nodules (Nodule 7d) and 14-days old nodules (Nodule 14d) that correspond to main steps of the symbiotic process. In addition, aerial plant material was harvested in mature plants grown in the greenhouse: stems, leaves, flowers and green pods. Samples were snap-frozen in liquid nitrogen, ground using a mortar and pestle and RNA was isolated using a CTAB extraction combined to purifications with LiCl and sodium acetate precipitations. RNA quality was assessed using an Agilent 2100 BioAnalyzer. Sequencing libraries were prepared using the TruSeq Stranded mRNA Kit for Illumina. 150-bp paired end sequencing was performed on one lane of Illumina HiSeq 3000 with the TruSeq Rapid SBS kit v4 (Illumina) chemistry at the GeT-PlaGe Platform (Toulouse, France). Raw Illumina RNA-seq data from the 9 libraries were trimmed for quality using trim_galore (https://www.bioinformatics.babraham.ac.uk/projects/ trim_galore/) v0.4.0 using standard parameters, providing an average 24x coverage for each (Supplementary Table 5).

Sub-samples of RNA material described above were combined in an equimolar pool for PacBio Iso-Seq. Pooled RNA was fractionated into three libraries by BluePippin (Sage Science), consisting of different sized RNA (1-2 kb, 2-3 kb and 3-6 kb). A total of 8 SMRT cells (3 for the 1-2 kb and 2-3 kb libraries, and 2 for the 3-6 kb library) was run on the PacBio RS II system with the P6-C4 chemistry by the Cold Spring Harbor Laboratory (New York, USA). Sequencing reads were processed with the RS_IsoSeq protocol of SMRT Analysis (v2.2) which includes the consensus sequence polishing Quiver pipeline (Supplementary Table 5).

Additional genome annotation

Non-coding RNAs, including microRNAs, small nuclear RNAs and ribosomal RNAs, were identified by using INFERNAL[6] v1.1.2 to search the Rfam database. RNAmmer v1.2 was additionally used to classify rRNA in more detail subclasses (Supplementary Table 11). The tRNA genes were searched by tRNAscan-SE[7] v1.3.1 (Supplementary Table 11). The telomeric

repeats were identified by searching for short (~7bp) high copy number repeats using Tandem Repeat Finder[8] v4.04 (Supplementary Table 13).

Resequencing of additional *A. evenia* accessions

Twelve *A. evenia* lines were selected for DNA re-sequencing since they represent seven identified genotypes (Supplementary Table 14). All sequencing was performed with an Illumina HiSeq 3000 machine at the GeT-PlaGe Platform (Toulouse, France), using 150-bp PE libraries. The reads were processed as also explained in the section method "Mapping-by-Sequencing". For this, they were mapped to the reference using bwa mem v0.7.12-r1039 with standard parameters. The sam files were compressed, sorted and indexed with samtools (v1.3.1). The variations were searched using freebayes v0.9.7 with standard parameters (Supplementary Table 15). The SNP distribution patterns for the re-sequenced accessions were visualized as SNP density value in a 1-Mb window along the reference genome using Circos[9] software. SNPs were also used to calculate a genetic distance matrix based on the identity-by-state similarity method and a maximum-likelihood phylogenetic tree was constructed based on 2,880,599 parsimony-informative SNPs with 1,000 bootstraps using IQ-TREE[10,11]. A phylogenetic tree was then prepared using iTOL[12] v 4.3.

Synteny analysis

To analyze intragenomic colinearity blocks inside the *A. evenia* genome and synteny colinearity with *A. duranensis* and *A. ipaiensis*, we used SynMap (CoGe, www.genomevolution.org) using homologous CDS pairs with the following parameters: maximum distance between two matches (-D): 20; minimum number of aligned pairs (-A): 10; "Quota Align Merge" algorithm with maximum distance between two blocks (-Dm): 50. We analyzed the Ks distribution among pairs of orthologous and paralogous gene pairs between and within the *A. evenia-A. duranensis-A. ipaiensis* genomes as proportion (%) of genes pairs in Ks bin sizes of 0.05. For analysis of synteny blocks within the *A. evenia* genome, clusters containing at least 10 collinear genes with Ks values ≤ 1.5 were retained and illustrated in Circos[9]. JCVI[13] was used to represent orthologous gene relationships between the *A. evenia*, *A. duranensis* and *A. ipaiensis* genomes.

Transcriptome assembly for different *Aeschynomene* species

Available Illumina single reads data were used for *A. evenia* ssp. *evenia* and *A. evenia* ssp. *serrulata*. For 10 additional diploid *Aeschynomene* species, belonging to the Nod factor-independent clade (Supplementary Table 16), transcriptomes were prepared using tissue samples corresponding to roots at 0 dpi and nodules at 4 and 7 dpi following inoculation with the photosynthetic *Bradyrhizobium* strain ORS278. For each species, RNAs extracted from the three samples were pooled equally and sent to the CEES Platform (Oslo, Norway) where 200-bp short insert libraries were prepared and sequenced on two lanes of Hi-Seq 4000 to obtain 2x150 bp reads. Raw Illumina RNA-seq datasets were assembled with DRAP[14] using runDrap module with standard parameters). Transcriptome assemblies' completeness was assessed using the Benching Universal Single Copy Ortholog approach (BUSCO v3) (Supplementary Table 17).

Supplementary Note 2. *In silico* gene analysis and symbiotic properties

Search for nodulation related genes

Genes from Medicago and Lotus served to retrieve orthogroups generated with OrthoFinder, where the presence of *Aeschynomene* genes were searched. Since the different taxa included in this analysis share the same ancestral legume whole genome duplication, the orthogroups can comprise both orthologous and paralogous genes. Therefore, the presence or absence of *Aeschynomene* orthologs or paralogs was assessed through visual analysis of the topology of the phylogenies obtained. A BLAST analysis in the Legume Information System (https://legumeinfo.org/) was also performed to ascertain genetic relationships. Phylogenetic patterns for genes of interest were summarized in datasests (Supplementary Data S1-8) and combined to the consensus species tree generated by OrthoFinder and showing the phylogenetic relationships between the analyzed *Aeschynomene* and other legume species.

Synteny analysis between nodulation genes

Microsynteny analysis of symbiotic genes present in *A. evenia*, *Arachis* spp. and *M. truncatula* was performed in the Legume Information System with the Genome Viewer

Context (https://legumeinfo.org/lis_context_viewer). Synteny was built using the *A. evenia* genome as reference and macrosyntenic genome regions were centered on *A. evenia* genes of interest. Microsynteny was used to comfort orthologous/paralogous relationships for the two copies of *SYMRK* and *PUB1* resulting from the Papilionoid WDG by evidencing that they rely in blocks of synteny with a conserved gene collinearity. For the genes missing in *A. evenia* but present in *Arachis* spp*.,* the *Arachis* protein sequences were used as BLAST queries on the *A. evenia* genome to exclude to possibility that the genes are actually present but not annotated. Finally, genomic losses in *A. evenia* were investigated by synteny analysis with *A. duranensis* in the Accelerating Comparative Genomic Database (CoGe: (https://genomevolution.org/coge/). *A. duranensis* symbiotic genes were used to find syntenic regions in the *A. evenia* genome. The GEvo (genome evolution analysis) tool was then applied to visualize the collinearity between syntenic regions and to evaluate gene loss. The Genome Viewer Context tool from the Legume Information System was used in parallel to draw the schematic representation of the syntenic analysis.

Analysis of the LysM-RLK/RLP gene family

For the identification and phylogenetic analysis of the LysM-RLK/RLP gene family in *A. evenia*, a database with protein sequences from *M. truncatula* and *L. japonicus* models was used to identify orthogroups produced by OrthoFinder and containing LysM-RLK/RLP homologs. Protein sequences were aligned and compared to detect anomalous truncated or fused forms probably resulting from incorrect gene annotation. Erroneous sequences were manually corrected based both on RNA-Seq evidence and expected conserved protein structure, leading to a LysM-RLK/RLP dataset for *A. evenia*, *A. duranensis*, *M. truncatula* and *L. japonicus*. To facilitate cross-species comparisons, we adopted the established nomenclature that distinguishes LysM-RLKs with a functional kinase (the LYK group), a dead or WALK-like kinase (the LYR group) and LysM-RLP that are anchored to the plasma membrane through a GPI anchor site. Phylogenetic trees were inferred independently with predicted protein sequences of the LYK, LYM and LYRs.

Analysis of the Topoisomerase VI complex

Two components were previously described as being involved in the endoreplication of cortical cells of nodules in Lotus, SUNERGOS1, coding for the A subunit, and VAG1, an

interactor. To identify additional components of this complex, we mined the Arabidopsis data and retrieved two genes, BIN3 (At3g20780), corresponding to the B subunit, and BIN4 (At5g24630), representing a second interactant. The Arabidopsis genes served for a BLAST search in the Legume Information System (https://legumeinfo.org/) to identify closest homologs in Medicago and Lotus. Orthogroups for SUNERGOS1, VAG1, BIN3 and BIN4 were then identified and analyzed for the presence of orthologs in *Aeschynomene* species. *VAG1* gene completeness was analyzed by sequence alignment in Multalin (http://multalin.toulouse.inra.fr/), and when truncated in *Aeschynomene* spp, transcriptome assemblies were mined to try to recover missing sequences.

Analysis of the NCR gene family

The automated annotation pipeline failed to predict NCR coding genes correctly. Therefore, they were manually searched and annotated by combining homology search with NCRs sequences that were previously identified in RNAseq data from *A. indica* and *A. afraspera* and transcript evidences for *A. evenia*. This resulted in the annotation of 58 NCR-coding bi-exonic genes. The NCR genes were plotted on the *A. evenia* chromosomes using SpiderMap (Jean-François Rami, CIRAD, France). All the *AeNCR* genes felled in the same orthogroup generated with OrthoFinder, that also contained 10 NCRs from *A. duranensis* (Aradu.ICC64, Aradu.RQG82, Aradu.H0Y95, Aradu.FUC53, Aradu.2CS1T, Aradu.25ZW3, Aradu.816EN, Aradu.55RTG, Aradu.GI8LM, Aradu.C8Z10) and 5 NCRs from *A. ipaiensis* (Araip.486ZT, Araip.7C3ZL, Araip.SG3VK, Araip.WY8RA, Araip.Y9YA6). Six additional NCR genes were found in the *A. duranensis* genome by direct Blast search (Aradu.2CS1T, Aradu.25ZW3, Aradu.816EN, Aradu.55RTG, Aradu.GI8LM, Aradu.C8Z10). Theorical pI were calculated for mature peptides from Expasy website (https://web.expasy.org/compute_pi/). Sequence logos were generated using NCR sequences aligned in Jalview[15] and WebLogo software[16].

Analysis of globin and lehemoglobin genes

They were searched by using the globin repertoire described for different legume species[17,18]. They were distributed in three orthogroups generated by OrthoFinder, which corresponded to globins of class 1, 2 and 3, respectively. *A. evenia* contains three class 1 globins (Ae04g33090, Ae04g33100 and Ae04g33130), one class 2 globin (Ae09g18610) and two class 3 globins (Ae07g15760 and Ae08g12400). A. Gene structure analysis indicates that

they all six contain four exons interrupted by three introns at the same positions as occurs with other globin genes. Sequence alignment indicated that some class 1 globin genes of *A. duranensis* were not correctly annotated. Based on both gene structure and sequence conservation, the predicted Aradu.9IJ7G.1 gene could be putatively split into three genes named here. Aradu.9IJ7G.1-A, -B and –C, respectively.

Analysis of nodule cell endoreplication

To evaluate of ploidy levels in roots and nodules, *Arachis hypogaea* plantlets were cultured in vermiculite and inoculated or not with *Bradyrhizobium* strain ORS3257 and different *Aeschynomene* spp. cultured in liquid BNM medium and inoculated or not with *Bradyrhizobium* strain ORS278. Two to three weeks post inoculation, roots and nodules were collected from each species. Plant samples were analyzed by flow cytometry at the Imagerie-Gif Cytometry Facility. For this, roots or nodules nuclei were isolated by chopping tissue with a razor blade in a plastic Petri dish with 300 µl of Gif nuclei-isolation buffer (45 mM MgCl2, 30 mM sodium citrate, 60 mM MOPS, 1% (w/v) polyvinylpyrrolidone 10,000, pH 7.2) containing 0.5% (w/v) Triton X–100, supplemented with 5 mM sodium metabisulphite and RNAse (2.5 U/ml) (19). The suspension was filtered through 50-µm nylon mesh. The isolated nuclei were stained with 5 µg/ml DAPI, a specific DNA fluorochrome, and kept 5 min at 4°C. Endopolyploidy of at least 10,000 stained isolated nuclei was determined for each sample using a cytometer (CytoFLEX S, Beckman Coulter. Excitation 405 nm, 85 mW; emission through a 450/45 nm band-pass filter). The frequency values were calculated from measurements of samples comprising 3 individuals.

Globin purification and spectral analyses

The four globins were cloned into Champion pET-11a(+) (Invitrogen) and expressed with an N-terminal *Strep*-tag in *Escherichia coli* C41(DE3) cells (Lucigen)[20]. Cells were precultured at 37°C with mild agitation overnight in 250 ml of LB medium with 100 µM ampicillin. One liter of TB medium containing 100 µM ampicillin was inoculated with 10 ml of preculture, and cells were incubated under the same conditions until an optical density at 660 nm of 0.6-0.7 was reached. Cells transformed with AeLb1 and AeLb2 were then grown at 28°C for 16 h. For AeGlb1 and AeGlb2, 0.25 mM isopropyl β-D-1-thiogalactopyranoside was added to the medium. Transformed cells were washed in 50 mM potassium phosphate buffer (pH 7.0) and

stored at -80°C for no longer than three weeks. For purification, cells were resuspended in 20 mM Tris (pH 8.0) and 150 mM NaCl, sonicated (3 x 2 min), and cleared by centrifugation.

The supernatant was loaded on a StrepTactin Sepharose High Performance column (GE Healthcare), previously equilibrated with the same buffer. After the protein was loaded, the column was washed with at least five column volumes of buffer, and the recombinant proteins were eluted with buffer containing 2.5 mM desthiobiotin (Sigma). After desalting on NAP-5 mini-columns (GE Healthcare) and concentration by ultrafiltration, protein purity was examined on CoomassieBlue-stained SDS gels. Spectra were taken with 17-22 µM of protein in 50 mM potassium phosphate buffer (pH 7.0) following published protocol (17). Briefly, the proteins were oxidized with potassium ferricyanide, desalted and quantified based on the Soret band, and the spectra of the ferric forms were recorded. The ferrous forms were obtained by adding a trace of sodium dithionite, the oxyferrous forms by immediately passing the ferrous globin onto a NAP-5 mini-column and the nitrosyl forms by immediately adding a few crystals of sodium nitrite to the ferrous form.

Supplementary Note 3. Forward genetic screen in *A. evenia*

EMS mutagenesis on *A. evenia*

To determine the optimal EMS dosage, a small scale mutagenesis was performed on lots of 260 seeds of *A. evenia* (inbred line CIAT22838) using a range of EMS concentrations. Seeds were scarified with concentrated sulfuric acid (96%) for 40 min and cleansed five times in distilled water. Seeds were then incubated in 0, 0.30, 0.35, 0.40, 0.50 and 0.06% EMS (Sigma-Aldrich) overnight under gentle agitation on a rotary shaker set at 80 rpm. The treated seeds were washed every ten times over a period of 5 hours and then germinated on Petri dishes filled with 0.8% water agar and put upside-down at 34°C overnight. Germinated seedlings (M_1 plants) were transferred to square pots (12x12 cm) containing atapulgite (Sorbix US Special) (six plants per pot) and grown until maturity under tropical greenhouse conditions (28°C temperature, 70% relative humidity, natural sunlight). Early radicle outgrowth, seedling survival, frequency of fertile individuals and seed production were assessed as measures of plant reproductive capacity and EMS mutagenicity (*SI Appendix*, Table S19). As all parameters were significantly impacted by treatments of 0.35% EMS and

above, the optimal EMS dose to use was determined to be of 0.30%. In a three year-effort, a large scale mutagenesis was performed by treating 9,000 seeds subdivided in lots of 500 seeds with 0.30% EMS as described. Germinated seedlings (M_1 plants) were transferred in square pots (12x12 cm) containing atapulgite. One tray contained 72 plants distributed in 12 pots, with ~60 out of the 72 plants effectively producing seeds. These M_1 plants were allowed to self and 4-6 M_2 pods corresponding approximately 40 seeds were collected from individual M_1 plants. Seeds collected from a same tray were pooled and defined a bulk. 116 bulks of M_2 seeds were thus developed and constituted the EMS-mutagenized population above (Supplementary Fig. 21).

Screening for nodulation mutants

M_2 plants of the mutagenized population were screened for their nodulation properties as follows: 600 seeds per bulk were scarified, germinated and subsequently transferred in 30x45 cm pots (300 plants/pot) filled with atapulgite. Just planted seedling were root-inoculated with the photosynthetic *Bradyrhizobium* strain ORS278 and grown for 4 weeks in the greenhouse. Plants were then visually inspected for root nodulation. Those with visible changes in their nodulation phenotype were transferred to new pots filled with fertilizer-enriched compost (45% Nehaus N2/45% Nehaus S/10% pozzolan) and grown to maturity. The phenotype of the putative M_2 *A. evenia* mutants was evaluated in the M_3 generation obtained by self-pollination (approximately 40 M_3 plants/mutant) using the same culture conditions as for the screening. Finally, 250 mutated lines with stable and homogeneous nodulation phenotypes and sufficient vigor to grow normally were kept. These mutants were named according to the bulk they belong and their order of discovery (e.g., A40 is the first mutant isolated from the bulk A4). Some mutants were grown in BNM medium, inoculated with the photosynthetic *Bradyrhizobium* strain ORS278 and analyzed for their root nodulation phenotype at 14 dpi. Whole inoculated roots were examined using a stereomicroscope (Niko AZ100; Campigny-sur-Marne, France) in order to specify the alterations in nodulation and establish phenotypic groups (Supplementary Fig. 21).

Genetic analysis

Crossing the *A. evenia* nodulation mutants with the CIAT22838 WT parental line was performed with the established manual emasculation and pollination procedure. Because

the WT line was usually more vigorous than the mutated lines, it was used as female partner in most segregation analyses performed. The resulting F_1 plants were phenotyped for root nodulation and segregation of the nodulation phenotype was subsequently examined in the F_2 population to determine the genetic determinism of the observed symbiotic phenotypes, four weeks post-inoculation and growth in atapulgite substrate in the greenhouse. 300 to 600 segregant F_2 plants derived from each mutant x WT cross were analyzed. Statistical analysis of the segregation patterns was performed using the Khi-2 test with $\alpha<5\%$ to validate a monogenic and recessive determinism. The resulting mutant x WT F2 plants were also used for two additional analyses. First, to investigate whether the root nodulation mutants are also altered for stem nodulation and if these two nodulation phenotypes co-segregate, several F2 plants per mutant x WT crossing were treated as follow: plants phenotyped for root nodulation 4 weeks post-inoculation were distributed into a bulk of 6 WT plants and a bulk of 12 mutant plants and transferred into pots (six plants/pot) with atapulgite for an additional one week-growth. These plants were then inoculated on the stem by submerging them in a diluted culture of *Bradyrhizobium* for 24 h. Co-segregation of the root and stem phenotypes in the F2 plants was suggestive of a single-gene control. Second, for mutants with a mutation in a candidate symbiotic gene found by the Targeted Sequence Capture analysis, F2 plants were genotyped. Typically, DNA was isolated from roots of 16 homozygous mutant F2 plants using the CTAB extraction method. PCR primers flanking the mutation site were designed with Primer3 and used to amplify the selected fragments for Sanger sequencing. Co-segregation of the mutation with the nodulation phenotype was then analyzed to infer the probable involvement of the symbiotic gene in the control of the observed nodulation phenotype. To further validate the candidate genes by allelism tests, monogenic mutants displaying independent mutations in the same gene were crossed and their F1 hybrids analyzed for the nodulation phenotype. A mutant phenotype observed in the F1 progeny was indicative that the parental mutated lines belong to the same complementation group.

Preparation of captured libraries for the Targeted Sequence Capture

DNA samples extracted from mutant roots were sheared to approximately 100-600 bp length with an average 300 bp length on a Biorupter Standard (Diagenod Cat No. UCD-200, Woburn, MA). After end repair using NEB Next End Repair Enzyme Mix (New England

Biolabs) to generate blunt ends, DNA fragments were purified using AMPure beads (Beckman Coulter), ligated to P5 and P7 adapters and purified on AMPure beads. Ligated DNA fragments were treated with the Bst enzyme (DNA Polymerase Large Fragment) (New England Biolabs) to modify adapter ends and purified on AMPure beads. Dual indexing and pre-capture enrichment was then performed using varied number of PCR cycles for each capture. Amplified samples were purified on AMPure beads, quantified on TapeStation and by quantitative PCR (Q-PCR) using the Clontech kit. Equal amount of library products from 30 to 48 genomic libraries were pooled to obtain at least 500 ng DNA. Sequence capture by hybridization was performed according to the manufacturer's protocol for the MYbaits kit, with the custom oligonucleotide library corresponding to selected genes of A. evenia (baits). Genomic DNA-bait hybrids were captured using Streptavidin magnetic beads, washed, and amplified by PCR using postcapture primers. The final captured libraries were quantified by Q-PCR before the Illumina sequencing.

Analysis of the *AeCRK* gene

AeCRK (Ae05g12380) protein domains were identified and annotated using InterProscan and further refined by comparative structural analysis with other CRKs. Homologous genes were then identified in other legume species by mining the orthogroups database generated with OrthoFinder. Chimeric sequences were manually split and separated genes are tagged with _A or _B or _C. Genes of the CRK gene cluster on Ae05 were further analyzed combining synteny analysis with other legume species and by inferring gene relationships from the phylogenetic tree.

Functional complementation experiment

To prove that mutations in *AeCRK* are causative of the Nod$^-$ phenotype, genetic complementation assays were carried out. A construct containing the AeCRK coding sequence (2 Kb) with the 1.4-kb native promoter and the 0.2 kb T35S terminator was synthetized and cloned into pCambia1302 by GeneCust (www.genecust.com) using the *KpnI* and *SalI* restriction sites, and confirmed by DNA sequencing. The pCambia1302 empty vector and the pCambia1302 vector carrying the p*AeCRK*-A*eCRK*-*T35S* construct were introduced into *A. rhizogenes* ARqua1 by the freeze-thaw method and then introduced in roots of the I10 mutant by hairy root transformation as previously described[21]. In short, seedling were

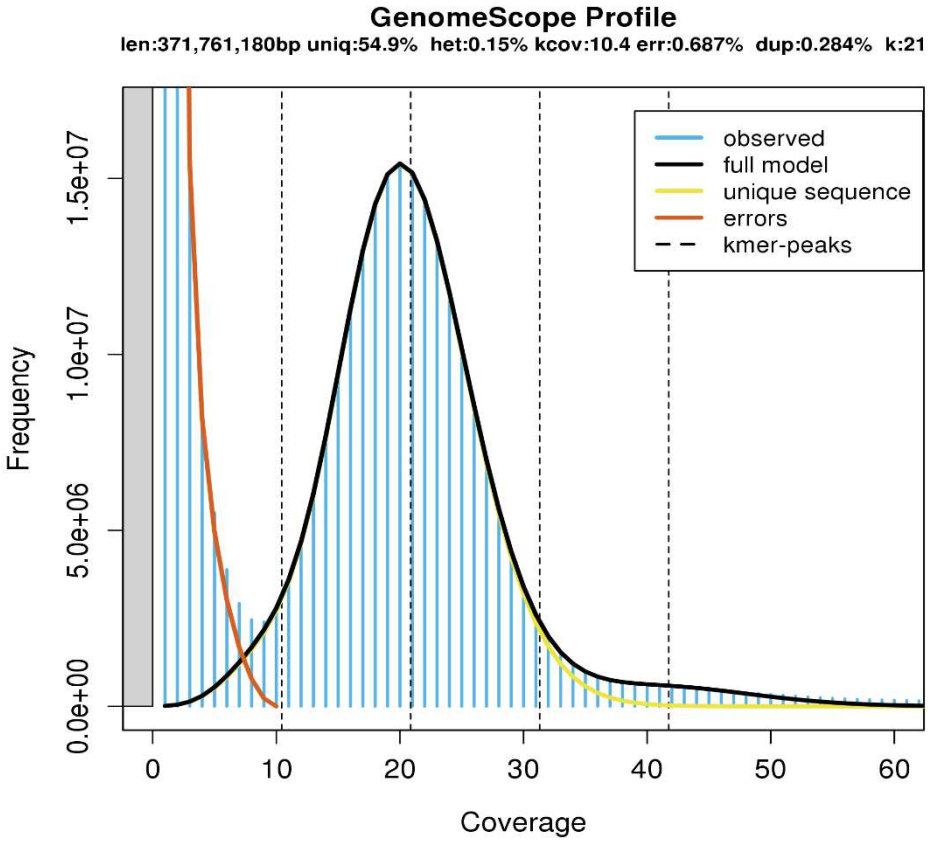
infected with *A. rhizogenes* on freshly sectioned radicles, grown on half-strength MS medium (Murashige and Skoog basal salt mixture) at 25°C for one week, then transferred twice on solid half-strength MS supplemented with a cefotaxime at 300 µg ml/L and cultivated at 25°C with a 16-h light and 8-h dark photoperiod. Transgenic hairy roots were identified on GFP fluorescence and non-transformed roots removed for the plants. Dissected plants were transferred to Falcon tubes filled with liquid buffered nodulation media supplemented with 0.5 mM KNO3. Plants were grown in a 28°C growth chamber with a 16-h light and 8-h dark regime and 70% humidity for one week and then inoculated with 1 mL of a 5-day-old *Bradyrhizobium* ORS278 culture grown in YM medium. Plants were regularly observed for nodule formation, and nodulation was quantified 14 days after inoculation.

Signature of selection on *AeCRK*

We investigated the selection acting on *CRK* in the *Aeschynomene* clade (foreground branch). We adopted branch models to identify differential signatures of selective pressure acting on the sequences between the foreground branch and the rest of the tree (i.e. background branches) and branch-site models to identify differential signatures of selective pressures acting on specific sites between foreground and background branches. These models are implemented in the codeml program using ete-evol wrapper. These methods calculate different synonymous and nonsynonymous substitution rates ($\omega=\frac{dN}{dS}$) using the phylogenetic tree topology for both foreground and background branches. CDS sequences from CRK orthologs were aligned using MUSCLE v3.8.3. Likelihood-ratio test (LRT) was used to assess significance between models ran with *codeml*. For the branch model, we compare likelihoods from the "b_free" and "M0" models to determine if the ratios are different between background and foreground branches77 (*p-val*>0.05 : ωforeground and ωbackground are not different, *pval*<0.05 : ωforeground and ωbackground are different). We also compared likelihoods from the "b_free" and "b_neut" models to determine if the ratio on the foreground branch is not neutral (*p-val*>0.05: no signature of selection on foreground (not different from neutral evolution), *p-val*<0.05 and ωforeground<1: signature of negative/relaxed selection on foreground, *p-val*<0.05 and ωforeground>1: signature of positive selection on foreground). For the branch-site model, we compared likelihoods from the "bsA" and "M1" models to determine if the selection pressure is relaxed on specific sites

on the foreground branch (*p-val*<0.05: signature of relaxed selection on specific foreground sites). We also compared likelihoods from the "bsA" and "bsA1" models to determine if there are sites under positive selection on the foreground branch (*p-val*<0.05: signature of positive selection on specific foreground sites).

Supplementary Figures



Supplementary Figure 1. Genome size estimation using *k*-mer distribution (k = 21). The *x* axis represents the *k*-mer peak coverage and the *y*-axis depicts the *k*-mer frequency. Genome size is estimated from the depth of the main peak.

Supplementary Tables

Supplementary Table 1. Comparison of the different PacBio assemblers

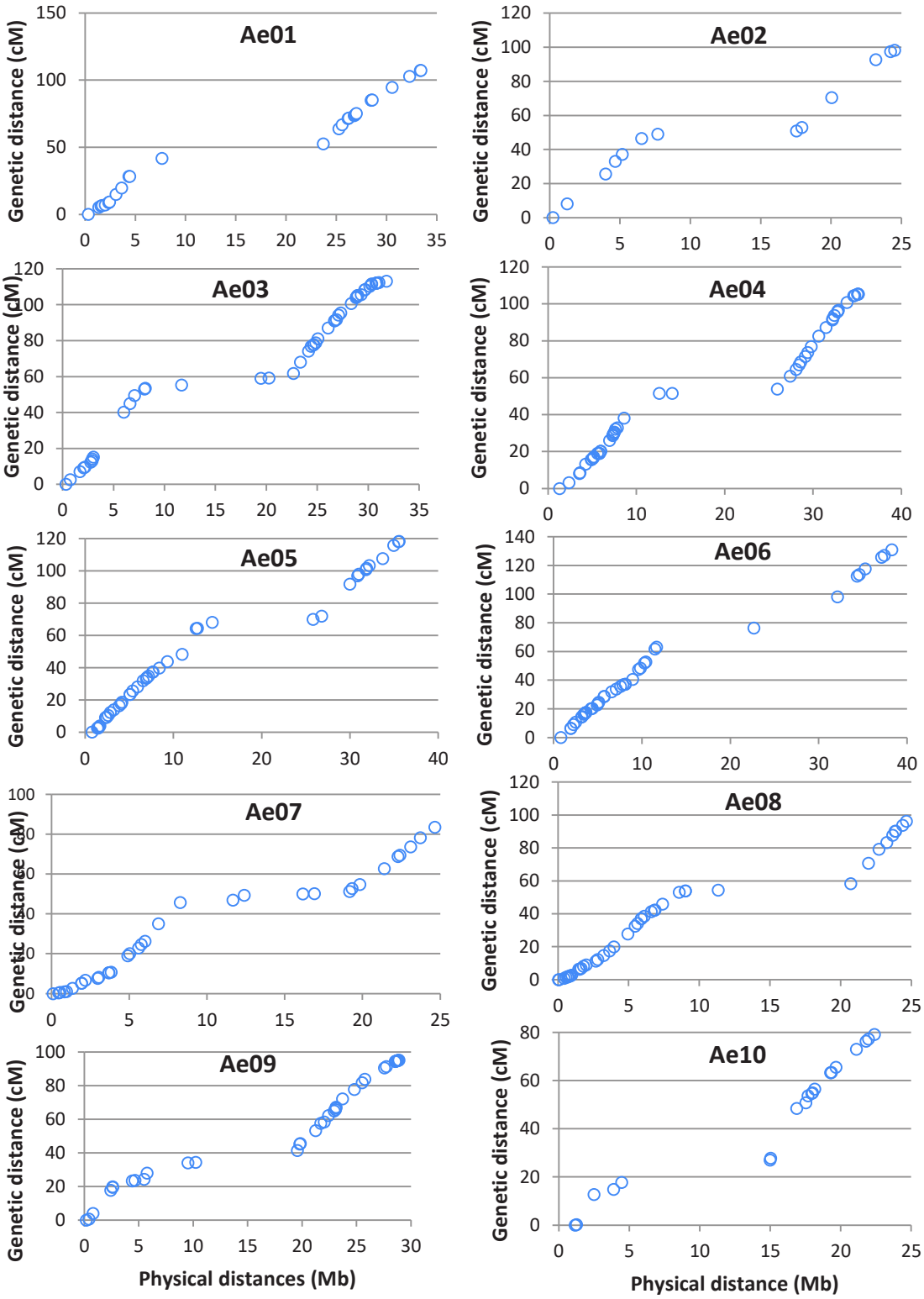
	HGAP (48x)	HGAP (78x)	Falcon (78x)	MHAP (78x)
Assembly size	384 Mb	381 Mb	320 Mb	415 Mb
Scaffolds	2,312	2,099	1,350	5,597
Largest scaffolds	5.04 Mb	11.8 Mb	36.4 Mb	2.22 Mb
N50	0.84 Mb	1.05 Mb	1.09 Mb	0.23 Mb
L50	80	65	76	318

Supplementary Table 2. Data metrics obtained with the MiSeq reads

Raw input		Output (SPADES assembler)		Output (Filtered)	
Coverage	24x	Contigs	241,444	Contigs	73,804
n reads	16,198,053 x2	Length	373 Mb	Length	265 Mb
total length	4,875 Gb x2	N50	23,327	N50	51,310
read size	300 bp	L50	2,721	L50	1,152
match rate	99,79	insert size	764 bp	insert size	764 bp

Supplementary Table 3. Genome sequence assembly in the chromosome level pseudomolecules

Pseudo-molecule	Scaffolds	Scaffolds placed with markers	Scaffolds placed using synteny	Gene based-markers	GBS-based markers	Size (bp)	GC%
Ae01	46	34	12	29	247	32,760,111	34,61
Ae02	35	30	5	13	150	24,242,855	33,61
Ae03	45	41	4	49	300	31,946,466	35,06
Ae04	42	39	3	46	267	35,017,145	34,00
Ae05	39	35	4	48	325	36,294,835	34,24
Ae06	42	35	7	41	318	38,544,573	34,29
Ae07	34	32	2	34	160	24,667,243	33,81
Ae08	39	28	11	43	214	26,254,046	35,04
Ae09	40	33	7	36	195	28,845,408	34,26
Ae10	38	29	9	25	207	23,578,249	35,18
Unknown	1,446	0	0	0	0	73,801,912	38,80
Total	1,846	336	64	364	2,383	375,952,843	35.26



Supplementary Figure 2. Plots of genetic-by-physical distances using gene-based markers. Dots show the locations of gene markers from the *A. evenia* genetic map[10] on the chromosome sequence. There are 364 SSR and SNP markers shown in these comparisons.

Supplementary Table 4. ***Aeschynomene evenia*** **genome assembly statistics**

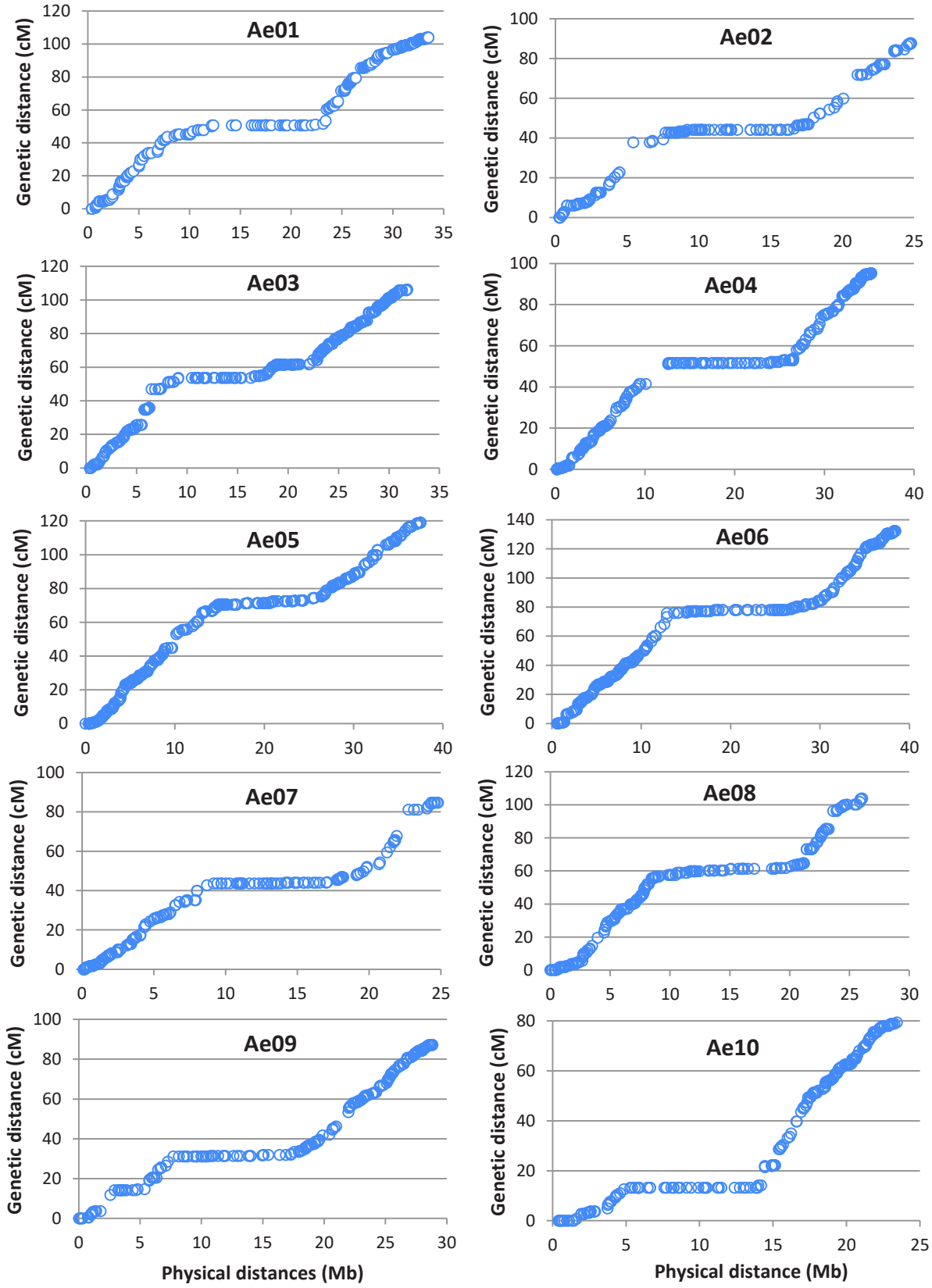
Number of polished scaffolds	1,846
N50 scaffold lentgh	0.985 Mb
N90 scaffold lentgh	0.111 Mb
L50 scaffold count	69
L90 scaffold count	538
Minimum scaffold size	1,100 bp
Longest scaffold	11,436,131 bp
Scaffolds > 1kb	1,846
Scaffolds > 10kb	1,508
Scaffolds > 100kb	570
Scaffolds > 1 Mb	66
Scaffolds > 10 Mb	1
Ns	5,170,011
Scaffolds anchored	336
Scaffolds placed by synteny	64
Scaffolds unanchored	1,446
Total length of scaffolds on chromosomes	302,150,931 bp
Total length of unanchored scaffolds	73,801,912 bp
Total length of scaffolds	375,952,843 bp

Supplementary Table 5. Summary of Illumina transcriptome sequencing (raw data)

Sample	Description	Coverage	Total reads	Total length	Insert size
Root N-	15d-old roots without KNO3	24x	32,481,391 x2	4.90 Gb x2	214 bp
Root N+	15d-old roots with KNO3	26x	34,069,118x2	5.14 Gb x2	208 bp
Nodule 4d	Young nodules	22x	29,309,566 x2	4.43 Gb x2	210 bp
Nodule 7d	Mature nodules	20x	26,981,020 x2	4.07 Gb x2	208 bp
Nodule 14d	Mature nodules	26x	34,289,397 x2	5.18 Gb x2	215 bp
Stem	Upper part of the stems	26x	34,143,493 x2	5.16 Gb x2	214 bp
Leaf	Fully developped leaves	24x	31,697,706 x2	4.79 Gb x2	217 bp
Flower	Open flowers	22x	29,422,999 x2	4.44 Gb x2	231 bp
Pod	Green pods	24x	31,087,498 x2	4.69 Gb x2	225 bp

Supplementary Table 6. Summary of Iso-Seq transcriptome sequencing

Library	SMRT cells	Total reads	Total length	n isoforms
1-2 kb	3	2,203,829	4,391,781,054 bp	37,615
2-3 kb	3	2,124,014	4,090,397,158 bp	27,345
3-6 kb	2	1,242,966	2,261,431,519 bp	9,771
Total	8	5,570,809	10,743,609,731 bp	-



Supplementary Figure 3.Plots of genetic-by-physical distances using genome-based markers. Dots show the locations of genomic markers generated by GBS analysis in *A. evenia* on the chromosome sequence. There are 3,189 genomic SNPs shown in these comparisons.

Supplementary Table 7. Gene prediction statistics for *Aeschynomene evenia* and comparison with other legume species

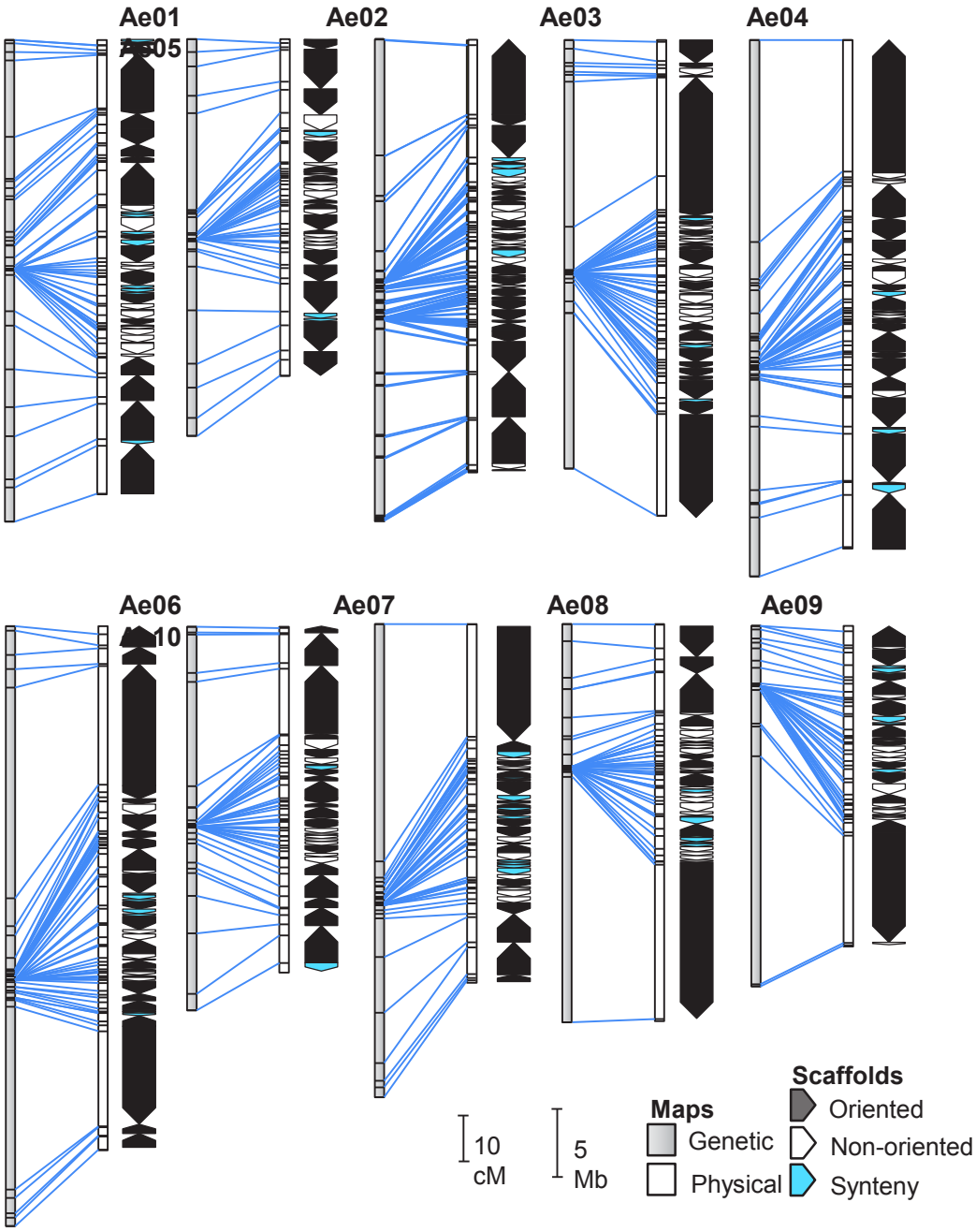
Legume species	Total genes count	Average gene length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
***A. evenia* (2n=20)**	**32,667**	**3,196**	**1,221**	**5.04**	**241**	**530**
L. angustifolius (2n=40)	38,688	4,384	1,456	6.79	293	614
A. duranensis (2n=20)	36,734	3,363	1,102	5.81	246	400
A. ipaiensis (2n=20)	41,840	3,269	1,050	5.66	242	407
A. hypogaea (2n=40)	83,709	5,077	1,590	6,82	233	600
G. max (2n=40)	46,290	3,673	1,253	5.80	216	419
P. vulgaris (2n=22)	27,433	5,693	1,332	5.69	290	457
C. cajan (2n=22)	31,716	4,304	1,374	6.29	305	636
V. radiata (2=22)	28,822	4,219	1,430	6.65	324	597
C. arietanum(2n=16)	28,88	3,970	1,392	6.35	317	577
M. truncatula (2n=16)	37,734	3,749	1,426	5.8	348	556

Supplementary Table 8. RNAseq mapping results and transcriptome assemblies

Library	Total read pairs	Total mapped	% hit	Total genes	Gene coverage
Root N-	32,472,115	31,766,466	97,83	22,991	70,4 %
Root N+	34,055,125	33,078,564	97,13	22,716	69,5 %
Nodule 4d	29,297,488	28,754,643	98,15	22,391	68,5 %
Nodule 7d	26,970,778	26,121,792	96,85	22,277	68,2 %
Nodule 14d	34,275,689	32,655,873	95,27	22,025	67,4 %
Stem	34,132,254	32,942,927	96,52	22,045	67,5 %
Leaf	31,682,921	30,849,149	97,37	21,395	65,5 %
Flower	29,413,019	28,763,221	97,79	22,245	68,1 %
Pod	31,078,257	30,113,451	96,90	22,948	70,2 %
Total	283,377,646	275,046,086	97,70	25,301	77,5 %

Supplementary Table 9. BUSCO analysis of genome annotation for *Aeschynomene evenia*

	A. evenia
Complete total BUSCOs	1360 (94.4%)
Complete sinlge-copy BUSCOs	1256 (87.2%)
Complete duplicated BUSCOs	104 (7.2%)
Fragmentted BUSCOs	30 (2.1%)
Missing BUSCOs	50 (3.5%)
Total BUSCO groups searched	1440



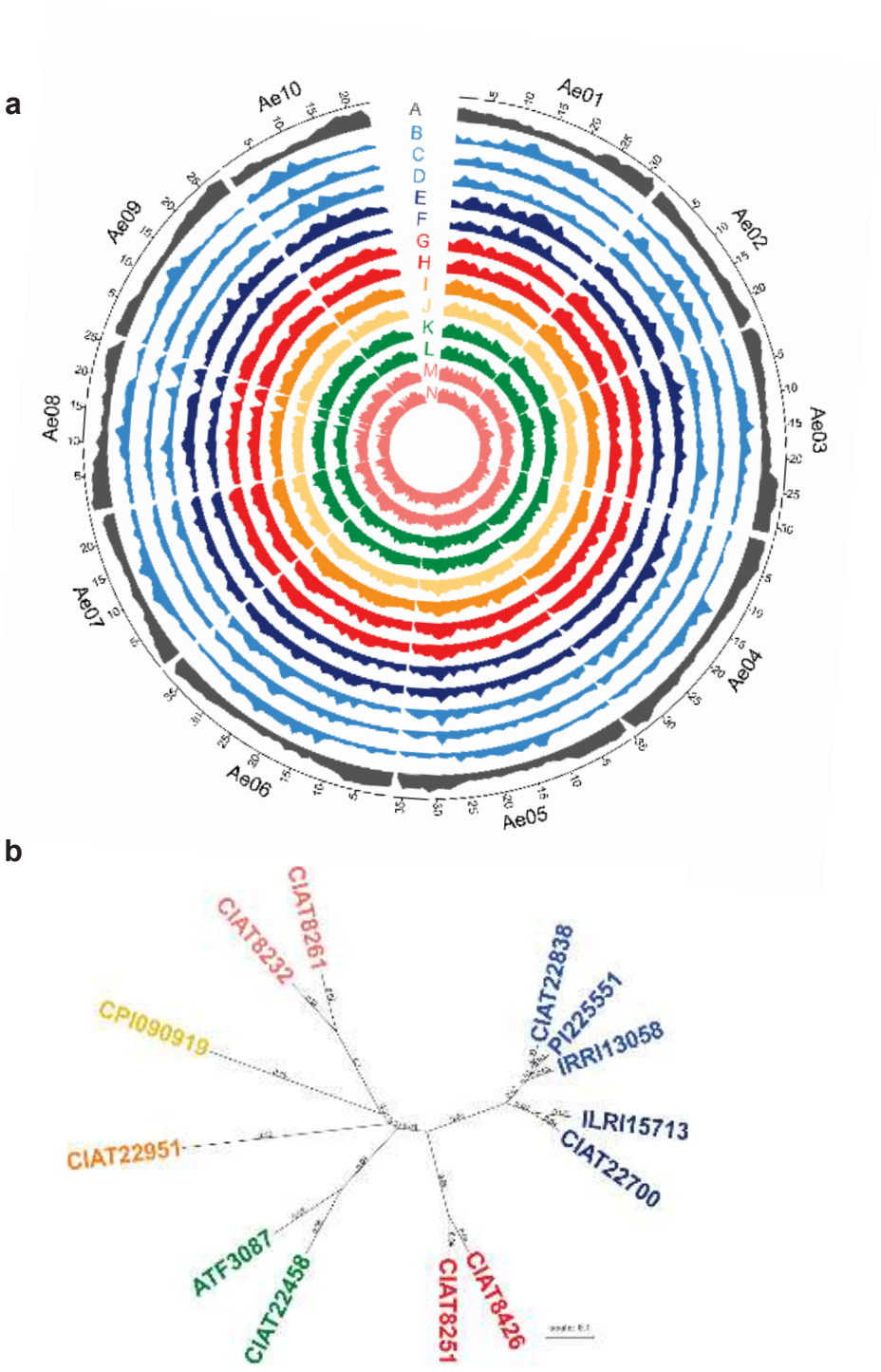
Supplementary Figure 4. Graphical representation of *A. evenia* pseudomolecules in reference to the genomic SNP-based map. *Left*, correspondence between the genetic map resulting from the GBS analysis and the physical map. Blue connecting lines represent the first and last SNP marker of each scaffold. *Right*, Scaffolds assembled into pseudomolecules with origin (anchored to the genetic map or placed by synteny) and orientation indicated.

Supplementary Table 10. Functional annotation of predicted genes for *Aeschynomene evenia*

	Number	Percent (%)
Swissprot	20,865	63,8
InterPro	22,268	68,1
GO	16,353	50,0
KEGG	9,786	29,9
Annotated	23,544	72
Un-annotated	9,156	28
Total	32,667	100

Supplementary Table 11. Summary of non-coding RNA genes predicted in the *Aeschynomene evenia* genome

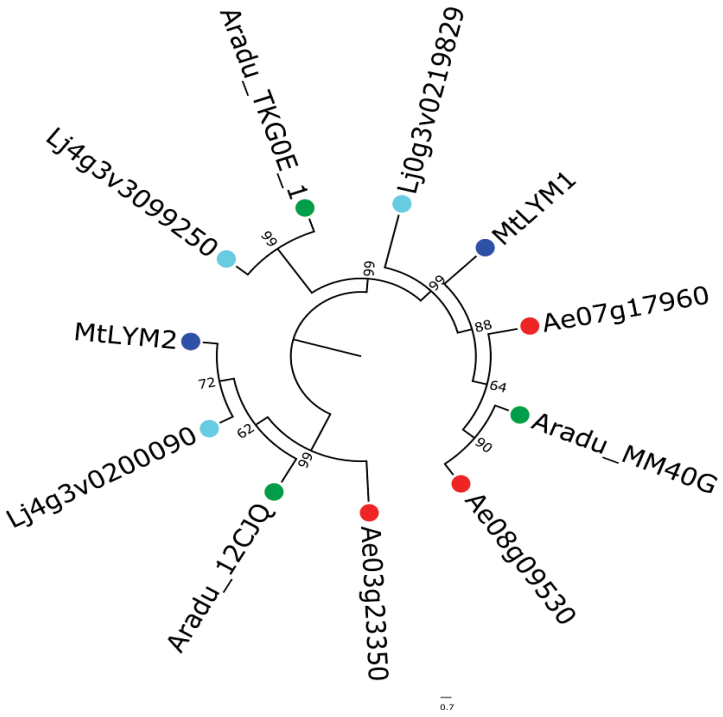
Type		Copy	Average length (bp)	Total length (bp)	% of genome
miRNA		119	137,73	16390	0,004360 %
tRNA		671	73,66	49427	0,013147 %
rRNA	rRNA	5134	753,84	3870207	1,029439 %
	18S	436	1830,86	798257	0,212329 %
	28S	445	5753,73	2560410	0,681046 %
	5.8S	423	153,36	64871	0,017255 %
	5S	3830	116,62	446669	0,118810 %
snRNA	snRNA	634	112,30	71199	0,018938 %
	CD-box	506	104,58	52917	0,014075 %
	HACA-box	45	133,04	5987	0,001592 %
	splicing	83	148,13	12295	0,003270 %



Supplementary Figure 5. *Aeschynomene evenia* diversity. **a**, SNP density identified by resequencing of 13 accessions of *A. evenia*. Gene density is in the outer track (A in grey). From the outer to inner track: (B) CIAT22838, (C) PI225551, (D) IRRI13058, (E) CIAT22700, (F) ILRI15713, (G) CIAT8426, (H) CIAT8251, (I) CIAT22951, (J) CPI090919, (K) ATF3087, (L) CIAT22458, (M) CIAT8232, (N) CIAT8261. The SNP density is represented in 1 Mb bins. **b**, *A. evenia* genetic diversity. Maximum-likelihood phylogenetic tree of *A. evenia* accessions based on SNPs. b and c, colors refer to previously identified genotypes: East Africa (pale blue), West Africa (dark blue), Brazil I (red), Peru (orange), Mexico (yellow), Argentina (green) and Brazil II (pink). .

Supplementary Table 12. Summary of repetitive sequences in the *Aeschynomene evenia* genome

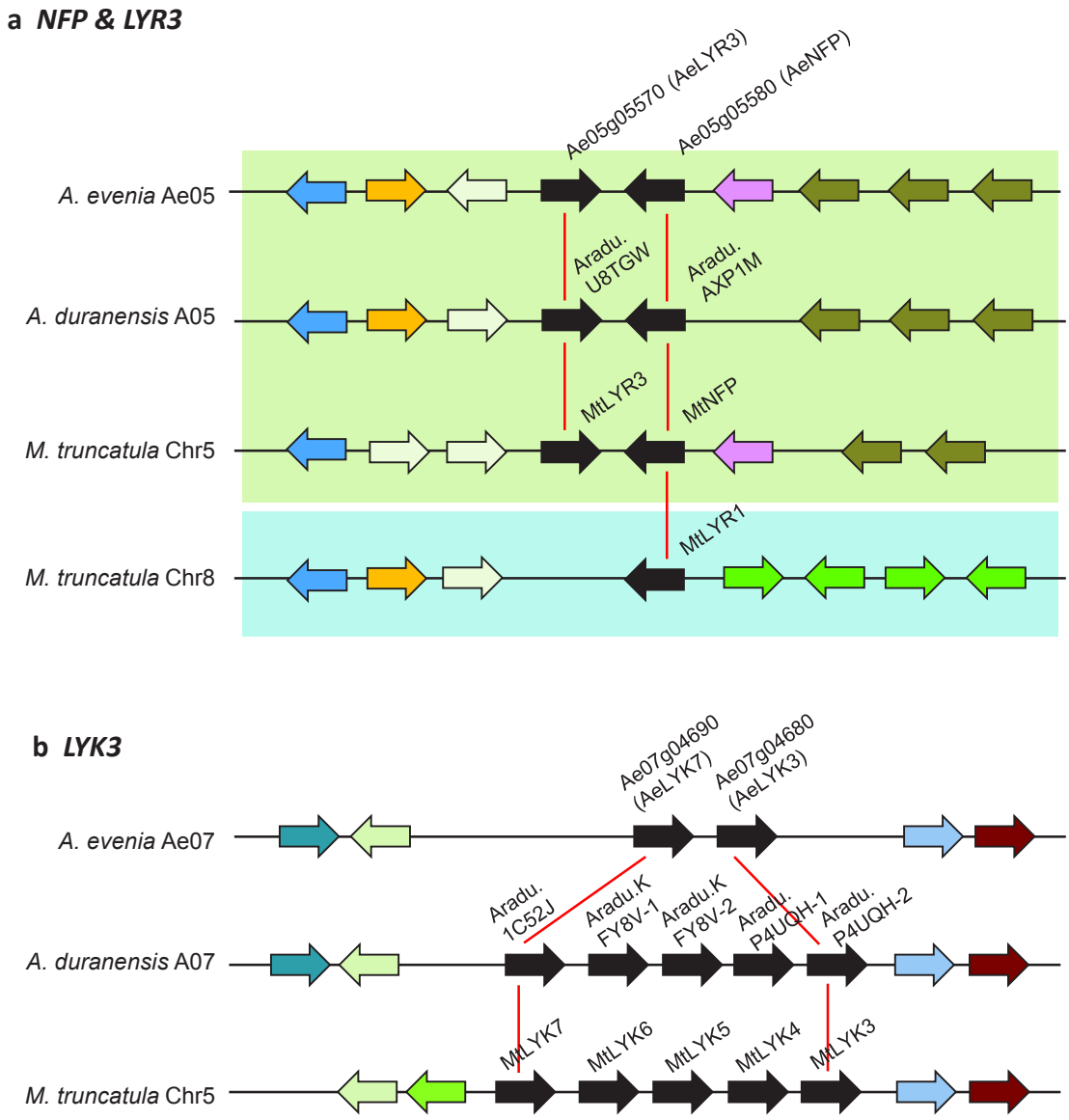
Super families of Transposable Elements	Occupied length (bp)	In total repeat (%)	In genome (%)	Repeat number
Class I (Retroelements)	91,609,865	63,67	24,34	85,853
LINEs	1,897,634	3,84	0,50	5,174
SINEs	337,483	2,14	0,09	2,892
LTR-Gypsy	40,297,264	26,53	10,71	35,780
LTR-Copia	49,077,484	31,15	13,04	42,007
Class II (DNA transposons)	18,143,366	36,33	4,82	48,993
Unclassified	84,038,871	-	22,33	246,847
Satellites	19,187	-	0,01	150
Low complexity	1,025,012	-	0,27	19,847
Simple repeats	6,398,582	-	1,70	123,546



Supplementary Figure 6. Phylogenetic tree of the LYM genes in *A. evenia* (red), *Arachis duranensis* (orange), *M. truncatula* (blue) and *Lotus japonicus* (green). Node numbers represent boostrap values (% of 1000 replicates). The scale bar represents substitutions per site.

Supplementary Table 13. Telomere repeat location and organization in the *Aeschynomene evenia* genome

Chromosome	Position of telomere on chromosome	Start of telomeric array	End of telomeric array	Size of telomeric array (bp)	Telomeric repeat sequence	Number of telomeric repeats
Ae01	BEGIN	-	-	-	-	-
	END	32757029	32760111	3083	TTTTAGG	444
Ae02	BEGIN	1	2099	2099	AACCCTA	299
	END	24237933	24242854	4922	TTTAGGG	694
Ae03	BEGIN	1	1914	1914	AACCCTA	274
	END	31944852	31946466	1615	TTAGGGT	230
Ae04	BEGIN	1	2018	2018	AAACCCT	288
	END	35015777	35017145	1369	AGGGTTT	196
Ae05	BEGIN	1	1505	1505	CCTAAAC	218
	END	36292128	36294833	2706	TTTAGGG	388
Ae06	BEGIN	1	2718	2718	AAACCCT	393
	END	38542264	38544573	2310	AGGGTTT	333
Ae07	BEGIN	-	-	-	-	-
	END	24664443	24667243	2801	TAGGGTT	401
Ae08	BEGIN	-	-	-	-	-
	END	26251397	26254046	2650	TTTAGGG	385
Ae09	BEGIN	-	-	-	-	-
	END	28843023	28845408	2386	TTTAGGG	340
Ae10	BEGIN	1	4033	4033	AACCCTG	577
	END	23575706	23578249	2544	TAGGGTT	367
Unknown		964419	967348	2930	AAACCCT	420
		2637359	2640006	2648	TTTAGGG	379
		15147461	15149374	1914	TTTAGGG	274
		22681401	22683802	2402	CCTAAAC	346
		66231206	66233269	2064	TTTAGGG	295
		70589138	70591736	2599	TTTAGGG	372



Supplementary Figure 7. Syntenic localization of Nod factor receptor genes in *A. evenia*. a and **b**, schematic representation of microsynteny analysis for *NFP* and *LYK3* between *A. evenia*, *A. duranensis* and *M. truncatula*. Note that these genes are present in gene tandem or cluster. Orthologous/paralogous gene pairs are indicated through the use of a common colour. Orphan genes are not represented for clarity. Green and blue rectangles highlight duplicated regions derived from the ~58 MYA WGD event. Note that some genes of the LYK cluster were manually reannotated.

Supplementary Table 14. Information on the *Aeschynomene evenia* lines

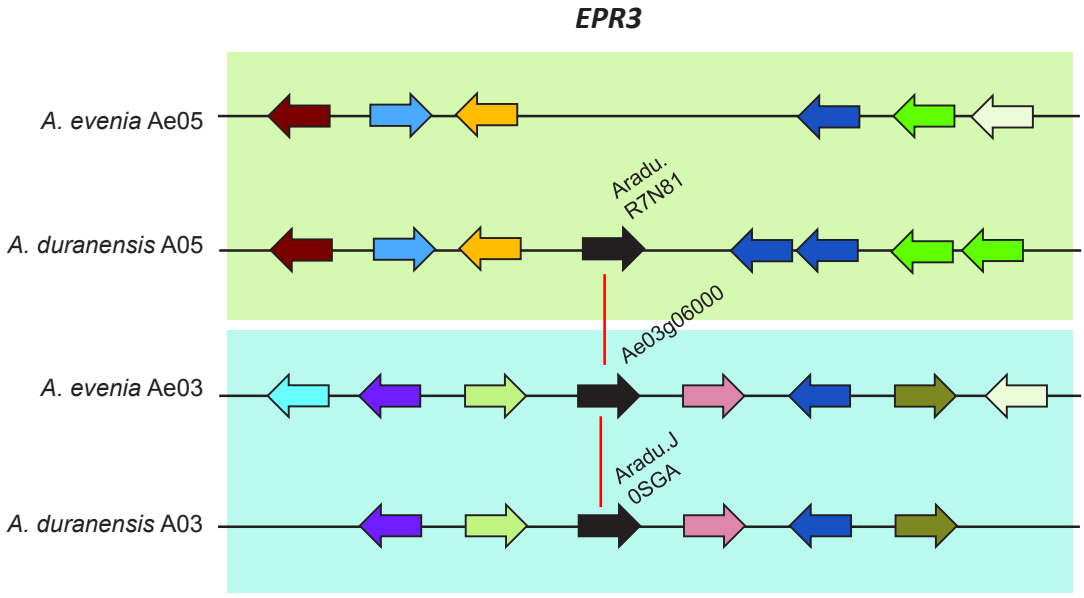
Accession	Genotype	Genome size (1C)	Origin	Seedbank
ATF3087	Argentina	403 Mb	Argentina	AusPGRIS
CIAT22458	Argentina	417 Mb	Argentina	CIAT
CIAT8251	Brazil I	417 Mb	Brazil	CIAT
CIAT8426	Brazil I	417 Mb	Brazil	CIAT
CIAT8261	Brazil II	403 Mb	Brazil	CIAT
CIAT8232	Brazil II	400 Mb	Brazil	CIAT
PI225551	Eastern Africa	412 Mb	Zambia	USDA
IRRI13058	Eastern Africa	-	Madagascar	IRRI
CIAT22838*	Eastern Africa	400 Mb	Malawi	CIAT
CPI090919	Mexico	412 Mb	Mexico	AusPGRIS
CIAT22951	Peru	436 Mb	Peru	CIAT
CIAT22700	Western Africa	417 Mb	Senegal	CIAT
ILRI15713	Western Africa	-	Chad	ILRI

* Reference line used for the whole genome sequencing

Supplementary Table 15. Resequencing statistics on *Aeschynomene evenia* lines

Line	Number of reads	Total length (bp)	Genome coverage	Reads mapped (%)	Number of SNPs
ATF3087	58,746,587	8,811,988,050	22 x	99,1	2773247
CIAT22458	56,569,871	8,485,480,650	21 x	97,8	2737861
CIAT22838*					105568
CIAT8251	49,072,110	7,360,816,500	18 x	97,3	2393403
CIAT8426	44,992,492	6,748,873,800	17 x	99,0	2415358
CIAT8261	51,754,517	7,763,177,550	19 x	97,1	2847902
CIAT8232	47,269,407	7,090,411,050	18 x	98,5	2855470
PI225551	45,392,112	6,808,816,800	17 x	97,8	707837
IRRI13058	58,663,284	8,799,492,600	22 x	99,2	784002
CPI090919	48,200,765	7,230,114,750	18 x	97,0	2880438
CIAT22951	62,565,710	9,384,856,500	23 x	98,9	3050650
CIAT22700	63,526,697	9,529,004,550	24 x	99,1	1303397
ILRI15713	49,168,143	7,375,221,450	18 x	97,5	1415653

* Mapping of the Miseq sequences (Supplementary Table 2) obtained for the reference line on the genome sequence



Supplementary Figure 8. Syntenic localization of EPS receptor genes in *A. evenia*. Schematic representation of microsynteny analysis for the *EPR3* gene (black arrow) between *A. evenia* and *A. duranensis*. Orthologous/paralogous gene pairs are indicated through the use of a common colour. Orphan genes are not represented for clarity. Green and blue rectangles highlight duplicated regions derived from the ~58 MYA WGD event. Note the absence of an *EPR3* ortholog but the retention of a paralog in *A. evenia*.

Supplementary Table 16. Information on the *Aeschynomene* species of the Nod-independent clade

Species	Accession	Numbers of chromosomes	Genome size (1C)	Origin	Seedbank
A. ciliata	IRR 013078	2n=20	520 Mb	Colombia	IRRI
A. deamii	LSTM24	2n=20	930 Mb	Mexico	LSTM
A. denticulata	IRRI013003	2n=20	590 Mb	Brazil	IRRI
A. evenia ssp. *evenia*	CIAT22838*	2n=20	400 Mb	Malawi	CIAT
A. evenia ssp. *serrulata*	IRFL6945	2n=20	460 Mb	USA	USDA
A. filosa	CIAT22466	2n=20	400 Mb	Mexico	CIAT
A. rudis	Matt Lavin #82	2n=20	500 Mb	USA	LSTM
A. scabra	LSTM26	2n=20	501 Mb	Mexico	LSTM
A. selloi	CPI104040	2n=20	580 Mb	Argentina	AusPGRIS
A. sensitiva	LSTM28	2n=20	720 Mb	Costa Rica	CIAT
A. sp (328)	CIAT8499	2n=20	730 Mb	Brazil	CIAT
A. tambacoundensis	LSTM60	2n=20	370 Mb	Senegal	LSTM

* Accession used for the genome sequencing

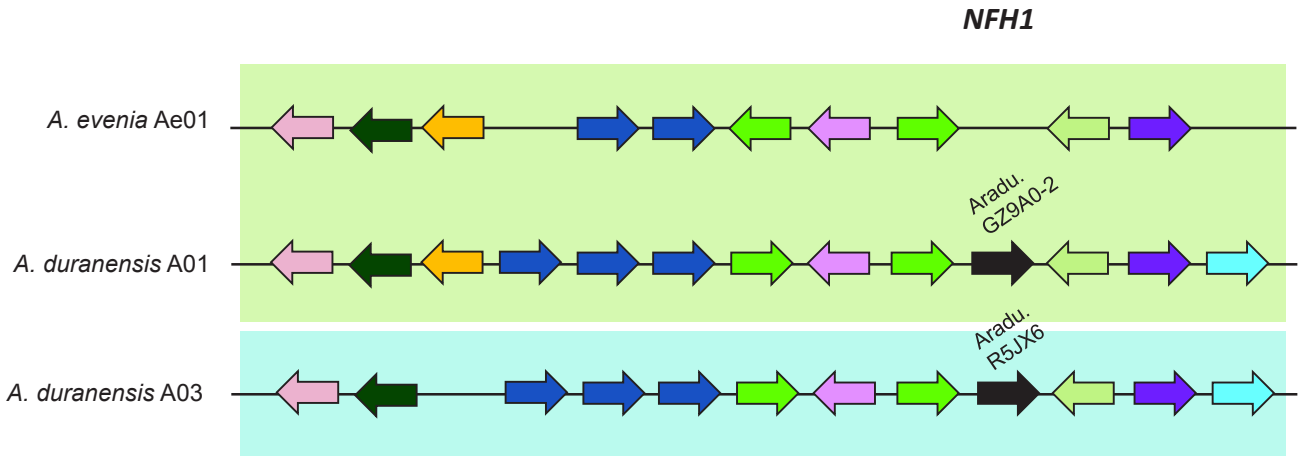
* Mapping of the Miseq sequences (Supplementary Table 2) obtained for the reference line on the genome sequence

Supplementary Table 17. Summary of the *Aeschynomene* species transcriptome assemblies

Species	Illumina sequencing	Number of reads	Numbers of contigs	N50	Number of genes	BUSCO score (n:1440)
A. ciliata	PE	55873472	37958	2274	33101	C:86.5%[S:67.1%,D:19.4%],F:5.6%,M:7.9%
A. deamii	PE	75693011	49237	2257	40246	C:84.7%[S:64.1%,D:20.6%],F:6.5%,M:8.8%
A. denticulata	PE	60642100	38668	2263	33464	C:86.3%[S:66.7%,D:19.6%],F:4.9%,M:8.8%
A. evenia ssp. *evenia*	PE	*	45932	1999	25221	C:89.5%[S:72.8%,D:16.7%],F:3.8%,M:6.7%
A. evenia ssp. *serrulata*	Single	**	138675	1599	28494	C:93.4%[S:85.4%,D:8.0%],F:1.2%,M:5.4%
A. filosa	PE	60176167	38358	2196	33769	C:86.9%[S:66.1%,D:20.8%],F:4.7%,M:8.4%
A. rudis	PE	56727897	41478	2268	34878	C:85.9%[S:65.3%,D:20.6%],F:6.0%,M:8.1%
A. scabra	PE	59473825	42710	2211	35756	C:86.4%[S:63.4%,D:23.0%],F:4.9%,M:8.7%
A. selloi	PE	76792277	39924	2325	34971	C:85.3%[S:63.6%,D:21.7%],F:5.9%,M:8.8%
A. sensitiva	PE	77550240	41187	2292	35861	C:83.9%[S:62.9%,D:21.0%],F:6.4%,M:9.7%
A. sp (328)	PE	80262940	48651	2211	40878	C:85.6%[S:61.3%,D:24.3%],F:6.0%,M:8.4%
A. tambacoundensis	PE	116821813	42307	2380	38564	C:80.0%[S:60.7%,D:19.3%],F:9.5%,M:10.5%

*: total data obtained for the gene expression analysis

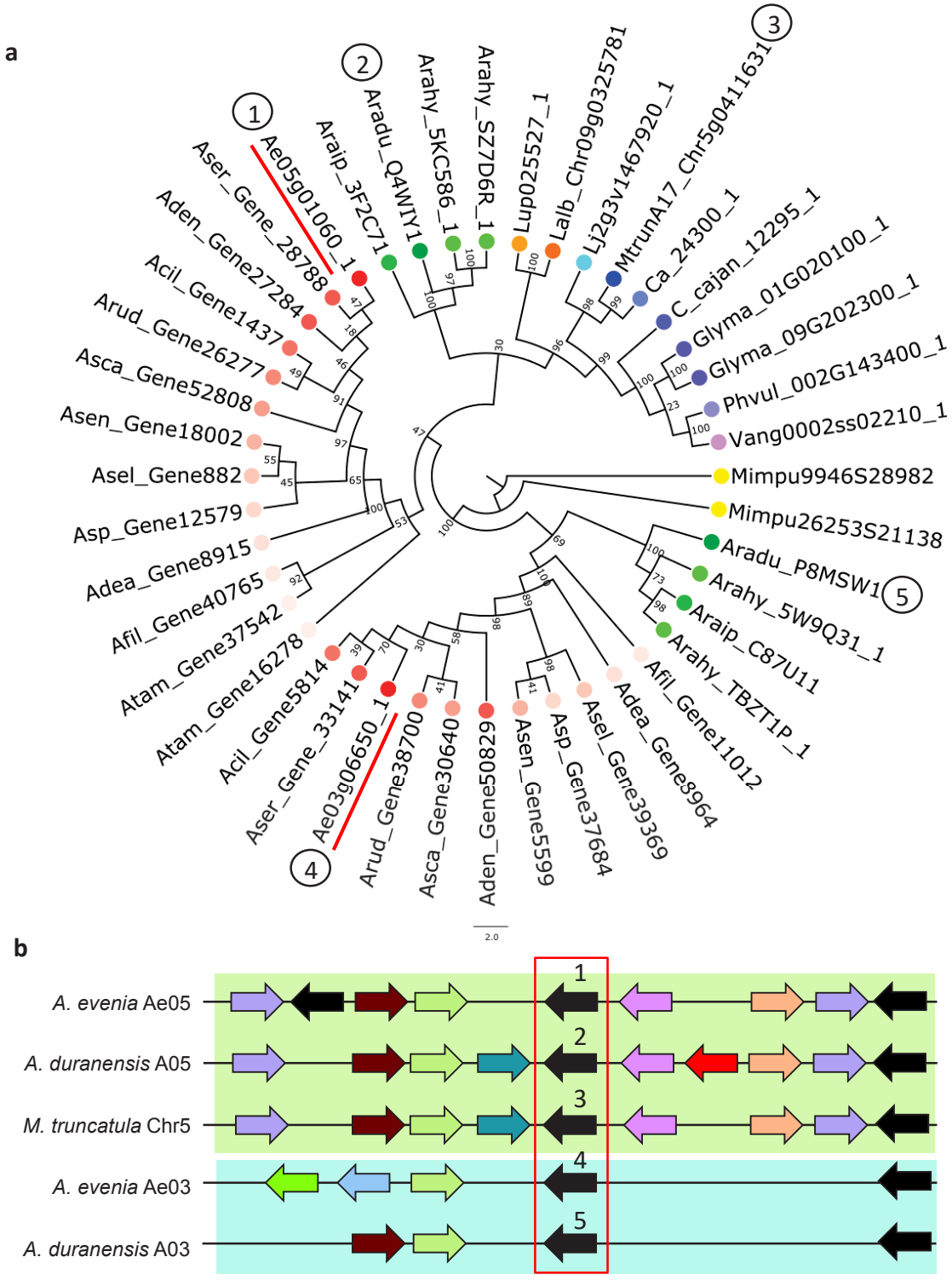
**: total data from ref. 17.



Supplementary Figure 9. Genomic evolution of the genes coding for Nod factor hydrolases. Schematic representation of microsynteny analysis for the (**a**) ortholog and (**b**) paralog of *NFH1* (black arrow) between *A. evenia* and *A. duranensis*. Orthologous/paralogous gene pairs are indicated through the use of a common colour. Orphan genes are not represented for clarity and some the Aradu.GZ9A0 gene was manually re-annotated. Green and blue rectangles highlight duplicated regions derived from the ~58 MYA WGD event. Note the absence of an *NFH1* ortholog and also of the paralogous genomic region in *A. evenia*.

Supplementary Table 18. Statistics of OrthoFinder analysis

Species	Sequences	Genes number	Genes in orthologous groups	Unclustered genes	Total orthologous groups
Aeschynomene ciliata	Transcriptome	33101	32397	704	16746
A. deamii	Transcriptome	40246	38481	1765	17479
A. denticulata	Transcriptome	33464	32696	768	16796
A. evenia var. *evenia*	Genome	32667	31115	1552	16556
A. *evenia* var. *serrulata*	Transcriptome	28494	27358	1136	16587
A. filosa	Transcriptome	33769	32798	971	16417
A. rudis	Transcriptome	34878	34068	810	16998
A. scabra	Transcriptome	35756	34854	902	17115
A. selloi	Transcriptome	34971	34139	832	16975
A. sensitiva	Transcriptome	35861	34909	952	17042
A. sp 328	Transcriptome	40878	39599	1279	17924
A. tambacoundensis	Transcriptome	38564	37469	1095	17528
Arachis hypogaea	Genome	84714	79994	4720	23161
Arachis duranensis	Genome	36734	34870	1864	19114
Arachis ipaiensis	Genome	41840	39919	1921	20337
Cicer arietanum	Genome	30686	29149	1537	15095
Cajanus cajan	Genome	40071	38791	1280	16974
Chamaecrista fasciculata	Genome	32832	31330	1502	14900
Glycine max	Genome	88647	82885	5762	18496
Lupinus albus	Genome	38258	35213	3045	16779
Lupinus angustifolius	Genome	33072	31788	1284	15356
Lotus japonicus	Genome	48105	43364	4741	18939
Mimosa pudica	Genome	33108	31417	1691	14015
Medicago truncatula	Genome	44624	38835	5789	17993
Phaseolus vulgaris	Genome	36995	36182	813	15997
Vigna angularis	Genome	36692	34810	1882	16309

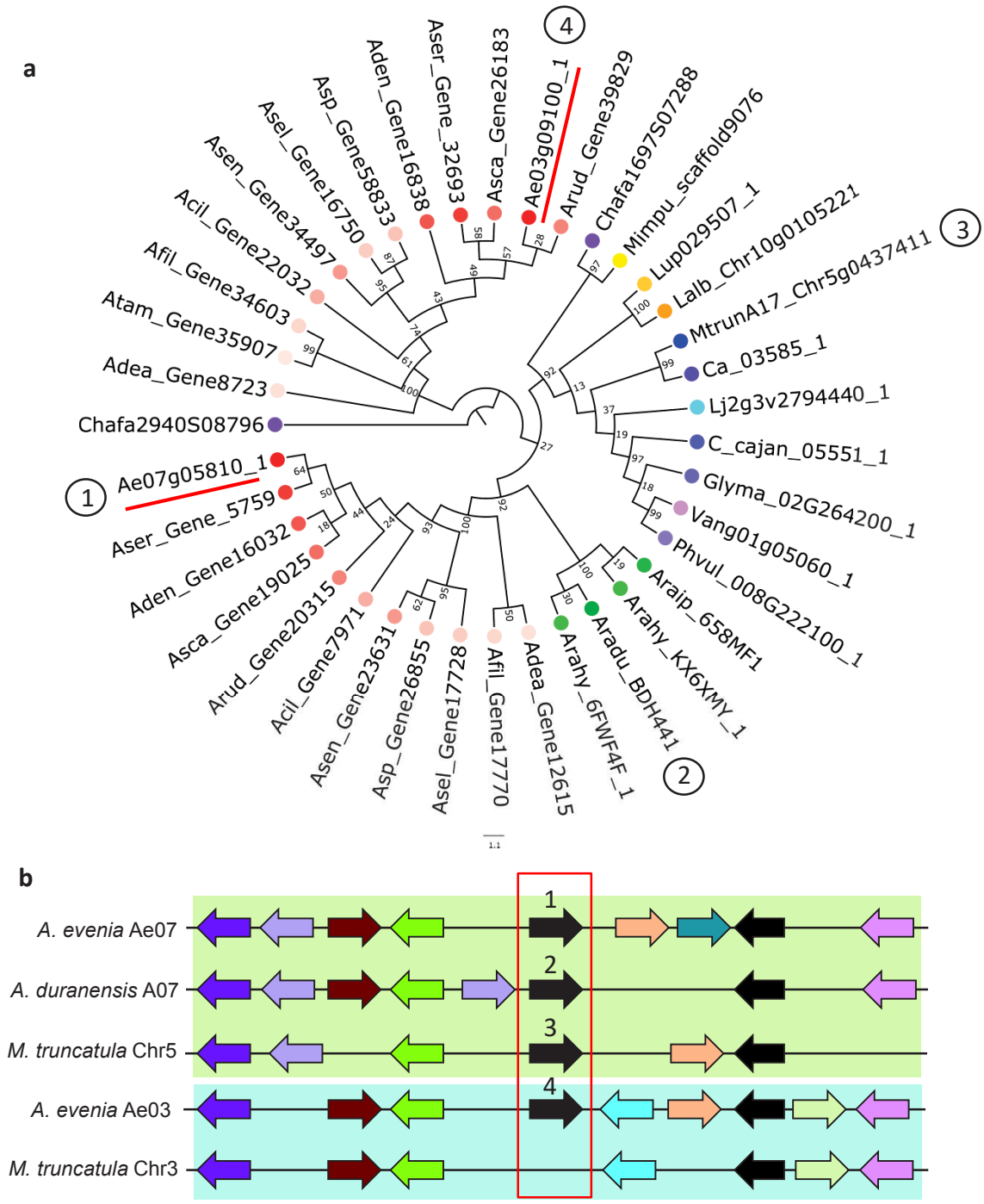


Supplementary Figure 10. Variation in orthologous/paralogous relationships in *SYMRK*. (**a**) Phylogenetic tree showing duplicated *SYMRK* genes for *Aeschynomene* spp. (pink) and *Arachis* spp (green). SYMRK genes of *A. evenia* are underlined in red. Numbers in circles refer to the genes analysed for microsynteny. (**b**) Schematic representation of microsynteny analysis of *SYMRK* (black arrow with numbers referring to the gene IDs in the phylogenetic tree) between *A. evenia*, *A. duranensis* and *M. truncatula*. Orthologous/paralogous gene pairs are indicated through the use of a common colour. Orphan genes are not represented for clarity. Green and blue rectangles highlight duplicated regions.

Supplementary Table 19. Somatic and germinal effects of EMS dosages in *A. evenia*

	EMS dose (%)					
	0	0,3	0,35	0,4	0,5	0,6
Early radicle outgrowth[a]	0 %	4 %	9 %	14 %	20 %	31 %
Seedling survival[b]	94 %	92 %	78 %	76 %	69 %	43 %
Fertile plants[c]	100 %	90 %	47 %	30 %	30 %	0 %
n seed/pod[d]	8,2	8,4	7,2	6,5	3,5	0,0

a, % of imbibed seeds showing a radicle outgrowth following the EMS treatment and before germination induction. b, Seedling survival scored as effective cotyledon opening 7 days after the EMS treatment. c, Plant fertility scored as the ability to produce at least one pod in three months. d, Number of seeds per pod were scored individually on >40 randomly selected pods for each EMS treatment. Dose 0% corresponds to water treatment. Each treatment was performed on 260 seeds.



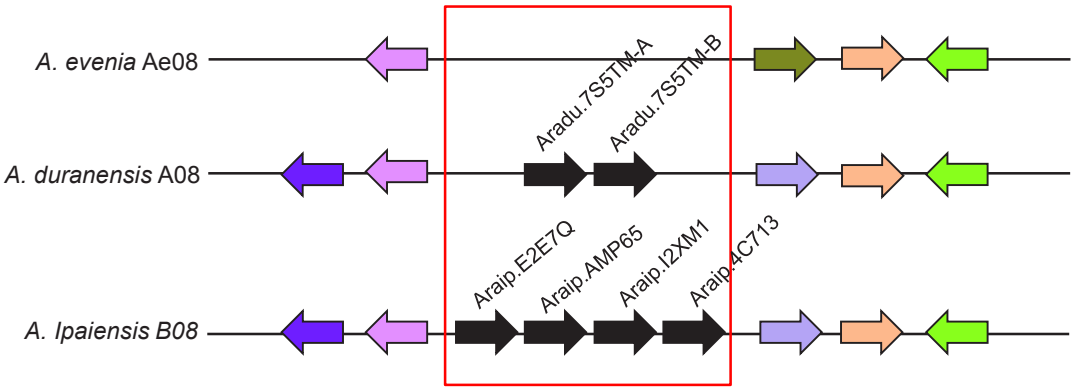
Supplementary Figure 11. Variation in orthologous/paralogous relationships in *PUB1*. (**a**) Phylogenetic tree showing duplicated *PUB1* genes for *Aeschynomene* spp. (pink) and *Arachis* spp (green). *PUB1* genes of *A. evenia* are underlined in red. Numbers in circles refer to the genes analysed for microsynteny. (**b**) Schematic representation of microsynteny analysis of *PUB1* (black arrow with numbers referring to the gene IDs in the phylogenetic tree) between *A. evenia*, *A. duranensis* and *M. truncatula*. Orthologous/paralogous gene pairs are indicated through the use of a common colour. Orphan genes are not represented for clarity. Green and blue rectangles highlight duplicated regions.

Supplementary Table 20. ***Aeschynomene evenia*** **Nod- mutants with their phenotypic, molecular and genetic data**

Gene	Mutant	Mutation effect		Genetic determinism		Stem/root nodulation on F2 Nod+:Nod- plants	Mutation sequencing on F2 Nod- plants	Complementation groups
		Nucleotide change	Amino-acid change	F2 Nod+:Nod-	(*, P> 0.05)			
AeCCaMK (Ae08g13330)	H34	G_{3905} to A	A_{450} to STOP	F2 381:149	3:1*	12 -/- ; 6 +/+	100%**	a
	J31	G_{2405} to A	S_{341} to STOP	F2 437:143	3:1*	12 -/- ; 6 +/+	13/13*	
	T26	C_{1204} to T	P_{218} to L	F2 226:71	3:1*	12 -/- ; 6 +/+	15/15*	
	AK7	C_{1965} to T	Splice site	F2 219:65	3:1*	12 -/- ; 6 +/+	-	
AeCRK (Ae05g12380)	I10	G_{2228} to A	Splice site	F2 151:60	3:1*	12 -/- ; 6 +/+	100%**	b
	J42	G_{1062} to A	G_{354} to E	F2 144:42	3:1*	12 -/- ; 6 +/+	100%**	
AeCYCLOPS (Ae05g02230)	L33	G_{2372} to A	Splice site	F2 159:58	3:1*	12 -/- ; 6 +/+	-	c
	S31	G_{2325} to A	Splice site	F2 88:153	2:1*	12 -/- ; 6 +/+	16/16*	
AeNIN (Ae07g00100)	E26	C_{1919} to T	Q_{352} to STOP	nt		nt	-	nt
	G40	A_{2800} to G C_{2768} to T	K_{647} to E P_{636} to L	F2 239:68	3:1*	12 -/- ; 6 +/+	16/16*	d
	S11	A_{339} to T	R_{114} to STOP	F2 229:69	3:1*	12 -/- ; 6 +/+	15/15*	
	V20	T_{142} to A	L_{48} to Q	F2 206:61	3:1*	12 -/- ; 6 +/+	16/16*	
	Y11	C_{3173} to T	R_{729} to STOP	F2 441:134	3:1*	12 -/- ; 6 +/+	14/14*	
	AD8	C_{3048} to T	L_{687} to F	F2 245:61	3:1	12 -/- ; 6 +/+	-	
AeNSP2 (Ae08g09660)	C10	C_{692} to T	P_{231} to L	F2 423:147	3:1*	12 -/- ; 6 +/+	100%**	e
	J25	A_{430} to T	R_{144} to STOP	F2 220:68	3:1*	12 -/- ; 6 +/+	15/15*	
	M30	G_{1545} to A	W_{515} to STOP	F2 553:84	7:1*	12 -/- ; 6 +/+	100%**	
	AK8	C_{806} to T	S_{269} to F	F2 230:70	3:1*	8 -/- ; nt	20/20*	
AePOLLUX (Ae04g09220)	C15	G_{1699} to A	G_{237} to R	F2 222:68	3:1*	12 -/- ; 6 +/+	-	f
	E28	G_{5039} to A	W_{716} to STOP	F2 236:66	3:1*	12 -/- ; 6 +/+	16/16*	
	H10	G_{5476} to T	E_{833} to STOP	F2 408:134	3:1*	12 -/- ; 6 +/+	16/16*	
	J33	G_{3444} to A	G_{520} to E	F2 241:72	3:1*	12 -/- ; 6 +/+	16/16*	
	P34	G_{1931} to A	S_{314} to N	F2 359:123	3:1*	12 -/- ; 6 +/+	-	
	R16	A_{1944} to T	E_{318} to D	F2 229:75	3:1*	12 -/- ; 6 +/+	15/15*	

Note: Mutations were searched and identified on F2 Nod- plants either (*) by PCR amplication and sequencing or (**) by Mapping-by-Sequencing.

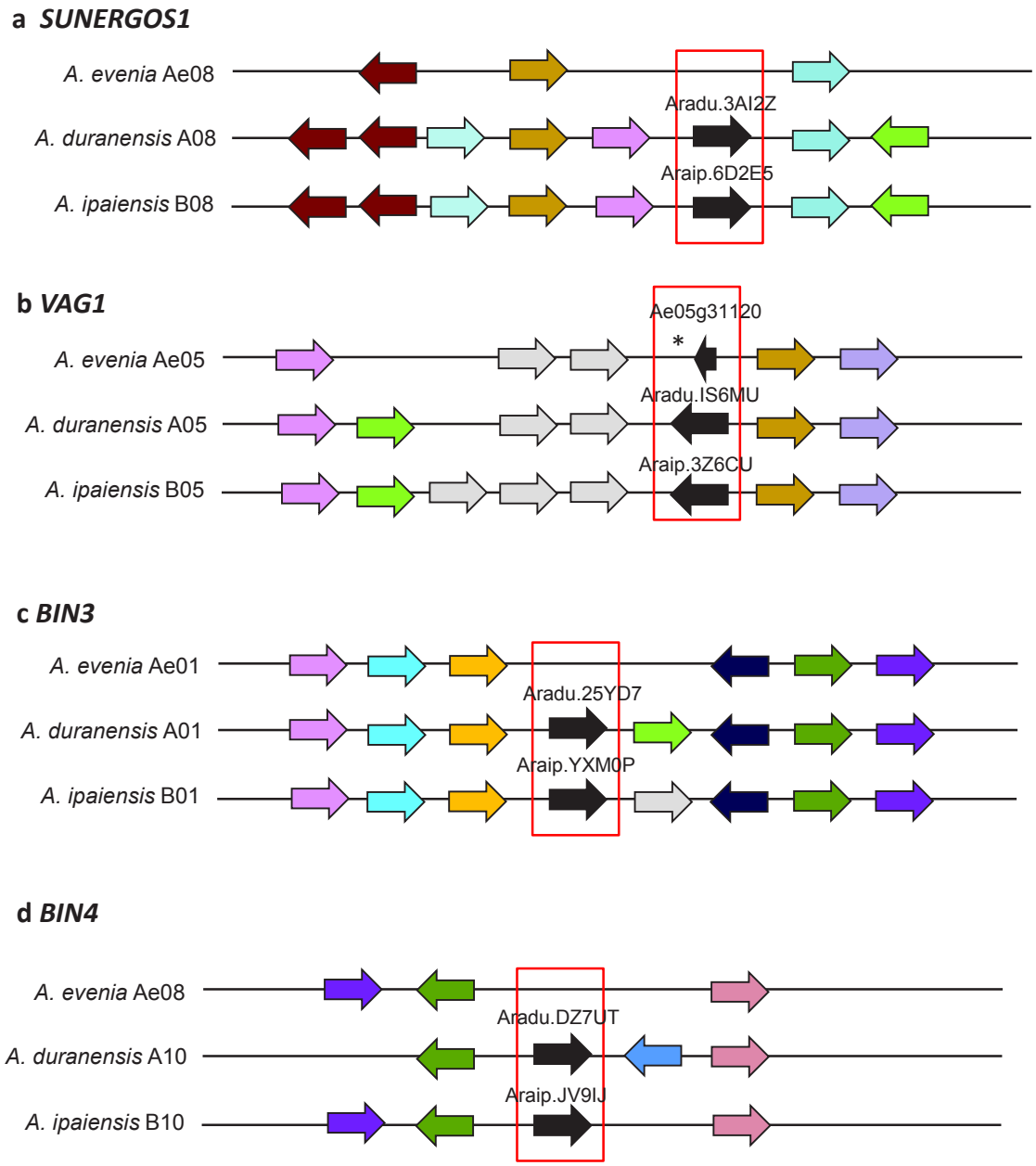
-: not determined.



Supplementary Figure 12. Genomic evolution of the *FLOT* genes. Syntenic relationships for the chromosome regions containing the *FLOT* genes (black arrows) between *Aeschynomene evenia, Arachis duranensis* and *Arachis ipaiensis*. Orthologous/paralogous gene pairs are indicated through the use of a common colour. Orphan genes are not represented for clarity.

Supplementary Table 21. Allelism tests performed on Nod- mutants

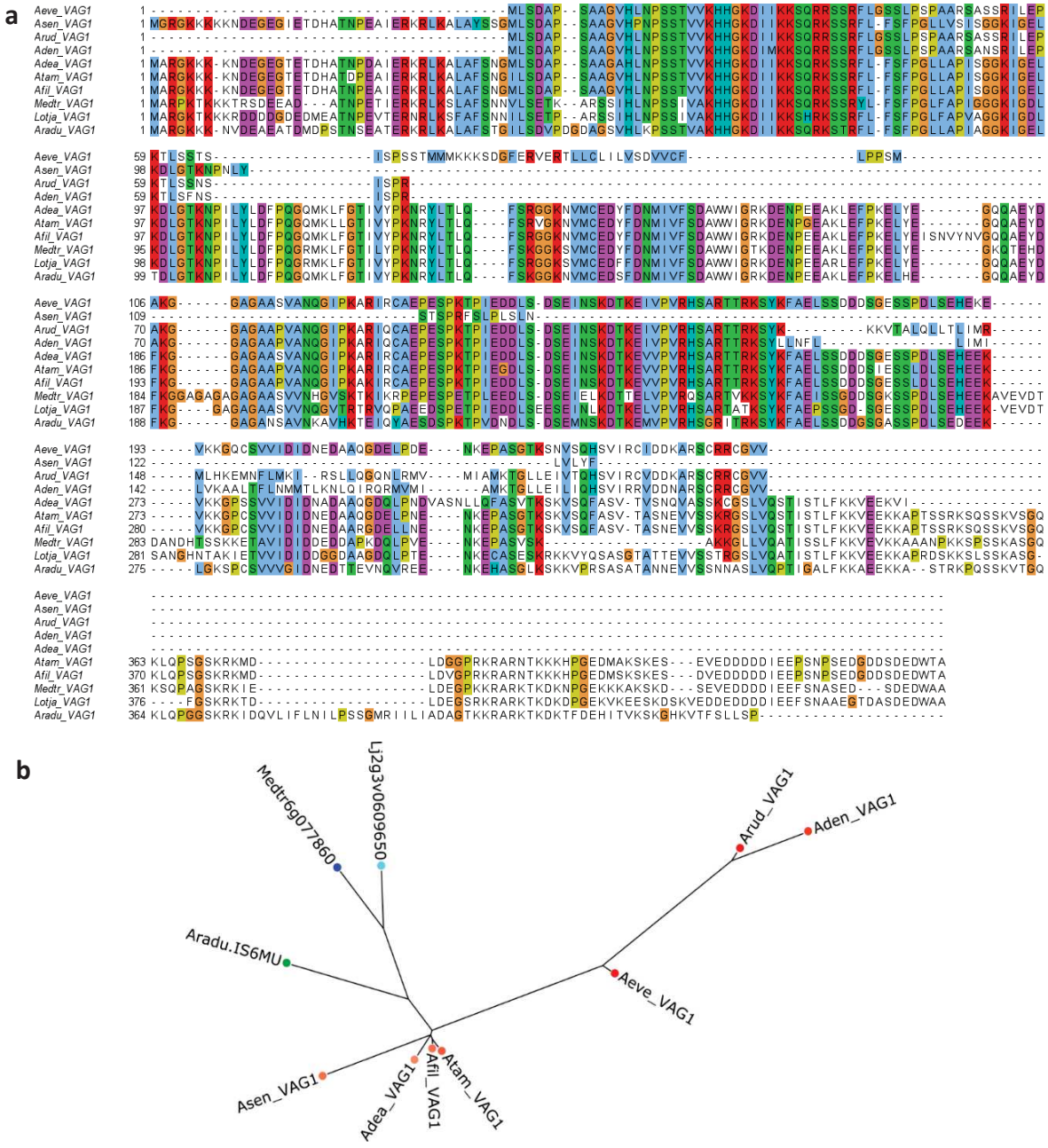
Gene	Crossing (♂ x ♀)	n F1 plants (n pods)	F1 phenotype	Complementation group
AeCCaMK	H34 x AK7	3(1)	Nod-	a
	H34 x T26	17(2)	Nod-	
	J31 x T26	18(3)	Nod-	
	T26 x H34	11(1)	Nod-	
AeCRK	I10 x J42	12(2)	Nod-	b
	J42 x I10	18(3)	Nod-	
AeCYCLOPS	S31 x L33	42(6)	Nod-	c
AeNIN	G40 x Y11	30(4)	Nod-	d
	S11 x AD8	7(1)	Nod-	
	S11 x V20	8(1)	Nod-	
	V20 x G40	6(1)	Nod-	
	V20 x Y11	6(1)	Nod-	
	Y11 x G40	26(3)	Nod-	
	Y11 x V20	8(1)	Nod-	
AeNSP2	AK8 x C10	11(3)	Nod-	e
	C10 x AK8	17(3)	Nod-	
	C10 x M30	14(2)	Nod-	
	J25 x M30	28(4)	Nod-	
AePOLLUX	C15 x E28	6(1)	Nod-	f
	C15 x J33	4(1)	Nod-	
	C15 x R16	12(2)	Nod-	
	E28 x H10	8(1)	Nod-	
	E28 x J33	9(1)	Nod-	
	E28 x R16	7(1)	Nod-	
	H10 x E28	5(1)	Nod-	
	H10 x R16	5(1)	Nod-	
	P34 x E28	11(2)	Nod-	
	J33 x E28	2(1)	Nod-	
	Q20 x J33	1(1)	Nod-	
	R16 x E28	19(3)	Nod-	



Supplementary Figure 13. Genomic evolution of the genes coding for components of the Topoisomerase VI complex. Syntenic relationships for (**a**) *SUNERGOS1*, (**b**) *VAG1*, (**c**) *BIN3* and (**d**) *BIN4* genes (black arrows) between *Aeschynomene evenia, Arachis duranensis* and *Arachis ipaiensis*. Orthologous/paralogous gene pairs are indicated through the use of a common colour. Orphan genes are not represented for clarity. * indicates the gene is present in a truncated form.

Supplementary Table 22. Mutations and their effect on the genetic linkage identified on chromosome 5 for the I10 and J42 mutants

Chromosome	Position (bp)	WT nucleotide	Mutant nucleotide	Quality	Mutation category	Gene	Effect	% Mutant allele	
								I10 mutant	J42 mutant
Ae05	6840402	G	A	648,384	Upstream_gene_variant	Ae05g09700	MODIFIER	92	0
	7045016	G	A	926,244	Downstream_gene_variant	Ae05g10010	MODIFIER	86	0
	7827316	C	T	717,814	Downstream_gene_variant	Ae05g11170	MODIFIER	0	86
	8028224	G	A	685,776	Upstream_gene_variant	Ae05g11410	MODIFIER	96	0
	8714708	G	A	519,198	Missense_variant	**Ae05g12380**	**Gly354Glu**	0	**100**
	8715482	G	A	594,911	Splice_donor_variant	**Ae05g12380**	**No splicing**	**100**	0
	9470738	G	A	322,809	intron_variant	Ae05g13280	MODIFIER	0	100
	10502432	T	A	456,098	Intergenic_region	Ae05g14240-Ae05g14250	MODIFIER	81	0
	10910970	G	A	381,325	Missense_variant	Ae05g14620	Pro140Ser	0	87

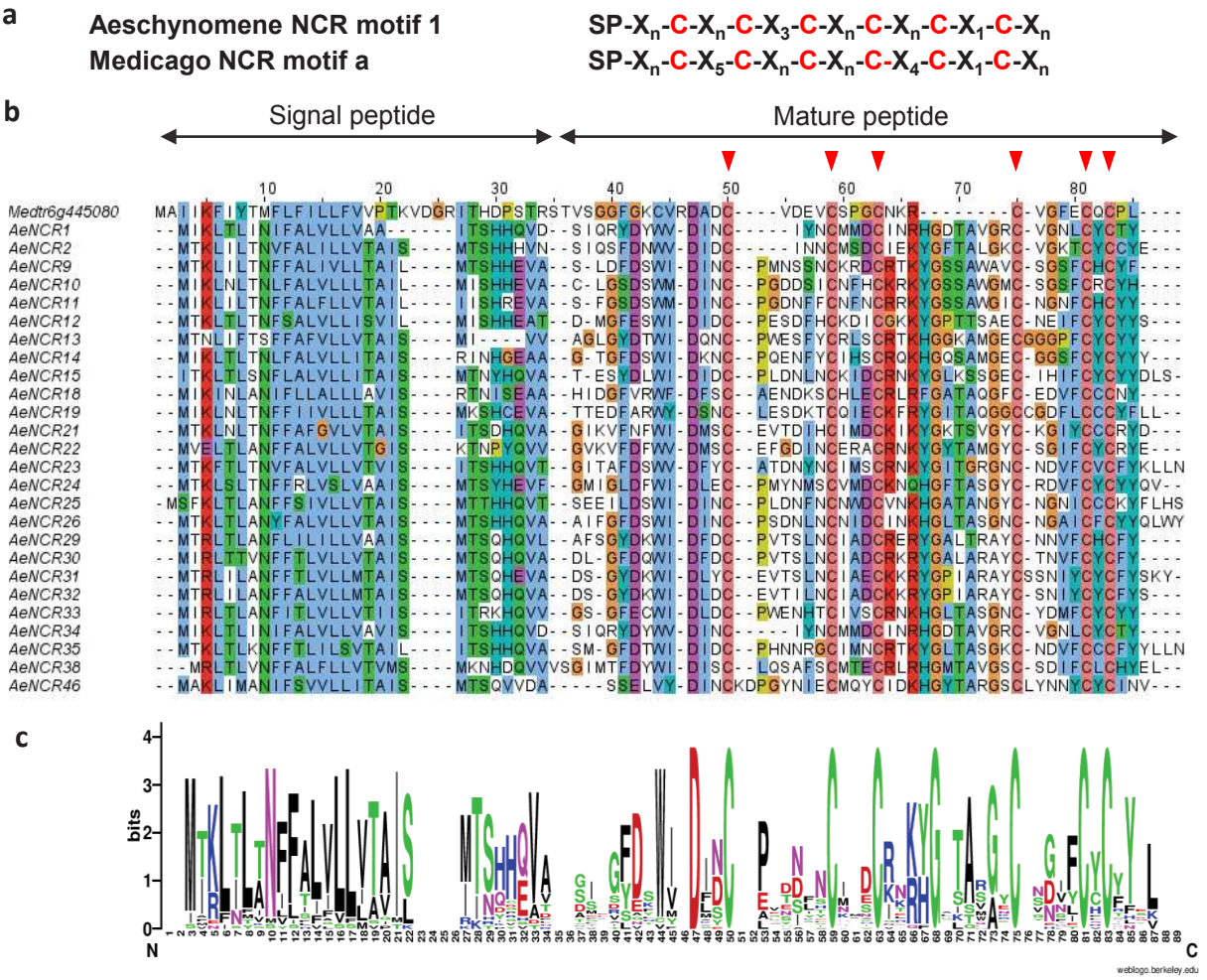


Supplementary Figure 14. Sequence analysis of the Topoisomerase VI-component VAG1 in *Aeschynomene* species. a, Sequence alignment (Muscle) of the predicted VAG1 protein from *A. evenia* (Aeve), *A. tambacoundensis* (Atam), *A. filosa* (Afil), *A. deamii* (Adea), *A. sensitiva* (Asen), *A. denticulata* (Aden), *A. rudis* (Arud), *Medicago truncatula* (Medtr*), Lotus japonicus* (Lj) and *Arachis duranensis* (Aradu), showing partial deletions in VAG1 for several *Aeschynomene* species. **b**, Bayesian tree of the predicted VAG1 proteins for the same legume species as in (**a**). Note distorted branches in the *Aeschynomene* group in link with the presence of VAG1 truncated forms.

Supplementary Table 23. Functional complementation test for nodulation with *AeCRK*

A. evenia mutant	Transformation construct	Transformed plants	Nodulated plants	Nodules/ nodulated plant*
I10	pCambia1302	13	0	0
I10	pCambia1302-p*AeCRK-AeCRK-T35S*	18	16	10±6.75

* values correspond to mean nodule number per nodulated plant ± standard deviation



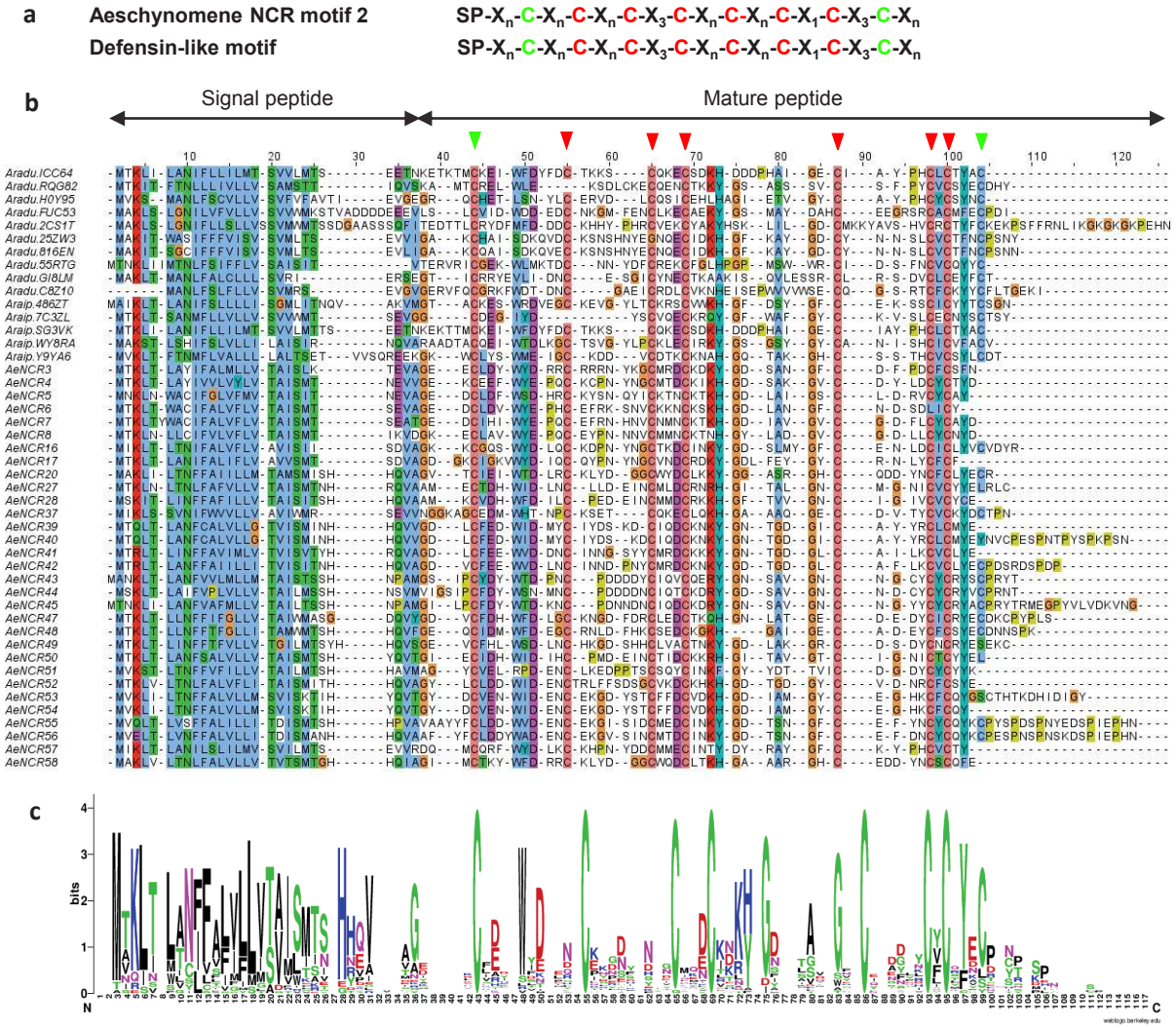
Supplementary Figure 15. Structure and alignment of the NCR peptides with the cysteine-rich motif 1. a, Cysteine-rich motif 1 and Medicgo NCR structure as previously presented[Czernic,2015]. SP, Signal Peptide, Xn, length of conserved spacings between cysteines. In red, conserved cysteines in the motif 1. **b**, Sequence alignment of NCR peptides with motif 1 found in *A. evenia*, with one Medicago NCR (Medtr6g445080) used for comparison. The putative signal peptide, the mature NCR peptide and cysteine positions are indicated above the sequence alignment. **c**, Sequence logos generated by processing the above alignment with the WebLogo software and showing conserved motifs in NCRs.

Supplementary Table 24. Results of the branch and branch site analysis for selection signature acting on CRK in *Aeschynomene* clade on full length protein and different domains.

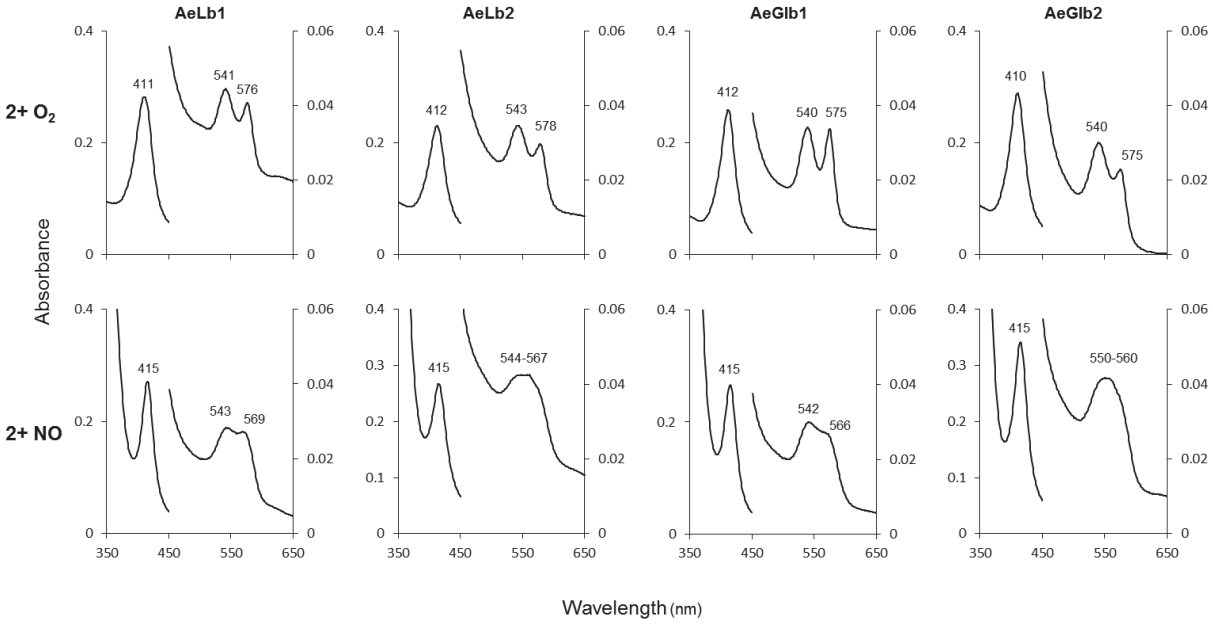
		Branch models						Branch site models	
Domain	Positions on AeCRK	Average ω for all tree (M0)	ω_{BG}	ω_{FG}	b_free *vs* M0	b_free *vs* b_neut	Comments	bsA *vs* M1	bsA *vs* bsA1
Full protein	1 - 668	**0,285**	0,294	0,202	ω_{FG} and ω_{BG} are not different p =0.147088	ω is not neutral p<1E-22***	General negative selection on CRK protein	No specific sites on FG are under relaxed selection p =1.0	No specific sites on FG are under positive selection p =1.0
Signal peptide	1 - 28	**1,021**	1,189	0,436	ω_{FG} and ω_{BG} are not different p =0.289591	ω is not different from neutral p =0.329225	Peptide signal is not different from neutral evolution	No specific sites on FG are under relaxed selection p =1.0	No specific sites on FG are under positive selection p =1.0
Extracellular domain	29 - 262	0,450	**0,480**	**0,187**	ω_{FG} and ω_{BG} are not different p =0.017214*	ω_{FG} is not neutral p =0.000034**	ECD of CRK is under stronger negative selection in *Aeschynomene* clade	No specific sites on FG are under relaxed selection p =1.0	No specific sites on FG are under positive selection p =1.0
Transmembrane domain	263 - 285	**0,319**	0,664	0,000	ω_{FG} and ω_{BG} are different p =0.169851	ω is not different from neutral p =0.169849	TM domain is not different from neutral evolution	No specific sites on FG are under relaxed selection p =1.0	No specific sites on FG are under positive selection p =1.0
Kinase domain	286 - 668	**0,164**	0,161	0,218	ω_{FG} and ω_{BG} are different p =0.432145	ω is not neutral p =0.001208**	General negative selection acting on kinase domain of CRK protein	No specific sites on FG are under relaxed selection p =1.0	No specific sites on FG are under positive selection p =1.0

FG: foreground branch (branch supporting *Aeschynomene* clade) BG: background branches (other Papilionaceae).

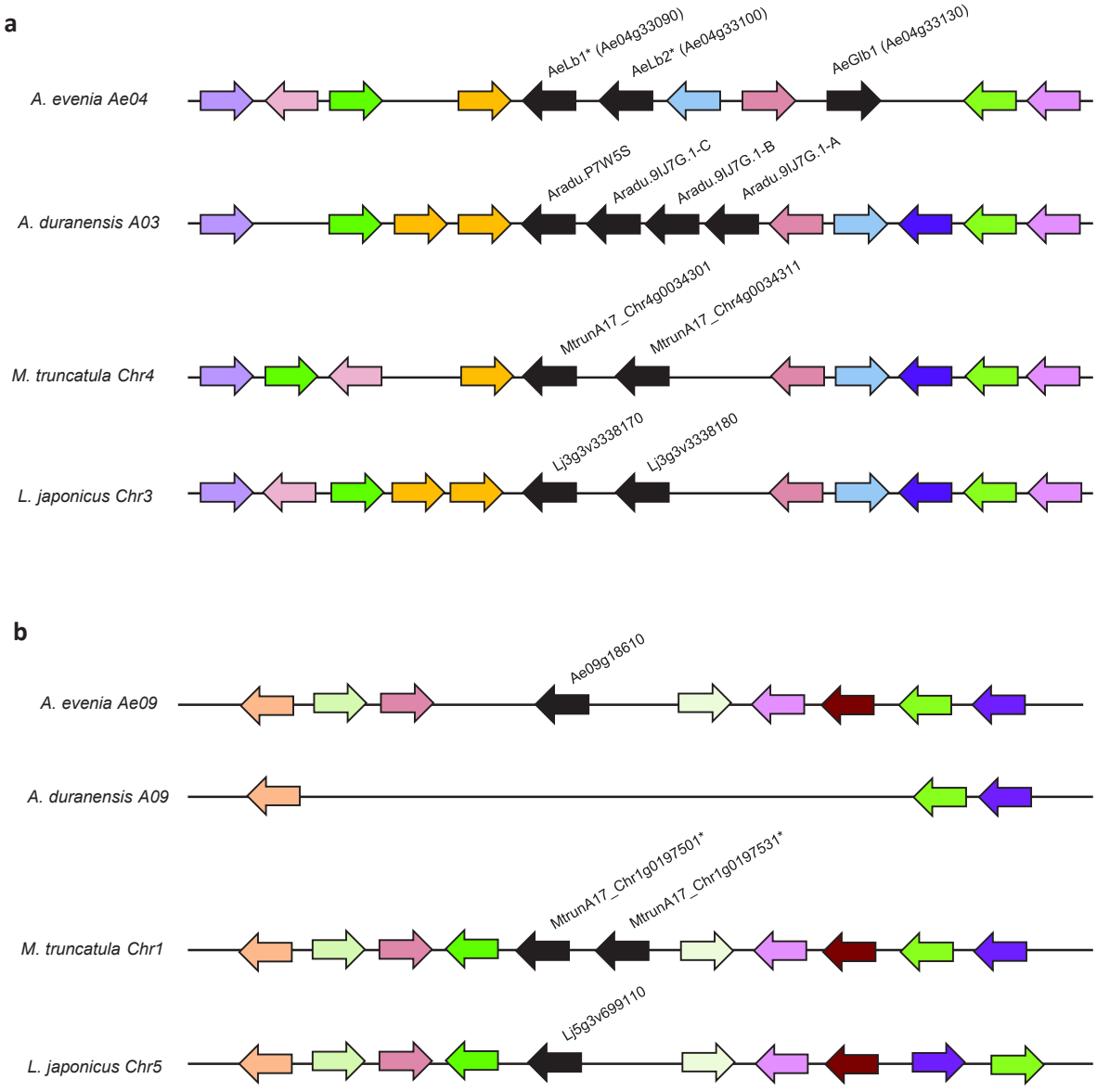
Significance of LRT: *0.05>*p-val*>0.01, **0.01 >*p-val*>0.001, ***p-*val*<0.001, absence of symbols: not significant.



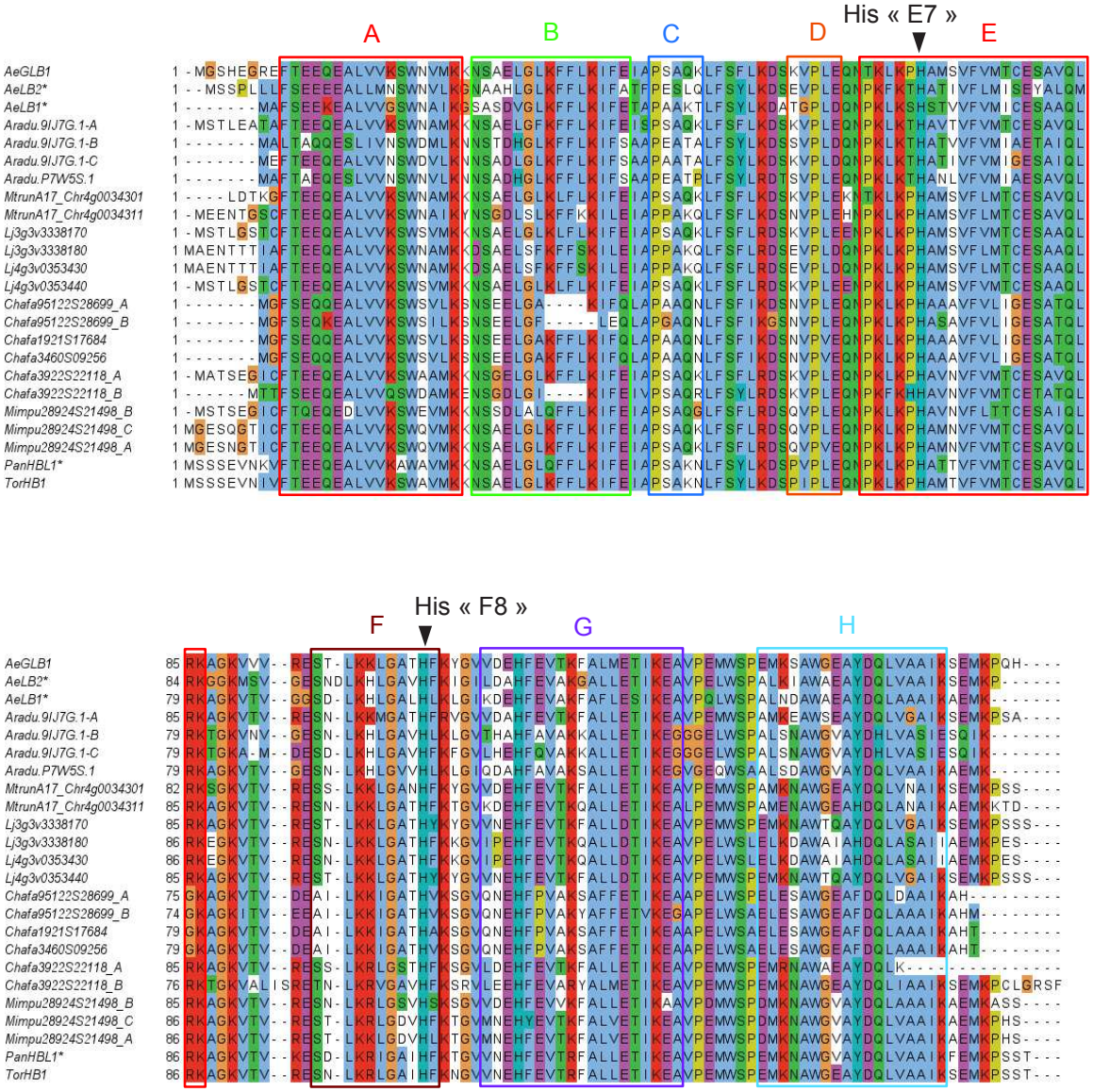
Supplementary Figure 16. Structure and alignment of the NCR peptides with the cysteine-rich motif 2. a, Cysteine-rich motif 2 and defensin structure as previously presented[Czernic,2015]. SP, Signal Peptide, Xn, length of conserved spacings between cysteine. In red, cysteines shared with the motif 1, in green, cysteine shared with the defensin signature. **b**, Sequence alignment of NCR peptides with motif 2 found in *A. evenia*, *A. duranensis* and *A. ipaiensis*. The putative signal peptide, the mature NCR peptide and cysteine positions are indicated above the sequence alignment. **c**, Sequence logos generated by processing the above alignment with the WebLogo software and showing conserved motifs in NCRs.



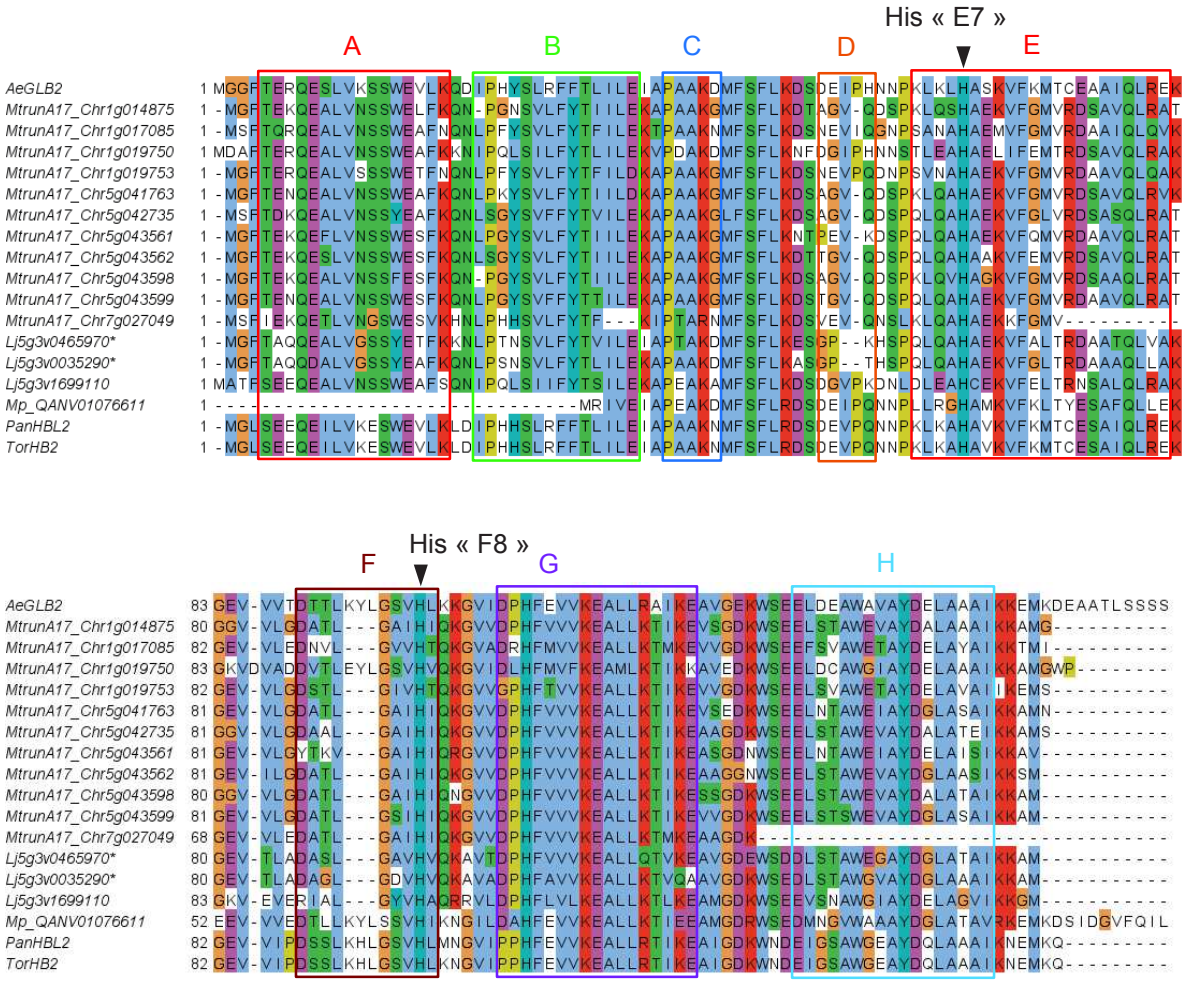
Supplementary Figure 17. UV-visible spectra of the oxyferrous (2+O_2) and nitrosyl (2+NO) complexes of *A. evenia* globins. The spectra indicate that all four globins are able to combine with O_2 and NO. The nitrosyl complexes show the same Soret peak at 415 nm but significant differences in the visible region. Thus, the nitrosyl complex of AeLb1 is typical, with α and β peaks at 569 and 543 nm, respectively, whereas AeLb2 and especially AeGlb2 display poorly defined bands. The physiological relevance of these variations is unknown.



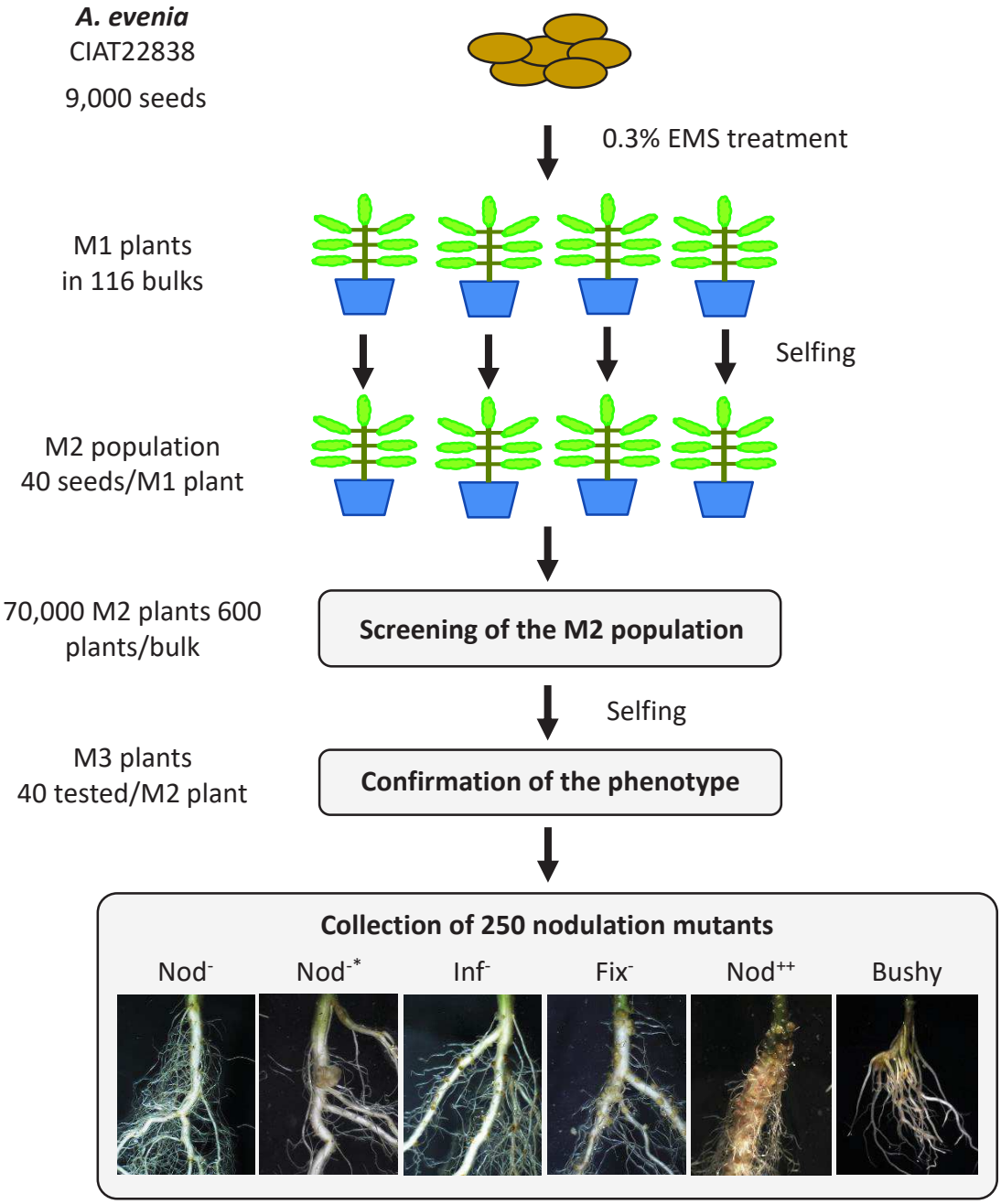
Supplementary Figure 18. Syntenic localization of Glb and Lb coding genes in *A. evenia*. Schematic representation of microsynteny analysis for the class 1 genes in **a** and for class 2 genes in **b** (black arrow) between *A. evenia* and *A. duranensis*. Orthologous genes are indicated through the use of a common colour. Orphan genes are not represented for clarity. * indicate Lbs. Note an inversion impacting gene arrangement on chromosome Ae04 in *A. evenia* and the absence of Glb gene in syntenic region on chromosome A09 in *A. duranensis*.



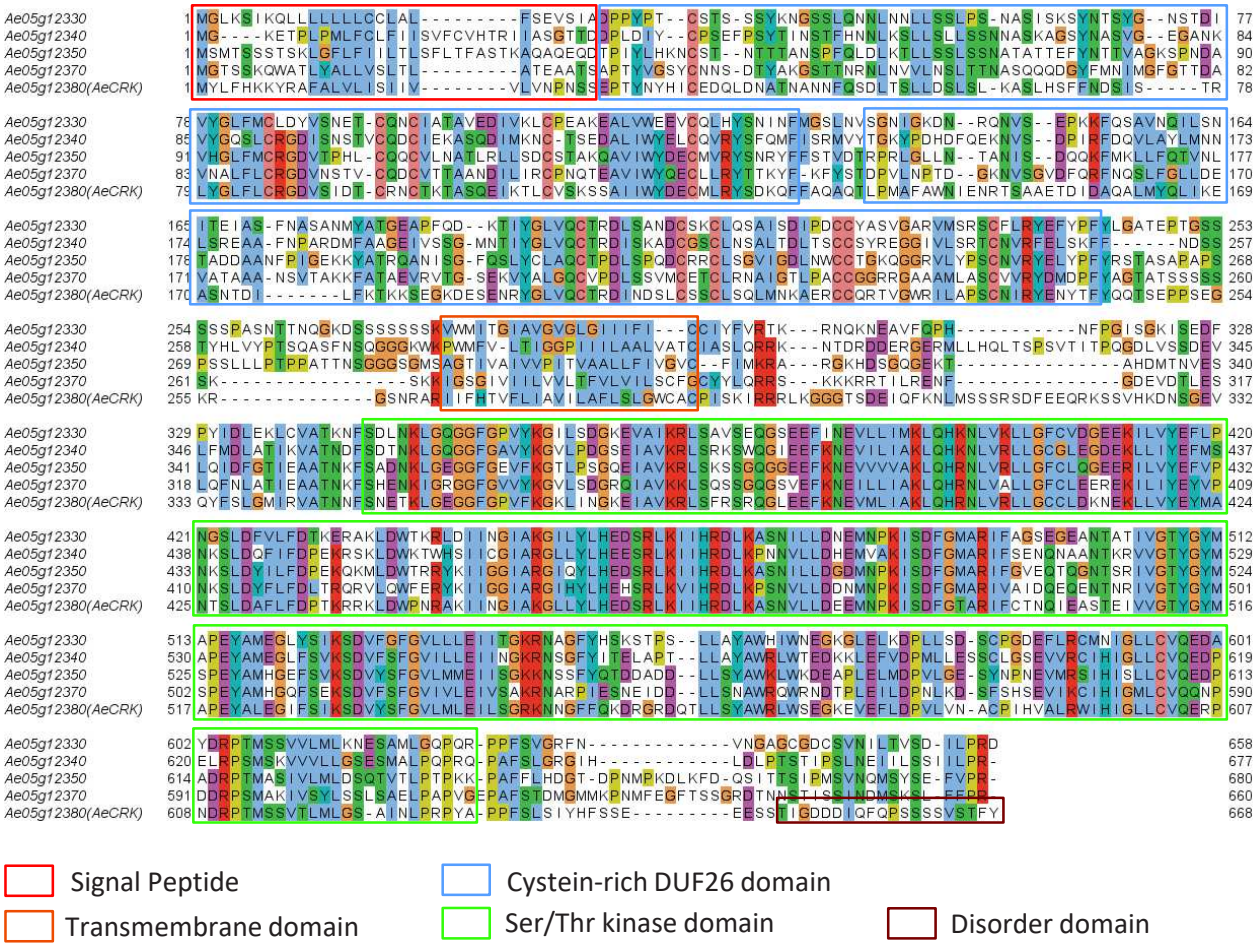
Supplementary Figure 19. Amino acid alignment of class 1 Glbs and Lbs from *A. evenia*, *A. duranensis*, *M. truncatula* and *L. japonicus*. Predicted proteins were aligned with Muscle and the conserved amino acids highlighted in colour in Jalview. The different putative helices A to H were delineated based on comparison with other Glbs and Lbs (37). Also indicated on the sequence alignment are the distal (E7) and proximal (F8) histidines which are involved in heme coordination or interaction with oxygen.



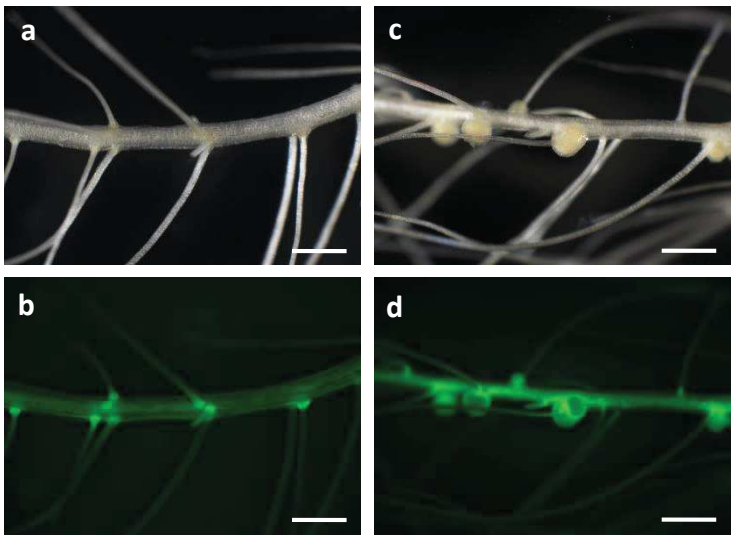
Supplementary Figure 20. Amino acid alignment of class 2 Glbs and Lbs. Predicted proteins were aligned with Muscle and the conserved amino acids highlighted in colour in Jalview. The different putative helices A to F were delineated based on comparison with other Glbs and Lbs (37). Also indicated on the sequence alignment are the distal (E7) and proximal (F8) histidines which are involved in heme coordination or interaction with oxygen.



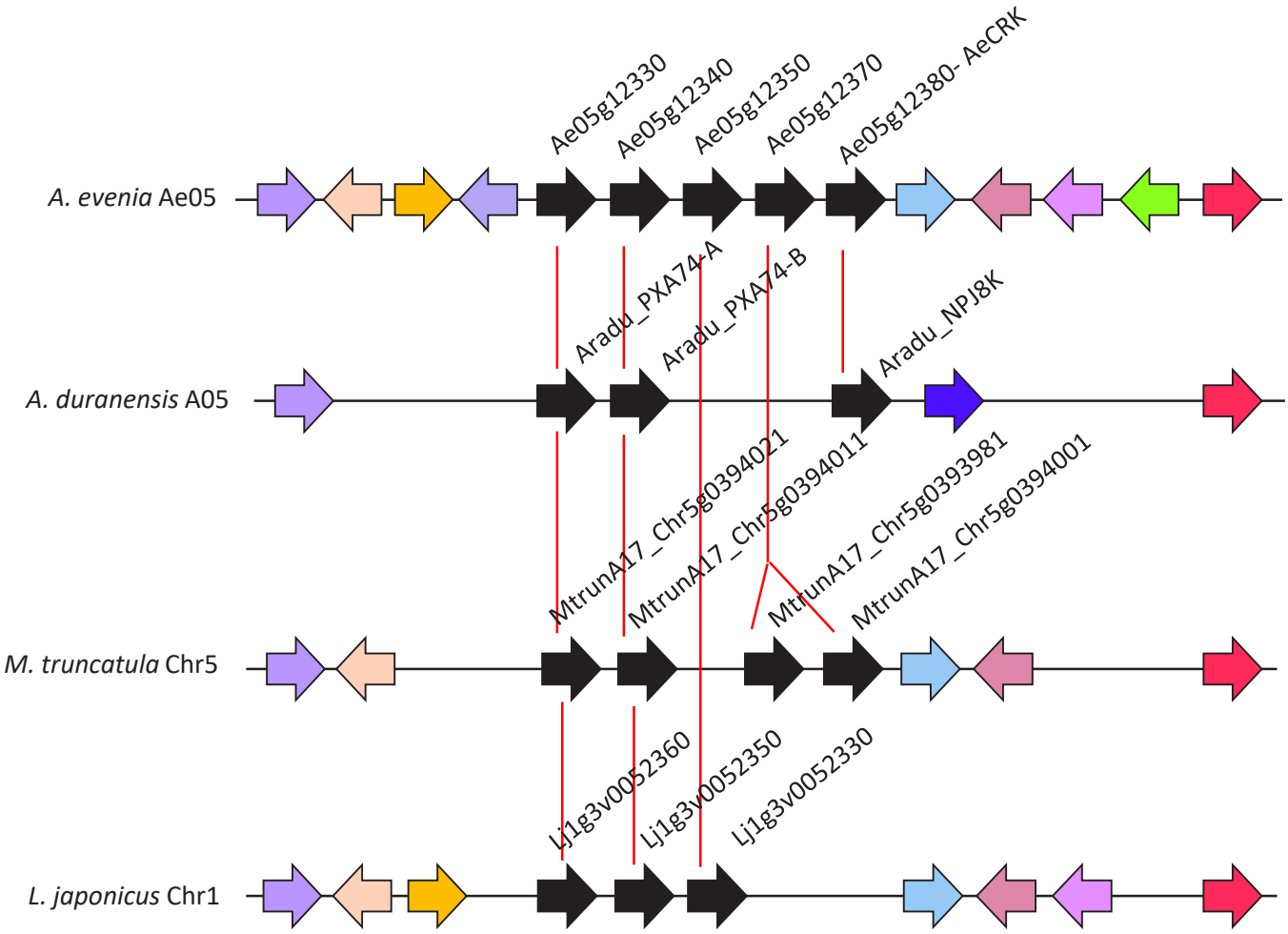
Supplementary Figure 21. Generation of a collection of nodulation mutants in *A. evenia*. A mutagenized population was produced by EMS treatment of seeds followed by selfing of the M1 individuals. Bulks of M2 plants were screened for mutants altered in nodulation. Mutants were confirmed for their nodulation defects in the M3 progeny and grouped in 6 main phenotypic categories. [Nod-], complete absence of nodulation; [Nod-*], occasional nodule formation; [Inf-], defects in infection; [Fix-], defect in fixation; [Nod++], hypernodulation; [Bushy]; altered nodulation and bushy root morphology.



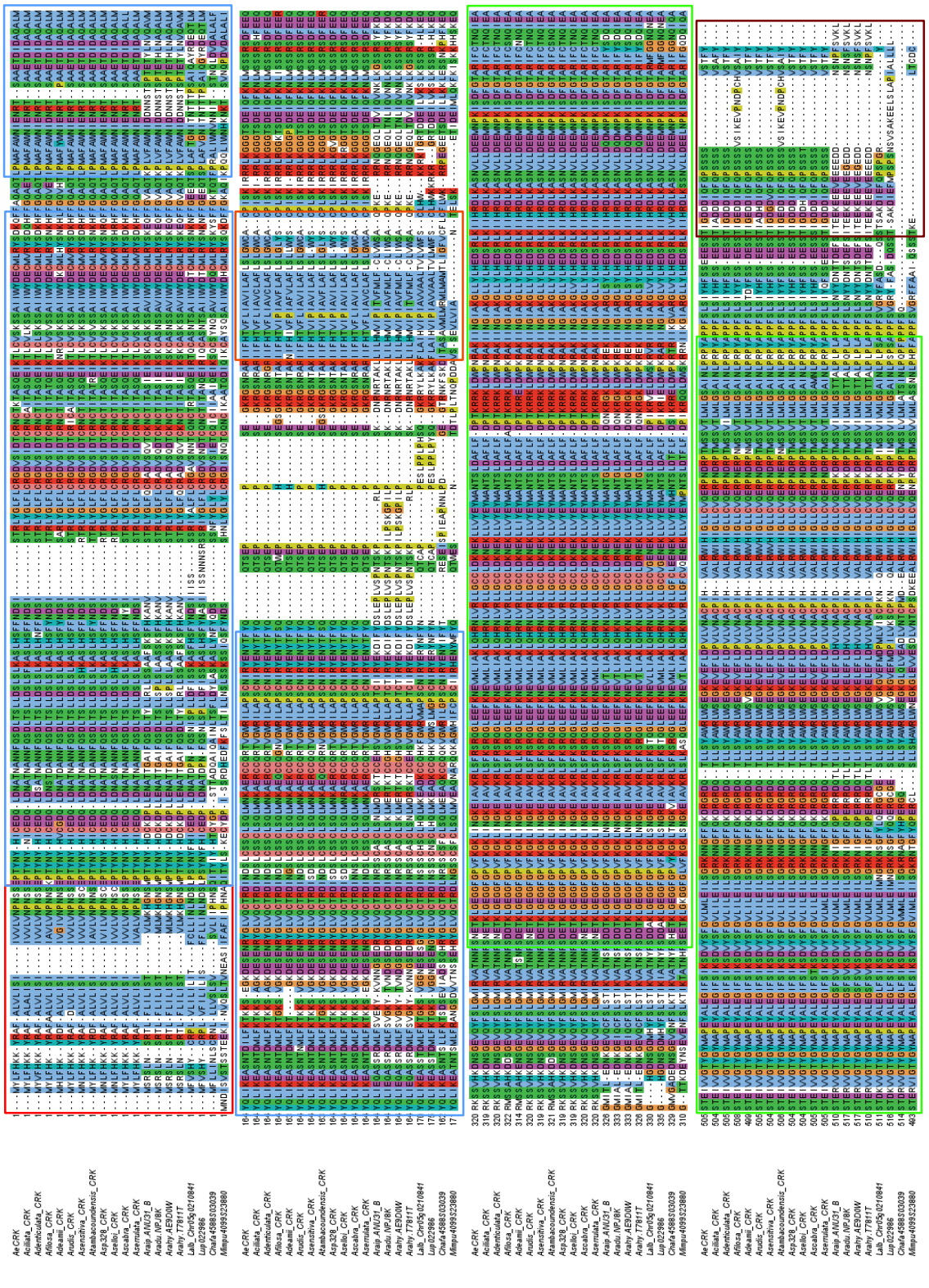
Supplementary Figure 22. Amino acid alignment of AeCRK and other CRKs encoded by the *CRK* cluster on the Ae05 chromosome in *A. evenia*. Predicted proteins were aligned with Muscle and the conserved amino acids highlighted in colour in Jalview. The different domains were identified with Interproscan. Note the conservation of Cysteines in all the DUF26 domain and the Disorder domain found only at the C-ter end of AeCRK.



Supplementary Figure 23. Functional complementation of the I10 mutant with *AeCRK.* Hairy roots were induced via *A. rhizogenes*-mediated transformation, inoculated with *Bradyrhizobium* strain ORS278 and nodulation phenotypes observed at 15 dpi. **a** and **b** The I10 mutant was transformed with an empty vector. None of the 13 transformed plant plants developed nodules. **c** and **d** A vector containing the *AeCRK* coding sequence fused to its native promoter was introduced into I10 mutant plants. Nodules were observed in 16 of 18 transformed plants. **b** and **d** GFP fluorescence indicating transformed hairy roots. Scale bar: 100 µm.



Supplementary Figure 24. Microsynteny analysis of the *CRK* gene cluster containing *AeCRK* in *A. evenia*, *A. duranensis*, *M. truncatula* and *L. japonicus*. Black arrows correspond to *CRK* genes, and red lines indicate orthologous relationships inferred from phylogenetic analyses. Genes with the same colour are homolog.



Supplementary Figure 25. Amino acid alignment of AeCRK ortholog proteins. Predicted proteins were aligned with Muscle and the conserved amino acids highlighted in colour in Jalview. The different domains were identified with Interproscan. Note the conservation of Cysteines in all the DUF26 domain and the Disorder domain found only at the C-ter end of AeCRK orthologs.

Supplementary References

1. Vurture, G.W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
2. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-70 (2011).
3. Oueslati, A. *et al.* Genotyping by sequencing reveals the interspecific *C. maxima / C. reticulata* admixture along the genomes of modern citrus varieties of mandarins, tangors, tangelos, orangelos and grapefruits. *PLoS One* **12**, e0185618 (2017).
4. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–5 (2007).
5. Martin, G. *et al.* Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics* **17**, 243 (2016).
6. Nawrocki, E.P. & Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933-5 (2013).
7. Lowe, T.M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. **25**, 955-64 (1997).
8. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573-580 (1999).
9. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res*. **19**, 1639-45 (2009).
10. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
11. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587 (2017).
12. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
13. Tang, H. *et al.* tanghaibao/jcvi: JCVI v0.7.5. doi:10.5281/ZENODO.846919 (2017).
14. Cabau, C. *et al.* Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. *PeerJ* **5**:e2988 (2017).
15. Crooks GE, *et al.* (2004) WebLogo: A sequence logo generator. *Genome Research*, 14:1188-1190.
16. Schneider TD, Stephens RM (1990) Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res*. 18:6097-6100
17. Sainz, M. et al (2013) Plant hemoglobins may be maintained in functional form by reduced flavins in the nuclei, and confer differential tolerance to nitro-oxidative stress. *Plant J*. 76, 875-87.

18. Bustos-Sanmamed, P. et al (2011) Regulation of non-symbiotic and truncated hemoglobin genes of *Lotus japonicus* in plant organs and in response to nitric oxide and hormones. *New Phytol*. 189, 765-76.
19. Bourge, M., Brown, S.C. and Siljak-Yakovlev S. Flow cytometry as tool in plant sciences, with emphasis on genome size and ploidy level assessment. *Genetics & Applications* **2.2**: 1-12 (2018).
20. Schmidt, T.G.M. & Skerra, A. The *Strep*-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols* **2,** 1528-1535 (2007).
21. Fabre, S. *et al.* Nod factor-independent nodulation in *Aeschynomene evenia* required the common plant-microbe symbiotic toolkit. *Plant Physiol*. **169**, 2654-64 (2015).

Chapitre 2

2.1 - Contexte de l'étude

Le crible génétique mené chez *Aeschynomene evenia* a montré que plusieurs gènes de la voie de signalisation Nod décrite chez les légumineuses modèles sont conservés, à savoir : *AePOLLUX*, *AeCCaMK*, *AeCYCLOPS*, *AeNSP2* et *AeNIN*. Cette approche mutant a également conduit à la découverte d'un gène nouveau, *AeCRK*, qui pourrait être impliqué dans la signalisation Nod-independante mais dont le pattern phylogénétique au sein des légumineuses suggère qu'il est ancestral même s'il est absent chez les modèles. Ces gènes de signalisation contrôlent à la fois la nodulation de racine et de tige, indiquant qu'ils participent au même programme d'activation sur les deux parties de la plante (Chapitre 1).

A. evenia se distingue de *M. truncatula* et *L. japonicus* par une activation de la nodulation qui ne passe pas par la reconnaissance des facteurs Nod ni par la formation de cordons d'infection via les poils absorbants. Aussi, une caractérisation plus approfondie des mutants altérés dans les gènes de la voie de signalisation Nod sera informatif de leur rôle symbiotique dans un contexte symbiotique sans facteurs Nod et sans cordon d'infection. De même, *AeCRK* ne possède pas d'orthologue évident chez les légumineuses modèles. Une caractérisation des mutants correspondants est requise pour tenter de positionner ce gène par rapport à ceux conservés de la voie de signalisation symbiotique.

Pour avancer dans notre compréhension des mécanismes symbiotiques chez *A. evenia*, j'ai déterminé le stade de blocage dans l'infection et l'organogenèse nodulaire de mutants de gènes de signalisation. Ces mutants ont également été analysés pour leur altération potentielle dans la symbiose endomycorhizienne par Rémi Guyonnet (LSTM). L'altération de réponses précoces chez ces mutants Nod^- a été analysée très récemment par Q-RT-PCR grâce au travail conjoint avec Marjorie Pervent et Djamel Gully (LSTM). Ces travaux ont précisé le rôle symbiotique de ces gènes de signalisation et révélé un rôle non symbiotique clé pour *NSP2* ; ils sont présentés dans ce Chapitre 2. Des tests de nodulation spontanée sur les mutants d'*AeCRK* et des analyses statistiques adéquates sur les données de Q-PCR sont prévues pour compléter ce travail qui fera l'objet à terme d'une publication.

Article 2 : Caractérisation de la voie de signalisation Nod-indépendante chez la légumineuse *Aeschynomene evenia*

2.2 - Article 2 :

Caractérisation de la voie de signalisation Nod-indépendante chez la légumineuse *Aeschynomene evenia*

Johan Quilbé[1], Marjorie Pervent[1], Rémi Guyonnet[1], Djamel Gully[1], Laurent Brottier[1], Thomas Benichou[1], Christophe Klopp[2], Fabienne Cartieaux[1], Eric Giraud[1] et Jean-François Arrighi[1]

1-IRD, Laboratoire des Symbioses Tropicales et Méditerranéennes (LSTM), UMR IRD/ SupAgro/INRAE/ UM2 /CIRAD, TA-A82/J – Campus de Baillarguet 34398 Montpellier cedex 5, France

2-Plateforme bioinformatique Genotoul, BioinfoMics, UR875 Biométrie et Intelligence Artificielle, INRAE, Castanet-Tolosan, France.

Titre court : Signalisation symbiotique Nod-indépendante chez *Aeschynomene evenia*

Mots clé : Symbiose, signalisation, activation facteurs Nod-indépendante, infection, *Aeschynomene*

Introduction

La symbiose rhizobium-légumineuse est initiée par la reconnaissance par la plante-hôte de lipochitooligosaccharides rhizobiens spécifiques appelés facteurs Nod. En réponse à ces molécules signal, une myriade de changements structurels et fonctionnels se produit dans la racine de la plante, avec le développement des cordons d'infection au niveau des poils absorbants et la formation à distance de divisions cellulaires dans le cortex de la racine à l'origine du primordium nodulaire (pour revue : Roy *et al.*, 2020). La dissection génétique de la voie de signalisation conduisant aux différentes réponses symbiotiques a conduit à l'identification de récepteurs des facteurs Nod, MtNFP/LjNFR5 et MtLYK3/LjNFR1. La cascade de signalisation en aval de ces récepteurs inclut un autre récepteur symbiotique MtDMI2/LjSYMRK, plusieurs composants de l'enveloppe nucléaire, parmi lesquels se trouvent MtDMI1/LjPOLLUX et LjCASTOR, une kinase décodant des oscillations calciques MtDMI3/LjCCaMK, qui interagit avec MtIPD3/LjCYCLOPS. Ces deux régulateurs nucléaires contrôlent l'expression de plusieurs facteurs de transcription, NSP1, NSP2, ERN1, ERN2, ERN3 et NIN, qui orchestrent les réponses symbiotiques au niveau transcriptionnel (pour revue : Roy *et al.*, 2020). Plusieurs acteurs qui la composent cette voie de signalisation Nod s'avèrent être aussi impliqués, de façon plus ou moins importante, dans la symbiose mycorhizienne (pour revue : Gobbato, 2015). Une fonction non symbiotique a également été mise en évidence pour le couple de facteurs de transcription NSP1/NSP2 dans la voie de synthèse des caroténoïdes et des strigolactones qui est conservée avec les non-légumineuses (Liu *et al.*, 2011).

Alors que l'étude des mécanismes moléculaires de la nodulation est bien avancée chez les deux légumineuses modèles, *M. truncatula* et *L. japonicus*, celle de mécanismes alternatifs présents dans d'autres systèmes symbiotiques commence tout juste à se développer. Parmi eux, certaines espèces du genre *Aeschynomene* sont capables d'établir une symbiose avec des *Bradyrhizobium* photosynthétiques qui ne produisent pas de facteurs Nod pourtant considérés comme les molécules signal clé de la symbiose rhizobienne (Giraud *et al.*, 2007; Giraud and Fleischman, 2004). Si la nature des signaux potentiels produits par ces souches de *Bradyrhizobium* reste énigmatique, il a néanmoins été établi que dans ce système symbiotique, l'infection commence par une colonisation par les bactéries de couronnes de

poils axillaires situés à la base des racines latérales, qui représentent les sites d'infection (Bonaldi *et al.*, 2011). La pénétration intercellulaire des bactéries conduit dans le cortex à un « cell collapse » puis à leur endocytose dans une ou quelques cellules. Ces cellules infectées se divisent alors abondamment pour donner un nodule (Bonaldi *et al.*, 2011). Des analyses transcriptomiques et une approche de génétique inverse menées sur le modèle d'étude *Aeschynomene evenia* ont suggéré que la voie de signalisation symbiotique est, au moins en partie, conservée tandis que les mécanismes moléculaires de perception des rhizobiums et l'infection présentent des divergences notables avec les légumineuses modèles (Chaintreuil *et al.*, 2016; Fabre *et al.*, 2015; Gully *et al.*, 2018).

Ces observations ont été confortées par une approche de génétique directe menée chez *A. evenia*. Le crible d'une population de mutagenèse EMS a en effet permis de sortir des mutants altérés dans les toutes premières étapes de la nodulation. La caractérisation génétique de ces mutants et la recherche des mutations causales a conduit à l'identification de plusieurs gènes de la voie de signalisation symbiotique, *AePOLLUX*, *AeCCaMK*, *AeCYCLOPS*, *AeNSP2* et *AeNIN* et à la découverte d'un gène nouveau, *AeCRK*, appartenant à la famille des Cysteine-rich Receptor-like Kinase dont plusieurs membres sont des acteurs de la signalisation des ROS (Reactive Oxygen Species) (Quilbé *et al.*, accepté ; Bourdais *et al.*, 2015). Ce gène ne possède pas d'orthologue évident chez les légumineuses modèles et sa fonction symbiotique exacte n'est pas encore déterminée. L'existence de mutants pour ces gènes de signalisation chez *A. evenia* offre la possibilité de mieux comprendre le rôle de ces gènes dans un contexte symbiotique indépendant des facteurs Nod et de la formation de cordon d'infection. L'implication des gènes de la voie de signalisation symbiotique est-elle le même chez *A. evenia* et les légumineuses modèles ? Quelle est la relation hiérarchique entre *AeCRK* et les autres gènes de signalisation ? Pour répondre à ces questions, nous avons combiné des analyses phénotypiques de mutants altérés dans ces gènes de signalisation à des analyses transcriptomiques chez *A. evenia*.

Résultats

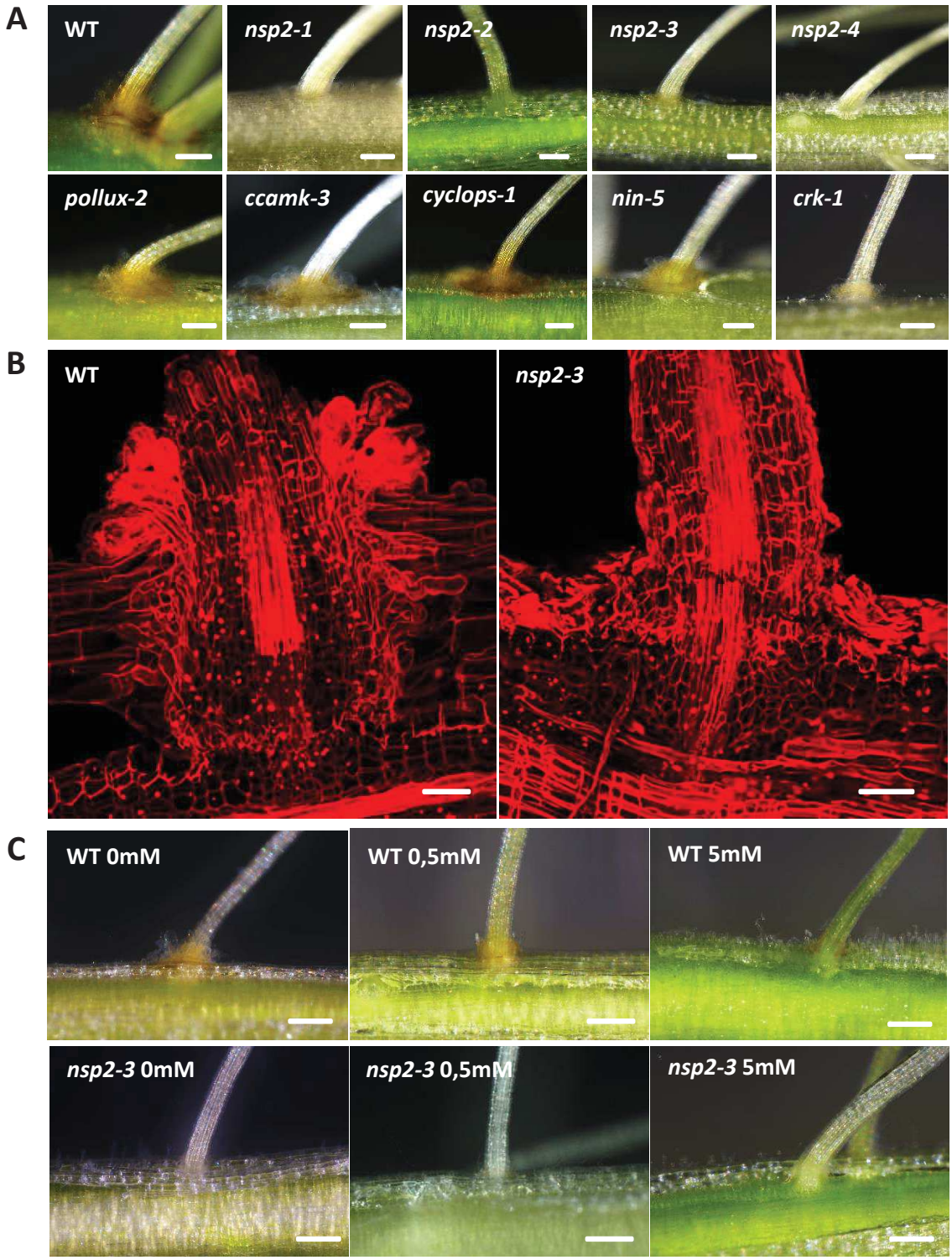


Figure 1 : Absence de couronnes de poils axillaires chez les mutants *nsp2* d'*A. evenia*. A. Phénotype de la racine non-inoculée et cultivée en tube chez la lignée sauvage, les mutants alléliques *nsp2* et un mutant pour chaque gène de signalisation. Les barres d'échelles représentent 250 µm. **B.** Coupes d'émergences de racines latérales en microscopie confocale (paroi marquée au calcofluor blanc) chez la lignée sauvage et du mutant *nsp2-3*. Les barres d'échelles représentent 50 µm. **C.** Phénotypes de la lignée WT et du mutant *nsp2-3* cultivés sur une gamme croissante d'azote. Les barres d'échelles représentent 250 µm.

Les mutants *nsp2* sont déficients pour la formation de couronnes de poils axillaires

Des mutants déficients pour la nodulation ont été isolés du crible d'une population de mutagenèse EMS mené chez *A. evenia*. Les analyses génétiques couplées à des approches de re-séquençage ont permis de constituer des séries de mutants alléliques pour six gènes de signalisation symbiotique (Tableau S1) (Quilbé *et al.*, accepté). Lors de ce crible réalisé sur des plantes cultivées dans de l'attapulgite et inoculées avec la souche de *Bradyrhizobium* ORS278, la lignée sauvage a présenté des nodules roses en haut de la racine principale tandis que les mutants des gènes de signalisation en sont dépourvus. Néanmoins, les mutants *crk*, *pollux*, *ccamk*, *cyclops* et *nin* présentent normalement le long de leur système racinaire des couronnes de poils axillaires à la base des racines latérales, qui constituent le premier site de colonisation par les *Bradyrhizobium*. De façon frappante, les quatre mutants alléliques *nsp2* diffèrent des autres mutants par l'absence notable de ces rosettes (données non montrées). Pour s'assurer que ce phénotype n'est pas lié aux conditions de culture, ces mêmes mutants ont été cultivés en milieu BNM liquide et sans inoculation. Les mutants *nsp2* sont les seuls à ne pas développer de couronnes de poils axillaires (Fig. 1A). Le fait que les quatre mutants alléliques *nsp2* présentent le même phénotype indique qu'*NSP2* est requis pour le développement des poils axillaires.

Pour préciser les altérations du développement de poils axillaires liées à *AeNSP2* et l'impact sur la colonisation de la racine par les bactéries, le mutant *nsp2-3* non inoculé a été comparé à la lignée sauvage. Les observations de microscopie confocale montrent pour la plante sauvage une couronne de poils axillaires de forme bulbeuse à la base de la racine latérale. En revanche, pour le *nsp2-3*, aucune émergence de poils axillaires n'est visible à la base des racines latérales (Fig. 1B). Nous avons ensuite regardé si les mutants *Aensp2* sont altérés dans d'autres réponses développementales de la racine. Des observations encore préliminaires suggèrent que les mutants *Aensp2* développent normalement des poils absorbants mais que la structure des coiffes de racine est altérée par rapport à la lignée sauvage (données non montrées).

La disponibilité en azote de la plante étant connue pour réguler négativement la nodulation chez les légumineuses, nous avons testé la possibilité que l'induction de la formation des couronnes de poils axillaires soit dépendante de l'azote. Pour cela, la lignée sauvage a été cultivée dans du milieu BNM liquide complémenté avec une gamme de concentration d'azote (0, 0,5 et 5 mM de KNO_3). L'observation des plantes âgées de 10 jours a révélé que

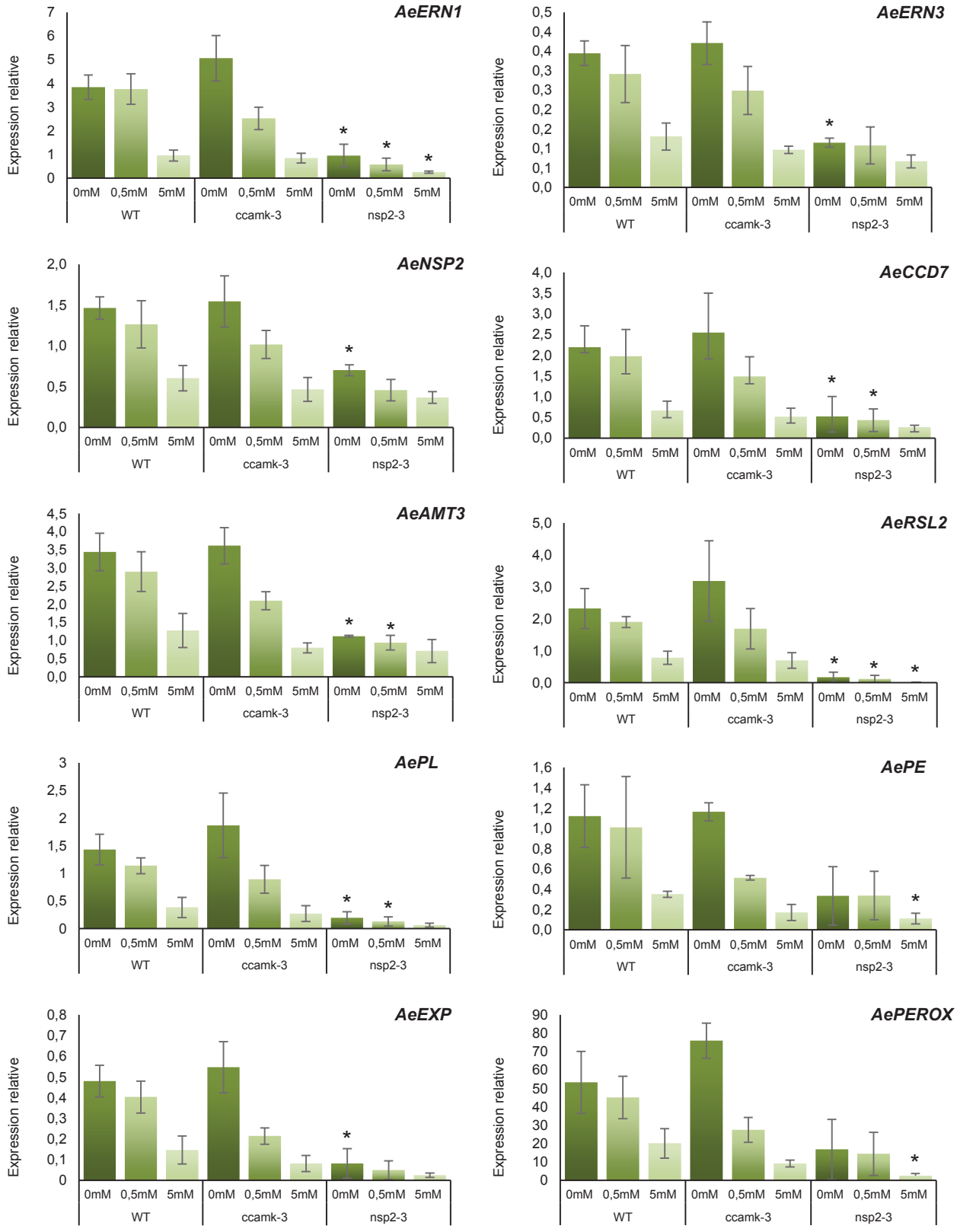


Figure 2 : Rôle de l'azote et d'*AeNSP2* dans le contrôle de l'expression de gènes réprimés au cours de la nodulation. Expression de gènes impliqués dans l'infection (*ERN1* et *ERN3*), la biosynthèse de caroténoïdes et de strigolactones (*CCD7*), la formation des poils absorbants (*RSL2*), la modification de la paroi végétale (*EXP, PEROX, PE et PL*) et le transport d'azote (*AMT3*) chez la lignée sauvage, les mutants *ccamk-3* et *nsp2-3*. cultivés pendant 10 jours avec 0mM, 0,5mM et 5mM de KNO_3^-. L'expression est normalisée avec celles d'*AeEF1α* et l'ubiquitine (trois réplicats biologiques). Les étoiles correspondent à une valeur significativement différente du WT avec un test Kruskal-Wallis ($p<0,05$).

les poils axillaires se développent avec 0 et 0,5 mM KNO_3 mais que leur formation est fortement réprimée avec 5 mM de KNO_3 (Fig. 1C). L'ensemble de ces observations suggère un lien entre *AeNSP2*, la formation des poils axillaires et le statut azoté de la plante.

***NSP2* contrôle l'expression d'un set de gènes en condition non-symbiotique**

Pour mieux comprendre le ou les rôles remplis par *AeNSP2*, nous avons d'abord analysé son profil d'expression qui est disponible dans le Gene Atlas d'*A. evenia* (Tableau S2). La comparaison des conditions racines avec et sans azote indique qu'*AeNSP2* est exprimé de façon constitutive dans la racine mais que son expression est réprimée par l'azote. De même, son niveau d'expression diminue très rapidement au cours de la nodulation. *AeNSP2* présente un profil d'expression très différent de celui des autres gènes de signalisation symbiotique qui sont soit exprimés de façon similaire dans la racine et le nodule, soit induits au cours de la nodulation (Quilbé *et al.*, accepté). Nous avons ensuite recherché les gènes co-exprimés avec *AeNSP2* et qui sont donc potentiellement impliqués dans les mêmes fonctions biologiques. Pour cela, une analyse de Clustering Hiérarchique a été conduite pour extraire les principaux patterns d'expression de gènes chez *A. evenia*. Un groupe contenant *AeNSP2* et des gènes avec un profil d'expression très similaire a été identifié. En plus d'*AeNSP2*, ce groupe contient des gènes symbiotiques connus comme *AeERN1* et *AeERN3*, les autres gènes sont impliqués potentiellement dans d'autres processus biologiques (Tableau S2).

Pour confirmer que l'expression des gènes de ce groupe est bien régulée par l'azote et pour tester si une mutation dans *AeNSP2* a un effet dessus, un set de dix gènes de ce groupe a été analysé par Q-RT-PCR chez la lignée sauvage, les mutants *nsp2-3 et ccamk-3* cultivés sur une gamme de concentration d'azote croissante (0, 0,5 et 5mM de KNO_3) pendant 10 jours. Pour ces 10 gènes testés, nous avons obtenu des résultats similaires (Fig. 2). Chez la plante sauvage, le niveau d'expression est comparable entre les conditions 0 et 0,5 mM KNO_3 mais il est réduit dans la condition 5 mM de KNO_3, confirmant l'effet régulateur de l'azote sur l'expression de ces gènes. Chez le mutant *ccamk-3*, le profil d'expression est similaire à celui obtenu avec la lignée sauvage tandis que des niveaux d'expression réduits sont retrouvés chez le mutant *nsp2-3* dans les trois conditions azote (Fig. 2). Ce niveau d'expression est comparable à la condition 5 mM de KNO_3 chez la lignée sauvage. Ces données indiquent qu'*AeNSP2* contrôle de façon directe ou indirecte l'expression de ces gènes en condition non

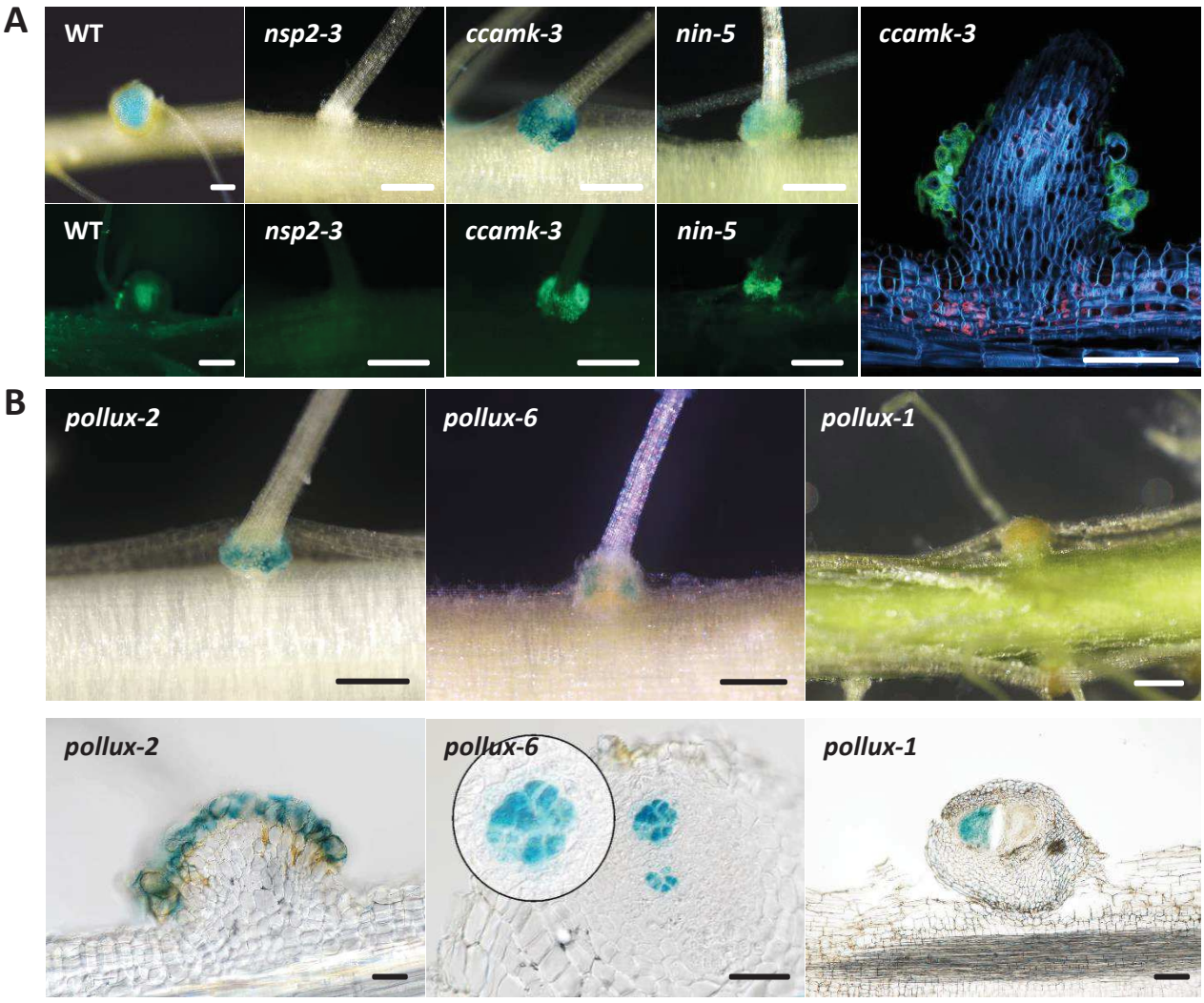


Figure 3 : Phénotypes de nodulation des mutants *pollux*, *ccamk*, *nsp2* et *nin* d'*A. evenia*. **A.** Phénotypes racinaires des lignées WT, *nsp2-3*, *caamk-3* et *nin-5* 14 jours après inoculation de la souche de *Bradyrhizobium* ORS278-GUS ou ORS285-GFP. La coupe de racine du mutant ccamk-3 a été observée en microscopie confocale. Les barres d'échelles représentent 250 µm pour les racines entières et 50 µm pour les coupes **B.** Phénotypes racinaires des lignées *pollux-2*, *pollux-6* et *pollux*-1 14 jours après inoculation de la souche de *Bradyrhizobium* ORS278-GUS et coupes de Bumps observés en microscopie de phase. Les barres d'échelles représentent 500 µm pour lers racines entières et 50 µm pour les coupes.

symbiotique. En revanche, l'absence d'effet d'une mutation dans le gène *AeCCaMK* sur l'expression de ces gènes suggère que cette régulation ne dépend pas de la voie de signalisation symbiotique qui est essentielle pour l'activation de la nodulation.

Nous nous sommes ensuite intéressés à la fonction biologique des gènes dont l'expression est régulée par *AeNSP2*. Chez *M. truncatula*, *ERN1* code un facteur de transcription symbiotique qui se fixe sur des boites « NF » présent sur les promoteurs de gènes induits par les facteurs Nod (Cerri *et al.*, 2012). L'expression d'*ERN1* est régulée positivement par *NSP1* et *NSP2*, et négativement par *ERN3* (Andriankaja *et al.*, 2007; Cerri *et al.*, 2012). Chez *A. evenia* qui utilise un processus symbiotique indépendant des facteurs Nod, l'expression d'*AeERN1* et *AeERN3* est régulée par *AeNSP2*, et ces trois gènes sont réprimés au cours de la nodulation. Un gène, *AeCCD7*, code une protéine qui catalyse une conversion dans la voie de synthèse des carotéoïdes et des strigolactones (van Zeijl *et al.*, 2015). Chez *M. truncatula*, plusieurs gènes de cette voie de biosynthèse se sont montrés être régulés par *NSP2* (Liu *et al.*, 2011). Un autre gène, *AeMT3*, code un transporteur d'ammonium et souligne le lien avec la régulation par le statut azoté de la plante. Plusieurs gènes codent des enzymes intervenant dans la modification de la paroi végétale (pectate lyase, pectate esterase, peroxydase et expansine) et un gène code un facteur de transcription requis pour le développement des poils absorbants chez *Arabidopsis thaliana*, *AtRLS2* (Mangano *et al.*, 2018).

Les mutants de gènes de signalisation présentent différents stades de blocage de l'infection rhizobienne et de l'organogenèse nodulaire

Pour préciser les stades de blocage de la nodulation chez les mutants des différents gènes de signalisation et pour analyser l'homogénéité des phénotypes symbiotiques entre mutants alléliques, les différentes lignées mutantes ont été inoculées avec la souche de *Bradyrhizobium* ORS278-GUS et analysées en microscopie de lumière à 14 dpi. Pour les quatre mutants *nsp2*, aucun changement au niveau de la racine n'est visible (Fig. 3A). Pour les quatre lignées mutantes d'*AeCcamK*, et les six mutants d'*AeNIN*, une colonisation des couronnes de poils axillaires par la bactérie est révélée par une coloration GUS (Fig. 3A). L'inoculation de ces lignées avec la souche de *Bradyrhizobium* ORS285-*GFP* conduit au même phénotype (Fig. 3A). L'analyse de coupes de racines de la lignée *ccamk-3* en microscopie confocale montre que les bactéries restent agglutinées au niveau des poils axillaires sans

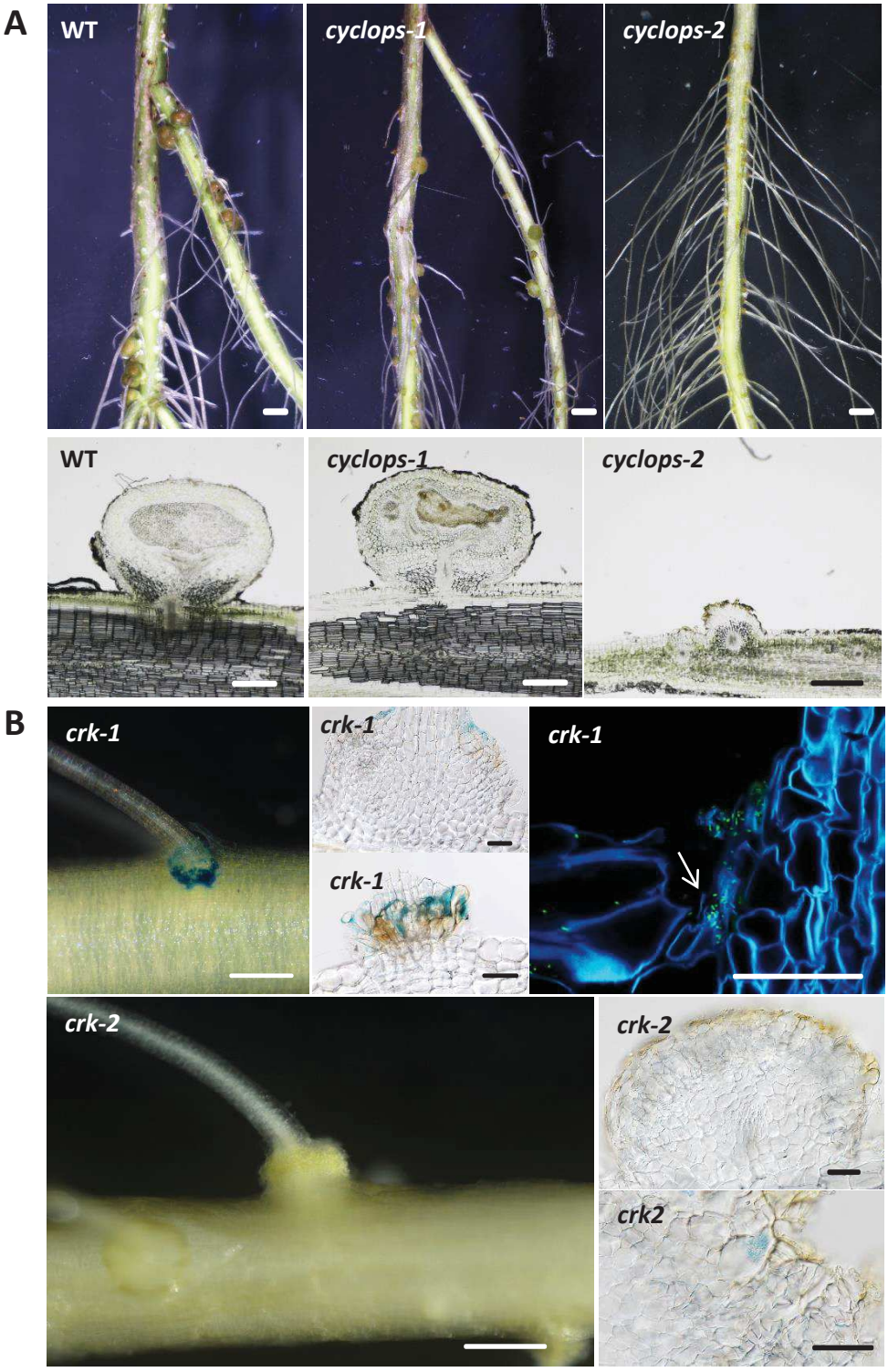


Figure 4 : Phénotypes de nodulation des mutants *cyclops* et *crk* d'*A. evenia*. **A.** Phénotypes racinaires des lignées WT, *cyclops-1* et *cyclops-2* 14 jours après inoculation de la souche de *Bradyrhizobium* ORS278. Les barres d'échelles représentent 500 µm pour les racines entières et 100 µm pour les coupes. **B.** Phénotypes racinaires des lignées *crk-2* et *crk-1* 14 jours après inoculation de la souche de *Bradyrhizobium* ORS278-GUS ou ORS278-GFP. Les coupes de racines ont été observées en microscopie de phase ou confocale. Les barres d'échelles représentent 500 µm pour les racines entières et 50 µm pour les coupes.

pénétrer dans l'espace intercellulaire de la racine. Cette observation indique que le processus d'infection intercellulaire est bloqué aux toutes premières étapes chez les mutants *ccamk*, *nsp2* et *nin*.

L'inoculation des six mutants *pollux* avec la souche de *Bradyrhizobium* ORS278-*GUS* a révélé trois phénotypes (Fig. 3B). Chez le mutant *pollux-2*, seule la colonisation des poils axillaires par la bactérie est visible. En revanche, chez le mutant *pollux-6*, des renflements discrets et surmontés de la couronne de poils axillaires sont présents à la base des racines latérales. La coloration GUS révèlent que ces poils axillaires sont aussi colonisés par les bactéries. Des sections de ces renflements (ou « bumps ») montrent la présence d'un spot d'infection coloré en bleu tandis que le reste du tissu n'est pas infecté (Fig. 3B). Ce phénotype suggère que l'infection intercellulaire est avortée. Enfin, le mutant *pollux-1* est capable de développer de petits nodules mais qui ne sont pas roses, indiquant qu'ils ne sont pas fonctionnels (Fig. 3B). Ces différences de sévérité du phénotype symbiotique sont certainement dues à des effets plus ou moins importants des mutations affectant *AePOLLUX* (Tableau S1). Des premières analyses réalisées sur les deux mutants *cyclops* indiquent que le mutant *cyclops-2* ne présente aucune réponse visible à l'inoculation avec la souche de *Bradyrhizobium* ORS278 tandis que le mutant *cyclops-1* est capable de développer de petits nodules non fixateurs dans lequel le tissu central est nécrosé (Fig. 4A). Ces phénotypes de nodulation restent à être analysés plus en détail mais il est possible que les différences observées soient aussi dues à un effet allélique comme pour *AePOLLUX* (Tableau S1).

Les deux mutants *Aecrk-1* et *-2* présentent un phénotype de nodulation similaire en montrant la capacité de développer des « bumps » discrets après inoculation avec la souche de *Bradyrhizobium* ORS278-GUS (Fig. 4B). Néanmoins, le mutant *crk-2* est capable de produire des « bumps » à la base de la majorité des racines latérales tandis que chez le mutant *crk-1*, l'apparition de petits « bumps » est plus rare. La différence de fréquence de formation de ces « bumps » entre les deux mutants alléliques est possiblement due à la nature des mutations qui impactent le gène *AeCRK* (Tableau S1). Mais pour les deux mutants, les coupes de « bumps » montrent que seulement quelques cellules à l'intérieur sont infectées (Fig. 4B). L'analyse de coupes de racines en microscopie confocale de la lignée *crk-1* inoculée avec la souche de *Bradyrhizobium* ORS278-*GFP* a révélé de façon récurrente la présence des bactéries au niveau du cortex racinaire (Fig. 4B). Ces données montrent que les mutants *crk* sont capables d'aller un peu plus loin dans le processus d'infection.

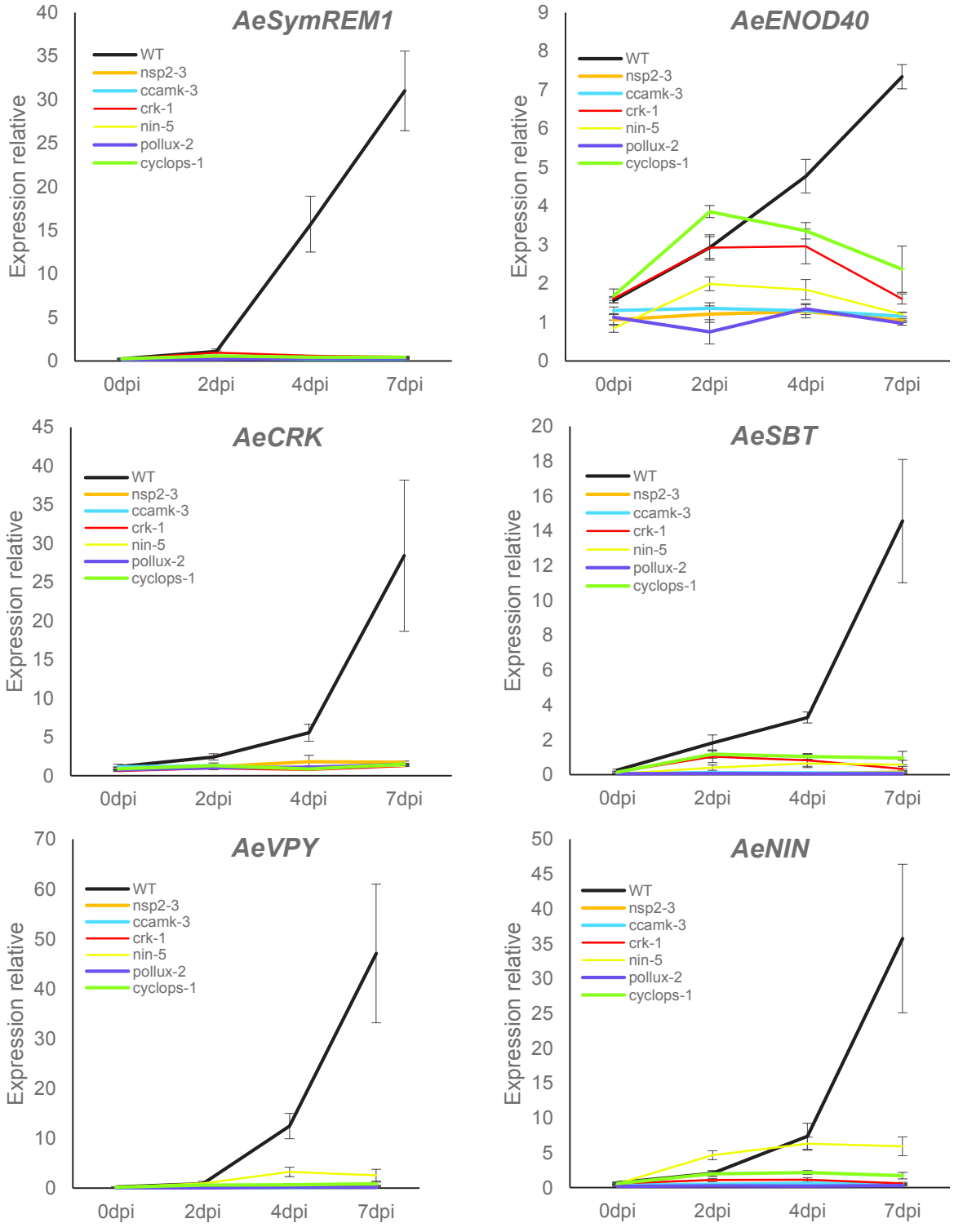


Figure 5 : Rôle des gènes de signalisation symbiotique dans l'induction de gènes de nodulines précoces chez *A. evenia*. L'expression des gènes *AesymREM1, AeENOD40*, *AeCRK, AeSBT*, *AeVPY* et *AeNIN* a été analysé par q-RT-PCR à 0, 2, 4 et jours après inoculation avec la souche de *Bradyrhizobium* ORS278 chez la lignée WT et différents mutants de gènes de signalisation. L'expression est normalisée avec celles d'*AeEF1a* et de l'ubiquitine. Les moyennes et les erreurs standards proviennent de trois réplicats biologiques.

Les gènes de signalisation symbiotique sont requis pour l'expression de gènes de nodulines

La nodulation implique la mise en place de deux programmes symbiotiques, celui d'infection et celui d'organogenèse nodulaire qui sont accompagnés par des changements transcriptionnels importants. Les gènes de nodulines précoces peuvent être utilisés pour disséquer les étapes précoces de l'interaction symbiotique chez *A. evenia*. Sur la base de données transcriptomiques obtenues précédemment pour *A. evenia* (Quilbé *et al.*, accepté), nous avons sélectionné quatre gènes induits au cours de la nodulation et orthologues à des gènes symbiotiques bien connus chez les légumineuses modèles : *NIN* code un régulateur clé de la voie de signalisation Nod qui contrôle l'expression d'un grand nombre de gènes, *MtsymREM* et *MtVPY* codent des acteurs de l'infection, *MtENOD40* intervient dans les divisions cellulaires lors de l'organogenèse nodulaire. Nous avons également choisi deux autres gènes d'*A. evenia* : *AeSBT* est induit très précocement après inoculation avec bradyrhizobium (Gully *et al.*, 2018) et *AeCRK* est requis pour la nodulation chez *A. evenia* (Quilbé *et al.*, accepté). Les profils d'expression, présentés dans la Tableau S2, montrent les cinétiques d'induction de leur expression au cours de la nodulation. Les analyses de Clustering Hiérarchique indiquent que ces gènes sont répartis en trois groupes contenant : 1) *AeSymREM*, 2) *AeVPY* et *AeSBT*, 3) *AeNIN* et *AeCRK* (*AeENOD40* n'est pas annoté dans le génome et pas inclus dans cette analyse) soulignant des différences de comportement.

Nous avons suivi l'expression de ces gènes symbiotiques marqueurs par Q-RT-PCR à 0, 2, 4 et 7 jours après inoculation avec la souche de *Bradyrhizobium* ORS278 chez la lignée sauvage et les mutants *pollux-2*, *ccamk-3*, *cyclops-1*, *nsp2-3*, *nin-5 et crk-1* (Fig. 5). Ces lignées ont été choisis car correspondant à des allèles mutants à effet fort pour les gènes étudiés. Chez la lignée sauvage, tous ces gènes sont induits à 2 dpi et atteignent un pic d'expression à 7 dpi. En revanche, l'expression de ces gènes est fortement altérée chez tous les mutants testés, en accord avec le blocage précoce du processus symbiotique observé chez ces mutants. Néanmoins, une analyse plus détaillée révèle plusieurs profils d'expression. L'induction de l'expression d'*AesymREM1*, *AeVPY*, *AeSBT* et *AeCRK* est complètement absente dans les différents fonds mutants (Fig. 5). L'expression d'*AeNIN* est très similaire mais une induction réduite est encore visible dans son propre fond mutant (Fig. 5). Enfin, le comportement d'*AeENOD40* diffère de celui des autres gènes avec une indiction totalement bloquée chez les mutants *pollux-2*, *ccamk-3* et *nsp2-3* et une induction partielle chez les mutants *cyclops-1* et *crk-1* (Fig. 5). Ces données suggèrent que tous les gènes de signalisation symbiotique

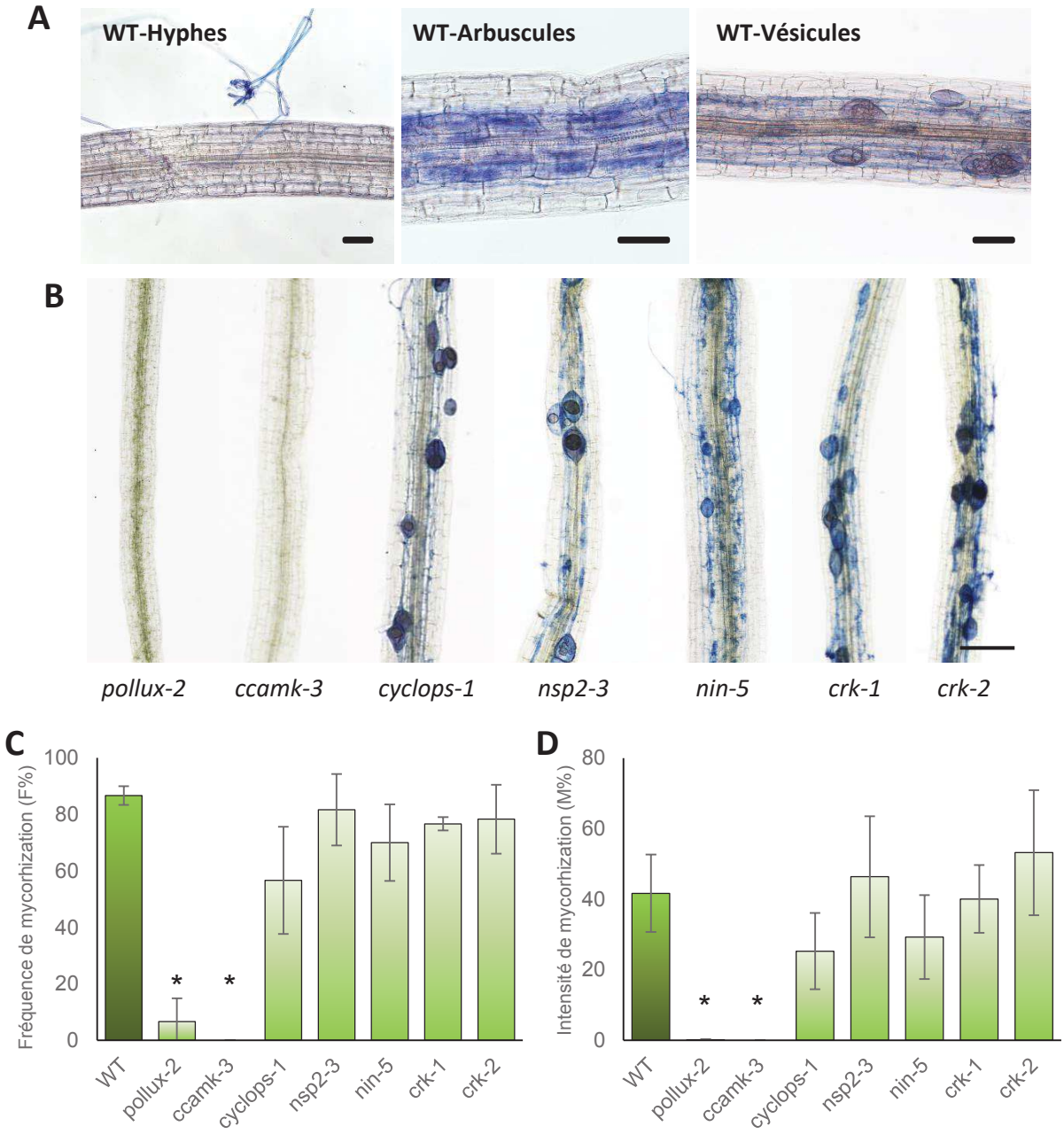


Figure 6 : Phénotypes de mycorhization des mutants des gènes de signalisation d'*A. evenia*. La lignée WT et des différents mutants de gènes de signalisation ont été cultivés pendant 8 semaines avec *R. irregularis* et les racines analysées pour la présence de structures fongiques. **A.** Les trois types de structures fongiques retrouvées chez la lignée WT avec une colonisation par les hyphes, la présence de vésicules et la présence d'arbuscules (de gauche a droite). Les barres d'échelles représentent 50 µm. **B.** Phénotypes de mycorhization des mutants de gènes de signalisation. La barre d'échelle représentent 100 µm. **C.** Fréquence de mycorhization (F%) de la lignée WT et des différents mutants cultivés pendant 8 semaines avec *R. irregularis*. **D.** Intensité de mycorhization (M%) de la lignée WT et des différents mutants cultivés pendant 8 semaines avec *R. irregularis*. En C et D, les moyennes et erreurs standards proviennent de trois réplicats biologiques. Les valeurs significativement différentes de celle du WT avec un F test sont notées par une *.

testés sont requis pour l'induction de l'expression d'*AesymREM1*, *AeVPY*, *AeSBT*, *AeCRK* et *AeNIN*. En revanche, l'induction de l'expression d'*AeENOD40* fait probablement intervenir des régulations indirectes ou parallèles de celle contrôlée par les acteurs symbiotiques testés.

Implication des gènes de signalisation dans la symbiose endomycorhizienne

La dissection génétique de la voie de signalisation Nod chez les légumineuses modèles a montré un chevauchement avec la symbiose endomycorhizienne. Les mutants des gènes de signalisation montrent des phénotypes de mycorhization variables, suggérant différents niveaux d'implication dans la symbiose endomycorhizienne. Nous avons cherché à déterminer si les mutants des gènes de signalisation d'*A. evenia*, qui sont conservés avec *M. truncatula* et *L. japonicus*, présentent des phénotypes de mycorhization similaires dans un contexte symbiotique différent. Aussi, *AeCRK* représentant un gène de nodulation nouveau, nous avons voulu investiguer son rôle potentiel dans cette symbiose. Pour cela, nous avons réalisé des tests de mycorhization en inoculant différentes lignées d'*A. evenia* avec des spores de *Rhizophagus irregularis* : la lignée sauvage, *pollux-2*, *ccamk-3*, *cyclops-1*, *nsp2-3*, *nin-5* et les deux mutants alléliques d'*AeCRK*. Huit semaines après inoculation, les racines des différentes lignées ont été colorées à l'encre de Sheaffer pour observer et quantifier la mycorhization suivant la méthode MycoCalc (https://www2.dijon.inra.fr/mychintec/Mycocalc-prg/download.html). Les racines de la plante sauvage sont fortement colonisées avec des hyphes mycéliens intra-racines, des arbuscules et des vésicules au niveau des cellules du cortex (Fig. 6A). Ces mêmes structures fongiques ont été retrouvées dans les racines des mutants *cyclops-1*, *nsp2-3*, *nin-5, crk-1* et *crk-2*, alors qu'elles sont complètement manquantes dans les racines des mutants *pollux-2*, *ccamk-3* (Fig. 6B). La quantification des structures fongiques dans les racines révèle une fréquence de mycorhization comparable entre la lignée sauvage et les mutants *cyclops-1*, *nsp2-3*, *nin5x, crk-1* et *crk-2* (Fig. 6C). L'intensité de mycorhization est également similaire entre ces différentes lignées. Les mutant *cyclops-1* et *nin-5* présentent des intensités un peu plus faibles que le WT, mais ces différences ne sont pas supportées par les tests statistiques (Fig. 6D). A l'inverse, la fréquence et l'intensité de mycorhization se sont révélées être nulles chez le mutant *ccamk-3* et quasiment nulle chez *pollux-2* (Fig. 6C,D).

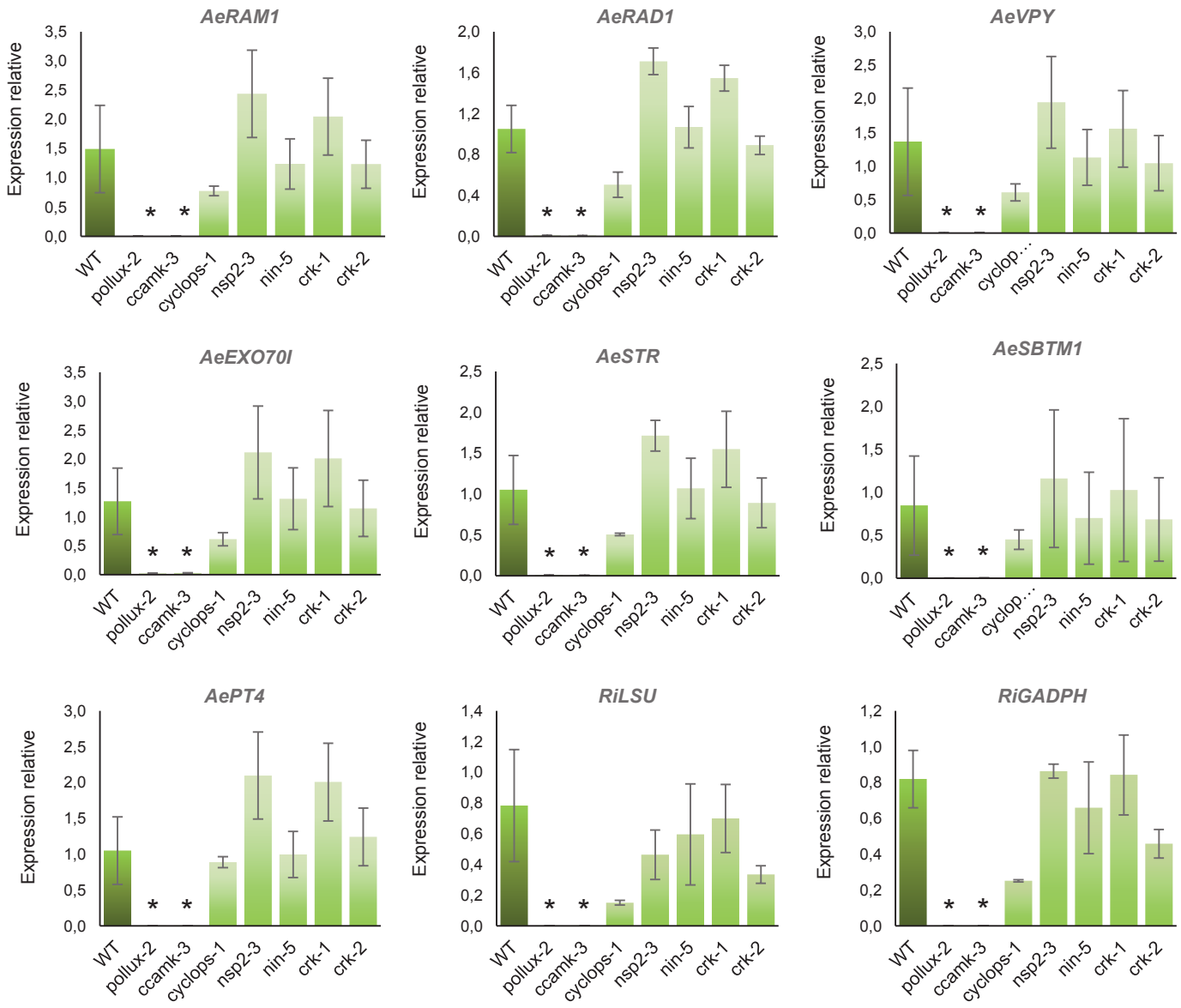


Figure 7 : Analyse de l'expression de gènes marqueurs de la mycorhization chez *A. evenia*. L'expression des sept gènes d'*A. evenia* (*AeRAM1, AeRAD1*, *AeVPY, AeEXO70I*, *AeSTR*, *AeSBTM1* et AePT4) et de deux gènes de *R. irregularis* (*RiLSU* et *RiGADPH*) a été analysée par q-RT-PCR huit semaines après inoculation avec les spores fongiques des racines de la lignée WT et différents mutants de gènes de signalisation. L'expression est normalisée avec celles d'*AeEF1a* et de l'ubiquitine. Les moyennes et les erreurs standards proviennent de trois réplicats biologiques. Les valeurs significativement différentes de celle du WT avec un test Kruskal-Wallis sont notées par une * ($p<0,05$).

Pour pouvoir caractériser d'avantage l'interaction symbiotique avec le champignon mycorhizien chez *A. evenia*, nous avons recherché les orthologues probables de gènes connus pour être induits au cours de la mycorhization chez les légumineuses modèles (Gibelin-Viala *et al.*, 2019; Handa *et al.*, 2015; Park *et al.*, 2015) : *RAM1* et *RAD1* codent des facteurs de transcription impliqués dans la signalisation symbiotiques propre à la mycorhization, *VPY* et *EXO70I* sont des acteurs de l'infection, *STR* et *SBTM1* sont liés au développement des arbuscules intracellulaires, tandis que *PT4* code un transporteur de phosphate lié au fonctionnement des arbuscules. Le niveau de colonisation des racines par *Rhizophagus irregularis* a également été évalué en utilisant deux gènes fongiques : *RiLSU* codant la grande sous-unique ribosomale (Xue *et al.*, 2015) et *RiGADPH* codant une Glycéraldehyde 3-phosphate déhydrogenase (Buendia *et al.*, 2016). La quantification de l'expression de ces gènes de plante et de champignon par Q-RT-PCR chez la lignée sauvage et les différents mutants révèle des niveaux d'expression similaires entre le WT et les mutants *nsp2-3*, *nin-5, crk-1* et *crk-2*, un niveau d'expression réduit pour tous les gènes testés chez *cyclops-1* et des valeurs nulles ou quasi-nulles chez les mutants *ccamk-3* et *pollux-2* (Fig. 7).

L'ensemble des données phénotypiques et d'expression révèlent qu'*A. evenia* est mycorhizable par *Rhizophagus irregularis* et cette interaction symbiotique est fonctionnelle. Parmi les acteurs de signalisation symbiotique, *AePOLLUX* et *AeCCAMK* jouent un rôle essentiel dans le processus de mycorhization et *AeCyclops* est également impliqué même si les altérations observées chez le mutant testé sont plus modérées. En revanche, aucun rôle d'*AeNIN*, *AeNSP2* et *AeCRK* dans la mycorhization n'a été mis en évidence ici.

Discussion

La caractérisation phénotypique des mutants de gènes de signalisation chez *A. evenia* a mis en évidence qu'*AeNSP2* est requis pour le développement des couronnes de poils axillaires. Chez *Arachis hypogaea* (arachide), une approche génétique basée sur la variabilité naturelle a révélé récemment que certaines lignées, obtenues par croisement de différents cultivars et qui sont déficientes à la fois pour la formation de poils axillaires et la nodulation, sont altérées pour les gènes homéologues de *NSP2* (Peng *et al.*, 2020). Cette découverte

parallèle chez une autre légumineuse Dalbergioïde souligne le rôle clé de *NSP2* dans la formation de couronnes de poils axillaires. Ces structures situées à la base des racines latérales sont recouvertes d'une matrice épaisse qui semble attirer par chimiotactisme les bradyrhizobiums et qui constituent le site d'initiation de l'infection (Bonaldi *et al.*, 2011). Cependant, de jeunes plantules d'*Aeschynomene* peuvent être nodulées sur la racine avant que ces poils axillaires n'aient eu le temps de se former. De telles structures ne sont pas présentes sur la tige où se forment de nodules caulinaires. Cela suggère que l'absence de poils axillaires chez les mutants *nsp2* n'est pas la cause directe de la perte de capacité à noduler. De plus, la formation de ces poils axillaires n'est pas dépendante des bradyrhizobiums mais du statut azoté de la plante, indiquant qu'elle n'est induite que si la plante est en carence d'azote. Il semble donc qu'*AeNSP2*, en plus de son rôle dans la voie de signalisation Nod, possède un rôle non symbiotique. Cette hypothèse fait écho à des observations réalisées chez *L. japonicus* et *A. thaliana* que *NSP2* joue un rôle dans le contrôle du développement du système racinaire et des poils absorbants de façon dépendante de l'azote (Murakami *et al.*, 2013 ; Teft, 2018). Une autre étude menée chez *M. truncatula* a révélé que chez des mutants *nsp1* et *nsp2*, plusieurs gènes intervenant dans la voie de synthèse des caroténoïdes et des strigolactones sont réprimés. La synthèse des strigolactones est connue pour être induite par une carence en phosphate (Liu *et al.*, 2011). Il a été proposé que *NSP2* ait une double fonction, une dans la voie de signalisation, l'autre non symbiotique et dépendante du statut nutritionnel de la plante. Cette autre fonction n'est pas encore bien comprise alors qu'elle est considérée comme correspondant à la fonction ancestrale de ce gène (Liu *et al.*, 2011). Chez *A. evenia*, la démonstration expérimentale que l'expression de gènes impliqués dans des processus biologiques différents est dépendant du statut azoté de la plante et d'*AeNSP2* ouvre de nouvelles pistes d'exploration de cette fonction non symbiotique. Les poils axillaires constituent un marqueur morphologique de choix à utiliser pour décortiquer les mécanismes de cette régulation par le statut nutritionnel de la plante et pour mettre à jour le contrôle génétique de leur formation chez les Dalbergioïdes et d'autres légumineuses comme *Sesbania rostrata* chez lesquelles elles sont également présentes (D'Haeze *et al.*, 2003).

Les mutants altérés dans des gènes de signalisation autre qu'*AeNSP2* développent normalement des couronnes de poils axillaires qui sont colonisées par les bradyrhizobiums. Les analyses de microscopie menées sur différentes espèces d'*Aeschynomene* ont montré

qu'à partir de ce site de colonisation, les bactéries pénètrent de façon intercellulaire entre ces poils puis dans le cortex de la racine. Chez les *Aeschynomene* Nod-indépendantes comme *A. evenia*, un « cell collapse » est observée dans le cortex, puis les bactéries sont internalisées dans une ou quelques cellules corticales qui se divisent abondamment pour donner le nodule (Arrighi *et al.*, 2012; Bonaldi *et al.*, 2011). Cependant, les molécules signal supposées être produites par les *Bradyrhizobium* photosynthétiques n'ont pas encore été identifiées, des réponses symbiotiques très précoces comme des flux ioniques et des oscillations calciques n'ont pas encore été caractérisées chez *A. evenia*. L'analyse fine de l'infection intercellulaire et de l'organogenèse nodulaire est également compliquée par l'absence de structures morphologiques faciles à suivre, comme les poils absorbants déformés en crosse de berger et les cordons d'infection, et des gènes rapporteurs restent à être développés comme fait chez *M. truncatula* avec le suivi de l'expression de *MtENOD11* et *MtENOD40* (Boisson-Dernier *et al.*, 2005; Charon *et al.*, 1997). Cependant, la disponibilité de mutants de gènes de signalisation d'*A. evenia* permet maintenant d'entreprendre la dissection génétique de l'établissement de la nodulation et d'identifier les étapes de ce processus soumises à un contrôle génétique (Quilbé *et al.*, accepté). Dans cette étude, nous avons montré que les mutants affectés dans les gènes *AeCCaMK* et *AeNIN* montrent les blocages les plus précoces, avec les bactéries restant en surface des poils axillaires. Ce phénotype fait écho à celui observé chez les mutants des gènes correspondants identifiés chez *M. truncatula* et *L. japonicus*. Néanmoins, des réponses transcriptionnelles, quoique très altérées par rapport à la plante sauvage, sont visibles chez le mutant d'*AeNIN* testé, indiquant qu'il répond encore à l'inoculation avec le partenaire rhizobien comme chez les légumineuses modèles (Schauser *et al.*, 1999). La série de mutants alléliques d'*AePOLLUX* est aussi informative car elle montre trois niveaux de blocage du processus symbiotique avec des mutants colonisés juste en surface des poils axillaires, d'autres présentant des « bumps » discrets avec une infection avortée et un mutant capable de développer des nodules non fixateurs. Cette gradation des phénotypes suggère qu'*AePOLLUX,* et par extension probablement la voie de signalisation symbiotique, intervient dans ces différentes étapes, comme montré aussi chez les légumineuses modèles. Des deux mutants d'*AeCYCLOPS*, un semble être bloqué de façon précoce tandis que le deuxième garde la capacité de développer quelques nodules qui restent petits et non fonctionnels. Ce phénotype est similaire à celui observé chez les mutants *ipd3* de *M. truncatula*, tandis que

les mutants *cyclops* de *L. japonicus* présentent un phénotype plus sévère avec la présence chez ces mutants de « bumps » et une infection bloquée au niveau des poils absorbants (Horváth *et al.*, 2011; Yano *et al.*, 2008). Cependant la surexpression de formes dérégulées du gène *LjCCaMK* permet d'induire le développement de nodules spontanés dans un fond mutant *cyclops*, indiquant que contrairement aux autres gènes de la voie de signalisation symbiotique, *LjCYCLOPS* ne régule que l'infection rhizobienne (Yano *et al.*, 2008).

Le crible d'une population de mutagenèse EMS chez *A. evenia* a permis d'isoler aussi deux mutants alléliques pour le gène *AeCRK* qui représente un gène symbiotique nouveau car n'ayant pas d'orthologue évident chez les légumineuses modèles (Quilbé *et al.*, accepté). La caractérisation phénotypique réalisée dans cette étude a révélé que les deux mutants *crk* sont capables de développer avec des fréquences différentes des « bumps » discrets dans lesquels l'infection est avortée. Ce phénotype est similaire à celui de certains mutants *pollux* chez *A. evenia* et de mutants de gènes symbiotiques décrits chez les légumineuses modèles comme contrôlant l'infection rhizobienne : *MtLYK3* (récepteur des facteurs Nod), *LjCYCLOPS* (signalisation symbiotique), *LjRINRK1* et *MtLIN* (infection) (Catoira *et al*, 2001.; Smit *et al.*, 2007; Yano *et al.*, 2008; Li *et al.*, 2020; Kiss *et al.*, 2009). Les analyses transcriptomiques ont montré qu'*AeCRK* présente le même profil d'expression que le gène *AeNIN* et que son induction est dépendante de ce régulateur. Chez les légumineuses modèles, NIN régule l'expression de nombreux gènes symbiotiques en se fixant directement sur un motif NBS (NIN-Binding Site) (Liu *et al.*, 2019a; Soyano *et al.*, 2014). Il est possible que ce soit également le cas pour *AeCRK*. Comme le gène *NIN*, *AeCRK* semble intervenir spécifiquement dans la symbiose rhizobienne car aucun phénotype lié à la mycorhization n'a pu être observé chez les deux mutants *crk* (cette étude ; Kumar *et al.*, 2020). Néanmoins, on ne peut pas exclure la possibilité qu'il participe à la mycorhization sans être essentiel, à l'image de *NSP2* (Gobbato, 2015). Alors quel est le rôle d'*AeCRK* dans la nodulation ? Les patterns phylogénétiques et moléculaires d'*AeCRK* ont suggéré que ce gène est présent à l'état ancestral chez les légumineuses bien qu'il soit absent chez les modèles et qu'il présente des adaptations à l'activation sans facteurs Nod (Quilbé *et al.*, accepté). Les données phénotypiques et transcriptionnelles obtenues dans cette étude vont dans le sens qu'*AeCRK* est impliqué dans l'infection et/ou la signalisation symbiotique. Un rôle éventuel dans l'organogenèse nodulaire pourrait être testé par des expériences de surexpression de formes dérégulées de la *CCaMK* pour tenter d'induire la formation de nodule spontanés

(Fabre *et al.*, 2015; Yano *et al.*, 2008; Kiss *et al.*, 2009). AeCRK fait partie de la famille des Cystein Receptor-like Kinases dont les membres sont fréquemment associés à la signalisation des ROS (Reactive Oxygen Species) (Bourdais *et al.*, 2015). L'hypothèse dominante est que lors d'un stress abiotique ou biotique, des ROS sont produits dans l'espace extracellulaire, les CRKs en assureraient la perception et la transduction. Ils semblent intervenir aussi dans des boucles de régulation positive pour amplifier la signalisation des ROS (Lee *et al.*, 2017). Des études futures chez *A. evenia* permettront de préciser d'avantage le rôle d'*AeCRK* dans la voie de signalisation symbiotique et/ou l'infection et si ce rôle est associé à la production de ROS lors de l'interaction avec les *Bradyrhizobium* photosynthétiques.

Matériels & Méthodes

Matériel végétal et conditions de germination

Les différents mutants de nodulation d'*A. evenia* et leurs caractéristiques sont présentées dans le Tableau S1. Les graines sont scarifiées à l'acide sulfurique pendant 40 min puis mises une nuit sous agitation (70 tour/min) dans une solution d'éthrel concentré à 5.10^{-5} %. Les graines germées sont par la suite déposées sur de l'eau gélosée et incubées à 37°C pendant 24h avant d'être transféré sur un milieu de culture

Tests de nodulation

Les plantules sont soit transférées dans des pots remplis d'attapulgite et cultivé en serre en condition tropicale, soit dans des tubes contenant du milieu BNM liquide additionné d'azote (0,5mM de KNO_3) et refermés par du papier aluminium et mis à pousser en chambre de culture *in vitro* comme décrit dans Arrighi *et al.* (2012). Les différentes souches de Bradyrhizobium photosyntéthiques ORS278, ORS278-GUS, ORS285 et ORS285-GFP sont cultivées en milieu YM liquide pendant 5 jours à 37°C, milieu dans lequel on ajoute ou non des antibiotiques dépendamment des résistances des souches utilisées (Bonaldi *et al.*, 2011; Giraud *et al.*, 2007).

Tests de mycorhization

Les plantules sont déposées sur un milieu MS gélosé (Murashige-Skoog). Après 5 jours de croissance *in vitro*, 10 plantes/lignée sont transférées sur substrat solide (1:2 terreau:sable) et inoculées avec 100 spores de *Rhizophagus irregularis* DAOM197198 (Agronutrition, Carbone, France) par plante. Les plantes sont alors cultivées pendant 8 semaines en chambre de culture *in vitro* (température de 28°C et un cycle lumineux 16/8h) et arrosées avec un milieu H modifié (milieu Hoagland : K_2SO_4 625µM, $CaSO_4$,$2H_2O$ 499µM, $CaCl_2$, $2H_2O$ 119µM, $MgSO_4$,$7H_2O$ 499µM, $(NH_4)2SO_4$ 128,6µM, $Ca(H_2PO_4)_2$ 5µM, FeEDTA 25µM et une solution d'oligoéléments (H_3BO_3 11,5µM, $MnCl_2$,4H2O 2,25mM, $ZnSO_4$,$7H_2O$ 0,17µM, $CuSO_4$,$5H_2O$ 0.075µM ET NaMo,$2H_2O$ 0.025µM) avec un pH de 5) faible en phosphate (Gherbi *et al.*, 2008b). Les racines des plantes inoculées avec le champignon mycorhizien sont ensuite collectées pour des observations macroscopiques et pour des analyses d'expression de gènes marqueurs de la mycorhization par Q-PCR.

Pour l'analyse des phénotypes de mycorhization, tout le système racinaire secondaire de chaque plant a été excisé, coupé en morceaux et décoloré dans une solution d'hydroxyde de potassium (KOH 10%) chauffée à 90°C pendant 10 min. L'opération est renouvelée une ou plusieurs fois selon l'état de l'échantillon. La décoloration est inactivée dans une solution d'acide acétique 5% qui permet aussi de conserver l'échantillon plusieurs mois à 4°C. La coloration des racines est réalisée à l'encre de Sheaffer en chauffant l'échantillon dans une solution d'acide acétique 5% plus encre de Sheaffer 5% à 90°C pendant 10min, puis 2 rinçages dans l'eau sont réalisés à 20 min d'intervalle (Chabaud *et al.*, 2002). Les racines sont alors montées sur lames dans une solution de PBS/glycérol à 50%. Pour chaque lignée, le système racinaire de 3 à 5 plantes a été observé en lumière blanche au macroscope (Nikon AZ100, Champignysur-Marne, France) et le niveau de colonisation fongique des racines déterminé par la méthode MYCO-CALC (https://www2.dijon.inra.fr/mychintec/Mycocalc-prg/download.html). Les analyses statistiques ont été réalisées en utilisant un test de Student (ns, not significant; *, P-value ≤ 0.05, ***, P-value ≤ 0.01).

Analyses de microscopie et de macroscopie

Racines et nodules sont observés en utilisant le macroscope (Nikon AZ100, Champigny sur-Marne, France) qui permet une vue générale ainsi qu'une vue précise des zones d'intérêt et

le microscope Nikon DS-Ri2 (Tokyo, Japan) qui possède une qualité optique supérieure pour des grossissements importants. Les photos sont prises en fluorescence et en lumière blanche avec le logiciel Nikon Advanced. Les coupes de nodules ont été réalisées à l'aide du vibratome Leica® VT1000S (Mussloch, Germany) réglé sur une épaisseur de coupe de 40 µm. Les coupes sont récupérées puis montées sur lame de microscope afin d'être observées.

Pour la coloration GUS, les échantillons de racines sont couverts d'une solution tampon de X-GLUC (adapté de Jefferson, 1987) : 0,075% de X-GLUC; 0,1M de Na_2HPO_4/KH_2PO_4 (pH : 7); 20mM NaEDTA ; 0,02% Triton X- 100, 1mM de K-ferricyanide et 1mM de K-ferrocyanide. Les échantillons sont ensuite placés dans une pompe à vide avec une pression de -0,9 bar pendant 15 min et sont finalement incuber à 37°C pendant une nuit. Les échantillons sont ensuite traités à l'éthanol 70% à 65°C pendant 2 h pour décolorer la racine et pouvoir mieux observer la coloration.

Analyses de microscopie confocale

Les coupes de racines d'environ 50µm d'épaisseur (émergences de racine latérale ou nodules) sont incubées dans une solution de PBS et calcofluor blanc à 0,01% (M2R de chez Sigma) sous agitation pendant 15min. Après rinçage au PBS, les coupes sont montées sur lame/lamelle avec une solution contenant 50:50 (volume) de PBS:glycérol. Les observations sont réalisées avec le microscope confocal Carl Zeiss LSM700. Le calcofluor est excité à 405nm et la GFP à 488nm. Le signal d'émission à 488nm est collecté par deux canaux (490-520nm pour la GFP et 520-700nm pour l'autofluorescence). Les images obtenues sont ensuite traitées si besoin avec le logiciel Carl Zeiss Confocal Software.

Extraction d'ARN et RT-PCR

L'extraction des ARN totaux est réalisée avec le kit d'extraction RNeasy® Plant Mini Kit (Qiagen®) selon les recommandations du fournisseur suivi d'un traitement à la DNase I (RNase-Free DNase Set, Qiagen®) et d'une purification à l'aide du kit RNeasy MinElute Cleanup Kit (Qiagen®). Les ARNs sont dosés par spectrophotométrie à 260nm (NANODROP 1000 spectromètre, Thermo Scientific) et sont conservés à -20°C. Une appréciation de la qualité des ARN est effectuée en mesurant DO(260nm)/DO(280nm) qui doit être >2 en absence de contamination par des protéines. La rétro-transcription est effectuée à partir de 1 µg d'ARN totaux extraits. Les ARNs sont incubés pendant 5 min à 65°C en présence

d'oligonucléotides poly-dT ancrés (5'-T(18)V3') avant de les mettre dans la glace pour permettre l'hybridation avec les ARNm. La rétro-transcription est catalysée par la SuperScript II (Invitrogen™ Transcriptase inverse SuperScript™ II) dans le tampon commercial pendant 1h à 43°C puis inactivée à 70°C pendant 15 min, selon les instructions du fournisseur. Les ADNc ainsi obtenus sont dilués au 1/50 ème dans de l'eau.

Analyses Q-PCR

Les amorces utilisées pour les gènes symbiotiques ont été dessinées en utilisant le logiciel Beacon Designer (Premier Biosoft) à partir du génome annoté d'*A. evenia* pour les gènes d'intérêt (Tableau S3). Les oligonucléotides ont été synthétisés par Eurofins (Eurofins France). La PCR quantitative est réalisée avec le kit Takyon (Eurogentec) SYBR®Master Mix dTTP Blue en plaque 96 puits dans un LightCycler® Stratagene MX3005P (Agilent Technologies). Les conditions d'amplification sont : 15' à 95°C + (10'' à 95°C + 30'' à 60°C + 60'' à 95°C + 30'' à 60°C)x40 + 30'' à 95°C. Les amorces sont dessinées pour avoir un Tm proche de 60°C. Deux couples d'amorces spécifiques des ADNc de *AeEF1-α* et *AeUbi*, dont l'accumulation est non affectée par les traitements sont utilisés comme témoins internes afin de normaliser les mesures. La concentration d'amorce optimale pour chaque échantillon a été définie lors de tests préliminaires (environ 300nM). Les résultats sont analysés grâce au logiciel MXPro en utilisant la valeur de cycle seuil ou "treshold cycle value" (CT) moyennés sur 2 réplicats techniques. L'accumulation des transcrits est calculée selon la formule : N=10 ((CT-b)/a), où a et b sont des paramètres qui dépendent de l'efficacité de chaque couple d'amorce. Pour chaque couple d'amorce, des dilutions d'ADNc (mélange des différents échantillons de l'expérience) sont utilisées pour obtenir la gamme étalon qui permet de déterminer deux paramètres qui sont la pente et l'ordonnée à l'origine de la courbe Cp=a.log10(dilution)+b. Les résultats correspondants aux transcrits d'intérêts sont normalisés par rapport à la moyenne des deux transcrits *AeEF1-α* et *AeUbi* et l'induction relative des gènes de trois réplicats biologiques indépendants est moyennée pour créer des histogrammes/courbes sur XLStat.

Analyse de Clustering Hiérarchique (HCA)

Pour identifier des groupes de gènes co-exprimés, une analyse de clustering hiérarchique utilisant la corrélation de Pearson a été mise en œuvre. Les données brutes du RNAseq d'A.

evenia ont été transformées en FPKM (fragments per million fragments) en divisant le nombre de comptes par le nombre de millions de fragments par condition. Les lignes de la table de données normalisées ont été filtrées pour ne garder que les lignes ayant au moins un FPKM au-dessus de 10 pour enlever les gènes faiblement exprimés. Un clustering hiérarchique hclust a été réalisé sur une matrice de distance calculée avec la méthode Pearson incluse dans le package amap (version 0.8.17). L'arbre de clustering a été initialement fragmenté en 12 puis 100 clusters. Les clusters ont ensuite été analysés manuellement en fonction des profils d'expression et de la présence de gènes symbiotiques.

Références bibliographiques

Andriankaja, A., A. Boisson-Dernier, L. Frances, L. Sauviac, A. Jauneau, D. G. Barker, and F. de Carvalho-Niebel. 2007. AP2-ERF Transcription Factors Mediate Nod Factor–Dependent Mt *ENOD11* Activation in Root Hairs via a Novel *cis* -Regulatory Motif. *The Plant Cell* 19: 2866–2885.

Arrighi, J.-F., F. Cartieaux, S. C. Brown, M. Rodier-Goud, M. Boursot, J. Fardoux, D. Patrel, *et al.* 2012. *Aeschynomene evenia* , a Model Plant for Studying the Molecular Genetics of the Nod-Independent Rhizobium-Legume Symbiosis. *Molecular Plant-Microbe Interactions®* 25: 851–861.

Boisson-Dernier, A., A. Andriankaja, M. Chabaud, A. Niebel, E.-P. Journet, D. G. Barker, and F. de Carvalho-Niebel. 2005. *MtENOD11* Gene Activation During Rhizobial Infection and Mycorrhizal Arbuscule Development Requires a Common AT-Rich-Containing Regulatory Sequence. *Molecular Plant-Microbe Interactions®* 18: 1269–1276.

Bonaldi, K., D. Gargani, Y. Prin, J. Fardoux, D. Gully, N. Nouwen, S. Goormachtig, and E. Giraud. 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by Photosynthetic *Bradyrhizobium* Sp. Strain ORS285: The Nod-Dependent Versus the Nod-Independent Symbiotic Interaction. *Molecular Plant-Microbe Interactions®* 24: 1359–1371.

Bourdais, G., P. Burdiak, A. Gauthier, L. Nitsch, J. Salojärvi, C. Rayapuram, N. Idänheimo, *et al.* 2015. Large-Scale Phenomics Identifies Primary and Fine-Tuning Roles for CRKs in Responses Related to Oxidative Stress N. M. Springer [ed.],. *PLOS Genetics* 11: e1005373.

Buendia, L., T. Wang, A. Girardin, and B. Lefebvre. 2016. The LysM receptor-like kinase SlLYK10 regulates the arbuscular mycorrhizal symbiosis in tomato. *New Phytologist* 210: 184–195.

Catoira, R. Medicago truncatula root hair curling mutants. 12.

Cerri, M. R., L. Frances, T. Laloum, M.-C. Auriac, A. Niebel, G. E. D. Oldroyd, D. G. Barker, *et al.* 2012. *Medicago truncatula* ERN Transcription Factors: Regulatory Interplay with NSP1/NSP2 GRAS Factors and Expression Dynamics throughout Rhizobial Infection. *Plant Physiology* 160: 2155–2172.

Chabaud, M., C. Venard, A. Defaux-Petras, G. Becard, and D. G. Barker. 2002. Targeted inoculation of Medicago truncatula in vitro root cultures reveals MtENOD11 expression during early stages of infection by arbuscular mycorrhizal fungi. *New Phytologist* 156: 265–273.

Chaintreuil, C., R. Rivallan, D. J. Bertioli, C. Klopp, J. Gouzy, B. Courtois, P. Leleux, *et al.* 2016. A gene-based map of the Nod factor-independent *Aeschynomene evenia* sheds new light on the evolution of nodulation and legume genomes. *DNA Research* 23: 365–376.

Charon, C., C. Johansson, E. Kondorosi, A. Kondorosi, and M. Crespi. 1997. enod40 induces dedifferentiation and division of root cortical cells in legumes. *Proceedings of the National Academy of Sciences* 94: 8901–8906.

D'Haeze, W., R. De Rycke, R. Mathis, S. Goormachtig, S. Pagnotta, C. Verplancke, W. Capoen, and M. Holsters. 2003. Reactive oxygen species and ethylene play a positive role in lateral root base nodulation of a semiaquatic legume. *Proceedings of the National Academy of Sciences* 100: 11789–11794.

Fabre, S., D. Gully, A. Poitout, D. Patrel, J.-F. Arrighi, E. Giraud, P. Czernic, and F. Cartieaux. 2015. The Nod factor-independent nodulation in Aeschynomene evenia required the common plant-microbe symbiotic 'toolkit'. *Plant Physiology*: pp.01134.2015.

Gherbi, H., K. Markmann, S. Svistoonoff, J. Estevan, D. Autran, G. Giczey, F. Auguy, *et al.* 2008. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiabacteria. *Proceedings of the National Academy of Sciences* 105: 4928–4932.

Gibelin-Viala, C., E. Amblard, V. Puech-Pages, M. Bonhomme, M. Garcia, A. Bascaules-Bedin, J. Fliegmann, *et al.* 2019. The *Medicago truncatula* LysM receptor-like kinase LYK9 plays a dual role in immunity and the arbuscular mycorrhizal symbiosis. *New Phytologist* 223: 1516–1529.

Giraud, E., and D. Fleischman. 2004. Nitrogen-fixing symbiosis between photosynthetic bacteria and legumes. *Photosynthesis Research* 82: 115–130.

Giraud, E., L. Moulin, D. Vallenet, V. Barbe, E. Cytryn, J.-C. Avarre, M. Jaubert, *et al.* 2007. Legumes Symbioses: Absence of Nod Genes in Photosynthetic Bradyrhizobia. *Science* 316: 1307–1312.

Gobbato, E. 2015. Recent developments in arbuscular mycorrhizal signaling. *Current Opinion in Plant Biology* 26: 1–7.

Gully, D., P. Czernic, S. Cruveiller, F. Mahé, C. Longin, D. Vallenet, P. François, *et al.* 2018. Transcriptome Profiles of Nod Factor-independent Symbiosis in the Tropical Legume Aeschynomene evenia. *Scientific Reports* 8: 10934.

Handa, Y., H. Nishide, N. Takeda, Y. Suzuki, M. Kawaguchi, and K. Saito. 2015. RNA-seq Transcriptional Profiling of an Arbuscular Mycorrhiza Provides Insights into Regulated and Coordinated Gene Expression in *Lotus japonicus* and *Rhizophagus irregularis*. *Plant and Cell Physiology* 56: 1490–1511.

Horváth, B., L. H. Yeun, Á. Domonkos, G. Halász, E. Gobbato, F. Ayaydin, K. Miró, *et al.* 2011. *Medicago truncatula IPD3* Is a Member of the Common Symbiotic Signaling Pathway Required for Rhizobial and Mycorrhizal Symbioses. *Molecular Plant-Microbe Interactions®* 24: 1345–1358.

Jefferson, R. A. 1987. Assaying chimeric genes in plants: The GUS gene fusion system. *Plant Molecular Biology Reporter* 5: 387–405.

Kiss, E., B. Oláh, P. Kaló, M. Morales, A. B. Heckmann, A. Borbola, A. Lózsa, *et al.* 2009. LIN, a Novel Type of U-Box/WD40 Protein, Controls Early Infection by Rhizobia in Legumes. *Plant Physiology* 151: 1239–1249.

Kumar, A., D. R. Cousins, C.-W. Liu, P. Xu, and J. D. Murray. 2020. Nodule Inception Is Not Required for Arbuscular Mycorrhizal Colonization of Medicago truncatula. *Plants* 9: 71.

Lee, D. S., Y. C. Kim, S. J. Kwon, C.-M. Ryu, and O. K. Park. 2017. The Arabidopsis Cysteine-Rich Receptor-Like Kinase CRK36 Regulates Immunity through Interaction with the Cytoplasmic Kinase BIK1. *Frontiers in Plant Science* 8: 1856.

Li, X., Z. Zheng, X. Kong, J. Xu, L. Qiu, J. Sun, D. Reid, *et al.* 2020. Atypical Receptor Kinase RINRK1 Required for Rhizobial Infection But Not Nodule Development in Lotus japonicus1[OPEN]. 181: 13.

Liu, C.-W., A. Breakspear, D. Guan, M. R. Cerri, K. Jackson, S. Jiang, F. Robson, *et al.* 2019. NIN Acts as a Network Hub Controlling a Growth Module Required for Rhizobial Infection. *Plant Physiology* 179: 1704–1722.

Liu, W., W. Kohlen, A. Lillo, R. Op den Camp, S. Ivanov, M. Hartog, E. Limpens, *et al.* 2011. Strigolactone Biosynthesis in *Medicago truncatula* and Rice Requires the Symbiotic GRAS-Type Transcription Factors NSP1 and NSP2. *The Plant Cell* 23: 3853–3865.

Mangano, S., S. P. Denita-Juarez, E. Marzol, C. Borassi, and J. M. Estevez. 2018. High Auxin and High Phosphate Impact on RSL2 Expression and ROS-Homeostasis Linked to Root Hair Growth in Arabidopsis thaliana. *Frontiers in Plant Science* 9: 1164.

Murakami, Y., H. Yokoyama, R. Fukui, and M. Kawaguchi. 2013. Down-Regulation of NSP2 Expression in Developmentally Young Regions of Lotus japonicus Roots in Response to Rhizobial Inoculation. *Plant and Cell Physiology* 54: 518–527.

Park, H.-J., D. S. Floss, V. Levesque-Tremblay, A. Bravo, and M. J. Harrison. 2015. Hyphal branching during arbuscule development requires RAM1. *Plant Physiology*: pp.01155.2015.

Peng, Z., H. Chen, L. Tan, H. Shu, R. K. Varshney, Z. Zhou, Z. Zhao, *et al.* 2020. Natural Polymorphisms in a Pair of NSP2 Homoeologs Can Cause Loss of Nodulation in Peanut. *Journal of Experimental Botany*: eraa505.

Roy, S., W. Liu, R. S. Nandety, A. Crook, K. S. Mysore, C. I. Pislariu, J. Frugoli, *et al.* 2020. Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *The Plant Cell* 32: 15–41.

Schauser, L., A. Roussis, J. Stiller, and J. Stougaard. 1999. A plant regulator controlling development of symbiotic root nodules. *Nature* 402: 191–195.

Smit, P., E. Limpens, R. Geurts, E. Fedorova, E. Dolgikh, C. Gough, and T. Bisseling. 2007. Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling. *Plant Physiology* 145: 183–191.

Soyano, T., H. Hirakawa, S. Sato, M. Hayashi, and M. Kawaguchi. 2014. NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. *Proceedings of the National Academy of Sciences* 111: 14607–14612.

Teft M. (2018) Characterising the role of ATSCL26 in root development. PhD thesis, University of Warwick.

Xue, L., H. Cui, B. Buer, V. Vijayakumar, P.-M. Delaux, S. Junkermann, and M. Bucher. 2015. Network of GRAS Transcription Factors Involved in the Control of Arbuscule Development in *Lotus japonicus*. *Plant Physiology* 167: 854–871.

Yano, K., S. Yoshida, J. Muller, S. Singh, M. Banba, K. Vickers, K. Markmann, *et al.* 2008. CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proceedings of the National Academy of Sciences* 105: 20540–20545.

Données supplémentaires

Tableau S1 : Liste des mutants alléliques pour les gènes de signalisation d'*A. evenia*.

Gène	Mutant	Allèle	Effet de la mutation	
			Changement nt	Changement aa
AeCCaMK (Ae08g13330)	H34	*ccamk-1*	G_{3905} to A	A_{450} to STOP
	J31	*ccamk-2*	G_{2405} to A	S_{341} to STOP
	T26	*ccamk-3*	C_{1204} to T	P_{218} to L
	AK7	*ccamk-4*	C_{1965} to T	Splice site
AeCRK (Ae05g12380)	I10	*crk-1*	G_{2228} to A	Splice site
	J42	*crk-2*	G_{1062} to A	G_{354} to E
AeCYCLOPS (Ae05g02230)	L33	*cyclops-1*	G_{2372} to A	Splice site
	S31	*cyclops-2*	G_{2325} to A	Splice site
AeNIN (Ae07g00100)	E26	*nin-1*	C_{1919} to T	Q_{352} to STOP
	G40	*nin-2*	A_{2800} to G C_{2768} to T	K_{647} to E P_{636} to L
	S11	*nin-3*	A_{339} to T	R_{114} to STOP
	V20	*nin-4*	T_{142} to A	L_{48} to Q
	Y11	*nin-5*	C_{3173} to T	R_{729} to STOP
	AD8	*nin-6*	C_{3048} to T	L_{687} to F
AeNSP2 (Ae08g09660)	C10	*nsp2-1*	C_{692} to T	P_{231} to L
	J25	*nsp2-2*	A_{430} to T	R_{144} to STOP
	M30	*nsp2-3*	G_{1545} to A	W_{515} to STOP
	AK8	*nsp2-4*	C_{806} to T	S_{269} to F
AePOLLUX (Ae04g09220)	C15	*pollux-1*	G_{1699} to A	G_{237} to R
	E28	*pollux-2*	G_{5039} to A	W_{716} to STOP
	H10	*pollux-3*	G_{5476} to T	E_{833} to STOP
	J33	*pollux-4*	G_{3444} to A	G_{520} to E
	P34	*pollux-5*	G_{1931} to A	S_{314} to N
	R16	*pollux-6*	A_{1944} to T	E_{318} to D

Tableau S2 : Données d'expression de gènes d'A. evenia induits ou réprimés au cours de la nodulation

Profil d'expression	Gene	Gene ID	Gene Atlas d'*A. evenia* CIAT22838 (FPKM)								
			Racine -N	Racine +N	Nod 4dpi	Nod 7dpi	Nod 14dpi	Tige	Feuille	Fleur	Gousse
Induits au cours de la nodulation	*AeCRK*	Ae05g12380	2,28	2,85	16,04	49,92	119,68	0	0,03	0	0
	AeNIN	Ae07g00100	2,89	4,92	24,48	73,28	174,15	0,08	0,02	0,14	0,04
	AeSBT	Ae05g09230	4,47	6,46	96,73	574,03	793,15	0,13	0,57	1,28	0,26
	AeSYMREM1	Ae03g30480	0,29	0	41,07	71,57	9,14	0,15	0	0	0,32
	AeVPY	Ae05g16930	0,54	1,05	17,63	53,53	54,84	0,04	0,08	0,55	0,31
Réprimés par l'azote et au cours de la nodulation	*AeAMT3*	Ae03g29540	14,4	32,15	5,67	1,48	1,18	1,53	9,56	3,14	1,6
	AeCCD7	Ae02g22440	18,79	36,87	1,37	1,71	0,1	1,11	0,17	0	0,07
	AeERN1	Ae06g15750	22,8	44,35	8,77	8,75	4,38	0,15	0,08	0,17	2,53
	AeERN3	Ae03g25920	1,55	4,33	0	0	0	0	0	0	0
	AeEXP	Ae05g09290	17,92	55,02	0,19	0	0,18	0	0	0	0,18
	AeNSP2	Ae08g09660	14,92	28,55	0,97	1,64	0,68	0,8	0,5	0,34	0,32
	AePEROX	Ae04g12180	8,75	21,36	0,07	0	0	0	0	0	0
	AePE	Ae08g13350	17,4	50,7	0,38	0,08	0	0	0	0,04	0
	AePL	Ae01g30710	14,48	35,02	0,06	0,13	0	0	0	0	0
	AeRSL2	Ae09g24590	18,13	45,28	0,21	0	0,13	0,13	0	0	0

Tableau S3 : Liste des gènes et des amorces utilisées pour la Q-RT-PCR

Gene	Gene ID	Amorces	Séquence 5'-3'
AeAMT3	Ae03g29540	AeAMT3-F	TGAGAAGCTCTTACCATTT
		AeAMT3-R	AAACCACCATTCTTGTAGT
AeCCD7	Ae02g22440	AeCCD7-F	AGAGACTGGTGAGAATGATT
		AeCCD7-R	CTCCTTGGTGGCATCTTA
AeCRK	Ae05g12380	AeCRK-F	CCCCAATAGTTCTGAGCCAACC
		AeCRK-R	AGGAAGAGGCCATAGAGCCTTG
AeEF1a	Ae09g20140	AeEF1-F	TGCTGGTATGGTTAAGATGGTTCC
		AeEF1-R	TTCTTCTTCTGTGCTGCCTTGG
AeENOD40	non annoté	AeNOD40-F	CACACTTCTCCTCCATTCACTTTTC
		AeNOD40-R	TTGCCATACTTGTAGCCAAAAGC
AeERN1	Ae06g15750	AeERN1-F	ACTCGCACCAACTTCATA
		AeERN1-R	TGTTATTAAGAAGATTCCTGATCC
AeERN3	Ae03g25920	AeERN3-F	GATGTGGAGAGGCAGTTA
		AeERN3-R	GGAGAGAGCGTAATTTCC
AeEXO70I	Ae08g04110	AeEXO70I-1-F	CGTGTATCAGAAGCAAGC
		AeEXO70I-1-R	CCAAGCACTTCAAGAATGT
AeEXP	Ae05g09290	AeEXP-F	GGCTTCCACAATAGTTCAA
		AeEXP-R	GACTGGGTCGAAATATAATTTG
AeNIN	Ae07g00100	AeNIN-F	CAACAGAACAAGGGGAAAGGGG
		AeNIN-R	TAATGAGGCAGAGGCGGAAGTG
AeNSP2	Ae08g09660	AeNSP2-F	ATAACAATGGACATGCTGGAT
		AeNSP2-R	GGTAGTCGTCGGAAGAGG
AePE	Ae08g13350	AePE-F	TTCTTCTCTGGTGATGGT
		AePE-R	AGATCCTGGCAATTCTTC
AePEROX	Ae04g12180	AePEROX-F	ATGGTGAACTTGTCAAACT
		AePEROX-R	CCTTCTTGGGATGAACTTT
AePL	Ae01g30710	AePL-F	ACCTCACTGCCTATTCAAC
		AePL-R	AAGCCAATAGCACAGTCA
AePT4	Ae08g08360	AePT4-F	GCTGTTTGCTCTGCTTTA
		AePT4-R	CATCCACCGTGTAGTTCT
AeRAD1	Ae01g19180	AeRAD1-F	TATGCAAACAGTTTAGGAATCA
		AeRAD1-R	ATCATCTGAAGTACCGAGTT
AeRAM1	Ae06g18380	AeRAM1-F	GTAATTCCCTCGGTAATCTGTT
		AeRAM1-R	TGGCTTCCTGCTCAACTA

Tableau S3 suite : Liste des gènes et des amorces utilisées pour la Q-RT-PCR

Gene	Gene ID	Amorces	Séquence 5'-3'
AeRSL2	Ae09g24590	AeRSL2-F	GTCTCAGAGAAGCACTGT
		AeRSL2-R	CATTGAAAACCATGTGAACTAA
AeSBT	Ae05g09230	AeSBT-F	ATGAAGGAATGCCACCTCCACC
		AeSBT-R	TGTGTGTGCCGTGTCCATTATC
AeSTR	Ae05g35200	AeSTR-F	TTCCTCGTATGTCATATCTTCA
		AeSTR-R	GCATTGGTTGTGATAAGTGA
AeSBTM1	Ae05g09240	AeSBTM1-F	TGATATAGGCGGAGGAGT
		AeSBTM1-R	ACATGAACAGCAGGAAGTA
AeSYMREM1	Ae03g30480	AeSYMREM1-F	TGATGATGCTGCTGATGA
		AeSYMREM1-R	TTTGGGTTGTGAAGAAGTG
AeVPY	Ae05g16930	AeVPY-F	CTGAGGCTTCTCTTGCTTA
		AeVPY-R	CTATGGCTGCTGCTATGT
AeUbi	Ae10g10900	AeUbi-F	TCAAAGTGAAGACTCTAACCG
		AeUbi-R	CAAGTGAAGCACGGAACC
RiGADPH	*Rhizophagus irregularis*	RiGADPH-F	GACGTCTCAGTTGTTGATTTA
		RiGADPH-R	TTTGGCATCAAAAATACTAGA
RiLSU		RiLSU-F	GCATATCAATAAGCGGAGGA
		RiLSU-R	ACTCCTCACGCTCCACAGA

2.3 - ADDENDUM : NSP2, a key symbiotic regulator in the spotlight

Lors du crible de la population de mutagenèse EMS chez *A. evenia*, l'isolement de plusieurs mutants ne développant pas de couronnes de poils axillaire et totalement déficients pour la nodulation a éveillé notre curiosité. En effet, si les poils axillaires situés à la base des racines latérales sont connus pour être le premier site de colonisation par la bactérie chez les légumineuses Dalbergioïdes comme *Aeschynomene* spp. et *Arachis* spp. (Bonaldi *et al.*, 2011 ; Sharma *et al.*, 2020), et qu'ils sont également retrouvés chez quelques autres légumineuses comme *Sesbania rostrata* (D'Haeze *et al.*, 2003), leur rôle exact lors des premières étapes de l'interaction symbiotique n'est pas bien compris. Chez l'arachide (*A. hypogagea*), l'existence de lignées obtenues par croisement de différents cultivars, qui sont incapables de noduler et qui sont également dépourvues de couronnes de poils axillaires est connue depuis longtemps (Gorbet et Burton, 1979). Ces lignées ont fait l'objet de nombreuses études génétiques, transcriptomiques et phénotypiques (pour les plus récentes : Peng *et al.*, 2017 et 2018). Le phénotype de ces lignées de l'arachide fait écho à celui observé chez certains mutants de nodulation d'*A. evenia* et des approches de génétique directe menées en parallèle sur ces deux espèces ont conduit à l'identification du même gène symbiotique, *NSP2*. Les travaux de l'équipe de Jiangping Wang (Université de Floride) sur l'arachide qui ont abouti à la découverte de *NSP2* a fait l'objet d'une publication dans le Journal of Experimental Botany (Peng *et al.*, 2020). A cette occasion, nous avons publié un Insight associé à cet article. Cet Insight met en exergue, d'une part l'émergence de nouveaux modèles d'étude comme l'arachide et *A. evenia* pour aborder la diversité des mécanismes de nodulation chez les légumineuses, et d'autre part les rôles symbiotiques et non-symbiotiques de *NSP2*. Nous proposons aussi un modèle qui unifie les différents rôles de *NSP2*, ce qui fournit une base de travail pour continuer à caractériser la fonction de ce gène clé.

eXtra Botany

Insight

NSP2, a key symbiotic regulator in the spotlight

Johan Quilbé1 and Jean-François Arrighi1,*

1LSTM, Univ. Montpellier, CIRAD, INRA, IRD, SupAgro, 34398 Montpellier, France

*Correspondence: jean-francois.arrighi@ird.fr

This article comments on:

Peng Z, Tan L, Chen H, Shu H, Varshney R.K., Zhou Z, Zhao Z, Luo Z, Chitikineni A, Wang L, Maku J, López Y, Gallo M, Zhou H and Wang J. 2020. Mendelian and Non-Mendelian Inheritance of a Pair of Homoeologous GRAS Transcription Factor Genes Coordinately Controlling Nodulation in Peanut. Journal of Experimental Botany **XX,** XXXX–XXXX.

Keywords: peanut, legume, nodulation, symbiosis, signaling, nitrogen, NSP2

Legumes have the remarkable ability to develop root nodules with nitrogen-fixing rhizobium bacteria. This symbiosis enables them to overcome nitrogen limitation affecting their growth, and such nutritional advantage is key in the ecological and agronomic success of legumes. Intriguingly, inbreeding work in cultivated peanut can lead to the obtaining of non-nodulating lines. Using elegant genetic approaches, Peng *et al.* (2020) have elucidated this 40 years-old mystery by identifying the underpinning genes. They correspond to the two homoeologs of the symbiotic signaling gene *NSP2*, pinpointing *NSP2* as being of main adaptative importance for nodulation in legumes.

Understand the genetics of nodulation to cope with plant nitrogen supply

Nitrogen is an essential nutrient for plant growth but, in many eco- and agro-systems, its availability in the soil is limited. To overcome nitrogen limitation, most legumes are able to interact symbiotically

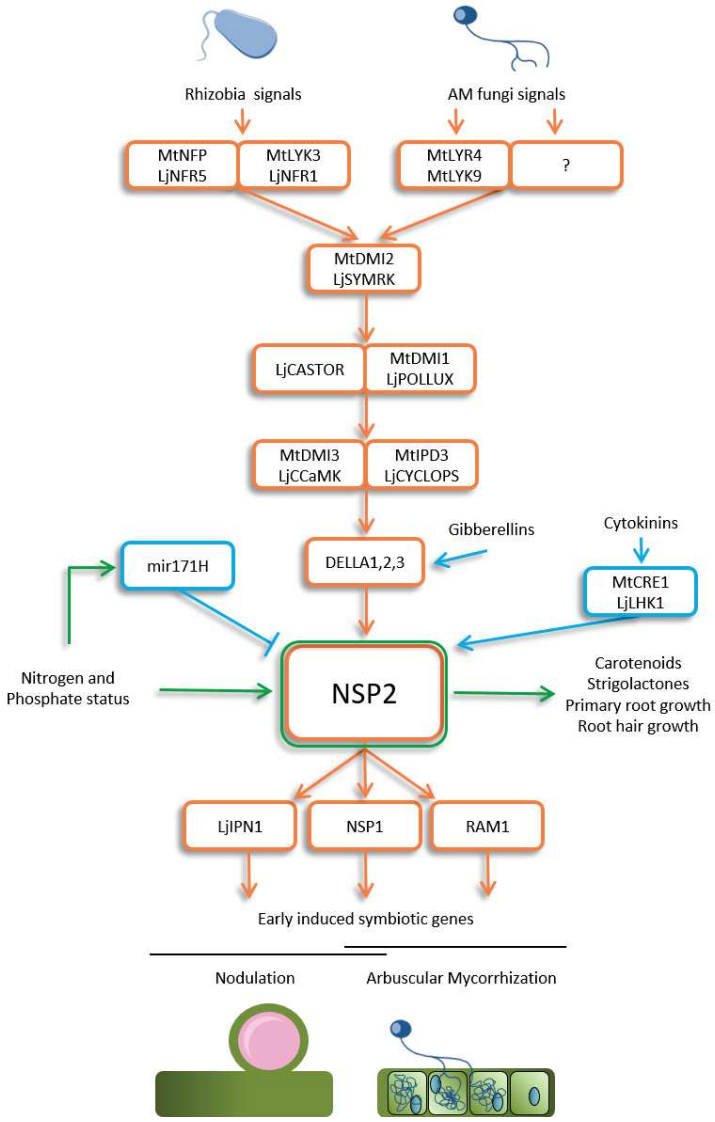
with rhizobial bacteria. They are housed in root nodules in which they find suitable conditions to reduce atmospheric nitrogen into ammonium that is used by the host plant. Promoting this biological nitrogen fixation with legume crops, instead of chemical fertilizers as inputs, is expected to lead to a more sustainable agriculture. Such green evolution also includes the improving of nitrogen-fixation in some cultivated legumes and its engineering in non-legume crops.

To achieve these goals, genetic studies were initiated 20 years ago on two legume models *Medicago truncatula* and *Lotus japonicus*. They successfully uncovered many genes underpinning nodulation, most specially those involved in a symbiotic signalling pathway that is activated upon recognition of rhizobial Nod factors and that stretches from plasma membrane receptors to a nuclear transcriptional network. In turn, symbiotic transcription factors trigger rhizobial infection through root epidermis and coordinated nodule organogenesis across the root cortical cells (reviewed in Roy *et al.*, 2020). Entrance of Nod factor-producing rhizobia via root hairs and infection thread formation to deliver them into distantly induced root nodule primordia is the most common process found in legumes and also the best understood because studied in legume models. However, alternative nodulation processes occur in legumes, notably an intercellular invasion that is observed in 25% of all legume genera (mostly Genistoid and Dalbergioid legumes) (Masson-Boivin, 2009). In that case, rhizobia directly access to root cortical cells to develop nodule primordia, but the genetic and molecular study of this intercellular infection-mediated nodulation is still in infancy (Sharma *et al.*, 2020).

Fortunately, study of the rhizobial symbiosis has recently broadened with the rising of additional symbiotic systems, among which are two Dalbergioid legume species: *Arachis hypogaea* (peanut) and *Aeschynomene evenia*. On one hand, peanut (2n=4*x*=40, AABB polyploid genome of ~2.7 Gb) is a major oilseed crop and is second after soybean among cultivated legumes. On the other hand, *A. evenia* (2n=2*x*=20, 400 Mb) has attracted attention due to its original ability to develop stem and roots nodules in a Nod factor-independent fashion (Arrighi *et al.*, 2012). Transcriptomic analyses and reverse genetics performed on these Dalbergioid species evidenced that some symbiotic mechanisms described *M. truncatula* and *L. japonicus* are conserved and others not (Sharma *et al.*, 2020; Karmakar *et al.*, 2019; Gully *et al.*, 2018; Fabre *et al.*, 2015). Therefore, research in these legume species is warranted to provide critical insights into nodulation diversity.

Exploitation of natural genetic variation in peanut identifies the symbiotic *NSP2* gene

Peng *et al.* (2020) have stepped forward this nodulation research by making use of forward genetic strategies on non-nodulating inbred lines in peanut. Initial report of non-nodulating lines date back to 1979 when Gorbet and Burton fortuitously obtained them by crossing normally-nodulating peanut accessions. Multiple genetic analyses indicated that the non-nodulation trait was under the control



Box 1. Model of NSP2 as a central regulatory hub integrating symbiotic and nutritional cues to mediate symbiotic and non-symbiotic responses. This schematic representation gathers diverse findings on NSP2 functions. Given that most studies were made on M. truncatula (Mt) and Lotus japonicus (Lj), names of proteins from these model legumes are used (no prefix if the same name is used). In the nitrogen-fixing symbiosis, rhizobial Nod factors are perceived by LysM-RLK receptor kinases. This recognition triggers a symbiotic signalling pathway, in which NSP2 is an essential transcriptional regulator by interacting with the NSP1 and IPN1 transcription factors to upregulate the expression of several key symbiotic genes. NSP2 is required for nodulation in legumes using an infection thread-mediated infection as in legume models, or an intercellular infection as revealed in peanut by Peng et al. (2020). This symbiotic signaling pathway is in part shared with mycorrhizal symbiosis, to which NSP2 contributes by interacting with NSP1 and RAM1 transcription factors. NSP2 symbiotic functions are dependent on cross-talks between the symbiotic, the DELLA-mediated gibberellin, and the MtCRE1/LjLHK1-dependent cytokinin signallings, and are tightly regulated by mir171h in a symbiotic, N- and P-dependent fashion. The non-symbiotic functions of NSP2 include regulation of the metabolic pathway leading to carotenoid and strigolactone biosynthesis and modulation of root and root hair growth. These biological processes are also dependent on plant nutrient status. Considered together, all of these data position NSP2 as a probable regulatory hub integrating symbiotic and nutritional cues to coordinate different adaptive responses. Orange boxes indicate symbiotic components, blue boxes gather signify hormonal and post-transcriptional regulatory cues and green boxes designate nutrient-dependent non-symbiotic functions.

of two genes or more in peanut (discussed in Peng *et al.*, 2018). Peng *et al.* (2020) successfully disentangled this messy genetic determinism by combining map-based cloning and QTL-seq approaches that enabled the identification in the allopoplyploid peanut genome of deleterious mutations in the two homoeologous *NSP2* genes. Functional complementation of non-nodulating lines by overexpressing the WT homoeolog *NSP2* versions confirmed that they are conjointly responsible of the non-nodulation trait. Intriguingly, their segregation patterns are different, *AhNSP2-A08* follows a Mendelian inheritance (with a typical 3:1 ratio) while *AhNSP2-B07* exhibits a non-Mendelian inheritance (5:3∼1:1 ratio). What's more, contribution of their respective deleterious alleles to the non-nodulation phenotype is unequal and modulated by a parental effect. The whole makes peanut *NSP2* genes a complex genetic case.

From an symbiotic point of view, one can wonder why *NSP2* alleles leading to gene loss-of-function are present in peanut. Indeed *NSP2* encodes a key GRAS-type transcriptional regulator acting downstream of the CCAMK/CYCLOPS complex in the nodulation signalling pathway, that integrates phytohormone cues and controls the expression of several symbiotically-induced genes (detailed in Roy *et al.*, 2020) (Box 1). As a matter of fact, cytokinins activate *NSP2* expression and this activation requires the CRE1 receptor (Ariel *et al.*, 2012). Cytokinins interplay with gibberellins that also regulates the nodulation signaling pathway (Maekawa *et al.*, 2009). This regulation involves gibberellin signaling-mediating DELLA proteins that interact with NSP2 to enhance the expression of Nod factor-inducible genes (Jin *et al.*, 2016; Fonouni-Farde *et al.*, 2016). To ensure its function of transcriptional regulator, NSP2 interacts with NSP1, another GRAS protein, and also IPN2, a MYB transcription factor (Xiao *et al.*, 2020 ; Hirsch *et al.*, 2009). Together they form a protein complex where, contrary to NSP2, NSP1 and IPN2 have a DNA-binding domain allowing them to associate directly with specific *cis* elements present in the promoter of some symbiotic genes including *NIN*, *ERN1* and *ENOD11* (Xiao *et al.*, 2020 ; Hirsch *et al.*, 2009).

Stepping forward the characterization of *NSP2* multiple functions

NSP2 was originally thought to function specifically in nodulation, but latter studies extended its involvement to the arbuscular mycorrhizal (AM) symbiosis that shares several other signalling- and infection-players with the rhizobial symbiosis (Box 1). In the AM symbiosis, NSP2 is not essential but it participates with NSP1 to signaling and to AM root colonization, and it also interacts with the GRAS protein, RAM1, to control the mycorrhizal-induced expression of *RAM2* (reviewed in Gobbato, 2015). As another common feature of the mycorrhizal and rhizobial symbioses, *NSP2* transcripts are target of the microRNA miR171h and they show a mainly mutual exclusive expression pattern in the root (Ariel *et al.*, 2012 ; Hofferek *et al.*, 2014). This miRNA-mediated negative regulation of *NSP2* likely

serves as regulatory mechanism to restrict symbiosis development depending on the symbiotic and also the nutritional (phosphate and nitrogen) status of the plant (Hofferek *et al.*, 2014).

It is worthy to note that *Arabidopsis thaliana*, which is unable to develop any symbiosis, has lost components of the symbiotic signaling pathway but has retained NSP1 and NSP2 orthologous genes, suggesting that these GRAS proteins have a non-symbiotic function (Liu *et al.*, 2011). The fact that legume *NSP1* and *NSP2* can be functionally replaced by orthologs notably from rice further indicates that their function is conserved (Yokota *et al.*, 2010). This more generic function is poorly understood, but a study led conjointly in *M. truncatula* and rice revealed that *NSP1* and *NSP2* regulate genes of the carotenoid and strigolactone biosynthetic pathway in a phosphate-dependent manner (Liu *et al.*, 2011). Other analyses performed in *L. japonicus* and *A. thaliana* also pinpointed a role of *NSP2* in the nitrogen-dependent regulation of root and root hair elongation (Murakami *et al.*, 2013 ; Teft, 2018).

All these scattered puzzle pieces show that NSP2 has multiple functions. Liu *et al.* (2011) proposed that NSP2 fulfills dual regulatory functions: a conserved non-symbiotic function in higher plants and an additional function in the symbiotic signalling where it is activated in a different manner to control downstream genes. However, one might consider that NSP2 generic function depends on the plant nutritional status and that NSP2 is involved in symbioses improving the nitrogen and phosphate nutrition of the plant. This suggests a link between the generic and symbiotic functions. Therefore, it is tempting to speculate that NSP2 constitutes a central regulatory hub integrating both nutrient homeostasis and symbiotic signaling to coordinate adaptative responses (Box 1).

The discovery by Peng *et al.* (2020) that non-nodulating lines are altered in *NSP2* in peanut provides now the opportunity to expand the study of this key gene: can Dalbergioid legumes serve to uncover additional nutrient status-dependent regulations? What is the role of *NSP2* during intercellular infection? Is it also involved in stem nodulation and in Nod factor-independent signaling in the other Dalbergioid legume *A. evenia*? Novel findings would significantly contribute to our understanding of the panoply of *NSP2* regulatory functions.

References

Ariel F, Brault-Hernandez M, Laffont C, *et al.* 2012.Two direct targets of cytokinin signaling regulate symbiotic nodulation in Medicago truncatula. The Plant Cell **24**, 3838-52.

Arrighi JF, Cartieaux F, Brown SC, *et al.* 2012. *Aeschynomene evenia*, a model plant for studying the molecular genetics of the nod-independent rhizobium-legume symbiosis. Molecular Plant Microbe Interactions **25**, 851-61.

Fabre S, Gully D, Poitout A, Patrel D, Arrighi JF, Giraud E, Czernic P, Cartieaux F. 2015. Nod Factor-Independent Nodulation in *Aeschynomene evenia* Required the Common Plant-Microbe Symbiotic Toolkit. Plant Physiology **169**, 2654-64.

Fonouni-Farde C, Tan S, Baudin M, Brault M, Wen J, Mysore KS, Niebel A, Frugier F, Diet A. 2016. DELLA-mediated gibberellin signalling regulates Nod factor signalling and rhizobial infection. Nature Communications **7**, 12636.

Gobbato E. 2015. Recent developments in arbuscular mycorrhizal signaling. Current Opinion in Plant Biology **26**, 1-7.

Gorbet DW, Burton JC. 1979. A non-nodulating peanut. Crop Sciences **19**, 727-728.

Gully D, Czernic P, Cruveiller S, *et al.* 2018. Transcriptome Profiles of Nod Factor-independent Symbiosis in the Tropical Legume *Aeschynomene evenia*. Scientific Reports **8**, 10934.

Hirsch S, Kim J, Muñoz A, Heckmann AB, Downie JA, Oldroyd GE. 2009. GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in Medicago truncatula. The Plant Cell **21**, 545-57.

Hofferek V, Mendrinna A, Gaude N, Krajinski F, Devers EA. 2014. *MiR171h* restricts root symbioses and shows like its target *NSP2* a complex transcriptional regulation in *Medicago truncatula*. BMC Plant Biology **14**, 199.

Jin Y, Liu H, Luo D, Yu N, Dong W, Wang C, Zhang X, Dai H, Yang J, Wang E. 2016. DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. Nature Communications **7**, 12433.

Karmakar K, Kundu A, Rizvi AZ, Dubois E, Severac D, Czernic P, Cartieaux F, DasGupta M. 2019. Transcriptomic Analysis With the Progress of Symbiosis in 'Crack-Entry' Legume *Arachis hypogaea* Highlights Its Contrast With 'Infection Thread' Adapted Legumes. Molecular Plant Microbe Interactions **32**, 271-285.

Liu W, Kohlen W, Lillo A, *et al.* 2011. Strigolactone biosynthesis in *Medicago truncatula* and rice requires the symbiotic GRAS-type transcription factors NSP1 and NSP2. The Plant Cell **23**, 3853-65.

Maekawa T, Maekawa-Yoshikawa M, Takeda N, Imaizumi-Anraku H, Murooka Y, Hayashi M. 2009. Gibberellin controls the nodulation signaling pathway in *Lotus japonicus*. The Plant Journal **58**, 183-94.

Masson-Boivin C, Giraud E, Perret X, Batut J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends in Microbiology **17**, 458–66.

Murakami Y, Yokoyama H, Fukui R, Kawaguchi M. 2013. Down-regulation of *NSP2* expression in developmentally young regions of *Lotus japonicus* roots in response to rhizobial inoculation. Plant & Cell Physiology **54**, 518-27.

Peng Z., Tan LB, Lopez Y, *et al.* 2018. Morphological and genetic characterization of nonnodulating peanut recombinant inbred lines. Crop Science **58**, 540-550.

Plet J, Wasson A, Ariel F, Le Signor C, Baker D, Mathesius U, Crespi M, Frugier F. 2011. MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in *Medicago truncatula*. Plant Journal **65**, 622-33.

Roy S, Liu W, Nandety RS, Crook A, Mysore KS, Pislariu CI, Frugoli J, Dickstein R, Udvardi MK. 2020. Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. The Plant Cell **32**, 15-41.

Sharma V, Bhattacharyya S, Kumar R, *et al.* 2020. Molecular Basis of Root Nodule Symbiosis between Bradyrhizobium and 'Crack-Entry' Legume Groundnut (*Arachis hypogaea* L.). Plants (Basel) **9**, 276.

Teft M. (2018) Characterising the role of ATSCL26 in root development. PhD thesis, University of Warwick.

Xiao A, Yu H, Fan Y, Kang H, Ren Y, Huang X, Gao X, Wang C, Zhang Z, Zhu H, Cao Y. 2020. Transcriptional regulation of NIN expression by IPN2 is required for root nodule symbiosis in *Lotus japonicus*. The New Phytologist **227**, 513-528.

Yokota K, Soyano T, Kouchi H, Hayashi M. 2010. Function of GRAS proteins in root nodule symbiosis is retained in homologs of a non-legume, rice. Plant & Cell Physiology **51**, 1436-42.

Chapitre 3

3.1 - Contexte de l'étude

La caractérisation des mutants Nod$^-$ d'*A. evenia* a abouti à l'identification de plusieurs gènes symbiotiques initialement décrits chez les légumineuses modèles, *AePOLLUX*, *AeCCaMK*, *AeCYCLOPS*, *AeNSP2* et *AeNIN*, indiquant que la voie de signalisation symbiotique est au moins en partie conservée chez *A. evenia*. Ce crible génétique a également conduit à la découverte d'un gène symbiotique nouveau, *AeCRK*, dont les patterns phylogénétiques et moléculaires le lient à la fois à l'activation sans facteurs Nod et à l'infection intercellulaire. Cette découverte appuie fortement l'hypothèse de l'existence de mécanismes moléculaires propres à la symbiose Nod-indépendante.

Dans l'objectif d'identifier d'autres gènes symbiotiques nouveaux qui puissent rendre compte des spécificités de la symbiose Nod-indépendante, j'ai continué à exploiter la collection de mutants de nodulation d'*A. evenia*. En particulier, je me suis consacré à l'étude de mutants qui possèdent un phénotype symbiotique singulier que l'on a baptisé « OBN » pour One Big Nodule. Les analyses de Targeted Sequence Capture menée sur la collection de mutants de nodulation n'a pas permis de relier ce phénotype symbiotique à des gènes connus chez les légumineuses modèles.

Pour mener à bien la caractérisation phénotypique et génétique de ces mutants OBN, j'ai encadré deux étudiants, Thomas Bénichou (stage de L3 Pro) et Rémi Guyonnet (stage de M2). En collaboration avec Christophe Klopp (GenoToul BioInfo), j'ai également mené une recherche sans *a priori* de gènes candidats par Mapping-by-Sequencing. Nous avons montré que chez les mutants OBN l'infection est bloquée de façon très précoce et que le même gène est altéré. Ce gène code une « Receptor-like Cytoplasmique Kinase » et a été nommé *AeRLCK*. Dans ce chapitre 3, je présente les analyses menées sur les mutants OBN et le gène *AeRLCK* qui indiquent que nous avons identifié un nouvel acteur clé de l'infection intercellulaire dans un contexte de signalisation Nod-indépendante chez *A. evenia*. Pour compléter ce travail, des analyses moléculaires et fonctionnelles sont en cours dans l'équipe (stage de M2 de Natasha Horta). Il fera l'objet à terme d'une publication.

Article 3 : Une "Receptor-like Cytoplasmic Kinase" contrôle l'infection pendant la symbiose Nod-independante chez *Aeschynomene evenia*

3.2 - Article 3 :

Une “Receptor-like Cytoplasmic Kinase” contrôle l’infection pendant la symbiose Nod-independante chez *Aeschynomene evenia*

Johan Quilbé[1], Laurent Brottier[1], Rémi Guyonnet[1], Thomas Benichou[1], Joël Fardoux[1], Christophe Klopp[2], Eric Giraud[1] et Jean-François Arrighi[1]

1-IRD, Laboratoire des Symbioses Tropicales et Méditerranéennes (LSTM), UMR IRD/ SupAgro/INRAE/ UM2 /CIRAD, TA-A82/J – Campus de Baillarguet 34398 Montpellier cedex 5, France

2-Plateforme bioinformatique Genotoul, BioinfoMics, UR875 Biométrie et Intelligence Artificielle, INRAE, Castanet-Tolosan, France.

Titre court: Une “ Receptor-like Cytoplasmic Kinase” requise pour l’infection

Mots clé: Symbiose, RLCK, infection, activation Nod-indépendante, *Aeschynomene*

Introduction

L'établissement de la symbiose Nod-indépendante chez les *Aeschynomene* repose sur un processus d'activation par les *Bradyrhizobium* photosynthétiques ne passant pas par la production de facteurs Nod, qui sont les molécules clé de la symbiose rhizobium-légumineuse, ni par l'intervention d'un T3SS (Type 3 Secretion System) présent chez de nombreux *Bradyrhizobium* (Giraud *et al.*, 2007; Okazaki *et al.*, 2016). Si la nature des signaux produits par les *Bradyrhizobium* photosynthétiques reste énigmatique, il est établi que ces bactéries colonisent les racines de la plante-hôte aux sites d'émergence des racines latérales qui sont caractérisés par la présence d'une couronne de poils axillaires. Au niveau de ces sites d'infection, les bactéries pénètrent de façon intercellulaire entre les cellules des assises épidermiques et corticales, provoquant localement une « cell collapse », avant d'être internalisées dans une ou quelques cellules du cortex. Ces cellules infectées se divisent alors abondamment pour aboutir au développement d'un nodule dans lequel toutes les cellules du tissu central contiennent des bactéries différentiées en bactéroïdes fixateurs d'azote (Arrighi *et al.*, 2012; Bonaldi *et al.*, 2011).

Un crible génétique mené chez *Aeschynomene evenia* a montré que plusieurs gènes de la voie de signalisation Nod décrite chez les légumineuses modèles sont conservés chez *A. evenia*. Ces gènes de signalisation symbiotique -*AePOLLUX*, *AeCCaMK*, *AeCYCLOPS*, *AeNSP2* et *AeNIN*- sont nécessaires à la fois pour l'infection intercellulaire et l'organogénèse nodulaire (Chapitres 1 & 2). Cette approche mutant a également permis d'identifier un gène nouveau, *AeCRK*, dont la fonction biologique est probablement liée dans la signalisation Nod-indépendante ou l'infection intercellulaire (Chapitres 1 et 2). Comprendre les mécanismes et les liens entre la reconnaissance des *Bradyrhizobium* photosynthétiques, la signalisation symbiotique et l'infection intercellulaire représente un enjeu important.

Pour avancer dans notre compréhension des mécanismes de la symbiose Nod-indépendante en mettant à jour d'autres acteurs symbiotiques, nous avons caractérisé une nouvelle catégorie de mutants d'*A. evenia* présentant un phénotype symbiotique duel, les plantes étant essentiellement Nod^- mais étant capable occasionnellement de développer un nodule fixateur. Cette caractérisation a conduit à l'identification d'un gène nouveau, *AeRLCK*, qui code une Receptor-like Cytoplasmic Kinase.

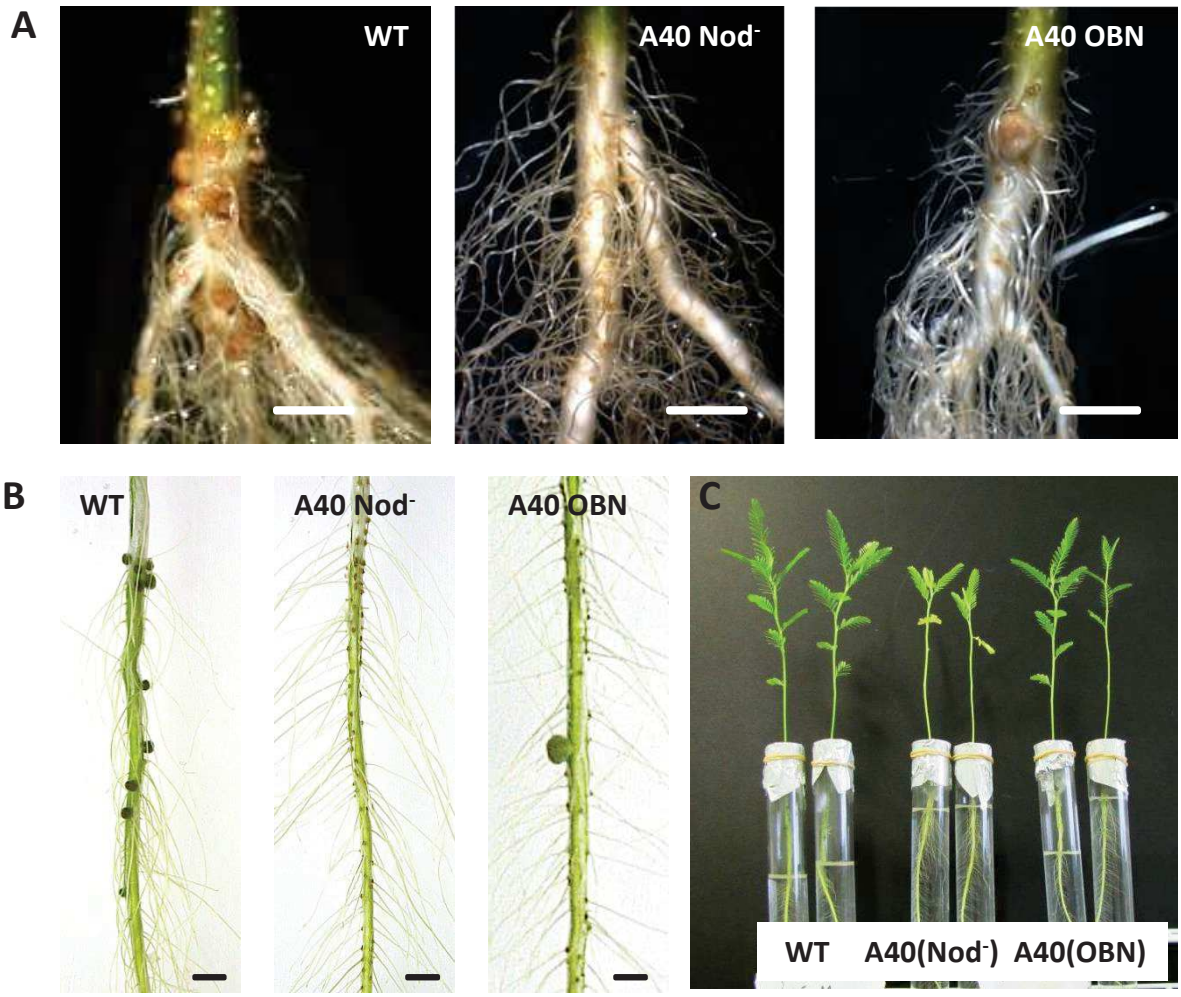


Figure 1 : Phénotype de nodulation des mutants « One Big Nodule ». A. Phénotype racinaire observé 4 semaines après inoculation de la souche de *Bradyrhizobium* ORS278 sur des plantes cultivées dans l'attapulgite pour la lignée sauvage et le mutant A40 montrant un phénotype Nod$^-$ et OBN,. **B.** Phénotype racinaire observé 14 jours après inoculation de la souche de *Bradyrhizobium* ORS278 sur des plantes cultivées en milieu liquide pour lignée sauvage et le mutant A40 montrant un phénotype Nod$^-$ et OBN. **C.** Développement foliaire de la plante sauvage et de plantes de la lignée mutante A40 ayant un phénotype racinaire Nod$^-$ et OBN, respectivement. Les barres d'échelles représentent 500 µm.

Résultats

Identification de mutants avec un phénotype « One Big Nodule »

Pour comprendre les mécanismes moléculaires de l'interaction symbiotique entre *A. evenia* et les *Bradyrhizobium* photosynthétiques, qui ne passe pas par la reconnaissance des facteurs Nod et la formation de cordon d'infection, nous avons précédemment criblé une population mutagénéisée à l'éthyl-methane-sulfonate (EMS). Ce crible a permis d'isoler de nombreux mutants défectueux pour la fixation d'azote avec la souche de *Bradyrhizobium* ORS278. Ces mutants sont caractérisés par l'absence de nodules (mutants Nod^-) ou la présence de nodules blancs ou jaunes (mutants Inf^- ou Fix^-). Lors de ce crible, des mutants ne possédant en général qu'un seul nodule en haut de la racine ont été également isolés. Ce nodule est alors de couleur rose, indiquant la présence de leghémoglobine, et il est également hypertrophié par rapport aux nodules présents chez la lignée sauvage (Fig. 1A). En conséquence, de tels mutants ont été baptisés OBN pour « One Big Nodule ».

Les mutants M_2 sortis du crible ont été menés à graine. La stabilité et l'homogénéité du phénotype symbiotique ont alors été analysées à la fois sur des plantes M_3 cultivées en attapulgite*, id est* dans les mêmes conditions que le crible, et en culture liquide en condition *in vitro*, et inoculées avec la souche de *Bradyrhizobium* ORS278. Cette analyse a révélé que pour tous les mutants OBN, la plupart des plantes M_3 sont Nod^- et une petite partie seulement présente le phénotype « One Big Nodule », indépendamment du type de culture (Fig. 1A, B). De même, pour certains mutants M_2 Nod^-, le phénotype « One Big Nodule » a été retrouvé chez quelques plantes M_3. Dans ces populations de mutants, les individus Nod- sont chlorotiques tandis ceux portant des « One Big Nodule » présentent un feuillage vert comme la plante sauvage, indiquant que ces nodules sont fonctionnels (Fig. 1C). Pour déterminer si ce changement de phénotype est dû à la plante ou à la bactérie, nous avons isolé chez le mutant A40 les bradyrhizobiums contenus dans un « One Big Nodule » et ces bradyrhizobiums ont servi à réinoculer des plantes de la même lignée mutante. La même proportion de plantes Nod^-/OBN a été observée, suggérant que l'apparition du phénotype « One Big Nodule » n'est pas dû à la bactérie (données non montrées).

Au final, une série de 12 mutants OBN -A40, D22, E27, L10, L22, L24, L31, L37, N10, V21, X30 et AE4- a été identifiée à partir de différents lots de crible (Tableau 1). Ces mutants OBN

Tableau 1 : Données phénotypiques, génétiques et moléculaires des mutants OBN d'*A. evenia*

Lignée mutante	Déterminisme génétique		Allèles	Fréquence d'apparition OBN chez les mutants F2	Effet de la mutation		Séquençage
	F2 WT:mutant	(*, P>0,05)			Changement nt	Changement aa	
A40	F2 458:164	3:1*	*rlck-1*	14%	G_{1932} to A	E_{278} to K	MbS
D22	F2 431:162	3:1*	*rlck-2*	6%	G_{516} to A	G_{108} to E	MbS
E27	F2 415:159	3:1*	*rlck-3*	3%	C_{857} to T	L_{170} to F	PCR
L10	F2 466:151	3:1*	*rlck-4*	6,60%	G_{1957} to A	G_{286} to E	MbS
L22	F2 418:157	3:1*	*rlck-5*	4,50%	G_{1066} to A	Splice site	PCR
L24	F2 496:163	3:1*	*rlck-6*	12,30%	G_{525} to A	G_{110} to E	MbS
L31	F2 379:118	3:1*	*rlck-7*	15%	G_{1129} to A	G_{207} to R	PCR
L37	F2 398:138	3:1*	*rlck-8*	20%	G_{1951} to A	S_{284} to N	PCR
N10	F2 137:58	3:1*	*rlck-9*	8%	G_{776} to A	E_{143} to K	PCR
V21	F2 418:153	3:1*	*rlck-10*	6,30%	G_{1920} to A	G_{274} to R	MbS
X30	F2 394:155	3:1*	*rlck-11*	14%	C_{749} to T	S_{131} to STOP	MbS
AE4	F2 245:72	3:1*	*rlck-12*	5,50%	G_{1292} to A	G_{261} to E	PCR

* test statistique de conformité (Khi2) ; MbS : Mapping-by-Sequencing ; PCR : Polymerase Chain Reaction

Tableau 2 : Tests d'allélismes entre les mutants OBN d'*A. evenia*

Croisements (♂ x ♀)	n Plantes F1 (n pods)	Phénotype F1
A40 x L10	5(1)	Nod^-
A40 x V21	6(1)	Nod^-
D22 x L10	3(1)	Nod^-
D22 x M11	4(1)	Nod^-
E27 x V21	18(3)	Nod^-/1 OBN
L10 x V21	6(1)	Nod^-
L22 x V21	18(3)	Nod^-
L24 x L10	10(2)	Nod^-
L24 x V21	3(1)	Nod^-
L31 x X30	7(1)	Nod^-/1 OBN
L37 x V21	23(3)	Nod^-
N10 x V21	12(2)	Nod^-
N10 x X30	6(1)	Nod^-
V21 x E27	6(2)	Nod^-
V21 x L10	9(1)	Nod^-
X30 x L10	1(1)	Nod^-
X30 x L31	15(3)	Nod^-/2 OBN
X30 x N10	6(1)	Nod^-
X30 x V21	7(1)	Nod^-
X30 x AE4	10(2)	Nod^-

Nod^- : incapable de noduler ; OBN : One Big Nodule

Tableau 3 : Test de complémentation fonctionnelle du mutant *rlck*

Lignée	Construction pour la transformation	Plantes transformées	Plantes nodulées	Nodules/plantes nodulées*
X30	pUB-GW-GFP	14	0	0
X30	pUB-GW-GFP-*AeRLCK*	15	9	17,4±11,8
WT	Non-transformé	5	5	19,6±5,9

* Les valeurs correspondent à la moyenne du nombre de nodule par plante nodulées ± déviation standard

présentent un phénotype symbiotique duel qui est déterminé par la plante, avec la présence à la fois du phénotype Nod$^-$ et du phénotype « One Big Nodule ».

Les mutants OBN définissent un seul et même gène symbiotique

Pour clarifier le déterminisme génétique qui sous-tend le phénotype « One Big Nodule », nous avons réalisé le croisement des 12 mutants OBN avec la lignée CIAT22838 ayant servi pour faire la mutagenèse EMS. Nous avons ensuite analysé la ségrégation du phénotype symbiotique dans une population de ~600 individus F_2, quatre semaines après inoculation avec la souche de *Bradyrhizobium* ORS278. Cela nous a permis de montrer pour chaque mutant OBN une ségrégation des phénotypes symbiotiques avec 75% d'individus F_2 WT et 25% des individus F_2 mutants (ratio 3 :1 avec $p<0.05$), indiquant que le caractère de nodulation est monogénique et récessif dans tous les cas (Tableau 1). Parmi les F_2 mutants, la proportion d'individus OBN observée varie de 3 à 20 %, donnant une fréquence moyenne d'apparition du phénotype « One Big Nodule » de 10% (Tableau 1).

Ensuite, pour déterminer combien de gènes sont définis par les mutants OBN, une série de tests d'allélisme a été entreprise en croisant les mutants OBN entre eux. Toutes les plantes F1 issues de ces croisements ont présenté soit un phénotype Nod$^-$ soit un phénotype OBN, quatre semaines après inoculation avec la souche de *Bradyrhizobium* ORS278 (Tableau 2). Ces résultats indiquent que les 12 mutants OBN appartiennent au même groupe de complémentation et donc qu'ils définissent une série de mutations alléliques pour le même gène.

Identification de mutations dans un gène codant une Receptor-like Cytoplasmic Kinase

L'approche de Targeted Sequence Capture qui a été menée précédemment sur l'ensemble de la collection de mutants de nodulation d'*A. evenia* n'a pas permis d'identifier de mutation dans les orthologues de gènes symbiotiques décrits chez les légumineuses modèles. Nous avons donc employé sur 8 mutants OBN (A40, D22, L10, L24, L31, L37, V21 et X30) une approche de séquençage sans *a priori*, dite de Mapping-by-Sequencing, pour trouver le gène responsable du phénotype « One Big Nodule ». Pour chaque mutant, 100 à 130 individus F_2 OBN issus du croisement avec la lignée sauvage ont été poolés, l'ADN extrait du pool et séquencé. La séquence obtenue pour ces pools de F_2 mutants a ensuite été comparée à celle de la lignée sauvage, pour identifier toutes les mutations EMS et les utiliser comme

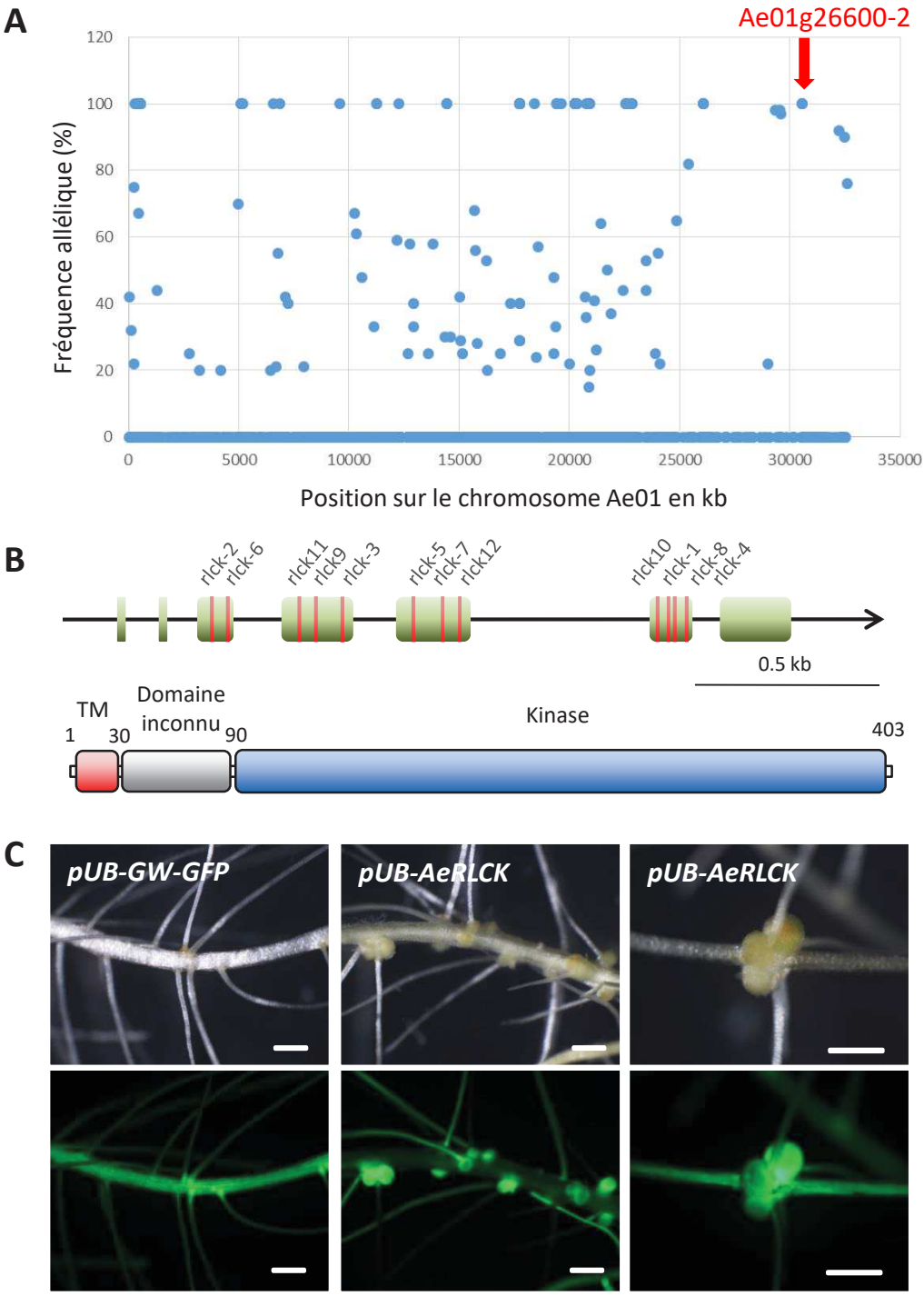


Figure 2 : Identification et validation du gène *AeRLCK* comme étant responsable du phénotype OBN. A. Fréquence des allèles mutants en fonction de leur position sur le chromosome Ae01 d'*A. evenia* ayant permis l'identification d'une mutation dans le gène Ae01g26600-2 chez le mutant D22. **B.** En haut, structure intron/exon du gène Ae01g26600-2, renommé *AeRLCK*, avec la position des mutations identifiées. En bas, structure de la protéine prédite contenant un domaine transmembranaire (TM), un domaine de fonction inconnue et un domaine kinase. **C.** Complémentation fonctionnelle du mutant X30 avec le gène *AeRLCK*. Racines du mutant X30 transformées avec le vecteur vide ou contenant le gène *AeRLCK* sous contrôle d'un promoteur Ubiquitine, et observées à 14 dpi après inoculation avec la souche ORS278. En dessous, les mêmes racines montrant de la fluoresence GFP. Les barres d'échelles représentent 500 µm.

marqueurs de cartographie. L'analyse des fréquences alléliques des mutations EMS, le long des 10 paires de chromosomes d'*A. evenia*, a révélé pour les 8 mutants OBN, la même liaison génétique sur le chromosome Ae01 (Fig. 2A, Figure S1). Dans la région génomique où la fréquence allélique est égale ou proche de 100% d'allèle mutant, pour chacun des 8 mutants OBN, une mutation a été trouvée dans le gène Ae01g26600-2 codant un Receptor-Like Cytoplasmic Kinase et qui a été nommé *AeRLCK* (Fig. 2B, Tableau 1).

Pour valider *AeRLCK* comme étant responsable du phénotype « One Big Nodule », le gène candidat a été amplifié par PCR et séquencé chez les 4 mutants OBN non passés au Mapping-by-Sequencing (E27, N10, L22 et AE4). Ce séquençage ciblé a permis d'identifier quatre allèles mutants additionnels pour le gène Ae01g26600-2 (Tableau 1). Ensuite, une approche de complémentation fonctionnelle a été entreprise. Pour cela, le cDNA du gène *AeRLCK* a été amplifié et cloné derrière le promoteur Ubiquitine dans le vecteur pUB-GW-GFP. La construction a été introduite, en utilisant le système « hairy root » par transformation avec *Agrobacterium rhizogenes* dans les racines du mutant X30, choisi car possédant un codon stop prématuré dans la CDS d'*AeRLCK*. Les racines transformées et repérables par l'expression du gène de la *GFP*, ont été inoculées avec la souche de *Bradyrhizobium* ORS278 et observées à 14 dpi (Fig. 2C). Les racines transformées avec la construction pUB-*AeRLCK* ont développé des nodules de couleur rosée, suggérant qu'ils sont fonctionnels avec en moyenne 17,4 nodules par plante (n=9) (Tableau 3). A l'inverse, aucun nodule ne s'est développé sur les racines du mutant X30 transformées avec le vecteur vide (n=14) (Tableau 3). Ainsi, la présence d'allèles mutants multiples et les données de complémentation fonctionnelle indiquent sans ambiguïté que le gène *AeRLCK* est responsable du phénotype « One Big Nodule ». En conséquence, les mutations alléliques ont été nommées *rlck-1* à *-12* (Fig. 2B, Tableau 1).

AeRLCK semble requis pour l'initiation de l'infection au niveau de la racine

Pour préciser le rôle d'*AeRLCK* au cours de l'interaction symbiotique, nous avons analysé le phénotype symbiotique des mutants OBN en culture *in vitro*, après inoculation avec deux souches de *Bradyrhizobium* photosynthétiques, ORS278 et ORS285, tagguées GUS et GFP, respectivement. Chez la plante sauvage, les bactéries colonisent rapidement les couronnes de poils axillaires situées à la base des racines latérales (1 dpi), pénètrent de façon intercellulaire dans le cortex (2 dpi) et leur endocytose conduit à la formation de bumps (3

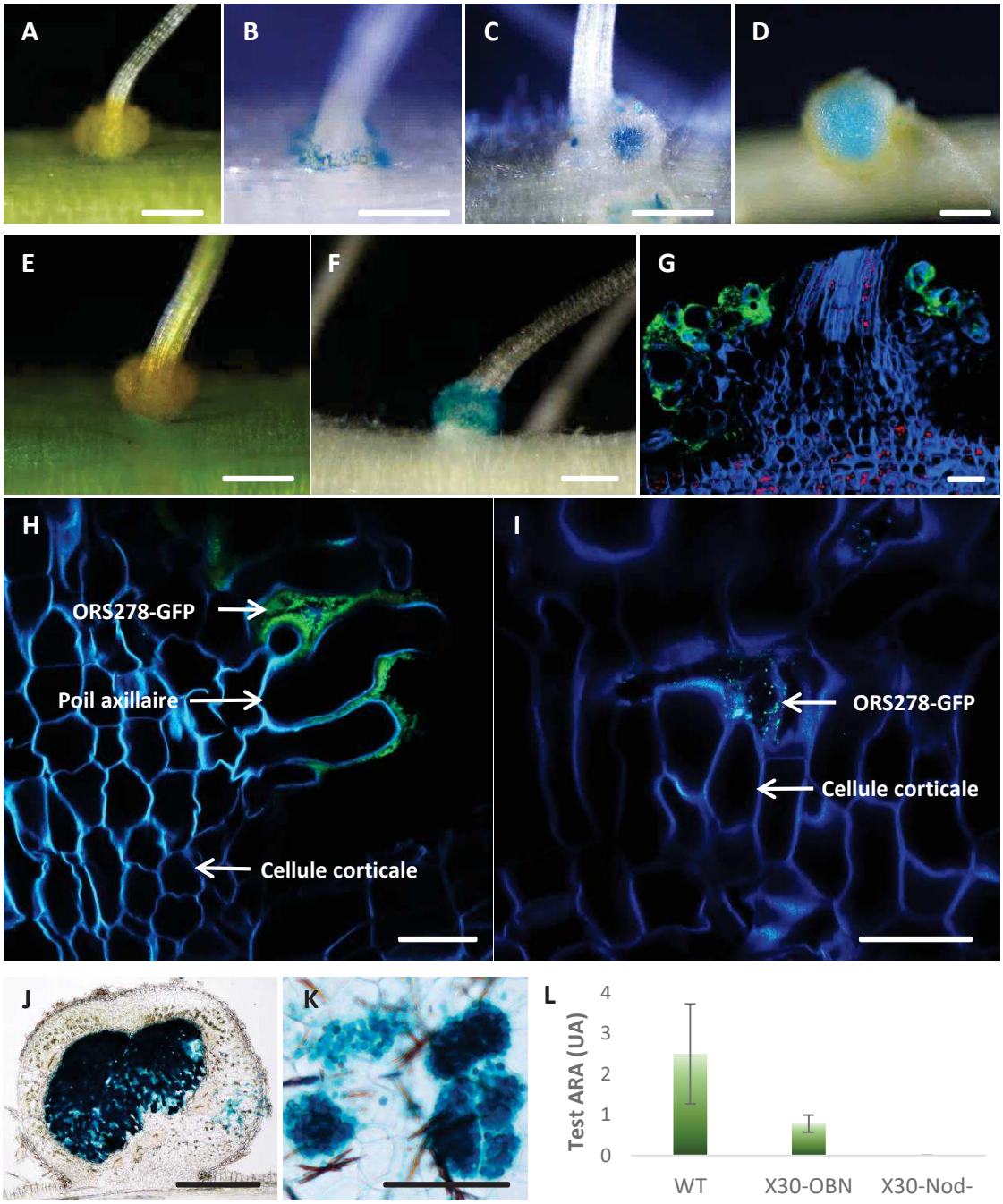


Figure 3 : Infection et organogenèse nodulaire chez les mutants OBN. A-D. Cinétique de nodulation chez la lignée sauvage. (A) Racine non-inoculée. (B,C,D) Racines inoculées avec la souche de *Bradyrhizobium* ORS278-GUS et observées à 1, 3 et 5 dpi, respectivement. **E-I.** Infection chez le mutant X30. (E) Racine non-inoculée. (F) Racine inoculée avec la souche de *Bradyrhizobium* ORS278-GUS et observées à 14 dpi. (G) Coupe de racine du mutant X30 inoculée avec la souche ORS285-GFP à 14 dpi montrant la colonisation de la couronne de poils axillaires par les bactéries de l'infection. (H,E) Zoom sur des coupes de racines montrant des bactéries restant localisées autour des poils axillaires (d) et ayant pénétré dans le cortex (e). **J-L.** Structure et activité des « One Big Nodule ».(J) Coupe longitudinale d'un « One Big Nod leule » chez mutant X30, 14 jours après inoculation de la souche ORS278-GUS. (K) Zoom sur les cellules infectées contenant des bactéroïdes sphériques. (L) Test ARA (unité arbitraire –UA-) réalisé sur les racines de la lignée sauvage et de plantes de lignée mutante X30 avec un « One Big Nodule » ou sans. Les barres d'échelles représentent 500 µm pour les photos en A-GB, 50 µm en I et K.

dpi) aboutissant à la formation de nodules différenciés (5 dpi) (Fig. 3B,C,D). A l'inverse, pour les 12 mutants OBN, la majorité des plantes (~90%) est strictement Nod$^-$ car ne présentant aucun bump ni nodule sur la racine à 14 dpi, seule une petite proportion de plantes (~10%) présente un seul ou un nombre très réduit de nodules le long de la racine (Fig. 1B). Les observations en microscopie sur racine entière révèlent qu'il y a une forte colonisation par les bradyrhizobiums au niveau de la couronne de poils axillaires qui se traduit par une coloration bleu (GUS) ou une fluorescence GFP intense, suivant la souche utilisée (Fig. 3F,G,H). Les analyses en microscopie confocale de sections transversales de racines inoculées de mutants OBN montrent que les bactéries sont présentes en surface ou entre les poils axillaires (Fig. 3H). Parmi le grand nombre de coupes de racines analysées, un seul évènement de pénétration intercellulaire dans le cortex racinaire a pu être observé (Fig. 3I). Cet évènement précède potentiellement la formation d'un nodule. Les plantes présentant des « One Big Nodule » sont plus développées et leurs feuilles sont plus vertes comparées à celles qui sont strictement Nod$^-$ (Fig. 1B). L'analyse de coupes transversales de ces nodules révèle une vascularisation périphérique et un tissu central infecté contenant des bactéroïdes différenciés, comme observé dans les nodules du WT (Fig. 3J,K). De plus, les tests ARA (Acetylene Reduction Assay) montrent une activité pour la nitrogénase, renforçant l'idée que les nodules des mutants OBN sont capables de fixer l'azote au profit de la plante (Fig. 3L).
Ces observations, prises ensemble, suggèrent que le gène *AeRLCK* n'est pas essentiel ou pas requis pour l'organogenèse nodulaire. La rareté des nodules développés chez les mutants OBN est probablement due à un blocage précoce de l'infection lors de l'invasion de la racine par les *Bradyrhizobium* photosynthétiques. La taille importante des nodules des mutants OBN correspond certainement à un mécanisme de compensation à leur très faible nombre.

AeRLCK n'est pas essentiel à la nodulation caulinaire

A. evenia est capable de produire des nodules à la fois au niveau de la racine et de la tige. Les mutants Nod$^-$ précédemment caractérisés comme étant mutés dans différents gènes de la voie de signalisation symbiotique chez *A. evenia*, *AePOLLUX*, *AeCCaMK*, *AeCYCLOPS*, *AeNSP2* et *AeNIN*, ainsi que dans le nouveau gène symbiotique *AeCRK*, conservent leur phénotype de nodulation entre la racine et la tige (Quilbé *et al.*, accepté). Pour déterminer si les mutants OBN sont également altérés pour la nodulation caulinaire, des plantes âgées de 5 semaines ont été inoculées avec la souche de *Bradyrhizobium* ORS278 au niveau de leur

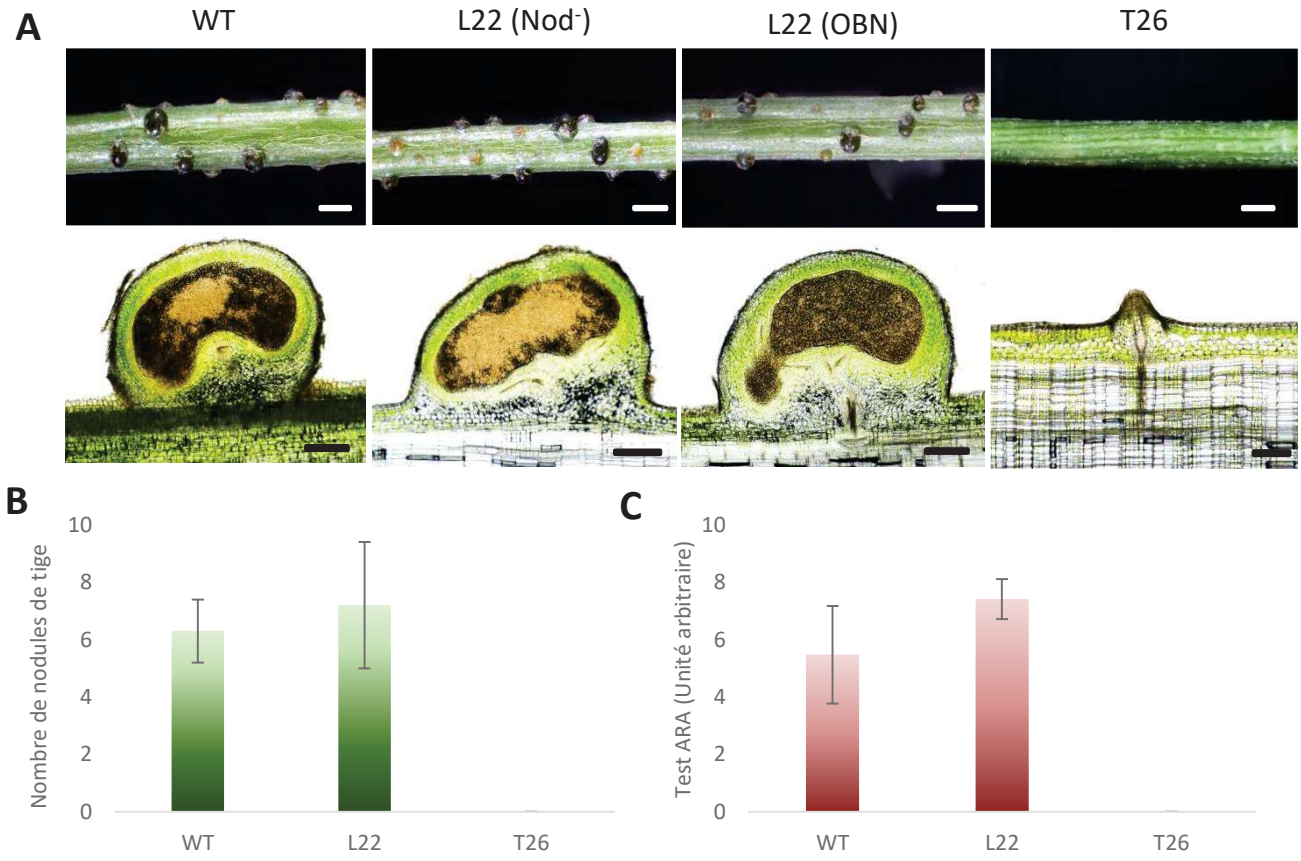


Figure 4 : Nodulation caulinaire chez les mutants OBN. Comparaison du phénotype de nodulation de tige chez la lignée sauvage, le mutant L22 (*rlck-5*) le mutant T26 (*ccamk-3*). **A.** Aspect et structure des nodules caulinaire des trois lignées. En haut, vue d'ensemble des tiges ; en bas, coupes de nodules caulinaires pour le WT et le mutant L22, et coupe de tige pour le mutant T26 au niveau d l'émergence d'une racine adventive. **B.** Nombre de nodules de tiges développés chez les trois lignées. **C.** Activité ARA en unité arbitraire obtenues pour les nodules de tiges des trois lignées. Les barres d'échelles représentent 500 µm pour les photos de tige entières et 200 µm pour les coupes.

tige et observées à 14 dpi. Dans ces conditions, des nodules caulinaires parfaitement développés ont été retrouvés chez les 12 mutants OBN. Ces nodules sont comparables en taille et en structure avec ceux observés sur les tiges des plantes WT, et le développement de nodules caulinaires ne dépend de la formation ou pas de « One Big Nodule » au niveau de la racine (Fig. 4A). A l'inverse aucun nodule ne se développe au niveau des sites d'émergence des racines adventives chez un mutant *ccamk* (Fig. 4A). Le comptage de nodules de tige chez les mutants OBN et les tests ARA révèlent un niveau de nodulation et d'activité de la nitrogénase similaire à la lignée sauvage (Fig. 4B, C). Cette absence de phénotype caulinaire suggère que le gène *AeRLCK* n'est pas impliqué ou pas essentiel pour la nodulation caulinaire. Ce type de nodulation est beaucoup moins bien caractérisé au niveau cytologique que la nodulation racinaire, il est possible que les modalités d'infection soient différentes entre la racine et la tige.

Structure d'AeRLCK, un nouvel acteur de signalisation symbiotique

La séquence d'*AeRLCK* contient 8 exons représentant 1209 nucléotides et codant une protéine prédite de 403 acides aminés qui appartient au groupe des Receptor-like Cytoplasmic Kinases (RLCK) (Fig. 2B). Ces RLCKs sont homologues aux Receptor-like Kinases (RLK) mais se différencient par le fait qu'en général seule la partie intracellulaire est présente. De nombreux RLCKs sont connus pour être des médiateurs clé de signalisation précoce en interagissant avec des RLKs qui perçoivent des ligands extracellulaires et en assurant la transduction du signal dans le cytoplasme (Liang and Zhou, 2018; Lin *et al.*, 2013). La séquence déduite d'AeRLCK révèle qu'elle est constituée d'un domaine transmembranaire (TM) (résidus 5 à 30), suivi d'un domaine inconnu (résidus 30 à 90) et finalement d'un domaine kinase (résidus 90 à 403) (Fig. 2B). Les RLCKs peuvent être ancrés à la membrane plasmique via une ancre GPI. Dans le cas d'AeRLCK, la prédiction d'un domaine transmembranaire est indicatif d'un ancrage direct à la membrane plasmique. Le domaine de fonction inconnue (position 30-97) présente des similitudes avec les domaines « KELCH » d'après le prédicteur SMART, suggérant qu'il pourrait correspondre à un domaine d'interaction protéine-protéine. Par ailleurs, la comparaison du domaine kinase avec celui de RLKs révèle que les différents sous-domaines et motifs clé sont conservés chez AeRLCK suggérant qu'il est fonctionnel.

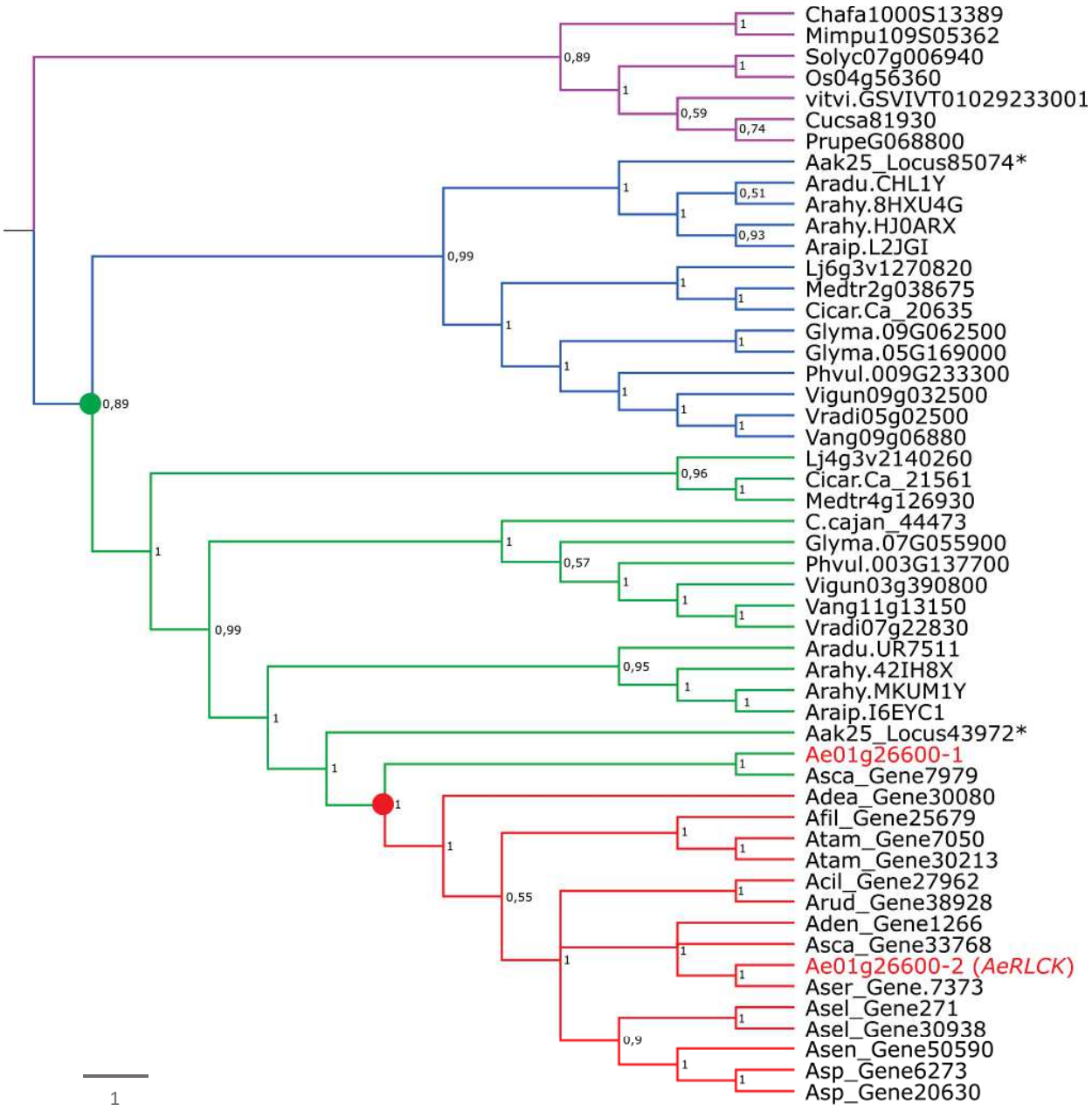


Figure 5 : Phylogénie d'*AeRLCK*. L'arbre phylogénétique a été généré par inférence Bayésienne. Les branches violettes représentent des espèces chez lesquelles un seul homologue d'*AeRLCK* est présent. Le point vert indique la WGD datée de 58 MA des Papilionacées résultant en la présence d'une paire de gènes paralogues (branches bleues et vertes). Les astérisques montrent la présence de ces deux paralogues chez *A. afraspera*. Le point rouge marque la duplication de gènes trouvée chez les *Aeschynomene* Nod-indépendantes. Les deux copies présentes en tandem chez *A. evenia* sont indiquées en rouge. Les branches rouges correspondent aux orthologues du gène *AeRLCK* présent uniquement chez les *Aeschynomene* Nod-indépendantes. Les valeurs représentent la probabilité postérieure (PP) des nœuds. La barre d'échelle représente le nombre de substitutions par site.

L'ensemble des 12 mutations identifiées chez les mutants OBN tombent dans le domaine kinase d'AeRLCK (Tableau 1). Certaines mutations affectent un site d'épissage (L22) ou génèrent un codon stop prémature (X30), résultant en des formes tronquées d'AeRLCK. D'autres mutations correspondent à des substitutions et révèlent ainsi des acides aminés clé pour la structure ou la fonction d'AeRLCK. Parmi elles, on retrouve six substitutions de Glycine (G) en d'autres acides aminés, notamment au niveau de la « Glycin-rich loop » qui est essentielle pour l'activité kinase.

***AeRLCK* est issu d'une duplication en tandem chez les *Aeschynomene* Nod-indépendantes**

Pour étudier l'évolution d'*AeRLCK*, une analyse phylogénétique a été menée en recherchant les gènes homologues chez plusieurs espèces de plantes, que ce soit des légumineuses ou non, dont le génome est disponible. Nous avons également utilisé les données transcriptomiques disponibles pour plusieurs espèces d'*Aeschynomene*, parmi lesquelles 11 sont Nod-indépendantes (*A. ciliata*, *A. deamii*, *A. denticulata*, *A. filosa*, *A. rudis*, *A. scabra*, *A. selloi*, *A. sensitiva*, A. serrulata, *A.* sp. 328, *A. tambacoundensis*, et une Nod-dépendendante, *A. afraspera*. L'arbre phylogénétique obtenu par inférence Bayésienne montre la présence d'un seul gène homologue chez les non-légumineuses comme le riz (mais pas *Arabidopsis*) et des légumineuses non Papilionaceae (Fig. 5). Chez les Papilionaceae, deux clades sont présents et correspondent probablement à des groupes de gènes paralogues issus de la WGD (Whole Genome Duplication) datée de 58 MA (Fig. 5). On note qu'*A. afraspera* possède les deux gènes paralogues à l'instar des autres légumineuses papilionacées tandis qu'un des deux gènes est manquant chez toutes les *Aeschynomene* Nod-indépendantes. En revanche, chez toutes les *Aeschynomene* Nod-indépendantes, deux sous-clades sont visibles (Fig. 5). Un sous-clade contient le gène Ae01g26600-1 et l'autre le gène *AeRLCK* (Ae01g26600-2). *AeRLCK* apparait ainsi comme étant la deuxième copie d'un tandem de gènes (fusionnés en un seul gène, Ae01g26600, dans la version annotée actuelle du génome d'*A. evenia*), alors qu'une seule copie est présente chez les autres Papilionacées (Fig. 5). Les analyses de microsynténie montrent que tous ces gènes sont présents dans des régions synténiques chez les légumineuses papilionacées (Fig. S2). Le très fort taux d'identité de séquence, proche de 92,37%, pour les deux copies présentes chez *A. evenia* suggère qu'elles résultent d'un évènement de duplication en tandem qui a eu lieu de façon relativement récente (Fig. S3).

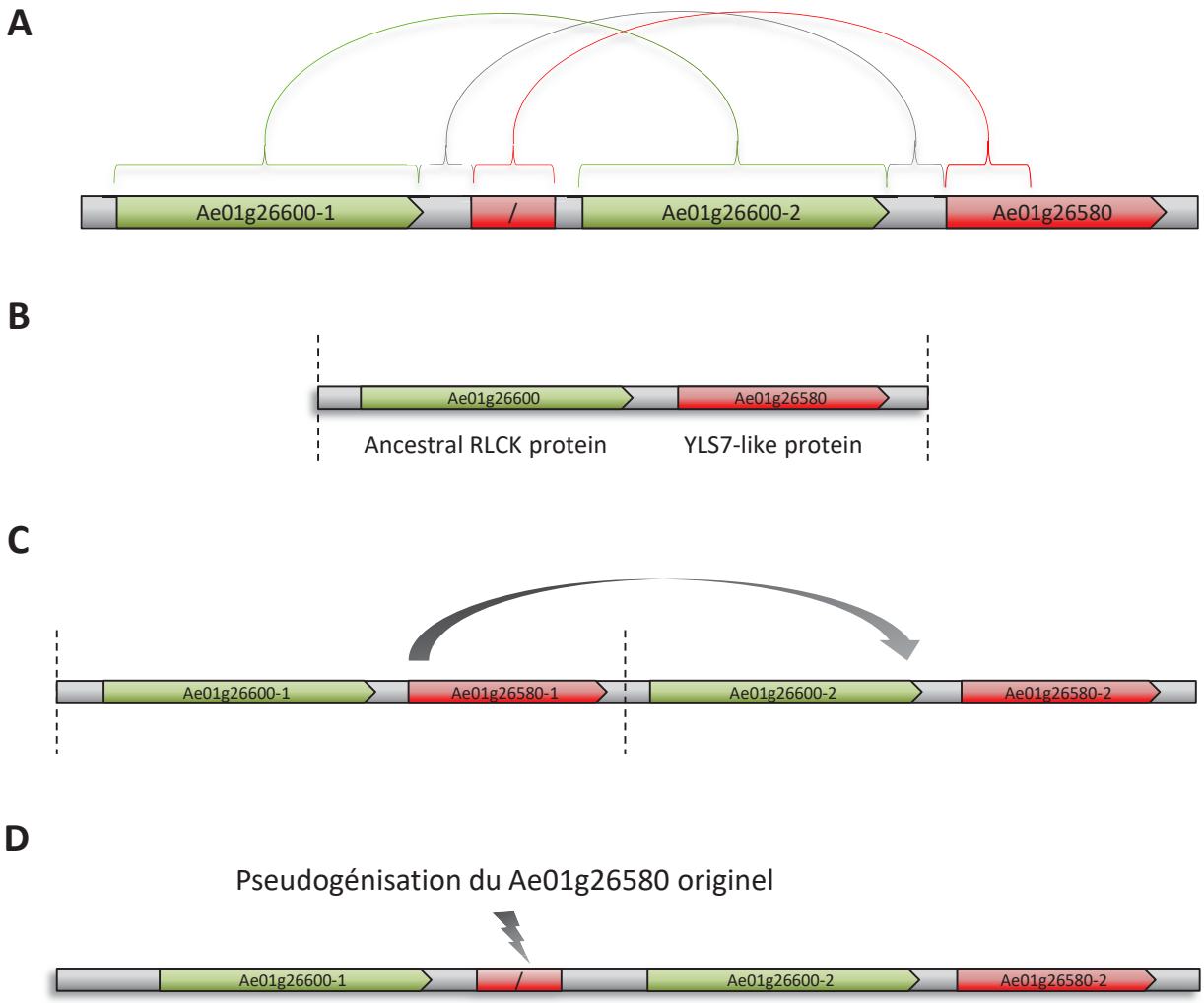


Figure 6 : Structure du locus Ae01g26600 et reconstitution de l'évènement de duplication en tandem à l'origine d'AeRLCK. **A**. Vue schématique de 3 gros blocs de séquence conservés au niveau du locus Ae01g26600. **B.** Structure génomique initiale supposée avec une version ancestrale d'Ae01g26600 et le gène adjacent Ae01g26580 sur le chromosome Ae01. **C.** Structure génomique supposée à l'état transitoire, avec une duplication d'un bloc contenant une partie du promoteur d'Ae01g26600 plus sa séquence codante ainsi que le promoteur du gène Ae01g26580 plus sa séquence codante. D) Structure génomique actuelle avec deux copies du gène Ae01g26600 (-1 et -2). La séquence en aval de la copie -2 a subi des remaniement ou a accumulé des mutations ayant conduit a une forte modification du promoteur de la copie -2 par rapport a la copie 1.

Il est à noter toutefois pour les *Aeschynomene* Nod-indépendantes, que la copie 2 a été retrouvée dans la grande majorité des cas et plus rarement la copie 1 (Fig. 5). Ce résultat est peut-être dû au fait que la copie 1 est plus faiblement exprimée que la copie 2 et n'est pas toujours détectée ou que les deux copies ont été fusionnées lors de l'assemblage *de novo* des transcriptomes car très similaires (Fig. S3). Pour déterminer combien de copies d'*AeRLCK* sont présentes chez les différentes *Aeschynomene*, des tests d'amplification PCR ont été réalisés en ciblant des régions conservées du gène. Ce travail a révélé qu'un seul gène est présent chez les *Aeschynomene* Nod-dépendantes et deux copies sont identifiables chez les *Aeschynomene* Nod-indépendantes (Tableau S1). Ces données prises ensembles indiquent que l'évènement de duplication en tandem d'*AeRLCK* a très probablement eu lieu à la base du clade des *Aeschynomene* Nod-indépendantes.

***AeRLCK* représente une innovation génomique**

Pour comprendre comment cette duplication en tandem s'est passée et si elle a eu un impact sur la structure génique d'*AeRLCK*, nous avons analysé plus en détail le locus Ae01g26600 contenant les deux copies. Tout d'abord, l'alignement de séquences protéiques déduites des différents gènes présents chez les *Aeschynomene*, montre que la copie 1 des *Aeschynomene* Nod-indépendantes est très similaire en taille à la copie unique présente chez *A. afraspera* (Fig. S3). En revanche la copie 2 diffère en étant plus courte de 8 aa dans la partie N-ter et 16 aa dans la partie C-ter (Fig. S3). L'analyse de la séquence génomique du locus Ae01g26600, contenant les deux gènes en tandem, met en évidence des modifications complexes de la région 5' d'*AeRLCK* qui aboutissent à des changements de structure intron/exon et du codon start d'*AeRLCK* (copie 2) par rapport à la copie 1 (Fig. S4). En revanche, la partie 3' d'*AeRLCK* se révèle être très conservée avec celle de la copie 1, mais un SNP dans cette partie crée un codon stop prématuré (Fig. S4). Il faut aussi noter que la séquence intergénique entre les deux copies correspond aux 1000 premiers nucléotides du gène Ae01g26580, situé à proximité du locus Ae01g26600 (Fig. 6). L'ensemble de ces observations nous permet de formuler une hypothèse sur l'origine génomique d'*AeRLCK* : (1) au départ le gène Ae01g26600 (RLCK) et Ae01g26580 (YLS7-like protein) sont positionnés côte-à-côte. (2) Un évènement de duplication en tandem a lieu créant une nouvelle copie d'Ae01g26600 et emportant tout (ou partie) du gène Ae01g26580. (3) Suite à

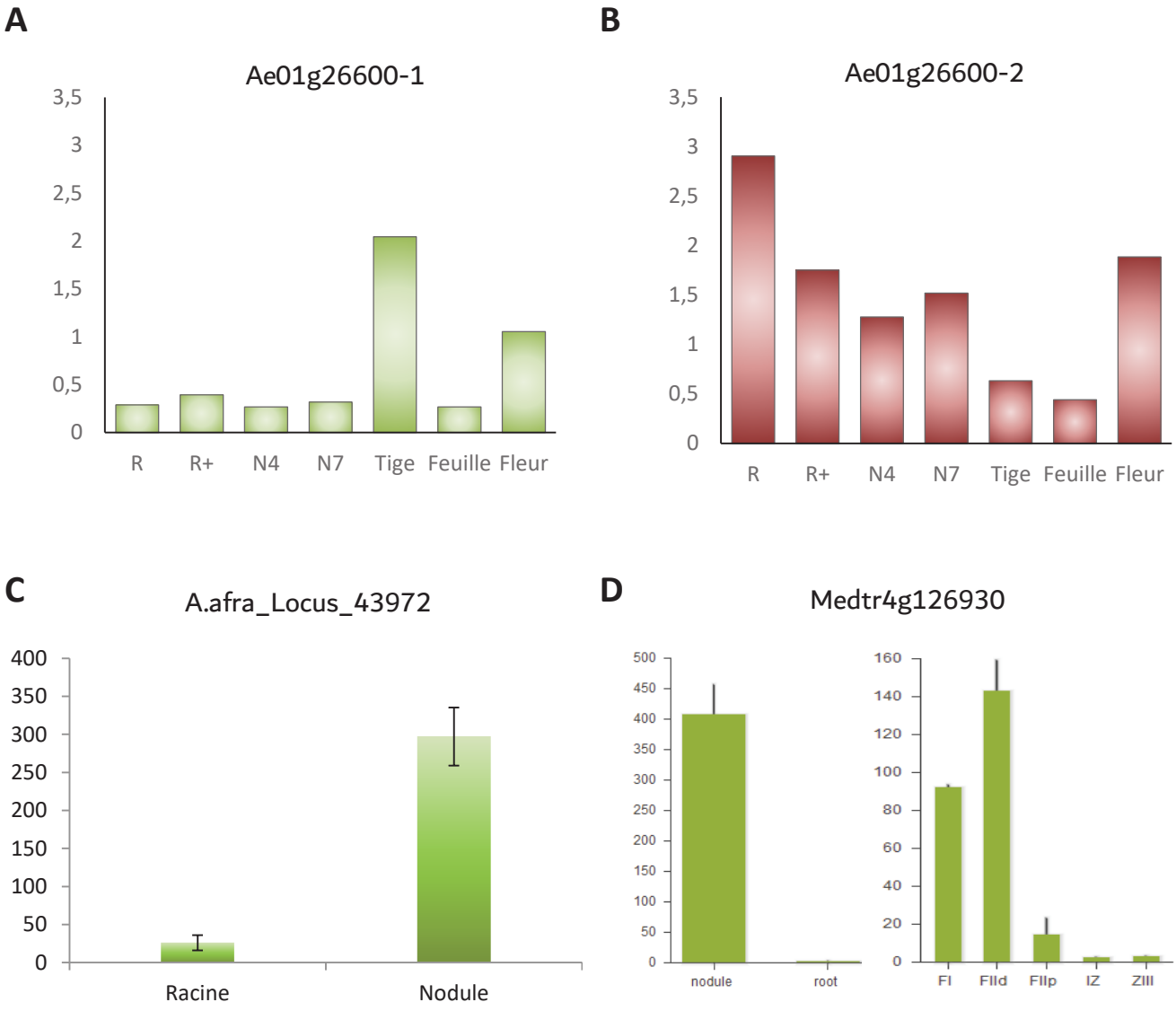


Figure 7 : Profils d'expression d'*AeRLCK* et des gènes les plus proches chez *A. afraspera* et *M. truncatula*. **A et B.** Profil d'expression d'Ae01g26600-1 et B) d'*AeRLCK* (Ae01g26600-2) obtenus par Q-RT-PCR dans plusieurs conditions (R=racine seule; R+=racine avec azote; N4=nodule de 4 jours; N7=nodule de 7 jours; T=tige; F=Feuille et FL=Fleurs) standardisée sur l'expression du gène de ménage *AeEF1α*. **C.** Expression du gène Aak25_Locus_43972 chez *Aeschynomene afraspera* déduites des données RNAseq générée par D. Gully. **D.** Profil d'expression du gène Medtr4g126930 chez *M. truncatula* (d'après Roux *et al.*, 2014).

réarrangements, la région promotrice et la structure d'*AeRLCK* est modifiée par rapport à la version originale (Fig. 6).

Pour déterminer l'impact de ces réarrangements génomiques sur l'expression d'*AeRLCK*, nous avons comparé son profil d'expression avec celui des gènes présents en simple copie chez d'autres légumineuses. Les analyses de Q-PCR menées chez *A. evenia* avec des paires d'amorces spécifiques de chaque copie, indiquent que le niveau d'expression de la copie 1 est extrêmement faible par rapport à celle d'*AeRLCK* (Fig. 7A, B). En revanche, *AeRLCK* est exprimé dans la racine et le nodule mais apparemment pas dans les parties aériennes de la plante (Fig. 7B). Son niveau d'expression est important dans la racine sans azote et dans le nodule en développement (4dpi) alors qu'il est moins exprimé dans la racine avec azote et dans le nodule mâture (7dpi) (Fig. 7B). L'analyse de l'expression du gène le plus proche chez *Aeschynomene afraspera* (données RNAseq de l'équipe) qui est dépendant des facteurs Nod montre au contraire que une expression plus forte dans le nodule que dans la racine (Fig. 7C). De même, le Gene Atlas disponible pour *M. truncatula*, qui utilise un processus symbiotique dépendant des facteurs Nod et des cordons d'infection, montre que ce gène est induit au cours de la nodulation et qu'il est très fortement exprimé dans la zone d'infection du nodule (Fig. 7D). Bien que les données d'expression obtenues chez *A. evenia* nécessitent d'être confortées, elles suggèrent très fortement que la duplication en tandem a conduit à la création d'*AeRLCK*, qui est original à la fois de par sa structure génique et son profil d'expression.

Discussion

Grâce à l'approche mutant couplée à la méthode de séquençage de Mapping-by-Sequencing, nous avons pu identifier un gène nouveau codant une « Receptor-like cytoplasmic Kinase » et qui il n'est retrouvé que chez les *Aeschynomene* utilisant un processus d'activation Nod-indépendante. Ce gène, *AeRLCK*, semble intervenir très tôt dans l'établissement de la symbiose Nod-indépendante.

Les RLCKs appartiennent à la même grande famille que les « Receptor-like Kinases » dont ils ne se distinguent que par l'absence du domaine extracellulaire et souvent également du domaine transmembranaire. On compte 149 RLCKs chez *Arabidopsis thaliana* et 349 RLCKs chez *Oryza sativa* qui sont dispersés dans la phylogénie des RLKs (Shiu *et al.*, 2004; Vij *et al.*,

2008). Les RLKs (Receptor-Like Kinase) et les RLPs (Receptor-Like Proteins) peuvent former des complexes récepteurs responsables de la perception de signaux extracellulaires puis de la transduction du signal aboutissant à l'expression de programmes génétiques impliqués dans la régulation du développement de la plante ou de réponses aux stress biotiques et abiotiques. Les RLKs et RLPs possèdent un domaine extracellulaire et un domaine transmembranaire (ou parfois une ancre GPI pour les RLPs). Les RLKs possèdent également un domaine kinase intracellulaire qui assure la transduction du signal par phosphorylation (He *et al.*, 2018). Parmi les intermédiaires de signalisation qui interagissent avec les RLKs, les RLCKs jouent souvent un rôle clé. Les RLCKs peuvent s'associer avec différents RLKs ou complexes récepteurs pour activer une variété de réponses signalétiques (Liang and Zhou, 2018). Cette transduction en aval des RLCKs peut passer par une cascade MAP-Kinase comme c'est le cas chez le riz avec *OsRLCK185* qui est activé par OsCERK1, le récepteur de chitine (He *et al.*, 2018; Lin *et al.*, 2013). Elle peut aussi impliquer la phosphorylation d'autres protéines membranaires comme les RBOH qui interviennent dans la production de ROS, comme montré pour AtBIK1 qui est activé par FLS2, le récepteur de flagellines (Zipfel and Oldroyd, 2017). Récemment, une RLCK nommée LjNICK4, a été identifiée chez la légumineuse modèle *L. japonicus* (Wong *et al.*, 2019). Ce RLCK interagit et est capable de phosphoryler le domaine kinase de LjNFR5, un des récepteurs des facteurs Nod. La perception des facteurs Nod par le complexe de récepteurs LjNFR5-LjNFR1 entraine une relocalisation de LjNICK4 au niveau du noyau puis promeut le programme d'organogenèse nodulaire. La caractérisation fonctionnelle d'AeRLCK permettra aussi de mettre à jour des mécanismes moléculaires symbiotiques nouveaux chez *A. evenia*. Pour initier cette caractérisation fonctionnelle d'AeRLCK et faciliter la recherche future d'interactants, il est prévu de vérifier la prédiction qu'elle est ancrée à la membrane plasmique via un domaine transmembranaire (TM) en analysant la localisation subcellulaire de fusions AeRLCK-GFP et AeRLCKΔTM-GFP (Stage de Natasha Horta). Une collaboration avec Julie Cullimore (LIPM, INRAE, Toulouse) a également été établie pour tester l'activité kinase d'AeRLCK.

Les résultats obtenus chez *A. evenia* ont déjà livré des indications précieuses sur le rôle d'*AeRLCK* pendant le processus de nodulation Nod-indépendant. En effet, chez les mutants OBN, les bactéries ne sont plus capables de pénétrer de manière intercellulaire dans les tissus de la racine, sauf dans de rares cas. Ce blocage très précoce de l'infection indique qu'*AeRLCK* est impliqué dans les toutes premières étapes de l'interaction symbiotique. Mais

la capacité des mutants à produire de façon ponctuelle un gros nodule qui est fonctionnel montre que la plante n'est pas affectée dans l'organogénèse nodulaire. Les deux hypothèses les plus vraisemblables pour expliquer ce phénotype symbiotique seraient soit qu'*AeRLCK* contrôle l'initiation de d'infection soit qu'il intervient dans l'immunité de la plante. Néanmoins, les mutants OBN ne montrent aucun signe de réaction de défense ni au niveau de la couronne de poils axillaires colonisée par les bradyrhizobiums ni quand un « One Big Nodule » se forme. Il est donc tentant de supposer qu'*AeRLCK* est plutôt impliqué dans le contrôle du processus d'infection. De plus, le phénotype « One Big Nodule » est similaire à celui décrit chez les mutants *epr3* de *L. japonicus* (Kawaharada *et al.*, 2015). Ces mutants sont altérés dans *LjEPR3*, un gène de LysM-RLK impliqué dans la perception des EPS (exopolysaccharides) bactériens. Lorsque ces mutants de plante sont inoculés avec la souche mutante de *Mesorhizobium loti R7AexoU*, qui produit des EPS de structure anormale, on observe la formation de seulement quelques gros nodules qui sont soit blancs soit rosés, mais qui sont dans tous les cas infectés par la bactérie. Il a été suggéré que LjEPR3, via la surveillance de la structure de l'EPS rhizobiens, contôle l'infection lors de son initiation au niveau de l'épiderme et qu'il permet aussi d'amplifier la signalisation symbiotique à un niveau nécessaire pour soutenir l'infection (Kawaharada *et al.*, 2015). AeRLCK pourrait être impliqué chez *A. evenia* dans une fonction symbiotique similaire à celle de LjEPR3 chez *L. japonicus*. Pour renforcer l'hypothèse d'un rôle d'*AeRLCK* dans le contrôle de l'infection, le profil d'expression spatio-temporel d'*AeRLCK* pourra être suivi en utilisant une fusion promoteur-GUS ou -GFP. De même, pour déterminer si des réponses symbiotiques sont encore induites ou non chez les mutants OBN et pour tenter de positionner *AeRLCK* par rapport aux gènes de la voie de signalisation symbiotique, des analyses RNAseq pourraient être réalisées sur les mutants correspondants et comparés. Le fait que les gènes de la voie de signalisation symbiotique sont aussi requis pour la nodulation de tige mais pas *AeRLCK* pose question sur l'existence d'une redondance fonctionnelle possible ou de mécanismes d'infection différents entre la racine et la tige.

L'étude de l'histoire évolutive du gène *AeRLCK* a révélé qu'il est présent uniquement dans le clade des *Aeschynomene* Nod-indépendantes et qu'il est issu d'une duplication en tandem. Cette duplication s'est accompagnée d'un changement de structure du gène mais également de sa région promotrice, laissant penser qu'*AeRLCK* représente une innovation génomique et qu'il pourrait avoir subi une néo-fonctionnalisation. Le fait que la version ancestrale,

présente chez les légumineuses qui utilisent un processus Nod-dépendant, ait une expression induite au cours de la nodulation avec un lien clair avec l'infection chez *Medicago* truncatula suggère que la fonction originelle de ce gène est liée à l'infection. Mais le changement de profil d'expression observé chez *A. evenia*, avec *AeRLCK* étant exprimé de façon constitutive dans la racine, indique que ce gène intervient de façon plus précoce dans la symbiose Nod-indépendante. Il sera donc important de déterminer si cette diversification fonctionnelle potentielle correspond juste à un changement de profil d'expression spatio-temporel ou également à une activité biologique nouvelle. Pour cela, des analyses similaires à celles menées chez *M. truncatula* avec le couple de gènes paralogues ERN1/ERN2 (Cerri *et al.*, 2012) pourront être menées chez *A. evenia* : 1) en menant des tests de complémentation fonctionnelle de mutants OBN par surexpression de la copie 1 et de la version simple copie présente chez les *Aeschynomene* Nod-dépendantes, 2) en utilisant pour ces tests de complémentation interspécifique soit le promoteur endogène d'*AeRLCK* soit celui de la version simple copie des *Aeschynomene* Nod-dépendantes. Si cette néo-fonctionnalisation a eu lieu pour AeRLCK, est-ce également le cas du ou des RLKs avec qui il doit interagir ? En l'absence d'implication de LysM-RLK dans le processus Nod-indépendant (pas de mutant sortis du crible pour *AeNFP*, *AeLYK3* n'est pas exprimé et *EPR3* est absent du génome d'*A. evenia*), la nature de ces récepteurs symbiotiques reste à découvrir. Une telle question représente un enjeu important des recherches à développer pour caractériser les mécanismes moléculaires de la symbiose Nod-indépendante et comprendre comment elle a évolué.

Matériels & méthodes

Matériel végétal et bactérien

Les mutants de nodulation ont été obtenus par mutagenèse EMS chez la lignée CIAT22838 d'*A. evenia* et criblés pour leur phénotype symbiotique racinaire quatre semaines après inoculation avec la souche de *Bradyrhizobium* photosynthétique ORS278. Le phénotype symbiotique a été vérifié à la génération suivante dans les mêmes conditions que pour le crible, comme détaillé dans Quilbé *et al.* (accepté). Les tests de nodulation sont réalisés en

utilisant les souches sauvages de *Bradyrhizobium* sp.ORS278 et ORS285, les souches tagguées ORS278-GUS et ORS285-GFP (Bonaldi *et al.*, 2011; Giraud *et al.*, 2007).

Conditions de culture

Les graines sont scarifiées à l'acide sulfurique 98% et la germination induite par un traitement à l'éthrel comme décrit dans Arrighi *et al.*, 2012. Pour la culture de plantes en serre, les plantules sont repiquées dans des bacs remplis d'attapulgite (Dry Oil, US Sorbix Special) et placées en serre tropicale (28°C avec une hygrométrie de 70%). Pour la culture *in vitro*, les plantules sont cultivées dans du milieu BNM liquide et mis à pousser en chambre 24°C et un cycle lumineux 16h/8h comme expliqué dans Arrighi *et al.* (2012)

Tests de nodulation

Pour les tests de nodulation de racines, les plantes cultivées in vitro et âgées de 6 jours sont inoculées avec 1 mL de suspension bactérienne préalablement lavée et dont la DO600nm mesurée à l'aide d'un spectrophotomètre (Varian, UV-visible spectrophotometer Cary 50 scan) est ajustée à 1. Pour les tests de nodulation de tige, les bacs en serre contenant les plantes âgées de 5 semaines sont placés dans un contenant en plastique de 10 L remplis d'eau et inoculés avec 800 mL de *Bradyrhizobium* (culture de 5 jours), de sorte à submerger les tiges d'*A. evenia* pendant 24h.

Pour les tests de ré-inoculation, les nodules développés sur les mutants d'*A. evenia* après inoculation avec la souche ORS278 sont excisés avec un scalpel stérile. Ils sont désinfectés 30s dans de l'éthanol absolu puis rincé 5 fois à l'eau stérile. Après désinfection, les nodules sont broyés dans un tube Eppendorf 1,5 mL à l'aide d'un pilon dans 500 µL d'eau stérile. La solution du broyat obtenue est alors déposée en strie à l'aide d'une oese sur boite de milieu YM solide pour vérifier la pureté de l'échantillon. Parallèlement 400 µL du broyat est ajouté à 10 mL de milieu BNM liquide. Cet inoculum est directement injecté dans les tubes contenant les mutants de plantes.

Test ARA

L'activité de fixation des nodules est déterminée par des tests ARA (Acetylene Reduction Assay) en suivant le protocole décrit par Bonaldi *et al.* (2010). Pour cela, 6 plantes entières (pour les nodules de racine) ou 6 tronçons de tige de 10 cm (pour les nodules caulinaires) de

chaque lignée mutante ont été placé dans un flacon en verre de 125 ml fermé par un bouchon hermétique. Ensuite, 15 ml d'air sont retirés du flacon puis remplacés par le même volume d'acétylène. Après, trois heures d'incubation à 28°C en chambre de culture in vitro, la quantité d'éthylène produite par chaque échantillon est dosée par chromatographie en phase gazeuse (CPG).

Observations macroscopiques et microscopiques

Les nodules sont observés entre 14 et 21 dpi en utilisant le macroscope (Nikon AZ100, Champigny-sur-Marne, France) qui permet une vue générale ainsi qu'une vue précise des zones d'intérêt et le microscope Nikon DS-Ri2 (Tokyo, Japan) qui possède une qualité optique supérieure pour des grossissements importants. Les photos sont prises en fluorescence et en lumière blanche avec le logiciel Nikon Advanced. Les coupes de nodules ont été réalisées à l'aide du vibratome Leica® VT1000S (Mussloch, Germany) réglé sur une épaisseur de coupe de 30 µm. Les coupes sont récupérées puis montées sur lame de microscope afin d'être observées. Pour la coloration GUS, les échantillons de racines sont couverts d'une solution tampon de X-GLUC (adaptée de Jefferson *et al.*, 1989) : 0,075% de X-GLUC; 0,1M de Na_2HPO_4/KH_2PO_4 (pH7); 20mM NaEDTA ; 0,02% Triton X- 100, 1mM de K-ferricyanide et 1mM de K-ferrocyanide. Les échantillons sont ensuite placés dans une pompe à vide avec une pression de -0,9 bar pendant 15 min et sont finalement incuber à 37°C pendant une nuit. Les échantillons sont ensuite traités à l'éthanol 70% à 65°C pendant 2 h pour décolorer la racine et pouvoir mieux observer les éventuelles colorations.

Analyses génétiques

Les croisements entre mutants et avec la lignée sauvage, ainsi que les analyses de déterminisme génétique et les tests d'allélisme ont été réalisés comme décrits dans Quilbé *et al.* (accepté).

Mapping-by-Sequencing

Pour cette approche, 100 à 130 individus F2 ayant un phénotype mutant et issus d'un croisement avec la lignée sauvage ont été poolés et leur ADN extrait au CTAB comme détaillé dans Quilbé *et al.* (accepté). Le séquençage Illumina (HiSeq 2x150 pb) avec une profondeur moyenne de 40X des échantillons d'ADN a été réalisé par les plateformes de

séquençage Get-Plage (INRAE, Toulouse) et CEES (Université d'Oslo, Norvège). Les analyses bio-informatiques ont été réalisées par la plateforme GenoToul (INRAE, Toulouse) en utilisant le programme SnpEff comme expliqué dans Quilbé *et al.* (accepté). En résumé, les séquencées Illumina sont alignées sur le génome de référence et les variations de séquence (SNP et INDEL) correspondant potentiellement aux mutations EMS sont identifiées et leur fréquence allélique calculée. Pour chaque mutant, une liaison génétique se traduisant par un déplacement des valeurs de fréquence allélique vers 100% est recherchée dans le génome. Dans cet intervalle, les mutations sont analysées pour leur position (gène/intergène, exon/intron) et leur impact (mutation silencieuse, changement d'acide aminé, création d'un codon stop ou épissage alternatif).

Clonage du gène *AeRLCK*

La CDS du gène *AeRLCK* (1212 pb) a été amplifié la paire d'amorces listée dans le Tableau S2, des cDNA provenant d'un pool de racines et nodules comme matrice et la Taq Phusion. Des dATPs sont ajoutés aux extrémités -5' du produit PCR pour insertion dans le premier vecteur (pCR8/GW/TOPO). Le vecteur pCR8-*AeRLCK* est introduit dans la souche d'*E. coli* compétente TOP10 par HeatShock qui est cultivée sur milieu LB Sp^{50}. Le plasmide est extrait grâce au kit Miniprep (Qiagen) et la séquence de l'insert vérifié par séquençage Sanger. La recombinaison avec le plasmide de destination pUB-GW-GFP est réalisée avec le kit LR clonase (Invitrogen) puis le produit transformé par HeatShock dans la souche d'*E. coli* compétente TOP10 cultivée sur milieu LB Kn^{50}. A nouveau, le plasmide pUB-GW-*AeRLCK* est extrait grâce au kit Miniprep (Qiagen) et la construction vérifiée par Sanger. La souche Arqua1 d'*Agrobacterium rhizogenes* est transformée par électroporation (12.5kV/cm) puis cultivées sur milieu AG Kn^{150}.

Transformation par *Agrobacterium rhizogenes*

Le mutant X30 a été transformé selon la technique décrite dans Arrighi *et al.* (2012) et Quilbé *et al.* (accepté). Pour cela, les radicules d'*A. evenia* obtenues après 24h de germination, sont coupées au scalpel et la zone de coupure est mise en contact avec *A. rhizogenes*. Les plants inoculés sont ensuite déposés sur milieu solide 0,5X MS (Murashige and Skoog basal salt mixture) et placées à 20°C pendant 7 jours. Les plantes sont alors rincées dans une solution d'eau stérile additionnée de céfotaxime (300mg/l) afin d'éliminer

les Agrobacterium. Les plantes sont repiquées deux fois à 7 jours d'intervalle sur milieu solide 0,5X MS contenant de la céfotaxime (300mg/l). Les plantes sont cultivées en chambre de culture avec une photopériode 16h/8h, une température de 22°C/18°C. Trois semaines après la transformation des racines, les plantes chimères obtenues sont analysées sous macroscope à fluorescence (Nikon AZ100). Les racines co-transformées sont sélectionnées grâce à l'observation en fluorescence de la GFP et les racines non transformées enlevées. Les plantes chimériques sont transférées en tubes Falcon entourés d'aluminium contenant 40 ml de milieu BNM liquide complémenté avec 0.5 mM de KNO3.

Analyses phylogénétiques

Toutes les analyses sont réalisées à partir de la séquence protéique des différents gènes étudiés. Pour la recherche des homologues chez les autres légumineuses, les groupes de gènes orthologues, précédémment générés avec OrthoFinder (Quilbé *et al.*, révision), et ceux disponibles sur le « Legume Information System » (https://www.legumeinfo.org/) ont été utilisés. Les données transcriptomiques d'*A. afraspera* ont été fournies par Djamel Gully (LSTM, non publié). Les séquences sont alignées grâce à la fonctionnalité MUSCLE du logiciel MEGA X. Une première analyse par Neighbor-Joining et MaximumLikelihood est réalisée pour avoir une idée de la topologie de l'arbre et pour vérifier si tous les homologues identifiés appartiennent bien au clade étudié. Après cette analyse préliminaire, les arbres phylogénétiques sont générés par inférence bayésienne en utilisant le logiciel MrBayes avec les paramèters suivants : lset rates=gamma, prset statefreqpr=fixed(empirical), mcmc nchains=2 ngen=1.000.000 (ou 2.000.000 quand besoin d'une meilleure résolution). Puis les arbres sont visualisés et modifiés avec le logiciel FigTree.

Analyses d'expression

La quantification de l'expression d'*AeRLCK* et de la copie 1 chez *A. evenia* a été réalisée par Q-RT-PCR comme décrit en détail dans le chapitre 2, en utilisant les échantillons d'ARN qui ont servi pour générer le Gene Atlas d'A. evenia (Quilbé *et al.*, accepté). Les amorces ont été désignées pour amplifier de façon spécifique au niveau du locus Ae01g26600 la copie 1 et la copie 2 (Tableau S2). Les données d'expression chez *Aeschynomene afraspera* ont été fournies par Djamel Gully (LSTM, non publié). Les données d'expression chez *Arachis duranensis* ont été obtenus avec l'outil « Legume Mine » du « Legume Information System »

(https://www.legumeinfo.org/). Celles de *Medicago truncatula* proviennent du site « *Medicago truncatula* A17 r5.0 genome" (https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR).

Tests d'amplification et analyses génomiques comparatives

Pour détecter la présence des différentes versions d'*AeRLCK*, les couples d'amorces suivant ont été utilisés Tableau ST2 et ont servi pour des tests d'amplication PCR avec des échantillons d'ADN de différentes espèces d'*Aeschynomene* listées dans Brottier *et al.*, 2018. Les produits PCR ont été séquencées par Sanger et alignées sur les séquences de gènes d'*A. evenia* et d'*A. afraspera* pour comparaison.

Les analyses génomiques ont été réalisées en utilisant le GBrowse disponible pour le génome d'*A. evenia* sur le site AeschynomeneBase (http://aeschynomenebase.fr), le Genome Context Viewer du Legume Information System (https://www.legumeinfo.org/) et le programme d'alignement MULTALIN (http://multalin.toulouse.inra.fr/multalin/).

Références bibliographiques

Arrighi, J.-F., F. Cartieaux, S. C. Brown, M. Rodier-Goud, M. Boursot, J. Fardoux, D. Patrel, *et al.* 2012. *Aeschynomene evenia* , a Model Plant for Studying the Molecular Genetics of the Nod-Independent Rhizobium-Legume Symbiosis. *Molecular Plant-Microbe Interactions®* 25: 851–861.

Bonaldi, K., D. Gargani, Y. Prin, J. Fardoux, D. Gully, N. Nouwen, S. Goormachtig, and E. Giraud. 2011. Nodulation of *Aeschynomene afraspera* and *A. indica* by Photosynthetic *Bradyrhizobium* Sp. Strain ORS285: The Nod-Dependent Versus the Nod-Independent Symbiotic Interaction. *Molecular Plant-Microbe Interactions®* 24: 1359–1371.

Bonaldi, K., B. Gourion, J. Fardoux, L. Hannibal, F. Cartieaux, M. Boursot, D. Vallenet, *et al.* 2010. Large-Scale Transposon Mutagenesis of Photosynthetic *Bradyrhizobium* Sp. Strain ORS278 Reveals New Genetic Loci Putatively Important for Nod-Independent Symbiosis with *Aeschynomene indica*. *Molecular Plant-Microbe Interactions®* 23: 760–770.

Cerri, M. R., L. Frances, T. Laloum, M.-C. Auriac, A. Niebel, G. E. D. Oldroyd, D. G. Barker, *et al.* 2012. *Medicago truncatula* ERN Transcription Factors: Regulatory Interplay with NSP1/NSP2 GRAS Factors and Expression Dynamics throughout Rhizobial Infection. *Plant Physiology* 160: 2155–2172.

Giraud, E., L. Moulin, D. Vallenet, V. Barbe, E. Cytryn, J.-C. Avarre, M. Jaubert, *et al.* 2007. Legumes Symbioses: Absence of Nod Genes in Photosynthetic Bradyrhizobia. *Science* 316: 1307–1312.

He, Y., J. Zhou, L. Shan, and X. Meng. 2018. Plant cell surface receptor-mediated signaling – a common theme amid diversity. *Journal of Cell Science* 131: jcs209353.

Kawaharada, Y., S. Kelly, M. W. Nielsen, C. T. Hjuler, K. Gysel, A. Muszyński, R. W. Carlson, *et al.* 2015. Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* 523: 308–312.

Liang, X., and J.-M. Zhou. 2018. Receptor-Like Cytoplasmic Kinases: Central Players in Plant Receptor Kinase–Mediated Signaling. *Annual Review of Plant Biology* 69: 267–299.

Lin, W., X. Ma, L. Shan, and P. He. 2013. Big Roles of Small Kinases: The Complex Functions of Receptor-Like Cytoplasmic Kinases in Plant Immunity and Development: Multiple Functions of RLCKs. *Journal of Integrative Plant Biology* 55: 1188–1197.

Okazaki, S., P. Tittabutr, J. Thouin, J. Fardoux, C. Chaintreuil, D. Gully, J.-F. Arrighi, *et al.* 2016. Rhizobium–legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *The ISME Journal* 10: 64–74.

Shiu, S.-H., W. M. Karlowski, R. Pan, Y.-H. Tzeng, K. F. X. Mayer, and W.-H. Li. 2004. Comparative Analysis of the Receptor-Like Kinase Family in Arabidopsis and Rice. *The Plant Cell* 16: 1220–1234.

Vij, S., J. Giri, P. K. Dansana, S. Kapoor, and A. K. Tyagi. 2008. The Receptor-Like Cytoplasmic Kinase (OsRLCK) Gene Family in Rice: Organization, Phylogenetic Relationship, and Expression during Development and Stress. *Molecular Plant* 1: 732–750.

Wong, J. E. M. M., M. Nadzieja, L. H. Madsen, C. A. Bücherl, S. Dam, N. N. Sandal, D. Couto, *et al.* 2019. A *Lotus japonicus* cytoplasmic kinase connects Nod factor perception by the NFR5 LysM receptor to nodulation. *Proceedings of the National Academy of Sciences* 116: 14339–14348.

Zipfel, C., and G. E. D. Oldroyd. 2017. Plant signalling in symbiosis and immunity. *Nature* 543: 328–336.

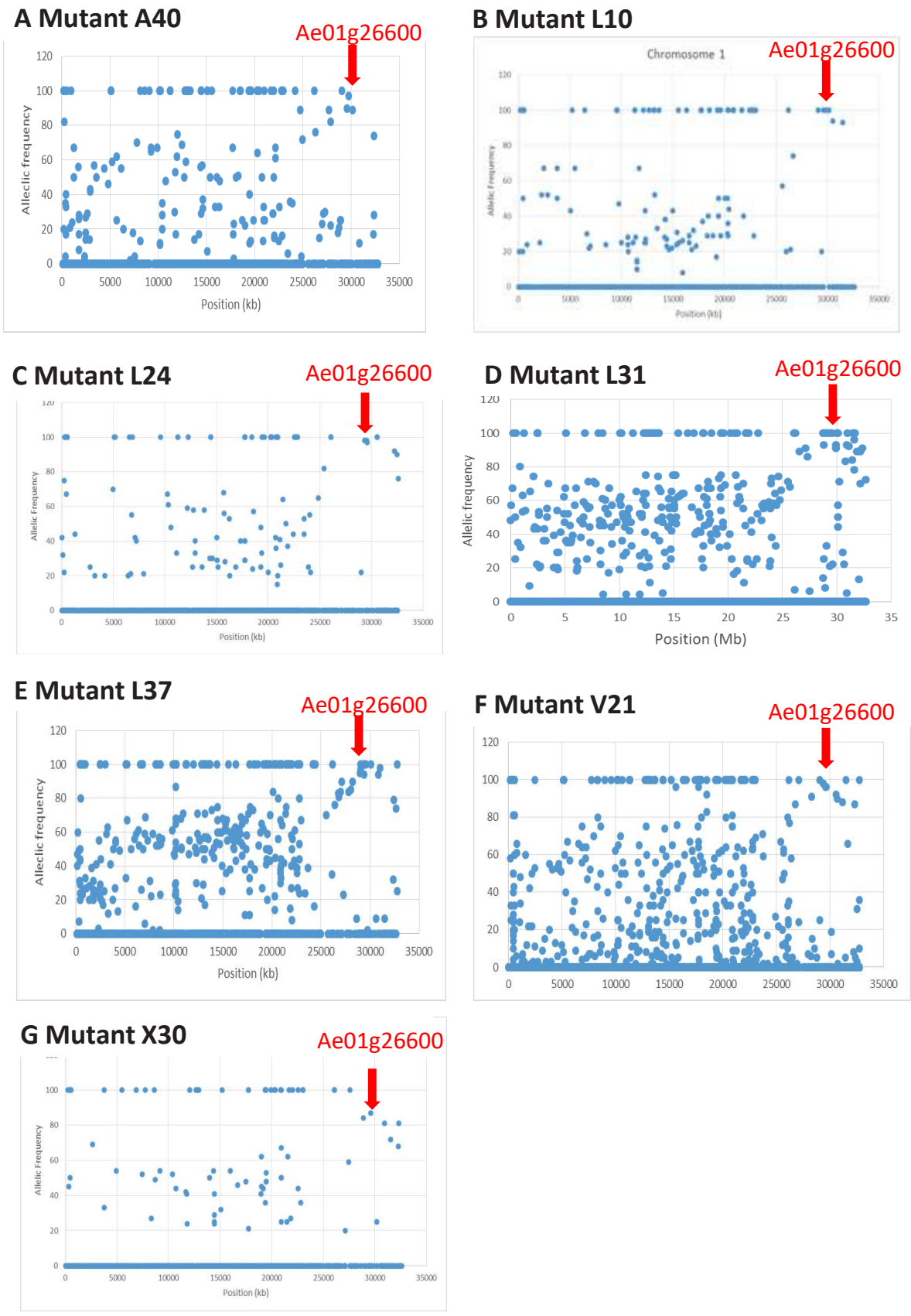


Figure S1: Identification de mutations dans le gène Ae01g26600 chez les mutants OBN par l'approche de Mapping-by-Sequencing. Graphes représentant la fréquence des allèles mutants (%) en fonction de leur position sur le chromosome Ae01 pour les mutants D22 (A), L10 (B), L24 (C), L31 (D), L37 (E), V21 (F) et X30 (G).

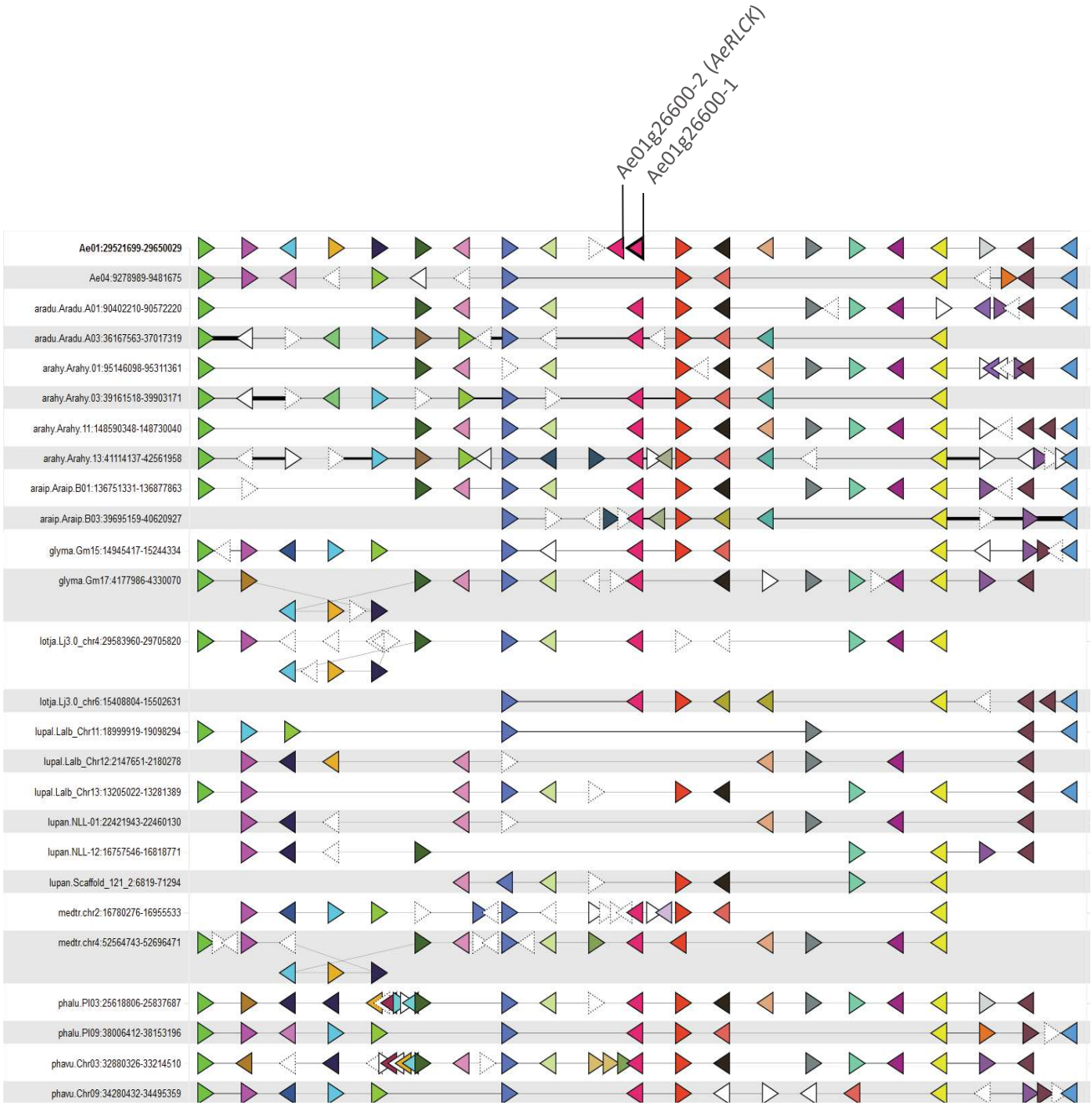


Figure S2 : Analyse de la synténie d'*AeRLCK* chez différentes légumineuses papilionacées.
Représentation schématique de la conservation de la colinéarité des gènes au niveau du locus contenant le gène *AeRLCK* (Ae01g26600-1). Les gènes homologues sont indiqués par l'utilisation de la même couleur. Les données ont été obtenues dans le Legume Information System (https://legumeinfo.org/) en utilisant le Genome Context Viewer.

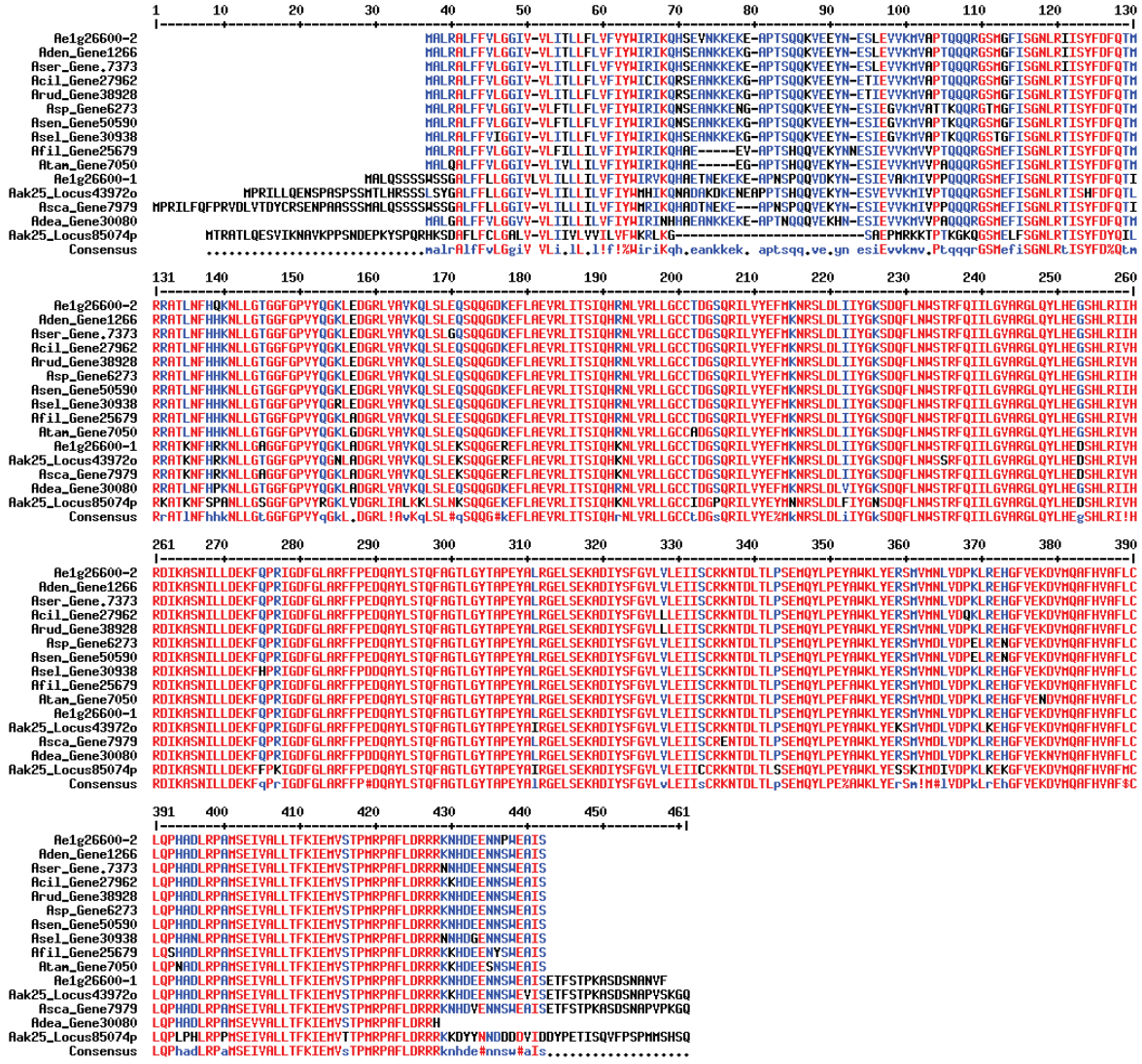


Figure S3: Alignement de séquence des différentes versions d'*AeRLCK.* Les séquences protéiques proviennent des données génomiques et transcriptomiques disponibles pour les *Aeschynomene* et elles ont été alignées dans Multalin.

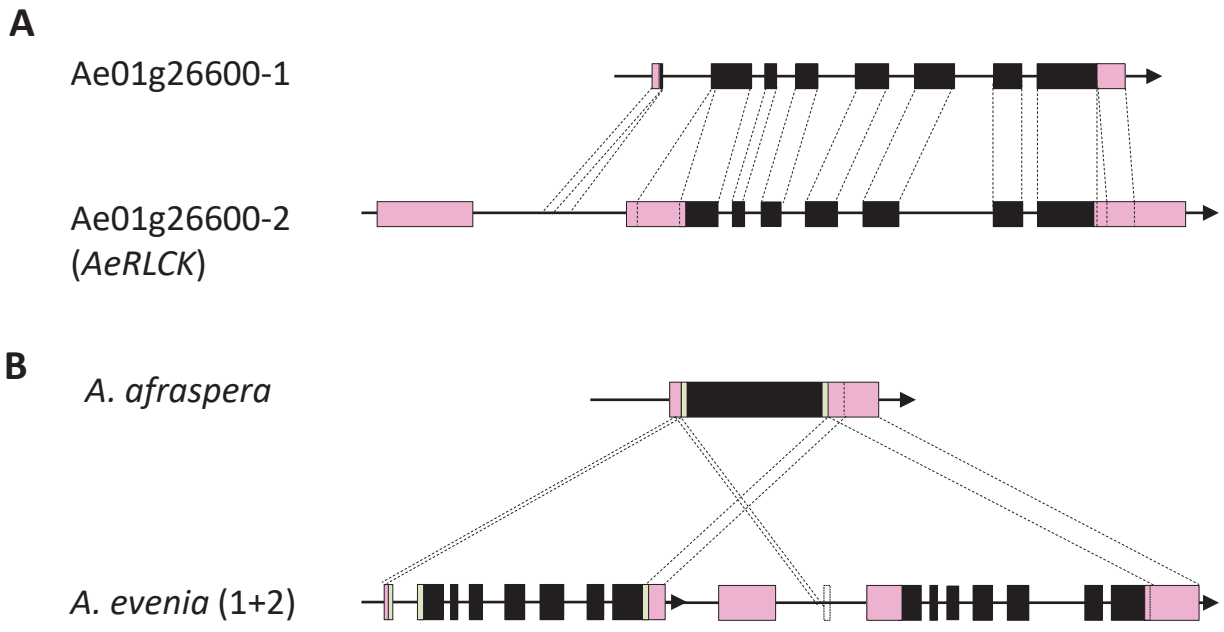


Figure S4 : Structure du gène *AeRLCK*, comparaison de la copie 1 d'*A. evenia* et la copie simple d'*A. afraspera*. A. Comparaison de la structure génomique des deux copies présentes au niveau du locus Ae01g26600 chez *A. evenia*. **B.** Comparaison du transcrit identifié chez *A. afraspera* (Aak25_Locus_43972) avec les deux copies présentes dans le génome d'*A. evenia*. Les exons sont représentés en noir et les régions UTRs en orange. Les blocs de conservations (séquence) entre les différents gènes sont représentés par des lignes pointillées.

Tableau S1 : Analyse des versions d'*AeRLCK* présentes chez les *Aeschynomene*

Lignée	Espèce	Donnée	*RLCK-1*	*RLCK-2*
Arachis	*Arachis ipaiensis*	Génomique	1	
	Arachis duranensis	Génomique	1	
Aeschynomene Nod-dépendantes	*A. americana*	PCR	1	
	A. fluminensis	PCR	1	
	A. parviflora	PCR	1	
	A. montevidensis	PCR	1	
	A. patula	PCR	1	
	S. semperflorens	PCR	1	
	A. afraspera (8x)	RNAseq	1	
Aeschynomene Nod-indépendantes 1	*A. tambacoundensis*	RNAseq + PCR	1	1
	A. filosa	RNAseq + PCR	2	1
Aeschynomene Nod-indépendantes 2	*A. deamii*	RNAseq + PCR	1	1
Aeschynomene Nod-indépendantes 3	*A. sp*	RNAseq + PCR	1	1
	A. selloi	RNAseq + PCR	1	1
	A. sensitiva	RNAseq + PCR	1	1
Aeschynomene Nod-indépendantes 4	*A. evenia*	Génomique	1	1
	A. serrulata	RNAseq + PCR	1	1
	A. ciliata	RNAseq + PCR	1	1
	A. denticulata	RNAseq + PCR	1	1
	A. scabra	RNAseq	1	1
	A. rudis	RNAseq + PCR	1	1

En noir les données obtenues in silico et en vert les données par PCR + séquençage
Les sous-clades 1 à 4 des *Aeschynomene* Nod-indépendantes sont indiqués.

Tableau S2 : Liste des amorces utilisées pour le séquençage, la complémentation et l'analyse q-RT-PCR

Nom d'amorce	Fonction	Séquence 5'-3'
RLCK1-F1	**Séquençage**	TGGAGCATTGTTCTTTSTCCT
RLCK1-F2		ATTGCTGATYCTTGTGTTCATA
RLCK1-F3		AGCATTGTTCTTTGTCCTTGG
RLCK1-F4		TGGAGTTCATTAGTGGGAACCT
RLCK1-F5		TGATTTCCAGACAATGAGAAGG
RLCK1-F6		GAATTTCTGGCAGAGGTGAGA
RLCK1-F9		TGGTGGATTTGGACCTGTCT
RLCK1-F10		TGGAGTTCATTAGTGGGAACCT
RLCK1-R1		GTTTTTGCGCCTCCTATCAA
RLCK1-R2		GGCCTCATTGGTGTYGAAA
RLCK1-R3		GCCTCCTATCAAGGAATGCT
RLCK1-R5		CAAATCTGCATGAGGCTGAA
AeRLCK2-F1c	**Complémentation fonctionnelle**	ATGGCATTACGAGCATTGTTCT
AeRLCK2-R1c		CTAGGATATGGCTTCCCAAG
AeRLCK1-F1-qp	**q-RT-PCR**	TTCGGCCATGGCAATATTGA
AeRLCK1-R1-qp		CAATAGGATGAGCACTAGTA
AeRLCK2-F1qp		ATTACGAGCATTGTTCTT
AeRLCK2-R1-qp		ATCTTCACCACCTCTAAT

3.3 - ADDENDUM : Identification de trois gènes requis pour l'infection intercellulaire chez *Aeschynomene evenia*

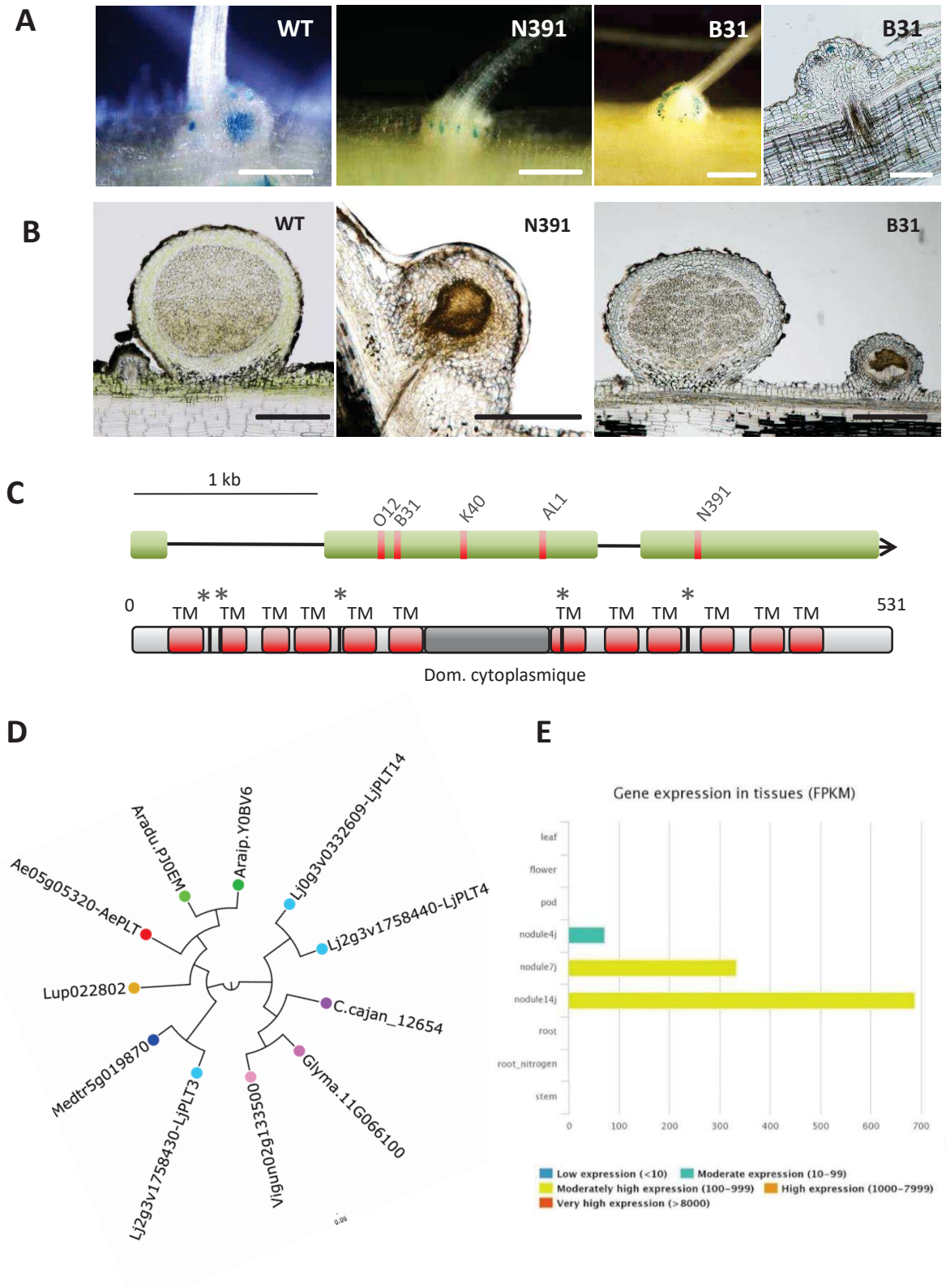


Figure 1 : *AePLT*, phénotype des mutants et caractéristiques du gène. A. Phénotype symbiotique observé 3 jours après inoculation chez le sauvage et 14 jours après inoculation de la souche de *Bradyrhizobium* ORS278-GUS chez les mutants N391 et B31 **B.** Phénotype symbiotique observé 14 jours après inoculation de la souche de *Bradyrhizobium* ORS285 chez la lignée WT et les mutants N391 et B31 WT. **C.** Structure intron/exon du gène *AePLT* (Ae05g05320) et de la protéine traduite avec en rouge les hélices transmembranaires et en gris le domaine cytoplasmique. Les étoiles correspondent aux mutations identifiées. **D.** Arbre phylogénétique généré par inférence bayésienne des homologues d'*AePLT*. **E.** Profil d'expression du gène *AePLT* provenant du Gene Atlas d'*A. evenia*. Les barres d'échelles représentent 500 µm sauf pour la photo de coupe zoomée où elle représente 250 µm.

Au cours de ma thèse, je me suis focalisé sur l'étude des mutants de nodulation d'*A. evenia* qui sont affectés dans les étapes les plus précoces de l'interaction symbiotique pour rechercher les gènes sous-jacents. Ainsi, les gènes de signalisation symbiotique *AePOLLUX* *AeCCaMK*, *AeCYCLOPS*, *AeNSP2*, *AeNIN* et *AeCRK* ont été trouvés avec les mutants classés Nod$^-$ (pas de nodulation) tandis qu'*AeRLCK* a été mis à jour avec les mutants OBN (One Big Nodule). Cinq mutants présentant un phénotype Bump$^+$ discret, similaire à celui observé chez certains mutants Nod$^-$, et deux mutants avec un phénotype proche du « One Big Nodule » ont également été sortis du crible de la population de mutagenèse. Ces mutants ont en commun d'avoir des phénotypes de nodulation conditionnels de sorte qu'ils ont aussi des caractéristiques phénotypiques typiques de mutants d'infection. Les approches de séquençage par Targeted Sequence Capture et de Mapping-by-Sequencing appliquées à ces mutants ont permis d'identifier trois acteurs de l'infection intercellulaire chez *A. evenia*.

Identification d'*AePLT* codant un transporteur de polyols

Plusieurs mutants (B30, K40, O12, N391 et AL1) présentent un phénotype Bump$^+$ 14 jours après inoculation avec la souche de *Bradyrhizobium* ORS278 (Fig. 1A). Ces bumps correspondent à des petites protubérances au niveau de la base des la racines latérales et elles sont colonisées en surface par les bradyrhizobiums. Ils sont souvent annulaires à la différence des bumps observés chez la lignée sauvage au cours d'une cinétique de nodulation où ils correspondent à des nodules en formation qui commencent à émerger des structures de la racine 3 jours après inoculation de la bactérie (Fig. 1A). Mais des coupes fines de ces bumps chez les mutants révèlent des spots d'infection, ce qui indique un blocage de l'invasion par les bradyrhizobiums à l'intérieur du cortex (Fig. 1A). Lorsqu'inoculée avec la souche de *Bradyrhizobium* ORS285, la plante sauvage nodule parfaitement tandis que ces mêmes mutants présentent des réponses variables : le mutant E28 est strictement Nod$^-$, le mutant N391 développe de petits nodules jaunes avec des réactions de défense tandis que le mutant B31 est capable de former à la fois des nodules fonctionnels et de petits nodules jaunes (Fig. 1B).

Nous avons tiré profit des données obtenues par l'approche de Targeted Sequence Capture sur la collection de mutants de nodulation pour près de 400 gènes symbiotiques sélectionnées sur la bases des gènes caractérisés chez les légumineuses modèles et de gènes induits fortement au cours de la nodulation chez *A. evenia* (Quilbé *et al.*, accepté). Nous

Tableau 1 : Données phénotypiques, génétiques et moléculaires des mutants d'infection d'*A. evenia.*

Lignée mutante	Déterminisme		Fréquence gros nodules	Effet de la mutation		Génotypage des F2 mutants	Type de séquençage
	F2 WT:mutant	(*, P>0,05)		Changement nt	Changement aa		
B31	F2 215:89	3:1*	-	G_{1745} to A	G_{93} to E	14/16	TSC
K40	F2 256:69	3:1*	-	G_{1946} to A	G_{160} to E	16/16	TSC
N391	F2 229:77	3:1*	-	G_{3052} to A	G_{386} to R	16/16	TSC
O12	F2 496:141	3:1*	-	G_{1720} to A	A_{85} to T	16/16	TSC
AL1	F2 453:157	3:1*	-	G_{2429} to A	E_{321} to K	20/20*	TSC
D41	F2 437:87	7:1*	52%	G_{1167} to A	W_{389} to STOP	-	MbS
AB3	F2 392:70	7:1*	100%	G_{239} to A	W_{80} to STOP	-	MbS

*nombre d'individus portant la mutation/nombre d'individus total séquencés

Tableau 2 : Tests d'allélismes des mutants d'*AePLT.*

Gène	Croisement (♂ x ♀)	n plante F1 plants (n gousses)	Phénotype F1
AePLT	B31 x K40	24(3)	mutant
	B31 x N391	13(3)	mutant
	K40 x N391	5(2)	mutant
	O12 x K40	26(4)	mutant

avons réussi à identifier chez ces 5 lignées mutantes, à chaque fois une mutation à effet fort dans le gène Ae5g05320 codant un transporteur de polyols. Le Tableau 1 résume les changements de nucléotides identifiés et l'impact sur la séquence protéique qui en découle. Ce gène candidat a été nommé *AePLT*. Les analyses génétiques menées sur ces mutants ont montré pour chacun un déterminisme monogénique et récessif ($p<0.05$) et le génotypage de 20 individus mutants pour chaque population mutant x WT F_2 a montré une co-ségrégation parfaite des mutations avec le phénotype Bump$^+$ (Tableau 1). De plus, les tests d'allélisme, réalisés pour quatre des cinq mutants Bump$^+$, ont confirmé qu'ils définissent le même groupe de complémentation (Tableau 2). Ces données génétiques valident *AePLT* comme étant responsable du phénotype Bump$^+$ chez ces mutants.

Le gène *AePLT* possède une séquence génomique de 3492 nucléotides donnant une protéine prédite de 531 acides aminés. La structure d'AePLT consiste en un premier enchainement de 6 hélices transmembranaires (position 39-215), puis d'un domaine cytoplasmique central (position 216-340) et se termine par 6 hélices transmembranaires (position 341-492), comme les autres membres de la famille des MFS (Major Facilitator Superfamily) qui sont impliqués dans le transport passif de petits solutés comme les ions ou des monosaccharides. Les analyses phylogénétiques révèlent que le gène *AePLT* appartient au même clade que 3 transporteurs précédemment caractérisés chez *L. japonicus* comme étant induits pendant la nodulation et la mycorhization : *LjPLT3*, *LjPLT4* et *LjPLT14* (Tian *et al.*, 2017). Les données du Gene Atlas disponible pour *A. evenia* montrent que ce gène est faiblement exprimé dans les organes aériens de la plante mais que son expression est induite au niveau de la racine après inoculation de la bactérie pour atteindre un maximum d'expression 14 dpi. Ce profil d'expression conforte l'idée qu'*AePLT* intervient au cours de l'infection avec les bradyrhizobiums.

Identification d'*AeRINRK* codant une Receptor-like Kinase

Le mutant D41 a été retenu en raison de son phénotype de nodulation très différent de ceux communément observés lors du crible de la population de mutagenèse d'*A. evenia*. En effet, ce mutant développe plusieurs gros nodules rappelant le phénotype « One Big Nodule » précédemment décrit. L'analyse du phénotype de nodulation sur une population d'individus mutants, cultivés en serre dans de l'attapulgite et inoculés avec la souche de *Bradyrhizobium* ORS278, indique que ce phénotype est aussi duel : on retrouve environ la moitié des

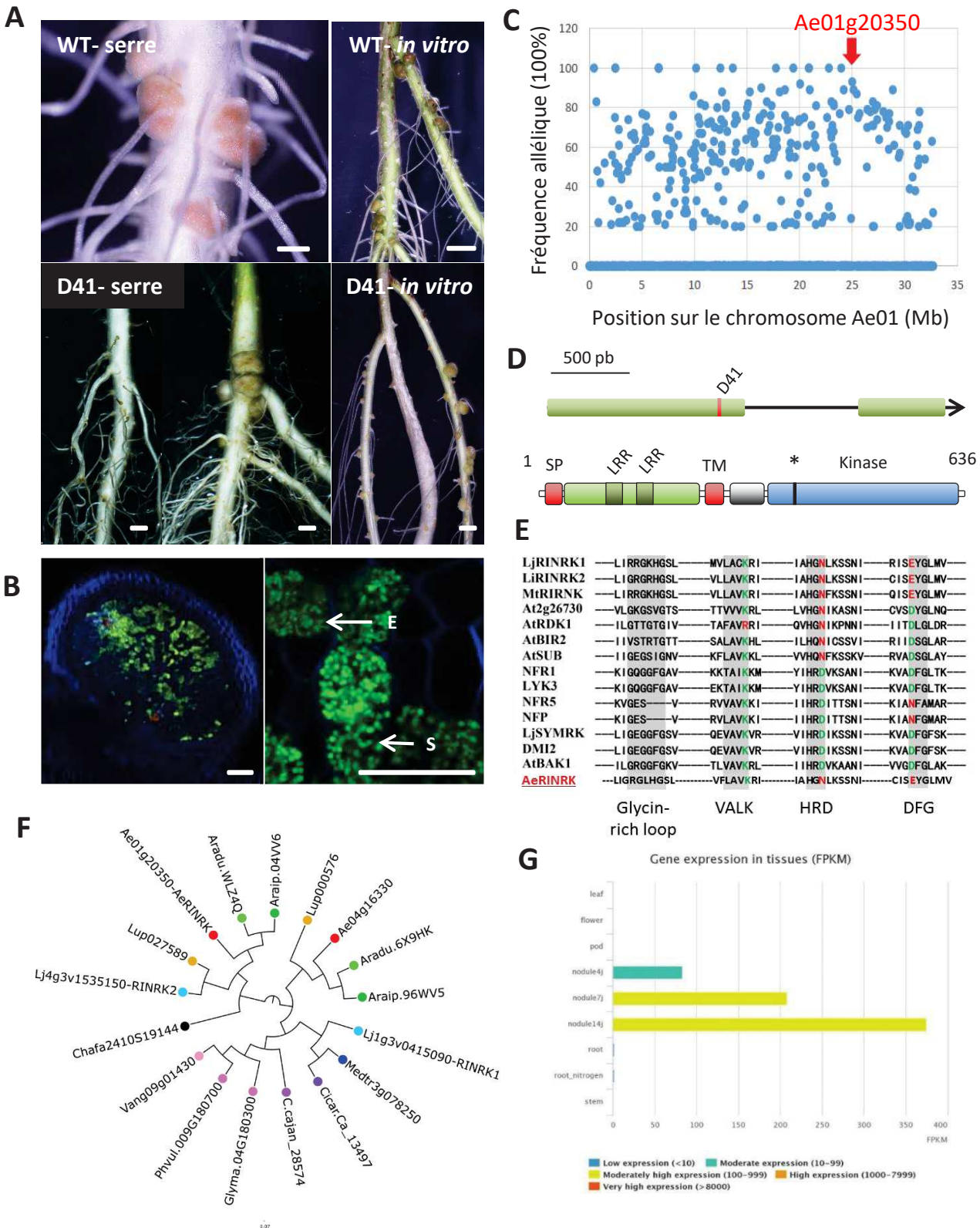


Figure 2 : *AeRINRK*, phénotype des mutants et caractéristiques du gène. A. Phénotype symbiotique observé après inoculation avec la souche ORS278 chez le WT et le mutant D41 cultivés en serre (4wpi) et en tube (14 dpi). **B.** Observation en microscopie confocale d'une coupe de nodule jaune du mutant D41 cultivé en tube 14 jours après inoculation de la souche ORS278-GFP. Le panneau de droite montre des bactéries différenciées (S: sphérique) ou non (E: « Elongated »). **C.** Fréquence des allèles mutants en fonction de leur position sur le chromosome Ae01 chez le mutant D41. **D.** Structure intron/exon du gène *AeRINRK* (Ae01g20350) et de la protéine traduite. L'étoile représente la mutation identifiée **E.** Alignement de la séquence protéique du domaine kinase de plusieurs homologues révélant l'absence de certains motifs clé (en rouge, adapté de Wong *et al.*, 2019). **F.** Phylogénie d'*AeRINRK.* F) Profil d'expression du gène *AeRINRK* provenant du Gene Atlas d'*A. evenia*. Les barres d'échelles représentent 500 µm pour les racines entières sauf pour la microscopie confocale où elle représentent 100 µm

individus strictement Nod$^-$ et l'autre moitié produisant de gros nodules regroupés en haut de la racine (Fig. 2A). Cependant, en culture liquide *in vitro*, les plantes mutantes développent un mélange de petits nodules jaunes et de nodules verts le long de la racine principale (Fig. 2A). Ces plantes ont des symptômes de chlorose indiquant que les nodules ne sont pas fonctionnels. Le mutant D41 présente donc un phénotype de nodulation conditionnel au mode de culture solide (attapulgite) et liquide. Les observations en microscopie confocale révèlent des réactions de défense sur des coupes de nodules du mutant D41 cultivé en milieu liquide, témoignant d'une infection anormale (Fig. 2B). La différenciation des bactéries en bactéroïdes semble également affectée dans les nodules jaunes chez le mutant D41 où sont présents à la fois des bactéries en forme de longs bâtonnets (forme non différenciée) et d'autres sphériques (forme différenciée) (Fig. 2B).

Le croisement du mutant D41 avec la lignée sauvage a révélé en F2 une ségrégation du phénotype de nodulation avec 437 WT et 87 mutants correspondant à un ratio 7:1 (Tableau 1). Ce ratio suggère un déterminisme monogénique et récessif mais avec un problème de transmission des allèles mutants, de sorte que la ségrégation mendélienne 3:1 n'est pas retrouvée. Pour ce mutant, l'approche de Targeted Sequence Capture n'a pas permis d'identifier de mutations dans le set de gènes symbiotiques séquencés. Aussi nous avons mis en œuvre l'approche de Mapping-by-Sequencing pour identifier le gène sous-jacent. L'analyse des fréquences alléliques a révélé la présence d'une liaison génétique sur le chromosome Ae01. Parmi les gènes contenus dans cet intervalle, Ae01g20350 présente une mutation non-sens dans sa CDS (Fig. 2C). Ce gène possède une séquence génomique de 2809 nucléotides et code une protéine prédite de 636 acides aminés homologue à deux Receptor-like Kinase (RLK) identifiés récemment chez *L. japonicus, LjRINRK1* et *LjRINRK2* (pour Rhizobial INfection Receptor-like Kinase) (Fig. 2D) (Li *et al.*, 2020). En conséquence Ae01g20350 a été renommé *AeRINRK*. Un test de complémentation fonctionnel reste à faire pour valider ce gène candidat, mais le fait que les mêmes gènes aient été identifiés chez *L. japonicus* comme étant impliqués dans l'infection par les rhizobiums suggère que *AeRINRK* est bien le gène responsable du phénotype symbiotique observé chez le mutant D41.

AeRINRK se compose d'un peptide signal (position 1-30), un domaine extracellulaire (position 36-245) contenant 2 motifs LRRs, un domaine transmembranaire (position 256-280), un domaine de fonction inconnue (position 280-350) et un domaine kinase (position 359-617) (Fig. 2D). Chez le mutant D41, la mutation EMS (G_{1167}->A) génère un codon stop au

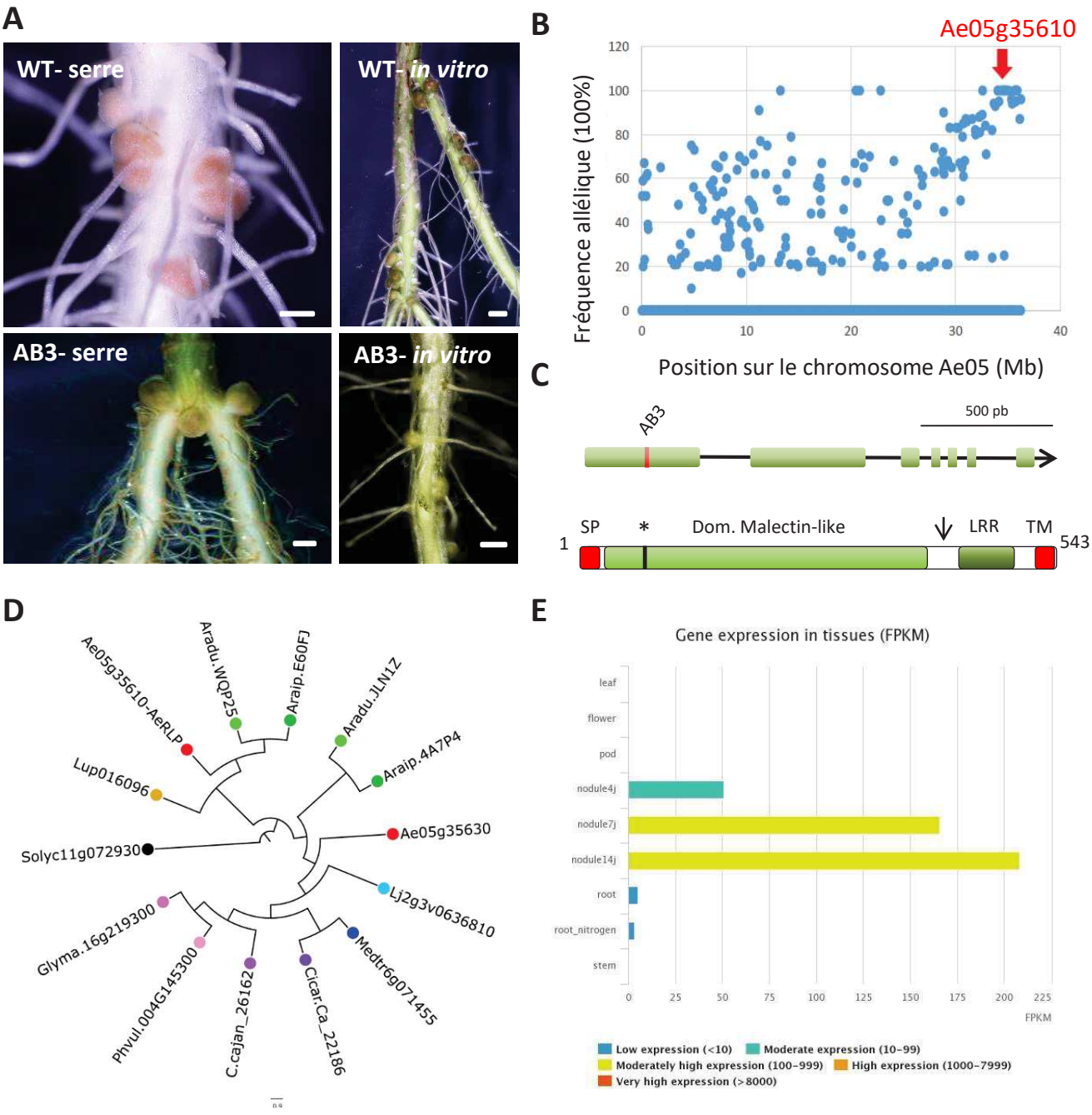


Figure 3 : *AeRLP*, phénotype des mutants et caractéristiques du gène. A. Phénotype symbiotique observé après inoculation avec la souche ORS278 chez le WT et le mutant AB3 cultivés en serre (4wpi) et en tube (14 dpi). **B.** Fréquence des allèles mutants en fonction de leur position sur le chromosome Ae05 chez le mutant AB3. **C.** Structure intron/exon du gène *AeRLP* (Ae05g35610) et de la protéine traduite. L'étoile représente la mutation identifiée et la flèche le motif de clivage GDPC. **D.** Arbre phylogénétique des homologues d'*AeRLP* généré par inférence bayésienne. **E.** Profil d'expression du gène *AeRLP* provenant du Gene Atlas d'*A. evenia*. Les barres d'échelles représentent 500 µm

début du domaine kinase, prédisant que la protéine tronquée qui en résulte est inactive (Tableau 1). Chez *L. japonicus*, LjRINRK1 et LjRINRK2 sont décrits comme étant des kinases atypiques car certains motifs importants pour l'activité kinase ne sont pas conservés (Li *et al*, 2019). Une analyse comparative du domaine kinase d'AeRINRK avec celles de LjRINRK1 et LjRINRK2 et d'autres kinases ayant une structure classique révèle les motifs DFG du KSD VII et Glycin-rich loop du KSD II ne sont également pas conservés chez AeRINRK (Fig. 2E).

L'analyse phylogénétique réalisée avec *AeRINRK* indique qu'il est l'orthologue probable de *LjRINRK2* (Fig. 2F). On remarque également la présence d'un gène paralogue chez *A. evenia,* Ae4g16330, qui est l'orthologue putatif de *LjRINRK1* (Fig. 2F) Il est intéressant de noter que chez *L. japonicus*, *LjRINRK1* a un effet majeur sur l'infection et *LjRINRK2* un rôle très mineur alors que c'est l'orthologue de *LjRINRK2* qui est sorti du crible mutant chez *A. evenia*. Cette observation couplée à la présence des orthologues de *LjRINRK2* chez *L. japonicus* et surtout les légumineuses utilisant un processus d'infection intercellulaire (*Arachis*, *Lupinus*, *Aeschynomene*) suggèrent que RINRK1 pourrait être impliqué plutôt dans l'infection via les cordons d'infection et RINRK2 dans l'infection intercellulaire. Les données du Gene Atlas disponible pour *A. evenia* montrent que *AeRINRK* est faiblement exprimé dans les organes aériens de la plante mais que son expression est induite après inoculation de la bactérie pour atteindre un maximum d'expression 14 dpi (Fig. 2G). Ce profil d'expression conforte l'idée qu'*AeRINRK* intervient au cours de l'infection avec les bradyrhizobiums.

Identification d'*AeRLP* codant une Receptor-like Protein

Au même titre que le mutant D41, le mutant AB3 a été identifié lors du crible de la population de mutagenèse comme développant quelques gros nodules groupés en haut de la racine principale (Fig. 3A). Mais à la différence du mutant D41, les plantes mutantes, cultivées dans de l'attapulgite dans la serre et inoculées avec la souche de *Bradyrhizobium* ORS278, produisent toutes de gros nodules multilobés et les plantes n'ont pas de symptôme foliaire de chlorose, suggérant que ces nodules sont fonctionnels pour la fixation d'azote (Fig. 3A, donnée non montrée). En culture liquide, le mutant AB3 développe un mélange de petits nodules jaunes et verts le long de la racine (Fig. 3A). Il présente ainsi un phénotype de nodulation conditionnel au mode de culture similaire à celui observé avec le mutant D41.

Le croisement du mutant AB3 avec la lignée sauvage a révélé en F2 une ségrégation du phénotype de nodulation avec 392 WT et 70 mutants correspondant à un ratio 7:1 (Tableau

1). Comme observé avec le mutant D41, ce ratio indique un déterminisme monogénique et récessif mais avec un problème de transmission des allèles mutants, de sorte que la ségrégation mendélienne 3:1 n'est pas retrouvée. Pour ce mutant, l'approche de Targeted Sequence Capture n'a pas permis non plus d'identifier de mutations dans le set de gènes symbiotiques séquencés, aussi nous avons utilisé l'approche sans *a priori* de Mapping-by-Sequencing. L'analyse des fréquences alléliques a révélé la présence d'une liaison génétique à une extrémité du chromosome Ae05 (Fig. 3B). Parmi les gènes contenus dans cet intervalle, Ae05g35610 présente une mutation non-sens dans sa CDS et représente le meilleur gène candidat. Ae05g35610 possède une séquence génomique de 2678 nucléotides et code une protéine de 543 acides aminés homologue aux Receptor-like Proteins (RLP). Pour cette raison, elle a été renommée *AeRLP*. La structure de la protéine prédite consiste en un peptide signal (position 1-25), un domaine malectin-like qui représente la majorité de la protéine (position 30-370), un motif de clivage « GDPC » (position 401-404), deux motifs LRR (position 446-494), et un domaine transmembranaire (position 523-542) (Fig. 3C). Chez le mutant AB3, la mutation EMS (G_{239}>A) génère un codon STOP dans le domaine malectine, prédisant que la protéine résiduelle est totalement inactive (Tableau 1). De façon intrigante, la partie extracellulaire d'AeRLP possède la même structure que celle du récepteur symbiotique SYMRK qui appartient aussi à la famille des « Malectin-like containing LRR-RLKs ». Cependant, l'absence de domaine kinase indique que cette protéine appartient à une sous-famille associée de RLPs (Bellande *et al.*, 2017).

L'arbre phylogénétique contenant *AeRLP* montre la présence de deux clades. Un premier clade contient des représentants de la plupart des légumineuses et le second dans lequel *AeRLP* est situé ne contient que des espèces nodulant par un processus d'infection intercellulaire (*Arachis*, *Lupinus*, *Aeschynomene*) (Fig. 3D). Il est aussi intéressant de noter que les légumineuses Dalbergioïdes possèdent deux gènes organisés en tandem (Ae5g35610-*AeRLP* et Ae5g35630 chez *A. evenia*) et qui sont répartis dans les deux clades (Fig. 3D). En écho de l'hypothèse formulée pour *AeRINRK*, il est tentant de spéculer que les gènes du premier clade sont impliqués dans l'infection via les cordons d'infection et ceux du deuxième clade plutôt dans l'infection intercellulaire. Les données du Gene Atlas d'*A. evenia* indiquent que *AeRLP* est faiblement exprimé dans les organes aériens de la plante mais que son expression est induite dans la racine après inoculation de la bactérie pour atteindre un maximum d'expression 14 dpi (Fig. 3E). Ce profil d'expression est compatible avec un rôle

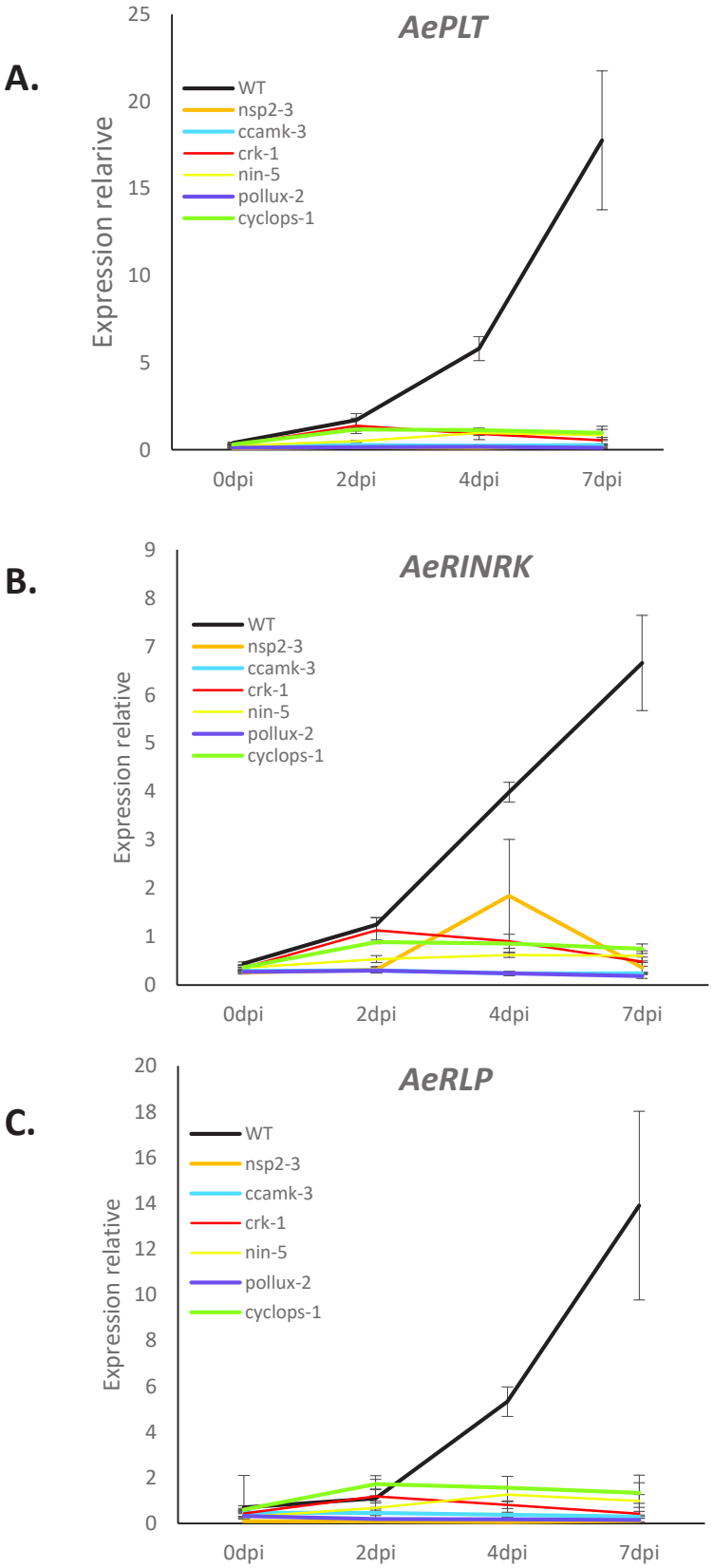


Figure 4 : Rôle des gènes de la voie de signalisation Nod-indépendante dans l'induction de l'expression de gènes d'infection au cours de la nodulation. **A**. Expression du gène *AePLT*. **B.** Expression du gène *AeRINRK*. **C.** Expression du gène *AeRLP*. **C.** Expression du gène *AeRINRK.* Le suivi d'expression a été réalisé chez le sauvage et différents mutants affectés dans des gènes de signaisation symbiotique à 0, 2, 4 et 7 jours après inoculation (dpi) avec la souche de Bradyrhizobium ORS278. Les valeurs ont été normalisées celles des gènes de *AeEF1* et *AeUBI*. Les moyennes et l'erreur standard. Sont dérivées de trois répétitions biologiques.

d'*AeRLP* dans l'infection avec les bradyrhizobiums. Ce profil d'expression et la présence d'un codon stop prématuré dans ce gène chez le mutant AB3 laissent à penser que le bon gène candidat a été identifié. Néanmoins, il conviendra de valider le rôle symbiotique d'*AeRLP* par des expériences de complémentation fonctionnelle.

L'expression d'*AePLT*, *AeRINRK* et d'*AeRLP* est contrôlée par la voie de signalisation symbiotique chez *A. evenia*

Les trois gènes identifiés sont probablement impliqués dans l'infection intercellulaire chez *A. evenia*. Ils ont en commun de présenter un profil d'expression similaire à celui observé pour *AeNIN*, codant un facteur de transcription clé de l'infection et de l'organogenèse nodulaire, et *AeCRK*, codant un récepteur kinase intervenant de façon précoce dans l'établissement de l'interaction symbiotique (Chapitres 1 et 2). A savoir, leur expression est induite au cours de la nodulation et pas d'expression notable n'est retrouvé dans les parties aériennes de la plante. Pour déterminer dans quelle mesure l'expression d'*AePLT*, *AeRINRK* et d'*AeRLP* est contrôlée par la voie de signalisation symbiotique, nous avons quantifié par Q-RT-PCR leur expression chez la lignée sauvage et les mutants *crk*, *pollux*, *ccamk*, *cyclops*, *nsp2* et *nin*, à 0, 2, 4 et 7 jours après inoculation avec la souche de *Bradyrhizobium* ORS278. Chez la lignée sauvage, on observe que les trois gènes étudiés sont faiblement exprimés au début de la cinétique (0 et 2 dpi) mais que leur expression croît fortement au cours du temps (4 et 7dpi) (Fig. 4A-C). Au contraire, chez tous les mutants qui sont altérés dans des gènes de la voie de signalisation symbiotique, une très faible induction de leur expression peut être observée à 2 dpi, mais il n'y a pas d'augmentation significative aux temps plus tardifs (4 et 7 dpi) (Fig. 4A-C). L'expression d'*AePLT*, *AeRINRK* et d'*AeRLP* est donc dépendante de l'activité de la voie de signalisation symbiotique.

Discussion

La découverte de trois gènes d'infection ouvre la voie à l'étude des mécanismes moléculaires de l'infection intercellulaire chez *A. evenia*, tant au niveau signalisation (*AeRINRK* et *AeRLP*) que du métabolisme (*AePLT*). De par leurs profils d'expression, ils représentent aussi d'excellents marqueurs potentiels pour caractériser plus finement l'infection en développant des fusions promoteur-gène rapporteur.

Le gène *AePLT* code un transporteur de monosaccharides qui appartient à une famille multigénique. Chez *L. japonicus,* plusieurs gènes de PLT sont induits au cours de la nodulation, *LjPLT3*, *LjPLT14* et *LjPLT4* (Tian *et al.*, 2017). Mais aucun mutant dans ces gènes avec un phénotype symbiotique n'a été décrit jusqu'à présent, certainement en raison d'une redondance fonctionnelle entre ces gènes homologues. L'identification de plusieurs mutants de nodulation altérés pour *AePLT* indique que ce problème n'est pas présent chez *A. evenia*. *AePLT* étant dans le même sous-clade phylogénétique que *LjPLT3* et *LjPLT4*, cela suggère que le rôle de ces transporteurs est conservé entre l'infection via les cordons d'infection et l'infection intercellulaire. Des études sur LjPLT4 ont montré une forte spécificité pour le transport actif du xylitol (Kalliampakou *et al.*, 2011). On peut donc se demander si *AePLT* est également capable de transporter du xylitol ou si le substrat correspond à un autre polyol comme le mannitol ou le sorbitol, qui sont des sucres communément métabolisés par les bactéries rhizosphériques. Chez *A. evenia*, le phénotype des mutants *plt* est différent suivant le partenaire *Bradyrhizobium*. On peut émettre l'hypothèse que les souches de *Bradyrhizobium* photosynthétiques n'utilisent pas les mêmes sucres pendant l'infection. Il sera donc utile de caractériser quels sucres sont métabolisés par ces souches au cours du processus symbiotique.

Les gènes *AeRINRK* et *AeRLP* codent deux récepteurs, eux aussi impliqués dans l'infection à double titre. En effet, le phénotype symbiotique en culture liquide des mutants correspondants est similaire à celui observé chez de nombreux mutants d'infection de la collection mais qui n'ont pas été encore caractérisés au niveau moléculaire. En revanche, le phénotype symbiotique observé en milieu solide est proche de celui décrit pour les mutants d'*AeRLCK* qui est proposé contrôler l'initiation de l'infection. Ce double phénotype ouvre la possibilité qu'*AeRINRK* et *AeRLP* interagissent à la fois avec *AeRLCK* et des gènes de composants de la machinerie d'infection. Au même moment où nous avons identifié *AeRINRK*, les gènes correspondant *LjRINRK1* et *LjRINRK2* ont été décrits chez *L. japonicus* et proposés avoir un rôle spécifique dans l'infection soit en coordonnant des réponses de la voie de signalisation symbiotique soit en percevant un signal inconnu et requis pour l'infection rhizobienne (Li *et al.*, 2020). Le gène *AeRLP* est également un acteur original car il appartient à une petite famille de gènes codant des récepteurs de type RLP et dont les fonctions biologiques restent largement inconnues (Bellande *et al.*, 2017; Jamieson *et al.*, 2018). Sous réserve de validation fonctionnelle, *AeRLP* représenterait le premier RLP avec

une fonction symbiotique décrite. Les phénotypes très similaires des mutants d'*AeRINRK* et d'*AeRLP* permettent de suggérer qu'ils pourraient être des co-récepteurs. Néanmoins, la présence d'une kinase prédite être inactive chez AeRINRK et l'absence de partie intracellulaire chez AeRLP indiquent qu'ils doivent interagir avec d'autres acteurs de signalisation, qui restent à identifier, pour assurer la transduction du signal.

MATERIELS & METHODES

La nature du matériel végétal, les analyses de caractérisation phénotypique et les analyses de séquences sont identiques à celles décrites dans la partie 3.2.

Les analyses de Q-RT-PCR ont été réalisées comme décrit dans le Chapitre 2 avec les amorces listées dans le Tableau S1.

Références bibliographiques

Bellande, K., J.-J. Bono, B. Savelli, E. Jamet, and H. Canut. 2017. Plant Lectins and Lectin Receptor-Like Kinases: How Do They Sense the Outside? *International Journal of Molecular Sciences* 18: 1164.

Jamieson, P. A., L. Shan, and P. He. 2018. Plant cell surface molecular cypher: Receptor-like proteins and their roles in immunity and development. *Plant Science* 274: 242–251.

Kalliampakou, K. I., E. D. Kouri, H. Boleti, O. Pavli, L. Maurousset, M. K. Udvardi, P. Katinakis, *et al.* 2011. Cloning and functional characterization of *Lj* PLT4, a plasma membrane xylitol H $^{+}$ - symporter from *Lotus japonicus*. *Molecular Membrane Biology* 28: 1–13.

Li, X., Z. Zheng, X. Kong, J. Xu, L. Qiu, J. Sun, D. Reid, *et al.* 2020. Atypical Receptor Kinase RINRK1 Required for Rhizobial Infection But Not Nodule Development in Lotus japonicus1[OPEN]. 181: 13.

Tian, L., L. Liu, Y. Yin, M. Huang, Y. Chen, X. Xu, P. Wu, *et al.* 2017. Heterogeneity in the expression and subcellular localization of POLYOL/MONOSACCHARIDE TRANSPORTER genes in Lotus japonicus H. Gerós [ed.],. *PLOS ONE* 12: e0185269.

Tableau S1 : Liste des amorces utilisées pour la q-RT-PCR

Nom d'amorce	Fonction	Séquence 5'-3'
AeRLP-F	q-RT-PCR	GCACAACAACACCATAGTA
AeRLP-R		TTGAACCAGATTAGTATTCCTTA
AeRINRK-F		TCAACTCTTGTCGTCCTT
AeRINRK--R		TTGCCAATCCTTCTCTGA
AePLT-F		AATAGAGAAGAGGAGCAGTT
AePLT-R		CAGTGGAAGCAATGAGTATC

Discussion générale

Le genre *Aeschynomene* contient un grand nombre d'espèces possédant des propriétés de nodulation originales comme la capacité à former des nodules caulinaires et racinaires, de s'associer avec des *Bradyrhizobium* photosynthétiques ainsi que la possibilité pour certaines d'entre elles d'être nodulées sans le besoin de reconnaissance des facteurs Nod. Si les études sur le système symbiotique *Bradyrhizobium-Aeschynomene* ont longtemps été et continuent d'être menées sur le partenaire bactérien depuis les travaux d'Alazard (1985), des efforts de recherche ont été fournis début des années 2010 pour faire émerger un axe de recherche côté plante, conduisant au choix de l'espèce *Aeschynomene evenia* comme modèle d'étude de la symbiose Nod-indépendante (Arrighi *et al.*, 2012).

Des analyses transcriptomiques couplées d'une approche de génétique inverse chez *A. evenia* ont été le point de départ pour avancer dans la compréhension des mécanismes moléculaires intervenant dans le processus Nod-indépendant, notamment en révélant que plusieurs gènes de signalisation symbiotiques sont conservés comme *SYMRK*, *CCaMK* et *LHK1* (Chaintreuil *et al.*, 2016; Fabre *et al.*, 2015; Gully *et al.*, 2018). Une période clé dans l'étude de la symbiose Nod-indépendante s'est ouverte avec le développement d'une approche de génétique directe sur *A. evenia* grâce à la création d'une population de mutagénèse EMS et au séquençage complet de son génome (Projet ANR AeschyNod, 2014-2019). C'est à cette étape du projet conduit sur *A. evenia* que ma thèse a débuté.

J'ai pu participer à la fin du crible de la population de mutagénèse EMS et à la confirmation des mutants de la collection, m'amenant à observer une grande diversité de phénotypes symbiotiques. J'ai ensuite conduit une analyse génomique, génétique et fonctionnelle de la nodulation Nod-indépendante chez *A. evenia*. Ce travail a permis de montrer l'implication de plusieurs gènes de la voie de signalisation Nod, mais pas de ceux codant les récepteurs des facteurs Nod, dans la symbiose Nod-indépendante. J'ai également réussi à mettre à jour deux gènes nouveaux, *AeCRK* et *AeRLCK*, car sans orthologue évident chez les légumineuses modèles. Ces découvertes ouvrent la voie à de nouvelles analyses de génétique fonctionnelle des mécanismes de nodulation chez les *Aeschynomene*. Je discute ici des résultats phares de mon travail de thèse, des possibilités de poursuite et de l'apport de ce travail dans le domaine de la symbiose fixatrice d'azote.

1. Les LysM-RLK ne semblent pas impliqués dans la symbiose NF-indépendante

Chez les légumineuses modèles, le point de départ de la voie de signalisation Nod correspond à la perception des facteurs Nod par deux récepteurs : MtNFP/LjNFR5 et MtLYK3/LjNFR1 (Gough *et al.*, 2018). Ils appartiennent à la famille des LysM-RLK qui est largement impliquée dans les interactions plante-microorganisme (Buendia *et al.*, 2018). Pour cette raison, ils sont largement étudiés dans un contexte de nodulation, de mycorhization mais aussi de l'immunité. L'exploration du rôle de ces récepteurs chez les légumineuses et les non-légumineuses conduit à ré-évaluer constamment le modèle établi chez *M. truncatula* et *L. japonicus*. Une étude récente menée chez la seule non-légumineuse capable de noduler avec des rhizobiums, *Parasponia*, a révélé la présence de deux copies à la fois pour *NFP* et *LYK3* qui sont toutes nécessaires à la nodulation (Rutten *et al.*, 2020). Une autre étude menée chez l'espèce 4x *A. hypogaea* (arachide), qui appartient au même clade des Dalbergioïdes que les *Aeschynomene,* vient de montrer que les 2 gènes homéologues de *NFP* sont impliqués dans la nodulation mais cette démonstration reste a faire pour les 2 gènes *LYK3* (Shu *et al.*, 2020). Chez *A. evenia*, pour qui l'activation de la nodulation ne repose pas sur la reconnaissance des facteurs Nod, aucune mutation causale n'a été retrouvée dans les gènes de LysM-RLK parmi les mutants de nodulation sortis du crible. Les gènes *AeNFP* et *AeLYK3* sont présents dans le génome mais *AeNFP* possède un motif lié aux facteurs Nod qui n'est pas conservé et *AeLYK3* n'est pas exprimé. Par ailleurs l'orthologue du gène *LjEPR3*, qui est important chez *L. japonicus* pour la reconnaissance des exopolysaccharides de rhizobiums au cours de l'infection, est absent du génome d'*A. evenia*. Ces données suggèrent fortement que ces trois LysM-RLKs ne sont pas impliqués dans le processus de nodulation Nod-indépendant et que la perception du partenaire bactérien met en jeu d'autres types de récepteurs chez *A. evenia*. Pour comprendre l'évolution des LysM-RLKs symbiotiques chez les *Aeschynomene*, une analyse de génétique comparative doit être menée entre *A. evenia* (Nod-indépendant) et *A. patula* (Nod-dépendant). D'une part ces gènes de LysM-RLK chez *A. patula* seront inactivés par RNAi ou CRISPR-Cas9, pour déterminer leur rôle symbiotique, d'autre part, les gènes d'*A. patula* seront surexprimés chez *A. evenia* pour tester la possibilité qu'ils interfèrent avec la nodulation Nod-indépendante en présence de la souche de *Bradyrhizobium* photosynthétique ORS285 qui produit des facteurs Nod.

2. La voie de signalisation Nod est (au moins partiellement) conservée

Chez les légumineuses modèles, en aval des récepteurs des facteurs Nod, différents composants participent à une voie de signalisation Nod aboutissant à l'activation de l'expression de gènes de nodulines dans le noyau (pour revue, Roy *et al.*, 2020)). La majorité des orthologues probables des gènes de cette voie de signalisation sont présents chez *A. evenia* et le rôle dans le processus Nod-indépendant de plusieurs d'entre eux a été mis en évidence avec l'approche mutant, à savoir : *AePOLLUX*, *AeCCAMK*, *AeCYCLOPS*, *AeNSP2* et *AeNIN*. Tous les gènes connus de cette voie ne sont pas sortis via le crible de mutants de nodulation. C'est le cas notamment de *SYMRK* qui intervient au tout début de cette voie et de *PUB1* qui code un interactant de SYMRK. Les analyses génomiques ont révélé que chez *A. evenia*, ces deux gènes présentent deux paralogues probablement issus de la WGD datée de 58 MA et ces paralogues présentent des profils d'expression compatibles avec un rôle symbiotique. Il sera donc intéressant de tester s'il y a de la redondance fonctionnelle ou de la sub-fonctionnalisation comme investigué chez les légumineuses avec les paires de paralogues symbiotiques *MtERN1/ERN2* et *LjEIN1/EIN2* (Cerri *et al.*, 2016; Miyata *et al.*, 2013). Nous n'avons pas retrouvé de mutation causale non plus pour les gènes *AeCASTOR* et *AeNSP1* alors qu'ils sont présents en simple copie chez *A. evenia* et que chez les légumineuses modèles CASTOR est connu pour fonctionner avec POLLUX et NSP1 avec NSP2 (pour revue, Roy *et al.*, 2020). Une inactivation ciblée par RNAi ou CrisPR-Cas9 devrait permettre de statuer sur leur implication éventuelle dans le processus Nod-indépendant. Parmi les acteurs de la voie de signalisation symbiotique chez *A. evenia*, *AeNSP2* s'est révélé être nécessaire pour le développement des couronnes de poils axillaires colonisés par les bradyrhizobiums et pour induire l'expression d'un set de gènes, poils axillaires et gènes étant régulés négativement par le statut azoté de la plante. Ce contrôle génétique pourrait donc correspondre à une fonction non symbiotique d'*AeNSP2*. Une telle fonction non-symbiotique, dépendante du statut azoté ou phosphaté de la plante, a également été évoquée pour les *NSP2* de *M. truncatula*, *L. japonicus* et *A. thaliana*, mais les connaissances dessus restent très parcellaires (Lin *et al.*, 2013; Murakami *et al.*, 2013; Teft, 2018). Avec les poils axillaires comme marqueur morphologique et un set de 60 gènes candidats comme marqueurs génétiques, *A. evenia* représente un bon modèle d'étude pour clarifier comment NSP2 remplit sa fonction non-symbiotique en fonction du statut nutritionnel de la plante.

3. Un nouveau récepteur-kinase, AeCRK, est requis pour la symbiose Nod-indépendante

Le crible de mutants de nodulation chez *A. evenia* nous a aussi permis d'identifier le gène *AeCRK* qui est requis pour les premières étapes de la symbiose Nod-indépendante. La caractérisation phénotypique et moléculaire des mutants *crk* suggère qu'*AeCRK* est impliqué dans la voie de signalisation symbiotique et/ou l'infection. Les analyses phylogénétiques et moléculaires révélent qu'*AeCRK* est ancestral chez les légumineuses mais qu'il a été perdu chez les légumineuses modèles et que ce gène est soumis à une pression négative chez les *Aeschynomene* Nod-indépendantes. Les études réalisées sur d'autres membres de la famille des CRKs (Cysteine-rich Receptor-like Kinases) apportent des pistes sur le rôle qu'*AeCRK* peut remplir chez *A. evenia*. Une étude récente a mis en exergue la similitude de forme 3D des deux domaines DUF26 présents dans la partie extracellulaire des CRKs avec les domaines lectines de plantes, suggérant qu'ils puissent reconnaître des sucres produits par les micro-organismes (Vaattovaara *et al.*, 2019). Le fait que certaines protéines sécrétées à domaine DUF26, comme Gnk2 du gingko, sont capables de lier mannose *in vitro* et possèdent une activité antifongique, appuie cette possibilité (Miyakawa *et al.*, 2009, 2014). Néanmoins les CRKs sont plutôt associés à la signalisation des ROS. Il est supposé que dans le domaine DUF26, trois paires de cystéines sont la cible de modifications RedOx et que la modification des ponts disulfures permet la perception des ROS (Bourdais *et al.*, 2015). Les CRKs interviennent souvent en aval d'autres récepteurs qui activent la production de ROS via les protéines RBOHs (Respiratory Burst Oxidase Homolog). Les études menées sur AtCRK2 et AtCRK36 montent que les CRKs peuvent interagir avec un complexe récepteur activé par la perception de ligands extracellulaires et les RBOHs par phosphorylation, régulant positivement la production de ROS et assurant l'amplification du signal (Lee *et al.*, 2017; Kimura *et al.*, 2020). Les ROS sont connus pour intervenir dans les étapes précoces de la symbiose rhizobienne, à la fois au niveau signalisation et infection, et plusieurs RBOHs ont un rôle symbiotique démontré (Marino *et al.*, 2011; Zipfel and Oldroyd, 2017; Yu *et al.*, 2018). Peu est encore connu chez *A. evenia*, mais des « cell collapse » ont lieu au cours de l'infection intercellulaire et plusieurs gènes de RBOHs ont des profils compatibles avec un rôle symbiotique. Aussi, la caractérisation fonctionnelle d'*AeCRK* et l'étude de la signalisation par les ROS devrait donc apporter un éclairage important sur leur rôle conjoint dans la mise en place de la symbiose Nod-indépendante.

4. Un nouveau RLCK, AeRLCK, est aussi essentiel pour initier la symbiose Nod-indépendante

L'identification du gène *AeRLCK* constitue l'autre grande découverte issue de l'approche mutant chez *A. evenia*. En effet, il a été identifié à partir de mutants « One Big Nodule » dont le phénotype suggère qu'il contrôle l'initiation de l'infection par les *Bradyrhizobium* photosynthétiques. De plus, les analyses génomiques et phylogénétiques indiquent qu'*AeRLCK* est issu d'une duplication en tandem propre aux *Aeschynomene* Nod-indépendantes et que ce gène a suivi une évolution particulière dans ce clade. Pour déterminer si le rôle d'*AeRLCK* est spécifique à la symbiose Nod-indépendante, des analyses comparatives et de complémentation interspécifique pourront être réalisées avec la version simple copie présente chez des légumineuses utilisant un processus dépendant des facteurs Nod, comme l'espèce proche, *A. patula*, et le modèle *M. truncatula*. *AeRLCK* est probablement un déterminant clé des mécanismes symbiotiques chez *A. evenia* car il appartient à une famille d'acteurs de signalisation, les RLCK (Receptor-like Cytoplasmic Kinases) qui sont centraux dans de nombreuses voies de signalisation, en interagissant avec des RLKs percevant des signaux extracellulaires et en transmettant le signal par phosphorylation d'autres acteurs moléculaires (Liang and Zhou, 2018; Lin *et al.*, 2013). La transduction du signal par les RLCK peut passer par l'activation d'une cascade MAPK, la phosphorylation de protéines ancrées dans la membrane plasmique, comme les CRKs et les RBOHs, ou de facteurs de transcription nucléaires (Lu *et al.*, 2010; Li *et al.*, 2014; Lal *et al.*, 2018; Tian *et al.*, 2019; Wang *et al.*, 2020). Pour mieux comprendre le rôle symbiotique d'AeRLCK, la recherche d'interactants par des approches de biochimie représente donc enjeu important. Des stratégies comme le Turbo-ID, qui permet un marquage de proximité basée sur la biotine, et le « Multiplexed Assay for Kinase Specificity » (MAKS), pour d'identifier les cibles phosphorylées, ont été récemment adaptées aux récepteurs symbiotiques de *M. truncatula* et *L. japonicus* (Jayaraman *et al.*, 2017; Arora *et al.*, 2020). Ces approches appliquées à AeRLCK pourraient permettre de déterminer s'il interagit avec AeCRK et/ou des récepteurs intervenant en amont. En effet, les récepteurs responsables de la perception des *Bradyrhizobium* photosynthétiques n'ont pas été trouvés avec les mutants Nod^- d'*A. evenia*. La mutagénèse EMS peut ne pas avoir été saturante, mais il est également possible qu'il existe de la redondance fonctionnelle ou que plusieurs signaux bactériens soient perçus de façon concomitante par des récepteurs différents. La caractérisation de l'interactome d'AeRLCK sera donc déterminante dans cette recherche.

5. Mécanismes de l'infection intercellulaire dans un contexte Nod-indépendant

Les processus d'infection chez les légumineuses peuvent être divisés en deux catégories avec d'un côté l'infection intracellulaire décrite chez les légumineuses modèles, et de l'autre, l'infection intercellulaire présente chez près de 25% des légumineuses. Plusieurs gènes importants pour l'infection intracellulaire ont été identifiés chez *M. truncatula* et *L. japonicus*, leur expression est plus la plupart régulée par *NIN*, un acteur central de l'infection par rhizobium (Liu *et al.*, 2019a; Roy *et al.*, 2020). A l'inverse, on en connait peu sur les mécanismes de l'infection intercellulaire. Une étude récente a tiré parti de la capacité de *L. japonicus* à mettre en œuvre une infection soit intracellulaire soit partiellement intercellulaire, suivant le partenaire bactérien (Montiel *et al.*, 2020). En utilisant à la fois des mutants de nodulation et une approche RNAseq, il a été montré que les gènes *RPG*, *ERN1* et *RINRK1* sont associés au processus intracellulaire, tandis que *LHK1* ou *IPT4* sont plus importants pour l'infection intercellulaire. L'émergence d'*A. evenia* comme modèle d'étude d'un processus symbiotique pour lequel l'infection intercellulaire est la règle doit permettre d'identifier des acteurs de ce type d'infection alternatif à celui décrit chez les légumineuses modèles. Nos analyses génomiques ont déjà montré que certains gènes d'infection, comme *RPG* et *FLOT*, sont absents chez *A. evenia.* L'approche mutant a permis d'identifier sans *a priori*, en plus d'*AeRLCK* et *AeCRK*, trois gènes de l'infection intercellulaire, deux impliqués dans la signalisation, *AeRINRK* et *AeRLP*, et un dans le transport de métabolites, *AePLT*. Pour continuer à caractériser l'infectome d'*A. evenia*, les autres mutants d'infection de la collection seront prochainement analysés par Mapping-by-Sequencing. Cette approche est escomptée permettre d'identifier des récepteurs symbiotiques ou d'autres acteurs de signalisation, des composant de la machinerie d'infection et des transporteurs. Nos connaissances sur les mécanismes de signalisation et d'infection chez *A. evenia* sont encore très parcellaires, mais il est possible de formuler une hypothèse sur l'ordre d'intervention des acteurs identifiés. AeRLCK serait l'un des premiers acteurs à intervenir : après perception du signal non-Nod par le récepteur partenaire, il permettrait aux bactéries de progresser dans l'espace intercellulaire. Dans un deuxième temps, AeCRK pourrait percevoir des ROS dont la production est activée par les bactéries et qui induit un collapse au niveau de la cellule du cortex infectée. Finalement, AeRINRK et AeRLP pourraient former un complexe pour assurer l'infection et/ou l'endocytose des bactéries dans la cellule fondatrice du nodule.

6. Les *Aeschynomene* : modèles d'étude de la diversité des mécanismes symbiotiques

A. evenia a émergé comme modèle d'étude pour comprendre les mécanismes de la nodulation en absence de reconnaissance de facteurs Nod et de formation de cordons d'infection. Cependant, *A. evenia* est aussi d'intérêt pour étudier d'autres processus symbiotiques. Elle présente avec certaines légumineuses semi-aquatiques, comme *S. rostrata*, la capacité à développer des nodules de tige dont la régulation est restée largement inconnue jusqu'à la démonstration que les gènes de la voie signalisation symbiotique contrôlent à la fois la nodulation de racine et de tige chez *A. evenia* (Masson-Boivin *et al.*, 2009). En revanche les mutants affectés dans *AeRLCK*, *AeRINRK* et *AeRLP* conservent la capacité à développer des nodules caulinaires (chapitre 3 et données non montrées). Cette dichotomie a été retrouvée chez plusieurs mutants d'infection d'*A. evenia* dont les gènes sous-jacents n'ont pas encore été identifiés, suggérant des mécanismes d'infection différents au niveau de la racine et de la tige. Au cours de nos analyses génomiques, nous avons aussi mis en évidence des traits symbiotiques intéressants lors des étapes de différenciation et de fonctionnement des nodules, la plus remarquable étant la mise évidence qu'*A. evenia* et d'autres légumineuses comme l'arachide ont recruté des globines de classe 1 comme des leghémoglobines (Lbs) pour maitriser la pression en oxygène dans le nodule, alors que toutes les Lbs étaient supposées être dérivées de globine de classe 2 (Becana *et al.*, 2020). Cette découverte ouvre la voie à l'analyse comparative des deux types de Lbs et à l'étude de leur convergence évolutive. La caractérisation des étapes précoces et tardives de la nodulation menée actuellement chez *A. evenia* peut être potentialisée en la combinant à une analyse comparative avec une *Aeschynomene* qui possède des propriétés de nodulation contrastées. *A. patula*, qui présente une activation symbiotique dépendante des facteurs Nod et un fonctionnement du nodule différent mais qui utilise le même processus d'infection intercellulaire, a récemment émergé d'une analyse phylogénétique du genre *Aeschynomene* comme modèle d'étude complémentaire (Brottier *et al.*, 2018). En effet, son génome diploïde et sa bonne maniabilité expérimentale doivent faciliter les analyses comparatives avec *A. evenia* pour éclairer l'évolution récente de traits symbiotiques. Le développement du système génétique *A. evenia*-*A. patula* s'inscrit ainsi pleinement dans l'essor récent d'approches de phylogénomique visant à mieux appréhender l'origine et l'évolution de la symbiose fixatrice d'azote (Griesmann *et al.*, 2018; van Velzen *et al.*, 2018).

Bibliographie

Adams, D.G., Duggan, P.S., 2008. Cyanobacteria-bryophyte symbioses. Journal of Experimental Botany 59, 1047–1058. https://doi.org/10.1093/jxb/ern005

Alazard, D., 1985. Stem and Root Nodulation in Aeschynomene spp. Applied and Environmental Microbiology 50, 732–734. https://doi.org/10.1128/AEM.50.3.732-734.1985

Alazard, D., Becker, M., 1987. Aeschynomene as green manure for rice. Plant Soil 101, 141–143. https://doi.org/10.1007/BF02371043

Alazard, D., Duhoux, E., 1988. Diversity of Stem Nodulation Sites in Aeschynomene spp. Journal of Plant Physiology 132, 123–125. https://doi.org/10.1016/S0176-1617(88)80195-0

Allen, O.N., Allen, E.K., 1981. The Leguminosae: a source book of characteristics, uses and nodulation. Macmillan [u.a.], London.

Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D.G., de Carvalho-Niebel, F., 2007. AP2-ERF Transcription Factors Mediate Nod Factor–Dependent Mt *ENOD11* Activation in Root Hairs via a Novel *cis* -Regulatory Motif. Plant Cell 19, 2866–2885. https://doi.org/10.1105/tpc.107.052944

Antolín-Llovera, M., Ried, M.K., Parniske, M., 2014. Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE Ectodomain Promotes Complex Formation with Nod Factor Receptor 5. Current Biology 24, 422–427. https://doi.org/10.1016/j.cub.2013.12.053

Ardourel, M., Demont, N., Debellé, F., Maillet, F., de Billy, F., Promé, J.C., Dénarié, J., Truchet, G., 1994. Rhizobium meliloti lipooligosaccharide nodulation factors: different structural requirements for bacterial entry into target root hair cells and induction of plant symbiotic developmental responses. Plant Cell 6, 1357–1374. https://doi.org/10.1105/tpc.6.10.1357

Arora, D., Abel, N.B., Liu, C., Van Damme, P., Yperman, K., Eeckhout, D., Vu, L.D., Wang, J., Tornkvist, A., Impens, F., Korbei, B., Van Leene, J., Goossens, A., De Jaeger, G., Ott, T., Moschou, P.N., Van Damme, D., 2020. Establishment of Proximity-Dependent Biotinylation Approaches in Different Plant Model Systems. Plant Cell 32, 3388–3407. https://doi.org/10.1105/tpc.20.00235

Arrighi, J.-F., Barre, A., Ben Amor, B., Bersoult, A., Soriano, L.C., Mirabella, R., de Carvalho-Niebel, F., Journet, E.-P., Ghérardi, M., Huguet, T., Geurts, R., Dénarié, J., Rougé, P., Gough, C., 2006. The *Medicago truncatula* Lysine Motif-Receptor-Like Kinase Gene Family Includes *NFP* and New Nodule-Expressed Genes. Plant Physiol. 142, 265–279. https://doi.org/10.1104/pp.106.084657

Arrighi, J.-F., Cartieaux, F., Brown, S.C., Rodier-Goud, M., Boursot, M., Fardoux, J., Patrel, D., Gully, D., Fabre, S., Chaintreuil, C., Giraud, E., 2012. *Aeschynomene evenia* , a Model Plant for Studying the Molecular Genetics of the Nod-Independent Rhizobium-Legume Symbiosis. MPMI 25, 851–861. https://doi.org/10.1094/MPMI-02-12-0045-TA

Arrighi, J.-F., Cartieaux, F., Chaintreuil, C., Brown, S., Boursot, M., Giraud, E., 2013. Genotype Delimitation in the Nod-Independent Model Legume Aeschynomene evenia. PLoS ONE 8, e63836. https://doi.org/10.1371/journal.pone.0063836

Arrighi, J.-F., Godfroy, O., de Billy, F., Saurat, O., Jauneau, A., Gough, C., 2008. The RPG gene of Medicago truncatula controls Rhizobium-directed polar growth during infection. Proceedings of the National Academy of Sciences 105, 9817–9822. https://doi.org/10.1073/pnas.0710273105

Azani, N., Babineau, M., Bailey, C.D., Banks, H., Barbosa, ArianeR., Pinto, R.B., Boatwright, JamesS., Borges, LeonardoM., Brown, GillianK., Bruneau, A., Candido, E., Cardoso, D., Chung, K.-F., Clark, RuthP., Conceição, A. deS., Crisp, M., Cubas, P., Delgado-Salinas, A., Dexter, KyleG., Doyle, JeffJ., Duminil, J., Egan, AshleyN., De La Estrella, M., Falcão, MarcusJ., Filatov, DmitryA., Fortuna-Perez, A.P., Fortunato, RenéeH., Gagnon, E., Gasson, P., Rando, J.G., Azevedo Tozzi, A.M.G. de, Gunn, B., Harris, D., Haston, E., Hawkins, JulieA., Herendeen, PatrickS., Hughes, ColinE., Iganci, JoãoR.V., Javadi, F., Kanu, S.A., Kazempour-Osaloo, S., Kite, GeoffreyC., Klitgaard, BenteB., Kochanovski, FábioJ., Koenen, ErikJ.M., Kovar, L., Lavin, M., Roux, M. le, Lewis, GwilymP., de Lima, HaroldoC., López-Roberts, M.C., Mackinder, B., Maia, V.H., Malécot, V., Mansano, VidalF., Marazzi, B., Mattapha, S., Miller, JosephT., Mitsuyuki, C., Moura, T., Murphy, DanielJ., Nageswara-Rao, M., Nevado, B., Neves, D., Ojeda, DarioI., Pennington, R.T., Prado, DariénE., Prenner, G., de Queiroz, L.P., Ramos, G., Ranzato Filardi, FabianaL., Ribeiro, PétalaG., Rico-Arce, M. de L., Sanderson, MichaelJ., Santos-Silva, J., São-Mateus, WallaceM.B., Silva, MarcosJ.S., Simon, MarceloF., Sinou, C., Snak, C., de Souza, ÉlviaR., Sprent, J., Steele, KellyP., Steier, JuliaE., Steeves, R., Stirton, CharlesH., Tagane, S., Torke, BenjaminM., Toyama, H., Cruz, D.T. da, Vatanparast, M., Wieringa, JanJ., Wink, M., Wojciechowski, MartinF., Yahara, T., Yi, T., Zimmerman, E., 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny – The Legume Phylogeny Working Group (LPWG). taxon 66, 44–77. https://doi.org/10.12705/661.3

Becana, M., Yruela, I., Sarath, G., Catalán, P., Hargrove, M.S., 2020. Plant hemoglobins: a journey from unicellular green algae to vascular plants. New Phytol 227, 1618–1635. https://doi.org/10.1111/nph.16444

Behm, J.E., Geurts, R., Kiers, E.T., 2014. Parasponia: a novel system for studying mutualism stability. Trends in Plant Science 19, 757–763. https://doi.org/10.1016/j.tplants.2014.08.007

Bellande, K., Bono, J.-J., Savelli, B., Jamet, E., Canut, H., 2017. Plant Lectins and Lectin Receptor-Like Kinases: How Do They Sense the Outside? IJMS 18, 1164. https://doi.org/10.3390/ijms18061164

Bensmihen, S., de Billy, F., Gough, C., 2011. Contribution of NFP LysM Domains to the Recognition of Nod Factors during the Medicago truncatula/Sinorhizobium meliloti Symbiosis. PLoS ONE 6, e26114. https://doi.org/10.1371/journal.pone.0026114

Bergman, B., Johansson, C., Soderback, E., 1992. The Nostoc-Gunnera symbiosis. New Phytol 122, 379–400. https://doi.org/10.1111/j.1469-8137.1992.tb00067.x

Berrabah, F., Bourcy, M., Eschstruth, A., Cayrel, A., Guefrachi, I., Mergaert, P., Wen, J., Jean, V., Mysore, K.S., Gourion, B., Ratet, P., 2014. A nonRD receptor-like kinase prevents nodule early senescence and defense-like reactions during symbiosis. New Phytol 203, 1305–1314. https://doi.org/10.1111/nph.12881

Bhattacharya, S., Sen, A., Thakur, S., Tisa, L.S., 2013. Characterization of haemoglobin from Actinorhizal plants – An in silico approach. J Biosci 38, 777–787. https://doi.org/10.1007/s12038-013-9357-0

Boisson-Dernier, A., Andriankaja, A., Chabaud, M., Niebel, A., Journet, E.-P., Barker, D.G., de Carvalho-Niebel, F., 2005. *MtENOD11* Gene Activation During Rhizobial Infection and Mycorrhizal Arbuscule Development Requires a Common AT-Rich-Containing Regulatory Sequence. MPMI 18, 1269–1276. https://doi.org/10.1094/MPMI-18-1269

Boivin, C., Ndoye, I., Molouba, F., de Lajudie, P., Dupuy, N., Dreyfus, B., de Bruijn, F.J., 1997. Stem Nodulation in Legumes: Diversity, Mechanisms, and Unusual Characteristics. Critical Reviews in Plant Sciences 16, 1–30. https://doi.org/10.1080/07352689709701944

Bonaldi, K., Gargani, D., Prin, Y., Fardoux, J., Gully, D., Nouwen, N., Goormachtig, S., 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. MPMI 24, 1359–1371. https://doi.org/10.1094/MPMI-04-11-0093

Bonaldi, K., Gourion, B., Fardoux, J., Hannibal, L., Cartieaux, F., Boursot, M., Vallenet, D., Chaintreuil, C., Prin, Y., Nouwen, N., 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*. MPMI 23, 760–770. https://doi.org/10.1094/MPMI-23-6-0760

Boogerd, F.C., van Rossum, D., 1997. Nodulation of groundnut by *Bradyrhizobium* : a simple infection process by crack entry. FEMS Microbiol Rev 21, 5–27. https://doi.org/10.1111/j.1574-6976.1997.tb00342.x

Bourcy, M., Berrabah, F., Ratet, P., Gourion, B., 2013. Failure of self-control: Defense-like reactions during legume/rhizobia symbiosis. Plant Signaling & Behavior 8, e23915. https://doi.org/10.4161/psb.23915

Bourdais, G., Burdiak, P., Gauthier, A., Nitsch, L., Salojärvi, J., Rayapuram, C., Idänheimo, N., Hunter, K., Kimura, S., Merilo, E., Vaattovaara, A., Oracz, K., Kaufholdt, D., Pallon, A., Anggoro, D.T., Glów, D., Lowe, J., Zhou, J., Mohammadi, O., Puukko, T., Albert, A., Lang, H., Ernst, D., Kollist, H., Brosché, M., Durner, J., Borst, J.W., Collinge, D.B., Karpiński, S., Lyngkjær, M.F., Robatzek, S., Wrzaczek, M., Kangasjärvi, J., on behalf of the CRK Consortium, 2015. Large-Scale Phenomics Identifies Primary and Fine-Tuning Roles for CRKs in Responses Related to Oxidative Stress. PLoS Genet 11, e1005373. https://doi.org/10.1371/journal.pgen.1005373

Bravo, A., York, T., Pumplin, N., Mueller, L.A., Harrison, M.J., 2016. Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. Nature Plants 2, 15208. https://doi.org/10.1038/nplants.2015.208

Brewin, N.J., 2004. Plant Cell Wall Remodelling in the Rhizobium–Legume Symbiosis. Critical Reviews in Plant Sciences 23, 293–316. https://doi.org/10.1080/07352680490480734

Broghammer, A., Krusell, L., Blaise, M., Sauer, J., Sullivan, J.T., Maolanon, N., Vinther, M., Lorentzen, A., Madsen, E.B., Jensen, K.J., Roepstorff, P., Thirup, S., Ronson, C.W., Thygesen, M.B., Stougaard, J., 2012. Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. Proceedings of the National Academy of Sciences 109, 13859–13864. https://doi.org/10.1073/pnas.1205171109

Brottier, L., Chaintreuil, C., Simion, P., Scornavacca, C., Rivallan, R., Mournet, P., Moulin, L., Lewis, G.P., Fardoux, J., Brown, S.C., Gomez-Pacheco, M., Bourges, M., Hervouet, C., Gueye, M., Duponnois, R., Ramanankierana, H., Randriambanona, H., Vandrot, H., Zabaleta, M., DasGupta, M., D'Hont, A., Giraud, E., Arrighi, J.-F., 2018. A phylogenetic framework of the legume genus Aeschynomene for comparative genetic analysis of the Nod-dependent and Nod-independent symbioses. BMC Plant Biol 18, 333. https://doi.org/10.1186/s12870-018-1567-z

Bu, F., Rutten, L., Roswanjaya, Y.P., Kulikova, O., Rodriguez-Franco, M., Ott, T., Bisseling, T., Zeijl, A., Geurts, R., 2020. Mutant analysis in the nonlegume *Parasponia andersonii* identifies NIN and NF-YA1 transcription factors as a core genetic network in nitrogen-fixing nodule symbioses. New Phytol 226, 541–554. https://doi.org/10.1111/nph.16386

Buendia, L., Girardin, A., Wang, T., Cottret, L., Lefebvre, B., 2018. LysM Receptor-Like Kinase and LysM Receptor-Like Protein Families: An Update on Phylogeny and Functional Characterization. Front. Plant Sci. 9, 1531. https://doi.org/10.3389/fpls.2018.01531

Buendia, L., Wang, T., Girardin, A., Lefebvre, B., 2016. The LysM receptor-like kinase SlLYK10 regulates the arbuscular mycorrhizal symbiosis in tomato. New Phytol 210, 184–195. https://doi.org/10.1111/nph.13753

Buist, G., Steen, A., Kok, J., Kuipers, O.P., 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol 68, 838–847. https://doi.org/10.1111/j.1365-2958.2008.06211.x

Busset, N., De Felice, A., Chaintreuil, C., Gully, D., Fardoux, J., Romdhane, S., Molinaro, A., Silipo, A., Giraud, E., 2016. The LPS O-Antigen in Photosynthetic Bradyrhizobium Strains Is Dispensable for the Establishment of a Successful Symbiosis with Aeschynomene Legumes. PLoS ONE 11, e0148884. https://doi.org/10.1371/journal.pone.0148884

Cannon, S.B., McKain, M.R., Harkess, A., Nelson, M.N., Dash, S., Deyholos, M.K., Peng, Y., Joyce, B., Stewart, C.N., Rolf, M., Kutchan, T., Tan, X., Chen, C., Zhang, Y., Carpenter, E., Wong, G.K.-S., Doyle, J.J., Leebens-Mack, J., 2015. Multiple Polyploidy Events in the Early Radiation of Nodulating and Nonnodulating Legumes. Molecular Biology and Evolution 32, 193–210. https://doi.org/10.1093/molbev/msu296

Capoen, W., Oldroyd, G., Goormachtig, S., Holsters, M., 2010. Sesbania rostrata: a case study of natural variation in legume nodulation. New Phytologist 186, 340–345. https://doi.org/10.1111/j.1469-8137.2009.03124.x

Catoira, R., n.d. Medicago truncatula root hair curling mutants 12.

Catoira, R., Galera, C., de Billy, F., Penmetsa, R.V., Journet, E.-P., Maillet, F., Rosenberg, C., Cook, D., Gough, C., Dénarié, J., 2000. Four Genes of *Medicago truncatula* Controlling Components of a Nod Factor Transduction Pathway. Plant Cell 12, 1647–1665. https://doi.org/10.1105/tpc.12.9.1647

Cebolla, A., 1999. The mitotic inhibitor ccs52 is required for endoreduplication and ploidy-dependent cell enlargement in plants. The EMBO Journal 18, 4476–4484. https://doi.org/10.1093/emboj/18.16.4476

Cerri, M.R., Frances, L., Kelner, A., Fournier, J., Middleton, P.H., Auriac, M.-C., Mysore, K.S., Wen, J., Erard, M., Barker, D.G., Oldroyd, G.E., de Carvalho-Niebel, F., 2016. The Symbiosis-Related ERN Transcription Factors Act in Concert to Coordinate Rhizobial Host Root Infection. Plant Physiol. pp.00230.2016. https://doi.org/10.1104/pp.16.00230

Cerri, M.R., Frances, L., Laloum, T., Auriac, M.-C., Niebel, A., Oldroyd, G.E.D., Barker, D.G., Fournier, J., de Carvalho-Niebel, F., 2012. *Medicago truncatula* ERN Transcription Factors: Regulatory Interplay with NSP1/NSP2 GRAS Factors and Expression Dynamics throughout Rhizobial Infection. Plant Physiol. 160, 2155–2172. https://doi.org/10.1104/pp.112.203190

Chabaud, M., Gherbi, H., Pirolles, E., Vaissayre, V., Fournier, J., Moukouanga, D., Franche, C., Bogusz, D., Tisa, L.S., Barker, D.G., Svistoonoff, S., 2016. Chitinase-resistant

hydrophilic symbiotic factors secreted by *Frankia* activate both Ca $^{2+}$ spiking and *NIN* gene expression in the actinorhizal plant *Casuarina glauca*. New Phytol 209, 86–93. https://doi.org/10.1111/nph.13732

Chabaud, M., Venard, C., Defaux-Petras, A., Becard, G., Barker, D.G., 2002. Targeted inoculation of Medicago truncatula in vitro root cultures reveals MtENOD11 expression during early stages of infection by arbuscular mycorrhizal fungi. New Phytol 156, 265–273. https://doi.org/10.1046/j.1469-8137.2002.00508.x

Chaintreuil, C., Arrighi, J.-F., Giraud, E., Miché, L., Moulin, L., Dreyfus, B., Munive-Hernández, J.-A., Villegas-Hernandez, M. del C., Béna, G., 2013. Evolution of symbiosis in the legume genus *Aeschynomene*. New Phytol 200, 1247–1259. https://doi.org/10.1111/nph.12424

Chaintreuil, C., Perrier, X., Martin, G., Fardoux, J., Lewis, G.P., Brottier, L., Rivallan, R., Gomez-Pacheco, M., Bourges, M., Lamy, L., Thibaud, B., Ramanankierana, H., Randriambanona, H., Vandrot, H., Mournet, P., Giraud, E., Arrighi, J.-F., 2018. Naturally occurring variations in the nod-independent model legume Aeschynomene evenia and relatives: a resource for nodulation genetics. BMC Plant Biol 18, 54. https://doi.org/10.1186/s12870-018-1260-2

Chaintreuil, C., Rivallan, R., Bertioli, D.J., Klopp, C., Gouzy, J., Courtois, B., Leleux, P., Martin, G., Rami, J.-F., Gully, D., Parrinello, H., Séverac, D., Patrel, D., Fardoux, J., Ribière, W., Boursot, M., Cartieaux, F., Czernic, P., Ratet, P., Mournet, P., Giraud, E., Arrighi, J.-F., 2016. A gene-based map of the Nod factor-independent *Aeschynomene evenia* sheds new light on the evolution of nodulation and legume genomes. DNA Res 23, 365–376. https://doi.org/10.1093/dnares/dsw020

Charon, C., Johansson, C., Kondorosi, E., Kondorosi, A., Crespi, M., 1997. enod40 induces dedifferentiation and division of root cortical cells in legumes. Proceedings of the National Academy of Sciences 94, 8901–8906. https://doi.org/10.1073/pnas.94.16.8901

Charpentier, M., Bredemeier, R., Wanner, G., Takeda, N., Schleiff, E., Parniske, M., 2008. *Lotus japonicus* CASTOR and POLLUX Are Ion Channels Essential for Perinuclear Calcium Spiking in Legume Root Endosymbiosis. Plant Cell 20, 3467–3479. https://doi.org/10.1105/tpc.108.063255

Charpentier, M., Sun, J., Martins, T.V., Radhakrishnan, G.V., Findlay, K., Soumpourou, E., Thouin, J., Very, A.-A., Sanders, D., Morris, R.J., Oldroyd, G.E.D., 2016. Nuclear-localized cyclic nucleotide-gated channels mediate symbiotic calcium oscillations. Science 352, 1102–1105. https://doi.org/10.1126/science.aae0109

Chen, T., Zhu, H., Ke, D., Cai, K., Wang, C., Gou, H., Hong, Z., Zhang, Z., 2012. A MAP Kinase Kinase Interacts with SymRK and Regulates Nodule Organogenesis in *Lotus japonicus*. Plant Cell 24, 823–838. https://doi.org/10.1105/tpc.112.095984

Cissoko, M., Hocher, V., Gherbi, H., Gully, D., Carré-Mlouka, A., Sane, S., Pignoly, S., Champion, A., Ngom, M., Pujic, P., Fournier, P., Gtari, M., Swanson, E., Pesce, C., Tisa, L.S., Sy, M.O., Svistoonoff, S., 2018. Actinorhizal Signaling Molecules: Frankia Root Hair Deforming Factor Shares Properties With NIN Inducing Factor. Front. Plant Sci. 9, 1494. https://doi.org/10.3389/fpls.2018.01494

Clavijo, F., Diedhiou, I., Vaissayre, V., Brottier, L., Acolatse, J., Moukouanga, D., Crabos, A., Auguy, F., Franche, C., Gherbi, H., Champion, A., Hocher, V., Barker, D., Bogusz, D., Tisa, L.S., Svistoonoff, S., 2015. The *Casuarina NIN* gene is transcriptionally activated

throughout *Frankia* root infection as well as in response to bacterial diffusible signals. New Phytol 208, 887–903. https://doi.org/10.1111/nph.13506

Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J., Udvardi, M.K., 2004. Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. The Plant Journal 39, 487–512. https://doi.org/10.1111/j.1365-313X.2004.02150.x

Cook, D.R., 1999. Medicago truncatula — a model in the making! Current Opinion in Plant Biology 2, 301–304. https://doi.org/10.1016/S1369-5266(99)80053-3

Couzigou, J.-M., Zhukov, V., Mondy, S., Abu el Heba, G., Cosson, V., Ellis, T.H.N., Ambrose, M., Wen, J., Tadege, M., Tikhonovich, I., Mysore, K.S., Putterill, J., Hofer, J., Borisov, A.Y., Ratet, P., 2012. *NODULE ROOT* and *COCHLEATA* Maintain Nodule Development and Are Legume Orthologs of *Arabidopsis BLADE-ON-PETIOLE* Genes. Plant Cell 24, 4498–4510. https://doi.org/10.1105/tpc.112.103747

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., Barrière, Q., Arrighi, J.-F., Mergaert, P., Giraud, E., 2015. Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides. Plant Physiol. 169, 1254–1265. https://doi.org/10.1104/pp.15.00584

Das, D.R., Horváth, B., Kundu, A., Kaló, P., DasGupta, M., 2019. Functional conservation of CYCLOPS in crack entry legume Arachis hypogaea. Plant Science 281, 232–241. https://doi.org/10.1016/j.plantsci.2018.12.003

de Billy, F., Grosjean, C., May, S., Bennett, M., Cullimore, J.V., 2001. Expression Studies on *AUX1* -like Genes in *Medicago truncatula* Suggest That Auxin Is Required at Two Steps in Early Nodule Development. MPMI 14, 267–277. https://doi.org/10.1094/MPMI.2001.14.3.267

De Mita, S., Streng, A., Bisseling, T., Geurts, R., 2014. Evolution of a symbiotic receptor through gene duplications in the legume-rhizobium mutualism. New Phytol 201, 961–972. https://doi.org/10.1111/nph.12549

Debellé, F., Rosenberg, C., Vasse, J., Maillet, F., Martinez, E., Dénarié, J., Truchet, G., 1986. Assignment of symbiotic developmental phenotypes to common and specific nodulation (nod) genetic loci of Rhizobium meliloti. Journal of Bacteriology 168, 1075–1086. https://doi.org/10.1128/JB.168.3.1075-1086.1986

Del Pero Martínez, M.A., Martínez, A.J., 1993. Flavonoid distribution in Tradescantia. Biochemical Systematics and Ecology 21, 255–265. https://doi.org/10.1016/0305-1978(93)90043-Q

Delaux, P.-M., Varala, K., Edger, P.P., Coruzzi, G.M., Pires, J.C., Ané, J.-M., 2014. Comparative Phylogenomics Uncovers the Impact of Symbiotic Associations on Host Genome Evolution. PLoS Genet 10, e1004487. https://doi.org/10.1371/journal.pgen.1004487

Demont, N., Debellé, F., Aurelle, H., Dénarié, J., Promé, J.C., 1993. Role of the Rhizobium meliloti nodF and nodE genes in the biosynthesis of lipo-oligosaccharidic nodulation factors. J Biol Chem 268, 20134–20142.

Dénarié, J., Debellé, F., Promé, J.-C., n.d. RHIZOBIUM LIPO-CHITOOLIGOSACCHARIDE NODULATION FACTORS: Signaling Molecules Mediating Recognition and Morphogenesis 35.

D'Haeze, W., De Rycke, R., Mathis, R., Goormachtig, S., Pagnotta, S., Verplancke, C., Capoen, W., Holsters, M., 2003. Reactive oxygen species and ethylene play a positive role in

lateral root base nodulation of a semiaquatic legume. Proceedings of the National Academy of Sciences 100, 11789–11794. https://doi.org/10.1073/pnas.1333899100

Dombrecht, B., Tesfay, M., Verreth, C., Heusdens, C., Nápoles, M., Vanderleyden, J., Michiels, J., 2002. The Rhizobium etli gene iscN is highly expressed in bacteroids and required for nitrogen fixation. Mol Gen Genomics 267, 820–828. https://doi.org/10.1007/s00438-002-0715-0

Downie, J.A., 2005. Legume Haemoglobins: Symbiotic Nitrogen Fixation Needs Bloody Nodules. Current Biology 15, R196–R198. https://doi.org/10.1016/j.cub.2005.03.007

Doyle, J.J., 1998. Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. Trends in Plant Science 3, 473–478. https://doi.org/10.1016/S1360-1385(98)01340-5

Duhoux, E., 1984. Ontogenèse des nodules caulinaires du *Sesbania rostrata* (légumineuses). Can. J. Bot. 62, 982–994. https://doi.org/10.1139/b84-137

Duhoux, E., Diouf, D., Gherbi, H., Franche, C., Ahée, J., Bogusz, D., 1996. Le nodule actinorhizien. Acta Botanica Gallica 143, 593–608. https://doi.org/10.1080/12538078.1996.10515361

Eaglesham, A.R.J., Ellis, J.M., Evans, W.R., Fleischman, D.E., Hungria, M., Hardy, R.W.F., 1990. The first photosynthetic N2-fixing Rhizobium: Characteristics, in: Gresshoff, P.M., Roth, L.E., Stacey, G., Newton, W.E. (Eds.), Nitrogen Fixation. Springer US, Boston, MA, pp. 805–811. https://doi.org/10.1007/978-1-4684-6432-0_69

Escudero, V., Abreu, I., Tejada-Jiménez, M., Rosa-Núñez, E., Quintana, J., Prieto, R.I., Larue, C., Wen, J., Villanova, J., Mysore, K.S., Argüello, J.M., Castillo-Michel, H., Imperial, J., González-Guerrero, M., 2020. *Medicago truncatula* Ferroportin2 mediates iron import into nodule symbiosomes. New Phytol 228, 194–209. https://doi.org/10.1111/nph.16642

Fabre, S., Gully, D., Poitout, A., Patrel, D., Arrighi, J.-F., Giraud, E., Czernic, P., Cartieaux, F., 2015. The Nod factor-independent nodulation in Aeschynomene evenia required the common plant-microbe symbiotic "toolkit." Plant Physiol. pp.01134.2015. https://doi.org/10.1104/pp.15.01134

Falcón, L.I., Carpenter, E.J., Cipriano, F., Bergman, B., Capone, D.G., 2004. N2 Fixation by Unicellular Bacterioplankton from the Atlantic and Pacific Oceans: Phylogeny and In Situ Rates. AEM 70, 765–770. https://doi.org/10.1128/AEM.70.2.765-770.2004

Farnelid, H., Turk-Kubo, K., Muñoz-Marín, M., Zehr, J., 2016. New insights into the ecology of the globally significant uncultured nitrogen-fixing symbiont UCYN-A. Aquat. Microb. Ecol. 77, 125–138. https://doi.org/10.3354/ame01794

Ferguson, B.J., Ross, J.J., Reid, J.B., 2005. Nodulation Phenotypes of Gibberellin and Brassinosteroid Mutants of Pea. Plant Physiol. 138, 2396–2405. https://doi.org/10.1104/pp.105.062414

Fields, S., 2004. Global Nitrogen: Cycling out of Control. Environmental Health Perspectives 112. https://doi.org/10.1289/ehp.112-a556

Fisher, R.F., Long, S.R., 1993. Interactions of NodD at the nod Box: NodD Binds to Two Distinct Sites on the Same Face of the Helix and Induces a Bend in the DNA. Journal of Molecular Biology 233, 336–348. https://doi.org/10.1006/jmbi.1993.1515

Fonouni-Farde, C., Tan, S., Baudin, M., Brault, M., Wen, J., Mysore, K.S., Niebel, A., Frugier, F., Diet, A., 2016. DELLA-mediated gibberellin signalling regulates Nod factor signalling and rhizobial infection. Nat Commun 7, 12636. https://doi.org/10.1038/ncomms12636

Fournier, J., Teillet, A., Chabaud, M., Ivanov, S., Genre, A., Limpens, E., de Carvalho-Niebel, F., Barker, D., 2015.Remodeling of the infection chaber before infection thread formation reveals a two-step mechanism for rhizobial entry into the host legume root hair. Plant Physiology Apr 2015, 167 (4) 1233-1242 ; DOI: 10.1104/pp.114.253302

Freeling, M., 2006. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Research 16, 805–814. https://doi.org/10.1101/gr.3681406

Gaudioso-Pedraza, R., Beck, M., Frances, L., Kirk, P., Ripodas, C., Niebel, A., Oldroyd, G.E.D., Benitez-Alfonso, Y., de Carvalho-Niebel, F., 2018. Callose-Regulated Symplastic Communication Coordinates Symbiotic Root Nodule Development. Current Biology 28, 3562-3577.e6. https://doi.org/10.1016/j.cub.2018.09.031

Gauthier-Coles, C., White, R.G., Mathesius, U., 2019. Nodulating Legumes Are Distinguished by a Sensitivity to Cytokinin in the Root Cortex Leading to Pseudonodule Development. Front. Plant Sci. 9, 1901. https://doi.org/10.3389/fpls.2018.01901

Geurts, R., Xiao, T.T., Reinhold-Hurek, B., 2016. What Does It Take to Evolve A Nitrogen-Fixing Endosymbiosis? Trends in Plant Science 21, 199–208. https://doi.org/10.1016/j.tplants.2016.01.012

Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., Auguy, F., Peret, B., Laplaze, L., Franche, C., Parniske, M., Bogusz, D., 2008a. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiabacteria. Proceedings of the National Academy of Sciences 105, 4928–4932. https://doi.org/10.1073/pnas.0710618105

Gherbi, H., Markmann, K., Svistoonoff, S., Estevan, J., Autran, D., Giczey, G., Auguy, F., Peret, B., Laplaze, L., Franche, C., Parniske, M., Bogusz, D., 2008b. SymRK defines a common genetic basis for plant root endosymbioses with arbuscular mycorrhiza fungi, rhizobia, and Frankiabacteria. Proceedings of the National Academy of Sciences 105, 4928–4932. https://doi.org/10.1073/pnas.0710618105

Ghodhbane-Gtari, F., Nouioui, I., Salem, K., Ktari, A., Montero-Calasanz, M. del C., Tisa, L.S., Klenk, H.-P., Gtari, M., 2014. Nocardia casuarinae sp. nov., an actinobacterial endophyte isolated from root nodules of Casuarina glauca. Antonie van Leeuwenhoek 105, 1099–1106. https://doi.org/10.1007/s10482-014-0168-6

Gibelin-Viala, C., Amblard, E., Puech-Pages, V., Bonhomme, M., Garcia, M., Bascaules-Bedin, A., Fliegmann, J., Wen, J., Mysore, K.S., Signor, C., Jacquet, C., Gough, C., 2019. The *Medicago truncatula* LysM receptor-like kinase LYK9 plays a dual role in immunity and the arbuscular mycorrhizal symbiosis. New Phytol 223, 1516–1529. https://doi.org/10.1111/nph.15891

Gibson, K.E., Kobayashi, H., Walker, G.C., 2008. Molecular Determinants of a Symbiotic Chronic Infection. Annu. Rev. Genet. 42, 413–441. https://doi.org/10.1146/annurev.genet.42.110807.091427

Girardin, A., Wang, T., Ding, Y., Keller, J., Buendia, L., Gaston, M., Ribeyre, C., Gasciolli, V., Auriac, M.-C., Vernié, T., Bendahmane, A., Ried, M.K., Parniske, M., Morel, P., Vandenbussche, M., Schorderet, M., Reinhardt, D., Delaux, P.-M., Bono, J.-J., Lefebvre, B., 2019. LCO Receptors Involved in Arbuscular Mycorrhiza Are Functional for Rhizobia Perception in Legumes. Current Biology 29, 4249-4259.e5. https://doi.org/10.1016/j.cub.2019.11.038

Giraud, E., Fleischman, D., 2004. Nitrogen-fixing symbiosis between photosynthetic bacteria and legumes. Photosynth Res 82, 115–130. https://doi.org/10.1007/s11120-004-1768-1

Giraud, E., Hannibal, L., Fardoux, J., Vermeglio, A., Dreyfus, B., 2000. Effect of Bradyrhizobium photosynthesis on stem nodulation of Aeschynomene sensitiva. Proceedings of the National Academy of Sciences 97, 14795–14800. https://doi.org/10.1073/pnas.250484097

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., Vermeglio, A., Medigue, C., Sadowsky, M., 2007. Legumes Symbioses: Absence of Nod Genes in Photosynthetic Bradyrhizobia. Science 316, 1307–1312. https://doi.org/10.1126/science.1139548

Gleason, C., Chaudhuri, S., Yang, T., Muñoz, A., Poovaiah, B.W., Oldroyd, G.E.D., 2006. Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. Nature 441, 1149–1152. https://doi.org/10.1038/nature04812

Gobbato, E., 2015. Recent developments in arbuscular mycorrhizal signaling. Current Opinion in Plant Biology 26, 1–7. https://doi.org/10.1016/j.pbi.2015.05.006

Gonzalez-Sama, A., Lucas, M.M., de Felipe, M.R., Pueyo, J.J., 2004. An unusual infection mechanism and nodule morphogenesis in white lupin (Lupinus albus). New Phytol 163, 371–380. https://doi.org/10.1111/j.1469-8137.2004.01121.x

Goormachtig, S., Capoen, W., Holsters, M., 2004. infection: lessons from the versatile nodulation behaviour of water-tolerant legumes. Trends in Plant Science 9, 518–522. https://doi.org/10.1016/j.tplants.2004.09.005

Gough, C., Cottret, L., Lefebvre, B., Bono, J.-J., 2018. Evolutionary History of Plant LysM Receptor Proteins Related to Root Endosymbiosis. Front. Plant Sci. 9, 923. https://doi.org/10.3389/fpls.2018.00923

Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M.B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., Roswanjaya, Y.P., Kohlen, W., Pujic, P., Battenberg, K., Alloisio, N., Liang, Y., Hilhorst, H., Salgado, M.G., Hocher, V., Gherbi, H., Svistoonoff, S., Doyle, J.J., He, S., Xu, Y., Xu, S., Qu, J., Gao, Q., Fang, X., Fu, Y., Normand, P., Berry, A.M., Wall, L.G., Ané, J.-M., Pawlowski, K., Xu, X., Yang, H., Spannagl, M., Mayer, K.F.X., Wong, G.K.-S., Parniske, M., Delaux, P.-M., Cheng, S., 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science eaat1743. https://doi.org/10.1126/science.aat1743

Grilli Caiola, M., 1980. ON THE PHYCOBIONTS OF THE CYCAD CORALLOID ROOTS. New Phytol 85, 537–544. https://doi.org/10.1111/j.1469-8137.1980.tb00769.x

Gruber, N., Galloway, J.N., 2008. An Earth-system perspective of the global nitrogen cycle. Nature 451, 293–296. https://doi.org/10.1038/nature06592

Guillotin, B., Couzigou, J.-M., Combier, J.-P., 2016. NIN Is Involved in the Regulation of Arbuscular Mycorrhizal Symbiosis. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.01704

Gully, D., Czernic, P., Cruveiller, S., Mahé, F., Longin, C., Vallenet, D., François, P., Nidelet, S., Rialle, S., Giraud, E., Arrighi, J.-F., DasGupta, M., Cartieaux, F., 2018. Transcriptome Profiles of Nod Factor-independent Symbiosis in the Tropical Legume Aeschynomene evenia. Sci Rep 8, 10934. https://doi.org/10.1038/s41598-018-29301-0

Handa, Y., Nishide, H., Takeda, N., Suzuki, Y., Kawaguchi, M., Saito, K., 2015. RNA-seq Transcriptional Profiling of an Arbuscular Mycorrhiza Provides Insights into Regulated and Coordinated Gene Expression in *Lotus japonicus* and *Rhizophagus irregularis*. Plant Cell Physiol 56, 1490–1511. https://doi.org/10.1093/pcp/pcv071

Handberg, K., Stougaard, J., 1992. Lotus japonicus, an autogamous, diploid legume species for classical and molecular genetics. Plant J 2, 487–496. https://doi.org/10.1111/j.1365-313X.1992.00487.x

Haney, C.H., Long, S.R., 2010. Plant flotillins are required for infection by nitrogen-fixing bacteria. Proceedings of the National Academy of Sciences 107, 478–483. https://doi.org/10.1073/pnas.0910081107

Haney, C.H., Riely, B.K., Tricoli, D.M., Cook, D.R., Ehrhardt, D.W., Long, S.R., 2011. Symbiotic Rhizobia Bacteria Trigger a Change in Localization and Dynamics of the *Medicago truncatula* Receptor Kinase LYK3. Plant Cell 23, 2774–2787. https://doi.org/10.1105/tpc.111.086389

Hause, B., Fester, T., 2005. Molecular and cell biology of arbuscular mycorrhizal symbiosis. Planta 221, 184–196. https://doi.org/10.1007/s00425-004-1436-x

He, J., Zhang, C., Dai, H., Liu, H., Zhang, X., Yang, J., Chen, X., Zhu, Y., Wang, D., Qi, X., Li, W., Wang, Z., An, G., Yu, N., He, Z., Wang, Y.-F., Xiao, Y., Zhang, P., Wang, E., 2019. A LysM Receptor Heteromer Mediates Perception of Arbuscular Mycorrhizal Symbiotic Signal in Rice. Molecular Plant 12, 1561–1576. https://doi.org/10.1016/j.molp.2019.10.015

He, Y., Zhou, J., Shan, L., Meng, X., 2018. Plant cell surface receptor-mediated signaling – a common theme amid diversity. J Cell Sci 131, jcs209353. https://doi.org/10.1242/jcs.209353

Heckmann, A.B., Sandal, N., Bek, A.S., Madsen, L.H., Jurkiewicz, A., Nielsen, M.W., Tirichine, L., Stougaard, J., 2011. Cytokinin Induction of Root Nodule Primordia in *Lotus japonicus* Is Regulated by a Mechanism Operating in the Root Cortex. MPMI 24, 1385–1395. https://doi.org/10.1094/MPMI-05-11-0142

Hirsch, S., Kim, J., Muñoz, A., Heckmann, A.B., Downie, J.A., Oldroyd, G.E.D., 2009. GRAS Proteins Form a DNA Binding Complex to Induce Gene Expression during Nodulation Signaling in *Medicago truncatula*. Plant Cell 21, 545–557. https://doi.org/10.1105/tpc.108.064501

Hocher, V., Alloisio, N., Auguy, F., Fournier, P., Doumas, P., Pujic, P., Gherbi, H., Queiroux, C., Da Silva, C., Wincker, P., Normand, P., Bogusz, D., 2011. Transcriptomics of Actinorhizal Symbioses Reveals Homologs of the Whole Common Symbiotic Signaling Cascade. Plant Physiol. 156, 700–711. https://doi.org/10.1104/pp.111.174151

Honma, M.A., Asomaning, M., Ausubel, F.M., 1990. Rhizobium meliloti nodD genes mediate host-specific activation of nodABC. Journal of Bacteriology 172, 901–911. https://doi.org/10.1128/JB.172.2.901-911.1990

Horváth, B., Yeun, L.H., Domonkos, Á., Halász, G., Gobbato, E., Ayaydin, F., Miró, K., Hirsch, S., Sun, J., Tadege, M., Ratet, P., Mysore, K.S., Ané, J.-M., Oldroyd, G.E.D., Kaló, P., 2011. *Medicago truncatula IPD3* Is a Member of the Common Symbiotic Signaling Pathway Required for Rhizobial and Mycorrhizal Symbioses. MPMI 24, 1345–1358. https://doi.org/10.1094/MPMI-01-11-0015

Hwang, J.H., Ellingson, S.R., Roberts, D.M., 2010. Ammonia permeability of the soybean nodulin 26 channel. FEBS Letters 584, 4339–4343. https://doi.org/10.1016/j.febslet.2010.09.033

Ivanov, S., n.d. The formation of endosymbiotic membrane compartments: membrane identity merkers and the regulation of vesicle trafficking.

Jaffré, T., Rigault, F., Sarrailh, J.M., n.d. La végétalisation des anciens sites miniers 14.

Jaftha, J.B., Strijdom, B.W., Steyn, P.L., 2002. Characterization of Pigmented Methylotrophic Bacteria which Nodulate Lotononis bainesii. Systematic and Applied Microbiology 25, 440–449. https://doi.org/10.1078/0723-2020-00124

Jamieson, P.A., Shan, L., He, P., 2018. Plant cell surface molecular cypher: Receptor-like proteins and their roles in immunity and development. Plant Science 274, 242–251. https://doi.org/10.1016/j.plantsci.2018.05.030

Jayaraman, D., Richards, A.L., Westphall, M.S., Coon, J.J., Ané, J.-M., 2017. Identification of the phosphorylation targets of symbiotic receptor-like kinases using a high-throughput multiplexed assay for kinase specificity. Plant J 90, 1196–1207. https://doi.org/10.1111/tpj.13529

Jefferson, R.A., 1987. Assaying chimeric genes in plants: The GUS gene fusion system. Plant Mol Biol Rep 5, 387–405. https://doi.org/10.1007/BF02667740

Jiang, G., Krishnan, A.H., Kim, Y.-W., Wacek, T.J., Krishnan, H.B., 2001. A Functional myo-Inositol Dehydrogenase Gene Is Required for Efficient Nitrogen Fixation and Competitiveness of Sinorhizobium fredii USDA191 To Nodulate Soybean (Glycine max [L.] Merr.). J. Bacteriol. 183, 2595–2604. https://doi.org/10.1128/JB.183.8.2595-2604.2001

Jiménez-Guerrero, I., Pérez-Montaño, F., Monreal, J.A., Preston, G.M., Fones, H., Vioque, B., Ollero, F.J., López-Baena, F.J., 2015. The *Sinorhizobium* (*Ensifer*) *fredii* HH103 Type 3 Secretion System Suppresses Early Defense Responses to Effectively Nodulate Soybean. MPMI 28, 790–799. https://doi.org/10.1094/MPMI-01-15-0020-R

Jin, Y., Liu, H., Luo, D., Yu, N., Dong, W., Wang, C., Zhang, X., Dai, H., Yang, J., Wang, E., 2016. DELLA proteins are common components of symbiotic rhizobial and mycorrhizal signalling pathways. Nat Commun 7, 12433. https://doi.org/10.1038/ncomms12433

Kalliampakou, K.I., Kouri, E.D., Boleti, H., Pavli, O., Maurousset, L., Udvardi, M.K., Katinakis, P., Lemoine, R., Flemetakis, E., 2011. Cloning and functional characterization of *Lj* PLT4, a plasma membrane xylitol H^+ - symporter from *Lotus japonicus*. Molecular Membrane Biology 28, 1–13. https://doi.org/10.3109/09687688.2010.500626

Kangatharalingam, N., Priscu, J.C., Paerl, H.W., 1992. Heterocyst envelope thickness, heterocyst frequency and nitrogenase activity in Anabaena flos-aquae: influence of exogenous oxygen tension. Journal of General Microbiology 138, 2673–2678. https://doi.org/10.1099/00221287-138-12-2673

Karmakar, K., Kundu, A., Rizvi, A.Z., Dubois, E., Severac, D., Czernic, P., Cartieaux, F., DasGupta, M., 2019. Transcriptomic Analysis With the Progress of Symbiosis in 'Crack-Entry' Legume *Arachis hypogaea* Highlights Its Contrast With 'Infection Thread' Adapted Legumes. MPMI 32, 271–285. https://doi.org/10.1094/MPMI-06-18-0174-R

Kawaharada, Y., James, E.K., Kelly, S., Sandal, N., Stougaard, J., 2017. The Ethylene Responsive Factor Required for Nodulation 1 (ERN1) Transcription Factor Is Required for Infection-Thread Formation in *Lotus japonicus*. MPMI 30, 194–204. https://doi.org/10.1094/MPMI-11-16-0237-R

Kawaharada, Y., Kelly, S., Nielsen, M.W., Hjuler, C.T., Gysel, K., Muszyński, A., Carlson, R.W., Thygesen, M.B., Sandal, N., Asmussen, M.H., Vinther, M., Andersen, S.U., Krusell, L., Thirup, S., Jensen, K.J., Ronson, C.W., Blaise, M., Radutoiu, S., Stougaard, J., 2015.

Receptor-mediated exopolysaccharide perception controls bacterial infection. Nature 523, 308–312. https://doi.org/10.1038/nature14611

Kaye, J.P., Hart, S.C., 1997. Competition for nitrogen between plants and soil microorganisms. Trends in Ecology & Evolution 12, 139–143. https://doi.org/10.1016/S0169-5347(97)01001-X

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, G., Kondorosi, E., Ané, J.-M., 2007. 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase1 Interacts with NORK and Is Crucial for Nodulation in *Medicago truncatula*. Plant Cell 19, 3974–3989. https://doi.org/10.1105/tpc.107.053975

Kim, S., Zeng, W., Bernard, S., Liao, J., Venkateshwaran, M., Ane, J.-M., Jiang, Y., 2019. Ca2+-regulated Ca2+ channels with an RCK gating ring control plant symbiotic associations. Nat Commun 10, 3703. https://doi.org/10.1038/s41467-019-11698-5

Kimura, S., Hunter, K., Vaahtera, L., Tran, H.C., Citterico, M., Vaattovaara, A., Rokka, A., Stolze, S.C., Harzen, A., Meißner, L., Wilkens, M.M.T., Hamann, T., Toyota, M., Nakagami, H., Wrzaczek, M., 2020. CRK2 and C-terminal Phosphorylation of NADPH Oxidase RBOHD Regulate Reactive Oxygen Species Production in Arabidopsis. Plant Cell 32, 1063–1080. https://doi.org/10.1105/tpc.19.00525

Kiss, E., Oláh, B., Kaló, P., Morales, M., Heckmann, A.B., Borbola, A., Lózsa, A., Kontár, K., Middleton, P., Downie, J.A., Oldroyd, G.E.D., Endre, G., 2009. LIN, a Novel Type of U-Box/WD40 Protein, Controls Early Infection by Rhizobia in Legumes. Plant Physiol. 151, 1239–1249. https://doi.org/10.1104/pp.109.143933

Kneip, C., Lockhart, P., Voß, C., Maier, U.-G., 2007. Nitrogen fixation in eukaryotes – New models for symbiosis. BMC Evol Biol 7, 55. https://doi.org/10.1186/1471-2148-7-55

Krusell, L., Krause, K., Ott, T., Desbrosses, G., Krämer, U., Sato, S., Nakamura, Y., Tabata, S., James, E.K., Sandal, N., Stougaard, J., Kawaguchi, M., Miyamoto, A., Suganuma, N., Udvardi, M.K., 2005. The Sulfate Transporter SST1 Is Crucial for Symbiotic Nitrogen Fixation in *Lotus japonicus* Root Nodules. Plant Cell 17, 1625–1636. https://doi.org/10.1105/tpc.104.030106

Kryvoruchko, I.S., Sinharoy, S., Torres-Jerez, I., Sosso, D., Pislariu, C.I., Guan, D., Murray, J., Benedito, V.A., Frommer, W.B., Udvardi, M.K., 2016. MtSWEET11, a Nodule-Specific Sucrose Transporter of *Medicago truncatula*. Plant Physiol. 171, 554–565. https://doi.org/10.1104/pp.15.01910

Kumar, A., Cousins, D.R., Liu, C.-W., Xu, P., Murray, J.D., 2020. Nodule Inception Is Not Required for Arbuscular Mycorrhizal Colonization of Medicago truncatula. Plants 9, 71. https://doi.org/10.3390/plants9010071

Kundu, A., DasGupta, M., 2018. Silencing of Putative Cytokinin Receptor Histidine Kinase1 Inhibits Both Inception and Differentiation of Root Nodules in *Arachis hypogaea*. MPMI 31, 187–199. https://doi.org/10.1094/MPMI-06-17-0144-R

Laffont, C., Rey, T., André, O., Novero, M., Kazmierczak, T., Debellé, F., Bonfante, P., Jacquet, C., Frugier, F., 2015. The CRE1 Cytokinin Pathway Is Differentially Recruited Depending on Medicago truncatula Root Environments and Negatively Regulates Resistance to a Pathogen. PLoS ONE 10, e0116819. https://doi.org/10.1371/journal.pone.0116819

Lal, N.K., Nagalakshmi, U., Hurlburt, N.K., Flores, R., Bak, A., Sone, P., Ma, X., Song, G., Walley, J., Shan, L., He, P., Casteel, C., Fisher, A.J., Dinesh-Kumar, S.P., 2018. The Receptor-like Cytoplasmic Kinase BIK1 Localizes to the Nucleus and Regulates

Defense Hormone Expression during Plant Innate Immunity. Cell Host & Microbe 23, 485-497.e5. https://doi.org/10.1016/j.chom.2018.03.010

Lavin, M., Pennington, R.T., Klitgaard, B.B., Sprent, J.I., de Lima, H.C., Gasson, P.E., 2001. The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. Am. J. Bot. 88, 503–533. https://doi.org/10.2307/2657116

Lee, D.S., Kim, Y.C., Kwon, S.J., Ryu, C.-M., Park, O.K., 2017. The Arabidopsis Cysteine-Rich Receptor-Like Kinase CRK36 Regulates Immunity through Interaction with the Cytoplasmic Kinase BIK1. Front. Plant Sci. 8, 1856. https://doi.org/10.3389/fpls.2017.01856

Lefebvre, B., Timmers, T., Mbengue, M., Moreau, S., Hervé, C., Tóth, K., Bittencourt-Silvestre, J., Klaus, D., Deslandes, L., Godiard, L., Murray, J.D., Udvardi, M.K., Raffaele, S., Mongrand, S., Cullimore, J., Gamas, P., Niebel, A., Ott, T., 2010. A remorin protein interacts with symbiotic receptors and regulates bacterial infection. Proc Natl Acad Sci USA 107, 2343–2348. https://doi.org/10.1073/pnas.0913320107

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J.C., Dénarié, J., 1990. Symbiotic host-specificity of Rhizobium meliloti is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature 344, 781–784. https://doi.org/10.1038/344781a0

Levy, J., 2004. A Putative Ca2+ and Calmodulin-Dependent Protein Kinase Required for Bacterial and Fungal Symbioses. Science 303, 1361–1364. https://doi.org/10.1126/science.1093038

Lewis, G.P. (Ed.), 2005. Legumes of the world. Royal Botanic Gardens, Kew, Richmond, UK.

Li, L., Li, M., Yu, L., Zhou, Z., Liang, X., Liu, Z., Cai, G., Gao, L., Zhang, X., Wang, Y., Chen, S., Zhou, J.-M., 2014. The FLS2-Associated Kinase BIK1 Directly Phosphorylates the NADPH Oxidase RbohD to Control Plant Immunity. Cell Host & Microbe 15, 329–338. https://doi.org/10.1016/j.chom.2014.02.009

Li, X., Zheng, Z., Kong, X., Xu, J., Qiu, L., Sun, J., Reid, D., Jin, H., Andersen, S.U., Oldroyd, G.E.D., Stougaard, J., Downie, J.A., Xie, F., 2020. Atypical Receptor Kinase RINRK1 Required for Rhizobial Infection But Not Nodule Development in Lotus japonicus1[OPEN] 181, 13.

Liang, J., Klingl, A., Lin, Y.-Y., Boul, E., Thomas-Oates, J., Marín, M., 2019. A subcompatible rhizobium strain reveals infection duality in *Lotus*. Journal of Experimental Botany 70, 1903–1913. https://doi.org/10.1093/jxb/erz057

Liang, P., Stratil, T.F., Popp, C., Marín, M., Folgmann, J., Mysore, K.S., Wen, J., Ott, T., 2018. Symbiotic root infections in *Medicago truncatula* require remorin-mediated receptor stabilization in membrane nanodomains. Proc Natl Acad Sci USA 115, 5289–5294. https://doi.org/10.1073/pnas.1721868115

Liang, X., Zhou, J.-M., 2018. Receptor-Like Cytoplasmic Kinases: Central Players in Plant Receptor Kinase–Mediated Signaling. Annu. Rev. Plant Biol. 69, 267–299. https://doi.org/10.1146/annurev-arplant-042817-040540

Liang, Y., Cao, Y., Tanaka, K., Thibivilliers, S., Wan, J., Choi, J., Kang, C. h., Qiu, J., Stacey, G., 2013. Nonlegumes Respond to Rhizobial Nod Factors by Suppressing the Innate Immune Response. Science 341, 1384–1387. https://doi.org/10.1126/science.1242736

Lievens, S., Goormachtig, S., Den Herder, J., Capoen, W., Mathis, R., Hedden, P., Holsters, M., 2005. Gibberellins Are Involved in Nodulation of *Sesbania rostrata*. Plant Physiol. 139, 1366–1379. https://doi.org/10.1104/pp.105.066944

Limpens, E., 2003. LysM Domain Receptor Kinases Regulating Rhizobial Nod Factor-Induced Infection. Science 302, 630–633. https://doi.org/10.1126/science.1090074

Limpens, E., Moling, S., Hooiveld, G., Pereira, P.A., Bisseling, T., Becker, J.D., Küster, H., 2013. Cell- and Tissue-Specific Transcriptome Analyses of Medicago truncatula Root Nodules. PLoS ONE 8, e64377. https://doi.org/10.1371/journal.pone.0064377

Lin, W., Ma, X., Shan, L., He, P., 2013. Big Roles of Small Kinases: The Complex Functions of Receptor-Like Cytoplasmic Kinases in Plant Immunity and Development: Multiple Functions of RLCKs. J. Integr. Plant Biol. 55, 1188–1197. https://doi.org/10.1111/jipb.12071

Liu, C.-W., Breakspear, A., Guan, D., Cerri, M.R., Jackson, K., Jiang, S., Robson, F., Radhakrishnan, G.V., Roy, S., Bone, C., Stacey, N., Rogers, C., Trick, M., Niebel, A., Oldroyd, G.E.D., de Carvalho-Niebel, F., Murray, J.D., 2019a. NIN Acts as a Network Hub Controlling a Growth Module Required for Rhizobial Infection. Plant Physiol. 179, 1704–1722. https://doi.org/10.1104/pp.18.01572

Liu, C.-W., Breakspear, A., Stacey, N., Findlay, K., Nakashima, J., Ramakrishnan, K., Liu, M., Xie, F., Endre, G., de Carvalho-Niebel, F., Oldroyd, G.E.D., Udvardi, M.K., Fournier, J., Murray, J.D., 2019b. A protein complex required for polar growth of rhizobial infection threads. Nat Commun 10, 2848. https://doi.org/10.1038/s41467-019-10029-y

Liu, J., Bisseling, T., 2020. Evolution of NIN and NIN-like Genes in Relation to Nodule Symbiosis. Genes 11, 777. https://doi.org/10.3390/genes11070777

Liu, W., Kohlen, W., Lillo, A., Op den Camp, R., Ivanov, S., Hartog, M., Limpens, E., Jamil, M., Smaczniak, C., Kaufmann, K., Yang, W.-C., Hooiveld, G.J.E.J., Charnikhova, T., Bouwmeester, H.J., Bisseling, T., Geurts, R., 2011. Strigolactone Biosynthesis in *Medicago truncatula* and Rice Requires the Symbiotic GRAS-Type Transcription Factors NSP1 and NSP2. Plant Cell 23, 3853–3865. https://doi.org/10.1105/tpc.111.089771

Lohar, D.P., Sharopova, N., Endre, G., Peñuela, S., Samac, D., Town, C., Silverstein, K.A.T., VandenBosch, K.A., 2006. Transcript Analysis of Early Nodulation Events in *Medicago truncatula*. Plant Physiol. 140, 221–234. https://doi.org/10.1104/pp.105.070326

Lu, D., Wu, S., Gao, X., Zhang, Y., Shan, L., He, P., 2010. A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. Proceedings of the National Academy of Sciences 107, 496–501. https://doi.org/10.1073/pnas.0909705107

Madsen, E.B., Antolín-Llovera, M., Grossmann, C., Ye, J., Vieweg, S., Broghammer, A., Krusell, L., Radutoiu, S., Jensen, O.N., Stougaard, J., Parniske, M., 2011. Autophosphorylation is essential for the in vivo function of the Lotus japonicus Nod factor receptor 1 and receptor-mediated signalling in cooperation with Nod factor receptor 5: NFR1 and NFR5 phosphorylation and signalling. The Plant Journal 65, 404–417. https://doi.org/10.1111/j.1365-313X.2010.04431.x

Madsen, E.B., Madsen, L.H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., 2003. A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. Nature 425, 637–640. https://doi.org/10.1038/nature02045

Madsen, L.H., Tirichine, L., Jurkiewicz, A., Sullivan, J.T., Heckmann, A.B., Bek, A.S., Ronson, C.W., James, E.K., Stougaard, J., 2010. The molecular network governing nodule

organogenesis and infection in the model legume Lotus japonicus. Nat Commun 1, 10. https://doi.org/10.1038/ncomms1009

Maekawa, T., Maekawa-Yoshikawa, M., Takeda, N., Imaizumi-Anraku, H., Murooka, Y., Hayashi, M., 2009. Gibberellin controls the nodulation signaling pathway in *Lotus japonicus*. The Plant Journal 58, 183–194. https://doi.org/10.1111/j.1365-313X.2008.03774.x

Magne, K., George, J., Berbel Tornero, A., Broquet, B., Madueño, F., Andersen, S.U., Ratet, P., 2018. *Lotus japonicus NOOT-BOP-COCH-LIKE1* is essential for nodule, nectary, leaf and flower development. Plant J 94, 880–894. https://doi.org/10.1111/tpj.13905

Maillet, F., Poinsot, V., André, O., Puech-Pagès, V., Haouy, A., Gueunier, M., Cromer, L., Giraudet, D., Formey, D., Niebel, A., Martinez, E.A., Driguez, H., Bécard, G., Dénarié, J., 2011. Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. Nature 469, 58–63. https://doi.org/10.1038/nature09622

Mangano, S., Denita-Juarez, S.P., Marzol, E., Borassi, C., Estevez, J.M., 2018. High Auxin and High Phosphate Impact on RSL2 Expression and ROS-Homeostasis Linked to Root Hair Growth in Arabidopsis thaliana. Front. Plant Sci. 9, 1164. https://doi.org/10.3389/fpls.2018.01164

Marino, D., Andrio, E., Danchin, E.G.J., Oger, E., Gucciardo, S., Lambert, A., Puppo, A., Pauly, N., 2011. A Medicago truncatula NADPH oxidase is involved in symbiotic nodule functioning. New Phytologist 189, 580–592. https://doi.org/10.1111/j.1469-8137.2010.03509.x

Markmann, K., Parniske, M., 2009. Evolution of root endosymbiosis with bacteria: how novel are nodules? Trends in Plant Science 14, 77–86. https://doi.org/10.1016/j.tplants.2008.11.009

Marvel, D.J., Torrey, J.G., Ausubel, F.M., 1987. Rhizobium symbiotic genes required for nodulation of legume and nonlegume hosts. Proceedings of the National Academy of Sciences 84, 1319–1323. https://doi.org/10.1073/pnas.84.5.1319

Masalkar, P., Wallace, I.S., Hwang, J.H., Roberts, D.M., 2010. Interaction of Cytosolic Glutamine Synthetase of Soybean Root Nodules with the C-terminal Domain of the Symbiosome Membrane Nodulin 26 Aquaglyceroporin. J. Biol. Chem. 285, 23880–23888. https://doi.org/10.1074/jbc.M110.135657

Masson-Boivin, C., Giraud, E., Perret, X., Batut, J., 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends in Microbiology 17, 458–466. https://doi.org/10.1016/j.tim.2009.07.004

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.-E., Barloy-Hubler, F., Galibert, F., Kondorosi, A., Kondorosi, E., 2006. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. Proceedings of the National Academy of Sciences 103, 5230–5235. https://doi.org/10.1073/pnas.0600912103

Milligan, A.J., Berman-Frank, I., Gerchman, Y., Dismukes, G.C., Falkowski, P.G., 2007. Light-dependent oxygen consumption in nitrogen-fixing cyanobacteria plays a key role in nitrogenase protection [1]. Journal of Phycology 43, 845–852. https://doi.org/10.1111/j.1529-8817.2007.00395.x

Mimmack, M.L., Hong, G.F., Johnston, A.W.B., 1994. Sequence and regulation of psrA, a gene on the Sym plasmid of Rhizobium leguminosarum biover phaseoli which inhibits transcription of the psi genes. Microbiology 140, 455–461. https://doi.org/10.1099/00221287-140-3-455

Miya, A., Albert, P., Shinya, T., Desaki, Y., Ichimura, K., Shirasu, K., Narusaka, Y., Kawakami, N., Kaku, H., Shibuya, N., 2007. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. Proceedings of the National Academy of Sciences 104, 19613–19618. https://doi.org/10.1073/pnas.0705147104

Miyahara, A., Richens, J., Starker, C., Morieri, G., Smith, L., Long, S., Downie, J.A., Oldroyd, G.E.D., 2010. Conservation in Function of a SCAR/WAVE Component During Infection Thread and Root Hair Growth in *Medicago truncatula*. MPMI 23, 1553–1562. https://doi.org/10.1094/MPMI-06-10-0144

Miyakawa, T., Hatano, K. -i., Miyauchi, Y., Suwa, Y. -i., Sawano, Y., Tanokura, M., 2014. A Secreted Protein with Plant-Specific Cysteine-Rich Motif Functions as a Mannose-Binding Lectin That Exhibits Antifungal Activity. PLANT PHYSIOLOGY 166, 766–778. https://doi.org/10.1104/pp.114.242636

Miyakawa, T., Miyazono, K., Sawano, Y., Hatano, K., Tanokura, M., 2009. Crystal structure of ginkbilobin-2 with homology to the extracellular domain of plant cysteine-rich receptor-like kinases. Proteins 77, 247–251. https://doi.org/10.1002/prot.22494

Miyata, K., Kawaguchi, M., Nakagawa, T., 2013. Two Distinct EIN2 Genes Cooperatively Regulate Ethylene Signaling in Lotus japonicus. Plant and Cell Physiology 54, 1469–1477. https://doi.org/10.1093/pcp/pct095

Mohd-Radzman, N.A., Laffont, C., Ivanovici, A., Patel, N., Reid, D., Stougaard, J., Frugier, F., Imin, N., Djordjevic, M.A., 2016. Different Pathways Act Downstream of the CEP Peptide Receptor CRA2 to Regulate Lateral Root and Nodule Development. Plant Physiol. 171, 2536–2548. https://doi.org/10.1104/pp.16.00113

Moling, S., Pietraszewska-Bogiel, A., Postma, M., Fedorova, E., Hink, M.A., Limpens, E., Gadella, T.W.J., Bisseling, T., 2014. Nod Factor Receptors Form Heteromeric Complexes and Are Essential for Intracellular Infection in *Medicago* Nodules. Plant Cell 26, 4188–4199. https://doi.org/10.1105/tpc.114.129502

Molouba, F., Lorquin, J., Willems, A., Hoste, B., Giraud, E., Dreyfus, B., Gillis, M., de Lajudie, P., Masson-Boivin, C., 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. https://doi.org/10.1128/AEM.65.7.3084-3094.1999

Montiel, J., Downie, J.A., Farkas, A., Bihari, P., Herczeg, R., Bálint, B., Mergaert, P., Kereszt, A., Kondorosi, É., 2017. Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. Proc Natl Acad Sci USA 114, 5041–5046. https://doi.org/10.1073/pnas.1704217114

Montiel, J., Reid, D., Grønbæk, T.H., Benfeldt, C.M., James, E.K., Ott, T., Ditengou, F.A., Nadzieja, M., Kelly, S., Stougaard, J., 2020. Distinct signalling routes mediates intercellular and intracellular rhizobial infection in *Lotus japonicus* (preprint). Plant Biology. https://doi.org/10.1101/2020.05.29.124313

Mulder, L., Lefebvre, B., Cullimore, J., Imberty, A., 2006. LysM domains of Medicago truncatula NFP protein involved in Nod factor perception. Glycosylation state, molecular modeling and docking of chitooligosaccharides and Nod factors. Glycobiology 16, 801–809. https://doi.org/10.1093/glycob/cwl006

Mulligan, J.T., Long, S.R., 1985. Induction of Rhizobium meliloti nodC expression by plant exudate requires nodD. Proceedings of the National Academy of Sciences 82, 6609–6613. https://doi.org/10.1073/pnas.82.19.6609

Muñoz, J.A., Coronado, C., Pérez-Hormaeche, J., Kondorosi, A., Ratet, P., Palomares, A.J., 1998. *MsPG3* , a *Medicago sativa* polygalacturonase gene expressed during the alfalfa– *Rhizobium meliloti* interaction. PNAS 95, 9687–9692. https://doi.org/10.1073/pnas.95.16.9687

Murakami, Y., Yokoyama, H., Fukui, R., Kawaguchi, M., 2013. Down-Regulation of NSP2 Expression in Developmentally Young Regions of Lotus japonicus Roots in Response to Rhizobial Inoculation. Plant and Cell Physiology 54, 518–527. https://doi.org/10.1093/pcp/pct008

Murray, J.D., Karas, B.J., Sato, S., Tabata, S., Amyot, L., Szczyglowski, K., 2007. A Cytokinin Perception Mutant Colonized by Rhizobium in the Absence of Nodule Organogenesis. Science 315, 101–104. https://doi.org/10.1126/science.1132514

Murray, J.D., Muni, R.R.D., Torres-Jerez, I., Tang, Y., Allen, S., Andriankaja, M., Li, G., Laxmi, A., Cheng, X., Wen, J., Vaughan, D., Schultze, M., Sun, J., Charpentier, M., Oldroyd, G., Tadege, M., Ratet, P., Mysore, K.S., Chen, R., Udvardi, M.K., 2011. Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of Medicago truncatula: VAPYRIN is required for rhizobial and mycorrhizal symbioses. The Plant Journal 65, 244–252. https://doi.org/10.1111/j.1365-313X.2010.04415.x

Nambiar, P.T.C., Nigam, S.N., Dart, P.J., Gibbons, R.W., 1983. Absence of Root Hairs in Non-Nodulating Groundnut, *Arachis hypogaea* L. J Exp Bot 34, 484–488. https://doi.org/10.1093/jxb/34.4.484

Nash, T.H. (Ed.), 1996. Lichen biology. Cambridge University Press, Cambridge ; New York.

Niemietz, C.M., Tyerman, S.D., 2000. Channel-mediated permeation of ammonia gas through the peribacteroid membrane of soybean nodules. FEBS Letters 465, 110–114. https://doi.org/10.1016/S0014-5793(99)01729-9

Nishimura, R., Hayashi, M., Wu, G.-J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., Kawasaki, S., Akao, S., Ohmori, M., Nagasawa, M., Harada, K., Kawaguchi, M., 2002. HAR1 mediates systemic regulation of symbiotic organ development. Nature 420, 426–429. https://doi.org/10.1038/nature01231

Okamoto, S., Shinohara, H., Mori, T., Matsubayashi, Y., Kawaguchi, M., 2013. Root-derived CLE glycopeptides control nodulation by direct binding to HAR1 receptor kinase. Nat Commun 4, 2191. https://doi.org/10.1038/ncomms3191

Okazaki, S., Kaneko, T., Sato, S., Saeki, K., 2013. Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. Proceedings of the National Academy of Sciences 110, 17131–17136. https://doi.org/10.1073/pnas.1302360110

Okazaki, S., Tittabutr, P., Thouin, J., Fardoux, J., Chaintreuil, C., Gully, D., Arrighi, J.-F., Furuta, N., Miwa, H., Yasuda, M., Nouwen, N., Teaumroong, N., Giraud, E., 2016. Rhizobium–legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. ISME J 10, 64–74. https://doi.org/10.1038/ismej.2015.103

Oono, R., Schmitt, I., Sprent, J.I., Denison, R.F., 2010. Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. New Phytologist 187, 508–520. https://doi.org/10.1111/j.1469-8137.2010.03261.x

Ott, T., van Dongen, J.T., Gu¨nther, C., Krusell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P., Udvardi, M.K., 2005. Symbiotic Leghemoglobins Are Crucial for Nitrogen Fixation in Legume Root Nodules but Not for General Plant Growth and Development. Current Biology 15, 531–535. https://doi.org/10.1016/j.cub.2005.01.042

Pacios-Bras, C., Schlaman, H.R.M., Boot, K., Admiraal, P., Mateos Langerak, J., Stougaard, J., Spaink, H.P., 2003. Auxin distribution in *Lotus japonicus* during root nodule development. Plant Molecular Biology 52, 1169–1180. https://doi.org/10.1023/B:PLAN.0000004308.78057.f5

Park, H.-J., Floss, D.S., Levesque-Tremblay, V., Bravo, A., Harrison, M.J., 2015. Hyphal branching during arbuscule development requires RAM1. Plant Physiol. pp.01155.2015. https://doi.org/10.1104/pp.15.01155

Pawlowski, K., Demchenko, K.N., 2012. The diversity of actinorhizal symbiosis. Protoplasma 249, 967–979. https://doi.org/10.1007/s00709-012-0388-4

Peck, M.C., Fisher, R.F., Long, S.R., 2006. Diverse Flavonoids Stimulate NodD1 Binding to nod Gene Promoters in Sinorhizobium meliloti. JB 188, 5417–5427. https://doi.org/10.1128/JB.00376-06

Peiter, E., Sun, J., Heckmann, A.B., Venkateshwaran, M., Riely, B.K., Otegui, M.S., Edwards, A., Freshour, G., Hahn, M.G., Cook, D.R., Sanders, D., Oldroyd, G.E.D., Downie, J.A., Ané, J.-M., 2007. The *Medicago truncatula* DMI1 Protein Modulates Cytosolic Calcium Signaling. Plant Physiol. 145, 192–203. https://doi.org/10.1104/pp.107.097261

Peng, Z., Chen, H., Tan, L., Shu, H., Varshney, R.K., Zhou, Z., Zhao, Z., Luo, Z., Chitikineni, A., Wang, L., Maku, J., López, Y., Gallo, M., Zhou, H., Wang, J., 2020. Natural Polymorphisms in a Pair of NSP2 Homoeologs Can Cause Loss of Nodulation in Peanut. Journal of Experimental Botany eraa505. https://doi.org/10.1093/jxb/eraa505

Peng, Z., Liu, F., Wang, L., Zhou, H., Paudel, D., Tan, L., Maku, J., Gallo, M., Wang, J., 2017. Transcriptome profiles reveal gene regulation of peanut (Arachis hypogaea L.) nodulation. Sci Rep 7, 40066. https://doi.org/10.1038/srep40066

Penmetsa, R.V., 1997. A Legume Ethylene-Insensitive Mutant Hyperinfected by Its Rhizobial Symbiont. Science 275, 527–530. https://doi.org/10.1126/science.275.5299.527

Penmetsa, R.V., Cook, D.R., 2000. Production and Characterization of Diverse Developmental Mutants of *Medicago truncatula*. Plant Physiol. 123, 1387–1398. https://doi.org/10.1104/pp.123.4.1387

Perret, X., Freiberg, C., Rosenthal, A., Broughton, W.J., Fellay, R., 1999. High-resolution transcriptional analysis of the symbiotic plasmid of Rhizobium sp. NGR234. Mol Microbiol 32, 415–425. https://doi.org/10.1046/j.1365-2958.1999.01361.x

Perzl, M., Muller, P., Poralla, K., Kannenberg, E.L., 1997. Squalene-hopene cyclase from Bradyrhizobium japonicum: cloning, expression, sequence analysis and comparison to other triterpenoid cyclases. Microbiology 143, 1235–1242. https://doi.org/10.1099/00221287-143-4-1235

Peters, G.A., Evans, W.R., Toia, R.E., 1976. *Azolla-Anabaena azollae* Relationship: IV. Photosynthetically Driven, Nitrogenase-catalyzed H_2 Production. Plant Physiol. 58, 119–126. https://doi.org/10.1104/pp.58.2.119

Plet, J., Wasson, A., Ariel, F., Le Signor, C., Baker, D., Mathesius, U., Crespi, M., Frugier, F., 2011. MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in Medicago truncatula: Hormonal interactions in nodulation. The Plant Journal 65, 622–633. https://doi.org/10.1111/j.1365-313X.2010.04447.x

Podlešáková, K., Fardoux, J., Patrel, D., Bonaldi, K., Novák, O., Strnad, M., Giraud, E., Spíchal, L., Nouwen, N., 2013. Rhizobial Synthesized Cytokinins Contribute to But Are Not

Essential for the Symbiotic Interaction Between Photosynthetic Bradyrhizobia and *Aeschynomene* Legumes. MPMI 26, 1232–1238. https://doi.org/10.1094/MPMI-03-13-0076-R

Radutoiu, S., Madsen, L.H., Madsen, E.B., Felle, H.H., Umehara, Y., Grønlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., Stougaard, J., 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature 425, 585–592. https://doi.org/10.1038/nature02039

Rasmussen, S.R., Füchtbauer, W., Novero, M., Volpe, V., Malkov, N., Genre, A., Bonfante, P., Stougaard, J., Radutoiu, S., 2016. Intraradical colonization by arbuscular mycorrhizal fungi triggers induction of a lipochitooligosaccharide receptor. Sci Rep 6, 29733. https://doi.org/10.1038/srep29733

Remigi, P., Zhu, J., Young, J.P.W., Masson-Boivin, C., 2016. Symbiosis within Symbiosis: Evolving Nitrogen-Fixing Legume Symbionts. Trends in Microbiology 24, 63–75. https://doi.org/10.1016/j.tim.2015.10.007

Rey, T., Nars, A., Bonhomme, M., Bottin, A., Huguet, S., Balzergue, S., Jardinaud, M.-F., Bono, J.-J., Cullimore, J., Dumas, B., Gough, C., Jacquet, C., 2013. NFP, a LysM protein controlling Nod factor perception, also intervenes in *Medicago truncatula* resistance to pathogens. New Phytol 198, 875–886. https://doi.org/10.1111/nph.12198

Rikkinen, J., 2015. Cyanolichens. Biodivers Conserv 24, 973–993. https://doi.org/10.1007/s10531-015-0906-8

Robledo, M., Jimenez-Zurdo, J.I., Velazquez, E., Trujillo, M.E., Zurdo-Pineiro, J.L., Ramirez-Bahena, M.H., Ramos, B., Diaz-Minguez, J.M., Dazzo, F., Martinez-Molina, E., Mateos, P.F., 2008. Rhizobium cellulase CelC2 is essential for primary symbiotic infection of legume host roots. Proceedings of the National Academy of Sciences 105, 7064–7069. https://doi.org/10.1073/pnas.0802547105

Rosendahl, L., Dilworth, M.J., Glenn, A.R., 1992. Exchange of Metabolites Across the Peribacteroid Membrane in Pea Root Nodules. Journal of Plant Physiology 139, 635–638. https://doi.org/10.1016/S0176-1617(11)80385-8

Roy, S., Liu, W., Nandety, R.S., Crook, A., Mysore, K.S., Pislariu, C.I., Frugoli, J., Dickstein, R., Udvardi, M.K., 2020. Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. Plant Cell 32, 15–41. https://doi.org/10.1105/tpc.19.00279

Rutten, L., Miyata, K., Roswanjaya, Y.P., Huisman, R., Bu, F., Hartog, M., Linders, S., van Velzen, R., van Zeijl, A., Bisseling, T., Kohlen, W., Geurts, R., 2020. Duplication of Symbiotic Lysin Motif Receptors Predates the Evolution of Nitrogen-Fixing Nodule Symbiosis. Plant Physiol. 184, 1004–1023. https://doi.org/10.1104/pp.19.01420

Sagan, M., Morandi, D., Tarenghi, E., Duc, G., 1995. Selection of nodulation and mycorrhizal mutants in the model plant Medicago truncatula (Gaertn.) after γ-ray mutagenesis. Plant Science 111, 63–71. https://doi.org/10.1016/0168-9452(95)04229-N

Schauser, L., Roussis, A., Stiller, J., Stougaard, J., 1999. A plant regulator controlling development of symbiotic root nodules. Nature 402, 191–195. https://doi.org/10.1038/46058

Schiessl, K., Lilley, J.L.S., Lee, T., Tamvakis, I., Kohlen, W., Bailey, P.C., Thomas, A., Luptak, J., Ramakrishnan, K., Carpenter, M.D., Mysore, K.S., Wen, J., Ahnert, S., Grieneisen, V.A., Oldroyd, G.E.D., 2019. NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in Medicago truncatula. Current Biology 29, 3657-3668.e5. https://doi.org/10.1016/j.cub.2019.09.005

Schubert, K.R., 1986. Products of Biological Nitrogen Fixation in Higher Plants: Synthesis, Transport, and Metabolism. Annu. Rev. Plant. Physiol. 37, 539–574. https://doi.org/10.1146/annurev.pp.37.060186.002543

Schultze, M., Kondorosi, A., 1998. REGULATION OF SYMBIOTIC ROOT NODULE DEVELOPMENT. Annu. Rev. Genet. 32, 33–57. https://doi.org/10.1146/annurev.genet.32.1.33

Shen, D., Xiao, T.T., van Velzen, R., Kulikova, O., Gong, X., Geurts, R., Pawlowski, K., Bisseling, T., 2020. A Homeotic Mutation Changes Legume Nodule Ontogeny into Actinorhizal-Type Ontogeny. Plant Cell 32, 1868–1885. https://doi.org/10.1105/tpc.19.00739

Shiu, S.-H., Karlowski, W.M., Pan, R., Tzeng, Y.-H., Mayer, K.F.X., Li, W.-H., 2004. Comparative Analysis of the Receptor-Like Kinase Family in Arabidopsis and Rice. Plant Cell 16, 1220–1234. https://doi.org/10.1105/tpc.020834

Shu, H., Luo, Z., Peng, Z., Wang, J., 2020. The application of CRISPR/Cas9 in hairy roots to explore the functions of AhNFR1 and AhNFR5 genes during peanut nodulation. BMC Plant Biol 20, 417. https://doi.org/10.1186/s12870-020-02614-x

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., Molinaro, A., 2014. Covalently linked hopanoid-lipid A improves outer-membrane resistance of a Bradyrhizobium symbiont of legumes. Nat Commun 5, 5106. https://doi.org/10.1038/ncomms6106

Simpson, F., Burris, R., 1984. A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase. Science 224, 1095–1097. https://doi.org/10.1126/science.6585956

Singh, S., Katzer, K., Lambert, J., Cerri, M., Parniske, M., 2014. CYCLOPS, A DNA-Binding Transcriptional Activator, Orchestrates Symbiotic Root Nodule Development. Cell Host & Microbe 15, 139–152. https://doi.org/10.1016/j.chom.2014.01.011

Sinharoy, S., DasGupta, M., 2009. RNA Interference Highlights the Role of CCaMK in Dissemination of Endosymbionts in the Aeschynomeneae Legume *Arachis*. MPMI 22, 1466–1475. https://doi.org/10.1094/MPMI-22-11-1466

Sinharoy, S., Liu, C., Breakspear, A., Guan, D., Shailes, S., Nakashima, J., Zhang, S., Wen, J., Torres-Jerez, I., Oldroyd, G., Murray, J.D., Udvardi, M.K., 2016. A *Medicago truncatula* Cystathionine-β-Synthase-like Domain-Containing Protein Is Required for Rhizobial Infection and Symbiotic Nitrogen Fixation. Plant Physiol. 170, 2204–2217. https://doi.org/10.1104/pp.15.01853

Sinharoy, S., Torres-Jerez, I., Bandyopadhyay, K., Kereszt, A., Pislariu, C.I., Nakashima, J., Benedito, V.A., Kondorosi, E., Udvardi, M.K., 2013. The C_2H_2 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. https://doi.org/10.1105/tpc.113.114017

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C., Bisseling, T., 2007. Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling. Plant Physiol. 145, 183–191. https://doi.org/10.1104/pp.107.100495

Smith, S.E., Read, D.J., 2009. Mycorrhizal symbiosis, 3. ed., Repr. ed. Elsevier/Acad. Press, Amsterdam.

Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., Martin, P.G., 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proceedings of the National Academy of Sciences 92, 2647–2651. https://doi.org/10.1073/pnas.92.7.2647

Sørensen, K.K., Simonsen, J.B., Maolanon, N.N., Stougaard, J., Jensen, K.J., 2014. Chemically Synthesized 58-mer LysM Domain Binds Lipochitin Oligosaccharide. ChemBioChem 15, 2097–2105. https://doi.org/10.1002/cbic.201402125

Soyano, T., Hirakawa, H., Sato, S., Hayashi, M., Kawaguchi, M., 2014. NODULE INCEPTION creates a long-distance negative feedback loop involved in homeostatic regulation of nodule organ production. Proc Natl Acad Sci USA 111, 14607–14612. https://doi.org/10.1073/pnas.1412716111

Soyano, T., Kouchi, H., Hirota, A., Hayashi, M., 2013. NODULE INCEPTION Directly Targets NF-Y Subunit Genes to Regulate Essential Processes of Root Nodule Development in Lotus japonicus. PLoS Genet 9, e1003352. https://doi.org/10.1371/journal.pgen.1003352

Sprent, J.I., 2008. 60Ma of legume nodulation. What's new? What's changing? Journal of Experimental Botany 59, 1081–1084. https://doi.org/10.1093/jxb/erm286

Sprent, J.I., 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation: *Tansley review*. New Phytologist 174, 11–25. https://doi.org/10.1111/j.1469-8137.2007.02015.x

Sprent, J.I., Sprent, P., 1990. Nitrogen fixing organisms: pure and applied aspects, 1st ed. ed. Chapman and Hall, London ; New York.

Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K., Parniske, M., 2002. A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417, 959–962. https://doi.org/10.1038/nature00841

Stranges, A.N., 2001. Vaclav Smil. *Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production* . xx + 338 pp., illus., figs., tables, apps., indexes. Cambridge, Mass./London: MIT Press, 2001. $34.95. Isis 93, 329–330. https://doi.org/10.1086/345022

Suzaki, T., Ito, M., Yoro, E., Sato, S., Hirakawa, H., Takeda, N., Kawaguchi, M., 2014. Endoreduplication-mediated initiation of symbiotic organ development in Lotus japonicus. Development 141, 2441–2445. https://doi.org/10.1242/dev.107946

Suzuki, W., Konishi, M., Yanagisawa, S., 2013. The evolutionary events necessary for the emergence of symbiotic nitrogen fixation in legumes may involve a loss of nitrate responsiveness of the NIN transcription factor. Plant Signaling & Behavior 8, e25975. https://doi.org/10.4161/psb.25975

Svistoonoff, S., Benabdoun, F.M., Nambiar-Veetil, M., Imanishi, L., Vaissayre, V., Cesari, S., Diagne, N., Hocher, V., de Billy, F., Bonneau, J., Wall, L., Ykhlef, N., Rosenberg, C., Bogusz, D., Franche, C., Gherbi, H., 2013. The Independent Acquisition of Plant Root Nitrogen-Fixing Symbiosis in Fabids Recruited the Same Genetic Pathway for Nodule Organogenesis. PLoS ONE 8, e64515. https://doi.org/10.1371/journal.pone.0064515

Swensen, S.M., 1996. The evolution of actinorhizal symbioses: Evidence for multiple origins of the symbiotic association. American Journal of Botany 83, 1503–1512. https://doi.org/10.1002/j.1537-2197.1996.tb13943.x

Szczyglowski, K., Shaw, R.S., Wopereis, J., Copeland, S., Hamburger, D., Kasiborski, B., Dazzo, F.B., de Bruijn, F.J., 1998. Nodule Organogenesis and Symbiotic Mutants of the Model

Legume *Lotus japonicus*. MPMI 11, 684–697. https://doi.org/10.1094/MPMI.1998.11.7.684

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P.X., Chabaud, M., Ratet, P., Mysore, K.S., 2008. Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume Medicago truncatula. Plant J 54, 335–347. https://doi.org/10.1111/j.1365-313X.2008.03418.x

Teulet, A., Busset, N., Fardoux, J., Gully, D., Chaintreuil, C., Cartieaux, F., Jauneau, A., Comorge, V., Okazaki, S., Kaneko, T., Gressent, F., Nouwen, N., Arrighi, J.-F., Koebnik, R., Mergaert, P., Deslandes, L., Giraud, E., 2019. The rhizobial type III effector ErnA confers the ability to form nodules in legumes. Proc Natl Acad Sci USA 116, 21758–21768. https://doi.org/10.1073/pnas.1904456116

Thompson, A.W., Foster, R.A., Krupke, A., Carter, B.J., Musat, N., Vaulot, D., Kuypers, M.M.M., Zehr, J.P., 2012. Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga. Science 337, 1546–1550. https://doi.org/10.1126/science.1222700

Tian, L., Liu, L., Yin, Y., Huang, M., Chen, Yanbo, Xu, X., Wu, P., Li, M., Wu, G., Jiang, H., Chen, Yaping, 2017. Heterogeneity in the expression and subcellular localization of POLYOL/MONOSACCHARIDE TRANSPORTER genes in Lotus japonicus. PLoS ONE 12, e0185269. https://doi.org/10.1371/journal.pone.0185269

Tian, W., Hou, C., Ren, Z., Wang, C., Zhao, F., Dahlbeck, D., Hu, S., Zhang, L., Niu, Q., Li, L., Staskawicz, B.J., Luan, S., 2019. A calmodulin-gated calcium channel links pathogen patterns to plant immunity. Nature 572, 131–135. https://doi.org/10.1038/s41586-019-1413-y

Timmers, A.C., Auriac, M.C., Truchet, G., 1999. Refined analysis of early symbiotic steps of the Rhizobium-Medicago interaction in relationship with microtubular cytoskeleton rearrangements. Development 126, 3617–3628.

Tirichine, L., Sandal, N., Madsen, L.H., Radutoiu, S., Albrektsen, A.S., Sato, S., Asamizu, E., Tabata, S., Stougaard, J., 2007. A Gain-of-Function Mutation in a Cytokinin Receptor Triggers Spontaneous Root Nodule Organogenesis. Science 315, 104–107. https://doi.org/10.1126/science.1132397

Uheda, E., Daimon, H., Yoshizako, F., 2001. Colonization and invasion of peanut (*Arachis hypogaea* L.) roots by *gusA*- marked *Bradyrhizobium* sp. Can. J. Bot. 79, 733–738. https://doi.org/10.1139/b01-050

Urbański, D.F., Małolepszy, A., Stougaard, J., Andersen, S.U., 2012. Genome-wide LORE1 retrotransposon mutagenesis and high-throughput insertion detection in Lotus japonicus: FST poolit detection of LORE1 insertions. The Plant Journal 69, 731–741. https://doi.org/10.1111/j.1365-313X.2011.04827.x

Usher, K.M., Bergman, B., Raven, J.A., 2007. Exploring Cyanobacterial Mutualisms. Annu. Rev. Ecol. Evol. Syst. 38, 255–273. https://doi.org/10.1146/annurev.ecolsys.38.091206.095641

Vaattovaara, A., Brandt, B., Rajaraman, S., Safronov, O., Veidenberg, A., Luklová, M., Kangasjärvi, J., Löytynoja, A., Hothorn, M., Salojärvi, J., Wrzaczek, M., 2019. Mechanistic insights into the evolution of DUF26-containing proteins in land plants. Commun Biol 2, 56. https://doi.org/10.1038/s42003-019-0306-9

Van de Velde, W., Guerra, J.C.P., Keyser, A.D., De Rycke, R., Rombauts, S., Maunoury, N., Mergaert, P., Kondorosi, E., Holsters, M., Goormachtig, S., 2006. Aging in Legume Symbiosis. A Molecular View on Nodule Senescence in *Medicago truncatula*. Plant Physiol. 141, 711–720. https://doi.org/10.1104/pp.106.078691

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A., Tiricz, H., Satiat-Jeunemaitre, B., Alunni, B., Bourge, M., Kucho, K. -i., Abe, M., Kereszt, A., Maroti, G., Uchiumi, T., Kondorosi, E., Mergaert, P., 2010. Plant Peptides Govern Terminal Differentiation of Bacteria in Symbiosis. Science 327, 1122–1126. https://doi.org/10.1126/science.1184057

van Velzen, R., Doyle, J.J., Geurts, R., 2019. A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation. Trends in Plant Science 24, 49–57. https://doi.org/10.1016/j.tplants.2018.10.005

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D., Roswanjaya, Y., Wardhani, T.A.K., Kalhor, M.S., Jansen, J., van den Hoogen, J., Güngör, B., Hartog, M., Hontelez, J., Verver, J., Yang, W.-C., Schijlen, E., Repin, R., Schilthuizen, M., Schranz, M.E., Heidstra, R., Miyata, K., Fedorova, E., Kohlen, W., Bisseling, T., Smit, S., Geurts, R., 2018. Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. Proc Natl Acad Sci USA 115, E4700–E4709. https://doi.org/10.1073/pnas.1721395115

van Velzen, R., Holmer, R., Bu, F., Rutten, L., van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D., Roswanjaya, Y.P., Wardhani, T.A.K., Kalhor, M.S., Jansen, J., van den Hoogen, D.J., Güngör, B., Hartog, M., Hontelez, J., Verver, J., Yang, W.-C., Schijlen, E., Repin, R., Schilthuizen, M., Schranz, M.E., Heidstra, R., Miyata, K., Fedorova, E., Kohlen, W., Bisseling, T., Smit, S., Geurts, R., 2017. Parallel loss of symbiosis genes in relatives of nitrogen-fixing non-legume *Parasponia* (preprint). Plant Biology. https://doi.org/10.1101/169706

van Zeijl, A., Liu, W., Xiao, T.T., Kohlen, W., Yang, W.-C., Bisseling, T., Geurts, R., 2015. The strigolactone biosynthesis gene DWARF27 is co-opted in rhizobium symbiosis. BMC Plant Biol 15, 260. https://doi.org/10.1186/s12870-015-0651-x

van Zeijl, A., Wardhani, T.A.K., Seifi Kalhor, M., Rutten, L., Bu, F., Hartog, M., Linders, S., Fedorova, E.E., Bisseling, T., Kohlen, W., Geurts, R., 2018. CRISPR/Cas9-Mediated Mutagenesis of Four Putative Symbiosis Genes of the Tropical Tree Parasponia andersonii Reveals Novel Phenotypes. Front. Plant Sci. 9, 284. https://doi.org/10.3389/fpls.2018.00284

Venkateshwaran, M., Cosme, A., Han, L., Banba, M., Satyshur, K.A., Schleiff, E., Parniske, M., Imaizumi-Anraku, H., Ané, J.-M., 2012. The Recent Evolution of a Symbiotic Ion Channel in the Legume Family Altered Ion Conductance and Improved Functionality in Calcium Signaling. Plant Cell 24, 2528–2545. https://doi.org/10.1105/tpc.112.098475

Venkateshwaran, M., Jayaraman, D., Chabaud, M., Genre, A., Balloon, A.J., Maeda, J., Forshey, K., den Os, D., Kwiecien, N.W., Coon, J.J., Barker, D.G., Ané, J.-M., 2015. A role for the mevalonate pathway in early plant symbiotic signaling. Proc Natl Acad Sci USA 112, 9781–9786. https://doi.org/10.1073/pnas.1413762112

Vij, S., Giri, J., Dansana, P.K., Kapoor, S., Tyagi, A.K., 2008. The Receptor-Like Cytoplasmic Kinase (OsRLCK) Gene Family in Rice: Organization, Phylogenetic Relationship, and Expression during Development and Stress. Molecular Plant 1, 732–750. https://doi.org/10.1093/mp/ssn047

Vinardell, J.M., Fedorova, E., Cebolla, A., Kevei, Z., Horvath, G., Kelemen, Z., Tarayre, S., Roudier, F., Mergaert, P., Kondorosi, A., Kondorosi, E., 2003. Endoreduplication Mediated by the Anaphase-Promoting Complex Activator CCS52A Is Required for

Symbiotic Cell Differentiation in *Medicago truncatula* Nodules. Plant Cell 15, 2093–2105. https://doi.org/10.1105/tpc.014373

Voisin, A.-S., Guéguen, J., Huyghe, C., Jeuffroy, M.-H., Magrini, M.-B., Meynard, J.-M., Mougel, C., Pellerin, S., Pelzer, E., 2014. Legumes for feed, food, biomaterials and bioenergy in Europe: a review. Agron. Sustain. Dev. 34, 361–380. https://doi.org/10.1007/s13593-013-0189-y

Wang, R., He, F., Ning, Y., Wang, G.-L., 2020. Fine-Tuning of RBOH-Mediated ROS Signaling in Plant Immunity. Trends in Plant Science 25, 1060–1062. https://doi.org/10.1016/j.tplants.2020.08.001

Wilf, P., Johnson, K.R., n.d. Land plant extinction at the end of the Cretaceous: a quantitative analysis of the North Dakota megafloral record 23.

Wong, J.E.M.M., Gysel, K., Birkefeldt, T.G., Vinther, M., Muszyński, A., Azadi, P., Laursen, N.S., Sullivan, J.T., Ronson, C.W., Stougaard, J., Andersen, K.R., 2020. Structural signatures in EPR3 define a unique class of plant carbohydrate receptors. Nat Commun 11, 3797. https://doi.org/10.1038/s41467-020-17568-9

Wong, J.E.M.M., Nadzieja, M., Madsen, L.H., Bücherl, C.A., Dam, S., Sandal, N.N., Couto, D., Derbyshire, P., Uldum-Berentsen, M., Schroeder, S., Schwämmle, V., Nogueira, F.C.S., Asmussen, M.H., Thirup, S., Radutoiu, S., Blaise, M., Andersen, K.R., Menke, F.L.H., Zipfel, C., Stougaard, J., 2019. A *Lotus japonicus* cytoplasmic kinase connects Nod factor perception by the NFR5 LysM receptor to nodulation. Proc Natl Acad Sci USA 116, 14339–14348. https://doi.org/10.1073/pnas.1815425116

Wopereis, J., Pajuelo, E., Dazzo, F.B., Jiang, Q., Gresshoff, P.M., de Bruijn, F.J., Stougaard, J., Szczyglowski, K., 2000. Short root mutant of Lotus japonicus with a dramatically altered symbiotic phenotype. Plant J 23, 97–114. https://doi.org/10.1046/j.1365-313x.2000.00799.x

Xie, F., Murray, J.D., Kim, J., Heckmann, A.B., Edwards, A., Oldroyd, G.E.D., Downie, J.A., 2012. Legume pectate lyase required for root infection by rhizobia. Proceedings of the National Academy of Sciences 109, 633–638. https://doi.org/10.1073/pnas.1113992109

Xue, L., Cui, H., Buer, B., Vijayakumar, V., Delaux, P.-M., Junkermann, S., Bucher, M., 2015. Network of GRAS Transcription Factors Involved in the Control of Arbuscule Development in *Lotus japonicus*. Plant Physiol. 167, 854–871. https://doi.org/10.1104/pp.114.255430

Yano, K., Aoki, S., Liu, M., Umehara, Y., Suganuma, N., Iwasaki, W., Sato, S., Soyano, T., Kouchi, H., Kawaguchi, M., 2016. Function and evolution of a *Lotus japonicus* AP2/ERF family transcription factor that is required for development of infection threads. DNA Res dsw052. https://doi.org/10.1093/dnares/dsw052

Yano, K., Yoshida, S., Muller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S., Asamizu, E., Tabata, S., Murooka, Y., Perry, J., Wang, T.L., Kawaguchi, M., Imaizumi-Anraku, H., Hayashi, M., Parniske, M., 2008. CYCLOPS, a mediator of symbiotic intracellular accommodation. Proceedings of the National Academy of Sciences 105, 20540–20545. https://doi.org/10.1073/pnas.0806858105

Yokota, K., Fukai, E., Madsen, L.H., Jurkiewicz, A., Rueda, P., Radutoiu, S., Held, M., Hossain, M.S., Szczyglowski, K., Morieri, G., Oldroyd, G.E.D., Downie, J.A., Nielsen, M.W., Rusek, A.M., Sato, S., Tabata, S., James, E.K., Oyaizu, H., Sandal, N., Stougaard, J., 2009. Rearrangement of Actin Cytoskeleton Mediates Invasion of *Lotus japonicus*

Roots by *Mesorhizobium loti*. Plant Cell 21, 267–284. https://doi.org/10.1105/tpc.108.063693

Yoon, H.J., Hossain, M.S., Held, M., Hou, H., Kehl, M., Tromas, A., Sato, S., Tabata, S., Andersen, S.U., Stougaard, J., Ross, L., Szczyglowski, K., 2014. *Lotus japonicus SUNERGOS 1* encodes a predicted subunit A of a DNA topoisomerase VI that is required for nodule differentiation and accommodation of rhizobial infection. Plant J 78, 811–821. https://doi.org/10.1111/tpj.12520

Young, N.D., Debellé, F., Oldroyd, G.E.D., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F.X., 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.T., 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., Roe, B.A., 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480, 520–524. https://doi.org/10.1038/nature10625

Yu, H., Xiao, A., Dong, R., Fan, Y., Zhang, X., Liu, C., Wang, C., Zhu, H., Duanmu, D., Cao, Y., Zhang, Z., 2018. Suppression of innate immunity mediated by the CDPK-Rboh complex is required for rhizobial colonization in *Medicago truncatula* nodules. New Phytol 220, 425–434. https://doi.org/10.1111/nph.15410

Zhang, X., Dong, W., Sun, J., Feng, F., Deng, Y., He, Z., Oldroyd, G.E.D., Wang, E., 2015. The receptor kinase *CERK1* has dual functions in symbiosis and immunity signalling. Plant J 81, 258–267. https://doi.org/10.1111/tpj.12723

Zhu, H., Chen, T., Zhu, M., Fang, Q., Kang, H., Hong, Z., Zhang, Z., 2008. A Novel ARID DNA-Binding Protein Interacts with SymRK and Is Expressed during Early Nodule Development in *Lotus japonicus*. Plant Physiol. 148, 337–347. https://doi.org/10.1104/pp.108.119164

Zipfel, C., Oldroyd, G.E.D., 2017. Plant signalling in symbiosis and immunity. Nature 543, 328–336. https://doi.org/10.1038/nature22009