

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE



Université Constantine 1
Faculté des Science de la Nature et de la Vie
Département de Biologie Animale



Mémoire présenté en vue de l'obtention du Diplôme de Master
Domaine : Sciences de la Nature et de la Vie
Filière : Microbiologie
Spécialité : Ecologie Microbienne

N° d'ordre :
N° de série :

Intitulé :

Analyses bioinformatique et phylogénétique des gènes *nod*
et *nif* de *Rhizobium sp.*

Présentée et soutenu par : SAIB Rayane Houda

Le : 22/06/2014

HABIBATNI Imene

Jury d'évaluation :

Président du jury : BENHIZIA Y. Professeur Université Constantine 1

Rapporteur : HAMIDECHI M. A. Professeur Université Constantine 1

Examinatrice : BOUZERAIB L. Maitre assistante Université Constantine 1

Année universitaire
2013/2014

Nous exprimons tout d'abord, nos profonds remerciements à DIEU tout puissant, qui nous a guidé sur le droit chemin et nous a donné le courage et la volonté de finir ce mémoire.

REMERCIEMENTS

Au terme de ce travail, nous tenons à exprimer toutes nos profondes gratitude et remerciements à notre encadreur Monsieur HAMIDECHI. M. A. Professeur à l'université Constantine 1 pour avoir accepté la responsabilité de diriger ce travail malgré ces nombreuses obligations. Merci pour ses conseils avisés et pour ses encouragements, ainsi que ses critiques constructives pour nous diriger vers un esprit scientifique dont nous avons pu bénéficier. Merci Monsieur.

Nous tenons à exprimer notre grande considération et nos profonds remerciements aux membres de jury qui nous ont fait l'honneur de juger notre travail.

Nos sincères remerciements et gratitude vont à Monsieur Le Professeur BENCHIZIA Y. pour avoir accepté de présider le jury de ce mémoire.

Nos sincères remerciements vont également à Madame BOUZERAIB L. de nous avoir fait l'honneur d'examiner ce travail.

C'est avec un très grand plaisir que nous remercions infiniment, Monsieur Le Professeur BENGUEDOUAR A., responsable de notre filière. Qu'il trouve ici notre très profonde gratitude.

Nos sentiments de reconnaissance et nos remerciements vont également à l'ensemble de tous nos enseignants qui ont contribué dans notre formation le long de notre parcours pédagogique. Nous spécifions ceux qui nous ont dirigé et encadré principalement en licence et en Master.

A tous, nous dirons MERCI.

DEDICACES

A nos très chers parents, nous n'oublions jamais vos sacrifices exprimés à notre égard,
votre attention corrective et votre dévouement pour notre éducation.

A nos chères sœurs, nos chers frères, leurs enfants et leurs épouses que Dieu les préserve.

A nos chères amies, pour tous les moments que nous avons partagés.

A tous les membres des familles : HABIBATNI et SAIB.

RAYANE HOUDA

IMENE

Résumé

Plusieurs approches sont utilisées pour l'identification et la classification des espèces bactérienne. Cependant, la plupart de ces méthodes présentent certaines insuffisances. L'objectif de ce travail est d'établir une analyse phylogénétique de huit espèces de *Rhizobium*, basée sur quatre familles de gènes (ADNr 16S, House Keeping Genes, *nodC* et *nifH*). L'analyse phylogénétique de ces souches a été réalisée par le logiciel MEGA 4 suivant la méthode de Neighbor-joining (NJ) en utilisant les paramètres de correction de Kimura. La robustesse des phylogrammes à été évaluée à l'aide de la méthode de bootstrap. L'analyse des phylogrammes basée sur le gène *nifH* a montré que celui-ci peut être utilisée avec les autres types de gènes afin d'aboutir à des résultats phylogénétiques plus appropriés.

Mots clés : Phylogénie, alignement multiple, MLST, *nifH*, *nodC*

Abstract

Several approaches are used to identification and classification of kinds or species of bacteria. However, most of these methods are in insufficient. The objective of this work is to establish a phylogenetic analysis of eight species of *Rhizobium*, based on four families of genes (rDNA 16S, House Keeping Genes, *nodC*, *nifH*). The phylogenetic analysis of these stumps was realized by the software MEGA 4.0 following the method of Neighbor-joining (NJ) by using the Kimura correction parameters. The robustness of phylogrammes was evaluated by means of bootstrap method. The analysis of phylogramme based on the *nifH* gene showed that this latter can be used with other types of genes to get more appropriate phylogenetic results.

Key words: Phylogeny, multiple alignment, MLST, *nifH*, *nodC*.

الملخص

العديد من النظريات المستخدمة لتصنيف و تحديد أنواع البكتيريا العضوية، غير أن أغلبية هذه الطرق أظهرت بعض النقائص. الهدف من هذه الدراسة هو وضع تحليل وراثي جيني لثمانية أنواع من *Rhizobium* بالاعتماد على أربع عائلات من الجينات (*ADNr 16S*, *House Keeping Genes*, *nodC*, *nifH*). إن تحليل التاريخ الوراثي لهذه السلالات الثمانية تم باستعمال برنامج MEGA 4.0 بإتباع طريقة Neighbor- joining (NJ)، مع استخدام معايير التصحيح لـ Kimura. كما أن دقة phylogramme قد قيمت بالاستعانة بطريقة bootstrap. التحليل الوراثي القائم على الجين *nifH* اظهر أن هذا الأخير يمكن استخدامه مع جينات أخرى لغرض الحصول على نتائج تحليل وراثي جيني مناسبة.

الكلمات المفتاحية :

فلسفة الجينات، *nifH*، *nodC*، *MLST*، *alignement multiple*.

Liste des figures

Figure 1 : classification des <i>Rhizobium</i> spp	6
Figure 2 : Régulation de l'opéron <i>nod</i> en présence des flavonoïdes	10
Figure 3 : Structure et synthèse du squelette de base des facteurs Nod	13
Figure 4 : la régulation de l'expression des gènes <i>nif</i>	15
Figure 5 : différentes étapes de construction de la phylogénie	19
Figure 6 : Arbre phylogénétique du gène ADNr 16S construit par la méthode NJ	29
Figure 7 : Arbre phylogénétique des gènes ménagers construit par la méthode NJ	30
Figure 8 : Arbre phylogénétique de gène <i>nod C</i> construit par la méthode NJ	31
Figure 9 : arbre phylogénétique de gène <i>nif H</i> construit par la méthode NJ	32

Liste des tableaux

Tableau 1 : caractères biochimiques de genre <i>Rhizobium</i>	4
Tableau 2 : Evolution du nombre d'espèces au sein de différents genres de rhizobia	5
Tableau 3 : Espèces de <i>Rhizobium</i> recensées en 2005	7
Tableau 4 : la fonction biochimique de quelques gènes <i>nod</i> et leur espèces	11
Tableau 5 : les différents types de gènes <i>nif</i> et leurs fonctions	16
Tableau 6 : Qualité des alignements multiples des différentes familles de gènes	28

Liste des annexes

Annexe 1 : Classification du *Rhizobium sp.*

Annexe 2 : Les séquences nucléotidiques (GenBank) des différents gènes de ménage du *Rhizobium spp*

Suite annexe 2 : Les séquences nucléotidiques (GenBank) des différents gènes du *Rhizobium spp*

Annexe 3 : Codes d'accèsion (GenBank) des différents gènes de ménage de *Rhizobium spp*

Annexe 4 : Le code des nucléotides selon UIB

Annexe 5 : Interface principale du programme Mega 4.0

Annexe 6 : Insertion des huit séquences sur Mega 4.0

Annexe 7 : Résultat de l'alignement multiple des séquences par Mega 4.0

Annexe 8 : La matrice distances d'arbre ARN 16s

Liste des abréviations

ADN : acide désoxyribonucléique.
ADNr : ADN ribosomal.
AM : Alignment multiple.
ARN : acide ribonucléique.
ARNm : ARN messenger.
ARNr : ARN ribosomal.
°C : degrés Celsius.
Fe : fer .
FeMo-co: Cofacteur Fer molybdène.
KDa : kilo Daltons.
gyrB : sous unité β de l'ADN gyrase.
LCOs : lipo-chito-oligosaccharidiques.
MEGA : Analyse de Génétique Evolutionnaire Moléculaire.
MLST : Multilocus sequence typing.
NCBI : National Center for Biotechnology Information.
Ntr : nitrogène régulateur.
Nif : " nitrogen fixation "gène de fixation d'azote.
nod : gène de nodulation.
NJ : Neighbor-joining.
Pb : paires de bases.
pSym : plasmide symbiotique.
recA : recombinaise A.
rpoA : facteur α de l'ARN polymérase.
rpoB : sous unité β de l'ARN polymérase.
rrs : ribosomal RNA small subunit.
R.leguminosarum : *Rhizobium leguminosarum*.
sp : espèce non déterminée.
SSU: Small subnité unite.
UPGMA : Unweight Pair Group Method with Arithmetic mean
 μm : micrometre.
YMA : Yeast Mannitol Agar

SOMMAIRE

Remerciements

Dédicaces

Résumé

INTRODUCTION	01
SYNTHESE BIBLIOGRAPHIQUE	
1- BIOLOGIE DE <i>RHIZOBIUM</i>	02
1.1. Description de genre <i>Rhizobium</i>	02
1.2. Classification	05
2. ASPECTS MOLECULAIRES DE LA SYMBIOSE	08
2.1. Le dialogue moléculaire et gènes de la symbiose	08
2.2. Les gènes de nodulation ou gènes <i>Nod</i>	08
2.3. Les gènes de la nitrogénase (<i>Nif</i>)	13
2.4. Le gène ribosomique ARN 16S	17
3- LA PHYLOGENIE MOLECULAIRE	17
3.1. Les données de la phylogénie	18
3.2. Les gènes ménagers (MLST)	21
4- METHODES DE CONSTRUCTION DE PHYLOGRAMME	22
4.1. La méthode Neighbor-joining (NJ)	22
4.2. La méthode de parcimonie	23

METHODOLOGIE EXPERIMENTALE	
1- MATERIEL	24
1.1. Origine des séquences	24
1.2. Outils d'exploitation des séquences	25
2- METHODES	25
2.1. L'alignement multiple	25
2.2. Paramétrage de ClustalW2	26
2.3. Construction des arbres phylogénétiques	26
RESULTATS ET DISCUSSIONS	
1. RESULTATS	28
1.1. Arbre phylogénétique du gène ADNr 16S	28
1.2. Arbre phylogénétique des gènes de ménage	30
1.3. Arbre phylogénétique de gène <i>Nod C</i>	31
1.4. Arbre phylogénétique de gène <i>Nif H</i>	32
2. DISCUSSION	33
CONCLUSION	36
REFERENCES BIBLIOGRAPHIQUES	37

Annexes

INTRODUCTION

Les progrès rapides de la technologie de séquençage de l'ADN ont conduit à des changements majeurs dans les études de taxonomie et d'identification bactériennes. Les outils bioinformatiques contribuent considérablement dans l'analyse des gènes voire même des génomes complets. L'utilisation de tels programmes devient de plus en plus une nécessité dans le domaine de la recherche biologique moléculaire, plus particulièrement dans la systématique et la phylogénie bactériennes. Leur emploi permet une interprétation plus adéquate et plus rationnelle des données moléculaires. Ainsi, plusieurs germes ont du être reclassés dans de nouveaux genres auxquels ils n'appartenaient pas, et les différentes éditions du Bergey en sont l'exemple fulgurant.

Dans cette analyse, huit espèces de *Rhizobium* ont été sélectionnées pour une phylogénie qui sera réalisée sur une série de plusieurs gènes différents. Le gène qui code l'ARNr16S est la macromolécule de choix pour les analyses phylogénétiques, mais seule n'est pas discriminante pour les analyses taxonomiques. L'exploration d'autres molécules notamment les gènes ménagers (House Keeping Genes) codant pour des protéines conservées peuvent être une alternative intéressante au gène ARNr 16S (Aguileta, 2008).

Deux gènes symbiotiques sont également utilisés dans cette analyse vue leurs importances dans le processus de fixation d'azote et leur contribution dans la caractérisation des *Rhizobium* : Les gènes *nod C* responsables de la nodulation et les gènes *nif H* de la fixation. L'objectif de ce travail est de comparer la phylogénie de quatre familles de gènes au sein d'une population bactérienne composée de huit espèces du genre *Rhizobium*. Les familles de gènes sont :

1. ARNr 16S,
2. Séquences concaténées de quatre gènes ménagers :
 - La sous-unité α de l'ADN polymérase (*rpoA*),
 - la sous-unité β de l'ADN polymérase (*rpoB*),
 - la sous-unité β de l'ADN gyrase (*gyrB*),
 - la Recombinase A (*recA*).
3. Gènes *nodC*,
4. Gènes *nifH*.

1- BIOLOGIE DE *Rhizobium*

Rhizobium forme un groupe de bactéries du sol aérobies appartenant à la famille des *Rhizobiaceae* et le terme de rhizobia, dans son *stricto sensu*, fait référence à tous les membres de ce genre. Cependant, au fil des années, le terme est venu à être utilisé pour toutes les bactéries capables de nodulation et de fixation d'azote en association symbiotique avec les légumineuses et qui appartiennent à des genres étroitement liés à *Rhizobium* (Willems, 2006). Ces bactéries présentent la capacité de rentrer en symbiose avec des plantes en formant des nodosités, qu'on retrouve spécifiquement chez les légumineuses telles que : pois (*Pisum sp.*), haricot (*Phaseolus sp.*), soja (*Glycine sp.*), arachide (*Arachis hypogaea*), trèfle (*Trifolium*), luzerne (*Medicago sativa*)... ou même chez certaines céréales (*Triticum sp.*) (Rosenberg, 1997). Se sont des fixateurs symbiotiques de l'azote atmosphérique dans les nodules des racines ou des tiges des plantes légumineuses où elles se différencient en bactéroïdes (Graham, 1991 ; Haukka *et al.*, 1998 ; Gage, 2004).

Après colonisation du système racinaire par les rhizobia de la plante hôte, il y a formation de nodules (ou nodosités) à l'intérieur desquels les rhizobia endosymbiotiques fixeront l'azote atmosphérique au bénéfice de la plante. Le développement des nodules sur les racines des plantes hôtes et l'infection des cellules végétales par ces bactéries est appelé "processus de nodulation". Les bactéries du genre *Rhizobium* se trouvent sous deux formes :

1. une forme végétative (non bactéroïde) : ce sont des microorganismes réguliers que l'on trouve dans la rhizosphère et/ou dans le cordon d'infection, très mobiles quand ils sont jeunes.
2. une forme bactéroïde : Les individus sont irréguliers et ont une taille à peu près dix fois plus grande que celle de la forme végétative (Somasegaran et Hoben, 1994). Cette forme peut se retrouver par exemple chez *Rhizobium trifoli*, *Rhizobium meliloti* et *Rhizobium leguminosarum*

1.1. Description de genre *Rhizobium*

Selon le Bergey's manual® of systematic bacteriology, le genre *Rhizobium* comprend des bacilles de taille 0,5-1,0 x 1,2-3,0 µm. Ne formant pas de spores, Gram négatif, mobiles grâce à des flagelles péritriches. Jordan (1984) et Somasegaran et Hoben (1994) affirment la mobilité grâce un flagelle ou de deux à six flagelle péritriches. Sont aérobies avec l'oxygène comme accepteur final d'électrons mais peuvent bien croître en faibles pression d'oxygène. En anaérobiose, elles sont capables d'utiliser le nitrate comme accepteur final d'électrons. La température optimale de croissance se situe dans l'intervalle de 25 à 30°C. Certaines espèces peuvent, cependant croître à des

températures de plus de 40°C. Le pH optimum de croissance étant de 6 à 7. Les colonies sont généralement de couleur blanchâtre ou beige, de forme circulaire, convexe, translucides ou opaques, mucilagineuses et bombées. La croissance sur milieux (YMA) enrichis de carbohydrates est accompagnées de la synthèse de polysaccharides extracellulaires.

Au point de vue nutritionnel, ce sont des chimio-organotrophes utilisant pour leur croissance une large gamme de sucres (sauf l'amidon et la cellulose) et de sels d'acides organiques sans production de gaz (Tableau 1). Les *Rhizobium* utilisent des carbohydrates relativement simples comme le glucose, le mannitol, le saccharose et des composés aminés; mais la majorité des souches à croissance lente préfère le pentose (Somasegaran et Hoben, 1994). Certaines souches peuvent se développer en chimiolithotrophie en présence de H₂O, CO₂, d'une faible pression en oxygène et de faibles concentrations en sels d'ammonium ou de nitrate. Les sels d'ammonium, les nitrates, les nitrites et beaucoup d'acides aminés sont une source d'azote. Les *Rhizobium* n'assimilent pas l'azote en dehors de la plante et ont besoin d'une source d'azote ammoniacal (NH₃) ou aminé pour se développer à l'état libre. Des souches de certaines espèces peuvent croître sur des milieux simples de sels minéraux additionnés de vitamine ou d'hydrolysats de caséine en tant que seule source de carbone et d'azote à la fois (Somasegaran et Hoben, 1994). D'autres souches exigent, pour leur croissance, un ou plusieurs facteurs de croissance tels que la biotine, le pantothénate ou l'acide nicotinique. La peptone est faiblement utilisée. La caséine, l'amidon, la chitine et l'agar ne sont pas hydrolysés. Toutes les espèces de *Rhizobium* induisent un hypertrophisme chez les plantes se traduisant par des nodosités racinaires avec ou sans fixation symbiotique de l'azote.

Les *Rhizobium* peuvent assimiler un large spectre de formes organiques ou minérales d'azote et il existe une relative diversité au sein des souches d'une même espèce quant à leurs préférences, particulièrement chez les bactéries à croissance lente :

- L'urée peut être assimilée par toutes les espèces du genre *Rhizobium* (Jensen et Schroeder, 1965).
- Parmi la grande variété d'acides aminés utilisés, le glutamate est le plus adéquat. Elkan et Kurik (1968) constatent que les autres acides aminés largement utilisés à travers les différentes espèces sont l'aspartate, la proline et l'histidine, alors que Tan et Broughton (1981) ajoutent que l'ion ammonium est utilisé par tous les *Rhizobium*.
- Vincent (1977) constate que la peptone constitue une très mauvaise source d'azote à l'exception de certains *Rhizobium*. L'extrait de levure est la source d'azote fournie habituellement au *Rhizobium*. Sa capacité à stimuler la croissance de nombreuses souches

résulte autant de l'apport de substrats azotés assimilables que de divers composés qu'il contient (vitamines, facteurs de croissance etc.). Sa concentration doit cependant rester inférieure à 0,35% car au-delà, elle pose des problèmes de perte de viabilité et de déformation des cellules de *Rhizobium* (Urban, 1979 ; Chakrabarti *et al*, 1981).

Tableau 1: Caractères biochimiques de genre *Rhizobium* (Sadowsky *et al*, 1983 ; Shahzad *et al*, 2012 ; Young *et al* 2001)

	<i>R. leguminosarum</i>	<i>R. etli</i>	<i>R. gallicum</i>	<i>R. giardinii</i>	<i>R. tropici</i>	<i>R. multihospitium</i>	<i>R. pisi</i>	<i>R. leucaenae</i>
Catalase	+	+	+	+	+	+	+	+
Oxydase	+	+	+	+	+	+	+	+
Production d'uréase	+	+	+	+	+	+	+	+
Citrate	-	-	-	-	-	-	-	-
Production d'H ₂ S	-	-	-	-	-	-	-	-
Indole	-	-	-	-	-	-	-	-
Rouge de méthyle	-	-	-	-	-	-	-	-
Gélatinase	+	+	+	+	+	+	+	+
pH de croissance	4 - 9	I	4 - 8	4 - 8,5	4 - 10	6,0 - 8,0	I	I
28 °C	+	+	+	+	+	+	+	+
35 °C	I	+	I	I	+	+	I	+
40 °C	-	-	-	-	+	+		+
1% NaCl	-	-	I	I	I	I	+	I
2% NaCl	-	I	-	+	-	I	I	-
Nitrate réductase	+	+	+	+	+	+	+	+

+ : positive, - : négative, I : Inconnue

Sur le plan biomoléculaire, les gènes de nodulation (*nod*) et ceux de fixation d'azote (*nif*) sont regroupés sur des plasmides ou sur des mégaplasmides comme le plasmide symbiotique *pSym*. De nombreux travaux ont démontré la possibilité de transfert du *pSym* entre les souches de rhizobia et d'autres souches incapables d'induire la nodulation dans les conditions de laboratoire (Laguerre *et al*, 1993 ; Louvrier *et al*, 1996).

1.2. Classification

La taxonomie des rhizobia a considérablement changé au cours des 20 dernières années. Dans l'édition de 1984 du *Bergey's manual*® of systematic bacteriology, la famille des *Rhizobiaceae* était composée des rhizobia (qui comprenaient les genres *Rhizobium* et *Bradyrhizobium*), d'*Agrobacterium* et de *Phyllobacterium*. Durant cette période, les progrès de biologie moléculaire tels que l'hybridation ADN/ADN et le séquençage des gènes ont révélé des caractéristiques génétiques nouvelles qui se traduisent par une diversité au sein des rhizobia avec notamment un plus grand nombre de genres comme indiqué dans le tableau suivant :

Tableau 2 : Evolution du nombre d'espèces au sein de différents genres de rhizobia (Willems, 2006)

Genres	Auteurs	Nombres d'espèces					
		< 1980	81 - 85	86 - 90	91 - 95	96 - 00	01 - 06
<i>Agrobacterium</i>	Cohn, 1942	4	4	5	5	5	5
<i>Rhizobium</i>	Frank, 1889	4	5	5	10	10	16
<i>Bradyrhizobium</i>	Jordan, 1982			1	3	3	7
<i>Sinorhizobium</i>	Chen <i>et al</i> , 1988			2	5	8	11
<i>Azorhizobium</i>	Dreyfus <i>et al</i> , 1988					1	2
<i>Mesorhizobium</i>	Jarvis <i>et al</i> , 1997					7	11
<i>Allorhizobium</i>	De Lajudie <i>et al</i> , 1998						1
Total		8	9	13	23	34	53

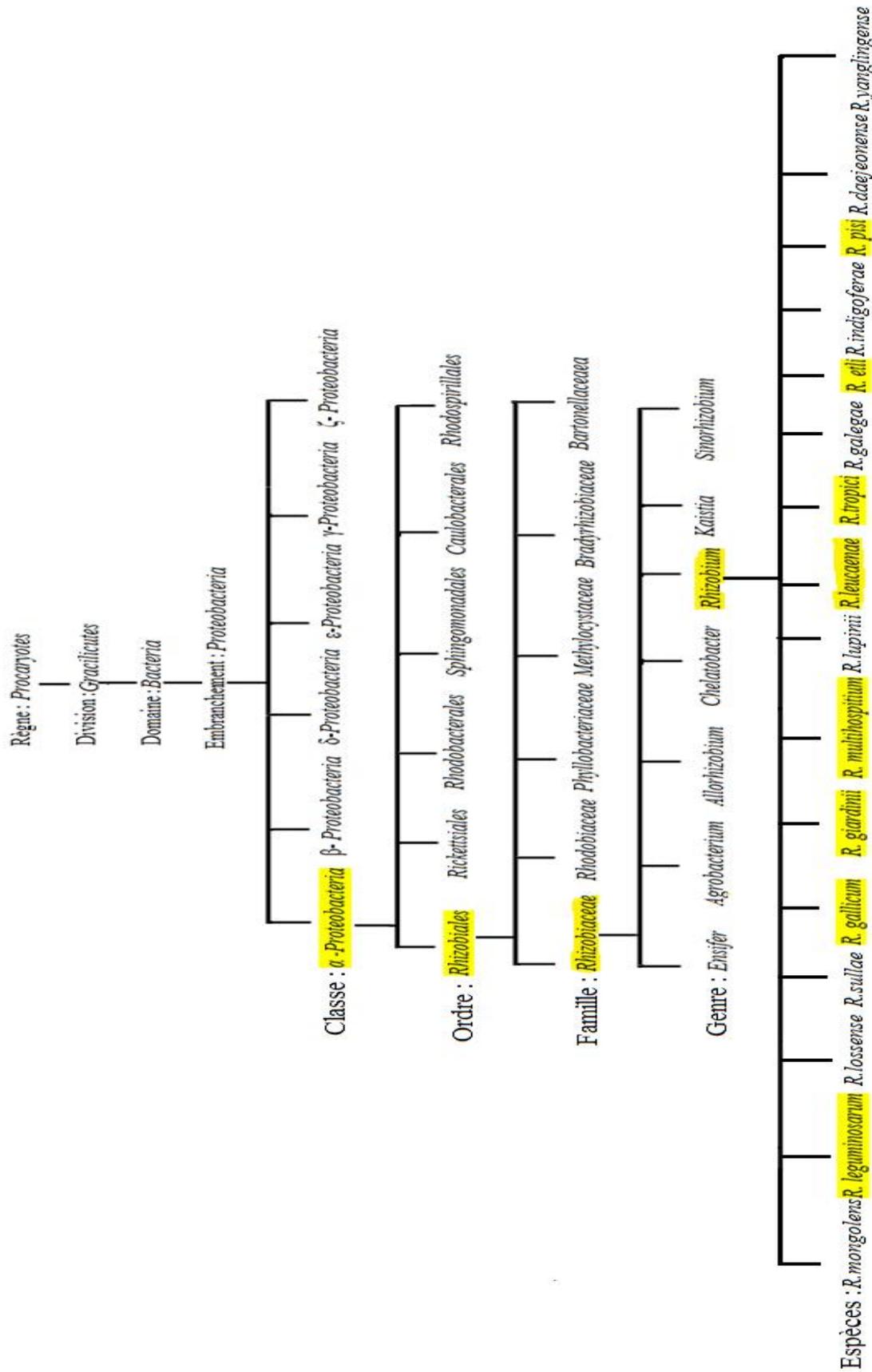


figure 1: Classification des *Rhizobium* spp

Le développement des méthodes phylogénétiques et polyphasiques ont aussi fait évoluer le nombre de souches rhizobia et le séquençage du génome total présente, indubitablement, un impact sur la taxonomie de ces souches. Les études taxonomiques faisant usage de l'approche polyphasique en incluant les caractères génétiques, phénotypiques, la chimiotaxonomie et les données phylogénétiques ont permis une meilleure vue de l'image du réseau interrelationnel des rhizobia et une classification appropriée. Les *Rhizobium* sont divisés en fonction d'un ensemble de caractères morphologiques, biochimiques et moléculaires (Jordan, 1984). Leur place dans le monde bactérien est illustrée dans la figure 1 présentée dans la page ci-contre.

L'évolution du nombre d'espèces de *Rhizobium* n'a cessé de croître. En 2005, on en comptait une quinzaine (Tableau 3) et depuis, le nombre a complètement explosé pour doubler.

Tableau 3 : Espèces de *Rhizobium* recensées en 2005 (Willems, 2006)

Nom	Année	Nom	Année	Nom	Année
<i>R. leguminosarum</i>	1879	<i>R. gallicum</i>	1997	<i>R. yanglingense</i>	2001
<i>R. lupinii</i>	1886	<i>R. giardinii</i>	1997	<i>R. sullae</i>	2002
<i>R. galegae</i>	1989	<i>R. hainanense</i>	1997	<i>R. indigoferae</i>	2002
<i>R. tropici</i>	1991	<i>R. mongolens</i>	1998	<i>R. lossense</i>	2003
<i>R. etli</i>	1993	<i>R. huautlense</i>	1998	<i>R. daejeonense</i>	2005

La consultation (Juin 2014) du site Internet du NCBI (<http://www.ncbi.nlm.nih.gov/taxonomy>) fait apparaître un nombre beaucoup plus grand d'espèces du genre *Rhizobium* (voir annexe N° 1).

2- ASPECTS MOLECULAIRES DE LA SYMBIOSE

2.1. Le dialogue moléculaire et gènes de la symbiose

La symbiose fixatrice d'azote se traduit par la capacité des *Rhizobium* à induire la formation de nodosités au niveau des racines ou même des tiges de légumineuses particulières (processus de nodulation). Lors de cette symbiose un véritable dialogue moléculaire s'installe entre les deux partenaires. Ce dialogue est effectué grâce à un échange de messagers libérés sous forme de signaux biochimiques dont font partie les flavonoïdes excrétés par les racines en conditions de carence en ions ammonium. Une induction de l'expression des gènes *nod* permet alors aux bactéries de synthétiser des molécules signales appelées facteurs Nod (lipo-chitooligosaccharides) responsables de la nodulation (Tereferwork, 2002). Les gènes *nod* sont localisés pour la plupart sur des plasmides de grande taille appelés les *pSym* et sont hautement conservés entre les différentes espèces.

L'analyse génétique de diverses espèces de *Rhizobium* a permis d'identifier trois types de gènes symbiotiques qui entre dans le dialogue moléculaire plante/*Rhizobium* :

- Les gènes *nod* contrôlant la reconnaissance spécifique, l'infection et la nodulation,
- les gènes *fix* indispensables pour la fixation d'azote,
- les gènes *nif* codant pour le complexe de la nitrogénase

2.2. Les gènes de nodulation ou gènes *nod*

Les fonctions des protéines codées par les différents gènes *nod* chez certaines espèces de *Rhizobium* diffèrent d'une espèce à une autre. Sur le plasmide *pSym*, les gènes *nod* sont organisés en opérons. Il s'agit des opérons *nod ABCIJ*, *nod FEL*, *nod MNT*, *nod O* et *nod D*. Ces opérons sont localisés entre deux régions géniques responsables de la fixation d'azote (gènes *nif*).

La structure fine des gènes *nod* comporte les régions suivantes :

- Un gène régulateur *nodD* qui code pour des protéines qui activent et régulent l'expression des autres gènes *nod* de la bactérie (Davet, 1996),
- des gènes communs à tous les *Rhizobium* (*nodA*, *nodB*, *nodC*, *nod I* et *nod J*) qui jouent un rôle essentiel dans la nodulation,
- une série de gènes de spécificité vis à vis de l'hôte à infecter (*nodE*, *nodF*, *nodG*, *nodH* etc.). Ils codent pour les différents radicaux (R1, R2, R3, etc.) des facteurs Nod (Hopkins, 2003).

2.2.1. Gène *nodD*

De rôle régulateur, il est le premier gène transcrit lors du processus de nodulation. L'expression de ce gène est constitutive, c'est à dire qu'il est exprimé quelque soit les conditions de croissance de la bactérie. Par contre, l'affinité de ses produits pour l'opérateur sera modifiée en présence de flavonoïdes. Il code pour la protéine nod D qui joue le rôle de récepteur des flavonoïdes excrétés par la plante. Une fois l'association établie, sous forme de complexe "flavonoïdes – protéines D", la liaison au promoteur Nod-Box peut alors avoir lieu (Tereferwork, 2002) (Voir figure 2).

2.2.2. Gènes *nod* communs (ABCIJ)

Les gènes *nod A, B, C, I* et *J* sont communs à tous les *rhizobia*. Les gènes A, B et C sont physiquement et fonctionnellement conservés chez toutes les espèces des rhizobia et leurs mutations provoquent une perte complète de la capacité à noduler les plantes hôtes. En effet, ces gènes sont indispensables pour induire la déformation des poils absorbants en forme de crosse (*curling*) et inciter les cellules végétales à se diviser (Rosenberg, 1997) sous l'effet de la production d'oligosaccharides (facteurs nod).

Les gènes *nod I* et *J* ont été identifiés chez de nombreux rhizobia (Fellay *et al*, 1995 ; van Rhijn et Vanderleyden, 1995). Ces gènes situés en aval du gène *nod C*, codent pour les protéines de transport externe Nod I et Nod J de la membrane bactérienne lesquelles présentent des homologies de structure avec des protéines de système de transport membranaire ATP-dépendant, impliquées dans l'exportation de polysaccharides. Une mutation dans l'un de ces deux gènes se traduit par une diminution importante de la quantité de facteurs Nod (Rosenberg, 1997).

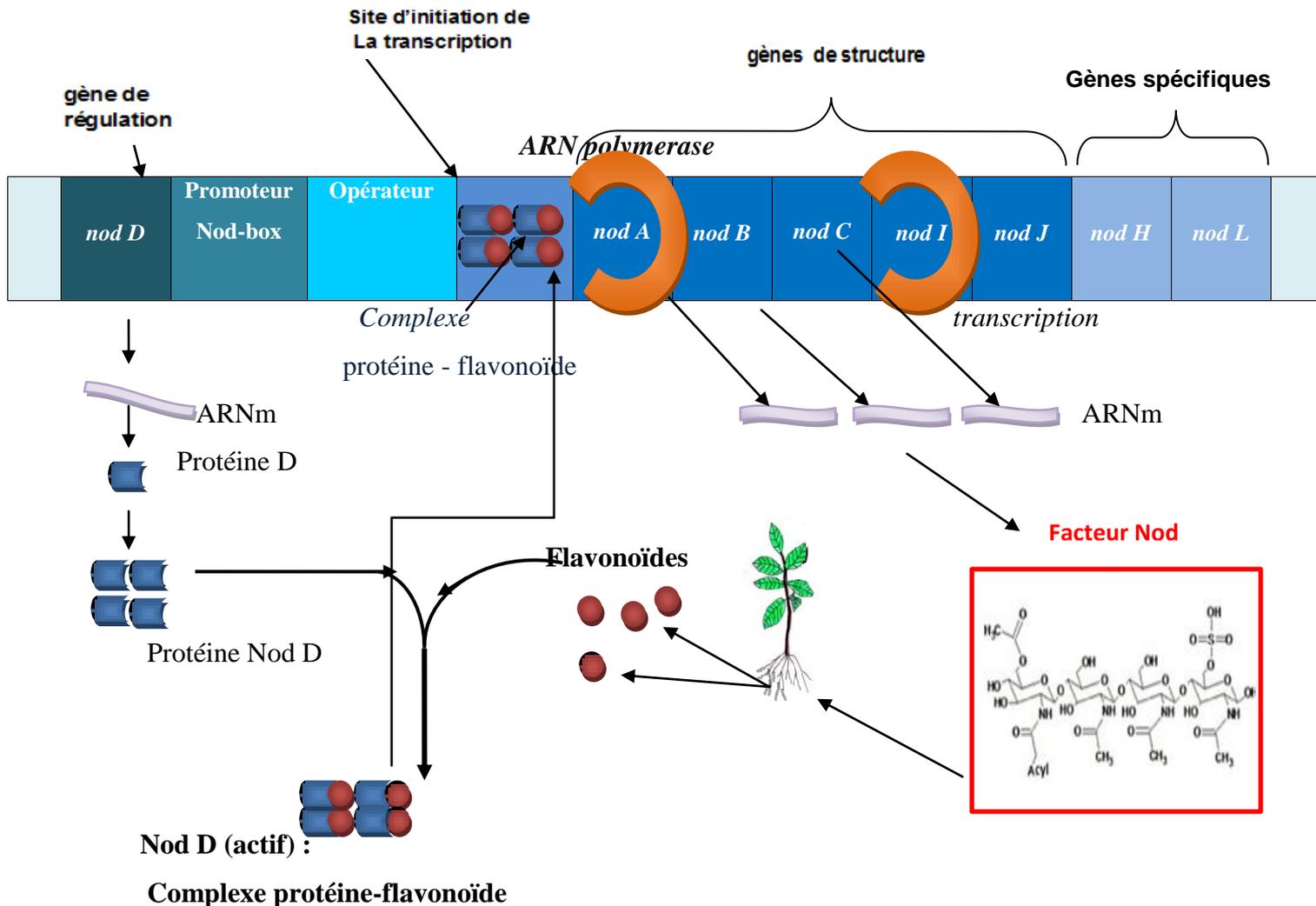


Figure 2 : Régulation de l'opéron *nod* en présence des flavonoïdes

2.2.3. Les gènes spécifiques

On rencontre différents gènes spécifiques présents chez certaines espèces et non chez d'autres. Leur rôle étant de coder pour des enzymes catalysant les modifications entre les différents radicaux du facteur Nod. On peut citer :

- Le gène *nodH*, présent chez *R.meliloti*, code pour une sulfotransférase qui va permettre l'introduction d'un groupement sulfate sur l'unité réductrice.
- Le gène *nodL* code pour la protéine Nod L qui permet, chez *R.meliloti* l'acétylation sur l'unité non réductrice (Terefework, 2002).

Tableau 4 : la fonction biochimique de quelques gènes *nod* et leur espèces (Dénarié *et al*, 1996).

Gènes <i>nod</i>		Fonctions	espèces
Gène régulateur	<i>nod D</i>	Régulateur de type Lys-R	
Gènes communs	Les gènes de synthèse de squelette lipo-chito-oligosaccharidique : <i>nod C</i> <i>nod A</i> <i>nod B</i> Les gènes d'exportation des facteurs Nod a l'extérieur : <i>nod I</i> <i>nod J</i>	N-acétyl-D-glucosaminyl transférase N-acyltransférase N-D-acétylase ATP-dépendant Protéine membranaire	Toutes les espèces de <i>Rhizobium</i>
Gènes spécifiques	Les substitutions de coté non-réducteur : <i>nod L</i> <i>nod U</i> Les substitutions de coté réducteur : <i>nod P</i> <i>nod H</i> <i>nod Z</i>	Acetyl-transférase Carbamoyl transférese ATP sulfurylase Sulfotransférese Glycusyl transférase	<i>R.tropici</i> <i>R.fredii</i> <i>R.leguminosarum</i> <i>R.melilloti</i>

2.2.4. Les facteurs Nod (structure et synthèse)

Après la transcription, les gènes *nod* produisent des enzymes qui contrôlent la synthèse de molécules indispensables au processus de nodulation. Ces signaux, généralement connus sous le nom de facteurs Nod, sont des molécules lipo-chito-oligosaccharidiques (Terefework, 2002). Ils sont constitués d'un squelette de trois à six résidus N-acétyl-D-glucosamine réunies par des liaisons β (1-4) (structure de la chitine), substitué par une chaîne d'acétyl au niveau de l'extrémité non réductrice et portant divers motifs structuraux aux deux extrémités de la chaîne oligosaccharidique (figure 3).

Les facteurs Nod sont responsables de la spécificité, de la reconnaissance entre les symbiotes, et du déclenchement de la nodulation. En effet, même émis à des concentrations minimales, ces signaux peuvent déclencher des réponses symbiotiques chez la plante telles que la déformation des poils radiculaires, la division corticale des cellules et la formation de nodule primordial (Debelle *et al*, 2001).

Les gènes communs *nod ABC* déterminent la synthèse de la structure de base commune à tous les facteurs Nod (Denarie *et al*, 1996). La protéine NodC catalyse la synthèse du chito-oligomère et en contrôle la longueur. L'unité de glucosamine à l'extrémité non réductrice est désacétylée par *nodB*, et substituée par une chaîne d'acide gras par *nodA*. Les gènes *nod* spécifiques modifient le squelette oligochitinique à l'aide de substitutions particulières et confèrent aux facteurs Nod leurs spécificités.

Aucune corrélation n'a été observée entre le type de facteur Nod produit par le *Rhizobium* et les plantes nodulés. Par exemple : les facteurs Nod produits par *R.etli* et *R.loti* sont identiques mais les deux espèces ont des gammes d'hôtes distinctes.

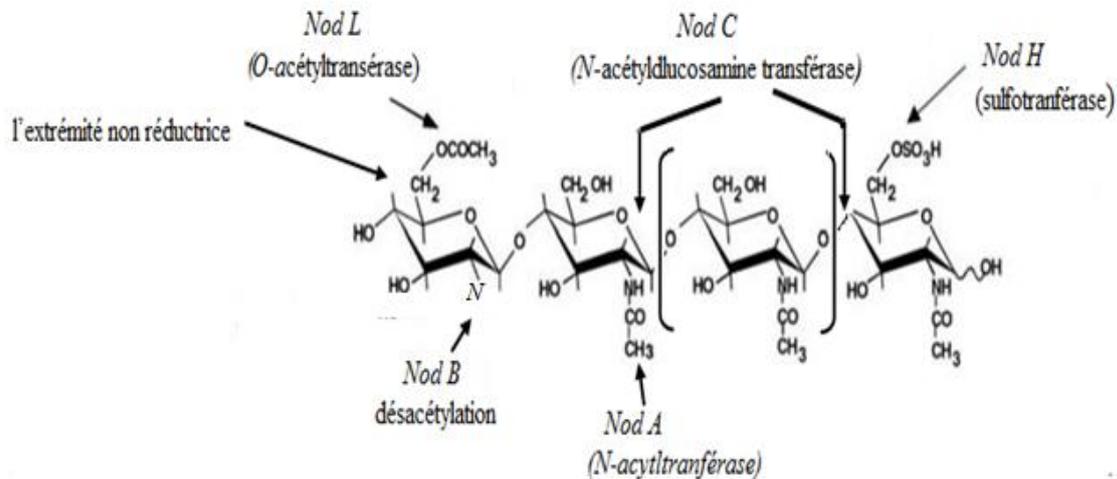


Figure 3: Structure et synthèse du squelette de base des facteurs Nod

nodC est une N-acétyl glucosaminyl transférase synthétisant un oligomère de chitine. *nodA* est une acyltransférase impliquée dans le greffage d'une chaîne d'acyl sur le squelette oligosaccharidique désacétylé par la protéine *NodB*. Des résidus de nature différente (R1 à R5) sont ajoutés au squelette de base et modulent la spécificité d'hôte de la souche.

2.3. Les gènes de la nitrogénase (*nif*)

En plus des gènes *nod*, les rhizobia possèdent d'autres gènes interviennent dans les étapes tardives de la symbiose (après la formation du nodule). Ils sont plus de 17 gènes organisés en opérons (Dupuy et Nougier, 2005).

Ils sont découverts sur les plasmides bactériens et codent pour la synthèse et le fonctionnement d'un complexe enzymatique catalysant la réduction de l'azote et connu sous le nom de nitrogénase (Hopkins, 2003). Cette enzyme a été mise en évidence uniquement chez les procaryotes. Elle est très conservée chez les bactéries fixatrices d'azote tant au niveau de sa séquence que de sa structure. Cette enzyme est constituée de deux métalloprotéines de tailles différentes :

- le site de la réduction du substrat est la protéine MoFe ou dinitrogénase (245 KDa),
- le donneur d'électrons est la protéine Fe ou dinitrogénase réductase (64 KDa).

L'expression des gènes *nif* dépend d'un activateur de la transcription appelé *nif A* et d'un facteur sigma spécifique produit de *rpoN*, qui reconnaît des sites promoteurs particuliers en amont des opérons *nif* (régulation positive). Il existe aussi une protéine de régulation négative *nif L* qui à une activité antagoniste à celle de *nif A*. elle consiste à inhiber la synthèse des gènes *nif* en condition d'aérobie (Spaink *et al*, 1998) (voir figure 4).

La régulation de la fixation d'azote chez les *Rhizobium* impose deux systèmes différents qui répondent à des variations de concentration de l'azote fixé ou de l'oxygène externe.

2.3.1. La régulation par le système central Ntr

Dans des conditions de limitation en azote le système de régulation central Ntr (nitrogen regulation) aura lieu. Il est composé d'un système régulateur à deux composantes NtrB, NtrC, Des protéines PII (GlnK et GlnB) et le GlnD (Abdelmadjid, 2010).

Lorsqu'il y a une carence en azote au niveau du sol le système Ntr va activer la phosphorylation de NtrB qui fournit à son tour un groupement phosphorylé à NtrC. Dès que la phosphorylation de NtrC est activée, la transcription des autres gènes *nif* aura lieu (Abdelmadjid, 2010).

2.3.2. La régulation par le système Fix L J

Lorsqu'il y a absence d'O₂ le système FixL J joue un rôle très important dans l'expression des gènes *nif*. La protéine régulatrice FixL située sur la membrane des bactéroïdes détecte le taux d'oxygène (condition d'anoxie), Cette dernière s'autophosphoryle afin que la protéine FixJ est phosphorylée aussi par l'utilisation de FixL-Phosphate comme donneur. Ceci permet l'activation de la transcription des différents gènes et particulièrement celle de *nifA* (activateur de la transcription des gènes *nif*) (Domergue, 2006).

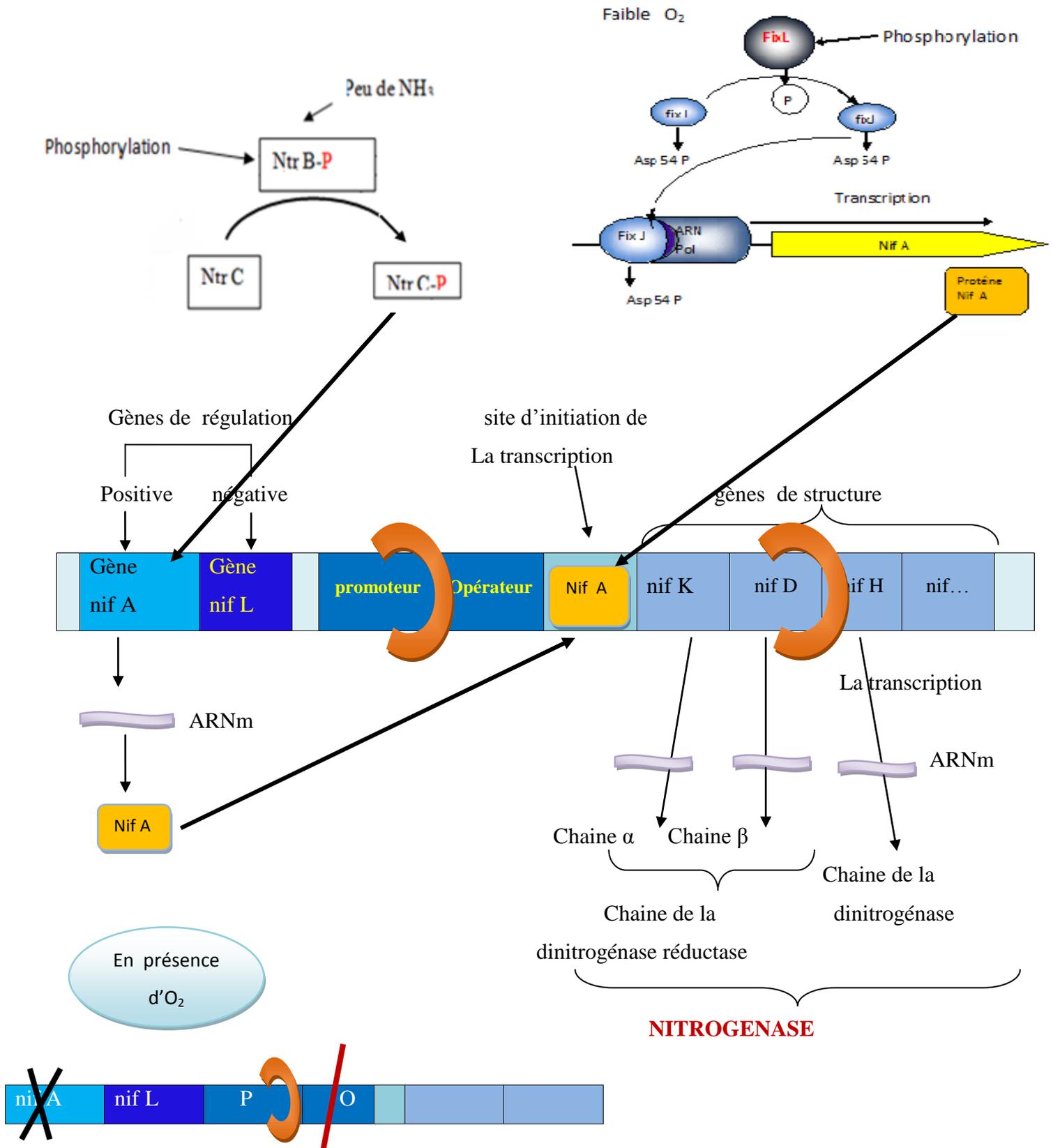


Figure 4 : la régulation de l'expression des gènes *nif*

Les gènes *nif K*, *nif D*, *nif H* sont les gènes de structure de la dinitrogénase réductase et de la dinitrogénase. Le produit de gène *nifH* code pour la réductase, alors que le produit des gènes *nifD*, *K* codent pour les polypeptides de la protéine à cofacteur FeMo (Hopkins, 2003).

Les gènes *nif Q*, *B*, *V*, *N* et *E* codent pour des protéines dont les produits sont impliqués dans la biosynthèse et l'assemblage du cofacteur Fe-Mo. Alors que le produit des gènes *nif F* et *J* sert au transport des électrons. Le produit des gènes *nif S* et *U* prend une place dans le processus de la dinitrogénase et celle de *nif M* dans le processus de la dinitrogénase réductase (Gharzouli, 2006). (Tableau 5).

Tableau 5 : les différents types de gènes *Nif* et leurs fonctions

Gènes <i>nif</i>		Fonctions
Gènes régulateurs	<i>nif A</i>	Régulation positive ; régulateur de la transcription de tous les gènes <i>nif</i> , <i>fix</i> et autres gènes
	<i>nif L</i>	Régulation négative ; Inhibe l'expression de gènes <i>nif</i>
Gène de structure	<i>nif H</i>	Code la Protéine Fe nitrogénase
	<i>nif D</i>	Code la sous unité α protéine MoFe
	<i>nif K</i>	Code sous unité β protéine MoFe
	<i>nif E</i>	Code la biosynthèse cofacteur MoFe
	<i>nif N</i>	Code la biosynthèse cofacteur MoFe
	<i>nif B</i>	Code la biosynthèse cofacteur MoFe
	<i>nif S</i>	Code pour la Cystéine désulfurase
	<i>nif W</i>	Requis pour l'activité de la protéine MoFe
<i>nif X</i>	inconnue	

2.4. Le gène ribosomique ARNr 16S

De nombreux travaux ont montré que les ARNr étaient de bons outils phylogénétiques pour établir les liens de parenté entre tous les êtres vivants et ce en raison de leur universalité, leur abondance, leur taille convenable aux analyses comparatives.

L'ARNr contient des régions de séquences hautement conservées. Chez les procaryotes les gènes codant pour les ARNr sont organisés en opérons appelés rrs (ribosomal RNA small subunit). Le gène qui code l'ARNr 16S est codé pour la petite sous unité ribosomique SSU (Small subnité unite) d'environ 1540 pb.

Woese (1987) a montré que le gène de l'ARNr 16S était présent chez tous les organismes procaryotes, qu'il a la même fonction et que sa structure primaire est conservée. Plusieurs chercheurs l'ont utilisé pour une approche rapide afin d'évaluer la variabilité génétique entre les *Rhizobium* (Laguerre *et al*, 1994 ; Nour *et al*, 1994). Cependant, la présence de plusieurs copies du gène au sein d'un génome peut conduire à des hétérogénéités de séquences entre les copies. Le gène de l'ARNr 16S représente en moyenne 0,05 % du génome du procaryote (Rodriguez-Valera, 2002). Phylogénétiquement, il est admis que des séquences partageant moins de 97% de similarité correspondent à des espèces différentes. Le pouvoir résolutif de l'ARNr 16S, *a priori* suffisant pour séparer des espèces différentes, ne permet pas, par contre, de regrouper des espèces identiques. Si le pourcentage d'identité des ARNr 16S de deux souches est supérieur ou égal à 97%, leur placement dans une unique espèce ou dans deux espèces différentes doit reposer sur les résultats des hybridations ADN/ADN (kitouni, 2007).

3- LA PHYLOGENIE MOLECULAIRE

La phylogénie est une discipline émergente. Elle permet d'étudier les espèces afin de les classer en fonction de leurs ressemblances phénotypiques (phylogénie phénétique) ou en fonction de leurs séquences géniques (phylogénie moléculaire). Les relations de parenté sont généralement représentées sous forme d'un arbre phylogénétique ou dendrogramme (Luchetta *et al*, 2005).

L'objectif de la phylogénie est d'aboutir à une représentation statistique (arbre) qui permet aux systématiciens de bien visualiser les relations phénotypiques et/ou génotypiques entre les différents taxons analysés, en se basant sur des théories de construction d'arbres bien établies.

Dans le cas de comparaison des séquences ADN (séquences codantes ou non), les mutations par substitution ou par insertion/délétion fournissent beaucoup d'informations sur l'évolution des séquences comparées.

La phylogénie introduit deux concepts permettant de différencier deux gènes très proches (homologues) du point de vue de leurs séquences primaires :

1. Les gènes dérivant d'un gène ancestral par spéciation sont appelés *orthologues* et présentent de fortes probabilités pour posséder des fonctions similaires.
2. Les gènes produits par une duplication à l'intérieur du génome d'une même espèce sont appelés *paralogues* et s'orientent vers une divergence fonctionnelle, puisque la copie à travers le processus de l'évolution semble avoir adopté une nouvelle fonction (Philippe, 2009).

3.1. Les données de la phylogénie

La construction de la phylogénie est basée sur deux grands types de données :

- Les données liées aux caractères phénotypiques : comprennent les caractères observables (différents états : morphologiques, biochimiques et physiologiques; par exemple enzyme actif / non actif) et les patterns binaires (de type présence d'un caractère donné / absence de ce même caractère ; exemple présence d'une bande électrophorétique / son absence).
- Les données moléculaires : telles que les séquences nucléotidiques dont les séquences ADNr 16S, les séquences des cytochromes C, la ribulose 1,5-biphosphate carboxylase et les séquences du facteur d'élongation alpha (*tuf*) (Djekoun, Hamidechi, 2013).

La figure 5 ci-dessus représente les différentes étapes de la construction des arbres phylogénétiques :

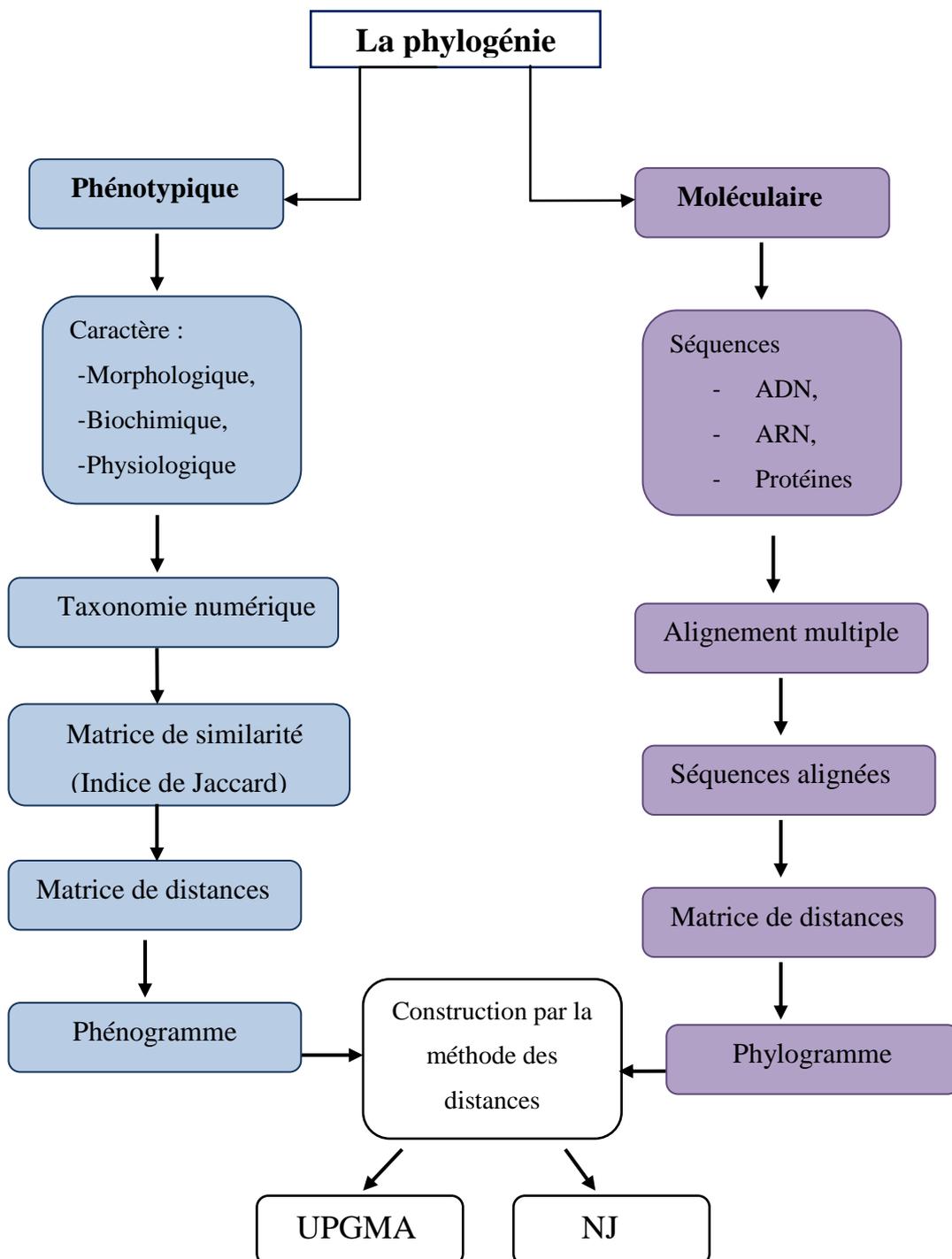


Figure 5 : les différentes étapes de construction de la phylogénie

La phase capitale avant toute construction phylogénétique étant l'alignement multiple. Celui-ci est réalisé dans le but de comparer plusieurs séquences géniques (ou protéiques) en même temps et de fournir une matrice de distances qui constituera la base de la construction du dendrogramme.

L'alignement multiple peut être réalisé grâce à plusieurs programmes informatiques dont ClustalW, Multalin, phylip, PhyML, etc.

Le résultat de l'alignement multiple est un ensemble de séquences présentées chacune dans une ligne pour pouvoir les comparer position par position et voir s'il ya concordance (identité parfaite des nucléotides) ou dis-concordance entre les nucléotides des séquences alignées. Ces dis-concordances sont des mutations par substitutions. La matrice des distances obtenue après alignement multiple mesure le pourcentage de ces dis-concordances entre l'ensemble des séquences.

Dans l'exemple suivant, il est possible de visualiser le résultat d'un alignement multiple :

	Positions de l'alignement								
	1	2	3	4	5	6	7	8	9
Séquence 1	G	T	A	T	C	A	C	C	G
Séquence 2	A	T	A	C	C	A	C	T	G
Séquence 3	T	-	A	A	C	A	C	C	G
Séquence 4	T	T	A	T	C	T	C	T	G
Séquence 5	T	T	A	T	C	A	C	C	G
Séquence 6	T	A	A	C	C	A	T	C	G

Dans cet exemple de lecture d'un alignement multiple, il est à constater la mutation par substitution entre les deux séquences 1 et 2 dans la première position de cet alignement ou encore entre les séquences 4 et 5 en position 8 avec une substitution entre les nucléotides T de la séquence 4 et le nucléotide C de la séquence 5.

La séquence 3, en position N°2 ne présente aucun nucléotide. Ceci revient au fait que cet individu aurait perdu son nucléotide T par délétion. On parle à ce niveau de gap ou de brèche ou

encore Indels (Insertion/Délétion). Il faut noter que le nombre des Indels affecte négativement la qualité des alignements.

En fin, des zones de fortes conservations sont à indiquer; il s'agit des sites N°3, 5 et 9 où l'on note une forte identité des nucléotides entre l'ensemble des séquences. De tels sites ne font qu'améliorer la qualité de l'alignement.

3.2. Les gènes ménagers (MLST)

L'expression d'un gène cible est exprimée par rapport à des gènes de référence appelés gènes d'entretien ménager ou gènes domestiques. Ils sont toujours exprimés car codent pour la synthèse des protéines dont l'activité et la fonction sont indispensables à la vie cellulaire. Ils sont donc toujours présents dans toutes les conditions (Romaric. 2006).

Ces gènes sont considérés comme un outil de choix dans les expériences de typage moléculaire afin de permettre la différenciation entre les espèces très proches.

Il existe plusieurs types de gènes de ménage

3.2.1. La sous-unité α de l'ADN polymérase (*rpoA*)

Le gène *rpoA* codant pour la sous-unité α de l'ARN polymérase dépendante de l'ADN. C'est une enzyme qui responsable a L'initiation de la transcription nécessite que la sous-unité σ de l'ARN polymérase qui se fixe sur l'oligomère $\alpha 2\beta\beta'$ (appelé "co-enzyme") pour former l'holoenzyme $\alpha 2\beta\beta'\sigma$.

3.2.2. La sous-unité β de l'ADN polymérase (*rpoB*)

Le gène *rpoB*, codant la sous-unité β de l'ARN polymérase, fourni une résolution phylogénétique comparable, en particulier dans l'analyse des isolats étroitement liés avec le gène *ARNr 16S*, donc il permet de révéler de nouvelles espèces bactériennes et d'affiner l'analyse de la communauté bactérienne.

3.2.3 La sous-unité β de l'ADN gyrase (*gyrB*)

Le *gyrB* est un gène qui code la protéine de la sous -unité β de l'ADN gyrase (ou la topoisomérase II). C'est une enzyme qui est responsable de la présence des surenroulements négatifs dans les chromosomes bactériens et qui joue un rôle crucial dans leur répliation (watt et higgins, 1994). La sous-unité de l'ADN gyrase est essentielle pour la survie des cellules bactériennes. Le transfert horizontale de ce gène est rare ce qui explique son utilisation dans la

classification taxonomique des bactéries (Sawada *et al*, 1999 ; Watanabe *et al*, 2001 ; Fucushima *et al*, 2002 ; Yanez *et al*, 2003).

3.2.4. Recombinase A (*recA*)

Le *recA* est un gène qui code pour la protéine Recombinase A (*recA*) qui joue un rôle essentiel dans la réparation et la maintenance de l'ADN. La protéine *recA* a un intérêt particulier pour la phylogénie moléculaire en raison de ses propriétés stables (Pleckaityte, 2010).

4- METHODES DE CONSTRUCTION DE PHYLOGRAMME

Plusieurs méthodes ont été développées pour construire un arbre phylogénétique à partir d'un ensemble de données mesurées ou observées sur un ensemble d'individus. Ces données peuvent être phénotypiques (caractères morphologiques, physiologiques, etc.) ou génotypiques (séquences d'ADN ou de protéines). De manière générale, les méthodes de construction phylogénétique peuvent être divisées en :

	Méthodes	Algorithmes de classification
Types de données	Parcimonie	
	Caractères	Maximum de vraisemblance Pas d'algorithme
		Minimum d'évolution - UPGMA
	Distances	Carrés moyens - Neighbor-Joining (NJ)

4.1. La méthode Neighbor-joining (NJ)

C'est un procédé de construction des arbres phylogénétiques basé sur le calcul de distances entre les séquences prises deux à deux (Saitou et Nei, 1987). Ces distances expriment les différences génétiques entre les séquences comparées en mesurant le pourcentage de mutations entre toutes ces paires de séquences (Substitutions et Insertions/Délétions). Il permet d'identifier les paires de séquences voisines avec des faibles taux de mutations qui minimisent la longueur totale des branches de l'arbre à construire (Kitouni, 2007).

C'est une méthode rapide qui s'intéresse aux séquences de fortes identités dites séquences homologues. Cette méthode ne tient pas compte de la théorie de l'horloge moléculaire pour la construction des dendrogrammes. Elle se base par contre sur le principe d'attribuer au dendrogramme des branches dont la longueur totale est la plus petite possible (minimum d'évolution moléculaire) (Coutouly, 2006).

4.2. La méthode de parcimonie

La méthode de parcimonie utilise comme données, pour la construction phylogénétique, le résultat d'un alignement multiple. Seuls les sites informatifs de l'alignement multiple sont pris en considération. On appelle site informatif dans un alignement multiple toute position de cet alignement qui comprend au moins deux nucléotides différents et dont chacun est répété au moins deux fois (Djekoun et Hamidechi, 2013). L'arbre construit par la méthode de parcimonie ne tient compte que de ces sites informatifs.

L'exemple suivant indique les sites informatifs dans un alignement multiple dans les positions 2, 3 et 7 :

	1	2	3	4	5	6	7
individu 1	C	G	T	G	A	A	A
individu 2	C	G	T	G	A	C	A
individu 3	C	C	A	A	A	C	C
individu 4	C	G	A	A	C	-	G
individu 5	C	C	C	C	T	G	C

La méthode de parcimonie construit tous les arbres possibles entre les cinq individus, mais ne retient que l'arbre le plus parcimonieux. Celui-ci contient le nombre le plus petit de mutations entre ces individus.

1- MATERIEL

1.1. Origine des séquences

Les huit espèces retenues dans cette étude ont été sélectionnées sur la base de la disponibilité de leurs séquences nucléotidiques soumises dans la banque de données nucléotidiques Genbank hébergée sur le portail du National Center for Biotechnology Information (NCBI¹). Ces espèces sont :

<i>R.leguminosarum</i>	<i>R.etli</i>
<i>R.gallicum</i>	<i>R.multihospitium</i>
<i>R.giardinii</i>	<i>R.pisi</i>
<i>R.leucaenae</i>	<i>R.tropici</i>

Un total de 56 séquences nucléotidiques des différents gènes a concerné cette analyse. Les gènes retenus sont :

1. Gène ARNr 16S
2. Gène *nodC*
3. Gène *nifH*
4. Gènes ménager :
 - a. la sous-unité α de l'ADN polymérase (*rpoA*),
 - b. la sous-unité β de l'ADN polymérase (*rpoB*),
 - c. la sous-unité β de l'ADN gyrase (*gyrB*),
 - d. et la recombinaise A (*recA*).

Les structures primaires des séquences de ces gènes sont portées sur l'annexe N° 2 avec leurs codes d'accèsion de Genbank (Annexe N°3).

¹ <http://www.ncbi.nlm.nih.gov/genbank/>

1.2. Outils d'exploitation des séquences

1.2.1. Algorithme ClustalW

ClustalW est un programme de réalisation d'alignement multiple disponible en ligne mais non téléchargeable. À partir d'un ensemble de séquences (taxons), l'algorithme mesure les meilleures identités (score de l'alignement). Il fait ressortir la meilleure paire de séquences ayant le meilleur score d'alignement, puis passe à la troisième séquence qui présente le meilleur score d'alignement par rapport à la paire déjà alignée. L'algorithme continue toutes les comparaisons pour finir avec la dernière séquence à inclure dans l'alignement multiple (Tagu et Risler, 2010). Ainsi, les séquences les plus identiques sont alignées en premier puis l'alignement progresse vers les séquences les plus distantes ou les plus divergentes. Le nombre d'alignements possibles entre n séquences est donné par la relation :

$$\frac{n \times (n - 1)}{2}$$

1.2.2. MEGA 4.0

Le programme MEGA a été utilisé pour construire les arbres phylogénétiques entre les différents taxons. C'est un logiciel libre de droits et téléchargeable à partir de plusieurs sources Internet disponibles sur différents moteurs de recherches (google, yahoo, altavista, ...) (Annexe N°5).

2- METHODES

2.1. L'alignement multiple

L'alignement multiple fait ressortir des zones de forte similarité et / ou des zones de faible similarité. Celles-ci s'expliquent par des taux de mutations (substitutions ou insertions / délétions) très élevées entre les séquences alignées. Au contraire, les zones de forte similarité retrouvées dans l'alignement multiple indiquent une forte conservation des nucléotides (ou des acides aminées) entre ses séquences (Djekoun et Hamidechi, 2013).

La qualité de l'AM de nos séquences a été appréciée positivement lorsque les zones de forte conservation sont plus importantes en taille (paire de base) par rapport à celles de faible conservation.

L'alignement multiple a concerné les séquences de chacun des gènes retenus pour cette étude. Ces séquences de gènes ont été enregistrées sous le format FASTA pour pouvoir les utiliser sur le

logiciel MEGA. Seules, les séquences des gènes de ménage ont été additionnées les unes aux autres pour former une seule séquence et ce pour chacune des espèces retenues pour cette analyse (Annexe N°7).

2.2. Paramétrage de ClustalW2²

Dans une première étape, les huit séquences de chacun des gènes ont été soumises au programme de clustalW2 en tenant compte, dans une deuxième étape, des réglages de ses paramètres comme suit :

- Alignement type : slow (lent) au lieu de Fast (rapide) car ce dernier est approximatif.
- DNA Weight Matrix (matrice de poids d'ADN) : choisir la matrice IUB qui tient compte de l'ambiguïté du code génétique (voir Annexe N°4) alors que la matrice clustalw ne tient compte que des identités (match scoré à 1) et des différences (mismatch scoré à 0)
- Gap Open ou Ouverture de gap : scorée à la valeur de -10. Une ouverture de gap est la première brèche d'une série de trous consécutifs observés sur un alignement : A T T C G - - - C T A CTT : Le premier gap après le nucléotide G est l'ouverture de gap. Les trois suivants sont appelés extension de gap.
- extension de gap : Dans l'exemple précédent, ce sont les indels qui suivent immédiatement l'ouverture de gap. Chacun d'eux est pénalisé avec la valeur de score de - 0.2
- Itération : alignement
- Numiter : 1
- Clustering : NJ
- Format : clustalW / numbers
- Ordre : aligné

Les séquences alignées de ces procédures sont revérifiées manuellement et des ajustements sont apportés à cet alignement.

2.3. Construction des arbres phylogénétiques

Après avoir aligné les séquences nucléotidiques des différents gènes, les arbres phylogénétiques sont construits par le programme MEGA 4.0 en utilisant la méthode neighbor-joining (NJ). La validité des différents clades des phylogrammes a été testée par le bootstrap (1000 répétitions ont été réalisées). La correction des distances a été effectuée par la méthode de Kimura à deux paramètres avant de procéder à la construction phylogénétique

² <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

2.3.1. Méthode de Kimura à deux paramètres

Cette méthode tient compte de la proportion entre le nombre de transitions (α) et transversions (β) qui sont différents ($\alpha \neq \beta$), les transitions et les transversions ne sont pas équivalentes en termes de proportions. Les transitions étant plus nombreuses (Djekoun et Hamidechi, 2013). Si \mathbf{P} est la fréquence des transitions, \mathbf{Q} la fréquence des transversions et \mathbf{d} la distance, la formule de cette méthode est la suivante :

$$d_{ij} = -\frac{1}{2} \text{Ln}(1 - 2\mathbf{P} - \mathbf{Q}) + \frac{1}{4} \text{Ln}(1 - 2\mathbf{Q})$$

2.3.2. Test de robustesse des phylogrammes (bootstrap)

L'analyse de la robustesse des dendrogrammes par bootstrap est utilisée pour compléter la construction de l'arbre phylogénétique. Elle permet de déterminer si l'ensemble des données (séquences alignées) confirme les relations de parenté existant entre les différents taxons analysés (Efron, 1979 ; Felsenstein, 1985).

Le bootstrap consiste à construire plusieurs arbres à partir de plusieurs échantillons aléatoires tirés de l'alignement multiple, avec remise des caractères. Chaque échantillon de l'alignement multiple aura la même taille que le sous-ensemble de départ mais ne comportera qu'une partie des caractères (Vervoort, 2005). L'opération de construction est donc réalisée 100 à 1000 fois. La valeur de bootstrap correspond à sa fréquence d'obtention sur les différentes topologies obtenues. (Kitouni, 2007).

1. RESULTATS

L'alignement multiple a montré que les séquences des gènes diffèrent par rapport à leurs structures primaires et par rapport à leurs longueurs. Ces constats expliquent la présence des gaps (brèches) dans les différents alignements. Les nombres de gaps sont différents d'un alignement multiple à un autre en fonction de la nature des familles de gènes comparés lors de cette analyse (Tableau 6).

Nous constatons un minimum dans le nombre de gaps dans la famille du gène *NifH* ($n = 4$) comparativement au reste des séquences. Cependant, nous constatons un nombre important de gaps dans le cas de l'alignement multiple des gènes ménagers ($n = 292$).

Tableau 6 : Qualité des alignements multiples des différentes familles de gènes

Familles de gènes	Nombres de gaps
ARNr 16S	41
Ménagers	292
<i>nodC</i>	16
<i>nifH</i>	4

1.1. Arbre phylogénétique du gène ARNr 16S

La figure6 montre l'arbre phylogénétique obtenu à partir des séquences du gène ADNr 16S construit selon l'algorithme NJ. L'arbre est divisé en deux clades A et B.

Le clade A comprend trois groupes de gènes différents (I, II, III). Le groupe I renferme les souches *R. pisi*, *R. etli* et *R. leguminosarum*. A l'intérieur de ce groupe, notons la forte identité des deux individus *R. pisi* et *R. etli* ($d = 0,0018 + 0,0021 = 0,0039$; soit 0,39% de différence entre leurs deux séquences) par rapport à *R. leguminosarum*. Celui-ci présente une distance de 0,0092 (soit $0,0045 + 0,0026 + 0,0021 = 0,0092$) par rapport à *R. etli* et une distance de 0,0089 ($0,0045 + 0,0026 + 0,0018 = 0,0089$) par rapport à *R. pisi*.

La valeur de bootstrap du clade *R. pisi*, *R. etli* est de 69%. Ceci exprime que sur 1000 arbres construits aléatoirement à partir de différentes portions de l'alignement multiple, le clade *R. pisi*, *R. etli* revient dans 690 cas. Cette valeur de bootstrap exprime la robustesse de l'arbre obtenu d'une part et la forte identité des deux souches d'autre part.

Le groupe II est constitué des individus *R. leucaenae*, *R. multihospitium* et *R. tropici*. Ce groupe de gènes est éloigné du groupe I d'une distance de 0,0160 (soit 0,0113 + 0,0047).

Le groupe III est constitué de la seule espèce *R. giardinii*. Sa distance avec le groupe I est de 0,0436 (soit 0,0299 + 0,0024 + 0,0113) et est de 0,0370 avec le groupe II.

Le clade B, renferme *R. gallicum* qui présente la séquence ADNr 16S la plus divergente de toutes les *Rhizobium* ayant concerné cette analyse.

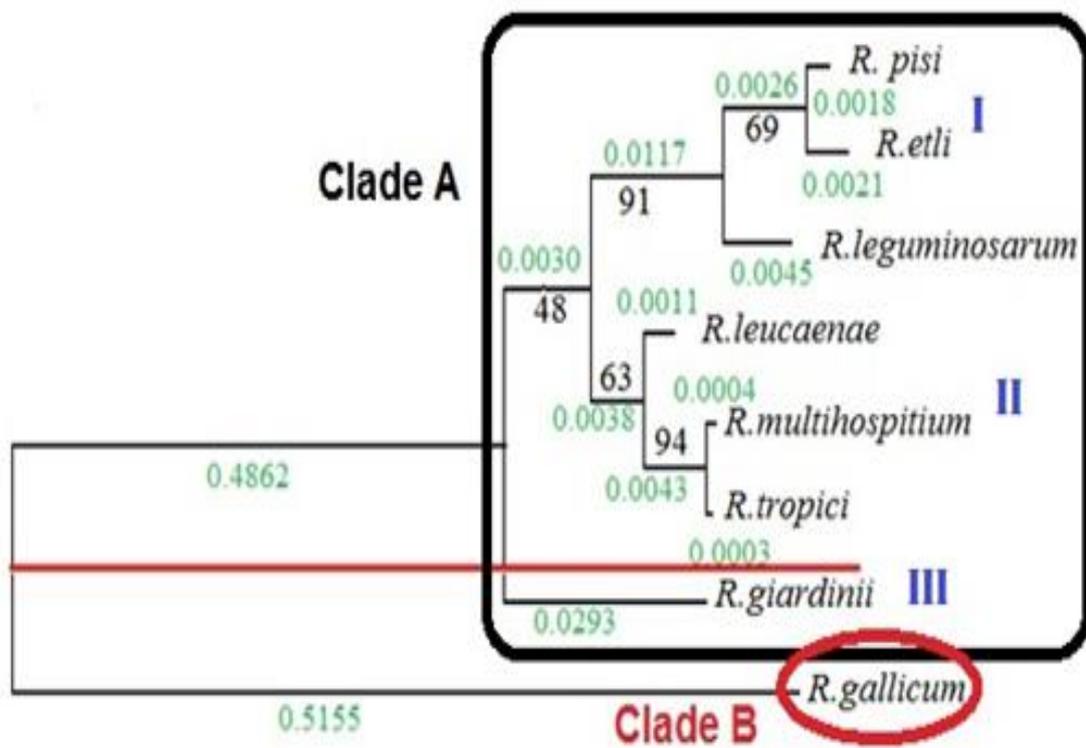


Figure 6 : Arbre phylogénétique du gène ADNr 16S construit par la méthode NJ

1.2. Arbre phylogénétique des gènes de ménage

L'arbre de séquences concaténées des gènes ménagers (*rpoA*, *rpoB*, *gyrB*, *recA*) est illustré dans la figure 7.

Cet arbre est également partagé en deux clades différents A et B. Le premier clade regroupe cinq espèces qui sont : *R. etli*, *R. gallicum*, *R. tropici*, *R. leucaenae* et *R. multihospitium*. La souche *R. etli* et *R. gallicum* sont dans le même clade. La souche *R. tropici* est plus proche du clade *R. etli* et *R. gallicum* comparativement aux deux souches *R. leucaenae* et *R. multihospitium*.

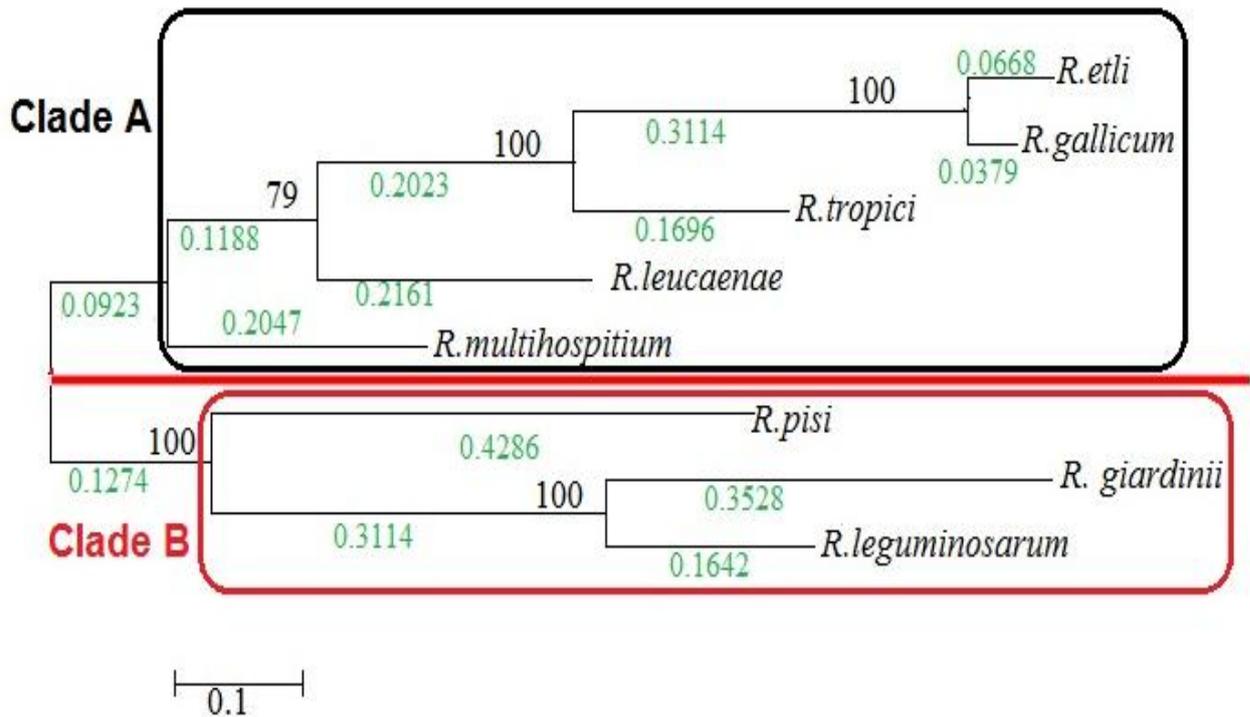


Figure 7: Arbre phylogénétique des gènes ménagers construit par la méthode NJ

1.3. Arbre phylogénétique de gène *Nod C*

L'arbre de gène *Nod C* (figure 8) est divisé en deux clades (A et B). Le clade A comprend *R. pisi*, *R. multihospitium*, *R. etli*, *R. leucaenae* et *R. tropici*. Le clade B est constitué d'une association de *R. gallicum* avec *R. leguminosarum* et *R. giardinii* avec la valeur de bootstrap de 100%.

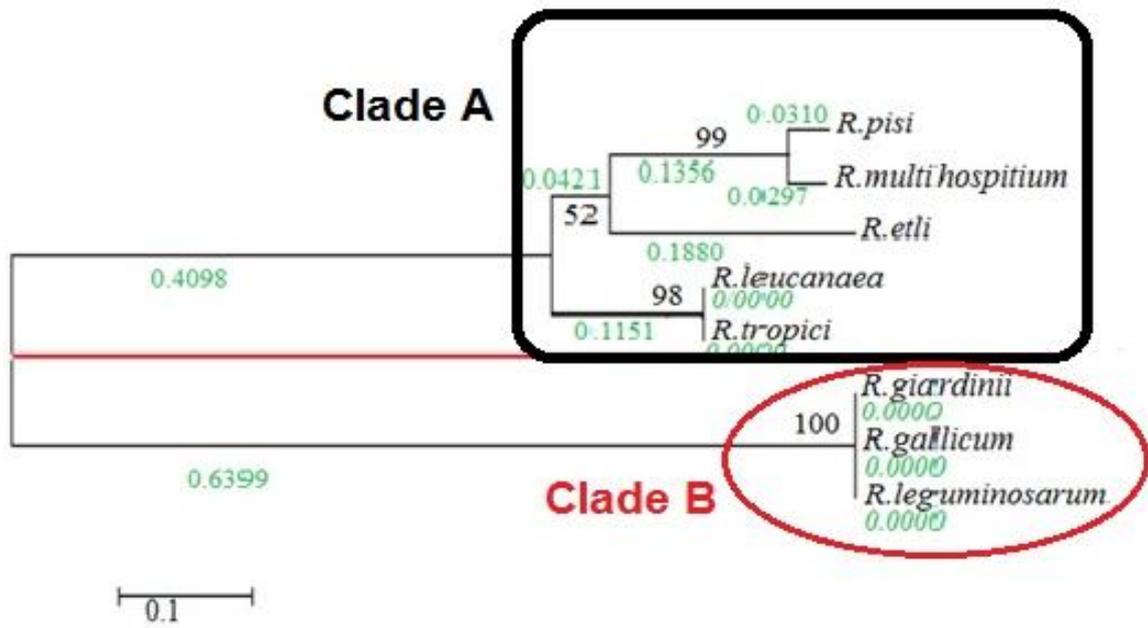


Figure 8 : Arbre phylogénétique de gène *nod C* construit par la méthode NJ

1.4. Arbre phylogénétique de gène *nif H*

La figure 9 représente l'arbre phylogénétique de gène *nif H* qui se divise en deux clades (A, B). Le clade A regroupe sept souches. Le deuxième clade renferme la seule espèce *R. leguminosarum*.

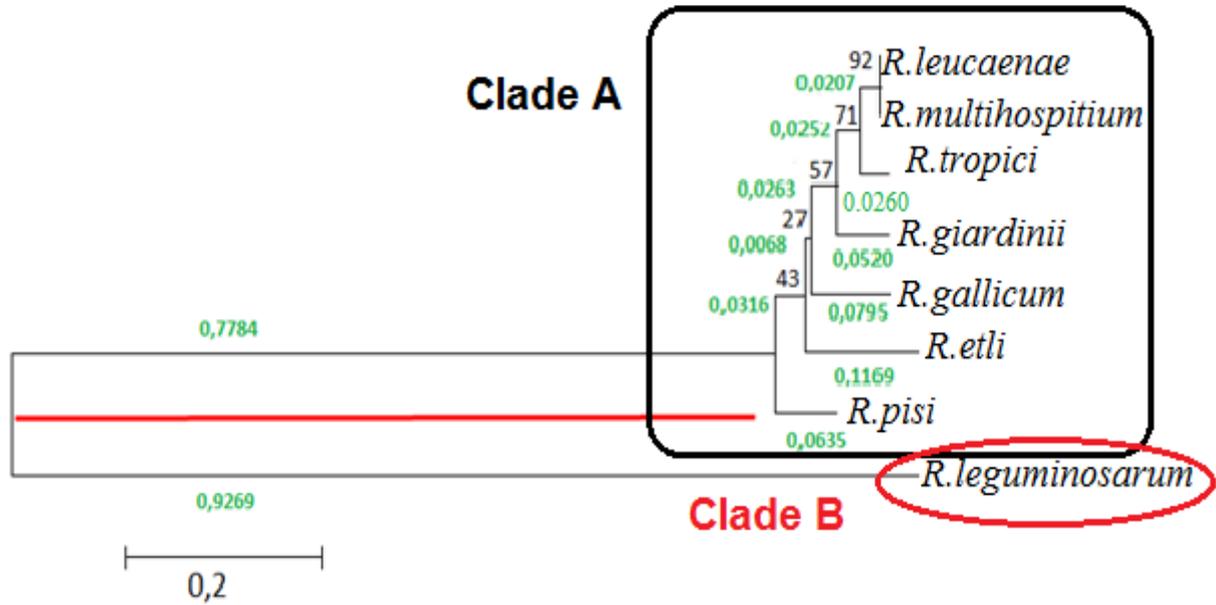


Figure 9 : arbre phylogénétique de gène *nif H* construit par la méthode NJ

2. DISCUSSION

Les constructions phylogénétiques apportent un supplément d'informations dans la taxonomie du fait qu'elles s'intéressent à des données moléculaires, et donc plus spécifiques, telles que les ADNr 16S ou encore les gènes de certaines régions très conservées le long de l'évolution des microorganismes (Cytochrome C, facteurs de transcription, les gènes de ménage, etc.). Cependant, il n'y a pas encore de consensus à proprement dire pour utiliser telle ou telle famille de gènes (Willems, 2006).

Les familles de gènes s'expriment différemment entre les microorganismes non seulement à cause de leur rang systématique, mais également à cause des influences que leur font subir leurs environnements respectifs.

Dans le cas de cette analyse, il a été constaté que les séquences expriment des topologies différentes selon les familles de gènes considérées. La comparaison de l'arbre 16S avec celui des gènes de ménage fait ressortir deux clades principaux dans les deux cas. Les clades A et B de l'arbre 16S sont totalement disproportionnés avec sept souches contre une seule *R. gallicum*. Cependant, les clades A et B de l'arbre des gènes de ménage sont plus équilibrés. Nous observons cinq souches en clade A contre trois en clade B.

Des valeurs de bootstrap de 100% ont été observées dans le cas de l'arbre des gènes de ménages. La plus forte valeur de bootstrap dans le cas de l'arbre 16S est de 94% et concerne le clade *R. multihospitium* et *R. tropici*.

Ces valeurs marquent une différence dans la qualité des deux arbres, mais ne confirment pas encore une réalité systématique, car la valeur de bootstrap, à elle seule, reste insuffisante pour confirmer l'homologie des espèces (Margulies, 2003).

Les deux arbres construits par les gènes *nifH* et *nodC* ont montré une ressemblance dans la disposition de la souche *R. leguminosarum* qui se situe dans les deux clades B des deux arbres. De plus *R. giardinii* et *R. gallicum* sont relativement conservées dans les deux arbres; ainsi que les deux espèces *R. etli* et *R. pisi*.

Globalement, ces différences seraient dues à la qualité des séquences. En effet, les longueurs des séquences soumises dans la base de données GenBank ne présentent pas les mêmes longueurs ; ce

qui pourrait affecter considérablement la qualité des alignements avec l'apparition des gaps qui ne font que pénaliser la topologie de l'arbre construit (Perrière, 2012). Plusieurs séquences partielles sont soumises dans les bases de données et acceptées sans le moindre contrôle. Le nombre de séquences soumises a augmenté de manière exponentielle ces dernières années vues le développement spectaculaire des outils du génie génétique (dont la PCR et le séquençage). En consultant les statistiques éditées sur le site du NCBI, on constate que plus de neuf milliards de nucléotides constituent l'ensemble des séquences partielles et totales soumises à GenBank.

De plus, il est à noter que la qualité "nucléotidique" des séquences n'est pas contrôlée au moment de la soumission des séquences sur GenBank et sur les deux autres bases de données, à savoir European Molecular Biology Laboratory (EMBL) et Dna Data Base of Japan (DDBJ). En effet, beaucoup de séquences montrent des ambiguïtés dans leurs nucléotides: Autres les caractères fondamentaux des quatre nucléotides (A, C, G et T), les séquences soumises contiennent d'autres types de codification tels que le caractère Y : Confusion entre C et T, ou encore R : Confusion entre A et G, etc. Ces caractères, non universels, affectent les scores des alignements et donc conduisent à une mauvaise interprétation des résultats des alignements et des arbres construits.

Un autre fait peut affecter la fiabilité des arbres phylogénétiques. Les méthodes de construction sont variables et adaptées spécifiquement aux différents types de données phylogénétiques. Dans la méthode des distances (adoptée dans cette analyse), les arbres obtenus ne sont pas racinés. Les taxons comparés par leurs seules séquences, n'ont pas une séquence "out group" ou racine avec laquelle une comparaison supplémentaire pourrait donner plus d'informations et de crédibilité à l'arbre phylogénétique (Aguileta, 2008).

Cette faible concordance entre les topologies des arbres obtenus n'est pas une différence absolue entre les huit taxons analysés. Elle exprime une divergence des différentes familles de gènes. Ainsi, si un groupe monophylétique de taxons est observé dans un clade d'un arbre donné (de 16S par exemple), il peut être dissocié et partagé en différents clades dans un deuxième arbre (gènes de ménage).

L'une des meilleures utilisations des séquences nucléotidiques est la méthode MLST (Multi Locus Sequence Typing). Dans cette technique d'alignement, il est question de concaténer plusieurs séquences de gènes de ménage en plus du gène 16S.

Pour une plus significative phylogénie, une autre technique est à utiliser : l'approche MALDI-TOF. Cette branche de métagénomique apporte plus d'informations car elle tient compte du séquençage total du génome des espèces bactériennes analysées dans un échantillon donné.

CONCLUSION

La phylogénie moléculaire est une approche émergente dans l'étude de la distribution des individus biologiques. Son utilisation devient une nécessité pour la mise en place des différents groupes (taxons) et la mise en évidence de leurs interrelations génétiques.

Lors de cette analyse, nous avons constaté une différence structurale entre les différentes familles de gènes : 16S, Keeping House Gene, *nodC* et *nifH*. Ainsi, le gène 16S, par exemple ne présente pas la même structure primaire chez les huit espèces retenues pour cette analyse. L'alignement multiple a révélé des disconcordanances entre les huit séquences comparées. Hormis l'espèce *R. gallicum* qui a exprimé de fortes identités avec les autres espèces (annexe 8), les distances entre les séquences 16S n'ont pas révélé de fortes homologues entre les espèces de *Rhizobium*

La construction des arbres a conduit à des topologies distinctes d'une famille de gènes à une autre. Seuls les gènes *nodC* et *nifH* ont conduit à une similarité des clades. Ainsi, il serait plus approprié de se pencher sur le gène *nifH* pour des constructions phylogénétiques en plus des gènes classiquement utilisés dont 16S et *nodC*.

Il est plus adéquat de procéder également à la concaténation d'au moins sept séquences House Keeping Genes au lieu de quatre comme dans le cas de cette analyse, afin d'aboutir à une topologie plus crédible, surtout dans le cas des isolats non identifiés systématiquement.

Au terme de cette analyse, nous préconisons le respect des procédures microbiologiques standards depuis les techniques d'échantillonnage jusqu'aux étapes de la biologie moléculaire ; ce qui aura permis une identification polyphasique des isolats.

REFERENCE BIBLIOGRAPHIQUES

Abdelmadjid I. (2010). *Fonction de l'AmtB dans la régulation de la nitrogénase chez Rhodobacter capsulatus*. Mémoire de maitre ès science. Faculté de médecine. Université Montréal. pp : 14 – 15

Aguileta G. (2008). *Introduction à la phylogénie*. Ed. MIG (Mathématique, Informatique et génome). INRA. France. 92p.

Amarger N., Macheret V., Laguerre G. (1997). *Rhizobium gallicum* sp. nov. and *Rhizobium giardinii* sp. nov. from *Phaseolus vulgaris* nodules. *Int. J. Syst. Bacteriol.* **47** (4) : 996 - 1006. In : El-Hilali I. (2006). *La symbiose Rhizobium – Lupin : Biodiversité des Microsymbiotes et mise en évidence d'une Multi-Infection Nodulaire chez Lupinus luteus*. Thèse Doctorat. Université Mohammed V – Agdal faculté des sciences rabat. p : 51

Aude J. C. (2010). *Alignement multiple : ClustalW*. Fiche 21. p : 88. In : Tagu D., Risler J. L. *Bio-informatique : principe d'utilisation des outils*. Ed. Quae. Versailles Cedex. France. ISBN : 978-2-7592-0870-8. 270 p

Chakrabarti S., lee M.S., Gibson H. (1981). Diversity in the nutritional requirements of strains of various Rhizobium species. *Soil Biol. Biochem.* **13** : 349 - 354. In : Lakhal A. (2011). *Effets de certains inducteurs de gènes nod (composés phénoliques) sur la croissance de Rhizobium en symbiose avec Vicia faba caractérisation et lutte biologique*. Thèse Doctorat. Université Abou bakr belkaid. Tlemcen p : 22

Chen T. H., Murata N. (2002). Enhancement of tolerance of abiotic stress by metabolic engineering of betaines and other compatible solutes. *Curr. Opin. Plant Biol.* **5** (3) : 250 -257. In : wlodarczyk A. (2010). *Recherche de signaux moléculaires végétaux impliqués dans l'induction de gènes chez la bactérie phytopathogène Erwinia chrysanthemi (Dickeya dadantii)*. Thèse Doctorat. Université de Lyon. p : 23

Coutouly G., Klein E., Barbieri E. et al. (2006). *Travaux dirigés de biochimie, biologie moléculaire et bioinformatique*. Ed. Doin. ISBN. 2-7040-1215-6. 372 p

Davet P. (1996). *Vie microbienne du sol et production végétale*. Ed INRA. Paris. ISBN : 2-7380-0648-5. pp : 149 - 152

Debellé F., Moulin L., Mangin B., Dénarié J., Boivin C. (2001). Nod Genes and Nod signals and the evolution of the *Rhizobium* legume symbiosis. *Acta Biochimica Polonica. Mini.review.* **48** (2) : 359 - 365

Denarie J., Debelle F. (1996). *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochemis.* **65** : 503 - 35. In : Mbengue M. (2010). *Perception et transduction du signal bacterien facteur Nod dans l'établissement de la symbiose Rhizobium-légumineuse : recherche et caractérisation de partenaires du Lys M-RLK LYK3, un récepteur putatif des facteurs Nod chez Medicago truncatula.* Thèse Doctorat. Université Toulouse III. P : 15

Dey P. M., Harborne J. B. (1997). *Methodes In Plant biochemistry.* Academic Press. Ed. San Diego. U S A. ISBN: 0122146743. 554 p. In : wlodarczyk A. (2010). *Recherche de signaux moléculaires végétaux impliqués dans l'induction de gènes chez la bactérie phytopathogène Erwinia chrysanthemi (Dickeya dadantii).* Thèse Doctorat. Université de Lyon. p : 22

Dupuy Y., Nougier P. (2005). *Les microorganismes. Du gène à la biosphère.* Ed Ellipses. Paris. ISBN : 9782729821937. 256 p

Efron B. (1979). Bootstrap Methods: Another look at the jackknife. *Ann. Statist.* **7** (1) : 1 – 28

Elkan G. H., Kurik I. (1968). Nitrogen, energy and vitamin nutrition of *Rhizobium japonicum*. *J. Appl. Bactériol.* **31** : 399 - 404. In : Lakhal A. (2011). *Effets de certains inducteurs de gènes nod (composés phénoliques) sur la croissance de Rhizobium en symbiose avec Vicia faba caractérisation et lutte biologique.* Thèse Doctorat. Université Abou bakr Belkaid. Tlemcen. p : 22

Fellay R., Rochepeau P., Broughton W.J. (1995). Signals to emanating from *Rhizobium* largely control symbiotic specificity. In : Singh S., Singh R. P., Kohmoto K. Pathogenesis and host specificity in plant diseases. *Histopathological. Biochimical. Genetic and Molecular Bases. Procaryotes.* **1** : 199 - 220

Felsenstein J. (1981). Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution.* **35** (6) : 1229 – 1242

Fisher H. M. (1994). Genetic regulation of nitrogen fixation in *Rhizobia*. *Microbiol. Rev.* **58** : 352 - 386

Frank B. (1889). Uber die Pilzsymbiose der Leguminosen. *Ber. Dtsc. Bot. Ges.* **7**: 332 - 346. In : El-Hilali I. (2006). *La symbiose Rhizobium – Lupin : Biodiversité des Microsymbiotes et mise en évidence d'une Multi-Infection Nodulaire chez Lupinus luteus*. Thèse Doctorat. Université Mohammed v– AGDAL. Rabat. p : 51

Fukushima M., Kakinuma K., Kawaguchi R. (2002). phylogenetic analysis of *Salmonella*, *Shigalla*, and *Escherichia coli* strains on the basis of *gyrB* gene sequence. *j.Clin. Microbiol.* **40** (8): 2779 - 2785

Gage D. J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev.* **68**(2) : 280 - 300

Gagnon H., Ibrahim R. K. (1998). Aldonic acids: a novel family of *nod* gene inducers of *Interact.* **11** (10) : 988 - 998. In : Wlodarczyk A. (2010). *Recherche de signaux moléculaires végétaux impliqués dans l'induction de gènes chez la bactérie phytopathogène Erwinia chrysanthemi (Dickeya dadantii)*. Thèse Doctorat. Université de Lyon. p : 23

Gharzouli R. (2006). *Influence d'agents mutagènes, les rayons Ultra-violet, sur la nodulation et les caractères phénotypiques de quelques espèces de Rhizobium sp.* Mémoire de Magister. Université Constantine 1. p : 13

Graham P. H., Sadowsky M. J., Keyser H. H., Barnet Y. M., Bradley R. S., Cooper J. E., *et al.* (1991). Proposed minimal standards for the description of new genera and species of root-and stem- nodulating bacteria. *Int. J. sys. bacteriol.* **41**(4). p : 582 - 587

- Hamidechi M. A., Djekoun A. (2013). *Cours de phylogénie moléculaire*. [En ligne]. <http://www.umc.edu.dz/index.php/2013-01-21-15-33-40/cours-en-ligne> . Consulté le Mai 2014.
- Han T.X., Wang E.T., Wu L. J., Chen W. F., Gu J. G., Gu C.T. *et al.* (2008). *Rhizobium multihospitium* sp. nov., isolated from multiple legume species native of Xinjiang, China. *Int. J. Sy. Evol. Microbiol.* **58** (7) : 1693 - 1699
- Haukka k., Lindstrom K., Young P.W. (1998). Three Phylogenetic Groups of *nodA* and *nifH* Genes in *Sinorhizobium* and *Mesorhizobium* Isolates from Leguminous Trees Growing in Africa and Latin America. *Appl. Environ. Microbiol. Rev.* **64** (2) : 419 - 426
- Hopkins W. G. (2003). *Les plantes et l'azote*. Chapitre 6. P : 99 - 120. *Physiologie végétale*. Ed. De boeck. Bruxelles. *Physiologie végétale*. 495 p
- Jensen H. L., Schroeder M. (1965). Urea and biuret as nitrogen sources for *Rhizobium* spp. *J. Appli. Bacteriol.* **28** : 473 - 478
- Jordan D.C., (1984). *Rhizobiaceae*. In : *Bergey's Manual of Systematic Bacteriology*. Krieg N. R., Holt J. G. Vol : 1. Ed. Williams and Wilkins. ISBN. 10: 0-387-24145-0. Baltimore. pp. 234 - 245
- Kitouni M. (2007). *Isolement de bactéries actinomycétales productrices d'antibiotiques à partir d'écosystèmes. Identification moléculaire des souches actives et caractérisation préliminaire de substances élaborées*. Thèse Doctorat. Université Constantine 1. pp : 42 – 71
- Louvrier P., Laguerre G., Amarger N. (1996). Distribution of symbiotic genotypes in *Rhizobium leguminosarum* biovar *viciae* populations isolated directly from soils. *Appl. Env. Microbiol.* **62** : 4202 - 4205. In : BA S. (2001). *Caractérisation moléculaire de la diversité symbiotique des souches isolées d'Acacia tortilis subsp. Raddiana*. Thèse Doctorat. Université Cheikh Anta Diop Dakar. La république du Sénégal. p : 19
- Margulies E. H., Blanchette M, HausslerD, Gren ED. (2003). Comparative Sequencing Program. Identification and characterization of multi-species conserved sequences. *Genome Res.* 13(12):2507-18. PMID: 14656959

Martinez-Romero E., Segovia L., Mercante F. M., Franco A. A., Graham P., Pardo M. A. (1991). *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris* L. Beans and *Leucaena* sp. Trees. *Int. J. Syst. Bacteriol.* **41** (3) : 417 - 426

Nour S. M., Cleyet-Marel J. C., Beck D., Effose A., Fernandez M. P. (1994). Genotypic and phenotypic diversity of *Rhizobium* isolated from chickpea (*Cicer arietinum* L). *Can J. Microbiol.* **44** : 345 – 354

Perrière G. (2012). Phylogé nie Molé culaire. Laboratoire de Biométrie et Biologie Evolutive. UMR CNRS n° 5558. Université Claude Bernard. Lyon 1.

Philippe G. (2009). *Automatisation de processus d'annotation génomique contrôlée par système expert*. Thèse Doctorat. Université de Provence. Marseille. p : 13

Pleckaityte M. (2010). Alzheimer's disease: a molecular mechanism, new hypotheses, and therapeutic strategies. *Medicina (Kaunas)*. **46** (1) : 70 - 6

Ramyrez-Bahena M. H., Garcia-Fraile P., Peix A., Valverde A., Rivas R., Igual J. M. *et al.* (2008). Revision of the taxonomic status of the species *Rhizobium leguminosarum* (Frank 1879) Frank 1889AL, *Rhizobium phaseoli* Dangeard 1926AL and *Rhizobium trifolii* Dangeard 1926AL. *R. trifolii* is a later synonym of *R. leguminosarum*. Reclassification of the strain *R. leguminosarum* DSM 30132 (5NCIMB 11478) as *Rhizobium pisi* sp. nov. *Int. J. Sy. Evol. Microbiol.* **58** (11) : 2484 - 2490

Ribeiro R. A., Rogel M. A. T., Lopez-Lopez A. T., Ormeno-Orrillo E., Barcellos F.G., Martinez J. *et al.* (2012). Reclassification of *Rhizobium tropici* type A strains as *Rhizobium leucaenae* sp. nov. *Int. J. Sy. Evol. Microbiol.* **62** (5) : 1179 - 1184

Rodriguez-Valera F. (2002). Approaches to prokaryotic biodiversity: a population genetics perspective. *Env. Microbiol.* p : **4** (11) : 628 - 633

Romarc F. (2006). *Dico de Bio*. Ed. De beock. Bruxelles. ISBN : 2-8041-5248-0. 257 p.

- Rosenberg C. (1997). *Signaux symbiotiques chez Rhizobium*. Chapitre 8. pp : 149 - 161. In : Gudry J. F. M. *Assimilation de l'azote chez les plantes : aspects physiologique, biochimique*. Ed. INRA. Paris. ISBN : 2-7380-0716-3. 419 p
- Sadowsky J. M., Keyser H. H., Bohloulh B. (1983). Biochemical Characterization of fast- and Slow-Growing Rhizobia that Nodulate Soybeans. *Int. J. Syst. Bacteriol.* **33** (4) : 716 – 722
- Saitou N., Nei M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Boil. Evol.* **4** : 406 - 425
- Segovia L., Young J. P. W., Martinez-Romero E. (1993). Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli* sp. nov. *Int. J. Syst. Bacteriol.* **43** (2) : 374 - 377
- Shahzad F., Shafee M., Abbas F., Babar S., Tariq M. M., Ahmad Z. (2012). Isolation and biochemical characterization of *Rhizobium meliloti* from root nodules of alfalfa (*Medicago sativa*). *J. Anim. Plant Sci.* **22** (2) : 522 - 524
- Sawada H., Suzuki P., Matsuda I., Saitou N. (1999). Phylogenetic Analysis of *Pseudomonas syringae* Pathovars Suggests the Horizontal Gene Transfer of *argK* and the Evolutionary Stability of *hrp* Gene Cluster. *J. Mol. Evol.* **49** : 627 - 644
- Somasegaran P., Hoben H. J. (1994). *Handbook for Rhizobia*. Ed. Springer- verlag. New York. Inc. ISBN: 3-540-94134-7. 450 p
- Spaink H. P., Kondorosi A., Hooykaas P. J. J. (1998). *The Rhizobiaceae: molecular biology of model plant associated bacteria*. Kluwer Academic Publishers. 566 p. In : Domergue O. (2006). *Diversité des rhizobia associés à ononis repens : une légumineuse adaptée aux milieux méditerranéens*. Mémoire. École Pratique des Hautes Études. p : 11
- Tan I. K. P., Broughton W. J. (1981). Rhizobia in tropical legumes XIII Biochemical basis of acid and alkali reactions. *Soil. Biology. Biochem.* **13** : 389 - 393

- Terefework Z. (2002). *Diversity and Phylogeny of Rhizobium galegae, and reflections on molecular evolution of Rhizobium-legume symbiosis. Department of Applied Chemistry and Microbiology. Université Helsinki.* 52 p
- Urban J. E. (1979). Nondividing, bacteroid-like *Rhizobium trifolii*: in vitro induction via nutrient enrichment. *Appl. Environ. Microbiol.* **38** : 1173 - 1178
- Van Rhijn P., Vanderleyden J. (1995). The *Rhizobium* plant symbiosis. *Microbiol. Rev.* **59** (1): 124 - 142
- Vervoort M. (2005). *Phylogénie moléculaire*. Chapitre 11. pp : 231 - 239. In : Luchetta P., Maurel M.C., Higué D. *et al.* *Evolution moléculaire*. Ed Dunod. ISBN : 2 10 006880. 331 p
- Vincent J. M. (1977). *Rhizobium General microbiology*. In Silver W.S., Hardy R. W. F. *A Treatise on dinitrogen Fixation*. Section III. Biology. Ed. Wiley- Interscience. New York. pp. 277 - 366. In : Lakhal A. (2011). *Effets de certains inducteurs de gènes nod (composés phénoliques) sur la croissance de Rhizobium en symbiose avec Vicia faba caractérisation et lutte biologique*. Thèse Doctorat. Université Abou bakr belkaid. Tlemcen. p : 22
- Watt P.M., Hickson D.I. (1994). Structure and function of DNA II topoisomérase. *biochem. J. Nov.* **303** (3) : 681 - 695
- Willems A. (2006). The taxonomy of rhizobia: an overview. *Plant and Soil.* **287** : 3 - 14
- Woese C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **512** : 221 - 271
- Watanab K., Nelson J. S., Harayama S., Kasai H. (2001). Le B Database : The gyrB database for identification and classification of bacteria. *Nucleic Acids Res.* **29** : 337 - 341. In : Bouheroum K. (2013). *Etude phylogénétique du clade Streptomyces griseus basée sur le gène 16S ARNr et Multilocus Sequence Typing (MLST)*. Mémoire. Université Constantine 1. p : 09

Yanez M. A., Catalan V., Apraiz D., Figueras M. J., Martinez Murcia A. J. (2003). Phylogenetic analysis of members of the genus *Aeromonas* based on *gyrB* gene sequences. *int. syst. Evol. Microbiol.* **53** : 875 - 883. In : Bouheroum K. (2013). *Etude phylogénétique du clade Streptomyces griseus basée sur le gène 16S ARNr et Multilocus Sequence Typing (MLST)*. Mémoire. Université Constantine 1. p : 09

Young J. M., Kuykendall L. D., Martínez-Romero E., Kerr A., Sawada H. (2001). A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie *et al.* 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. *Int. J. Sys. Evol. Microbiol.* **51** : 89 - 103

Annexe 1
Classification de *Rhizobium* sp

- *Candidatus Rhizobium massiliae*
- *Rhizobium alamii*
- *Rhizobium aggregatum*
 - non-classé *Rhizobium aggregatum* LMG 23059
- *Rhizobium alkalisoli*
- *Rhizobium arachis*
- *Rhizobium borbori*
- *Rhizobium endophyticum*
 - non-classé *Rhizobium endophyticum* CCGE 2052
- *Rhizobium etli*
 - non-classé *Rhizobium etli* 8C-3
 - non-classé *Rhizobium etli* Brasil 5
 - non-classé *Rhizobium etli* bv. *mimosae*
 - non-classé *Rhizobium etli* bv. *phaseoli*
 - non-classé *Rhizobium etli* CFN 42
 - non-classé *Rhizobium etli* CIAT 652
 - non-classé *Rhizobium etli* CIAT 894
 - non-classé *Rhizobium etli* CNPAF512
 - non-classé *Rhizobium etli* GR56
 - non-classé *Rhizobium etli* IE4771
 - non-classé *Rhizobium etli* Kim 5
- *Rhizobium fabae*
- *Rhizobium cellulosilyticum*
 - non-classé *Rhizobium cellulosilyticum* ALA10B2
- *Rhizobium cnuense*
- *Rhizobium daejeonense*
 - non-classé *Rhizobium daejeonense* L61
- *Rhizobium* genosp. AI
- *Rhizobium* genosp. AJ
- *Rhizobium* genosp. AK
- *Rhizobium* genosp. Q
- *Rhizobium* genosp. R

- *Rhizobium* genosp. TUXTLAS-24
- *Rhizobium* genosp. TUXTLAS-25
- *Rhizobium freirei*
 - non-classé *Rhizobium freirei* PRF 81
- *Rhizobium galegae*
 - non-classé *Rhizobium galegae* bv. *officinalis*
 - non-classé *Rhizobium galegae* bv. *orientalis*
- *Rhizobium gallicum*
 - non-classé *Rhizobium gallicum* bv. *gallicum*
 - non-classé *Rhizobium gallicum* bv. *gallicum* R602sp
 - non-classé *Rhizobium gallicum* bv. *phaseoli*
- *Rhizobium lusitanum*
- *Rhizobium* genosp. TUXTLAS-26
- *Rhizobium* genosp. TUXTLAS-27
- *Rhizobium* genosp. TUXTLAS-28
- *Rhizobium giardinii*
 - non-classé *Rhizobium giardinii* bv. *giardinii*
 - non-classé *Rhizobium giardinii* bv. *giardinii* H152
 - non-classé *Rhizobium giardinii* bv. *phaseoli*
- *Rhizobium grahamii*
 - non-classé *Rhizobium grahamii* CCGE 502
- *Rhizobium hainanense*
- *Rhizobium helanshanense*
- *Rhizobium herbae*
- *Rhizobium huautlense*
- *Rhizobium indigoferae*
- *Rhizobium leguminosarum*
 - non-classé *Rhizobium leguminosarum* bv. *phaseoli*
 - non-classé *Rhizobium leguminosarum* bv. *trifolii*
 - non-classé *Rhizobium leguminosarum* bv. *viciae*
- *Rhizobium leucaenae*
 - non-classé *Rhizobium leucaenae* USDA 9039
- *Rhizobium loessense*
- *Rhizobium lupini*

- non-classé *Rhizobium lupini* HPC(L)
- *Rhizobium mongolense*
 - non-classé *Rhizobium mongolense* USDA 1844
- *Rhizobium multihospitium*
- *Rhizobium nepotum*
 - non-classé *Rhizobium nepotum* 39/7
 - non-classé *Rhizobium nepotum* AGR18
 - non-classé *Rhizobium nepotum* AGR19
- *Rhizobium oryzae*
- *Rhizobium petrolearium*
- *Rhizobium phaseoli*
 - non-classé *Rhizobium phaseoli* Ch24-10
- *Rhizobium phenanthrenilyticum*
- *Rhizobium pisi*
- *Rhizobium pongamiae*
- *Rhizobium pseudoryzae*
- *Rhizobium pusense*
 - non-classé *Rhizobium pusense* L2/2/1
 - non-classé *Rhizobium pusense* M2/1
- *Rhizobium rosettiformans*
 - non-classé *Rhizobium rosettiformans* W3
- *Rhizobium selenitireducens*
 - non-classé *Rhizobium selenitireducens* ATCC BAA-1503
- *Rhizobium skierniewicense*
 - non-classé *Rhizobium skierniewicense* Ch11
- *Rhizobium soli*
 - non-classé *Rhizobium soli* DS-42
- *Rhizobium sphaerophysae*
- *Rhizobium sullae*
 - non-classé *Rhizobium sullae* WSM1592
- *Rhizobium taeanense*
- *Rhizobium taibaishanense*
 - non-classé *Rhizobium taibaishanense* CCNWSX 0483
- *Rhizobium tarimense*

- *Rhizobium tibeticum*
- *Rhizobium tropici*
 - non-classé *Rhizobium tropici* CIAT 899
- *Rhizobium tubonense*
- *Rhizobium mesoamericanum*
 - non-classé *Rhizobium mesoamericanum* CCGE 501
 - non-classé *Rhizobium mesoamericanum* STM3625
- *Rhizobium mesosinicum*
- *Rhizobium miluonense*
- *Rhizobium undicola*
 - non-classé *Allorhizobium undicola* ORS 992
 - non-classé *Rhizobium undicola* ATCC 700741
- *Rhizobium vallis*
- *Rhizobium vignae*
- *Rhizobium yanglingense*

Annexe 2

Les séquences nucléotidiques (GenBank) des différents gènes de ménage du *Rhizobium spp*

Gènes Espèces	<i>rpoA</i>	<i>rpoB</i>	<i>recA</i>	<i>gyrB</i>
R. pisi	<p>CGGTTCATCCCGTFCACAGCCT CTATTCGCGGTCAAGAAGGTGTC CTACAAGGTTGAAAATACCCGCGA AGGACAGGTTCTCGACTACGACAA GCTGAACATGACCATCGAAACCGA TGGCTCGATCACCGGCGAAGACGC CGTCGCTTTCGCGGCGCGCATCCT CCAGGATCAGCTTGGCGTCTTCGT CAACTTCGACGAGCCGAGAAGGA AACCGAAGAGGAAGCAGTACC ACTCGCTTCAACCCGGCTCTCCT CAAGAAGTTCGACGAACCTCGA GTCGGTCCGTTCCGCAAACTGCCT GAAGAACGACAACATCGTCTACAT CGGCGACCTCATTGAGAAGACCGA AGCAGAAATGCTCCGACGCGCGAA TTTTGGTTCGAAAGTTCGTAACGA AATC</p>	<p>AGAGCCCGAAAAGCGGTCTGCC GACGAAGGTTCTCAGGCCGCTTC AAGTCGGTTTTCCCGATCACCATT TCTCCGGCGTTCATGCTGGAAT TCGTGCTCCTATGAGTTCGAGCCGC CGAAGTTCGACGTCGACGATGCCG CCAGCGCGACCTGACCTATGCGGC GCCGCTGAAGGTGACGCTGCGCCT CATCGTGTTCGATATGACGAGGAT ACCGGCGCAAGTCGATCAAGGAC ATCAAGGAACAGTCCGCTACATG GGCGACATGCCGTCATGACCAACA ACGGCACGTTTATCGTCAAGCGCA CCGAGCGGCTGATCGTCTCCAGA TGCACCGTTGCCGGGCTCTTCT CGACCACGACAAGGGCAAGAGCCA CTCGTCTGGCAAACTGCTTTTTGC TGCCCGTCAATCCCTATCGCGGT TCCTGGCTCGATATCGAATTCGAC GCCAAGGACATGTTTACGCCCGT ATTACCGCCGCGCAAGATCCGG TGACGTCGCTGCTGATGCGCTCG GCATGGACGGCGAGGAAATCCFCG CACCTTCTACACGAAGTCTGCTA CAAGCGCGACGGCGAAGGCTGGCG CATTCCCTTCAAGCCGGAACCTG AAGGTTGCCAAGGCATCACCAG ATGGTTCGACGCCGATACCGCGAA GTCTGCTGGAAGCCGCGAGAAGC TGACCCCGGCTTGCTGCGCCAGC TGTCGGATAAGGGCTGAAGGCGT TGAAGGCGCGGACCGATCTTTAC GGCAACTACTCGCCGGGACATC GTCAACTACTCGACGGGTGAAATC TATCTCGAGGCGGCGACGAAATCG ACGAAAAGACGCTCGGCATCATCC TGTCGACGCGTTTCGACGAGATCC CGGTCTCGCATCGACCACATCA</p>	<p>GCAAGGCGCTTTGAAGCGCACTTC ACAGATAGAGCGGTCTCGGCAAGG GCTCGATCATGAACTCGCTCCAACG AGAAATGTTGTCGAGATCGAAACGAT TCGACCGGCTCGCTTGGCTCGATAT TGCACCTCGGTTGGTGGCTGCGGAG GGGGCGCATCATCGAGATCTACGGCC CGGAAAGCTCGGTTAAAACGACGCTG CTTTGCGAGACATTGCCGAAGCGCAG AAGAAGGGTGGCATCTGCGCTTCGT CGATGCCGAACATGCGTCTCGGCT CTATGCCCGAAGCTTGGCGTCGATC TGCAGAACCTTCTGATCTCACAGCCC GATACCGCGAACAGGCGCTTGAAATP ACCGATACGCTGGTGGCTCCGGCGC TGTCAGAACCTTCTGCTCGCTCGT CGCCGCGCTGACGCCAGCGCGGAAA TCCGAAGCGAGATGGGCGACAGCCTC CCCGCCTGACGGCCGCTGATGAGC CAGGCGCTGCGCAAGCTCACCGCTTC GATCTCGAATGCGAACA</p>	<p>GGCGTCGCGGTTTCCGTCGTAAT GCGCTTTCGCTGTGGCTGAAGCTG AAGATCCGCGCCACGACAAAGATC CACGAAATGAGTTTACCACATGGC GTGGCCGATGCTCCGCTGAAGGTT ACGGGCGAAGCGCCAAATGAGACC GGCACAGAAGTAACTTCATGCGC AGCACCGACACCTTACTATGACG GAATTCGACTACGGAACGCTGGAG CATCGCTTCGCGAGCTCGCTTTC CTGAACTCGGGCTGGGTATCCCTG CTGACTGACAAGCGCCATTCGGAT ATCAAGCAGGAAGAAGTGGTAT GACCGCGGCTCGAGGCTTTCGTC GCCTATCTCGACGAGCCAAAGAG TCGCTCGTCAAGCGGCTCGCC ATCCGCGCGAAAAAGACGGCATC ACCGTCGAAGTCGCGATGTGGTGG AACGACAGCTATCATGAGAAGCTG CTTTGCTTACCACAAACATTC CAGCGCGACGGCGCACCGCATATG GCGCGCTTTCGCGCGGCGCTGACG CGTCAGGTCGTTTCTATGCGGAC AGTTCGGGATCACCAAAAGGGAA AAGGTGACGCTGACGGCGAAGAT TGCCGTGAAGCCCTGACGGCCGTC TTGTCGGTCAAGTGGCGGATCCG AAATTCCTCGCAGACCAAGGGA CAAGCT</p>
R. leucaenae	<p>GGCGATGACAGCAAGCGCATGGTC GTGCGCAAGCAGGCGCCAGGCGTG GTAACGGTGGCGACATCCAGACG GTCGGCGACATCGAAATCCTCAAC CCCGAGCATGTATCTGACGCTC GACGAGGGCGCGAAATCCGCATG GAATTCACCGTCAACAACGGCAAG GGCTACGTCGCCGCGAGCGTAAT CGTGCAGAAATGCTCCGATCGGC CTCATCCCGTTCGACAGCCTTAT TCGCGCGTCAAGAAGGTGCTTAC AAGGTTGAAAACCCCGGAAGGA CAGGTTCTGACTACGACAAGCTG AGCATGTGATCGAAACTGATGGT TCGATCACCGCGAAGATGCGGTC GCTTTTGGCGTTCGATCCTCCAG GATCAGTTCGGCTCTTCGTCAAC TTCGACGAGCCGCGAGAAGGAGCC GAAGAAGAAGCAGTACCAGACTT GCTTCAACCCGGCGCTCCTCAAG AAGGTGGACAACTGGAAGTGTCT GTTCCGCTCGGCAACTGCCTGAAG AACGACAACATCGTCTATATCGGC GACCTCATTGAGAAGACCGAAGCA GAAATGCTCCGACGCGGAATTTT GGTCGCAAGTTCGCTGAACGAAATC AAGGAAGTTCGCTTCCATGGGC CTGCAC</p>	<p>GGCGACAAGATGGCGGCGCTCAC GGCAACAAGGTTGCTATCGCGC ATCGTGGCGGTGAGGACATGCGT TCCTCGAAGCGGTACGATGTCG ACATCGTGTGAACCCGCTCGGCG TGCCGTCGCGCATGAACGCGGTCA GATCCTCGAGACGCACTTGGCTG GGCTTTCGCCCGCATGGTTCGCCA GATCGCGAGCTGATGAAGCCTAC AAGGCAAGCGGCAACATCGAGCCT CTGCCAGACGATCGCATGGTC GTCCGCGCGTCAAGAGCGACG ACGTTACAGATGATGACGATCGT CGGTCTCGCTTGGCTGATCAGT GGAAGCGTGGTGTCCATCGCGAC CCCGTCTTTCGACGGCGCGCGCA AGTGGATGTCAACGAGATGCTCGA AATGGCGGTCTGAAACAGACCGGT CAGTCGACGCTTATGACGGTCTG ACCGCGAGCAGTTTCGACCCGAG GTCCGGTTGGCTACATCTACATGC TGAAGCTCAACCACTGCTCGACG ACAAGATCCACGCCGCTCGATCG CCCGTACTCGTCTGTCACCCAGCA GCCGCTGGGCGCAAGGCGAGTT CGGCGGTGACGCTTTCGCGAATG GAAGTCTGGGCGCTCGAAGCTTAC GGCGAGCCTACACGCTGCAG</p>	<p>TTGAAGCGCACTCTCAAATGAG CGGTGTTTCGCGAAGGCTCGATCAT GAAGCTTGGCTCGAACGAGACGTTG TCGAAATCGAAACAGTTTCCACCGGC TCTCTGAGCCTCGATATCGCGCTCG CATCGCGGTTTCCGGAAGGGCGCA TCATCGAGATCTACGACCCGAAAGC TCGGTAAGACGACGCTGGCGTTGA GACCATTGCGAAGCCAGAAGAAGG GCGGATCTGCGCTTCGTCGACCGC GAACATGCGCTCGATCCGGTCTATGC CCGCAAGCTCGGCTCGACCTGCAGA ACCTTCTGATTCGCGAGCCGATAACC GGTGAACAGGCGCTTGAATCACCGA TACACTGGTACGCTCCGCGCCGCTCG ACGTTCTCGTCTGACTCCGTTGCT GCGCTGACGCGCGCGCGAAATCGA AGGCGAAATGGCGACAGCCTGCCGG GCCTGACGCGCGCTGATGAGCCAG GGCTACCGCAAGCTCACCGCTTCGAT CTCGAAGTCC</p>	<p>TCCGCGGCTCGATGGCGTCCGTT GTTTCGCTCGTCAACGCTTTGTC GTCGCTGAAACTGAAGATCCCGC CGCCAGGCGAGGATCCATGAAATG AGCTTACCACGCGGCTCGCCGAT GCGCCGCTCAGGAGACCGCGGAC GCGCGCATGAGACGGCGACGGAA GTCAGTTCATGCCAAGCTCGCAG ACCTTACAGATGACGGAATTCGAC TATGCCACGCTGGAGCATCGCCTG CGGAGCTGGCGTTCCTGAATTCG GGGTCGGGATTCCTCCTCAGCGAC AAGCGTCACTCCGACGTCAAGCAG GAAGAAATGATCTACGATGGCGCG CTCGAGGCTTTCGCTCGCCTATCTC GATCGCGCAAGAAAGCGCTGGTC GAAAAGCGGCTCCTATCCGCGGC GAAAAGGATGGCATCACCGTCGAA GTGGCGATGTTGGTGAACGACAGC TACCAGAAAACGCTGCTCTGCTTC ACCAATAACATCCCGACGCGGAT GGCGGACGATATGGCGGGCTTC CGTGGCGACTGACGCGCCAGATC ACCTCCTATGCGACACCTCCGCG ATCACAAGAAGGAAAAGGTGACG CTGACGGGTGACGATTCGCGCGAA GGACTGA</p>

<i>R. multihospitium</i>	<p>ATGGTTCGTCGCAAGCAGGGCCCA GGCGTTGTAACGGCTGGCGACATC CAGACGGTCGGCGATATCGAAATC CTCAATCCCGAGCATGTCTATCTGC ACGCTFCGACGAGGGCCCGAAATC CGCATGGAATTCACCGTCAACAAC GGCAAGGGCTATGTGCCCGCCGAG CGCAATCGTCCGGAAGACGCTCCG ATCGGCCTTATCCCGGTTCGACAGC CTCTATTCCGCGGTCAAGAAGGTG TCCACAAAGTTGAAAAATACCCGC GAAGGACAGGTTCTCGACTACGAC AAGCTGAGCATGTCTGATCGAAACC GATGGTTCGATCAGCGGTGAAGAT GCCCTFCGTTTCGCGGCTFCGATC CTCCAGGATCAGCTTGGCGTGTTC GTCAACTTCGACGAGCCCGAGAAG GAAGCCGAAGAAGAAGCAGTACC GAGCTFCGTTTCAACCCGCGCTC CTCAAGAAGGTGGACGAACTCGAA CTGTCCGTTTCGTCGCAACTGC CTGAAGAACGACAACTCGTCTAC ATCGGCGACCTCATTCAGAAGACC GAAGCAGAAATGCTCCGACGCGG AATTTTGGTCGCAAGTTCGCTGAAC GAAATCAAGGAAGTTCTCGCTTCC ATGGCCCTGCAC</p>	<p>CGCTGTGAAGCGCAAGATCCAACCG GGGACAAGATGGCAGGCCGTACCG GGAAACAAGGGCGTGGTCTCCGCATT GTTCCGGTCGAAGACATGCCGTTCT CGAAGACGGCAGCATGTGACATC GTGCTGAACCCGTCGGCGTCCCTCG CGCATGAACCTCGGTGAGATTCTCGA AACGCACCTGGGCTGGGCTTGCGCC GGCAGGTCGCCAGATCGCGGAAT GATCGAAGCCTCAAGGCCAGCGGC AACATCGAGCCGTCGCGAAGACAT CGATCTGGTCTCGCGATGGCCCC AAGAGCGACGACGTTACAGATATG ACGATGCGTGGTCTTGGTCTGGCCG ACCAAGTGAAGCGTGGCGTTCAT CGGACCCCGTCTTCCAGCGTGA GGCGAAGCGCGTCAACCACATGCT GGCAATGGCAGGTCTCAAGCAGACC GGCCAGTCGACACTGTACGACGGT GTAGGCGAGCAGTTCGACCTCAGG TCACGGTTGGCTACATCTACATGCTG AAGCTGAACCACTGTGACACAAAG ATCCACGCCGTTGATCGGTCCGA CTCGCTCGTCAACCAGCAGCGTTGG GCGGCAAGGCGGTTGCGGCGCCAG CGCTTCCGCGAAATGGAAGTCTGGG CGCTCGAAGCTTACGGTGCAGCTAC ACGCTCAG</p>	<p>AACGGTTTCCACCGGCTCTCTGAGCC TCGATATCGCTCTCGGCATTGGCGGC TTGCCGAAGGGCGTATATCGAAATC TACGGGCGGAAAGCTCGGGTAAGAC GACGTTGGCGCTGCAGACATTGGCGG AAGCGCAGAGAAGGGCGCATCTGCG CCTTCGTCGATGCGGAACATGCGCTC GATCCGGTCTATGCCCGCAAGCTTGC GTCGATCTGCAGAATCTCTGATTTTC TCAGCCGGATACCGGCGAACAGGGCC TTGAAATCACCAGTACCTGGTACGCT CCGGCCCATCGACGTTCTCTGCTGTC GACTCCGTTGCGGCGTTCAGCCCGCG GCCCGAATCGAAGCGAAATGGGCGA CAGCCTGCCGGCCTGACGGCTCGCC TGATGAGCCAGGCGTTCGCAAGCCA CGGCTTCGATTCGCAAGTCCAAGTGC ATGGTGTATCTTCAACAGATCCG CATGAAGATCGGCG</p>	<p>GTCTGGCTACGCCTGAAGATCCGC CGTCACGGCAAGATCCATGAGATG AGCTTTACCCATGGTGGCCGAT GCGCCGCTGAGGAAACGGGTGAG GCAGGCGACGCGACCCGCCACCGAG GTCAGCTTACATGCCGAGCTCCGAG ACCTTACCATGATCGAGTTCGAT TACGGAACGCTGGAACATCGTCTG CGCGAGCTTGCCTTCCTGAATTC GGCGTCCGCATCTTTTGACCGAC AAGCGTCATTCCGACGTGAAGCAG GAAGAGTTGCTCTATGACGGCGGT CTGGAAGCCTTCGCTCTCTATCTC GACCGCGCCAAGAAGCCGCTGGTC CAGAAGCCGGTTCGATCCGCGAC GAAAAGGACGGCATCAGGTCGAA GTGGCGATGTGGTGAACGACAGC TACCACGAAATGTGCTCTGCTTC ACCAACACACTTCCAGCTGAAGCAG GGCGGCACACATATGGCTGGCTTC CGTCCGCTCTGACGCGTCAAGT ACCTCTATGCTGACACCTCGGGC ATCACAAGAAGGAAAGGTGACG CTGACCGGAGACGACTGCCGCGAA GGCTGA</p>
<i>R. gallicum</i>	<p>GGCCATGGAAGCGAGAATTCCTT GATTTTCGTTACGATTTGCGACC AAAAATTCGGCGTTCGGAGCATTT TGCTTCGCTTCTGAATGAGGTC GCCGATGTAGACGATGTTGTCGTT CTTACGGCAGTTTGGCAGCGGAC CGACAGTTCGAGTTCGCTACCTT CTTGAGGAGCGCCGGTTGAAAGC GAGCTCGGTGACTGCTTCTCTTC GGCTTCTTCGCGGCTCGTCGAA GTTGACGAAGACGCCAAGCTGGTC CTGGAGGATGCGCGCCGAAAAGC GACGGGCTTTCGCGGAAACCGA GCCATCGGTTTCGATGGTCATCGT CAGCTTGTGCTAATCGAGAACCTG TCCTTCGCGGTTGTTTCAACCTT GTAGGACACCTTCTGACCGCGA ATAGAGGCTGTCAGCCGGATCAG GCCGATCGGAGCATCTCCGACG ATTGCGTTCGCGCGAACAATAGCC CTTGCCTTGTGACGGTGAATTC CATGCGGATCTCGGCACCTTCGTC GAGCTGCGAGATGACATGCTCAGG GTTGAGGATTTGATATCGCCGAC CGCTGGATGTGCGCAGCCGTCAC AACGCCCGGCCCTGCTTGGCGAC GACCATGCGCTT</p>	<p>GTGAAGCGCAAGATCCAGCCGGGC GACAAGATGGCTGGCCGTCACGGC AACAAGGGTGTCTGTCTCCGCTCG TTCCGGTCGAAGACATGCCGTTCC TCGAAGACGGCAGCATGTGACG TCGTTCTGAACCCGCTGGCGTGC TTCGCGCATGAACGTCGGCCAGAT CCTCGAGACGCACCTCGGCTGGGC TTGCGCCGGCATGGGAAGCAGATC GGTGACCTGATCGAAGCCTACAAG GCGAACGGCAATATCGAGCCGTTG CGCAAGACGATCGCGACGTTGTCG GCGACGGTCCGAAGGCCAGCAGG TCAAGGACTTCGACGATGACTCGG TTCGCGCTTCGCCACAGTGGAA GCGCGGCTTTCGATCGCAACGCC GGTCTTTCGACGGCGCCATGAAGG CGATGTAACGAGATGCTGCGTCTG GCGGGTCTCAAGGATTCGCGTCA TCGACGCTGTATGACGGCCGTACC GGCAGCAGTTCGACCGTCAAGTGA CCGTGGGCTACATCTACATGCTGA AGCTCAACACCTGGTGCAGACA GATCCACGCCGTTTCGATCGGTCC TTACTCGCTCGTACCCAGCAGCC GCTGGCGGCAAGGCCAGTTGGC GGCCAGCGCTTCGGCGAAATGGAG GTCTGGGCTCTGGAA</p>	<p>CGAGATCGAATATTTTCGACCGGTT CCTCAGCCTCGATATCGCGCTCGGCA TCGGCGGTTTGGCCAAAGGGCCGATC ATCGAGATCTACGACCGGAAAGCTC TGGCAAGACGACGCTTTCGCTGCGA CGATCGCCGAAGCGCAGAAAAAGGC GGCATCTGCGCATTCGTCGATGAGA ACATGCGCTCGATCCGCTCTATGCC GCAAGCTCGGCGTTCGATCTCCAGA CTCCTGATCTCGACCGGATACCGG CGAGCAGGCGCTCGAAATCACCATA CGCTGGTGCCTTCGGCGCCGTCGAC GTTCTCGTCTGACTCCGTTGACG GCTGACACCGCGCCGAAATCGAGG GCGAGATGGGCGACAGCTTCCGGGC CTACAGGCGCGCTGATGAGCCAGG CCTGCGCAAGCTCACGGCGTCAATC CCAAGTCAACACCATGGTCACTTC ATCAACCAGATCCGATGAAGATCGG CGTCAATG</p>	<p>TCCGGCGGCTTCGATGGCGTGGC GTCTCGGTCGTAACCGCTGTGCG GTCTGGCTGAACTGAAGATCCCG CGCAGGGTAAGATTATGAAATG AGCTTACGATGCGGCTTGGCGAT GCGCCGCTCAAGGTGACGGGTGAG GCGGGCAGCGACACAGGCACGGAA GTCAGCTTACATGCCGAGCACCAG ACCTTACGATGACGGACTTCGAT TACGGCAGCTGGAGCACCCTTC CGCGAGCTTGCCTTCTTGAATTC GGTGTCCGATTTTGTGACAGAC AAGCGTCATTTCGACATCAAGCAG GAAGAATGCGTTATGACGGCGGC CTAGAGGCTTCGTTTCTATCTC GATCGTCCGAAGAACTCTGGTC GACAAGCTGTCTCCATTCGCGGC GAAAAGGACGGAATTACCGTCGAA GTCGCAATGTGGTGAACGACAGC TATCATGAAAATGTGCTCTGCTTC ACCAACAATATTCACAGCGCGAC GGCGGCACGATATGGCCGGCTTC CGCGCCGCTGACGCGCAAGTT GTGTCTATGCGGATAGCTCCGGT ATTACCAAGAAGGAGAAAGTCACA CTCCAGGGCGAAGACTGCGCGGAG GGCCTGACGGCTGCTTTTCGGTG AAGGTGCCGATCCGAAGTCTCG</p>

<i>R. etli</i>	<p>GATTTTCGTTACGCGACTTGCAGCC AAAATTCGGCGTGCAGGACATTTTC TGCTTCGGTCTTCTGAATGAGGTC GCCGATGTAGACGATGTTGTCGTT CTTCAGGCAGTTTGCCGAACGGAC CGACAGTTCGAGTTCGTCACCTT CTTGAGGAGACCGGGTTGAAAGC GAGTTCGGTGACTGCTTCTCTTC GGTTTCCTTCTGCGCTCGTCGAA GTTGACGAAGACGCCAAGCTGATC CTGGAGGATGCGCGCCGCAAAGC GACGGCGTCTTCGCGGGTATNGA GCCATCGGTTTCGATGGTCATATT CAGCTTGTCTAGTCGAGAACCCTG GTCTCAAGGACACCGGCCAGTGA CGTCTTGACGGCCGTACCGGCGA GCAGTTCGACCGCCAGGTCACGGT GGGCTACATCTACATGCTGAAGCT CAACACCTGGTTCGAGATAAGATC CATGCCCGCTCGATCGGTCCTTAC TCGCTCGTCAACCCAGCAGCCGTTG GCGGCAAGGCGCAGTTCGGCGGTC AGCGCTTC</p>	<p>GGGCGTGGTCTCGCGCATTGTGCC TGTCGAGGATATGCCGTTCTTGA AGACGGCAGCATGTGCGAGTGTT CTGAACCCGCTCGGGTTCCTTCG CGCATGAACGTGCGCCAGATCCTG GAAACGCATCTGGGCTGGCTGTG CCGGCATGGGTCCGAGATCGCGG AATTGATCGAAGCCTACAAAGCGA ACGGCAACATCGAGCGTGCAGAA GACCATCGGCGATGTCGTCGGTGC CGGTCCGAAGCCGAGCAGGTGCA CGAATTCGACGAGACTCGGTTCTG CGTCTCGCCGACCGTGGAAAGCGC GGGCTTTCGATTGCAACCGCTGTC TTCGACGGTCCAACGAAGGCGAGC TCAACGACATGCTGCGTTCGGCGG GTCTCAAGGACACCGGCCAGTGA CGTCTTGACGGCCGTACCGGCGA GCAGTTCGACCGCCAGGTCACGGT GGGCTACATCTACATGCTGAAGCT CAACACCTGGTTCGAGATAAGATC CATGCCCGCTCGATCGGTCCTTAC TCGCTCGTCAACCCAGCAGCCGTTG GCGGCAAGGCGCAGTTCGGCGGTC AGCGCTTC</p>	<p>GTTTCGACGGGCTCGCTTGGCCTCGA TATTGCACTCGGCATTGGCGGTCTGC GAAAGGGCCGTATCATTAATCTATG GGCCGAAAGCTCGGGCAAACAACG CTTGGCTGACAGACCATTCGCGAATC GCAGAAGAGGGCGGCATTTGCGCCTT CGTCGATGCCGAACACGCGCTCGACC CTGTCTATGCCCAGAGCTTGGCGTG ATCTGCAGAACCTCCTGATCTCACAG CCAGACACGGGCGAGCAGGCGTGA AATCACCGACACGCTGTGCGCTCCGG CGTGTGACGTTCTCGTCTCGACT CGGTGCGCGGCTGACGCGCGCGGCC GAAATCGAGGCGAGATGGGCGACAGC CTGCGGCGCATGCAGGCCGCTGAT GAGCCAGGCGCTGCGCAAACCTGACGC TCAATCTCCAAGTCGAATGATGG TGATATTCATCAACAGATCCGCATG AAGATC</p>	<p>TCCGGCGGCTGCACGGGGTCCGGC GTTCCGCTCGTCAACGCGCTTTCC GTCTGGCTGAAGCTGAAGATCCGC CGCCATGGCAAGATCCATGAAATG AGCTTACCACATGGCGTGGCTGAC GCGCCGCTGAAGTTCACGGCGGAA GCGCAGACGGGACCGGCACCGAG GTGAGCTTTCATGCCGAGCAGCGAA ACCTTACCATGACGGAGTTCGAC TACGGACCGCTGGAGCATCGCCTG CGCGAGCTTGCCTTCCTGAACCTG GGCGTGGCATCTGCTGACCGAC AAGCGCCATTCCGATATCAAGCAG GAAGAATCGCTTATGACGGCGGC CTCGAGGCTTGGTTCCTGCTATCTC GACCGAGCCAAGAAGTTCGCTCGTC GACAAGCCGCTGCCATCCGCGGC GAAAAGGACGGTATCACCTCGAA GTCCGATGTGGTGGAAACGACG TACCAGAGAACGTGCTCTGCTTC ACCAACAACATCCCGAGCGGAC GGCGCACCATATGGCCGGCTTC CGCGGGCCTTACGCGCCAGGTTG GTGCTCTATGCCACAATCCCGC ATCACAAGAAGGAAAAGGTGACG CTGAGGGCGAAGATTGCCGCGAA GGTCTGA</p>
<i>R. leguminosarum</i>	<p>GATTTTCGTTACGCGACTTGCAGCC AAAATTCGGCGTGCAGGACATTTTC TGCTTCGGTCTTCTGAATNAGGTC GCCGATGTAGACGATGTTGTCGTT CTTCAGGCAGTTTGCCGAACGGAC CGACAGTTCGAGTTCGTCACCTT CTTCAGGAGACCGGGTTGNAAGC GAGTTCGGTACTGCTTCTCTTC GGTTTCCTTCTGCGCTCGTCGAA GTTNACGAAGACGCCAAGCTGATC CTGGAGGATGCGCGCCGCAAAGC GACGGCGTCTTCGCGGTTNATCGA GCCATCGGTTTCGATGGTCATGTT CAGCTTGTCTAGTCGAGAACCCTG TCCTTCGCGGTTATTTCAACCTT GTAGGACACCTTCTTACCGCGGA ATAAAGGCTGTCGACCGGGATNAG ACCC</p>	<p>GAGACCGAGGGCGGCGACCCGA CGAAGGCTTTCAGGCGGTTTCAA GTCCGTTTTCCCGATCACCGATTC TCCGCTGCATCGATGCTGGAATTC GTGTCTACGAATTCGAACCGCCG AAGTTCGACGTTGACGAAGCCGTC AGCGCGACTGACCTATGCGCGGC CGCTGAAGGTGACGCTGCGCTCA TCGTGTTGATATCGCGAGGATAC CGGCGCAAGTCGATCAAGGACAT CAAGGAACAGTCCGTCTACATGGG CGACATGCCGCTATGACCAACAAC GGCAGTTCATGCTCAAGGCGACC GAGCGGTGATCGTTTCCAGATG CACCCTTCGCGGGCTTCTTTCG ACCATGACAAGGGCAAGAGCCATT CTTCGGCAAGCTGCTCTTTGCTG CGCGCGCATCCCTATCGCGGTTTC CTGGCTCGATATCGAATTCGACGC CAAGGACATCGTCTACGCCGAT CGACCGCGCCGCAAGATCCCTGTG ACGTGCTGCTGATGGCGCTCGGC ATGGACGGCGAGGAATCTCGAC CCTTCTACACGAAGTTCGCTTACA AGCGCGACGGCGAAGGCTGGCGCA TTCCTTCAAGCCGAAACGCGAA GGGTGCCAAGCGATCACCGAGAT GGTCGACCGCATACCGGCGAAGT CGTCTGGAAGCCGCAAAAGCTG ACCCCGCGCTGCTGCGCCAGCTT TCCGACAAGGGCCTGAAGCCCTC AAGCGCCGACGACATCTTTACG GCAATTACCTTCCCGGCGACATCG TCAACTACTGACGGGTGAGATCT ACCTCGAGGCGCGACGAAATCGA CGAGAAGCGCTTGGCGTATCCT GGCGAATGGTTTCGACGAGTCCCG GTCTC</p>	<p>TGGTAGAGGANAATTCGGTGGACAAA AGCAAGGGCTTGAAGCGGCACTCTC ACAGATAGAGCGGTCGTTTCGGCAAG GCTCGATCATGAAACTCGGCTCCAAC GAGAAATGTTGTCGAGATCGAGACGAT CTCGACCGGCTCGCTTGGCTCGATA TTGCACTGCGGCTTGGTGGCTGCGG AGGGGCGCATCATCGAGATCTACGG GCCGAAAAGCTCCGGTAAGACGACCG TTGCACTGCGAGCAGATTGCCGAAGCG CAGAAGAAGGGCGGCATTCGCGCTT CGTCGATGCCGAACATGCGCTCGATC CCGTCTATGCCGCAAGCTTGGCGTC GACCTGCAGAACCTTCTGATCTCGCA GCCCGATACCGCGAGCAGGCGCTCG AAATCACCGATACGCTGGTGCCTCC GGCGCGTTGACGTTCTCGTCTGCA CTCGTCCGCGCACTGACGCGCGCG CCGAAATCGAAGGCGAAATGGGCGAC AGCTTCCCGGCTGCAGGCGACGACT GATGAGCCAGGCGCTGCGCAAGCTCA CCGCTTCGATCCAAATCGAA</p>	<p>GTCAACGCGCTGTCCGCTGCTGGTTG AAGCTGAAGATCCGCGCTCAGGAC AAGATCCATGAAATGAGCTTACC CATGGCGTCCCGACGCTCCGCTG AAGGTTACGGCGCAGCGCCGAAT GAGACCGGCACGGAAGTGAAGCTTC ATGCCGAGACCGACACCTTACC ATGACGGAGTTCGACTATGGCAGC CTGGAGCATCGCTTCCGCGAGCTC GCCTTCTGAAATTCAGGCGTCCGC ATCCTGCTGACCGACAAGCGCCAT TCCGATATCAAGCAGGAAGAAGCTG CGTTATGACGGCGGCTCGAGGCT TTCGTCGCTTATCTAGACCGCGCC AAGAAGTTCGCTCGTTCGACAAGCCG GTTGCCATCCGCGCGGAAAAGGAC GGTATCACCGTCGAAGTTCGCGATG TGGTGAACGACAGCTATCATGAG AACGTGCTGTGCTTACCAACAAC ATCCCTCAGCGCGACGGCGGACG CATATGGCCGGTTTCCGCGCAGCG CTGACGCGCAGGTCGTTTCTTAT GCCGACAGTTCGGCATCACCAAAA AAGGAAAAGTACCGCTGACAGGCG GAAGACTGCCGGAAGGCTTGA</p>

<i>R. tropici</i>	<p>GATTTCGTTTCAGCGACTTGCAGC AAAATTCGGCGTGCAGGACATTTTC TGCTTCGGTCTTCTGAATGAGGTC GCCGATGTAGACGATGTTGTCGTT CTTCAGGCAATTTGCCGAGCGAAC GGACAGTTCAGCTCGTCCACCTT CTTGAGGAGCGCCGGTTGAAAGC GAGTTCGGTGACTGCTTCTTCTTC GGCTTCCTTCTGCGCTCGTCGAA GTTGACGAACACGCCAAGCTGATC CTGGAGGATGCGAGCGCGAAAGC GACGGCATCTTACCCTGATCGA ACCGTCGGTTTCGATCGACATGCT CAGCTGTGCTAGTCGAGAACCCTG TCCTTCGGGATFATTTTCAACCTT GTAGGACACCTTCTTGACCGCGA ATAGAGGCTATCGACCGGATGAG GCCG</p>	<p>GTTAAACGCAAGATCCAGCCGGG GACAAGATGGCAGGTCGTACGGG AACAAGGGCGTGGTCTCGCGTTG TTCCGGTCGAAGACATGCCGTTCC TTGAAGACGGTACGCATGTCGATA TCGTGCTGAACCCGCTGGCGTTCC CTCGCGTATGAACGTCGGTCAGAT TCTCGAAACGCACCTTGGCTGGGC TTGCGCCGGCATGGGCGTCAGATC GGCGAACTGATCGAAGCCTACAAG GCGAGCGCAACATCGAACCTCTG CGAAGACGATCATCTGGTCTGCG GCGACGGTCCGAAGAGCGACGACG TTCACGAGTATGACGATCGCTCGG TGCTGCGTGGCCGATCAATGGAA GCGTGGCGTGTCCATCGCGACCCC GGTCTTCGACGGTGTGGCGAAGC TGACGTAACCACATGCTGGCAATG GCAGGTCTAAGCAGACCGGTGAG TCGACGCTGTACGATGGTCTGACC GGCAGCATTCGACCCGAGGTCA CGGTTGGCTACATCTACATGCTGA AGCTGAACCACCTTGTGACGACA GATCCACGCCGCTCGATCGGTCC GTACTCGCTCGTTACCCAGCAGCC GCTTGGTGGTAAGGCGCAGTTGGC GGCCAGCCTTCGGCGAAATGGAA GTCTGGGCGCTCGAA</p>	<p>TCAAACGAGCCTGAAGGCCGGCAGG CTGTCGCCCATTTTCGCTTCGATTT GGCGCGGGCGTACGCGGGCAACGGA GTCGACGACGAGAACGTCGATGGCGC CGGAGCGCACAGCGTATCGGTGAT TCAAGCGCTGTTCGCGGTGTCCGGC TGCGAAATCAGAAGGTTCTGCAGGTC GACACCGAGCTTGGCGCATAGACCG ATCGAGCGCGTGTTCGGCGTCGACGA AGGCGCAGATGCCGCCCTTCTTCTGG GCTTCTGCAATGGTCTCAGCGCCAGC GTCGTCTTACCCGAGCTTTCGGGTCC GTAGATCTCTATGATACGCCCTTGTG GCAAGCCCCGATGCCGAGCGGATAT GAGGCTCAGAGAGCGGTGGAAACT GTTTCGATTTTCAGCACATTTCTGTC GAGCCGAGCTTATGATCGAGCCCTT GCCGAACGACCGCTCAATTTGAGAGA GTGCCGCTTCAAGTCTTGCTTT</p>	<p>TCCGGTGGCTGCACGGCGTCCGGC GTCTCGGTGCTCAATGCGCTTTCC GTCTGGCTGAAGTGAAGATCCGC CGCCACGGCAAGATCCATGAAATG AGCTTACCACATGGCGTCGCCGAT GCGCCGCTGAGGAAACGGCGGAT GCCGCAATCTCACCGCACCGAA GTTAGCTTACATGCCGAGTGTGAA ACCTTACCATGACGGAATTCGAT TACGGAACGCTGGAGCATCGCCTG CGCGAACTGCTTTCCTGAATTC GGTGTCCGATTTCTTTGACCGAT AAGCGCATTCGACGTCAGGACG GAAGAAATGCTCTATGACGGCGC GTGGAACTGCTTCTCTTATCTC GATCGCGCTAAGAAGCGCTGGTT CAGAAGCCGGTGTGATCCGACG GAGAAGGACGGTATCACGGTCGAA GTGGCATGTGGTGGAAACGACG TACCAGAAAACGTGCTCTGCTTC ACCAACAACATCCCGACGGCGAC GGCGTACCATATGGCCGGCTTC CGCGGGCTGACGCGCCAGATC ACCTCTATGCGGATTCATCGGGC ATACCAAGAAGGAGAAGGTGACG CTGACGGCGACGACTGCCCGGAA GGCTGA</p>
<i>R. giardinii</i>	<p>GGCCCATGGAAGCGAGAATTCCT TGATTCGTTTCAGGACTTGCAGC CCAAGTTCGGAGTGCGAAGCATTT CGGCCCTCGGTCTCTGAATGAGAT CGCCGATATAAACGATGTTGTCTG TCTTCAGGCAAGTGGCCGAACGGA CGGAAAGTTCAGTTCGTGACCT TCTTGAGGAGCGCCGGTTGAACG CCAGTTCGGTACGCTTCTTCTT CGGCTTCTTCTGCGGTTCTGTCGA AGTTGACGAAGACGCCGAGCTGGT CCTGAAGGATGCGGGCCGGAAG CGACTGCATCTTACCCTGACCG AACCATCGGTTTCGATCTGCATGG ACAGCTTGTCTAGTCGAGAACCT GACCTTCGCGGTTGTTTCCACCT TGTAGGACACTTCTTGACCCGGC AATACAGGCTGTCCAGCCGGATGA GGCCGATCGCGCATTTCCGAAC GGTTGCGATCGCCGGCACGTAAC CCTTGCCTTATTTGACGGTGAAC CGATCGGATCTCGCGCCCTCGT CGAGGGTGCAGATCACATGTTTCG GGTTGAGGATTTTCGATATCGCCAA CCGCTGATGTCACCGCGGTTA CAACGCCGACCTGCTTGCACA CGACCATGCGCTT</p>	<p>GCGCGCTCCGACGAGGACTTC AATCCGTTTTCAAGTCCGTTTTCC CGATCACGATTTCTCCGGCTTC GATGCTCGAATTCGTTCTCTACGA GTTGCAACAGCCAAAGTTCGACGT GGAGAAATGCCGTACGGGATCTG ACCTACGCTGCACCCTCAAGGTG ACGCTGCGCCTCATCGTGTTCGAT ATCGATGAGGATACGGCCCAAGT CCATCAAGGACATCAAGGAACAGA ACGCTACATGGGCGACATGCCG TCATGACCAACACGGTACGTTTAT CGTCAACGGCACCGGCGGTAT CGTCTCCAGATGCACCCTTCGCG GGGCGTGTTCGATACGACAAG GGCAAGAGCCACTCTCCGGCAAG CTGCTGTTGCTGCCCGGCTCATC CCGATGCGGGCTCCTGGCTCGACA TCGAGTTCGATGCCAAGGACGTTG TCCATGCGGATATGACCCGCC GCAGATCCCTGTACGTCGCTGT GATGGCGCTCGGCATGAGCGGCGA GGAAATCCTCGAGACGTTCTACAC AAGTCTTACACGCGCGACGGC AAGGGCTGGCGGATTCGTTCCAG CCGGACACGCTGAAGGGCCAGAG CTCTGCCGACTGATCGACGCCG ATACTGGCGAAGTCTGTTGCAAT CCGGCAAGAACTGACCGCGCT GTTGAAGACGCTGAAGGAAAAGGG CCTGAAGCCATCAAGCGGTCTGA CGACGATCTTACGGAATCTCTG GCCGAAGACCTCGTCAACATGACG ACGGGCGAAATCTTCTCGAAGCC GGCGACGAAATCACGAGAAGACG TCGGCGTATCTCGGTGCTGGCT TCGACGAAAT</p>	<p>CGAGATCGAACCGTTTCGACGGGCTC TTTGAAGTCTCGACATCGGACTCGGGA TCGGCGGGCTGCCAAGGGGCGCATC ATCGAAATCAGCGACCGGAAAGCTC GGGCAAGACGACGCTGGCGCTCGACA CCATTGCCGAAGCGCAGAAAAAGGGC GGTATCTGCGCTCTCGTGCACGCGGA ACACGCGCTCGATCCGGTCTATGCC GCAAGCTCGGCGTTCGATCGGAAAT CTTCTGATCTCGACCGGATACCGG CGAACAGGCGCTGGAATCACCAGTA CGCTGGTGCCTCCGGCGCCATCGAC GTTCTCGTTCGATTCGGTTCGCGGC GCTGGTGCCTCGGCTGAAATCGAAG GTGAAATGGGCGACGCTGCCCGC CTGACGGCCCGCTGATGAGCCAGG GCTGCGCAAGCTGACCGCTCGATCT CCAAGTCAAGTGCATGGTCACTTC ATCAACCAGATCCGATGAAGATCGG CGTCAATG</p>	<p>TCCGGTGGCTGCACGGCGTCCGGC GTCTCGGTGCTCAATGCGCTTCTCG GTCAAGCTCAAGCTCAAGATCAAG CGCCAGGGCAAGTCCACGAAATC AACTTACGATGGTGTGCGGAC GCGCCGCTGGCGTTCATCGCGCAC AGCAACGGGAGACCGGCACGGAA GCTACCTTCTGCGGACACCGAA ACCTTACCATGATCGAATTCGAC TACGGCACGCTGGAAACCGGTTG CGCGAACTCGCTTCTGAAATTC GGCTTTCGATCCTGCTTACCGAC AAGCGCATTCGATATCCGCCAG GACGAGATGATCTATGACGGCGC CTTGAAGCATTCGTGCGCTATCTC GACCGGGCAAGAAAGCCGCTGGTC TCCAAGCTGTCGCGATCCCGGGC GAAAAGGACGGCATACCGTCGAA GTCGCGATGTGGTGGAAACGACAGC TACCAGGAAAGTGTCTGCTTTC ACCAACAACATTCGCGACGGCGAC GGCGGACCCATATGGCCGGCTTC CGTGGCGACTGACTCGCCAGATC GTCTCTACGCGACAGTTCGGC ATCATGAAGAAGGAAAAGGTAACG CTGACCGGTGACGACTGCCGTGAA GGGTTGACGGCGGTTCTCTCGGTC AAGGTGCCGATCCGAAATCTCC</p>

Suite Annexe 2

Les séquences nucléotidiques (GenBank) des différents gènes du *Rhizobium spp*

Gènes Espèces	<i>Nif H</i>	<i>Nod C</i>	<i>ARNr 16S</i>
<i>R. pisi</i>	<p>GACGTGCTCGGTGACGTTGTGTGCGG CGGCTTCGCGATGCCGATCCGTGAAA ACAAAGCTCAGGAAATCTACATCGTC ATGTCCGGCGAAATGATGGCACCTTA TGCCGCCAACAAATATCGCCAGGGGCA TCCTGAAATATGCTGCAGGCGGGCAGC GTACGCTGGGTGGCCTCATTGTGTA CGAGCGTCAGACCGACCTGAACTCG ACCTCGCCGAAGCGCTGGCTGCTAAA CTCAATCCAAGCTCATCCAATTCGT GCCACGCGACAACATTGTCCAACATG CCGAGCTTAGAAAGATGACGGTAATC CAATATGCGCCGGACTCTCAGCAAGC GGCCGAGTACCGGACACTTGCCCAAA GAATACATGACAATTCGGCAAAAGGC ACCGTCCCACCCCGATCACCA</p>	<p>TGGCTTGCCGTGAACGAGGAGCGCGGCACA AGCTCGCTTCGGTGCAGTTATGTGTTGCTGCG GCCATGTGCGATGTACCCTCGGCTGCTATG CTTTCGCTGCTCGATCAGTACGAGACGACGCT TTATCGCGGCAAGCCGAGTACTTCGGTGAAG ATCGCCATTTGACGATTTCTCATGTCAGCGCA GGCTTTCGAACTGAGTATGTCCGAGCGCCAT CGCGGGACAGTCTTCTGACACAATGGGTG TTTATTTGCGTCAACAACCTACGGTGGGCACGC AGCACCTTTCGGGATACTTTGCTTGGCTCCC CATACTGCTGGCTCGATCGGTATCTCACTC TGGACGTAATCGGGCAAAATGGCGGCTTCTG CTTCTTGGCGTGTGCGTATTGACGGGTATTGG CCAGTTTGGCTGACCGCCACAGTACCATGGT GGACGATCCTGGTGTGATCGGATCCATGACTCTT GTACGATGCAAGCTGGCTGCTATCGCGCCCG CGAATTTAGGTTTCTGGGTTTGTCTTCCACA CGCTCGTGAACATCTTCTCTTAATCCCTTG AAGCCCTATGCCCTTTGTACTTTGTCCTAATAG CGA</p>	<p>TAGAGTTTGATCTGGCTCAGAACGAACTGGCGGAGGCTTAAAC ACATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGACAGCGGTGAGT AACCGGTGGGAACGTACCCTTTACTACGGAATAACGACAGGAAACT TGTGCTAATACCCTATGTGCCCTTTGGGGGAAAGATTATTCGGTAA AGGATCGGCCCCGTTGGATTAGCTAGTGTGGTGGGTAAAGGCCA CCAAGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACAT TGGGACTGAGACACGGCCCCAACTCCTACGGGAGGACAGCAGTGGG AATATTTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCTGAG TGATGAAGGCCCTAGGGTTGTAAGCTCTTTCACCGGAGAAGATAA TGACGGTATCCGGAGAAGAAGCCCGGCTAACTTCGTGCCAGCAGC CGCGTAAATACGAAGGGGCTAGCGTGTGTCGGAAATTAAGTGGCGT AAAGCGCACGTAGGCGGATCGATCAGTACAGGGTGAATCCCAGGG CTCAACCTTGAACCTGCCTTTGATAGTCTGATCTGGAGTATGGAA GAGGTGAGTGGAAATCCGAGTGTAGAGGTGAATTCGTAGATATTC GGAGAACACAGTGGCGAAGGCGGCTCACTGGTCCATTAAGTACG CTGAGGTGCGAAGCGTGGGGAGCAACAGGATAGATACCTGGT AGTCCACGCGCTAAACGATGAATGTTAGCCGTGGGCGATATACTG TTCGGTGGCGCAGTAAACGATTAACATTCGCCCTGGGGAGTACG GTCCGAAGATTAACACTCAAAGGAATTGACGGGGGCCGACCAAGC GGTGGAGCATGTGGTTAATTCGAAGCAACGCGCAGAACCTTACCA GCCCTTGACATGCCCGGCTACTTGCAGAGATGCAAGGTTCCCTTCG GGGACCGGACACAGGTGCTGATGGCTGTGTCAGCTCGTGTGCTG GAGATGTTGGTTAAGTCCCGCAACGAGCGCAACCTCGCCCTTAG TTGCCAGCATTCAGTTGGGCACTTAAGGGGACTGCCGGTGATAAG CCGAGAGGAAGGTGGGATGACGTCAAGTCCCTATGGCCCTTACGG GCTGGGCTACACACGTGCTACAATGGTGGTACAGTGGGCGAGGAG CACCGGAGTGTGAGCTAATCTCCAAGGCCATCTCAGTTCGGATTG CACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCG GATCAGCATGCCCGGCTGAATACGTTCCCGGGCTTGTACACACCG CCGCTCACACCATGGGAGTTGGTTTACCCGAAGGTAGTGGCGTAA CCGCAAGGAGGACGTAACACCGGTAGGGTACGAGCTGGGGTGA GTCGTAACAAGGTAGCCGTAGGGGAACCTGCGGCTGGATCACCTC TTA</p>
<i>R. leucaenae</i>	<p>GACGTGCTCGGCGACGTCGTGTGCGG CGGCTTCGCTATGCCGATCCGCGAGA ACAAGGCCAGGAAATCTACATCGTC ATGTCCGGCGAGATGATGGCGCTTTA CGCGGCAAAACAACATCGCCAAGGGCA TCCTGAAATATGCCATGCAGGCGGC GTGAGGCTCGGGGCTGTGATCTGCAA CGAGCGCCAGACCGGACCGGAGATCG ACCTCGCCGAGGCCCTGGCTCCAGG CTCAATCCAAGCTCATCCATTTTCGT GCCGCGGACAACATCGTCCAGCACG CCGAGCTCAGGAAGATGACGGTCAATC CAGTATGCGCCGATTCGCAACAGGC CGGAGAATATCGACCCTGGCTGAGA AAATTCATGCCAATTCGGGTCAAGGG ACCGTCCCACCCCGATCACAA</p>	<p>TGGCTAGCCTGTAATGAGGAGCGCGGCACA GGCTCGTTTCGGTGTGCTATGTGCTGCTGTG GCCATGCGCCATGTACCCTCGGCTGCTGCTG ACCATGCTACTCGACAGTACGAAACGCAAA GTTCCGGGAAAGCGAAGCGATTTCCGGCAAG ATCGTCACTTACGATTTCTCATGTTGAAGCC GGATTTGCAACCGAATATGTTCCAACCGCCAT CGCGGCAACTGTTGTTCCAACAAGCTGCGCC CTTATCTGCGTCAACAACCTCGCTGGGACGC AGCACTTCCGTCAGCAGTTGCTTGAATGAA CCTTCTGCCGTGCTTACCGCTTTTCTCACGT TGGACGTCATTTGGTCAAGAACCTCGGACCGCTC TTGCTCGCCCTATCGGTGCTAACCGGACTAGC GCAGTTCGCGCTACCGGGACCGTGCCTGGT GGACATGCCGTGATGATCGCATCGATGACCATG ATCCGCTGCAGCGTCCGAGCGGTTCTGCTCG CCAATTTGATTTATCGGCTTCTCCCTGCATA CCTTCATCAACATCTTTTCTGCTGCCCTTG AAGGC</p>	<p>TCGAGCGCCCTCCTTCAAGCAAGCTTGAAGGATTTACTCCTTGG AAAGAAGATCAAGAGCGCTTCTAAGAGGCTTTCTGTAGGAGGGGA CGGCGACAGCGGTGAGTAAACCGTGGGAATCTACCCTTTTCTACGG AATAACCGAGGAAACTGTGCTAATACCGTATGTGCTCTCCGGGA GAAAGATTTATCGGAAAGGATGAGCCCGCTTGGATTAGCTAGTT GGTGGGGTAAAGGCCCTACCAAGCGCAGATCCATAGCTGGTCTGAG AGGATGATCAGCCACTTGGGACTGAGACAGGCCCCAACTCCTAC GGGAGGACGAGTGGGAAATTTGGACAATGGGCGCAAGCCTGATC CAGCCATGCCCGGTGAGTGTGAAAGGCCCTAGGGTTGTAAGCTCT TTCACCGAGAAGATAATGACGGTATCCGGAGAAGAAGCCCGGCT AACTTCGTGCCAGCAGCCCGGTAATACGAAGGGGCTAGCGTTGT TCGGAATTAAGTGGGCTAAAGCGCAGTACGGGATCGATCAGTCA GGGGTGAATCCAGGGCTCAACCTTGAACCTGCTTTGATAGTGT CGATCTGGAGTATGGAAGAGGTGAGTGAATTCGAGTGTAGAGGT GAAATTCGTAGATATTCGGAGGAACACAGTGGCGAAGGCGGCTCA CTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGAGCAAA GGATTAGATACCTGGTGTCCACGCGCTAAACGATGAATGTTAGC CGTCCGGCAGTATGTTGCGGTGGCGCAGTAAACGATTAACAT TCCGCTGGGAGTACGGTCCGAAGATTAACACTCAAAGGAATTGA CGGGGCCCCGACAAGCGGTGGAGCATGTGGTTAATTCGAAGCAA CGCGCAGAACCTTACAGCCCTTGCATCCTGTGTACCCTAGAG ATAGTGGGTCCACTTCGGTGGCGCAGAGACAGGTGCTGATGGCTG TCGTACGTCGCTGCTGAGATGTTGGGTTAAGTCCCGCAACGAGC GCAACCTCGCCCTTAGTTGCCAGCATTTAGTTGGGCACTTAAGG GGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTCAAGT CCTCATGGCCCTTACGGGCTGGGCTACACACGCTGCTACAATGGTGG TGACAGTGGGACGAGCAGCGGAGTGTGAGCTAATCTCCAAAAGC CATCTCAGTTTCGATTGCACTCTGCAACTCGAGTGCATGAAGTTGG AATCGCTAGTAATCGCGGATCAGCATGCCCGGCTGAATACGTTCCC GGGCTTGTACACACCGCCCGTACACCATGGGAGTTGGTTTACC CGAAGGTAGTGCCTAACCGCAAGGAGGACGCTAAACACCGGTAGGG TCAGC</p>

<p style="text-align: center;"><i>R. multihospiitium</i></p>	<p>GACGTGCTCGGCGACGTCGTGTGCGG CGGCTTCGCATGCCGATCCGCGAGAA CAAGGCCAGGAAATCTACATCGTCA TGTCCGGCAGATGATGGCGCTTTAC GCGGCAAAACAACATCGCCAAGGGCAT CCTGAAATATGCCATGCAGGGCGGG TGAGGCTCGGGGTCTGATCTGCAAC GAGCGCCAGACGGACCGGAGATCGA CCTCGCCGAGGCGCTGGCCTCCAGGC TCAATTCGAAGCTCATCCATTTCTGT CCGCGGACAACATCGTCCAGCACGC CGAGCTCAGGAAGATGACGGTTCATCC AGTATGCGCCGATTCCAACAGGCC GGAGAATATCGCACCTGGCTGAGAA AATTCATGCCAATTCGGGTCAAGGGA CGGTCCCGACCCGATAAACCATGGAA GCA</p>	<p>ATACAAAGGAACGCGCGGCGCAGCTCGCTTCG GTGCAGTTATGTGCTGCTGCGGCCATGCGCG ATGTACCGGGCTTCTGCTATGCTCTCGCTGCT GGATCAGTACGAGACGCAACTTTATCGCGCA AGCCGAGTACTTCGGTGAAGATCGCCATTTA ACGATTCTCATGCTGAGCGCAGGTTTTCAAC TGAGTATGTTCCGAGGCCATCGCGCGCAG TTGTTCTGACACAATAGGTGTTTATCTGCGT CAACAACACTACGGTGGGCACGACACTTTTTCG GGATACTTTGCTTGCCTTCCCGTACTGCGCTG GCCTTGATCGATATCTCACGCTGGACGTAATC GGGCAAAATGGCGGCCCTCTACTTCTTTCGCT GTCGGTGTGACGGGTATTGGCCAGTTTGCCT TGACCGCCACAGTACCATGGTGGACGATCATG GTCATCGTCTCCATGACTCTTGTGCGATGCAG TGTGGCTGCCATTCGCGCCCGCGAGCTTAGGT TTCTGGGTTTTGCTCTCCACAGCTCGTGAAC ATCTTCTCTTAATTCCTTGAAGCCATATGC CCTTTGTACCCTATCCAACAGCGACTGGCTGT CACGA</p>	<p>TTACACATCGAGTCGAGCGCCCGCAAGGGGAGCGGACGCGGTG AGTAACCGGTGGGAATCTACCTTTTGTCTACGGAATAACCGAGGGAA ACTTGTGCTAAATACCGTATGTCTCTTCGGGAGAAAGATTTATCGG CAAGAGATGAGCCCGTGTGGATTAGCTAGTTGGTGGGGTAAAGGC CTACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCA CATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGACGACGTG GGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCGCGGT GAGTGTGAAGGCCCTAGGGTTGTAAGCTCTTTACCCGGAGAAGA TAATGACGGTATCCGGAGAAGAGCCCGGGTAACTTCGTGCCAGC AGCCCGGTAATACGAAGGGGGCTAGCGTGTTCGGAATTACTGGG CGTAAAGCGCACGTAGGCGGATCGATCAGTCAGGGGTGAAATCCCA GGGCTCAACCTTGAAGTGCCTTTGATACTGTGATCTGGAGTATG GAAGAGGTGAGTGGAAATCCGAGTGTAGAGGTGAAATTCGTAGATA TTCCGAGGAACACCAGTGGCGAAGGGCGCTACTGGTCCATTACTG ACGCTGAGGTGCGAAGCGTGGGAGCAAAACAGGATTAGATACCTT GGTAGTCCACGCGTAAACGATGAATGTTAGCCGTGGGCGAGTATA CTGTTCCGTGGCGCAGCTAACGCATTAACATTCGCGCTGGGAGT ACGGTCCGCAAGATTAACACTCAAAGGAAATGACGGGGGGCCGCA AGCGGTGGAGCATGTGGTTAATTCGAAGCAACGCGCAGAACCCTTA CCAGCCCTTGACATCTGTGTTACCTCTAGAGATAGGGGTCCACT TCGGTGGCGCAGAGACAGGTGCTGCATGGCTGTCTGCTCAGCTCGTGT CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCTTCGCCCT TAGTTGCCAGCATTAGTTGGGCACCTAAGGGGACTGCCGTTGAT AAGCCGAGAGGAAGTGGGGATGACGTCGAAGTCTCATGGCCCTTA CGGGCTGGGCTACACACGTGCTACAATGGTGGTGAACGAGTGGGCG GAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCCGA TTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATC CGGATCAGCATGCGCGGTGAATACGTTCCCGGGCTTGTACACA CCGCCCGTACACCATGGGAGTGTGGTTTTACCCGAAGGTAGTCCGC TAACCGCAAGGAGCGACTAACCCAGGTA</p>
<p style="text-align: center;"><i>R. gallicum</i></p>	<p>AGTGGCGTGGAGTCCGGTGGCCGAGA GCCGGCGTGGCTGCGCCGGACGCG GCGTTATCACCTCGATCAACTTCCTG GAAGAGAAGCGGCTTACACGATGT CGATTACGTCTCATACGACGTGCTAG GGGACGTAGTATGCGCGGCTTTGCG ATGCCATTCGCGAAAACAAGGCTCA GGAAATCTACATCGTATGTCGGTG AGATGATGGCGCTCTATGCCGCCAAC AACATCGGAAGGATTCCTGAAGTA CGCCATGCGCGCGGCTGCGCCTGG GGGGTTGATTTGCAACGAGCGCCAG ACCGATCGGGAGCTCGACCTCGCCGA GGCACTTGCGCCCGCTCAATTTCA AGCTCATCCACTTCGTGCCGCGCGAC AATATCGTTGACGACGAGAGCTCAG AAAGATGACAGTATCCAATATGCGC CGAACTCTAAGCAAGCCGGGAATAT CGCGCCCTGGCTGAAAAGATCCATGC AAATTCGGCCGAGGACCGTCCCTA CACCGATCACTATGGAGGAAGTGGAG GACACGATGC</p>	<p>ATACAAAGGAACGCGCGGCGCAGCTCGCTTCG GTGCAGTTATGTGCTGCTGCGGCCATGCGCG ATGTACCGGGCTTCTGCTATGCTCTCGCTGCT GGATCAGTACGAGACGCAACTTTATCGCGCA AGCCGAGTACTTCGGTGAAGATCGCCATTTA ACGATTCTCATGCTGAGCGCAGGTTTTCAAC TGAGTATGTTCCGAGGCCATCGCGCGCAG TTGTTCTGACACAATAGGTGTTTATCTGCGT CAACAACACTACGGTGGGCACGACACTTTTTCG GGATACTTTGCTTGCCTTCCCGTACTGCGCTG GCCTTGATCGATATCTCACGCTGGACGTAATC GGGCAAAATGGCGGCCCTCTACTTCTTTCGCT GTCGGTGTGACGGGTATTGGCCAGTTTGCCT TGACCGCCACAGTACCATGGTGGACGATCATG GTCATCGTCTCCATGACTCTTGTGCGATGCAG TGTGGCTGCCATTCGCGCCCGCGAGCTTAGGT TTCTGGGTTTTGCTCTCCACAGCTCGTGAAC ATCTTCTCTTAATTCCTTGAAGCCATATGC CCTTTGTACCCTATCCAACAGCGACTGGCTGT CACGA</p>	<p>TCCTTGCGGGTAGCGCAGCGCTTCGGGTAACCAACTCCCAT GGTGTGACGGCGGTGTGTAACAGGCCCGGGAACGTAATTCACCGCG CGGTGCTGATCCGCGATTAAGCAGTTCAACTTCATGCACTCGA GTTGACAGTGAATCCGAACTGAGATGGCTTTTGGAGATTAGCTC ACACTCGCTGCTTGTCTGCCACTGTCCACCATTGTAGCACGTG TGTAGCCAGCCGTAAGGGCATGAGGACTTGACGTTCATCCCCAC CTTCTCTCGGCTTATCACCAGCAGTCCCTTAGAGTGCCTCACTG AATGATGGCAACTAAGGGCGAGGGTTCGCGCTCGTTGCGGGACTTAA CCCAACATCTCACGACAGAGCTGACGACAGCCATGCAGCACCTGT GTCCCGGTCGCCGAAGGAAACCATCCATCTCTGGATGATAGCGGGC ATGTCGAAGGGCTGGTAAGGTTCTGCGCGTTCGTAATTAACCA CATGCTCCACCCTTGTGCGGGCCCGCTCAATTCCTTGGAGTTTT AATCTTGCACCGTACTCCCGAGGCGGAATGTTAATGCGTTAGCT CGCTCACCAGCAGTAACTGCCCGAGCTAACATTCATCGTTTA CGCGGTGGACTACAGGGTATTAATCCTGTTTGTCCCCACGCTT TCGCACCTCAGCGTCAAGTTGTGGACAGTCAAGCCGCTTCGCGACT GGTGTCTTCGCAATATCTACGAATTTACCTCTACACTCGCAATT CCACTGACCTTCCACACTCAAGACACCCAGTATCAAAGGCGAGTT CCAGAGTTGAGCTCTGGGATTTACCCCTGACTTAAATGTCCGCTC ACGTGCGCTTACGCGCAAGTATTCGGAACAGCTAGCCCTTCTC GTATTACCGCGCTGCTGGCACGAAGTTAGCCGGGCTTCTTCTCC GGATACCGTCATATCTTCCGGTGAAGAACTTTACAACCCCTAA GGCCTTCATCATTACGCGGCATGGCTGGATCAGGCTTGCGCCAT TGTCCAATATTCCTCACTGCTGCTCCCGTAGGAGTTTGGGCCGTG TCTCAGTCCCAATGTGGCTGATCATCTCTCAGACAGCTATGGAT CGTCCGCTTGGTAGGCTTTACCCCACTAGCTAATCCAACGC GGGCCGATCCTTACCAGATAAATCTTCCCCAAAGGGCACATACG GTATTAGTCCAGTTTCCCGGAGTTGTTCCGTAGTAAAGGGTACGT TCCACGCGTTACTCACCGCTGCGCGTCCCTTGCGGGGCGCTC GACTTGCATGTGTTAAGCTGCGCCAGCGTTCGTTCTGAGCCA</p>

<p style="text-align: center;"><i>R. etli</i></p>	<p>TCGATCTAGCCGCAACGAAAGGTTCT GTCGAAGATCTGGAATCGGCATGT GCTCAAAACTGGCTACGGCGCATCA AATGTGTGGAGTCGGCGGCCCTGAA CCCGGCTCGGCTGCGCGGGCGAGG CGTCATCACATCGATAAAATTCCTGG AAGAAACGGCGCCTATGACGATGTC GACTAGCTTCCATGACGTGCTCGG AGACGTGGTGTGCGGAGGCTTTGCTA TGCCCATCCGCGAGAACAAGGCTCAG GAAATTTACATCGTCATGTCCGGTGA GATGATGGCCCTGTATGCTGCAACA ACATCGCGAGAGGCATCCTCAAAAT GCCAGCGCCGGAAGCGTGGCGCTGGG CGGCTGATCTGTAATGAGCGGCAGA CCGATCGAGAATTAGACCTTGGCTGAA GGCTGGCCGCCAAACTCAATTGCAA GCTTATTCACTTCGTGCCGCGGACA ACATTGTCCAGCAGCGGAGCTTAGA AAGATGACCGTATCCAATACGGCC ACGCTCCAAAC</p>	<p>CGTTCGGAGCCGTTATGTGCTGCTGCGGTTCC ATGTGCTATCTACCGTCTGCTGCTGCTGCTT CGCTGCTTGACCAGTACGAATCACAGTATTTT CGGGAAAGCCAAGCGATTTCCGGTGAGGATCG GCATCTCACCATTTCTCATGTGAAGGAGCT TTCGAACGGAGTACGTGCCGAGCCATCGCA GGACAGTCTTCCGAACAAGCTAGGACCGTA TCTGCGCAACAACCTACGCTGGGCGGGAGCA CGTTCGGGACACGTTGCTTGGGCTGCGCCTG CTGCCAACCTCAATCGCTTCCCTACGCTCGA CGTTGTCGGACAGAACCTCGGACCGCTGCTTC TGGCACTATCAGTGTGACGGGCTCGCACAG CTTGCATTGACGGGCAACGTGCCTTGGTTGCC AGCCCTGATGATTGTGGCCATGACGATGATCC ACTGCAGCGTTGTGCGCTTCCGGCCGCCAA CTACGGTTCCTCGGGTCTCTCTGACGCAATT TATCAGTATTTTCTGCTACTTCCCTTGAAAG CCTACGCGCTGTGACCGTGTGACCAATAGC</p>	<p>AACGAACGCTGGCGGAGGCTTAACACATGCAAGTCGAGCGCCCG CAAGGGAGCGGCAGACGGGTGAGTAACCGGTGGGAACGTACCCCTT TACTACGGAATAACCGAGGAACTTGTGCTAATACCGTATGTGCC CTTTGGGGAAAGATTTATCGTAAAGGATCGCCCGCGCTTGGATT AGCTAGTTGGTGGGTAAAGCCCTACAAGGCCACCATCCATAGCT GGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAA ACTCCTACGGGAGGAGCAGTGGGGAATATTGGACAATGGGCGCAA GCCTGATCCAGCCTGCGCGTGTGATGAAGGCCCTAGGGTTGT AAAGCTCTTTCACCGGAGAAGATAATGACGGTATCCGGAGAAGAAG CCCCGGCTAACTTCTGCGCAGCGCCGGTAAATACGAAGGGGGCT AGCGTTGTTCCGGAATTAAGTGGCGTAAAGCGCACGTAGCGGATCG ATCAGTCAGGGGTGAAATCCAGGGCTCAACCTGGAACCTGCCTTT GATACTGTGATCTGGAGTATGGAAGAGGTGAGTGAATTCGGAGT GTAGAGGTGAAATTCGTAGATATTCCGGAGAACACCAAGTGGCGAAG GCGGCTCACTGGTCCATTACTGACGCTGAGGTGCGAAAGCGTGGGG AGCAACAGGATTAGATACCTTGGTGTAGTCCACCGCTAAACGATGA ATGTTAGCGCTCGGGCAGTATACTGTTCGGTGGCGAGCTAACGCA TTAAACATTCGCGCTGGGGAGTACGGTTCGCAAGATTAACACTCAA GGAATTTGACGGGGGGCCGACAGCGGTGGAGCATGTGGTTTAAAT CGAAGCAACGCGCAGAACCTTACCAGCCCTGACATGCGCGGAC CTGCAGAGATGCAGGTTCCCTTCGGGGACCGGGACACAGGTGCTG CATGGCTGTGTCAGCTCGTGTGATGTTGGGTTAAGTCCCG CAACGAGCGCAACCTTCGCCCTTAGTTGCCAGCATTTGGTTGGGCA CTTAAGGGGACTGCCGTTGATAAGCCGAGAGGAAAGTGGGGATGA CGTCAAGTCTCATGGCCCTTACGGGCTGGGCTACACACGTGTCTAC AATGGTGGTACAGTGGCGAGCAGCAGCGAGTGTGAGCTAATCT CCAAAAGCCATCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCAT GAAGTTGGAATCGTAGTAATCGCGGATCAGCATGCCCGGGTGAAT ACGTTCCCGGGCTTGTACACACCGCCCGTACACCATGGGAGTTG GTTTTACCAGAGGTAGTGCCTAACCGCAAGGAGCAGCTAACCA CGGTAGGCTCAGCGACTGGGGTGAAG</p>
<p style="text-align: center;"><i>R. leguminosarum</i></p>	<p>TGTTTCGGCAAAGCATCTGCTCGTCCG TCTTATGATCCGAAATCGAGAAGC ATGTCCTCAAGCTCTTCCATGTTGAT GGGGTCGGGATGGTGCCTTTGCCCG AGTTGTCAATGATTTCTTTGGGCAAGT GTCCGGTACTCGCGCCTTGTCTGAGA GTCCGGCGCATATGGATCACCGTCA TCTTTCTAAGCTCGGCGTGTGGACG ATGTTGTGCGGTGGCACGAAATGGAT GAGCTTAGAATTTAGTTTGGCAGCCA CGCTTTCGGCGAGGTCGATTTCCGCG TCGGTCTGACGCTCGTTACAATGAG GCCACCCAGGCGCACGCTCCCGCCTG CAGCATATTTAGAAATGCCCTGGCG ATATTGTTGGCGCATAGAGGGCCAT CATTTCCGCGACATGACGATATAGA TTTCTGAGCTTTGTTTTACGGATC GGCATCGGAAGCCCGCGAGACAAC GTCACCGACACGTCATAAGACACGT AGTCGACATCGTTGTAGGCGCGTTC TCTTCGAGAAAGTTGATCGACGTGAT AACGCCGCT</p>	<p>GTGACGGCCGTTTATTCTCTGCCAGCCGATGT TTCATTGGCGTCCCAGAGCTTGGCCGCTTTG AGGAGCTGCCCAGCGTAGATGTCATCGTGCCA AGCTTCAACGAGGATCCCCGCACGCTTTCCGA GTGCTGGCTTCTATTGCGGGTCAGGAATACG GGGAAGGCTGACAGGTTTACTAGTTGATGAC GGTTCGAAAATCGCAGGCTTTGCGACCTGT GCACGAGGCTTTCGCACGAGACCCAGATTCA ATATTCTCTGCTTCCCAGAAATGTTGGTAAA CGGAAGGCACAGATCGCCGATACCGCGCTC TGCTGGAGATATGGTGTAAACGTCGACTCCG ACACGATCTCGCATCTGACGTCATCAGGAAG CTCGTGCCTAAAATGCAAGATCCGGCTGTCCG CGCGGCCATGGGACAGTTGACGGCCCGCAACC GAAACGATAGTTGGCTGACCCGTTTGTATCGAT ATGGAGTACTGGCTGGCTTGAACGAGGAGCG TGGGCAACAAGCTCGCTTCCGAGCCGTTATGT GCTGCTCGGTCATGTGCTATCTACCGTCCG</p>	<p>TAGAGTTTGTCTGCTCAGAACGCTGGCGGAGGCTTAAC ACATGCAAGTCGAGCGCCCGCAAGGGGAGCGGCAGACGGGTGAGT AACCGTGGGAATCTACCCTTACTACGGAATAACCGAGGAAACT TGTGCTAATACCGTATGTGCTCTTCGGGAGAAGATTTATCGGTCA AGGATGAGCCCGCTTGGATTAGCTAGTTGGTGGGTAAGGGCCTA CCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGATCAGCCACAT TGGGACTGAGACACGGCCAACTCTACGGGAGGCGAGTGGGG AATATTGGACAATGGCGCAAGCCTGATCCAGCCATGCCGCGTGAG TGATGAAGGCCCTAGGGTTGTAAGCTCTTTCACCGGAGAAGATAA TGACGGTATCCGGAGAAGAAGCCCGGCTAACCTTCGTGCCAGCAGC CGCGTAAATACGAAGGGGGCTAGCGTGTTCGGAATTAAGTGGGCT AAAGCGCAGTAGGCGGATCGATCAGTCAGGGGTGAAATCCCAGGG CTCAACCTTGAAGTCCCTTTGATCTGTGATCTGGAGTATGGAA GAGGTGAGTGGAAATCCGAGTGTAGAGGTGAAATTCGTAGATATTC GGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTAAGTACG CTGAGGTGCGAAAGCGTGGGGAGCAACAGGATAGATACCTTGGT AGTCCACGCGTAAACGATGAATGTTANCCGTCGGGCGAGTATACTG TTCGTTGGCGCAGCTAACGCATTAACATTTCCGCTTGGGAGTACG GTCGCAAGATTAACACTAAAGGAATTCAGGGGGCCCGCACAGC GGTGGAGCATGTGGTTTAAATTCGAAGCAACGCGCAGAACCTTACCA GCCCTTGACATCCCGGCTACTTGCAGAGATGCAAGGTTCCCTTCCG GGGACCGGACACAGGTGCTGATGGCTGTGTCAGCTCGTGTGCTCGT GAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCTTCGCCCTTAG TTGCCAGCATTCAGTTGGGCACTTAAGGGGACTGCCGGTATAAG CCGAGAGGAAGGTGGGATGACGTCAAGTCCCTCATGGCCCTTACGG GCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGGCGAGCAG CACGGGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTG CACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCG GATCAGCATGCCCGGTGAATACGTTCCCGGGCTTGTACACACCG CCCGTACACCATGGGAGTGGTTTTACCAGGAGTGTGCGCTAA CCGCAAGGAGGAGCTAACACGCTAGGGTACGCGACTGGGGTGAA GTCGTAACAAGGTAGCCGTAGGGGAACCTGNGGCTGGATCACCTCC TTA</p>

<p style="text-align: center;"><i>R. tropici</i></p>	<p>CCGGAGCCGGCGTAGGTTGCGCCGG TCGCGCGGTATACCTCCATCAATT TCCTCGAGAAAAACGGCGCTACGAC AATGTCGACTATGCTCCTATGACGT GCTCGGCGACGTGCTGTGGCGGCT TCGCCATGCCGATCCGCGAGAATAAG GCCCAGAAAATACATCGTCATGTC TGCGCGGATGATGGCGCTTATGCGG CAAACAACATCGCAAGGGCATCTTG AAATATGCCCATGACGGCGGCTGCG GCTCGGCGGTCTGATCTGCAACGAGC GTCAGACAGACCGGAAATCGATCTC GCCGAGGCGCTGGCTTCCAGGCTCAA TTCCAAGCTCATCCATTTCGTGCCGC GCGACAACATCGTCCAGCACGCCGAG CTCAGGAAAATGACAGTCCAGTA TGCGCGGAGTCCAAACAGGCTGGAG AATATCGCACGCTGGCGGAGAAGATC CATGCCAATTCGGGTCAAGGGACCGT CCGACCCCGATCACCATGGAAGAGC TCGAGGACATGCTGCTCGACTTCGGC ATCATGAAGACC</p>	<p>GCTGGCCTGTAATGAGGAGCGCGCCACAGG CTCGTTTCGGTGTGTATGCTGTGCTGGC CCATGCGCCATGTACCGCCGACTGCGCTTAC CATGCTACTCGACCAGTACGAAACGCAAAATGT TCCGGGAAAGCGAAGCGAATTCGGCGAAAT CGTCATCTTACGATTCATGTTGAAGGCCG ATTTGCAACCGAATATGTTCCAACCGCCATCG CGGCAACTGTTGTTCCAAACAAGCTGCGCCCT TATCTGCGTCAACAACCTCGCTGGGACCGCAG CACTTTCGGTGACACGTTGCTTGAATGAACC TTCTGCTGGTCTTGACCGTTTTCTCACGTTG GACGTCATTGGTCAGAACCCTGGACCGCTCTT GCTCGCCCTATCGGTGCTAACCGGACTAGCGC AGTTCGCGCTCACCGGGACCGTGCCTTGGTGG ACATGCTGATGATCGCATCGATGACCATGAT CCGCTGACGCGTGCAGCGGTTGCTGCTCGCC AATTCGATTTATCGGCTTTCCTCGCATAACC TTTATCAACATCTTTTTCTGCTGCCCTTGAA GGCTACGCGCTCTGTACGTTGAGCAATAGCG</p>	<p>GGGAGCCGAGACGGGTGAGTAACCGGTGGGAATCTACCTTTTGGCT ACGGAAACAGCAGGGAACTTGTGTAATACCGTATGTTGCTCTC GGGAGAAAGATTATTCGGCAAGAGATGAGCCCGGCTTGGATTAGCT AGTTGGTGGGGTAAAGGCCACCAAGCGCAGCATCCATAGCTGGCT TGAGAGGATGATCGACCATTTGGGACTGAGACACCGCCAACTC CTACGGGAGGCAGCAGTGGGGAATATGGACAATGGGCGCAAGCCT GATCCAGCCATGCCCGGTGAGTATGAAGCCCTAGGGTTGTAAGG CTCTTTCACCCGAGAGATAATGACGGTATCCGGAGAAAGGCC GGCTAACTTCTGTCAGCAGCCCGGTAATACGAAGGGGGCTAGCG TTGTTCCGAATTAAGCGGTAAGCGCACGATAGCGGATCGATCA GTCAGGGGTGAAATCCACGGGCTCAACCCCTGGAATGCTTTGATA CTGTCGATCTGGAGTATGGAAGAGGTGAGTGGAAATCCGAGTGTAG AGGTGAAATTCGTAGATATTCGGAGGAAACACAGTGGCGAAGGCGG CTCACTGGTCCATTACTGACGCTGAGGTGCGAAAGCTGGGGAGCA AACAGGATTAGATACCCTGGTAGTCCACGCGCTAAACGATGAATGT TAGCCGTCGGGAGTATGCTGTTGCTGGCGCAGCTAACCGCATTA ACATTCGCGCTGGGAGTACCGTCCGAAAGATTAAAACGCAAGGAA TTGACGGGGCCCGCACAAGCGGTGGAGCATGTGTTTAAATTCGAA GCAACGCGCAGAACCTTACCAGCCCTTGACATCCTGTGTTACCTCA AGAGATAGGGGTCCACTTCGGTGGCGCAGAGACAGGTGCTGATG GCTGTGCTGAGTCTGCTGAGATGTTGGGTTAAGTCCCGCAAC GAGCGCAACCCCTCGCCCTTAGTTGCCAGCATTAGTTGGGCACTCT AAGGGGACTGCGGTTGATAAGCCGAGAGGAAAGTGGGGATGACGTC AAGTCTCATGGCCCTTACGGGCTGGGCTACACACGCTGCTACAAATG GTGGTGACAGTGGGCGAGCAGCAGCGAGTGTGAGCTAATCTCCAA AAGCCATCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAG TTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGT TCCCGGGCTTGTACACACCGCCCTCACACCATGGGAGTGTGGTTT TACCCGAAAGGTAGTGCCTAACCGCAAGGAGGCAAGCTAACCCGCT AGGGTCAGGACTGGGTTAAGTCGTAACAGGTAGCCGTAGGGGA ACCTGC</p>
<p style="text-align: center;"><i>R. giardii</i></p>	<p>GACGTGCTCGCGATGTGGTGTGCGG TGGCTTTCGATGCCGATCCGTGAGA ACAAGGCCAGGAGATCTACATCGTG ATGTCGCGGAGATGATGGCGCTCTA TGCCGCCAACAAATCGCCAAGGGCA TCCTGAAATATGCCATTCCGGCGGC GTGCGGCTCGCGGCCCTGATCTGTAA CGAGCGCCAGACGGACCGGAGCTCG ACCTCTCCGAGGCGCTGGCTGCCAGG CTCAATTCGAAGCTCATCCACTTTGT GCCCGTGACAACATCGTCCAGCACG CCGAGCTCAGGAAGATGACGGTGATC CAGTACGCGCCGACTCCAAGCAGGC CGGGAAATATCGGGCGCTAGCCGAGA AGATCCATGCCAATTCGGGCAAGGG ACCATTCCGACCCCGATCACAA</p>	<p>GTGACGCGCTTTATTCTCTGCCACCGATGT TTCATTGGCGTCCCAGAGCTTGGCCGGCTTTG AGGAGCTGCCAGCGTAGATGTCATCGTGCCA AGCTTCAACGAGGATCCCCGACGCTTTCGGA GTGCTGGCTTCTATTGCGGGTCAGGAATACG GGGAAGGCTGCAGGTTTACCTAGTTGATGAC GGTTCGAAAATCGCAGGCTTTCGACCTGT GCACGAGGCTTCGCACGAGACCCAGATTCA ATATTCTCTGCTTCCCAGAATGTTGGTAAA CGGAAGCACAGATCGCCGATACGCCGCTC TGCTGGAGATATGGTGTAAACGTCGACTCCG ACACGATCCTCGCATCTGACGTATCAGGAAG CTCGTGCCTAAAATGCAAGATCCGGCTGTCGG CGCGCCATGGGACAGTTGACGGCCCGCAACC GAAACGATAGTTGGCTGACCCGTTTATGATGAT ATGGAGTACTGGCTGGCTTGAACGAGGAGCG TGCGGCACAAGCTCGCTTCGGAGCCGTTATGT GCTGCTGCGGTCATGTGCTATCTACCGTCCG</p>	<p>AGAGTTGATCCTGGCTCAGAACGAACGCTGGCGGACGGCTTAACA CATGCAAGTCGAGCCGCCCGCAAGGGGAGCGGACAGCGGTGAGTA ACGCGTGGGAACTCAACCATCTCTACGGAATACTCAGGGAACCT GTGCTAATACCCTATACGCCCTTCGGGGGAAAGATTATCGAGAT GGATGAGCCCGGTTGGATTAGCTAGTTGGTGGGGTAAAGGCTAC CAAGCGCAGATCCATAGCTGCTGAGAGGATGATCAGCCACAT GGGACTGAGACACGGCCAAACTCTACGGGAGGCAGCAGTGGGGA ATATTGGACAATGGGCGCAAGCTGATCCAGCCATGCCGCGTGTG GATGAAGGCTTAGGGTTGTAAGCTCTTTTACCCGAGAGAAGATAAT GACGGTATCCGGAGAAGAAGCCCGGCTAACTTCGTGCCAGCAGCC CGGTAATACGAAGGGGGCTAGCGTGTTCGGAATTAAGTGGGCGTA AAGCGCACGTAGCGGATCGATCAGTCAGGGGTTGAAATCCGAGC TCAACTCTGGAAGTGCCTTTGATACTGTGATCTAGAGTATGGAAG AGGTGAGTGAATTCGAGTGTAGAGGTGAAATTCGTAGATATTCG GAGGAACACCAGTGGCGAGGGCGGCTCACTGGTCCATTACTGACCG TGAGGTGCGAAAGCTGGGGAGCAACAGGATTAGATACCCTGGTA GTCCACCGCTAAACGATGAATGTAGCCGTCGGCAAGTTTACTTG TCGGTGGCGCAGTAACGATTAACATTCGCCCTGGGGAGTACGG TCGCAAGATTAACACTCAAAGGATTTGACGGGGCCCGCACAAGCG GTGGAGCATGTGGTTAATTGGAAGCAACCGGAGAACCTTACCGAG CTCTTGACATCCCGATCGCGACAGTGGAGACATTTGCTTTCAGTT AGGCTGGATCGGTGACAGGTGCTGATGGCTGTGTCAGCTCGTGT CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTCGCCCT TAGTTGCCAGCATTCAGTTGGGCACTTAAGGGGACTGCCGCTGAT AAGCCGAGAGGAAAGTGGGGATGACGTCAAGTCTCATGGCCCTTA CGGGCTGGGCTACACAGTGTACAAATGGTGGTGCAGTGGGCGAGC GAGACCGGAGGTCGAGTAATCTCCAAAAGCCATCTCAGTTCGGA TTGACTCTGCAACTCGAGTGCATGAAGTTGGAATCGTAGTAATC GCAGATCAGCATGCTGCGGTGAATACGTTCCCGGGCTTGTACACA CGCCCGTACACCATGGGAGTTGGTTTACTCGAAGGTAGTGGCG TAACCGCAAGGAGGAGCTAACCCAGGTAGGGTACGCGACTGGGGT GAAGTCGTAACAGGTAGCCGTAGGGGAACCTGCGGCTGGATCACC TCCTT</p>

Annexe 3

Codes d'accèsion (GenBank) des différents gènes de ménage de *Rhizobium spp*

Gènes Espèces	<i>rpoA</i>	<i>rpoB</i>	<i>recA</i>	<i>gyrB</i>	ARNr 16S
<i>R. pisi</i>	KC293535.1	KC293535.1	DQ431676.1	JQ795183.1	AY509899.1
<i>R. leucaenae</i>	JF318203.1	JF318203.1	JF318197.1	KC293524.1	JF318176.1
<i>R. multihospitium</i>	JF318207.1	JF318207.1	EF490039.1	KC293528.1	NR_044053.1
<i>R. gallicum</i>	EU488840.1	EU488840.1	EU518354.1	AM418828.1	FR728933.1
<i>R. etli</i>	JN129370.1	JN129370.1	KF483575.1	JN129340.1	AB680905.1
<i>R. leguminosarum</i>	JN129371.1	JN129371.1	JQ302792.1	KC293526.1	AY509900.1
<i>R. tropici</i>	JN129357.1	JN129357.1	EU488828.1	JN129327.1	JF318177.1
<i>R. giardinii</i>	EU488829.1	EU488829.1	EU532142.1	AM418827.1	U86344.1

Annexe 4

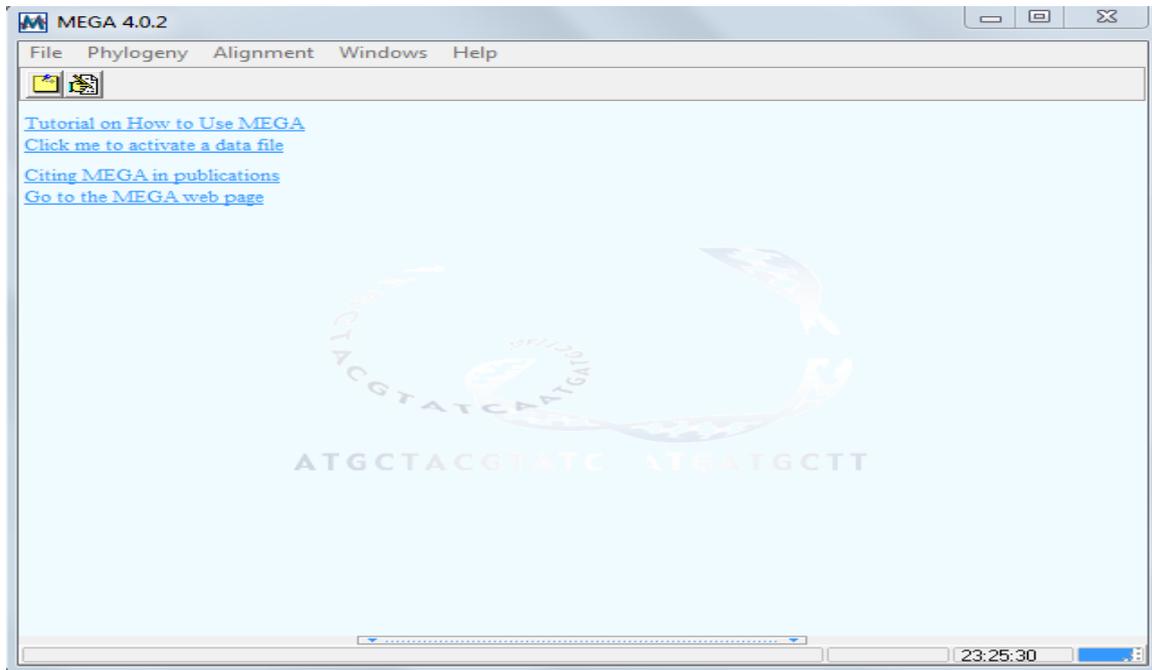
Le code des nucléotides selon UIB¹

Symbole	signifiant	acide nucléique
A	A	Adénine
C	C	Cytosine
G	G	Guanine
T	T	Thymine
U	U	Uracil
M	A ou C	
R	A ou G	
W	A ou T	
S	C ou G	
Y	C ou T	
K	G ou T	
V	A ou C ou G	
H	A ou C ou T	
D	A ou G ou T	
B	C ou G ou T	
X	G ou A ou T ou C	
N	G ou A ou T ou C	

¹ http://www.chick.manchester.ac.uk/SiteSeer/IUPAC_codes.html

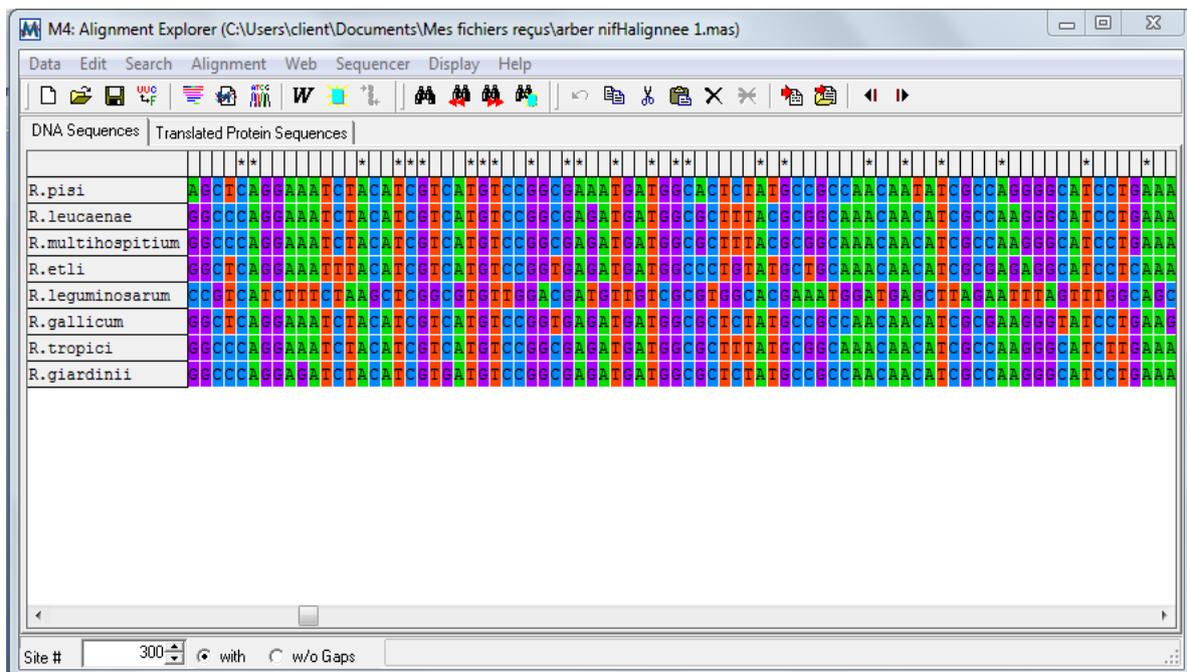
Annexe 5

Intérface principale du programme Mega 4.0



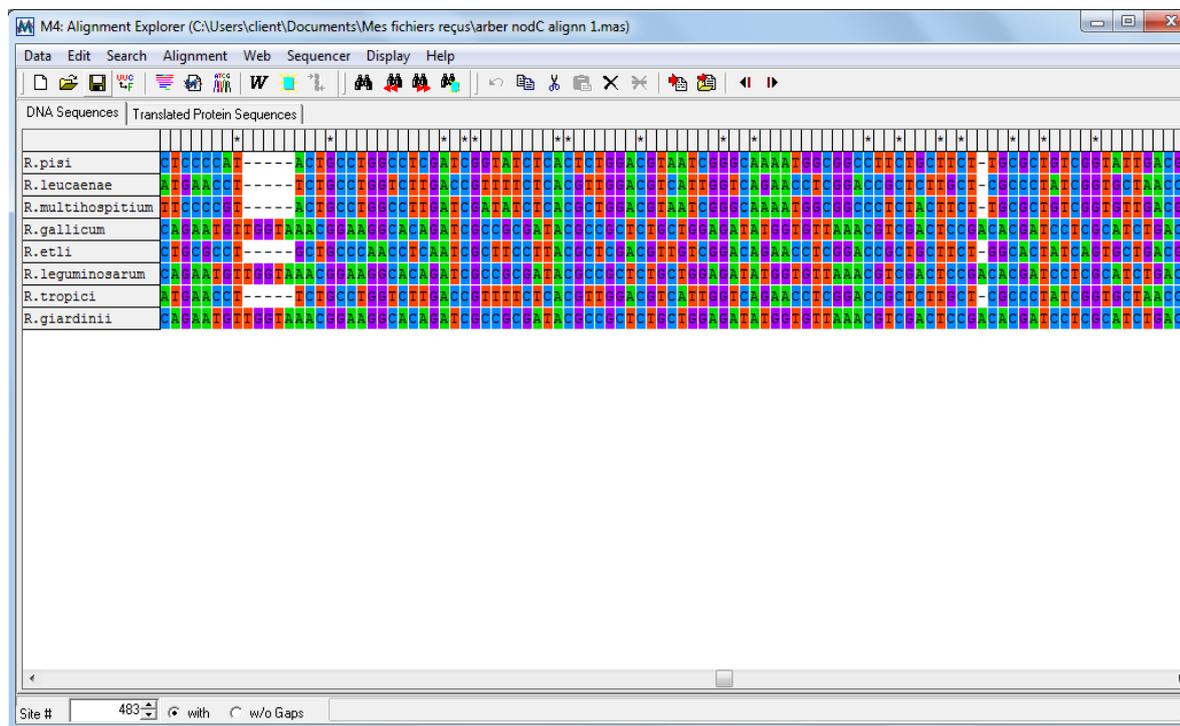
Annexe 6

Insertion des huit séquences sur Mega 4.0



Annexe 7

Résultat de l'alignement multiple des séquences par Mega 4.0



Annexe 8

La matrice distances d'arbre ARN 16S

	<i>R. gallicum</i>	<i>R. leguminosarum</i>	<i>R. pisi</i>	<i>R. multihospitium</i>	<i>R. tropici</i>	<i>R. giardinii</i>	<i>R. leucaenae</i>	<i>R. etli</i>
<i>R. gallicum</i>								
<i>R. leguminosarum</i>	0.90826							
<i>R. pisi</i>	0.90505	0.00547						
<i>R. multihospitium</i>	0.92374	0.01099	0.01651					
<i>R. tropici</i>	0.92374	0.01099	0.01651	0.00000				
<i>R. giardinii</i>	0.91524	0.04059	0.04624	0.03955	0.03955			
<i>R. leucaenae</i>	0.93610	0.01497	0.02050	0.00394	0.00394	0.03528		
<i>R. etli</i>	0.90505	0.00802	0.00251	0.01395	0.01395	0.04364	0.01794	

Année Universitaire : 2013/2014

Présenté par : SAIB Rayane Houda
HABIBATNI Imene

Analyses bioinformatique et phylogénétique des gènes *nod* et *nif* de *Rhizobium* sp.

**Mémoire pour l'obtention du diplôme de Master en microbiologie
Option : Ecologie Microbienne**

Résumé :

Plusieurs approches sont utilisées pour l'identification et la classification des espèces bactérienne. Cependant, la plupart de ces méthodes présentent certaines insuffisances. L'objectif de ce travail est d'établir une analyse phylogénétique de huit espèces de *Rhizobium*, basée sur quatre familles de gènes (ADNr 16S, House Keeping Genes, *nodC* et *nifH*). L'analyse phylogénétique de ces souches a été réalisée par le logiciel MEGA 4 suivant la méthode de Neighbor-joining (NJ) en utilisant les paramètres de correction de Kimura. La robustesse des phylogrammes à été évaluée à l'aide de la méthode de bootstrap. L'analyse des phylogrammes basée sur le gène *nifH* a montré que celui-ci peut être utilisé avec les autres types de gènes afin d'aboutir à des résultats phylogénétiques plus appropriés.

Mots clés : Phylogénie, alignement multiple, MLST, *nifH*, *nodC*

Membre de jury :

Président du jury : BENHIZIA Y.	Professeur	Université Constantine 1
Rapporteur : HAMIDECHI M. A.	Professeur	Université Constantine 1
Examinatrice : BOUZERAIB L.	Maitre assistante	Université Constantine 1