

T H E S E

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline : Biologie des populations et écologie

Formation Doctorale : Ressources Phytogénétiques et Interactions Biologiques

Ecole Doctorale : Biologie des systèmes intégrés : AGRONOMIE.ENVIRONNEMENT

présentée et soutenue publiquement

par

Xavier BAILLY

Le 31/05/2006

Recombinaison, spécialisation et spéciation chez les symbiotes du genre
Sinorhizobium associés aux plantes du genre *Medicago* : des patrons de
diversité aux hypothèses évolutives

JURY

Jean-Loup NOTTEGHEM, Professeur à l'Agro-Montpellier
Claire NEEMA, Professeur à l'Institut National Agronomique Paris-Grignon
Philippe NORMAND, Directeur de Recherche CNRS
Nicolas GALTIER, Chargé de Recherche CNRS
Isabelle OLIVIERI, Professeur à l'Université Montpellier II
Gilles BENA, Chargé de Recherche IRD

Président
Rapporteur
Rapporteur
Examineur
Directrice de Thèse
Directeur de Thèse

A mes parents

A Céline

Remerciements

Mes premiers remerciements s'adressent à Claire Neema et Philippe Normand, qui ont accepté de donner de leur temps pour juger ce travail. Je remercie également Jean-Loup Notteghem et Nicolas Galtier d'avoir gentiment pris part à la composition du jury en tant qu'examineurs.

Voilà bientôt dix ans que je collectionne avec acharnement les cartes d'étudiant de l'Université Montpellier II. De mes premiers cours de génétique mendélienne en DEUG à l'achèvement de cette thèse, en passant par le choix de mes sujets de recherche, je ne pourrais exprimer en quelques lignes tout ce qu'isabelle a pu m'apporter durant ces années. Ce travail n'aurait pas pu aboutir (ou même commencer !) sans ses compétences, son dynamisme, sa disponibilité... Merci !

Etant le premier thésard à plein temps de Gilles « Gilou » Béna, j'ai eu le privilège de l'initier au bonheur de l'encadrement « du jeune chercheur en folie ». Rien ne lui a été épargné : les questions métaphysiques sur la composante interpopulationnelle du déséquilibre de liaison (à huit heures du matin), ma tendance irrépressible à retarder les échéances (mais je me soigne...) et l'entraînement physique intensif (même si je dois avouer que je suis à présent à la traîne pour le footing). Merci pour ces quatre ans qui auront été des plus enrichissants pour moi !

Pour réaliser cette thèse, j'ai trouvé une terre d'asile au Laboratoire des Symbioses Tropicales et Méditerranéennes. De ce fait, je tiens à exprimer toute ma gratitude à Bernard Dreyfus qui m'a accueilli au sein de son laboratoire, qui m'a accordé sa confiance (parce que j'avais mis une photo de Plampinet en fond d'écran ?) et qui m'a fait partager sa passion pour des sujets aussi variés que la rhizobiologie tropicale ou la zootechnie alpine.

Parmi toutes les personnes qui travaillent au LSTM, je tiens à remercier plus particulièrement ceux qui m'ont aidé à réaliser ce travail sur les symbiotes de *Medicago* : Jean-Claude a témoigné un intérêt tout particulier pour mon sujet, n'hésitant pas à enfiler la blouse à la moindre occasion ; Brigitte m'a initié au modèle *M. laciniata* et m'a permis de discuter phylogénie avec les étudiants de l'AGRO-M ; Jean-Christophe a partagé ses chips avec moi (je veux dire ses puces à ADN !) et ses qualités tennistiques ; Lucette a bichonné mes cultures de luzernes et de *Sinorhizobium*. Enfin, je suis sûr que Cécile saura faire vivre ce sujet pendant les trois ans à venir.

J'adresse également un amical remerciement aux autres membres permanents du LSTM qui ont rendu si agréable ces quatre ans de travail grâce à leur bienveillance et à leur bonne humeur : Bruno m'a permis d'aller donner des cours dans le Gard, Fabienne ne m'a jamais tiré les oreilles lorsque je lui piquais sa place, les fameuses blagues de Marc,...

Comment ne pas également remercier mes camarades de la corporation des thésards, le revenant Lionel « Biquet » Moulin (même s'il est maintenant un brillant permanent, je ne voudrais pas qu'il prenne un coup de vieux...), ceux qui sont partis affronter le climat hostile du nord de l'Europe (Cyril, Abdala, Fred,...), ceux qui sont (ou qui vont) allés se dorer la pilule au soleil (Sophie, Rado, Chacha,...) et enfin ceux qui n'ont pas encore goûté au bonheur de la rédaction de thèse (Marianne, Laurie, Adeline, Céline(s),...).

Je tiens aussi à remercier tous ceux qui m'ont fait partager leur expérience, notamment à l'occasion de mes comités de thèse : Thomas Bataillon, Christophe Mougel, Joëlle Ronfort,...

Enfin, quelques mots pour les miens. Je remercie mes parents pour leur affection, leur soutien moral et matériel. Mon père, pour m'avoir fait découvrir les questionnements fondamentaux des sciences de la vie et de la terre (« et dire que c'est cette petite rivière... »). Ma mère, pour son optimisme obstiné et son encadrement énergique (« Xavier, as-tu bien fait ton travail ? »). Un petit mot pour mes grands-parents et mon oncle, parce que même si je n'ai pas fait « rougir les pipettes », j'ai fini par soutenir. Last but not least..., un immense merci à Céline, qui supporte ledit « thésard en folie » en dehors de son habitat naturel. Ton soutien et le bonheur d'être à tes côtés m'ont permis d'accomplir ce travail plein d'entrain et de bonne humeur.

Sommaire

I. Introduction générale	1
I.1. Transferts de matériel génétique entre taxons évolutivement éloignés	1
I.2. Transferts de matériel génétique entre individus apparentés	2
I.3. Présentation générale des rhizobium et de leur symbiose avec les légumineuses	4
II. Intérêts et questions relatives au modèle <i>S. meliloti</i> et <i>S. medicae</i>	5
II.1. Données génomiques disponibles pour le couple <i>Sinorhizobium</i> sp. / <i>Medicago</i> sp.	5
II.2. Spécificité de l'interaction <i>Sinorhizobium</i> sp. / <i>Medicago</i> sp.	6
II.3. Transferts horizontaux chez <i>S. meliloti</i> et <i>S. medicae</i>	6
II.4. Structure spatiale de la diversité de <i>S. meliloti</i> et <i>S. medicae</i>	7
II.5. Sexe, sélection et spéciation chez <i>S. meliloti</i> et <i>S. medicae</i>	8
III. Présentation du travail de thèse	9
IV. Taxonomie de <i>S. meliloti</i> et <i>S. medicae</i> {Article 1}	10
V. Echantillonnages de populations symbiotiques et outils de caractérisation génétiques en adéquation avec les questions posées	12
V.1. Echantillonnages de populations de <i>S. meliloti</i> et <i>S. medicae</i>	12
V.1.1. Echantillonnage E1 {article 2}	12
V.1.2. Echantillonnage E2 {article 3}	12
V.1.3. Echantillonnage E3 {article 4}	13
V.2. Caractérisation génétique des populations de <i>S. meliloti</i> et <i>S. medicae</i>	13
V.2.1. Définition d'un schéma de MLST pour caractériser l'échantillonnage E1	13
V.2.2. Caractérisation des échantillonnages E2 et E3	14
V.2.3. Caractérisation des groupes symbiotiques associés aux <i>Medicago</i>	15
VI. De la nécessité de prendre en compte la structure spatiale des données pour étudier la diversité génétique de <i>S. meliloti</i> et <i>S. medicae</i> {Article 2, article 3 et article 4}	16
VI.1. Polymorphisme de présence/absence de phénotypes symbiotiques	16
VI.2. Structuration spatiale de la diversité des populations de <i>Sinorhizobium</i>	16
VI.3. Conclusions	18
VII. Impact des transferts horizontaux sur la diversité génétique des lignées de <i>Sinorhizobium</i> sp.	19
VII.1. Transferts « à longue distance » taxonomique : l'origine des îlots symbiotiques nod de <i>S. meliloti</i> et <i>S. medicae</i> ? {Article 3}	19

VII.2. Transferts entre les groupes de spécificité chez les symbiotes des <i>Medicago</i> : vers une définition évolutionniste des concepts d'espèce et de biovars ? {Article 3}	20
VII.3. Transferts au sein de deux populations de symbiotes des <i>Medicago</i> : existe-t-il un lien entre architecture génomique et patron de recombinaison ? {Article 2}	22
VII.3.1. Etat de l'art sur le déséquilibre de liaison chez <i>S. meliloti</i> et <i>S. medicae</i>	22
VII.3.2. Inférence de transferts au sein d'une population de <i>S. meliloti</i> bv. <i>meliloti</i> et de <i>S. medicae</i> {Article 2}	23
VII.3.3. Congruence entre architecture des génomes et patrons de recombinaisons : la modularité représente-elle un avantage adaptatif ?	24
VII.4. Conclusions	25
VIII. Evolution de la spécificité et diversification des <i>Sinorhizobium</i> spp.	27
VIII.1. Evolution vers différents groupes de spécificité {Articles 3 et 4}	27
VIII.1.1. Origine et évolution de la variabilité inter-biovars chez <i>S. meliloti</i>	27
VIII.1.2. Origine et évolution de la variabilité inter-spécifique chez <i>S. meliloti</i>	30
VIII.2. Evolution des populations de <i>S. meliloti</i> et <i>S. medicae</i> interagissant avec des plantes hôtes présentant des spectres d'hôtes comparables : des spécialisations tous azimuts ? {Article 2}	33
VIII.3. Remarque conclusive	38
IX. Conclusions et (surtout) perspectives	40
IX.1. Transferts horizontaux	40
IX.2. Influence des plantes hôtes du genre <i>Medicago</i> sur la diversification des <i>Sinorhizobium</i> sp.	41
IX.2.1. Evolution expérimentale	41
IX.2.1. Modélisation	42
IX.3. Biogéographie	45
X. Conclusion générale	49
XI. Références	51
XII. Article 1	62
XIII Article 2 : accepté dans « Molecular Ecology »	74
XIII Article 3 : en préparation	99
XIV. Article 4	129

I. Introduction générale

I. Introduction générale

Depuis une quinzaine d'années, notamment sous l'impulsion de John Maynard Smith (Maynard Smith 1990; Maynard Smith et al. 1991; Maynard Smith 1993; Maynard Smith et al. 1993), les bactéries ne sont plus considérées comme des organismes strictement clonaux. De nombreux exemples illustrent à présent l'importance des transferts horizontaux de matériel génétique dans l'évolution des procaryotes. Ces transferts peuvent être classés en deux types d'événements : i) des transferts entre taxons évolutivement éloignés ; ii) des transferts horizontaux de matériel génétique au niveau sub-spécifique.

I.1. *Transferts de matériel génétique entre taxons évolutivement éloignés*

Les génomes procaryotes séquencés (316 sur la base des données fournies par le site du « National Center for Biotechnology Information ») ou en cours de séquençage (aujourd'hui 580 selon la même source) représentent une source d'information formidable pour documenter des événements de transferts entre bactéries évolutivement éloignées et étudier leur impact sur l'émergence de groupes génétiques *via* l'acquisition de nouvelles compétences. Ces transferts peuvent intervenir dans le cadre de trois mécanismes distincts (Mazodier and Davies 1991) : i) la transformation, *i.e.* par acquisition d'un fragment d'ADN exogène libre ; ii) la transduction, au cours de laquelle l'ADN exogène est transféré par l'intermédiaire d'un phage ; iii) la conjugaison, c'est-à-dire par transfert d'un plasmide (Grohmann et al. 2003) ou d'un fragment d'ADN intégratif (Burrus et al. 2002) entre deux bactéries .

La distribution des gènes de résistance aux antibiotiques ou des gènes de dégradation de composés xénobiotiques sont les exemples les plus utilisés pour illustrer ce type de transferts (Davison 1999). Les *rhizobium*, un groupe fonctionnel de bactéries capables de fixer l'azote atmosphérique en symbiose avec les légumineuses, peut lui aussi constituer un exemple à part entière. En effet, la diversification de ce groupe fonctionnel incluant des α et des β -protéobactéries est en partie liée au transfert des déterminants génétiques de l'interaction avec la plante hôte (Moulin et al. 2001; Chen et al. 2003).

La fréquence des transferts aboutissant à une intégration de l'ADN exogène dans le génome receveur semble diminuer de manière exponentielle avec la distance phylogénétique entre le donneur et le receveur (Vulic et al. 1997). Cette relation est due à deux contraintes

majeures liées d'une part à la nécessité d'une forte homologie entre l'ADN exogène et le génome receveur au niveau des extrémités du fragment transféré et d'autre part un niveau d'homologie suffisant sur l'ensemble de la séquence recombinée. D'après la littérature, l'intensité de ces contraintes dépend de l'espèce receveuse (Majewski and Cohan 1998; Majewski and Cohan 1999b; Majewski et al. 2000).

Après l'intégration, la probabilité de fixation d'un ADN exogène dans les populations du taxon receveur dépend nécessairement de ses conséquences éventuelles en terme de valeur sélective (Berg and Kurland 2002). En outre, un travail théorique récent indique que l'insertion d'ADN exogène dans un génome bactérien peut être à l'origine d'une spéciation (*i.e.* une différenciation sur l'ensemble du génome) chez l'espèce receveuse (Vetsigian and Goldenfeld 2005). Une telle insertion peut en effet être à l'origine d'une zone de faible homologie entre individus conspécifiques. Au sein de cette zone, la probabilité de recombinaison homologue avec un individu de la même espèce serait donc limitée. L'accroissement progressif de la zone où la probabilité de recombinaison est faible, liée à l'accumulation de mutations sur les régions avoisinantes, pourrait mener à l'émergence de deux espèces.

I.2. *Transferts de matériel génétique entre individus apparentés*

Les études sur l'impact des transferts horizontaux intraspécifiques se focalisent majoritairement sur des organismes entretenant des interactions durables avec des hôtes eucaryotes (allant du parasitisme au mutualisme). Elles cherchent notamment :

a) à comprendre quelle est l'importance des transferts horizontaux sur la diversification des bactéries. Dans un premier temps, l'utilisation de mesures d'association multilocus sur des données alloenzymatiques a permis de révéler que le niveau de déséquilibre de liaison des populations bactériennes pouvait varier d'un déséquilibre de liaison total à une situation d'équilibre de liaison en fonction de l'espèce étudiée (Maynard Smith et al. 1993). Plus récemment, des données de séquençage multilocus ont permis de montrer qu'il existait entre espèces bactériennes des variations d'intensité ainsi que des variations du ratio des taux de mutation et de recombinaison populationnels, θ et ρ , qui influencent leur diversité (Hanage et al. 2006).

b) à identifier des barrières génétiques liées à la spécificité d'hôte ou à l'isolement géographique. Ainsi, le comportement des populations hôtes ou la distance

géographique ont chacun été invoqués pour expliquer le niveau de différenciation entre populations bactériennes. Par exemple, la structure génétique des populations d'*Helicobacter pylori* dans la région du Ladakh est intimement liée à la structure sociale des populations humaines (Wirth et al. 2004) et celle des populations de *Rhizobium leguminosarum* semble être influencée par la distance géographique à l'échelle mondiale (Souza et al. 1992).

c) à savoir si la recombinaison peut limiter les interactions entre les pressions de sélection qui s'exercent sur différents gènes bactériens. Toutefois, cet aspect de recherche reste relativement limité dans la mesure où les travaux visant à estimer un taux de recombinaison moyen dans les populations bactériennes se sont intéressés à des locus de « ménage » situés sur une même unité de réplication et dépourvus de traces de sélection, du moins de sélection diversifiante. Par exemple, une étude portant sur des bactéries pathogènes du genre *Burkholderia* (Godoy et al. 2003) présente les locus choisis comme des « housekeeping genes ... that were flanked by other housekeeping genes and that appeared to be devoid of nearby genes that might be under diversifying selection from the host immune system or that might have been subject to high rates of horizontal gene transfer were selected », ces gènes de ménage étant positionnés sur une des trois unités de réplifications majeures. Les différences entre les résultats obtenus sur la base de ce type de marqueurs et ceux obtenus lors de travaux incluant des gènes influencés par une forte sélection diversifiante ont parfois abouti à des débats énergiques concernant l'influence des mécanismes de recombinaison sur la diversité des populations bactériennes, notamment en ce qui concerne l'impact des événements de recombinaison sur la diversité des populations de *Borrelia burgdorferi* (Dykhuizen and Baranton 2001; Stevenson 2001). Les résultats obtenus par le biais du séquençage de trois génomes de cette espèce et l'analyse de données de séquence, incluant les deux types de locus, obtenues chez plusieurs isolats (Qiu et al. 2004) ont permis de montrer l'influence des transferts horizontaux sur les patrons de diversité observés. Ceci souligne l'intérêt d'approches utilisant des locus potentiellement sujets à différentes pressions de sélection de manière à mettre en évidence d'éventuelles variations du mode de transmission au sein des génomes bactériens étudiés. En effet, le fait que des locus soient génétiquement liés peut entraîner des interférences qui limitent l'efficacité de la sélection sur chacun d'eux, à l'image de l'effet Hill-Robertson (Hill and Robertson 1966). Or, ce type

d'interférence pourrait être impliqué dans l'évolution du sexe (Otto and Lenormand 2002).

Comme indiqué ci-dessus, les études sur l'évolution des populations bactériennes peuvent maintenant s'appuyer sur le séquençage de plusieurs génomes au sein d'une espèce. Par exemple quinze génomes des pathogènes *Burkholderia mallei* et *Burkholderia pseudomallei* ou huit génomes de *Burkholderia cepacia sensu lato* (un groupe de bactéries incluant des pathogènes opportunistes animaux, des phytopathogènes, des bactéries rhizosphériques et des *rhizobium*) sont actuellement séquencés. Elles peuvent également se baser sur le séquençage d'importantes portions génomiques, comme par exemple 30 000 paires de bases séquencées chez 56 souches de *Bacillus anthracis* (Zwick et al. 2005) ou chez 50 souches de *Staphylococcus aureus* (van Leeuwen et al. 2003) (même s'il faut bien avouer que l'analyse des données obtenues dans ces deux cas n'est pas très aboutie), ou profiter des deux approches (Qiu et al. 2004).

I.3. Présentation générale des *rhizobium* et de leur symbiose avec les légumineuses

Ma thèse a été l'occasion de développer des recherches (à une échelle bien plus modeste que celles mentionnées ci-dessus) adressant les questions précédemment présentées à deux espèces de *rhizobium* qui interagissent avec les plantes du genre *Medicago* : *Sinorhizobium meliloti* et *Sinorhizobium medicae*. Les *rhizobium* fournissent un modèle d'intérêt pour l'étude des interactions mutualistes et de leur évolution. Ces bactéries sont, en partie, à l'origine des capacités de colonisation et à l'origine de l'importance agronomique des Fabaceae (parmi lesquelles les espèces du genre *Medicago*). Elles sont capables d'induire, chez leurs hôtes, la formation d'un organe spécialisé, la nodosité. Le développement de cette structure est contrôlé par un dialogue moléculaire complexe entre les deux partenaires. Le principal facteur de spécificité bactérien intervenant dans la mise en place de cette interaction est dénommé Facteur Nod. La synthèse de cette molécule, clef d'entrée de la bactérie au sein de la plante, est due aux protéines codées par un groupe de gènes appelés *nod*, *noe* et *nol* qui sont généralement regroupés au sein d'un « îlot symbiotique » (Spaink 2000). Au sein des nodosités, les *rhizobium* sont capables de fixer l'azote atmosphérique, produisant ainsi des composés azotés organiques qui sont utilisés par la plante hôte. En contrepartie, les bactéries bénéficient de conditions de croissance optimales, à condition que la quantité d'azote fixée disponible pour la plante dépasse un seuil limite (Kiers et al. 2003).

II. Intérêts et questions relatives au modèle *S. meliloti* et *S. medicae*

II. Intérêts et questions relatives au modèle *S. meliloti* et *S. medicae*

Les pressions sélectives liées aux hôtes et les transferts horizontaux de matériel génétique sont des mécanismes clefs qui influencent la structure génétique des populations des micro-organismes symbiotiques (Tibayrenc 1996). Comprendre leur influence respective sur la diversification des *rhizobium* est primordial à la fois d'un point de vue fondamental et à la fois pour mettre en place une gestion durable de ces ressources génétiques d'intérêt écologique et agronomique. Parmi les couples symbiotiques utilisés pour étudier l'interaction *rhizobium*/Fabaceae, *Medicago* sp. / *Sinorhizobium* sp. est un cas particulièrement intéressant (Cook 1999).

II.1. Données génomiques disponibles pour le couple *Sinorhizobium* sp. / *Medicago* sp.

Medicago truncatula est un modèle biologique émergent pour les études portant sur des plantes pouvant interagir avec des microorganismes mutualistes, *Arabidopsis thaliana* n'étant ni mycorhizée ni associée à des bactéries mutualistes telles que les *rhizobium* ou les *Frankia*. Le génome de *M. truncatula* est actuellement en cours de séquençage (Young et al. 2005). Du côté bactérien, le génome de la souche *S. meliloti* 1021 est disponible (Galibert et al. 2001) et le projet de séquençage d'une souche de *S. medicae* vient d'être accepté. De plus, des expériences d'électrophorèse en champ pulsé ont montré que *S. meliloti* et *S. medicae* partageaient une même architecture de génome, composée d'un chromosome et de deux mégaplasmines pSymA et pSymB (Roumiantseva et al. 1999). Les tailles respectives de ces unités de réplication sont 3,65 Mb, 1,35 Mb et 1,68 Mb pour le génome de *S. meliloti* souche 1021. Ces unités de réplifications possèdent des caractéristiques fonctionnelles différentes. En effet, alors que les gènes impliqués dans le métabolisme de base de la bactérie sont localisés sur le chromosome, les gènes impliqués dans des métabolismes secondaires, généralement appelés gènes accessoires, sont localisés sur les mégaplasmines. Le mégaplasmine pSymA contient notamment un îlot de gènes symbiotiques regroupant la majorité des gènes impliqués dans l'interaction avec les plantes hôtes du genre *Medicago*. Cette architecture de génome fournit un modèle d'intérêt pour étudier le niveau de sélection s'appliquant à des locus

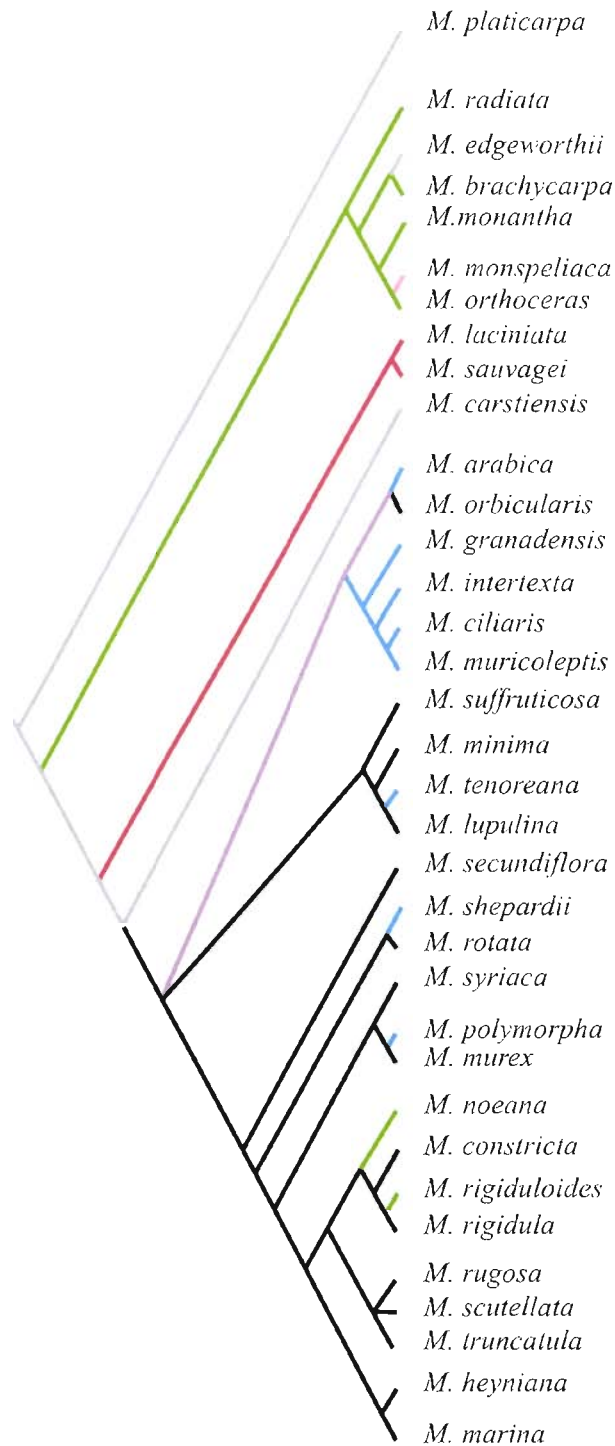


Figure 1. Evolution des capacités symbiotiques des espèces du genre *Medicago* inférée sur la base d'une phylogénie obtenues à partir des séquences d'espaces transcrits des ADNr nucléaires (d'après Béna et al. (2005)). Les branches figurées en couleur correspondent respectivement à des capacités symbiotiques : totalement indéterminées (gris), permettant une symbiose avec un écotype particulier de *S. meliloti* (vert); avec *S. meliloti* bv. *medicaginis* (rouge); avec *S. meliloti* bv. *meliloti* (rose), avec *S. meliloti* bv. *meliloti* et *S. medicae* (noir); avec *S. medicae* (bleu); avec *S. medicae* ou avec *S. medicae* et *S. meliloti* bv. *meliloti* (mauve).

impliqués dans différentes fonctions du métabolisme bactérien et à leur indépendance évolutive.

II.2. Spécificité de l'interaction *Sinorhizobium* sp. / *Medicago* sp.

Toutes les espèces de *Medicago*, dont les symbiotes ont été isolés jusqu'à présent, interagissent de manière spécifique avec des bactéries appartenant aux espèces *S. meliloti* et *S. medicae* à l'exception de *Medicago ruthenica* qui s'associe avec des souches de l'espèce *Rhizobium mongolense* (de Lajudie et al. 1994; Rome et al. 1996; Van Berkum et al. 1998). Malgré le fait que seules trois espèces bactériennes interagissent efficacement avec le genre *Medicago*, ces plantes présentent un polymorphisme de spécificité relativement important. Plus précisément, les espèces de *Medicago* peuvent être groupées en fonction de leurs partenaires symbiotiques (Bena et al. 2005) (Fig. 1): (i) certaines espèces comme *M. truncatula* sont capables de mettre en place une symbiose efficace avec des souches *S. meliloti* apparentées à la souche 1021 et des souches de *S. medicae*; (ii) des espèces comme *Medicago polymorpha* interagissent de manière efficace seulement avec les souches de *S. medicae*; (iii) *Medicago laciniata* et *Medicago sativa* entretiennent une relation mutualiste avec certaines souches de *S. meliloti* caractérisés par des allèles particuliers au niveau du gène *nodC* (Barran et al. 2002); (iv) des espèces telles que *Medicago rigiduloides* ou *Medicago noeana* établissent également leur symbiose avec des génotypes particuliers de *S. meliloti*, mais le déterminisme de cette spécificité n'a pas encore été mis en évidence. Compte tenu du polymorphisme de spécificité observé, la diversité génétique des isolats échantillonnés dans des nodosités de plantes du genre *Medicago* devrait dépendre de l'espèce hôte. De plus, le génotype au sein d'une espèce de *Medicago* peut également influencer la diversité génétique des *Sinorhizobium* sp. échantillonnés dans les nodosités, comme certaines études conduites sur la luzerne cultivée *Medicago sativa* le suggèrent (Paffetti et al. 1998). Malgré tout, peu de travaux se sont intéressés à l'influence de la diversité des plantes hôtes sur la diversité génétique des symbiotes échantillonnés. De plus, aucune étude n'a focalisé son attention sur les traces de sélection qui pourraient être mises en évidence sur les données de séquences obtenues sur la base de locus impliqués dans l'interaction symbiotique au sein de populations de *rhizobium*.

II.3. Transferts horizontaux chez *S. meliloti* et *S. medicae*

En dehors des plantes hôtes, le régime de reproduction peut également avoir un impact sur la structure génétique des populations de bactéries symbiotiques. De précédentes études ont décrit la différenciation génétique et l'isolement sexuel entre *S. meliloti* et *S. medicae* (Eardly et al. 1990; Maynard Smith et al. 1993). De plus, Maynard Smith *et al.* (1993) ont suggéré un taux de recombinaison important au sein de chaque espèce en décrivant des structures « quasi-panmictiques ». Toutefois, une étude récente a observé des niveaux de déséquilibre de liaison important au sein d'un ensemble de souches de *S. meliloti* isolées de différentes espèces de *Medicago* à partir de sols d'Asie Centrale (Roumiantseva et al. 2002). La différence entre les résultats obtenus par Maynard Smith *et al.* (1993) et par Roumiantseva *et al.* (2002) concernant le niveau de déséquilibre de liaison de *S. meliloti* et de *S. medicae* montre que le régime de reproduction de ces bactéries n'est pas parfaitement défini et pose deux questions : (i) les premiers résultats sont-ils biaisés par les problèmes de calcul soulevés par Haubold *et al.* (1998) ? (ii) Le plan d'échantillonnage utilisé dans la seconde étude est-il responsable du patron de déséquilibre de liaison observé ? En effet, si le déséquilibre de liaison est calculé à partir de génotypes issus d'un petit nombre de populations génétiquement différenciées, le niveau de déséquilibre de liaison observé sera à la fois le reflet du régime de reproduction et celui du niveau de différenciation entre les populations échantillonnées (Tachida and Cockerham 1986). Cette propriété du déséquilibre de liaison a d'ailleurs été utilisée pour discuter du niveau de différenciation des populations de *Rhizobium leguminosarum* (Souza et al. 1992) ou de *Helicobacter pylori* (Wirth et al. 2004). Pour étudier le régime de reproduction, il semble donc pertinent d'étudier le niveau de déséquilibre de liaison de populations de *S. meliloti* et *S. medicae* à une échelle locale.

II.4. Structure spatiale de la diversité de *S. meliloti* et *S. medicae*

Afin de savoir si les mesures de déséquilibre de liaison peuvent être affectées par le plan d'échantillonnage choisi, il faut connaître le niveau de structuration spatiale de la diversité génétique des populations symbiotiques de *Sinorhizobium*. S'il est notable que les aires de répartition des espèces de *Medicago* présentant différentes spécificités d'interaction sont structurées à l'échelle du bassin méditerranéen (Small and Jomphe 1988), peu d'études se sont intéressées à la distribution spatiale de la diversité génétique de *S. meliloti* et *S. medicae*. Différents travaux ont montré que *S. medicae* était absent de certains sols (Roumiantseva et al. 2002; Zribi et al. 2004), mais la différenciation géographique des populations de *S. meliloti* ou *S. medicae* a été peu documentée. L'échantillonnage de

populations bactériennes dans des stations distantes à l'échelle du bassin méditerranéen devrait donner des informations à ce sujet. Par ailleurs, ce type d'échantillonnage pourrait permettre d'apporter à terme un éclairage sur la relation entre les aires de répartition des espèces du genre *Medicago* et la structure des populations de *Sinorhizobium* qui leur sont associées.

II.5. Sexe, sélection et spéciation chez *S. meliloti* et *S. medicae*

Enfin, l'impact relatif des spécialisations écologiques et des transferts horizontaux de matériel génétique sur la diversification et la spéciation des bactéries est l'objet d'un débat qui connaît un regain d'intérêt suite à l'augmentation des capacités de séquençage et les analyses qui peuvent en découler (Cohan 2001; Cohan 2002; Lawrence 2002; Gevers et al. 2005). En plaçant les mutations adaptatives dans plusieurs fonds génétiques, la recombinaison (après acquisition de matériel génétique par transformation, conjugaison ou transduction) pourrait en effet représenter un processus capable de restreindre l'action d'une sélection disruptive à une portion de génome, et donc d'empêcher la spéciation (Majewski 2001). Toutefois, les modèles mathématiques disponibles suggèrent que la recombinaison devrait avoir un effet très limité sur le mécanisme de spéciation bactérienne (Cohan 1994a; Cohan 1994b; Cohan 1995; Majewski and Cohan 1999a; Vetsigian and Goldenfeld 2005). Aussi, peut-on se demander quelle part du génome présente une différenciation entre les écotypes de *S. meliloti* dans le but de déterminer sur la base d'approches évolutives le statut taxonomique adéquat (*i.e.* espèce ou biovar) pour les identifier.

III. Présentation du travail de thèse

III. Présentation du travail de thèse

Dans ce contexte, l'objet de cette thèse a été d'apporter des informations relatives aux interrogations décrites ci-dessus. Plus précisément, nous avons étudié : i) la répartition spatiale de la diversité chez *S. meliloti* en mesurant la différenciation entre des populations bactériennes issues de trois localités du bassin méditerranéen ; ii) le régime de reproduction de populations de *S. meliloti* et *S. medicae* par le biais de mesures de déséquilibre de liaison et par comparaison de phylogénies obtenues à l'aide de différents marqueurs génétiques ; iii) l'influence des plantes hôtes sur la diversité génétique et taxonomique des populations symbiotiques de *Sinorhizobium* en mesurant la différenciation entre des populations bactériennes isolées des nodosités de différents lots de plantes hôtes et par la recherche de traces de sélection. Dans cette synthèse, je me suis principalement attaché à discuter les résultats obtenus lors des différents travaux effectués (dont les publications associées sont fournies à la suite du manuscrit), à mettre en relief l'intérêt du genre *Sinorhizobium* comme modèle biologique pour effectuer des travaux sur l'évolution des bactéries mutualistes et à énoncer les questions de recherche qui, à mon avis, découlent de ce travail. Ainsi, après avoir défini taxonomiquement les groupes bactériens qui constituent le modèle biologique étudié {Article 1}, les matériels biologiques et les méthodes de caractérisation utilisés sont présentés. Les résultats obtenus à l'aide des données acquises sont ensuite discutés. Dans un souci de clarification, les informations contenues dans les articles 2, 3 et 4 ont été réorganisées de manière à présenter successivement l'ensemble des éléments de discussion relatifs à l'influence de la distance géographique, des transferts horizontaux et des plantes hôtes sur la structure génétique des populations de *S. meliloti* et *S. medicae*. Enfin, je conclus cette synthèse par une présentation de différentes perspectives de recherche.

IV. Taxonomie de *S. meliloti* et *S. medicae* { Article 1 }

IV. Taxonomie de *S. meliloti* et *S. medicae* { Article 1 }

Le genre *Sinorhizobium* a été défini par Chen *et al.* (1988) pour différencier *Sinorhizobium fredii* (auparavant *Rhizobium fredii*) du genre *Rhizobium*. Par la suite, différentes espèces ont été assignées au genre *Sinorhizobium* dont *S. meliloti* (anciennement *Rhizobium meliloti*) sur la base des critères utilisés dans le cadre de la taxonomie polyphasique bactérienne, basée sur la congruence de données acquises lors du séquençage de l'ADN 16S, d'hybridations ADN-ADN, et de caractérisations phénotypiques de différentes espèces (Encadré 1) (de Lajudie *et al.* 1994). La majorité des souches du genre *Sinorhizobium* est capable d'établir une interaction mutualiste avec différentes plantes de la famille des Fabaceae. La seule exception concerne certaines souches de l'espèce de *Sinorhizobium adhaerens*. En effet, cette espèce, précédemment décrite sous le nom d'*Ensifer adhaerens* (Casida Jr 1982), comprend à la fois des souches saprophytes (*i.e.* dépourvues de gènes *nod* et incapables de noduler) et des souches capables d'interagir avec certaines Fabaceae (Rogel *et al.* 2001). L'existence de la capacité à interagir avec des Fabaceae chez certaines souches a été à l'origine de nouvelles études taxonomiques qui ont abouti à la reclassification de cette espèce (Willems *et al.* 2003).

Au sein du genre *Sinorhizobium*, on constate une variation importante du spectre d'hôte en fonction des taxons rencontrés. Notamment, des changements de spectre d'hôte sont constatés à l'intérieur de différentes espèces du genre *Sinorhizobium* (espèces définies sur la base des techniques classiques de taxonomie polyphasique). Les écotypes, au sein de chaque espèce, sont alors appelés biovar (bv.). Malgré tout, on rencontre des cas où différents groupes de bactéries partagent certains hôtes. Un des cas les plus surprenant est celui des espèces *Sinorhizobium terangaie* et *Sinorhizobium sahelii* qui comprennent chacune deux biovars appelés *acaciae* et *sesbaniae* (de Lajudie *et al.* 1994), en référence aux plantes hôtes des genres *Acacia* et *Sesbania*. Il est notable que les souches de *S. terangaie* et *S. sahelii* qui partagent les mêmes plantes hôtes présentent des gènes *nodA* fortement apparentés. Dans le cas des souches de *Sinorhizobium* associées au genre *Medicago*, on trouve à la fois des groupes génétiques présentant des spectres d'hôtes chevauchants et des groupes génétiques présentant des spectres d'hôtes distincts. Les souches de *S. meliloti* présentant le même spectre d'hôte que la souche 1021 ne réalisent une symbiose efficace qu'avec un sous-ensemble des espèces de *Medicago* capables d'interagir avec les souches de *S. medicae*. Cette

Encadré 1. Taxonomie des bactéries et le problème de la définition des espèces

L'évolution de la taxonomie, surtout dans le cas de la description des espèces de bactéries, est guidée par deux objectifs distincts et parfois antagonistes. Le premier est de différencier les taxons qui présentent des propriétés particulières (virulence, spectre d'hôte,...) avec des implications évidentes dans le domaine médical ou agronomique. Le second est de délimiter des entités biologiques adéquates pour mener à bien des études évolutives à l'échelle sub-spécifique ou macroévolutive.

Pendant longtemps, l'absence de concepts permettant de déterminer les limites des espèces bactériennes, comme le concept d'espèce biologique utilisé chez les eucaryotes, a amené à privilégier l'intérêt pratique de la taxonomie bactérienne par rapport à son acception évolutive. Cette absence de concepts est liée d'une part au mode de reproduction des bactéries, car la clonalité ne permet pas d'appliquer les principes basés sur l'interfécondité, et à l'existence de transferts horizontaux entre individus évolutivement très éloignés, ce qui peut se révéler gênant pour mettre en évidence des différences phénotypiques ou génotypiques justifiant le statut des espèces.

A l'heure actuelle, la définition de l'espèce bactérienne tend à s'appuyer sur la détermination d'un groupe d'individus génétiquement et phénotypiquement homogène sur la base d'une approche polyphasique (*i.e.* mettant en jeux différentes approches méthodologiques). Toutefois, les méthodes utilisées pour définir ces entités et la volonté d'appliquer des critères généraux font que la taxonomie bactérienne reste basée sur des critères relativement arbitraires. Ainsi, un groupe de bactéries doit généralement remplir trois conditions pour accéder au statut d'espèce : (i) présenter un certain niveau d'homologie génomique, délimité par un seuil de 70% en hybridation ADN/ADN. Les principaux problèmes liés à ce critère sont d'une part qu'il n'est pas applicable aux bactéries qui ne sont pas cultivables (pour des raisons méthodologiques) et d'autre part qu'il ne prend pas en compte la diversité des niveaux de polymorphisme qui peut exister entre espèces (tant au niveau du polymorphisme de présence/absence de gènes, voir d'unités de réplication, qu'au niveau du niveau de la divergence nucléotidique moyenne). (ii) Présenter plus de 97% d'homologie nucléotidique au niveau de l'ADNr 16S. Les limites de cette approche sont liées à l'existence d'un seuil limite faisant l'hypothèse d'une vitesse d'évolution identique pour toutes les espèces et d'un ratio vitesse d'évolution du génome sur vitesse d'évolution du marqueur choisi constant. (iii) Présenter des caractères phénotypiques cohérents (capacités symbiotiques, résistance aux antibiotiques, ...), avec le risque de diviser des groupes présentant un fond génétique commun mais qui sont différenciés au niveau de gènes liés à des niches écologiques particulières (des écotypes). Les différents problèmes soulevés sont à l'origine de la remise en question de la discrimination des genres *Agrobacterium* et *Rhizobium* (Young et al. 2001; Young et al. 2003).

Avec la démocratisation des capacités de séquençage ou l'apparition d'outils de génomique comparative performants, il est maintenant possible de mener des études prenant en compte des considérations évolutives dans la délimitation des espèces bactériennes. Ces approches permettent également de poser des questions relatives aux processus liés aux divergences observées. Dans le cas de groupes génétiques où le sexe joue un rôle important dans la distribution de la variabilité génétique, on peut différencier le statut d'espèce de celui d'écotype (ou biovar) par l'analyse de données génétiques à différents locus.

différence phénotypique et les faibles valeurs d'hybridation ADN/ADN obtenues en comparant les génomes de ces deux types de souches sont à l'origine de la description des espèces *S. meliloti* et *S. medicae* (Rome et al. 1996). Au sein de l'espèce *S. meliloti*, on trouve plusieurs groupes de spécificité interagissant avec des plantes hôtes distinctes. Toutefois, les données d'hybridation ADN/ADN ou de séquençage des ADNr 16S ne sont pas compatibles avec la définition de différentes espèces. Ainsi, deux biovars ont été décrits {Article 1} : (i) *S. meliloti* bv. *meliloti* regroupe les souches capables d'interagir avec des espèces telles que *M. truncatula* ou la luzerne cultivée *Medicago sativa* (c'est notamment le cas de la souche *S. meliloti* 1021), (ii) *S. meliloti* bv. *medicaginis* regroupe les souches capables d'interagir avec *Medicago laciniata* et probablement *Medicago sativa*. Le statut taxonomique des souches capables d'interagir avec les espèces telles que *Medicago noeana* ou *Medicago rigiduloides* n'a pas encore été défini mais devrait aboutir à la description d'un nouveau biovar au sein de *S. meliloti*. En effet, ces souches semblent partager le fond génétique de l'espèce *S. meliloti*, du fait de l'absence de divergence au niveau des séquences des gènes 16S ou *recA*, mais présentent des capacités symbiotiques singulières (2005). Si le déterminisme de cette spécificité d'interaction est pour l'instant inconnu, l'absence de différenciation avec les populations sympatriques isolées de *M. truncatula* sur *IGS_{RKP}*, *IGS_{GAB}* et *IGS_{EXO}* (voir le chapitre VI.2) et le positionnement du groupe phénotypique sur le super-arbre *nodABC-nodEG* (voir le chapitre VII.1), suggèrent que leur spectre d'hôte est lié aux allèles de gènes *nod* qu'elles possèdent.

Si les méthodes de taxonomie polyphasique nous ont permis de délimiter des entités taxonomiques sur la base de critères qui peuvent paraître arbitraires (voir l'encadré 1), nous avons essayé par la suite d'étudier les patrons de diversité et les processus évolutifs justifiant ces regroupements.

V. Echantillonnages de populations symbiotiques et outils de caractérisation génétiques en adéquation avec les questions posées

V. Echantillonnages de populations symbiotiques et outils de caractérisation génétiques en adéquation avec les questions posées

V.1. Echantillonnages de populations de *S. meliloti* et *S. medicae*

Nous avons effectué trois échantillonnages de bactéries symbiotiques dans le but de répondre aux questions posées dans le cadre de cette thèse.

V.1.1. Echantillonnage E1 {article 2}

Le but du premier échantillonnage était de savoir si le niveau de diversité des populations de *S. meliloti* et de *S. medicae* dépendait de la diversité des plantes hôtes utilisées pour le piégeage. Nous souhaitons également estimer le niveau de déséquilibre de liaison entre et au sein des unités de réplication pour des populations de *S. meliloti* et *S. medicae* isolées d'un même site. Cet échantillonnage {Article 2} a été effectué avec 20 espèces de *Medicago*, 20 lignées de *M. truncatula* et 20 plantes d'une même lignée fixée de *M. truncatula* à partir d'un sol français échantillonné à Sainte colombe l'Eglise (Pyrénées orientales, longitude 02°45'26" Est – latitude 42°37'59" Nord). Il nous a permis d'obtenir 52 isolats de *S. meliloti* appartenant au groupe génétique capable d'interagir avec *M. truncatula* (*S. meliloti* bv. *meliloti*) et 64 isolats de *S. medicae*.

V.1.2. Echantillonnage E2 {article 3}

Les objectifs du second échantillonnage étaient de comparer les patrons de divergence interspécifiques et interécotypes et de savoir si des recombinaisons avaient lieu entre les écotypes de *S. meliloti*, afin de discuter des différents scénarios évolutifs pouvant être à l'origine des structures génétiques observées. Cet échantillonnage {Article 3} a été réalisé à l'aide de *M. laciniata* et *M. truncatula* à partir de deux sols échantillonnés dans l'aire de répartition des deux espèces de plantes hôtes. Ces sols sont originaires de Tunisie et ont été prélevés à Enfidha (longitude 10°23' Est – latitude 36°07' Nord) et à Hadjeb (longitude 9°33' Est – latitude 35°24' Nord). Il nous a permis d'obtenir 60 isolats de *S. meliloti* associés à

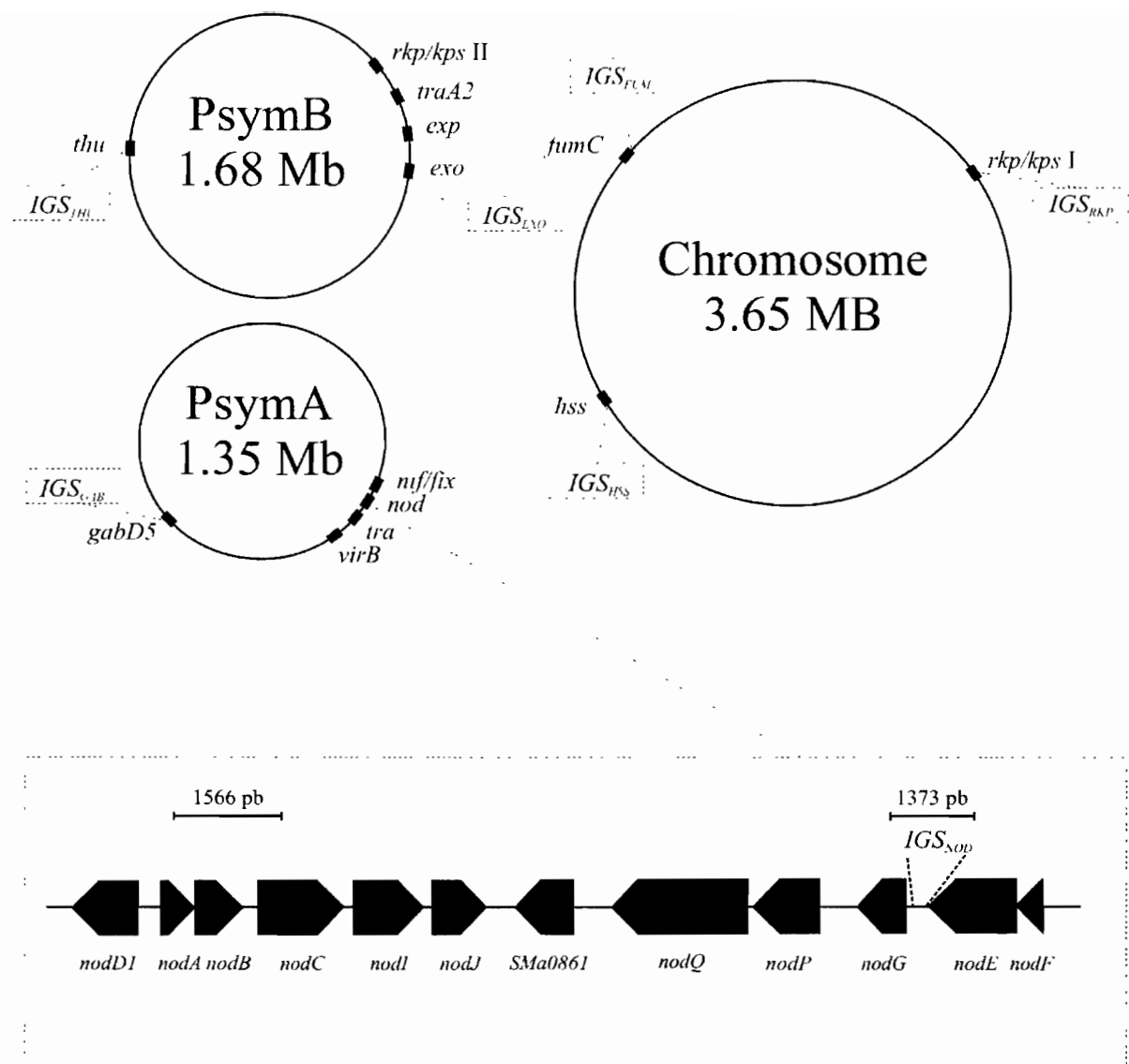


Figure 2. Localisation sur le génome de *Sinorhizobium meliloti* souche 1021 des différents marqueurs moléculaires utilisés pour caractérisés les différents isolats de *S. meliloti* et *Sinorhizobium medicae*.

M. truncatula (*S. meliloti* bv. *meliloti*), 40 isolats de *S. meliloti* associés à *M. laciniata* (*S. meliloti* bv. *medicaginis*), et 20 isolats de *S. medicae*.

V.1.3. Echantillonnage E3 {article 4}

(iii) Nous souhaitons obtenir à l'aide d'un troisième échantillonnage un ensemble de souches permettant de tester un nouvel outil de génotypage et vérifier, en s'appuyant sur les génotypes précédemment caractérisés, qu'il était possible de différencier sur cette base les populations isolées de plantes hôtes différentes ou de sites différents à l'échelle du bassin méditerranéen. Ce troisième échantillonnage {Article 4} a été effectué à l'aide de *M. truncatula* et *M. rigiduloides* à partir d'un sol libanais prélevé dans la vallée de la Becka. Ce dernier nous a permis d'isoler 17 souches associées à *M. rigiduloides* et 18 souches associées à *M. truncatula*.

V.2. Caractérisation génétique des populations de *S. meliloti* et *S. medicae*

Parmi les nombreuses méthodologies disponibles pour caractériser la diversité d'isolats bactériens à l'échelle intra-spécifique, le génotypage par séquençage de plusieurs locus (MLST) paraît la plus informative (Maiden et al. 1998). Nous avons utilisé cette approche pour caractériser les isolats de France et de Tunisie. Toutefois, cette méthode est coûteuse (en temps et en argent), ce qui nous a amené à envisager une alternative de génotypage à la fin de cette thèse dans l'optique de manipulations à plus grande échelle {Article 4}. Ainsi, les échantillons libanais ont été caractérisés par le biais d'une puce de génotypage.

V.2.1. Définition d'un schéma de MLST pour caractériser l'échantillonnage E1

Afin de caractériser les isolats issus de l'échantillonnage E1 {Article 2}, nous avons défini sept locus, en l'occurrence des espaceurs intergéniques (IGS), répartis de manière régulière sur l'ensemble du génome de *S. meliloti* souche 1021, ce qui autorisait le calcul de mesures de déséquilibre de liaison entre et au sein des unités de réplication (Fig. 2). Les pressions de sélection agissant sur les gènes physiquement liés à ces IGS, notamment les gènes impliqués dans la symbiose, peuvent avoir une influence sur le polymorphisme de ces séquences non codantes par le biais de mécanismes tels que l'auto-stop génétique (Maynard

Smith and Haigh 1974). Ceci devait permettre la mise en évidence d'évènement de sélection même si les séquences étudiées ne sont pas transcrites. Toutefois, nous espérons que les IGS se révèlent plus polymorphes que certains gènes tels que l'ADNr 16S ou *recA*, un gène codant pour une enzyme impliquée dans la réparation des molécules d'ADN et les mécanismes de recombinaison, qui s'étaient révélés monomorphes à l'échelle intraspécifique (Bena et al. 2005). Les sept IGS choisis incluent trois locus chromosomiques (*IGS_{RKP}*, *IGS_{FUM}*, et *IGS_{HSS}*), deux locus sur pSymA (*IGS_{NOD}*, *IGS_{GAB}*) et deux locus sur pSymB (*IGS_{EXO}*, *IGS_{THU}*). Les principales voies métaboliques bactériennes qui sont nécessaires pour la mise en place de la symbiose ont été particulièrement étudiées (Perret et al. 2000). Ces études ont débouché sur la description de deux groupes de gènes qui influencent le phénotype symbiotique : (i) les gènes impliqués dans l'organogenèse des nodosités *via* la synthèse des Facteurs Nod, tels que les gènes *nod*, *noe* et *nol* (Spaink 2000); (ii) les gènes codant pour la synthèse de polysaccharides de surface (LPS, EPS, KPS) comme les gènes *exo* (Reuber and Walker 1993) ou *rkp* (Kereszt et al. 1998) qui agissent lors d'étapes de reconnaissance tardives (*i.e.* quand la bactérie est entrée dans la plante), en limitant la mise en place de mécanismes de défense chez l'hôte (Mithofer 2002). Ainsi, nous avons choisi les locus *IGS_{RKP}*, *IGS_{NOD}* et *IGS_{EXO}* qui sont physiquement liés à des gènes impliqués dans l'interaction avec les plantes hôtes. Inversement, les locus *IGS_{FUM}*, *IGS_{HSS}*, *IGS_{GAB}*, et *IGS_{THU}* sont physiquement liés à des gènes impliqués dans le métabolisme de base ou dans des voies métaboliques secondaires qui n'interviennent pas directement dans la symbiose.

V.2.2. Caractérisation des échantillonnages E2 et E3

Compte tenu des résultats obtenus lors de la caractérisation du premier échantillonnage, notamment un fort déséquilibre de liaison entre les marqueurs chromosomiques, nous n'avons utilisé que quatre marqueurs pour caractériser les souches tunisiennes { Article 3 } : *IGS_{RKP}*, *IGS_{NOD}*, *IGS_{GAB}* et *IGS_{EXO}*. Du fait de la structure particulière reflétée par le marqueur *IGS_{NOD}* (liée aux plantes hôtes) et de l'utilisation d'une méthode de génotypage en cours de développement, les souches libanaises ont été caractérisées par le biais de puces de génotypage au niveau de polymorphismes de séquences identifiés sur *IGS_{RKP}*, *IGS_{GAB}* et *IGS_{EXO}* grâce aux données de séquences acquises lors de la caractérisation des échantillonnages E1 et E2. D'une manière générale, le niveau de polymorphisme, les patrons de déséquilibre de liaison et les structures de population observées suggèrent que les

marqueurs choisis était adaptés aux questions posées, même si on peut s'interroger sur la différence de polymorphisme entre les IGS et les séquences codantes avoisinantes.

V.2.3. Caractérisation des groupes symbiotiques associés aux *Medicago*

Parallèlement à l'approche multilocus, visant à caractériser des populations symbiotiques, nous avons étudié les patrons de divergence entre les séquences de gènes de nodulation des bactéries associées aux différents groupes de spécificité observés chez les plantes du genre *Medicago*. Dans ce but, nous avons obtenu un super-arbre, basé sur les séquences partielles des groupes de gènes *nodABC* et *nodEG*. Cette topologie inclut au moins un représentant pour chacun des groupes bactériens qui interagissent avec les plantes hôtes du genre *Medicago*. L'ensemble des souches utilisées inclut trois souches de *S. meliloti* bv. *medicaginis*, deux souches de *S. meliloti* bv. *meliloti*, trois souches de *S. meliloti* associées à *M. rigiduloides*, *M. noeana* et *M. radiata*, une souche de *S. medicae*, une souche de *R. mongolense* associée à *M. ruthenica*, une souche de *Rhizobium galegae*, une souche de *Rhizobium leguminosarum*, une souche de *Rhizobium tropici* ainsi que la souche *Sinorhizobium* NGR 234. Cet ensemble de souche sera appelé échantillonnage E4.

VI. De la nécessité de prendre en compte la structure spatiale des données pour étudier la diversité génétique de *S. meliloti* et *S. medicae* {Article 2, article 3 et article 4}

VI. De la nécessité de prendre en compte la structure spatiale des données pour étudier la diversité génétique de *S. meliloti* et *S. medicae* {Article 2, article 3 et article 4}

Il existe, tant dans la bibliographie que dans les expériences que nous avons effectuées, des résultats suggérant que les populations de *Sinorhizobium* associées aux plantes du genre *Medicago* sont spatialement structurées. Ce patron se retrouve à la fois à l'échelle inter-spécifique et à l'échelle intra-spécifique.

VI.1. Polymorphisme de présence/absence de phénotypes symbiotiques

A l'échelle inter-spécifique, nous n'avons obtenu des isolats de *S. medicae* que dans l'un des deux sols utilisés lors de l'échantillonnage E2 {Article 3}. Cette observation est concordante avec les résultats obtenus par Zribi *et al.* (2004) à partir d'un autre échantillonnage mené en Tunisie et Roumiantseva *et al.* (2002) sur la base d'un échantillonnage effectué à partir de sols du Caucase. A l'échelle intra-spécifique, l'absence de souches de *S. meliloti* bv. medicaginis et de l'écotype de *S. meliloti* associé à *M. rigiduloïdes* observée lors de l'échantillonnage E1 {Article 2} est en accord avec l'hypothèse d'un polymorphisme de type présence/absence des phénotypes de *Sinorhizobium* associées aux plantes du genre *Medicago*.

VI.2. Structuration spatiale de la diversité des populations de *Sinorhizobium*

A l'échelle intra-spécifique, les premières indications suggérant une structure spatiale des populations de *Sinorhizobium* sont issues d'études sur les capacités symbiotiques des isolats. Notamment, les données publiées par Matteron (1991) suggèrent que les fréquences des souches capables d'interagir efficacement avec six espèces de *Medicago* diffèrent entre les sols de plusieurs pays de l'est du bassin méditerranéen (Chypre, Jordanie, Liban, Syrie et Turquie). Plus récemment, une étude a décrit un déséquilibre de liaison significatif entre la majorité des marqueurs utilisés pour caractériser des populations de *S. meliloti* isolées de nodosités de différentes espèces de *Medicago* à partir de sols du Caucase (Roumiantseva *et al.* 2002). Toutefois, le déséquilibre de liaison observé dans cette étude pourrait être lié à l'échantillonnage de plusieurs populations génétiquement isolées. En effet, les populations

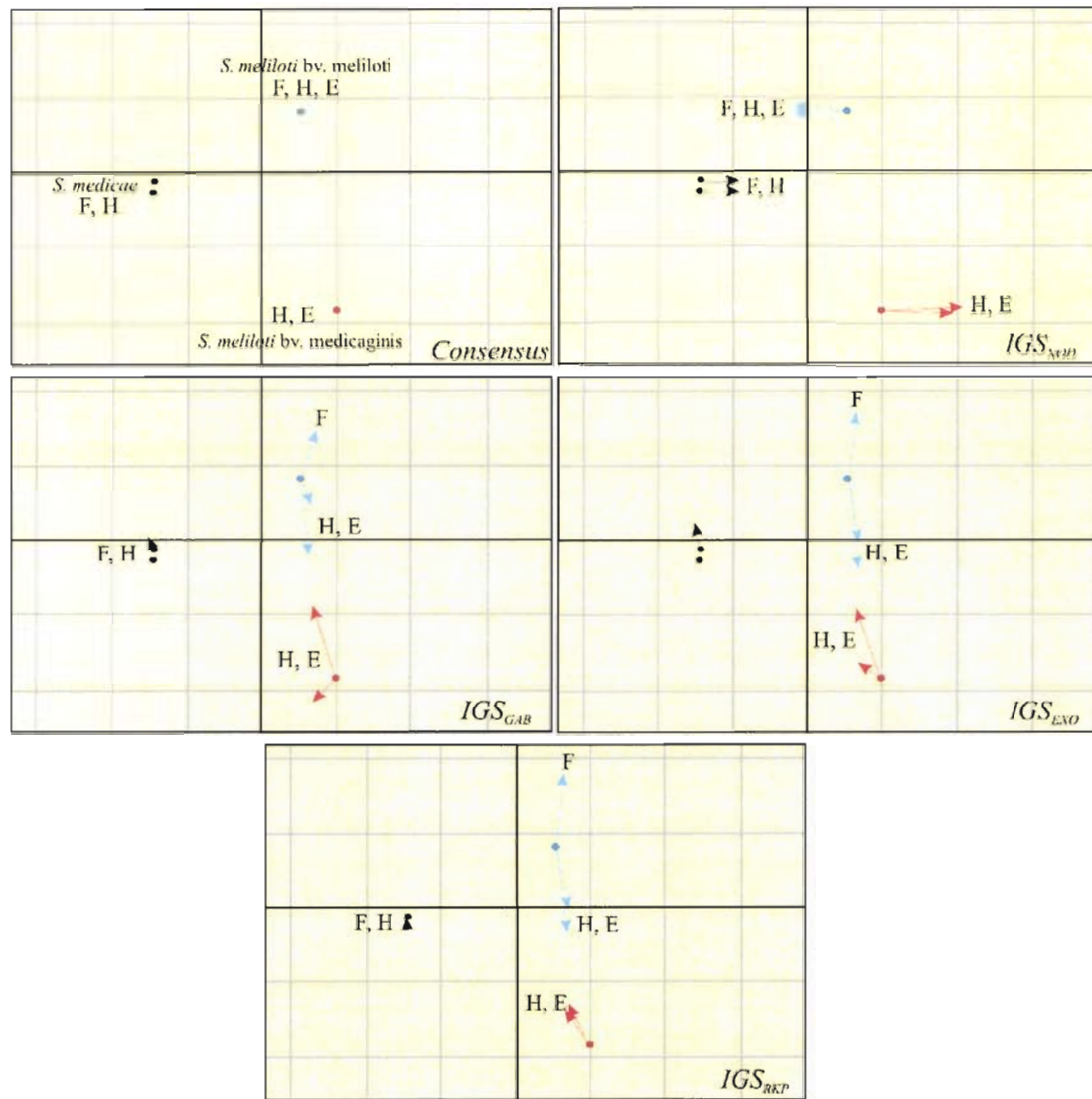


Figure 3. Exploration des jeux de données de séquence obtenus lors de la caractérisation des échantillonnages E1 et E2 pour les locus *IGS_{NOD}*, *IGS_{GAB}*, *IGS_{EXO}* et *IGS_{RKP}* par la méthode de DPCoA (Pavoine et al. 2004). Pour chaque marqueur, chaque allèle est positionné dans un espace traduisant la distance génétique entre allèles. Chaque population est alors placée au barycentre allèles qui la compose. Le consensus correspond à une typologie moyenne des populations prenant en compte l'information fournie par les différents locus. Pour chaque locus, la position consensus de chaque population est indiquée par le départ d'une flèche qui aboutit à la position de la population selon le locus étudié (pointe de la flèche). Les coordonnées sur chaque axe sont normées. En effet, les deux axes expliquent des pourcentages de variance génétique différents (95% pour l'axe 1 horizontal, 4.5% pour l'axe 2 vertical). Par rapport à l'image globale, les locus *IGS_{GAB}*, *IGS_{EXO}* et *IGS_{RKP}* ont tendance à grouper les populations de *S. meliloti* tunisiennes (Hadjeb, H et Enfidha, E) et à isoler la population française (F). Pour ces locus, le patron observé suggère donc que la distance géographique influence la structure génétique. Le locus *IGS_{NOD}* présente un patron différent qui semble particulièrement influencé par l'effet plante hôte. Ceci souligne l'indépendance évolutive de ce marqueur.

VI. Structure spatiale de la diversité génétique de *S. meliloti* et *S. medicae*

isolées dans les sites où l'effort d'échantillonnage est le plus important sont significativement différenciées ($p < 0,05$ pour le test global de différenciation génétique entre les populations 3, 4 et 5 de l'étude de Roumiantseva *et al.* (2002)), ce qui pourrait être à l'origine des déséquilibres de liaison significatifs observés.

Une analyse des jeux de données de séquence des échantillonnages E1 et E2 sur les marqueurs *IGS_{RKP}*, *IGS_{NOD}*, *IGS_{GAB}* et *IGS_{EXO}* a été réalisée à l'aide d'une extension de la méthode de double analyse en composante principale proposée par Pavoine *et al.* (2004) (Fig. 3). L'analyse montre que le jeu de données utilisé se structure en fonction de deux axes majeurs : i) un premier axe intègre près de 95% de la variabilité observée. Il illustre la différenciation et la divergence entre *S. meliloti* et *S. medicae* ; ii) un deuxième axe qui intègre environ 5% de la variabilité observée illustre la différenciation et la divergence entre souche à l'échelle intra-spécifique, plus particulièrement chez *S. meliloti*. Les comparaisons du patron de structuration globale du jeu de données avec l'image fournie par chaque locus suggèrent que la diversité de locus étudiés ne se structure pas de manière cohérente :

-au niveau du marqueur *IGS_{NOD}*, on trouve un « effet plante hôte » fort qui se traduit par une ségrégation des populations de *S. meliloti* isolées de *M. truncatula* (ou de plante présentant des spécificités équivalentes, bv. *meliloti*) et celles isolées de *M. laciniata* (bv. *medicaginis*). De plus, il est notable que le patron de diversité obtenu tend à réduire la distance génétique entre les populations de *S. meliloti* bv. *meliloti* et *S. medicae* par rapport à celle observée entre les biovars *meliloti* et *medicaginis* au sein de *S. meliloti*. Ce patron coïncide avec la monophylie des gènes *nod* de *S. meliloti* bv. *meliloti* et *S. medicae* observée sur le super-arbre obtenu à partir des séquences *nodABC-nodEG* présentée dans le chapitre VII.1.

-au niveau des marqueurs *IGS_{RKP}*, *IGS_{GAB}* et *IGS_{EXO}*, on constate également qu'un effet plante structure les populations de *S. meliloti* bv. *meliloti* et *S. meliloti* bv. *medicaginis*. Parallèlement, les graphiques obtenus suggèrent un effet lié à la structure spatiale de l'échantillonnage. En effet, ces marqueurs ont tendance, par rapport à l'image consensus, à minimiser la distance génétique entre les populations tunisiennes de *S. meliloti* et à isoler la population française de *S. meliloti* bv. *meliloti*.

De manière à confirmer ce dernier point, une topologie traduisant les distances génétiques entre les populations bactériennes de *S. meliloti* isolées lors des échantillonnages E1, E2 et E3, caractérisées aux niveaux de sites polymorphes des marqueurs *IGS_{RKP}*, *IGS_{GAB}* et

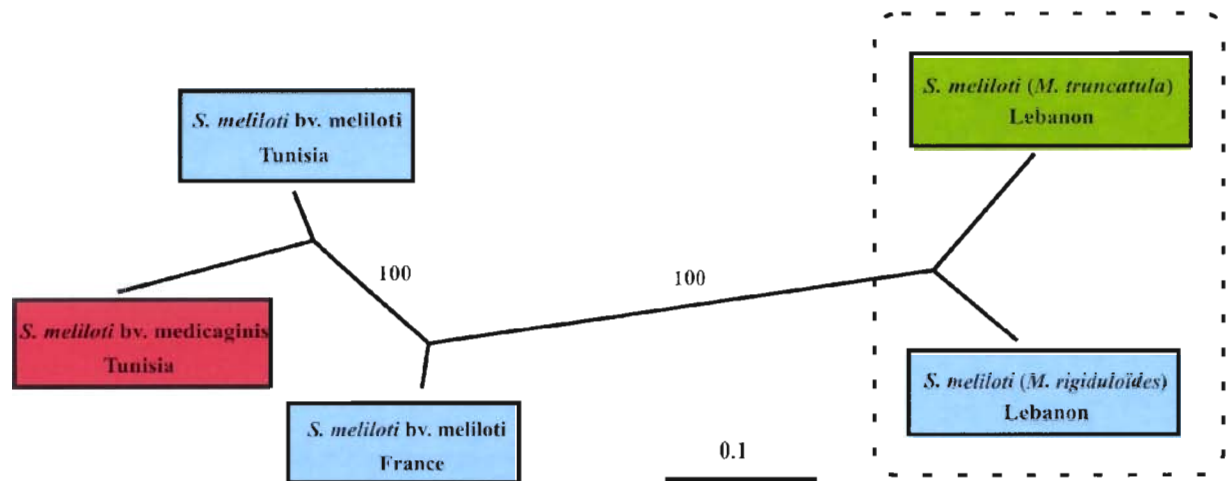


Figure 4. Arbre de Neighbor-Joining représentant la différenciation entre 5 populations de *Sinorhizobium meliloti* définies sur la base de l'origine géographique des souches et des plantes utilisées pour leur isolement. Les populations non différenciées ($p > 5\%$) sont entourées par une ligne pointillée. Les distances de Cavalli-Sforza utilisées pour obtenir l'arbre sont obtenues sur la base des fréquences des génotypes aux locus IGS_{RKP} , IGS_{EXO} et IGS_{GAB} inférés à l'aide de puces de génotypage { Article 4 }.

IGS_{EXO} {Article 4}, a été obtenue (Fig. 4). Cette topologie suggère qu'à l'échelle du bassin méditerranéen, la structure des populations de *S. meliloti* est plus sensible à l'effet d'isolement par la distance qu'à l'effet plante hôte. En effet, les populations de *S. meliloti* se regroupent en fonction de leur origine géographique, alors que les populations de *S. meliloti* bv. *meliloti* ne forment pas un groupe monophylétique.

VI.3. *Conclusions*

Conformément aux indices relevés dans la littérature, les données acquises montrent qu'il est essentiel de prendre en compte la composante spatiale pour mieux comprendre l'évolution des populations de *Sinorhizobium* spp. associées aux plantes du genre *Medicago*. A l'issue de cette analyse globale, il ressort d'ores et déjà que la distance géographique et le spectre d'hôte sont deux facteurs structurants primordiaux dont les effets se surimposent, et qui dépendent des locus. Dans ce contexte l'intérêt d'une étude phylogéographique visant à déterminer si la distribution spatiale de la variabilité génétique des souches de *S. meliloti* et *S. medicae* est liée aux aires de répartition des groupes de spécificité du genre *Medicago* sera discutée dans le chapitre consacré aux perspectives.

Enfin, l'incongruence des structures génétiques obtenues à partir des marqueurs étudiés suggère que les mécanismes de recombinaison jouent un rôle important dans la diversification des souches de *S. meliloti* et *S. medicae*.

VII. Impact des transferts horizontaux sur la diversité génétique des lignées de
Sinorhizobium sp.

VII. Impact des transferts horizontaux sur la diversité génétique des lignées de *Sinorhizobium* sp.

Suite aux analyses précédemment présentées, il nous est apparu nécessaire d'étudier l'influence des mécanismes de transferts horizontaux de matériel génétique sur la diversité des populations de *Sinorhizobium*.

VII.1. Transferts « à longue distance » taxonomique : l'origine des îlots symbiotiques *nod* de *S. meliloti* et *S. medicae* ? {Article 3}

Chez les bactéries, contrairement à la plupart des organismes eucaryotes, des transferts horizontaux peuvent se dérouler entre des organismes appartenant à des taxons phylogénétiquement éloignés. Ces transferts peuvent tout aussi bien se dérouler entre des bactéries appartenant à des espèces différentes qu'entre des bactéries qui n'appartiennent pas aux mêmes phylums. De tels transferts sont probablement impliqués dans la diversité de spectre d'hôte observée chez les souches de *Sinorhizobium*.

Au sein d'une reconstruction phylogénétique basée sur des séquences de gènes de ménage obtenues par l'intermédiaire de GeneBank tels que les gènes *recA* ou *atpD*, les souches des genres *Sinorhizobium* et *Rhizobium* forment respectivement des groupes monophylétiques (Fig. 5). Au contraire, le super-arbre obtenu à partir des séquences *nodABC-nodEG* de l'échantillonnage E4 {Article 3} supporte l'hypothèse que l'ensemble des souches capables de s'associer aux plantes du genre *Medicago* (i.e. les différents biovars de *S. meliloti*, *S. medicae* et *R. mongolense*) forme un groupe monophylétique (Fig. 6). Au sein de ce clade, la séquence de *R. mongolense* se positionne en groupe frère de l'ensemble des séquences de *Sinorhizobium* associées au genre *Medicago*. En outre, *M. ruthenica*, l'hôte spécifique de *R. mongolense*, se positionne comme la première espèce de *Medicago* à se différencier sur une phylogénie obtenue par Downie *et al.* (1998). L'ensemble de ces données suggère que le genre *Sinorhizobium* a acquis la capacité à interagir avec les plantes du genre *Medicago* suite au transfert de l'îlot symbiotique *nod* d'un ancêtre de *R. mongolense*.

Au sein du super-arbre *nodABC-nodEG*, les séquences des différents biovars de *S. meliloti* forment un groupe paraphylétique à cause de la position de la séquence de

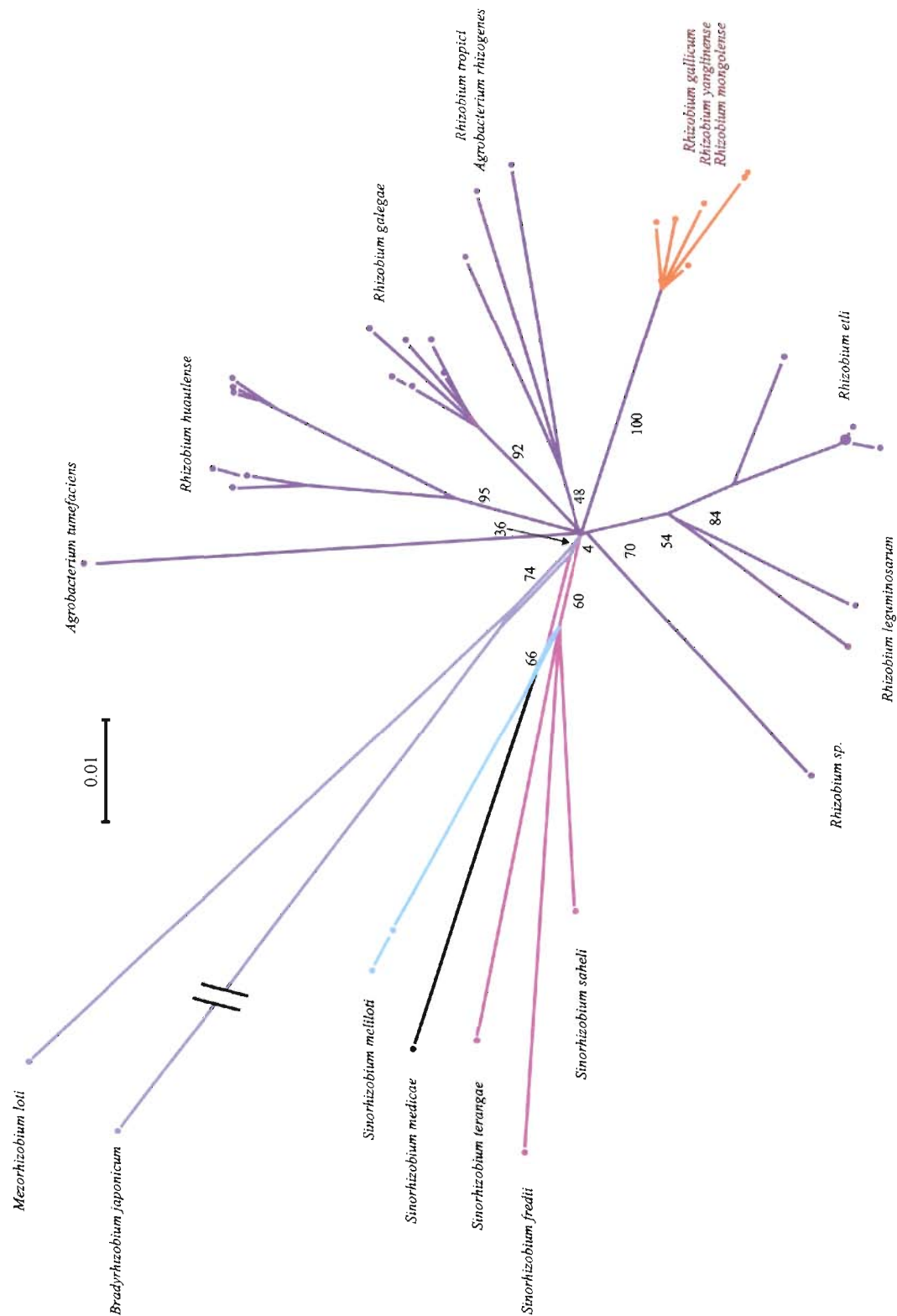


Figure 5. Réseau obtenu par Split décomposition (Bandelt and Dress 1992) sur la base d'un jeu de données obtenu en concaténant les séquences des gènes *recA* et *atpD*. Les robustesses des branches internes ont été obtenues par bootstrap.

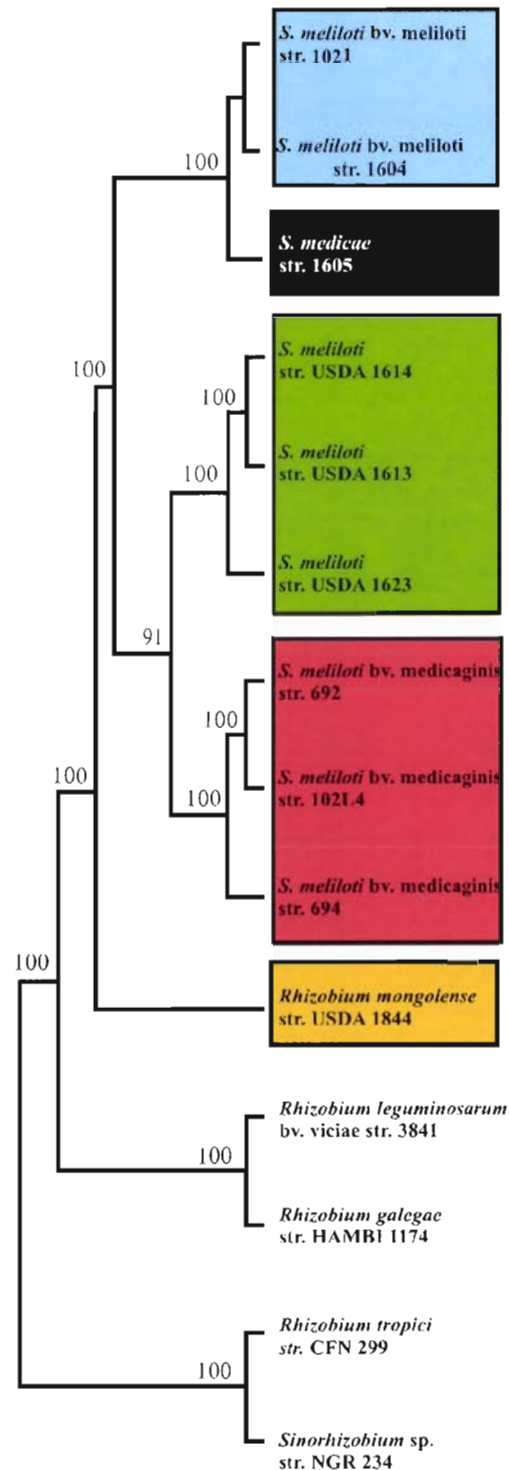


Figure 6. Super-arbre obtenu à partir des séquences partielles des groupes de gènes *nodABC* et *nodEFG* incluant les séquences des différents groupes bactériens capables d'interagir avec les plantes du genre *Medicago*. Plus précisément, *Sinorhizobium meliloti* USDA 1614, USDA 1613 et USDA 1623 interagissent efficacement avec des plantes telles que *Medicago rigiduloides*. Les souches de *S. meliloti* appartenant au biovars *meliloti* et *medicaginis* interagissent respectivement avec des plantes telles que *Medicago truncatula* et *Medicago laciniata*. *S. medicae* interagit notamment avec *M. truncatula*. Enfin, *Rhizobium mongolense* interagit avec *Medicago ruthenica*.

Encadré 2. Phylogénie et transferts horizontaux

Le but premier des reconstructions phylogénétiques est l'inférence des relations de parenté entre organismes en utilisant l'information apportée par la distribution des différents états des caractères (génétiques ou phénotypiques) polymorphes. Les relations phylétiques inférées sont alors généralement représentées par des arbres, qui représentent un scénario évolutif. Trois principales voies méthodologiques peuvent être utilisées pour obtenir ces arbres :

- les méthodes de parcimonie visent à trouver, parmi l'ensemble des arbres possibles, les scénarios qui minimisent le nombre d'hypothèses évolutives (mutation inférées) nécessaires pour faire coïncider l'évolution des organismes étudiés et l'évolution de l'ensemble des caractères d'intérêt.

- les méthodes de distance se déroulent en deux phases. Dans un premier temps, une matrice de distances entre paires de taxons est obtenue. La méthode de calcul de cette matrice peut être plus ou moins complexe en fonction des caractères choisis (caractères continus ou discrets) et des corrections que l'on souhaite affecter aux distances observées (pour des caractères nucléotidiques, on applique des corrections qui augmentent les distances observées sur la base d'un modèle d'évolution des séquences). On utilise ensuite un algorithme d'agglomération qui regroupe les taxons de proche en proche et ajuste les distances patristiques (celles de la matrice) afin qu'elle puissent être représentées dans un arbre.

- les méthodes probabilistes se basent sur l'estimation des probabilités associées aux différents événements mutationnels. Sur la base d'un modèle d'évolution des caractères étudiés (généralement des séquences nucléotidiques), le but est alors de calculer la probabilité des données pour chaque arbre afin de trouver l'arbre qui maximise la probabilité des données. D'une manière générale, l'augmentation exponentielle du nombre d'arbres envisageables, en fonction du nombre de taxons et l'importance des temps de calcul associés à la méthode impliquent une exploration heuristique de l'ensemble des scénarios évolutifs.

Lorsqu'un scénario évolutif est choisi, il est nécessaire d'obtenir une mesure de robustesse des relations phylétiques inférées. Ces mesures sont généralement obtenues par bootstrap sur le jeu de données, inférence d'une phylogénie pour chaque itération et calcul de la fréquence de chaque bipartition sur l'ensemble des arbres obtenus. Deux causes principales peuvent expliquer de faibles valeurs de bootstrap pour les bipartitions initialement inférées :

- La faible quantité d'information disponible pour inférer la phylogénie. Un déficit d'information peut être expliqué par deux mécanismes, une trop faible accumulation de mutations entre taxons ou une trop forte accumulation de mutations. Cette dernière cause est liée aux mutations récurrentes sur un même caractère qui font disparaître les synapomorphies et les symplesiomorphies, augmentent le nombre d'autapomorphies et peuvent être à l'origine d'homoplasies.

- L'existence de groupes de caractères possédant des histoires évolutives différentes (à force de vouloir l'évolution indépendante de chaque caractère, ça devait arriver...). C'est notamment le cas lorsque que l'on utilise des jeux de séquences nucléotidiques au sein desquels figurent des génotypes recombinants. En effet, l'utilisation d'un arbre pour retracer les relations phylétiques entre organismes ne permet la représentation que d'un seul scénario évolutif. Dans ce contexte, l'existence d'informations contradictoires dans le jeu de données peut se traduire par des biais dans le choix de la topologie optimale et la robustesse des bipartitions inférées. Trois voies méthodologiques peuvent être utilisées pour mettre en évidence la recombinaison dans le cadre

S. medicae. Deux scénarios alternatifs pourraient expliquer un tel patron : i) la spéciation de *S. medicae* s'est effectuée à partir d'une population ancestrale de *S. meliloti* caractérisée par son aptitude à interagir avec *M. truncatula* ; ii) la diversité actuelle des îlots *nod* de *S. meliloti* et *S. medicae* est liée à un transfert horizontal de matériel génétique entre les deux taxons. Même s'il ne nous a pas été possible de trancher entre les deux hypothèses nous reviendrons sur les problèmes rencontrés dans le chapitre VIII.1.2.

VII.2. Transferts entre les groupes de spécificité chez les symbiotes des *Medicago* : vers une définition évolutionniste des concepts d'espèce et de biovars ? {Article 3}

Si les transferts horizontaux entre des lignées divergentes de bactéries peuvent avoir des implications importantes sur la diversification de ces organismes (Ochman et al. 2000), les études menées sur les bases de données de séquences de génomes semblent indiquer que la majorité des échanges s'effectuent entre souches fortement apparentées (Ochman et al. 2005). Cette caractéristique devrait permettre la définition d'espèces bactériennes sur la base d'approches populationnelles chez les bactéries qui ne sont pas strictement clonales. La meilleure illustration de cette hypothèse se trouve probablement chez les symbiotes du genre *Medicago*. Une des premières études de déséquilibre de liaison entre souches de *rhizobium*, menée par Maynard Smith *et al.* (1993) sur la base d'un jeu de données produit par Eardly *et al.* (1990) sur des souches alors assignées à *Rhizobium meliloti*, a décrit une différenciation importante, associée à des mesures d'association multilocus (I_A) élevées, entre deux groupes de souches, alors que chacun des groupes présentait une structure quasi-panmictique (*i.e.* sans déséquilibre de liaison). Par la suite, deux espèces sœurs ont été définies sur la base de ces deux clades : *S. meliloti* et *S. medicae* (Rome et al. 1996).

Afin d'approfondir les résultats disponibles dans la littérature, nous avons cherché à savoir comment se structuraient les transferts horizontaux entre les biovars *meliloti* et *medicaginis* de l'espèce *S. meliloti*, comparativement à ce qui est observé entre *S. meliloti* et *S. medicae* {Article 3}. Cette étude se base principalement sur les données obtenues lors de la caractérisation de l'échantillonnage E2. Compte tenu du niveau de structuration au sein des populations de *S. meliloti* disponibles, nous avons opté pour une approche phylogénétique pour mettre en évidence d'éventuels transferts horizontaux de matériel génétique (Encadré 2).

d'approches phylogénétiques (complémentairement aux approches populationnelles présentées dans l'encadré 4) (voir Posada et Crandall (2001) et Posada (2002) pour des revues à ce sujet) :

-Si l'on connaît les groupes de caractères susceptibles de posséder des histoires évolutives différentes (par exemple deux gènes bactériens dont un aurait pu faire l'objet de transferts évolutifs), il est possible de mesurer l'adéquation entre les topologies obtenue et l'information évolutive contenue pour chaque lot de caractères. La mise en évidence d'incongruences peut être liée à des événements de recombinaison. Cette méthode a été utilisée dans l'article 3.

-Si l'on ne possède pas *d'a priori* sur les caractères sujets à recombinaison, il est possible de s'appuyer sur la distribution des polymorphismes afin de détecter des changements dans les patrons de distances observés entre des sous ensembles de caractères (par exemples les méthodes de fenêtres glissantes sur des séquences nucléotidiques) ou afin de comparer la distribution de différents types de mutations à un attendu théorique (par exemple la proportion d'homoplasie attendue). Cet type de méthode a été utilisée dans l'article 3.

-Enfin, il est envisageable de représenter les relations phylétiques sur l'ensemble des caractères par une structure plus complexe qu'un arbre, en l'occurrence un réseau. L'avantage des réseaux est qu'ils peuvent représenter plusieurs hypothèses évolutives mettant en jeu les mêmes taxons, en substituant certaines branches par des réticulations (voir les articles 2 et 3 pour une illustration). Toutefois, une quantité importante d'information est généralement perdue lors de la construction de ces réseaux du fait d'une utilisation de distances globales entre taxons, à cause de contraintes de constructions, ou du fait de la difficulté de mesurer la robustesse des différents liens évolutifs inférés.

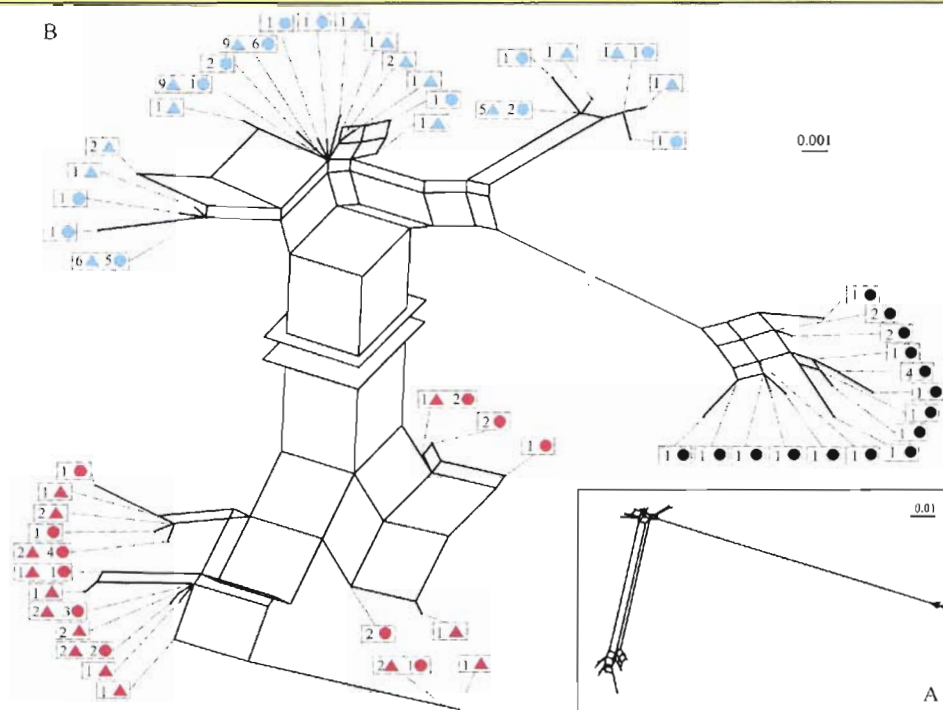


Figure 7. Présentation globale (A) et détaillée (B) du reticulogramme obtenu par l'algorithme Neighbor-Net sur la base des distances génétiques aux locus IGS_{RKP} , IGS_{EXO} , IGS_{NOD} et IGS_{GAB} entre les souches de *Sinorhizobium* isolées à partir des sols tunisiens {Article 3}. Les souches de *S. medicae*, *S. meliloti* bv *meliloti* et *S. meliloti* bv *medicaginis* sont respectivement représentées par des figurés noirs, bleus et rouges. Les génotypes isolés à partir du sol d'Enfidha sont figurés par des cercles alors que ceux isolés à partir du sol d'Hadjeb sont figurés par des carrés. Les numéros représentent le nombre de bactéries de l'échantillon possédant chaque génotype multilocus

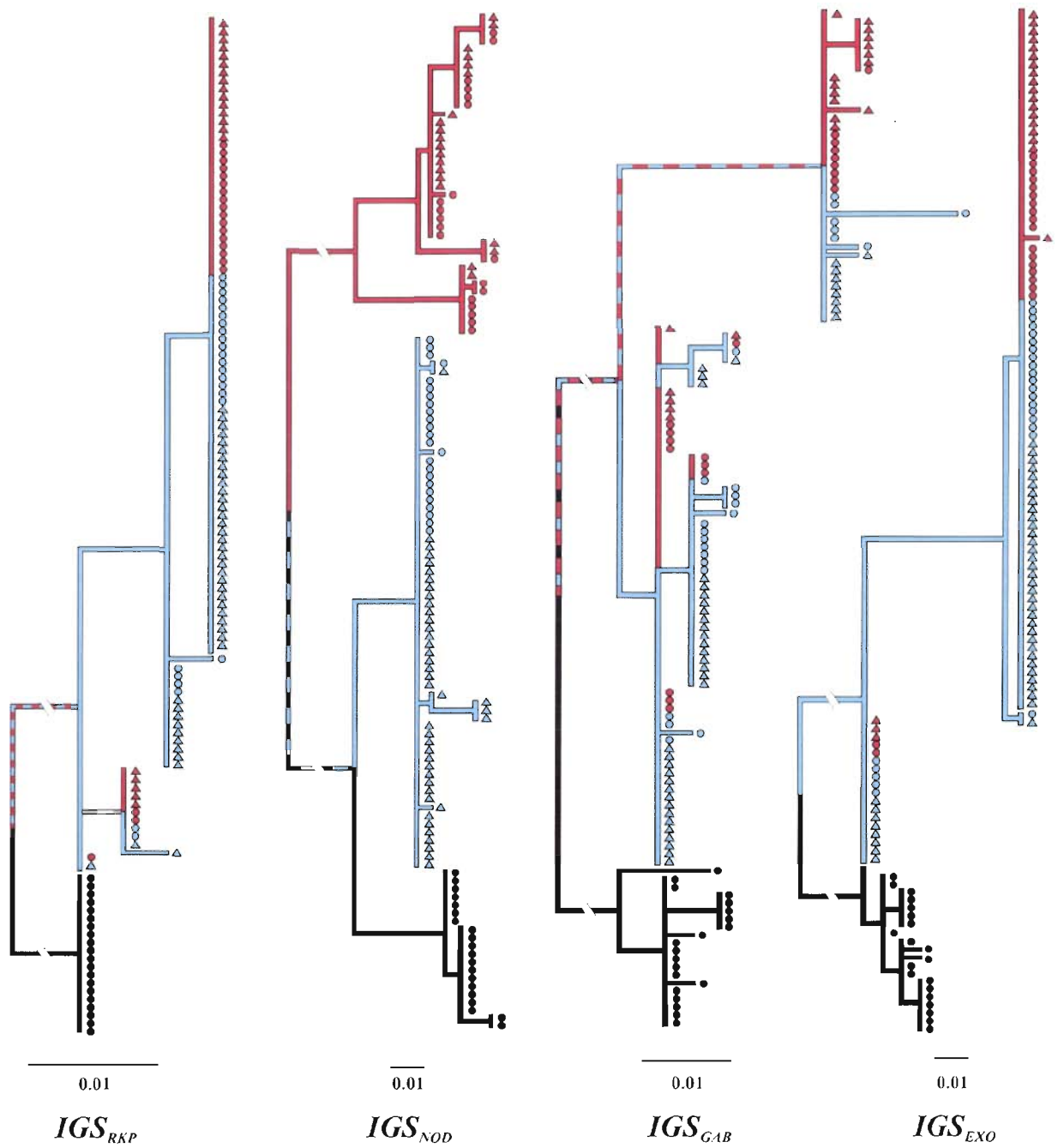


Figure 8. Phylogénies obtenues par maximum de vraisemblance pour chacun des quatre marqueurs utilisés pour caractériser les souches isolées à partir des sols tunisiens { Article 3}. Les souches de *S. medicae*, *S. meliloti* bv *meliloti* et *S. meliloti* bv *medicaginis* sont respectivement représentées par des figurés noirs, bleus et rouges. L'évolution des capacités symbiotiques bactériennes est inférée par une méthode de parcimonie et illustrée sur les phylogénies par les changements de nuance de gris. Les génotypes isolés à partir du sol d'Enfidha sont figurés par des cercles alors que ceux isolés à partir du sol d'Hadjeb sont figurés par des carrés.

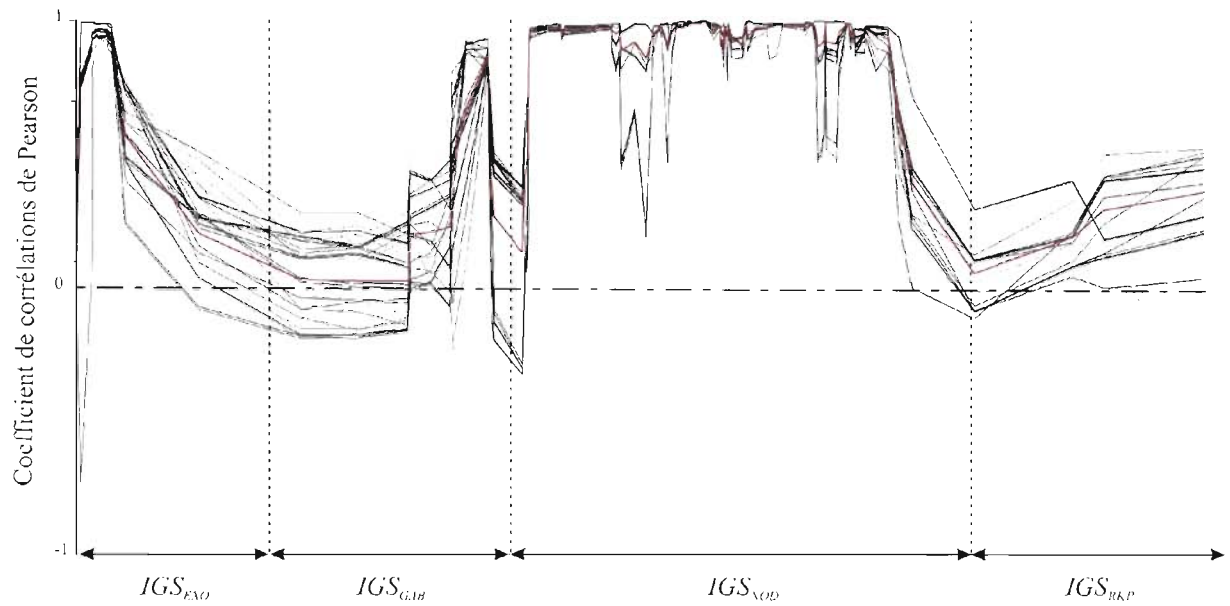


Figure 9. Profil obtenu à l'aide du logiciel PHYLPRO (Weiller 1998) illustrant les points de recombinaison le long d'un jeu de données obtenu en concaténant les séquences de souches de *S. meliloti* acquises lors de la caractérisation de l'échantillonnage E2. Les points de recombinaison les plus flagrants (*i.e.* les points où les mesures de corrélation sont les plus faibles) se trouvent à la marge des séquences de chaque locus. Ce patron indique que les événements de recombinaison inférés se déroulent entre les régions étudiées.

VII. Transferts horizontaux et diversité génétique de *S. meliloti* et *S. medicae*

Dans un premier temps, un réseau a été obtenu à partir des données de séquence aux marqueurs *IGS_{RKP}*, *IGS_{NOD}*, *IGS_{GAB}* et *IGS_{EXO}* (Fig. 7). Cette représentation montre que le groupe incluant l'ensemble des souches de *S. meliloti* et celui incluant les souches de *S. medicae* sont reliés par une branche unique. Ceci suggère qu'il existe un isolement sexuel entre les deux groupes génétiques. Au contraire, les groupes incluant respectivement les souches de *S. meliloti* bv. *meliloti* et les souches de *S. meliloti* bv. *medicaginis* sont reliés par des réticulations. Ce patron suggère qu'il existe une certaine ambiguïté dans le signal phylogénétique séparant les deux biovars, probablement liée à des transferts horizontaux de matériel génétique entre ces taxons.

Dans un second temps, l'analyse des phylogénies obtenues pour chacun des quatre marqueurs confirme les hypothèses posées (Fig. 8). Dans toutes les phylogénies, les séquences de l'espèce *S. medicae* sont groupées dans un clade, soulignant l'isolement reproducteur entre *S. meliloti* et *S. medicae*. Au contraire, les séquences des deux biovars de *S. meliloti* forment des groupes paraphylétiques sur l'ensemble des phylogénies, à l'exception de la topologie obtenue pour le marqueur *IGS_{NOD}*. Pour ce marqueur, les séquences des souches des biovars *meliloti* et *medicaginis* forment deux clades indépendants. L'arbre *IGS_{NOD}* est donc compatible avec l'image évolutive obtenue dans le cadre de la phylogénie *nodABC-nodEG*. Il est également notable que l'évolution des propriétés symbiotiques des souches semble parfaitement corrélée à la phylogénie *IGS_{NOD}*. Enfin, pour chaque jeu de données, la comparaison par test de Shimodaira-Hasegawa (Shimodaira and Hasegawa 1999) des vraisemblances associées aux topologies obtenues avec chacun des marqueurs a systématiquement abouti à la mise en évidence d'incongruences significatives ($p < 5\%$). Ceci suggère que les différences entre les topologies inférées à l'aide des quatre marqueurs, notamment les différences de positionnement des souches appartenant aux deux biovars de *S. meliloti*, sont dues à des événements de transferts horizontaux de matériel génétique.

Enfin, un jeu de séquences, obtenu en concaténant les données *IGS_{RKP}*, *IGS_{NOD}*, *IGS_{GAB}* et *IGS_{EXO}* pour les souches de *S. meliloti* de l'échantillonnage E2, a été analysé par différentes méthodes visant à inférer des événements de recombinaison en étudiant les différences de patrons de polymorphisme le long de fragments d'ADN homologues. Les résultats obtenus montrent : i) que les événements de recombinaison inférés se déroulent principalement entre les locus étudiés (Fig. 9) ; ii) que des événements de recombinaison ont lieu entre les souches de *S. meliloti* bv. *meliloti* et celles de *S. meliloti* bv. *medicaginis*. Cette

Encadré 3. Structure des populations bactériennes *sensu lato*

L'haploïdie du génome des bactéries fait qu'un grand nombre de termes et de concepts, communs dans le cadre d'études de génétique des populations des organismes diploïdes, revêtent un sens différent ou ne sont pas applicables aux populations de ces microorganismes. On peut d'ailleurs penser que ces différences ont constitué un frein au développement des approches populationnelles chez les bactéries (avec l'absence d'exemples illustrant l'influence du sexe sur l'évolution des populations bactériennes avant les années 1990, etc...).

Par exemple, dans les études se focalisant sur les eucaryotes, le terme panmictique décrit un régime de reproduction sous lequel les gamètes se rencontrent aléatoirement. Le régime de reproduction panmictique se caractérise dans ce contexte par des fréquences génotypiques décrites par la structure d'Hardy-Weinberg, par l'absence de différence entre hétérozygotie observée et l'hétérozygotie attendue (en espérance) et donc par un F_{IS} dont l'espérance est nulle.

Au sein de populations bactériennes, le calcul de la statistique F_{IS} n'est pas possible du fait de l'haploïdie.

Toutefois, si le calcul de l'hétérozygotie attendue à l'équilibre d'Hardy-Weinberg n'a pas plus de sens *sensu stricto*, la formule utilisée, revêt un intérêt particulier dans la mesure où elle fréquemment utilisée pour mesurer

le niveau de diversité des populations bactériennes. En effet, $H_e = 1 - \sum_{i=1}^t p_i^2$, où p_i est la fréquence de l'allèle i

parmi les t allèles observés pour le marqueur étudié, correspond à la formule utilisée pour obtenir l'indice de

diversité de Gini-Sympson (ou indice de Nei). Enfin, si le terme panmixie *sensu stricto* n'est pas compatible

avec le cycle de vie des procaryotes, il a été utilisé par Maynard Smith *et al.* (1993) pour définir la structure génétique des populations bactériennes en fonction du patron de déséquilibre de liaison observé. Dans ce cadre,

trois cas ont été différenciés : (i) une structure de population panmictique décrit une population bactérienne où

l'ensemble des marqueurs étudiés sont en équilibre de liaison. Le cas le plus utilisé pour illustrer ce type de

structure est celui d'*Helicobacter pylori* (Suerbaum *et al.* 1998). (ii) Une structure épidémique décrit une

population au sein de laquelle certaines souches connaissent un succès évolutif entraînant la détection d'un

déséquilibre de liaison au niveau de l'ensemble des individus (lié à la surreprésentation de certains génotypes) et

un équilibre de liaison lorsque l'on utilise un représentant par génotype multilocus pour tester l'association.

Cette structure est par exemple rencontrée au sein des populations du pathogène *Pseudomonas aeruginosa*

(Pirnay *et al.* 2002). (iii) Enfin, les populations présentant un déséquilibre de liaison aux deux échelles sont

appelées clonale. Cette structure est par exemple rencontrée dans les populations de *Mycobacterium bovis* (Smith

et al. 2003).

dernière hypothèse est cohérente avec l'image évolutive illustrée par l'arbre représentant les distances génétiques entre différentes populations de *S. meliloti* présenté dans le chapitre VI.2 (Fig. 4).

VII.3. Transferts au sein de deux populations de symbiotes des *Medicago* : existe-t-il un lien entre architecture génomique et patron de recombinaison ? {Article 2}

VII.3.1. Etat de l'art sur le déséquilibre de liaison chez *S. meliloti* et *S. medicae*

Outre la mise en évidence des espèces *S. meliloti* et *S. medicae*, l'étude menée par Maynard Smith *et al.* (Maynard Smith *et al.* 1993) sur la base du jeu de données produit par Eardly *et al.* (Eardly *et al.* 1990) a fourni les premières informations concernant l'impact des mécanismes de recombinaison sur la diversité des souches de ces deux taxons. Toutefois, une nouvelle analyse de ce jeu de données semblait nécessaire compte tenu des améliorations méthodologiques portant sur les tests de déséquilibre de liaison basés sur les mesures d' I_A . Contrairement à ce qui avait été observé pour d'autres jeux de données, les résultats des tests de déséquilibre de liaison réalisés à l'aide de la formule préconisée par Haubold *et al.* (1998) à partir des génotypes de *Sinorhizobium* publiés par Eardly *et al.* (1990) sont en accord avec ceux obtenus par Maynard Smith *et al.* (1993). Dans le cas des souches de *S. meliloti*, une structure épidémique est inférée par les deux méthodes. (voir l'encadré 3 sur la définition des structures de populations bactériennes). Sachant que treize des quatorze enzymes utilisées dans l'étude de Eardly *et al.* (1990) sont codées par des locus chromosomiques dans le génome de *S. meliloti* souche 1021, cela signifie qu'il existe un certain déséquilibre de liaison au niveau de cette unité de réplication chez cette espèce. Dans le cas des souches de *S. medicae*, Maynard Smith *et al.* (1993) ont décrit une structure de population panmictique. Toutefois, sur quatorze marqueurs utilisés (Eardly *et al.* 1990) pour calculer l'indice d'association, seuls trois présentent des patrons de polymorphisme incompatibles avec l'hypothèse de déséquilibre de liaison, alors que trois locus ne présentent pas de signe de recombinaison et que huit locus sont monomorphes ou non informatifs (*i.e.* le polymorphisme est lié à la présence de singletons). Enfin, on trouve parmi les trois premiers marqueurs une enzyme codée par un gène du plasmide pSymB sur le génome de *S. meliloti* souche 1021. Ceci laissait supposer que le niveau de déséquilibre de liaison au sein du chromosome de *S. medicae* n'a pas été parfaitement déterminé.

Encadré 4. Déséquilibre de liaison

Dans le cadre d'études sur l'évolution des populations bactériennes, déterminer le régime de reproduction représente une étape importante dans la compréhension des processus mis en jeu. Ainsi, différentes méthodologies ont été utilisées pour mettre en évidence le rôle de la recombinaison sur les patrons de diversité bactériens. D'une manière générale, deux types d'approches sont utilisées :

-des méthodes multilocus, mesurant l'association statistique entre les génotypes au niveau de plusieurs marqueurs. L'indice de ce type le plus couramment utilisé est l'indice d'association I_A développé initialement par Brown *et al.* (1980) et utilisé dans le cadre d'une étude du régime de reproduction de différentes populations bactériennes par Maynard Smith *et al.* (1993). Cet indice est basé sur la comparaison des variances de la distance génétique entre individus observée V_O et attendue sous l'hypothèse d'équilibre de liaison V_E ($I_A = (V_O/V_E) - 1$). Il existe deux manières de tester la significativité d'une valeur d' I_A : (i) sur la base de permutations des génotypes à chaque locus ; (ii) en obtenant analytiquement l'intervalle de confiance à 95% de V_E . Dans ce cadre, Haubold *et al.* (1998) ont publié la formule exacte permettant d'obtenir la variance de V_E ($\text{Var}(V_E)$), montrant à cette occasion le manque de robustesse du test effectué avec l'approximation de $\text{Var}(V_E)$ anciennement utilisée. Dans le cas de nos travaux, le calcul de mesures de déséquilibre de liaison multilocus est handicapant dans la mesure où il n'est pas possible de mettre en évidence des différences de taux de recombinaison entre marqueurs (*i.e.* une mesure d'association est obtenue pour l'ensemble des marqueurs, « moyennant » les différences potentielles).

-des méthodes mesurant l'association statistique entre génotypes par paires de locus. D'une manière générale, ces mesures sont toutes apparentées au χ^2 de Pearson qui vise à tester l'indépendance de deux variables dans un tableau de fréquence à double entrée. Parmi l'ensemble de mesures disponibles (voir Hedrick (1987) pour une revue), trois sont particulièrement utilisées. Soit une paire de locus diallélique A/B , pour chaque paire d'allèles A_i/B_j , on peut calculer $D_{ij} = P_{A_i/B_j} - P_{A_i} \times P_{B_j}$.

D'autre part, on peut définir :

$$D'_{ij} = D_{ij} / D_{\max} \text{ où } D_{\max} = \min \left[P_{A_i} \times P_{B_j} ; (1 - P_{A_i}) \times (1 - P_{B_j}) \right] \text{ si } D_{ij} \leq 0 \text{ ou}$$

$$D_{\max} = \min \left[P_{A_i} \times (1 - P_{B_j}) ; P_{B_j} \times (1 - P_{A_i}) \right] \text{ si } D_{ij} \geq 0.$$

Enfin, on peut utiliser la statistique : $r^2 = D_{ij}^2 / \left(P_{A_i} \times (1 - P_{B_j}) \times P_{B_j} \times (1 - P_{A_i}) \right)$.

Ces trois statistiques présentent différentes propriétés (Hedrick 1987) : a) D'_{ij} et r^2 sont des mesures normées (*i.e.* D'_{ij} est compris entre -1 et 1, r^2 entre 0 et 1) contrairement à D_{ij} ; b) D_{ij} et r^2 sont plus sensibles aux fréquences alléliques que D'_{ij} ; c) il est facile de calculer un niveau de déséquilibre de liaison moyen entre deux locus possédant plus de deux allèles sur la base des D_{ij} et D'_{ij} ; voir le schéma ci-dessous pour une illustration de certaines propriétés des trois statistiques présentées. Sur la base de ces propriétés, nous avons décidé d'utiliser D_{ij} et D'_{ij} pour mesurer le niveau de déséquilibre de liaison entre nos marqueurs {Article 2}. Dans ce contexte, l'hypothèse d'équilibre de liaison entre paires de locus peut être testée sur la base de la distribution des valeurs

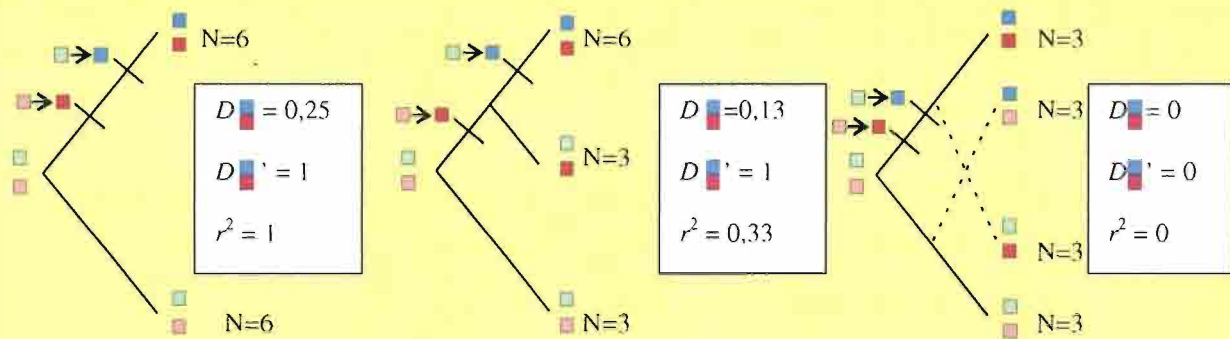
Plus récemment, une étude publiée par Roumiantseva *et al.* (2002) a décrit un déséquilibre de liaison significatif entre la majorité des marqueurs utilisés pour caractériser des populations de *S. meliloti* isolées de nodosités de différentes espèces de *Medicago* à partir de sols du Caucase. Toutefois, les auteurs suggèrent que le déséquilibre de liaison observé dans leur étude est lié à l'échantillonnage de plusieurs populations génétiquement isolées. En effet, nous avons indiqué dans le chapitre VI.2 que le jeu de données utilisé était composé de trois populations différenciées.

VII.3.2. Inférence de transferts au sein d'une population de *S. meliloti* bv. *meliloti* et de *S. medicae* {Article 2}

Les différents problèmes soulevés par les études de Maynard Smith *et al.* (1993) et Roumiantseva *et al.* (2002) ainsi que les résultats présentés dans le chapitre précédent nous ont amené à mesurer le niveau de déséquilibre de liaison entre les sept marqueurs utilisés pour caractériser les souches de *S. meliloti* bv. *meliloti* et de *S. medicae* isolées lors de l'échantillonnage E1. Une approche populationnelle (encadré 4) a été utilisée car aucune sous-structure n'a été détectée que se soit au sein du groupe de souches de *S. meliloti* bv. *meliloti* ou au sein du groupe de souches de *S. medicae*, même si ce point ne constitue pas une condition *sine qua non* puisqu'il est possible d'utiliser des mesures de déséquilibre de liaison intra et inter-populationnelle (Ohta 1982; Tachida and Cockerham 1986).

Les mesures d'association calculées (*i.e.* D et D') et les tests de déséquilibre de liaison effectués indiquent que le taux de recombinaison dépend de la localisation des marqueurs, au moins sur le génome de *S. meliloti* bv. *meliloti* (Fig. 10). Au sein de cette espèce, les marqueurs chromosomiques sont en effet en fort déséquilibre de liaison, ce qui semble en accord avec les données de Maynard Smith *et al.* (1993). Toutefois, nous n'avons pas testé l'existence d'un déséquilibre de liaison sur notre jeu de données réduit à un individu par génotype multilocus à cause du manque de puissance de cette approche lors de comparaisons par paires de locus. Au contraire, les calculs de déséquilibre de liaison mettant en jeu des locus plasmidiques, à la fois les comparaisons entre marqueurs plasmidiques et entre marqueurs plasmidiques et chromosomiques, indiquent que ces derniers sont sujets à des transferts horizontaux suffisamment fréquents pour casser le déséquilibre de liaison lié au mode de reproduction clonal de ces bactéries. Parallèlement, une étude récente indique que le nombre d'événements d'insertion/délétion de gènes entre souches de *S. meliloti* bv. *meliloti*

de déséquilibre de liaison obtenues par permutations des génotypes à chaque locus ou par un test dérivé du test exact de Fisher.



Valeurs prises par D , D' et r^2 pour différents jeux de fréquences alléliques à deux locus bialléliques (rouge/rose et bleu clair/foncé) ; Adapté de Rafalski (2002). D' et r^2 sont des mesures normées et seul D' est peu sensible aux différences de fréquences alléliques entre locus. N : effectif.

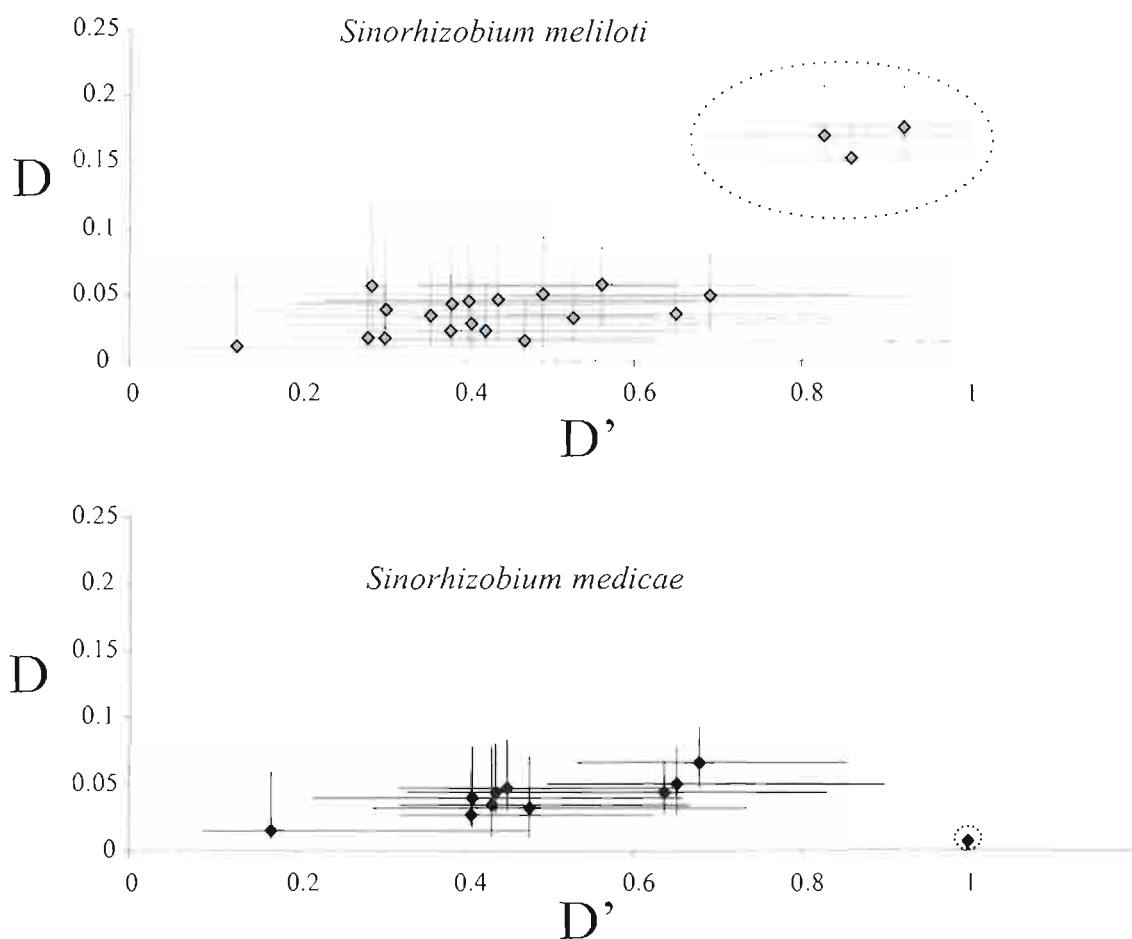


Figure 10. Distribution des mesures d'association D et D' pour l'ensemble des paires de locus chez *Sinorhizobium meliloti* et *Sinorhizobium medicae*. Les barres d'erreur définissent les intervalles de confiance à 95% des mesures. Les comparaisons entre paires de locus chromosomiques sont entourées par une ligne pointillée. Les résultats obtenus lors de telles comparaisons chez *S. medicae* ne sont pas précis du fait d'un déficit de polymorphisme.

est plus important sur pSymA que sur le reste du génome (Giuntini et al. 2005). Ceci semble indiquer que pSymA constitue un « point chaud » d'insertion de matériel génétique par recombinaison homologue ou par transfert horizontal. Toutefois, il est difficile de savoir s'il existe un lien direct entre ces deux observations (*i.e.* recombinaison et insertion/délétion).

Les marqueurs chromosomiques utilisés au sein de la population de *S. medicae* ne se sont pas révélés assez polymorphes pour apporter des informations complémentaires à celles obtenues par Maynard Smith *et al.* (1993). Par contre, les mesures de déséquilibre de liaison calculées sur la base des marqueurs plasmidiques chez *S. medicae* sont comparables (*i.e.* les valeurs sont du même ordre de grandeur) à celles observées chez *S. meliloti*. Ceci laisse supposer que les marqueurs plasmidiques évoluent de manière indépendante chez *S. medicae*.

De précédentes études ont montré que les mégaplasmines de *S. meliloti* comportaient différents déterminants génétiques impliqués dans les processus de conjugaison, dont plusieurs locus *tra* (Galibert et al. 2001). De plus, plusieurs origines de transfert impliquées dans le mécanisme de conjugaison (*oriT* and *mob*) ont également été identifiées sur les mégaplasmines pSymA et pSymB. Notamment, un de ces éléments a été localisé à proximité de l'îlot symbiotique de pSymA (Herrera-Cervera et al. 1998). Ces éléments pourraient participer à un système de conjugaison qui permette des transferts horizontaux de matériel génétique intra-spécifique. Malgré tout, le mécanisme qui contrôle ce type de transfert devrait être clarifié. Jusqu'à maintenant, peu d'études ont observé *ex-situ* des transferts horizontaux de mégaplasmine (Pretorius-Guth et al. 1990; Perez-Mendoza et al. 2005) et aucune étude n'a mis en évidence expérimentalement les déterminants génétiques impliqués dans les transferts, partiel ou complet, de pSymA et pSymB. De plus, ces unités de réplication semblent être de taille trop importante pour être transmises intégralement par conjugaison. Si des approches mécanistiques sont indispensables afin de mieux comprendre comment s'effectuent les transferts inférés, des informations plus précises sur leur nature et sur leur impact sur la diversité de *S. meliloti* et *S. medicae* pourraient être obtenues par des approches évolutives complémentaires. Certaines d'entre elles seront discutées dans la partie dédiée aux perspectives.

VII.3.3. Congruence entre architecture des génomes et patrons de recombinaisons : la modularité représente-elle un avantage adaptatif ?

Encadré 5. Evolution de l'architecture des génomes bactériens

La multiplication des études d'électrophorèse en champ pulsé et des projets de séquençage totaux de génomes a révélé que de nombreuses espèces bactériennes présentaient des organisations génomiques particulières. D'une part, il existe une diversité importante dans la taille des génomes bactériens. D'autre part, le nombre et la nature des unités de réplication qui composent ces génomes sont eux aussi variables (Casjens 1998).

Plusieurs auteurs ont suggéré que la taille des génomes était liée au mode de vie des bactéries, indépendamment de la nature de l'interaction à laquelle ils participent, ou de leur position taxonomique (Moran and Wernegreen 2000; Lio 2002; Konstantinidis and Tiedje 2004). Certains symbiotes obligatoires à transmission (majoritairement) verticale présentent des tailles de génome proches d'1 Mb. Notamment, les parasites du genre *Mycoplasma* (Firmicute ; 0,6-1 Mb), les bactéries du genre *Wolbachia*, qu'elles soient parasites ou mutualistes, (α -Proteobacteria ; 1-1,3 Mb), ou les bactéries du genre *Buchnera* (γ -Proteobacteria ; 0,6 Mb) illustrent cette tendance évolutive. Au contraire, les bactéries les plus versatiles, notamment les symbiotes à transmission horizontale, sont caractérisés par des tailles de génomes importantes. Ainsi, les génomes des parasites du complexe *Burkholderia cepacia* (β -Proteobacteria ; 8-8,7 Mb), ou de différents *rhizobium* tels que *Bradyrhizobium japonicum* (α -Proteobacteria ; 9,1 Mb) font partie des plus grands génomes bactériens séquencés. La variation de la taille du génome d'une espèce serait liée à deux contraintes évolutives : (i) la nécessité de posséder et de réguler un nombre variable de voies métaboliques en fonction des habitats rencontrés ; (ii) des processus populationnels (impact de la dérive génétique sur les symbiotes à transmission verticale) liés au mode de transmission des symbiotes. Notamment, les symbiotes à transmission verticale devraient être particulièrement sensibles au cliquet de Muller du fait d'un effectif efficace limité lors des transmissions, et de possibilités réduites de recombinaison.

Les variations dans le nombre et la nature (chromosome/megaplasmide/plasmide ; circulaire/linéaire) des unités de réplication posent différentes questions. Les premières sont liées à l'origine de telles structures. Les hypothèses avancées pour expliquer l'origine d'une telle variabilité sont l'acquisition horizontale d'unités de réplication surnuméraires et la segmentation d'unités de réplication préexistantes. Le nombre de génomes séquencés augmentant, les approches phylogénétiques visant à comprendre l'évolution de l'organisation des génomes permettront sûrement d'obtenir des informations pertinentes à ce sujet. Les secondes sont liées à la stabilité de telles structures. En effet, différentes approches ont permis de souligner l'extrême plasticité des génomes bactériens à l'échelle intra-spécifique. Sur la base des données acquises à ce jour, on constate qu'en moyenne 10% des gènes d'une souche bactérienne intégralement séquencée ne sont pas présents chez les individus conspécifiques. Par ailleurs différentes expériences (notamment chez les *rhizobium Sinorhizobium* NGR 234 et *S. meliloti*) ont montré que les mégaplasmides pouvaient s'intégrer au sein du chromosome. Dans ces conditions, trois hypothèses peuvent être avancées pour expliquer le maintien de structures de génomes complexes : (i) une diminution du temps de réplication pour les génomes multipartites, permettant un temps de génération plus court, pourrait constituer un avantage sélectif. (ii) De plus, les génomes multipartites pourraient bénéficier des avantages procurés par leur modularité à savoir une meilleure allocation des ressources aux différentes voies métaboliques et un potentiel évolutif plus important.

Le patron de déséquilibre de liaison observé pose des questions sur la nature adaptative de l'architecture des génomes bactériens. En effet, il existe une variabilité non négligeable dans le nombre et la nature (chromosome ou megaplasmide, linéaire ou circulaire) des unités de réplication entre les génomes des espèces bactériennes (Encadré 5 et Ochman (2002)) alors que la variabilité intraspécifique semble assez limitée. Pourtant, des études d'évolution expérimentale ont montré que ces structures étaient plastiques chez certaines espèces de *rhizobium* dont *S. meliloti* (Flores et al. 2000; Guo et al. 2003). Le génome de *S. meliloti*, et probablement celui de *S. medicae* (le séquençage du génome de la souche WSM419 de *S. medicae* pourra apporter certaines réponses à ce sujet), se compose de groupes de gènes impliqués dans les fonctions du métabolisme de base, dans l'interaction symbiotique ou dans des voies métaboliques secondaires. Ces groupes de gènes définissent des modules qui sont localisés sur différentes unités de réplication (Galibert et al. 2001), dont le fonctionnement est relativement intégré (*i.e.* à l'image d'un « super-opéron »). Jusqu'alors, deux raisons principales ont été avancées pour expliquer l'émergence de la modularité (voir (McAdams et al. 2004) pour une revue à ce sujet). D'une part, l'organisation modulaire des génomes, *i.e.* l'intégration des voies de régulations des gènes impliqués dans les mêmes fonctions, peut augmenter la valeur sélective des bactéries en facilitant l'allocation des ressources aux différentes fonctions codées par chaque module, en régulant leur expression. L'accumulation de gènes de régulation dans les grands génomes bactériens comme ceux des *rhizobium* (Konstantinidis and Tiedje 2004), est en accord avec cette hypothèse. D'autre part, les transferts horizontaux permettent l'évolution indépendante de chaque module, facilitant l'action de pressions de sélection spécifiques sur différentes parts du génome. Les patrons de déséquilibre de liaison et de polymorphisme observés dans l'article 2 sont en accord avec cette hypothèse. Ce type de patron pourrait se retrouver chez d'autres bactéries, notamment au sein du genre *Agrobacterium* où des données semblent indiquer une localisation non-aléatoire des marqueurs polymorphes dont la diversité semble être structurée par la spécificité d'hôte (Xavier Nesme, communication personnelle).

VII.4. Conclusions

Les transferts horizontaux ont un impact important sur la diversité des souches de *Sinorhizobium* tant à l'échelle macroévolutive qu'à l'échelle microévolutive. D'une part, les analyses effectuées semblent indiquer que les bactéries du genre *Sinorhizobium* ont acquis la capacité à interagir avec les plantes du genre *Medicago* suite à un transfert d'îlot symbiotique

provenant d'un ancêtre de *R. mongolense*. D'autre part, l'étude des patrons de transfert de matériel génétique chez *S. meliloti* bv. *meliloti*, *S. meliloti* bv. *medicaginis* et *S. medicae* nous a permis de distinguer les statuts d'espèces et de biovars sur la base d'approche évolutive. Par ailleurs, ceci suggère que les processus de spécialisation et de spéciation ne sont pas indissociables chez les bactéries du genre *Sinorhizobium*. Enfin, l'étude des patrons de recombinaison au sein d'une population de *S. meliloti* bv. *meliloti* nous a permis de mettre en évidence des différences de taux de recombinaison entre les différentes unités de réplication du génome de *S. meliloti*. Ce résultat est particulièrement intéressant dans la mesure où le génome de *S. meliloti* est structuré en trois réplicons possédant des caractéristiques fonctionnelles différentes. Dans ce contexte, il est intéressant d'étudier la structure des populations de *Sinorhizobium* au niveau de marqueurs situés sur chacune de ces unités de réplication.

VIII. Evolution de la spécificité et diversification des *Sinorhizobium* spp.

VIII. Evolution de la spécificité et diversification des *Sinorhizobium* spp.

Compte tenu de la différence de patron de structuration entre les locus utilisés dans le chapitre VI.2, les événements de recombinaison présentés dans le chapitre VII.2 semblent avoir un impact important sur la diversité génétique des populations de *Sinorhizobium* associées aux plantes du genre *Medicago*. Par conséquent, nous avons essayé d'aborder les questions relatives à l'influence de la diversité des plantes du genre *Medicago* sur la diversité des populations symbiotiques qui leur sont associées au niveau de locus impliqués dans la spécificité d'hôte et au niveau de locus de ménage. Nous avons dans un premier temps focalisé notre attention sur les processus agissant au niveau de populations bactériennes interagissant avec des plantes hôtes présentant des spécificités différentes puis sur les processus agissant au sein de populations bactériennes interagissant avec des plantes présentant des spécificités comparables.

VIII.1. Evolution vers différents groupes de spécificité {Articles 3 et 4}

Sur la base de l'échantillonnage E2 {Article 3}, nous avons étudié l'architecture de la différenciation génétique entre les souches de *S. meliloti* bv. *meliloti*, *S. meliloti* bv. *medicaginis* et *S. medicae*. Le but de cette approche était d'observer comment les différences de spécificité symbiotique entre ces trois groupes bactériens se traduisaient en terme de différenciation génétique au niveau de plusieurs locus répartis sur le chromosome, pSymA et pSymB. Les structures génétiques observées ont servi de base à une réflexion portant sur les processus qui influencent l'évolution des *Sinorhizobium* associés au genre *Medicago*.

VIII.1.1. Origine et évolution de la variabilité inter-biovars chez *S. meliloti*

Les analyses multilocus portant sur la variabilité génétique de l'échantillonnage E2 (Fig. 3, Fig. 7) pourraient sembler en accord avec la définition écologique des espèces bactériennes préconisée par Cohan, où chaque écotipe représente une espèce (2002). Sur la base de ces analyses, *S. meliloti* bv. *meliloti* et *S. meliloti* bv. *medicaginis* semblent en effet former des groupes génétiques discriminés. Toutefois, nous avons montré que les deux biovars de *S. meliloti* n'étaient pas affectés par un isolement sexuel. Ceci suggère qu'une

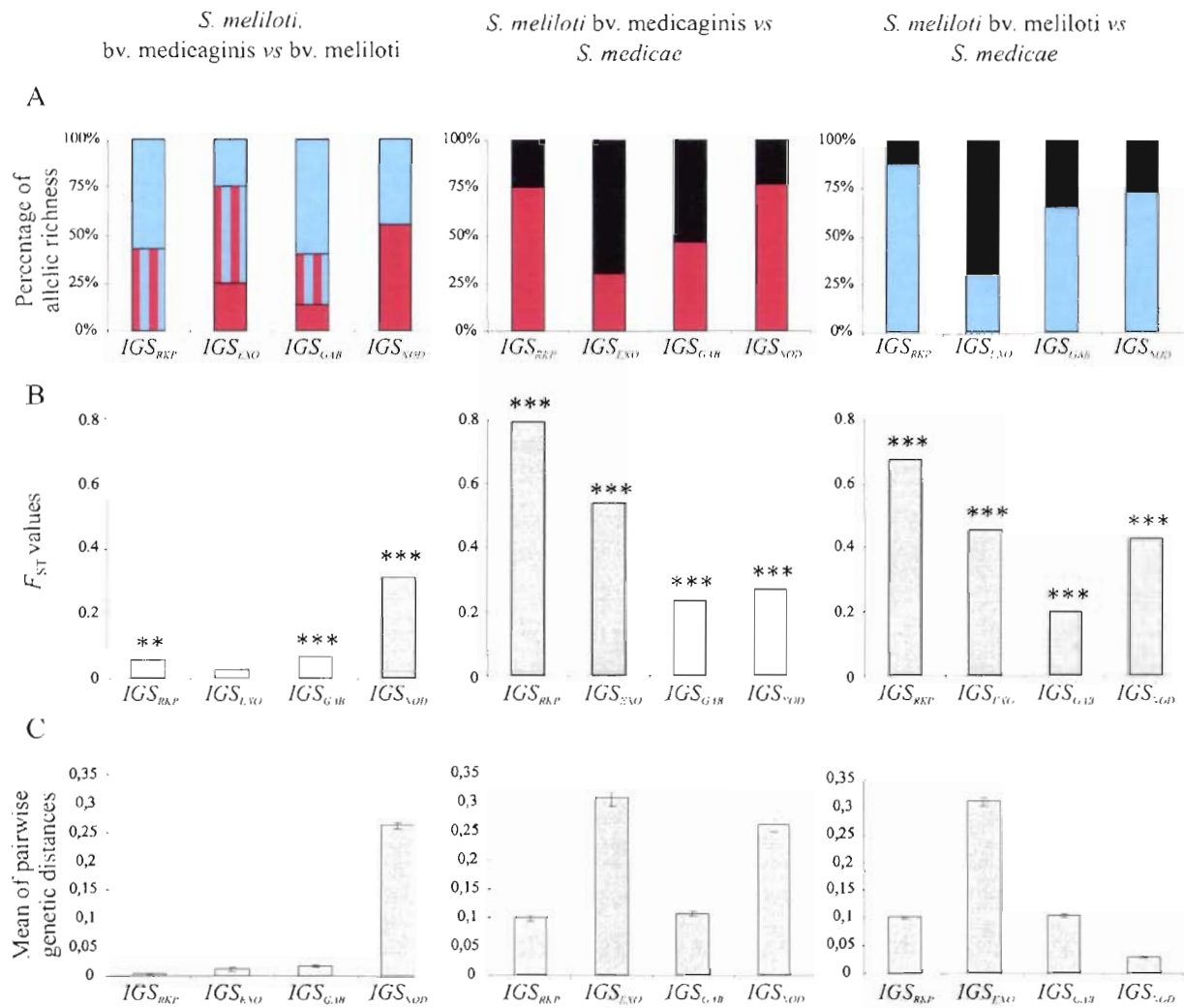


Figure 11. Distribution de la richesse allélique (A), F_{ST} et distance génétique moyenne (B et C) pour chaque locus, entre les isolats des populations tunisiennes, assignés à *Sinorhizobium medicae*, *Sinorhizobium meliloti* bv. *meliloti* et *S. meliloti* bv. *medicaginis* [Article 3]. Dans les graphiques consacrées à la distribution de la richesse allélique, les pourcentages d'allèles spécifiques de *S. medicae*, *S. meliloti* bv. *meliloti* et *S. meliloti* bv. *medicaginis* sont respectivement figurés en noir, bleu et rouge. Les allèles partagés par deux groupes sont représentés par un figuré hachuré. Dans les graphiques dédiés à la distance génétique moyenne entre isolats, les barres d'erreur représentent les intervalles de confiance à 95% des valeurs calculées.

approche locus par locus est nécessaire pour obtenir une description précise de la structure des populations de *S. meliloti*. Dans cette optique, nous avons calculé différentes mesures de différenciation entre les deux biovars pour les locus IGS_{RKP} , IGS_{NOD} , IGS_{GAB} et IGS_{EXO} (Fig. 11). Les locus étudiés présentent différents patrons de structuration.

Au niveau du locus IGS_{NOD} , les deux biovars ne partagent aucun allèle, ce qui se traduit par un fort F_{ST} et des tests de différenciation significatifs. En outre, les allèles des deux biovars forment des groupes monophylétiques très divergents sur les phylogénies IGS_{NOD} (Fig. 8) et *nodABC-nodEG* (Fig. 6). Ceci se traduit par une forte divergence génétique moyenne entre les allèles des deux biovars. Barran *et al.* (2002) ont mis en évidence que le déterminisme de la capacité à interagir efficacement avec *M. laciniata* et *M. truncatula* était lié à un polymorphisme au niveau du gène *nodC*. Il est donc fort probable qu'une sélection disruptive agisse au niveau de ce gène. Cette sélection pourrait influencer la variabilité génétique au niveau de l'ensemble de la région des gènes *nod* par un effet d'auto-stop génétique (Maynard Smith and Haigh 1974) à cause du faible taux de recombinaison au sein de cet îlot symbiotique.

Au contraire, les biovars *meliloti* et *medicaginis* ne sont pas différenciés pour le locus IGS_{EXO} . Par ailleurs, il est notable que le polymorphisme observé à ce locus se caractérise par deux lignées alléliques divergentes (Fig. 8). Nous reviendrons durant le chapitre VIII.2 sur cette caractéristique. Quoi qu'il en soit, l'absence de différenciation au locus IGS_{EXO} suggère que certaines régions du génome accessoire de *S. meliloti* évoluent sous des pressions de sélection cohésives.

Enfin, les locus IGS_{RKP} et IGS_{GAB} présentent des mesures de F_{ST} inférieures à celles observées pour IGS_{NOD} mais des tests de différenciation significatifs. Les faibles valeurs de F_{ST} sont expliquées par le fait qu'une part importante de la richesse allélique de *S. meliloti* est partagée par les deux biovars. Par ailleurs, alors que les allèles de *S. meliloti* à ces deux locus forment un groupe monophylétique, les allèles de chaque biovar forment des groupes paraphylétiques (Fig. 7). Ceci explique la faible divergence génétique moyenne observée entre les deux biovars.

Il est intéressant de noter que la structure génétique des populations sympatriques libanaises capables d'interagir soit avec *M. rigiduloïdes*, soit avec *M. truncatula* semble

comparable à celle observée dans le cas des populations sympatriques tunisiennes capables d'interagir soit avec *M. laciniata*, soit avec *M. truncatula*. En effet, les deux écotypes ne sont pas différenciés sur les marqueurs *IGS_{RKP}*, *IGS_{EXO}* et *IGS_{GAB}* (Fig. 4). Au contraire une nette divergence a été observée au niveau de l'îlot symbiotique *nod*, pour les cluster *nodABC* et *nodEG* (Fig. 6). Ceci pourrait donc indiquer que le déterminant génétique du différentiel de spécificité vis-à-vis de *M. rigiduloïdes* ou *M. truncatula* soit également situé dans l'îlot symbiotique *nod* (tout au moins à proximité).

Le fort niveau de divergence observé entre les différents clades de gènes *nod* de *S. meliloti* et la paraphylie des biovars de *S. meliloti* pour les autres marqueurs suggère que *S. meliloti* évolue selon un modèle de recombinaison cohésive, où les allèles conférant la spécificité d'hôte sont transférés aux différents fonds génétique de l'espèce (Gevers et al. 2005). On peut constater sur l'arbre des populations de *S. meliloti* échantillonnées durant cette thèse, basé sur les génotypes aux marqueurs *IGS_{RKP}*, *IGS_{EXO}* et *IGS_{GAB}* (Fig. 4), que la plante hôte n'est pas le principal facteur structurant la diversité des souches de *S. meliloti*, ce qui va à l'encontre de l'hypothèse d'une spéciation bactérienne liée à la spécificité d'hôte. Le fait que seule la région des gènes de nodulation soit responsable du différentiel de spécificité entre les différents écotypes de *S. meliloti* pourrait être impliqué dans le maintien de la structure génétique observée. En effet, l'îlot symbiotique (en tout cas, la région comprise entre *nodA* et *nodE*) semble former un module transféré dans son intégralité de manière fréquente. Une telle architecture génétique devrait faciliter le maintien des différents écotypes (ou biovars), le fond génétique de l'espèce évoluant sous des pressions de sélection indépendantes de celles qui agissent sur la région des gènes *nod*.

Toutefois, les processus à l'origine de la diversification fonctionnelle des gènes de nodulation chez *S. meliloti* restent indéterminés. Intuitivement, un mutant affecté dans le mécanisme de reconnaissance de son partenaire, chez l'un ou l'autre des organismes mutualistes, devrait être contre-sélectionné. Dans ces conditions, il semble difficile d'aboutir au patron de diversité observé autrement que par : i) une diversification allopatrique des systèmes de reconnaissance des *Medicago* alors que des flux de gènes existaient encore entre bactéries symbiotiques ; ii) une pression de sélection disruptive affectant le mécanisme de reconnaissance de couples symbiotiques sympatriques, à condition que les *Medicago* présentant un système de reconnaissance muté ne soient pas en compétition avec d'autres plantes du genre.

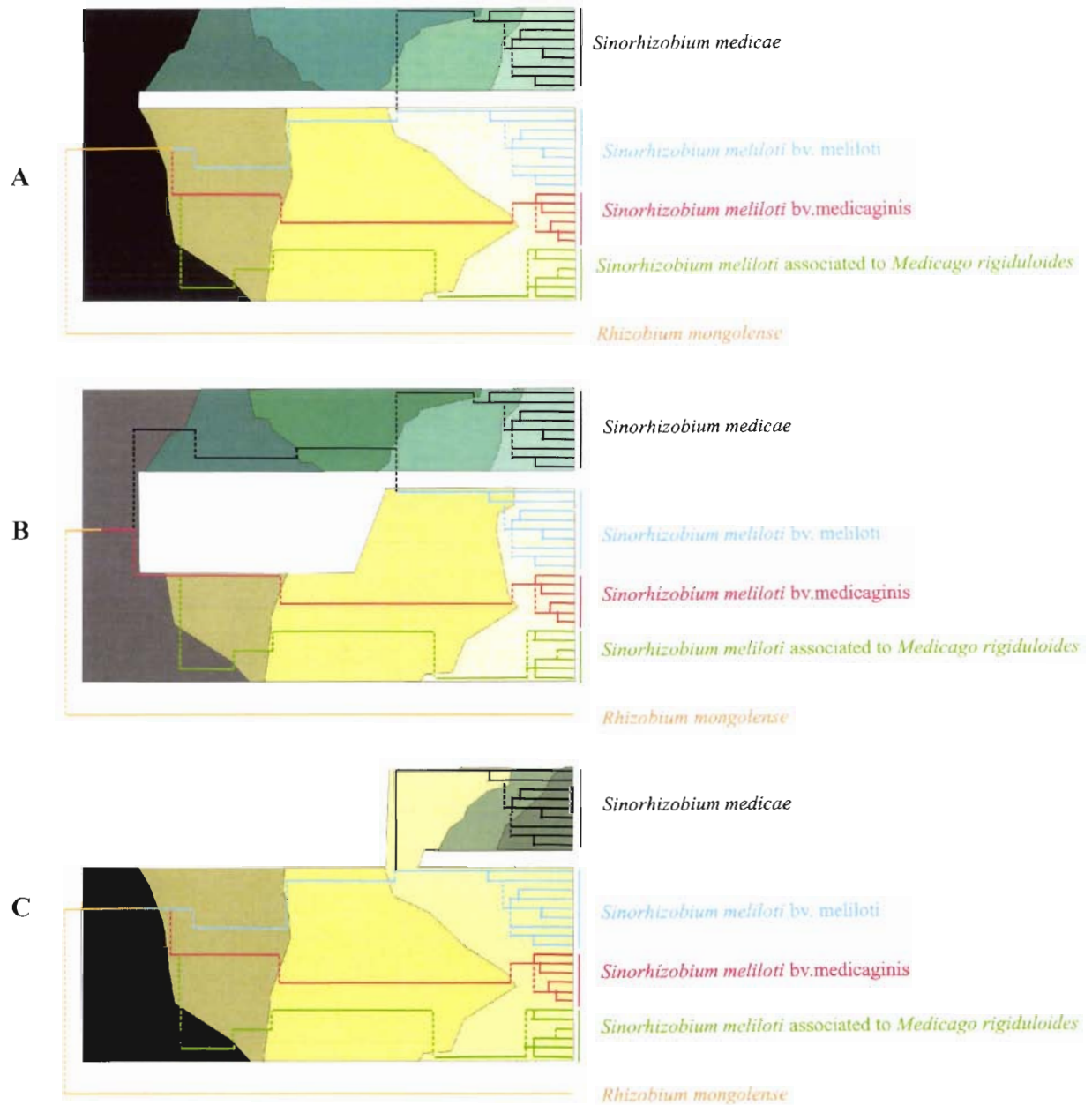


Figure 12. Représentation schématique des différents scénarios proposés pour expliquer les patrons de divergence entre *S. meliloti* et *S. medicae*. Les dégradés représentent l'évolution de séquence des fonds génétiques au cours du temps, l'arbre surimposé illustre l'évolution de la région *nod*. Dans l'arbre, les pointillés indiquent des événements de transfert horizontal.

VIII.1.2. Origine et évolution de la variabilité inter-spécifique chez *S. meliloti*

Inversement, les allèles de *S. medicae* sont clairement divergents de ceux rencontrés au sein des deux biovars de *S. meliloti* (Fig. 11). De plus, la monophylie des allèles de *S. medicae* dans l'ensemble des reconstructions phylogénétiques obtenues supporte l'hypothèse que le génome de base et le génome accessoire des deux espèces sont différenciés (Fig. 8). Ce résultat est en accord avec les études de génétique des populations conduites jusqu'alors (Eardly et al. 1990; Maynard Smith et al. 1993) et avec les méthodes de taxonomie bactérienne classique comme les mesures d'hybridation ADN/ADN qui ont mené à la description de *S. medicae* (Rome et al. 1996). Toutefois, si nos résultats confirment le statut d'espèce de *S. meliloti* et *S. medicae*, les questions de l'émergence et de la coexistence de ces deux espèces restent posées.

Alors que les phylogénies basées sur les locus *IGS_{RKP}*, *IGS_{GAB}* et *IGS_{EXO}* suggèrent que *S. meliloti* et *S. medicae* forment des groupes monophylétiques (*S. meliloti* et *S. medicae* sont décrites comme des espèces sœurs, notamment sur la base de phylogénies 16S), celles obtenues à l'aide de marqueurs localisés dans la région des gènes *nod* supportent l'hypothèse que *S. meliloti* forme un groupe paraphylétique à cause de la position des séquences de *S. medicae* (Fig. 6). En supposant que les îlots symbiotiques des souches associées aux plantes du genre *Medicago* forment un groupe monophylétique, deux scénarios évolutifs peuvent réconcilier les deux patrons que nous avons observés :

- un événement de spéciation à l'origine de *S. meliloti* et *S. medicae* puis un transfert d'îlot symbiotique interspécifique aurait pu se produire (Fig. 12 A, B). Par la suite, la séquence transférée aurait pu augmenter en fréquence et se répandre dans l'ensemble des fonds génétiques présents dans l'espèce receveuse par recombinaison intra-spécifique. Ainsi, l'émergence des trois lignées fonctionnelles *nod* serait liée à une sélection disruptive et la faible distance génétique entre les îlots de *S. meliloti* bv. *meliloti* et *S. medicae* serait expliqué par une date de divergence récente liée au transfert horizontal. Dans ce contexte, le fait qu'un seul écotype de *S. medicae* ait été découvert jusqu'à maintenant va à l'encontre de l'hypothèse d'un transfert horizontal d'îlot symbiotique de *S. meliloti* vers *S. medicae*. En effet, si *S. medicae* a acquis son îlot symbiotique par transfert horizontal de matériel génétique, il est possible que des populations dépourvues d'îlot symbiotique subsistent à l'image des données obtenues

pour *S. adhaerens* (Rogel et al. 2001). Toutefois, cet argument doit être relativisé du fait de notre méconnaissance de la diversité microbienne.

-après la diversification fonctionnelle de la région *nod* au sein de populations ancestrales de *S. meliloti*, une partie des souches, possédant un îlot symbiotique permettant d'interagir avec le groupe de spécificité de *M. truncatula*, aurait subi un isolement sexuel induisant la spéciation de *S. medicae* (Fig. 12 C). Une étude récente (Biondi et al. 2003), qui suggère que *S. medicae* serait originaire d'une population ancestrale de *S. meliloti*, est compatible avec ce scénario. Toutefois, cette hypothèse pose une contrainte quant au temps écoulé depuis la spéciation, la distance génétique estimée entre *S. meliloti* et *S. medicae* étant de 0,080+/-0,008 mutations par paire de base (μ /bp) sur la séquence des gènes *recA*, *atpD*, *glnA* et *glnII* et la distance génétique moyenne entre des individus de *S. meliloti* et *S. medicae* étant de l'ordre de 0,103+/-0,001 μ /bp et 0,103+/-0,008 μ /bp sur les marqueurs *IGS_{RKP}* et *IGS_{GAB}*. De plus, aucun polymorphisme partagé par les deux espèces n'a été observé. Ces données suggèrent donc que la spéciation de *S. meliloti* et *S. medicae* n'est pas particulièrement récente, ce qui remet en question l'image évolutive révélée par Biondi *et al.* (2003). Elle soulève également des questions sur une vitesse d'évolution spécifique de la région *nod* la distance génétique moyenne entre de *S. meliloti* bv. *meliloti* et *S. medicae* étant de 0,030+/-0,003 μ /bp sur *IGS_{NOD}* (contre 0,269+/-0,011 μ /bp entre *S. meliloti* bv. *meliloti* et *S. meliloti* bv. *medicaginis*). Malgré tout, une telle différence de vitesse d'évolution ne semble pas aberrante compte tenu de la vitesse d'évolution observée sur d'autres marqueurs. Par exemple le niveau de divergence entre les souches de *S. meliloti* et *S. medicae* est de 0,003 μ /bp au niveau de l'ADNr 16S. Dans ce contexte, la faible vitesse d'évolution observée pour le marqueur *IGS_{NOD}* pourrait être liée à une différence de régime de sélection, d'effectif efficace ou de taux de mutation affectant l'îlot symbiotique (Bromham and Penny 2003).

Dans ces conditions, mettre en évidence le scénario le plus probable sur la base des données actuelles semble difficile. L'acquisition de données sur une part plus importante du mégaplasme pSymA semble nécessaire pour affiner les hypothèses relatives à l'émergence de *S. meliloti* et *S. medicae*. Notamment, la recherche de traces de recombinaisons interspécifique sur pSymA pourra être menée après le séquençage du génome de la souche WSM 419 de *S. medicae* par le « Joint Genome Institut ».

Parallèlement aux scénarios proposés pour expliquer l'émergence du patron de diversité observé pour la région symbiotique *nod*, trois processus pourraient expliquer la mise en place de l'isolement sexuel entre *S. meliloti* et *S. medicae* : (i) un isolement géographique aurait pu induire une divergence génétique, limitant les transferts horizontaux une fois que les deux espèces se sont retrouvées en sympatrie. En effet, la probabilité d'observer un transfert horizontal de matériel génétique est positivement corrélée à l'homologie des génomes donneurs et receveurs (Vulic et al. 1997) ; (ii) les mécanismes impliqués dans la régulation et l'efficacité des transferts horizontaux de matériel génétique pourraient être à l'origine d'un mécanisme d'isolement sexuel précédant la phase de transfert. Notamment, les signaux émis par les plantes hôtes (comme décrit par Oger et Farrand chez *Rhizobium radiobacter* (2002)), la spécificité liée à l'organisme donneur (De Gelder et al. 2005) ou les mécanismes liés aux communications entre bactéries (voir Lyon et Novick (2004) et Whitehead *et al.* (2001) pour des revues sur les peptides-sinaux et le quorum-sensing) pourraient être impliqués dans la mise en place d'un isolement sexuel « actif » ; (iii) le taux de recombinaison efficace aurait pu diminuer si les souches qui subissent plusieurs transferts horizontaux présentent une valeur sélective inférieure à celle des génotypes « parentaux ». Cette hypothèse est utilisée dans différents modèles visant à mettre en évidence l'impact de la recombinaison homologue sur la spéciation d'écotypes bactériens (Cohan 1994a; Cohan 1994b; Cohan 1995). Une telle diminution de valeur sélective pourrait être liée à la perte de certaines fonctions ou à des allocations de ressource inadaptées, par exemple du fait de transferts de gènes de régulation. Ainsi, la complexité (*i.e.* le nombre et la localisation des modules impliqués) des mécanismes de reconnaissance (Mitra and Long 2004) et des interactions physiologiques (Kiers et al. 2003; Ludwig et al. 2003) entre les symbiotes pourrait également être à l'origine d'un isolement sexuel entre lignées bactériennes.

Parce que *S. meliloti* bv. *meliloti* peut fixer l'azote en interaction avec un sous-ensemble des espèces de *Medicago* qui interagissent de manière efficace avec *S. medicae*, il est difficile d'expliquer la coexistence de ces deux groupes génétiques sans émettre l'hypothèse de compromis différents entre les composantes de leur valeur sélective, à moins que les plantes ne constituent pas une ressource limitante. En effet, la compétition entre des symbiotes généralistes et spécialistes devrait conduire à la disparition des derniers, à moins qu'ils soient plus performants que les généralistes sur les plantes hôtes utilisées par les deux phénotypes. Cette différence pourrait être expliquée par une meilleure capacité à former des nodosités et/ou un nombre plus important de descendants résultants de chaque nodosité

(Parker 1999). Ainsi, on peut envisager qu'une adaptation différentielle à l'interaction symbiotique puisse être liée à la fois à l'émergence et à la coexistence de *S. meliloti* bv. *meliloti* et *S. medicae*. En outre, les travaux de modélisation publiés par Parker (1999) soulignent qu'une structuration spatiale des populations de plantes hôtes présentant des spécificités différentes est un facteur susceptible de favoriser la coexistence de *rhizobium* spécialistes et généralistes.

Si la discussion concernant l'émergence et la coexistence de *S. meliloti* et *S. medicae* peut sembler particulièrement spéculative, elle soulève un problème intéressant. La majorité des études de co-évolution dans les systèmes mettant en jeu des bactéries mutualistes (surtout dans le cas des *rhizobium*) ont essayé de mettre en relation, sans grande réussite, la phylogénie des hôtes et celle des gènes bactériens impliqués dans la reconnaissance par l'hôte (Wernegreen and Riley 1999; Bena et al. 2005). D'une part, les transferts horizontaux de gènes impliqués dans la spécificité d'hôte entre bactéries ou les événements de perte/capture de symbiotes font partie des processus qui limitent l'intérêt des recherches d'évènements de co-cladogenèse pour améliorer notre compréhension de l'évolution des mutualismes. D'autre part, l'évolution des signaux de spécificité n'est pas le seul mécanisme symbiotique pouvant influencer la diversification des *rhizobium*. Notamment, la spécificité des échanges métaboliques entre les partenaires symbiotiques pourrait être un moteur pour leur évolution. Dans ce contexte, étudier l'évolution de la physiologie des interactions mutualistes (*i.e.* l'évolution qualitative et quantitative des échanges de métabolites et des sanctions entre partenaires) par l'étude des phénotypes mis en jeu et/ou des déterminants génétiques sous-jacents ou bien étudier l'impact de la spécialisation physiologique sur la diversification génétique des symbiotes par l'intermédiaire d'approches théoriques semble être une alternative intéressante. Dans ce contexte, certaines pistes de recherche sont présentées dans le chapitre IX.2. Elle pourrait permettre de mieux comprendre l'influence des interactions symbiotiques sur la diversité génétique et fonctionnelle des organismes qui y participent, ceci à différentes échelles, de l'évolution des symbiotes du genre *Medicago* à la réussite de *rhizobium* atypiques (*i.e.* *rhizobium* photosynthétiques ou méthylotrophes, β -*rhizobium*, ...).

VIII.2. Evolution des populations de S. meliloti et S. medicae interagissant avec des plantes hôtes présentant des spectres d'hôtes comparables : des spécialisations tous azimuts ?
{Article 2}

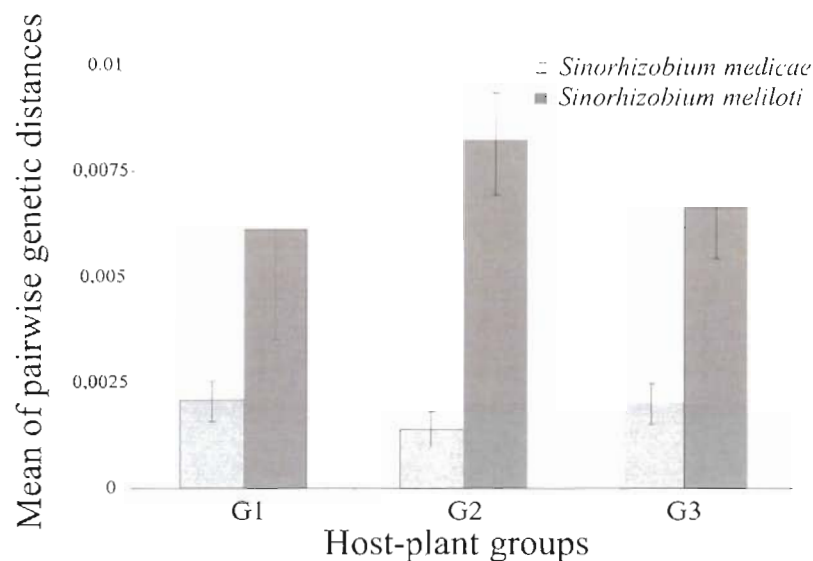


Figure 13. Valeurs et intervalles de confiance à 95% des distances génétiques moyennes entre les souches de *Sinorhizobium meliloti* et *Sinorhizobium medicae* isolées des lots de plante G1 (20 espèces du genre *Medicago*), G2 (20 lignées de l'espèce *Medicago truncatula*) et G3 (20 plantes de la lignée fixée *Medicago truncatula* Jemalong A17) {Article 2}.

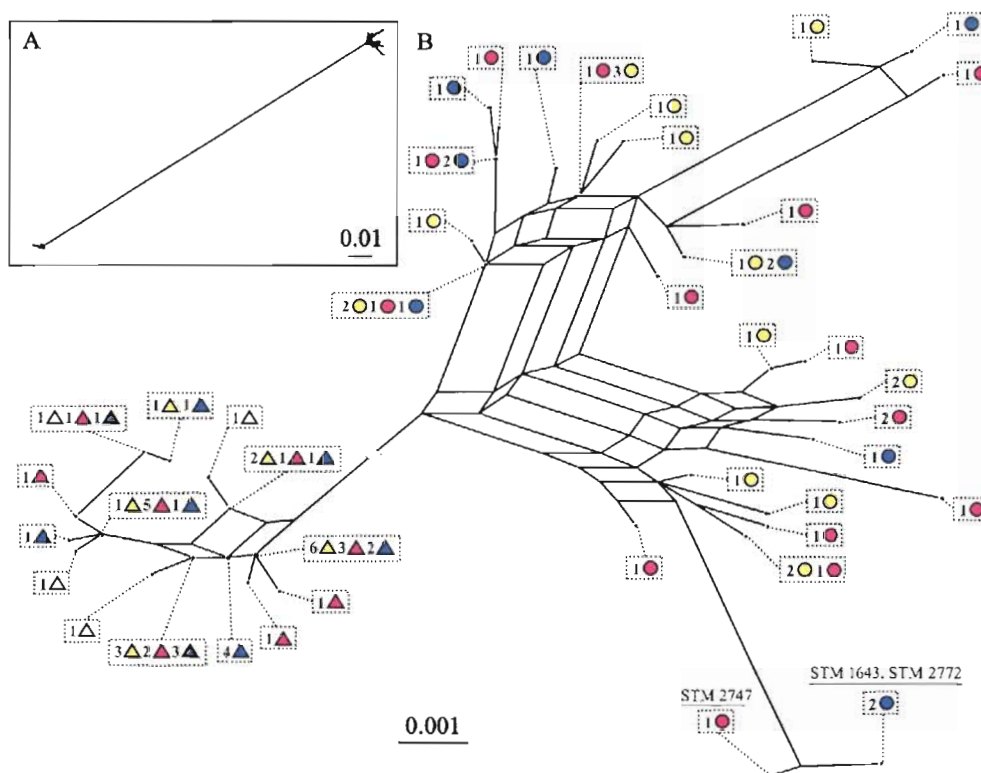


Figure 14. Vue globale (A) et détaillée (B) du réseau mettant en relation les génotypes de *Sinorhizobium medicae* (figurés par des triangles) et de *Sinorhizobium meliloti* (figurés par des cercles) isolés des 20 espèces de *Medicago* (figurés bleus), des 20 lignées de *Medicago truncatula* (figurés rouges) et des 20 plants de *Medicago truncatula* Jemalong A17 (figurés jaunes) {Article 2}. Les numéros indiquent le nombre de représentants du génotype multilocus.

Nous avons essayé de relier la diversité des plantes du genre *Medicago* à la diversité des souches de *Sinorhizobium* qui leurs sont associées. Dans ce but, une étude de génétique des populations a été conduite sur la base de l'échantillonnage E1 {Article 2}. Les analyses effectuées, basées sur la richesse allélique, les fréquences alléliques, le niveau de divergence entre génotypes (Fig. 13) et les relations phylétiques entre génotypes (Fig. 14), n'ont pas permis d'établir une relation entre la diversité des plantes hôtes utilisées pour le piégeage et celle des bactéries échantillonnées dans leurs nodosités, que ce soit pour *S. meliloti* bv. *meliloti* ou pour *S. medicae*. Ce résultat suggère que l'échantillonnage de la diversité symbiotique des deux taxa peut être effectué à partir d'un nombre limité de génotypes de plantes hôtes. Dans notre cas, échantillonner avec la lignée fixée *M. truncatula* Jemalong A17 aurait été suffisant pour récupérer toute la diversité génétique observée. Il est toutefois notable que *M. truncatula* Jemalong A17 peut être considéré comme un hôte présentant une spécificité assez large puisqu'il interagit à la fois avec *S. meliloti* bv. *meliloti* et *S. medicae*.

Cette conclusion peut sembler *a priori* différente de celle obtenue sur la base d'échantillonnages *in situ* à partir desquels il avait été montré que différents cultivars de *M. sativa* interagissaient avec des populations bactériennes différenciées (Paffetti et al. 1996; Paffetti et al. 1998) ou sur la base d'études physiologiques chez *S. meliloti* qui ont décrit des déterminants génétiques impliqués dans des différences d'aptitudes à interagir avec différents génotypes de *M. truncatula* (Kiss et al. 2004). Toutefois, il est important de noter que notre plan d'échantillonnage n'était pas fait pour détecter des préférences quantitatives de certains génotypes de plantes hôte envers des génotypes bactériens. Par ailleurs, le niveau de diversité des populations bactériennes contenues dans le sol utilisé peut avoir une influence sur le résultat de ce type d'étude. Par exemple, nous n'avons pas isolé lors de l'échantillonnage E1 de souches capables d'interagir efficacement avec *M. laciniata* ou *M. rigiduloides*.

Le manque de relation entre les diversités des deux partenaires symbiotiques pourrait être lié à deux autres aspects de l'évolution de la spécificité d'interaction entre les espèces de *Medicago* et leurs *Sinorhizobium*. Premièrement, même si le processus de co-évolution entre les partenaires symbiotiques est structuré dans l'espace (*i.e.* les populations de *rhizobium* sont localement adaptées à leurs hôtes), les bactéries d'un même biovar sont généralement en contact *in natura* avec différents génotypes de plantes hôtes voir plusieurs espèces de plantes hôtes, ce qui pourrait limiter les pressions de sélection pour une spécialisation des lignées de *rhizobium*. En effet, Carelli *et al.* (2000) ont décrit une homogénéisation progressive des

Encadré 6. Détection de la sélection

Déterminer la nature, l'intensité et l'influence relative des différentes pressions de sélection qui s'exercent sur les génomes des organismes symbiotiques est fondamental pour comprendre leur diversification. A cet effet, différentes méthodologies permettent de tester l'adéquation entre les patrons de polymorphisme nucléotidique observés et un modèle évolutif neutre. Ces méthodes peuvent être groupées en fonction de l'échelle évolutive à laquelle elles peuvent être appliquées :

-Un premier ensemble de méthodes s'intéresse aux patrons de polymorphismes à l'échelle intra-spécifique. Un des intérêts majeurs de ces méthodes est qu'elles peuvent s'appliquer aussi bien sur des locus codants que sur des espaceurs intergéniques. Au sein de cet ensemble trois approches principales, utilisant la diversité à un locus donné, peuvent être distinguées : a) les approches basées sur la réponse différentielle de plusieurs estimateurs du paramètre θ en cas d'écarts à un modèle d'évolution neutre. C'est notamment le cas du D de Tajima (Tajima 1989) utilisé dans le cadre de l'article 2 ; b) les approches basées sur les propriétés du spectre de fréquences haplotypiques attendu sous un régime de sélection neutre (Ewens 1972). On peut également utiliser des approches basées sur la comparaison des probabilités d'observer les données étudiées sous un modèle d'évolution neutre ou implémentant l'effet d'une pression de sélection comme celle développée par Galtier *et al.* (2000).

-Un second ensemble d'approches phylogénétiques étudie la distribution des mutations non-synonymes et est donc dédiée à l'analyse de locus codant pour des protéines. Dans le cas de l'approche la plus utilisée (Yang 1998), le but est de comparer les probabilités des données sous des modèles d'évolution prenant en compte différents *a priori* sur la distribution de ω (ou dN/dS, le ratio de mutations non-synonymes/synonymes). Les modèles comparés peuvent se focaliser soit sur la distribution de ω entre les différents sites de l'alignement ou les différentes branches de l'arbre. On peut penser que l'échelle évolutive la plus appropriée pour utiliser ce type d'approche est l'échelle interspécifique. Deux explications motivent cette position : (i) une étude récente indique que le ratio dN/dS pour un gène peut varier au cours du temps à l'échelle intra-spécifique au sein des populations bactériennes à cause d'effets indirects tels que l'auto-stop génétique (Rocha *et al.* 2006) ; (ii) ces méthodologies s'appuient sur l'obtention de phylogénies par maximum de vraisemblance. Or, obtenir des reconstructions phylogénétiques fiables à l'échelle intra-spécifique peut être compliqué par l'existence d'événements de recombinaison au sein des gènes.

-Enfin, un dernier ensemble d'approches s'intéresse à la comparaison des patrons de polymorphisme aux échelles intra et inter-spécifiques. Parmi les deux principales méthodologies utilisant cette propriété, certains tests comme celui développé par Hudson, Kreitman et Aguadé (Hudson *et al.* 1987) peuvent s'appliquer à des séquences non-codantes contrairement au test de Mac Donald et Kreitman (McDonald and Kreitman 1991) qui utilise les notions de mutation synonymes et non synonymes. Le test est effectué sur des données de séquence obtenues sur plusieurs locus vise à mettre en évidence des écarts entre les niveaux de divergence observés à chaque locus, au sein de chaque espèce et entre espèces, et les niveaux de divergence attendus sous un modèle neutre. La significativité des écarts observés est jugée sur la base d'un paramètre d'ajustement qui suit une loi de χ^2 . Des extensions de ce test basées sur le calcul de la vraisemblance des données sous différents modèles faisant

populations de *rhizobium* associées à différents cultivars de *M. sativa*, cultivés sur un même sol pendant quatre ans. Ce résultat est en accord avec l'hypothèse que les populations de *S. meliloti* ou de *S. medicae* évoluent pour devenir plus généralistes lorsqu'elles sont en contact avec différents génotypes de plantes hôtes. Deuxièmement, les groupes d'espèces de *Medicago* basés sur les capacités symbiotiques sont paraphylétiques dans la phylogénie des plantes du genre *Medicago* (Bena et al. 2005). Notamment, la capacité à interagir avec les seules souches de *S. medicae* ou avec à la fois les souches de *S. medicae* et les souches de *S. meliloti* bv. *meliloti* a évolué par plusieurs événements de réversion ou de convergence (Fig. 1). Ce patron d'évolution, par le jeu de la compétition entre souches spécialisées sur le même type d'hôte et par le mécanisme d'évolution vers des comportements généralistes au sein des populations bactériennes, devrait réduire la corrélation entre la diversité des plantes hôtes et celles de leurs symbiotes ainsi que la probabilité d'observer des événements de co-spéciation par le biais d'approches phylogénétiques.

Si la diversité génétique des plantes utilisées pour le piégeage n'a pas eu une influence directe sur la diversité des symbiotes étudiés (*i.e.* pas de corrélation entre les deux séries de mesures), la recherche de traces de sélection (voir l'encadré 6) suggère que l'interaction avec les populations naturelles de plantes hôtes a influencé la diversité des populations de *Sinorhizobium* étudiées. Étonnamment, nous avons observé une forte corrélation entre les valeurs de D de Tajima calculées pour chaque locus, chez les deux espèces ($R^2=0,70$, $p<0,05$) (Fig. 15). La corrélation observée est en dehors de la distribution attendue, cette dernière étant obtenue par des simulations de coalescents neutres conditionnées au nombre de sites polymorphes observés pour chaque locus (en l'absence d'information sur les taux de mutations, et donc à quel estimateur de θ se fier, il est préférable de travailler à nombre de sites polymorphes constant) (Depaulis et al. 2001). Elle suggère donc que des pressions de sélection similaires s'exercent aux mêmes locus chez *S. meliloti* et *S. medicae*. Les deux espèces présentent a priori la même architecture de génome, des niches écologiques proches (tant au niveau du sol qu'au niveau des plantes hôtes) et des comportements sexuels similaires ce qui pourrait les amener à évoluer sous les mêmes pressions de sélection.

La majorité des locus présente des patrons de polymorphisme qui ne s'écartent pas de l'attendu sous un modèle neutre, malgré la corrélation observée entre les D de Tajima chez les deux espèces. Au contraire, les locus *IGS_{NOD}* et *IGS_{EXO}* présentent des marques de sélection sur la base de la statistique D de Tajima (1983). Le locus *IGS_{NOD}* est localisé dans l'îlot

varier l'indépendance des vitesses d'évolution entre locus et/ou l'influence de la sélection sur la divergence nucléotidique aux niveaux intra et interspécifiques sont maintenant disponibles (Wright and Charlesworth 2004).

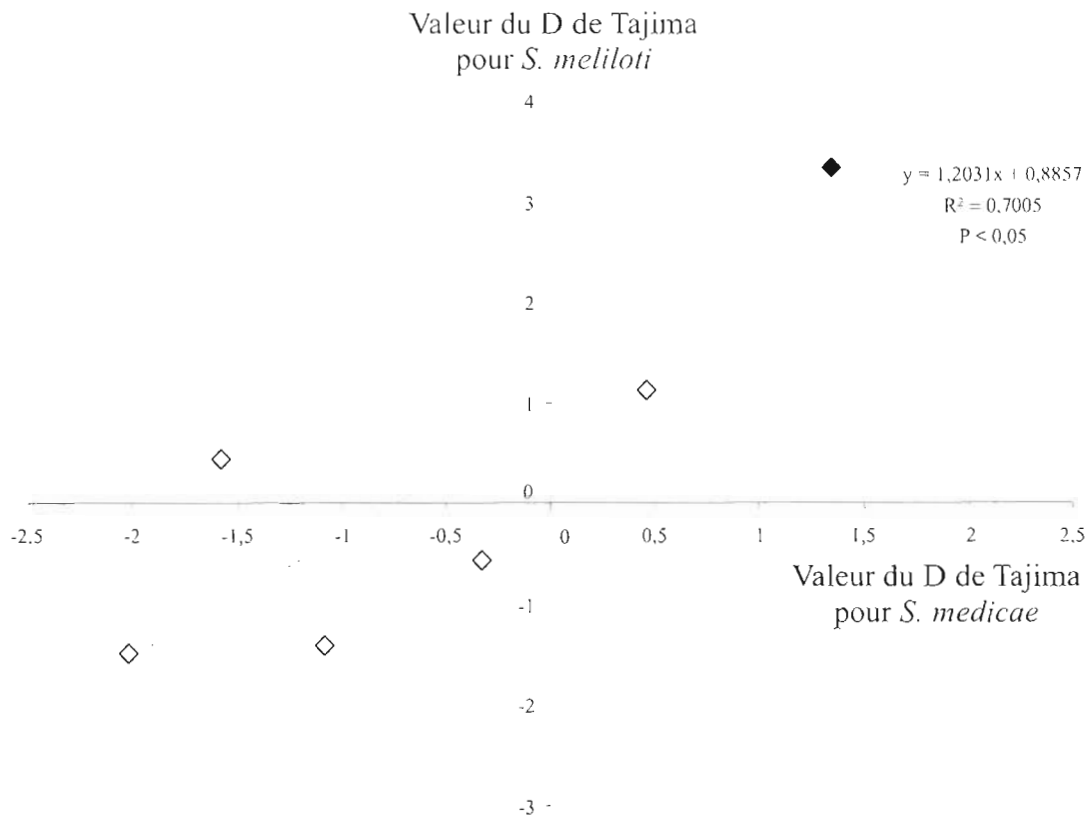


Figure 15. Distribution des valeurs de D de Tajima prise pour chaque locus pour la population de *Sinorhizobium meliloti* et pour la population de *Sinorhizobium medicae*. Les figurés gris indiquent que la valeur de D ne s'écarte pas de l'attendu neutre pour les deux espèces. Les figurés blanc et noir indiquent respectivement que la valeur de D de Tajima s'écarte de l'attendu neutre pour *S. medicae* (IGS_{NOD}) et pour *S. meliloti* (IGS_{EXO}) {Article 2}.

symbiotique du mégaplasmide pSymA, qui contient les gènes *nod*, *nif* et *fix*. Ce locus présente, à la fois chez *S. meliloti* et chez *S. medicae*, les valeurs de D les plus négatives parmi celles que nous avons calculées. De plus, la valeur négative rencontrée chez *S. medicae* diverge significativement de l'attendu neutre. Puisque un seul marqueur présente une valeur de D significativement différente de l'attendu neutre (et compte tenu de la distribution particulière des D de Tajima entre espèces), l'hypothèse qu'un événement démographique soit à l'origine du polymorphisme observé n'est pas la plus probable. Nous avons donc proposé qu'un balayage sélectif lié à une pression de sélection purificatrice ou positive est l'explication la plus probable pour les valeurs négatives de D observée sur le locus *IGS_{NOD}*. Parker (1999) a proposé que dans les mutualismes, contrairement à ce qui est attendu dans les interactions antagonistes, les individus présentant une mutation du système de reconnaissance devrait être désavantagés. Une étude phylogénétique récente chez les Rickettsiaceae (Jiggins et al. 2002), portant notamment sur la diversité du gène *wsp* qui code pour une protéine de surface, montre que certains sites de la séquence de ce gène évoluent sous sélection positive dans les lignées parasitaires (un modèle incluant une classe de site avec $\omega > 1$ est plus probable), alors que l'ensemble des sites des lignées mutualistes évoluent sous sélection purificatrice (modèle avec $\omega < 1$). Sur la base de ces travaux, nous avons supposé que les valeurs de D négatives observées pour le locus *IGS_{NOD}* étaient liées à une pression de sélection purificatrice. Toutefois, l'existence d'épisodes de sélection positive n'est pas improbable compte tenu du niveau de divergence entre les îlots symbiotiques *nod* des biovars de *S. meliloti*.

A l'inverse, le locus *IGS_{EXO}* présente les valeurs de D les plus fortes chez *S. meliloti* et *S. medicae*. De plus, le patron de polymorphisme observé s'écarte significativement de l'attendu neutre pour le jeu de séquence de *S. meliloti*. Pour les mêmes raisons que celles invoquées dans le cas du locus *IGS_{NOD}*, l'hypothèse qu'un événement démographique soit à l'origine des valeurs de D observées pour le locus *IGS_{EXO}* semble peu probable. De plus nous n'avons pas détecté de sous structure, qui aurait pu biaiser l'ensemble des valeurs de D de Tajima vers des valeurs positives, à partir du réticulogramme construit, des tests de différenciation effectués entre les groupes de plantes hôtes ou à l'aide du logiciel STRUCTURE (les analyses ont été effectuées en groupant les locus chromosomiques qui présentent du déséquilibre de liaison et qui pourraient biaiser l'analyse). Les observations effectuées suggèrent donc que la région incluant le locus *IGS_{EXO}* évolue sous un régime de sélection balancée. Or, plusieurs groupes de gènes impliqués dans la synthèse d'exopolysaccharides de

surface (EPS) sont physiquement liés à *IGS_{EXO}*, comme les gènes *exo* ou *exp*. Plusieurs études physiologiques ont mis en évidence le rôle de ces gènes dans la mise en place d'une symbiose efficace chez *S. meliloti* (Battisti et al. 1992; Gonzalez et al. 1996a; Gonzalez et al. 1996b), notamment de par leur action sur les mécanismes de défense de la plante hôte (voir Fraysse *et al.* (2003) pour une revue à ce sujet). Un travail récent sur l'évolution de nombreux gènes bactériens, à l'échelle intraspécifique, a montré que les protéines de surface des espèces parasites, qui peuvent être reconnues par les hôtes eucaryotes, évoluaient fréquemment sous sélection balancée (Hughes 2005). De fait, l'adaptation à différents génotypes d'hôtes est avancé pour expliquer le patron de sélection balancée au niveau du gène *ospC* dans les populations du parasite *B. burgdoferi* (Brisson and Dykhuizen 2004).

Le patron de recombinaison inféré à l'aide de mesures de déséquilibre de liaison a probablement facilité l'action des pressions de sélection différentielles inférée pour les locus plasmidiques *IGS_{NOD}* et *IGS_{EXO}*. Ces marqueurs étant en équilibre de liaison, les interactions entre les pressions de sélection qui s'exercent sur les deux régions génomiques sont limitées. Cette différence de pressions de sélection pourrait constituer une illustration du dilemme des organismes symbiotiques (van Baalen and Jansen 2001). En effet, la valeur sélective des *rhizobium* dépend de leur capacité : i) à interagir avec les plantes hôtes, cette dernière étant tributaire des étapes précoces de reconnaissance entre plantes et *rhizobium* liées aux gènes *nod* ; ii) à échapper aux sanctions *sensu lato* de la plante hôte qui sélectionnent des comportements plus coopératifs au détriment des bactéries symbiotiques (West et al. 2002b). Dans ce contexte, une explication possible pour les patrons de polymorphisme observé serait que les étapes précoces de reconnaissance par l'hôte (*i.e.* la reconnaissance par le biais des Facteurs-Nod) induisent des pressions de sélection purificatrice au niveau des gènes *nod*, alors que les étapes tardives de reconnaissance induisent des pressions de sélection fréquence-dépendantes sur les gènes impliqués dans la synthèse de polysaccharides de surface. Le mécanisme à l'origine de cette sélection fréquence-dépendante reste à éclaircir. La mise en place d'une sélection balancée au niveau des gènes intervenant dans la synthèse de polysaccharides de surface pourrait être liée :

- aux pressions de sélection sur le compromis d'allocation des ressources entre les fonctions symbiotiques et le taux de reproduction. Certains récepteurs de l'hôte, par exemple les défensines ou les lectines pourraient induire une préférence d'interaction avec un phénotype « polysaccharides de surface » donné. Or, il est envisageable qu'une augmentation de l'appareillement au sein des populations de

rhizobium puisse conduire à une diminution du taux de fixation optimal (West et al. 2002b). Dans ces conditions, on peut se demander si un polymorphisme protégé pourrait émerger tant au niveau des récepteurs des plantes hôtes qu'au niveau des gènes bactériens impliqués dans le phénotype « polysaccharides de surface ».

-à l'hétérogénéité des espèces de plantes hôtes qui interagissent avec les *rhizobium*. Si des différences de spécificité vis-à-vis des phénotypes « polysaccharides de surface » existent entre espèces de *Medicago* sympatriques, la fréquence des allèles codant pour les différents phénotypes bactériens sera fonction de la fréquence des différentes plantes hôtes. Cette hypothèse pose toutefois la question de l'origine d'un tel polymorphisme de spécificité.

Ce type de régime de sélection pourrait engendrer une diminution de l'efficacité de sanctions *a priori* de la plante hôte *via* des réactions de défense qui interviendraient avant les mécanismes de sanction *a posteriori* basées sur le niveau de fixation d'azote dans les nodosités (Kiers et al. 2003).

VIII.3. Remarque conclusive

Enfin, il est intéressant de constater que les patrons de divergence observés entre *S. meliloti* bv. *meliloti* et *S. medicae* (surtout abordés dans l'article 3) présentent une forte analogie avec les résultats de l'étude des patrons de sélection au sein de *S. meliloti* bv. *meliloti* et *S. medicae* {Article 2}, les valeurs de D de Tajima étant liées au niveau de divergence moyen entre les allèles. En effet, on distingue dans les deux approches trois patrons : (i) les patrons de diversité des locus *IGS_{RKP}* et *IGS_{GAB}* ne semblent pas s'écarter de l'attendu neutre à l'échelle intra-spécifique. Or, les niveaux de divergence moyens entre *S. meliloti* bv. *meliloti* et *S. medicae* à ces deux locus ne sont pas significativement différents. (ii) Au contraire, le locus *IGS_{EXO}* présente des valeurs de D de Tajima positives qui s'écartent du modèle neutre chez *S. meliloti* bv. *meliloti*. Ces valeurs de D sont liées à l'observation d'une faible richesse allélique, chaque allèle étant très différencié (dans le cas de *S. meliloti* bv. *meliloti*, 2 allèles avec 10 sites en ségrégation). Parallèlement, le niveau moyen de divergence entre les allèles *IGS_{EXO}* de *S. meliloti* bv. *meliloti* et *S. medicae* est significativement plus important que celui observé sur *IGS_{RKP}* et *IGS_{GAB}*. (iii) Enfin, le locus *IGS_{NOD}* présente des valeurs de D de Tajima négatives qui s'écartent du modèle neutre chez *S. medicae*. Ces valeurs de D sont liées à présence d'allèles peu divergents caractérisés par un spectre de fréquence très déséquilibré (typique d'une généalogie en « étoile »). Parallèlement,

VIII. Evolution de la spécificité et diversification des *Sinorhizobium* spp.

le niveau moyen de divergence entre les allèles IGS_{NOD} de *S. meliloti* bv. meliloti et *S. medicae* est significativement plus faible que celui observé sur IGS_{RKP} et IGS_{GAB} . La ressemblance des patrons observés aux échelles intra- et inter-spécifique pourrait indiquer que les même processus seraient à l'origine des deux types de polymorphisme.

IX. Conclusions et (surtout) perspectives

IX. Conclusions et (surtout) perspectives

Le travail effectué a permis de mieux cerner certains aspects de la structure et de l'évolution des populations de *Sinorhizobium* sp. associées aux plantes du genre *Medicago*. Toutefois, certaines questions restent posées. Le but de ce chapitre est essentiellement de discuter des perspectives de recherche qui découlent de ces questions.

IX.1. Transferts horizontaux

Dans le chapitre consacré à l'influence des transferts horizontaux sur la diversité des souches de *Sinorhizobium* sp., nous avons montré que les marqueurs localisés sur les mégaplasmines pSymA et pSymB étaient sujets à des transferts horizontaux de matériel génétique. Chez *S. meliloti* l'ensemble de ces marqueurs sont en effet en équilibre de liaison aussi bien entre eux qu'avec les marqueurs chromosomiques. Comme discuté précédemment, le mécanisme à l'origine de ces transferts horizontaux reste relativement obscur. Si des études permettant de caractériser les déterminants génétiques à l'origine des transferts observés sont indispensables pour une meilleure caractérisation du système, des études de génétique des populations pourraient permettre d'obtenir des informations complémentaires. En effet, les différents mécanismes de transfert qui peuvent intervenir entre bactéries (conjugaison typique ou d'îlots), transformation et transduction devraient engendrer des patrons de déséquilibre de liaison différents au sein des génomes (Hudson 1994) (Fig. 16). Ces patrons pourraient être mis en évidence par l'étude du niveau de déséquilibre de liaison entre des marqueurs polymorphes au sein du génome de *S. meliloti*. Le séquençage de marqueurs génétiques localisés de manière régulière le long d'un des mégaplasmines pourrait être utilisé à cet effet. A l'heure actuelle, les puces de réséquençage constitueraient l'outil le plus approprié pour mener à bien la caractérisation de ces marqueurs. Ce type d'outil permettant de déterminer environ 30000 pb au sein d'un génome, il serait à l'heure actuelle envisageable de séquencer 250 pb sur environ un gène sur 10 du mégaplasmine pSymA. Dans ce contexte, le patron de déséquilibre de liaison observé pourrait être comparé aux attendus théoriques correspondants aux différentes modalités de transfert horizontal envisageables pour notre système biologique. Les résultats obtenus pourraient orienter les recherches mécanistiques sur les déterminants qui pourraient focaliser notre attention sur les gènes impliqués dans le mécanisme de transfert inféré. Ainsi, si un mécanisme de type conjugal est inféré, il serait intéressant d'approfondir

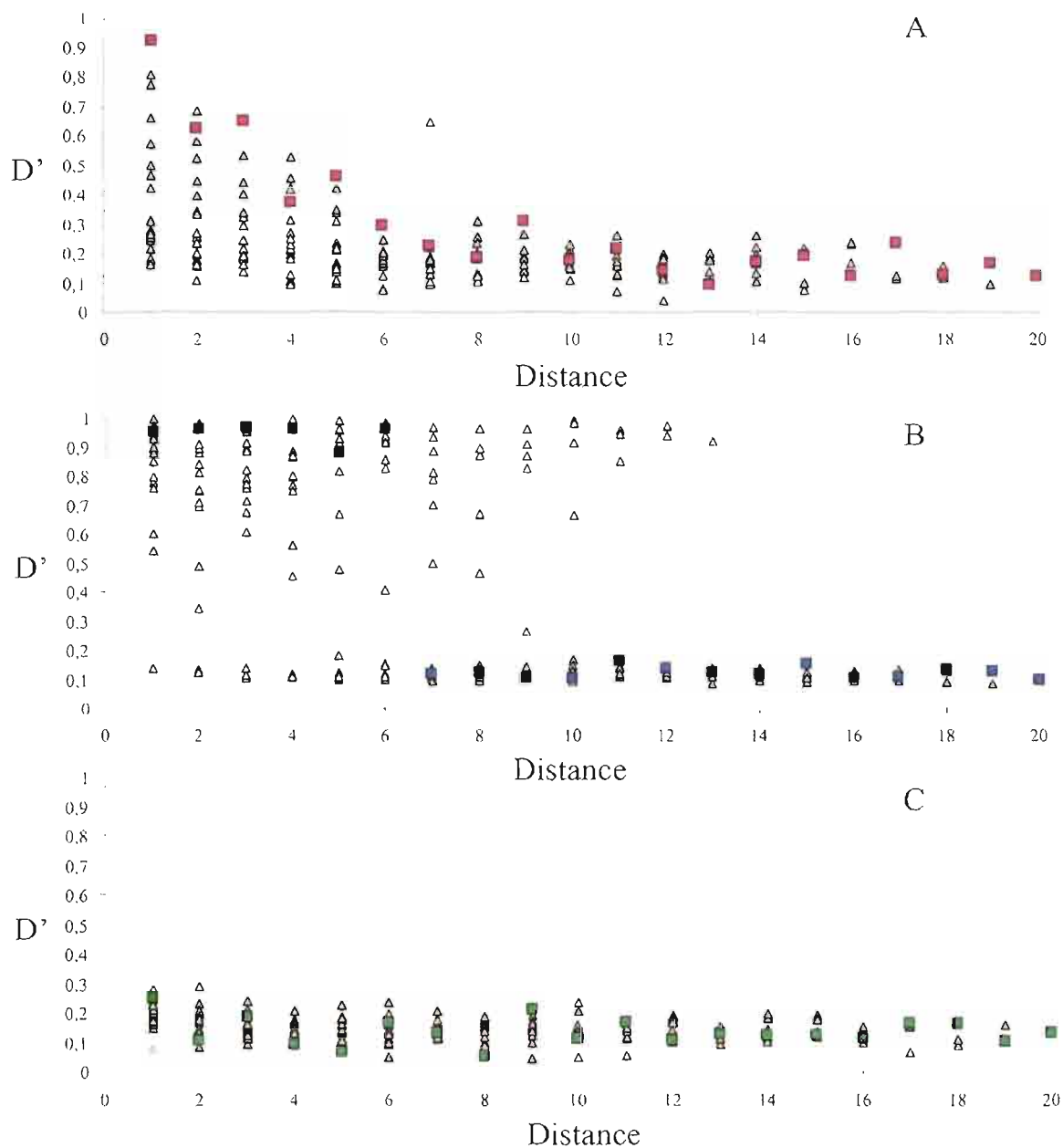


Figure 16. Simulation de la relation entre patrons de déséquilibre de liaison et distance physique pour une unité de réplication circulaire attendue sous trois modes de transfert horizontal différents. L'unité de réplication est balisée par 21 marqueurs équidistants. Pour chaque graphique, les triangles beiges correspondent à l'ensemble des valeurs de déséquilibre de liaison attendues, les carrés correspondent aux mesures d'association entre l'origine de transfert (*oriT*) de l'unité de réplication et les vingt autres marqueurs. A : cas de transferts conjugatifs classiques. Des fragments débutant à l'*oriT*, dont les tailles sont tirées dans une loi normale, sont échangés entre bactéries. B : cas de transferts conjugatifs d'élément intégratif. Des fragments de taille constante débutant à l'*oriT* sont échangés entre bactéries. C : cas de transfert par transformation. Des fragments débutant à une position aléatoire, dont les tailles sont tirées dans une loi normale, sont échangés entre bactéries. Il est notable que pour une même quantité moyenne d'ADN recombinée pour chaque transfert, un taux de recombinaison plus important est nécessaire pour casser le déséquilibre de liaison lié à la clonalité avec le mécanisme de transformation.

les résultats obtenus par Pretorius-Guth *et al.* (1990) en mesurant des taux de recombinaison entre souches de *Sinorhizobium* mutée au niveau des différents gènes *tra* annotés dans le génome de *S. meliloti* souche 1021.

IX.2. Influence des plantes hôtes du genre *Medicago* sur la diversification des *Sinorhizobium* *sp.*

Le chapitre concernant l'impact des plantes hôtes sur la diversité des bactéries symbiotiques associées nous a permis de mettre en exergue la complexité de ce problème qui, au-delà de la spécificité d'un dialogue moléculaire, pourrait impliquer une spécificité physiologique.

IX.2.1. Evolution expérimentale

Dans ce contexte, des expériences d'évolution expérimentales seraient intéressantes. L'objectif de telles manipulations pourrait être de savoir si la mise en contact récurrente avec un hôte particulier influence l'évolution phénotypique d'une population de *rhizobium*. Cette évolution pourrait se traduire à la fois par une adaptation à l'interaction avec le génotype de plante utilisé lors de l'expérience et par une modification du phénotype observable lors de l'interaction avec d'autres plantes hôtes. Ainsi, un plan d'expérience basé sur des inoculations indépendantes de différentes souches bactériennes sur plusieurs génotypes de plantes hôtes pourrait être réalisé. Chaque interaction génotype bactérien/génotype végétal pourrait être reconduite par broyage de nodosités et inoculation de la solution obtenue à de nouvelles germination présentant le même génotype. Au terme de l'expérience, il serait alors possible :

- i) de comparer le phénotype des souches bactériennes initiales à celui des populations bactériennes sélectionnées ;
- ii) de comparer le phénotype des populations bactériennes sélectionnées sur différentes plantes hôtes en opérant des inoculations croisées.

Afin de mesurer des différences phénotypiques entre traitements, trois caractères principaux pourraient être suivis :

- la valeur sélective globale de chaque population bactérienne inoculée pourrait être obtenue par comptage du nombre de bactéries formant colonie au sein du système racinaire de chaque type de plante hôte.

- le niveau d'investissement des populations bactériennes dans les fonctions symbiotiques pourrait être estimé en observant par PCR quantitative le ratio du nombres de transcrits des locus *nifH* et le nombre de copies d'ADNr 16S.

- la valeur sélective des plantes hôtes en interaction avec les différentes populations bactériennes pourrait être obtenue par des mesures directes (nombre de graines) ou indirectes (poids des parties aériennes).

Les données obtenues devraient aboutir à une meilleure connaissance du potentiel évolutif de ces caractères et à une meilleure compréhension de l'effet des plantes hôtes sur la diversité fonctionnelle des *rhizobium*.

IX.2.1. Modélisation

En outre, le développement d'une approche théorique pourrait permettre la mise en évidence de paramètres clefs, qui seraient susceptible d'induire les patrons de diversité observés dans les différents groupes génétiques de *Sinorhizobium* qui sont associées aux espèces de *Medicago*. Jusqu'à aujourd'hui, l'interaction entre *rhizobium* et légumineuse a été étudié par le biais d'approches théoriques pour aborder deux problèmes principaux :

- l'émergence et le maintien d'une interaction mutualiste entre les deux partenaires. Les modèles développés pour répondre à cette question se basent sur l'hypothèse d'un apparemment fort entre les bactéries interagissant avec les plantes hôtes et celles qui sont localisées à l'extérieur des nodules (Olivieri and Frank 1994; Bever and Simms 2000) ou sur l'hypothèse de sanctions appliquées par les plantes à l'encontre des *rhizobium* les moins coopératifs (West et al. 2002a; West et al. 2002b). Des expériences ont démontré par ailleurs que les plantes ont en effet la capacité de sanctionner les *rhizobium* les moins coopératifs (Kiers et al. 2003; Simms et al. 2006), ou d'influencer la différenciation des *rhizobium* en bactéroïdes (Mergaert et al. 2006).

- le maintien d'un polymorphisme de spécificité dans des métapopulations de plantes et de bactéries. En particulier, Parker (1999) a montré que moyennant certaines hypothèses concernant la répartition initiale des phénotypes symbiotiques, le taux de migration des deux partenaires ou l'efficacité symbiotique de différents types bactériens, il était possible de maintenir un polymorphisme de spécificité de type spécialiste/généraliste chez des couples symbiotiques.

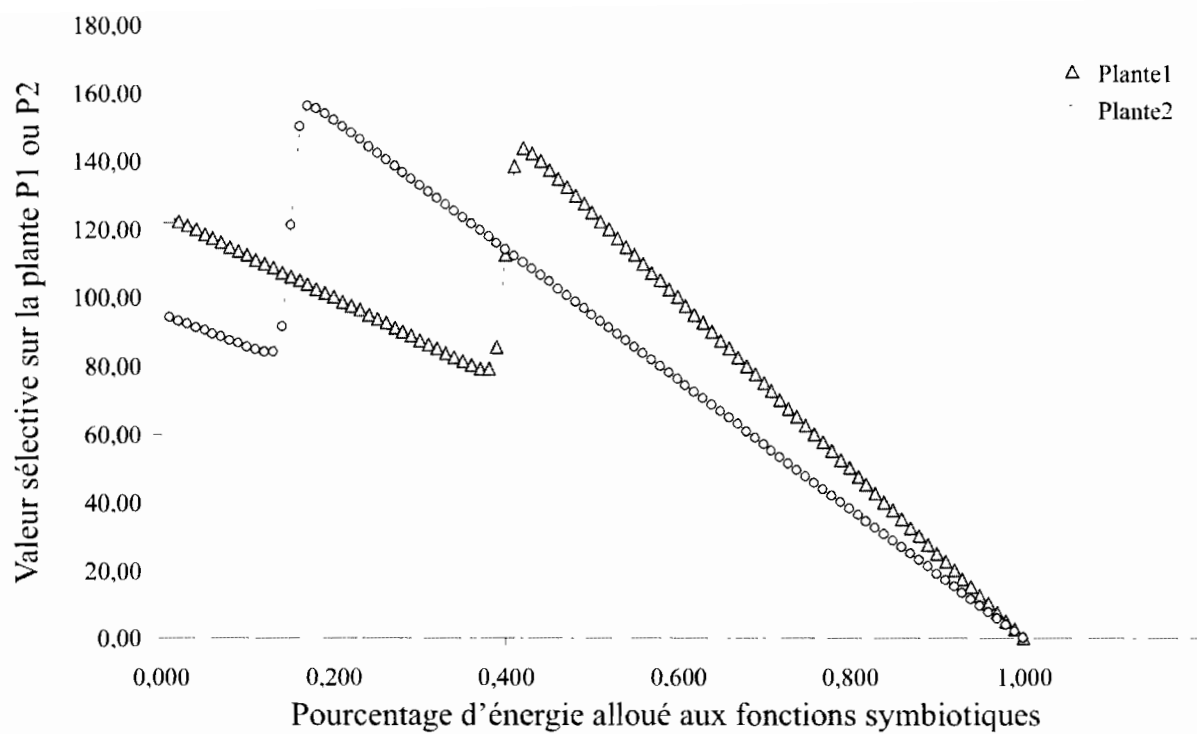


Figure 17. Modélisation de l'effet de deux stratégies de sanction différentes (plantes 1 et 2) sur la valeur sélective des symbiotes associés, en fonction de la part d'énergie qu'ils allouent à la fixation d'azote symbiotique établie à partir de l'équation 1.

En complément, il serait intéressant d'étudier l'impact des relations entre l'hôte et son symbiote pour comprendre comment cette interaction influence la diversification des bactéries symbiotiques. L'intérêt pour ce type d'approche est souligné par différents résultats. D'une part, Doeblé et Dieckmann (2000) ont montré que l'évolution d'un caractère impliqué dans l'interaction pour chaque espèce mutualiste pouvait déboucher sur des spéciations sympatriques parallèles chez les deux partenaires. Dans ce contexte, la stabilité des branchements observés serait liée au rapport des bénéfices mutualistes récupéré par chaque partenaire. D'autre part, Porcher et *al.* (2001) ont montré par une approche de simulation que la complexité métabolique pouvait être à l'origine de l'émergence de polymorphismes dans les populations bactériennes.

Ainsi, un modèle pourrait être développé afin d'étudier i) les processus de spécialisation entre groupes bactériens, ii) l'évolution du sexe chez les formes spécialisées et par la même la spéciation l'influence et iii) la possibilité d'une co-évolution entre plantes et bactéries. Dans un premier temps, une approche de simulation préliminaire a été mise en place. Sur la base des *a priori* disponibles, une fonction décrivant la valeur sélective d'un rhizobium en fonction de la part d'énergie allouée aux fonctions symbiotiques et de caractéristiques de la plante hôte :

$$W(F) = \left[E \times \left(D + (1-D) \times \left(\frac{1}{1 + e^{I \times S - I \times F}} \right) \right) \right] \times (1-F) \quad \text{Equation 1}$$

où W est la valeur sélective des bactéries, E la quantité d'énergie que la plante peut investir dans la symbiose, D la part d'énergie donnée par la plante sans sanction (formation d'une nodosité), S un seuil d'énergie alloué par la bactérie aux fonctions symbiotiques au dessous duquel la plante applique une sanction, F le phénotype d'allocation de la bactérie focus.

Les simulations effectuées jusqu'alors utilisent deux types de plantes hôtes ($P1$ et $P2$), en fréquences constantes, présentant chacun des paramètres propres qui influencent la forme de $W(F)$ (*i.e.* variation de E , D , I et S) (Fig. 17). Les simulations sont alors obtenues comme suit : une gamme de 100 phénotypes d'allocation équiprobables est créée et la fréquence de rencontre de chaque phénotype face à chaque hôte est calculée sous l'hypothèse d'association aléatoire des *rhizobium* aux hôtes $P1$ ou $P2$. Ensuite, on calcule la fréquence du phénotype X ($f(X)$) parmi les $i = 100$ phénotypes au temps $T+1$ par :

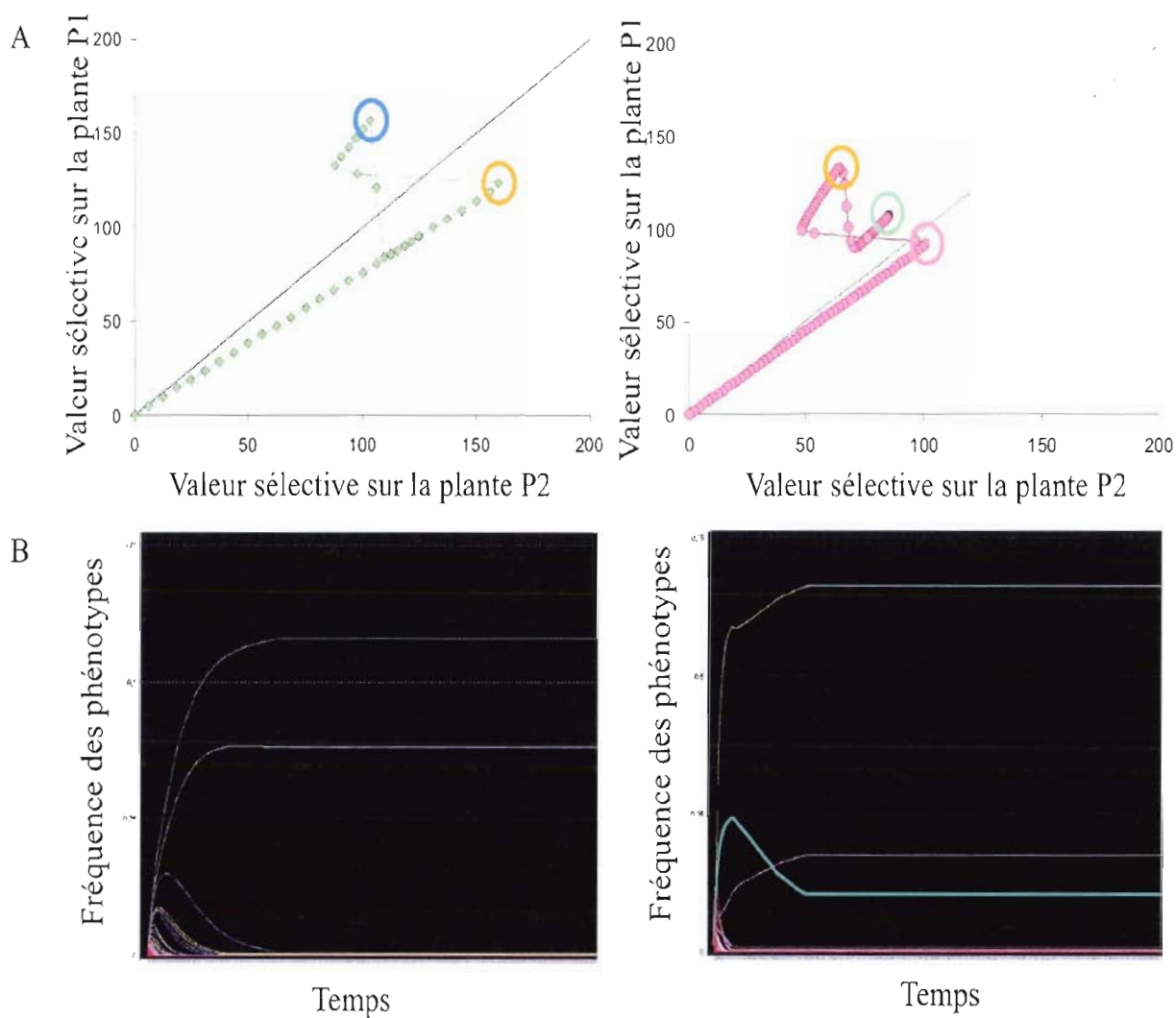


Figure 18. (A) Effet de deux modes de sanction pour deux plantes P1 et P2 sur la relation entre valeur sélective sur P1 et valeur sélective sur P2 des symbiotes bactériens en fonction du pourcentage d'énergie qu'ils allouent à la fixation d'azote symbiotique. (B) Résultats des simulations associées aux jeux de paramètres décrits dans la partie (A). Au sein de la partie A les phénotypes maintenus lors des simulations sont entourés par un cercle reprenant la couleur utilisée pour suivre son évolution de fréquence dans lors des simulations.

$$f_{i+1}(X) = f_i(X) \times \left(\frac{f(P1) \times W_{P1}(X)}{\sum_{i=1}^{100} W_{P1}(i) \times f_i(i)} \times \frac{f(P2) \times W_{P2}(X)}{\sum_{i=1}^{100} W_{P2}(i) \times f_i(i)} \right) \quad \text{Equation 2}$$

où $W_{PY}(i)$ est la valeur sélective du phénotype i sur la plante Y et $f(PY)$, la probabilité de rencontre de la plante Y . Il est notable qu'un modèle alternatif de régulation démographique inter-habitat ne permet pas de maintenir de polymorphisme dans les gammes de paramètre testées.

Les résultats obtenus suggèrent que sous certaines conditions, deux plantes hôtes caractérisées par des modes de sanction différents pouvaient être à l'origine d'une sélection disruptive conduisant à une coexistence stable de deux voire trois phénotypes caractérisés par des investissements symbiotiques différents (deux phénotypes fixateurs ou deux phénotypes fixateurs et un non-fixateur), et par-là même à une possibilité de spéciation sympatrique (Fig.18). Toutefois, deux problèmes se posent à ce niveau :

- Vérifier analytiquement si certains jeux de paramètres permettent effectivement de maintenir un polymorphisme à l'équilibre (si c'est possible...) et le cas échéant étudier les conditions qui permettent de conserver ce polymorphisme.

- Cette approche est fondée sur une fonction mathématique définie arbitrairement et décrivant la relation entre la valeur sélective d'une bactérie et le taux d'énergie qu'elle alloue à la fixation d'azote. Or, les résultats obtenus peuvent dépendre de certaines caractéristiques de cette fonction qui n'auraient pas de fondement biologique (notamment la possibilité de maintenir trois phénotypes en interaction avec deux hôtes). Il faudrait donc prendre en compte de manière structurée les interactions métaboliques afin de décrire de façon plus réaliste le mutualisme liant les bactéries fixatrices d'azote à leurs plantes hôtes (la rencontre, la formation de la nodosité, puis l'échange entre les partenaires et la sortie des nodules) afin de vérifier que l'existence d'un polymorphisme stable ne dépend pas de détails non biologiques liés à la fonction mathématique définie précédemment.

A terme, ce type de modèle pourrait évoluer afin de prendre en compte différents paramètres complémentaires :

- il serait intéressant d'étudier si des génotypes sexués seraient capables d'envahir les populations de *rhizobium*. Le compromis d'allocation aux fonctions symbiotiques devrait être un phénotype complexe lié à différents déterminants

génétiques. Si on suppose un tel déterminisme, ce phénotype peut-être modélisé comme la somme de différents caractères indépendants dont les effets sont additifs. Dans ces conditions, il devrait être possible de modéliser des mutants capables d'échanger/d'acquérir du matériel génétique contenant les déterminants aux caractères d'intérêt. De plus, il serait envisageable d'étudier l'évolution d'allèles permettant la mise en place d'un isolement reproducteur actif entre lignées spécialisées.

-il serait également intéressant d'étudier l'interaction entre la mise en place de systèmes de spécificité précoces (type Facteur-Nod) ou tardifs (type polysaccharides de surface) et l'évolution du phénotype d'allocation aux fonctions symbiotiques. Notamment, il serait intéressant de constater si des déterminants facilitant la reconnaissance par la plante hôte peuvent évoluer pour l'ensemble des phénotypes étudiés.

-enfin, il semble indispensable de prendre en compte la diversification fonctionnelle des plantes pour aboutir à un modèle de co-évolution. De ce fait, il serait intéressant de définir les conditions nécessaires à l'émergence de groupes de plantes hôtes présentant des mécanismes de sanction différents. Ces conditions pourraient notamment être liées à une structuration spatiale ou un polymorphisme dans la capacité à prélever de l'azote du sol.

IX.3. Biogéographie

Dans le cadre du (court) chapitre consacré à l'étude de la structure spatiale de la diversité génétique des souches de *Sinorhizobium* associées aux plantes du genre *Medicago*, nous avons mis en évidence que l'isolement géographique était un facteur important dans l'optique de la compréhension du maintien de la diversité génétique chez *S. meliloti* et *S. medicae*. De plus, nous avons montré que les populations de *S. meliloti* étaient structurées du fait de l'existence de différents groupes de spécificité au sein du genre *Medicago*. Dans ce cadre, une approche phylogéographique à l'échelle du bassin méditerranéen permettrait d'affiner nos connaissances sur les processus qui influencent la diversification des populations de *S. meliloti* et *S. medicae*. Ainsi, un échantillonnage a été initié afin de répondre à quatre questions :

-les souches de *S. meliloti* associées à *M. rigiduloides* constituent-elle un nouveau biovar ?

-la structure génétique des populations de *Sinorhizobium* associées aux *Medicago* est-elle corrélée à la distribution géographique des plantes hôtes appartenant aux différents groupes de spécificité ?

-est-il possible de déterminer à quelle échelle la distance géographique a une influence significative sur la structure des populations de *Sinorhizobium* symbiotiques ?

-quels sont les facteurs clefs qui permettent le maintien de différents biovars chez *S. meliloti* ?

Dans ce contexte, l'échantillonnage utilisé doit permettre d'analyser les relations génétiques entre les écotypes bactériens associés à différentes plantes du genre *Medicago*. Le piégeage *ex-situ* de bactéries symbiotiques a été envisagé de la façon suivante. Il est effectué à l'aide de quatre espèces de plantes hôtes qui présentent des spectres d'hôtes et des aires de répartition différents : *M. truncatula*, *M. laciniata*, *M. rigiduloides*, et *Medicago carstiensis*. Les plantes sont cultivées individuellement et sont mises en contact avec différents échantillons de sol provenant de l'ensemble du bassin méditerranéen. Au moins cinq plantes de chaque espèce sont utilisées pour chaque sol. Des plantes témoins de chaque espèce sont cultivées sur un substrat stérilisé. Après deux mois, les poids secs des parties aériennes des plantes sont mesurés afin de corréliser le génotype des bactéries à l'efficacité de l'interaction qu'elles ont initiée avec la plante utilisée pour le piégeage et dans le but de limiter le nombre de populations bactériennes à étudier. Les nodosités sont récoltées sur chaque plante, quatre d'entre elles sont stérilisées, puis la souche bactérienne présente à l'intérieur de chaque nodosité sera isolée. Le génotypage de l'ensemble des souches est envisagé par séquençage ou typage SNP (ça coûte cher de faire du séquençage...) sur la base de quatre locus, un marqueur chromosomique (*IGS_{RKP}*) et trois marqueurs plasmidiques (*IGS_{EXO}*, *IGS_{NOD}*, *IGS_{GAB}*).

Jusqu'à présent, un piégeage a été effectué sur dix échantillons de sols provenant du Portugal, d'Italie, de Bulgarie, du Liban et de Tunisie. Une série de piégeages supplémentaires est envisagée à l'aide de sols provenant du Maroc, de France, du Caucase, de Turquie, de Libye, de Grèce et d'Israël.

Les résultats déjà acquis indiquent pour *Medicago rigiduloides* et *Medicago truncatula* qu'un certain nombre de sols sur lesquels les plantes cultivées présentent des moyennes de poids sec des parties aériennes significativement supérieures aux valeurs obtenues pour le

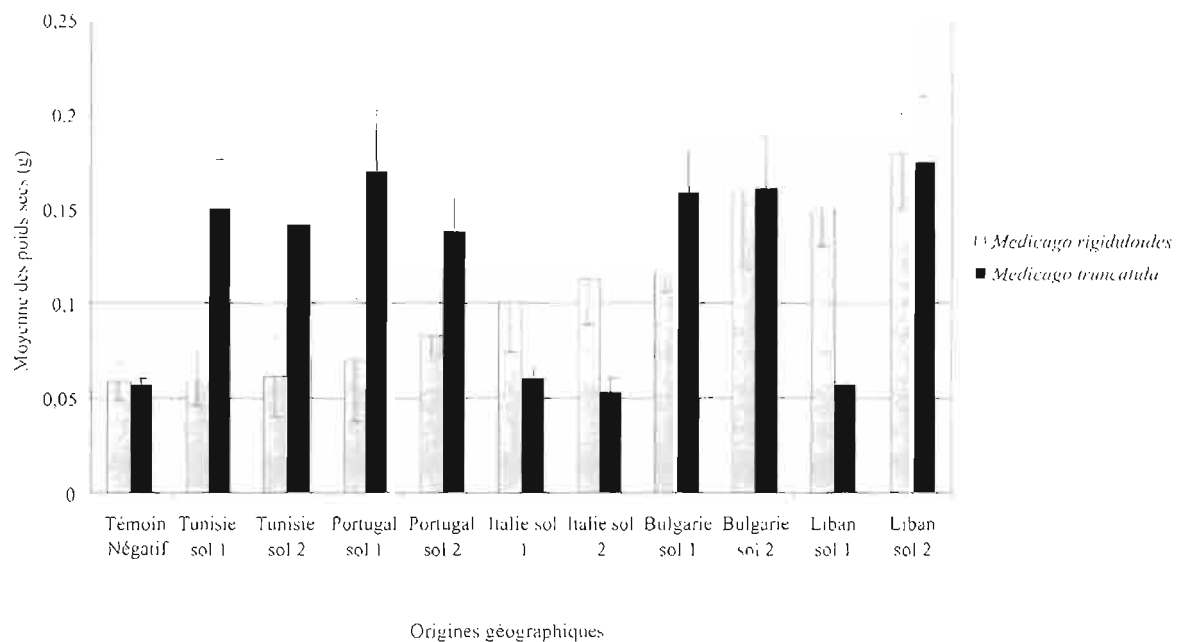


Figure 19 Moyennes et variances des masses sèches des parties aériennes des plantes hôtes utilisées dans le cadre du piégeage pour l'expérience de phylogéographie pour les espèces *M. truncatula* et *M. rigiduloides*. Les données sont classées en fonction de l'origine des sols.

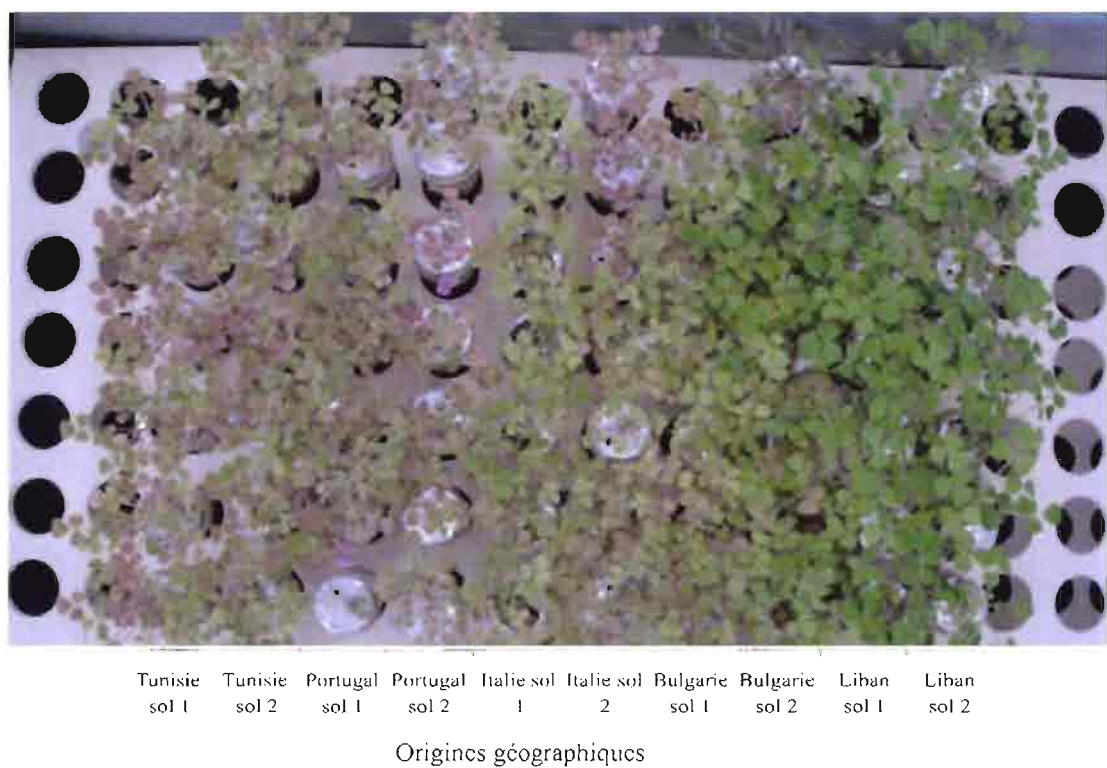


Figure 20. Photographie illustrant le piégeage de bactéries symbiotiques avec *Medicago rigiduloides* concernant l'expérience de phylogéographie.

témoin négatif (sans bactéries symbiotiques) (Fig. 19). Pour ces deux espèces, les lots de plantes qui ne présentent pas de différence significative avec le témoin ont montré une nodulation inefficace (nodosités de petite taille de coloration blanche). La répartition géographique des sols contenant les populations de bactéries symbiotiques efficaces est différente pour les deux espèces végétales :

- dans le cas de *Medicago rigiduloides*, seuls les sols bulgares et libanais contenaient une population symbiotique efficace ayant abouti à une augmentation significative du poids sec des parties aériennes des plantes hôtes. Cette observation, illustrée par le phénotype des plantes présenté dans la figure 20, peut être mise en relation avec l'aire de répartition naturelle de l'espèce hôte, située à l'est du bassin méditerranéen. Par ailleurs, nous avons observé des différences flagrantes dans le nombre et la forme des nodules obtenus sur les plantes cultivées sur les sols bulgares et libanais.

- dans le cas de *Medicago truncatula*, contrairement à ce que nous attendions au regard de l'aire de répartition de l'espèce, certains sols (les deux sols italiens et un sol libanais) ne contenaient pas une population bactérienne symbiotique efficace suffisante pour induire une augmentation des poids secs des parties aériennes des plantes hôtes par le biais d'une symbiose efficace. Cette absence de nodulation efficace n'est pas obligatoirement liée à l'absence totale de bactéries capables d'induire une symbiose efficace avec des plantes du genre *Medicago* comme en attestent les résultats obtenus sur *Medicago rigiduloides* pour les deux sols libanais.

Dans le cas de *Medicago carstiensis*, les tests d'inoculation et de piégeage de *rhizobium* que nous avons effectués jusqu'à présent n'ont pas permis de mettre en évidence de bactéries capables d'induire une symbiose efficace avec cette espèce. En effet, lors des expériences de piégeage, nous avons obtenu sur l'ensemble des sols une nodulation inefficace (nodosités de petite taille de coloration blanche) les différences avec les témoins étant non significatives. Un nouvel échantillonnage de sol a été effectué de manière à obtenir des populations de bactéries adaptées à cette plante hôte.

Enfin, dans le cas de *Medicago laciniata*, nous n'avons obtenu aucune nodulation sur la base des sols utilisés jusqu'à présent. Cependant nous disposons d'autres échantillons de sols tunisiens ou les symbiotes associés à cette plante ont été piégés (comme précisé dans l'introduction) et des échantillons de sols marocains, algériens et libyens qui devraient nous permettre, à terme de compléter notre échantillonnage de bactéries.

L'ensemble des données acquises suggèrent que la diversité fonctionnelle des bactéries associées aux plantes du genre *Medicago* est structurée spatialement, ceci au niveau inter-écotypes (différences de patrons de présence/absence des populations symbiotiques associées aux différentes plantes hôtes) et intra-écotype (différences dans le phénotype de nodulation des plantes de l'espèce *Medicago rigiduloides*). A l'heure actuelle, le piégeage a repris sur les sols complémentaires et la caractérisation génétique des populations bactériennes sera entamée prochainement. Gageons que cette expérience apportera des informations intéressantes...

X. Conclusion générale

X. Conclusion générale

Durant cette thèse, nous avons abordé plusieurs points touchant à l'évolution de la diversité des populations de *S. meliloti* et *S. medicae*. Du point de vue des transferts horizontaux de matériel génétique, nous avons essayé d'apporter des précisions sur l'impact de ce mécanisme sur la diversité des populations observées. Nos données semblent indiquer que les flux de gènes sont structurés en fonction de la localisation sur le génome des régions recombinantes et qu'ils jouent un rôle fondamental dans le potentiel évolutif des groupes bactériens étudiés. Les travaux effectués sur l'influence des populations de plantes hôtes sur la diversité des *Sinorhizobium* reflètent la complexité du mécanisme étudié : (i) La coexistence au sein du génome des souches de *Sinorhizobium* de patrons de sélection attendus dans le cadre de relations mutualistes et pathogènes souligne la diversité des pressions de sélection qui influencent la diversification de nos bactéries modèles. Il semble que cette diversité de pression soit en rapport avec des conflits d'intérêts entre les composantes du génome bactérien et/ou entre les partenaires symbiotiques (ii) L'observation concomitante d'espèces et de biovars de *Sinorhizobium* associés de manière différentielle aux plantes du genre *Medicago* pose la question du lien entre l'architecture génétique des déterminants de l'adaptation et le mode de diversification des génomes. Enfin, la mise en évidence d'une structure géographique des populations de *Sinorhizobium* associés aux plantes du genre *Medicago* ouvre des perspectives intéressantes tant d'un point de vue fondamental (influence de la répartition des plantes hôtes sur la diversité génétique des bactéries, estimation des taux de migration bactériens,...) que pratique (mise en évidence de génotypes adaptés localement, gestion des pratiques d'inoculation). Dans un cadre plus général, l'étude de la diversité et de l'évolution des bactéries est actuellement en train de vivre une révolution. L'augmentation exponentielle des capacités de séquençage permet d'envisager que dans quelques années, l'étude de l'évolution du patrimoine génétique de populations de bactéries appartenant à des espèces modèles se fasse à l'échelle des génomes (voir Margulies *et al.* (2005) pour voir comment une personne peut séquencer un génome bactérien en une journée à l'aide d'une seule expérience). En parallèle, les progrès réalisés sur les outils de caractérisation phénotypiques à haut débit (puces d'expression ou PCR quantitative pour des cas plus ciblés) devrait permettre une étude plus fine des relations entre phénotypes et génotypes. La capacité d'associer ces deux types de données devrait faire des bactéries des modèles de choix pour

des études de biologie intégrative. On pourra en effet envisager d'étudier en parallèle l'évolution des fonctions, des régulations et des organismes. Si l'étude de ces processus représente un enjeu majeur dans le cadre de bactéries pathogènes, certaines bactéries mutualistes d'intérêt agronomique, telles que les *rhizobium*, pourraient bénéficier de tels programmes de recherche. En effet, les perspectives de recherche sur *S. meliloti* et *S. medicae* s'intègrent dans cette démarche intégrative, tant au niveau de l'étude des mécanismes de transferts horizontaux de matériel génétique qu'au niveau de l'évolution du spectre d'hôte. On peut d'ailleurs souligner que les données préliminaires du séquençage d'une souche de *S. medicae* devraient être disponibles d'ici la fin de l'année et que des chercheurs de l'USDA (B. Eardly et P. Van Berkum) ont entrepris des études de MLST sur *S. meliloti* et *S. medicae* avec un nombre important de marqueurs (24 gènes dont 10 sur le chromosome, 7 sur pSymB et 7 sur psymA). Je ne rajouterai qu'une chose : longue vie au modèle *Sinorhizobium/Medicago*...

XI. Références

XI. Références

- Bandelt, H. J., and Dress, A. W. 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol Phylogenet Evol* **1**:242-252.
- Barran, L. R., Bromfield, E. S., and Brown, D. C. 2002. Identification and cloning of the bacterial nodulation specificity gene in the *Sinorhizobium meliloti*--*Medicago laciniata* symbiosis. *Can J Microbiol* **48**:765-771.
- Battisti, L., Lara, J., and Leigh, J. 1992. Specific oligosaccharide form of the *Rhizobium meliloti* exopolysaccharide promotes nodule invasion in alfalfa. *Proc Natl Acad Sci U S A* **89**:5625-5629.
- Bena, G., Lyet, A., Huguet, T., and Olivieri, I. 2005. *Medicago* - *Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. *J Evol Biol* **18**:1547-1558.
- Berg, O. G., and Kurland, C. G. 2002. Evolution of microbial genomes: sequence acquisition and loss. *Mol Biol Evol* **19**:2265-2276.
- Bever, J. D., and Simms, E. L. 2000. Evolution of nitrogen fixation in spatially structured populations of *Rhizobium*. *Heredity* **85**:366-372.
- Biondi, E. G. Pilli, E. Giuntini, E. et al. 2003. Genetic relationship of *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains isolated from Caucasian region. *FEMS Microbiol Lett* **220**:207-213.
- Brisson, D., and Dykhuizen, D. E. 2004. *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics* **168**:713-722.
- Bromham, L., and Penny, D. 2003. The modern molecular clock. *Nat Rev Genet* **4**:216-224.
- Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. 2002. Conjugative transposons: the tip of the iceberg. *Mol Microbiol* **46**:601-610.
- Carelli, M., Gnocchi, S., Fancelli, S., Mengoni, A., Paffetti, D., Scotti, C., and Bazzicalupo, M. 2000. Genetic diversity and dynamics of *Sinorhizobium meliloti* populations nodulating different alfalfa cultivars in Italian soils. *Appl Environ Microbiol* **66**:4785-4789.
- Casida Jr, L. E. 1982. *Ensifer adhaerens* gen. nov., sp. nov.: a bacterial predator of bacteria in soil. *Int J Syst Bacteriol* **32**:339-345.
- Casjens, S. 1998. The diverse and dynamic structure of bacterial genomes. *Annu Rev Genet* **32**:339-377.

- Chen, W. M., Moulin, L., Bontemps, C., Vandamme, P., Bena, G., and Boivin-Masson, C. 2003. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. *J Bacteriol* **185**:7266-7272.
- Chen, W. X., Yan, G. H., and Li, J. L. 1988. Numerical taxonomic study of fast-growing soybean rhizobia and a proposal that *Rhizobium fredii* be assigned to *Sinorhizobium* gen. nov. *Int J Syst Bacteriol* **38**:392-397.
- Cohan, F. M. 1994a. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am Nat* **143**:965-986.
- Cohan, F. M. 1994b. Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol Evol* **9**:175-180.
- Cohan, F. M. 1995. Does recombination constrain neutral divergence among bacterial taxa? *Evolution Int J Org Evolution* **49**:164-175.
- Cohan, F. M. 2001. Bacterial species and speciation. *Syst Biol* **50**:513-524.
- Cohan, F. M. 2002. What are bacterial species? *Annu Rev Microbiol* **56**:457-487.
- Cook, D. R. 1999. *Medicago truncatula*--a model in the making! *Curr Opin Plant Biol* **2**:301-304.
- Davison, J. 1999. Genetic exchange between bacteria in the environment. *Plasmid* **42**:73-91.
- De Gelder, L., Vandecasteele, F. P., Brown, C. J., Forney, L. J., and Top, E. M. 2005. Plasmid donor affects host range of promiscuous IncP-1beta plasmid pB10 in an activated-sludge microbial community. *Appl Environ Microbiol* **71**:5309-5317.
- de Lajudie, P., Willems, A., Pot, B., Dewettinck, D., Maestrojuan, G., Neyra, M., Collins, M. D., Dreyfus, B., Kersters, K., and Gillis, M. 1994. Polyphasic taxonomy of rhizobia: emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov. *Int J Syst Bacteriol* **44**:715-733.
- Depaulis, F., Mousset, S., and Veuille, M. 2001. Haplotype tests using coalescent simulations conditional on the number of segregating sites. *Mol Biol Evol* **18**:1136-1138.
- Doebeli, M., and Dieckmann, U. 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Am Nat* **156**:S77-S101.
- Downie, S. R., Katz-Downie, D., Rogers, E. J., Zujewski, H. L., and Small, E. 1998. Multiple independent losses of the plastid *rpoCI* intron in *Medicago* (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. *Can J Bot* **76**:791-803.

- Dykhuizen, D. E., and Baranton, G. 2001. *Borrelia burgdorferi*: a (somewhat) clonal bacterial species. Trends Microbiol **9**:472.
- Eardly, B. D., Materon, L. A., Smith, N. H., Johnson, D. A., Rumbaugh, M. D., and Selander, R. K. 1990. Genetic structure of natural populations of the nitrogen-fixing bacterium *Rhizobium meliloti*. Appl Environ Microbiol **56**:187-194.
- Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. Theor Popul Biol **3**:87-112.
- Flores, M., Mavingui, P., Perret, X., Broughton, W. J., Romero, D., Hernandez, G., Davila, G., and Palacios, R. 2000. Prediction, identification, and artificial selection of DNA rearrangements in *Rhizobium*: toward a natural genomic design. Proc Natl Acad Sci U S A **97**:9138-9143.
- Frayse, N., Couderc, F., and Poinso, V. 2003. Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. Eur J Biochem **270**:1365-1380.
- Galibert, F., Finan, T. M., Long, S. R. et al. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science **293**:668-672.
- Galtier, N., Depaulis, F., and Barton, N. H. 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics **155**:981-987.
- Gevers, D., Cohan, F. M., Lawrence, J. G. et al. 2005. Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol **3**:733-739.
- Giuntini, E., Mengoni, A., De Filippo, C., Cavalieri, D., Aubin-Horth, N., Landry, C. R., Becker, A., and Bazzicalupo, M. 2005. Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. BMC Genomics **6**:158.
- Godoy, D., Randle, G., Simpson, A. J., Aanensen, D. M., Pitt, T. L., Kinoshita, R., and Spratt, B. G. 2003. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. J Clin Microbiol **41**:2068-2079.
- Gonzalez, J. E., Reuhs, B. L., and Walker, G. C. 1996a. Low molecular weight EPS II of *Rhizobium meliloti* allows nodule invasion in *Medicago sativa*. Proc Natl Acad Sci U S A **93**:8636-8641.
- Gonzalez, J. E., York, G. M., and Walker, G. C. 1996b. *Rhizobium meliloti* exopolysaccharides: synthesis and symbiotic function. Gene **179**:141-146.
- Grohmann, E., Muth, G., and Espinosa, M. 2003. Conjugative plasmid transfer in gram-positive bacteria. Microbiol Mol Biol Rev **67**:277-301.

- Guo, X., Flores, M., Mavingui, P., Fuentes, S. I., Hernandez, G., Davila, G., and Palacios, R. 2003. Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures. *Genome Res* **13**:1810-1817.
- Hanage, W. P., Fraser, C., and Spratt, B. G. 2006. The impact of homologous recombination on the generation of diversity in bacteria. *J Theor Biol* **239**:210-219.
- Haubold, B., Travisano, M., Rainey, P. B., and Hudson, R. R. 1998. Detecting linkage disequilibrium in bacterial populations. *Genetics* **150**:1341-1348.
- Hedrick, P. W. 1987. Gametic disequilibrium measures: proceed with caution. *Genetics* **117**:331-341.
- Herrera-Cervera, J. A., Sanjuan-Pinilla, J. M., Olivares, J., and Sanjuan, J. 1998. Cloning and identification of conjugative transfer origins in the *Rhizobium meliloti* genome. *J Bacteriol* **180**:4583-4590.
- Hill, W. G., and Robertson, A. 1966. The effect of linkage on limits to artificial selection. *Genet Res* **8**:269-294.
- Hudson, R. R. 1994. Analytical results concerning linkage disequilibrium in models with genetic transformation and conjugation. *J Evol Biol* **7**:535-548.
- Hudson, R. R., Kreitman, M., and Aguade, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**:153-159.
- Hughes, A. L. 2005. Evidence for abundant slightly deleterious polymorphisms in bacterial populations. *Genetics* **169**:533-538.
- Jiggins, F. M., Hurst, G. D., and Yang, Z. 2002. Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. *Mol Biol Evol* **19**:1341-1349.
- Kereszt, A., Kiss, E., and Reuhs, B. L. 1998. Novel *rkp* gene clusters of *Sinorhizobium meliloti* involved in capsular polysaccharide production and invasion of the symbiotic nodule: the *rkpK* gene encodes a UDP-glucose dehydrogenase. *J Bacteriol* **180**:5426-5431.
- Kiers, E. T., Rousseau, R. A., West, S. A., and Denison, R. F. 2003. Host sanctions and the legume-rhizobium mutualism. *Nature* **425**:78-81.
- Kiss, E., Huguet, T., Poinot, V., and Batut, J. 2004. The *typA* gene is required for stress adaptation as well as for symbiosis of *Sinorhizobium meliloti* 1021 with certain *Medicago truncatula* lines. *Mol Plant Microbe Interact* **17**:235-244.
- Konstantinidis, K. T., and Tiedje, J. M. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc Natl Acad Sci U S A* **101**:3160-3165.

- Lawrence, J. G. 2002. Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61**:449-460.
- Lio, P. 2002. Investigating the relationship between genome structure, composition, and ecology in prokaryotes. *Mol Biol Evol* **19**:789-800.
- Lodwig, E. M., Hosie, A. H., Bourdes, A., Findlay, K., Allaway, D., Karunakaran, R., Downie, J. A., and Poole, P. S. 2003. Amino-acid cycling drives nitrogen fixation in the legume-Rhizobium symbiosis. *Nature* **422**:722-726.
- Lyon, G. J., and Novick, R. P. 2004. Peptide signaling in *Staphylococcus aureus* and other Gram-positive bacteria. *Peptides* **25**:1389-1403.
- Maiden, M. C., Bygraves, J. A., Feil, E. et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**:3140-3145.
- Majewski, J. 2001. Sexual isolation in bacteria. *FEMS Microbiol Lett* **199**:161-169.
- Majewski, J., and Cohan, F. M. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* **148**:13-18.
- Majewski, J., and Cohan, F. M. 1999a. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* **152**:1459-1474.
- Majewski, J., and Cohan, F. M. 1999b. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* **153**:1525-1533.
- Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F. M., and Dowson, C. G. 2000. Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *J Bacteriol* **182**:1016-1023.
- Margulies, M., Egholm, M., Altman, W. E. et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376-380.
- Materon, L. A. 1991. Symbiotic characteristics of *Rhizobium meliloti* in west asian soils. *Soil Biol Biochem* **23**:429-434.
- Maynard Smith, J. 1990. The evolution of prokaryotes: does sex matter? *Annu Rev Ecol Syst* **21**:1-13.
- Maynard Smith, J. 1993. The role of sex in bacterial evolution. *J Hered* **84**:326-327.
- Maynard Smith, J., Dowson, C. G., and Spratt, B. G. 1991. Localized sex in bacteria. *Nature* **349**:29-31.
- Maynard Smith, J., and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* **23**:23-35.

- Maynard Smith, J., Smith, N. H., O'Rourke, M., and Spratt, B. G. 1993. How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**:4384-4388.
- Mazodier, P., and Davies, J. 1991. Gene transfer between distantly related bacteria. *Annu Rev Genet* **25**:147-171.
- McAdams, H. H., Srinivasan, B., and Arkin, A. P. 2004. The evolution of genetic regulatory systems in bacteria. *Nat Rev Genet* **5**:169-178.
- McDonald, J. H., and Kreitman, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**:652-654.
- Mergaert, P., Uchiumi, T., Alunni, B. et al. 2006. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc Natl Acad Sci U S A* **103**:5230-5235.
- Mithofer, A. 2002. Suppression of plant defence in rhizobia-legume symbiosis. *Trends Plant Sci* **7**:440-444.
- Mitra, R. M., and Long, S. R. 2004. Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the *Medicago truncatula*/*Sinorhizobium meliloti* symbiosis. *Plant Physiol* **134**:595-604.
- Moran, N. A., and Wernegreen, J. J. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol* **15**:321-326.
- Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C. 2001. Nodulation of legumes by members of the beta-subclass of Proteobacteria. *Nature* **411**:948-950.
- Ochman, H. 2002. Bacterial evolution: chromosome arithmetic and geometry. *Curr Biol* **12**:R427-428.
- Ochman, H., Lawrence, J. G., and Groisman, E. A. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**:299-304.
- Ochman, H., Lerat, E., and Daubin, V. 2005. Examining bacterial species under the specter of gene transfer and exchange. *Proc Natl Acad Sci U S A* **102 Suppl 1**:6595-6599.
- Oger, P., and Farrand, S. K. 2002. Two opines control conjugal transfer of an *Agrobacterium* plasmid by regulating expression of separate copies of the quorum-sensing activator gene *traR*. *J Bacteriol* **184**:1121-1131.
- Ohta, T. 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proc Natl Acad Sci U S A* **79**:1940-1944.
- Olivieri, I., and Frank, S. A. 1994. The evolution of nodulation in *Rhizobium*: altruism in the rhizosphere. *J Hered* **85**:46-47.

- Otto, S. P., and Lenormand, T. 2002. Resolving the paradox of sex and recombination. *Nat Rev Genet* **3**:252-261.
- Paffetti, D., Daguin, F., Fancelli, S., Gnocchi, S., Lippi, F., Scotti, C., and Bazzicalupo, M. 1998. Influence of plant genotype on the selection of nodulating *Sinorhizobium meliloti* strains by *Medicago sativa*. *Antonie Van Leeuwenhoek* **73**:3-8.
- Paffetti, D., Scotti, C., Gnocchi, S., Fancelli, S., and Bazzicalupo, M. 1996. Genetic diversity of an Italian *Rhizobium meliloti* population from different *Medicago sativa* varieties. *Appl Environ Microbiol* **62**:2279-2285.
- Parker, M. A. 1999. Mutualism in metapopulations of legumes and rhizobia. *Am Nat* **153**:S48-S60.
- Pavoine, S., Dufour, A. B., and Chessel, D. 2004. From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. *J Theor Biol* **228**:523-537.
- Perez-Mendoza, D., Sepulveda, E., Pando, V. et al. 2005. Identification of the *rctA* gene, which is required for repression of conjugative transfer of rhizobial symbiotic megaplasmids. *J Bacteriol* **187**:7341-7350.
- Perret, X., Staehelin, C., and Broughton, W. J. 2000. Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol Rev* **64**:180-201.
- Pirnay, J. P., De Vos, D., Cochez, C., Bilocq, F., Vanderkelen, A., Zizi, M., Ghysels, B., and Cornelis, P. 2002. *Pseudomonas aeruginosa* displays an epidemic population structure. *Environ Microbiol* **4**:898-911.
- Porcher, E., Tenaillon, O., and Godelle, B. 2001. From metabolism to polymorphism in bacterial populations: a theoretical study. *Evolution Int J Org Evolution* **55**:2181-2193.
- Posada, D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol* **19**:708-717.
- Posada, D., and Crandall, K. A. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* **98**:13757-13762.
- Posada, D., and Crandall, K. A. 2002. The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* **54**:396-402.
- Pretorius-Guth, I. M., Puhler, A., and Simon, R. 1990. Conjugal transfer of magaplasmid 2 between *Rhizobium meliloti* strains in alfalfa nodules. *Appl Environ Microbiol* **56**:2354-2359.

- Qiu, W. G., Schutzer, S. E., Bruno, J. F., Attie, O., Xu, Y., Dunn, J. J., Fraser, C. M., Casjens, S. R., and Luft, B. J. 2004. Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci U S A* **101**:14150-14155.
- Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* **5**:94-100.
- Reuber, T. L., and Walker, G. C. 1993. Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Cell* **74**:269-280.
- Rocha, E. P. C., Smith, J. M., Hurst, L. D., Holden, M. T. G., Cooper, J. E., Smith, N. H., and Feil, E. J. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol* **239**:226-235.
- Rogel, M. A., Hernandez-Lucas, I., Kuykendall, L. D., Balkwill, D. L., and Martinez-Romero, E. 2001. Nitrogen-fixing nodules with *Ensifer adhaerens* harboring *Rhizobium tropici* symbiotic plasmids. *Appl Environ Microbiol* **67**:3264-3268.
- Rome, S., Fernandez, M. P., Brunel, B., Normand, P., and Cleyet-Marel, J. C. 1996. *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. *Int J Syst Bacteriol* **46**:972-980.
- Roumiantseva, M. L., Andronov, E. E., Sharypova, L. A., Dammann-Kalinowski, T., Keller, M., Young, J. P., and Simarov, B. V. 2002. Diversity of *Sinorhizobium meliloti* from the Central Asian Alfalfa Gene Center. *Appl Environ Microbiol* **68**:4694-4697.
- Roumiantseva, M. L., Yakutkina, V. V., Damman-Kalinowski, T., Sharypova, L. A., Keller, M., and Simarov, B. V. 1999. Comparative analysis of structural organization of the genome in alfalfa nodule bacteria *Sinorhizobium medicae* and *Sinorhizobium meliloti*. *Russ J Genet* **35**:128-135.
- Shimodaira, H., and Hasegawa, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**:1114-1116.
- Simms, E. L., Taylor, D. L., Povich, J., Shefferson, R. P., Sachs, J. L., Urbina, M., and Tausczik, Y. 2006. An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. *Proc R Soc Lond B Biol Sci* **273**:77-81.
- Small, E., and Jomphe, M. 1988. A synopsis of the genus *Medicago* (Leguminosae). *Can J Bot* **67**:3260-3294.
- Smith, N. H., Dale, J., Inwald, J., Palmer, S., Gordon, S. V., Hewinson, R. G., and Maynard Smith, J. 2003. The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc Natl Acad Sci U S A* **100**:15271-15275.

- Souza, V., Nguyen, T. T., Hudson, R. R., Pinero, D., and Lenski, R. E. 1992. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? *Proc Natl Acad Sci U S A* **89**:8389-8393.
- Spaink, H. P. 2000. Root nodulation and infection factors produced by rhizobial bacteria. *Annu Rev Microbiol* **54**:257-288.
- Stevenson, B. 2001. *Borrelia burgdorferi*: a (somewhat) clonal bacterial species. *Trends Microbiol* **9**:471.
- Suerbaum, S., Maynard Smith, J., Bapumia, K., Morelli, G., Smith, N. H., Kunstmann, E., Dyrek, I., and Achtman, M. 1998. Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **95**:12619-12624.
- Tachida, H., and Cockerham, C. C. 1986. Analysis of linkage disequilibrium in an island model. *Theor Popul Biol* **29**:161-197.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**:437-460.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595.
- Tibayrenc, M. 1996. Towards a unified evolutionary genetics of microorganisms. *Annu Rev Microbiol* **50**:401-429.
- van Baalen, M., and Jansen, V. A. A. 2001. Dangerous liaisons: the ecology of private interest and common good. *Oikos* **95**:211-224.
- Van Berkum, P., Beyene, D., Bao, G., Campbell, T. A., and Eardly, B. D. 1998. *Rhizobium mongolense* sp. nov. is one of the three rhizobial genotypes identified which nodulate and form nitrogen fixing symbiosis with *Medicago ruthenica* [(L.) Ledebour]. *Int J Syst Bacteriol* **48**:13-22.
- van Leeuwen, W. B., Jay, C., Snijders, S., Durin, N., Lacroix, B., Verbrugh, H. A., Enright, M. C., Troesch, A., and van Belkum, A. 2003. Multilocus sequence typing of *Staphylococcus aureus* with DNA array technology. *J Clin Microbiol* **41**:3323-3326.
- Vetsigian, K., and Goldenfeld, N. 2005. Global divergence of microbial genome sequences mediated by propagating fronts. *Proc Natl Acad Sci U S A* **102**:7332-7337.
- Vulic, M., Dionisio, F., Taddei, F., and Radman, M. 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc Natl Acad Sci U S A* **94**:9763-9767.
- Weiller, G. F. 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Mol Biol Evol* **15**:326-335.

- Wernegreen, J. J., and Riley, M. A. 1999. Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Mol Biol Evol* **16**:98-113.
- West, S. A., Kiers, E. T., Pen, I., and Denison, R. F. 2002a. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *J Evol Biol* **15**:830-837.
- West, S. A., Kiers, E. T., Simms, E. L., and Denison, R. F. 2002b. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc R Soc Lond B Biol Sci* **269**:685-694.
- Whitehead, N. A., Barnard, A. M. L., Slater, H., Simpson, N. J. L., and Salmond, G. P. C. 2001. Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol Rev* **25**:365-404.
- Willems, A., Fernandez-Lopez, M., Munoz-Adelantado, E., Goris, J., De Vos, P., Martinez-Romero, E., Toro, N., and Gillis, M. 2003. Description of new *Ensifer* strains from nodules and proposal to transfer *Ensifer adhaerens* Casida 1982 to *Sinorhizobium* as *Sinorhizobium adhaerens* comb. nov. Request for an opinion. *Int J Syst Evol Microbiol* **53**:1207-1217.
- Wirth, T., Wang, X., Linz, B., Novick, R. P., Lum, J. K., Blaser, M., Morelli, G., Falush, D., and Achtman, M. 2004. Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proc Natl Acad Sci U S A* **101**:4746-4751.
- Wright, S. I., and Charlesworth, B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* **168**:1071-1076.
- Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* **15**:568-573.
- Young, J. M., Kuykendall, L. D., Martinez-Romero, E., Kerr, A., and 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 Syst Evol Microbiol* **51**:89-103.
- Young, J. M., Kuykendall, L. D., Martinez-Romero, E., Kerr, A., and Sawada, H. 2003. Classification and nomenclature of *Agrobacterium* and *Rhizobium*. *Int J Syst Evol Microbiol* **53**:1689-1695.
- Young, N. D., Cannon, S. B., Sato, S., Kim, D., Cook, D. R., Town, C. D., Roe, B. A., and Tabata, S. 2005. Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. *Plant Physiol* **137**:1174-1181.

- Zribi, K., Mhamdi, R., Huguet, T., and Aouani, M. E. 2004. Distribution and genetic diversity of rhizobia nodulating natural populations of *Medicago truncatula* in tunisian soils. *Soil Biol Biochem* **36**:903-908.
- Zwick, M. E., McAfee, F., Cutler, D. J., Read, T. D., Ravel, J., Bowman, G. R., Galloway, D. R., and Mateczun, A. 2005. Microarray-based resequencing of multiple *Bacillus anthracis* isolates. *Genome Biol* **6**:R10.

XII. Article 1

XII. Article 1

Available online at www.sciencedirect.comSCIENCE @ DIRECT[®]

Systematic and Applied Microbiology ■■■■■

SYSTEMATIC AND
APPLIED MICROBIOLOGYwww.elsevier.de/syapm

Nitrogen-fixing sinorhizobia with *Medicago luciniata* constitute a novel biovar (bv. medicaginis) of *S. meliloti*

Maria Del Carmen Villegas^{a,1}, Sophie Rome^{a,2}, Lucette Mauré^a, Odile Domergue^a, Louis Gardan^b, Xavier Bailly^a, Jean-Claude Cleyet-Marel^a, Brigitte Brunel^{a,*}

^aLaboratoire des Symbioses Tropicales et Méditerranéennes Cirad Ensa-M Ird UmII TA101J Campus International de Baillarguet 34 398 Montpellier Cedex 5, France

^bUniv PaVe Inra - Inh Université Angers B.P. 60057, 49 071 Beaumont Cedex France

Received 9 September 2005

Abstract

Sixty-eight new rhizobial isolates were obtained from root-nodules of *Medicago luciniata* and from Mediterranean soils in Tunisia and France. All of them were identified as *Sinorhizobium meliloti* on the basis of PCR-RFLP analyses of 16S rDNA and the intergenic spacer sequence between 16S and 23S rDNAs. DNA:DNA hybridization, phenotypic characterization and 16S rRNA gene sequencing led to the conclusion that they belong the same taxon. All new isolates shared the ability to nodulate and fix nitrogen with *M. luciniata* except 11 of them not capable of fixing nitrogen with this plant and originating from French soils containing no efficiently adapted symbionts with *M. luciniata*. The nitrogen-fixing rhizobia on *M. luciniata* differed markedly from the other *S. meliloti* or *Sinorhizobium medicum* isolates and references in their symbiotic traits such as *nifDK* RFLP diversity, *nodA* sequences and nitrogen effectiveness with tree other different annual *Medicago* species (*M. truncatula*, *M. polymorpha* and *M. sativae*). Two infrasubspecific (biovar) divisions are therefore proposed within *S. meliloti*: bv. *medicaginis* for *Sinorhizobium* efficient on *M. luciniata* and bv. *meliloti* for the classically known *S. meliloti* group represented by the strains ATCC9930^T and RCR 2011 efficient on *M. sativa*.

© 2005 Elsevier GmbH. All rights reserved.

Keywords: *Sinorhizobium*; Medics; Polyphasic taxonomy; Mediterranean legumes; Bacterial diversity; Nitrogen fixation; *Medicago sativae*

Introduction

Medicago is an extensive genus of the Leguminosae, comprising approximately 83 different species and 18 infraspecific taxa [39]. Two-thirds are annuals and one-third perennials [24]. *Medicago* spp. are nodulated in soils by rhizobia that were classically grouped in the species *Rhizobium meliloti*, renamed *Sinorhizobium meliloti* [12]. *S. meliloti* is also described as a micro-symbiont of members of two other genera, *Trigonella*

*Corresponding author. Tel.: +33 467 593 786, fax: +33 467 593 802.

E-mail address: brunel@ensam.inra.fr (B. Brunel).

¹Present address: Centro de Investigación en Biotecnología Aplicada—Tlaxcala, Instituto Politécnico Nacional Ex-Hda. San Juan Molino Km 1.5 Carretera Estatal Sta. Ines Tlaxcoyacmac – Tepetitla, Topetitla de Lardizabal, Tlaxcala, Mexico.

²Present address: Inserm U449, Mécanismes Moléculaires du Diabète, Faculté de Médecine Laennec, Rue Guillaume Paradin, 69372 Lyon Cedex 8, France.

and *Melilotus*. The *Medicago-Melilotus-Trigonella* tribe thus forms a cross-inoculation plant group relevant for the definition of *S. meliloti*, although some unusual strains of *S. meliloti* were also found to be effective on *Glycine max* [15].

Nitrogen-fixing symbionts on *Medicago* are not restricted today to the species *S. meliloti* and can be genetically diverse. *Sinorhizolium medicae* [35] and *Rhizobium mongolense* [43] are two additional species displaying specialized nitrogen-fixation specificity with two species of *Medicago*, *M. polymorpha* (annual) and *M. ruthenica* (perennial), respectively. In addition, two *R. tropici* strains CIAI 899^T and CFN 299 were found to be effective on *M. ruthenica* [43]. The number of bacterial species associated with *Medicago* is probably underestimated as symbionts of various *Medicago* species have never been considered or sufficiently characterized for taxonomic definition. Indeed, at least eight groups showing various degrees of host plant affinity were reported among *Medicago* species in the preliminary study by Brockwell and Hely [9], and some may belong to diverse taxa. Our knowledge of rhizobial diversity associated with *Medicago* is therefore limited and the highly specific rhizobial symbionts require more detailed description to gain a comprehensive view of the taxonomy and evolution of rhizobial isolates of *Medicago* species.

Symbiotic properties are coded by genetic elements that can be transmitted between different bacteria and are not therefore considered suitable for species definition requiring reliable classification based on stable characters. Because of the value of rhizobia in the environmental and agricultural fields, rhizobial analysis based on symbiotic performances and cross inoculation range is essential for rhizobium users and should be performed in detail to produce a classification that is also useful for practical applications [16]. The symbiotic genes involved in symbiotic properties are therefore excellent markers for complete rhizobial characterization. Common nodulation genes such as *nodABC* or nitrogenase genes such as *nifHDK* have been useful in the study of interactions with host plants [22,23,27] and clarify relationships between rhizobial systematics and symbiotic properties.

For a better understanding of specific interactions and evolution among rhizobia and *Medicago* species, we have focused attention on the symbiotic diversity of rhizobia associated with *M. luciniata*. Together with other *Medicago* species (*M. polymorpha*, *M. noeana*, *M. sativae*), *M. luciniata* was reported to be highly specific in rhizobial requirements [9] with bacterial *nod* genes (specially *nodK*) involved in the establishment of their strict nodulation specificity [4,45]. *M. luciniata* (L.) Mill (cut leaf medic) is a winter medic indigenous to countries of the south Mediterranean Circle and east of the Mediterranean but occurs adventitiously in Europe

and Asia [20] where it was introduced as a ruderal weed. It is a pioneering plant that provides valuable winter fodder crop suitable for arid habitats and marginal cropping areas [49]. The objective of this study was to characterize the bacteria associated with *M. luciniata* and identify the symbiotic efficient ones. We isolated 68 root-bacteria associated with *M. luciniata* from soil samples collected in the Mediterranean basin where *M. luciniata* is native and compared their genetic and phenotypic relationships with other *Medicago*-nodulating bacteria that we had studied previously [10,34]. We considered it useful to describe chromosomal and symbiotic genetic diversity in order to understand the taxonomic and ecological aspects. We conclude that sinorhizobia, which form effective symbioses with *M. luciniata*, represent a novel biovar of *S. meliloti*.

Materials and methods

Bacterial strains

The sources of bacterial isolates and references are listed in Table 1.

Isolation of root-nodulating bacteria

A total of 68 new isolates were obtained from three French sites and from five Tunisian soils kindly sent by Dr. M.E. Aouani (INRST, Tunisia) [50]. The presence of rhizobia in soil samples was investigated by using *Medicago luciniata* as the trapping plant. Bacterial clones were isolated from 4-week-old nodules as previously described [10]. Bacteria were preserved at 80 °C in YEM broth [46] supplemented with 50% glycerol (v/v).

PCR/RFLP analyses on 16S rRNA-IGS and *nifK/D* regions

Bacteria were grown in TY agar Petri dishes at 28 °C for 2 days. Cells were suspended in sterile water. In total, 300 µl bacterial suspension was adjusted to an optical density (620 nm) of 1 and then pelleted. Cell pellet was resuspended in Tris-HCl (5 mM, pH 8.3) and proteinase K (0.2 µg/µl). Cells were incubated for 2 h at 55 °C and then 10 min in boiling water. Bacterial cell preparations were stored at -20 °C until use. A volume of 2 µl was used for PCR amplification as template DNA.

Amplification for 16S rDNA-IGS (2.8 kb) and *nifDK* (1.3 kb) regions was performed using a programmable heat block (Perkin-Elmer model 2400) as previously described [10], except for the amplification of the *nifKP* region for which the temperature cycle program was

Table 1. *Sinorhizobium* strains and isolates used in this study and their genotypes based on 16S rDNA-IGS and *nifDK* typing

Isolates or reference strains	Plant host	Geographical origin and source or reference	Genotype of	
			16S rDNA-IGS	nifKD
<i>Sinorhizobium</i> sp. (<i>M. laciniosa</i>)				
C11 to C15	<i>M. laciniosa</i>	Cézas, France (this study)	R6	N9
C111 to C115	<i>M. laciniosa</i>	Cézas, France (this study)	R6	N9
C1111 to C1114	<i>M. laciniosa</i>	Cézas, France (this study)	R6	N9
C1V1 to C1V2	<i>M. laciniosa</i>	Cézas, France (this study)	R6	N9
CV1 to CV4	<i>M. laciniosa</i>	Cézas, France (this study)	R6	N9
HL1, HL2	<i>M. laciniosa</i>	Hérault, France (this study)	R7	N12
HL3, HL4	<i>M. laciniosa</i>	Hérault, France (this study)	R3	N1
S2L1	<i>M. laciniosa</i>	Saint-Bauzille, France (this study)	R3	N1
S2L2	<i>M. laciniosa</i>	Saint-Bauzille, France (this study)	R1	N13
S2L3, S2L6	<i>M. laciniosa</i>	Saint-Bauzille, France (this study)	R1	N1
S2L4	<i>M. laciniosa</i>	Saint-Bauzille, France (this study)	R9	N11
S2L5	<i>M. laciniosa</i>	Saint-Bauzille, France (this study)	R8	N4
S2L7	<i>M. laciniosa</i>	Saint-Bauzille, France (this study)	R9	N4
T6L1	<i>M. laciniosa</i>	Enfidha, Tunisia (this study)	R3	N10
T6L2 to T6L5	<i>M. laciniosa</i>	Enfidha, Tunisia (this study)	R1	N10
T7L1	<i>M. laciniosa</i>	Kairouan, Tunisia (this study)	R3	N9
T7L2 to T7L5	<i>M. laciniosa</i>	Kairouan, Tunisia (this study)	R3	N10
T7L6, T7L7	<i>M. laciniosa</i>	Kairouan, Tunisia (this study)	R1	N9
T8L1, T8L2	<i>M. laciniosa</i>	Hadjeb, Tunisia (this study)	R3	N10
T8L3, T8L6	<i>M. laciniosa</i>	Hadjeb, Tunisia (this study)	R1	N10
T8L4, T8L5	<i>M. laciniosa</i>	Hadjeb, Tunisia (this study)	R3	N10
T8L7	<i>M. laciniosa</i>	Hadjeb, Tunisia (this study)	R3	N9
T9L1	<i>M. laciniosa</i>	Djelma, Tunisia (this study)	R1	N10
T9L2 to T9L4	<i>M. laciniosa</i>	Djelma, Tunisia (this study)	R3	N9
T9L5	<i>M. laciniosa</i>	Djelma, Tunisia (this study)	R3	N10
T9L6 to T9L10	<i>M. laciniosa</i>	Djelma, Tunisia (this study)	R3	N9
T10L1 to T10L3	<i>M. laciniosa</i>	Ben Aoune, Tunisia (this study)	R3	N9
T10L4	<i>M. laciniosa</i>	Ben Aoune, Tunisia (this study)	R1	N10
T10L5, T10L6	<i>M. laciniosa</i>	Ben Aoune, Tunisia (this study)	R3	N9
T10L7, T10L8	<i>M. laciniosa</i>	Ben Aoune, Tunisia (this study)	R1	N10
WSM253	<i>M. laciniosa</i>	South Australia, Dr. J. Howenson	R3	N9
WSM484	<i>M. laciniosa</i>	Australia, Dr. J. Howenson	R6	N9
WSM486	<i>M. laciniosa</i>	Western Australia, Dr. J. Howenson	R3	N10
<i>S. meliloti</i>				
RCR 2011	<i>M. sativa</i>	Australia [14]	R3	N1
1021	Sm ^R derivative of RCR 2011 [27]		R3	N1
ATCC 9930 ^T	<i>M. sativa</i>	United States [14]	R3	N1
CT11	<i>M. truncatula</i>	France [10]	R1	N4
CP711	<i>M. polymorpha</i>	France [10]	R1	N1
V11	<i>M. truncatula</i>	France [35]	R1	N1
A110B	<i>M. truncatula</i>	France [35]	R3	N4
V124A	<i>M. truncatula</i>	France [35]	R1	N4
<i>S. medicae</i>				
A321 ^T	<i>M. truncatula</i>	France [35]	R2	N2
CP41	<i>M. polymorpha</i>	France [10]	R2	N2
V22	<i>M. truncatula</i>	Aude, France [35]	R2	N3
M3	<i>M. orbicularis</i>	Syria [14]	R4	N3
M104	<i>M. truncatula</i>	Syria [14]	R5	N3

modified according to Don et al. [13]. Initial denaturation was performed at 95 °C for 5 min and was followed by 20 cycles consisting of denaturation (30 s at 94 °C),

annealing (30 s at a temperature decreasing by 0.5 °C/cycle from 65 to 55 °C), and extension (72 °C at 2 min). The annealing step was maintained at 55 °C for the

following 25 cycles. On completion of the 45 cycles, the samples were incubated for 7 min at 72 °C and then held at 4 °C.

Aliquots of about 10 µl PCR amplicats were restricted by *Cfo*I, *Hae*III, *Msp*I or *Ctr*I31 (for the *nifKD* region only) endonucleases as following the manufacturer's recommendations (Promega, Eurogentec). Restricted DNA was then analyzed by agarose gel electrophoresis as previously described [10]. Clustering from PCR-RFLP data was estimated with the Restsite program [28] based on the unweighted pair group method with arithmetic averages (UPGMA) [40] and neighbor-joining (NJ) [36] methods.

DNA/DNA hybridization

Bacterial DNA was extracted from 3-day-old 150-ml TY broth cultures as described by Brenner et al. [8]. Total DNA from the strain *S. meliloti* ATCC 9930^T was labeled by random-priming using a Megaprime kit (Amersham). Levels of DNA relatedness were determined using the S1 nuclease trichloroacetic acid method [17] at the optimal temperature for DNA reassociation (70 °C) for 16 h.

Phenotypic tests

- (i) *Growth characteristics*: Growth capacity at different pH values and temperatures and tolerance to sodium chloride were determined. Strains were pre-cultured in YEM broth at 28 °C to the log phase; the cells were then washed once with 0.8% NaCl solution and resuspended. Tubes containing 5-ml YEM broth were inoculated to achieve an initial optical density at 600 nm (OD_{600}) of between 0.08–0.1. The tubes were incubated for a week and OD_{600} values were measured. NaCl concentrations tested in YEM broth were 0.01%, 3.0% and 5.0%. pH values of YEM broth were adjusted to 4.0, 6.0, 7.0 and 9.0 by adding HCl or NaOH. Growth capacity at 37 and 42 °C was also determined.
- (ii) *Antibiotic susceptibility*: Intrinsic antibiotic resistance tests were performed by measuring inhibition zone diameters on YEM agar plates with the following antibiotic discs (bioMérieux, France): nalidixic acid (30 µg), chloramphenicol (30 µg), neomycin (30 µg), penicillin G (4 units), polymyxin B (300 units), streptomycin (10 µg), tetracycline (30 µg) and vancomycin (30 µg). Tests were performed in triplicate and the susceptibility of the strains was deduced from the antibiograms (bioMérieux) on the basis of the inhibition zone diameters.
- (iii) *Biochemical and auxanographic tests*: The assimilation of 49 different carbon sources was determined on API50CH strips with API50CHB medium (bioMérieux) with phenol red as the pH indicator.

Strains were inoculated as recommended by the manufacturer and incubated for 7 days at 28 °C.

- (iv) *Numerical analysis*: The similarity between each possible pair of isolates tested for auxanographic and biochemical characteristics was estimated by comparing the phenotypic variables by using the Jaccard similarity coefficient [40]. Clustering was performed by the UPGMA method.

Fixation and nodulation tests

Isolates and reference strains belonging to different PCR-RFLP types were tested for the ability to fix nitrogen on *M. lavinata* (ecotype from Tunisia), *M. polymorpha* (ecotype from Combaillaux, France), *M. truncatula* (ecotype from Var, France) and *M. sativae* (ecotype from Morocco). In addition, the strain C14 was tested on *Medicago sativa*, *M. murex*, *M. arborea*, *M. aculeata*, *Trigonella tenuifolium* and *T. balansae*. Seeds, plants and bacteria were prepared according to previously described procedures [10]. Seedlings were grown in glass tubes containing a nitrogen-free nutritive solution and kept in a growth chamber at 22 °C during the day and 18 °C at night, with a 16-h photoperiod and 60–70% relative humidity. Ten plants were tested for each isolate or strain. Plants were recorded for nodule formation for a 4–5-week period and effectiveness was estimated by weighing plant top dry matter. Uninoculated plants and plants fertilized with 5 mM $\text{Ca}(\text{NO}_3)_2$ were included as controls in all the plant tests. Analyses of variance were performed using Statistica. Differences were considered significant for *P*-values of <0.01.

Nucleotide sequencing of 16S rRNA and *nodA* genes, and sequence analysis

16S rDNA of strain C14 and *nodA* of strains C14 and TcL1 were amplified under the conditions described above. The primers for amplification of *nodA* region were *nodboxSme* (ATC GCA GAT GAT CGT TAT CC) and *NodB1rSme.Sfr.Rga* (CGG CCG TCR TCR AAC GTC AGG TA) according to Moulin et al. [29]. PCR products were purified from agarose gels using QIAquick spin columns (Qiagen Inc.). They were sequenced with a *Taq* dye-deoxy terminator cycle sequencing kit (Perkin-Elmer Corp.) in an automatic laser fluorescent DNA sequencer (Perkin-Elmer ABI PRISM model 310). Raw sequence data were displayed and assembled with the program Sequence Navigator (Perkin-Elmer). Database searches were conducted via the National Center for Biotechnology Information Web page using the BLAST-N 2.0.8 package programs [1]. DNA sequences were then aligned using the program

CLUSTAL-X version 1.83 [41]. Phylogenetic analysis was conducted by using maximum likelihood (ML) approach with PAUP* version 4.0b10 (from D.L. Swofford, Smithsonian Institution, Washington, DC, USA). For *nodA* phylogeny, the best fitting model of DNA evolution determined by using Modeltest 3.06 [32] was the HKY85 + G model [18]. Model parameters were as follows: transition/transversion ratio = 1.0974, nucleotide frequencies ($A = 0.2006$, $C = 0.2615$, $G = 0.3097$, $T = 0.2282$), shape parameter (α) = 0.7115. Reliability of nodes was estimated by ML bootstrap percentages obtained after 1000 pseudoreplications, performed with the previously estimated ML parameters, an NJ starting tree and tree-bisection-reconnection branch swapping.

The EMBL/GenBank accession numbers for the sequences *nodA* reported in this paper are AJ441287 and AJ441288.

Results

Isolation of rhizobia

Sixty-eight isolates were obtained and purified from nodules of *M. laciniata* after trapping from three French and five Tunisian soils. All were isolated from typical cylindrical nodules with apical and persistent meristem, except for the 11 isolates obtained from the two soils from Hérault and Saint-Bauzille (HLI-4, S2LI-7 in Table 1). Only small round nodules were obtained from these two French soils; 11 of them were crushed to achieve successful bacterial isolation.

PCR/RFLP analyses of 16S rDNA-IGS and *nifDK* regions

All *M. laciniata* isolates were identified on the basis of the PCR-RFLP of two regions: the 16S rDNA-IGS and the *nifDK* amplification DNAs. By combining all the restriction patterns obtained for the enzymes we tested (i.e. *CfoI*, *HaeIII*, *MspI* for 16S rDNA-IGS typing and *CfoI*, *HaeIII*, *MspI*, *CfrI* for *nifDK* typing), we found that the 68 bacterial isolates from *M. laciniata* could be grouped into six ribosomal types (R1, R3, R6, R7, R8 and R9 in Table 1) and seven *nif* types (N1, N4, and N9 to N13). Examples of each restriction pattern of digested ribosomal and *nif* DNA amplicons are provided in Figs. 1 and 2, respectively. The resulting ribosomal tree (Fig. 3a) showed that bacterial isolate types were closely related to *S. meliloti* references. In contrast, the *nifDK* relationships (Fig. 3b) exhibited by all the 57 *M. laciniata* isolates harboring types N9 and N10 are markedly divergent from the references of *S. meliloti* and *S. medicae*. The 11 remaining isolates (HLI-4 and S2LI-7) clustered with the *nif* types of *S. meliloti* references.

DNA/DNA hybridization

To confirm the taxonomic position of *M. laciniata* isolates, labeled DNA of strain *S. meliloti*^T was hybridized with DNAs extracted from a total of 15 isolates from *M. laciniata* (C14, C15, CIV2, HLI, S2LI, T6LI, T7LI-2, T8LI6, T9LI, T10LI, T10LI4, WSM253, WSM484, WSM486) plus 5 reference strains of *S. meliloti* (ATCC9930^T, RCR2011) and *S. medicae* (A321^T, M3, M104). The data showed high DNA homologies (73–90%) with all the 15 *M. laciniata* isolates and the strain RCR2011 (88%) whereas they displayed low DNA binding (37–41%) with DNA from the species *S. medicae*. Thus, *M. laciniata* isolates belonged to the same DNA homology group of *S. meliloti*.

Phenotypic tests

All the isolates and the *S. meliloti* or *S. medicae* reference strains tested were fast growers. They grew at 28, 37 °C but not at 42 °C and at pH 6.0 and 9.0 but not at pH 4.0, except for *S. meliloti* 2011. *S. meliloti* ATCC9930^T and isolate S2LI. They also grew in YEM broth with 3.0% NaCl but not with a 5% NaCl concentration. All isolates and strains tested were resistant to chloramphenicol and they all displayed sensitivity to tetracycline. Intrinsic resistance to penicillin G, streptomycin, neomycin, vancomycin and polymyxin B were variable, as shown in Table 2. Their resistance to streptomycin distinguished isolates from *M. laciniata*. All isolates and strains assimilated the 33 carbon sources tested: L-arabinose, ribose, D-xylose, adonitol, β -methyl-D-xyloside, galactose, glucose, fructose, mannose, sorbose, rhamnose, dulcitol, inositol, mannitol, arbutin, esculin, salicin, cellobiose, maltose, lactose, sucrose, trehalose, melezitose, raffinose, xylitol, turanose, lyxose, tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol and 2-keto-gluconate. None of them assimilated inulin. The assimilation of 15 other carbon sources revealed different bacterial responses, as shown in Table 2. The numerical analysis based on the 66 characters of phenotypic data showed that *M. laciniata* isolates were intermingled with the members of *S. meliloti* (data not shown) and could not be separated on the basis of biochemical and physiological characteristics.

16S rRNA and *nodA* gene sequence analysis

The 16S rRNA gene sequence of strain C14 (type R6), which displayed a different chromosomal pattern to the reference strains (RCR2011 and ATCC 9930^T) (Fig. 3a), was 100% identical to the sequence previously reported for *S. meliloti* 1021 (AL591782) with an R3 type pattern. Isolates C14 (type N9) and T6LI (type N10) were chosen as two distinct representative microsymbionts of the

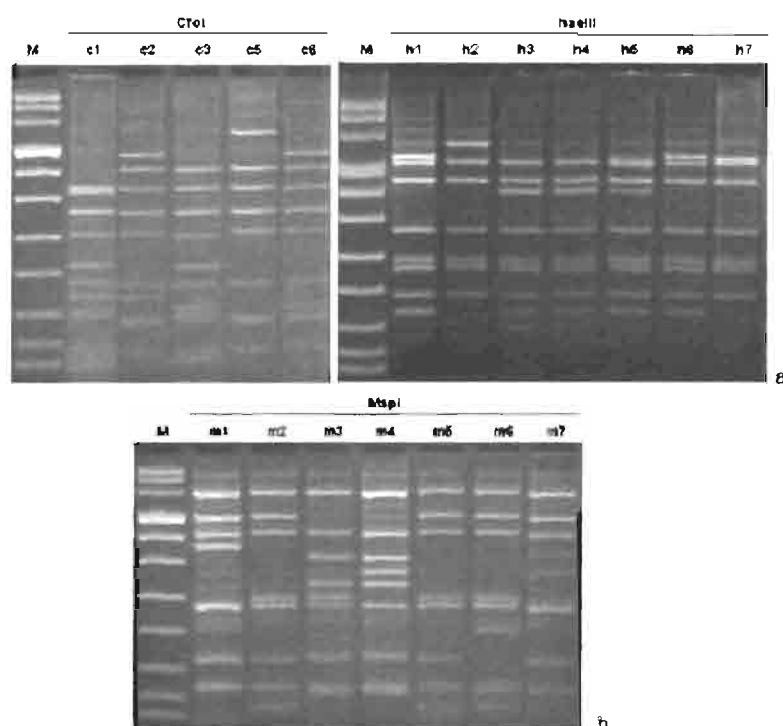


Fig. 1. Restriction patterns of PCR-amplified fragment of 16S rDNA and IGS regions digested by *CfoI*, *HaeIII* (a), and *MspI* (b). Lanes for (i) *CfoI* gel; c1: R1, R6, R8, and R9 types, c2: R4 type, c3: R3 and R7 types, c4: R5 type, c5: R2 type (ii) *HaeIII* gel; h1: R3, h2: R4, h3: R9, h4: R1 and R6, h5: R8, h6: R7, h7: R5 (iii) *MspI* gel; m1: R3, m2: R1, m3: R2 and R4, m4: R5, m5: R9, m6: R6, m7: R8. M: molecular mass marker VIII (Boehringer Mannheim), 1114, 900, 692, 501, 489, 404, 320, 242, 190, 147, 124, 110 bases.

M. laciniata bacterial dichotomy as revealed by *nifDK* typing (Fig. 3b) and their complete *nodA* sequences were determined. The *nodA* sequences were computed and a phylogenetic tree was constructed by the ML method. Isolates C14 and T6L1 grouped with USDA 1170 an other bacterium fixing nitrogen with *M. laciniata* and these three strains were markedly divergent from the other *S. meliloti* and *S. medicae* as supported by high bootstrap values (Fig. 4). This was consistent with clustering data obtained by PCR-RFLP data on the *nifKD* DNA region. All *nodA* genes were 591 bp long, corresponding to 196 amino-acids plus a stop codon. The protein sequence identities of isolates C14 and T6L1 were 95.4% and 94.4%, respectively, with the sequence of *S. meliloti* 1021 (spontaneous resistant mutant of RCR2011 streptomycin) and 97.4% and 95.9% with those of *S. medicae* type strain.

Fixation and nodulation tests

Whatever the chromosomal background, only isolates with N9 and N10 *nifKD* types were able to fix nitrogen effectively with *M. laciniata* (Table 3). In contrast,

strains of *S. meliloti* harboring plasmidic *nifKD* types other than N9 and N10, such as N1, N3, N12 and N13, and strains of *S. medicae*, induced inefficient symbiosis with *M. laciniata*, since shoot dry weights were not significantly different from uninoculated control plants. These inefficient bacteria did not elicit typical cylindrical root nodules. Strains of *S. meliloti* triggered small round nodules whereas *S. medicae* strains showed no nodulation or rare white swellings. Sinorhizobia isolated from rudimentary nodules on *M. laciniata* (HL1, S2L1, and S2L4) could only occasionally trigger small non-efficient nodules on their original host plant under the experimental conditions used (Table 3). Efficient isolates on *M. laciniata* did not share the symbiotic trait of *S. medicae* of fixing nitrogen with *M. polymorpha* (Table 3). To evaluate the host plant spectrum of strains efficient on *M. laciniata*, we tested the strain C14 in association with seven other *Medicago* and *Trigonella* plant species. It formed only non-efficient rudimentary nodules on *Medicago sativa*, *M. murex*, *M. arborea*, *M. aculeata*, *Trigonella fenuugrecum* and *T. balansae*, although high N-fixation was obtained on *M. sativae*. This particular fixation trait was confirmed with additional isolates and strains. Seven of the isolates

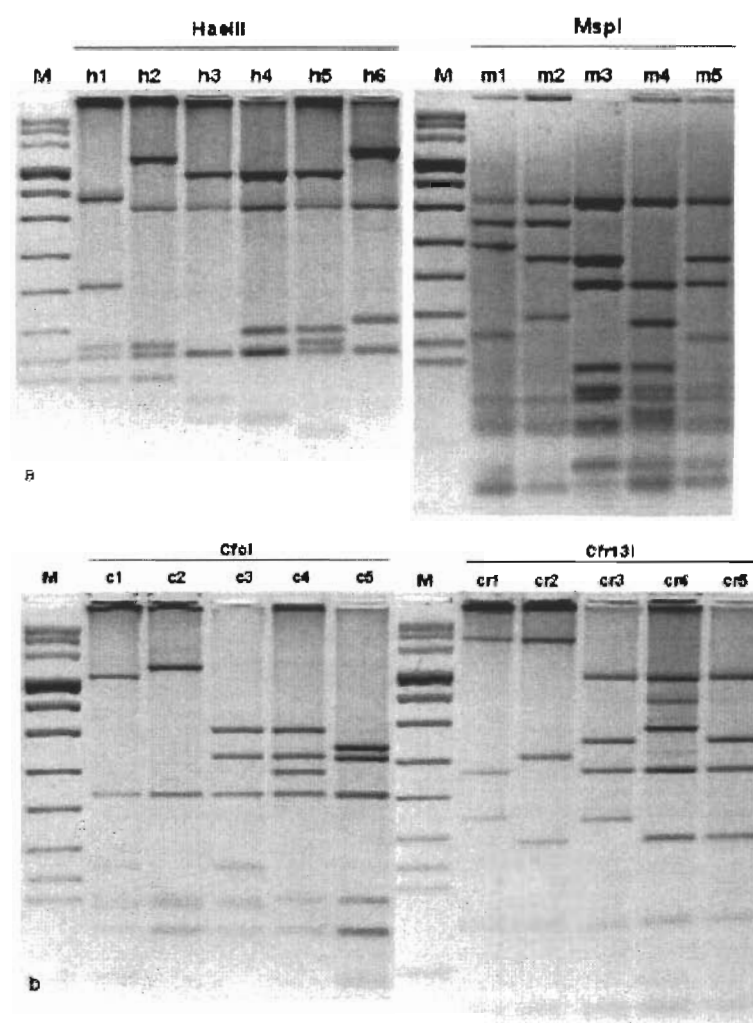


Fig. 2. Negative images of ethidium bromide-stained gels showing restriction patterns of PCR-amplified fragment of the *nifKD* region digested by *Hae*III and *Msp*I (a), *Cfo*I and *Cfr*I3I (b). Lanes for (i) *Hae*III gel: h1: N9, h2: N10, h3: N12, h4: N1, N4 and N11; h5: N13; h6: N2 and N3 (ii) *Msp*I gel: m1: N9, m2: N10, m3: N1, N2, N3, N12 and N13, m4: N11, 5: N4, (iii) *Cfo*I gel: c1: N9, c2: N10, c3: N1, N4 and N13, c4: N11, c5: N2 and N3 (iv) *Cfr*I3I gel: cr1: N9, cr2: N10, cr3: N1, N4, N12, N13 and N11, cr4: N2, cr5: N3; M: molecular mass marker VIII (Boehringer Mannheim).

effective on *M. laciniata* (C14, T6Li-2, T7L1, WSM253, WSM484 and 486) were also found to be effective with *M. savagei* whereas three isolates that were inefficient on *M. laciniata* (HL1, S2L1, S2L4) were not able to nodulate *M. savagei* either (Table 3). *S. meliloti* RCR 2011 and *S. medicae* A321^T were not effective in N-fixation with *M. laciniata* either.

Discussion

All the bacterial isolates obtained from *M. laciniata* belonged to the species *S. meliloti*. They displayed a

diagnostic pattern of RFLP on 16S rDNA-IGS that is characteristic of *S. meliloti*. DNA/DNA relatedness and common phenotypic features validated this result. This conclusion was consistent with the case of one strain isolated from *M. sativa* and symbiotically effective on *M. laciniata* (CC2003) which clustered with *S. meliloti* members in multilocus enzyme electrophoresis analysis [14]. The genetic relationship of the *M. laciniata* bacterial group was confirmed with *S. meliloti* by the absence of any variation along the 16S rDNA sequence of the isolate C14 compared to the sequence of strain 1021. Because the strain C14 displayed some variability in the 16S rDNA-IGS region compared with strain 1021

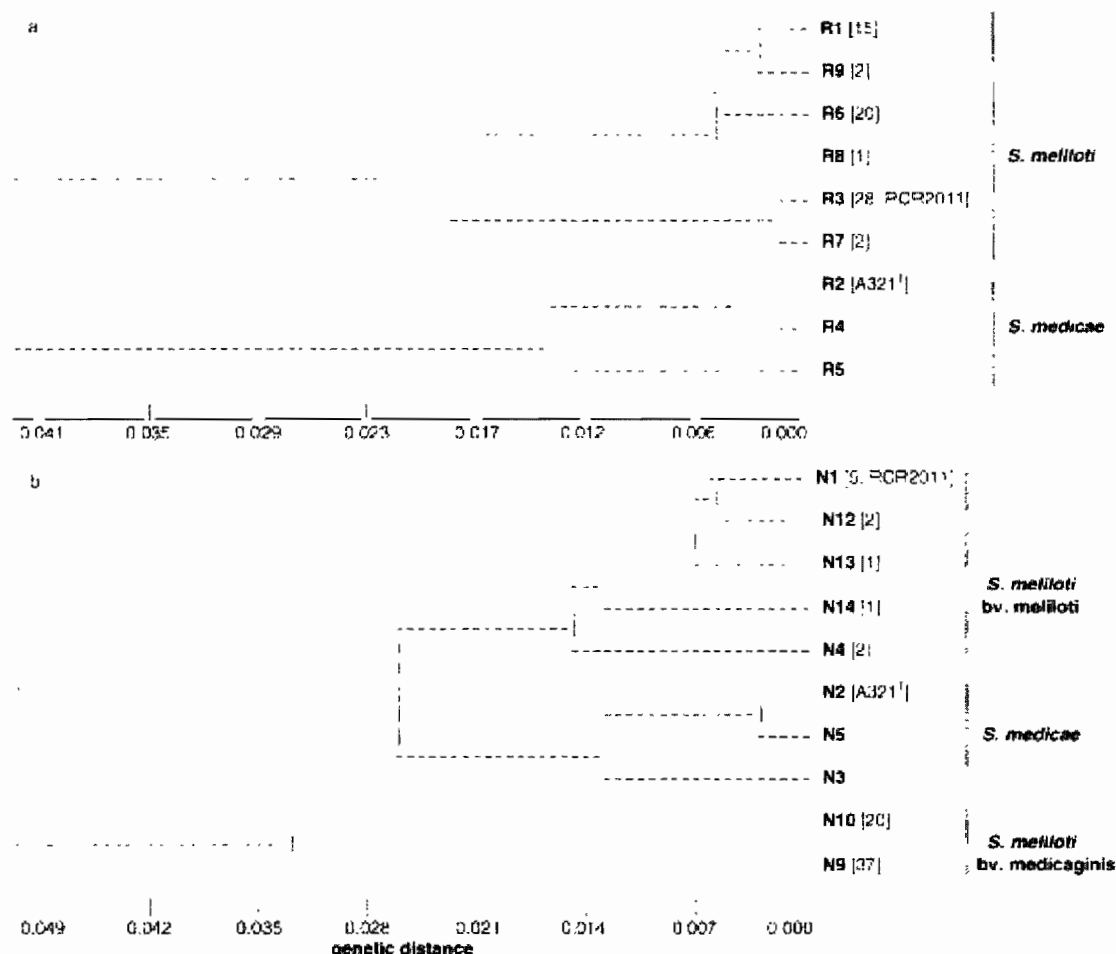


Fig. 3. Dendrograms (UPGMA) of genetic relationships between *Sinorhizobium* sp. (*M. laciniata*) isolates and *S. meliloti* and *S. medicae* reference strains estimated on the basis of the PCR-RFLP profiles from 16S rDNA-4GS amplified DNAs (a) and *nifK-D* amplified DNAs (b). Numbers of isolates from *M. laciniata* are given between brackets.

by PCR-RFLP (Fig. 3a), it can be concluded that the diversity we observed on 16S rDNA-IGS region of *S. meliloti* is harbored by the IGS part. In a similar manner to other *S. meliloti* strains, isolates effective on *M. laciniata* are reliably distinguished from members of the species *S. medicae* by symbiotic (Table 3) and genetic characteristics.

Although they were very similar to *S. meliloti* at the chromosomal level, N-fixing isolates on *M. laciniata* exhibited differences attributable to diverging symbiotic traits. They were clearly identified from the classically defined *S. meliloti* group represented by the two closely related strains ATCC9930^T and RCR2011 according to their *nifK-D* typing data, *nodA* divergences and symbiotic properties. They probably show divergences on other close symbiotic gene markers such *nodC* and *nodD1* as

described for the strain 102L4 or USDA1635 efficient on *M. laciniata* [4,6]. They were able to fix nitrogen with *M. laciniata*, unlike the typical *S. meliloti* strains. In contrast they were not effective with *M. truncatula* although members of the species *S. meliloti* and *S. medicae* were. In addition, the results suggest that the nitrogen-fixation capacities by sinorhizobia on *M. laciniata* and *M. sativae* are strongly linked. As *M. laciniata* and *M. sativae* are two closely related medics forming a coherent group according to cross hybridization potentialities [38] and ribosomal spacer analysis [5,6], we observe in this particular case, that a close host plant relationship does relate to rhizobial specificity. Symbiotic traits of sinorhizobia efficient with *M. laciniata* are distinguishable from these of *S. medicae* strains, which remain the only symbionts effective on *M. polymorpha*.

Table 2. Phenotypic differences between the three different groups of sinorhizobia nodulating *Medicago*.

Designation	Group I	Group II	Group III
Type strain	<i>S. meliloti</i> ATCC9930 ^T	<i>S. medicae</i> A321 ^T	<i>S. meliloti</i> bv. <i>medicaginis</i> C14
No. of strains tested	5	3	21 ^a
<i>Antibiotic sensibility^b</i>			
Streptomycin	(S)	I	R
Neomycin	I	(S)	I
Polymyxin B	D	S	(S)
Vancomycin	R	S	R
<i>Utilization of</i>			
Glycerol	—	d	(+)
Erythritol	—	(—)	(—)
D-arabinose	D	+	+
L-xylose	+	d	+
Sorbitol	D	+	+
α-Methyl-D-mannoside	—	+	(—)
α-Methyl-D-glucoside	(—)	—	d
N-acetyl-glucosamine	D	+	+
Amygdalin	—	d	d
Melbiiose	+	d	+
Starch	—	d	d
Glycogen	—	d	d
Gentiobiose	D	+	+
Glucanate	D	d	(+)
5-Keto glucanate	D	d	+

Results of tests are indicated as follows.

^a All the isolates tested had plasmid type N9 or N10.

^b R: resistant (diameter of the inhibition zone: <13 mm for streptomycin and neomycin, <15 mm for polymyxin B, <17 mm for vancomycin), S: sensitive (diameter of the inhibition zone: ≥15 mm for streptomycin and polymyxin B, ≥17 mm for neomycin and vancomycin), I: intermediate, (S): 76–89% of strains sensitive, d: up to 10% sensitive.

+ : at least 90% of strains positive, (+): 76–89% of strains positive; d: 26–75% of strains positive, (—): 11–25% of strains positive; —: up to 10% positive.

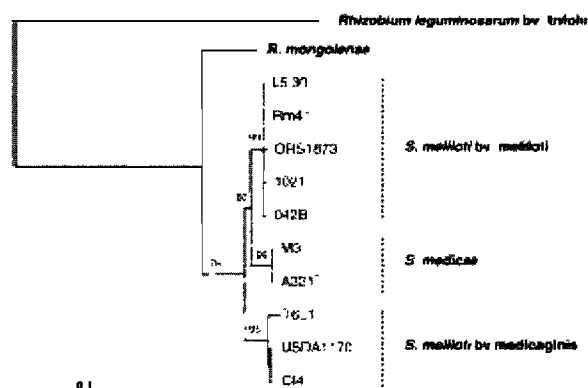


Fig. 4. Phylogenetic tree based on the *nodA* sequences and constructed by using the ML method. Bootstrap percentages are given at the branching points ($n = 1000$ replicates). Genbank accession numbers are as follows: AJ007237 (*S. meliloti* 1021), AF038577 (*S. meliloti* 042B), X01649 (*S. meliloti* Rm41), AJ300220 (*S. meliloti* L5-30), AJ300233 (*S. meliloti* ORS1873), AJ300223 (*S. meliloti* USDA1170), AJ300224 (*S. medicae* A321^T), AJ300225 (*S. medicae* M3), AJ300241 (*R. mongolense* USDA 1844), X03721 (*R. leguminosarum* bv. *trifolii* ANU643) was used as a root.

Although soils were sampled from sites of the Mediterranean Basin, not all of them contained N-fixing rhizobia with *M. laciniata*. North part of the Mediterranean area (France) seems to be only partially colonized compared to the southern part (Tunisia) where *M. laciniata* or *M. sativae* plants grow more frequently [24,39]. When N-efficient populations were absent in soil, *M. laciniata* plants trapped few strains of *S. meliloti* with new variant *nif* types closely related to the references strains of *S. meliloti* but which were not necessary efficient on *M. truncatula* as most of the typical strains of *S. meliloti* and *medicae*. Such new *S. meliloti* recruited by an exogenous legume and showing a parasitic behavior on *M. laciniata* could be more adapted to an other native *Medicago* legume. Yet, when N-efficient sinorhizobia were present in soils, only efficient *S. meliloti* bv. *medicaginis* were collected with two symbiotic types in Tunisian soils (N9 and N10) and one of them in French soils (N9). Populations of N-fixing sinorhizobia associated to *M. laciniata* are therefore geographically structured which could be a factor limiting the expansion of the host plant which cannot find valuable microsymbionts.

Table 3. Nodulation and nitrogen fixation of sinorhizobia isolates and reference strains with *M. loti*, *M. polymorpha*, *M. truncatula* and *M. savatzei*. Ribosomal and *nif* types are in brackets.

	<i>M. loti</i>	<i>M. polymorpha</i>	<i>M. truncatula</i>	<i>M. savatzei</i>
<i>Sinorhizobium</i> sp. (<i>M. loti</i>)				
C14 (R6N9)	E ^a	0	0	E
C115 (R6N9)	E	nt ^b	0	nt
C112 (R6N9)	E	nt	nt	nt
C1V2 (R6N9)	E	nt	nt	nt
C1V3 (R6N9)	E	nt	nt	nt
T6L1 (R3N10)	E	1 (0.2)	0	E
T6L2 (R1N10)	E	1 (0.9)	0	E
T7L1 (R3N9)	E	1 (0.4)	0	E
T7L2 (R3N10)	E	1 (0.7)	0	nt
T6L8 (R1N10)	E	1 (0.7)	nt	nt
T10L1 (R3N9)	E	1 (0.9)	nt	nt
HL1 (R7N12)	1 (0.1)	1 (1.0)	E	0
S2L1 (R3N1)	1 (0.1)	1 (1.0)	1 (1.0)	0
S2L4 (R9N11)	1 (0.1)	nt	1 (1.0)	0
WSM253 (R3N9)	E	E	nt	E
WSM484 (R1N9)	E	1 (1.0)	0	E
WSM486 (R3N10)	E	nt	nt	E
<i>S. meliloti</i>				
RCR2011 (R3N1)	1 (0.9)	E ^d	E ^d	0
1021 (R3N1)	1 (1.0)	1 (1.0)	nt	nt
AJCC9930 ^c (R3N1)	0	Nt	E	nt
A110B (R3N4)	1 (1.0)	E ^d	E ^d	Nt
V11A (R1N1)	1 (1.0)	E ^d	E ^d	nt
CT11 (R1N4)	1 (0.9)	E ^c	E ^c	nt
CP7H (R1N1)	1 (1.0)	E ^c	E ^c	nt
V124A	1 (1.0)	nt	nt	nt
<i>S. medicae</i>				
A321T (R2N2)	1 (0.1)	E ^d	E ^d	1 (0.1)
CP41 (R2N2)	1 (0.1)	E	E ^c	nt
V22 (R4N3)	1 (0.1)	E ^d	E ^d	nt
M3 (R4N3)	nt	E ^d	E ^d	nt
M104 (R5N3)	1 (0.2)	E	nt	nt

^aE, effective symbiosis (the shoot dry matter weights of plants were significantly different from the inoculated control ($P = 0.01$); E, ineffective symbiosis (the shoot dry matter weights of plants inoculated did not differ from those of the uninoculated plants ($P = 0.01$); values in brackets are the proportion of nodulated plants; 0, no nodules. Uninoculated controls were without nodules.

^bnt, not tested.

^cAccording to N-fixation data of Brunel et al. [10].

^dAccording to N-fixation data of Rome et al. [15].

The term biovar in rhizobia species was coined to refer to taxonomic infrasubspecific divisions with distinct symbiotic properties and symbiotic plasmids or genes with a similar chromosome background. Biovars were based first on nodulating and non-nodulating abilities, and then they were extended to fixation abilities [25,33]. Now with progress in genotyping and phylogenies, biovars are strongly suggested by phylogeny comparison of house-keeping genes and symbiotic ones as the strain formerly named *M. loti* shows typical genetic properties of a "biovar loti" in the species *M. huakuii* [42]. Thus in current rhizobial

systematics, because of different host ranges present in a single species, a total of 13 biovars have been defined in 12 species: bv. viciae and tritoli in *R. leuuminosarum* [21]; bv. phaseoli in *R. leuuminosarum* [21]; *R. gallicum* [2]; *R. giardinii* [2] and *R. ethi* [37]; bv. sesbaniae in *S. teranga* [26] and *S. sahelense* [7,19]; bv. acaciae in *S. teranga* [26]; *S. sahelense* [7,19]; and *S. meliloti* [3]; bv. gallicum in *R. gallicum* [2]; bv. giardinii in *R. giardinii* [2]; bv. mimosa in *R. ethi* [48]; bv. orientalis and officinalis in *R. galegae* [33]; bv. rege in *M. huakuii* nodulating rege-sou (*Astragalus sinica* cv. Japan) [30,31]; bv. genisterarum in *B. japonicum* and

B. canariense [47] and bv. *glycinearum* in *B. japonicum* and *B. luoningense* [47]. This classification is important for the user who needs easy identification of rhizobia for use in legume crop optimizing as the systematic diversity of bacteria nodulating legumes is widening and not always related to host plants [11,29,44]. It is therefore convenient to create two new biovars for *S. meliloti*: (i) the biovar *medicaginis* [N.L. n. *Medicago*-inis, scientific name of a botanical genus; *medicaginis*, of a plant of the genus *Medicago*, symbiont of *Medicago laciniosa*] which encompasses *sinorhizobium* showing N-fixation specificity with *M. laciniosa* and *M. sativae* and represented by strain CI4 and (ii) the biovar *meliloti* which clusters the classically known *S. meliloti* strains [21] represented by the strains ATCC 9930^T and RCR 2011, efficient on *M. sativa* and *M. truncatula*, and forming no nodules or only ineffective ones on *M. laciniosa*. This subdivision is confirmed by different plasmid-encoded symbiotic genes present in a uniform chromosomal background belonging to the species *S. meliloti* *nifDK* and *nodA* sequence organizations illustrate the genetic subdivision within *S. meliloti*.

N-fixation specificity in *Medicago* annual species is currently performed by rhizobia which can belong to two different genera [12,43], three different species [12,35,43] or two biovars. Thus, as for other legumes such as soybean, acacia and common bean, rhizobial symbionts associated with *Medicago* spp. are taxonomically heterogeneous. The history of the evolution of microsymbionts with *Medicago* species is therefore polyphyletic and adaptation strategy seems variable according to their host plants. No particular single chromosome seems to be required for a predisposition to *Medicago* symbiosis and the latter may have originated at different times and for different *Medicago* species. Further taxonomic and phylogenetic studies will be helpful to elucidate coevolution between medics and their micro-symbionts.

Acknowledgments

MCV received a fellowship from CONACYT. We are grateful to Dr. J. Howieson (CLIMA, Australia) for sending us strains, Dr M. E. Aouani (INRST, Tunisia) for sending soils from Tunisia, the Australian *Medicago* genetic Resource Centre (Adelaide, Australia), Jean-Marie Prospéri (Inra France) for offering *Medicago* seeds, Gustave Gintzberger (Cirad France) for identifying *M. laciniosa* and Pr. Jean P. Euzéby (Ecole Nationale de Toulouse, France) for helping with bacterial nomenclature. We thank Valérie Miralles and Virginie Chareyron for their excellent technical help and Simon Barnard for a critical reading of the manuscript and correction of the English.

References

- [1] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucl. Acids Res.* 25 (1997) 3389–402.
- [2] N. Amarger, V. Macheret, G. Laguerre, *Rhizobium gallium* sp. nov. and *Rhizobium qiaranum* sp. nov. from *Phaseolus vulgaris* nodules, *Int. J. Syst. Bacteriol.* 47 (1997) 996–1006.
- [3] S. Ba, A. Willems, P. de Lajudie, P. Roche, H. Jéder, P. Quainni, M. Neyra, M. Ferro, J.-C. Prome, M. Gillis, C. Boyvin-Masson, J. Lorquin, Symbiotic and taxonomic diversity of *Rhizobia* isolated from *Acacia tortilis* subsp. *raddiana* in Africa, *Syst. Appl. Microbiol.* 25 (2002) 136–145.
- [4] L.R. Barran, E.S.P. Bromfield, D.C.W. Brown, Identification and cloning of the bacterial nodulation specificity gene in the *Sinorhizobium meliloti* – *Medicago laciniosa* symbiosis, *Can. J. Microbiol.* 48 (2002) 765–771.
- [5] G. Bena, Molecular phylogenetic support the morphologically based taxonomic transfer of the “*medicagoid*” *Trigonella* species to *Medicago*, *L. Plant Syst. Evol.* 229 (2001) 217–236.
- [6] G. Bena, A. Lyet, T. Huguet, I. Olivieri, *Medicago* – *Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*, *J. Evol. Biol.* (2005).
- [7] C. Boyvin, E. Giraud, Molecular symbiotic characterization of rhizobia: toward a polyphasic approach using *Nod* factor and *nod* genes, In: E. Martínez-Romero, G. Hernandez (Eds.), *Highlights of Nitrogen Research*, Kluwer Academic/Plenum Publishers, New York, 1999, pp. 295–299.
- [8] D.J. Brenner, A.C. McWhorter, J.K. Leete Knutson, A.G. Steigerwalt, *Escherichia vulneris*: a new species of *Enterobacteriaceae* associated with human wounds, *J. Clin. Microbiol.* 15 (1982) 1133–1140.
- [9] J. Brockwell, F.W. Hely, Symbiotic characteristics of *Rhizobium meliloti*: an appraisal of the systematic treatment of nodulation and nitrogen fixation interactions between hosts and rhizobia of diverse origins, *Aust. J. Agric. Res.* 17 (1966) 885–899.
- [10] B. Brunel, S. Rome, R. Ziani, J.C. Cleyet-Marel, Comparison of nucleotide diversity and symbiotic properties of *Rhizobium meliloti* populations from annual species of *Medicago*, *FEMS Microbiol. Ecol.* 19 (1996) 71–82.
- [11] W.M. Chen, S. Laevens, T.M. Lee, T. Coenye, P. De Vos, M. Mergeay, P. Vandamme, *Ralstonia taiwanensis* sp. nov., isolated from root nodules of *Mimosa* species and sputum of a cystic fibrosis patient, *Int. J. Syst. Evol. Microbiol.* 51 (2001) 1729–1735.
- [12] P. De Lajudie, A. Willems, B. Pot, D. Dewettinck, G. Maestrouan, M. Neyra, M.D. Collins, B. Dreyfus, K. Kersters, M. Gillis, Polyphasic taxonomy of rhizobia, emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov., *Int. J. Syst. Bacteriol.* 44 (1994) 715–733.
- [13] R.H. Don, P.T. Cox, B.J. Wainwright, K. Baker, J.S. Mattick, “Touchdown” PCR to circumvent spurious

- symbiont of *Aeschynomene indica*. Appl. Environ. Microbiol. 68 (2002) 1132–1136.
- [45] M.C. Villegas-Hernandez, Specieite symbiotique de *Sinorhizobium meliloti* associe a *Medicago laciniana*: Biodiversite et determinisme genetique, Ph.D. Thesis, University of Lyon, 2002, 174pp.
- [46] J.M. Vincent, A Manual for the Practical Study of Root-nodule Bacteria, IBP Handbook No. 15, Blackwell Scientific Publications Ltd., Oxford, 1970.
- [47] P. Vinuesa, M. Leon-Barrion, C. Silva, A. Willems, A. Jarabo-Lorenzo, R. Perez-Galdona, D. Werner, E. Martinez-Romero, *Bradyrhizobium canariense* sp. nov., an acid-tolerant endosymbiont that nodulates endemic genistoid legumes (Papilionoideae: Genisteae) from the Canary Islands, along with *Bradyrhizobium japonicum* bv. genistearum, *Bradyrhizobium* genospecies 2 and *Bradyrhizobium* genospecies 6, Int. J. Syst. Evol. Microbiol. 55 (2005) 569–575.
- [48] L.T. Wang, M.A. Rogel, A. Garcia-de los Santos, J. Martinez-Romero, M.A. Cevallos, E. Martinez-Romero, *Rhizobium eili* bv. minosae, a novel biovar isolated from *Mimosa affinis*, Int. J. Syst. Bacteriol. 49 (1999) 1479–1491.
- [49] R. R. Young, P.H. Croft, G.A. Sandra, Variation in flowering times and agronomic characteristics of *Medicago laciniana* (L.) Miller collected from diverse locations in New South Wales, Austral J. Exp. Agric. 32 (1992) 59–63.
- [50] K. Zribi, R. Mhamdi, T. Huguet, M.E. Aouani, Distribution and genetic diversity of rhizobia nodulating natural populations of *Medicago truncatula*, Soil Biol. Biochem. 36 (2004) 903–908.

XIII Article 2 : accepté dans « Molecular Ecology »

XIII Article 2 : accepté dans « Molecular Ecology »

Title: Recombination and selection shape the molecular diversity pattern of nitrogen-fixing *Sinorhizobium* sp. associated to *Medicago*.

Authors: Xavier Bailly,* Isabelle Olivier,* Stéphane De Mita,[†] Jean-Claude Cleyet-Marel,[‡] and Gilles Béné[†]

Current affiliation of authors:

[†]Laboratoire des Symbioses Tropicales et Méditerranéennes UMR 113 IRD-Cirad-Ensam-UM2 / USC INRA ; ^{*}Institut des Sciences de l'Évolution de Montpellier, UMR 5554 CNRS-UM2; [‡]Diversité et Génome des plantes cultivées UMR 1097 CIRAD-INRA-ENSAM-IRD

Keywords: symbiosis, mutualism, *Sinorhizobium*, *Medicago*, selection, recombination.

Submitted as a Full Paper

Corresponding author:

Xavier Bailly

Laboratoire des Symbioses Tropicales et Méditerranéennes UMR 113 IRD-Cirad-Ensam-UM2 / USC INRA; Campus de Baillarguet 34398 Montpellier Cedex 5

Phone number: 33 (0)4 67 59 38 01

Fax number: 33 (0)4 67 59 38 02

E-mail: bailly@isem.univ-montp2.fr

Running title: Diversity of *Medicago* nodulating bacteria.

Abstract

We investigate the genetic structure and molecular selection pattern of a sympatric population of *Sinorhizobium meliloti* and *Sinorhizobium medicae*. These bacteria fix nitrogen in association with plants of the genus *Medicago*. A set of 116 isolates were obtained from a soil sample, from root nodules of three groups of plants representing among-species, within-species and intra-line diversity in the *Medicago* genus. Bacteria were characterized by sequencing at seven loci evenly distributed along the genome of both *Sinorhizobium* species, covering the chromosome and the two megaplasmids. We first test whether the diversity of host-plants influence the bacterial diversity recovered. Using the same data set, we then analyze the selective pattern at each locus. There was no relationship between the diversity of *Medicago* plants which were used for sampling and the diversity of their symbionts. However, we found evidence of selection within each of the two main symbiotic regions, located on the two different megaplasmids. Purifying selection or a selective sweep was found to occur in the *nod* genomic region, which includes genes involved in nodulation specificity, whereas balancing selection was detected in the *exo* region, close to genes involved in exopolysaccharide production. Such pattern likely reflects the interaction between host-plants and bacterial symbionts, with a possible conflict of interest between plants and cheater bacterial genotypes. Recombination appears to occur preferentially within and among loci located on megaplasmids, rather than within the chromosome. Thus, recombination may play an important role in resolving this conflict by allowing different selection patterns at different loci.

Introduction

Host selective pressures and lateral gene transfers are key mechanisms that shape the genetic population structure of symbiotic microorganisms (Tibayrenc, 1996). Understanding their respective influence on the diversification of rhizobia, which fix nitrogen in symbiosis with Fabaceae plants, is critical to manage the genetic resources of these ecologically important bacteria. Among the various symbiotic couples that are used to study the Fabaceae / rhizobia interaction, *Medicago* sp. / *Sinorhizobium* sp. is a particularly interesting one (Cook, 1999). *Medicago truncatula* is an emerging model system for plants that interact with symbiotic micro-organisms, and the gene rich regions of the genome of *M. truncatula* are currently being sequenced (Young *et al.*, 2005). On the bacterial side, the complete genome sequence of the strain 1021 of *Sinorhizobium meliloti* is available (Galibert *et al.*, 2001), and the tripartite genome architecture of both *S. meliloti* and *Sinorhizobium medicae* has been described by Pulsed Field Gel Electrophoresis analyses (Roumiantseva *et al.*, 1999). Their genome comprises a chromosome and two megaplasms called pSymA and pSymB. The sizes of these different replication units are respectively 3.65 Mb, 1.35 Mb and 1.68 Mb in *S. meliloti* strain 1021. Furthermore, these replication units display distinct functional features. Indeed, whereas housekeeping genes are located on the chromosome, genes involved in secondary metabolic pathways are located on both megaplasms, pSymA harbouring also nearly all symbiotic genes.

Sinorhizobium meliloti and *S. medicae* are the sole symbionts of *Medicago* species with the exception of *Medicago ruthenica* (de Lajudie *et al.*, 1994; Rome *et al.*, 1996; Van Berkum *et al.*, 1998). More precisely, *Medicago* species can be clustered in four groups depending on their symbiotic partners: (i) species that establish efficient symbiosis with both *S. meliloti* biovar (bv) *meliloti* and *S. medicae*, such as alfalfa, *Medicago sativa*, or *M. truncatula*; (ii) species that interact solely with *S. medicae*, such as *Medicago polymorpha*; (iii) species that interact with specific *S. meliloti* strains, such as *Medicago laciniata* with *S. meliloti* bv *medicaginis* (Villegas *et al.*, 2006) or *Medicago rigiduloides* (Materon, 1991); (iv) *M. ruthenica*, which is specifically associated to a biovar of *Rhizobium gallicum* (Silva *et al.*, 2005), initially defined as *Rhizobium mongolense* (Van Berkum *et al.*, 1998). The diversity of bacteria sampled from nodules should thus depend on the plant species used. Moreover, within a species, the plant genotype may also influence the symbiotic diversity of *Sinorhizobium* sp. as suggested for *Medicago sativa* (Paffetti *et al.*, 1998). However, to our knowledge, no study has yet addressed the influence of the diversity of host plants on the rhizobial diversity that can be recovered in one soil sample.

Besides host plants, sexual behavior may also influence the genetic structure of bacterial populations. Previous studies have described genetic differentiation and restricted recombination between the two *Sinorhizobium* species (Eardly *et al.*, 1990; Maynard Smith *et al.*, 1993). More precisely, Maynard Smith *et al.* (1993) first suggested a rather limited recombination rate between two distinct clusters of *S. meliloti* strains, but a nearly panmictic (*i.e.* no linkage disequilibrium) population structure within each one. Later on, Rome *et al.* (1996) described these two clusters as two sister species, *S. medicae* and *S. meliloti*. However a recent study reported significant linkage disequilibrium among most of the molecular markers used to characterize *S. meliloti* strains isolated from *Medicago* species in Central Asia (Roumiantseva *et al.*, 2002). Such discrepancy between the patterns of linkage disequilibrium found in both studies indicates that the genetic structure of these bacterial species is not clearly assessed. Indeed the spatial distribution of rhizobial diversity is likely to have a strong impact on the estimation of their sexual behavior. Clearly, if genetic drift is

acting independently among isolated populations, it will increase linkage disequilibrium measures at the level of a group of populations (Tachida, Cockerham, 1986). For example, previous studies used this linkage disequilibrium property to study differentiation among populations of *Rhizobium leguminosarum* (Souza *et al.*, 1992) or *Helicobacter pylori* (Faush *et al.*, 2003; Wirth *et al.*, 2004). Thus, a local sampling of a rhizobium population may be useful to estimate the extent of genetic exchange at a local, ecological scale.

The purpose of the present study is to refine hypotheses about the population genetic structure of *Sinorhizobium meliloti* and *S. medicae* symbiotic bacteria, and to infer their sexual behavior at a local scale. A single soil sample was put into contact with three sets of plants including: (i) twenty different species of the genus *Medicago*, (ii) twenty different lines of the species *M. truncatula* and (iii) 20 individuals of the same fixed line of *M. truncatula*, thus representative of three levels of diversity within the *Medicago* genus. One hundred and sixteen bacterial strains were isolated from plant nodules and characterized by multilocus sequence analysis (MLSA), which is an approach akin to multilocus sequence typing (Maiden *et al.*, 1998). The aims of the survey were: (i) to examine the impact of the diversity of host-plants used to sample rhizobia on the diversity of symbiotic bacteria isolated from their nodules, (ii) to test whether selective or demographic constraints have influenced the patterns of genetic diversity in the bacterial population before the sampling experiment and (iii) to investigate the sexual behavior of isolated strains within and among replication units.

Materials and methods

Isolation of strains

One liter of top soil was sampled at Sainte Colombe l'Eglise in the south of France (longitude 02°45'26" East – latitude 42°37'59" North). The sampling site is located within a Mediterranean climate area, at the margin between an olive tree plantation and a *Quercus ilex* garrigue. The local site flora included *Medicago truncatula* and *Medicago minima* which both interacts efficiently with *S. meliloti* or *S. medicae* and *Medicago polymorpha* which forms nitrogen-fixing symbiosis solely with *S. medicae*. No *Medicago sativa* fields were observed nearby the sampling site. We selected three plants sets: (i) twenty plants belonging to twenty *Medicago* species chosen so as to represent a large phylogenetic, life-history and symbiotic diversity within the genus (Bena *et al.*, 1998a; Bena *et al.*, 2005; Bena *et al.*, 1998b) (Group 1: G1), (ii) twenty plants belonging to twenty *M. truncatula* lines chosen so as to represent a wide range of geographic origins (G2), (iii) twenty plants from a single fixed genotype, *Medicago truncatula* Jemalong A17 (G3). The soil sample was homogenized before the experiment, and aliquots were put into contact with seed germinations (Table 1). Seeds were surface sterilized with calcium hypochlorite 5% (weight/volume) for 5 minutes and rinsed with sterile water. Seeds were then germinated for 72 hours on 1% (w/v) agar medium. Each plant was transferred individually into a Gibson tube. Gibson tubes were previously nine tenth filled with vermiculite and 25 ml of nitrogen-free plant nutrient solution. Tubes were then autoclaved, and subsequently completed by addition of one tenth of soil aliquot. The tubes were then moved into a growth chamber at 22°C day / 18°C night, with a 16-hour photoperiod and 50-60% relative humidity. After two months, symbiotic effectiveness was estimated by visual observation of plant vigor, foliage and nodules colors. All nodules were harvested. For each plant, two nodules were surface sterilized using calcium hypochlorite 1% (w/v) for three minutes, then rinsed three times in sterile water. Isolation of a single bacterial strain per nodule was performed by three successive sub-cultures on YEMA solid medium (Vincent,

1970) of isolated colony starting from crushed nodules. Bacterial strains were preserved at -80°C in YEM medium supplemented with half volume of 60% (volume/volume) glycerol.

MLSA marker design

We focused on seven different intergenic spacers (IGS) evenly distributed along the *Sinorhizobium* genome (Fig. 1, Table 2), thus allowing the measurement of linkage disequilibrium both within and among replication units. Selection on closely-linked coding sequences, such as symbiotic genes, should influence IGS polymorphism through hitch-hiking effects (Maynard Smith, Haigh, 1974), allowing us to detect the traces of selection events occurring at linked coding sequences, even with non-coding sequences. Conversely, non-coding, IGS sequences, might be more polymorphic than coding loci ones, such as the *recA* gene previously used to study the diversity of *Sinorhizobium* sp. (Bena *et al.*, 2005).

The seven markers used here included three chromosomal sequences (*IGS_{RKP}*, *IGS_{FUM}*, and *IGS_{HSS}*), two pSymA sequences (*IGS_{NOD}*, *IGS_{GAB}*) and two pSymB sequences (*IGS_{EXO}*, *IGS_{THU}*). The main bacterial metabolic pathways involved in symbiosis have been extensively studied (Perret *et al.*, 2000). These pathways can be categorized in two classes that have been proven to influence symbiotic phenotype: (i) genes involved in nodule organogenesis and the synthesis of Nod-Factors (NF), such as *nod*, *noe*, *nol* genes (Spaink, 2000); (ii) genes coding for the synthesis of surface polysaccharides (LPS, EPS, KPS), such as *evo* (Reuber, Walker, 1993) or *rkp* (Kereszt *et al.*, 1998) genes which limit plant defences (Mithofer, 2002). Thus, we chose *IGS_{RKP}*, *IGS_{NOD}* and *IGS_{EXO}* loci, which are included in gene clusters involved in the symbiotic interaction. By contrast, *IGS_{FUM}*, *IGS_{HSS}*, *IGS_{GAB}*, and *IGS_{THU}* markers are physically close to genes involved in housekeeping or secondary metabolic pathways.

These loci, which lengths are nearly 600pb, are not *a priori* involved in regulatory functions. The primers we designed are usually located within the 50 base pairs (bp) of coding sequences adjacent to each IGS by using homologous regions among the full sequenced *S. meliloti*, *Brucella melitensis*, *Mesorhizobium loti*, *Rhizobium leguminosarum* or *Agrobacterium tumefaciens* genomes, using GeneBank database (Table 2). Consequently, no coding sequence was included in our dataset with the exception of *IGS_{EXO}*, *IGS_{THU}* and *IGS_{GAB}* which include respectively 43 bp of *exoP*, 94 bp of *thuR* and 173 bp of *Sma1850*.

Nucleotide sequencing

Each bacterial isolate was grown in 20 ml YEM medium. We also included in our study the *S. medicae* type strain A321^T. One milliliter of liquid culture was washed twice in an Eppendorf micro-tube by centrifugation (15000g, 4 min) and the pellet re-suspended in 750 µl of sterile water. One hundred microliters of this solution was incubated for two hours with 20 µl of 1 mg.ml⁻¹ proteinase K and 100 µl of tris-HCl (10 mM, pH 8.3). After boiling, this mixture was used as DNA template. DNA amplification was performed using a Perkin-Elmer 2400 thermocycler in 25 µl volume, including 1 µl of DNA template, 200 µM of each dNTP, 0.8 µM of each primer, 1.5 mM of MgCl₂, 1X buffer supplied by the *Taq* polymerase manufacturer, and 1.25 U of Invitrogen *Taq* polymerase. We used a touch-down program including an initial denaturation stage (96°C, 4 min), 20 cycles of denaturation (96°C, 30 s), annealing (annealing temperature decrease steadily from 60°C to 50°C in 20 cycles, 30 s), elongation (72°C, 1 min) and 20 cycles of denaturation (96°C, 30 s), annealing (50°C, 30 s) and elongation (72°C, 1 min). PCR products were cut out from 1% agarose electrophoresis gel, purified with a QIAquick Gel extraction kit (Quiagen) and sequenced on one strand.

Sequencing reactions were performed using the ABI Prism BigDye Terminator Cycle sequence Kit (Applied Biosystem) and analysed on an Applied Biosystem model 310 DNA sequencer.

Sequence analysis and distance based approaches

Nucleotide sequence alignments were performed using CLUSTAL X 1.83 (Thompson *et al.*, 1997) and edited manually with BIOEDIT 5.0.9 (Hall, 1999). We chose an appropriate model of sequence evolution for each locus using the likelihood ratio test (LRT) implemented in MODELTEST 3.7 (Posada, Crandall, 1998). Distance matrices were thus computed for each of the 7 loci with PAUP* 4b10 (Swofford, 2003). An averaged distance matrix was then computed by obtaining an unweighted average of the 7 matrices, and was used in two ways. First, mean pairwise divergence between isolates sampled from G1, G2 and G3 (see Table 1) were calculated. Ninety-five percent confidence interval of this measure was estimated by bootstrapping individuals 1000 times in each group. Second, a molecular network was obtained by Neighbor-Net (Bryant, Moulton, 2004) using SPLITSTREE 4 (Huson, Bryant, 2006). The Neighbor-Net method is an agglomerative algorithm derived from Neighbor-Joining (Saitou, Nei, 1987) in order to represent conflicting phylogenetic signals in sequence dataset, due to recombination events or homoplasy. Moreover, we also used the network to examine whether isolates clustering could be related to taxonomy and sampling-plant sets G1, G2 or G3. Neighbor-Net algorithm inferred a fully-resolved network, with no information on the robustness of each branch. To limit this drawback, we applied a filter on splits based on a length threshold, according to Wilkinson *et al.* point of view (2003). We chose the threshold value so that the length of the graphical representation represents 95% of the global network one.

Statistical analysis of diversity

The diversity of *S. melitoni* and *S. medicae* bacterial populations isolated from G1, G2, G3 and the overall dataset was estimated at each locus using two summary statistics: allelic richness R , and Nei's diversity index H (Nei, 1987). We used a randomization procedure to cope with differences in isolate number among groups. One thousand pseudo-replicates of the original dataset were obtained by bootstrapping $N_{\min}$ bacterial genotypes, where $N_{\min}$ is the minimum sample size observed among G1, G2 and G3. R and H expected values for G1, G2 and G3 and their 95% confidence intervals were then estimated to compare the diversity of each group.

Analysis of population differentiation

Differentiation level among bacterial populations isolated from G1, G2 and G3 was estimated by F_{ST} values, according to Weir and Cockerham's (1984) procedure implemented in FSTAT 2.9.3.2 (updated from Goudet, (1995). Exact-tests of genotypic differentiation (Raymond, Rousset, 1995a) were also performed with GENEPOP 3.3 (Raymond, Rousset, 1995b), to test the hypothesis that haplotype frequencies in each bacterial group were not significantly different.

Detection of selection

Because the different molecular markers assayed in this study are almost only non-coding intergenic loci, we used Tajima's D statistic (Tajima, 1989) to detect traces of selective pressures. The quantity D measures the difference between two estimators of θ , the mutation rate scaled by the effective population size. These estimators are respectively obtained from

the average number of pairwise nucleotide differences and from the number of segregating sites. The D statistic is standardized by the variance of this difference. The observed Tajima's D value is compared to the distribution of D expected under the neutral evolutionary model, which assumes that polymorphisms segregate at mutation drift equilibrium.

A significantly positive D value may suggest balancing selection, for which highly divergent variants display even frequencies in population. Conversely, a significantly negative D value may indicate purifying selection or a selective sweep, for which the allele frequency spectrum displays an excess of weakly divergent rare alleles. However, these two patterns (D positive or negative) may alternatively be due to demographic processes that would affect the whole genome. In this latter case, similarly positive or negative D values should be observed for all loci studied. Tajima's D was computed, and departure from the neutral model was tested for each species at each locus using DNASP 3.51 (Rozas *et al.*, 2003).

Linkage disequilibrium (LD) analyses

We compared the degree of non-random association of genotypes within and among replication units. In this context, we used association measures that compare pairs of loci rather than a multilocus statistic such as the association index (Maynard Smith *et al.*, 1993), in order to distinguish levels of recombination between each pair of loci, rather than an average measure over the whole genome. We used two different measures for estimating the LD, D and D' (Lewontin, 1964). These two pairwise LD statistics were calculated for each pair of loci A-B. To obtain these statistics, D_{ij} and D'_{ij} values were computed for each pair of alleles A_i - B_j at loci A and B respectively. Then, absolute values of D_{ij} and D'_{ij} were weighted by the product of A_i and B_j allele frequencies and summed over allele combinations to achieve computation of, respectively, D_{AB} and D'_{AB} . Ninety five percent confidence intervals of D_{AB} and D'_{AB} were estimated by bootstrapping bacterial genotypes. With this aim, distribution of D and D' values were obtained from 1000 pseudo-replicates. The classical "D" statistic requires both particular allele frequencies and completely correlated polymorphisms to reach its maximum value. It is thus strongly dependent upon recombination and mutation histories at both loci. Conversely, the D' statistic is a relative measure of LD that is weakly dependent upon allele frequencies, and instead strongly depends upon recombination history (Hedrick, 1987). We also used the exact test procedure implemented in GENEPOP (Raymond, Rousset, 1995b) to test the hypothesis that alleles at different loci were not associated, which is expected if there is a strong incidence of sex. A Bonferonni adjustment correction for multiple testing was applied based on the number of intraspecific comparisons.

Results

Sampling of bacteria

Two bacterial isolates were obtained from nearly each of the 60 plants used in the trapping experiments. The collection of isolates included a total of 116 strains (Table 1). With regards to symbiotic efficiency, three species (*Medicago laciniata*, *Medicago noeana* and *Medicago rigiduloides*) were unable to fix nitrogen with bacteria of the soil sample they were in contact with. However, bacterial strains isolated from their nodules were used in further analyses. All other genotypes of plant showed evidence of efficient nitrogen fixation after two months (green foliage, pink nodules), unlike negative controls with no soil.

Nucleotide sequence accession numbers

All sequences have been deposited in the GeneBank database (*IGS_{RKP}*: DQ405768-DQ405882; *IGS_{FUM}*: DQ406230-DQ406341; *IGS_{HSS}*: DQ405289-DQ405402; *IGS_{NOD}*: DQ405403-DQ405466; DQ405593-DQ405641; *IGS_{GAB}*: DQ406342-DQ406392; DQ406519-DQ406567; *IGS_{EXO}*: DQ405994-DQ406103; *IGS_{THU}*: DQ405883-DQ405993).

Sequence dataset

We obtained 790 sequences over all seven markers (Table 1). These sequences allowed us to assess 98 complete genotypes, fifteen genotypes for which a single sequence was missing and three genotypes for which more than one sequence were missing. The 116 strains clustered into two genetic groups based on sequence homology observed after sequence alignments. These groups were identified as *S. meliloti* and *S. medicae* by sequence comparison with *S. meliloti* strain 1021 and *S. medicae* strain A321^T respectively (data not shown). Overall, 52 and 64 isolates were assigned to *S. meliloti* and *S. medicae* respectively, with no evidence of inter-specific recombination. Sequence datasets displayed different patterns of polymorphism depending on the loci (alignments available in the supplementary material section). In particular, we observed three patterns of polymorphism. First, *IGS_{EXO}* alleles of *S. meliloti* and *S. medicae* are obviously divergent with several insertions / deletions (identity scores 20%). Indel blocs were actually not numerous enough to induce alignment uncertainty at both intraspecific and interspecific levels for all markers but *IGS_{EXO}*. The alignment of *S. meliloti* and *S. medicae* sequences at *IGS_{EXO}* might be affected by indel events. Second, within *S. meliloti*, two well divergent sequence clusters were found for *IGS_{EXO}* and *IGS_{GAB}* (respective identity scores 92% and 93%). Finally, nine *IGS_{GAB}* alleles of *S. medicae* isolates contained an insertion sequence (IS). BlastN analysis revealed that this IS was homologous to Atu4604 which was annotated in *Agrobacterium tumefaciens* strain C58 genome as an IS3 family transposase (91% of DNA sequence identity on 350 pb). These nine *IGS_{GAB}* alleles were removed for distance-based and selection analyses. Sequences that differed from others by at least one nucleotide position were considered as new alleles. We considered that a new allele which was shared by at least 2 individuals was a true allele (i.e. no sequencing error). No sequencing error was revealed by reamplifying and resequencing singletons. Among the 72 alleles we obtained, 40 were shared by at least two individuals (see table 1).

Distance based analyses

Likelihood ratio tests selected various models of sequence evolution for the different markers: JC69 (Jukes, Cantor, 1969) for *IGS_{RKP}*; K80 (Kimura, 1980) for *IGS_{FUM}*; HKY (Hasegawa *et al.*, 1985) for *IGS_{HSS}*, *IGS_{NOD}* and *IGS_{THU}*; HKY + Γ_4 for *IGS_{EXO}* and *IGS_{GAB}*. It is noteworthy that model choices are conditioned to the polymorphism of the different datasets. After averaging the seven distance matrices, a Neighbor-Net split graph was built. The distribution of network length versus split length threshold was obtained (Fig. 2). We decided to apply a threshold of 3.6×10^{-4} to the graphical representation for a substantial gain of clarity (124 splits out of 181 were removed) without losing much information (the length of the final network was 95% of the original split graph length).

Species delineation

The network included strains belonging to both *S. meliloti* and *S. medicae*. It illustrated the clear delineation of the two species (Fig. 3). The split between the two species displayed the highest weight within the network. In agreement, *S. meliloti* and *S. medicae* strains did not

share any alleles and the two species were significantly differentiated (global test p value < 0.001).

Population structure

In the split graph, isolates from the same species (either *S. meliloti* or *S. medicae*) did not cluster according to host plant set (Fig. 3). Moreover, average pairwise genetic distances within G1, G2 and G3 were equivalent although values for *S. meliloti* were higher than those for *S. medicae* (Fig. 4). In the same way, diversity statistics R and H did not show significant differences among the three bacterial groups (Table 3). Finally, overall F_{ST} values calculated among strains isolated from the three groups were -0.016 (Not Significant), and -0.001 (NS) for *S. meliloti* and *S. medicae* respectively. Within either *S. meliloti* or *S. medicae*, no pairwise F_{ST} value between groups was significantly different from 0 ($p > 0.05$). Locus-by-locus analysis gave similar results (data not shown).

Inferences of selection

Tajima's D statistics for IGS_{RKP} , IGS_{FUM} , IGS_{HSS} , IGS_{GAB} and IGS_{THU} ranged from -1.409 to 1.335 and did not depart significantly from the neutral model for either *S. meliloti* or *S. medicae* (Table 4). The most negative D value was found for IGS_{NOD} marker, for both *S. meliloti* and *S. medicae*, deviating significantly from neutrality for the latter species (Table 4). In contrast, the polymorphism pattern for IGS_{EXO} locus resulted in the most positive D values for both *Sinorhizobium* species, with *S. meliloti* D value differing significantly from the neutral model expectations (Table 4). Quite surprisingly, even though most D values were not significant, we observed a strong correlation ($R^2 = 0.70$ $p < 0.05$) between D values obtained at each locus between the two species.

Evidences for Recombination

The two species clusters were connected by a long edge in the network, suggesting the lack of recombination among the two species. In contrast, clusters of each species included reticulations (Fig. 3). Only three *S. meliloti* strains, namely STM 2747, STM 2772 and STM 1643, stood out from others in a dichotomic structure by a single edge. This atypical clustering might indicate that these strains have not yet recombined with other strains of the population, and might descend from a recent immigrant genotype. Since our purpose was to infer the impact of recombination on a population that has mostly reached migration/recombination equilibrium, these three strains were discarded before linkage disequilibrium analyses, since migration events can strongly bias LD estimates. For each bacterial species, we obtained LD measures D and D' and association exact test p -values (Table 5). D versus D' values are presented in Fig. 5.

Analyses performed on the *S. meliloti* dataset indicated that molecular markers clustered into two groups depending on their linkage disequilibrium pattern: first, a chromosomal marker group including IGS_{RKP} , IGS_{FUM} and IGS_{HSS} , for which pairwise tests revealed significant departure from the null hypothesis of linkage equilibrium; second, a plasmidic marker group including IGS_{NOD} , IGS_{GAB} , IGS_{EXO} and IGS_{THU} . Indeed, pairwise tests that involve plasmidic markers (either plasmidic / chromosomal or plasmidic / plasmidic comparisons) were almost consistent with the linkage equilibrium hypothesis. Despite the large dispersion of D and D' 95% confidence intervals, a clear differentiation appeared between the two groups. This difference is more obvious in D values than in D' values (Fig. 5).

Analyses of linkage disequilibrium on *S. medicae* dataset suffer from the lack of polymorphism observed at *IGS_{RKP}* (1 allele) and *IGS_{FUM}* (2 alleles including 1 singleton). Thus, linkage disequilibrium computations at the intra-chromosomal scale were unreliable. Linkage disequilibrium tests for the five other markers showed a tendency to linkage equilibrium, as observed within *S. meliloti*, excepting a highly significant and unexpected association among *IGS_{HSS}*, *IGS_{GAB}* and *IGS_{EXO}* (Table 5). The distribution of D and D' values of pairwise comparisons are similar to *S. meliloti* ones (Fig. 5).

Discussion

Is there a relationship between plant host and symbiont diversities?

Our sampling experiment suggests that some *Sinorhizobium* symbiotic lineages are rare or absent from the soil sample used in this study. Indeed, *M. laciniata*, *M. noeana* and *M. rigiduloides* did not interact with appropriate *S. meliloti* strains to initiate an effective symbiosis. In this context of a mixed *S. meliloti* by *meliloti* and *S. medicae* population, we did not observe any relationship between the diversity of plants used to sample rhizobia and the rhizobial diversity actually collected. Our analyses based on diversity statistics, measures of population differentiation and the reticulogram topology indicate that nitrogen-fixing bacterial populations isolated from 20 *Medicago* species (G1), 20 *M. truncatula* lines (G2) and *M. truncatula* Jemalong A17 (G3) displayed equivalent levels of molecular diversity for both *S. meliloti* and *S. medicae* species. This result has important consequences for the conservation genetics of *Medicago* symbionts. The exploration of symbiotic diversity within *S. meliloti* by *meliloti* or *S. medicae* at one locality might be achieved by using only a few *Medicago* genotypes as sampling host. In our case, sampling with only the fixed line *M. truncatula* Jemalong A17, which is a rather promiscuous host since it can interact with *S. meliloti* by *meliloti* and *S. medicae*, would have been sufficient to recover the entire symbiotic diversity that we observed.

This conclusion differs from the results of previous *in situ* experiments, in which *Medicago sativa* cultivars had a significant effect on the symbiont diversity recovered (Paffetti *et al.*, 1998) or physiological studies that described potential molecular determinants of differential symbiotic efficiency among various *M. truncatula* lines and *S. meliloti* strains (Kiss *et al.*, 2004). Nevertheless, it is important to note that the sampling scheme of this study was not designed to detect quantitative preferences of *Medicago* plants for different symbiont genotypes.

The lack of a relationship between diversities of the two symbiotic partners might be due to two other aspects of the evolution of specificity in the *Medicago* sp. / *Sinorhizobium* sp. interaction. On one hand, even if the co-adaptation process between symbiotic partners is spatially structured (*i.e.* local coadaptation), bacterial populations that belong to the same biovar are usually in contact with different plant genotypes and plant species found in sympatry, limiting the selective pressure for their specialization. Indeed, Carelli *et al.* (2000) describe a homogenization of rhizobium populations sampled in a soil from different *M. sativa* cultivars when these latter were grown together. Their results are consistent with the hypothesis that *S. meliloti* populations, when in contact with different *Medicago* genotypes, evolve to become less specialized, *i.e.* generalist. On the other hand, species groups based on symbiotic abilities are paraphyletic in the *Medicago* phylogeny (Bena *et al.*, 2005). Symbiotic abilities to fix nitrogen either solely with *S. medicae* or with both *S. medicae* and *S. meliloti*

evolved through several reversion or evolutionary convergence events. This may explain the lack of correlation between the genetic divergence among plant species and the genetic divergence of their symbiotic rhizobia.

*How is selection acting on *S. meliloti* and *S. medicae*?*

Most loci that we examined seem to evolve under a neutral evolutionary model, contrary to *IGS_{NOD}* and *IGS_{EXO}* that showed evidence of selection, based on Tajima's D statistic (Table 4). *IGS_{NOD}* locus is located on the symbiotic island of the pSymA megaplasmid, harboring symbiotic *nod*, *nif* and *fix* genes. *IGS_{NOD}* locus is characterized in both *S. meliloti* and *S. medicae* by negative Tajima's D values which depart significantly from neutral expectation for *S. medicae*. Since only one marker has a significantly negative value, purifying selection or a selective sweep is the most likely explanation for this result since a demographic event would affect the polymorphism of the whole genome. Analyses based on the R2 statistic (Ramos-Onsins, Rozas, 2002) performed with DNASP (Rozas *et al.*, 2003) also did not provide any evidence for a demographic event influencing the diversity of each species (analyses not shown). We thus hypothesize that purifying selection or selective sweeps on *nod* genes has influenced *IGS_{NOD}* polymorphism by hitch-hiking. Parker (1999) proposed that in mutualisms, contrarily to antagonistic interactions, individuals mutated in symbiotic recognition pathways would present a lower fitness compared to others. Furthermore, a recent study on *osp* gene of symbiotic Rickettsiaceae, which encodes outer membrane proteins, showed that some sites of sequences of parasitic lineages evolved under positive selection while sequences of mutualistic lineages evolve under purifying selection (Jiggins *et al.*, 2002). Thus, Parker's proposition and the results of the latter study suggest that the negative Tajima's D we observed at the *IGS_{NOD}* might reflect purifying selection, rather than a selective sweep.

Conversely, a positive Tajima's D statistic was estimated from *IGS_{EXO}* locus in both *S. meliloti* and *S. medicae*, and the polymorphism pattern of *S. meliloti* strains differed significantly from neutrality ($p < 0.001$). As for *IGS_{NOD}* locus, a demographic event scenario is not the most plausible explanation for such pattern since it is not shared by other loci. Furthermore, no population sub-structure, which could also bias Tajima's D at all loci towards positive values, was inferred from the split graph and differentiation tests. Running STRUCTURE software (Pritchard *et al.*, 2000), also did not provide any evidence for substructure within each species (analyses not shown). These observations suggest that balancing selection is acting on the *S. meliloti* *IGS_{EXO}* region. Several gene clusters involved in surface exopolysaccharides (EPS) production are physically linked to *IGS_{EXO}*, as *exo* or *exp* genes (Fig. 1). Previous physiological studies pointed out that these genes were involved in the symbiotic efficiency of *S. meliloti* (Battisti *et al.*, 1992; Gonzalez *et al.*, 1996a; Gonzalez *et al.*, 1996b), by limiting plant defenses during the interaction (see Frayssé *et al.* (2003) for a review). Moreover, a recent work on molecular evolution of numerous bacterial genes demonstrated that balancing selection is a common feature of genes that encode surface proteins in parasitic bacteria, which may be recognized by eukaryotic hosts (Hughes, 2005). For instance, adaptation to different host genotypes explains polymorphism patterns such as observed for *ospC* gene among populations of the parasitic bacteria *Borrelia burgdorferi* (Brisson, Dykhuizen, 2004).

The difference between selective pressures that were inferred from *IGS_{NOD}* and *IGS_{EXO}* might illustrate the 'dilemma' of symbiotic organisms (see Van Baalen, Jansen (2001) about the concept of alignment of interest of symbiotic organisms). Indeed, rhizobium fitness depends

on its ability to both (i) interact with plants, which may induce purifying selection and selective sweeps on *nod* genes and (ii) escape plant sanction (West *et al.*, 2002), which may involve balancing selection on EPS metabolic pathway genes. Thus, one possible scenario is that early recognition (i.e. NF mediated recognition) of the host induces purifying/positive selection, and that bacteria has subsequently to thwart host sanctions (Kiers *et al.*, 2003) via defense responses at a late recognition step through a negative frequency dependent selection process (i.e. balancing selection) acting on genes involved in EPS metabolic pathway.

We observed a rather high correlation between Tajima's D values calculated for each *Sinorhizobium* species ($R^2=0.70$, $p<0.05$). Such correlation suggests that similar selective pressures act on both species for each locus. Sharing nearly the same ecological niche (both soil conditions and host plants), similar genome architecture (Roumiantseva *et al.*, 1999) and perhaps similar gene location within replication units (the sequencing of the genome of *S. medicae* strain WSM 419 will provide insights into this point) lead the two species to evolve under similar selective constraints.

Influence of recombination on S. meliloti and S. medicae population structure:

The bacterial sampling scheme used in the present work allows studying *S. meliloti* and *S. medicae* sexual behavior at a local scale. In our dataset, *S. meliloti* by *meliloti* and *S. medicae* populations present strong evidence of recombination. Within each species, several pairs of loci did not depart significantly from linkage equilibrium. This is consistent with the genetic structure observed among *S. meliloti* biovars (Bailly, unpublished). Further, the sexual isolation between *S. meliloti* and *S. medicae* support the hypothesis that *bona-fide* evolutionary species can be delineated in the rhizobia (Vinuesa *et al.*, 2005).

More precisely, our analyses suggest that recombination rates depend on the location of markers in the genome, at least for the *S. meliloti* population (Fig. 5). Chromosomal markers were in strong linkage disequilibrium for *S. meliloti* but were not informative enough to provide reliable conclusions on chromosomal association patterns in *S. medicae*. In contrast, LD measures indicated that for both species, plasmidic loci are independent, both within and among replication units. This difference is more obvious in D values than in D'. This might be due to the linkage disequilibrium among chromosomal markers but also to the similarity of polymorphism patterns observed at chromosomal loci. Indeed, D values are influenced by polymorphisms patterns whereas D' values are not (Hedrick, 1987). Previous studies showed that *S. meliloti* megaplasmids harbour different molecular determinants involved in the conjugation process, such as several *tra* loci (Galibert *et al.*, 2001) (Fig. 1). Moreover, several conjugative transfer origins (*oriT* and *mob*) have been identified on the two megaplasmids (Herrera-Cervera *et al.*, 1998), especially near the symbiotic island on pSymA. These elements may be parts of a conjugative system that allows horizontal gene transfers. However, the mechanisms controlling such transfers remain unclear. Indeed, to our knowledge, no laboratory experiment has ever reported partial or complete transfers of megaplasmids between either *S. meliloti* or *S. medicae* strains, and these megaplasmids seem to be far too large to be transmitted by "classical" conjugation.

The linkage disequilibrium pattern that we observed raises questions about the adaptive nature of the genome architecture in bacteria. The *S. meliloti* and *S. medicae* genomes include gene clusters involved in housekeeping functions, symbiotic interactions or secondary metabolic pathways. These gene sets may be defined as modules located on different replications units. Two main reasons for the emergence of modularity have been suggested (see McAdams *et al.*

(2004) for a review). First, modular organization might enhance bacterial fitness by facilitating resource allocation to different physiological functions, by regulating their expression. The accumulation of regulatory genes in large bacterial genomes, such as those of rhizobia (Konstantinidis, Tiedje, 2004), is congruent with this hypothesis. Second, horizontal transfers allow the independent evolution of each module, and different selective pressures to act independently on different parts of the genome. The high rates of gene transfer suggested by our data, together with the contrasted patterns of polymorphism observed at *IGS_{NOD}* and *IGS_{EXO}* loci, are consistent with this hypothesis.

Until now, three major studies have focused on the impact of sex on *S. meliloti* and *S. medicae* diversity (Eardly *et al.*, 1990; Maynard Smith *et al.*, 1993; Roumiantseva *et al.*, 2002). Several discrepancies between these studies and ours lead to three main comments. First, linkage disequilibrium tests, based on the statistic I_A , performed with the formula improved by Haubold *et al.* (2000; 1998), from the *S. meliloti* genotypes data set published by Eardly *et al.* (1990), are in agreement with those computed by Maynard Smith *et al.* (1993). Both results suggest an epidemic population structure for *S. meliloti*, as evidenced by linkage disequilibrium among multilocus genotype on the entire data set, but linkage equilibrium when computations are performed on a dataset containing a single individual per multilocus genotype. Thirteen out of fourteen enzymes used by Eardly *et al.* (1990) are encoded by chromosomal genes. This pattern of linkage disequilibrium partly agrees with our results, since we observed significant linkage disequilibria among *S. meliloti* chromosomal markers. However, the limited number of multilocus genotypes obtained in the present studies prevents us from studying linkage disequilibrium on a dataset containing a single individual per multilocus genotype. Second, Maynard Smith *et al.* (1993) conversely found a putative panmictic structure in *S. medicae* populations. However, only three loci out of fourteen displayed possible horizontal gene transfer, eight being monomorphic or polymorphic due to singletons and three loci presenting a pattern of polymorphism consistent with a linkage disequilibrium hypothesis. Among the first three loci, one is located on pSymB in *S. meliloti* strain 1021 genome. Since we could not estimate any LD measure for chromosomal markers from our data set, the pattern of linkage disequilibrium within *S. medicae* chromosome remains unclear. Finally, Roumantsieva *et al.* (2002) found LD among chromosomal and plasmidic markers, but, as suggested by the authors, it likely reflected a cross-population sampling. Indeed, strain sets isolated from three main sampling sites displayed significant differentiation ($p < 0.05$ for global genotypic differentiation test performed on populations isolated from location 3, 4 and 5 in Roumantsieva *et al.* study (2002)), thus preventing the estimation of local levels of gene transfers.

Conclusion

We did not find any relationship between the diversity of *Medicago* plants which were used to sample nitrogen fixing bacteria and the diversity of their rhizobium populations. However, we inferred purifying selection or a selective sweep at *IGS_{NOD}* and balancing selection at *IGS_{EXO}*. We hypothesize that this pattern reflects the interaction between host-plants and bacterial symbionts with a possible conflict of interest among *Sinorhizobium* sp. genes and between symbiotic partners. Recombination plays an important role in resolving these conflicts, and thus in shaping genetic variability and the ability to respond to selection in both *S. medicae* and *S. meliloti*. Recombination appears to be effective preferentially within and among *Sinorhizobium* sp. megaplasmids, which carry genes important for symbiosis. It is possible that this pattern allows a more efficient response to selection pressures related to the symbiosis. Further studies on *S. meliloti* and *S. medicae* population genetics are needed to get

a better understanding in both species sexual behavior and evolutionary implications of the observed patterns of polymorphism.

References

- Battisti L, Lara J, Leigh J (1992) Specific oligosaccharide form of the *Rhizobium meliloti* exopolysaccharide promotes nodule invasion in alfalfa. *Proceedings of the national academy of sciences of the United States of America* **89**, 5625-5629.
- Bena G, Lejeune B, Prosperi JM, Olivieri I (1998a) Molecular phylogenetic approach for studying life-history evolution: the ambiguous example of the genus *Medicago* L. *Proceedings of the royal society of London, Series B* **265**, 1141-1151.
- Bena G, Lyet A, Huguet T, Olivieri I (2005) *Medicago* - *Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. *Journal of evolutionary biology* **18**, 1547-1558.
- Bena G, Prosperi JM, Lejeune B, Olivieri I (1998b) Evolution of annual species of the genus *Medicago*: a molecular phylogenetic approach. *Molecular phylogenetics and evolution* **9**, 552-559.
- Brisson D, Dykhuizen DE (2004) *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics* **168**, 713-722.
- Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution* **21**, 255-265.
- Carelli M, Gnecchi S, Fancelli S, et al. (2000) Genetic diversity and dynamics of *Sinorhizobium meliloti* populations nodulating different alfalfa cultivars in Italian soils. *Applied and environmental microbiology* **66**, 4785-4789.
- Cook DR (1999) *Medicago truncatula*--a model in the making! *Current opinion in plant biology* **2**, 301-304.
- de Lajudie P, Willems A, Pot B, et al. (1994) Polyphasic taxonomy of rhizobia: emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov. *International journal of systematic bacteriology* **44**, 715-733.
- Eardly BD, Materon LA, Smith NH, et al. (1990) Genetic structure of natural populations of the nitrogen-fixing bacterium *Rhizobium meliloti*. *Applied and environmental microbiology* **56**, 187-194.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.
- Frayse N, Couderc F, Poinot V (2003) Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. *European journal of biochemistry* **270**, 1365-1380.
- Galibert F, Finan TM, Long SR, et al. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293**, 668-672.
- Gonzalez JE, Reuhs BL, Walker GC (1996a) Low molecular weight EPS II of *Rhizobium meliloti* allows nodule invasion in *Medicago sativa*. *Proceedings of the national academy of sciences of the United States of America* **93**, 8636-8641.
- Gonzalez JE, York GM, Walker GC (1996b) *Rhizobium meliloti* exopolysaccharides: synthesis and symbiotic function. *Gene* **179**, 141-146.
- Goudet J (1995) FSTAT (vers. 1.2), a computer program to calculate F-statistics. *The journal of heredity* **86**, 485-486.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*, 95-98.
- Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of molecular evolution* **22**, 160-174.

- Haubold B, Hudson RR (2000) LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Linkage Analysis. Bioinformatics* **16**, 847-848.
- Haubold B, Travisano M, Rainey PB, Hudson RR (1998) Detecting linkage disequilibrium in bacterial populations. *Genetics* **150**, 1341-1348.
- Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* **117**, 331-341.
- Herrera-Cervera JA, Sanjuan-Pinilla JM, Olivares J, Sanjuan J (1998) Cloning and identification of conjugative transfer origins in the *Rhizobium meliloti* genome. *Journal of bacteriology* **180**, 4583-4590.
- Hughes AL (2005) Evidence for abundant slightly deleterious polymorphisms in bacterial populations. *Genetics* **169**, 533-538.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* **23**, 254-267.
- Jiggins FM, Hurst GD, Yang Z (2002) Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. *Molecular biology and evolution* **19**, 1341-1349.
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian protein metabolism* (ed. Munro HN). Academic press, New York.
- Kereszt A, Kiss E, Reuhs BL (1998) Novel *rkp* gene clusters of *Sinorhizobium meliloti* involved in capsular polysaccharide production and invasion of the symbiotic nodule: the *rkpK* gene encodes a UDP-glucose dehydrogenase. *Journal of bacteriology* **180**, 5426-5431.
- Kiers ET, Rousseau RA, West SA, Denison RF (2003) Host sanctions and the legume-rhizobium mutualism. *Nature* **425**, 78-81.
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution* **16**, 111-120.
- Kiss E, Huguet T, Poinsoit V, Batut J (2004) The *rypA* gene is required for stress adaptation as well as for symbiosis of *Sinorhizobium meliloti* 1021 with certain *Medicago truncatula* lines. *Molecular plant-microbe interactions* **17**, 235-244.
- Konstantinidis KT, Tiedje JM (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. *Proceedings of the national academy of sciences of the United States of America* **101**, 3160-3165.
- Lewontin RC (1964) The interaction of selection and linkage. I. General considerations, heterotic models. *Genetics* **49**, 49-67.
- Maiden MC, Bygraves JA, Feil E, *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the national academy of sciences of the United States of America* **95**, 3140-3145.
- Materon LA (1991) Symbiotic characteristics of *Rhizobium meliloti* in west asian soils. *Soil biology and biochemistry* **23**, 429-434.
- Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical research* **23**, 23-35.
- Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the national academy of sciences of the United States of America* **90**, 4384-4388.
- McAdams HH, Srinivasan B, Arkin AP (2004) The evolution of genetic regulatory systems in bacteria. *Nature reviews. Genetics* **5**, 169-178.
- Mithofer A (2002) Suppression of plant defence in rhizobia-legume symbiosis. *Trends in plant science* **7**, 440-444.

- Nei M (1987) *Molecular evolutionary genetics* Columbia university press, New York.
- Paffetti D, Daguin F, Fancelli S, *et al* (1998) Influence of plant genotype on the selection of nodulating *Sinorhizobium meliloti* strains by *Medicago sativa*. *Antonie van Leeuwenhoek* **73**, 3-8.
- Parker MA (1999) Mutualism in metapopulations of legumes and rhizobia. *The American naturalist* **153**, S48-S60.
- Perret X, Staehelin C, Broughton WJ (2000) Molecular basis of symbiotic promiscuity. *Microbiology and molecular biology reviews* **64**, 180-201.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular biology and evolution* **19**, 2092-2100.
- Raymond M, Rousset F (1995a) An exact test for population differentiation. *Evolution: international journal of organic evolution* **49**, 1280-1283.
- Raymond M, Rousset F (1995b) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *The journal of heredity* **86**, 248-249.
- Reuber TL, Walker GC (1993) Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Cell* **74**, 269-280.
- Rome S, Fernandez MP, Brunel B, Normand P, Cleyet-Marel JC (1996) *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. *International journal of systematic bacteriology* **46**, 972-980.
- Roumiantseva ML, Andronov EE, Sharypova LA, *et al*. (2002) Diversity of *Sinorhizobium meliloti* from the Central Asian Alfalfa Gene Center. *Applied and environmental microbiology* **68**, 4694-4697.
- Roumiantseva ML, Yakutkina VV, Damman-Kalinowski T, *et al*. (1999) Comparative analysis of structural organization of the genome in alfalfa nodule bacteria *Sinorhizobium medicae* and *Sinorhizobium meliloti*. *Russian journal of genetics* **35**, 128-135.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**, 406-425.
- Silva C, Vinuesa P, Eguarte LE, Souza V, Martinez-Romero E (2005) Evolutionary genetics and biogeographic structure of *Rhizobium gallicum sensu lato*, a widely distributed bacterial symbiont of diverse legumes. *Molecular Ecology* **14**, 4033-4050.
- Souza V, Nguyen TT, Hudson RR, Piner D, Lenski RE (1992) Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? *Proceedings of the national academy of sciences of the United States of America* **89**, 8389-8393.
- Spaink HP (2000) Root nodulation and infection factors produced by rhizobial bacteria. *Annual review of microbiology* **54**, 257-288.
- Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, Massachusetts.
- Tachida H, Cockerham CC (1986) Analysis of linkage disequilibrium in an island model. *Theoretical population biology* **29**, 161-197.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence

- alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876-4882.
- Tibayrenc M (1996) Towards a unified evolutionary genetics of microorganisms. *Annual review of microbiology* **50**, 401-429.
- van Baalen M, Jansen VAA (2001) Dangerous liaisons: the ecology of private interest and common good. *Oikos* **95**, 211-224.
- Van Berkum P, Beyene D, Bao G, Campbell TA, Eardly BD (1998) *Rhizobium mongolense* sp. nov. is one of the three rhizobial genotypes identified which nodulate and form nitrogen fixing symbiosis with *Medicago ruthenica* [(L.) Ledebour]. *International journal of systematic bacteriology* **48**, 13-22.
- Villegas MDC, Rome S, Maure L, *et al.* (2006) Nitrogen-fixing sinorhizobia with *Medicago laciniata* constitute a novel biovar (bv. *laciniata*) of *S. meliloti*. *Systematic and applied microbiology* in press. doi:10.1016/j.syapm.2005.1012.1008.
- Vincent JM (1970) *A manual for the practical study of root-nodule bacteria* Blackwell Scientific Publications Ltd., Oxford.
- Vinuesa P, Silva C, Werner D, Martinez-Romero E (2005) Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. *Molecular phylogenetics and evolution* **34**, 29-54.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution: international journal of organic evolution* **38**, 1358-1370.
- West SA, Kiers ET, Simms EL, Denison RF (2002) Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proceedings of the royal society of London. Series B* **269**, 685-694.
- Wilkinson M, Lapointe FJ, Gower DJ (2003) Branch lengths and support. *Systematic biology* **52**, 127-130.
- Wirth T, Wang X, Lutz B, *et al.* (2004) Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proceedings of the national academy of sciences of the United States of America* **101**, 4746-4751.
- Young ND, Cannon SB, Sato S, *et al.* (2005) Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. *Plant physiology* **137**, 1174-1181.

Acknowledgements

Bill Sherwin and Nicolas Galtier made very useful linguistic and scientific comments on a late version of the manuscript. We had very useful discussions with Joëlle Ronfort, Jean-Marie Prosperi and Thomas Bataillon in the early stage of this work. Jean-Marie Prosperi (INRA Montpellier) provided the plant material. Lucette Maure provided technical help for strain isolation. The research was funded by the Bureau des Ressources Génétiques, Programme « Recherches méthodologiques pour l'amélioration des processus de gestion et de conservation des ressources génétiques animales, végétales et microbiennes » to I.O. and G.B., X. Bailly was supported by a PhD fellowship from the French Ministry of Education and Research. This is publication ISEM 2006-015 of the Institut des Sciences de l'Evolution, Montpellier.

Table 1 Origin and genotypes of *S. medicae* and *S. meliloti* isolates.

Host group	Host plant	Isolates	Genotype at indicated locus ^a						
			<i>IGS_{MSI}</i>	<i>IGS_{PS}</i>	<i>IGS_{SS}</i>	<i>IGS_{MSL}</i>	<i>IGS_{MSR}</i>	<i>IGS_{MSL}</i>	<i>IGS_{MSR}</i>
G1		<i>S. medicae</i>							
		STM 2764	K	L	K	AF	U	N	Q
		STM 1647	K	L	L	AG	AB	K	T
		STM 2773	K	L	K	AE	Y	Q	T
		STM 2767	K	L	K	AF	Y	P	N
		STM 1651	K	L	L	AF	AB	K	T
		STM 2771	K	L	L	AE	U	P	P
		STM 1655	K	L	K	AF	-	N	T
		STM 1656	K	L	L	AF	AB	K	T
		STM 1649	K	L	L	AF	Y	Q	P
		STM 2770	K	L	L	AF	AB	K	T
		STM 1654	K	L	L	AD	T	K	T
		STM 2762	K	L	K	AF	U	N	T
		STM 2764	K	L	K	AE	Y	Q	T
		STM 1653	K	L	K	AE	Y	Q	S
		STM 2768	K	L	K	AF	U	P	T
		STM 1652	K	L	K	AF	U	P	T
		STM 2766	K	L	M	AF	Y	P	M
		STM 1650	K	L	K	AE	Y	Q	T
		STM 2761	K	L	L	AF	T	K	T
		STM 1644	K	L	K	AE	Y	P	S
G1		<i>S. meliloti</i>							
		STM 2776	D	E	A	K	E	D	H
		STM 1662	D	D	A	K	E	B	H
		STM 1659	C	D	A	K	-	-	H
		STM 1648	C	D	A	K	-	D	H
		STM 2775	C	D	E	I	E	B	H
		STM 1661	C	A	A	K	E	B	B
		STM 2763	D	E	E	K	E	B	H
		STM 1646	C	D	A	K	U	B	G
		STM 2765	C	D	A	-	E	B	H
		STM 1645	C	D	A	K	E	B	B
		STM 2777	D	E	E	K	E	B	-
		STM 1663	C	E	E	K	E	B	B
		STM 2772	D	F	E	J	N	D	G
		STM 1657	C	D	A	K	-	B	H
		STM 1658	D	E	E	K	E	-	B
		STM 2774	D	E	E	K	E	B	B
		STM 1660	D	E	E	K	E	B	B
G2		<i>S. medicae</i>							
		STM 2760	K	L	K	AF	AB	K	T
		STM 2756	K	L	L	AF	AB	Q	T
		STM 1639	K	L	K	AF	U	N	T
		STM 1636	K	L	K	AF	U	N	T
		STM 1637	K	L	L	AF	Y	P	T
		STM 1632	K	L	K	AF	U	P	T
		STM 1631	K	L	L	AF	Y	P	T
		STM 1630	K	L	-	AE	Y	Q	S
		STM 1633	K	L	K	AE	Y	-	Q
		STM 1628	K	L	L	AF	Y	P	K
		STM 2744	K	L	L	AC	T	K	T
		STM 1627	K	L	L	AE	Y	P	T
		STM 2746	K	-	L	AF	-	K	T
		STM 1629	K	L	K	AF	Y	P	R
		STM 2741	K	L	L	AF	Y	Q	-
		STM 2754	K	-	K	AF	Y	P	T
		STM 1642	K	L	K	AF	U	N	T
		STM 2753	K	L	K	AE	Y	K	T
		STM 2751	K	L	L	AE	AB	P	R
		STM 1634	K	L	L	AE	AB	K	T
		STM 1626	K	L	K	AF	U	N	T
		STM 2742	K	L	K	AF	U	P	T
		STM 1625	K	L	K	AF	U	N	T
G2		<i>S. meliloti</i>							
		STM 1643	D	F	F	J	N	D	G

	<i>M. truncatula</i> , DZA 045.5	STM 27.57	D	E	E	K	E	D	H
	<i>M. truncatula</i> , DZA 045.5	STM 1640	D	E	E	K	E	B	H
	<i>M. truncatula</i> , DZA 220.b	STM 27.58	C	D	A	K	E	B	-
	<i>M. truncatula</i> , DZA 220.b	STM 1641	C	D	A	K	E	D	B
	<i>M. truncatula</i> , DZA 315.16	STM 27.55	C	D	A	K	S	D	B
	<i>M. truncatula</i> , DZA 327.7	STM 27.54	D	E	E	K	F	-	B
	<i>M. truncatula</i> , ESP 105.i	STM 27.49	C	B	A	I	E	D	G
	<i>M. truncatula</i> , ESP 158.a	STM 27.45	B	C	A	G	E	D	B
	<i>M. truncatula</i> , ESP 171.f	STM 27.47	D	F	D	K	O	D	G
	<i>M. truncatula</i> , F11005.i	STM 27.50	D	E	E	H	E	D	B
	<i>M. truncatula</i> , F20047.a	STM 27.45	D	E	E	F	E	D	B
	<i>M. truncatula</i> , F83005.5	STM 1624	C	D	C	K	A	B	E
	<i>M. truncatula</i> , GRC 020.b	STM 1636	D	E	E	K	E	B	B
	<i>M. truncatula</i> , GRC 043.i	STM 27.52	C	D	A	I	E	B	H
	<i>M. truncatula</i> , GRC 043.i	STM 1635	C	D	A	K	O	B	F
	<i>S. medicae</i>								
G3	<i>M. truncatula</i> , JA 17	STM 27.23	K	L	K	AE	Y	P	S
	<i>M. truncatula</i> , JA 17	STM 27.24	K	L	L	AF	Y	P	T
	<i>M. truncatula</i> , JA 17	STM 27.25	K	L	L	AB	T	K	T
	<i>M. truncatula</i> , JA 17	STM 27.25	K	L	L	AF	Y	P	T
	<i>M. truncatula</i> , JA 17	STM 27.33	K	L	L	AE	U	Q	L
	<i>M. truncatula</i> , JA 17	STM 27.35	K	L	L	AF	-	K	T
	<i>M. truncatula</i> , JA 17	STM 27.36	K	L	L	AF	Y	P	U
	<i>M. truncatula</i> , JA 17	STM 27.37	K	L	L	AE	Y	P	T
	<i>M. truncatula</i> , JA 17	STM 27.39	K	L	K	AE	Y	P	S
	<i>M. truncatula</i> , JA 17	STM 27.40	K	L	K	AF	Y	P	N
	<i>M. truncatula</i> , JA 17	STM 1605	K	L	L	AF	-	K	T
	<i>M. truncatula</i> , JA 17	STM 1611	K	L	K	AF	U	N	V
	<i>M. truncatula</i> , JA 17	STM 1612	K	L	K	AF	Y	-	W
	<i>M. truncatula</i> , JA 17	STM 1613	K	L	L	AE	AB	Q	T
	<i>M. truncatula</i> , JA 17	STM 1615	K	L	L	AF	T	K	T
	<i>M. truncatula</i> , JA 17	STM 1616	K	L	K	AF	U	N	T
	<i>M. truncatula</i> , JA 17	STM 1617	K	K	K	AF	Y	P	R
	<i>M. truncatula</i> , JA 17	STM 1618	K	L	K	AE	Y	P	S
	<i>M. truncatula</i> , JA 17	STM 1620	K	L	L	AF	Y	S	L
	<i>M. truncatula</i> , JA 17	STM 1621	K	L	L	AF	Y	P	T
	<i>M. truncatula</i> , JA 17	STM 1622	K	L	K	AF	U	O	T
	<i>S. meliloti</i>								
G3	<i>M. truncatula</i> , JA 17	STM 27.22	D	E	E	K	E	B	G
	<i>M. truncatula</i> , JA 17	STM 27.26	C	D	A	K	C	D	H
	<i>M. truncatula</i> , JA 17	STM 27.27	C	D	A	K	C	D	H
	<i>M. truncatula</i> , JA 17	STM 27.29	D	E	E	H	E	D	B
	<i>M. truncatula</i> , JA 17	STM 27.30	D	E	E	K	E	D	C
	<i>M. truncatula</i> , JA 17	STM 27.31	C	-	A	-	E	B	-
	<i>M. truncatula</i> , JA 17	STM 27.32	C	D	A	I	E	B	H
	<i>M. truncatula</i> , JA 17	STM 27.34	D	E	E	K	E	B	H
	<i>M. truncatula</i> , JA 17	STM 27.38	C	D	A	I	O	B	H
	<i>M. truncatula</i> , JA 17	STM 1604	C	D	B	I	E	B	H
	<i>M. truncatula</i> , JA 17	STM 1606	C	D	A	K	E	B	B
	<i>M. truncatula</i> , JA 17	STM 1607	D	E	F	K	D	D	D
	<i>M. truncatula</i> , JA 17	STM 1608	C	D	A	H	E	B	H
	<i>M. truncatula</i> , JA 17	STM 1609	D	E	E	I	E	B	H
	<i>M. truncatula</i> , JA 17	STM 1610	C	D	A	K	E	B	H
	<i>M. truncatula</i> , JA 17	STM 1614	C	D	A	K	E	D	I
	<i>M. truncatula</i> , JA 17	STM 1619	D	E	E	H	E	D	B
	<i>M. truncatula</i> , JA 17	STM 1623	C	D	A	I	E	B	H

^a The absence of data is scored as a minus mark.

Table 2. Primers used for MLST of *S. meliloti* and *S. medicae*.

Surrounding ORF	Genome location ^a	Primers	Sequence 5'—3'	T _m (°C)	PCR product size (pb)
<i>rnpA</i> - <i>rnpE</i>	chromosome	RKP1	AGGCATGCACGLUUTATGA	58.8	599
		RKP2	GCCATCGACATCTACAATATCAA	57.1	
<i>hsl</i> - <i>ompP3</i>	chromosome	HSS1	TACACCGACTGGACACCGC	61.0	556
		HSS2	GATGCCGTTCCGATCGACCC	63.5	
<i>tunC</i> - <i>Y09150</i>	chromosome	FUM1	CGTGACAATCGCTGCCGA	61.4	612
		FUM2	CGCTGCGCKSGCGGCC	71.8	
<i>nodE</i> - <i>nodG</i>	pSymA	NOD1	CAGTTCTGGCATTCAAGC	53.7	579
		NOD2	CCCCCTCTATGGCTCCTGAT	61.4	
<i>gabD5</i> - <i>SmaI850</i>	pSymA	GAB1	CATGACCAAAGACCGCTTC	59.4	658
		GAB2	GCATGATCGGCCTCAACAC	58.8	
<i>exoP</i> - <i>rhiD</i>	pSymB	EXOmeliloti1	CAACAAGACGGATATGAACGAA	56.5	494
		EXOmeliloti2	GTGGTGGAAGGATTGACTGC	59.4	
		EXOmedicae1	CATGAACGAGCTGGGCAAAT	57.3	250
		EXOmedicae2	CTGGTCGAAGCGGCAAAA	56.0	
<i>rhuK</i> - <i>rhuE</i>	pSymB	THUmeliloti1	ACCAGGCGCACGGCGTTGAT	63.5	722
		THUmeliloti2	AATTCCGCGGCGGCAGCAAG	63.5	
		THUmedicae1	GCACCGCGTTGATATTG	56.0	720
		THUmedicae2	GACGATAGCGGCACAGGA	58.2	

^a Genome locations refer to the genome of *S. meliloti* strain 1021.

Table 3. Genetic diversity of *S. medicae* and *S. meliloti* population isolated from the three host plant groups used in this study^a

aps used in this study

Sample		N ^b of sequences	Genetic diversity			
			R		H	
			Obs	N _{max}	Obs	N _{max}
<i>S. medicae</i>						
IGS _{ACP}	G1	20	1	1.00 (1, 1)	0.000	0.000 (0.000; 0.000)
	G2	23	1	1.00 (1, 1)	0.000	0.000 (0.000; 0.000)
	G3	21	1	1.00 (1, 1)	0.000	0.000 (0.000; 0.000)
IGS _{HTT}	G1	20	3	2.64 (2, 3)	0.563	0.525 (0.409; 0.632)
	G2	22	2	2.00 (2, 2)	0.520	0.468 (0.404; 0.526)
	G3	21	2	2.00 (2, 2)	0.514	0.489 (0.351; 0.526)
IGS _{FTM}	G1	20	1	1.00 (1, 1)	0.000	0.000 (0.000; 0.000)
	G2	21	1	1.00 (1, 1)	0.000	0.000 (0.000; 0.000)
	G3	21	2	1.60 (1, 2)	0.095	0.094 (0.000; 0.281)
IGS _{WOB}	G1	20	4	3.22 (2, 4)	0.574	0.546 (0.351; 0.667)
	G2	23	3	2.55 (2, 3)	0.466	0.433 (0.281; 0.585)
	G3	21	3	2.63 (2, 3)	0.495	0.470 (0.292; 0.608)
IGS _{GAB}	G1	19	4	3.66 (3, 4)	0.737	0.701 (0.585; 0.772)
	G2	22	4	3.54 (3, 4)	0.688	0.649 (0.550; 0.749)
	G3	19	4	3.49 (3, 4)	0.573	0.537 (0.298; 0.696)
IGS _{ENC}	G1	20	4	3.93 (3, 4)	0.737	0.734 (0.661; 0.784)
	G2	22	4	3.92 (3, 4)	0.745	0.690 (0.602; 0.778)
	G3	20	6	5.00 (4, 6)	0.721	0.635 (0.530; 0.807)
IGS _{MTV}	G1	20	6	4.58 (3, 6)	0.579	0.549 (0.296; 0.731)
	G2	22	5	5.76 (4, 7)	0.407	0.401 (0.198; 0.602)
	G3	19	8	3.58 (2, 5)	0.719	0.685 (0.485; 0.836)
<i>S. meliloti</i>						
IGS _{ACP}	G1	17	2	1.99 (2, 2)	0.529	0.498 (0.363; 0.538)
	G2	17	3	2.58 (2, 3)	0.581	0.547 (0.385; 0.659)
	G3	18	2	1.99 (2, 3)	0.503	0.471 (0.264; 0.539)
IGS _{HTT}	G1	17	2	2.00 (2, 2)	0.640	0.606 (0.473; 0.714)
	G2	17	4	3.15 (2, 4)	0.640	0.601 (0.473; 0.714)
	G3	18	4	3.04 (2, 4)	0.608	0.577 (0.385; 0.714)
IGS _{FTM}	G1	17	4	3.15 (2, 4)	0.529	0.498 (0.362; 0.538)
	G2	17	5	3.95 (3, 5)	0.728	0.687 (0.538; 0.802)
	G3	17	2	2.00 (2, 2)	0.514	0.481 (0.363; 0.538)
IGS _{WOB}	G1	16	3	2.23 (1, 3)	0.242	0.233 (0.000; 0.473)
	G2	17	6	4.14 (3, 6)	0.588	0.554 (0.274; 0.769)
	G3	17	3	2.93 (2, 3)	0.640	0.600 (0.440; 0.703)
IGS _{GAB}	G1	14	3	2.31 (1, 3)	0.275	0.247 (0.000; 0.484)
	G2	17	5	3.56 (2, 5)	0.507	0.479 (0.143; 0.714)
	G3	18	4	2.89 (2, 4)	0.399	0.376 (0.142; 0.615)
IGS _{ENC}	G1	15	2	1.95 (1, 2)	0.343	0.318 (0.000; 0.495)
	G2	16	2	1.99 (2, 2)	0.525	0.492 (0.363; 0.538)
	G3	16	2	2.00 (2, 2)	0.503	0.474 (0.264; 0.538)
IGS _{MTV}	G1	16	3	2.85 (2, 3)	0.633	0.593 (0.440; 0.703)
	G2	16	8	4.66 (3, 6)	0.775	0.726 (0.560; 0.835)
	G3	17	6	4.23 (3, 6)	0.647	0.614 (0.385; 0.791)

^a R, allelic richness; H, Nei diversity indice; Obs, observed value; N_{min}, expected value for N_{min} sample size, limits of the 95% confidence interval are in brackets. N_{min} values are 19 for *S. medicae* and 14 for *S. meliloti*.

Table 4. Molecular diversity and Tajima's D value for genetic markers.

Species	Genetic marker	N° of aligned sites ^a	Number of			π	d_{fix}	Tajima's D
			isolates	alleles	segregating sites			
<i>S. medicae</i>	<i>IGSRKP</i>	321 (0)	54	1	0	0.0001	0.213	-1.080
	<i>IGSFUM</i>	334 (0)	62	2	1	0.0019	0.424	0.454
	<i>IGSHSS</i>	304 (0)	63	3	2	0.0025	2.740	-2.014*
	<i>IGSNOE</i>	335 (2)	64	6	13	0.0011	0.445	-0.334
	<i>IGSGAB</i>	354 (2)	51	3	2	0.0072	0.844	1.335
	<i>IGSEXO</i>	190 (1)	62	5	4	0.0023	1.893	-1.575
	<i>IGSTHU</i>	343 (3)	63	13	4			
<i>S. meliloti</i>	<i>IGSRKP</i>	302 (0)	52	3	3	0.0035	0.664	1.211
	<i>IGSFUM</i>	341 (0)	51	6	5	0.0025	1.778	-1.409
	<i>IGSHSS</i>	308 (1)	52	6	7	0.0072	1.5493	1.129
	<i>IGSNOE</i>	344 (0)	50	6	8	0.0023	1.786	-1.486
	<i>IGSGAB</i>	274 (4)	44	7	10	0.0068	2.243	-0.561
	<i>IGSEXO</i>	241 (1)	44	2	10	0.0207	2.243	3.332***
	<i>IGSTHU</i>	325 (1)	49	8	4	0.0072	2.015	0.445

^a Number of aligned sites. Number of indel blocs within intraspecific alignments is provided in brackets.

^{nt} Computation of Tajima's D were not tractable; *, ***, D values differed significantly from neutral expectations with $\alpha=5\%$ or $\alpha=0.1\%$.

Table 5. Linkage disequilibrium measures and tests of association between genetic markers^a

Genetic marker	Results for <i>S. medicae</i> ^b						
	<i>IGSRKP</i>	<i>IGSFUM</i>	<i>IGSHSS</i>	<i>IGSNOE</i>	<i>IGSGAB</i>	<i>IGSEXO</i>	<i>IGSTHU</i>
<i>IGSRKP</i>	nt	nt	nt	nt	nt	nt	nt
<i>IGSFUM</i>	nt	1.000 ^{nt}	1.000 ^{nt}	1.000 ^{nt}	1.000 ^{nt}	0.2680 ^{nt}	
<i>IGSHSS</i>	nt	1.000 ^{nt}	0.634	0.000 ^{***}	0.000 ^{**}	0.003 [*]	
<i>IGSNOE</i>	nt	1.000 ^{nt}	0.168	0.001 [*]	0.015	0.321	
<i>IGSGAB</i>	nt	1.000 ^{nt}	0.449	0.407	0.000 ^{***}	0.743	
<i>IGSEXO</i>	nt	1.000 ^{nt}	0.436	0.405	0.678	0.104	
<i>IGSTHU</i>	nt	1.000 ^{nt}	0.477	0.431	0.651	0.636	
Genetic marker	Results for <i>S. meliloti</i> ^b						
	<i>IGSRKP</i>	<i>IGSFUM</i>	<i>IGSHSS</i>	<i>IGSNOE</i>	<i>IGSGAB</i>	<i>IGSEXO</i>	<i>IGSTHU</i>
<i>IGSRKP</i>		0.000 ^{***}	0.000 ^{***}	0.008	0.239	0.198	0.376
<i>IGSFUM</i>	0.922		0.000 ^{***}	0.030	0.431	0.185	0.054
<i>IGSHSS</i>	0.827	0.860		0.500	0.004	0.786	0.018
<i>IGSNOE</i>	0.284	0.381	0.304		1.00	0.056	0.577
<i>IGSGAB</i>	0.407	0.692	0.529	0.471		0.040	0.019
<i>IGSEXO</i>	0.288	0.305	0.127	0.358	0.493		0.264
<i>IGSTHU</i>	0.382	0.562	0.439	0.424	0.651	0.404	

^a The upper half shows probabilities based on the null hypothesis of linkage equilibrium. The lower half shows linkage disequilibrium measures (D' values).

^b nt, nr, computations are respectively not tractable or not reliable due to a lack of polymorphism; *, **, ***, the null hypothesis is rejected at $\alpha=5\%$, $\alpha=1\%$ or $\alpha=0.1\%$ taking into account a Bonferroni adjustment for 15 comparisons.

^c *, **, ***, the null hypothesis is rejected at $\alpha=5\%$, $\alpha=1\%$ or $\alpha=0.1\%$ taking into account a Bonferroni adjustment for 21 comparisons.

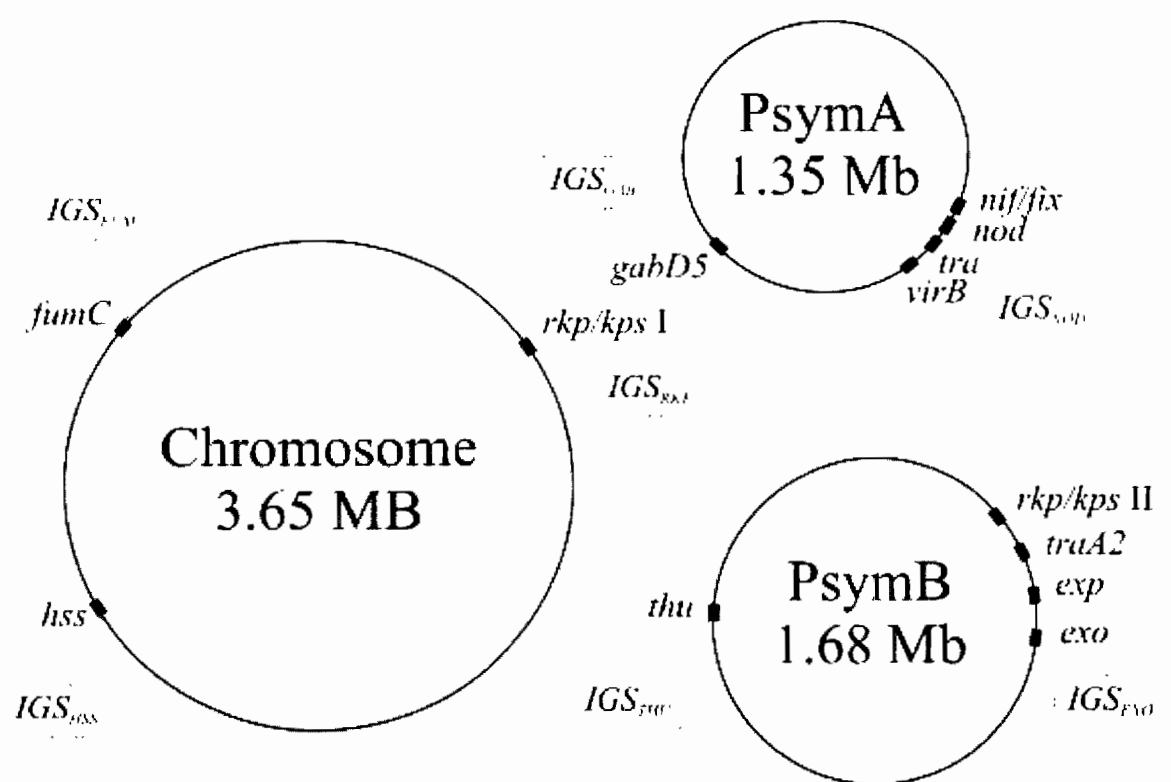


Figure 1. Location of the genetic markers used in this study on the genome of *Sinorhizobium meliloti* strain 1021. The seven loci analyzed are surrounded and their location is indicated by dotted lines. Genes clusters located nearby each genetic marker are indicated by black boxes. It is noteworthy that IGSNOD marker is located nearby genes involved in symbiosis specificity (*nod* genes), symbiosis efficiency (*nif/fix* genes), secretion (*virB* gene) and conjugation (*tra* genes) and that IGSEXO marker is located nearby gene clusters involved in surface polysaccharide production (*rkp*, *kps*, *exp* and *exo* genes) and conjugation (*tra* gene).

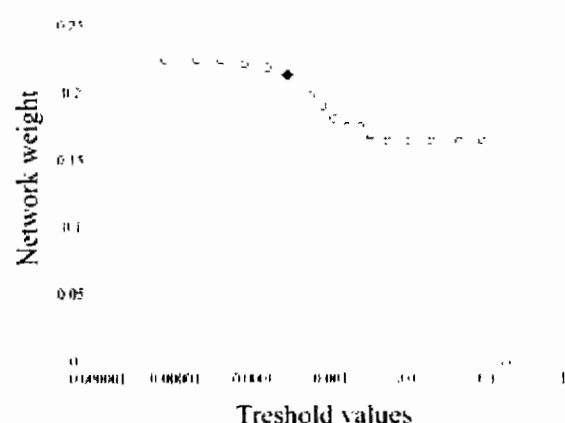


Figure. 2. Distribution of Neighbor-Net split graph length versus split length threshold values. The network length for 3.6×10^{-4} threshold value used in this study is indicated by a black rhomb. Using this threshold, 124 splits out of 181 were removed and the weight of the final network was 95% of the original split graph length.

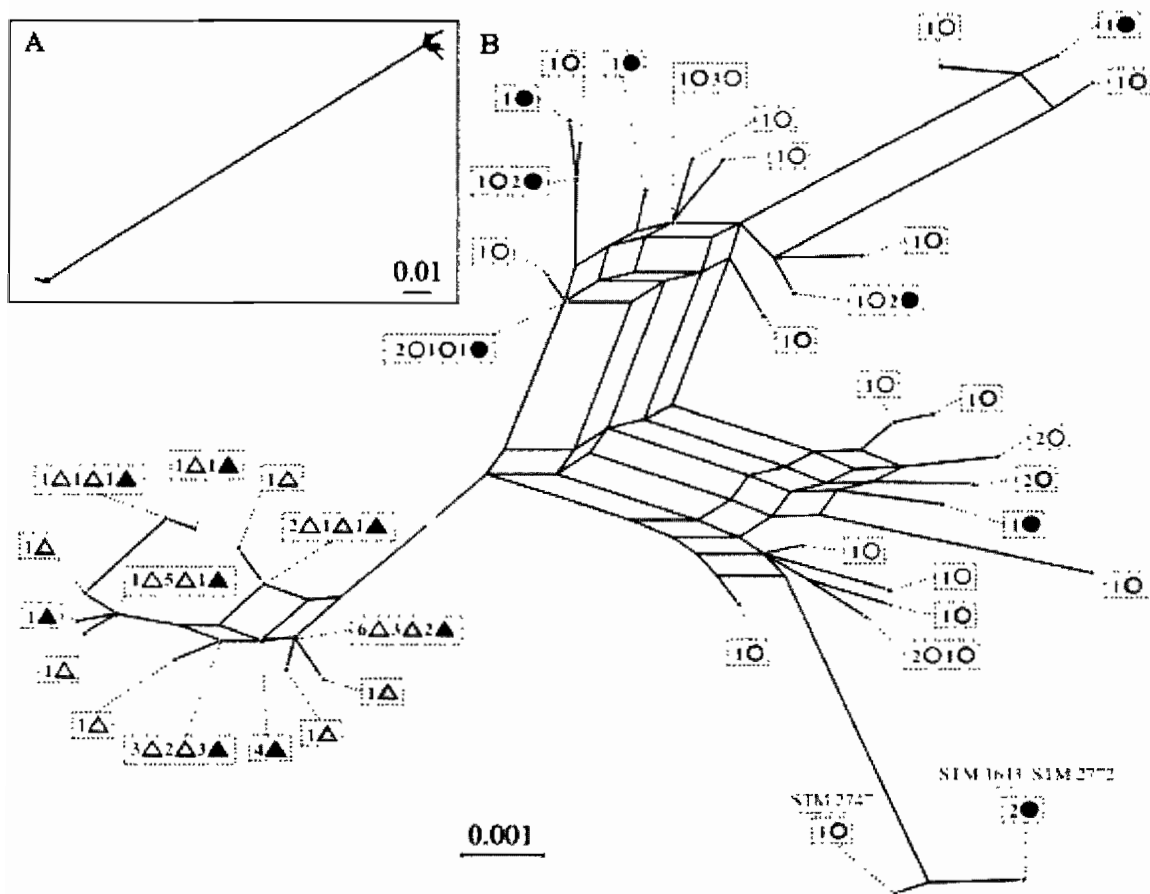


Figure. 3. Full view (A) and focus (B) on Neighbor-Net analysis obtained from 89 complete genotypes of *S. meliloti* and *S. medicae*. Several isolates within both *S. meliloti* and *S. medicae* are connected to each other by multiple pathways, forming an interconnected network and suggesting recombination. Conversely, *S. meliloti* and *S. medicae* clusters are connected by an edge, suggesting sexual isolation between the two species. Isolates did not cluster according to their sampling plant sets G1, G2 or G3. The analysis is based on an averaged distance matrix obtained from the seven sequences markers. Branch lengths are drawn to scale. Dotted lines indicate the position of haplotypes on the network. Triangles and circles indicate the isolates assigned respectively to *S. medicae* and *S. meliloti*. Isolates sampled from G1, G2 and G3 are indicated respectively by black, grey and white symbol colours. The numbering refers to the number of isolates sampled from the different groups.

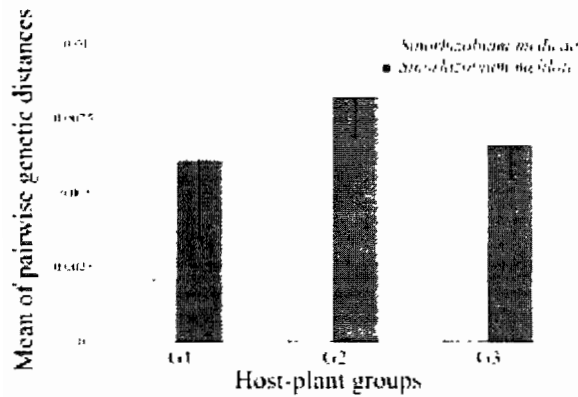


Figure. 4. Mean and 95% confidence interval of the genetic distance between *S. medicae* and *S. meliloti* isolates sampled from G1, G2 and G3. No significant difference was observed within both species, suggesting a lack of relationship between the diversity of host plants and the diversity of the rhizobia sampled from their nodules.

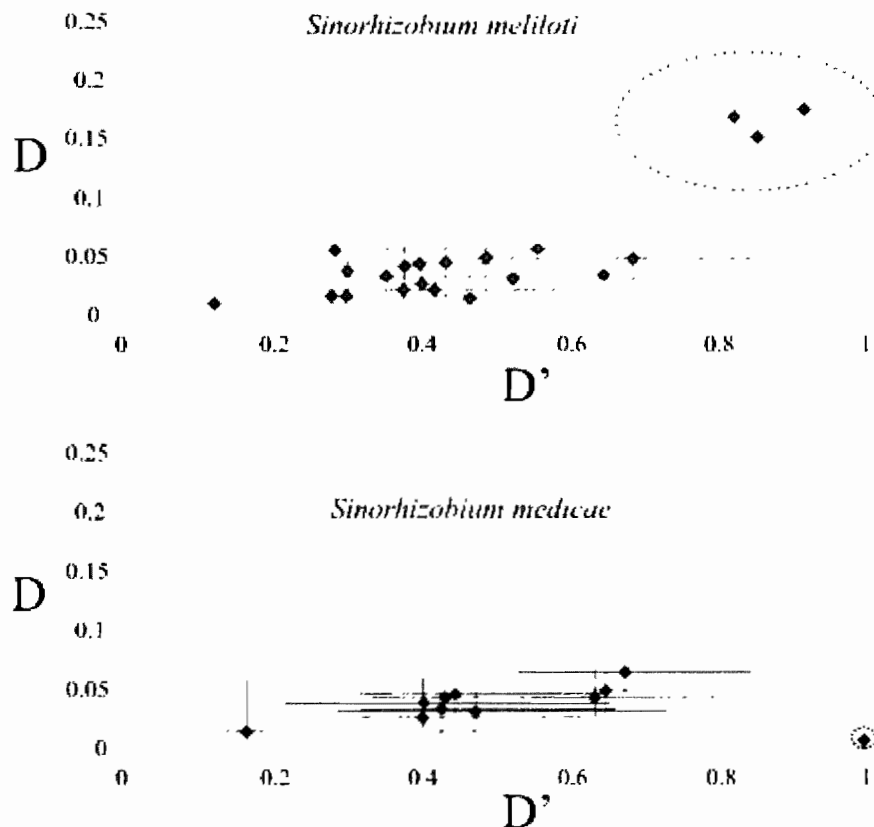


Figure. 5. Distribution the association measures D and D' among pairwise comparisons involving the genetic markers we used. Within *S. meliloti*, linkage disequilibrium measures obtained from chromosomal marker pairs are higher than others. This suggests that recombination rate depends on genome location. The difference is more obvious on D values than on D' values, which might be due to the similarity of polymorphism patterns observed among chromosomal markers. *S. medicae* chromosomal markers were not enough informative to obtain reliable linkage disequilibrium measures. However, others D and D' estimates are in agreement with *S. meliloti* ones. Each pairwise measure of linkage disequilibrium is illustrated by a rhomb. Ninety-five percent confidence interval of each measure is illustrated by horizontal and vertical lines. Association measures among chromosomal loci are surrounded by dotted lines.

XIII Article 3 : en préparation

XIII Article 3 : en préparation

Title: Host-specialization and speciation in the symbiotic nitrogen-fixing *Sinorhizobium* associated with *Medicago*.

Submitted as a Research article

Authors: Xavier Bailly,^{*,†} Isabelle Olivieri,[‡] Brigitte Brunel,[‡] Jean-Claude Cleyet-Marel,[‡] and Gilles Béné[‡]

Institution at which research was done:

Laboratoire des Symbioses Tropicales et Méditerranéennes UMR 113 IRD-Cirad-Ensam-UM2 / USC INRA Campus de Baillarguet 34398 Montpellier Cedex 5

Current affiliation of authors:

[†]Laboratoire des Symbioses Tropicales et Méditerranéennes UMR 113 IRD-Cirad-Ensam-UM2 / USC INRA : [‡]Institut des Sciences de l'Evolution de Montpellier, UMR 5554 CNRS-UM2; [‡]Diversité et Génome des plantes cultivées UMR 1097 CIRAD-INRA-ENSAM-IRD

Corresponding author: Xavier Bailly

Laboratoire des Symbioses Tropicales et Méditerranéennes UMR 113 IRD-Cirad-Ensam-UM2 / USC INRA: Campus de Baillarguet 34398 Montpellier Cedex 5

Phone number: 33 (0)4 67 59 38 01

Fax number: 33 (0)4 67 59 38 02

e-mail: bailly@mpl.ird.fr

Keywords: mutualism, *Sinorhizobium meliloti*, *Medicago*, biovar, disruptive selection, *nod* phylogeny.

Running head: Specialization and speciation of *Sinorhizobium* spp

Abstract

Using nitrogen-fixing *Sinorhizobium* species that interact with *Medicago* plants as a model system, we studied the influence of host specialization and speciation on bacterial diversity. The aims of this survey were: (i) to clarify how horizontal gene transfers might have shaped *Sinorhizobium* diversity at both species and biovar level; (ii) to compare the distribution of genetic diversity within and between *Sinorhizobium* species for core and accessory loci in order to distinguish speciation and host-specialization processes; (iii) to propose scenarios explaining the speciation and the coexistence of the sister-species *S. meliloti* and *S. medicae*. To gain insights into the diversification of *Medicago* symbionts both within and between species of *Sinorhizobium*, 126 isolates were obtained from two soil samples, from root nodules of *Medicago truncatula* and *Medicago luciniata*. Bacteria were characterized by single and multilocus sequence analysis at four loci, located on the chromosome and the two megaplasmids pSymA and pSymB. We also reconstructed a supertree based on sequences of nodulation gene clusters (*i.e.* *nod* genes clusters) including the different specificity groups which interact with *Medicago* species. Within *S. meliloti*, analyses suggest that *nod* genes have spread to different genetic backgrounds through recombination. This promotes variations in the observed genetic structure among loci. Further, two scenarios that might explain the pattern of genetic divergence between *S. meliloti* and *S. medicae* are proposed. These results are discussed in the light of theoretical models on bacterial speciation due to disruptive selection and on the maintenance of phenotypic variation in populations of rhizobia.

Introduction

Several recent theoretical studies have considered speciation in the context of *evolutionary branching* (Dieckmann et al. 2004), whereby local disruptive selection induced by density and frequency-dependent processes in a heterogeneous environment could lead to the emergence and coexistence of specialized phenotypes. These models predict that the coexistence of such phenotypes within species is not very robust to assumptions about their demography and their relative fitness (see Ravigne, Olivieri, and Dieckmann (2004) and references herein). Assuming such coexistence, speciation (*i.e.* whole genome divergence) will occur in only a limited set of conditions, unless strong linkage disequilibria initially occur between genes involved in local adaptation and in mating behaviour (Gavrilets 2003). Furthermore, organisms which are essentially clonal raise questions on the definition of species boundaries. The relative impact of ecological specialization and horizontal gene transfer on bacterial diversification and speciation is therefore a source of debate (Cohan 2001; Cohan 2002; Lawrence 2002; Gevers et al. 2005). By placing adaptive mutations in different genetic backgrounds, sexual promiscuity might indeed be a critical process that prevents bacterial speciation (Majewski 2001). The description of several biovars (*i.e.* ecotypes) indicate that recombination might be an efficient cohesion force in symbiotic rhizobium species (Silva et al. 2005; Vinuesa et al. 2005), contrary to predictions from several mathematical models for which recombination might not be significant in preventing bacterial speciation (Cohan 1994a; Cohan 1994b; Cohan 1995; Majewski and Cohan 1999b; Vetsigian and Goldenfeld 2005).

In this context, nitrogen-fixing bacteria which interact with *Medicago* plants appear as an ideal biological model for studying the link between specialization and speciation. *Sinorhizobium meliloti* and *Sinorhizobium medicae* are the main symbionts of *Medicago* plants. Each bacterial species and biovar (bv) interacts efficiently (*i.e.* fixes nitrogen) with a variable number of *Medicago* species, and *vice-versa* (Bena et al. 2005). For instance, *S. meliloti* bv. *medicaginis* (Villegas et al. 2006) can only fix nitrogen with the annual species *Medicago laciniata* and *M. satvagei*, and conversely, *M. laciniata* and *M. satvagei* cannot fix nitrogen with either *S. medicae* or *S. meliloti* bv. *meliloti* (Bena et al. 2005). Similarly, alfalfa (*Medicago sativa*) and *Medicago truncatula* interact with *S. medicae* and *S. meliloti* bv. *meliloti*, while *Medicago polymorpha* interacts solely with *S. medicae* (Rome et al. 1996).

Sinorhizobium meliloti and *S. medicae* genomes include three replication units (Roumiantseva et al. 1999). In *S. meliloti* strain 1021 genome, the chromosome harbors mainly core genes, encoding functions related to central metabolism and informational processing, whereas pSymA and pSymB megaplasmids harbor most of the accessory genes, which commonly encode supplementary metabolic pathways and symbiotic factors (Galibert et al. 2001). Specifically, bacterial genes involved in recognition by host-plants through Nod-Factors synthesis, such as *nod*, *noe* and *nol* genes (Perret, Staehelin, and Broughton 2000), are localized on pSymA. Furthermore, several genes that encode surface polysaccharidic compounds pathways (LPS, EPS, KPS), *e.g.* *exo* genes (Reuber and Walker 1993), are located on pSymB (Fig. 1). Studies based on genome sequence databases have shown that core bacterial genes are less prone to horizontal gene transfer than the accessory ones (Nakamura et al. 2004). Consistently, Bailly et al. (in press a) showed that horizontal gene transfers involving plasmidic loci are quite common within both *S. meliloti* bv. *meliloti* and *S. medicae*, leading to linkage equilibrium among the three replication units.

Whereas the sexual isolation between *S. meliloti* bv. *meliloti* and *S. medicae* has been described by Maynard Smith et al. (1993), nothing is known about the genetic differentiation among populations of *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis*. Populations of the two biovars of *S. meliloti* can be found in sympatry, as well as the two species *S. meliloti* and *S. medicae*. Further, Barran, Bromfield, and Brown (2002) described *nodC* gene as the main determinant of the specificity between *Medicago laciniata* and its symbiont. Such specificity could induce disruptive selection pressures promoting the coexistence of various ecotypes. Then, this process might lead to the divergence of ecotypes for genes not involved in host specificity (i.e. speciation) through clonal reproduction and / or mate choice. Conversely, homogenizing selection acting on the core genome (i.e. purifying selection and global selective sweeps) and frequent recombination events might prevent strains specialized on different *Medicago* species, e.g. *M. laciniata* or *M. truncatula*, to diverge for genes not involved in specificity, and thus to go to speciation.

In this context, our aim is to identify evolutionary processes that shaped the diversity of nitrogen-fixing symbionts which interact with *Medicago* species. We characterized by Multi Locus Sequence Analysis (MLSA), an approach derived from multilocus sequence typing (Maiden et al. 1998), two populations of bacteria, including isolates belonging to *S. medicae*, *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis*. Further, we obtained a phylogenetic reconstruction from partial sequences of the *nodABC-nodEG* gene clusters of symbionts of *Medicago* specificity groups. Based on these data, the goals of this study were: (i) to clarify how horizontal gene transfer shaped *Sinorhizobium* spp. diversity at both species and biovars level; (ii) to compare the distribution of genetic diversity within and between *Sinorhizobium* species at core and accessory loci in order to distinguish speciation and host-specialization processes; (iii) to propose different scenarios explaining the speciation and the coexistence of *S. meliloti* and *S. medicae*.

Materials and Methods

Isolation and culture of bacteria

Two samples of top soil were obtained from two localities in Tunisia which are included in a semi-arid bioclimatic area: Enfidha (longitude 10°23' East – latitude 36°07' North) and Hadjeb (longitude 9°33' East – latitude 35°24' North). These sampling sites fall within the natural geographic ranges of both *M. laciniata* and *M. truncatula*. Aliquots of each soil sample were put into contact with two plant species: i) 40 individuals of *M. truncatula* (Jemalong A17 line), ii) 20 individuals of *M. laciniata* (bulk of genotypes). Gibson tubes were previously nine tenth filled with vermiculite and 25 ml of nitrogen-free plant nutrient solution. Tubes were then autoclaved, and completed with a soil aliquot. Seeds were surface sterilized with calcium hypochlorite 5% (weight/volume) for 5 minutes and rinsed with sterile water. Seeds were germinated for 72 hours on 1% (w/v) agar medium. Each plant was transferred individually into a Gibson tube which was placed in growth chamber at 22°C during the day and 18°C at night, with a 16-hour photoperiod and 50-60% relative humidity. After two months, all nodules from each plant were harvested. One or two nodules per plant were sterilized using hypochlorite 1% (w/v) for three minutes, then rinsed three times in sterile water. The isolation of one bacterial strain per nodule was performed using three successive sub-cultures on YEMA solid medium (Vincent 1970) of an isolated colony starting from crushed nodules. Bacterial strains were preserved at -80°C in YEM medium supplemented with 20% (volume/volume) of glycerol.

Each bacterial isolate was grown in 20 ml YEM medium. We also grew, from our laboratory collection, one *S. medicae* and one *S. meliloti* bv. *meliloti* strains (STM 1605 and STM 1604 respectively, see Bailly et al. in press a), three *S. meliloti* strains which had been described as interacting specifically with *M. noeana*, *M. radiata*, and *M. rigiduloides* (USDA 1613, USDA 1614 and USDA 1623), and *Rhizobium mongolense* strain USDA 1844 (Van Berkum et al. (1998), recently renamed as *Rhizobium gallicum* bv. *orientale* (Silva et al. 2005)), which fix nitrogen with *M. ruthenica*.

DNA extraction and sequencing

MLSA scheme

We characterized symbiotic bacteria sampled from *M. truncatula* and *M. luciniata* by sequencing four loci. Previous studies (e.g. Bailly et al. in press a) have shown linkage disequilibrium among chromosomal markers in *S. meliloti* bv. *meliloti*. In the present study we thus only analyzed one locus on the chromosome rather than three loci as previously done in Bailly et al. (in press a). More precisely we investigated genetic polymorphism of one chromosomal sequence (*IGS_{RKP}*), one pSymB (*IGS_{EXO}*) and two pSymA (*IGS_{NOD}*, *IGS_{GAB}*) sequences (Fig. 1). Sequences of PCR/sequencing primers are provided in Table 1.

Design of primers within the nod gene cluster of the symbionts of Medicago plants

In addition to the MLSA approach described above, we aimed at obtaining two sequence datasets in *nod* gene region of rhizobia that interact with *Medicago* plants (Fig. 1): (i) partial sequence of the *nodABC* gene cluster, which is ubiquitous to the genome of all rhizobia. The large distribution of *nodABC* gene cluster allows us to obtain a phylogeny rooted by an appropriate outgroup (i.e. *Rhizobium leguminosarum* bv. *viciae*, *Rhizobium galegae*, *Rhizobium tropici* and *Sinorhizobium* NGR234); (ii) partial sequence of the *nodEG* gene cluster, which seems to be specific to *Medicago* symbionts. As *nodEG* gene cluster is not ubiquitous among rhizobia, we applied a supertree approach (see Bininda-Emonds (2004) for a review) to combine the information available from both datasets. Primers were designed based on homologous regions conserved among rhizobia, obtained from GeneBank database (Table 1).

Experimental procedures

One milliliter of liquid culture was washed twice in an Eppendorf micro-tube by centrifugation (15000 g, 4 min) and the pellet re-suspended in 750 µl of sterile water. One hundred microliters of this solution was incubated for two hours with 20 µl of 1 mg.ml⁻¹ proteinase K and 100 µl of tris-HCl (10 mM, pH 8.3). After boiling, this mixture was used as DNA matrix. DNA amplification was performed using a Perkin-Elmer 2400 thermocycler in 25 µl volume, including 1 µl of DNA matrix, 200 µM of each dNTP, 0.8 µM of each primer, 1.5 mM of MgCl₂, 1X buffer supplied by the *Taq* polymerase manufacturer, and 1.25 U of Invitrogen *Taq* polymerase. We used a touch-down program including an initial denaturation stage (96°C, 4 min), 20 cycles of denaturation (96°C, 30 s), annealing (annealing temperature decrease steadily from 60°C to 50°C in 20 cycles, 30 s), elongation (72°C, 1 min) and 20 cycles of denaturation (96°C, 30 s), annealing (50°C, 30 s) and elongation (72°C, 1 min). PCR products were cut out from 1% (w/v) agarose electrophoresis gel, purified with a QIAquick Gel extraction kit (Qiagen). Sequencing reactions were performed on one strand

using the DYEnamic ET Terminator Kit (Amersham Bioscience) and analyzed on an Amersham Bioscience model Megabace 1000 DNA sequencer.

Data analyses based on the nod gene cluster of the symbionts of Medicago plants

Alignment of sequence datasets

Nucleotide sequences of *nodABC-nodEG* gene cluster were obtained from strains STM 1605, STM 1604 (Bailly et al. in press a), *S. meliloti* isolates STM 2835 and STM 2836 of this study, USDA 1613, USDA 1614, USDA 1623, USDA 1844. These sequences were aligned using CLUSTAL X 1.83 (Thompson et al. 1997), together with those of *S. meliloti* strain 1021 (*nodABC-nodEG*) (GeneBank accession number (AN): NC_003037), *S. meliloti* bv. medicaginis strain 102L4 (*nodABC*) (AN: AF522456), *R. galegae* (*nodABC*) (AN: X87578), *R. tropici* (*nodABC*) (AN: X98514), *R. leguminosarum* bv. viciae (*nodABC*) (obtained from a blast search at http://www.sanger.ac.uk/Projects/R_leguminosarum/), and *Sinorhizobium* NGR234 (*nodABC*) (AN: U00090). Coding sequences were aligned based on amino-acid translation, while intergenic spacers were aligned directly. Gaps were manually reported in DNA sequence alignments which were edited with BIOEDIT 5.0.9 (Hall 1999).

Detection of recombination

We searched for traces of recombination in a concatenated dataset including *nodABC* and *nodEG* sequences of *Medicago* symbionts using three methods implemented in RDP2 (Martin, Williamson, and Posada 2005): MAXCHI, CHIMAERA and GENECONV, these methods being the most accurate to detect recombination events (Posada and Crandall 2001; Posada 2002). The two first methods use a sliding window to detect significant discrepancies in the segregation of polymorphisms on either side of each sequence site. These are based on the computation of χ^2 distributed statistics obtained from comparisons of pairs or triplets of sequences respectively. GENECONV is an extension of Sawyer's method (Sawyer 1989). It compares sequence pairs and scores unusually long series of polymorphic sites for which the two sequences shared a strong similarity. The null hypothesis being the lack of gene conversion events, it is tested through the randomization of the location of alignment sites. A p-value is given as the proportion of permuted alignments which result in a greater score than the original alignment. For all these methods, indel blocs were analysed as polymorphisms, we performed 1000 permutation steps and we used a P-value cut-off set to 0.05 using a Bonferroni correction for multiple tests. We also searched consensus daughter sequences and consensus recombination breakpoints. For MAXCHI and CHIMAERA, we used a variable window size including respectively 20 polymorphic sites and 10% of variable sites. For GENECONV, similarity scores were computed using a *gsclue* (i.e. a parameter which scale the impact of mismatches on similarity scores) set to 1.

Inference of a supertree

Models of sequence evolution were chosen separately for the *nodABC* dataset and for the *nodEG* dataset, using the nested likelihood ratio test (η LRT) implemented in MODELTEST 3.7 (Posada and Crandall 1998). Maximum likelihood (ML) topologies were obtained for the two datasets, through a heuristic search with Tree Bisection-Reconnection (TBR) swapping option, starting from a stepwise addition tree by using PAUP* 4b10 (Swofford 2003). Reliability of bipartitions was investigated using bootstrap proportions estimated from 1000 pseudo-replicates analysed by the same ML approach. The information obtained from the two

bootstrap computations were combined in a super-matrix in order to summarize the phylogenetic information included in both datasets as described by Daubin, Gouy, and Perriere (2002). An un-weighted parsimony phylogeny (*i.e.* a supertree) and bootstrap proportions were obtained from the super-matrix using PAUP[®] 4b10 (Swofford 2003).

Multilocus sequence analyses

Alignment of sequence datasets and phylogenetic inferences

Nucleotide sequence alignments were performed using CLUSTAL X 1.83 (Thompson et al. 1997) and manually edited with BIOEDIT 5.0.9 (Hall 1999). We built a ML phylogeny for each molecular marker using PAUP[®] 4b10 (Swofford 2003), applying models of sequence evolution determined by the η LRT implemented in MODELTEST 3.7 (Posada and Crandall 1998) and performing heuristic searches by TBR algorithm, starting from the Neighbor-Joining tree. The four phylogenies were used in two ways. First, Shimodaira-Hasegawa (SH) tests (Shimodaira and Hasegawa 1999) were performed to investigate likelihood differences among topologies obtained for the different markers for each sequence dataset. SH tests were performed using completely optimized models with PAUP[®] 4b10 (Swofford 2003). Second, in order to investigate the impact of horizontal gene transfer within *S. medicae*, *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis* genetic groups, we mapped the evolution of symbiotic specificity onto the four phylogenies. Ancestral character states were estimated using an unweighted parsimony model with MESQUITE 1.06 (Maddison and Maddison 2005).

Distance matrices were computed for the four loci with PAUP[®] 4b10 (Swofford 2003) according to each model of sequence evolution. These matrices were used to: i) compute for each locus the mean pairwise divergence between isolates that belong to two different groups among *S. medicae*, *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis*; ii) compute an unweighted average matrix of genetic distances between isolates over the four loci in order to obtain a molecular network by Neighbor-Net (Bryant and Moulton 2004) using SPLITSTREE4 (Huson and Bryant 2006). Neighbor-Net method is an agglomerative algorithm, derived from Neighbor-Joining (Saitou and Nei 1987), which can represent conflicting phylogenetic signals in sequence data. The Neighbor-Net method tends to infer fully resolved split graphs. In order to clarify the phylogenetic hypotheses supported by the network, measures of the robustness of each split should be used. Yet, bootstrap based procedures lead to poor results when applied to Neighbor-Net method (Huson and Bryant 2006). Since we mostly deal with intra-specific sequence polymorphisms (*i.e.* saturation of phylogenetic signal should not be observed), we thus applied a filter on splits based on a length threshold, according to Wilkinson, Lapointe, and Gower opinion (2003). We chose the threshold so that the length of the final network was nearly 95% of the global split lengths. The split graph was used not only to visualize conflicting phylogenetic signals but also isolates clustering in relation to the three symbiotic groups.

Detection of recombination

We screened *S. meliloti* MLSA datasets for evidence of recombination. After concatenating sequences of the four loci, alignments were scanned with MAXCHI, CHIMAERA and GENECONV using the setting described above. Moreover, we obtained a phylogenetic profile to illustrate recombination breakpoints by using PHYLPRO 0.96 (Weiller 1998). This method is based on a sliding window approach which measures the correlation between two vectors of pairwise genetic distances among any studied sequence and all others obtained from either side of each alignment site. Specific settings were as follow. Gaps were treated as

differences, analyses were performed on polymorphic sites using a variable window size including four polymorphisms and we measured the congruence among each pair of distance vectors with the Pearson's correlation coefficient.

Population structure

Sequences that differed from others by at least one nucleotide position were considered as different alleles. The diversity of each bacterial taxa (*i.e.* *S. medicae*, *S. meliloti* bv. *meliloti* or *S. meliloti* bv. *medicaginis*) was assessed using two summary statistics: allelic richness "R" and Nei diversity "H" indices (Nei 1987) which were computed for each of the four MLST dataset. We used a randomisation process to cope with differences in sample size. Expected allelic richness for each genetic group and 95% confidence intervals were estimated for the minimum sample size N_{min} observed among groups. Contribution of each symbiotic group to the allelic richness of pairs of ecotypes at each locus was also computed.

Differentiation among bacterial populations/taxa was estimated by F_{ST} values among populations according to Weir and Cockerham (1984) procedure implemented in FSTAT 2.9.3.2 (updated from Goudet, (1995). Exact-tests of genotypic differentiation (Raymond and Rousset 1995b) were performed with GENEPOP 3.3 (Raymond and Rousset 1995a). A population tree was obtained from pairwise distances between populations based on allele frequencies using Cavalli-Sforza algorithm with POPULATIONS 1.2.28 (Langella, O. 1999 unpublished). We performed 1000 bootstrap pseudo-replicates on individuals in order to assess robustness of the different nodes.

Results

Sampling experiment

One bacterial isolate was obtained from each of the 120 plants except for six *M. truncatula* plants for which two isolates per plant were used in the experiments. The bacterial sampling included a total of 126 isolates (Table 2). With regards to symbiotic efficiency, both *M. truncatula* and *M. lacinata* were able to fix nitrogen with bacteria of the soil sample they were in contact with (green foliage, pink nodules).

Nucleotide sequence accession numbers

All sequences have been deposited in the GeneBank database (*IGS_{RKP}*: DQ405642- DQ405767; *IGS_{NOD}*: DQ405467- DQ405592; *IGS_{GAB}*: DQ406393- DQ406518; *IGS_{EXO}*: DQ406104- DQ406229; *nodABC-nodEG* DQ406568- DQ406583).

Phylogenetic analysis on nodABC-nodEG gene clusters

We did not infer clear clues of recombination within the *nodABC* and *nodEG* concatenate sequence datasets. Indeed, no recombination breakpoints were evidenced by more than one of the three methods we used. Furthermore, no recombination event between specificity groups was inferred. We used a HKY+I+ Γ_4 model and HKY+ Γ_4 to perform phylogenetic inferences respectively on *nodABC* and *nodEG* sequence datasets, accordingly to the results obtained with the η LRT implemented in MODELTEST 3.7 (Posada and Crandall 1998). The supertree obtained from both *nodABC* and *nodEG* gene cluster indicates that *nod* region of symbionts associated to *Medicago* clustered in a monophyletic group (Fig. 2). Moreover, *S. meliloti* and *S. medicae* sequences grouped in a single clade. The *nod* gene region of *S. meliloti* isolates, including the three ecotypes defined within this species, formed a paraphyletic group with

S. medicae strains forming a sister clade to *S. meliloti* bv. *meliloti*. Bootstrap proportions that sustain these observations vary from 95% to 100%.

Multilocus sequence analyses

We obtained 504 sequences from the four loci, allowing us to define 126 complete genotypes (Table 2). The 126 isolates clustered into two genetic groups based on sequence homology. These groups were identified as *S. meliloti* and *S. medicae* by sequence comparison onto sequence alignments against available sequence datasets (Bailly et al. in press a). *S. meliloti* isolates sampled from *M. luciniata* were assigned to *S. meliloti* bv. *medicaginis* based on both their symbiotic abilities and their specific *IGS_{NOD}* alleles. Overall, the dataset of this study includes 20 *S. medicae* isolates, 66 *S. meliloti* bv. *meliloti* isolates and 40 *S. meliloti* bv. *medicaginis* isolates. Models of sequence evolution selected by the η LRT selected for each marker were: HKY85 (Hasegawa, Kishino, and Yano 1985) models for *IGS_{NOD}*, *IGS_{GAB}* or *IGS_{EXO}* datasets and JC69 (Jukes and Cantor 1969) for *IGS_{RKP}* dataset. It noteworthy that model choice is conditioned by the polymorphism that we observed at each locus. Sequence datasets of each locus displayed various patterns of polymorphism (Table 3). More precisely, we observed 3 particular patterns. *IGS_{EXO}* alleles of *S. meliloti* and *S. medicae* are strongly divergent and several insertions / deletions arise since the speciation (identity scores 20%). Within *S. meliloti*, two sequence clusters were obviously divergent in both *IGS_{EXO}* and *IGS_{GAB}* (respective identity scores 92% and 93%).

After obtaining an unweighted average of the four distance matrices, a Neighbor-Net split graph was built (Fig. 3). We applied a threshold of 4.5×10^{-4} to the graphical representation. This threshold resulted in a substantial gain of clarity (232 splits out of 299 were removed) without losing much information (the weight of the final network was 94.2% of the original split graph weight). The network displays three groups of strains: i) *S. medicae* isolates, ii) *S. meliloti* bv. *meliloti* isolates and iii) *S. meliloti* bv. *medicaginis* isolates. The clear delineation of *S. meliloti* bv. *meliloti* group and *S. meliloti* bv. *medicaginis* group is mainly due to *IGS_{NOD}* polymorphism (analyses not shown). This is consistent with locus by locus analyses presented below.

The Neighbor-Net network splits isolates into three groups identified as *S. medicae*, *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis* (Fig. 3). While *S. medicae* and *S. meliloti* bv. *meliloti* groups are connected by a long edge, *S. meliloti* bv. *medicaginis* and *S. meliloti* bv. *meliloti* groups are connected by several splits. Further, we did not detect any significant differentiation between Enfidha and Hadjeb populations within each *S. meliloti* biovar based on all loci ($p > 0.05$). Thus, Tunisian genotypes of each isolated from the two soil samples were pooled in further analyses.

Locus by locus analyses

S. medicae alleles clustered in monophyletic groups on the four phylogenies obtained from the molecular markers used in this study (Fig. 4). As no allele was shared between *S. medicae* and *S. meliloti* biovars (Fig. 5A), high F_{ST} values were measured among the two species and all differentiation tests between these groups rejected the null hypothesis (Fig. 5B). Moreover, average of pairwise genetic distance between isolates sampled from *S. medicae* and *S. meliloti* bv. *meliloti* or bv. *medicaginis* indicated high mean molecular divergence between the pairs of populations investigated (Fig. 5C). Conversely, mean molecular divergence between isolates belonging to different *S. meliloti* biovars was low for all loci except *IGS_{NOD}* (Fig. 5C).

Moreover, *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis* shared some allele for all loci excluding *IGS_{NOD}* (Fig. 5A). This explains that F_{ST} value observed at *IGS_{NOD}* are higher than those at other loci. Furthermore, allele sets of *S. meliloti* strains isolated from *M. truncatula* and *M. luciniata* formed paraphyletic groups in all phylogenies but *IGS_{NOD}* (Fig. 2, Fig. 4).

Recombination evidences

For each locus, SH test results indicated that the topology obtained is significantly more likely than those obtained from other loci (Table 4). At least two transitions in symbiotic abilities were inferred from *IGS_{RKP}*, *IGS_{GAB}* and *IGS_{EXO}* phylogenies within *S. meliloti* strain clusters (Fig. 4). No transition was inferred within the monophyletic groups that cluster *S. medicae* alleles onto the four phylogenies. No transition was observed in *IGS_{NOD}* phylogeny within *S. medicae*, *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis* groups. Finally, we found recombination breakpoints ($p < 0.05$) located between each locus based on the scan of the concatenated sequence alignment of the four loci with MAXCHI, CHIMAERA and GENECONV. Based on the three methods, significant assignments of parental sequences ($p < 0.05$) showed that these transfers took place both within and among *S. meliloti* biovars. These results are illustrated by a graph obtained using PHYLPRO 0.96 (Weiller 1998) (Fig. 6).

Discussion

Recombination occurs among *S. meliloti* biovars

Previous studies have demonstrated that the diversity of *S. meliloti* bv. *meliloti* is strongly influenced by recombination (Bailly et al. in press a; Maynard Smith et al. 1993). The present study shows that recombination also occurs between the two *S. meliloti* biovars, not only between *nod* and chromosomal markers but also between plasmidic markers. First, the Neighbor-Net network suggested that horizontal gene transfers (HGT) shaped the genetic diversity of both *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis*. Significant incongruence among the four phylogenies we built confirmed that network splits inferred resulted from recombination among *S. meliloti* biovars and not from a lack of phylogenetic signal based on the data we used. Second, the complex evolution of symbiotic abilities observed within *S. meliloti* clades (i.e. the two *S. meliloti* symbiotic groups are paraphyletic on all phylogenies with the exception of *IGS_{NOD}* for which each *S. meliloti* biovar forms a clade) is also consistent with the hypothesis that recombination took place among *S. meliloti* biovars. Finally, MAXCHI, CHIMAERA and GENECONV analyses explicitly localised recombination events between *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis*. Recombination promotes the evolution of loci which are horizontally transferred independently from their genetic background. Thus, these loci might display different genetic structures depending on selective pressures and barriers to gene flow that both shaped their evolution. In this study, we indeed observed three different patterns of genetic differentiation between the two *S. meliloti* biovars, among the four markers we investigated. First, *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis* are completely differentiated at *IGS_{NOD}* (i.e. no shared allele), and alleles of each biovar clustered in two distinct clades on both *IGS_{NOD}* and *nodABC-nodEG* topologies. Barran, Bromfield, and Brown (2002) reported that the efficiency of the interaction of *S. meliloti* bv. *medicaginis* strains with *M. luciniata* required a specific *nodC* allele. In this case, a clear disruptive selection pressure due to the host plant acted on *S. meliloti nodC* gene and influenced the variability of the entire *nod* gene region by hitchhiking (Maynard Smith and Haigh 1974) owing to the lack of recombination within the

symbiotic region. Second, *IGS_{EXO}* also appears as atypical among the four loci, since it displayed a clear differentiation in two major alleles which did not fit significantly with biovar boundaries. The major *IGS_{EXO}* alleles exist not only in two Tunisian populations but also within a *Sinorhizobium* population isolated from a French soil (Bailly et al. in press a). We previously showed that this locus evolved under balancing selection, maintaining two alleles at similar frequency in a *S. meliloti* bv. *meliloti* population. As we encountered a similar pattern, we can hypothesize that similar evolutionary pressures might have influenced the polymorphism of both biovars and prevented any differentiation between them. This pattern might be due to host plant selective pressures, as *exo* loci are involved in limiting plant defences (see Frayse, Couderc, and Poinot (2003) for a review). Finally, our analyses revealed a significant differentiation, mostly due to differences in allele frequencies, between the two biovars on *IGS_{RKP}* and *IGS_{GAB}*. Historical effects might be involved in the observed genetic differentiation. The recent admixture of the two biovars at a local scale may have not yet permitted recombination to homogenize allele frequencies, similarly to a Wahlund effect. Recombination might also be more frequent within than between biovars as sharing the same host plant could favor genetic exchanges. For instance, the two genes coding for the conjugal transfer protein *TruA*, which are respectively located on *S. meliloti* pSymA and pSymB, are over-expressed when bacteria are in contact with luteolin, a flavonoid produced by *Medicago* plants (Ampe et al. 2003). Moreover, Pretorius-Guth, Puhler, and Simon (1990) suggested that plant nodules create the most favorable natural conditions for recombination among *S. meliloti* isolates. Thus, a higher recombination rate within biovar might result from a higher mating probability due to host plant specificity and *tra* genes induction. This might result in a recombination-selection equilibrium characterized by some genetic differentiation between biovars.

The differentiation patterns we observed might also reflect an ongoing speciation process between the two biovars. Indeed, the observed differentiation on *IGS_{NOD}* might illustrate the selection of strains driven by their respective host plants, which could lead to divergence on other loci, e.g. *IGS_{GAB}* or *IGS_{RKP}*, and finally reach the overall genome. However, such speciation hypothesis is not the most likely since preliminary results at the Mediterranean basin scale suggested that *S. meliloti* populations preferentially clustered according to their geographical location rather than their original host plant species (see Fig. 7 based on genotyping results obtained from this study, Bailly et al. in press a and Bailly et al. in press b). This supports the hypothesis that HGT is frequent enough to prevent speciation of the different biovars within *S. meliloti*.

Speciation and specialization of S. meliloti and S. medicae

At *IGS_{NOD}* locus, the genetic divergence between the two *S. meliloti* biovars is 8.79 times higher than the divergence between *S. meliloti* bv. *meliloti* and *S. medicae*, and symbiotic regions of *S. meliloti* strains are paraphyletic on *nodABC-nodEG* supertree. Conversely, the genetic divergence between *S. meliloti* and *S. medicae* is five to sixteen times higher than the divergence between the two *S. meliloti* biovars at *IGS_{EXO}* (16.06 times), *IGS_{RKP}* (16.56 times) and *IGS_{GAB}* (5.22 times). The differentiation and the divergence between *S. meliloti* and *S. medicae* reflect their sexual isolation and illustrate the existence of *bona-fide* evolutionary species within the genus *Sinorhizobium*. This last statement is consistent with a previous study by Vinuesa et al. (2005) about the species concept within the genus *Bradyrhizobium*. In this context, host plant selective pressures constraining polymorphism at *nod* genes could explain a lower divergence between *S. meliloti* bv. *meliloti* and *S. medicae* at

these loci. However, this last hypothesis might also be challenged by HGT between species, leaving open multiple scenarios for speciation and specialization of *S. meliloti* and *S. medicae*.

Two scenarios might indeed explain the current genetic structure of *Medicago*'s symbiotic strains. i) A first explanation involves the speciation of *S. medicae* from one ancestral lineage of *S. meliloti* harbouring *nod* genes specific to *S. meliloti* bv. *meliloti*. Such speciation (either sympatric or allopatric) must have taken place for enough time to sort *S. meliloti* and *S. medicae* lineages on their whole genome leading to monophyletic groups except for the *nod* gene cluster of pSymA which evolve under host-plant selective constraints. ii) An alternative explanation involves an interspecific HGT of the *nod* gene cluster between *S. meliloti* and *S. medicae* after their speciation. Assuming this scenario, two possible issues have to be considered. First, following *S. meliloti* and *S. medicae* speciation, *S. meliloti* symbiotic region might have evolved under host plant selective pressures, inducing the emergence of three specificity groups adapted to either *M. truncatula*, *M. laciniata* or *M. rigiduloides*. Then, *nod* genes of *S. meliloti* bv. *meliloti* might have been transferred to *S. medicae*. Second, speciation between *S. meliloti* and *S. medicae* might have been due to specialization on two different plant species, resulting in divergent *nod* genes. Subsequently, *nod* genes of *S. medicae* adapted to host-plant displaying the same specificity than *M. truncatula* might have been transferred to a *S. meliloti* strain, replacing its own symbiotic cluster. In the same way, Silva et al. (2005) and Vinuesa et al. (2005) proposed that interspecific HGT of symbiotic island is the most likely explanation for the variability observed respectively within *Rhizobium gallicum* and *Bradyrhizobium canariense* populations.

Origin of Sinorhizobium meliloti and S. medicae nod genes

While *S. meliloti* and *S. medicae* are the main symbionts of *Medicago* plants, *Rhizobium mongolense* (recently renamed *Rhizobium gallicum* bv. *orientale* by Silva et al. (2005)) is the specific symbiont of *Medicago ruthenica* (Van Berkum et al. 1998). According to the supertree obtained from partial sequences of *nodABC* and *nodEG* gene clusters, *nod* sequences of symbionts of *Medicago* species cluster in a monophyletic group (Fig. 2). Within this group, a clade clusters all *nod* sequences of *S. meliloti* and *S. medicae*, whereas *R. mongolense* appears as a sister branch to this clade. These results are consistent with two other *nod* phylogenetic studies based on *nodB* (Silva et al. 2005) and *nodC* (Laguerre et al. 2001) genes which already support a monophyletic group including *R. mongolense*, *S. meliloti* and *S. medicae* sequences. In the same way, *S. meliloti* and *S. medicae* are considered as sibling species based on housekeeping genes phylogenies (Turner and Young 2000; Gaunt et al. 2001; Laguerre et al. 2001; Silva et al. 2005). However, based on chromosomal genes, *Medicago* symbionts belonging to *Rhizobium* and *Sinorhizobium* genera do not form a monophyletic group, while their branching is consistent with the monophyly of each genus (Turner and Young 2000; Gaunt et al. 2001; Laguerre et al. 2001; Silva et al. 2005). The incongruence between symbiotic and housekeeping genes phylogenies strongly suggests that HGT occurred between these two lineages. More precisely, these data suggest that the two *Sinorhizobium* species acquired the ability to nodulate *Medicago* species through a HGT of the symbiotic island from an ancestor of *R. mongolense*, either towards the common ancestor of the two *Sinorhizobium* species or towards one *Sinorhizobium* species followed by a spread to the other one (see above discussion on the different scenarios). In this context, it is noteworthy that *Medicago ruthenica* appears as the first divergent plant species in the strict consensus most parsimonious molecular phylogeny of the genus *Medicago* (Downie et al. 1998). This reinforce the idea that *R. mongolense* derived from the ancestral symbiotic

bacteria of the genus *Medicago*. Such HGT across different phyla of bacteria is usually invoked to explain the distribution of symbiotic ability among rhizobia, both within α -proteobacteria (Wernegreen, Harding, and Riley 1997; Moulin et al. 2004; Silva et al. 2005; Vinuesa et al. 2005) and between α - and β -proteobacteria (Moulin et al. 2001; Chen et al. 2003). A transfer of the entire symbiotic island (including both nodulation and N_2 fixation genes) has for instance been proved in *Mesorhizobium loti* by acquisition of a symbiosis island that integrates into a phe-tRNA gene (Sullivan and Ronson 1998).

Conclusions

Our data suggest that host-specialisation and speciation are two independent processes affecting the evolution of rhizobia. At a within-species scale, we reinforced the symbiotic classification of *S. meliloti* in two biovars (bv. *meliloti* and *medicaginis*) (Villeguas et al. 2006) since they shared a common genetic background not only on the chromosome but also on the two megaplasmids with the exception of the symbiotic region involved in nod-factor synthesis. Our analyses therefore support the hypothesis that *S. meliloti* evolved under a cohesive recombination model (see Gevers *et al.* (2005) for a review on theory based-concepts of prokaryotic taxonomy) because the different functional lineages of *nod* genes spread into the different genetic backgrounds of *S. meliloti* through recombination. Despite the fact that such model fits to many taxa, including rhizobia and some pathogens such as *Borrelia burgdorferi* (Qiu et al. 2004), available theoretical models predict that recombination has little impact on the speciation of bacterial ecotypes (Cohan 1994a; Cohan 1994b; Cohan 1995; Majewski and Cohan 1999b; Vetsigian and Goldenfeld 2005). Three main reasons may explain the discrepancy between theoretical expectations and *S. meliloti* population structure: i) *S. meliloti* genome is composed of several replication units, limiting the effect of propagating fronts of genetic divergence among ecotypes (Vetsigian and Goldenfeld 2005); ii) the modalities of gene exchange and DNA recombination might differ from those implemented in theoretical models which are fitted to transformation mechanism (Cohan 1994a; Cohan 1994b; Cohan 1995). For instance, the impact of sequence dissimilarity on gene conversion rate is species-specific (Majewski and Cohan 1999a; Majewski et al. 2000); iii) the diversity of *S. meliloti* population might be influenced by recurrent selective sweeps affecting all the biovars, which might prevent the speciation of ecotypes (Majewski and Cohan 1999b). The lack of differentiation we observed at *IGS_{Exo}* is consistent with this hypothesis. Thus, further studies on horizontal gene exchange mechanisms within *S. meliloti* are needed to get a better understanding on emergence and maintenance of its genetic diversity. At the interspecific level, two scenarios may explain both speciation and specialisation of *S. medicae* and *S. meliloti*. While our data suggest that *Sinorhizobium* strains acquired the ability to nodulate *Medicago* plants through an HGT of *nod* genes from an ancestor of *Rhizobium mongolense*, it is indeed not clear whether the low genetic divergence between *nod* genes of *S. meliloti* bv. *meliloti* and *S. medicae* is due to an interspecific HGT or to a budding speciation process leading to the emergence of *S. medicae*. For now, it seems difficult to exclude any of these hypotheses, but whatever the scenario, the speciation process between *S. meliloti* and *S. medicae* might have been due to geographic isolation or/and a differential adaptation to selective constraints (Majewski 2001). These last processes are of great interest in the context of the coexistence of *S. meliloti* bv. *meliloti* and *S. medicae* in natural populations. Generally, such coexistence is possible under particular sets of conditions (Egas, Dieckmann, and Sabelis 2004). Because *S. meliloti* bv. *meliloti* can fix nitrogen with only a subset of *Medicago* species that can interact efficiently with *S. medicae*, it is difficult to understand the coexistence of the two bacterial groups in the same sites without calling for trade-offs among the fitness components of each species, unless host plants are not a limiting

resource for rhizobia. A theoretical work published by Parker (1999) has shown that allopatric differentiation and better symbiotic abilities of the specialist (*i.e.* more nodules, faster nodulation and colonization, or nodules that yield more progeny due to a higher nitrogen fixation on the common host plant) would facilitate the coexistence of both generalist and specialist species of rhizobia. Estimation of these parameters for each species would be helpful to improve our understanding of the system.

Acknowledgements

Lionel Moulin made very useful comments on a late version of the manuscript. We thank Dr. Aouani for providing us Tunisian soil samples and Dr. Athalla for providing us Lebanese soil samples. Lucette Maure and Gaëlle Leonard provided technical help for strain isolation. The research was funded by the Bureau des Ressources Génétiques, Programme « Recherches méthodologiques pour l'amélioration des processus de gestion et de conservation des ressources génétiques animales, végétales et microbiennes » to I.O. and G.B.. X. Bailly was supported by a PhD fellowship from the French Ministry of Education and Research and by a "Jean and Marie-Louise Dufrénoy" fellowship from the French Academy of Agriculture. This is publication ISEM ***** of the Institut des Sciences de l'Evolution, Montpellier.

References

- Ampe, F., E. Kiss, F. Sabourdy, and J. Batut. 2003. Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis. *Genome Biol* 4:R15.
- Barran, L. R., E. S. Bromfield, and D. C. Brown. 2002. Identification and cloning of the bacterial nodulation specificity gene in the *Sinorhizobium meliloti*-*Medicago luciniata* symbiosis. *Can J Microbiol* 48:765-771.
- Bena, G., A. Lyet, T. Huguet, and I. Olivieri. 2005. *Medicago* - *Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. *J Evol Biol* 18:1547-1558.
- Bininda-Emonds, O. R. P. 2004. The evolution of supertrees. *Trends Ecol Evol* 19:315-322.
- Bryant, D., and V. Moulton. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21:255-265.
- Chen, W. M., L. Moulin, C. Bontemps, P. Vandamme, G. Bena, and C. Boivin-Masson. 2003. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. *J Bacteriol* 185:7266-7272.
- Cohan, F. M. 1994a. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am Nat* 143:965-986.
- Cohan, F. M. 1994b. Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol Evol* 9:175-180.
- Cohan, F. M. 1995. Does recombination constrain neutral divergence among bacterial taxa? *Evolution Int J Org Evolution* 49:164-175.
- Cohan, F. M. 2001. Bacterial species and speciation. *Syst Biol* 50:513-524.
- Cohan, F. M. 2002. What are bacterial species? *Annu Rev Microbiol* 56:457-487.
- Daubin, V., M. Gouy, and G. Perriere. 2002. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res* 12:1080-1090.
- Dieckmann, U., M. Doebeli, J. A. J. Metz, and D. Tautz. 2004. Adaptive Speciation. Cambridge University Press, Cambridge, UK.

- Downie, S. R., D. Katz-Downie, E. J. Rogers, H. L. Zujewski, and E. Small. 1998. Multiple independent losses of the plastid *rpoC1* intron in *Medicago* (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA interal transcribed spacer sequences. *Can J Bot* 76:791-803.
- Egas, M., U. Dieckmann, and M. W. Sabelis. 2004. Evolution restricts the coexistence of specialists and generalists: The role of trade-off structure. *Am Nat* 163:518-531.
- Frayse, N., F. Coudere, and V. Poinso. 2003. Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. *Eur J Biochem* 270:1365-1380.
- Galibert, F.T. M. FinanS. R. Long et al. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293:668-672.
- Gaunt, M. W., S. L. Turner, L. Rigottier-Gois, S. A. Lloyd-Macgilp, and J. P. Young. 2001. Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *Int J Syst Evol Microbiol* 51:2037-2048.
- Gavrilets, S. 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution Int J Org Evolution* 57:2197-2215.
- Gevers, D.F. M. CohanJ. G. Lawrence et al. 2005. Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* 3:733-739.
- Goudet, J. 1995. FSTAT (vers. 1.2): a computer program to calculate F-statistics. *J Hered* 86:485-486.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 95:95-98.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22:160-174.
- Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254-267.
- Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules in H. N. Munro, ed. *Mammalian protein metabolism*. Academic press, New York.
- Laguette, G., S. M. Nour, V. Macheret, J. Sanjuan, P. Drouin, and N. Amarger. 2001. Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. *Microbiology* 147:981-993.
- Lawrence, J. G. 2002. Gene transfer in bacteria: speciation without species? *Theor Popul Biol* 61:449-460.
- Maddison, W. P., and D. R. Maddison. 2005. Mesquite: a modular system for evolutionary analysis.
- Maiden, M. C.J. A. BygravesE. Feil et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95:3140-3145.
- Majewski, J. 2001. Sexual isolation in bacteria. *FEMS Microbiol Lett* 199:161-169.
- Majewski, J., and F. M. Cohan. 1999a. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* 153:1525-1533.
- Majewski, J., and F. M. Cohan. 1999b. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* 152:1459-1474.
- Majewski, J., P. Zawadzki, P. Pickerill, F. M. Cohan, and C. G. Dowson. 2000. Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *J Bacteriol* 182:1016-1023.
- Martin, D. P., C. Williamson, and D. Posada. 2005. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21:260-262.
- Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* 23:23-35.

- Maynard Smith, J., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria? *Proc Natl Acad Sci U S A* 90:4384-4388.
- Moulin, L., G. Bena, C. Boivin-Masson, and T. Stepkowski. 2004. Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus. *Mol Phylogenet Evol* 30:720-732.
- Moulin, L., A. Munive, B. Dreyfus, and C. Boivin-Masson. 2001. Nodulation of legumes by members of the beta-subclass of Proteobacteria. *Nature* 411:948-950.
- Nakamura, Y., T. Itoh, H. Matsuda, and T. Gojobori. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* 36:760-766.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia university press, New York.
- Parker, M. A. 1999. Mutualism in metapopulations of legumes and rhizobia. *Am Nat* 153:S48-S60.
- Perret, X., C. Stachelin, and W. J. Broughton. 2000. Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol Rev* 64:180-201.
- Posada, D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol* 19:708-717.
- Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817-818.
- Posada, D., and K. A. Crandall. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* 98:13757-13762.
- Pretorius-Guth, I. M., A. Puhler, and R. Simon. 1990. Conjugal transfer of magaplasmid 2 between *Rhizobium meliloti* strains in alfalfa nodules. *Appl Environ Microbiol* 56:2354-2359.
- Qiu, W. G., S. E. Schutzer, J. F. Bruno, O. Attie, Y. Xu, J. J. Dunn, C. M. Fraser, S. R. Casjens, and B. J. Luft. 2004. Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci U S A* 101:14150-14155.
- Ravigne, V., I. Olivieri, and U. Dieckmann. 2004. Implications of habitat choice for protected polymorphisms. *Evol Ecol Res* 6:125-145.
- Raymond, M., and F. Rousset. 1995a. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86:248-249.
- Raymond, M., and F. Rousset. 1995b. An exact test for population differentiation. *Evolution Int J Org Evolution* 49:1280-1283.
- Reuber, T. L., and G. C. Walker. 1993. Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Cell* 74:269-280.
- Rome, S., M. P. Fernandez, B. Brunel, P. Normand, and J. C. Cleyet-Marel. 1996. *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. *Int J Syst Bacteriol* 46:972-980.
- Roumiantseva, M. L., V. V. Yakutkina, T. Damman-Kalinowski, L. A. Sharypova, M. Keller, and B. V. Simarov. 1999. Comparative analysis of structural organization of the genome in alfalfa nodule bacteria *Sinorhizobium medicae* and *Sinorhizobium meliloti*. *Russ J Genet* 35:128-135.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Sawyer, S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol* 6:526-538.
- Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114-1116.

- Silva, C., P. Vinuesa, L. E. Eguiarte, V. Souza, and E. Martinez-Romero. 2005. Evolutionary genetics and biogeographic structure of *Rhizobium gallicum* sensu lato, a widely distributed bacterial symbiont of diverse legumes. *Mol Ecol* 14:4033-4050.
- Sullivan, J. T., and C. W. Ronson. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci U S A* 95:5145-5149.
- Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876-4882.
- Turner, S. L., and J. P. Young. 2000. The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Mol Biol Evol* 17:309-319.
- Van Berkum, P., D. Beyene, G. Bao, T. A. Campbell, and B. D. Eardly. 1998. *Rhizobium mongolense* sp. nov. is one of the three rhizobial genotypes identified which nodulate and form nitrogen fixing symbiosis with *Medicago ruthenica* [(L.) Ledebour]. *Int J Syst Bacteriol* 48:13-22.
- Vetsigian, K., and N. Goldenfeld. 2005. Global divergence of microbial genome sequences mediated by propagating fronts. *Proc Natl Acad Sci U S A* 102:7332-7337.
- Villegas, M. D. C., S. Rome, L. Maure, O. Domergue, L. Gardan, X. Bailly, J. C. Cleyet-Marel, and B. Brunel. 2006. Nitrogen-fixing sinorhizobia with *Medicago laciniata* constitute a novel biovar (bv. *laciniata*) of *S. meliloti*. *Systematic and applied microbiology in press*:doi:10.1016/j.syapm.2005.1012.1008.
- Villegas, M. D. C., S. Rome, L. Maure, O. Domergue, L. Gardan, X. Bailly, J. C. Cleyet-Marel, and B. Brunel. 2006. Nitrogen-fixing sinorhizobia with *Medicago laciniata* constitute a novel biovar (bv. *laciniata*) of *S. meliloti*. *Systematic and applied microbiology in press*:doi:10.1016/j.syapm.2005.1012.1008.
- Vincent, J. M. 1970. A manual for the practical study of root-nodule bacteria. Blackwell Scientific Publications Ltd., Oxford.
- Vinuesa, P., C. Silva, D. Werner, and E. Martinez-Romero. 2005. Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. *Mol Phylogenet Evol* 34:29-54.
- Weiller, G. F. 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Mol Biol Evol* 15:326-335.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution Int J Org Evolution* 38:1358-1370.
- Wernegreen, J. J., E. E. Harding, and M. A. Riley. 1997. Rhizobium gone native: unexpected plasmid stability of indigenous *Rhizobium leguminosarum*. *Proc Natl Acad Sci U S A* 94:5483-5488.
- Wilkinson, M., F. J. Lapointe, and D. J. Gower. 2003. Branch lengths and support. *Syst Biol* 52:127-130.

Table 1. Primers used in this study

Experiment	Genome location ^a	Primers ^b	Sequence 5' -3'	T _m (°C)	PCR product size (pb)
MLSA	<i>IGS_{RKP}</i> (<i>rkpA</i> – <i>rkpU</i>)	RKP1	AGGCATGCACGCCCTATGA	58.8	599
		RKP2	GCCATCGACATCTACAATATCAA	57.1	
	<i>IGS_{NOD}</i> (<i>nodE</i> – <i>nodG</i>)	NOD1	CAGTTCTGGCATTCAAGC	53.7	579
		NOD2	CCCCCTCCTATGGCTCCTGAT	61.4	
	<i>IGS_{GAB}</i> (<i>gabD5</i> – <i>SmaI</i> 850)	GAB1	CATGACCAAAGACCGCTTCC	59.4	658
		GAB2	GCATGATCGGCCTCAACAC	58.8	
	<i>IGS_{EXO}</i> (<i>exoP</i> – <i>thiD</i>)	EXOmelioti1	CAACAAGACGGATATGAACGAA	56.5	494
		EXOmelioti2	GTGGTGGAAGGATTGACTGC	59.4	
		EXOmedicue1	CATGAACGAGCTGGGCAAAT	57.3	250
		EXOmedicue2	CTGCTCGAAGCGGCAAAA	56.0	
Supertree	<i>nodABC</i>				
		XNA2 ^F	CGCCTTTGGGACAGTTCC	58.2	
		XNA3 ^F	CGGATCGGAGCTATGAAGCA	59.4	
		XNA4 ^R	AGTCCAGCACTGCATCAACAAT	58.4	
		XNA5 ^R	AGCGGCGGTGCTGGTTG	60.0	
		XNA6 ^R	TTCCGAGAACCATCATCAACGAC	60.6	
		XNA8 ^F	CGGCACAGTCTCGCTTCG	60.5	
		XNA7 ^R	CCGAAGCGAGACTGTGCC	60.5	
		XNA9 ^R	GCCTGCCTTCAACATGAGAAT	57.9	
	<i>nodEG</i>				
		XNE3 ^F	AAATGCCCTTCGGTTCGG	56.0	
		XNE4 ^F	GCGGTAGACCAGATCAAGTGC	61.8	
		XNE5 ^F	CGTGGTACTGGCGGAGGGT	63.1	
		XNE6 ^F	TCCATATCTTCCACCAAGTCCA	58.4	
		XNE7 ^F	GCGTGAGCGTAAGGTGCG	60.5	
		XNE8 ^R	CCGTCCAGTCCGACGAT	57.6	
		XNE9 ^R	GCATTGTTGACCAGGATGTCC	59.8	

^a Genome locations refer to the genome of *S. meliloti* strain 1021; ^b ^F, ^R refer to forward and reverse primers respectively.

TABLE 2. Origin and genotype of *S. medicae* and *S. meliloti* isolates.

Location	Taxa / Host plant	Isolates	Allele at indicated locus ^a			
			<i>IGS_{RKP}</i>	<i>IGS_{Exo}</i>	<i>IGS_{GAB}</i>	<i>IGS_{NOD}</i>
Enfidha	<i>S. medicae</i> / <i>M. truncatula</i>	STM 2778	K	P	Y	AH
		STM 2779	K	Q	V	AE
		STM 2780	K	N	Y	AE
		STM 2781	K	P	W	AF
		STM 2782	K	P	T	AF
		STM 2783	K	L	X	AF
		STM 2784	K	P	U	AE
		STM 2785	K	M	T	AF
		STM 2786	K	N	U	AF
		STM 2787	K	P	Y	AH
		STM 2788	K	R	U	AF
		STM 2789	K	P	T	AE
		STM 2790	K	K	Y	AE
		STM 2791	K	Q	Y	AE
		STM 2792	K	K	T	AE
		STM 2793	K	K	U	AE
		STM 2794	K	N	Y	AE
		STM 2795	K	K	U	AF
		STM 2796	K	Q	U	AE
		STM 2797	K	K	U	AE
	<i>S. meliloti</i> bv. <i>meliloti</i> / <i>M. truncatula</i>	STM 2798	F	B	E	K
		STM 2799	C	B	C	A
		STM 2800	E	D	G	K
		STM 2801	A	D	J	C
		STM 2802	C	B	G	A
		STM 2803	C	C	F	K
		STM 2804	C	B	O	K
		STM 2805	C	B	G	K
		STM 2806	C	B	O	K
		STM 2807	A	D	C	K
		STM 2808	A	D	E	K
		STM 2809	C	B	K	K
		STM 2810	C	B	E	K
		STM 2811	C	B	L	G
		STM 2812	C	B	C	A
		STM 2813	C	B	O	K
		STM 2814	C	B	C	A
		STM 2815	C	B	C	A
		STM 2816	C	B	C	A
		STM 2817	C	B	O	K
	<i>S. meliloti</i> bv. <i>medicaginis</i> / <i>M. laciniata</i>	STM 2818	A	B	C	K
		STM 2819	C	B	O	K
		STM 2820	C	B	M	K
		STM 2821	E	D	C	K
		STM 2822	C	B	E	Y
		STM 2823	C	B	E	Y
		STM 2824	C	B	O	R
		STM 2825	B	B	F	S
		STM 2826	C	B	E	Z
		STM 2827	E	D	B	X
		STM 2828	C	B	O	X
		STM 2829	C	B	O	W
		STM 2830	C	B	E	Z

		STM 2831	C	B	C	Z
		STM 2832	C	B	O	R
		STM 2833	C	B	O	U
		STM 2834	C	A	E	X
		STM 2835	E	B	C	V
		STM 2836	C	B	O	W
		STM 2837	C	B	C	Z
		STM 2838	C	B	O	X
		STM 2839	C	B	E	Y
		STM 2840	C	B	O	X
		STM 2841	C	B	E	Y
Hadjeb	<i>S. meliloti</i> bv. <i>meliloti</i> / <i>M. truncatula</i>	STM 2842	C	B	F	K
		STM 2843	A	D	E	D
		STM 2844	C	B	H	K
		STM 2845	C	B	P	K
		STM 2846	C	B	E	K
		STM 2847	A	D	C	K
		STM 2848	E	D	C	K
		STM 2849	A	D	C	K
		STM 2850	C	B	E	K
		STM 2851	C	B	E	K
		STM 2852	C	B	Q	K
		STM 2853	A	D	C	K
		STM 2854	C	B	O	A
		STM 2855	C	B	C	K
		STM 2856	C	B	O	A
		STM 2857	C	B	O	K
		STM 2858	C	B	R	K
		STM 2859	C	B	E	A
		STM 2860	C	B	C	K
		STM 2861	C	B	H	L
		STM 2862	C	B	E	B
		STM 2863	A	B	C	K
		STM 2864	C	B	O	B
		STM 2865	C	B	O	K
		STM 2866	G	B	E	K
		STM 2867	A	D	E	D
		STM 2868	C	B	E	A
		STM 2869	C	B	E	A
		STM 2870	C	B	H	K
		STM 2871	C	B	O	A
		STM 2872	C	B	E	A
		STM 2873	A	B	C	K
		STM 2874	A	D	E	E
		STM 2875	C	B	O	K
		STM 2876	C	B	C	A
		STM 2877	H	D	C	K
		STM 2878	B	B	E	C
		STM 2879	C	B	O	B
		STM 2880	C	B	C	K
		STM 2881	C	C	E	K
		STM 2882	A	B	C	K
		STM 2883	C	B	C	K
	<i>S. meliloti</i> bv. <i>medicaginis</i> / <i>M. lucmiara</i>	STM 2884	C	B	O	W
		STM 2885	C	B	B	U
		STM 2886	C	B	O	U
		STM 2887	C	B	O	Y
		STM 2888	C	B	E	Z

STM 2889	C	B	F	T
STM 2890	C	B	E	Y
STM 2891	C	B	E	X
STM 2892	E	D	B	X
STM 2893	E	D	B	X
STM 2894	C	B	O	X
STM 2895	C	B	O	X
STM 2896	C	B	I	X
STM 2897	E	B	B	V
STM 2898	C	B	E	X
STM 2899	E	B	B	X
STM 2900	C	B	E	Y
STM 2901	E	D	B	X
STM 2902	C	B	O	Y
STM 2903	C	B	O	W

^a Alleles previously described are identified by the same notation than in Bailly *et al.* (in press a).

Table 3. Genetic diversity of *S. medicae* and *S. meliloti* bv. meliloti and *S. meliloti* bv. medicaginis populations isolated from the three host plant groups used in this study^a

Sample	Number of			Genetic diversity ^a				
	sequences	aligned sites ^b	segregating sites	π	R		H	
					Obs	Nmin	Obs	Nmin
<i>S. medicae</i>								
<i>IGS_{RKP}</i>	20	321 (0)	0	nt	1	1.00 (1; 1)	0.000	0.000 (0.000; 0.000)
<i>IGS_{NOD}</i>	20	335 (0)	3	0.0026	3	2.87 (2; 3)	0.595	0.565 (0.426; 0.668)
<i>IGS_{GAB}</i>	20	334 (3)	8	0.0042	6	4.86 (3; 6)	0.779	0.739 (0.647; 0.816)
<i>IGS_{EXO}</i>	20	190 (0)	6	0.0087	7	5.85 (4; 7)	0.839	0.796 (0.711; 0.858)
<i>S. meliloti</i> bv. meliloti								
<i>IGS_{RKP}</i>	66	302 (1)	6	0.0032	7	3.65 (2; 5)	0.480	0.475 (0.279; 0.647)
<i>IGS_{NOD}</i>	66	349 (2)	7	0.0014	8	4.31 (3; 6)	0.544	0.549 (0.353; 0.711)
<i>IGS_{GAB}</i>	66	274 (6)	15	0.0136	13	6.50 (4; 9)	0.799	0.789 (0.700; 0.868)
<i>IGS_{EXO}</i>	66	241 (1)	11	0.0147	3	2.46 (2; 3)	0.369	0.373 (0.189; 0.542)
<i>S. meliloti</i> bv. medicaginis								
<i>IGS_{RKP}</i>	40	302 (0)	4	0.0041	3	2.37 (2; 3)	0.337	0.323 (0.100; 0.511)
<i>IGS_{NOD}</i>	40	415 (4)	24	0.0188	10	6.94 (5; 9)	0.823	0.803 (0.689; 0.884)
<i>IGS_{GAB}</i>	40	274 (5)	11	0.0150	6	4.86 (4; 6)	0.749	0.732 (0.626; 0.811)
<i>IGS_{EXO}</i>	40	241 (1)	11	0.0078	3	2.26 (1; 3)	0.229	0.219 (0.000; 0.416)

^a R, allelic richness; H, Nei diversity index; Obs, observed value; N_{min}, expected value for N_{min} sample size, limits of the 95% confidence interval are in brackets. N_{min} values are set to 20; ^b Number of indel blocs within intraspecific alignments is provided in brackets; ^{nt} Computation were not tractable.

Table 4. Results of Shimodaira-Hasegawa tests among phylogenies presented in Figure 4, obtained from each MLST loci.

Sequence dataset	Best tree	-ln(L)	Alternative tree	-ln(L)	$\Delta[-\ln(L)]^a$	p-value
<i>IGS_{RKP}</i>	<i>IGS_{RKP}</i>	613.65	<i>IGS_{EXO}</i>	911.90	298.25	0.01 [*]
			<i>IGS_{GAB}</i>	915.83	302.18	0.03 [*]
			<i>IGS_{NOD}</i>	1012.60	398.95	0.02 [*]
<i>IGS_{EXO}</i>	<i>IGS_{EXO}</i>	649.27	<i>IGS_{RKP}</i>	1257.67	608.39	0.00 ^{***}
			<i>IGS_{GAB}</i>	1548.31	899.03	0.00 ^{***}
			<i>IGS_{NOD}</i>	1543.96	894.69	0.00 ^{***}
<i>IGS_{GAB}</i>	<i>IGS_{GAB}</i>	769.49	<i>IGS_{RKP}</i>	2581.59	1812.09	0.00 ^{***}
			<i>IGS_{EXO}</i>	2609.34	1839.84	0.00 ^{***}
			<i>IGS_{NOD}</i>	2171.24	1401.74	0.00 ^{***}
<i>IGS_{NOD}</i>	<i>IGS_{NOD}</i>	1122.55	<i>IGS_{RKP}</i>	10786.54	9663.99	0.00 ^{***}
			<i>IGS_{EXO}</i>	11217.36	10094.81	0.00 ^{***}
			<i>IGS_{GAB}</i>	9777.26	8654.70	0.00 ^{***}

^a Difference between $-\ln(L)$ of the best and the alternative topologies: ^{*}, ^{***}, likelihoods of the two topologies differs significantly with $\alpha=5\%$ or $\alpha=0.1\%$ respectively.

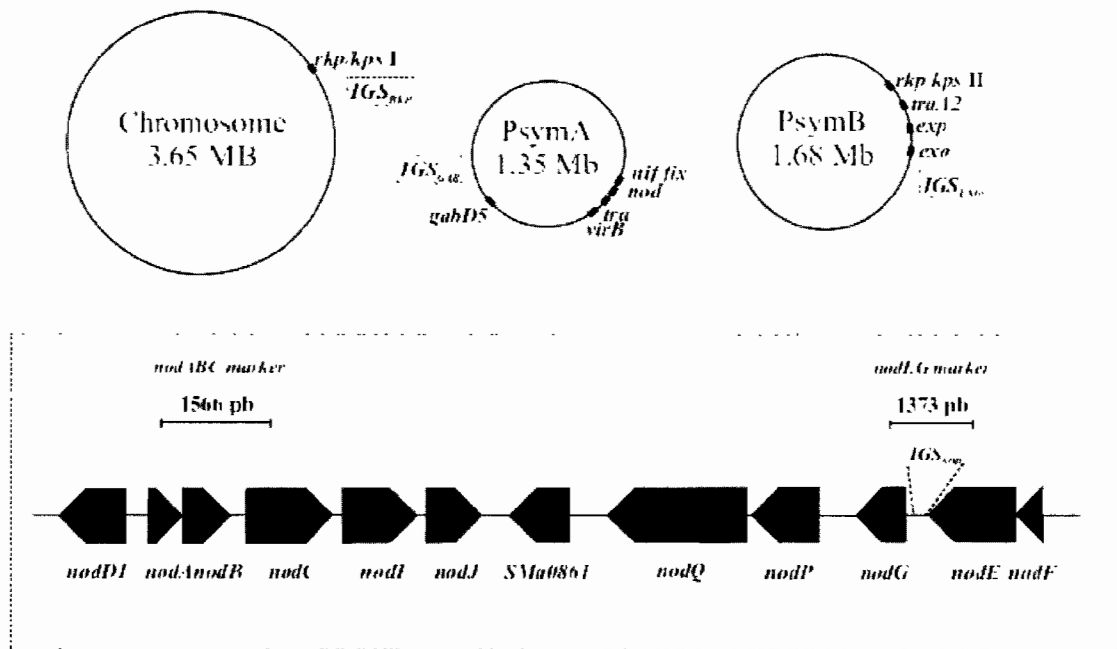


Figure 1. Location of the genetic markers used in this study on the genome of *Sinorhizobium meliloti* strain 1021. The four loci used to perform the MLSA study are framed by dotted lines and their location is indicated by dotted lines. Genes clusters located nearby each genetic marker are indicated by black boxes. It is noteworthy that *IGS_{Exo}* marker is located nearby gene clusters involved in surface polysaccharide production (*rkp*, *kps*, *exp* and *exo* genes) and conjugation (*tra* gene) and that *IGS_{NOD}* marker is located nearby genes involved in symbiotic specificity (*nod* genes), symbiotic efficiency (*nif/fix* genes), secretion (*virB* gene) and conjugation (*tra* genes). The locations of the partial sequences of *nodABC* and *nodEG* genes used to obtain a supertree are indicated above the schematic representation of the symbiotic gene area.

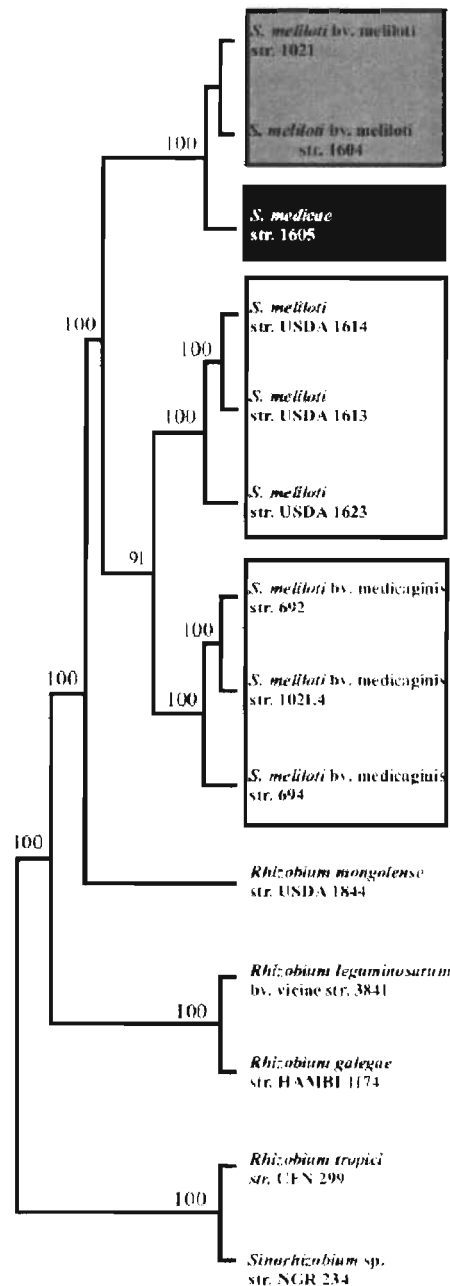


Figure 2. Supertree obtained from *nodABC-nodEG* gene clusters. The numbering below each node refers to bootstrap proportions. The topology indicates that symbionts of the genus *Medicago* grouped in a monophyletic group. Conversely, within *Sinorhizobium*, *S. meliloti* was paraphyletic due to the location of *S. medicae* sequence. *Sinorhizobium* spp associated to the genus *Medicago* clustered in three main groups among which bacterial strains associated to *M. truncatula* (i.e. *S. meliloti* bv. *meliloti* and *S. medicae*), *M. laciniata* (i.e. *S. meliloti* bv. *medicaginis*) or *M. rigiduloides* (i.e. *S. meliloti*).

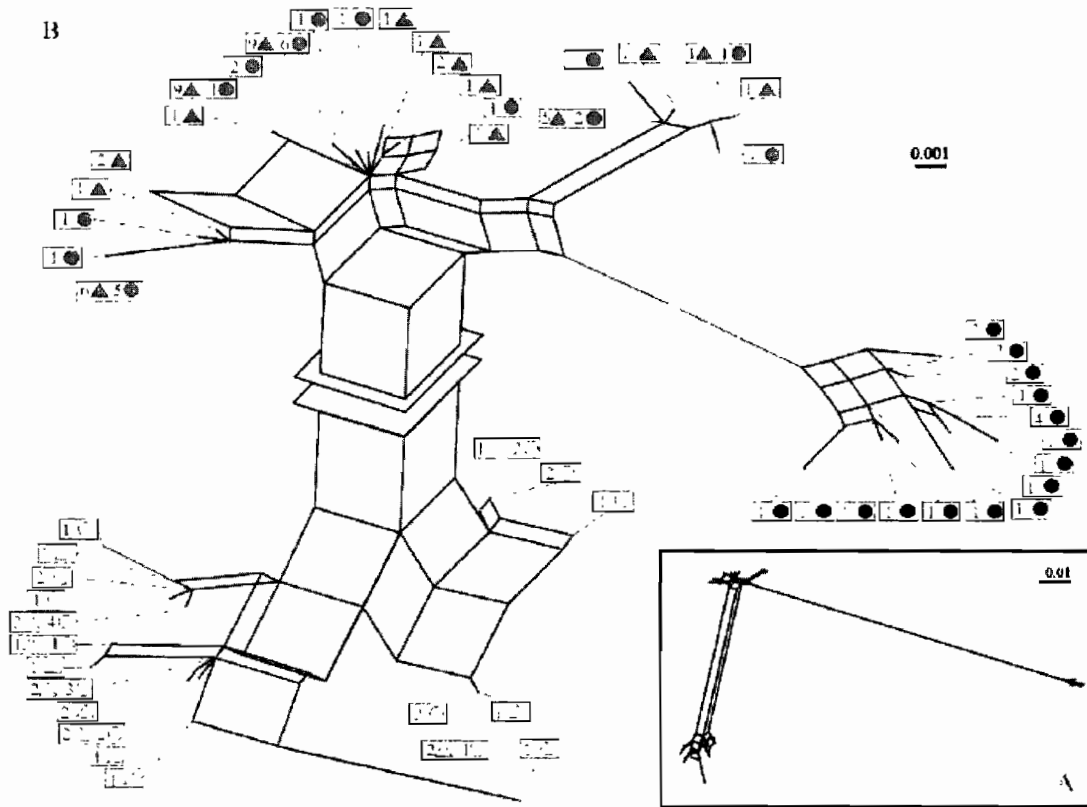


Figure. 3. Full view (A) and focus (B) on Neighbor-Net analysis obtained from 126 complete genotypes of *S. meliloti* bv. *meliloti*, *S. meliloti* bv. *medicaginis* and *S. medicae*. *Sinorhizobium meliloti* and *S. medicae* clusters are connected by a single edge, suggesting sexual isolation between the two species. Conversely, the clusters of the two *S. meliloti* biovars are connected by several edges, suggesting a lack of sexual isolation between the two biovars. Isolates did not cluster according to their geographical origin. The analysis is based on an averaged distance matrix obtain from the four MLSA markers. Branch lengths are drawn to scale. Dotted lines indicate the position of haplotypes on the network. Black, grey and white symbol colours indicate the isolates assigned respectively to *S. medicae* and *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis*. Isolates sampled from Hadjeb and Enfidha are indicated respectively by circles and triangles. The numbering refers to the number of isolates sampled from the different groups.

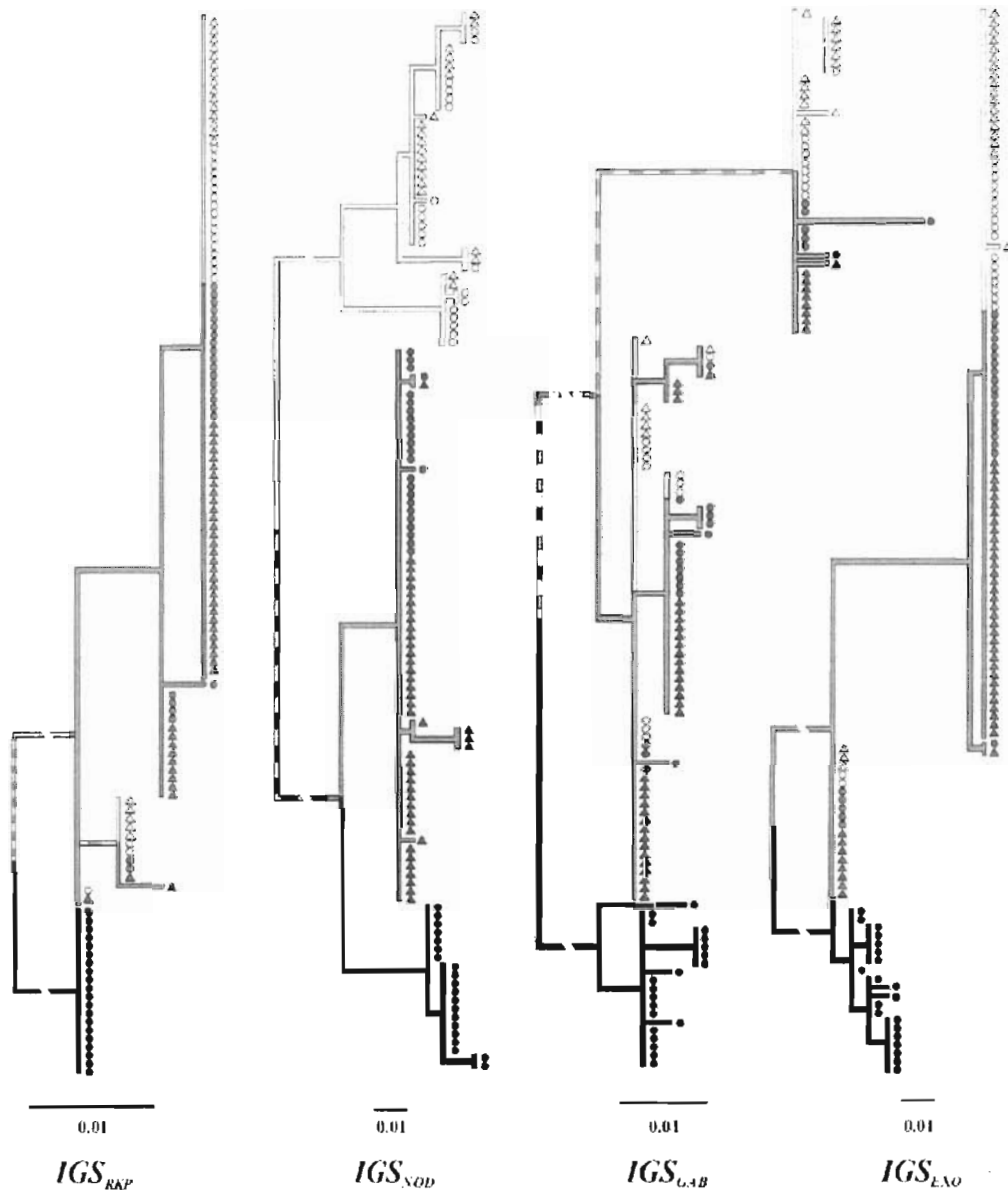


Figure. 4. Maximum likelihood trees obtained from each of the four markers used in the MLSA scheme. *Sinorhizobium medicae* isolates, which are indicated by black symbols colours, grouped in a clade on all phylogenies. This is consistent with the hypothesis of sexual isolation between *S. meliloti* and *S. medicae*. Conversely, isolates belonging to either *S. meliloti* bv. *meliloti* or *S. meliloti* bv. *medicaginis*, which are indicated respectively by grey and white symbol colours, do not cluster in monophyletic groups on all phylogenies but IGS_{NOD}. The three taxa display different symbiotic abilities. The colours of the branches on each phylogeny indicate the inference of ancestral symbiotic features. The incongruence among the four topologies and also the putative patterns of phenotypic evolution we inferred suggest that recombination occurred among *S. meliloti* biovars. Isolates sampled from Hadjeb and Enfidha are indicated respectively by circles and triangles.

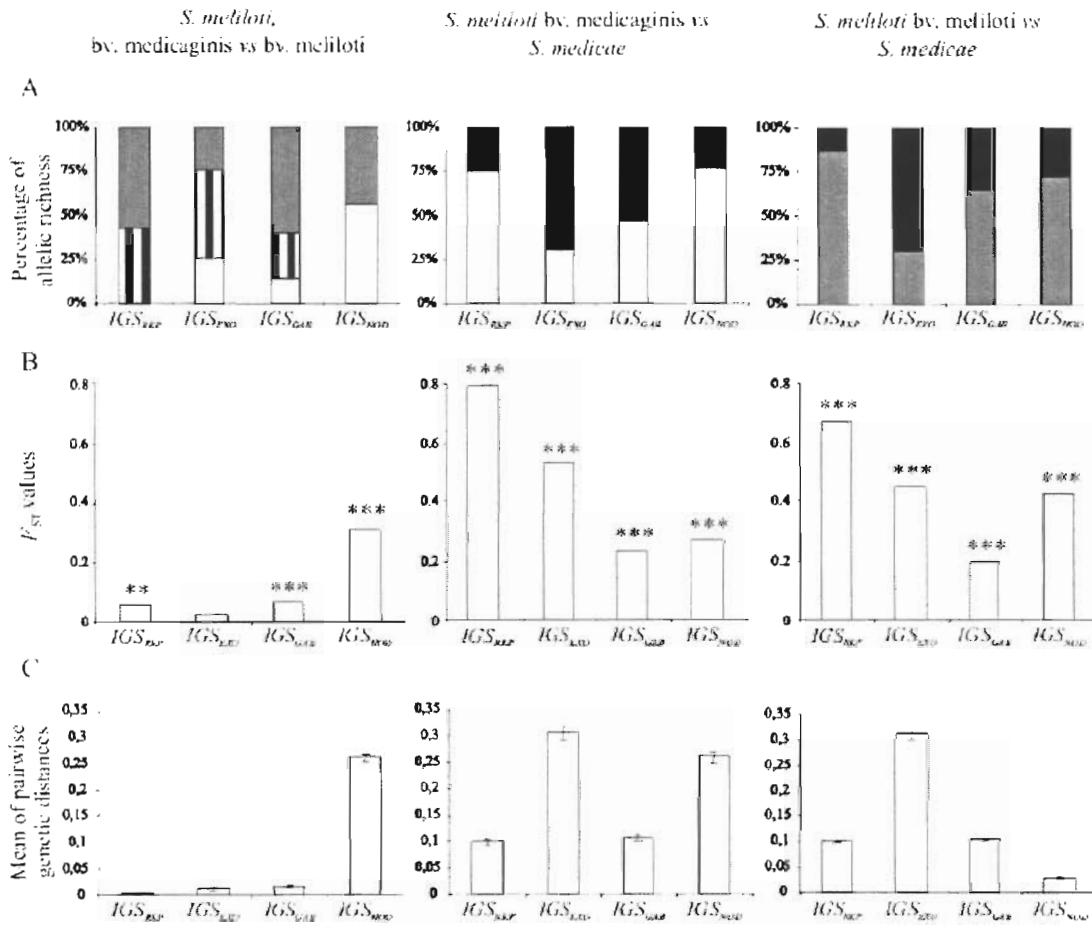


Figure. 5. Distribution of several measures illustrating the dissimilarity between populations pairs belonging either *S. meliloti* bv. *meliloti*, *S. meliloti* bv. *medicaginis* or *S. medicae*. Data are provided independently for each of the four loci used for the multilocus sequence analysis. A: Percentage of allelic richness specific to *S. medicae*, *S. meliloti* bv. *meliloti* and *S. meliloti* bv. *medicaginis* are indicated respectively by black, grey and white area. Striped area illustrates the percentage of allelic richness shared by pairs of taxa. B: distribution of pairwise F_{ST} values between taxa. NS, **, *** indicate respectively differentiation test p -values above 5% threshold, below 1% threshold and below 0.1% threshold. C: average and 95% confidence interval of the genetic distance among individuals belonging to two different taxa.

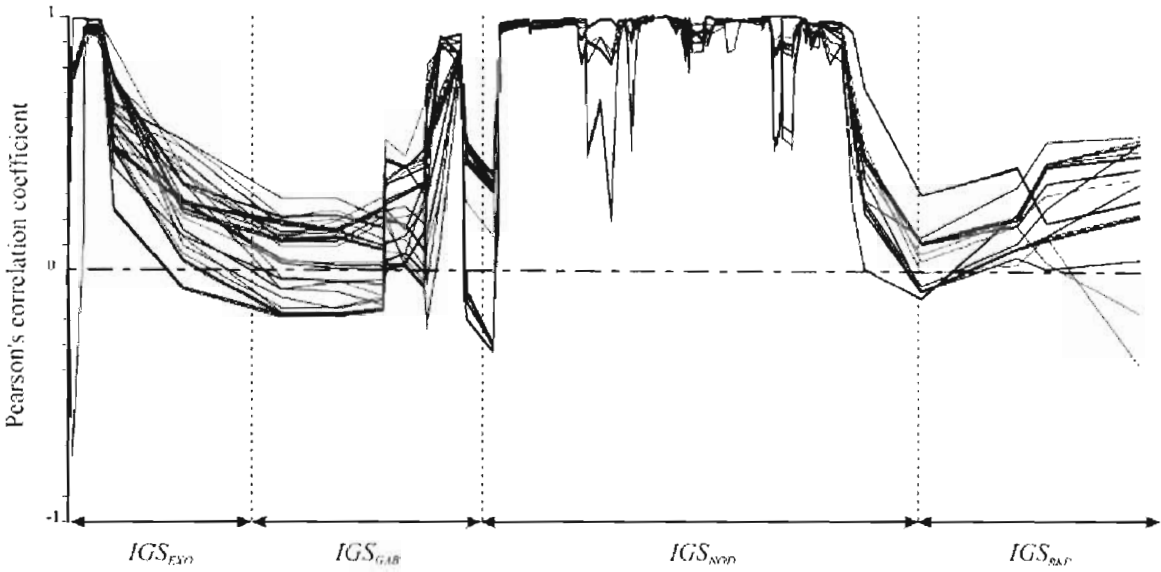


Figure. 6. Profile illustrating recombination breakpoints along a concatenate dataset of MLSA markers obtained from isolates belonging to both *S. meliloti* biovars. Most striking recombination evidences are found at marker pairs boundaries. This suggests that our dataset is mainly affected by recombination events that occur among loci. This pattern is consistent with those obtained with MAXCHI, CHIMAERA and GENECONV.

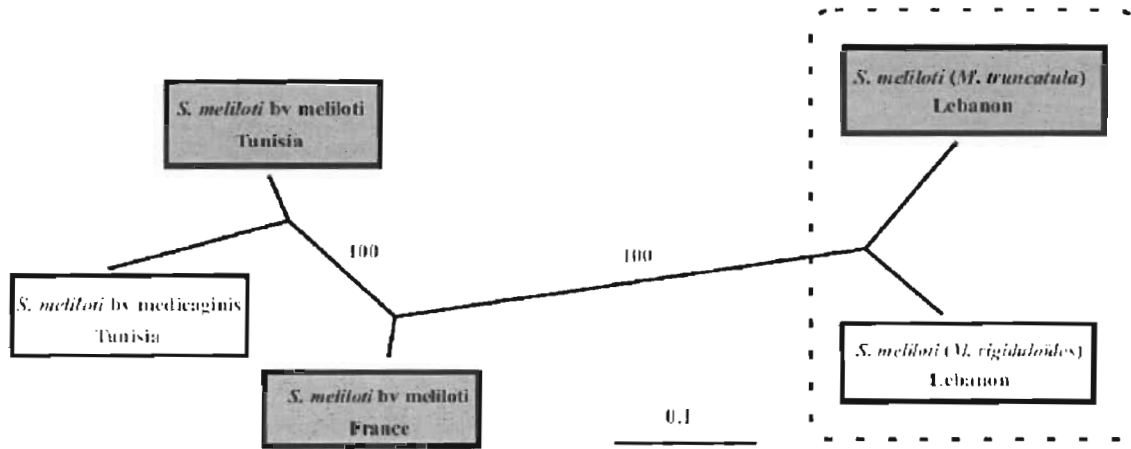


Figure. 7. Population tree based on five populations of *S. meliloti*, i.e. *S. meliloti* bv. *meliloti* populations isolated from France (Bailly et al. in press a) and Tunisia (this study), *S. meliloti* bv. *medicaginis* isolated from Tunisia (this study), and 33 *S. meliloti* isolates obtained from a Lebanese soil, sampled from either *M. truncatula* or *M. rigiduloides* (Bailly et al. in press b). Bacterial populations cluster according to their geographical origin rather than on symbiotic abilities. All global differentiation tests performed between population reject the null hypothesis ($p < 5\%$), excepting the test between Lebanese *S. meliloti* populations isolated from *M. truncatula* and *M. rigiduloides*. These last populations are surrounded by a dotted line. The neighbour joining tree was based on a Cavalli-Sforza distance matrix. This matrix was obtained from genotypes inferred at several polymorphic loci located within *IGS_{RKP}*, *IGS_{EXO}* and *IGS_{GAB}* (see Bailly et al. in press b, to obtain a description of polymorphism assayed).

XIV. Article 4

Available online at www.sciencedirect.comSCIENCE @ DIRECT[®]

Journal of Microbiological Methods xx (2006) xxx–xxx

**Journal
of Microbiological
Methods**
www.elsevier.com/locate/jmicmeth

Development of a lab-made microarray for analyzing the genetic diversity of nitrogen fixing symbionts *Sinorhizobium meliloti* and *Sinorhizobium medicae*

Xavier Bailly, Gilles Béna, Vanina Lenief, Philippe de Lajudie, Jean-Christophe Avarre^{*}

Laboratoire des Symbioses Tropicales et Méditerranéennes, UMR113 IRD/INRA/CIRAD/UM2-Agro M,
Campus International de Baillarguet, Montpellier F34398, France

Received 1 February 2006; received in revised form 9 March 2006; accepted 9 March 2006

Abstract

Some bacterial species, like nitrogen-fixing *Sinorhizobium* that interact with *Medicago* plants, are prone to frequent horizontal gene transfers. Investigation of their genetic structure requires to study polymorphism patterns at many loci. Although DNA microarrays represent a method of choice for high throughput analysis of polymorphisms, this technology yet remains an expensive and heavy approach, thus depriving most of research groups from this powerful tool. In an attempt to overcome this limitation, we have developed a simple genotyping procedure by DNA microarrays, and have evaluated its ability to characterize a *Sinorhizobium* population. Thirty 18- to 24-mer oligonucleotide probes were designed to target the most frequent mutations in three polymorphic loci of *Sinorhizobium meliloti* and *S. medicae*. Probe hybridization efficiency was compared on two spotting surfaces: nylon membranes and epoxy-coated glass slides. Epoxy-coated glass slides revealed more sensitive than nylon membranes and allowed discrimination of single mismatches. Using this procedure, an uncharacterized population consisting of 33 *S. meliloti*:*S. medicae* isolates was successfully genotyped.

© 2006 Published by Elsevier B.V.

Keywords: Microarray; Array surface; Genotyping; *Sinorhizobium*

1. Introduction

In the past two decades, several DNA-based genetic markers have been developed to explore bacterial population diversity. Two types of genotyping methods that differ according to the information they provide can be distinguished. First, fingerprinting methods revealed numerous polymorphisms through the overall genome. They include Random-Amplified Polymorphic DNA (RAPD) (Williams et al., 1990), Amplified Fragment

Length Polymorphism (AFLP) (Vos et al., 1995), and Pulsed-Field Gel Electrophoresis (PFGE) (Schwartz and Cantor, 1984). The main limitation of these techniques is the difficulty to assess the events that caused the observed polymorphisms (i.e. genomic rearrangements, insertion-deletion or punctual mutations). Moreover, RAPD is usually considered as a poor repeatable technique, whereas PFGE is extremely time-consuming and is difficult to use in routine. Secondly, locus-focusing methods examined allelic variation after PCR amplification of a specific DNA sequence. These techniques provide information on punctual variations such as Single Nucleotide Polymorphism (SNP), which

^{*} Corresponding author. Tel.: +33 467593801; fax: +33 467593802.
E-mail address: Jean-christophe.Avarre@mpl.ird.fr (J.-C. Avarre).

are of great interest in ecology, evolutionary or biogeography studies (Brumfield et al., 2003; Mornin et al., 2004). The survey of nucleotide variations from several loci is usually needed to explore and analyze the shape of the diversity from a bacterial population. Although Multi Loci Sequence Typing (MLST) (Maiden et al., 1998) is the most accurate approach to achieve this goal (Gevers et al., 2005), it requires high sequencing throughput and as such, is a relatively expensive approach. Several alternative methods have been developed, such as Multilocus Restriction Typing (MLRT) (Coenye and LiPuma, 2002) or Single Stranded Conformation Polymorphism (SSCP) (Sunnucks et al., 2000). However, genotype assessment is respectively conditioned by restriction enzyme availability or further sequencing, and any new allele may not be identified due to method discrimination limitations.

There is therefore a need for an alternative method which would provide reliable and sensitive DNA sequence information at a larger scale. To overcome these limitations, hybridization-based methods using nucleic acids immobilized on solid state surfaces have recently been developed. DNA microarray technology enables parallel nucleic acid hybridization for a large number of DNA fragments immobilized on a small surface area. It provides a high throughput means of analysis for a wide panel of applications (Drmanac et al., 2002; Lockhart et al., 1996; Pease et al., 1994; Wang et al., 1998). The use of DNA microarray technology for single mismatch discrimination has already been described through resequencing approaches (Hacia, 1999; Zwick et al., 2005). Yet, DNA microarray technology also remains an expensive and heavy approach, thus depriving most of research groups from this powerful tool. Besides, studies that use this technology concern a limited number of model organisms.

The two sister species *Sinorhizobium meliloti* and *Sinorhizobium medicae* (Rome et al., 1996) perform efficient symbiosis with *Medicago* plant species, and represent an interesting model to study the impact of horizontal gene transfer on the diversity of bacterial populations. Their genome includes three replication units, i.e. a chromosome and two megaplasmsids called pSymA and pSymB which display different functional features. While the chromosome harbors most of the housekeeping genes, genes involved in accessory functions are mainly located on megaplasmsids (Galibert et al., 2001). Recently, the population structure of *Sinorhizobium meliloti* and *S. medicae* species has been studied at a local scale in South of France and in Tunisia (Bailly et al., in press). In these studies, isolates were

characterized on several loci located on the three replication units by Multi Loci Sequence Typing (MLST). The obtained results showed that plasmidic markers evolve independently due to recombination within both *Sinorhizobium* species, allowing different selective pressure to act on each unlinked locus. Thus, several loci should be genotyped to obtain an accurate picture of the genetic structure of *S. meliloti* and *S. medicae* populations.

In this paper, we describe a simple procedure, accessible to every laboratory equipped for molecular biology, for genotyping bacterial strains by DNA microarrays. The aims of this study were: (i) to show that a lab-made protocol with a manual microarrayer enables single mismatch discrimination; (ii) to test and compare performances of two oligonucleotide spotting surfaces, in various labeling and hybridization conditions; (iii) to assess the ability of this method to genotype uncharacterized *Sinorhizobium* isolates sampled from a given soil.

2. Materials and methods

2.1. Microarray's oligonucleotide design

For the development of the array, only three distinct DNA regions, or loci, were investigated. These three loci, called "*IGS_{GAB}*", "*IGS_{LM}*" and "*IGS_{RKP}*", correspond to intergenic sequences of *Sinorhizobium meliloti* and *S. medicae* genomes, and are located on pSymA plasmid, pSymB plasmid and the chromosome, respectively. A set of sequences that were previously obtained from Tunisian and French *Sinorhizobium meliloti* and *S. medicae* sympatric bacterial populations, revealed the presence of 26 alleles for *IGS_{GAB}*, 13 for *IGS_{LM}* and 9 for *IGS_{RKP}* (Bailly et al., in press). Oligonucleotides were designed to cover most of the polymorphisms that were found in natural populations on these three chromosomal and plasmidic loci in the two species. Sequences were aligned with ClustalX 1.8 (Thompson et al., 1997), and alignments were edited with GeneDoc (Nicholas et al., 1997). Oligonucleotides were manually designed using Gene Runner 3.05 (Hastings software, Inc.) by seeking the best compromise between the mismatch position along the oligonucleotide and the oligonucleotide size, melting temperature (T_m), and probability of secondary structures (for reference, see Booth et al., 2003). Oligonucleotides were synthesized by Operon Biotechnologies GmbH (Cologne, Germany). Oligonucleotides were individually tested for their hybridization performances, and those that did not yield a sensitive enough signal were redesigned in order

to increase their length, synthesized and tested again. At the end, thirty oligonucleotides were accordingly obtained, with a length ranging from 18 to 24 nucleotides, and whose main features are listed in Table 1.

2.2 Oligonucleotide spotting

Two distinct types of arrays were tested: positively charged nylon membranes and epoxy-coated glass slides. For nylon membranes, lyophilized oligonucleotides were resuspended in TE buffer (100 mM Tris-HCl pH 7, 10 mM EDTA) at a final concentration of 500 μ M. They were manually spotted onto supercharged nylon membranes (Nytran N⁺, Schleicher and Schuell Bioscience, Dassel, Germany) with an 8-pin hand-held microarrayer (Microcaster, Schleicher and Schuell Bioscience, Dassel, Germany). Each oligonucleotide was spotted in triplicate. After arraying, oligonucleotides were covalently fixed to the membranes with an

automated cross-linker (Stratagene) at an energy of 120 mJ. Membranes were then rinsed with sterile water for 5 min, air dried and stored at RT till use. For glass slides, lyophilized oligonucleotides were resuspended in 1 $\times$ Nexterion Spot buffer solution (Schott, Jena, Germany) at a final concentration of 20 μ M. They were manually spotted onto Nexterion Slides E (Schott, Jena, Germany) with the same hand-held microarrayer (Microcaster, Schleicher and Schuell Bioscience, Dassel, Germany). For this array, each oligonucleotide was spotted in duplicate. Slides were subsequently treated according to the manufacturer's instructions. In both cases, positive controls consisted of an equimolar mix of all the oligonucleotides, and negative controls consisted of oligonucleotide resuspension buffer.

2.3. Target labeling and hybridization

PCR amplifications of each loci were carried out for each strain following Bailly et al. (in press) procedure.

Table 1
Oligonucleotide sequences and characteristics (sequence mutations are highlighted)

Probe #	Oligo name	Sequence 5' to 3'	Length (nt)	GC %	Tm* (°C)
1	exomed-93	gctgcctgaaggtatgccg	19	68.4	75.7
2	exomed-93bis	gctgcctgaaggtatgccg	19	73.7	77.9
3	exomed-123	gcttggtctgtacactaatcc	24	50.0	74.3
4	exomed-123bis	gcttggtctgtacactaatcc	24	54.2	76.6
5	exomed-123n2	tgggtctgtacactaatcc	22	45.5	69.3
6	exomed-123n2bis	tgggtctgtacactaatcc	22	50.0	71.8
7	exomed-96	gcttttgcctcttgcgc	19	63.2	79.0
8	exomed-96bis	atttgtgtgcgcgcgcct	18	55.6	74.8
9	exomed-113	tctgcctgaagccttgc	18	61.1	75.5
10	exomed-113bis	tctgcctgaagccttgc	18	66.7	75.1
11	exomed-132	tgtcactaatctctccacca	23	52.2	76.7
12	exomed-132bis	tgtcactaatctctccacca	23	56.5	76.8
13	gabmed-65	ggacgcgaatctcctgaagg	20	60.0	76.6
14	gabmed-65bis	ggacgcgaatctcctgaagg	20	55.0	74.4
15	gabmed-184	cagcgcggactataatgaagg	21	52.4	72.7
16	gabmed-184bis	ccagcgcggactataatgaagg	21	57.1	77.1
17	gabmed-166	gctggacaggaatcggtaga	20	60.0	72.3
18	gabmed-166bis	gctggacaggaatcggtaga	20	65.0	74.7
19	gabmed-192	gcagcctataaagggaaggcgc	22	54.5	76.8
20	gabmed-192bis	gcagcctataaagggaaggcgc	21	57.1	75.5
21	gabmed-192ter	gcagcctataaagggaaggcgc	21	61.9	78.3
22	gabmed-428	ccagacccctaggaccctagt	22	54.5	71.0
23	gabmed-428bis	ccagacccctaggaccctagt	22	59.1	75.8
24	gabmed-428ter	caagacccctaggaccctagt	21	61.9	74.4
25	rkpmel-58	attcctgctatcggtatcc	21	52.4	73.2
26	rkpmel-58bis	attcctgctatcggtatcc	21	57.1	75.4
27	rkpmel-160	ctctctgacacatccaatac	21	57.1	74.3
28	rkpmel-160bis	ctctctgacacatccaatac	21	52.4	72.0
29	rkpmel-193	acggctcgtatcaatcgaagg	23	43.5	72.7
30	rkpmel-193bis	acggctcgtatcaatcgaagg	23	47.8	74.7

*Tm was calculated using the nearest-neighbor thermodynamic values method (Breslauer et al., 1986), with the following conditions: oligonucleotide concentration of 1.25 nM (concentration of membrane-bound oligonucleotides during hybridization) and salt concentration of 395 mM (Na⁺ concentration in hybridization buffer).

2.3.1. Hybridization on nylon membranes

Specificity of the designed oligonucleotides was assessed on nylon membranes. Approximately 30 ng of each PCR product were independently labeled with [α - 32 P] dCTP by random priming with the Bioprime DNA labeling kit (Invitrogen) according to the manufacturer's instructions. The priming reaction mixture contained as final concentrations: 1 $\times$ Random Primers Solution, 100 μ M dATP, dGTP and dTTP, 0.33 μ M of [α - 32 P]dCTP (5 μ L, Amersham) and 40 U Klenow fragment. Labeling was carried out at 37 °C for 2 h, and reaction was stopped with 5 μ L stop buffer. Labeled DNA was cleaned from unincorporated nucleotides through Probe Quant G50 columns (Amersham). Membranes were prehybridized for two hours at 42 °C in a hybridization oven (Appligene) under rotation in 2 ml of 25% formamide, 2.5 $\times$ Saline Sodium Citrate (SSC, 375 mM NaCl, 37.5 mM Na citrate), 7% Sodium Dodecyl Sulfate (SDS), 1 $\times$ Denhardt's solution and 200 μ g ml $^{-1}$ denatured salmon sperm DNA (Invitrogen). Labeled target was heat-denatured and added to the hybridization buffer, and membranes were hybridized for 16 h in the same conditions. After hybridization, membranes were washed two times for 15 min at room temperature in 5 ml of 2 $\times$ SSC, 0.5% SDS and two times for 15 min at 42 °C in 5 ml of each of the following buffers: 0.5 $\times$ SSC, 0.5% SDS and 0.2 $\times$ SSC, 0.5% SDS. Membranes were air dried and exposed to a phosphor screen for several hours, and the phosphor screen was scanned with a phosphorimager (Typhoon, Amersham).

2.3.2. Hybridization on epoxy-coated glass slides

To compare the hybridization performances on another substrate, hybridizations were also carried out on epoxy-coated glass slides. Five nanograms of each of the 3 PCR products were mixed and used as template for labeling, which was carried out with the Bioprime DNA labeling kit (Invitrogen) as described above. Labeled DNA was purified with the Qiaquick PCR purification kit (Qiagen), eluted in 50 μ L H $_2$ O, evaporated by speed-vacuum, resuspended in 30 μ L of hybridization buffer and heat-denatured. Hybridization buffer was the same as described above for hybridizations on nylon membranes. Spotted slides were hybridized for 16 h at 42 °C with this preparation under coverslip. After hybridization, slides were washed exactly as described above, with 5 ml of washing solution per slide. Slides were then air dried and exposed to a phosphor screen for 2 to 6 h.

2.4. Hybridization signal analysis

Hybridization signals were quantified with Image Quant TL v2003.02 software (Amersham). Since the

arrayer allows to spot up to 768 probes on a single membrane, a grid consisting of 32 $\times$ 24 surface areas (as circles) was placed on each scanned slide and signal intensities were measured for each of these 768 circles. Background noise was subtracted by the spot edge average method. Raw results were expressed as pixels/cm 2 and analyzed as follow. Intensity distribution of all the "circles" corresponding to the noise, i.e. areas that were not spotted with an oligonucleotide, was calculated, and a detection threshold for positive signal was arbitrarily set for all the values exceeding 99.5% of the background noise values. Only signal spots which intensity was higher than this threshold were considered as significant.

In order to compare signal intensity between membranes and slides, a relative signal intensity was calculated as follow: relative signal intensity = signal intensity / slide average intensity, where slide average intensity consisted of the mean intensity of the 768 circles. A signal/background ratio was also calculated as signal intensity / background intensity, where background intensity consisted of the mean intensity of all the non-spotted surface areas (circles) of the membranes or slides.

2.5. Melting temperature analysis

In order to verify the accuracy of T $_m$ prediction, the melting temperature was calculated for each oligonucleotide. For this purpose, complementary sequences to the 30 oligonucleotides were synthesized by Operon Biotechnologies GmbH (Cologne, Germany). Oligonucleotides were diluted to 25 μ M in H $_2$ O. Two microliters of each oligonucleotide was mixed with its complementary (final concentration of 2.5 μ M) together with 2 μ L of a solution containing SybrGreen (Molecular Probes) diluted 1:3000 in DMSO+BSA at 500 μ g ml $^{-1}$ in a final volume of 20 μ L. Samples were transferred to a 96-well plate (Stratagene) and subjected to the following melting profile in a Mx3000P thermocycler (Stratagene). DNA was denatured for 5 min at 98 °C, renatured for 15 min at 30 °C and DNA melting was measured by increasing temperature to 95 °C at a rate of 2 °C/min. Resulting T $_m$ s are the average of 5 separate measurements.

2.6. Bacterial population sampling and DNA extraction

Bacterial sampling was done by trapping symbiotic nitrogen fixing bacteria with the Fabaceae *Medicago truncatula* and *Medicago riguloides* species, from a soil collected near Terbol (Bekaa valley, Lebanon).

Trapping was performed in Gibson tubes, following the procedure previously described (Brunel et al., 1996). Seventeen *M. rigiduloides* and 16 *M. truncatula* were grown individually in tube, and about 2 months after, nodules were harvested and stored in 25% glycerol at -80°C until use. For bacterial isolation, nodules were thawed, surface sterilized by immersion in 3% CaClO for 3 min, rinsed 3 times for 1 min in sterile distilled water and crushed in a drop of sterile distilled water. The resulting suspension was streaked on Yeast Mannitol (YM) agar plates (Vincent, 1970). Colonies were checked for purity by repeated streaking of single-colony isolates on YM plates. Pure bacterial isolates were grown in 10 ml of YM medium for 3 days at 28°C under shaking. Bacteria were pelleted and washed twice with sterile water. Total genomic DNA was extracted according to the procedure previously described (Chen and Kuo, 1993). Any trace of RNA was removed by treating the extracted DNA with $50\text{ }\mu\text{g ml}^{-1}$ RNase A (Sigma) for 30 min at 37°C , and DNA integrity was verified by OD measurements and agarose gel electrophoresis.

3. Results and discussion

3.1. Hybridization optimization

Thirty oligonucleotide probes were designed to target most of the polymorphisms previously reported at three loci (*IGS_{GAB}*, *IGS_{EXO}* and *IGS_{RKP}*) corresponding to intergenic sequences of *Sinorhizobium meliloti* and *S. medicae* genomes. PCR products of these regions were obtained and hybridized on arrays containing these 30 oligonucleotides.

The length of the oligonucleotide probes varied between 18 and 24 nucleotides. Ideally, all oligonucleotides should display nearly the same size and very close T_m and GC content, in order to yield an optimal hybridization signal under similar hybridization conditions. However, such “ideal” conditions were impossible to fulfill in most cases, which is the reason why the oligonucleotide length fluctuated around 21 nucleotides, their T_m varied between 69.3 and 79.0°C and their GC content was comprised between 43.5% and 73.7% (see Table 1). For point mutation detection, (Hacia et al., 1996) proposed that mismatches should be located in the center of the oligonucleotide. In our case, mismatches were not always located in the center of the probes because of the aforementioned constraints. Yet, a dissociation temperature (T_d) analysis previously performed (Urakawa et al., 2003) indicated that the largest T_d differences between perfect match and mismatch

probes were not necessarily observed for probes with mismatches located in the middle but rather for probes with mismatches located slightly off center (for illustration, see mismatch-harboring probes s3, s7 and s13 of Table 1 of this study).

Several hybridization temperatures as well as hybridization buffers with varying SDS and formamide contents were tested (data not shown), and best results were obtained with the conditions described in the MM section. Although the effect of many chemical compounds on nucleic acid interactions have been well studied (Chang et al., 1974; Hutton, 1977; Wetmur and Davidson, 1968), the effect of SDS concentration is not documented, except for the recent study of Rose et al. (2002). The use of a high concentration greatly improved hybridization signals in our case. This suggests that SDS may play an important role in DNA/DNA duplex formation and stability, in addition to reducing background (Rose et al., 2002).

We also tested several probe and target concentrations. On membrane arrays, oligonucleotides were spotted at concentrations varying between 50 and $500\text{ }\mu\text{M}$ and hybridized with 30, 50 or 100 ng of PCR product. Best results, in terms of signal intensity and background noise, were obtained for $500\text{ }\mu\text{M}$ of probe and 30 ng of labeled target. On slide arrays, oligonucleotides were spotted at 10, 20 and $40\text{ }\mu\text{M}$ and were hybridized with 5, 25 and 50 ng of PCR products. Best results were obtained with 10 and $20\text{ }\mu\text{M}$ of probes and 5 ng of target. No difference was observed between 10 and $20\text{ }\mu\text{M}$ of probes, and the concentration of $20\text{ }\mu\text{M}$ was kept.

3.2. Assessment of oligonucleotide hybridization efficiency

Several authors have demonstrated that nylon microarrays are a reliable tool for expression profiling studies (Bertucci et al., 1999, 2003; Richmond et al., 1999). Compared to glass slides, membrane microarrays are expected to yield a higher sensitive hybridization signal due to their higher DNA binding capacity coupled to the use of radioactively labeled targets. In addition, they also present the great advantage over slide microarrays to be cheaper. For all these reasons, we first decided to carry out hybridizations on nylon membranes.

Hybridization efficiency of each oligonucleotide was tested by hybridizing spotted nylon membranes with a single PCR product corresponding to one locus amplified from a *S. meliloti* or a *S. medicae* strain which sequence was already known, and which

resulting hybridization pattern was hence predictable. Each oligonucleotide was tested in at least two separate hybridization experiments. Fig. 1 shows the resulting relative hybridization intensity for each oligonucleotide with its perfect match target. Except for oligonucleotide 20, all of the oligonucleotides yielded detectable hybridization signals under the same conditions. Signal intensities were comparable between the three replicates of a membrane, and also between the replicates of distinct membranes, as revealed by relatively low standard deviations. Calculation of relative signal intensity enabled to overcome the possible discrepancies that could originate from labeling efficiency and thus to compare signal intensity between distinct membranes. Labeling efficiency was controlled by measuring the number of counts of labeled targets and proved to be rather constant over samples (data not shown).

Fig. 1 also highlights that signal intensities may be highly variable for oligonucleotides displaying nearly similar length and/or GC content, indicating that signal intensity does not only depend on oligonucleotide length and base composition, but also depends on the oligonucleotide sequence itself (compare Table 1 with Fig. 1). It is commonly admitted that the oligonucleotide hybridization yields greatly depend on their T_m . There exists a variety of methods to calculate the T_m of short oligonucleotide sequences, which most of the time lead to large and significant differences in the T_m values (Panjkovich and Melo, 2005). It is therefore important to choose the most accurate T_m determination method. In our case, we used the nearest-neighbor thermodynamic model (Breslauer et al., 1986), which enables to predict the melting behavior of a DNA duplex from its base sequence. Yet, very recently, (Panjkovich and

Melo, 2005) compared the most frequently used models for T_m calculation and proposed a consensus model for short oligonucleotides. We therefore submitted our oligonucleotide sequences to their algorithm (<http://dna.bio.puc.cl/tm.html>, (Panjkovich et al., 2005)) to obtain the resulting consensus T_m values. We also measured the T_m values of the oligonucleotides by hybridizing them in solution with their complementary sequences (see MM section). We then compared the 3 T_m values obtained for each oligonucleotide with their corresponding relative signal intensities, in order to evidence a possible correlation between relative signal intensity and T_m (Fig. 2). None of these T_m values, either calculated or measured, did allow a good prediction of hybridization yields. Very recently, the hybridization kinetics of short DNA sequences in solution and on microspheres was compared (Sekar et al., 2005). The authors reported important differences in the kinetics for a given sequence in solid versus solution phase and found that sequences do not impact reaction rates in the same way on surface than they do in solution. Analysis of data from commercial microarrays also suggested that solid-phase hybridization is less thermodynamically favored than hybridization of the same sequences in solution (Held et al., 2003). These results therefore confirm that hybridization signals with tethered probes are difficult to predict a priori, and that T_m values measured or inferred from solutions are not an accurate tool.

3.3. Membrane arrays versus slide arrays

Although nylon membranes yielded detectable and reproducible hybridization signals for most probes,

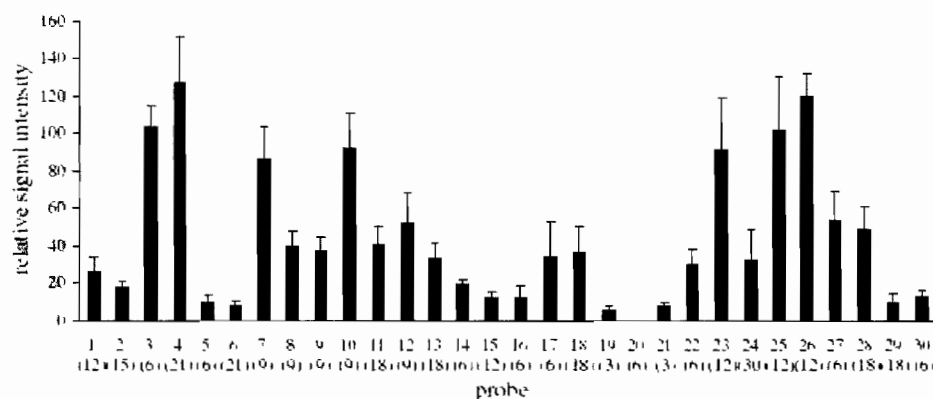


Fig. 1. Relative signal intensities of each probe to its perfect match target. Signal was quantified as described in MM section. Numbers in parentheses below the probe labels indicate the number of spots that have been quantified for each oligonucleotide. Results correspond to the means $\pm$ standard deviations.

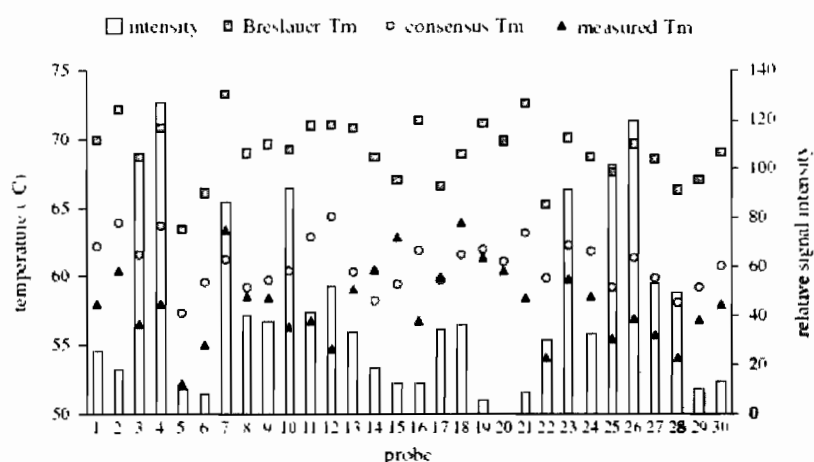


Fig. 2. Comparison of the relative hybridization signal intensity of each oligonucleotide with their measured and calculated Tm (according to Breslauer tables (Breslauer et al., 1986) and consensus tables (Panjkovich and Melo, 2005)). For clarity, and since standard deviations of measured Tm (corresponding to the average of 5 independent measurements) were very low, they are not presented. Relative signal intensities are the same as in Fig. 1, and for clarity standard deviations are not presented.

intensities were rather weak, and background noise rather high. We therefore decided to test another surface to try to improve the hybridization yields. In an attempt of optimizing protocols for glass slide-based DNA microarrays, it was reported that over all the surfaces tested, epoxy-coated glass slides appeared as the best compromise in terms of signal intensity and sensitivity

(Wrobel et al., 2003). Accordingly, we tested hybridizations on epoxy-coated glass slides, and compared hybridization profiles with those obtained on nylon membranes. Fig. 3 presents the hybridization results of an equimolar mix of the three PCR products corresponding to the three loci of a *S. meliloti* strain which genotype was known. Nylon membranes were

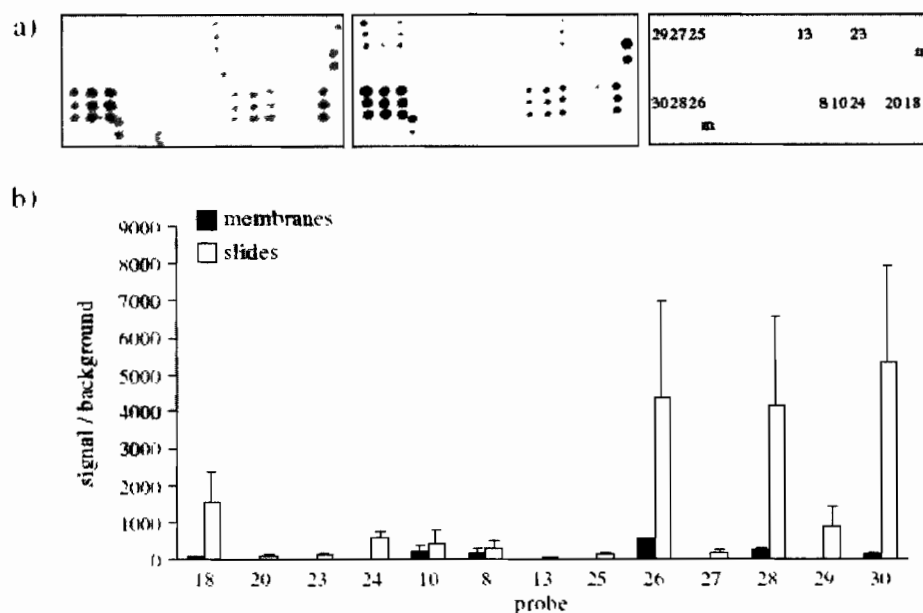


Fig. 3. Two nylon membranes and 2 epoxy-coated glass slides were hybridized with an equimolar mix of three PCR products corresponding to the three studied loci of one *S. meliloti* strain. (a) Resulting hybridization profile of one replicate on nylon membrane (left), on epoxy-coated glass slide (center), and corresponding spotting pattern with probe numbers (right). (b) Resulting signal/background ratios, expressed as means ($n=6$) $\pm$ standard deviations.

exposed for 24 h whereas epoxy-coated glass slides were only exposed for 2 h. Fig. 3a shows that hybridization profiles between the two surfaces were not exactly identical. According to the genotype of the tested *S. meliloti* strain, probes 18, 24, 10, 8, 26, 28 and 30 were expected to give a signal. These probes were detected and quantified on both surfaces. However, cross-hybridizations were also observed. Probe 13, which is specific to *S. medicae* strains but differs from *S. meliloti* sequences by one nucleotide located at position 8 from the 3' end, was detected on nylon membranes. On glass slides, several unspecific probes, in addition to oligonucleotide 13, were also detected: oligonucleotides 23, 25, 27 and 28, which all differ from oligonucleotides 24, 26, 28 and 30, respectively, by one single nucleotide (see Table 1); in some cases and at a lower extent, oligonucleotide 20, which differs from the tested sequence by two nucleotides and one gap. Furthermore, relative signal intensities of each oligonucleotide were also differing between nylon and epoxy (Fig. 3b): for example, probe 26 yielded the highest signal on nylon membranes, whereas the highest intensity was obtained with probe 30 on epoxy slides. Lastly, signal intensity was far stronger on glass slides, leading to a much lower signal/background ratio (Fig. 3b). Also, oligonucleotide 20, which was never detected on nylon membranes, yielded a detectable signal on epoxy-coated glass slides.

Altogether, these results indicate that epoxy-coated glass slides are far more sensitive than nylon membranes, and yield a higher signal/background ratio. They also indicate that a given oligonucleotide affixed to two different surfaces behaves in a pretty different manner. This observation suggests that a specific attention should be brought to the choice of the surface to use in microarray experiments, which might greatly interfere with the final result. Information on the performance of the large variety of substrates available for microarray experiments is crucially lacking. Such comparative studies, yet, should reveal extremely valuable for experimentalists dealing with complex sample mixtures.

Though signal intensities on nylon membranes were weaker, they showed less variation within and among membranes than signal intensities on epoxy slides, as revealed by lower standard deviations (Fig. 3b). Measurement of the spot diameter on both surfaces showed much larger variations on slides than on membranes, since the average diameter was indeed $385 \pm 104 \mu\text{m}$ on slides and $445 \pm 13 \mu\text{m}$ on membranes ($N = 74$ and 64 , respectively). The manual arrayer used in this study delivers oligonucleotides by contact of

slotted pins to the surface area. When the surface is soft, it is more likely that liquid deposition will be facilitated and more accurate than on a hard surface. This may probably explain why spot shape and intensity was more homogeneous on nylon membranes than on glass slides. This is probably one of the limits of using such a manual arrayer on hard surfaces such as epoxy-coated glass slides.

3.4 Probe concentration

For membrane arrays, we found that the optimal probe concentration was $500 \mu\text{M}$, whereas the optimal concentration was $20 \mu\text{M}$ for slide arrays. If we approximate that the average molecular weight of a nucleotide is 330 g mol^{-1} , that the average pin delivery volume of the hand microarrayer is 50 nl (communicated from Schleicher and Schuell), and that all the DNA spotted binds to its substrate and is available for hybridization, then the amount of DNA per spot for a 20-mer is $\sim 165 \text{ ng}$ with a $500\text{-}\mu\text{M}$ solution, and $\sim 6.6 \text{ ng}$ with a $20\text{-}\mu\text{M}$ solution, which corresponds to $1.5 \cdot 10^{13}$ molecules/spot for membrane arrays and $6 \cdot 10^{11}$ molecules/spot for slide arrays. If we consider that after random priming, the average target length is 200 bases (data not shown), the target amount for hybridization was $2.8 \cdot 10^{11}$ and $4.6 \cdot 10^{10}$ molecules for membranes and for slides, respectively. In both cases, oligonucleotides were hence in excess as compared to their targets. As recently pointed out (Lexick and Horgan, 2005), although end-attachment of probes is often intended, some surface chemistries might lead to cross-linking at internal sites, which can hinder association with target molecules. According to the manufacturers, oligonucleotides are likely to be covalently linked to the surface of supercharged nylon membranes along most of their length (covalent bonds form between the nucleotide phosphate groups and the ammonium groups of supercharged nylon), whilst they are likely to be covalently linked to epoxy-coated glass slides at one or two sites only (covalent bonds form between the amine groups of the bases and the epoxy). Consequently, probe accessibility for its target seems to play a greater role in DNA–DNA hybridization than probe amount itself.

3.5 Target length

It is generally believed that ideally, targets and probes should have the same length for best hybridization yields (Southern et al., 1999). In our case, target length varied from 250 to 658 bp. Apparently, such

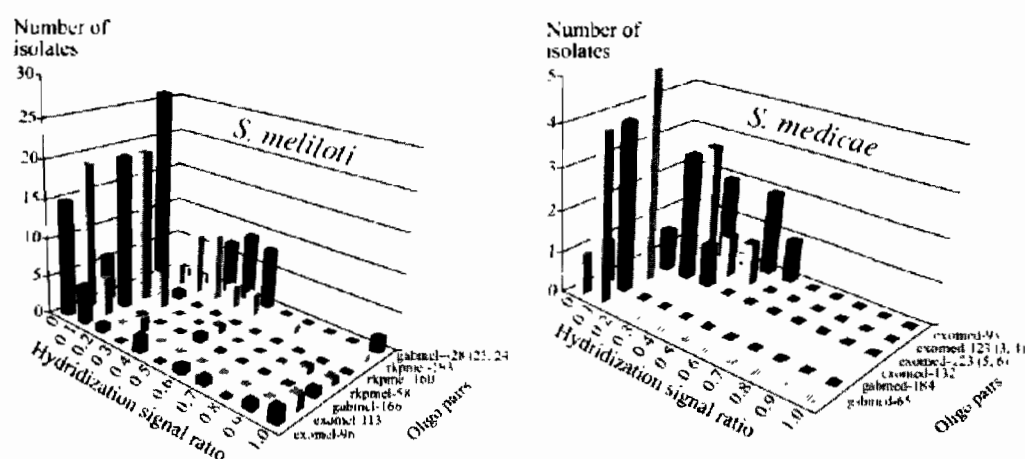


Fig. 4. Distribution of signal hybridization ratios for each set of oligonucleotides for *S. meliloti* and *S. medicae*. For *S. meliloti* gabmel-428, only probes 23 and 24 were considered since probe 22 was never detected. Also, the probe set corresponding to *S. meliloti* gabmel-192 was not included as a new allele was evidenced in the majority of isolates.

long targets did not hamper the probe-target duplex formation yield, demonstrating that under stringent conditions, good hybridization yields can be obtained with targets up to 30 times longer than their probes. This can probably be explained by the labeling procedure that was used, since random priming reduces the target size to ~200-bp fragments. Therefore, the use of random priming labeling on long PCR products presents two main advantages: (1) it simplifies the target preparation procedure since there is no need for nested PCR or DNA fragmentation, and (2) it allows screening of a larger DNA region, and consequently increases the cost-effectiveness of such genotyping microarrays, as it multiplies the number of polymorphic sites that can be simultaneously screened during one single hybridization.

3.6. Genotype analysis of a population of *S. meliloti* and *S. medicae*

Since hybridization intensities were higher on epoxy-coated glass slides, the analysis was performed on this substrate. Furthermore, signal intensity ratios were calculated to discriminate between perfect-match and mismatch products. The bacterial content of 33 nodules isolated from *Medicago truncatula* (16 nodules) and *M. rigiduloides* (17 nodules) was analyzed. The three studied loci (*IGSGAB*, *IGSEXO* and *IGSRKP*) were amplified by PCR for these 33 bacterial isolates. After purification, PCR products corresponding to the 3 loci were pooled for each isolate and hybridized to spotted epoxy-coated glass slides. Over 33 bacterial isolates, 28

belonged to *S. meliloti* and 5 belonged to *S. medicae*. Except for probe sets 23/24 (gabmel-428) and 29/30 (rkpmel-193), signal intensity ratio between perfect-match and mismatch targets was above 70:30 in most cases. Over the 229 genotypes that were determined, only six hybridization patterns were ambiguous with a ratio ~0.5. Since the distribution of intensity ratios was almost bimodal for all the loci, genotypes could be confidently assigned (Fig. 4). Sequencing of *IGSGAB* for 7 isolates, of *IGSEXO* for 3 isolates and of *IGSRKP* for 5 isolates taken randomly, always confirmed the genotype revealed by the array (data not shown). Furthermore, except for three isolates, hybridization signals with probes 19, 20 and 21 (gabmel-192) were always null or close to the background level. We suspected that this lack of signal could result from a sequence difference between the analyzed strains and the oligonucleotides spotted on the arrays. The sequencing of this region on eight isolates indeed revealed the presence of a new genotype that was unknown so far. Therefore, our approach not only permitted to reveal the full genotype at 8 and 5 sites for *S. meliloti* and *S. medicae* strains respectively, but also to detect new variants, for which we plan to design new specific oligonucleotides and spot them on the array together with the others.

4. Conclusions

We have described a simple procedure of DNA microarrays for genotyping analysis, and have compared two different substrates for oligonucleotide tethering. From our results, we can conclude that,

though both surfaces enable single mismatch discrimination, epoxy-coated glass slide arrays are more sensitive than nylon membrane arrays. For this reason, we illustrated the reliability of this procedure by the genotyping analysis of a population of symbiotic bacteria with epoxy-coated glass slides. This simple system proved very sensitive, specific and reproducible, thus avoiding to design several probes for the same polymorphism, or probes on both DNA strands (Hacia, 1999). Over all the protocols described in the literature, this protocol constitutes a low-cost alternative for “medium-scale” projects of population genetics, which is accessible to any laboratory equipped for molecular biology. The use of a non-autofluorescent substrate like epoxy glass slides, will allow to label the targets with fluorescence instead of radioactivity, which will render this procedure even more accessible.

In the near future, we intend to enlarge the number of polymorphic loci represented on the arrays in order to increase both the cost-effectiveness and time-effectiveness of this procedure. Indeed, the ability to simultaneously investigate polymorphism of many regions on the genome will be a clear advantage over sequencing techniques. We will also test other substrates for oligonucleotide spotting to evaluate if it is possible to improve the results regarding mismatch discrimination.

Acknowledgements

This work was supported by the French and Belgian Embassies through the Programme d'Actions intégrées franco-belge Tournesol (contract 03006ZJ), and by the European Commission (RT-program, contract BACDI-VERS QLRT-2001-02097). Xavier Bailly was supported by a grant from the French ministry of research. We thank Mrs. Lucette Maure for her technical assistance, and Dr. Thérèse Atallah (Faculté d'Agronomie, Université Libanaise, Beyrouth, Lebanon) for soil collection. We are also grateful to Dr. Eric Giraud and Dr. Lionel Moulin for their critical reading of this manuscript.

References

- Bailly, X., Olivier, I., de Mita, S., Cleyet-Marel, J.-C., Bena, G., in press. Recombination and selection shape the molecular diversity pattern of nitrogen-fixing *Sinorhizobium* sp. associated to *Medicago*. *Mol. Ecol.*
- Bertucci, F., Bernard, K., Liorod, B., Chang, Y.C., Granjeaud, S., Birnbaum, D., Nguyen, C., Peck, K., Jordan, B.R., 1999. Sensitivity issues in DNA array based expression measurements and performance of nylon microarrays for small samples. *Hum. Mol. Genet.* 8, 1715–1722.
- Bertucci, F., Liorod, B., Nasser, V., Granjeaud, S., Tagen, R., Braud, A.C., Patriez, V., Houlgatte, R., Dancie, B., Nguyen, C., 2003. Gene expression profiling of breast carcinomas using nylon DNA arrays. *C. R. Biol.* 326, 1031–1039.
- Booth, S.A., Drebot, M.A., Martin, J.E., Ng, I.K., 2003. Design of oligonucleotide arrays to detect point mutations: molecular typing of antibiotic resistant strains of *Neisseria gonorrhoeae* and hantavirus infected deer mice. *Mol. Cell Probes* 17, 77–84.
- Bresnahan, K.J., Frank, R., Bonker, H., Markey, L.A., 1986. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3746–3750.
- Brumfield, R., Beerli, T.P., Nakerson, D.A., Edwards, S.V., 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol.* 18, 249–256.
- Brunel, B., Rome, S., Ziani, R., Cleyet-Marel, J.C., 1996. Comparison of nucleotide diversity and symbiotic properties of *Rhizobium meliloti* populations from annual *Medicago* species. *FEMS Microbiol. Ecol.* 19, 71–82.
- Chang, C.T., Hair, T.C., Hutton, J.R., Weimar, J.G., 1974. Effects of microscope and macroscopic viscosity on the rate of renaturation of DNA. *Biopolymers* 13, 1847–1858.
- Chen, W.P., Kuo, T.T., 1993. A simple and rapid method for the preparation of gram negative bacterial genomic DNA. *Nucleic Acids Res.* 21, 2260.
- Coenye, T., dePuma, J.J., 2002. Multilocus restriction typing: a novel tool for studying global epidemiology of *Burkholderia cepacia* complex infection in cystic fibrosis. *J. Infect. Dis.* 185, 1454–1462.
- Drmanac, R., Dumanac, S., Chui, G., Diaz, R., Hou, A., Jin, H., Jin, P., Kwon, S., Lacy, S., Mocur, B., Shatto, J., Swanson, D., Ukrainczyk, T., Xu, C., Little, D., 2002. Sequencing by hybridization (SBH): advantages, achievements, and opportunities. *Adv. Biochem. Eng. Biotechnol.* 77, 75–101.
- Galibert, F., Finat, T.M., Long, S.R., Puhler, A., Aboia, P., Ampe, F., Barioy, Hubler, F., Barnett, M.J., Becker, A., Boistard, P., Botha, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capcia, D., Chain, P., Cowie, A., Davis, R.W., Dreano, S., Federspiel, N.A., Fisher, R.F., Gloux, S., Godre, T., Goffeau, A., Golding, B., Gouzy, J., Guat, M., Hernandez Lucas, J., Hong, A., Huizar, L., Hymann, R.W., Jones, T., Kahn, D., Kahn, M.L., Kaiman, S., Keating, D.H., Kiss, E., Komp, C., Lelaure, V., Massey, D., Palm, C., Peck, M.C., Pohl, T.M., Portetelle, D., Purnelle, B., Ramsperger, L., Surzycki, R., Thibault, P., Vandenbol, M., Vorholter, F.J., Weidner, S., Wells, D.H., Wong, K., Yeh, K.C., Batet, J., 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293, 668–672.
- Crevers, D., Cohen, E.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, E.L., Swings, J., 2005. Opinion: re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3, 733–739.
- Hacia, J.G., 1999. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat. Genet.* 21, 42–47.
- Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P., Collins, F.S., 1996. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two colour fluorescence analysis. *Nat. Genet.* 14, 441–447.
- Held, G.A., Grinstein, G., Tu, Y., 2003. Modeling of DNA microarray data by using physical properties of hybridization. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7575–7580.
- Hetton, J.R., 1977. Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. *Nucleic Acids Res.* 4, 3537–3555.

- Levicky, R., Horgan, A., 2005. Physicochemical perspectives on DNA microarray and biosensor technologies. *Trends Biotechnol.* 23, 143–149.
- Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.A., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., Brown, E.L., 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14, 1675–1680.
- Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3140–3145.
- Morm, P.A., Luikart, G., Wayne, R.K., the SNP workshop group, 2004. SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19, 208–216.
- Nicholas, K.B., Nicholas, H.B.J., Deerfield, D.W., 1997. GeneDoc analysis and visualization of genetic variation. *EMBNEW NEWS* 4, 14.
- Panjikovich, A., Melo, F., 2005. Comparison of different melting temperature calculation methods for short DNA sequences. *Bioinformatics* 21, 711–722.
- Panjikovich, A., Norambuena, T., Melo, F., 2005. *dmMATE*, a consensus melting temperature prediction server for short DNA sequences. *Nucleic Acids Res.* 33, W570–W572.
- Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., Fodor, S.P., 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5022–5026.
- Richmond, C.S., Glasner, J.D., Mau, R., Jin, H., Blattner, F.R., 1999. Genome wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* 27, 3821–3835.
- Rome, S., Fernandez, M.P., Brunel, B., Normand, P., Cleyet-Marel, J.C., 1996. *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. *Int. J. Syst. Bacteriol.* 46, 972–980.
- Rose, K., Masor, J.O., Lath, R., 2002. Hybridization parameters revisited: solutions containing SDS. *Biotechniques* 33, 54–58.
- Schwartz, D.C., Cantor, C.R., 1984. Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 37, 67–75.
- Sekar, M.M., Bloch, W., St John, P.M., 2005. Comparative study of sequence-dependent hybridization kinetics in solution and on microspheres. *Nucleic Acids Res.* 33, 366–375.
- Southern, E., Mir, K., Shepelinov, M., 1999. Molecular interactions on microarrays. *Nat. Genet.* 21, 5–9.
- Sunnucks, P., Wilson, A.C., Beheregaray, L.B., Zenger, K., French, J., Taylor, A.C., 2000. SSOP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. *Mol. Ecol.* 9, 1699–1710.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Urakawa, H., El Fantroussi, S., Smidt, H., Smoot, J.C., Tribou, E.H., Kelly, J.J., Noble, P.A., Stahl, D.A., 2003. Optimization of single base-pair mismatch discrimination in oligonucleotide microarrays. *Appl. Environ. Microbiol.* 69, 2848–2856.
- Vos, P., Hogers, R., Bleeker, M., Reijers, M., van de Lee, T., Hornes, M., Fryters, A., Pot, J., Peleman, J., Kuiper, M., 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414.
- Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, F., Mittmann, M., Morris, M.S., Sher, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M., Lander, E.S., 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077–1082.
- Wetmur, J.G., Davidson, N., 1968. Kinetics of renaturation of DNA. *J. Mol. Biol.* 31, 349–370.
- Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18, 6531–6535.
- Wrobel, G., Schlingemann, J., Hummerich, L., Kramer, H., Lichter, P., Hahn, M., 2003. Optimization of high-density cDNA-microarray protocols by 'design of experiments'. *Nucleic Acids Res.* 31, e67.
- Zwick, M.F., McAfee, F., Cutler, D.J., Read, T.D., Ravel, J., Bowman, G.R., Galloway, D.R., Mateczun, A., 2005. Microarray-based resequencing of multiple *Bacillus anthracis* isolates. *Genome Biol.* 6, R10.

Recombinaison, spécialisation et spéciation chez les symbiotes du genre *Sinorhizobium* associés aux plantes du genre *Medicago* : des patrons de diversité aux hypothèses évolutives

RESUME

Les *rhizobium* sont des bactéries du sol capables d'établir une symbiose fixatrice d'azote avec les Fabaceae. Le couple *Sinorhizobium/Medicago* est un modèle intéressant pour étudier l'évolution de ce mutualisme. Les *Medicago* interagissent avec les « espèces-sœur » *Sinorhizobium meliloti* et *Sinorhizobium medicae*, différents groupes de spécificité ayant été décrits chez les deux partenaires.

Des souches de *S. meliloti* et *S. medicae* ont été isolées à l'aide d'espèces du genre *Medicago* présentant des spécificités symbiotiques variées. Si l'étude de leur diversité génétique a confirmé l'isolement sexuel entre *S. meliloti* et *S. medicae*, les transferts horizontaux de matériel génétique semblent jouer un rôle fondamental dans la diversification de chacun des taxons. Chez *S. meliloti*, ces transferts affecteraient préférentiellement les locus plasmidiques.

L'isolement géographique semble représenter une limite aux flux de gènes chez ces bactéries. En effet, les populations de *S. meliloti* isolées à partir des mêmes plantes hôtes en différents sites sont significativement différenciées.

L'interaction avec des plantes hôtes présentant des spécificités différentes semble également avoir une influence sur la structure génétique des populations de *S. meliloti*. Toutefois, le niveau de différenciation observé dépend du locus considéré. Ainsi, une différenciation génétique maximale est observée pour le marqueur lié aux gènes impliqués dans la spécificité d'hôte (i.e. les gènes *nod*). Au contraire, le niveau de différenciation observé à d'autres locus s'est révélé faible ou non significatif.

Du fait d'événements de recombinaisons, les processus de spécialisation et de spéciation ne semblent donc pas indissociables chez ces bactéries. De plus, différentes hypothèses peuvent expliquer l'évolution de la spécificité symbiotique chez *S. meliloti* ou l'émergence de *S. medicae*. Enfin, plusieurs perspectives de recherche découlant de ce travail sont présentées.

Recombination, specialisation and speciation of *Sinorhizobium* symbiotic bacteria that are associated to the genus *Medicago* : from diversity patterns to evolutionary hypotheses

ABSTRACT

Rhizobia are soil bacteria that are able to establish a nitrogen fixing symbiosis with Fabaceae plants. *Sinorhizobium/Medicago* is an appropriate model to study the evolution of such mutualism. *Medicago* plants interact with the sibling-species *Sinorhizobium meliloti* and *Sinorhizobium medicae*, but several specificity groups have been defined for each partner.

S. meliloti and *S. medicae* strains have been isolated from *Medicago* plants displaying various symbiotic specificity. While the study of their diversity confirmed the sexual isolation between *S. meliloti* and *S. medicae*, horizontal gene transfers seem to have a significant impact on the diversification of each species. Moreover, these transfers affect preferentially plasmidic loci of *S. meliloti*.

Geographic isolation seems to be an efficient barrier to gene flow for these bacteria. Populations of *S. meliloti* sampled from the same host-plants at several locations are indeed significantly differentiated.

The interaction with host plants which feature different specificity have also an impact on the genetic structure of *S. meliloti* populations. However, the intensity of genetic differentiation depends on the loci surveyed. We obtained maximal measures of differentiation for molecular markers linked to genes involved in host specificity (i.e. *nod* genes). Conversely, we observed low or non-significant differentiation measures at other loci.

Due to recombination events, specialization and speciation do not seem to be joined for these bacteria. Furthermore, various hypotheses are discussed in order to explain the evolution of symbiotic specificity within *S. meliloti* or to explain the speciation of *S. medicae*. Finally, several research prospects resulting from this work are presented.

DISCIPLINE : Biologie des populations et écologie

MOTS-CLES : *Sinorhizobium*, *Medicago*, sélection, recombinaison, spécialisation, spéciation

INTITULE ET ADRESSE DES LABORATOIRES :

Laboratoire des Symbioses Tropicales et Méditerranéennes ; UMR 1063 ; Campus de Baillarguet ; TA 10/J ; 34398 MONTPELLIER Cedex 05

Institut des Sciences de l'Evolution – Montpellier ; UMR 5554 ; CC 065 BAT. 22 ; Université de Montpellier II ; Place Eugène Bataillon ; 34095 MONTPELLIER Cedex 05