

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche scientifique
Université des frères Mentouri Constantine

**Faculté des sciences de la Nature et de Vie
Département de Microbiologie**



N° de série : 11/D3L/2015

N° d'ordre : 02/microbio/2015

**Thèse de Doctorat 3^{ème} cycle LMD
En Biotechnologies microbiennes, Génomes et Environnement**

Thème :

**Application des outils biomoléculaires 16S, MALDI-TOF et
métagénomique pour la détermination des variantes
taxonomiques bactériennes telluriques**

Présentée par :
Kaoutar KHELILI

Soutenue le: 31 mai 2015

Devant le jury :

Président :	Mr. BOULAHROUF Abderrahmane	Professeur. Université Constantine 1
Rapporteur :	Mr. HAMIDECHI Mohamed Abdelhafid	Professeur. Université Constantine 1
Examineurs :	Mr. ARHAB Rabah	Professeur. Université Oum El Bouaghi.
	Mr. BOUDEMAGH Allaou Eddine	Maître de Conférences A. Université Constantine 1.
	Mr. DJEBRI Belgacem	Maître de Conférences A. Université Tébessa.

Année Universitaire 2014/2015

Dédicaces

Afin d'être reconnaissante envers ceux qui m'ont appuyé et encouragé à effectuer ce travail de recherche, je dédie cette thèse :

À ma très chère maman « Zahia » et à mon très cher papa « Abdelkader » qui ont fait de moi ce que je suis.

À ma chère sœur « Khaoula » et mes deux chers frères « Loqman » et « Aymene » qui n'ont pas cessé de me combler par leur amour et leur tendresse.

À mon conseiller « cher époux Abdelghani » pour son soutien moral, et pour tous les sentiments d'affection et d'amour.

À mon petit cœur « mon fils Kynane » qui m'a donné un coup de pouce pour compléter ma thèse.

À mes chères amies « Zeinouba », « Amoura », « Hinoucha », « Asouma » et « Ouideda » pour leur gentillesse et leur encouragement.

À tous les membres de ma famille.

Et à tous ceux que ma réussite leur tient à cœur.

Remerciements

J'aimerai, en premier lieu, remercier mon dieu « Allah » qui m'a donné la volonté et le courage pour la réalisation de se travail.

Avant tout, je remercie les membres du jury, M^r Boulahrouf Abderrahmane, M^r Arhab Rabah, M^r BOUDEMAGH Allaou Eddine et M^r Djebri Belgacem pour avoir accepté de juger mon travail et pour la discussion qu'ils ont apportée sur ce projet.

Un projet de thèse ne se réalise jamais seul, c'est pourquoi je souhaiterai ici remercier toutes les personnes qui ont participé à ces cinq années d'aventure scientifique et humaine.

Je souhaiterai tout d'abord remercier M^r Hamidechi Mohamed Abdelhafid, mon directeur de thèse, qui a cru en mes capacités pour relever le défi de ce projet. Je le remercie aussi pour m'avoir transmis son goût pour la recherche et pour m'avoir fait profiter de son expérience dans le domaine.

Au sein des deux laboratoires, de microbiologie et de biochimie, situés au sein de la faculté des sciences de la nature et de la vie de l'université Constantine 1, je souhaiterai remercier particulièrement toutes les techniciennes et tous les techniciens pour leur aide et leur sympathie durant mes travaux pratiques. Ainsi que le staff administratif de notre département.

J'aimerai bien remercier M^{me} Soltani Mounira, chef de service à l'ANRH (Constantine), de m'avoir effectué les analyses physicochimiques de mes échantillons de sol. De plus, je souhaiterai lui dire merci aussi pour sa gentillesse et sa générosité.

Durant mon premier stage pratique, j'ai eu l'occasion de travailler avec M^r Michel Drancourt du Laboratoire de microbiologie, hygiène, épidémiologie hospitalisation à l'hôpital de la Timone (Marseille, France). A son côté j'ai appris à utiliser le MALDI-TOF MS, la PCR et le séquençage de l'ADNr 16S. Je le remercie pour son accueil dans son laboratoire, et pour les discussions enrichissantes que nous avons eu sur le projet. Cela m'a permis d'explorer les données selon de nouvelles perspectives et de développer mon esprit critique sur les méthodes moléculaires que nous utilisons.

Je voudrai dire un grand merci à M^r Emmanuel Lemichez du Centre Méditerranéen de Médecine Moléculaire (Nice, France) de m'avoir appris beaucoup de connaissance dans le domaine de la biologie moléculaire durant mon deuxième stage. Merci à toute l'équipe de recherche (Daniel, Emilie, Patrick et Anne) de m'avoir apporté une vision différente au projet et pour les discussions très intéressantes que nous avons eu. Cela m'a permis d'appréhender des aspects auxquels je n'avais pas forcément pensé.

Je remercie aussi M^{me} Dominique Reichert –Thüiring de la plateforme génomique « The Swiss DNA Company ». Grâce à son travail, nous avons pu obtenir un formidable jeu de données : les librairies et les séquences de gènes d'ARNr 16S. Merci à lui de m'avoir fait partager son expérience dans le domaine de la métagénomique.

Enfin, en dehors du laboratoire, au cours de ces cinq années, j'ai eu la chance de rencontrer d'autres jeunes chercheurs en biologie, qui sont des gens formidables, auprès des quels j'ai non seulement développé mon expérience du management et de la gestion de projet, mais j'ai aussi passé de très bons moments de détente post-réunion. Je tiens à remercier particulièrement : Soumya, Sarah, Sofiane, Mehdi, Narjese, Rita, Mamadou et Fatima.

Table des matières :

Liste des figures	<i>i</i>
Liste des tableaux	<i>iv</i>
Liste des abréviations et des acronymes	<i>v</i>
Introduction générale	1
Synthèse bibliographique	
1. Objet sol : une matrice vivante complexe, hétérogène mais structurée	4
1.1. Fraction minérale	5
1.2. Matière organique	6
1.3. Structure du sol	6
1.4. Eau et l'air	7
2. Le sol : un réservoir de la diversité bactérienne	8
2.1. Facteurs influençant la structure des communautés bactériennes telluriques	9
2.1.1. Facteurs abiotiques	10
2.1.2. Facteurs biotiques	11
3. Caractérisation des bactéries telluriques	12
3.1. Approche phénotypique, le fondement de la microbiologie du sol	12
3.2. Approche moléculaire	13
3.2.1. Ere protéomique et l'identification bactérienne	14
3.2.1.1. Spectrométrie de masse	14
3.2.1.2. MALDI-TOF MS en microbiologie : une histoire récente	15
4. Etude de la diversité bactérienne tellurique:	19
4.1. Diversité bactérienne dans le sol	19
4.2. Paramètres influençant la diversité bactérienne dans le sol	20
4.3. Approche métagénomique	23
4.3.1. Définition	23
4.3.2. Contexte et origine	23
4.3.3. La notion d'espèce : de la microbiologie à la microbiologie de l'environnement	25
4.3.4. Contraintes de la définition d'une OTU en métagénomique	26
4.3.5. Méthodes et applications	27
4.3.6. Défis	30
4.3.6.1. Environnement analysé	30

4.3.6.2. Obtention du métagénome de sol	31
4.3.6.3. Etape de clonage	34
4.3.6.4. Séquençage à haut débit : la nouvelle technologie " pyroséquençage 454 "	35
4.3.6.5. Traitement des données et analyse de la diversité	40
Matériels et méthodes	
Partie 1 : Isolement et caractérisation phénotypique et moléculaire des bactéries telluriques	42
1. Échantillonnage de sol	42
2. Isolement et purification des souches	42
2.1. Réalisation des dilutions	42
2.2. Ensemencement et incubation	42
3. Identification phénotypique des isolats	42
4. Identification moléculaire des isolats	43
4.1. Identification protéomique par MALDI-TOF MS	43
4.1.1. Préparation de la matrice	43
4.1.2. Dépôt des isolats	44
4.1.3. Introduction de la cible MALDI dans le spectromètre de masse Microflex LT	44
4.1.4. Lancement de l'acquisition MALDI-TOF MS et interrogation de la base de données	46
4.2. Identification génotypique (amplification et séquençage de l'ADNr 16S)	48
Partie 2 : Caractérisation de la diversité bactérienne par analyse de l'ADN extrait des sols (approche métagénomique)	50
1. Prélèvement des échantillons de sol	50
2. Analyses physicochimiques des sols	50
3. Approche métagénomique	50
3.1. Extraction des acides nucléiques	50
3.2. Amplification d'ADNr 16S et pyroséquençage 454	50
3.3. Traitement et analyse des données	52
Résultats et discussion	
Partie 1 : Le MALDI-TOF MS fournit un outil utile pour l'identification rapide des bactéries telluriques	53

1. Caractérisation des isolats bactériens	53
1.1. Approche phénotypique	53
1.2. Approche moléculaire	57
1.2.1. Identification protéomique (MALDI-TOF MS)	57
1.2.2. Identification génotypique (amplification et séquençage de l'ADNr 16S)	65
Partie 2 : L'approche métagénomique : accès à la diversité bactérienne tellurique et l'influence des caractéristiques physicochimiques	70
1. Caractérisation des propriétés physicochimiques des trois types de sol	70
1.1. Couleur	70
1.2. Répartition granulométrique	70
1.3. Teneur en calcaire (CaCO ₃)	71
1.4. Teneur en matière organique	71
1.5. pH	72
2. Mesure de la diversité bactérienne tellurique	72
2.1. Assignment des séquences d'ADNr 16S des bactéries	72
2.2. Mesure de la diversité bactérienne	79
2.2.1. Courbe de raréfaction	79
2.2.2. Indices de la diversité bactérienne	81
2.2.2.1. Indice de Shannon-Weaver	81
2.2.2.2. Inverse de l'indice de Simpson	82
2.2.2.3. Indice Chao1	83
3. Effet des caractéristiques physicochimiques sur la diversité bactérienne	83
Conclusion générale et perspectives	85
Références bibliographiques	87
Annexes	

Liste des figures :

Figure 1 : Le positionnement et la complexité externe et interne du sol (Gobat *et al.*, 2003).

Figure 2: Proportion des principaux composants du sol en volume (White, 2006).

Figure 3 : Schéma d'identification par MALDI-TOF à partir des colonies.

Figure 4 : Relation entre le pH des sols et la diversité bactérienne, en utilisant la diversité phylogénétique (A) et le nombre de phylotypes (B) défini à 97% de similarité. La courbe pleine représente la courbe de tendance. Les indices de diversité ont été calculés en utilisant 1200 séquences par sol (Lauber *et al.*, 2009).

Figure 5 : Contribution des séquences d'ARNr 16S bactériens aux phylums les mieux représentés dans le sol. Les données ont été compilées à partir de 21 banques (2920 clones) de clones d'ARNr 16S obtenus par amplification à partir d'ADN du sol. La barre horizontale représente le pourcentage moyen de séquences attribuées au phylum, le rectangle représente l'écart type et les barres verticales représentent les valeurs maximales et minimales (North *et al.*, 2004).

Figure 6 : Représentation schématique de la variabilité le long de l'ADNr.

La variabilité est exprimée par l'entropie de Shannon en fonction des régions hypervariables (en rouge) sur les gènes codant pour l'ARNr 16S (Andersson *et al.*, 2008).

Figure 7 : Schématisation de plusieurs méthodologies courantes dans une analyse métagénomique (figure adaptée d'Ederer, 2011). Le chemin constitué de cadres et de flèches rouges symbolise la méthode métagénomique utilisée dans les différentes analyses présentées dans notre thèse.

Figure 8 : Etapes du pyroséquençage (Ahmadian *et al.*, 2006).

Figure 9 : Représentation du dépôt d'une microbille dans un puits d'une plaque en fibre optique contenant 1,4 millions de puits (Shendure et Ji, 2008).

Figure 10 : Principe du pyroséquençage 454 (figure adaptée de 454 Life sciences, © 1996-2012 Roche Diagnostics Corporation, www.454.com).

Figure 11 : Cible MALDI type Microflex, modèle " MSP 96 " (11 colonnes et 8 lignes).

Figure 12 : Trappe d'insertion en position ouverte sans (a) et avec (b) cible.

Figure 13 : Introduction de la cible Microflex LT.

Figure 14 : Interface du logiciel MALDI BioTyper™ 2.0 (Bruker Daltonics).

Figure 15 : Interface du logiciel d'acquisition FlexControl.

Figure 16 : Code couleur des analyses et identifications BioTyper™ 2.0 (Bruker Daltonics).

Figure 17 : Constitution des amorces fusionnées utilisées pour le pyroséquençage.

Figure 18 : Spectre de masse de l'isolat S1.

Figure 19 : Spectre de masse de l'isolat S2.

Figure 20 : Spectre de masse de l'isolat S3.

Figure 21 : Spectre de masse de l'isolat S4.

Figure 22 : Spectre de masse de l'isolat S5.

Figure 23 : Spectre de masse de l'isolat S6.

Figure 24 : Spectre de masse de l'isolat S7.

Figure 25 : Spectre de masse de l'isolat S8.

Figure 26 : Spectre de masse de l'isolat S9.

Figure 27 : Dendrogramme des neuf isolats bactériens réalisé par le logiciel MALDI BioTyper™ 2.0 (Bruker Daltonics).

Figure 28 : Produits d'amplification PCR du gène d'ARNr 16S des neuf isolats bactériens, sur gel d'agarose à 1,5 %. M) marqueur moléculaire 1 kpb.

Figure 29 : Arbre phylogénétique construit avec la méthode Neighbor-joining montrant les liens phylogénétiques entre 9 séquences basées sur les gènes d'ARNr 16S des isolats bactériens. Les nombres figurant au niveau des nœuds indiquent les taux de Bootstrap (1000 réplicats), la barre 0.02 indique le nombre de substitution par position de nucléotide. Les numéros d'accèsion sont indiqués entre parenthèse.

Figure 30 : Histogramme montrant les pourcentages des séquences identifiées au niveau du *phylum* dans les trois échantillons de sol.

Figure 31: Carte phylogénétique ("heatmap") des *phyla* identifiés dans les trois sols.

Figure 32: Histogramme montrant les pourcentages des séquences identifiées au niveau de famille dans les trois échantillons de sol.

Figure 33: Carte phylogénétique ("heatmap") des familles identifiées dans les trois sols.

Figure 34: Histogramme montrant les pourcentages des séquences identifiées au niveau du genre dans les trois échantillons de sol.

Figure 35 : Carte phylogénétique ("heatmap") des genres identifiés dans les trois sols.

Figure 36: Courbes de raréfaction (nombre d'OTU identifiées par rapport au nombre de séquences analysées) pour les trois échantillons de sol.

Figure 37: Histogramme des nombres des OTUs identifiées dans les trois échantillons de sol.

Figure 38: Histogramme représentant les valeurs de l'indice de diversité " Shannon-Wiener ".

Figure 39: Histogramme représentant les valeurs de " l'inverse de l'indice de diversité de Simpson ".

Figure 40: Histogramme représentant l'indice de la richesse spécifique " Chao1 ".

Liste des tableaux :

Tableau 1 : Score de concordance des spectres obtenus à partir d'une bactérie d'intérêt avec ceux de la base de données Biotyper, reflétant le degré de confiance à accorder à l'identification.

Tableau 2 : Amorces utilisées pour l'amplification et le séquençage du gène de l'ARNr 16S (Adékambi et Drancourt, 2004).

Tableau 3 : Liste des séquences code à barres (Hamady *et al.*, 2008).

Tableau 4 : Aspect macroscopique des colonies bactériennes isolées sur gélose au sang. (Les photos sont réalisées au laboratoire de cultures spéciales à la faculté de médecine de Marseille).

Tableau 5 : Résultats de l'identification présomptive des isolats bactériens.

Tableau 6 : Identification des isolats bactériens par MALDI-TOF MS.

Tableau 7 : Confrontation et correspondance biomoléculaire avec GenBank.

Tableau 8 : La répartition granulométrique des trois échantillons de sol.

Tableau 9 : Norme d'interprétation du taux du calcaire du sol (Baize, 1988).

Tableau 10 : Echelle d'interprétation du pH (Gagnard *et al.*, 1988).

Tableau 11 : Les résultats du pyroséquençage 454 d'ADNr 16S bactérien.

Liste des abréviations et des acronymes :

ACE : Abundance based coverage estimator.

ANRH : Agence Nationale des Ressources Hydrauliques.

bcPCR : Barcode Polymerase Chain Reaction.

BLAST : Basic Local Alignment Search Tool.

BLASTn : BLAST Nucleotide.

BTS : Bruker Bacterial Test Standard.

CCD : Charge-Coupled Device.

DHB : Acide 2,5-dihydroxybenzoïque.

dNTP : déoxyribonucléotides-triphosphate.

EMBL : European Molecular Biology Laboratory.

EMP : Earth Microbiome Project.

emPCR : Emulsion Polymerase Chain Reaction.

GOS : Global Ocean Sampling.

gyrB : gyraseB gene.

HMP : Human Microbiome Project.

HPLC : High-performance liquid chromatography.

MALDI-TOF MS: Matrix Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry.

MID : Multiplex Identifier.

MLST : Multilocus sequence typing.

MSP: Main Spectra Projection.

NCBI : National Center for Biotechnology Information.

NGS : Next Generation Sequencing.

OTU : Operational Taxonomic Unit.

PTP : PicoTiterplate.

PVPP : Polyvinylpolypyrrolidone.

RDP : Ribosomal Database Project.

SARAMIS : Spectral ARchive And Microbial Identification System.

SBS : sequencing by synthesis.

TFA : Trifluoroacetic Acid.

α -CHCA : Acide α -cyano-4-hydroxy-cinnamique.

Introduction générale

L'identification des bactéries est une priorité absolue surtout lorsqu'il s'agit de germes pathogènes ou à intérêt industriel. La description phénotypique permet, certes, de "cadre" le germe, mais est-elle vraiment suffisante ? Tous les microbiologistes diront que non. L'identification au niveau taxonomique de l'espèce a fait couler beaucoup d'encre et soulevé plus d'un débat, à cause des différentes méthodes développées jusque-là par plusieurs microbiologistes.

D'un point de vue écologique, le sol est l'un des plus importants réservoirs de diversité microbienne et surtout l'un des derniers remparts pour cette biodiversité (Swift, 1998) résultant de la complexité des assemblages taxonomiques et fonctionnels des communautés microbiennes et des interactions entre eux (Coleman et Whitman, 2005). Ceci est d'autant plus vrai pour les communautés bactériennes qui sont encore considérées, sur la base d'un postulat proposé par Beijerinck (1913) : "tout est partout" (everything is everywhere), comme des organismes ubiquistes présentant une forte redondance fonctionnelle. A ce titre, les communautés bactériennes telluriques apparaissent encore de façon récurrente comme une "boîte noire" fonctionnelle génératrice de flux dont l'intensité dépend de plusieurs facteurs biotiques et abiotiques.

Malgré que ce compartiment fondamental de la biosphère constitue une mosaïque d'habitats avec pour chacun des populations bactériennes propres, notre niveau de connaissance reste limité. Il en résulte un manque de prise en compte des communautés bactérienne en terme de diversité et de distribution spatiale à petite comme à grande échelle (Ranjard *et al.*, 2010). Ce qui peut être expliqué par le retard méthodologique sur la caractérisation des bactéries du sol *in situ* qui s'explique aussi par les caractéristiques propres du modèle bactérien (i) petite taille des organismes, (ii) accessibilité plus ou moins importante des populations au sein d'une matrice hétérogène et structurée comme le sol, (iii) difficulté à résoudre une information constituée de 100 000 à 1 000 000 d'espèces différentes par gramme de sol (Curtis et Sloan, 2005 ; Maron *et al.*, 2007). Toutefois, depuis une vingtaine d'années, les importantes avancées en biologie moléculaire ont permis de mettre au point des techniques qui permettent d'appréhender et de résoudre la diversité des communautés bactériennes du sol *in situ* et sans *a priori* (Ranjard *et al.*, 2000).

Jusqu'au début des années 2000, au sein des laboratoires de microbiologie du sol, l'identification bactérienne reposait essentiellement sur l'analyse de caractères phénotypiques dont l'interprétation nécessitait une expertise. Pour la plupart des cas, le résultat n'était

obtenu qu'au bout de 18 heures. En cas d'échec, l'identification était acquise par des analyses de biologie moléculaire comme celles des séquences du gène de l'ARNr 16S. Depuis près de dix ans, des automates utilisant la spectrométrie de masse de type MALDI-TOF (Matrix Assisted Laser Desorption/Ionization Time-Of-Flight) (Mellmann et Müthing, 2013) ont révolutionné l'identification bactérienne dans le domaine clinique. À partir de colonies, des résultats d'identification fiables sont acquis en moins de cinq minutes avec une facilité d'utilisation technique. Ces atouts ont largement contribué à son succès. Dès lors qu'un laboratoire s'équipe avec ce type d'automate, l'essentiel des identifications est assuré par cette nouvelle approche. Mais malheureusement, cette technique est très peu exploitée dans le domaine de la microbiologie du sol (Uhlik *et al.*, 2011 ; Koubek *et al.*, 2012) d'où la nécessité d'autres travaux de recherche. Pour cette raison, nous avons choisi de se focaliser, dans la première partie de nos études expérimentales, sur l'applicabilité du MALDI-TOF MS couplé à l'algorithme Biotyper dans ce domaine assez complexe.

Cependant, cette technique est dépendante des conditions de culture. En effet, l'étude des populations par isolement de bactéries sur milieux de culture ne permet la mise en évidence que d'une très faible proportion des espèces bactériennes présentes (Nocker *et al.*, 2007) car seulement 0,1 % à 1 % des bactéries telluriques peuvent être cultivées (Torsvik *et al.*, 1998 ; Kirk *et al.*, 2004). Ces limites imposées par la culture *in vitro* ont permis le développement d'approches s'en affranchissant. L'arrivée de la métagénomique a restreint l'étude des bactéries à leur ADN directement extrait de l'environnement (McHardy et Rigoutsos, 2007). L'ADN alors collecté à partir d'échantillons environnementaux (le métagénome) peut être analysé par des techniques permettant de caractériser la structure d'une communauté bactérienne complexe (Ramette, 2009). Ces méthodes nécessitent souvent une étape d'amplification. Récemment les nouvelles techniques de séquençage d'ADN ont permis d'obtenir des gigabases de données de séquences venant de bactéries non cultivées (Roesch *et al.*, 2007), permettant une variété d'applications génétiques.

Par rapport à l'absence d'un organigramme microbiologique consensus pour l'identification formelle des bactéries, nous proposons, dans cette thèse, d'étudier la diversité bactérienne du sol d'El-Meridj (Constantine, Algérie) en mettant en exergue l'importance des techniques moléculaires actuelles dans l'identification bactérienne ; à savoir :

- Evaluer la capacité du MALDI-TOF MS, par rapport aux gènes d'ARNr 16S en tant que référence, à identifier les différentes variantes taxonomiques bactériennes telluriques en confrontant leurs résultats sous forme de dendrogrammes.

- Mesurer la diversité bactérienne tellurique par la technique de métagénomique et étudier l'effet des caractéristiques physicochimiques du sol sur cette biodiversité.

Synthèse bibliographique

1. Objet sol : une matrice vivante complexe, hétérogène mais structurée

Le sol est considéré comme l'un des environnements les plus complexes de la biosphère et est à ce titre un réservoir majeur de la diversité microbienne. Il est à l'interface entre la lithosphère, l'atmosphère et l'hydrosphère (Alexander, 1977) et sert de support à une partie de la biosphère (Figure 1 ; Gobat, 2003).

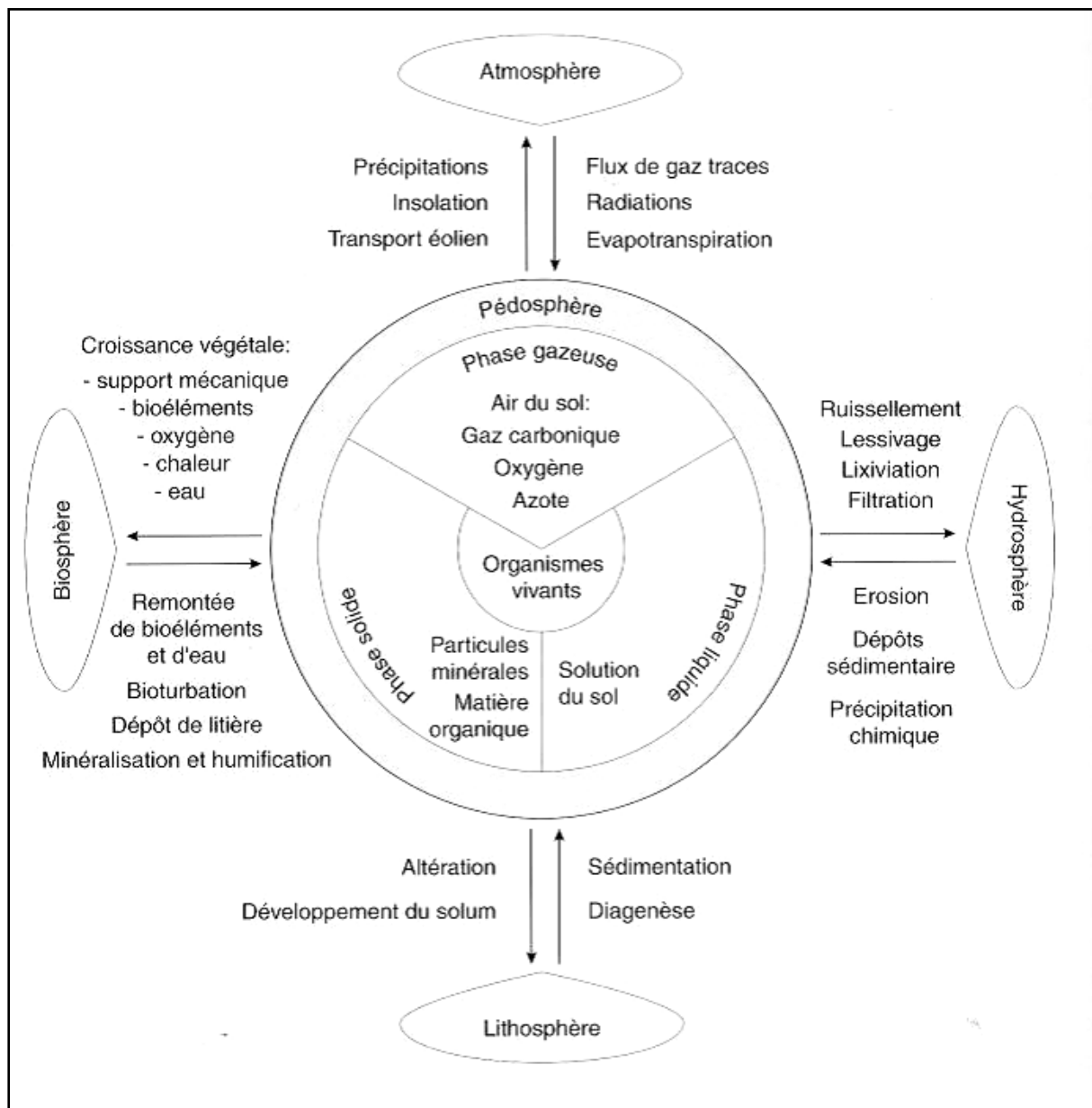


Figure 1 : Le positionnement et la complexité externe et interne du sol (Gobat *et al.*, 2003).

Il est constitué de cinq composants majeurs : fraction minérale, matière organique, eau, air et organismes vivants (Figure 2 ; Alexander, 1977). Les matières organiques et minérales s'organisent de manière à créer des vides alors occupés par l'air et l'eau (les pores).

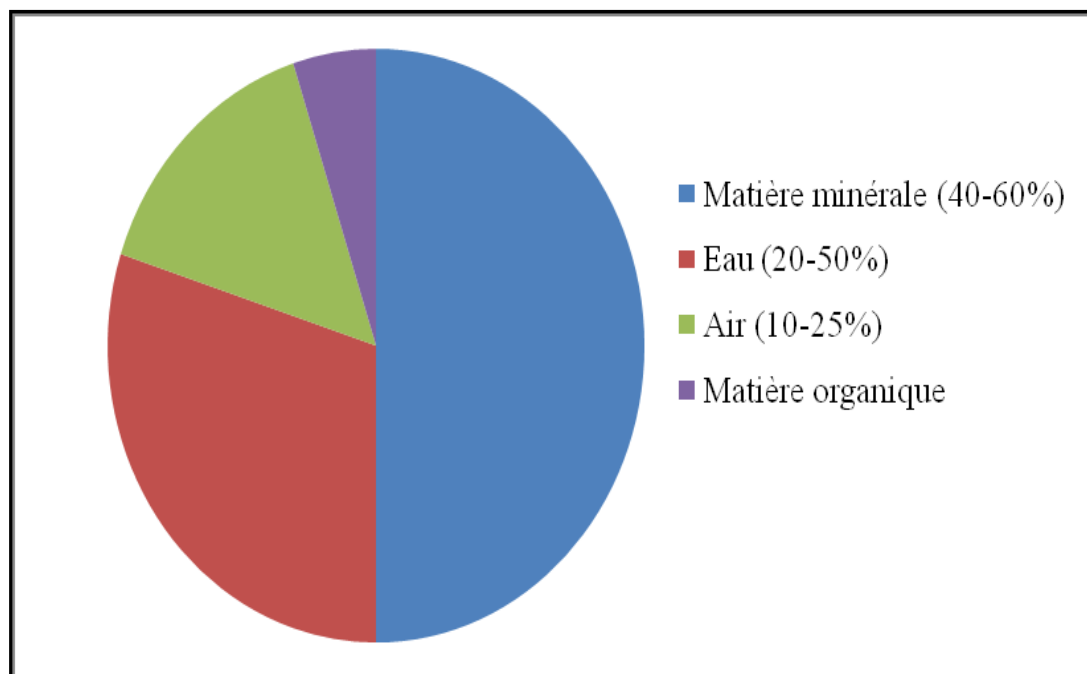


Figure 2: Proportion des principaux composants du sol en volume (White, 2006).

1.1. Fraction minérale

Elle dépend du type de roche mère à partir duquel le sol s'est formé. La fraction minérale des sols résulte de l'action de deux principaux phénomènes (Best, 1995). La désagrégation physique due aux alternances chaud/froid, le gel, le vent ou l'eau fractionnant la roche en morceaux de taille de plus en plus faible, les particules générées conservant la composition minéralogique initiale (Gobat, 2003). L'altération due à l'eau, associée ou non à l'oxygène, au gaz carbonique CO_2 ou à des acides organiques (Gobat, 2003) provoque une transformation des minéraux primaires avec formation de minéraux secondaires (comme l'argile) dont l'ensemble constitue le complexe d'altération (Duchaufour, 1977).

Cette fraction minérale est constituée de particules de différentes tailles, classées généralement selon leur diamètre : la fraction grossière ($>2\text{mm}$) et la terre fine ($<2\text{mm}$) (sables, limons et argiles). La proportion relative des éléments de la terre fine détermine la texture d'un sol conditionnant directement la structure d'un sol. Les argiles sont le résultat de l'altération des roches par hydrolyse des minéraux silicatés (Gobat, 2003). Les minéraux argileux formés sont des phyllosilicates, constitués de feuillets organisés en couches. Certaines argiles sont capables d'expansion suite à l'hydratation des couches interfoliaires (Gobat, 2003). Une autre composante de la réactivité des particules argileuses est leur surface spécifique. Plus leur surface disponible est élevée, plus leur réactivité sera prédominante

(Alexander, 1977). L'adsorption de molécules et ions à la surface des argiles augmente avec leur surface disponible par unité de masse (White, 2006). Par ailleurs les phyllosilicates argileux sont des colloïdes possédant d'autres propriétés qui vont influencer la structure, la porosité ou la capacité d'échange ionique du sol. En effet les argiles sont chargées électro-négativement, sont hydrophiles, mais aussi lipophiles présentant un pouvoir de dispersion et de floculation. Ces propriétés varient suivant les structures minéralogiques et les surfaces développées respectives, régissent la formation d'agrégats, cimentent les particules et favorisent l'habitat microbien (Gobat, 2003).

1.2. Matière organique

Par ses propriétés physicochimiques, la matière organique interagit avec les particules de sol et participe ainsi à sa structuration. La matière organique fraîche provient des déchets des organismes, de leurs sécrétions ou des tissus morts. Mais elle provient majoritairement de débris végétaux (cellulose, hémicellulose, lignine, tanins) (Chenu, 1993). Ces composés organiques humifiés sont séparés en trois fractions suivant leur solubilité : les acides fulviques très solubles à tout pH, les acides humiques insolubles à pH acide, et l'humine insoluble. Les acides fulviques sont électronégatifs et ont la capacité de complexer les cations. Les acides humiques forment des complexes avec les argiles, les minéraux, et ions métalliques ainsi qu'avec les composés organiques hydrophobes (Parent et Velegol, 2004). Des composés lipidiques (acides gras, cires, résines) et azotés (ADN, chitine, muréline, urée, amines, protéines) sont également présents dans le sol. Ces composés provenant des bactéries, plantes ou animaux peuvent se fixer à des composés humifiés (Gobat, 2003) ou à des colloïdes minéraux. Le sol contient également des êtres vivants appartenant à chacun des grands groupes du vivant : *Bacteria*, *Archaea* et *Eucarya*. Nous allons nous intéresser plus particulièrement aux bactéries, même si les autres groupes participent également à la structuration du sol interagissant donc avec les bactéries.

1.3. Structure du sol

Les agrégats du sol sont des assemblages opérationnels hétérogènes de particules organiques et minérales qui se distinguent par la taille : les macros (>250µm) et les micros (<250µm) (Tisdall, 1982). Les macro-agrégats sont formés par des associations temporaires de micro-agrégats, de minéraux et de particules de matière organique principalement par enchevêtrement des hyphes fongiques et des racines de plantes (Oades, 1984). Les micro-

agrégats typiquement formés par des processus microbiens à l'intérieur de macro-agrégats sont largement dépendants de la persistance des agents liants organiques pour la stabilité structurelle (Tisdall, 1982). Des perturbations physiques du sol résultent généralement en une diminution des macro-agrégats, qui deviennent alors des blocs de construction pour le prochain cycle de formation des macro-agrégats (Oades, 1984 ; Six, 1999 ; Six, 2000). Le sol peut donc présenter de nombreux niveaux d'organisation, chaque niveau ayant une stabilité variable. La structure du sol représente une mosaïque de microenvironnements qui diffèrent par leurs propriétés physiques, chimiques et biologiques. Le volume des espaces des sols dépend plus de la taille, de la forme et de la stabilité des agrégats que des particules minérales elles mêmes (Ranjard et Richaume, 2001).

1.4. L'eau et l'air

L'eau circule et est retenue dans le sol par le réseau des pores. Ces pores sont occupés par un réseau aqueux discontinu, sauf quand le sol est saturé, portant des substances inorganiques ou organiques dissoutes et comprennent une phase gazeuse. L'eau gravitationnelle circule librement dans les macro-pores de diamètre supérieur à 10 μ m, constitués entre les agrégats et à l'intérieur des macro-agrégats. Une partie de l'eau est fortement retenue par capillarité à l'intérieur des micropores (<10 μ m) (Robert, 1992) constitués à l'intérieur des micro-agrégats. La disponibilité de l'eau résulte de la combinaison des potentiels matriciel et osmotique. La rétention et potentiellement l'immobilisation de l'eau sont plus importantes dans les microporosités dues à un potentiel matriciel plus élevé que dans les macroporosités, maintenant les microporosités dans un état plus humide que les macroporosités quand le sol est soumis à de fortes pluies. L'eau du sol est très importante dans l'écologie des micro-organismes du sol car la disponibilité des nutriments et l'intégrité des membranes bactériennes nécessitent que la solution du sol soit largement disponible et circulante (Ranjard et Richaume, 2001).

La diffusion des gaz, comme la circulation de l'eau, dépendent de la distribution de l'espace poreux. La diffusion de gaz est faible dans les micropores (Baver, 1961). Le faible renouvellement de l'atmosphère dans ces microenvironnements combiné avec la consommation biologique de l'oxygène, peuvent conduire à un développement rapide et à la persistance de conditions anaérobies. La survie des bactéries dépend donc de leur capacité pour une respiration alternée, remplaçant l'oxygène par un autre accepteur final d'électrons. Il a été montré par exemple que les zones anaérobies à l'intérieur d'un agrégat artificiel sont

préférentiellement colonisées par des populations bactériennes capables d'utiliser le nitrate comme accepteur final d'électrons (Philippot *et al.*, 1996). La proportion relative de liquide et de gaz affectant les phénomènes aérobies et anaérobies, dépend de la saturation des sols et de la circulation à l'intérieur du réseau des pores (Robert, 1992).

2. Le sol : un réservoir de la diversité bactérienne

L'importance de la variabilité spatiale en microbiologie du sol a depuis longtemps été reconnue et la distribution spatiale des microorganismes a été décrite à des échelles allant de quelques millimètres à plusieurs mètres (Grundmann et Debouzie, 2000 ; Nunan *et al.*, 2002). Des différences de populations bactériennes sont associées à des tailles différentes de particules de sol ou des fractions d'agrégats, incluant les populations associées avec les sous-fractions des micro-agrégats. Mummey *et al.* (2004) ont clairement démontré une stratification spatiale suivant la localisation à l'intérieur et à l'extérieur des micro-agrégats. Le fractionnement des micro-agrégats du sol par des irradiations UV permet de comparer les structures de communautés bactériennes au niveau de la micro-échelle (Mummey, 2004). A micro-échelle, les organisations structurales du sol fournissent par conséquent différents habitats dans lesquels les bactéries indigènes sont distribuées de manière hétérogène (Ranjard et Richaume, 2001). Les bactéries sont localisées soit dans les pores du sol qui représentent jusqu'à 50% du volume total du sol, soit enlisées dans l'argile (Foster, 1988 ; Grundmann, 2004). La distribution quantitative hétérogène des bactéries dans le sol résulte en une plus grande abondance des bactéries dans les micropores (Ranjard et Richaume, 2001). Le diamètre maximum des pores les plus fréquemment colonisés par les bactéries est estimé entre 2.5 μm et 9 μm pour respectivement les textures fines et grossières du sol (Hattori, 1988 ; Postma, 1989). 80% des bactéries sont préférentiellement localisées dans la partie interne des micro et macro porosités, quel que soit le type de sol (Hattori, 1988 ; Ranjard, 1997). Les bactéries localisées dans les macropores sont sujettes à des alternances plus intenses entre sécheresses et inondations. Par conséquent, leur survie dépend de leurs capacités à s'adapter aux fluctuations des conditions d'humidité. La dessiccation d'un sol conduit à une nette diminution du nombre total de bactéries localisées dans la partie extérieure par rapport à celle intérieure (Hattori, 1988). La microporosité à l'intérieur de micro-agrégats stables est préférentiellement colonisée, ce qui suggère une combinaison de conditions favorables pour la vie bactérienne dans ces environnements. Les micro-agrégats sont les habitats les plus favorables pour les bactéries dans la plupart des types de sol. En effet, les régions intérieures des micro-agrégats peuvent constituer un habitat relativement

stable et isolé pour les microorganismes, caractérisé par une pression de prédateurs faible, une disponibilité d'eau relativement stable et de forts gradients de diffusion résultant en une faible disponibilité de nutriments et d'oxygène (Hattori, 1988 ; Ranjard, 1997). Ces micro-agrégats protègent donc les bactéries qui les colonisent de la dessiccation et d'autres substances solubles toxiques exogènes. Il est aussi probable que les microorganismes habitant l'intérieur des micro-agrégats contribuent peu à l'ensemble des activités microbiennes mesurables. Ces microorganismes peuvent cependant jouer un rôle clé dans le fonctionnement de l'écosystème dans des environnements anoxiques et en aidant à la formation et à la stabilisation des agrégats (Mummey, 2004). Il est également probable que, à cause de leur isolement, les organismes occupant ces microhabitats soient beaucoup moins disponibles pour leur isolement *in vitro*, leur extraction ou celle de leur ADN que les microorganismes colonisant la surface des agrégats (Hattori, 1988). Mummey *et al.* ont montré que ces cellules peuvent être classées comme des ultra microbactéries (un volume de $0.1\mu\text{m}^3$), qui utilisent préférentiellement des sucres comme substrats de croissance. Une petite taille et un métabolisme simple peuvent être un avantage pour la vie à l'intérieur des microagrégats. Hassink *et al.* ont montré une corrélation positive entre la taille bactérienne et le volume des pores (Hassink, 1993).

La localisation des bactéries à l'intérieur de la matrice sol est un facteur clé affectant leur survie, leur activité et leurs fonctions écologiques. Déterminer la structure de la communauté bactérienne dans l'espace qui concerne la structure du sol est d'abord fondamental pour élucider les fonctions écologiques. Définir les populations associées avec ces niches est important pour comprendre comment des perturbations influencent la diversité microbienne des sols, leur niveau d'activité et les fonctions susceptibles d'être réalisées.

2.1. Facteurs influençant la structure des communautés bactériennes telluriques

Beaucoup de paramètres biotiques et abiotiques interviennent pour modéliser la structure des communautés bactériennes des sols tant en termes de diversité que d'effectifs (Hassink, 1993 ; Borneman *et al.*, 1996 ; Borneman et Triplett, 1997). La dynamique bactérienne est aussi très importante, l'écosystème sol étant soumis en permanence à des perturbations de tous ordres qui vont avoir de profondes répercussions sur la structure des communautés. Il n'est pas encore possible aujourd'hui de prendre en compte tous les paramètres influant sur la composition de la microflore tellurique et surtout leurs interactions qui confèrent une myriade

de conditions différentes (Tarlera *et al.*, 2008). Un certain nombre d'études ont toutefois permis d'identifier les facteurs présentant une influence marquée sur la structure des communautés bactériennes.

2.1.1. Facteurs abiotiques

Les structures des communautés bactériennes sont variables et dépendent de nombreux facteurs (Noguez, 2005). En effet, nous savons que la structure des communautés bactériennes ainsi que sa composition peuvent varier considérablement suivant l'espace (Martiny *et al.*, 2006 ; Ramette et Tiedje, 2007). Plusieurs études suggèrent même que les caractéristiques des sols (Girvan *et al.*, 2003 ; Singh, 2006), qui conditionnent la distribution spatiale (Nunan *et al.*, 2005 ; Ritz, 2004), sont les facteurs les plus importants dans le modelage de la structure des communautés bactériennes. Par exemple le pH influence fortement la composition des *Acidobacteria* et des *Actinobacteria*. Le pH est l'un des forts prédicateurs de la composition des communautés bactériennes et de la diversité, mais suivant le type de sol d'autres facteurs peuvent agir sur la structure des communautés comme la disponibilité des nutriments, la solubilité des métaux (Muller *et al.*, 2002), le contenu en carbone (Asuming-Brempong, 2008) et en azote (Fierer *et al.*, 2007), l'humidité des sols, la salinité (Rajendhran et Gunasekaran, 2008), les variations climatiques (Lauber, 2009).

Au cours du temps, le sol subit des changements, qui augmentent la complexité globale de l'environnement sol (Nunan *et al.*, 2001 ; Zhou *et al.*, 2002). Tarlera *et al.* (2008) ont montré que le développement de l'écosystème sol favorisait le développement de communautés bactériennes distinctes. Les changements d'utilisation des terres, particulièrement la conversion des forêts en pâturage ou en champs cultivés, est un événement fréquent et un facteur affectant la biodiversité et le fonctionnement des écosystèmes terrestres (Sala *et al.*, 2000). Les effets de l'utilisation des terres sur l'altération des propriétés physiques et chimiques des sols ont été beaucoup étudiés. En effet, des changements d'utilisation des terres (au niveau de la composition des espèces de plantes et des pratiques de gestion utilisées) peuvent avoir des impacts significatifs et de long terme sur le carbone du sol, le contenu en nutriments, la texture du sol et le pH (Murty, 2002). La variabilité des facteurs édaphiques à travers différentes utilisations des terres peut avoir un effet significatif sur la structure des communautés bactériennes (Lauber *et al.*, 2008).

Les communautés sont dominées par 5 groupes majeurs *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, *Bacteroidetes* et *Firmicutes* (Janssen, 2006). L'effet d'une perturbation sur

les communautés bactériennes dépend de sa durée et de sa spécificité. Après une perturbation transitoire, le système peut retrouver son état, tandis qu'une perturbation permanente résulte en un nouvel état altéré (Rykiel, 1985). Les perturbations avec un mode spécifique d'action altèrent seulement quelques groupes d'organismes (exemple : le groupe des *Proteobacteria* réagit de manière sensible au changement des niveaux de carbone dans le sol (Asuming-Brempong, 2008)) tandis que celles qui agissent non spécifiquement affectent un grand nombre de bactéries (Muller *et al.*, 2002).

2.1.2. Facteurs biotiques

Pour comprendre les changements de structure des communautés bactériennes il faut prendre en compte différents paramètres incluant le groupe taxonomique en question (Lauber *et al.*, 2008), le niveau de résolution phylogénétique avec lequel sont étudiées les communautés (exemple : *Pseudomonas*) (Cho et Tiedje, 2000). Il a été montré que les plantes peuvent avoir un impact identique ou même supérieur à celui des facteurs abiotiques, tel que le climat, dans le contrôle des fonctions des écosystèmes terrestres mais aussi au niveau de la structure des communautés bactériennes du sol (Verville *et al.*, 1998). La présence de plantes résulte en une augmentation significative de l'humidité du sol, de la biomasse et de la respiration microbienne (Singh *et al.*, 2009). La présence de plantes joue sur la croissance des communautés bactériennes des sols et influe aussi sur les propriétés abiotiques du sol. La présence de plantes provoque des changements profonds au sein de la communauté bactérienne du sol. Il a été observé que beaucoup de plantes sélectionnent des groupes de micro-organismes spécifiques via l'exsudation de composés dans la rhizosphère, qui représente une association positive où les plantes fournissent le carbone pour la croissance microbienne et les microorganismes en retour fournissent des éléments majeurs tels que l'azote et le phosphore, ainsi qu'une protection contre les attaques des pathogènes et parasites (Singh *et al.*, 2004).

Des changements significatifs des structures des communautés bactériennes dans le sol ont été montrés lors de changements du couvert végétal. Ces changements induits par la végétation ont été observés aussi bien au niveau des groupes universels bactériens (Grayston *et al.*, 1998 ; Nusslein et Tiedje, 1999 ; Tiedje *et al.*, 1999) qu'au niveau des groupes fonctionnels (Singh *et al.*, 2007).

Bien que les communautés bactériennes et fongiques aient été largement étudiées indépendamment l'une de l'autre dans de nombreux habitats, quelques études ont examiné

simultanément ces deux groupes de microorganismes (Costa *et al.*, 2006 ; Singh *et al.*, 2007), sans toutefois tenter de comprendre les interactions entre ces deux communautés qui demeurent très difficiles à prendre en compte, bactéries et champignons vivant en étroite relation dans le sol. Les interactions champignons bactéries sont très certainement de tous types, positive négative ou neutre. Singh *et al.* (2009) ont observé une forte corrélation entre les communautés bactériennes et fongiques indépendamment des facteurs abiotiques du sol (humidité, C et N) à l'échelle du champ (Singh *et al.*, 2009). Des résultats précédents avaient déjà montré qu'un champignon *Glomus mossae* (Murty *et al.*, 2002) avait un impact direct sur la structure de la communauté bactérienne au niveau des racines de plantes (Artursson *et al.*, 2005). Singh *et al.* (2008) avaient montré aussi que la communauté fongique influence la structure de la communauté bactérienne au niveau du rhizoplan des graminées.

3. Caractérisation des bactéries telluriques

3.1. Approche phénotypique, le fondement de la microbiologie du sol

Les méthodes phénotypiques comprennent toutes les techniques ne faisant pas appel aux acides nucléiques et reposants sur la détermination de caractéristiques morphologiques, physiologiques et biochimiques des bactéries via des techniques standardisées. Les tests sont basés soit sur les critères classiques ou traditionnels utilisés dans les schémas d'identification pratiqués dans la plupart des laboratoires de microbiologie, soit sur des tests spécialisés réservés à certains laboratoires (Denis *et al.*, 2007). L'aspect des colonies observées sur la boîte de culture permet dans certains cas d'orienter l'identification vers un groupe de bactéries. La taille, la forme, la couleur (opaque, translucide, pigmentée...), l'état de la surface (lisse ou rugueuse), l'odeur éventuelle ainsi que la présence et le type d'hémolyse sur gélose au sang constituent des critères orientatifs d'identification. L'examen microscopique fournit les renseignements concernant la morphologie des bactéries (cocci, bacilles, spiralées, à bords parallèles ou non, à extrémités arrondies ou effilées ...) ainsi que leur taille. Un examen à l'état frais entre lame et lamelle indiquera la présence et le type de mobilité. La coloration de Gram ou de Ziehl-Neelsen permettra de les classer en bactéries à Gram positif ou à Gram négatif ou acido alcoolrésistantes ainsi que d'observer des arrangements particuliers (diplocoques, palissades, chaînettes, cordes...), la présence ou non de flagelles, et la localisation d'une éventuelle endospore.

3.2. Approche moléculaire

Les techniques de biologie moléculaire ont bouleversé l'identification des bactéries, et ont mis en lumière les insuffisances et les erreurs d'identifications phénotypiques qui étaient jusque là seules disponibles (Ronsin, 2005 ; Denis *et al.*, 2007).

L'identification moléculaire des bactéries repose essentiellement sur l'analyse de la séquence du gène ARNr 16S. La séquence obtenue au laboratoire est comparée via le réseau internet, à l'aide de logiciel spécialisé avec les banques électroniques de séquences comme " Nucleotide " du portail NCBI (appelée GenBank), ou EMBL en Europe.

L'identification bactérienne va reposer sur le pourcentage de similarité des séquences. Nous admettons généralement, qu'une similarité de 99% est identifiante au niveau de l'espèce. Une similarité de 97% identifiante au niveau du genre et qu'en dessous de 97%, il y a une possibilité de nouvelle espèce non décrite dans les banques de données. Par ailleurs, certains logiciels permettent de positionner la séquence obtenue au laboratoire parmi l'ensemble des séquences ARNr 16S disponibles sous forme d'arbres phylogénétiques (Sanger *et al.*, 1977 ; Thomas, 1980 ; Smith *et al.*, 1986 ; Swerdlow *et al.*, 1991 ; Zhang *et al.*, 2003). Cette possibilité est particulièrement intéressante lorsque nous déterminons au laboratoire une séquence ARNr 16S de similarité inférieure à 97% ; cela permet alors d'avoir une idée du groupe bactérien dans lequel se situe la bactérie nouvellement identifiée (Staden, 1979).

L'identification moléculaire des bactéries est utile dans plusieurs cas de figure : elle est utile pour les bactéries fastidieuses de croissance lente, parmi lesquelles se trouvent évidemment les bactéries intracellulaires, qui sont souvent des bactéries d'identification difficile exprimant peu de caractères phénotypiques, mais également le cas de certains genres bactériens cultivant en milieu axénique comme les mycobactéries. Elle est également utile lorsque l'identification phénotypique n'est pas performante (cas des *Acinetobacter*, exprimant très peu de caractères phénotypiques) ou dans le cas où il existe une discordance entre cette identification phénotypique et la sensibilité aux antibiotiques par exemple. Elle est enfin utile pour toutes les souches bactériennes qui sont isolées dans une circonstance inhabituelle pour laquelle il convient absolument de confirmer l'identification.

3.2.1. Ere protéomique et l'identification bactérienne

3.2.1.1. Spectrométrie de masse

a. Historique : Apparue au début du 20^{ème} siècle, la spectrométrie de masse a apporté des éléments clés dans la compréhension du noyau atomique. Elle est issue de la découverte par Euge Goldstein en 1886, des rayons canaux dus aux ions positifs et de leur analyse par un champ magnétique, par Wilhelm Wien en 1898.

En 1912, Joseph John Thomson obtient des spectres de masses de plusieurs composés gazeux : N₂, O₂, CO, CO₂, etc. Il met en évidence les ions négatifs et les ions multichargés. L'année suivante, il découvre les isotopes A (nombre de masse) égale à 20 et A égale à 22 du néon (Hoffmann *et al.*, 1994). Les innovations qui suivent tentent principalement d'augmenter la sensibilité du dispositif.

En 1918, Dempster construit un spectromètre à focalisation en direction. L'année suivante, Francis William Aston introduit le tri des ions en fonction de leur vitesse, ce qui lui permet de déterminer les abondances isotopiques du néon et de mettre en évidence les écarts entre les masses atomiques réelles et les nombres entiers.

En 1932, Bainbridge, en associant au dispositif un filtre de vitesse de Wien, vérifie expérimentalement l'équivalence entre masse et énergie. Il propose alors la double focalisation en direction et en vitesse, principe qui sera mis en œuvre par Jordan, Mattauch et Herzog.

En 1936, Penning propose de combiner un champ électrique et un champ magnétique pour piéger les ions dans un très faible volume : c'est le piège de Penning.

En 1940, Nier perfectionne la source d'ions à impact électronique. Ensuite, en 1948, Cameron invente le spectromètre de masse à temps de vol. Puis, entre 1953 et 1960, Paul et Steinwedel mettent au point l'analyseur quadripolaire, puis le piège ionique quadripolaire qui diffère de celui de Penning par l'absence de champ magnétique. Le piège quadripolaire aura un tel retentissement en physique fondamentale que le prix Nobel sera attribué en 1989 à Paul et à Dehmelt.

Pendant les années 1960, Smith met au point le spectromètre de masse à radiofréquence.

Pour ce qui est des applications, depuis 1950, les techniques d'ionisation se diversifient de façon à s'adapter à la diversité des échantillons à analyser, particulièrement pour la chimie organique. Cette technique d'analyse bénéficie alors du couplage avec un chromatographe en phase gazeuse puis liquide. Les progrès de l'informatique ont permis l'automatisation des réglages de l'appareil ainsi que la mise en œuvre des analyseurs à quadripôle et des pièges ioniques dans de nombreuses applications de la spectrométrie de masse (Hoffmann *et al.*, 1994 ; Constantin, 1996).

b. Principe de fonctionnement : La spectrométrie de masse est une technique d'analyse de la matière en fonction de la masse de ses constituants : particules subatomiques, atomes, molécules, macromolécules biologiques ou non, agrégats, etc. Elle offre trois fonctions principales : la résolution des constituants atomiques ou moléculaires, la mesure de leur abondance relative et la mesure précise de leurs masses atomiques ou moléculaires. Contrairement à ce qui se passe à l'échelle macroscopique, où la gravitation sert à déterminer la masse des objets, à l'échelle atomique, c'est la masse inertielle qui est déterminée à travers un processus dynamique (Andrade *et al.*, 2008). La matière à analyser est injectée, à l'aide d'un système d'introduction, dans le vide poussé de l'appareil où la source d'ions la transforme en particules électriquement chargées "les ions". Ces derniers sont analysés en masse, à l'aide d'un système analyseur qui met en jeu des combinaisons de champs électromagnétiques. Les ions transmis par le dispositif sont détectés, comptés, voire identifiés, à l'aide d'un détecteur. Un ordinateur pilote l'appareil, effectue l'acquisition des données réduites sous forme de spectres de masse qu'il permet d'analyser. Les performances d'un spectromètre de masse se caractérisent par sa limite en masse, la masse la plus élevée qu'il est capable d'analyser, son pouvoir de résolution, son aptitude à séparer des ions de masses voisines et sa sensibilité. La plus petite quantité de matière qu'il peut détecter (< 10-15 g) (De Lara *et al.*, 2008). Lorsqu'il s'agit de mesurer la masse atomique ou moléculaire, il est caractérisé par sa justesse.

3.2.1.2. MALDI-TOF MS en microbiologie : une histoire récente

Depuis sa mise au point au début des années 1990, la spectrométrie de masse de type MALDI-TOF (Matrix Assisted Laser Desorption/Ionization Time-Of-Flight) s'est développée dans différents domaines de la recherche fondamentale (Hemmersbach, 2008 ; Toyo'oka, 2008). Grâce à une ionisation douce, il est devenu possible de détecter de larges fragments et/ou complexes protéiques, ouvrant ainsi de nouvelles perspectives d'étude impossible jusqu'alors (Karas et Hillenkamp, 1988). Rapidement, les microbiologistes se

sont intéressés à cette technique et les premières identifications bactériennes à partir de cellules intactes, sans utilisation de protocoles complexes de préparation des échantillons, ont été obtenus avec succès (Claydon *et al.*, 1996 ; Krishnamurthy et Ross, 1996). Depuis ces premiers travaux, l'ensemble des publications concernant l'identification des bactéries, permet d'affirmer que les empreintes spectrales des principales espèces cultivables identifiées en routine ont été obtenues. Seng *et al.* ont récemment publié une revue exhaustive sur l'ensemble des bactéries identifiables par cette approche (Seng *et al.*, 2010).

L'identification par cette technique repose sur le fait que le spectre obtenu varie d'un genre à l'autre, d'une espèce à l'autre voire même d'une sous-espèce à l'autre, permettant ainsi de discriminer les bactéries entre elles. La réalisation des banques de données doit donc sélectionner parmi l'ensemble des pics, ceux qui sont spécifiques de la souche de référence étudiée. En effet, parmi les peptides détectés, il a été démontré que certains pics sont spécifiques du genre et de l'espèce et qu'ils sont fréquemment retrouvés à condition que les principaux critères culturels soient maîtrisés et comparables d'une acquisition à l'autre. Malgré l'efficacité de cette technique, le développement et la généralisation de cette technique au sein des laboratoires n'ont pas connu un succès immédiat.

En effet, identifier un nombre restreint d'espèces et si possible le plus éloignées taxonomiquement parlant comme c'est le cas dans la plupart des premières études réalisées est une chose, une utilisation en routine nécessitant des identifications d'espèce souvent proches les unes des autres en est une autre.

Le développement des banques de données permettant l'identification en routine repose donc sur les différents points évoqués. Plusieurs stratégies existent actuellement, mais elles ont toutes pour objectifs de sélectionner les pics spécifiques d'une espèce donnée. Emonet *et al.* ont récemment publié une revue détaillant les principes des trois principales banques de données (MALDI Biotyper Bruker, SARAMIS Anagnostec BioMérieux, ANDROMAS) permettant l'identification par MALDI-TOF-MS (Emonet *et al.*, 2010).

L'identification de la bactérie étudiée repose sur la comparaison du spectre obtenu à ceux de références contenus dans la base de données. Comme pour les analyses des séquences nucléotidiques par BLAST ou FASTA, la plus forte concordance (« matches ») est retenue et les résultats sont rendus avec un coefficient de similarité.

Les progrès réalisés dans le domaine de la génomique avec l'augmentation du nombre de génomes séquencés disponibles ont permis d'identifier certains pics au sein des empreintes

spectrales. Il est maintenant établi que la grande majorité des pics détectés entre 2 et 20 kDa correspond aux protéines ribosomales et à des protéines des gènes de ménage (« house keeping » gene) ce qui explique en partie le caractère spécifique et constant des spectres par espèce malgré les variations rencontrées lors de différentes acquisitions. Ces données laissent envisager le développement de stratégies permettant, au même titre que le MLST, de typer les bactéries et de réaliser des comparaisons phylogénétiques. Parmi les autres protéines identifiées, on retrouve des protéines de choc thermique (cold shock proteins), des DNA binding proteins ou encore des RNA chaperones (Ryzhov et Fenselau, 2001 ; Jones *et al.*, 2003 ; Dieckmann *et al.*, 2008).

Afin d'obtenir des identifications fiables, il est important de maîtriser les différents paramètres expérimentaux qui influencent la qualité des spectres. Les milieux de culture, la concentration en NaCl, la nature des tampons utilisés influencent la qualité des spectres. Une fois la majorité de ces paramètres standardisés, la reproductibilité et la précision des identifications sont parfaitement fiables au sein d'un laboratoire et d'un laboratoire à l'autre (Valentine *et al.*, 2005 ; Wunschel *et al.*, 2005). Ces données soulignent l'importance lors de la réalisation des bases de données, d'avoir à disposition un large panel d'espèces différentes représentatif des principales espèces rencontrées en routine avec si possible le plus grand nombre de souches d'une même espèce. Il est ensuite nécessaire de faire varier pour chaque souche les conditions expérimentales afin de prendre en compte les variations inhérentes aux conditions expérimentales.

Le schéma général de l'identification par MALDI-TOF MS est illustré dans la figure 2 et un protocole détaillé des différentes étapes peut être retrouvé dans la publication de Freiwald et Sauer (Freiwald et Sauer, 2009). Les bactéries à identifier sont obtenus à partir des colonies poussant sur des milieux de culture solides ou à partir de cultures en milieu liquide. Cette approche simplifie au maximum la préparation de l'échantillon à analyser, on parle alors d'acquisition sur bactéries intactes. Cette méthode de préparation de l'échantillon est simple et très rapide pouvant être réalisé par du personnel non spécialisé en spectrométrie de masse. Les bactéries sont déposées directement sur une plaque métallique support (figure 3). Aucune étape complexe de purification de l'échantillon n'est nécessaire.



Figure 3 : Schéma d'identification par MALDI-TOF à partir des colonies.

L'obtention d'un spectre de qualité dépend essentiellement de la qualité du dépôt effectué. Peu de matériel est nécessaire pour obtenir une empreinte spectrale de qualité et bien souvent les difficultés d'identification rencontrées sont liées à de mauvais dépôts avec excès de matériel déposé. Une fois l'échantillon séché sur la plaque cible, il est recouvert par la matrice dont la nature est adaptée au type d'analyse que l'on souhaite réaliser. Le rôle de la matrice est d'absorber l'énergie provenant du laser ce qui provoque la vaporisation de l'échantillon avec formation d'ions. Ces ions de masse et de charges différentes sont soumis à un champ électrique. Ils « volent » ensuite jusqu'à un détecteur situé à l'extrémité du tube de vol et la distance parcourue en un temps donné (Time-Of-Flight) est fonction du rapport de leur masse sur la charge (m/z). Les informations essentielles utilisées pour l'identification sont donc contenues dans une liste de pics contenant les rapports m/z et les intensités relatives de chaque pic, l'ensemble de ces données caractérisant l'empreinte spectrale de l'échantillon. Les matrices les plus courantes pour ce type d'analyse sont l'acide 2,5-dihydroxybenzoïque (acide gentisique ou DHB), l'acide trans-3,5-diméthoxy-4-hydroxycinnamique (acide sinapinique ou SA) et l'acide α -cyano-4-hydroxy-cinnamique (α -CHCA). Elles permettent entre autre l'analyse de substrats peptidiques, glycopeptidiques et lipidiques. Actuellement, le dépôt en « goutte épaisse » utilisant l' α -CHCA sur l'échantillon sec est largement utilisé pour les identifications de routine. En effet, la cristallisation obtenue est homogène contrairement à la DHB, permettant une utilisation automatique des acquisitions par le spectromètre avec une excellente qualité de spectres.

D'autres techniques de dépôt existent : les préparations en « couches minces » pour lesquelles l'échantillon est déposé sur une couche de matrice préalablement déposée sur la plaque cible pour former de larges films minces polycristallins, les préparations en "sandwich" où l'échantillon est déposé sur un film de matrice avant d'être lui-même

recouvert par une dernière couche de matrice. Après évaporation des solvants, la matrice co-cristallise avec l'échantillon et l'ensemble peut être analysé.

Des protocoles plus complexes ont été proposés afin d'augmenter la qualité des empreintes spectrales mais dans la grande majorité des cas, ils ne sont pas nécessaires pour une utilisation de routine (Liu *et al.*, 2007). En revanche, pour certains micro-organismes comme les levures ou les mycobactéries, certains auteurs préconisent de réaliser auparavant une étape d'extraction (Amiri-Eliasi et Fenselau, 2001).

4. Etude de la diversité bactérienne tellurique

Le sol est un environnement hétérogène complexe qui contient une grande variété de communautés et d'espèces bactériennes (Daniel, 2005).

De ce fait, accéder à la diversité bactérienne de manière exhaustive constitue un des défis majeurs de ces dernières décennies en microbiologie du sol et les techniques qui y sont dédiées n'ont cessé d'évoluer (Fierer et Lennon, 2011).

Avant le développement des techniques moléculaires, l'identification et la classification taxonomique de bactéries nécessitaient leur isolement, leur mise en culture et leur caractérisation sur la base de critères morphologiques, physiologiques, métaboliques, biochimiques et écologiques. Ces critères de classification ont rapidement trouvé leurs limites du fait de la diversité phénotypique et physiologique des bactéries mais aussi de par la simplicité de la classification engendrée.

Palliant ce problème, le développement d'approches moléculaires a révolutionné le monde de la microbiologie du sol en permettant d'affiner la classification bactérienne et en augmentant de manière significative le nombre d'espèces considérées, par rapport aux méthodes basées sur les seules bactéries cultivables.

4.1. Diversité bactérienne dans le sol

Les bactéries sont de loin les plus abondants microorganismes du sol à la fois en termes de biomasse et de diversité taxonomique (Buckley et Schmidt, 2002). En 1990, par observation microscopique en utilisant l'acridine orange (fluorochrome qui se fixe sur les acides nucléiques), Torsvik *et al.*, (1990) ont estimé qu'un gramme de sol forestier naturel contient près de $1,5 \cdot 10^{10}$ bactéries. Plus tard, Torsvik *et al.*, (1998) ont extrapolé des données

issues d'hybridation ADN-ADN et estimé qu'un gramme de sol forestier se compose de près de 6000 génomes bactériens. Dans cette même étude, les auteurs soulignent la limite des approches culturales qui, dans ce même gramme de sol révèlent la présence de 35 génomes bactériens (chiffre établi à partir de la taille du génome d'*E. coli* : $4,1.10^6$ pb).

Plus récemment, l'étude de métagénomés de sols contrastés en termes de caractéristiques physicochimiques et géographiques a permis d'estimer entre 2000 et 10000 le nombre d'espèces bactériennes (et donc de génomes) par gramme de sol (Roesch *et al.*, 2007).

D'une manière générale, nous considérons aujourd'hui qu'un gramme de sol héberge, en fonction de ses caractéristiques physicochimiques, plusieurs milliers d'espèces et que l'abondance de ces espèces peut varier de 10^8 à 10^{11} cellules par gramme de sol (Curtis *et al.*, 2002, Roesch *et al.*, 2007).

4.2. Paramètres influençant la diversité bactérienne dans le sol

La diversité bactérienne est exprimée par le nombre d'espèces différentes ainsi que par leur abondance relative dans la microflore du sol (Kennedy et Smith, 1995).

La diversité des espèces dans un écosystème local est considérée comme la diversité alpha. La variation de la diversité alpha des écosystèmes dans un même environnement est considérée comme la diversité bêta, et, lorsqu'elle est mesurable, la diversité gamma représente la richesse en espèces à l'échelle régionale et mondiale. La diversité gamma est sensible principalement à des phénomènes qui ont un impact environnemental à l'échelle mondiale (par exemple des changements majeurs sur le climat), par opposition aux impacts à l'échelle locale. La diversité bactérienne peut donc s'observer à différentes échelles, révélant d'une part la complexité des communautés bactériennes dans le sol mais aussi la conservation de certaines caractéristiques.

Le pH apparaît ainsi comme un paramètre qui influence de manière majeure la diversité des communautés bactériennes dans les sols. L'analyse de 88 échantillons de sol a permis de montrer une corrélation entre la diversité des bactéries et le pH de ces sols (Figure 4) (Lauber *et al.*, 2009). L'analyse de la structure des communautés bactériennes montre que l'abondance relative de certains phylums bactériens comme par exemple les *Actinobacteria*, les *Bacteroidetes* et les *Acidobacteria*, est corrélée au pH.

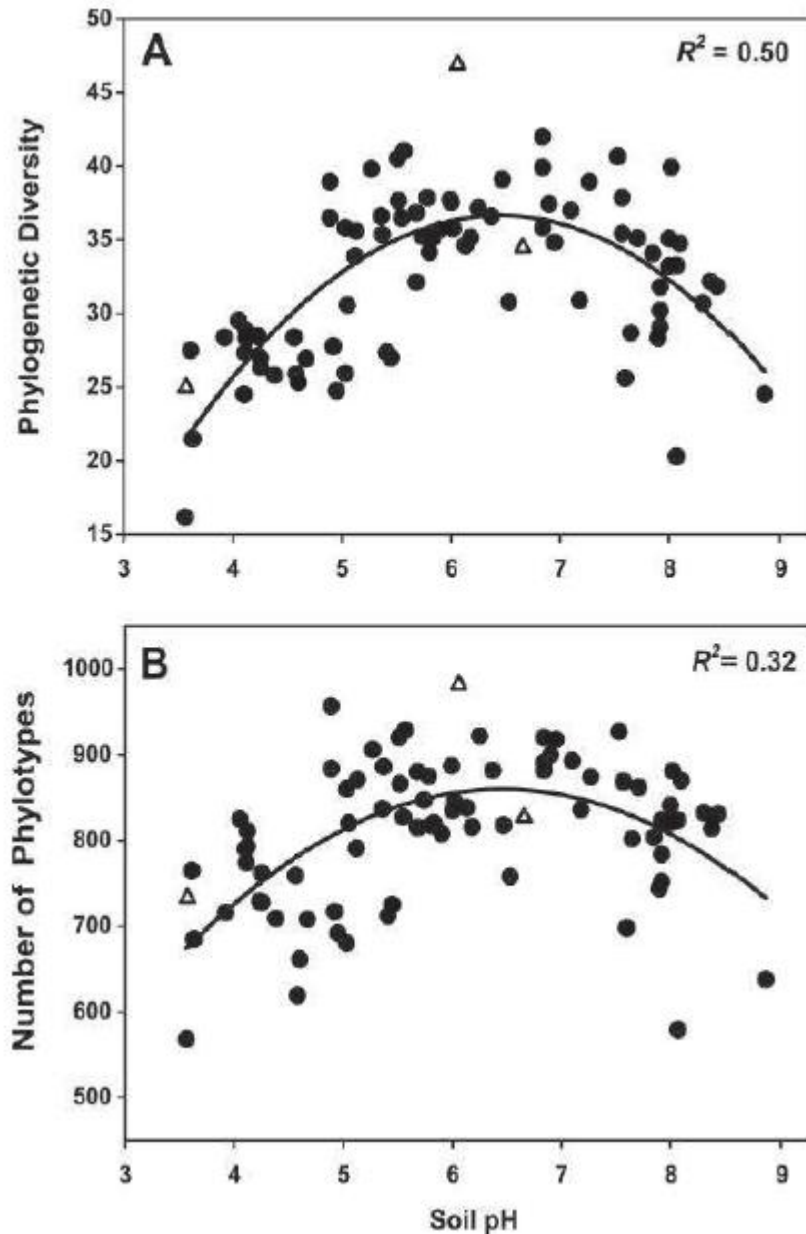


Figure 4 : Relation entre le pH des sols et la diversité bactérienne, en utilisant la diversité phylogénétique (A) et le nombre de phylotypes (B) défini à 97% de similarité. La courbe pleine représente la courbe de tendance. Les indices de diversité ont été calculés en utilisant 1200 séquences par sol (Lauber *et al.*, 2009).

Cette corrélation entre pH et diversité s'explique par le fait que le pH est directement corrélé à plusieurs paramètres du sol comme la disponibilité des nutriments, la solubilité des cations métalliques, la teneur en carbone organique, l'humidité et la salinité. Une seconde explication à cette corrélation est que le pH impose directement une contrainte physiologique sur les bactéries du sol.

Outre le pH, la structure du sol, la teneur en matière organique et la végétation influencent la diversité et l'abondance des bactéries. L'étude metagénomique des sols montre que quelle que soit la méthode d'analyse, certains phylums bactériens prédominent. C'est en tout cas le constat fait par Janssen et son équipe en 2006, qui se sont intéressés à l'analyse de différentes banques de séquences d'ADNr 16S provenant de sols différents (prairies, forêts, sols arides, agricoles...). Le résultat de cette étude montre que malgré les différences entre sols, certains phylums sont systématiquement dominants (Figure 5).

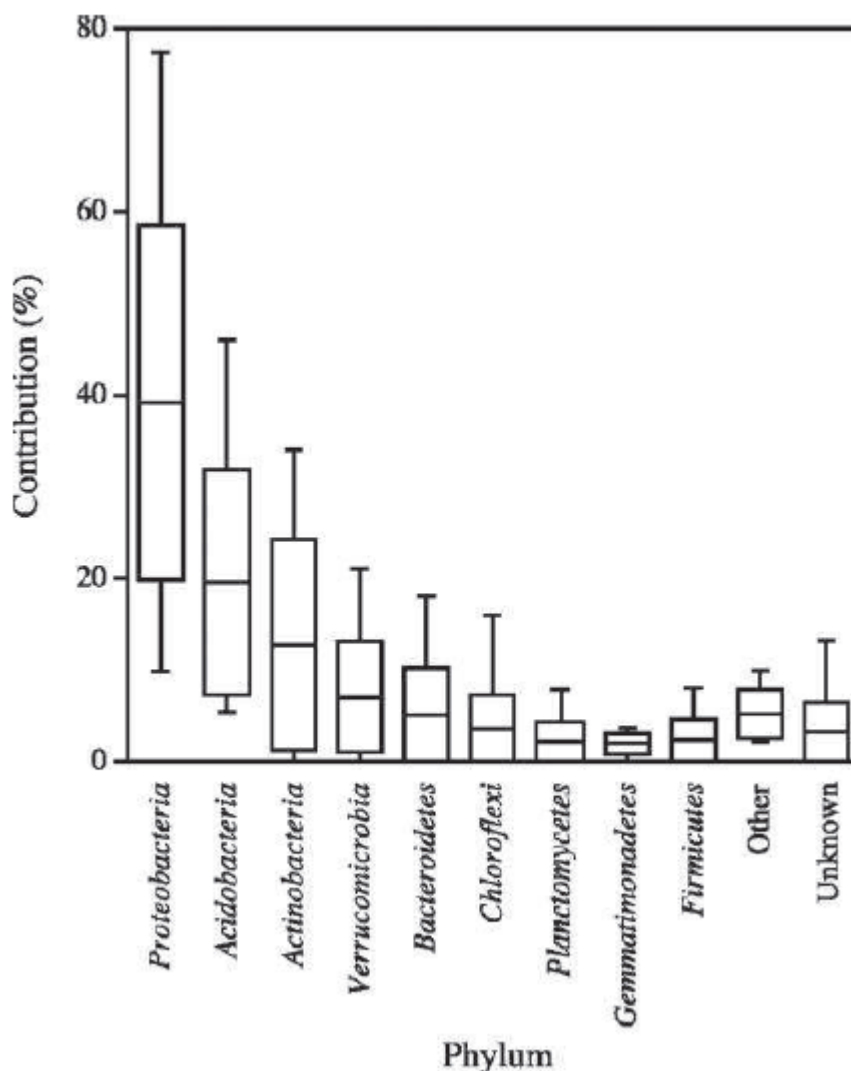


Figure 5 : Contribution des séquences d'ARNr 16S bactériens aux phylums les mieux représentés dans le sol. Les données ont été compilées à partir de 21 banques (2920 clones) de clones d'ARNr 16S obtenus par amplification à partir d'ADN du sol. La barre horizontale représente le pourcentage moyen de séquences attribuées au phylum, le rectangle représente l'écart type et les barres verticales représentent les valeurs maximales et minimales (North *et al.*, 2004).

En moyenne, 40% des séquences identifiées dans des échantillons de sol appartiennent aux *Proteobacteria*, 20% aux *Acidobacteria* et 13% aux *Actinobacteria*. Ce résultat a pu être retrouvé dans d'autres études (Delmont *et al.*, 2012, Roesch *et al.*, 2007).

4.3. Approche métagénomique

4.3.1. Définition

La métagénomique consiste à analyser l'ADN génomique d'une communauté bactérienne dans son ensemble. En d'autres mots, c'est une approche basée sur l'isolation directe de l'intégralité des acides nucléiques présents dans un échantillon prélevé dans un environnement donné, et ceci sans aucun isolement ou culture de bactéries au préalable (Handelsman, 2004; Simon et Daniel, 2011).

Le préfixe " méta " qui en grec veut dire littéralement « au-delà », induit une distinction majeure entre les termes " métagénomique " et " génomique ", ce dernier représentant l'étude de l'ADN génomique issu d'un seul microorganisme ou d'une cellule unique (Gilbert et Dupont, 2011).

4.3.2. Contexte et origine

Durant la majeure partie de son histoire, la microbiologie est restée centrée sur le fait que la culture des bactéries était une étape inhérente à toutes études dignes de ce nom.

En effet, les microbiologistes ont longtemps cru que les organismes bactériens pouvaient être classifiés, si et seulement si, ils étaient cultivés auparavant (Society of American Bacteriologists, 1923).

Cependant, dans les années 1980, les scientifiques ont réalisé que les bactéries étaient beaucoup plus ubiquitaires, diversifiées, et nombreuses que ce qu'ils pensaient (Ederer, 2011); et nous savons désormais que les techniques classiques de culture utilisées en microbiologie ont permis jusqu'à maintenant de caractériser seulement environ 1% de la diversité microbienne planétaire (D'Onofrio *et al.*, 2010).

Des estimations calculées par Whitman et collaborateurs (Whitman *et al.*, 1998) indiquent que la terre est l'habitat d'environ 5×10^{30} bactéries avec plus de 10^7 espèces de bactéries différentes, ces dernières étant présentes dans tous les environnements planétaires, capables de coloniser les habitats les plus extrêmes allant des sources hydrothermales à 340°C des profondeurs océaniques, aux tapis bactériens polaires (Whitman *et al.*, 1998).

Mais pourquoi l'étude de la diversité bactérienne devrait susciter autant d'intérêt ?

Les bactéries ne sont pas seulement omniprésentes, elles sont essentielles à la vie sur terre car elles représentent le plus grand réservoir de nutriments de la planète en séquestrant 50% du carbone total de tous les organismes vivants et en constituant le plus grand réservoir terrestre de nitrogène et de phosphore cellulaire (Whitman *et al.*, 1998).

De plus, les bactéries sont les principaux recycleurs de matières en décomposition permettant de rendre à nouveau disponible divers composés sous forme organique (Whitman *et al.*, 1998). Elles jouent donc un rôle critique dans les processus biogéochimiques en maintenant la survie des écosystèmes (Falkowski *et al.*, 2008).

Au vu du peu de connaissances de la communauté scientifique en termes de biodiversité bactérienne, les microbiologistes avaient besoin de nouvelles approches moins réductrices, et indépendantes de la culture pour la caractérisation des bactéries, prenant ainsi en compte l'écosystème dans son ensemble. Le développement de nouvelles méthodologies devaient permettre ainsi de caractériser les organismes identifiés dans leur environnement, ceci impliquant de pouvoir déceler leurs possibles interactions avec les autres membres de la communauté et leur potentiel métabolique.

C'est dans ce contexte que Pace et collaborateurs (Pace *et al.*, 1986), qui aspiraient à étudier la diversité et la classification des bactéries par d'autres méthodes que celles basées sur les traits morphologiques et physiologiques de ces organismes, furent les premiers à analyser directement (sans culture au préalable) les gènes codant pour l'ARN ribosomique (ARNr) 16S présents dans un échantillon environnemental afin de déterminer sa diversité microbienne. Depuis l'étude novatrice de Pace et collaborateurs en 1986 (Pace *et al.*, 1986), l'utilisation de l'ARNr 16S (couplée à la technologie de séquençage de Sanger (Sanger *et al.*, 1977) déjà très populaire à l'époque) s'est rapidement imposée au fil des années comme outil de prédilection pour la discrimination des différentes espèces bactériennes. Ce sont donc des recherches de ce type qui ont permis ultimement la naissance de la métagénomique, c'est-à-dire, une discipline nouvelle qui se situe à l'interface de la microbiologie, la génomique et l'écologie (Ederer, 2011). Ce terme ayant été mentionné officiellement pour la première fois par Handelsman et collaborateurs (Handelsman *et al.*, 1998) lors d'une étude visant à découvrir des nouveaux antibiotiques à partir de communautés microbiennes vivant dans le sol.

Cette nouvelle approche qu'est la métagénomique vise ainsi, en analysant le contenu en acides nucléiques d'une communauté microbienne quelconque, à caractériser la diversité bactérienne, la complexité génétique, mais aussi ses diverses interactions dont celles de nature métabolique. C'est sur ce dernier point que la métagénomique se distingue, en permettant d'aller bien plus loin que les analyses basées seulement sur l'amplification PCR des gènes codant pour l'ARNr 16S qui permettent de se concentrer uniquement sur des études taxonomiques. Ainsi, avec le soutien essentiel de la bio-informatique, le couplage de la métagénomique avec le séquençage à haut débit, offre dorénavant la possibilité d'accéder à des domaines inexplorés du monde bactérien.

La complexité et le nombre sans cesse grandissant de projets métagénomiques ont engendré la création par la communauté scientifique de consortiums visant à aider les chercheurs utilisant des données métagénomiques, en offrant des cyber-infrastructures permettant essentiellement le catalogage des génomes (complets) de référence issus de métagénomiques et des outils d'analyse et de consultation de ces données. Nous pouvons distinguer à l'heure actuelle quatre principales ressources de ce type :

- Le "*Human Microbiome Project (HMP)*" (<http://www.hmpdacc.org/>);
- Le "*Global Ocean Sampling (GOS) expedition*" (<http://www.jcvi.org/cms/research/projects/gos/>);
- Le "*Terragenome*" (<http://www.terragenome.org/>);
- Le "*Earth Microbiome Project (EMP)*" (www.earthmicrobiome.org/).

4.3.3. Notion d'espèce : de la microbiologie à la microbiologie de l'environnement

Le concept le plus largement utilisé pour définir une espèce est celui de l'espèce biologique proposée par Mayr (1942), bien que dans la réalité de nombreuses espèces (eucaryotes) soient définies sur la base de différences phénotypiques, les croisements étant difficilement observables dans la nature. L'application du concept biologique de l'espèce trouve ses limites dans la caractérisation des espèces microbiennes du fait de l'absence de reproduction sexuée associée à l'existence de transferts horizontaux entre individus évolutivement distants. Aussi, pour les procaryotes, les espèces sont actuellement définies de façon pragmatique en prenant en compte à la fois des caractères génotypiques et phénotypiques (Achtman et Wagner, 2008).

Lorsque les traits phénotypiques ne peuvent être décrits, ce qui est le cas pour la majorité des microorganismes sur terre, une désignation provisoire de l'espèce candidate peut

être proposée sur la base d'une discrimination exclusivement génétique (*Candidatus sp.*). Celle-ci est généralement basée sur l'analyse de la similitude entre des séquences de la petite sous unité de l'ARNr. Le pourcentage d'identité de ce marqueur phylogénétique est ainsi utilisé pour regrouper les séquences dans des unités taxonomiques opérationnelles ou OTUs (Operational Taxonomic Unit).

La majorité des études de diversité basées sur l'ADNr 16S définissent les OTUs comme étant des ensembles de séquences présentant au moins 97% d'identité entre elles. Ce seuil a été défini par Stackebrandt et Goebel (1994) comme équivalent au seuil d'hybridation ADN-ADN de 70% observé dans des expériences de réassociation réalisée entre les membres d'espèces bactériennes préétablies, issues d'organismes mis en culture.

La notion d'espèce est souvent vue comme l'unité de base de la biodiversité. Les OTUs sont devenues l'unité de mesure de la richesse spécifique en microbiologie du sol.

4.3.4. Contraintes de la définition d'une OTU en métagénomique

Comme nous l'avons vu, les études de diversité basées sur les gènes marqueurs analysent souvent la composition des communautés microbiennes en terme d'OTUs, et ce pour pallier à l'ambiguïté de la définition d'une espèce en microbiologie (Cohan, 2002).

Le seuil traditionnellement utilisé pour regrouper les séquences d'une même espèce en OTUs est 97%. Or, ce seuil ne peut pas être appliqué à tous les organismes et ce pour différentes raisons. D'une part parce qu'il a été défini à partir de bactéries isolées en culture et celles-ci ne représentent qu'une minorité des bactéries présentes dans l'environnement, d'autre part parce qu'il a été défini sur la séquence complète de la petite sous unité de l'ADNr 16S alors que la grande majorité des séquences générées, notamment avec les NGS sont de taille inférieure. Or, l'information biologique portée dépend de la taille du fragment et de la région amplifiée.

En effet, la présence de régions hypervariables (v1-v9) (Goebel et Stackebrandt, 1994) le long de la petite sous unité de l'ARNr, chacune avec des taux de variabilité différents (Figure 6), conduit à des résolutions taxonomiques différentes (Liu *et al.*, 2008 ; Wang *et al.*, 2007). Il est donc nécessaire d'adapter les seuils de clustérisation en OTUs à la région amplifiée. D'après une étude réalisée par Kim *et al.* (2011), le seuil de clustérisation varie de

96% à 98% pour l'ARNr 16S selon la région étudiée ; et ils préconisent l'utilisation des régions v1-v3 et v1-v4 pour une meilleure affiliation des bactéries.

Une méthode alternative pour la définition de la notion d'espèce, est l'intégration des informations phylogénétiques et des relations évolutives entre les organismes, et ce dans le but de s'affranchir des biais dus à la vitesse d'évolution différentielle entre les espèces. En effet, selon Koeppel et Wu (2013), les séquences contenues dans une OTU générée selon un seuil d'identité ne sont pas forcément monophylétiques, ce qui peut se traduire par une hétérogénéité écologique dans une OTU. Ces auteurs préconisent par ailleurs de remplacer les OTUs par la notion d'écotypes, qui au-delà de définir une unité pour les mesures de la diversité bactérienne, incorporent les modèles d'évolution et peuvent être discriminés sur la base de paramètres environnementaux.

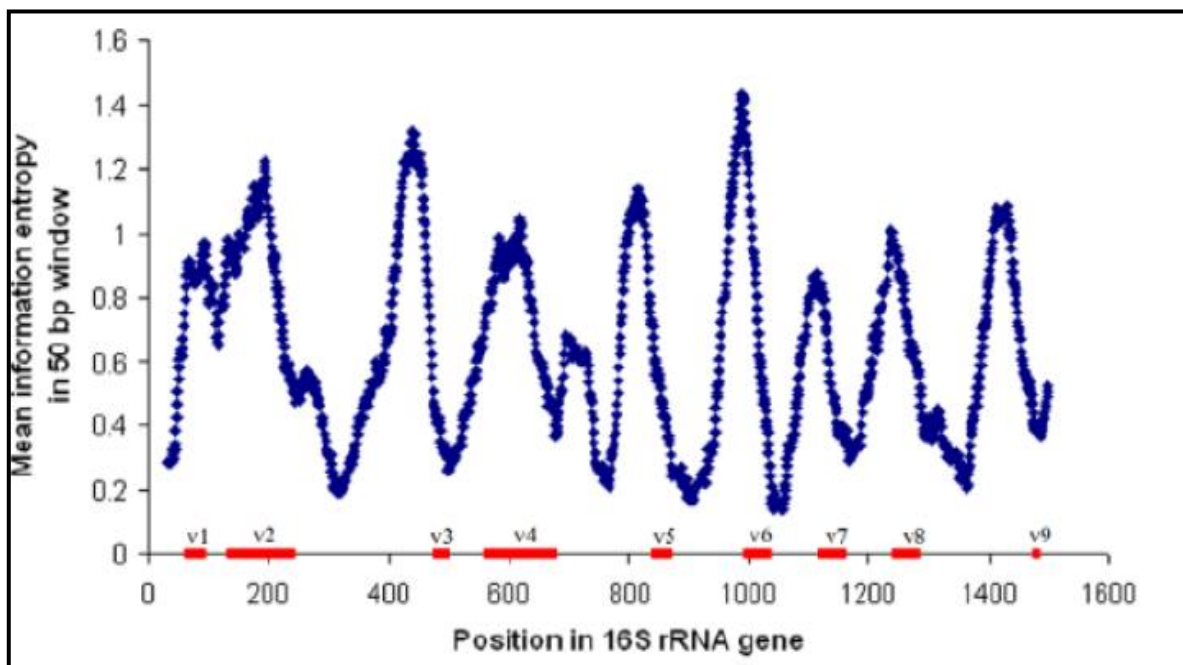


Figure 6 : Représentation schématique de la variabilité le long de l'ADNr. La variabilité est exprimée par l'entropie de Shannon en fonction des régions hypervariables (en rouge) sur les gènes codant pour l'ARNr 16S (Andersson *et al.*, 2008).

4.3.5. Méthodes et applications

La métagénomique représente un attrait indéniable pour les scientifiques dans le sens où elle leur permet d'accéder à un certain nombre des membres d'une communauté bactérienne donnée (indépendamment du fait de savoir si ces organismes peuvent être

cultivés ou pas), tout en ayant la possibilité d'accéder à une vue d'ensemble des voies métaboliques régnant en son sein.

Ceci étant dit, quels sont les types d'approches métagénomiques et leurs principales étapes en termes de protocole ?

Cela dépend bien évidemment du but de l'étude en question mais pouvons résumer ces étapes à la manière de celles illustrées dans la figure 7.

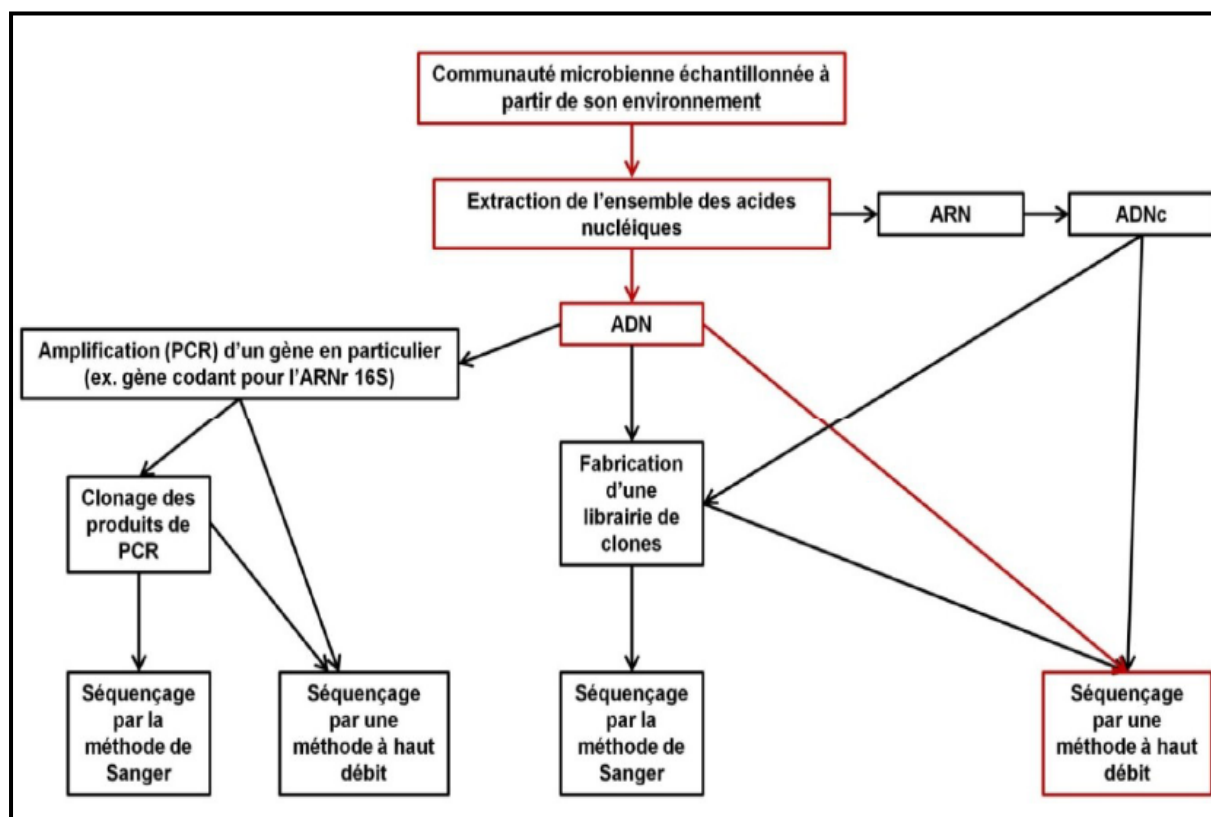


Figure 7 : Schématisation de plusieurs méthodologies courantes dans une analyse métagénomique (figure adaptée d'Ederer, 2011). Le chemin constitué de cadres et de flèches rouges symbolise la méthode métagénomique utilisée dans les différentes analyses présentées dans notre thèse.

Dans tous les cas, nous devons procéder après échantillonnage de l'écosystème en question, à une extraction de l'ensemble de l'ADN ou l'ARN génomique présent dans le ou les échantillon(s). Si l'estimation de la diversité bactérienne est le seul but de l'étude en question, une amplification PCR spécifique des gènes codant pour l'ARNr 16S sera réalisée afin de séquençer uniquement ces derniers en bout de ligne; nous obtiendrons ainsi une librairie à grande échelle de gènes d'ARNr. Une telle librairie peut ensuite être soumise à une biopuce de type PhyloChip afin de pouvoir comparer la diversité de cette dernière avec un ensemble

de gènes codant pour des ARNr 16S connus (Brodie *et al.*, 2006; DeSantis *et al.*, 2007; Yergeau *et al.*, 2009). Par contre, si nous s'intéressons cette fois au potentiel métabolique d'un écosystème, nous allons s'efforcer d'analyser l'ensemble des acides nucléiques de l'échantillon. L'ADN ou l'ADNc (ADN complémentaire obtenu par rétrotranscription de l'ARN isolé) extrait peut être alors traité essentiellement de trois manières différentes :

- Il peut être fragmenté aléatoirement, puis séquencé directement en utilisant des technologies de séquençage à haut débit. Il faut noter que ce séquençage direct d'acides nucléiques après extraction est de plus en plus privilégié à l'heure actuelle, étant donné qu'il permet de se soustraire à toute étape de clonage. À noter cependant que la plupart des techniques de séquençage à haut débit nécessitent la création d'une librairie de fragments d'ADN, néanmoins ces derniers ne sont pas clonés, et il ne faut donc pas confondre ce type de librairie avec celui mentionné dans le paragraphe suivant;

- Après fragmentation aléatoire, l'ADN peut également être cloné afin de constituer une librairie de clones à petits ou larges inserts qui sera ensuite séquencée par la technique de Sanger (Sanger *et al.*, 1977) ou par une technologie à haut débit; c'est ce que nous appelons le séquençage en aveugle. Cette librairie peut ensuite être parcourue pour rechercher puis séquencer des clones comportant un gène d'intérêt, ou bien séquencée en totalité pour permettre la reconstitution de génomes complets (Venter *et al.*, 2004);

- Troisièmement, l'ADN génomique extrait peut aussi être analysé par des biopuces de type GeoChip 3.0 contenant un ensemble de sondes spécifiques à des marqueurs phylogénétiques (tel que *gyrB*) et à des gènes d'intérêts connus (appartenant à des cycles biogéochimiques par exemple) (He *et al.*, 2007; 2010).

Dans tous les cas, le volume conséquent de fragments séquencés (appelés " reads " en anglais) lors d'une étude de métagénomique, fait en sorte que l'utilisation de la bio-informatique est indispensable afin d'analyser ces derniers.

D'un point de vue général, toute étude faisant appel à la métagénomique doit s'orienter vers l'une des approches suivantes (DeLong, 2007) :

- La première est centrée sur le " génome " afin de déterminer les membres composant une communauté bactérienne, avec le but ultime de tenter d'assembler leur génome complet;

- La seconde aspire quant à elle, à réaliser une analyse fonctionnelle de la communauté échantillonnée afin de déceler son potentiel métabolique. Elle est donc basée cette fois-ci non pas sur le " génome " mais sur le " gène ".

- La combinaison des deux approches citées précédemment est également envisageable.

Les études métagénomiques effectuées jusqu'à présent ont porté sur trois principaux types de communautés microbiennes :

- Les communautés que l'on retrouve à l'état naturel (environnements tempérés ou extrêmes);
- Les communautés vivant dans un environnement modifié par l'activité humaine;
- Les communautés qui dépendent d'un hôte pour survivre.

4.3.6. Défis

Conceptuellement, une approche métagénomique semble plutôt simple (Figure 5); il suffit d'extraire, puis de séquencer les acides nucléiques d'un échantillon environnemental, pour ensuite analyser le tout afin de caractériser la communauté bactérienne de l'échantillon en question. Mais en réalité, l'utilisation de la métagénomique amène certains défis et limites dont il faut tenir compte lors de la conception de la méthodologie d'un projet et dans l'interprétation des résultats obtenus. Les éventuels problèmes associés à une étude de métagénomique peuvent être répertoriés selon les aspects détaillés dans les sous-sections 4.3.6.1 à 4.3.6.6.

4.3.6.1. Environnement analysé

Il faut toujours être conscient que l'échantillonnage d'un environnement donné se fait en un point géographique bien particulier, et à un moment précis dans le temps, et que par conséquent, la composition d'une communauté peut varier grandement si l'un ou l'autre de ces paramètres change (Morris *et al.*, 2002; Kunin *et al.*, 2008).

De plus, la notion de microenvironnement vient compliquer le processus d'échantillonnage, qui généralement n'est pas assez précis pour différencier les sous-écosystèmes qui peuvent coexister dans un même environnement. Un gramme de sol forestier par exemple, peut contenir de 6400 à 38000 espèces différentes de bactéries (Ederer, 2011), cependant, ces dernières ne sont pas distribuées de façon égale dans le sol, il est bien plus probable qu'elles forment des mini-communautés organisées en sous-écosystèmes pouvant être en contact ou séparés les uns des autres. Ainsi, l'isolement de l'ADN d'un échantillon contenant ce type de microenvironnement ne permettra vraisemblablement pas de reconstituer la sous-distribution des membres qui composent ce genre de mini-communautés.

D'autre part, le principe même de la métagénomique fait en sorte que les fragments séquencés à partir d'un échantillon d'ADN génomique sont issus d'un plus ou moins grand nombre d'espèces différentes dont, pour la majeure partie, le génome complet n'est pas présent. L'identification de l'espèce d'origine auquel appartient un fragment donné n'est donc pas chose aisée.

En effet, la difficulté d'identifier potentiellement toutes les espèces d'un écosystème augmente avec la complexité et la diversité taxonomique de ce dernier (Sogin *et al.*, 2006; Huber *et al.*, 2007; Wooley *et al.*, 2010). Même un séquençage "en profondeur" d'un environnement naturel tel que le sol permet d'accéder seulement à une fraction infime de la variabilité génétique disponible dans le milieu en question (Gilbert et Dupont, 2011). Par conséquent, une analyse métagénomique permet habituellement d'identifier seulement les membres les plus abondants d'une communauté (Tyson *et al.*, 2004), car les espèces les plus nombreuses masquent celles qui sont rares, en fournissant une part beaucoup plus importante du matériel génétique total de la population bactérienne en question. Ainsi, les chances de retrouver des fragments appartenant aux espèces moins représentées dans l'échantillon sont plus faibles. Néanmoins, certaines techniques comme l'hybridation différentielle permettent de soustraire l'ADN le plus abondant d'un échantillon afin de permettre d'avoir accès aux génomes des espèces plus rares (Galbraith *et al.*, 2007).

D'une manière générale, il faut donc être prudent quant à la représentativité des résultats obtenus lors d'une étude métagénomique. En d'autres termes, ce qui est observé dans l'échantillon ne le sera pas forcément dans l'environnement dont ce dernier est issu. D'autant plus que Bent et Forney (Bent et Forney, 2008) ont mis en évidence que la majorité des communautés bactériennes complexes affichent une distribution taxonomique caractéristique où seulement quelques espèces dominent alors que la vaste majorité de la population est inusitée.

4.3.6.2. Obtention du métagénome de sol

Les approches métagénomiques ont nécessité le développement de méthodes recherchant la meilleure efficacité en termes d'exhaustivité et de représentativité de l'ADN extrait. Des protocoles ont alors été développés afin d'extraire et de purifier l'ADN à partir des échantillons environnementaux, mais aucun n'a encore été démontré suffisamment universel pour être accepté comme standard (Lakay *et al.*, 2007).

L'efficacité d'extraction et de purification de la totalité des acides nucléiques de la communauté demeure toujours un défi que tentent de résoudre les microbiologistes. Il existe deux types de méthodes d'extraction de l'ADN :

- Extraction directe : les cellules sont directement soumises à un traitement lytique dans le sol avant extraction et purification de l'ADN
- Extraction indirecte : les cellules bactériennes sont d'abord extraites de la matrice tellurique, purifiées puis soumises à un traitement lytique et de purification de l'ADN.

Les extractions directes d'ADN à partir d'échantillons environnementaux ont un fort rendement, contiennent un niveau moindre de contamination par des acides humiques, sont plus rapides et introduisent moins de biais que les méthodes où les cellules sont d'abord séparées de l'échantillon matrice avant l'extraction de l'ADN (von Wintzingerode *et al.*, 1997). De ce fait les méthodes d'extraction directe vont être plus couramment utilisées pour l'étude des communautés bactériennes. Nous allons donc, dans cette partie, nous intéresser plus particulièrement à ce type de méthodes d'extraction directe.

Les différentes méthodes d'extraction de l'ADN sont basées sur une dégradation physique ou enzymatique des enveloppes cellulaires : les plus courantes emploient un broyage à l'azote liquide (Volossiouk *et al.*, 1995), une rupture par microondes (Orsini, 2001), un traitement enzymatique (Zhou *et al.*, 1996 , Stach *et al.*, 2001), une disruption physique avec des billes en mouvement " bead beating " (Miller *et al.*, 1999 , Burgmann *et al.*, 2001). L'isolement de l'ADN du sol résulte cependant en la co-extraction des acides humiques qui interfèrent avec la détection et la quantification de l'ADN (Zhou *et al.*, 1996). Les impuretés humiques du sol sont difficiles à séparer de l'ADN, du fait de propriétés chimiques comparables à celles de l'ADN (Saano, 1995).

La plupart des contaminants dont les acides humiques, les polysaccharides ou l'urée, présentent des propriétés de solubilisation similaires à l'ADN, ne pouvant donc pas être totalement éliminés durant les protocoles classiques d'extraction tels que ceux utilisant l'extraction protéique par le phénol-chloroforme ou un traitement enzymatique avec des protéases. Le niveau de contamination des solutions d'ADN provenant d'un échantillon de sol demeure élevé (Moreira, 1998).

De plus, beaucoup de sols organiques contiennent des acides humiques avec des groupes phénoliques qui se lient de manière covalente à l'ADN ou aux protéines (Tsai et Olson, 1992). Récemment Sebastianelli *et al.* ont mis au point une méthode d'extraction de

l'ADN du sol en utilisant des nanoparticules pour une bio-séparation magnétique (Sebastianelli *et al.*, 2008).

Il y a trois structures tridimensionnelles qui ont la capacité de se lier à d'autres composés par leurs groupes fonctionnels réactifs : l'eau, les ions et les molécules organiques (Stevenson, 1976). Donc, la plupart des composés organiques naturels peuvent se lier ou être absorbés aux substances humiques (Fortin *et al.*, 2004). Les traitements nécessitent alors de combiner différentes méthodes de purification pour éliminer les composés phénoliques afin d'obtenir une solution d'ADN suffisamment pure pour un traitement par les outils enzymatiques utilisés en biologie moléculaire (Young *et al.*, 1993).

Les étapes de purification incluent par exemple une chromatographie par exclusion de taille (Kuske *et al.*, 1998 ; Hurt *et al.*, 2001), différentes filtrations sur colonne de gel de silice (Miller, 2001), une séparation isopycnique sur gradient de chlorure de césium (Holben *et al.*, 1988), etc. Cependant les substances humiques peuvent rivaliser avec les acides nucléiques au niveau des sites d'adsorption durant la purification sur les micro-colonnes (Harry *et al.*, 1999). Selon Miller (2001), la résine sépharose 2B est plus efficace que la résine séphadex pour purifier l'ADN du sol ou de sédiments des acides humiques.

Chaque type d'environnement, du fait de ses caractéristiques propres, nécessite une optimisation de la méthode d'extraction utilisée. De ce fait l'efficacité de l'extraction de l'ADN dépend de la qualité du sol, particulièrement de son contenu en argile et matière organique, car les micro-organismes peuvent interagir avec les colloïdes du sol.

Un nombre plus élevé de cellules est directement proportionnel au taux de carbone organique et d'azote, au contenu en argile ainsi qu'au contenu en acides humiques (Fortin *et al.*, 2004 , Sagova- Mareckova *et al.*, 2008). La variation de rendement de l'ADN peut être expliquée aussi par une différence de sensibilité ou plutôt « d'agressivité » des méthodes individuelles comparant à la fois la communauté microbienne et l'ADN extrait. Premièrement, différentes méthodes peuvent accéder à différentes parties de la communauté bactérienne, les bactéries peuvent contenir des quantités d'ADN différentes et l'ADN de diverses espèces libéré différemment (Kauffmann *et al.*, 2004). Le rendement peut alors être affecté par le type de méthode utilisée. Deuxièmement, une différence de rendement d'ADN peut être expliquée par la présence d'organismes autres que des bactéries et une fois encore les différents niveaux de sensibilité d'accession à l'ADN des méthodes respectives. La recommandation est d'évaluer les caractéristiques des sols avant de sélectionner le protocole

optimal pour l'extraction et la purification d'ADN (Sagova-Mareckova *et al.*, 2008). Pour des sols riches en acides humiques, la méthode d'extraction bead beating est la plus efficace, tandis que la méthode de broyage dans l'azote liquide donne des quantités négligeables d'ADN, elle est donc révélée comme insuffisante. Mais il apparaît que la méthode bead beating extrait plus de substances humiques par rapport à la méthode basée sur l'action de micro-ondes. Dans ces cas là, une simple purification de l'ADN est nécessaire (PVPP et sépharose 2B) (Lakay *et al.*, 2007).

Donc, les caractéristiques du sol qu'il faut prendre en compte pour une extraction optimale d'ADN sont le pH, la matière organique, l'argile et les limons, qui indiquent aussi une forte activité microbienne.

La plupart des difficultés ont été rencontrées pour des sols forestiers acides où des nouvelles méthodes d'extraction et de purification ont été mises en place pour augmenter le rendement d'extraction de l'ADN ainsi que l'efficacité de la PCR : prétraitement du sol au CaCO_3 ou purification de l'ADN extrait par CaCl_2 (Sagova-Mareckova *et al.*, 2008). Ces méthodes sont comparées à celle du kit Mobio PowerSoil et celle du phénol/chloroforme (Miller *et al.*, 1999). Au fil du temps, le bead beating apparaît comme la technique la plus efficace pour les agrégats du sol et la perturbation des cellules, elle est aussi utilisée dans les kits commerciaux. Pourtant le broyage du sol modifie fortement la structure et peut de ce fait affecter les propriétés d'adsorption et de désorption (Frostegard *et al.*, 1999), pouvant aussi affecter l'intégrité physique des molécules d'ADN. Cependant il n'y a toujours pas de moyens de quantifier l'ADN total contenu dans un sol (Burgmann *et al.*, 2001).

En effet, la lyse complète des cellules provenant d'échantillons environnementaux est toujours inaccessible. Une plus grande quantité d'ADN obtenue n'est pas toujours équivalente d'une plus grande richesse d'espèces, par la possibilité que l'ADN extrait provienne de cellules facilement lysées (Stach *et al.*, 2001). Par exemple les cellules bactériennes de petite taille (1.2 à 0.3 μm) sont plus difficiles à lyser que les cellules plus volumineuses (2 à 10 μm) (More *et al.*, 1994).

4.3.6.3. Etape de clonage

La fabrication d'une librairie de clones préalablement au séquençage peut entraîner certains biais inhérents au processus de clonage (niveau d'efficacité, nombres suffisants de clones produits), mais aussi à l'étape de PCR qui est souvent nécessaire après un clonage, et

qui comporte certains désavantages comme les biais introduits par l'utilisation des sondes et/ou par l'amplification exponentielle (Suzuki et Giovannoni, 1996; Wang et Wang, 1997; Kanagawa, 2003; Acinas *et al.*, 2005; Sipos *et al.*, 2010).

Par ailleurs, des fragments d'ADN pauvres en bases GC peuvent être moins stables et affecter l'efficacité du clonage (Temperton *et al.*, 2009). De plus, certains gènes ne peuvent être incorporés dans les vecteurs de librairie en raison de leur instabilité ou de leur toxicité pour la cellule hôte (Sorek *et al.*, 2007).

Néanmoins, l'ensemble de ces biais peuvent être évités en supprimant tout simplement l'étape de clonage, et donc en séquençant directement l'ADN génomique obtenu après extraction. Cette approche a d'ailleurs été proposée comme la méthode la plus précise pour l'estimation de la diversité taxonomique de métagénomés (von Mering *et al.*, 2007).

4.3.6.4. Séquençage à haut débit : La nouvelle technologie " pyroséquençage 454 "

La présence de certaines substances inhibitrices contenues dans l'environnement prélevé, comme l'acide humique souvent associée aux échantillons provenant du sol, peut affecter l'efficacité du séquençage. Par ailleurs, l'essor de la métagénomique n'aurait pu avoir lieu sans les remarquables avancées technologiques que le séquençage de l'ADN a connu lors de la dernière décennie, notamment avec l'avènement des techniques de séquençage de deuxième génération, qui permettent de produire un nombre beaucoup plus important d'ADN séquencés, en moins de temps et à moindre coût.

a. Historique du séquençage: Au cours du 20^{ème} siècle, le séquençage de l'ADN a révolutionné la recherche dans le domaine des sciences de la vie, en réalisant un grand pas pour l'humanité avec le séquençage complet du génome humain en 2001 (Lander *et al.*, 2001; Venter *et al.*, 2001). Historiquement, ceci a été rendu possible grâce aux prouesses de deux pionniers travaillant sur le séquençage, Gilbert et Sanger qui se partagèrent le prix Nobel de chimie en 1980 pour leurs co-découvertes dans ce domaine. Tout a commencé en 1975 avec Sanger et Coulson qui sont parvenus à développer la méthode de terminaison des chaînes, permettant le séquençage d'environ 50 nucléotides contigus sur plusieurs jours (Sanger et Coulson, 1975).

Deux ans plus tard, Maxam et Gilbert présentent alors une autre méthode de séquençage basée sur la dégradation chimique de l'ADN par coupures sélectives (Maxam et Gilbert, 1977).

Cependant, la même année Sanger publie une méthode complémentaire à celle qu'il avait déjà créé en 1975, en introduisant l'utilisation des didéoxynucléotides, des inhibiteurs de l'ADN polymérase spécifiques à chaque base nucléotidique (Sanger *et al.*, 1977). La même année, le développement de cette technologie a ainsi permis de séquencer le tout premier génome complet, celui du phage ϕ X (Sanger *et al.*, 1977).

Dans les deux décennies suivantes, la méthode de séquençage de Sanger a été améliorée en termes de procédés et de chimie, avec la capacité de produire des séquences d'une longueur de 1000 nucléotides avec une précision pouvant atteindre 99,999% (Shendure *et al.*, 2004).

Mais l'avancée majeure depuis cette optimisation reste son automatisation, qui a ouvert la voie à la démocratisation du séquençage à partir des années 1990 (Hutchison, 2007). Ces avancées dans la technique de Sanger ont rendu possible le séquençage d'une variété de génomes bactériens et eucaryotes, dont les plus marquants sont : *Saccharomyces cerevisiae* (Goffeau *et al.*, 1996), *Escherichia coli* (Blattner *et al.*, 1997), le riz (Goff *et al.*, 2002; Yu *et al.*, 2002), la souris (Waterston *et al.*, 2002), et enfin l'homme (Lander *et al.*, 2001; Venter *et al.*, 2001).

Néanmoins, la démocratisation du séquençage en aveugle de génomes dans les années 1990, a conduit à la génération d'un volume de plus en plus important de données à séquencer, rendant ainsi la technologie de Sanger souvent inadéquate, car jugée trop longue et particulièrement onéreuse (Shendure et Ji, 2008).

C'est la raison pour laquelle en 2004, le National Human Research Institute a lancé " le programme de séquençage du génome à 1000\$ " dans le but de réduire à cette somme le prix du séquençage des génomes de mammifère, et ceci d'ici une dizaine d'années. Ainsi après cette annonce, plusieurs compagnies se sont lancées dans l'aventure afin de créer des nouveaux instruments de séquençage permettant de produire plus rapidement, et pour moins cher, des quantités massives de données (Meldrum, 2001; Meldrum et Holl, 2002). C'est dans cet esprit que sont nées les technologies de séquençage à haut débit de deuxième et troisième génération.

b. Séquençage à haut débit : Apparues à partir de 2005, les méthodes de séquençage à haut débit de l'ADN, dites de "nouvelle génération" (NGS), ont gagné beaucoup de terrain sur la technologie de Sanger pour le séquençage des petits génomes et des écosystèmes environnementaux. Ceci est principalement dû au fait qu'elles sont grandement parallélisées permettant la production de plusieurs millions de séquences (*reads*) différentes par expérience sans avoir nécessairement besoin d'une étape de clonage au préalable.

Ainsi, ces méthodes permettent d'obtenir un plus grand nombre de séquences, plus rapidement et pour un prix inférieur à la méthode de Sanger. Néanmoins, les " *reads* " produits sont plus courts pour la plupart des technologies existantes en 2012: de 35 à 250 pb (paires de bases), excepté pour le pyroséquençage 454 qui depuis peu, serait en mesure de séquencer des fragments d'une longueur moyenne de 700 pb égalant ainsi plus ou moins la méthode de Sanger (Siqueira *et al.*, 2012). Les techniques de séquençage de deuxième génération les plus utilisées à l'heure actuelle utilisent des méthodologies différentes, cependant, elles présentent des similarités dans leur protocole, plus particulièrement elles comprennent toutes les étapes suivantes :

- Etape 1 : Construction d'une librairie de séquences, qui consiste à fragmenter l'ADN génomique pour ensuite lier des adaptateurs à chacune des extrémités de ces fragments;
- Etape 2 : Chaque fragment de la librairie est amplifié de manière à ce que chaque produit d'amplification soit séparé les uns des autres;
- Etape 3 : Séquençage effectué par plusieurs cycles de réactions enzymatiques, le tout étant mesuré en temps réel.

c. Pyroséquençage : Le pyroséquençage est la technique qui connaît actuellement le plus grand succès concurrençant la méthode de Sanger. Cette technique de séquençage d'ADN introduite depuis 1988, par Hyman *et al.* (Hyman, 1988), a été améliorée par un groupe suédois (Ronaghi *et al.*, 1996 , Ronaghi *et al.*, 1998) par introduction d'une étape PCR. Il s'agit d'un séquençage par synthèse (SBS) qui se caractérise par la révélation en temps réel de l'activité de l'ADN polymérase (*real time sequencing*) et qui ajoute un seul nucléotide non fluorescent à la fois.

La première étape consiste à préparer le mélange réactionnel, avec les enzymes clés et les différents substrats. Ici, les nucléotides ne sont pas ajoutés tous ensemble comme dans une réaction de séquençage Sanger mais successivement. Si le nucléotide ajouté dans le milieu

réactionnel correspond à celui attendu par la polymérase, il sera incorporé dans le brin en cours de synthèse (d'élongation) libérant un pyrophosphate. L'ATP sulfurylase vient alors transformer ce pyrophosphate (Ludwig *et al.*, 2004) en ATP qui est alors utilisé, couplé à une Luciférine, par une luciférase avec production d'oxyluciférine et d'un signal lumineux. L'apyrase dégrade les nucléotides en surplus. Le signal lumineux est capté par un capteur CCD (Charge-Coupled Device) puis reproduit sous forme d'un pic sur le Pyrogramme. La hauteur de ce pic est fonction de l'intensité du signal lumineux, elle-même proportionnelle au nombre de nucléotides incorporés en même temps. Nous pouvons donc déduire la séquence à partir de la taille des pics obtenus. Par ailleurs, en cas de mélange de nucléotides à une même position (polymorphisme de séquence), la taille des pics permet d'avoir une quantification de la proportion de brins porteurs de l'un ou l'autre des nucléotides (Figure 8).

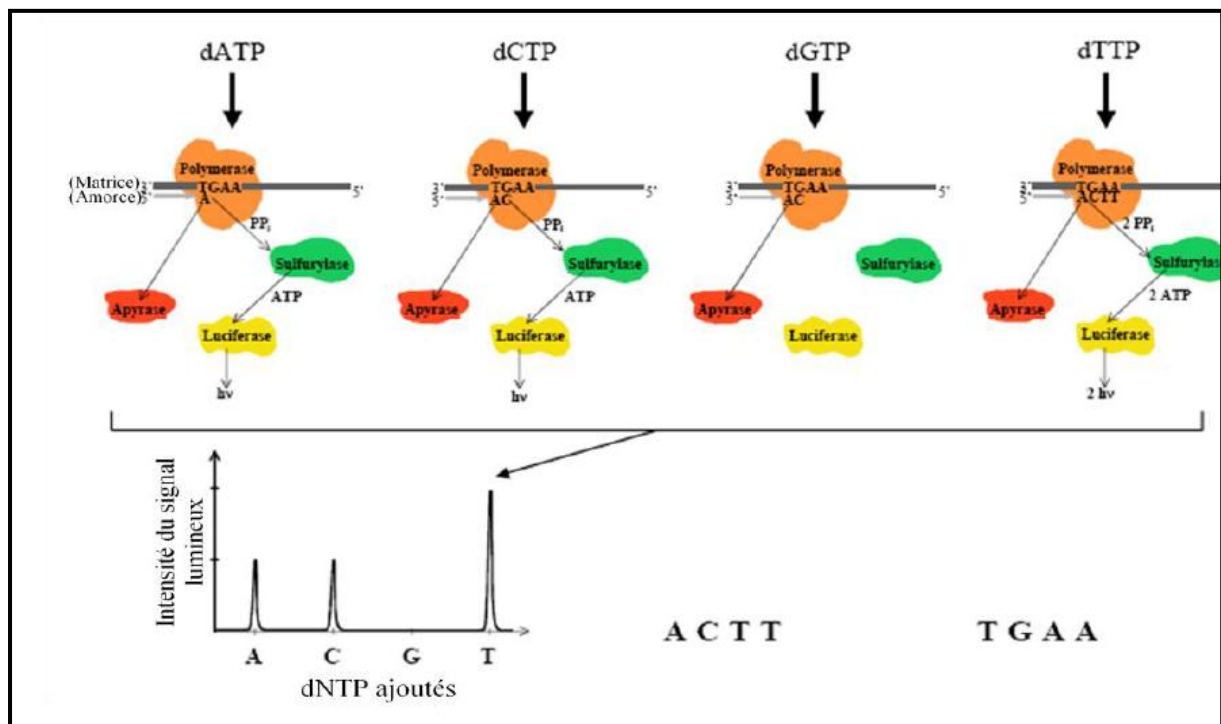


Figure 8 : Etapes du pyroséquence (Ahmadian *et al.*, 2006).

d. La nouvelle technologie 454 : Cette technique permet de traiter avec un seul instrument le GS20, plus de 20 millions de bases nucléotidiques par cycle de quatre heures. Cette technique ne nécessite pas de clonage et permet une lecture directe de la séquence obtenue après séquençage. Une seule machine (GS20) génère en fait, avec plus de 99% d'exactitude, autant de données que 100 séquenceurs capillaires à haut débit. La technologie de 454 Life sciences conçue par Rothberg (Margulies *et al.*, 2005) est fondée sur l'intégration

de plusieurs techniques : le pyroséquençage, les technologies des plaques en fibre optique picotitré [PicoTiterplate (PTP)] qui contiennent 1.6 million de puits, la PCR en émulsion (emPCR) (Figures : 9 et 10) dans des microréacteurs (gouttelettes d'une émulsion) (300 000 réactions PCR en parallèle), ainsi que des technologies informatiques de pointe pour l'acquisition, le traitement et l'analyse des images.

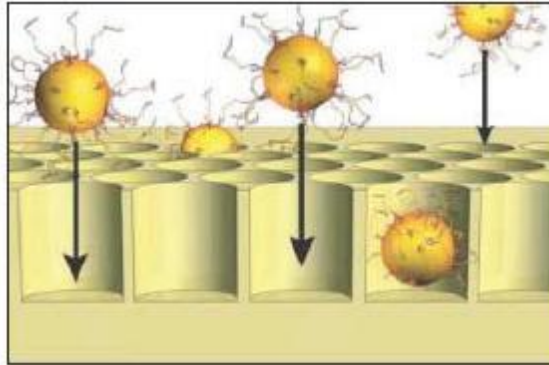


Figure 9 : Représentation du dépôt d'une microbille dans un puits d'une plaque en fibre optique contenant 1,4 millions de puits (Shendure et Ji, 2008).

- Etape1: Préparation d'une banque ADN simple brin (ADN sb) avec deux adaptateurs. L'ADN génomique est fragmenté par nébulisation, deux adaptateurs (A et B) sont fixés par ligation aux deux extrémités. Un adaptateur est lié à la biotine pour permettre la purification des molécules d'ADN porteurs des adaptateurs A et B.
- Etape 2 : Amplification clonale des molécules d'ADN simple brin. Des microbilles avec en surface des amorces complémentaires à un des adaptateurs permettent de fixer une molécule d'ADN sb à la fois. Les microbilles porteuses des brins d'ADN sb sont mises en émulsion en présence des réactifs pour PCR (Dressman *et al.*, 2003). Chaque goutte (microréacteurs) englobe une microbille et donc une molécule d'ADN, ce qui permet une amplification clonale de chaque fragment (Figures : 9 et 10).
- Etape 3 : Séquençage dans des plaques picotitrées. Après amplification, les microgouttelettes (microréacteurs) sont dissociées, et les microbilles porteuses de l'ADN simple brin largement amplifié sont transférées dans une plaque en fibre optique contenant 1,4 million de puits. Les puits possèdent un diamètre qui assure le dépôt d'une microbille par puits. Avec ce système, 400 000 réactions de séquençage peuvent être réalisées en parallèle. Le séquençage se fait selon le principe de pyroséquençage décrit ci-dessus. Une caméra CCD permet de capturer les images après addition de chaque nucléotide.
- Etape 4 : Analyse des images et détermination de la séquence d'ADN. Connaissant l'ordre dans lequel les 4 nucléotides sont ajoutés automatiquement, l'analyse des différentes images

capturées permet la déduction de la séquence des différents fragments d'ADN (illustré par Pyrogramme). Des logiciels bioinformatiques sont ensuite utilisés pour reconstituer la séquence initiale grâce au groupage des contigs.

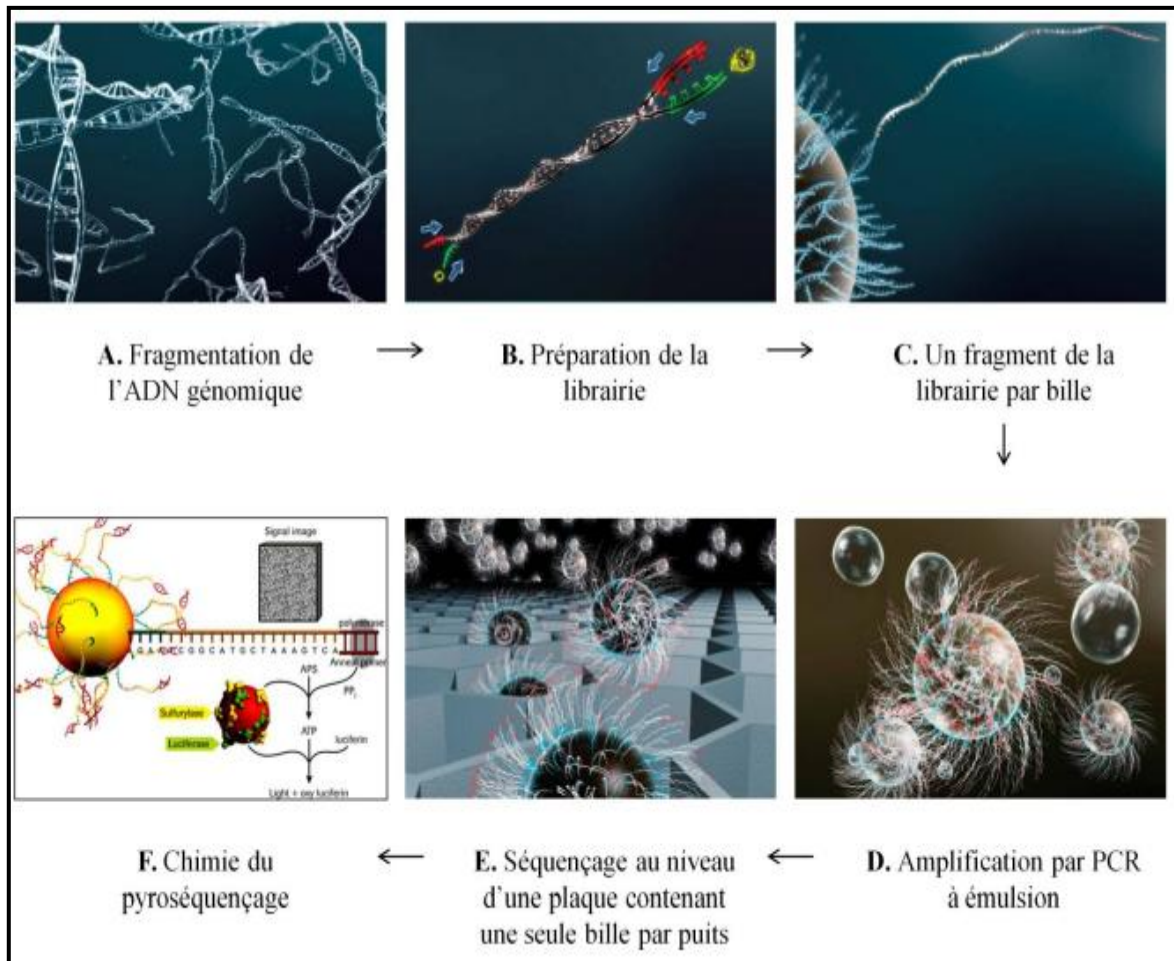


Figure 10 : Principe du pyroséquençage 454 (figure adaptée de 454 Life sciences, © 1996-2012 Roche Diagnostics Corporation, www.454.com).

Considérant, la taille relativement importante des fragments produits et la rapidité d'exploitation de cette technologie, c'est la méthode de séquençage que nous avons choisie pour réaliser les analyses métagénomiques de cette thèse.

4.3.6.5. Traitement des données et analyse de la diversité

Parmi les différentes méthodes développées dans le but de caractériser la diversité taxonomique d'un milieu, nous pouvons distinguer à ce jour :

- Les indices de diversité, qui permettent de quantifier l'hétérogénéité de la biodiversité de l'environnement étudié, comme par exemple l'indice de Shannon (Shannon, 1948), ou encore l'indice de Simpson (Simpson, 1949);

- Les estimateurs de richesse spécifique (en espèces) tels que Chao1 ou Chao2 (Chao, 1984; 1987), ou encore l'estimateur ACE (Chao et Lee, 1992);
- Les courbes de raréfaction utilisées pour déterminer le taux de couverture en espèces de l'échantillonnage. Elles sont matérialisées par des courbes indiquant la croissance du nombre d'organismes appartenant à un groupe taxonomique donné en fonction du nombre d'organismes échantillonnés.

Il existe de nombreux logiciels dédiés à l'analyse de diversité de communautés échantillonnées. Nous pouvons citer par exemple, MOTHUR (Schloss *et al.*, 2009) qui a l'avantage de donner accès à une combinaison de plusieurs autres outils d'analyse capables de calculer des indices de diversité tels que : DOTUR (Schloss & Handelsman, 2005), SONS (Schloss et Handelsman, 2006), S-Libshuff (Schloss *et al.*, 2003), et Unifrac (Lozupone *et al.*, 2006).

Plus récemment, des logiciels similaires mais adaptés aux caractéristiques des données métagénomiques ont vu le jour, c'est le cas d'ESPRIT (Sun *et al.*, 2009) qui est téléchargeable localement, et du « Pyrosequencing pipeline » (Cole *et al.*, 2009) de RDP qui est utilisable en ligne seulement; à noter cependant que ces deux logiciels fonctionnent uniquement à partir de bibliothèques métagénomiques de gènes codant pour l'ARN ribosomique 16S obtenus par pyroséquençage seulement. QIIME (Caporaso *et al.*, 2010) quant à lui, est un outil plus complet et plus flexible, puisqu'il est capable d'exécuter des analyses de diversité en utilisant aussi bien les gènes codant pour l'ARNr 16S, que les marqueurs non-ribosomiques. De plus, il est compatible avec les " reads " issus des technologies de Sanger et de 454, et permet d'exécuter des analyses taxonomiques comparatives de métagénomes.

Matériels et méthodes

Partie 1 : Isolement et caractérisation phénotypique et moléculaire des bactéries telluriques

1. Échantillonnage de sol

Après avoir écarté les débris de surface, 5 points de prélèvement de sol ont été réalisés (sur les 10 premiers centimètres) à partir de la forêt " El Meridj ", situé au sud-est de Constantine, est de l'Algérie (coordonnées: 36 ° 20'37.04 "N, 6 ° 41'21.36" E). Par la suite, les échantillons élémentaires ont été rassemblés et tamisés sur 2 mm pour les homogénéiser.

2. Isolement et purification des souches

2.1. Réalisation des dilutions

1g de sol fraîchement prélevée a été en suspension dans 9 ml d'eau physiologique stérile (NaCl 9g/l), ce qui représente la dilution 10^{-1} . Après homogénéisation et agitation à l'aide d'un vortex pendant 5 min, des dilutions décimales en cascades ont été réalisées jusqu'à la dilution 10^{-6} , en prélevant 1 ml de chaque dilution et en le transférant dans 9 ml d'eau physiologique stérile (Pochon, 1954 ; Rapilly, 1968 ; Bastide *et al.*, 1986).

2.2 Ensemencement et incubation

L'ensemencement a été réalisé en étalant, à l'aide d'un râteau stérile, 0,1 ml de chacune des trois dernières dilutions (10^{-4} , 10^{-5} , 10^{-6}) en triplicata sur des boîtes de Pétri contenant la gélose nutritive. Les cultures sont, ensuite, incubées à 37°C pendant 18-24h. Les colonies ont été prélevées et repiquées plusieurs fois jusqu'à l'obtention de souches pures.

3. Identification phénotypique des isolats

L'identification phénotypique des isolats a été réalisée en utilisant une gamme de critères morphologiques, physiologiques et biochimiques. Ces derniers ont été comparés, par la suite, aux caractéristiques connues du Manuel de Bergey de Bactériologie systématique (Holt et Williams, 1989).

Les tests et les examens d'identification phénotypique ayant été réalisés sont :

1. Examen de l'aspect macroscopique des colonies.
2. Coloration de Gram.

3. Mobilité à 25 et à 37°C.
4. Présence d' Ortho-Nitrophényl-β-galactoside.
5. Présence d'oxydase.
6. Présence de catalase.
7. Hydrolyse du lactose.
8. Hydrolyse du saccharose.
9. Hydrolyse du mannose.
10. Production d'H₂S (test sur milieu TSI).
11. Production du gaz (test sur milieu TSI).
12. Présence d'uréase.
13. Production d'indole.
14. Utilisation du citrate.
15. Rouge de méthyle (test RM).
16. Réaction de Voges-Proskauer (test VP).
17. Test d' Ornithine décarboxylase.
18. Test de la Lysine décarboxylase.
19. Test de l'Hormone anti-diurétique.

4. Identification moléculaire des isolats

4.1. Identification protéomique par MALDI-TOF MS

Cette technique a été réalisée selon le protocole de Seng *et al.* (2009) et Bizzini *et al.* (2010). Son principe général est basé sur l'étude du déplacement d'entités ioniques dans des champs électromagnétiques, ce qui permet l'identification des bactéries grâce à l'analyse de leur contenu protéique (Claydon *et al.*, 1996). L'ionisation de l'échantillon est essentielle et réalisée grâce à une matrice.

4.1.1. Préparation de la matrice

La matrice doit être préparée sous hotte chimique avant chaque série d'analyses et conservée à 4 °C à l'abri de la lumière. Le protocole de sa préparation selon Williams *et al.* (2003) est le suivant :

- Dans un tube polypropylène 1.5 ml, introduire 2 spatules d'alpha-cyano-4-hydroxy-cinnamic acid (HCCA),
- Ajouter 500 µl d'acétonitrile HPLC,

- Ajouter 250 μ l de TFA à 10% (trifluoroacetic acid),
- Ajouter 250 μ l d'eau HPLC,
- Bien mélanger (vortexer ou secouer vigoureusement)
- Soniquer 10 min (la solution doit devenir trouble),
- Centrifuger 5 min à 13000 g,
- Transférer le surnageant dans un tube polypropylène 1.5 ml propre : la matrice est prête.

4.1.2. Dépôt des isolats

L'étape de dépôt est essentielle pour une bonne identification et elle renferme les étapes suivantes :

- Déposer 2 spots du témoin positif " T+ " (1.5 μ l de " BTS ") et recouvrir chaque dépôt de 1.5 μ l de matrice.
- Sur la ligne suivante dans la même colonne, déposer 2 spots du témoin négatif " T- " (1.5 μ l de la matrice).
- Sur les lignes suivantes de la cible MALDI, déposer les échantillons en deux exemplaires dans les zones définies (cercles) de la cible MSP 96 (figure 11), en prélevant une petite quantité de chaque colonie des isolats poussés sur gélose au sang à l'aide d'une pointe de cône de pipette 10 μ l.
- Une fois tous les dépôts effectués, déposer 1.5 μ l de la matrice sur chaque spot.
- Laisser sécher pendant 5 min sous la hotte.

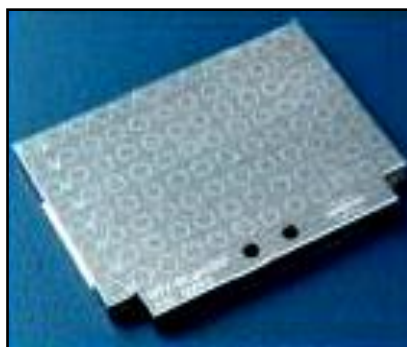


Figure 11 : Cible MALDI type Microflex, modèle " MSP 96 " (11 colonnes et 8 lignes).

4.1.3. Introduction de la cible MALDI dans le spectromètre de masse Microflex LT

- Appuyer sur le bouton vert en façade du Microflex LT.
- Soulever la trappe d'insertion de la cible MALDI (Figure 12).



Figure 12 : Trappe d'insertion en position ouverte sans (a) et avec (b) cible.

- Déposer délicatement la cible à analyser. Passer le doigt (sans gants) sur le joint noir de la trappe pour éliminer les poussières et refermer la trappe d'insertion.
- Appuyer sur le bouton vert pour introduire la cible dans le spectromètre. Pour atteindre le vide rapidement, maintenir une pression sur la trappe avec la main jusqu'à ce que le voyant "TARGET" devienne vert (Figure 13).
- Patienter le temps que le vide soit atteint (onglet " Vacuum " dans FlexControl). Lancer les analyses seulement une fois l'appareil prêt.



Figure 13 : Introduction de la cible Microflex LT.

4.1.4. Lancement de l'acquisition MALDI-TOF MS et interrogation de la base de données

- S'assurer que le logiciel d'acquisition FlexControl est ouvert pour créer le projet d'identification des bactéries. Ce logiciel est piloté par le programme MALDI BioTyper™ 2.0 (Bruker Daltonics) (figure 14).

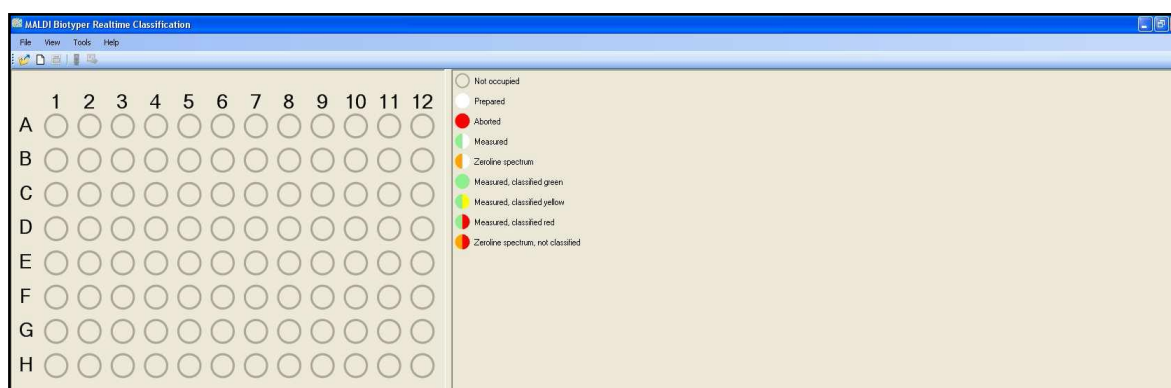


Figure 14 : Interface du logiciel MALDI BioTyper™ 2.0 (Bruker Daltonics).

- Sélectionner les spots à analyser par colonnes.
- Cliquer sur Finish lorsque le spectromètre est prêt (voyants verts). L'acquisition se lance automatiquement. Cliquer sur le logiciel d'acquisition FlexControl pour visualiser l'acquisition des spectres (figure 15).
- Les résultats s'affichent avec un code couleur (Figure 16).

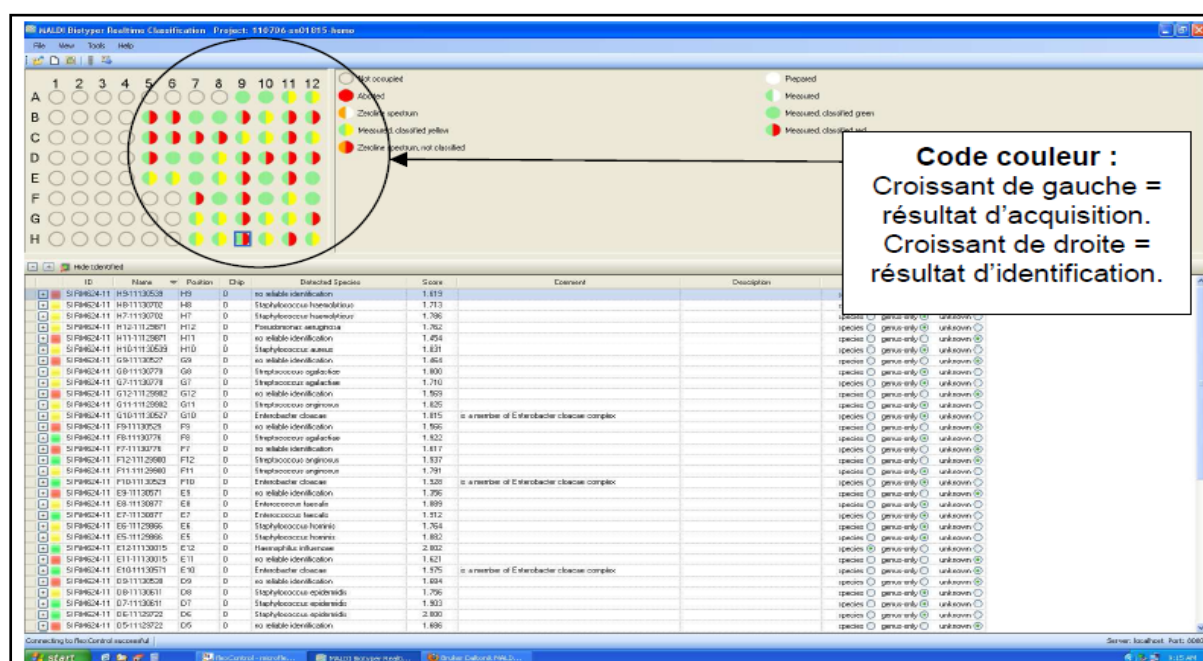


Figure 15 : Interface du logiciel d'acquisition FlexControl.










 Not occupied	Position non occupée sur la cible
 Aborted	Analyse annulée
 Zeroline spectrum	Spectre sans signal
 Measured, classified yellow	Spectre mesuré, identification jaune
 Zeroline spectrum, not classified	Spectre sans signal, identification rouge
 Prepared	Préparé
 Measured	Mesuré
 Measured, classified green	Spectre mesuré, identification verte
 Measured, classified red	Spectre mesuré, identification rouge

Figure 16 : Code couleur des analyses et identifications BioTyperTM 2.0 (Bruker Daltonics).

- Les spectres générés à partir de bactéries entières sont ensuite comparés aux spectres de référence présents dans la base de données BioTyper. Ces spectres de référence ont été obtenus à partir des souches types et d'isolats issus de collections internationales (Nagy *et al.*, 2009). La concordance des spectres obtenus à partir des isolats bactériens étudiés avec ceux des souches de référence est traduite par des scores (de 0 jusqu'à 3) qui indiquent le degré de confiance à accorder à l'identification (Lartigue *et al.*, 2009; Cherkaoui *et al.*, 2010 ; tableau 1).

Tableau 1 : Score de concordance des spectres obtenus à partir d'une bactérie d'intérêt avec ceux de la base de données Biotyper, reflétant le degré de confiance à accorder à l'identification.

Score	Description	Symbole
2,300 – 3,000	Forte probabilité d'identification à l'espèce	+++
2,000 – 2,299	Identification du genre sécurisée, identification à l'espèce probable	++
1,700 – 1,999	Identification au genre probable	+
0,000 – 1,699	Degré de confiance insuffisant pour l'identification	-

- Les résultats d'une cible sont exploitables seulement si les témoins T+ et T- sont valides :

- T+ doit correspondre à *Escherichia Coli* avec un score > 1.9,
- T- doit correspondre à un score sans identification < 1.7.
- Les résultats d'identification des bactéries sont, sous BioTyper, supérieur à 1,9 pour les 2 dépôts (même souche retrouvée).

- En fin, le logiciel BioTyper regroupe l'ensemble des spectres obtenus dans un dendrogramme.

4.2. Identification géotypique (amplification et séquençage de l'ADNr 16S)

Pour tous les isolats, le gène de l'ARNr 16S a été amplifié avec un jeu d'amorces : 16S_FD1 et 16S_rP2 (Weisburg *et al.*, 1991).

Les réactions d'amplification ont été réalisées dans un thermocycleur Biometra (LABGENE scientific Instruments, Archamps, France). Le mélange réactionnel d'amplification d'un volume final de 50 µl contient, pour 5 µL d'ADN extrait, 30.75 µl d'eau ultra pure ; 5 µl de tampon de PCR 10X contenant MgCl₂ 15mM (Applied biosystems, Courtaboeuf, France); 5 µl de désoxyribonucléotides-triphosphate (dNTP) (1,25mM) (dATP, dCTP, dGTP et dTTP); 1 µl d'amorce sens (FD1, 40nmol/l) ; 1 µl d'amorce anti-sens (rP2, 40nmol/l) ; 0.25 µl de Taq-ADN polymérase (Qiagen, Courtaboeuf, France).

Les résultats de l'amplification ont été révélés par électrophorèse (135V durant 15 minutes) sur un gel d'agarose à 1,5% réalisé dans du TAE 0,5 x coloré au bromure d'ethidium, visualisé sous U.V. L'ADN amplifié a été purifié en utilisant des plaques Multiscreen de PCR (Millipore, Molsheim, France).

Le cycle d'amplification consiste en une première étape de dénaturation de l'ADN à 95°C pendant 5 min, suivi d'un cycle de 30 sec à 95°C, 30 sec à 52°C et 5 min à 72°C. Ce cycle est répété 35 fois. Une étape d'élongation finale de 5 min à 72°C achève l'amplification par PCR. Les témoins négatifs contenant le mix d'amplification sont restés négatifs.

Ensuite nous avons procédé au séquençage du gène 16S. Les solutions de séquençage pour les amorces directes et reverses (tableau 2) contiennent 3µl de tampon (Big Dye V1, Applied Biosystems), 2µl de Big Dye, 5µl d'eau distillée et 1µl d'une solution d'amorce à 40nmol/l pour un volume final de 15µl. La réaction de séquençage comprend une étape de dénaturation initiale d'une minute à 95°C suivie de 25 cycles de dénaturation à 96°C pendant 10 secondes, hybridation pendant 5 secondes à 50°C et d'élongation à 60°C pendant 3 minutes. Les produits de séquençage sont purifiés en utilisant des plaques de sephadex (Amersham biosciences, Uppsala, Suède), centrifugées à 910g pendant 6 minutes et les produits de séquençage sont déposés sur des plaques de réaction de 96 puits Micro Amp Optical (Applied Biosystems).

Les séquences nucléotidiques d'ADNr 16S ont été corrigées à l'aide du logiciel ChromasPro 1.34 (Copyright (c) 2003-2006 par Technelysium Pty Ltd) et elles ont été déposées dans GenBank sous les numéros d'accèsion HG421009-HG421017.

Au final nous avons dressé un arbre phylogénétique montrant les relations phylogénétiques entre les isolats en utilisant le logiciel MEGA 6.0 (Tamura *et al.*, 2013).

Tableau 2 : Amorces utilisées pour l'amplification et le séquençage du gène de l'ARNr 16S (Adékambi et Drancourt, 2004).

	Primers	Sequence (5'-3')	Tm* (°C)
Amplification	FD1	AGAGTTTGATCATGGCTCAG	56
	rP2	ACGGCTACCTTGTTACGACTT	62
Sequencing	357F	TACGGGAGGCAGCAG	50
	357R	CTGCTGCCTCCCGTA	50
	536F	CAGCAGCCGCGGTAATAC	58
	536R	GTATTACCGCGGCTGCTG	58
	800F	ATTAGATACCCTGGTAG	48
	800R	CTACCAGGGTATCTAAT	48
	1050F	TGTCGTCAGCTCGTG	48
	1050R	CACGAGCTGACGACA	48

*: La Tm a été calculée en utilisant la formule suivante : $Tm = (2 \times AT + 4 \times GC) - 4$.

Partie 2 : Caractérisation de la diversité bactérienne par analyse de l'ADN extrait des sols (approche métagénomique)

1. Prélèvement des échantillons de sol

Nous avons réalisés des prélèvements dans la même région décrite précédemment (El Meridj), sauf que l'échantillonnage a été effectué dans trois sites différents : la forêt, le lac et le champ de blé suivant le même protocole détaillé dans la première partie.

2. Analyses physicochimiques des sols

La caractérisation physicochimique des trois échantillons de sol (sol forestier, sol cultivé et le sol du lac) a porté essentiellement sur la composition granulométrique, le pH, la teneur en calcaire et la teneur en matière organique. Toutes ces analyses ont été effectuées dans les laboratoires de l'ANRH de Constantine.

3. Approche métagénomique

3.1. Extraction des acides nucléiques

Les acides nucléiques ont été extraits selon le protocole du Kit " PowerSoil-DNA Isolation Kit (Mobio) " (annexe A). Cette extraction repose sur la combinaison d'une double lyse cellulaire : une lyse mécanique et une lyse chimique. De plus, le protocole intègre la purification des extraits, ce qui permet de passer directement à l'étape d'amplification par PCR.

3.2. Amplification d'ADNr 16S et pyroséquençage 454

Les amorces de fusion (Figure 17) sont des amorces particulières utilisées pour la constitution des bibliothèques d'amplicons destinées au pyroséquençage. Ce sont des amorces constituées des amorces spécifiques d'ADNr 16S, d'un marqueur de multiplexage (MID), d'une clé et d'un adaptateur spécifique du protocole de pyroséquençage.

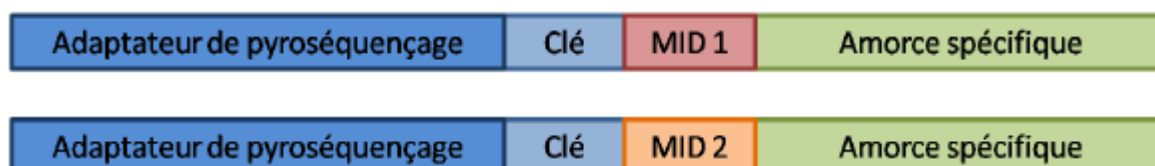


Figure 17 : Constitution des amorces fusionnées utilisées pour le pyroséquençage.

Les amorces de PCR (341F: CCTACGGGRSGCAGCAG et 909R: TTTCAGYCTTGCGRCCGTAC) ont été utilisées pour l'amplification du gène de l'ARNr 16S de la plupart des bactéries. Pour la bcPCR, de longs oligonucléotides ont été utilisés constitué des séquences d'amorces spécifiques d'ADNr 16S marqués avec les adaptateurs de séquençage pour GS FLX Titanium chemistry, qui ont été conçus comme recommandé par le fabricant (454 Life sciences) pour y inclure (5'-3'): adaptateur de titane, huit bases code à barres (uniquement sur l'amorce inverse) (tableau 3), une séquence de liaison (TA), et l'amorce d'ADNr 16S.

Tableau 3 : Liste des séquences code à barres (Hamady *et al.*, 2008).

Code à barres ID	Code à barres (5' -3')
709	CTGAACAC
755	CTTCTGCA
756	CTTCTGGT

Toutes les PCR ont été réalisées en volume de 50 μ l en utilisant 5 μ L d'ADN extrait, 30.75 μ l d'eau ultra pure ; 5 μ l MgCl₂ ; 5 μ l de déoxyribonucléotides-triphosphate (dNTP) (dATP, dCTP, dGTP et dTTP); 1 μ l de chaque amorce; 0.25 μ l de la Q5 Hot Start haute fidélité DNA polymérase de NEB (New England Biolabs). Pour une étape bcPCR, a été amplifié en utilisant des amorces à code-barres pour 20 cycles. Pour les deux étapes bcPCR, l'ADNr 16S a été amplifié en utilisant des amorces sans code à barres pour 20 cycles, suivi par une dilution au 1:50 du produit de PCR et 5 cycles supplémentaires d'amplification avec des amorces de PCR à code à barres. Toutes les réactions PCR ont été réalisées en triple exemplaire et sont ensuite rassemblées. Le programme thermique des deux PCR est composé d'une étape initiale de dénaturation (4 min à 95 ° C), un programme de cycles de 95 ° C pendant 30 s, 52 ° C pendant 30 s et 72 ° C pendant 60 s, enfin, une étape d'élongation (7 min à 72 ° C).

Pour le pyroséquençage, les amplicons, issus des deux PCR, ont été purifiés en utilisant des billes Agencourt AMPure (Beckman Coulter Genomics) et quantifiés avec un kit fluorescent à base de teinture (Quant-iT PicoGreen, Invitrogen). Les amplicons ont été séquencés du côté inverse. Les amplicons ont été séquencés sur un instrument GS FLX Titanium en utilisant la nouvelle technologie "454 Life sciences de Roche".

3.3. Traitement et analyse des données

Avant d'être analysées, les séquences ont été soumises à cinq étapes de traitement :

- Étape 1 : Les séquences ont été soumises à deux étapes successives de filtrage : (i) par les filtres des logiciels Roche (filtres lors du traitement du signal), (ii) par les filtres implémentés dans les outils de la base RDP (pas de nucléotides qui diffèrent dans les amorces, pas de nucléotides indéterminés dans les séquences, taille minimale de la séquence de 150 nucléotides, score Q de qualité supérieur ou égal à 20 (Cole *et al.*, 2008).
- Étape 2 : Les séquences ont été triées en fonction des échantillons grâce à la reconnaissance des marqueurs de multiplexage (Outils de RDP, Cole *et al.*, 2008).
- Étape 3 : Les séquences d'amorces spécifiques ont été taillées et des séquences < 200 pb ont été écartés de l'ensemble de données.
- Étape 4 : Les séquences chimères et les alignements aberrants ont été identifiées et écartées de l'ensemble de données (Edgar *et al.*, 2011).
- Étape 5 : Les séquences restantes ont été alignées à l'alignement de référence Greengenes 16S (DeSantis *et al.*, 2006) en utilisant le logiciel mothur (Schloss *et al.*, 2009).

Les mesures de la diversité (l'estimateur de la richesse "Chao1", l'inverse de "Simpson" et l'indice de diversité "Shannon") ont été calculées en utilisant le logiciel mothur (en regroupant les séquences alignées en OTUs à des distances différentes (pourcentage de similitude)).

En outre, les courbes de raréfaction ont été calculées pour chaque échantillon en utilisant les distances des OTUs : 0,03, 0,05 et 0,1. La classification taxonomique pour chaque lecture a été effectuée à l'aide du logiciel mothur basé sur la méthode bayésienne (Wang *et al.*, 2007). Comme base de données de référence de la taxonomie de référence, greengenes a été utilisée (mise à jour Octobre 2011). Des valeurs inférieures à un seuil de bootstrap de 70% ont été marquées comme non classés.

Résultats et discussion

Partie 1 : Le MALDI-TOF MS fournit un outil utile pour l'identification rapide des bactéries telluriques

L'identification bactérienne est essentielle pour confirmer l'espèce et orienter sa classification. En effet, les laboratoires de microbiologie de sol se font sentir le besoin de techniques d'identification de plus en plus courtes par rapport au temps et surtout fiables pour la confirmation de l'espèce. C'est ainsi qu'une gamme de techniques phénotypiques et moléculaires sont utilisées pour cet objectif qu'est l'identification des espèces bactériennes (Sintchenko *et al.*, 2007).

Le sol se voit profiter des nouvelles technologies moléculaires pour l'identification précise et rapide des bactéries dont le MALDI-TOF MS, introduit récemment dans la taxonomie bactérienne (Fenselau et Demirev, 2001 ; Lay et Liyanage, 2006 ; Van Baar, 2000) et appliquée avec succès à un bon nombre de taxons (Barbuddhe *et al.*, 2008 ; Fagerquist *et al.*, 2007 ; Grosse-Herrenthey *et al.*, 2008 ; Moura *et al.*, 2008 ; Vanlaere *et al.*, 2006 ; Vargha *et al.*, 2006).





1. Caractérisation des isolats bactériens

Neuf isolats bactériens, désignés par une codification allant de S1 à S9 ont été identifiés présomptivement (au rang taxonomique du genre) selon la démarche expérimentale suivante :

1.1. Approche phénotypique

L'identification phénotypique des bactéries repose dans un premier temps sur des tests d'orientation simples comme l'aspect des colonies (tableau 4) qui permet d'adapter dans un second temps le choix des tests phénotypiques à réaliser. C'est en confrontant ces premiers résultats morphologiques aux principaux caractères physiologiques, ainsi qu'aux caractères biochimiques que nous avons obtenu une première identification appréciable mais non définitive (tableau 5).

Tableau 4: Aspect macroscopique des colonies bactériennes isolées sur gélose au sang. (Les photos sont réalisées au laboratoire de cultures spéciales à la faculté de médecine de Marseille).

Isolats	Aspect macroscopique des colonies	Description
S1		<p>Les colonies sont lisses, brillantes, de structure homogène (type " smooth " ou S). Cet aspect peut évoluer après cultures successives pour donner des colonies à surface sèche rugueuse (type " rough " ou R).</p>
S2		<p>Les colonies sont petites, rondes et lisses.</p>
S3		<p>Les colonies sont rondes, lisses, blanches, opaques et atteignant 2 à 3 mm de diamètre.</p>
S4		<p>Les colonies sont grandes et leur couleur varie du gris terne à brun muqueux, caractérisées par une odeur forte et une apparence très humide au grain non colonisé.</p>



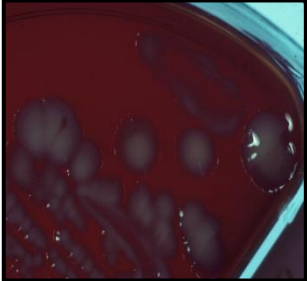


S5		<p>Les colonies sont de couleur blanche. Elles ont une surface brillante et un contour irrégulier.</p>
S6		<p>Les colonies sont de gris terne au marron, caractérisées par un aspect visqueux et une odeur moyennement à fortement déplaisante.</p>
S7		<p>Les colonies sont lisses et brillantes.</p>
S8		<p>Les colonies sont rondes à contour régulier, plates et visqueuses.</p>
S9		<p>Les colonies sont lisses, brillantes et possèdent une structure homogène (type "smooth" ou S).</p>

Tableau 5: Résultats de l'identification présomptive des isolats bactériens.

Isolats	Tests															Identification des isolats			
	Morphologiques		Physiologiques			Biochimiques													
	Gram	Mob ^a	ONPG ^b	OX ^c	Cat ^d	Lac ^e	Saccharose	Mannose	H ₂ S	Gaz	Uréase	Ind ^f	Cit ^g	RM ^h	VP ⁱ		ODC ^j	LDC ^k	ADH ^l
S1	-	+	+	-	+	+	+	+	-	+	+	-	+	-	+	+	-	+	<i>Enterobacter spp.</i>
S2	-	+	+	-	+	+	+	+	-	-	-	-	+	-	+	-	-	-	<i>Pantoea spp.</i>
S3	+	-	-	-	+	-	-	-	-	+	+	+	-	+	+	+	+	+	<i>Staphylococcus spp.</i>
S4	+	+	+	+	+	+	+	+	+	+	+	+	+/-	-	+	+	+	+	<i>Bacillus spp.</i>
S5	+	+	+	+	+	+	-	-	-	+	-	+	+/-	+	-	+	+	+	<i>Lysinibacillus spp.</i>
S6	-	+	-	+	+	+	+	-	-	-	-	-	+	-	-	-	-	+	<i>Pseudomonas spp.</i>
S7	-	+	+	-	+	+	+	+	-	+	-	-	+	-	+	+	+	-	<i>Enterobacter spp.</i>
S8	-	+	+	+	+	+	-	-	-	+	-	+	+	-	+	+	-	+	<i>Stenotrophomonas spp.</i>
S9	-	+	+	-	+	+	+	+	-	+	-	-	+	-	+	+	-	+	<i>Enterobacter spp.</i>

(+): Isolats positifs; (-): Isolats négatifs ; (+/-) : Réaction variable selon les souches.

(a): mobilité; (b): Ortho-Nitrophényl-β-galactoside; (c): Oxidase; (d): Catalase; (e): Lactose; (f): Production d'indole; (g): Citrate; (h): Rouge de méthyle; (i): Voges-Proskauer; (j): Ornithine décarboxylase; (k): Lysine décarboxylase; (l): Hormone anti-diurétique.

Cette méthode classique est encore admise comme étape primordiale pour l'identification présomptive des isolats bactériens (Carroll et Weinstein, 2007).

Comme le montre le tableau 2, tous les neuf isolats bactériens (100%) sont identifiés au niveau du genre seulement, ce qui n'est pas suffisant pour avoir une identification complète et précise. Malgré que cette méthode représente une des plus anciennes méthodes utilisées pour l'identification bactérienne tellurique (Dauga *et al.*, 2005) mais elle a montré plusieurs limites dont les principaux sont :

- La difficulté à déloger les bactéries à partir de particules de sol (Tabacchioni *et al.*, 2000),
- La sélection d'un milieu de croissance approprié,
- Le respect de conditions de croissance spécifiques (température, pH, lumière),
- Le potentiel d'inhibition de la propagation de colonies autres que celles d'intérêt et surtout l'incapacité de cultiver un grand nombre d'espèces bactériennes en utilisant les techniques disponibles à l'heure actuelle (Trevors, 1998) ce qui conduit à la description d'une très faible partie de la diversité bactérienne existante et à la sous-estimation même de la richesse du monde bactérien qui nous entoure.

Cependant, les tests morphologiques, physiologiques et biochimiques ne doivent pas être sous-évalués; ils ont encore un rôle irremplaçable dans l'obtention et la caractérisation de nouvelles souches. Leurs avantages importants comprennent le faible coût et la performance facile. Par conséquent, nous avons utilisé des méthodes moléculaires (comme décrit ci-dessous) pour compléter l'identification des isolats bactériens (au niveau de l'espèce).

1.2. Approche moléculaire

1.2.1. Identification protéomique (MALDI-TOF MS)

Les neuf isolats bactériens (S1 à S9) ont été soumis à une analyse de cellule entière de type MALDI-TOF MS. Cette méthode discrimine les bactéries sur la base du criblage de pics caractéristiques observés en tant que biomarqueurs pour l'identification des bactéries (Hollande *et al.*, 1996). Cette stratégie est améliorée par l'utilisation de plusieurs souches de référence pour chacune des espèces, qui doivent être incluses dans la base de données (Williams *et al.*, 2003 ; Chen, 2008).

Dans notre cas, le traitement de données a été réalisé par le logiciel MALDI BioTyper™ 2.0 (Bruker Daltonics) qui peut utiliser différentes banques de données pour produire, à partir des spectres, des listes de pics ou des références de spectres. L'identification des spectres inconnus des isolats bactériens a été réalisée à l'aide d'un algorithme de reconnaissance qui considère les positions de la totalité des pics dont l'intensité est comprise dans une échelle de 1 à 1000 (Bright *et al.*, 2002 ; Sauer, 2008 ; Freiwald et Sauer, 2009). Le logiciel génère automatiquement les listes de pics à partir de tous les spectres et extrait les pics typiques qui sont présents dans les spectres répertoriés pour une espèce donnée.

Les figures 18 à 26 représentent les spectres de masse recueillis des neuf isolats bactériens, réalisé par le logiciel MALDI Biotyper de Bruker.

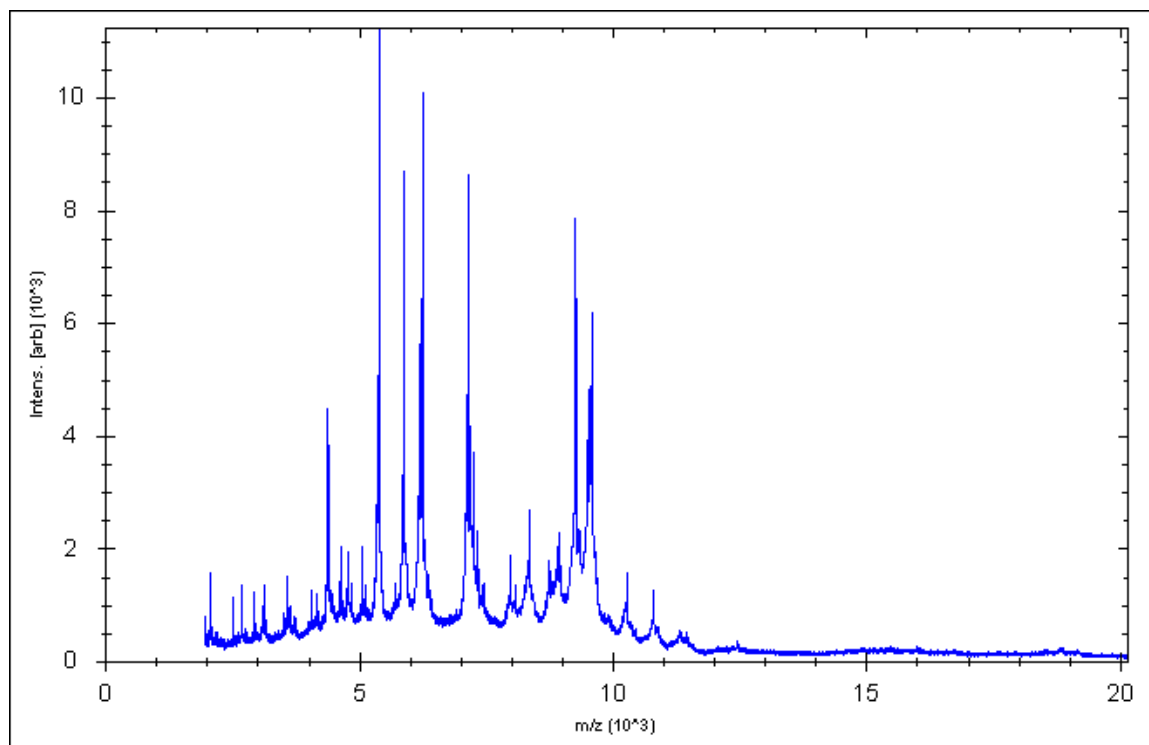


Figure 18 : Spectre de masse de l'isolat S1.

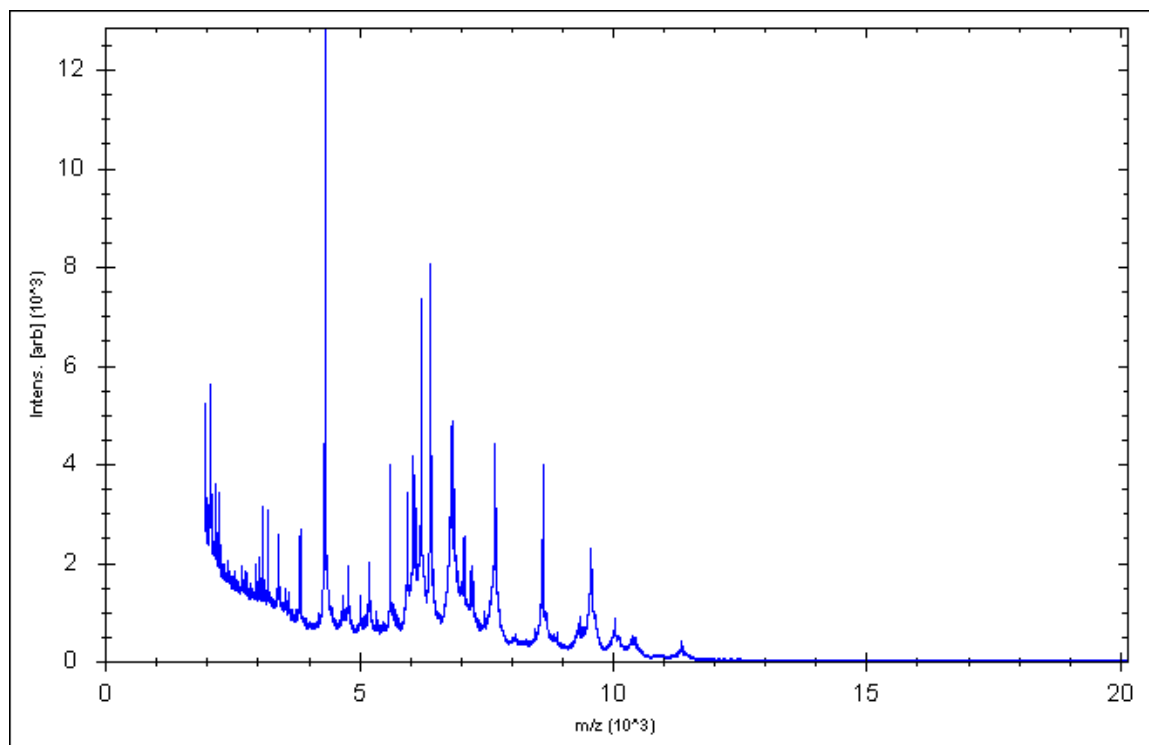


Figure 19 : Spectre de masse de l'isolat S2.

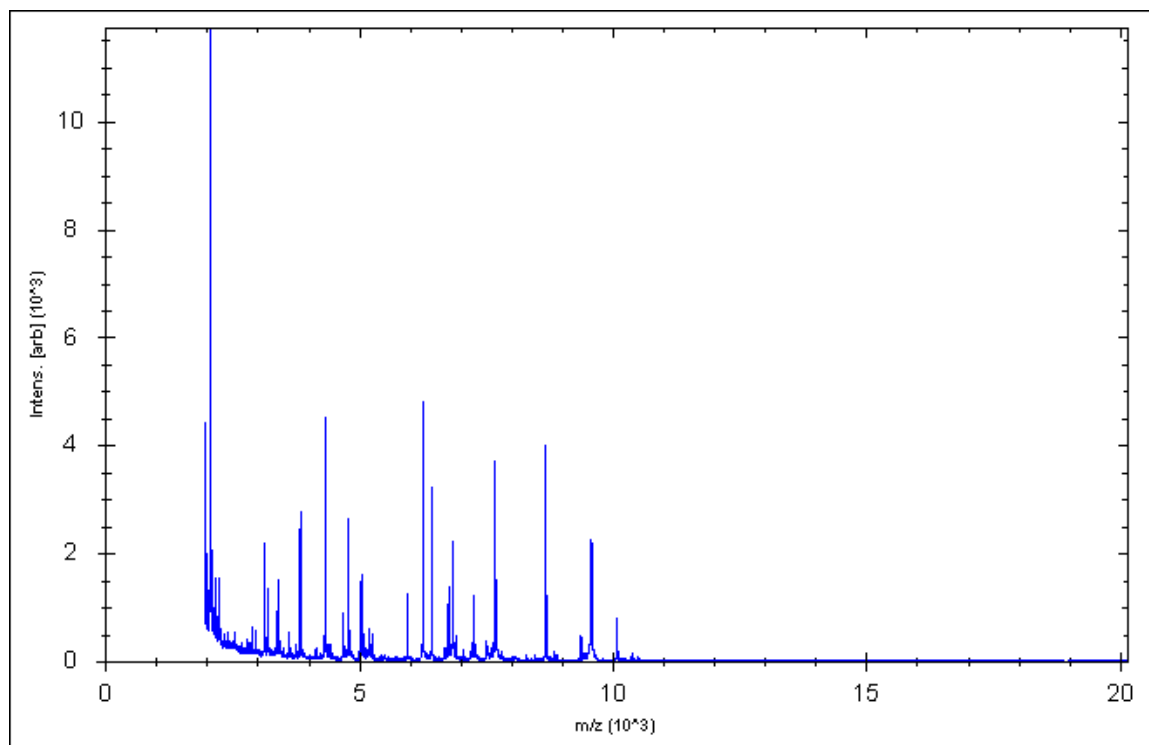


Figure 20 : Spectre de masse de l'isolat S3.

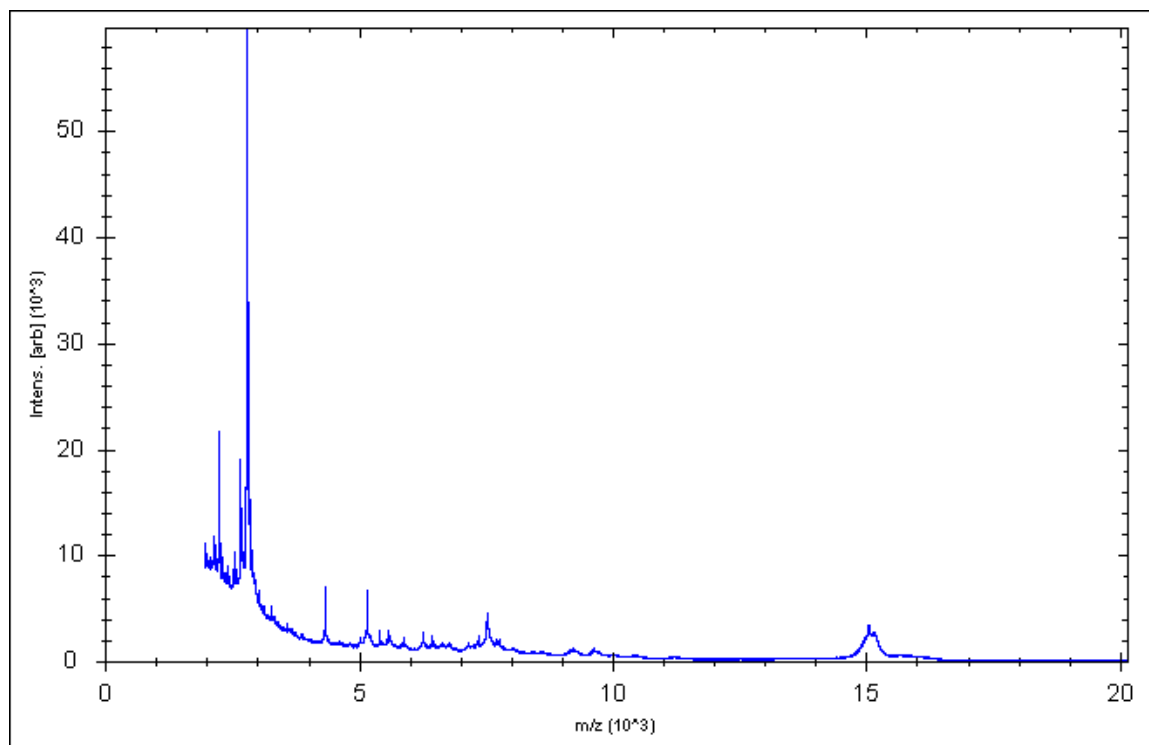


Figure 21 : Spectre de masse de l'isolat S4.

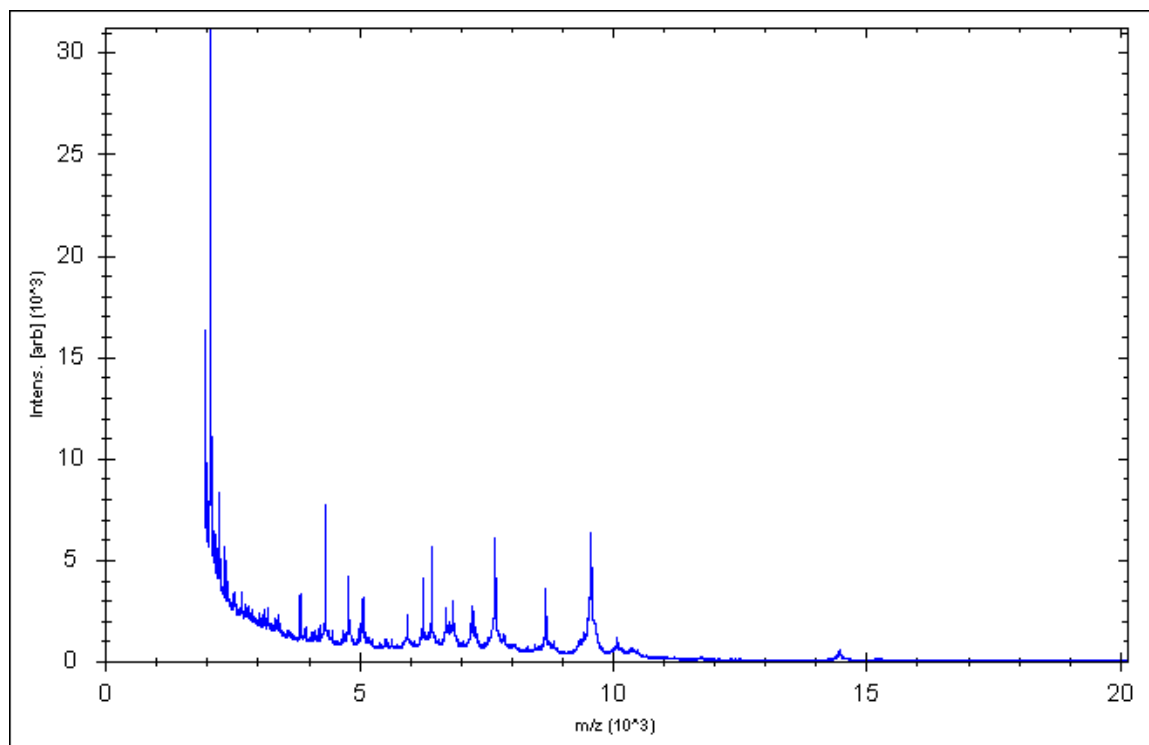


Figure 22 : Spectre de masse de l'isolat S5.

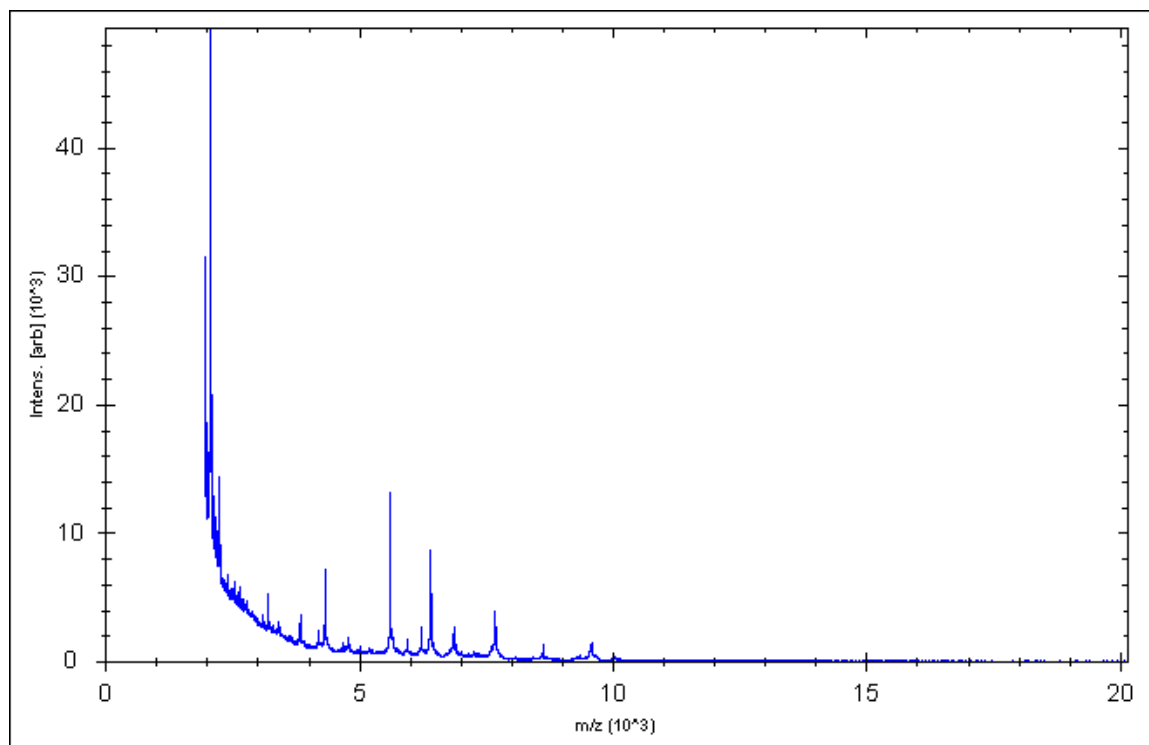


Figure 23 : Spectre de masse de l'isolat S6.

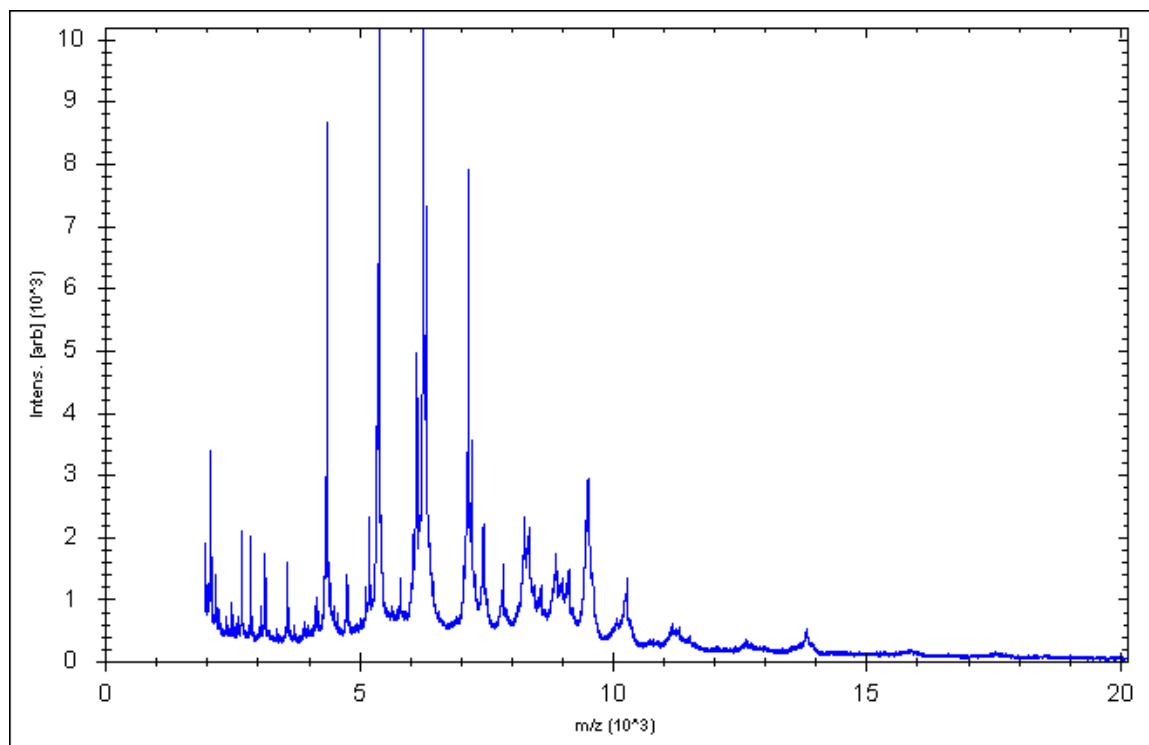


Figure 24 : Spectre de masse de l'isolat S7.

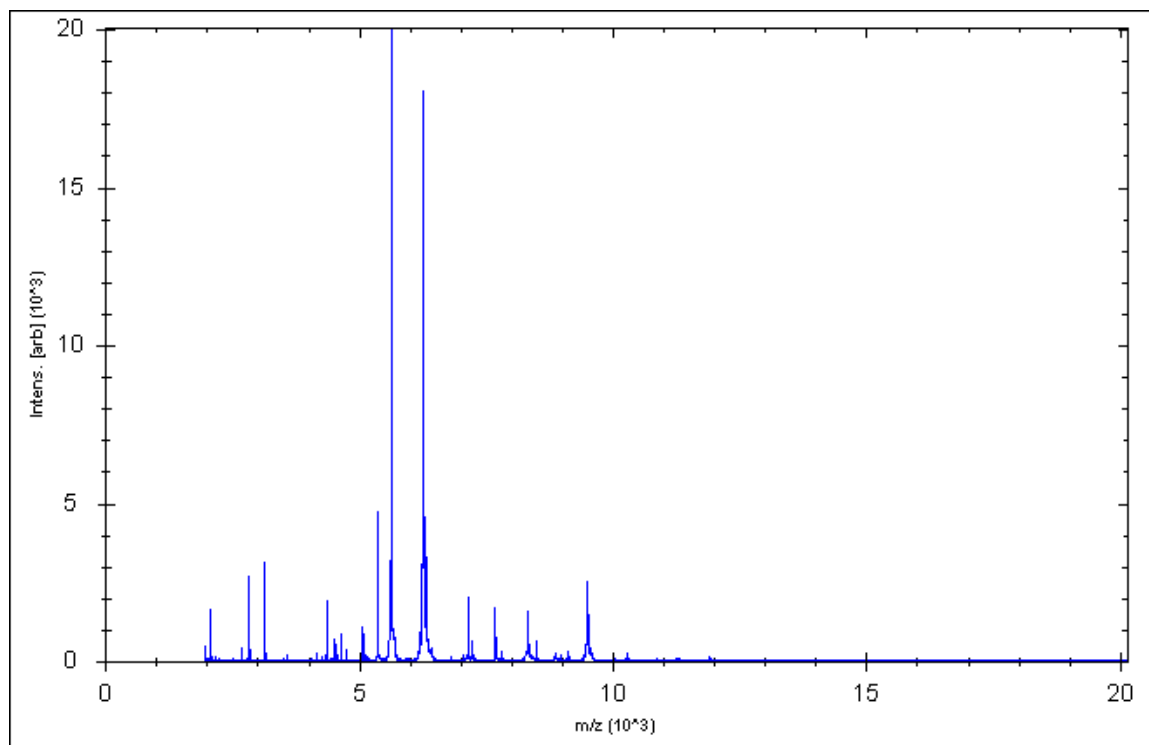


Figure 25 : Spectre de masse de l'isolat S8.

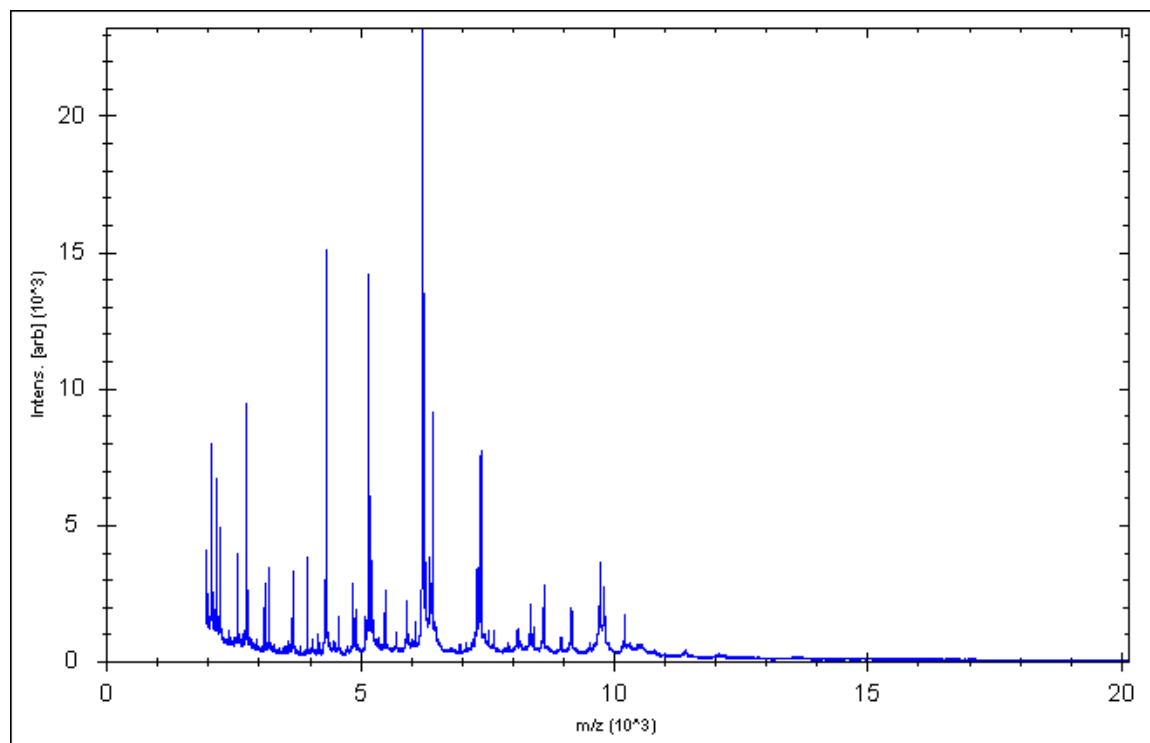


Figure 26 : Spectre de masse de l'isolat S9.

Un "matching score" ou score d'appariement et de fiabilité d'identification, basé sur les masses identifiées et la corrélation de leurs intensités, est généré et utilisé pour classer les résultats. Un score supérieur à 2,000 (ou compris entre 2,000 et 3,000) indique une bonne fiabilité et une forte probabilité de d'identification l'espèce (tableau 6). Le rang du genre, à ce score étant totalement sécurisé (Gidenn S, 2011 ; Blondiaux, 2013).

Tableau 6: Identification des isolats bactériens par MALDI-TOF MS.

Isolats	Identification selon MALDI Biotyper	Scores
S1	<i>Enterobacter hormaechei</i>	2,340
S2	<i>Pantoea vagans</i>	2,240
S3	<i>Staphylococcus caprae</i>	2,160
S4	<i>Bacillus cereus</i>	2,120
S5	<i>Lysinibacillus fusiformis</i>	2,115
S6	<i>Pseudomonas putida</i>	2,380
S7	<i>Enterobacter aerogenes</i>	2,300
S8	<i>Stenotrophomonas rhizophila</i>	2,202
S9	<i>Enterobacter cloacae</i>	2,320

Les spectres et les scores recueillis nous ont permis de construire un dendrogramme présenté dans la figure 27.

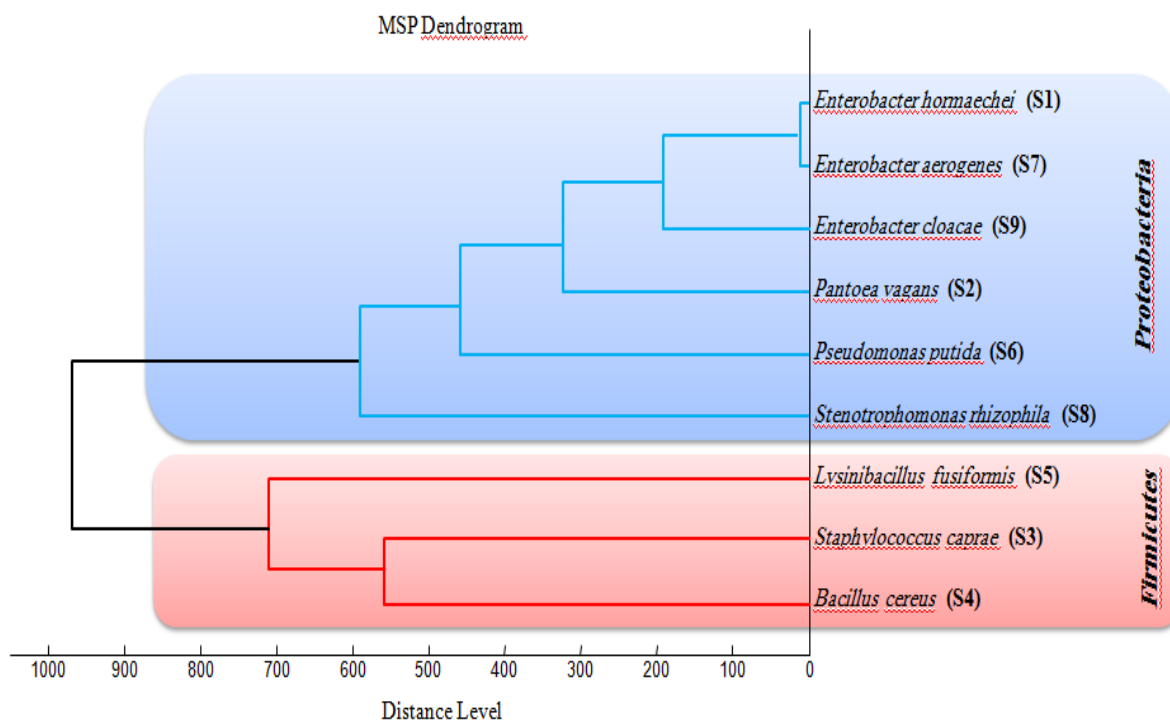


Figure 27 : Dendrogramme des neuf isolats bactériens réalisé par le logiciel MALDI BioTyper™ 2.0 (Bruker Daltonics).

Selon les données relevées sur le tableau 6 et sur la figure 27, nous déduisons trois points essentiels :

- Les neuf isolats bactériens ont été identifiés au rang de l'espèce grâce au MALDI-TOF (car score > 2,000). L'on peut conclure à ce stade que la spectrométrie de masse de type MALDI-TOF est bien établie comme étant une méthode précise et rentable pour l'identification des espèces bactériennes.
- Deux *phyla* ressortent de cette identification :
 - *Proteobacteria* représenté par les isolats S1, S2, S6, S7, S8 et S9
 - *Firmicutes* représenté par S3, S4 et S5.
- Chaque espèce appartient au même genre préalablement identifié par la méthode phénotypique ce qui implique l'existence d'une parfaite concordance entre les deux techniques. Cette concordance a été également observée chez d'autres auteurs où ils mettent en évidence les performances du MALDI-TOF comparés à certains systèmes d'identifications de référence. Blondiaux *et al.* (2013) retrouvent 264 isolats sur un

total de 362 (73%) en parfaite concordance entre la technologie MALDI-TOF et le système vitek 2. Dauxalder et al (2011), montrent une identification cohérente avec l'identification biochimique dans 94,4% (soit 305 isolats sur 323). Ils ajoutent que, seulement "quatre souches se révélèrent impossibles à identifier par le MALDI-TOF et 14 présentèrent des identifications discordantes".

En fait, depuis sa commercialisation au milieu des années 2000, le MALDI-TOF MS a fait couler beaucoup d'encre. Des études, comme celles de Bizzini et Greub (2010), ont démontré que les performances d'identification des bactéries par cette nouvelle technologie ont pu atteindre des pourcentages élevés variant de 88,8 à 98,2 % et de 75,6 à 97,7 % d'identification correcte à l'espèce pour les bactéries Gram-négatif et Gram-positif respectivement ; ce qui rassure sur l'utilisation du MALDI-TOF MS. De plus, cette technique a été utilisée dans des études spécifiques qui ont essentiellement évalué sa capacité à identifier les différentes espèces bactériennes chez les bacilles à Gram négatif tels que :

- *Pantoea vagans* et d'autres membres de la famille des entérobactéries (Conway *et al.*, 2001 ; Camara et Hays, 2007 ; Rezzonico *et al.*, 2010),
- des bacilles à Gram négatif tels que *Pseudomonas putida* (Degand *et al.*, 2008),
- des cocci à Gram positif, tels que *Staphylococcus aureus* (Edwards-Jones *et al.*, 2000 ; Kumar *et al.*, 2004),
- certains bacilles à Gram positif tels que *Bacillus cereus* (Ryzhov *et al.*, 2000 ; Demirev *et al.*, 2001 ; Barbuddhe *et al.*, 2008).

Ajoutons, enfin, que nombreuses études ont comparé la technique spectrométrique aux techniques utilisées traditionnellement dans les laboratoires de microbiologie, en particulier du sol. Toutes concluent que le MALDI-TOF MS surpasse les techniques conventionnelles tant dans la qualité de l'identification des bactéries, qu'elles soient courantes ou moins fréquentes, que dans le taux de mauvaise identification observé, généralement négligeable pour le MALDI-TOF MS (Martiny *et al.*, 2010 ; Cherkaoui *et al.*, 2011). Les résultats d'identification obtenus par le MALDI-TOF MS (95 % - 97,4 % d'identifications correctes) sont comparables, voire meilleurs, à ceux des méthodes phénotypiques (75,2 % - 92,6 %) (Basile *et al.*, 1998 ; Demirev *et al.*, 1999) malgré l'existence de quelques différences entre les deux techniques citées par Wieser *et al.* (2012).

Pour s'assurer de la fiabilité d'identification de nos isolats bactériens par MALDI-TOF MS, nous avons utilisé une méthode alternative basée sur le séquençage du gène qui code l'ARNr 16S (abusivement appelé ADNr 16S). Parmi les raisons du choix de cette dernière :

- Technique la plus utilisée pour l'identification bactérienne tellurique (Ward *et al.*, 1992 ; Chanama, 1999),
- Structure génique très conservée et longueur pouvant atteindre 1500 pb, l'ARNr 16S étant considéré comme étant le gène le plus apprécié pour l'identification des bactéries au niveau de l'espèce (Song *et al.*, 2003 ; Janda et Abbott, 2007),
- Les banques de données du gène 16S sont aujourd'hui très développées et très riches en séquences : jusqu'à 200 000 séquences sur la base de données EMBL (Barghoutti, 2011 ; Mizrahi-Man *et al.*, 2013).

1.2.2. Identification génotypique (amplification et séquençage de l'ADNr 16S)

Dans un premier temps, le gène de l'ARNr 16S des neuf isolats bactériens a été amplifié par PCR. Les résultats de chaque amplification ont été observés par électrophorèse sur gel d'agarose à 1,5 % (figure 28).

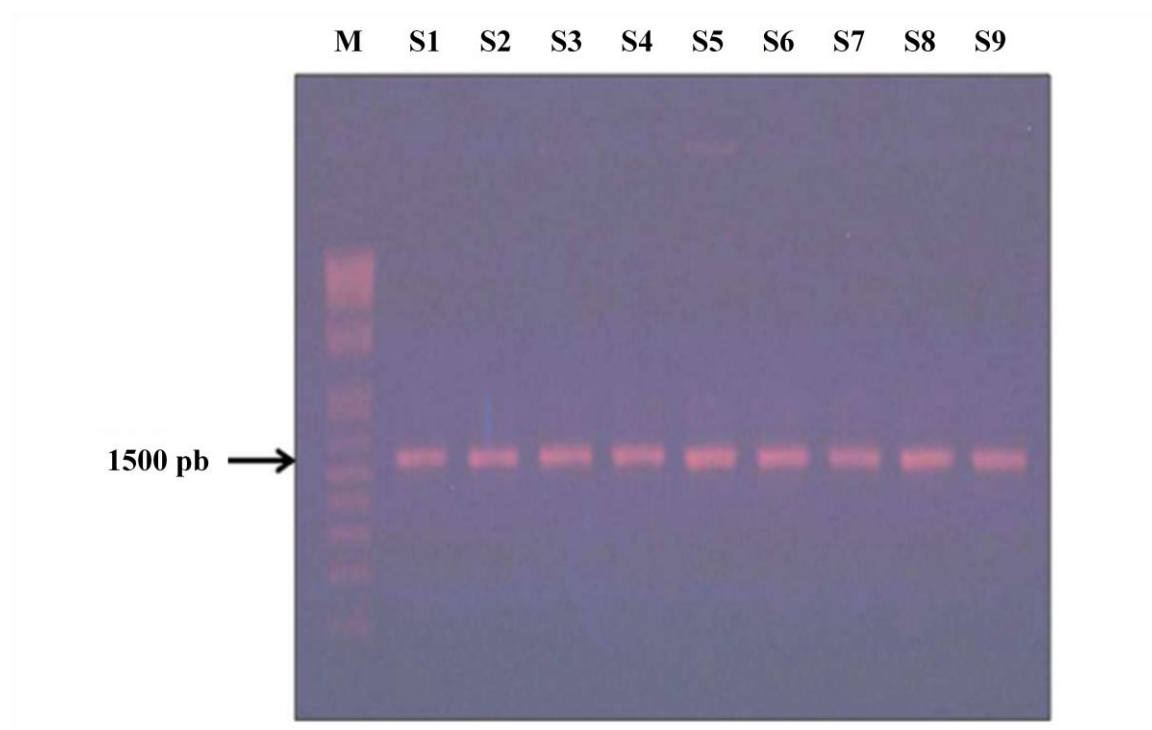


Figure 28 : Produits d'amplification PCR du gène d'ARNr 16S des neuf isolats bactériens, sur gel d'agarose à 1,5 %. M) marqueur moléculaire 1 kpb.

Une bande d'amplification unique à 1500 pb a été obtenue pour tous les isolats. Cette taille correspond bien à celle de l'ADNr 16S des bactéries, mise en évidence par Weisburg *et al.* (1991).

Les résultats du séquençage des neuf isolats peuvent être consultés en annexe B. Nous avons traité les séquences d'ADNr 16S au moyen d'outils bioinformatiques tels que Blastn qui nous a permis de comparer nos séquences à l'ensemble des séquences existantes sur GenBank. Une bonne identification est conséquente d'un pourcentage d'identité supérieur ou égal à 97 % entre notre souche et celle de la GenBank (Stackebrandt et Goebel, 1994 ; tableau 7).

Tableau 7 : Confrontation et correspondance biomoléculaire avec GenBank.

Isolats	Souches proposées par GenBank	score	E-value	identité
S1	<i>Enterobacter hormaechei</i> strain LRC5	2741	0.0	99 %
S2	<i>Pantoea vagans</i> C9-1	2734	0.0	99 %
S3	<i>Staphylococcus caprae</i> strain 11G	2771	0.0	99 %
S4	<i>Bacillus cereus</i> strain YR2-1	2804	0.0	100 %
S5	<i>Bacillus fusiformis</i> isolate LLP *	2789	0.0	99 %
S6	<i>Pseudomonas putida</i> W619	2760	0.0	99 %
S7	<i>Enterobacter aerogenes</i> strain CTSP48	2772	0.0	100 %
S8	<i>Stenotrophomonas rhizophila</i>	2793	0.0	100 %
S9	<i>Enterobacter cloacae</i> strain SB 3013	2745	0.0	99 %

* : Cette souche est transférée du genre *bacillus* au genre *Lysinibacillus* (Ahmed *et al.*, 2007).

GenBank, via le programme Blastn, réalise un alignement (concordance de l'enchaînement des nucléotides entre les séquences de GenBank et celle proposée dans notre étude) et propose une séquence qui présente les meilleurs bits score et pourcentage d'identité avec la nôtre.

Les valeurs du bit score mesurent l'identité entre les deux séquences : celle de GenBank et la nôtre. Si la valeur de % identité, renvoyée par le programme Blastn, est supérieure ou égale à 97%, cela indique que notre souche est identique à celle de GenBank et nous pouvons alors proposer à notre souche le nom systématique proposé pour celle de GenBank.

Les valeurs de E-value (ou Expect value) par contre, indiquent la probabilité de tomber sur une séquence autre que celle proposée par GenBank. Biologiquement, les bits scores

significatifs doivent avoir des E-values très inférieures à 1,0 ; voire nulles. Le calcul de cette valeur est fonction de trois paramètres :

- le bit score (S) : les grandes valeurs de S correspondent à de très petites E-values,
- la longueur de la séquence à comparer avec GenBank : les valeurs de E-value peuvent augmenter dans le cas des longues séquences
- la taille de la base de données GenBank : plus la taille de la banque (nombre de nucléotides) est importante, plus la valeur de E risque d'être élevée.

Les 9 souches alignées ont été identifiées comme suit :

1. S1 correspond à *Enterobacter hormaechei* strain LRC5 ,
2. S2 à *Pantoea vagans* C9-1,
3. S3 à *Staphylococcus caprae* strain 11G,
4. S4 à *Staphylococcus caprae* strain 11G,
5. S5 à *Lysinibacillus fusiformis* isolate LLP,
6. S6 à *Pseudomonas putida* W619,
7. S7 à *Enterobacter aerogenes* strain CTSP48,
8. S8 à *Stenotrophomonas rhizophila*
9. S9 à *Enterobacter cloacae* strain SB 3013.

Par conséquent, les mêmes espèces identifiées par la technique spectrométrique MALDI-TOF sont proposées par la méthode de séquençage ADNr 16s.

Les liens phylogénétiques, sous forme d'un arbre phylogénétique, entre les neuf séquences ADNr 16S sont mis en évidence au moyen du logiciel MEGA 6.0 (figure 29).

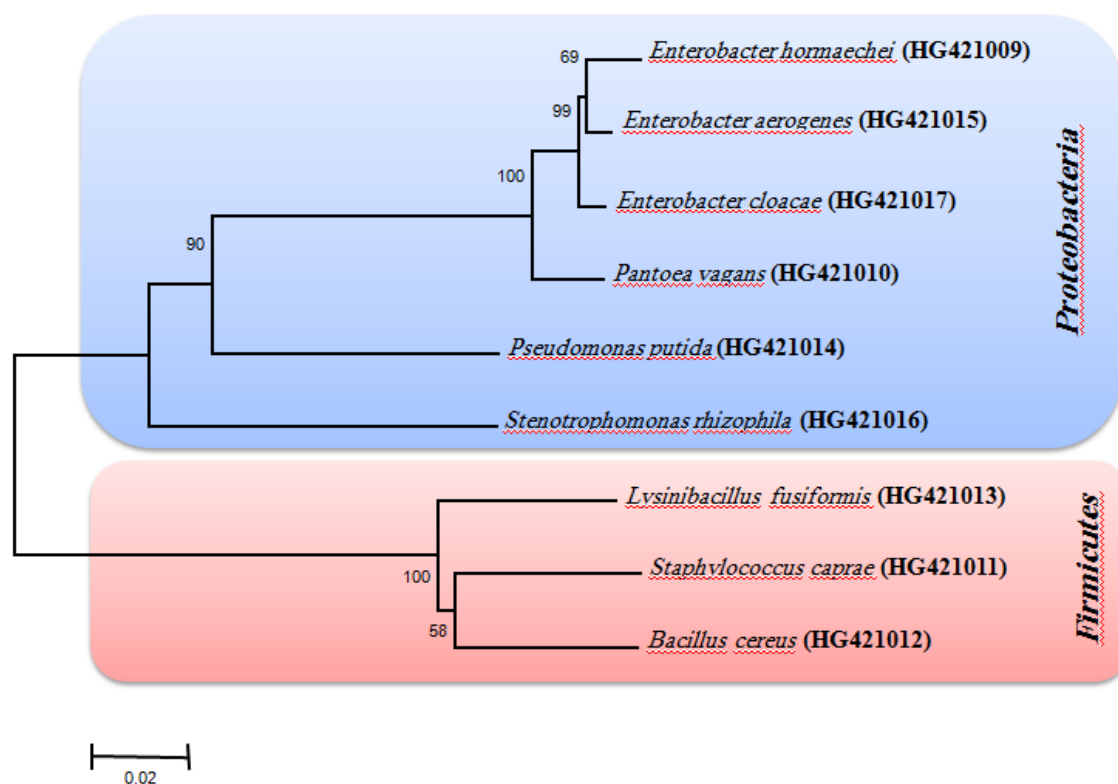


Figure 29 : Arbre phylogénétique construit avec la méthode Neighbor-joining montrant les liens phylogénétiques entre 9 séquences basées sur les gènes d'ARNr 16S des isolats bactériens. Les nombres figurant au niveau des nœuds indiquent les taux de Bootstrap (1000 réplicats), la barre 0.02 indique le nombre de substitution par position de nucléotide. Les numéros d'accèsion sont indiqués entre parenthèse.

Nous constatons que l'arbre phylogénétique a regroupé les espèces bactériennes dans deux clades correspondant aux deux *phyla* *Proteobacteria* et *Firmicutes*. La topologie de cet arbre est identique à celle de la figure 10, réalisée par le logiciel MALDI BioTyper™ 2.0 (Bruker Daltonics).

Par conséquent, nos résultats génotypiques sont en accord avec les résultats protéomiques. Le MALDI-TOF MS pourrait donc être une alternative intéressante aux techniques moléculaires basées sur le séquençage de l'ADNr 16S, généralement coûteuses. Ces constats sont en accord avec d'autres travaux (van Baar, 2000; Conway *et al.*, 2001; Fenselau et Demirev, 2001; Lay, 2001; Kumar *et al.*, 2004 ; Hettick *et al.*, 2006 ; Maier *et al.*, 2006 ; Dubois *et al.*, 2010 ; Bizzini *et al.*, 2011 ; Welker et Moore, 2011). Néanmoins, les résultats doivent être abordés avec une certaine prudence. Nous constatons que les bactéries à Gram-positif (S3, S4 et S5) fournissent moins de biomarqueurs que celles à Gram négatif (S1, S2, S6, S7, S8 et S9). Ce phénomène a également été décrit par d'autres groupes de recherche.

Smole *et al.* (2002) ont suggéré le traitement des échantillons avec le lysozyme pour faire face à cette limitation. Cependant, Grosse-Herrenthey *et al.* (2008) utilisent l'acide trifluoroacétique et l'acétonitrile ; un procédé qui a également été suggéré par Vargha *et al.* (2006).

De plus, en raison de sa rapidité, de la préparation facile de l'échantillon et de la taille réduite de ce dernier (Krishnamurthy *et al.*, 1996 ; Cherkaoui *et al.*, 2010), cette technique pourrait changer l'approche des laboratoires de microbiologie de sol et se positionner comme le test de première ligne, pour tous les isolats, à la place des techniques phénotypiques.

Partie 2 : L'approche métagénomique : accès à la diversité bactérienne tellurique et l'influence des caractéristiques physicochimiques

Les techniques modernes de séquençage à haut débit couplées à la métagénomique ont désormais la capacité de caractériser la diversité des communautés bactériennes d'une manière sans précédent. Ainsi, l'utilisation de ces technologies comme moyen d'étude des bactéries telluriques revêt un caractère tout à fait novateur à cette thèse, permettant ainsi d'accéder à des connaissances plus approfondies sur ces communautés bactériennes. Mais, tout d'abord, il faut essayer de comprendre l'environnement qui les entoure ; à ces fins, des analyses physicochimiques ont été réalisées.

1. Caractérisation des propriétés physicochimiques des trois types de sol

1.1. Couleur

Selon Webster et Butler (1976), Krishnan *et al.* (1980) et Stoner et Baumgardner (1981), la couleur des sols est une caractéristique clé qui a été utilisée de longue date pour l'identification, la détermination qualitative des caractéristiques des sols et leur classification. Notre échantillon est caractérisé par une couleur marron foncé (observée à l'œil nu) ce qui implique, d'après Schultze *et al.* (1993), une teneur élevée en matière organique, le classant comme un sol fertile.

1.2. Répartition granulométrique

Selon Soltner (2000), Pansu et Gautheyrou (2003), cette analyse a pour but de définir la texture du sol. Le pourcentage de ses différents éléments constitutifs explique ses propriétés physiques et mécaniques liées à la teneur en eau et son mouvement, à son comportement vis à vis de l'air, des racines et des bactéries. Le tableau 8 montre clairement la répartition granulométrique des trois échantillons de sol.

Tableau 8: La répartition granulométrique des trois échantillons de sol.

Composés minéraux	Teneur (%)		
	Sol forestier	Sol du lac	Sol cultivé
Argile	25	32	27
Limon fin	30	32	35
Limon grossier	20	15	15
Sable fin	8	6	8
Sable grossier	12	5	10

Le triangle de texture retenu est celui du GEPPA simplifié par Baize (1988). Il se présente sous la forme d'un triangle rectangle isocèle dont les côtés représentent les teneurs en argile (axe vertical), limon (axe horizontal) et sable (axe oblique), chaque axe étant gradué de 0 à 100%.

Nous devons positionner dans ce triangle trois points correspondant à la répartition granulométrique des trois sols afin de déduire leurs textures (annexe C). Par conséquent, Le sol forestier et le sol cultivé possèdent une texture LAS (limon argilo-sableux) tandis que le sol du lac à une texture ALS (argile limono-sableuse).

1.3. Teneur en calcaire (CaCO₃)

Le sol forestier possède 7,98 % du calcaire total, le sol du lac possède 13,30 % et enfin le sol cultivé possède 9,12 %. En comparant ces valeurs à celles signalées par Baize (1988) (tableau 9), nous constatons que nos trois sols sont modérément calcaires.

Tableau 9 : Norme d'interprétation du taux du calcaire du sol (Baize, 1988).

Taux du calcaire total	< 1%	1 à 5 %	5 à 25 %	25 à 50 %	50 à 80 %	> 80 %
Appréciation	Non calcaire	Peu calcaire	Modérément calcaire	Fortement calcaire	Très fortement calcaire	Excessivement calcaire

1.4. Teneur en matière organique

Les deux sols (forestier et cultivé) possèdent la même teneur en matière organique (3,440 %) tandis que le sol du lac possède 2,374 % de matière organique. Selon Baize (1988), un sol est considéré comme riche en matière organique si :

$$\frac{\text{matière organique (\%)} * 100}{\text{argile (\%)}} > 7$$

Par conséquent, l'application de cette formule nous donne les valeurs suivantes : 13.76 pour le sol forestier ; 12,74 pour le sol cultivé et 7,41 pour le sol du lac. Ce qui implique que les trois types de sol sont riches en matière organique (valeur > 7). Soltner (2000) affirme que plus le sol est calcaire plus il bloque la matière organique, ce qui est, exactement le cas de nos échantillons.

1.5. pH

Le pH des trois échantillons de sol est de 7,8. Selon Huetz De Lempis (1970), le pH est un facteur complexe qui représente en fait un véritable bilan de l'activité chimique du sol. Il est exprimé par la concentration des ions H_3O^+ dissociés en ions libres dans les solutions aqueuses du sol.

En fonction de l'échelle d'interprétation du pH signalé par Gagnard *et al.* (1988) (tableau 10), les trois sols sont alcalin.

Tableau 10 : Echelle d'interprétation du pH (Gagnard *et al.*, 1988).

pH	< 5,5	5,5 à 6,5	6,5 à 6,8	6,8 à 7,2	7,2 à 7,5	7,5 à 8,5	> 8,5
Appréciation	Fortement acide	Acide	Très légèrement acide	Voisin de la neutralité	Légèrement alcalin	alcalin	Fortement alcalin

D'après tout ces résultats, nous pouvons déduire que les trois types de sol (forestier, cultivé et le sol du lac) possèdent globalement les mêmes caractéristiques physicochimiques.

2. Mesure de la diversité bactérienne tellurique

Il est courant de vouloir estimer la diversité des espèces bactériennes par analyse de leurs métagénomomes par l'emploi de différentes moléculaires. Ces outils sont basés sur l'analyse de gènes spécifiques qui jouent le rôle de marqueurs phylogénétiques dont les plus couramment utilisés sont les gènes qui codent pour l'ARNr 16S chez les bactéries.

2.1 Assignation des séquences d'ADNr 16S des bactéries

Nous avons analysé un jeu de séquences issues des données du pyroséquençage 454 de l'ADNr 16S bactérien provenant des 3 échantillons de sol (tableau 11).

Tableau 11 : Les résultats du pyroséquençage 454 d'ADNr 16S bactérien.

Echantillons de sol	Nombre total de séquences	Longueur moyenne des séquences (pb)	Nombre des alignements aberrants	Nombre des séquences chimères	Nombre de séquences utilisées pour les analyses taxonomiques	Longueur moyenne des séquences (pb)
Sol du lac	9327	511	1119	28	8180	246
Sol forestier	9241	510	1100	32	8109	283
Sol cultivé	10227	511	1585	42	8640	286

Le nombre de séquences varie légèrement entre les trois échantillons de sol. Cependant, les répartitions du nombre de séquences en fonction des échantillons présentent des tendances similaires dont le nombre de séquences obtenues pour le sol cultivé est un peu plus élevé que pour le sol du lac et le sol forestier (8640 contre 8180 et 8109 séquences). L'identification de ces séquences révèle la présence de : 11 *Phyla* (figure 30), 20 familles (figure 32 et 33) et 13 genres (figure 34 et 35).

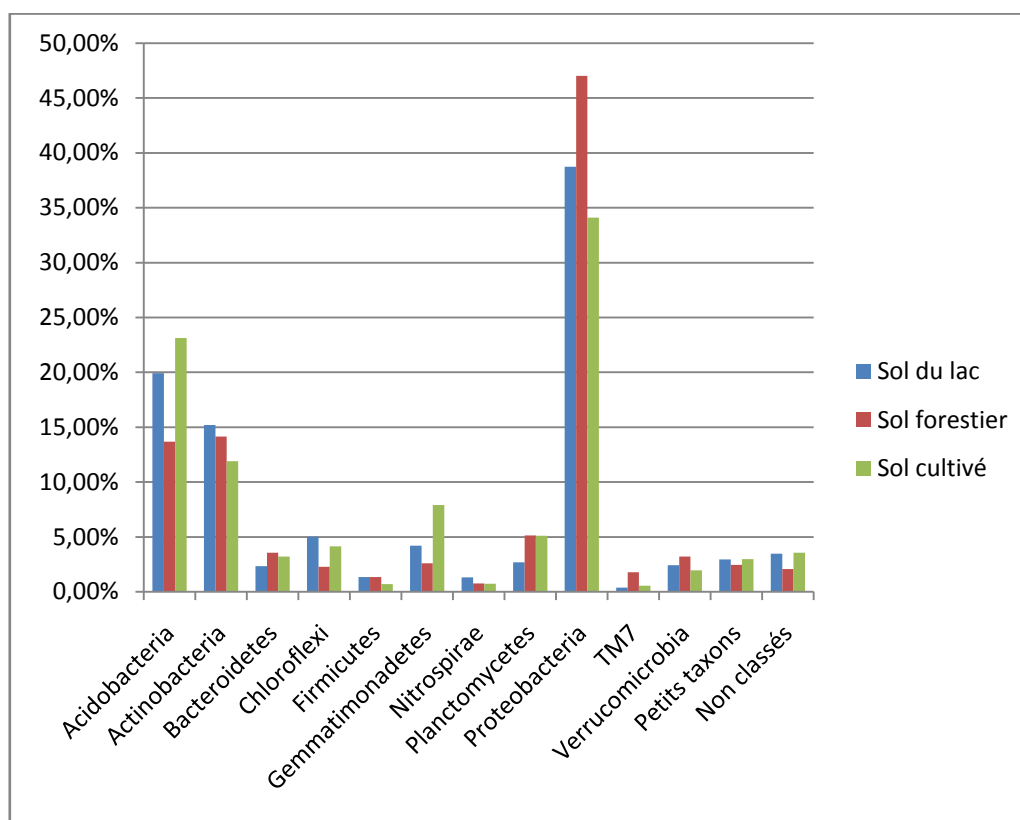


Figure 30 : Histogramme montrant les pourcentages des séquences identifiées au niveau du *phylum* dans les trois échantillons de sol.

Le *phylum* des *Proteobacteria* ressort nettement et domine les trois échantillons de sol (47,03% de la totalité des séquences pour le sol forestier, 38,75% pour le sol du lac et 34,10% pour le sol cultivé). D'autres *phyla* sont également retrouvés avec des proportions importantes :

- *Acidobacteria* (23,14 % sol cultivé, 19,91 % sol du lac et 13,67 % sol forestier)
- *Actinobacteria* (15,21 % sol du lac, 14,16 % forestier et 11,89 % cultivé).

Enfin, un faible pourcentage des séquences obtenues ne peuvent être assignées à aucun des *phyla* connus aujourd'hui (les *phyla* non classés représentent 3,46 % de séquences pour le sol du lac, 2,07 % pour le sol forestier et 3,57 % pour le sol cultivé). La carte phylogénétique (heatmap) illustre mieux ces pourcentages (figure 14).

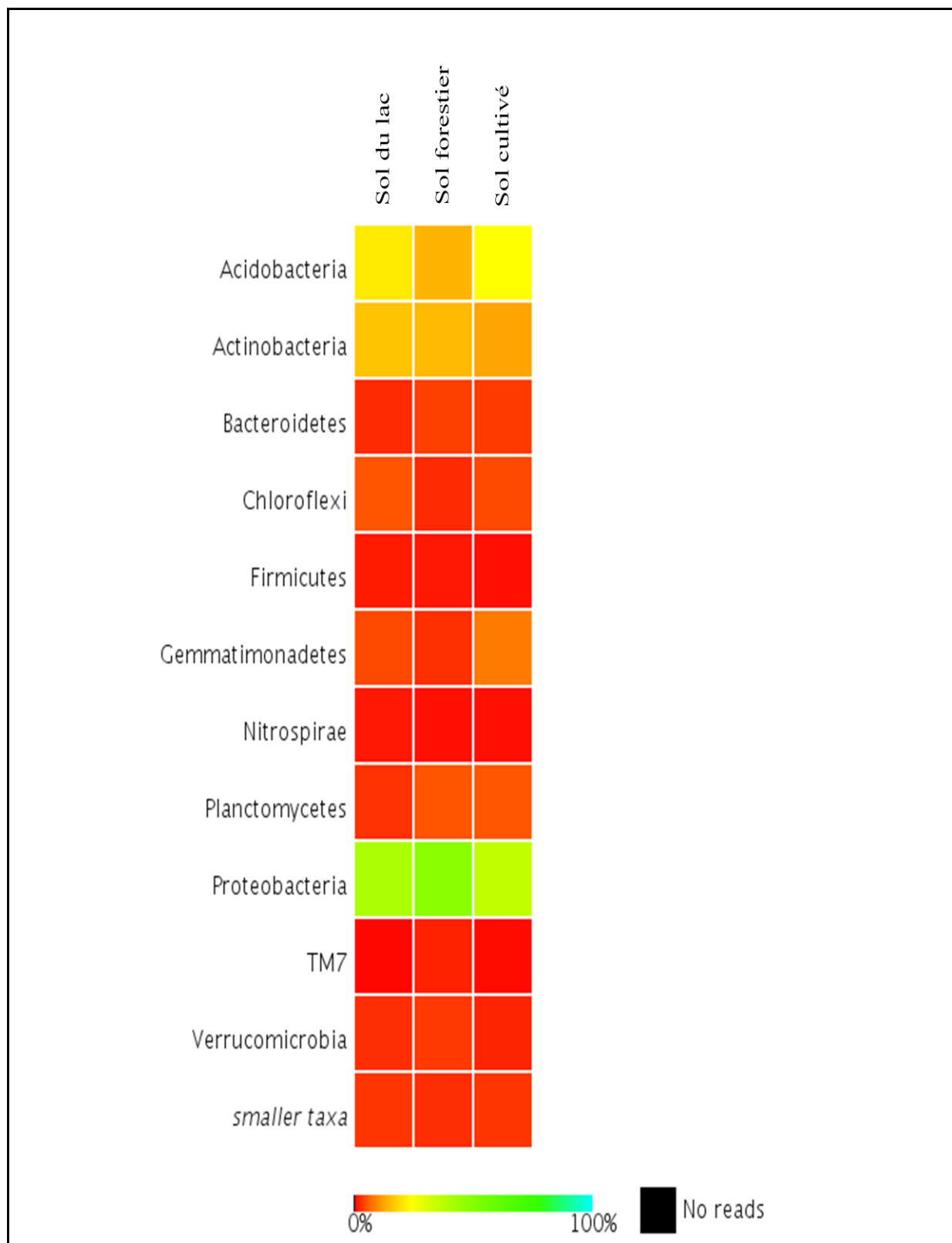


Figure 31: Carte phylogénétique ("heatmap") des *phyla* identifiés dans les trois sols.

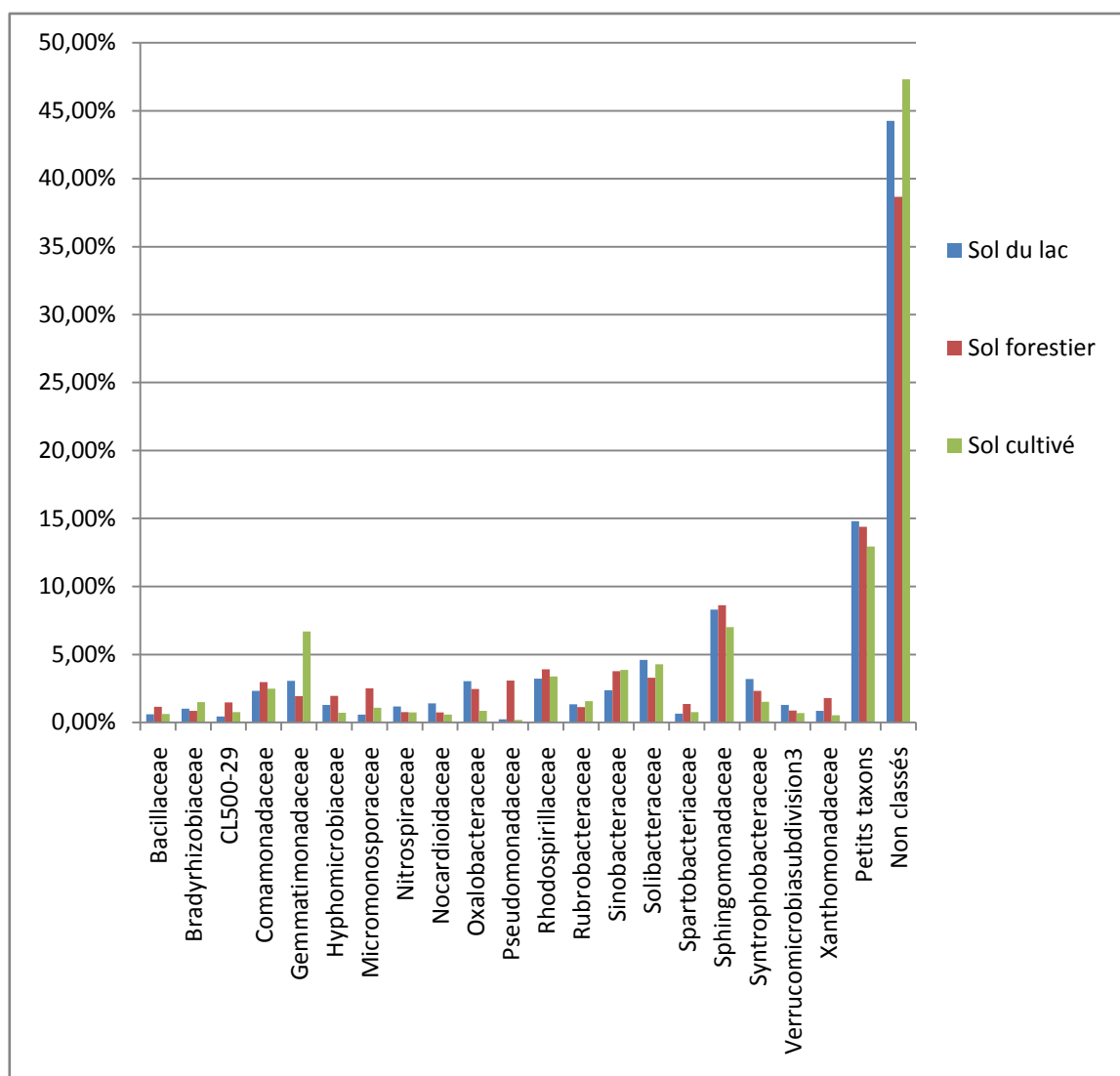


Figure 32: Histogramme montrant les pourcentages des séquences identifiées au niveau de famille dans les trois échantillons de sol.

Les familles non classées sont les plus dominantes dans les trois sols avec un pourcentage de 47,31 % pour le sol cultivé, 44,25 % pour le sol du lac et 38,68 % pour le sol forestier. Les petits taxons ont également un pourcentage élevé (14,39 % pour le sol forestier, 14,80 % pour le sol du lac et 12,93 % pour le sol cultivé). La famille des *Sphingomonadaceae* est la plus abondante parmi les taxons classés avec 8,63 % pour le sol forestier, 8,29 % pour le sol du lac et 7,01 % pour le sol cultivé.

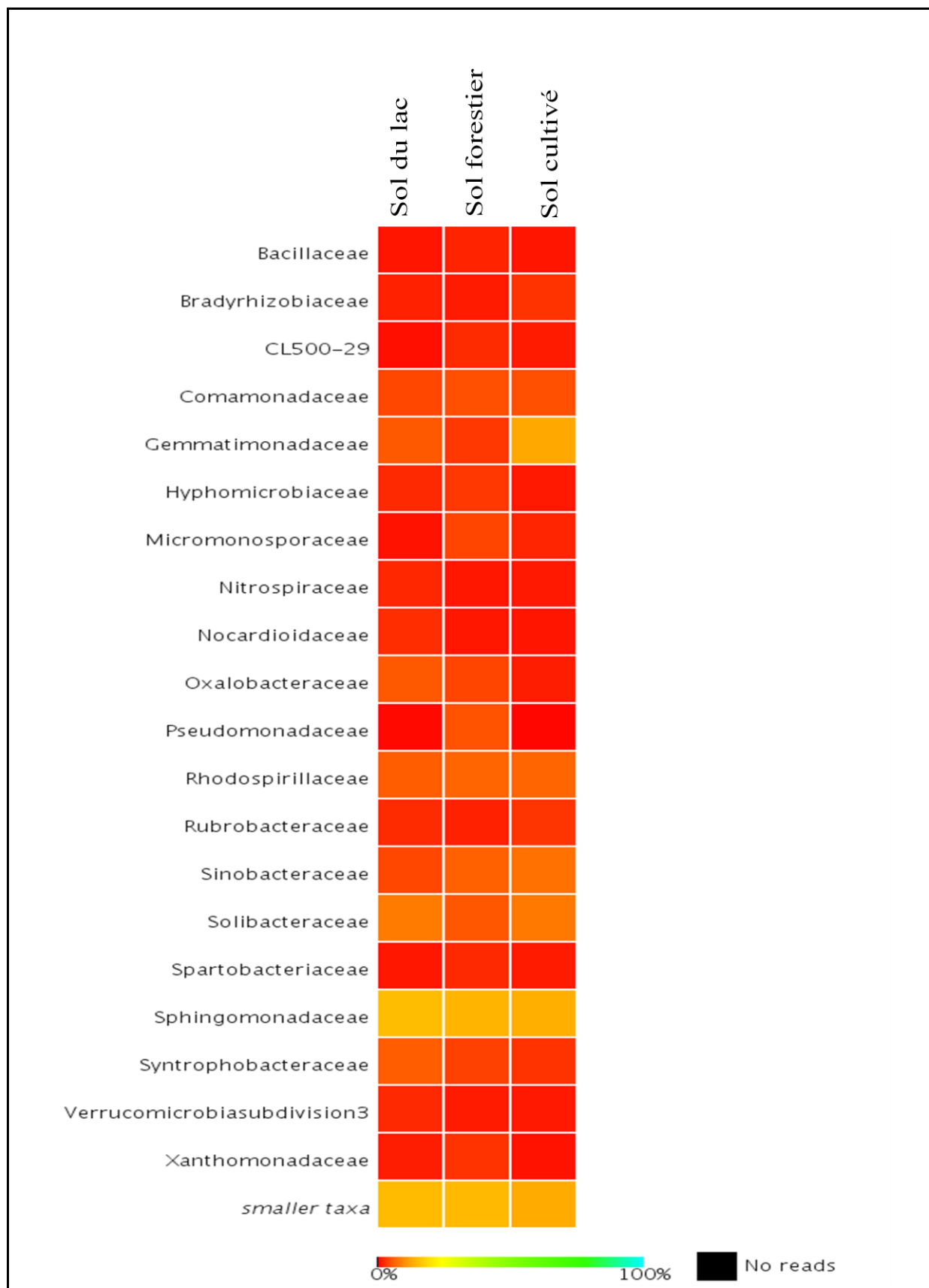


Figure 33: Carte phylogénétique ("heatmap") des familles identifiées dans les trois sols.

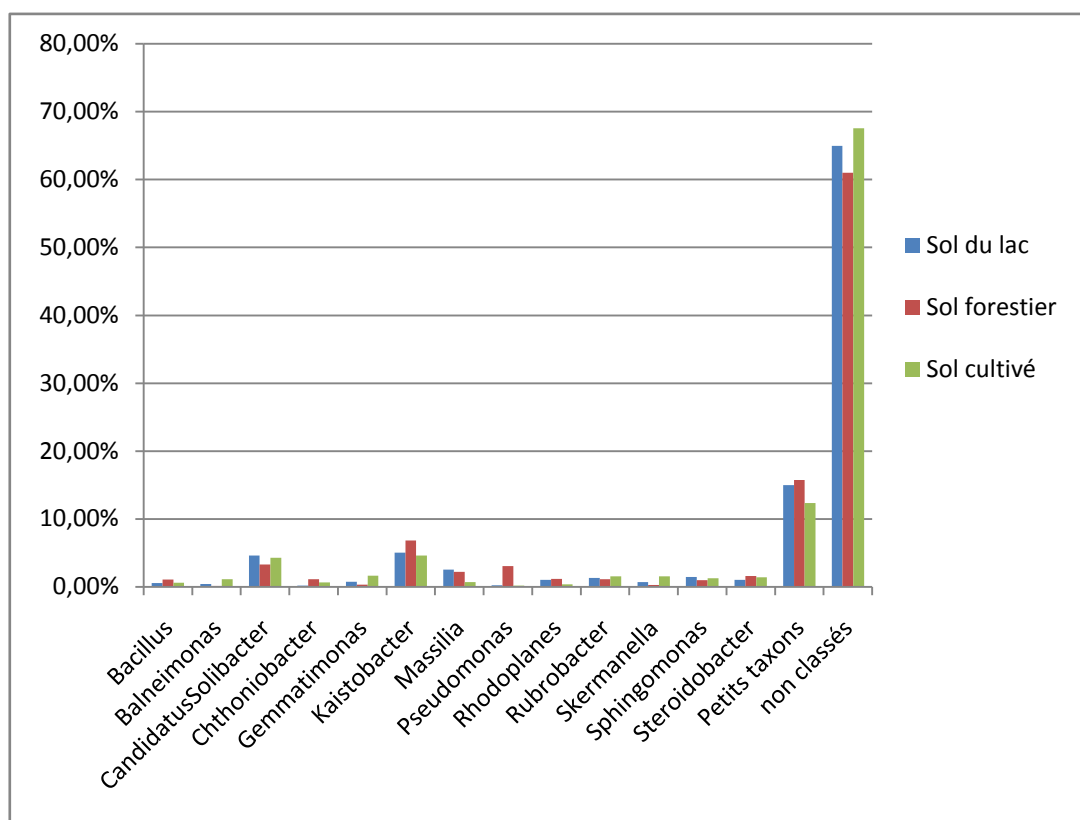


Figure 34: Histogramme montrant les pourcentages des séquences identifiées au niveau du genre dans les trois échantillons de sol.

Les genres non classés sont les plus dominants dans les trois échantillons de sol (67,57 % pour le sol cultivé, 64,98 % pour le sol forestier et 61,01 % pour le sol du lac). Les petits taxons sont en deuxième position de dominance avec des pourcentages élevés (environ 15 % pour les trois sols). Parmi les genres classés, *Kaistobacter* et *CandidatusSolibacter* sont les plus dominants dans les trois sols.

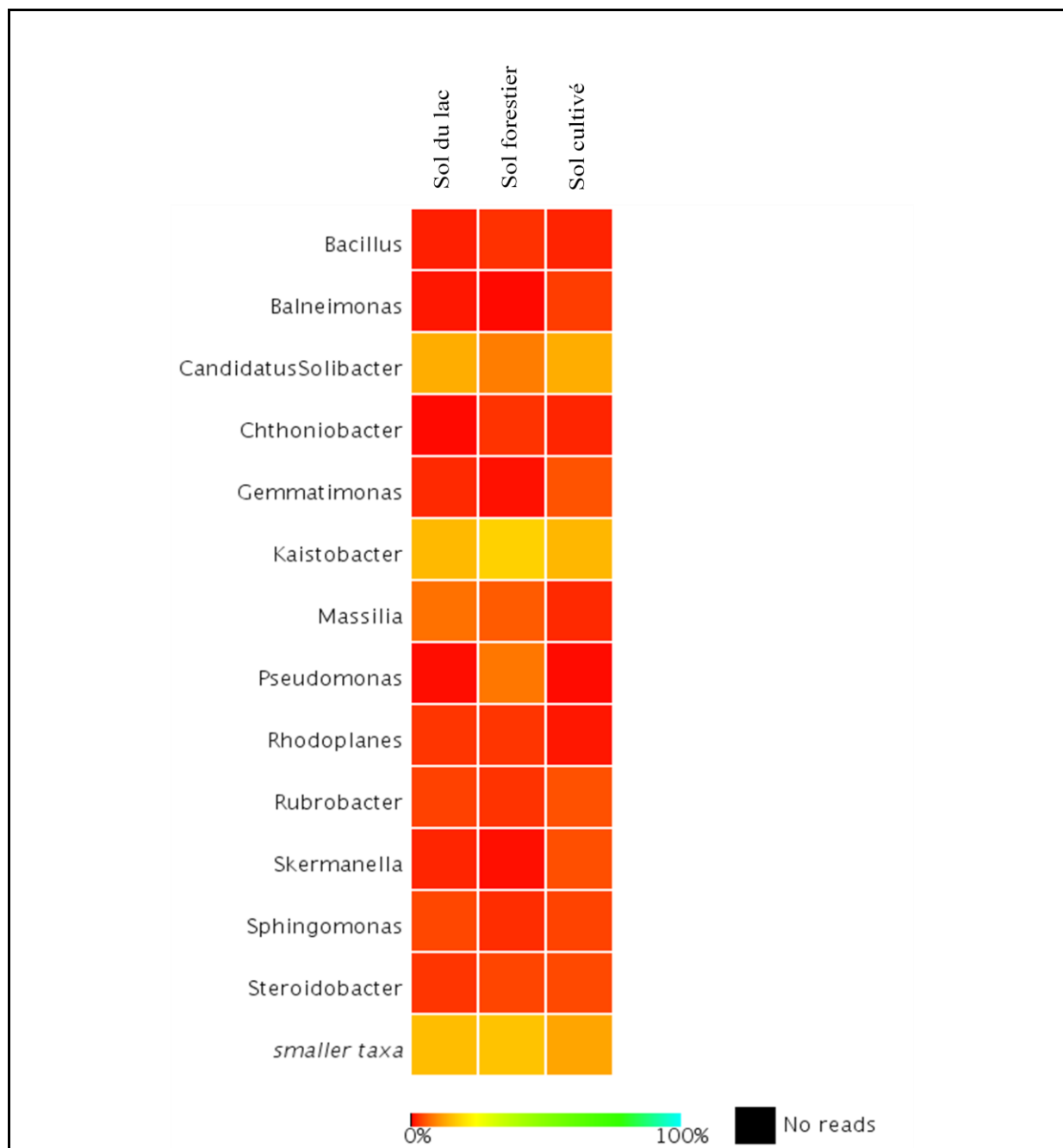


Figure 35 : Carte phylogénétique ("heatmap") des genres identifiés dans les trois sols.

2.2. Mesure de la diversité bactérienne

2.2.1. La courbe de raréfaction

Les courbes de raréfaction (Figure 36) montrent une saturation forte de la diversité bactérienne en termes d'OTU pour les trois types de sol (le nombre d'OTU augmente de 500 unités quand le nombre de séquences passent de 4000 à environ 8000), ce qui signifie que l'effort de séquençage appliqué est suffisant pour envisager la diversité des communautés bactériennes globale présente dans les différents échantillons analysés.

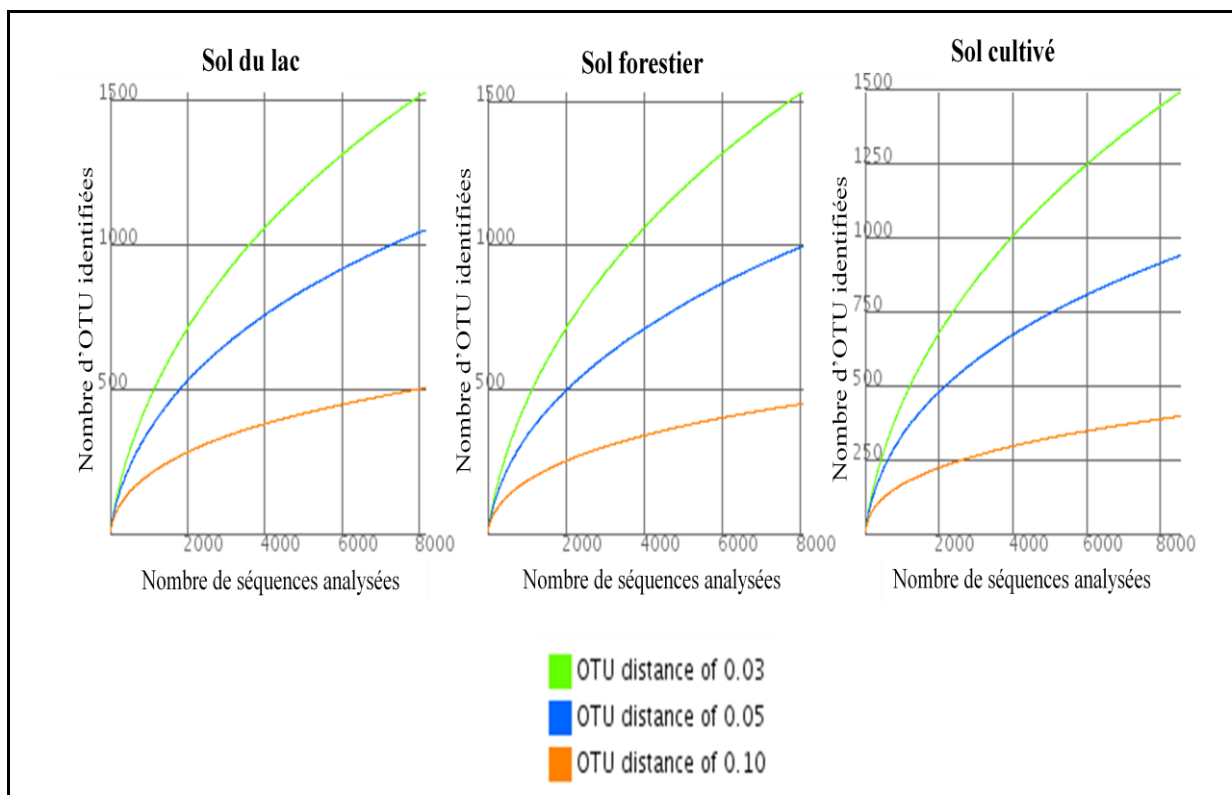


Figure 36: Courbes de raréfaction (nombre d’OTU identifiées par rapport au nombre de séquences analysées) pour les trois échantillons de sol.

L’histogramme ci-dessous (figure 37) montre que le nombre des OTUs est le même pour les trois échantillons de sol (environ 1500 OTU par échantillon).

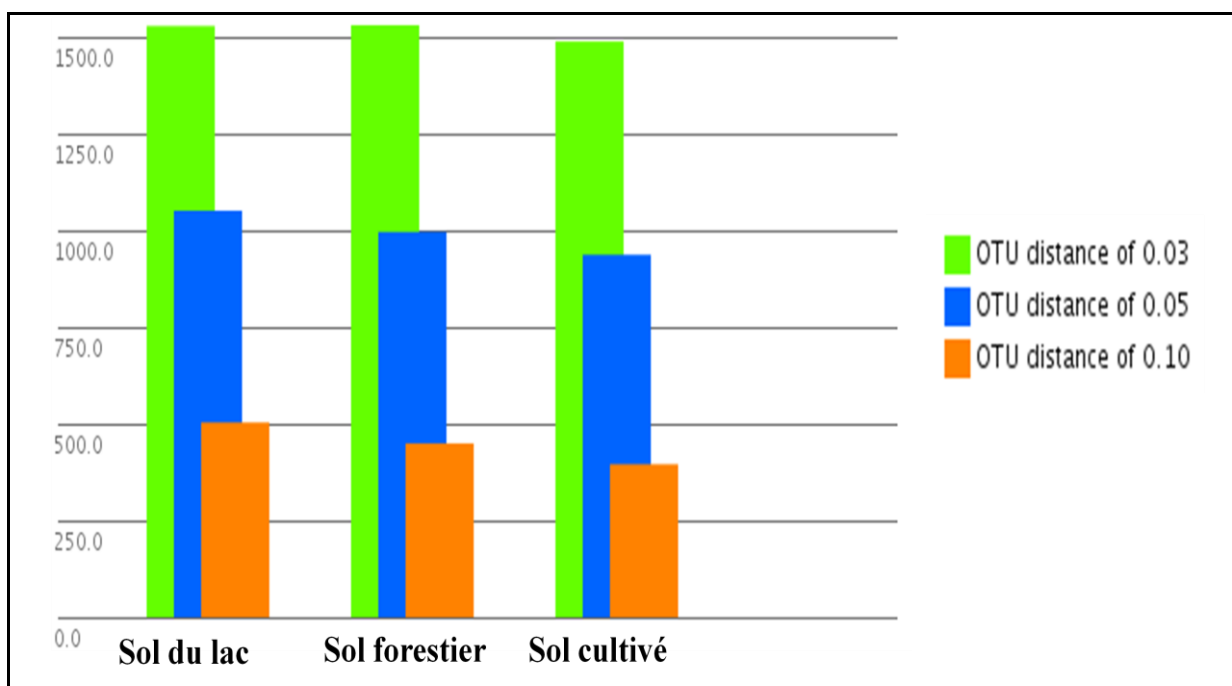


Figure 37: Histogramme des nombres des OTUs identifiées dans les trois échantillons de sol.

2.2.2. Indices de la diversité bactérienne

Ces estimateurs se répartissent en deux catégories : les indices paramétriques (estiment le nombre des OTUs non identifiées par rapport au nombre total de séquences analysées) et les indices non paramétriques (estiment le nombre des OTUs identifiées par rapport au nombre de séquences analysées) (Shannon et Weaver, 1949 ; Chao, 1984 ; Magguran, 1988 ; Frontier et Pichot-Viale, 1998 ; Curtis *et al.*, 2002 ; Colwell *et al.*, 2004 ; Kemp et Aller, 2003). Par ailleurs, nous avons utilisé les indices non paramétriques car ils sont les mieux adaptés pour traduire le degré de la diversité bactérienne tellurique.

2.2.2.1. Indice de Shannon-Weaver

L'indice de Shannon-Wiener est le plus couramment utilisé et est recommandé par différents auteurs (Gray *et al.*, 1992). Il permet d'exprimer la diversité en prenant en compte le nombre d'OTUs et l'abondance des individus au sein de chacune de ces espèces. Ainsi, une communauté dominée par une seule OTU aura un coefficient moindre qu'une communauté dont toutes les OTUs sont codominantes. La valeur de l'indice varie de 0 (une seule OTU, ou bien une OTU dominant très largement toutes les autres) à $\log S$ (lorsque toutes les OTUs ont même abondance). Les valeurs de l'indice de Shannon-Wiener sont représentées dans la figure 38.

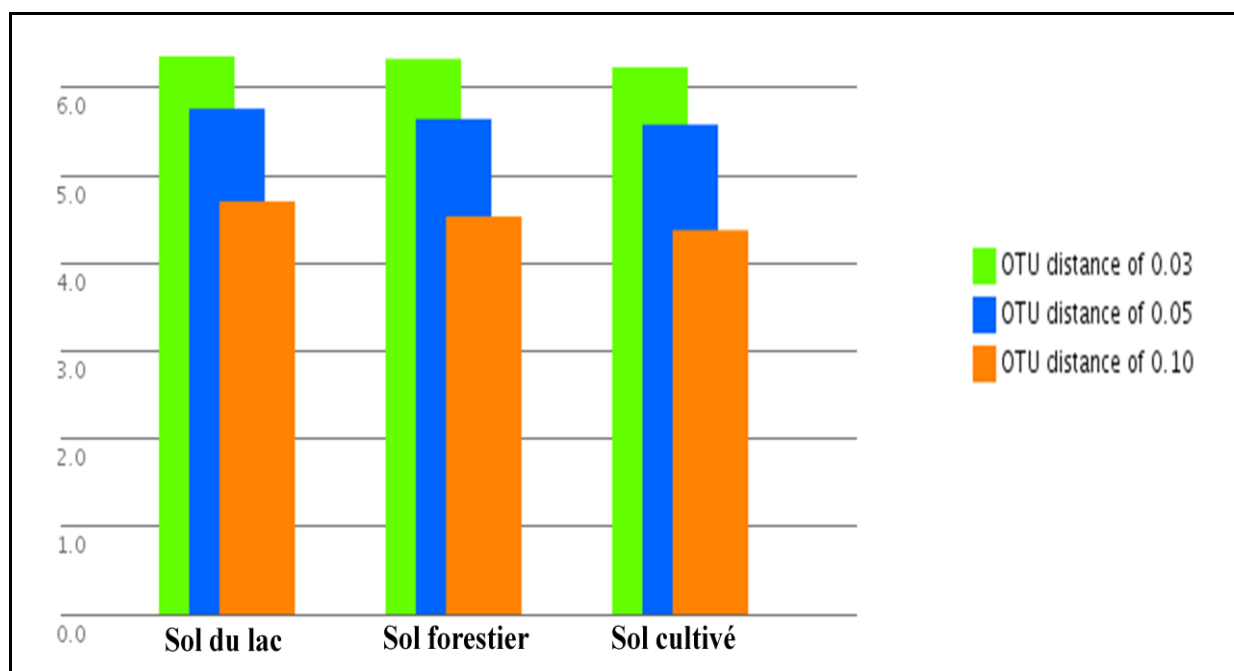


Figure 38: Histogramme représentant les valeurs de l'indice de diversité "Shannon-Wiener".

La valeur de l'indice de Shannon-Wiener (environ 6) est la même pour les trois échantillons de sol ce qui exprime une même abondance de toutes les OTUs identifiées précédemment.

2.2.2.2. Inverse de l'indice de Simpson

Cet indice a été proposé par Simpson en 1949. Il mesure la probabilité que deux individus sélectionnés au hasard appartiennent à la même espèce. L'indice est inversement proportionnel à la diversité. De ce fait, une autre formulation a été proposée afin d'établir un indice directement représentatif de l'hétérogénéité en retranchant l'indice de Simpson à sa valeur maximale : 1 (Piélou, 1969 ; Pearson et Rosenberg, 1978). Cet indice varie donc de 0 (diversité minimum) à 1 (diversité maximum). La figure 39 représente les valeurs de cet indice pour les trois échantillons de sol.

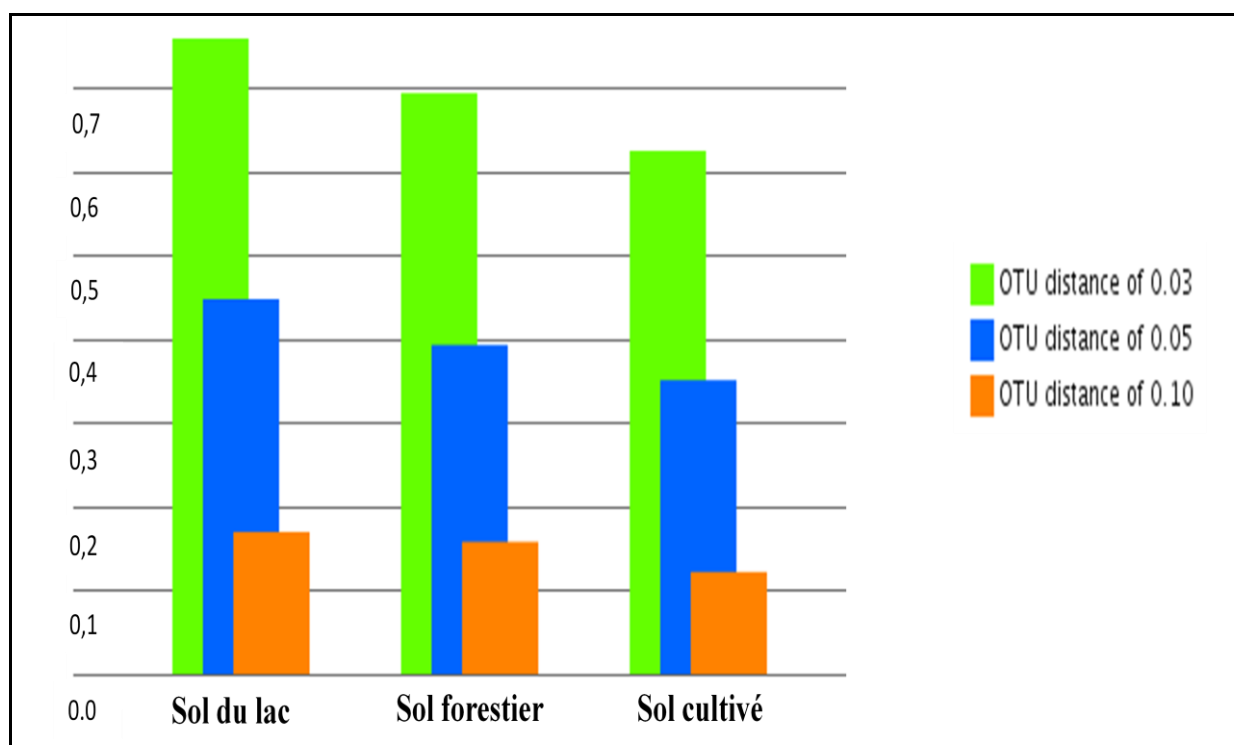


Figure 39: Histogramme représentant les valeurs de " l'inverse de l'indice de diversité de Simpson ".

Les valeurs des trois sols sont presque les mêmes et elles sont proches de " 1 " ce qui traduit l'existence d'une forte hétérogénéité de la population bactérienne dans ces écosystèmes.

2.2.2.3. Indice Chao1

Il permet d'estimer la richesse spécifique de la communauté bactérienne dans un échantillon donné (Chao, 1984). Toutes les valeurs de l'indice Chao1 sont représentées par l'histogramme ci-dessous (figure 40).

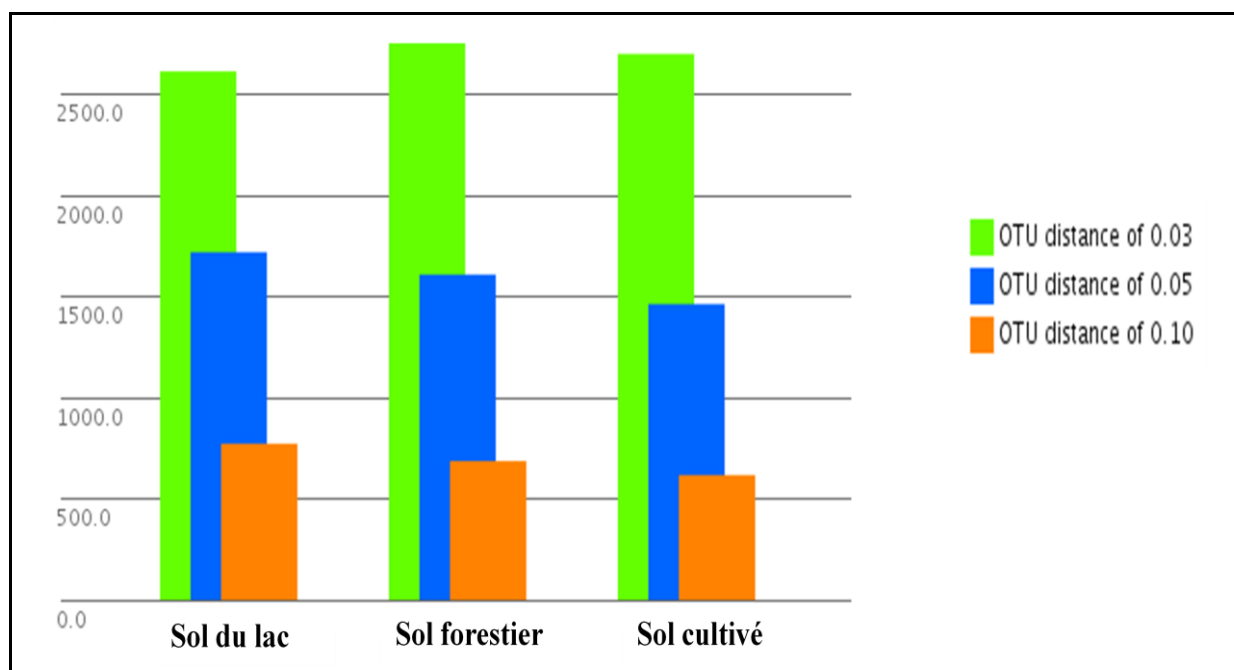


Figure 40: Histogramme représentant l'indice de la richesse spécifique " Chao1 ".

D'après ces résultats, nous constatons que les trois échantillons de sol possèdent la même richesse spécifique.

3. Effet des caractéristiques physicochimiques sur la diversité bactérienne

La richesse des sols leur confèrent une variabilité bactérienne taxonomique importante. Notre analyse à l'échelle métagénomique a permis d'estimer la nature et le nombre des communautés bactériennes relatives aux trois types de sols. Cette diversité est expliquée en termes d'abondance des différents OTU identifiées aux différents étages taxonomiques. En effet, à partir de 28795 séquences géniques initiales 24929 (soit 86,57%) ont été triées pour l'analyse métagénomique (pyroséquençage 454 d'ADNr 16S).

Les résultats métagénomiques ont révélé l'existence de 10 *phyla* en plus du candidat TM7 (figure 30), 20 familles (figure 32) et 13 genres (figure 34). Cette diversité est reliée, entre autre, à la somme des facteurs physicochimiques des sols comme indiquée dans les travaux de Lauber *et al.* (2009) et de Jacquiod (2012).

La diversité bactérienne ainsi que le comportement physiologique sont fonction du pH comme cela a été noté par Lauber *et al.* (2009) où ils ajoutent que cette corrélation ($r=0,79$) est due à la disponibilité des nutriments, l'humidité et la salinité. Les bactéries préfèrent les sols neutres ou légèrement alcalins (Davet, 1996); ce qui est le cas de nos trois sols qui sont classés alcalins. Les pH alcalins mesurés dans les différents sols de notre étude laisseraient supposer que ces valeurs alcalines n'altèrent pas directement la structure de la communauté bactérienne, voire microbienne. Ces valeurs basiques pourraient jouer le rôle de variable d'état qui fournit un indice d'intégration des conditions des sols car les valeurs de pH des sols sont directement ou indirectement liées à différents facteurs tels que la solubilité des ions cationiques, la salinité et la disponibilité des nutriments, comme cela a été indiqué dans les travaux de lauber *et al.* (2009). Il reste que la seule détermination des changements dans le pH du sol ne peut conduire aux tendances observées dans nos résultats. Il faut envisager la totalité des facteurs physicochimiques pour prétendre à une explication rationnelle de leurs effets (positifs ou négatifs) vis-à-vis de la variabilité spécifique bactérienne. Dans ce sens, Fortin (1972) a noté une étroite relation entre l'abondance des microorganismes et la nature basique du sol dans le Québec, ainsi que la composition de la végétation. Bedel *et al.* (2004), quant à eux, ont analysé les aspects granulométrique, microbiologique et physicochimique d'un sol d'un bassin d'infiltration et ont observé une différence entre les échantillons de surface (0 à 5cm) et ceux de l'horizon situé entre 5 et 20cm, surtout du point de vue matière organique et microorganismes.

La teneur en matière organique contribue, elle aussi, énormément dans la distribution des bactéries au sein du sol. Lauber *et al.* (2009) ont isolé des *Acidobacteria*, *Actinobacteria* et *Bacteroidetes*, comme dans le cas de notre étude. Ils ont, de plus, noté une prédominance des *Firmicutes* et *Bacteroides* et ont associé ces observations à la teneur en matière organique et aux ions métalliques.

L'outil de la métagénomique s'avère primordial dans la détermination de la nature des communautés bactériennes surtout celles qualifiées de non cultivables. Delmont (2011) et Shames (2012) ont noté la fiabilité de la métagénomique pour l'étude de l'effet de l'environnement sur les structures génétiques des bactéries.

Conclusion générale et perspectives

Ce travail de thèse devait, par ses objectifs, évaluer l'apport de techniques biomoléculaires que sont le MALDI TOF MS et la métagénomique dans l'étude de la biodiversité bactérienne tellurique.

Ainsi, la première partie de ce travail a concerné l'applicabilité du MALDI-TOF MS dans la caractérisation des communautés bactériennes présentes dans l'échantillon de sol collecté au niveau de la forêt "El Meridj". Tous les isolats bactériens ont été correctement identifiés au niveau de l'espèce par cette technique spectrométrique. La combinaison d'approches phénotypiques et moléculaires (MALDI-TOF MS et 16S) a permis de mettre en évidence la présence de communautés bactériennes diversifiées (obtention de neuf espèces bactériennes différentes). La concordance entre ces deux approches confirme la fiabilité de l'identification par MALDI-TOF MS. A travers l'analyse de nos résultats, nous avons constaté que la technique protéomique constitue un outil majeur dans la routine de microbiologie du sol pour deux raisons : d'une part il est maintenant possible de travailler directement sur des bactéries cultivables intactes (sans traitement préalable) et d'autre part, les bases de données permettant de comparer les spectres de masses sont disponibles. Cette méthode, peu coûteuse, permet en quelques minutes, d'obtenir une identification précise et fiable et au final d'optimiser la prise en charge des échantillons.

La seconde partie s'est orientée vers la mesure de la diversité bactérienne dans trois échantillons de sol provenant de la région d'EL- Meridj. Pour atteindre cet objectif, nous avons appliqué la nouvelle technologie de séquençage 454. Cette dernière a permis de révéler la présence de 11 *phyla*, 20 familles et 13 genres qui sont fortement hétérogènes. De plus, tous les OTUs sont également réparties (même abondance) sur les trois sols. En outre, l'indice de la richesse spécifique "Chao1" a montré que la région d'El-Meridj est extrêmement riche en bactéries dont *Proteobacteria* représente le *phylum* le plus abondant. Ceci peut être expliqué par l'influence des facteurs abiotiques des sols (notamment, le pH alcalin et la teneur élevée en matière organique).

Ces techniques, certes, apportent une précision supplémentaire dans l'identification bactérienne ; mais ne peuvent être catégoriques et formelles à notre avis du fait que les bases de données associées au MALDI ou au 16S (GenBank et autres) ne sont pas totalement fiables voire complètes. Elles sont "personnalisées" au niveau des laboratoires. Leur contenu n'est donc pas uniforme d'un laboratoire à un autre. Les bases de données associées à la technique 16S sont principalement Genbank et EMBL. Leur contenu est certes identique car

appartiennent à International Nucleotide Sequence Database Collaboration mais abritent énormément de redondances et de séquences incontrôlées au moment de la soumission des séquences par les auteurs chercheurs. La comparaison d'une séquence contre ces banques (BLAST) peut biaiser le résultat de comparaison.

Cependant, plusieurs études complémentaires pourraient être envisagées afin de mieux gérer la difficulté de l'identification bactérienne. Il serait nécessaire d'établir un lien entre la chimiotaxonomie, la phénotypie classique et les approches moléculaires dont les hybridations ADN/ADN ou ADN/ARN. En effet, les méthodes utilisées dans cette thèse ont permis de caractériser les communautés bactériennes d'un point de vue phylogénétique, mais apportent peu de renseignements sur le rôle de ces microorganismes polymorphes dans l'écosystème. D'autre part, Il serait particulièrement intéressant d'élargir les banques métagénomiques afin de compléter l'identification de nos OTUs à des niveaux taxonomiques plus fins (espèces et/ou sous espèces).

Références bibliographiques

Achtman M., Wagner M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* **6**: 431–440.

Acinas S.G., Sarma-Rupavtarm R., Klepac-Ceraj V., Polz M.F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* **71**:8966-8969.

Adékambi T., Drancourt M. (2004). Dissection of phylogenetic relationships among 19 rapidly growing Mycobacterium species by 16S rRNA, hsp65, sodA, recA and rpoB gene sequencing. *Int. J. Syst. Evol. Microbiol.* **54**: 2095–2105.

Ahmadian A., Ehn M., Hober S. (2006). Pyrosequencing: history, biochemistry and future. *Clin. Chim. Acta.* **363**: 83-94.

Alexander M. (1977). Introduction to soil microbiology. John Wiley and Sons, USA.
Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl. Environ. Microbiol.* **63**: 2647-2653.

Amiri-Eliasi B., Fenselau C. (2001). Characterization of protein biomarkers desorbed by MALDI from whole fungal cells. *Anal. Chem.* **73(21)**: 5228-31.

Andersson A.F., Lindberg M., Jakobsson H., Backhed F., Nyre´n P. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PLoS ONE.* **3**: e2836.

Andrade F., Shelley J., Wetzel W., Webb M., Gamez G., Ray S., Hieftje G. (2008). Atmospheric pressure chemical ionization source. 2. Desorption-ionization for the direct analysis of solid compounds. *Anal. Chem.* **80(8)**:2654-63.

Artursson V., Finlay R.D., Jansson J.K. (2005). Combined bromodeoxyuridine immunocapture and terminal-restriction fragment length polymorphism analysis highlights differences in the active soil bacterial metagenome due to Glomus mosseae inoculation or plant species. *Environ. Microbiol.* **7**: 1952-1966.

Asuming-Brempong S., Gantner S., Adiku S.G.K., Archer G., Edusei V., Tiedje J.M. (2008). Changes in the biodiversity of microbial populations in tropical soils under different fallow treatments. *Soil Biol. Biochem.* **40**: 2811-2818.

Baize D. (1988). Guide des analyses courantes en pédologie. 2^{ème} Ed. Revue et augmentée. INRA, Paris. ISBN: 978-2-7380-0892-3. 266p.

Barbuddhe S.B., Maier T., Schwarz G., Kostrzewa M., Hof H., Domann E. et al. (2008). Rapid identification and typing of listeria species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* **74 (17)**: 5402–5407.

Barghoutti S.A. (2011). A Universal Method for the Identification of Bacteria Based on General PCR Primers. *Indian J. Microbiol.* **51**: 430- 444.

Basile F., Beverly M.B., Voorhees K.J. (1998). Pathogenic bacteria: their detection and differentiation by rapid lipid profiling and pyrolysis mass spectrometry. *Trends Anal. Chem.* **17**:95-109.

- Bastide A., De Méo M., Andriantsoa M., Laget M. et Duménil G. (1986).** Isolement et sélection de souches d'actinomycètes productrices de substances antifongiques de structure non-polyénique. *Mircen J.* **2**: 453-466.
- Baver L.D. (1961).** Soil Physics. John Wiley and Sons, New York. 280p.
- Bedell J. P., Neto M., Delome C. et al. (2004).** Etude des paramètres physicochimiques et microbiologiques du sol d'un bassin d'infiltration rénové des eaux pluviales de l'est lyonnais. 5^{ème} Conférence internationale. Lyon. France. 260p.
- Beijerinck M.W. (1913).** De infusies en de ontdekking der bacterien. In: Jaarboek van de Koninklijke Akademie van den Wetenschappen. Müller, Amsterdam. 230p.
- Bent S.J., Forney L.J. (2008).** The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J.* **2**:689-695.
- Best M. (1995).** Classification of rocks and their abundances on the earth. In rocks physics and phase relation: a handbook of physical contents. pp:1-7.
- Bizzini A., Durussel C., Bille J., Greub G., Prod'hom G. (2010).** Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a clinical microbiology laboratory. *J. Clin. Microbiol.* **48**:1549–1554.
- Bizzini A., Greub G. (2010).** MALDI-TOF MS, a revolution in clinical microbial identification. *Clin Microbiol Infect.* **16**:1614-9.
- Blattner F.R., Plunkett G., Bloch C.A., Perna N.T., Burland V., Riley M. et al. (1997).** The complete genome sequence of *Escherichia coli* K-12. *Sci.* **277**:1453-1462.
- Blondiaux N., Gaillot O., Courcol R. J. (2013).** Identification bactérienne par spectrométrie de masse de type MALDI-TOF : évaluation au CHU de Lille. *Pathol. Biol.* **58** : 55 – 57.
- Borneman J., Skroch P.W., O'Sullivan K.M. (1996).** Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl. Environ. Microbiol.* **62**: 1935-1943.
- Borneman J., Triplett E.W. (1997).** Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and population shifts associated with deforestation. *Appl. Environ. Microbiol.* **63**:2647–2653.
- Bright J. J., Claydon M. A., Soufian M., Gordon D. B. (2002).** Rapid typing of bacteria using matrix-assisted laser desorption ionization time-of-flight mass spectrometry and pattern recognition software. *J. Microbiol. Methods* **48**: 127–138.
- Brodie E.L., DeSantis T.Z., Joyner D.C., Baek S.M., Larsen J.T., Andersen G.L. et al. (2006).** Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl. Environ. Microbiol.* **72**:6288-6298.

- Buckley D.H., Schmidt T.M. (2002).** Exploring the diversity of soil - a microbial rainforest In Biodiversity of Microbial Life: Foundation of Earth's Biosphere, ed. AL Reysenbach. New York, USA: Wiley. 300p.
- Burgmann H., Pesaro M., Widmer F., Zeyer J. (2001).** A strategy for optimizing quality and quantity of DNA extracted from soil. *J. Microbiol. Methods.* **45:** 7-20.
- Camara J.E., Hays F.A. (2007).** Discrimination between wild-type and ampicillin-resistant *Escherichia coli* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Bio. Anal. Chem.* **389:** 1633–1638.
- Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K. et al. (2010).** QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods.* **7:**335-336.
- Carroll K. Weinstein M. (2007).** Manual and automated systems for detection and identification of microorganisms. In: Murray, P.R. Baron, E.J. Jorgensen, J.H. Landry, M.L. Pfaller, M.A. eds. Manual of clinical microbiology 9th ed. Washington, DC. *American Soc. Microbiol.* 192-217.
- Chanama S. (1999).** Comparative 16S rRNA sequence analysis. Warasan Wichai Witthayasat Kanphaet. **13:** 107-117.
- Chao A. (1984).** Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* **11:** 265-270.
- Chao A. (1987).** Estimating the population size for capture–recapture data with unequal cathability. *Biometrics.* **43:** 783-791.
- Chao A., Lee S.M. (1992).** Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **87:**210-217.
- Chen P., Lu Y., Harrington P. B. (2008).** Biomarker profiling and reproducibility study of MALDI-MS measurements of *Escherichia coli* by analysis of variance-principal component analysis. *Anal. Chem.* **80:** 1474–1481.
- Chenu C. (1993).** Clay or sand polysaccharide associations as models for the interface between micro-organisms and soil: water related properties and microstructure. *Geoderma.* p.56.
- Cherkaoui A., Emonet S., Fernandez J., Schorderet D., Schrenzel J. (2011).** Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of *Beta-hemolytic streptococci*. *J. Clin. Microbiol.* **49(8):**3004-5.
- Cherkaoui A., Hibbs J., Emonet S., Tangomo M., Girard M., Francois P. et al. (2010).** Comparison of two matrix-assisted laser desorption ionization time-offlight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. *J. Clin. Microbiol.* **48:** 1169–1175.

- Cho J.C., Tiedje J.M. (2000).** Biogeography and degree of endemism of fluorescent *Pseudomonas* strains in soil. *Appl. Environ. Microbiol.* **66(12)**:5448-56.
- Claydon M.A., Davey S.N., Edwards-Jones V., Gordon D.B. (1996).** The rapid identification of intact microorganisms using mass spectrometry. *Nat. Biotechnol.* **14(11)**: 1584-1586.
- Cohan F.M. (2002).** Sexual isolation and speciation in bacteria. In Genetics of Mate Choice : From Sexual Selection to Sexual Isolation. *Springer*. pp:359-370.
- Cole J.R., Wang Q., Cardenas E., Fish J., Chai B., Farris R.J. et al. (2008).** The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acid Res.* **37**: 141-145.
- Coleman D.C., Whitman W.B. (2005).** Linking species richness, biodiversity and ecosystem function in soil systems. *Pedobiol.* **49**: 479-497.
- Colwell R. K., Mao C.X., Chang J. (2004).** Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecol.* **85** : 2717-2727.
- Constantin E. (1996).** Spectrométrie de masse, principes et applications. Tec Doc, Paris.189p.
- Conway G. C., Smole S. C., Sarracino D. A., Arbeit R. D., Leopold P. E. (2001).** Phyloproteomics: species identification of *Enterobacteriaceae* using matrix- assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mol. Microbiol. Biotechnol.* **3**:103–112.
- Costa R., Gotz M., Mroczek N., Lottmann J., Berg G., Smalla K. (2006).** Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. *FEMS Microbiol. Ecol.* **56**: 236-249.
- Curtis T.P., Sloan W.T. (2005).** Exploring microbial diversity - a vast below. *Sci.* **309**: 1331-1333.
- Curtis T.P., Sloan W.T., Scannell J.W. (2002).** Estimating prokaryotic diversity and its limits. *Proc. Nat. Acad. Sci. USA.* **99**: 10494–10499.
- Daniel R. (2005).** The metagenomics of soil. *Nat. Rev. Microbiol.* **3**: 470-8.
- Dauga C., Doré J., Sghir A. (2005).** La diversité insoupçonnée du monde microbien. *Med. Sci.* **21**: 290-296.
- Dauwalder O., Freydiere A. M., Meugnier M., Benito Y., Chomar M., Ginevra C. (2011).** Utilisation de la spectrométrie de masse pour l'identification bactérienne : évaluation du système Axima[®]/Saramin[®]/SirWeb MALDI-TOF[®]. *Biotribune.* **37** : 6-11.
- Davet P. (1996).** Vie microbienne du sol et production végétale. Ed. Quae. INRA. Paris. ISBN : 2-7380-0648-5. 383p.

- De Lara D., Ejmaes M., Casaburi A., Lisitskiy M., Cristiano R., Pagano S. et al. (2008).** Feasibility investigation of NbN nanowires as detector in time-of-flight mass spectrometers for macromolecules of interest in biology (proteins). *J. Low. Temp. Phys.* **151**:771-76.
- Degand N., Carbonnelle E., Dauphin B., Beretti J-L., Le Bourgeois M., Sermet-Gaudelus I. et al. (2008).** Matrix-Assisted Laser Desorption Ionization–Time of Flight mass spectrometry for identification of nonfermenting Gram-negative *Bacilli* isolated from cystic fibrosis patients. *J. Clin. Microbiol.* **46**: 3361–3367.
- Delmont T. (2011).** Description des communautés microbiennes du sol par une approche métagénomique. Thèse Doctorat. Ecully, Ecole centrale Lyon. France. 180p.
- Delmont T.O., Simonet P., Vogel T.M. (2012).** Describing microbial communities and performing global comparisons in the omicera. *ISME J.* **6**: 1625-28.
- Delong E.F. (2007).** Sea change for metagenomics?. *Nat. Rev. Microbiol.* **5**:326-326.
- Demirev P.A., Ho Y.P., Ryzhov V., Fenselau C. (1999).** Microorganism identification by mass spectrometry and protein database searches. *Anal. Chem.* **71**:2732-8.
- Demirev P.A., Ramirez J., Fenselau C. (2001).** Tandem mass spectrometry of intact proteins for characterization of biomarkers from *Bacillus cereus* T spores. *Anal. Chem.* **73**: 5725-5731.
- Denis F., Poly M.C., Martin C. et al. (2007).** Bacteriologie médicale, technique usuelles, Elsevier Masson. pp : 14-20.
- DeSantis T. Z. (2006).** Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **72**:5069-72.
- DeSantis T., Brodie E., Moberg J, Zubieta I., Piceno Y., Andersen G. (2007).** High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb. Ecol.* **53**:371-383.
- Dieckmann R., Helmuth R., Erhard M., Malorny B. (2008).** Rapid classification and identification of salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Appl. Environ. Microbiol.* **74(24)**: 7767-78.
- D'Onofrio A., Crawford J.M., Stewart E.J., Witt K., Gavrish E., Epstein S., Clardy J., Lewis K. (2010).** Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* **17**:254-264.
- Dressman D., Yan H., Traverso G., Kinzler K.W., Vogelstein B. (2003).** Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Nat. Acad. Sci. USA.* **100**: 8817-8822.
- Dubois D., Leysse D., Chacornac J.P., Kostrzewa M., Schmit P.O., Talon R. et al. (2010).** Identification of a variety of *Staphylococcus* species by matrix-assisted laser desorption ionization–time of flight mass spectrometry. *J. Clin. Microbiol.* **48**: 941–945.

- Duchaufour P. (1977).** Pédogénèse et classification. Masson, Paris. 290p.
- Ederer M.M. (2011).** Metagenomics: methods, applications and challenges. In: Li R.W. (ed.) *Metagenomics and its Applications in Agriculture, Biomedicine, and Environmental Studies. Nova. Sci. Inc.* New York, USA. 250p.
- Edgar R.C. (2011).** UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. **27**:2194-2200.
- Edwards-Jones V., Claydon M.A., Evason D.J., Walker J., Fox A.J., Gordon D.B. (2000).** Rapid discrimination between methicillin-sensitive and methicillin-resistant *Staphylococcus aureus* by intact cell mass spectrometry. *J. Med. Microbiol.* **49**:295–300.
- Emonet S., Shah H.N., Cherkaoui A., Schrenzel J. (2010).** Application and use of various mass spectrometry methods in clinical microbiology. *Clin. Microbiol. Infect.* **16(11)**: 1604-13.
- Fagerquist C.K., Yee E., Miller W.G. (2007).** Composite sequence proteomic analysis of protein biomarkers of *Campylobacter coli*, *C. lari* and *C. concisus* for bacterial identification. *Analyst*. **132 (10)**: 1010–1023.
- Falkowski P.G., Fenchel T., Delong E.F. (2008).** The microbial engines that drive earth's biogeochemical cycles. *Sci.* **320**:1034-1039.
- Fenselau C., Demirev P. A. (2001).** Characterization of intact microorganisms by MALDI mass spectrometry. *Mass Spectrom. Rev.* **20 (4)**: 157–171.
- Fierer N., Lennon J.T. (2011).** The generation and maintenance of diversity in microbial. *Am. J. Botany*. **98**: 439-448.
- Fierer N., Bradford M.A., Jackson R.B. (2007).** Toward an ecological classification of soil bacteria. *Ecology*. **88**: 1354-1364.
- Fortin J. A. (1974).** Observations microbiologiques de quelques sols forestiers du Québec. *Rev. Can. Microbiol.* **18 (1)** : 1379 – 1388.
- Fortin N., Beaumier D., Lee K., Greer C.W. (2004).** Soil washing improves the recovery of total community DNA from polluted and high organic content sediments. *J. Microbiol. Methods*. **56**: 181-191.
- Foster R.C. (1988).** Microenvironments of soil microorganisms. *Biol. Fert. Soils*. **6**:189-203.
- Freiwald A., Sauer S. (2009).** Phylogenetic classification and identification of bacteria by mass spectrometry. *Nat. protocols*. **4(5)**: 732-42.
- Frontiers P. V. (1998).** Écosystèmes. Structure, fonctionnement, évolution. Paris, Dunod, 2eme édition, 447 p.
- Frostegard A., Courtois S., Ramisse V. (1999).** Quantification of bias related to the extraction of DNA directly from soils. *Appl. Environ. Microbiol.* **65**: 5409-5420.

- Gagnard J., Huguet C., Ryser J.P. (1988).** L'analyse du sol et du végétal dans la conduite de la fertilisation, le contrôle de la qualité des fruits. Secrétariat générale OILB/SROP. (Ed.).
- Galbraith E.A., Antonopoulos D.A., White B.A. (2007).** Application of suppressive subtractive hybridization to uncover the metagenomic diversity of environmental samples, p. 295-334. In: Martin C.C. (ed.) Environmental Genomics. Humana Press, Totowa, USA.
- Gidenne S. (2011).** Identification bactérienne par la technologie MALDI-TOF. *Rev. Tech. Luxembourg*. **1** : 38-43.
- Gilbert J.A., Dupont C.L. (2011).** Microbial metagenomics: beyond the genome. *Annu. Rev. Mar. Sci.* **3**:347-371.
- Girvan M.S., Bullimore J., Pretty J.N., Osborn A.M., Ball A.S. (2003).** Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. *Appl. Environ. Microbiol.* **69**: 1800-1809.
- Gobat J.M., Aragno M., Matthey W. (2003).** Le sol vivant. Collection " Gérer l'environnement " 14, Presses Polytechniques et Universitaires Romandes. 290p.
- Goebel B.M., Stackebrandt E. (1994).** Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. *Appl. Environ. Microbiol.* **60(5)**: 1614-1621.
- Goff S.A., Ricke D., Lan T.H., Presting G., Wang R., Dunn M. et al. (2002).** A draft sequence of the rice genome (*Oryza sativa L. ssp. japonica*). *Sci.* **296**:92-100.
- Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H. et al. (1996).** Life with 6000 genes. *Sci.* **274**:546-567.
- Gray J. S., McIntyre A. D., Stirn J. (1992).** Manuel des méthodes de recherche sur l'environnement aquatique. Onzième partie. Evaluation biologique de la pollution marine, eu égard en particulier au benthos. FAO Document technique sur les pêches. 300p.
- Grayston S.J., Wang S.Q., Campbell C.D., Edwards A.C. (1998).** Selective influence of plant species on microbial diversity in the rhizosphere. *Soil Biol. Biochem.* **30**: 369-378.
- Grosse-Herrenthey A., Maier T., Gessler F., Schaumann R., Bohnel H., Kostrzewa, M. et al. (2008).** Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Anaerobe.* **14 (4)**: 242-249.
- Grundmann G.L. (2004).** Spatial scales of soil bacteria diversity - the size of a clone. *FEMS Microbiol. Ecol.* **48**: 119-127.
- Grundmann G.L., Debouzie D. (2000).** Geostatistical analysis of the distribution of NH_4^+ and NO_2^- oxidizing bacteria and serotypes at the millimeter scale along a soil transect. *FEMS Microbiol. Ecol.* **34**: 57-62.

- Hamady M., Walker J.J., Harris J.K., Gold N.J., Knight R. (2008).** Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods.* **5:** 235-237.
- Handelsman J. (2004).** Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68:** 669-685.
- Handelsman J., Rondon M.R., Brady S.F., Clardy J., Goodman R.M. (1998).** Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5:**245-249.
- Harry M., Gambier B., Bourezgui Y., Garnier-Sillam E. (1999).** Evaluation of purification procedures for DNA extracted from organic rich samples: interference with humic substances. *Analysis.* **27:** 439-442.
- Hassink J., Bouvman L.A., Zwart K.B., Brussaard L. (1993).** Relationships between habitable pore space, soil biota and mineralization rates in grasslands soils. *Soil Biol. Biochem.* **25:** 47-55.
- Hattori T. (1988).** Soil aggregates as microhabitats of microorganisms. *Biol. Fert. Soils.* **6:** 189-203.
- He Z., Deng Y., Van Nostrand J.D., Tu Q., Xu M., Hemme C.L. et al. (2010).** GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J.* **4:**1167-1179.
- He Z., Gentry T.J., Schadt C.W., Wu L., Liebich J., Chong S.C. et al. (2007).** GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* **1:**67-77.
- Hemmersbach P. (2008).** Special feature: Historical - History of mass spectrometry at the Olympic Games. *J. Mass Spectrom.* **43(7):**839-53.
- Hettick J. M., Kashon M.L., Slaven J.E., Ma Y., Simpson J.P., Siegel P.D. et al. (2006).** Discrimination of intact mycobacteria at the strain level: a combined MALDI-TOF MS and biostatistical analysis. *Proteomics.* **6:** 6416-6425.
- Hoffmann E., Charette J., Stroobant V. (1994).** Spectrométrie de masse. Masson, Paris. 220p.
- Holben W.E., Jansson J.K., Chelm B.K., Tiedje J.M. (1988).** DNA Probe Method for the Detection of Specific Microorganisms in the Soil Bacterial Community. *Appl. Environ. Microbiol.* **54:** 703-711.
- Holland R.D., Wilkes J.G., Rafii F., Sutherland J.B., Persons C.C., Voorhees K.J. (1996).** Rapid identification of intact whole bacteria based on spectral patterns using matrix assisted laser desorption ionization with time of flight mass spectrometry. *Rapid Commun Mass Spectrom.* **10:** 1227-1232.
- Holt J. G., Williams S. T. (1989).** Bergey's Manual of Systematic Bacteriology. Baltimore USA.

- Huber J.A., Mark Welch D.B., Morrison H.G., Huse S.M., Neal P.R., Butterfield D.A. et al. (2007).** Microbial population structures in the deep marine biosphere. *Sci.* **318**:97-100.
- Huetz De Lempis A. (1970).** La végétation de la terre. Masson et C^{ie}. (Ed.). 275p.
- Hurt R.A., Qiu X., Wu L., Roth Y., Palumbo A.V., Tiedje J.M., Zhou J. (2001).** Simultaneous recovery of RNA and DNA from soils and sediments. *Appl. Environ. Microbiol.* **67**: 4495-503.
- Hutchison C.A. (2007).** DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res.* **35**:6227-6237.
- Hyman E.D. (1988).** A new method of sequencing DNA. *Anal. Biochem.* **174**: 423-436.
- Iftikhar Ahmed I., Yokota A., Yamazoe A. et Fujiwara T. (2007).** Proposal of *Lysinibacillus boronitolerans* gen. nov. sp. nov., and transfer of *Bacillus fusiformis* to *Lysinibacillus fusiformis* comb. nov. and *Bacillus sphaericus* to *Lysinibacillus sphaericus* comb. nov. *Int. J. Syst. Evol. Microbiol.* **57** : 1117–1125.
- Jacquioud S. (2012).** Métagénomique et approches alternatives pour l'étude fondamentale et l'exploitation de la flore tellurique. Thèse Doctorat. Ecully Ecole centrale de Lyon. France.
- Janda J.M., Abbot S.L. (2007).** 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* **45**: 2761-2764.
- Janssen P.H. (2006).** Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* **72**: 1719-1728.
- Jones J.J., Stump M.J., Fleming R.C., Lay J.O., Wilkins C.L. (2003).** Investigation of MALDI-TOF and FT-MS techniques for analysis of *Escherichia coli* whole cells. *Anal. Chem.* **75**(6): 1340-7.
- Kanagawa T. (2003).** Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J. Bio. Sci. Bioeng.* **96**:317-323.
- Karas M., Hillenkamp F. (1988).** Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**(20): 2299-301.
- Kauffmann I.M., Schmitt J., Schmid R.D. (2004).** DNA isolation from soil samples for cloning in different hosts. *Appl. Microbiol. Biotechnol.* **64**: 665-670.
- Kemp P. F., Aller J.Y. (2003).** Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiol. Ecol.* **1600**:1-17.
- Kennedy A.C., Smith K.L. (1995).** Soil microbial diversity and the sustainability of agricultural soils. *Plant Soil.* **170**: 75-86.
- Kim M., Morrison M., Yu Z. (2011).** Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J. Microb. Methods.* **84**: 81–87.

- Kirk J.L., Beaudette L.A., Hart M., Moutoglis P., Klironomos J.N., Lee H. et al. (2004).** Methods of studying soil microbial diversity. *J. Microbiol. Methods*. **58**:169-188.
- Koeppel A.F., Wu M. (2013).** Surprisingly extensive mixed phylogenetic and ecological signals among bacterial operational taxonomic units. *Nucleic Acids Res.* **41(10)**:5175-88.
- Koubek J., Uhlik O., Jecna K., Junkova P., Vrkoslavova J., Lipov J. et al. (2012).** Whole-cell MALDI-TOF: Rapid screening method in environmental microbiology. *Int. Biodeterior Biodegrad.* **69** : 82-86.
- Krishnamurthy T., Ross P.L. (1996).** Rapid identification of bacteria by direct matrix assisted laser desorption/ionization mass spectrometric analysis of whole cells. *Rapid Commun Mass Spectrom.* **10 (15)**: 1992-1996.
- Krishnan P., John D., Alexander B.J., Butler B.E., Hummel J.W. (1980).** Reflectance Technique for Predicting Soil Organic Matter. *Soil Sci. Soc. Am.* **44**:1282-1285.
- Kumar M. P., Vairamani M., Raju R. P., Lobo C., Anbumani N., Kumar C. P., Menon T., Shanmugasundaram S. (2004).** Rapid discrimination between strains of *beta haemolytic streptococci* by intact cell mass spectrometry. *Indian J. Med. Res.* **119**: 283–288.
- Kunin V., Copeland A., Lapidus A., Mavromatis K., Hugenholtz P. (2008).** A bioinformatician's guide to metagenomics. *Microbiol. Mol. Biol. Rev.* **72**:557-578.
- Kuske C.R., Banton K.L., Adorada D.L., Stark P.C., Hill K., Jackson P.J. (1998).** Small-Scale DNA Sample Preparation Method for Field PCR Detection of Microbial Cells and Spores in Soil. *Appl. Environ. Microbiol.* **64**: 2463-2472.
- Lakay F.M., Botha A., Prior B.A. (2007).** Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J. Appl. Microbiol.* **102**: 265-273.
- Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J. et al. (2001).** Initial sequencing and analysis of the human genome. *Nat.* **409**:860-921.
- Lartigue M.F., Héry-Arnaud G., Haguenoer E., Domelier A.S., Schmit P.O., vander Mee Marquet N. et al. (2009).** Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by matrix-assisted laser desorption ionization time-offlight mass spectrometry. *J. Clin. Microbiol.* **47**: 2284–2287.
- Lauber C.L., Hamady M., Knight R., Fierer N. (2009).** Soil pH as a predictor of soil bacterial community structure at the continental scale: a pyrosequencing-based assessment. *Appl. Environ. Microbiol.* **75**: 5111-20.
- Lauber C.L., Strickland M.S., Bradford M.A., Fierer N. (2008).** The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol. Biochem.* **40**: 2407-2415.
- Lay J.O. (2001).** MALDI-TOF mass spectrometry of bacteria. *Mass Spectrom Rev.* **20**: 172–194.

- Lay J.O., Jr. Liyanage R. (2006).** MALDI-TOF Mass Spectrometry of intact bacteria. In: Wilkins C.L., Lay J.O., Jr. (Eds.), *Identification of Microorganisms by Mass Spectrometry*, John Wiley and Sons, Hoboken, NJ, p. 352.
- Liu H., Du Z., Wang J., Yang R. (2007).** Universal sample preparation method for characterization of bacteria by matrix-assisted laser desorption ionization- time of flight mass spectrometry. *Appl. Environ. Microbiol.* **73(6)**: 1899-907.
- Liu Z., DeSantis T.Z., Andersen G.L., Knight R. (2008).** Accurate taxonomy assignments from 16s rna sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res.* **36(18)** : 120-120.
- Loy A., Wagner M., Ben Mahfoudh K. et Berry D. (2011).** Barcoded Primers Used in Multiplex Amplicon Pyrosequencing Bias Amplification. *Appl. Environ. Microb.* **77**: 7846–7849.
- Lozupone C., Hamady M., Knight R. (2006).** Online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**:371.
- Ludwig W., Strunk O., Westram. (2004).** ARB: a software environment for sequence. *Nucleic Acids Res.* **32**:1363–1371.
- Magurran A.E. (1988).** Ecological diversity and its measurement. Princeton University Press, Princeton, New Jersey, USA. 256p.
- Maier T., Klepel S., Renner U., Kostrzewa M. (2006).** Fast and reliable MALDI-TOF MS-based microorganism identification. *Nat. Methods.* **3**: i–ii.
- Margulies M., Egholm M., Altman W.E. (2005).** Genome sequencing in microfabricated high-density picolitre reactors. *Nat.* **437**: 376-380.
- Maron P.A., Ranjard L., Mougél C., Lemanceau P. (2007).** Metaproteomics: A New Approach for Studying Functional. *Microb. Ecol.* **53**: 486–493.
- Martiny D., Dediste A., Debruyne L., Vlaes L., Ben Haddou N., Vandamme P. (2010).** Accuracy of the API Campy System, the Vitek 2 *Neisseria Haemophilus* (NH) Card and the matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF MS) for the identification of *Campylobacter* and related organisms. *Clin. Microbiol. Infect.* **17**: 1001-6.
- Martiny J.B., Bohannan B.J., Brown J.H. (2006).** Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4(2)**:102-12.
- Maxam A.M., Gilbert W. (1977).** A new method for sequencing DNA. *Nat. Acad. Sci. USA.* **74**:560-564.
- Mayr E. (1942).** Systematics and the origin of species, from the viewpoint of a zoologist . Harvard University Press. 280p.

- McHardy A.C., Rigoutsos I. (2007).** What's in the mix: phylogenetic classification of metagenome sequence samples. *Curr. Opin. Microbiol.* **10**: 499-503.
- Meldrum D.R. (2001).** Sequencing genomes and beyond. *Sci.* **292**:515-517.
- Meldrum D.R., Holl M.R. (2002).** Microscale bioanalytical systems. *Sci.* **297**:1197-1198.
- Mellmann A., Müthing J. (2013).** MALDI-TOF Mass Spectrometry-Based Microbial Identification. *Adv. Tech. Diagn. Microbiol.* 187-207.
- Miller D.N. (2001).** Evaluation of gel filtration resins for the removal of PCR-inhibitory substances from soils and sediments. *J. Microbiol. Methods.* **44**: 49-58.
- Miller D.N., Bryant J.E., Madsen E.L., Ghiorse W.C. (1999).** Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol.* **65**: 4715-4724.
- Mizrahi-Man O., Davenport E.R., Gilad Y. (2013).** Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *PLoS ONE.* **8**(1).
- More M.I., Herrick J.B., Silva M.C., Ghiorse W.C., Madsen E.L. (1994).** Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl. Environ. Microbiol.* **60**: 1572-1580.
- Moreira D. (1998).** Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucleic Acids Res.* **26**: 3309-3310.
- Morris R.M., Rappe M.S., Cannon S.A., Vergin K.L., Siebold W.A., Carlson C.A. et al. (2002).** SAR11 clade dominates ocean surface bacterioplankton communities. *Nat.* **420**:806-810.
- Moura H., Woolfitt A.R., Carvalho M.G., Pavlopoulos A., Teixeira L.M., Satten G.A. et al. (2008).** MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates. *FEMS Immunol. Med. Microbiol.* **53** (3): 333–342.
- Muller A.K., Westergaard K., Christensen S., Sorensen S.J. (2002).** The diversity and function of soil microbial communities exposed to different disturbances. *Microb. Ecol.* **44**: 49-58.
- Mummey D.L., Stahl P.D. (2004).** Analysis of soil whole and inner microaggregate bacterial communities. *Microb. Ecol.* **48**: 41-50.
- Murty D., Kirschbaum M.F.U., McMurtrie R.E., McGilvray A. (2002).** Does conversion of forest to agricultural land change soil carbon and nitrogen? *Global Change Biology.* **8**: 105-123.

- Nagy E., Maier T., Urban E., Terhes G., Kostrzewa M. (2009).** Species identification of clinical isolates of *Bacteroides* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin. Microbiol. Infect.* **15**: 796–802.
- Nocker A., Burr M., Camper A. (2007).** Genotypic microbial community profiling: a critical technical review. *Microb. Ecol.* **54**: 276-289.
- Noguez A.M., Arita H.T., Escalante A.E., Forney L.J., Garcia-Oliva F., Souza V. (2005).** Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Global Ecol. Biogeogr.* **14**: 241-248.
- North N.N., Dollhopf S.L., Petrie L., Istok J.D., Balkwill D.L., Kostka J.E. (2004).** Change in Bacterial Community Structure during In Situ Biostimulation of Subsurface Sediment Cocontaminated with Uranium and Nitrate. *Appl. Environ. Microbiol.* **70**: 4911-20.
- Nunan N., Daniell T.J., Singh B.K., Papert A., McNicol J.W., Prosser J.I. (2005).** Links between plant and rhizoplane bacterial communities in grassland soils, characterized using molecular techniques. *Appl. Environ. Microbiol.* **71**: 6784-6792.
- Nunan N., Ritz K., Crabb D., Harris K., Wu K.J., Crawford J.W., Young I.M. (2001).** Phyloproteomics: species identification of *Enterobacteriaceae* using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mol. Microbiol. Biotechnol.* **3**:103–112.
- Nunan N., Wu K., Young I.M., Crawford J.W., Ritz K. (2002).** In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. *Microb. Ecol.* **44**: 296-305.
- Nusslein K., Tiedje J.M. (1999).** Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. *Appl. Environ. Microbiol.* **65**: 3622-3626.
- Oades J.M. (1984).** Soil organic matter and structural stability: mechanisms and implications for management. *Plant Soil.* **76**: 319-337.
- Orsini M. (2001).** A micro-wave-based method for nucleic acid isolation from environmental samples. *Lett. Appl. Microbiol.* **33**: 17-20.
- Pace N.R., Stahl A.D., Lane J.D., Olson J.G. (1986).** Analyzing natural microbial populations by ribosomal RNA sequences. *Adv. Microb. Ecol.* **9**:1-55.
- Pansu M., Gautheyrou J. (2003).** L'analyse du sol, minéralogique, organique et minérale. Springer-verlag, France. (Ed.).
- Parent M.E., Velegol D. (2004).** *E. coli* adhesion to silica in the presence of humic acid. *Colloids Surf B .Biointerfaces.* **39**: 45-51.
- Pearson T. H., Rosenberg R. (1978).** Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceano. Mar. Biol. Ann. Rev.* **16**: 230-306.

- Phillipot L., Renault P., Sierra J., Henault C., Clays-Josserand A., Chenu C. et al. (1996).** Dissimilatory nitrite-reductase provides a competitive advantage to *Pseudomonas* sp. RTC01 to colonise the centre of soil aggregates. *FEMS Microbiol. Ecol.* **21**: 175-185.
- Pielou E. C. (1969).** An introduction to Mathematical Ecology (Witley-Interscience ed.). New York: Witley-InterScience. 230p.
- Pochon J. (1954).** Manuel technique d'analyse microbiologique du sol. Masson Et C^{ie}, Paris (Ed.). 210p.
- Postma J., van Veen J.A., Walter S. (1989).** Influence of different initial soil moisture contents on the distribution and population dynamics of introduced *Rhizobium leguminosarum* biovar trifolii. *Soil Biol. Biochem.* **21**: 437-442.
- Rajendhran J. Gunasekaran P. (2008).** Strategies for accessing soil metagenome for desired applications. *Biotechnol. Adv.* **26**: 576-590.
- Ramette A. (2009).** Quantitative community fingerprinting methods for estimating the abundance of operational taxonomic units in natural microbial communities. *Appl Environ. Microbiol.* **75**: 2495-2505.
- Ramette A., Tiedje J.M. (2007).** Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb. Ecol.* **53**: 197-207.
- Ranjard L., Dequiedt S., Jolivet C., Saby N.P.A., Thioulouse J., Harmand J. et al. (2010).** Biogeography of soil microbial communities: a review and a description of the ongoing French national initiative. *Agron. Environ.* **30**: 359-365.
- Ranjard L., Poly F., Nazaret S. (2000).** Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. *Res. Microbiol.* **151**: 167-177.
- Ranjard L., Richaume A. (2001).** Quantitative and qualitative microscale distribution of bacteria in soil. *Res. Microbiol.* **152**: 707-716.
- Ranjard L., Richaume A., Jocteur Monoroziere L., Nazaret S. (1997).** Response of soil bacteria to Hg(II) in relation to soil characteristics and cell location. *FEMS Microbiol. Ecol.* 321-331.
- Rapilly F. (1968).** Techniques de mycologie en pathologie végétale. Annales des Epiphyties, n° 19. INRA. (Ed.). 189p.
- Rezzonico F., Vogel G., Duffy B., Tonolla M. (2010).** Application of Whole-Cell Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry for Rapid Identification and Clustering Analysis of *Pantoea* Species. *Appl. Environ. Microbiol.* **76** : 4497–4509.

- Ritz K., McNicol W., Nunan N., Grayston S., Millard P., Atkinson D. et al. (2004).** Spatial structure in soil chemical and microbiological properties in an upland grassland. *FEMS Microbiol. Ecol.* **49**: 191-205.
- Robert M., Chenu C. (1992).** Interactions between soil minerals and micro-organisms. *Soil Biochem.* 307-404.
- Roesch L.F.W., Fulthorpe R.R., Riva A., Casella G., Hadwin A.K.M., Kent A.D. et al. (2007).** Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* 1: 283-90.
- Ronaghi M., Karamohamed S., Pettersson B., Uhlen M., Nyren P. (1996).** Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**: 84-89.
- Ronaghi M., Uhlen M., Nyren P. (1998).** A sequencing method based on real-time pyrophosphate. *Sci.* **281**: 363, 365.
- Ronsin C. (2005).** L'histoire de la biologie moléculaire (Pionniers & héros), De Boeck university, 1er Ed. 295p.
- Rykiel E.J. (1985).** Towards a definition of ecological disturbance. *Aus. J. Ecol.* **10**: 361-365.
- Ryzhov V., Fenselau C. (2001).** Characterization of the protein subset desorbed by MALDI from whole bacterial cells. *Anal. Chem.* **73**(4): 746-50.
- Ryzhov V., Hathout Y., Fenselau, C. (2000).** Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. *Appl Environ. Microbiol.* **66**:3828-3834.
- Saano A. (1995).** Small-scale extraction of DNA from soil with spun column cleanup. *Mol. Microb. Ecol.* 1341-1346.
- Sagova-Mareckova M., Cermak L., Novotna J., Plhackova K., Forstova J., Kopecky J. (2008).** Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Appl. Environ. Microbiol.* **74**: 2902-2907.
- Sala O.E., Chapin F.S., Armesto J.J. (2000).** Global biodiversity scenarios for the year 2000. *Sci.* **287**: 1770-1774.
- Sanger F., Air G., Barrell B., Brown N., Coulson A., Fiddes C., Hutchison C., Slocombe P., Smith M. (1977).** Nucleotide sequence of bacteriophage phi X174 DNA. *Nat.* **265** (5596):687-95.
- Sanger F., Coulson A.R. (1975).** A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**:441-448.
- Sanger F., Nicklen S., Coulson A.R. (1977).** DNA sequencing with chain-terminating inhibitors. *Nat. Acad. Sci. USA.* **74**:5463-5467.
- Sauer S. (2008).** Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS ONE.* **3**: 2843.

- Schloss P.D., Handelsman J. (2003).** Biotechnological prospects from metagenomics. *Curr. Opin. Biotech.* **14**:303-310.
- Schloss P.D., Handelsman J. (2005).** Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**:1501-1506.
- Schloss P.D., Handelsman J. (2006).** Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl. Environ. Microbiol.* **72**:6773-6779.
- Schloss P.D., Larget B.R., Handelsman J. (2004).** Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl. Environ. Microbiol.* **70**:5485-5492.
- Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B. et al. (2009).** Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**:7537-7541.
- Schultze U., Penson K.A., Perera A., Moreau M. (1993).** Wind-tree model in two dimensions with internal degrees of freedom: Exact solution. *J. Am. Phys. Soc.* **47** : 4589 - 4592.
- Sebastianelli A., Sen T., Bruce I.J. (2008).** Extraction of DNA from soil using nanoparticles by magnetic bioseparation. *Lett. Appl. Microbiol.* **46**: 488-491.
- Seng P., Drancourt M., Gouriet F., La S.B., Fournier P.E., Rolain J.M. et al. (2009).** Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin. Infect. Dis.* **49**:543-551.
- Seng P., Rolain J.M., Fournier P.E., La Scola B., Drancourt M., Raoult D. (2010).** MALDI-TOF-mass spectrometry applications in clinical microbiology. *Future microbial.* **5(11)**: 1733-54.
- Shannon C.E. (1948).** A mathematical theory of communication. *Sys. Tech. J.* **27**:379-423 and 623-656.
- Shannon C.E., Weaver W. (1949).** The Mathematical Theory of Communication. University of Illinois Press, Urbana, USA. 420p.
- Shendure J., Ji H. (2008).** Next-generation DNA sequencing. *Nat. Biotech.* **26**:1135-1145.
- Shendure J., Mitra R.D., Varma C., Church G.M. (2004).** Advanced sequencing technologies: methods and goals. *Nat. Rev. Genet.* **5**:335-344.
- Simon C., Daniel R. (2011).** Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.* **77**:1153-1161.
- Simpson E.H. (1949).** Measurement of diversity. *Nat.* 163:688.

- Singh B.K., Dawson L.A., McDonald C.A., Buckland S.M. (2009).** Impact of biotic and abiotic interaction on soil microbial communities and functions: A field study. *Appl. Soil Ecol.* **41**: 239-248.
- Singh B.K., Millard P., Whiteley A.S., Murrell J.C. (2004).** Unravelling rhizosphere-microbial interactions: opportunities and limitations. *Trends Microbiol.* **12**: 386-393.
- Singh B.K., Munro S., Potts J., Millard P. (2008).** Influence of grass species and soil type on rhizosphere microbial community structure in grassland soils. *Appl. Soil Ecol.* **36**: 147-155.
- Singh B.K., Nazarie L., Munro S., Anderson I., Campbell C.D. (2006).** Use of multiplex terminal restriction fragment length polymorphism for rapid and simultaneous analysis of different components of the soil microbial community. *Appl. Environ. Microbiol.* **72**: 7278-7285.
- Singh B.K., Tate K.R., Kolipaka G., Hedley C.B., Macdonald C.A., Millard P. et al. (2007).** Effect of afforestation and reforestation of pastures on the activity and population dynamics of methanotrophic bacteria. *Appl. Environ. Microbiol.* **73**: 5153-5161.
- Sintchenko V., Iredell J.R., Gilbert G.L. (2007).** Pathogen profiling for disease management and surveillance. *Nat. Rev. Microbiol.* **5** (6): 464-470.
- Sipos R., Székely A., Révész S., Márialigeti K. (2010).** Addressing PCR biases in environmental microbiology studies, p. 37-58. In: Cummings SP (ed.) *Methods in Mol Biol Biorem.* Humana Press, Totowa, USA.
- Siqueira J.F., Fouad A.F., Rôças I.N. (2012).** Pyrosequencing as a tool for better understanding of human microbiomes. *J. Microbiol.* **4**:10743.
- Six J., Elliot E.T., Paustian K. (1999).** Aggregate and soil organic matter dynamics under conventional and no-tillage systems. *Soil Sci. Soc. Am. J.* **63**: 1350-1358.
- Six J., Elliot E.T., Paustian K. (2000).** Soil macroaggregate turnover and microaggregate formation: a mechanism for C sequestration under no-tillage agriculture. *Soil Biol. Biochem.* **32**: 2099-2103.
- Smith L., Sanders J., Kaiser R., Hughes P., Dodd C., Connell C. et al. (1986).** Fluorescence detection in automated DNA sequence analysis. *Nat.* **321**(6071):674-9.
- Smole S. C., King L. A., Leopold P. E., Arbeit R. D. (2002).** Sample preparation of gram-positive bacteria for identification by matrix assisted laser desorption/ionization time-of-flight. *J. Microbiol. Methods.* **48**:107-115.
- Society of American Bacteriologists. (1923).** *Bergey's manual of determinative bacteriology.* Williams and Wilkins. Baltimore.
- Sogin M.L., Morrison H.G., Huber J.A., Welch D.M., Huse S.M., Neal P.R., Arrieta J.M., Herndl G.J. (2006).** Microbial diversity in the deep sea and the underexplored "rare biosphere". *Nat. Acad. Sci. USA.* **103**:12115-12120.

- Soltner D. (2000).** Les bases de la production végétale. Tome I. le sol et son amélioration. sciences et techniques agricoles. (22^e Ed.). 326p.
- Song Y., Liu C., McTeague M., Finegold S.M. (2003).** 16S ribosomal DNA sequence-based analysis of clinically significant gram-positive anaerobic cocci. *J. Clin. Microbiol.* **41**:1363-1369.
- Sorek R., Zhu Y., Creevey C.J., Francino M.P., Bork P., Rubin E.M. (2007).** Genome-wide experimental determination of barriers to horizontal gene transfer. *Sci.* **318**:1449-1452.
- Stach J.E., Bathe S., Clapp J.P. Burns R.G. (2001).** PCR-SSCP comparison of 16S rDNA sequence diversity in soil DNA obtained using different isolation and purification methods. *FEMS Microbiol. Ecol.* **36**: 139-151.
- Stackebrandt E., Goebel B.M. (1994).** Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44 (4)**: 846-9.
- Staden R. (1979).** A strategy of DNA sequencing employing computer programs. *Nucleic Acids Res.* **6(7)**:2601-10.
- Stevenson F.J. (1976).** Stability constants of Cu, Pb and Cd complexes with humic acids. *Soil Sci. Soc. Am. J.* **40**: 665-672.
- Stoner E.R., Baumgardner M.F. (1981).** Characteristic Variations in Reflectance of Surface Soils. *Soil Sci. Soc. Am.* **45**:1161-1165.
- Sun Y., Cai Y., Liu L., Yu F., Farrell M.L., McKendree W. et al. (2009).** ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Res.* **37**:76.
- Suzuki M.T., Giovannoni S.J. (1996).** Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* **62**:625-630.
- Swerdlow H., Zhang J., Chen D., Harke H., Grey R., Wu S. et al. (1991).** Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence. *Anal. Chem.* **63(24)**:2835-41.
- Swift M.J., Andrén O., Brussaard L., Briones M., Couteaux M.M., Ekschmitt K. et al. (1998).** Global change, soil biodiversity, and nitrogen cycling in terrestrial ecosystems: three case studies. *Global Change Biol.* **4**: 729-743.
- Tabacchioni S. Chiarini L. Bevivino A. Cantale C., Dalmastrri C. (2000).** Bias caused by using different isolation media for assessing the genetic diversity of a natural microbial population. *Microb. Ecol.* **40**: 169– 176.
- Tamura K., Stecher G., Peterson D., Filipinski A., Kumar S. (2013).** MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**: 2725-2729.

- Tarlera S., Jangid K., Ivester A.H., Whitman W.B., Williams M.A. (2008).** Microbial community succession and bacterial diversity in soils during 77,000 years of ecosystem development. *FEMS Microbiol. Ecol.* **64**: 129-140.
- Temperton B., Field D., Oliver A., Tiwari B., Muhling M., Joint I. et al. (2009).** Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *ISME J.* **3**:792-796.
- Thomas P. (1980).** Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Nat. Acad. Sci. USA.* **77(9)**:5201-5.
- Tiedje J.M., Assuming-Bremppong S., Nusslein K., Marsh T.L., Flynn S.J. (1999).** Opening the black box of soil microbial diversity. *Appl. Soil Ecol.* **13**: 109-122.
- Tisdall J.M., Oades J.M. (1982).** Organic matter and water-stable aggregates in soils. *J. Soil Sci.* **33**: 141-163.
- Torsvik V., Salte K., Sorheim R., Goksoyr J. (1990).** Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. *Appl. Environ. Microbiol.* **56**: 776-81.
- Torsvik V., Daae F.L., Sandaa R., Oslash A., Vreas L. (1998).** Novel techniques for analyzing microbial diversity in natural and perturbed environments. *J. Biotechnol.* **64**: 53-62.
- Toyo'oka T. (2008).** Determination methods for biologically active compounds by ultra-performance liquid chromatography coupled with mass spectrometry: Application to the analyses of pharmaceuticals, foods, plants, environments, metabonomics, and metabolomics. *J. Chromatogr. Sci.* **43**: 1291-1300.
- Trevors J.T. (1998).** Bacterial biodiversity in soil with an emphasis on chemically-contaminated soils. *Water, Air, Soil Pollut.* **101**: 45– 67.
- Tsai Y.L., Olson B.H. (1992).** Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Appl. Environ. Microbiol.* **58**: 2292-2295.
- Tyson G.W., Chapman J., Hugenholtz P., Allen E.E., Ram R.J., Richardson P.M. et al. (2004).** Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nat.* **428**:37-43.
- Uhlik O., Strejcek M., Junkova P., Sanda M., Hroudova M., Vlcek C. et al. (2011).** Matrix-Assisted Laser Desorption Ionization (MALDI)–Time of Flight Mass Spectrometry and MALDI Biotyper-Based Identification of Cultured Biphenyl-Metabolizing Bacteria from Contaminated Horseradish Rhizosphere Soil. *Appl. Environ. Microbiol.* **77**: 6858–6866.
- Valentine N., Wunschel S., Wunschel D., Petersen C., Wahl K. (2005).** Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. *Appl. Environ. Microbiol.* **71(1)**: 58-64.

- Van Baar B.L. (2000).** Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. *FEMS Microbiol. Rev.* **24**: 193–219.
- Vanlaere E., Sergeant K., Dawyndt P., Devreese B., Vandamme P. (2006).** Identification of Burkholderia cepacia complex using MALDI-TOF mass spectrometry. *J. Cyst. Fibr.* **5 (1)**: 34–134.
- Vargha M., Takats Z., Konopka A., Nakatsu C.H. (2006).** Optimization of MALDI-TOF MS for strain level differentiation of *Arthrobacter* isolates. *J. Microbiol. Methods.* **66**: 399–409.
- Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G. et al. (2001).** The sequence of the human genome. *Sci.* **291**:1304-1351.
- Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A. et al. (2004).** Environmental genome shotgun sequencing of the Sargasso Sea. *Sci.* **304**:66-74.
- Verville J.H., Hobbie S.E., Chapin F.S., Hooper D.U. (1998).** Response of tundra Ch₄ and CO₂ flux to manipulation of temperature and vegetation. *Biogeochem.* **41**: 215-235.
- Volossiuk T., Robb E.J., Nazar R.N. (1995).** Direct DNA extraction for PCR-mediated assays of soil organisms. *Appl. Environ. Microbiol.* **61**: 3972-3976.
- Von Mering C., Hugenholtz P., Raes J., Tringe S.G., Doerks T., Jensen L.J. et al. (2007).** Quantitative phylogenetic assessment of microbial communities in diverse environments. *Sci.* **315**:1126-1130.
- Von Wintzingerode F., Gobel U.B., Stackebrandt E. (1997).** Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* **21**: 213-229.
- Wang G.C., Wang Y. (1997).** Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl. Environ. Microbiol.* **63**:4645-4650.
- Wang Q. (2007).** Naïve Bayesian Classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**:5261-5267.
- Wang Q., Garrity G.M., Tiedje J. M., Cole J.R. (2007).** Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73(16)** :5261-5267.
- Ward D.M., Bateson M.M., Weller R., Ruff-Roberts A.L. (1992).** Ribosomal RNA analysis of microorganisms as they occur in nature. *Adv. Microb. Ecol.* 219–286.
- Waterston R.H., Lindblad-Toh K., Birney E., Rogers J., Abril J.F., Agarwal P. et al. (2002).** Initial sequencing and comparative analysis of the mouse genome. *Nat.* **420**:520-562.

- Webster R., Butler B.E. (1976).** Soil classification and survey studies at Ginninderra. *Australian J. Soil Res.* **14**: 1-24.
- Weisburg W.G., Barns S.M., Pelletier D.A., Lane D.J. (1991).** 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **173**:697-703.
- Welker M., Moore E.R.B. (2011).** Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst. Appl. Microbiol.* **34**: 2–11.
- White R.E. (2006).** Principles and practice of soil science. The soil as a natural resource. Blackwell Publishing.
- Whitman W.B., Coleman D.C., Wiebe W.J. (1998).** Prokaryotes: the unseen majority. *Nat. Acad. Sci. USA.* **95**:6578-6583.
- Wieser A., Schneider L., Jung J. et Schubert S. (2012).** MALDI-TOF MS in microbiological diagnostics identification of microorganisms and beyond (mini review). *Appl. Microbiol. Biotechnol.* **93**:965–974.
- Williams T. L., Andrzejewski D., Lay, J. O., Musser, S. M. (2003).** Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. *J. Am. Soc. Mass Spectrom.* **14**: 342–351.
- Williams T.L., Andrzejewski D., Lay J.O., Musser S.M. (2003).** Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. *J. Am. Soc. Mass Spectrom.* **14** (4): 342–351.
- Wooley J.C., Godzik A., Friedberg I. (2010).** A primer on metagenomics. *PLoS Comput Biol.* **6**: 660-667.
- Wunschel S.C., Jarman K.H., Petersen C.E. (2005).** Bacterial analysis by MALDI-TOF mass spectrometry: an inter-laboratory comparison. *J. Am. Soc. Mass Spectrom.* **16**(4): 456-62.
- Yergeau E., Schoondermark-Stolk S.A., Brodie E.L., Dejean S., DeSantis T.Z., Goncalves O. et al. (2009).** Environmental microarray analyses of Antarctic soil microbial communities. *ISME J.* **3**:340-351.
- Young C.C., Burghoff R.L., Keim L.G., Minak-Bernero V., Lute J.R., Hinton S.M. (1993).** Polyvinylpyrrolidone-Agarose Gel Electrophoresis Purification of Polymerase Chain Reaction-Amplifiable DNA from Soils. *Appl. Environ. Microbiol.* **59**: 1972-1974.
- Yu J., Hu S., Wang J., Wong G.K.S., Li S., Liu B. et al. (2002).** A draft sequence of the rice genome (*Oryza sativa L. ssp. indica*). *Sci.* **296**:79-92.
- Zhang J.H., Wu L.Y., Zhang X.S. (2003).** Reconstruction of DNA sequencing by hybridization. *Bioinformatics.* **19**(1):14-21.

Zhou J., Bruns M.A., Tiedje J.M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**: 316-322.

Zhou J., Xia B., Treves D.S. (2002). Spatial and resource factors influencing high microbial diversity in soil. *Appl. Environ. Microbiol.* **68**: 326-334.

Annexes

Annexe A: DNA Extraction Protocol Kit " PowerSoil-DNA Isolation Kit (Mobio) "

Items not provided in the extraction kit:

1. DNA decontaminating solution (DNAaway, 10% bleach, etc.)
2. Large volume 8-channel pipette + tips
3. 100% **Molecular** grade ethanol
4. Sterile reagent reservoirs
5. Water bath set to 65°C
6. Ice bath
7. Swing bucket centrifuge capable of 4500 x g
8. Plate shaker with 4 metal plate adapters. This protocol uses the 96 well plate shaker (catalog #11996) listed on the MoBio website.

Items included in the extraction kit (for 1-96 well extraction):

1. (1) Bead plate
 2. (1) Spin plate (filter)
 3. (1) 0.5ml collection
 4. (4) 1.0ml collection plates
 5. (2) 2.0ml collection plates
 6. (1) Microplate (DNA elution)
 7. Sealing Tape
 8. Centrifuge Tape
 9. Elution Sealing Mat
10. Labeled solutions

Before you start:

1. Clean all surfaces and pipettors to remove DNA
2. Label Plates
 1. Plate #1 – 1ml collection plate
 2. Plate #2 – 1ml collection plate
 3. Plate #3 – 1ml collection plate
 4. Plate #4 – 1ml collection plate
 5. Plate #5 – 2ml collection plate
 6. Plate #6 – 2ml collection plate
7. UV Sterilize and Label reservoirs
 1. Bead solution – 750ul/well
 2. C1 – 60ul/well
 3. C2 – 250ul/well
 4. C3 – 200ul/well
 5. C4 – 650ul/well
 6. C5 – 500ul/well
 7. C6 – 100ul/well

Please wear gloves at all times.

Annexe A (suite):

1. **BEFORE THE FIRST USE ONLY, Solution C5-D must be prepared. Add an equal amount of 100% Ethanol to Solution C5-D (for the 4 prep kit = 120 ml, or for the 12 prep kit = 360 ml). Mix well. Put a check mark in the “ethanol added” box on the bottle cap label.**

2. Centrifuge Bead Plate for 1 min at 2500 x g to pellet the beads. Remove the Square Well Mat from the PowerSoil®-htp Bead Plate and set aside. Add 0.1 to 0.25 grams of soil sample or sample swab.

Note: This is an appropriate stopping point and you can store the PowerSoil®-htp Bead Plate at 4°C covered with the Square Well Mat. This is the most time consuming step of the protocol. Care must be taken to avoid cross contamination between sample wells.

3. Add 750 µl of PowerSoil®-htp Bead Solution to the wells of the PowerSoil®-htp Bead Plate.

4. Check Solution C1. If Solution C1 has precipitated, heat solution at 60°C until the precipitate has dissolved.

Note: Solution C1 contains SDS. If it gets cold, it will precipitate. Heating at 60°C will dissolve the SDS. Solution C1 can be used while it is still warm.

5. Add 60 µl of Solution C1. Secure the Square Well Mat (from step 2) tightly to the plate. Place sealed plates in 65°C water bath for 10 min. **DO NOT SUBMERGE THE PLATES.**

6. Place PowerSoil®-htp Bead Plate between the aluminum plate adapters and securely fasten to the 96 Well Plate Shaker.

7. Shake at speed 20 for 20 minutes.

8. Centrifuge at room temperature for 6 minutes at 4500 x g.

a. While centrifuging, aliquot 250 µl of Solution C2 into each well of Plate #1 and cover with Sealing Tape. The Sealing Tape can be re-used when centrifuging Plate #1 in step 11 if handled carefully.

9. Remove and discard the Square Well Mat from the Bead Plate.

10. Carefully remove the Sealing Tape from Plate #1 and transfer the supernatant (~ 400-500µl) from the Bead Plate to Plate #1 and pipette up and down 4 times.

Note: The supernatant may still contain some particles.

11. Re-apply the Sealing Tape to Plate #1. Incubate at 4°C for 10 minutes. Centrifuge Plate #1 at room temperature for 6 minutes at 4500 x g.

Annexe A (suite):

a. While centrifuging, aliquot 200 µl Solution C3 into each well of Plate #3, then cover with Sealing Tape. The Sealing Tape can be re-used when centrifuging Plate #3 in step 13 if handled carefully.

12. After centrifugation, carefully remove and discard Sealing Tape from Plate #1.

13. Avoiding the pellet, transfer the entire volume (~600 µl depending on sample type) of supernatant in Plate #1 to Plate #2.

14. Apply Sealing Tape to Plate #2 and centrifuge at room temperature for 6 minutes at 4500 x g.

15. Carefully remove Sealing Tape from Plate #2 and Plate #3. Avoiding the pellet, transfer the entire volume of supernatant (~600 µl) from Plate #2 to Plate #3 and pipette up and down 4 times.

16. Re-apply Sealing Tape to Plate #3. Incubate at 4°C for 10 minutes. Centrifuge at room temperature for 6 minutes at 4500 x g.

17. Carefully remove and discard Sealing Tape from Plate #3. Avoiding the pellet, transfer the entire volume of supernatant (~750 µl) to Plate #4.

18. Apply Sealing Tape to Plate #4 and centrifuge at room temperature for 6 minutes at 4500 x g.

a. While centrifuging, add 650 µl of Solution C4 to Plate #5.

19. Avoiding any residual pellet, transfer up to 650 µl of supernatant in Plate #4 to Plate #5.

20. Add a second 650 µl (1300 µl C4 total) aliquot of Solution C4 to each well of Plate #5.

Note: It is safe to stop the protocol at this step and store the samples covered with Sealing Tape at 4°C. Make sure to briefly centrifuge the plate to collect any condensate on the plate seal after overnight storage.

21. Pipet samples “up and down” to mix.

22. Place Spin Plate onto Plate #6.

23. Load approximately 650 µl from Plate #5 into each well of the Spin Plate and apply Centrifuge Tape.

24. Centrifuge at room temperature for 5 minutes at 4500 x g. Discard the flow through and place the Spin Plate back on Plate #6. Carefully remove and discard the Centrifuge Tape.

25. Repeat steps 23-24 until all the supernatant has been processed. Discard the final flow through.

Annexe A (suite):

26. Place the Spin Plate back on Plate #6.

27. **Confirm that ethanol has been added to Solution C5-D (see step 1).** Add 500 µl of Solution C5-D to each well of the Spin Plate. Apply Centrifuge Tape to the Spin Plate.

28. Centrifuge at room temperature for 5 minutes at 4500 x g. Discard the flow through and place the Spin Plate back on Plate #6.

29. Centrifuge again at room temperature for 6 minutes at 4500 x g. Discard the flow through.

30. Carefully place the Spin Plate onto the Microplate. Remove Centrifuge Tape from the Spin Plate and discard.

31. Add 100 µl of Solution C6 to the center of each well of the Spin Plate. Apply Centrifuge Tape. Let C6 sit on the filter for 10 minutes at room temperature before final centrifugation step.

32. Centrifuge at room temperature for 7 minutes at 4500 x g. Remove Centrifuge Tape and discard.

33. Cover wells of Microplate with the Elution Sealing Mat provided. DNA is now ready for any downstream application. No further steps are required.

Prolonged storage at 4°C will result in the evaporation of eluted DNA. We recommend storing DNA frozen (-20°C or -80°C). Solution C6 does not contain EDTA. To concentrate the DNA see the Hints and Troubleshooting Guide provided in the MoBio protocol.

Important Considerations

1. Normal diameter 1 ml pipet tips are too large for some of the pipetting steps. To get around this problem we use a Rainin 300 ul 8-channel with filtered tips (Rainin #SR-L300F) and a Rainin 1000 ul 8-channel with extended length filtered tips (Rainin #RTS-L1000XF).

2. Make sure that all of the necessary consumables and reagents are in place before you start the extraction. Remember, each pipetting step will require 1 box of 96 tips per plate.

3. We use individually wrapped reagent reservoirs and expose them to UV light for 30 minutes prior to usage.

4. In Step 7, it is important that the plate not rub against any surfaces in the shaker.

5. Make sure that the alpha-numeric grid is in the same orientation across all the plates. On a few occasions we have noticed that the sticker is not always in the same position on the plate.

Annexe B : Les séquences d'ADNr 16S des neuf isolats bactériens

>S1

TAGAGTTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGTAACAGG
AAGCAGCTTGCTGCTTCGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAAACTGCCTGATGGAGGG
GGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCCT
CTTGCCATCGGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCC
TAGCTGTCTGAGAGGATGACCAGCCACACTGAACCTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAG
TGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTT
GTAAAGTACTTTTCAGCGGGGAGGAAGGCGACAGGGTTAATAACCCTGTCGATTGACGTTACCCGCAGAAG
AAGCACC GGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGC AAGCGTTAATCGGAATTACTGG
GCGTAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATT
CGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCAGGTGTAGCGGTGAAATGCGTAGAGAT
CTGGAGGAATACCGGTGGCGAAGGCGGCCCTTGACAAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGA
GCAAACAGGATTAGATAACCTGGTAGTCCACGCCGTAACGATGTGACTTGGAGGTTGTGCCCTTGAGG
CGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCTGGGGAGTACGGCCGAAGGTTAAAACCTCAAATG
AATTGACGGGGGCCCCGACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTA
CTCTTGACATCCAGAACTTAGCAGAGATGCTTTGGTGCCTTCGGGAACTCTGAGACAGGTGCTGCATG
GCTGTGCTCAGCTCGTGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTTATCCTTTGTTG
CCAGCGGTTAGGCCCGGGAACCTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAA
GTCATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGC
GAGAGCAAGCGGACCTCATAAAGTGCCTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCG
GAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCCGTC
ACACCATGGGAGTGGGTTGCAAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACCTTGTGAT
TCATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTA

>S2

CTATAGAGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGGACGGTAGC
ACAGAGAGCTTGCTCCTGGGTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGGATCTGCCCGATAGAGG
GGGATAACC ACTGGAAACGGTGGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCC
TCTCACTATCGGATGAACCCAGATGGGATTAGCTAGTAGGCGGGGTAATGGCCACCTAGGCGACGATCC
CTAGCTGGTCTGAGAGGATGACCAGCCACACTGGA ACTGAGACACGGTCCAGACTCCTACGGGAGGCAGC
AGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGG
TTGTAAAGTACTTTTCAGCGGGGAGGAAGGCGATGTGGTTAATAACC GCGTCGATTGACGTTACCCGCAGA
AGAAGCACC GGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGC AAGCGTTAATCGGAATTACT
GGGCGTAAAGCGCACGCAGGCGGTCTGTAAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCA
TTTGAAACTGGCAGGCTTGAGTCTCGTAGAGGGGGGTAGAATTCAGGTGTAGCGGTGAAATGCGTAGAG
ATCTGGAGGAATACCGGTGGCGAAGGCGGCCCTTGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGG
GAGCAAACAGGATTAGATAACCTGGTAGTCCACGCCGTAACGATGTGACTTGGAGGTTGTTCCCTTGA
GGAGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCTGGGGAGTACGGCCGCAAGGTTAAAACCTCAA
TGAATTGACGGGGGCCCCGACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACC
TACTCTTGACATCCAGCGAACTTTCCAGAGATGGATTGGTGCCTTCGGGAACTGAGACAGGTGCTGCA
TGGCTGTGCTCAGCTCGTGTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTTATCCTTTGT
TGCCAGCGATTCCGGTCGGGAACTCAAAGGAGACTGCCGGTGATAAACGGAGGAAGGTGGGGATGACGTC
AAGTCATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTC
GCGAGAGCAAGCGGACCTCACA AAGTGCCTCGTAGTCCGGATCGGAGTCTGCAACTCGACTCCGTGAAGT
CGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCC
TCACACCATGGGAGTGGGTTGCAAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACCTTGTG
ATTCATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTACC

Annexe B (suite):

>S3

TAGAGTTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAACAGACG
AGGAGCTTGCTCCTCTGACGTTAGCGGCGGACGGGTGAGTAACACGTTGGATAACCTACCTATAAGACTGG
GATAACTTCGGGAAACCGGAGCTAATACCGGATAACATGTTGAACCGCATGGTTCAACAGTGAAAGACGG
TCTTGCTGTCACTTATAGATGGATCCGCGCCGCATTAGCTAGTTGGTAAGGTAACGGCTTACCAAGGCAA
CGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGAAGTGAACACGGTCCAGACTCCTACGGGA
GGCAGCAGTAGGGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCAACGCCGCGTGAGTGAAGAAGGTC
TTCGGATCGTAAAACCTCTGTTATTAGGGAAGAACAATGTGTAAGTAACTATGCACGCTTTGACGGTACC
TAATCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGA
ATTATTGGGCGTAAAGCGCGCTAGGCGGTTTTTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGGA
GGGTCAATTGGAAACTGGAAAACCTTGAGTGCAGAAGAGGAAAAGTGGAAATTCATGTGTAGCGGTGAAATGC
GCAGAGATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAAGTGAACGCTGATGTGCGAAAG
CGTGGGGATCAAACAGGATTAGATAACCTGGTAGTCCACGCCGTAACCGATGAGTGCATAGTGTAGGGG
GTTTTCCGCCCTTAGTGTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAA
ACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAAATTCGAAGCAACGCCGAAGAA
CCTTACCAAATCTTGACATCCTCTGACCCCTCTAGAGATAGAGTTTTCCCTTCGGGGACAGAGTGACA
GGTGGTGCATGGTTGTCGTCAGCTCGTGTGTCGTCGATGTTGGTTAAGTCCCGCAACGAGCGCAACCCCTT
AAGCTTAGTTGCCATCATTAAAGTTGGGCACTCTAAGTTGACTGCCGTTGACAAACCGGAGGAAGGTGGGG
ATGACGTCAAATCATCATGCCCCCTTATGATTTGGGCTACACACGCTGCTACAATGGACAATACAAAGGGTA
GCGAAACCGCGAGGTCAAAGCAAATCCCATAAAAGTTGTTCTCAGTTCGGATTGTAGTCTGCAACTCGACTA
CATGAAGCTGGAATCGCTAGTAATCGTAGATCAGCATGCTACGGTGAATACGTTCCCGGGTCTTGTACAC
ACCGCCCGTACACCACGAGAGTTTGTAAACCCGAAGCCGGTGGAGTAACCTTTTTGGAGCTAGCCGTCG
AAGGTGGGACAAATGATTGGGGTGAAGTCGTAACAAGGTAGCCGTA

>S4

CGGAGAGTTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAATGGA
TTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACACGTTGGGTAACCTGCCATAAGACT
GGGATAACTCCGGGAAACCGGGGCTAATACCGGATAGCATTTTGAACCGCATGGTTCGAAATTGAAAGGC
GGCTTCGGCTGTCACTTATGGATGGACCCGCGTCGCATTAGCTAGTTGGTGAAGTAACGGCTCACCAAGG
CAACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCAGACTCCTACG
GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGATGAAG
GCTTTCCGGTTCGTA AAACTCTGTTGTTAGGGAAAGAACAAAGTCTAGTTGAATAAGCTGGCACCTTGACGG
TACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATC
CGGAATTATTGGGCGTAAAGCGCGCAGGTGGTTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCG
TGGAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAAAAGTGGAAATTCATGTGTAGCGGTGAA
ATGCGTAGAGATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAAGTGAACACTGAGGCGCG
AAAGCGTGGGGAGCAAACAGGATTAGATAACCTGGTAGTCCACGCCGTAACCGATGAGTGCATAAGTGTTA
GAGGGTTTTCCGCCCTTGTAGTGTGAAGTTAACGCATTAAGCACTCCGCCTGGGGAGTACGGCCGCAAGGC
TGAAACTCAAAGGAATTGACGGGGCCCCGCACAAGCGGTGGAGCATGTGGTTTAAATTCGAAGCAACCGGA
AGAACCTTACCAGGTCTTGACATCCTCTGAAAACCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGTGA
CAGGTGGTGCATGGTTGTCGTCAGCTCGTGTGTCGTCGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCC
TTGATCTTAGTTGCCATCATTAAAGTTGGGCACTCTAAGGTGACTGCCGGTGAACAAACCGGAGGAAGGTGG
GGATGACGTCAAATCATCATGCCCCCTTATGACCTGGGCTACACACGTCGCTACAATGGACGGTACAAAGAG
CTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCC
TACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGTGAATACGTTCCCGGGCCTTGTAC
ACACCGCCCGTACACCACGAGAGTTTGTAAACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCG
CCTAAGGTGGGACAGATGATTGGGGTGAAGTCGTAACAAGGTAGCCGT

Annexe B (suite):

>S5

TAGAGTTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCCTAATACATGCAAGTCGAGCGAACAGAAA
AGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAGTAACACGTTGGGCAACCTACCCCTATAGTTTGG
GATAACTCCGGGAAACCGGGGCTAATACCGAATAATCTCTTTTGGCTTCATGGTAAAAGACTGAAAGACGG
TTTTCGGCTGTGCTATAGGATGGGCCCGCGGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCC
ACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGG
AGGCAGCAGTAGGGAATCTTCCACAATGGGCGAAAGCCTGATGGAGCAACGCCGCGTGAGTGAAGAAGGT
TTTTCGGATCGTAAAACCTCTGTTGTAAGGGAAGAACAAGTACAGTAGTAACTGGCTGTACCTTGACGGTAC
CTTATTAGAAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGTAATACGTAGGTGGCAAGCGTTGTCCGG
AATTATTGGGCGTAAAGCGCGCGCAGGCGGTCTTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGG
AGGGTCATTGGAAAACCTGGGGGACTTGAGTGCAGAAGAGGAAAAGTGGAAATCCAAAGTGTAGCGGTGAAATG
CGTAGAGATTTGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACGACGCTGAGGCGCGAAA
GCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACGATGAGTGCATAAGTGTTAGGG
GGTTTCCGCCCTTAGTGTGCTGACGCTAACGCATTAAGCACTCCGCCCTGGGGAGTACGGTCGCAAGACTGA
AACTCAAAGGAATTGACGGGGGCGCACAAAGCGGTGGAGCATGTGGTTTAAATTCGAAGCAACGCGAAGA
ACCTTACCAGGCTTGACATCCCGTTGACCCTGTAGAGATATAGTTTCCCCTTCGGGGGCAACGGTGAC
AGGTGGTGCATGGTTGTCGTCAGCTCGTGTGAGATGTTGGTTAAGTCCCGCAACGAGCGCAACCCCT
TGATCTTAGTTGCCATCATTTAGTTGGGCACTCTAAGGTGACTGCCGTTGACAAACCGGAGGAAGGTGGG
GATGACGTCAAATCATCATGCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGATACAAACGGT
TGCCAACTCGCGAGAGGGAGCTAATCCGATAAAAGTCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCT
ACATGAAGCCGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACA
CACCGCCCGTACACCACGAGAGTTTGTAAACCCGAAAGTCGGTGAGGTAACCTTTTGGAGCCAGCCGCC
GAAGGTGGGATAGATGATTGGGGTGAAGTCGTAACAAGGTAGCCGTA

>S6

TAGAGTTTTGATCCTGGCTCAGATTGAACGCTGGCGGACGGCCTAACACATGCAAGTCGAGCGGATGACGG
GAGCTTGCTCCTTGATTACAGCGGCGGACGGGTGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGACAA
CGTTTTCGAAAGGAACGCTAATACCGCATAACGTCCTACGGGAGAAAAGCAGGGGACCTTCGGGCCCTTGCCT
ATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGGGGTAATGGCTCACCAAGGCGACGATCCGTAACCTG
GTCTGAGAGGATGATCAGTCACACTGGAAGTGGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG
AATATTGGACAATGGGCGAAAAGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTA
GCACTTTAAGTTGGGAGGAAGGGCATTAAACCTAATACGTTAGTGTTTTGACGTTACCGACAGAATAAGCA
CCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTA
AAGCGCGCGTAGGTGGTTTGTAAAGTTGGATGTGAAAGCCCCGGGCTCAACCTGGGAACTGCATCCAAAA
CTGGCAAGCTAGAGTACGGTAGAGGGTGGTGGAAATTTCTGTGTAGCGGTGAAATGCGTAGATATAGGAA
GGAACACCAGTGGCGAAGGCGACACCTGGACTGATACTGACACTGAGGTGCGAAAAGCGTGGGGAGCAAA
CAGGATTAGATAACCTGGTAGTCCACGCGGTAACGATGTCAACTAGCCGTTGGAATCCTTGAGATTTTA
GTGGCGCAGCTAACGCATTAAGTTGACCGCTGGGGAGTACGGCCGCAAGGTAAAACTCAAATGAATTG
ACGGGGGCGCACAAAGCGGTGGAGCATGTGGTTTAAATTCGAAGCAACGCGAAGAACCCTTACCAGGCCTT
GACATGCAGAGAACCTTCCAGAGATGGATTGGTGCCTTCGGGAACTCTGACACAGGTGCTGCATGGCTGT
CGTCAGCTCGTGTGAGATGTTGGGTTAAGTCCCCTAACGAGCGCAACCCCTTGTCCCTTAGTTACCAGC
ACGTTATGGTGGGCACTCTAAGGAGACTGCCGGTGCACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCA
TCATGGCCCTTACGGCCTGGGCTACACACGTGCTACAATGGTTCGGTACAGAGGGTTGCCAAGCCGCGAGG
TGGAGCTAATCTCACAAAACCGATCGTAGTCCGGATCGCAGTCTGCAACTCGACTGCGTGAAGTCGGAAT
CGCTAGTAATCGCGAATCAGAATGTGCGGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTACAC
CATGGGAGTGGGTTGCACCAGAAGTAGCTAGTCTAACCTTCGGGAGGACGGTTACCACGGTGTGATTCAT
GACTGGGGTGAAGTCGTAACAAGGTAGCCGT

Annexe B (suite):

>S7

AGAGTTTGGATCCTGGCTCAGATTGAACTCTGGCGGCAGGCCTCACACATACAAGTCGAGCGGTAGCACAG
AGAGCTTGCTCTCGGCTGACGAGCGGCGGACGGGTGAGTAATGTCTGGGAACTGCCGTGATGGAGGGGGA
TAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTT
GCCATCAGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAG
CTGGTCTGAGAGGATGACCAGCCACACTGGAAGTACGACACGGTCCAGACTCCTACGGGAGGCAGCAGTG
GGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGT
AAAGTACTTTTCAGCGGGGAGGAAGGCGATAAGGTTAATAACCTTGTTCGATTGACGTTACCCGCAGAAGAA
GCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGC
GTAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAAGTGCATTCCG
AAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCT
GGAGGAATACCGGTGGCGAAGGCGGCCCTTGGACAAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGC
AAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACGATGTTCGACTTGGAGGTTGTGCCCTTGAGGCG
TGGCTTCCGGAGCTAACGCGTAAAGTCGACCGCTGGGGAGTACGGCCGCAAGGTTAAAACCTCAAATGAA
TTGACGGGGGCCCCGACAAGCGGTGGAGCATGTGGTTAATTCGATGCAACGCGAAGAACCCTTACCTACT
CTTGACATCCAGAGAACTTTCCAGAGATGGATTGGTGCCTTCGGGAACTCTGAGACAGGTGCATGGC
TGTTCGTGAGTTCGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTGGC
AGCGGTTAGGCCCGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGT
CATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGCGA
GAGCAAGCGGACCTCATAAAGTGCCTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGA
ATCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCAC
ACCATGGGAGTGGGTTGCAAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACCTTGTGATTC
ATGACTGGGGTGAAGTCGTAACAAGGTAGCC

>S8

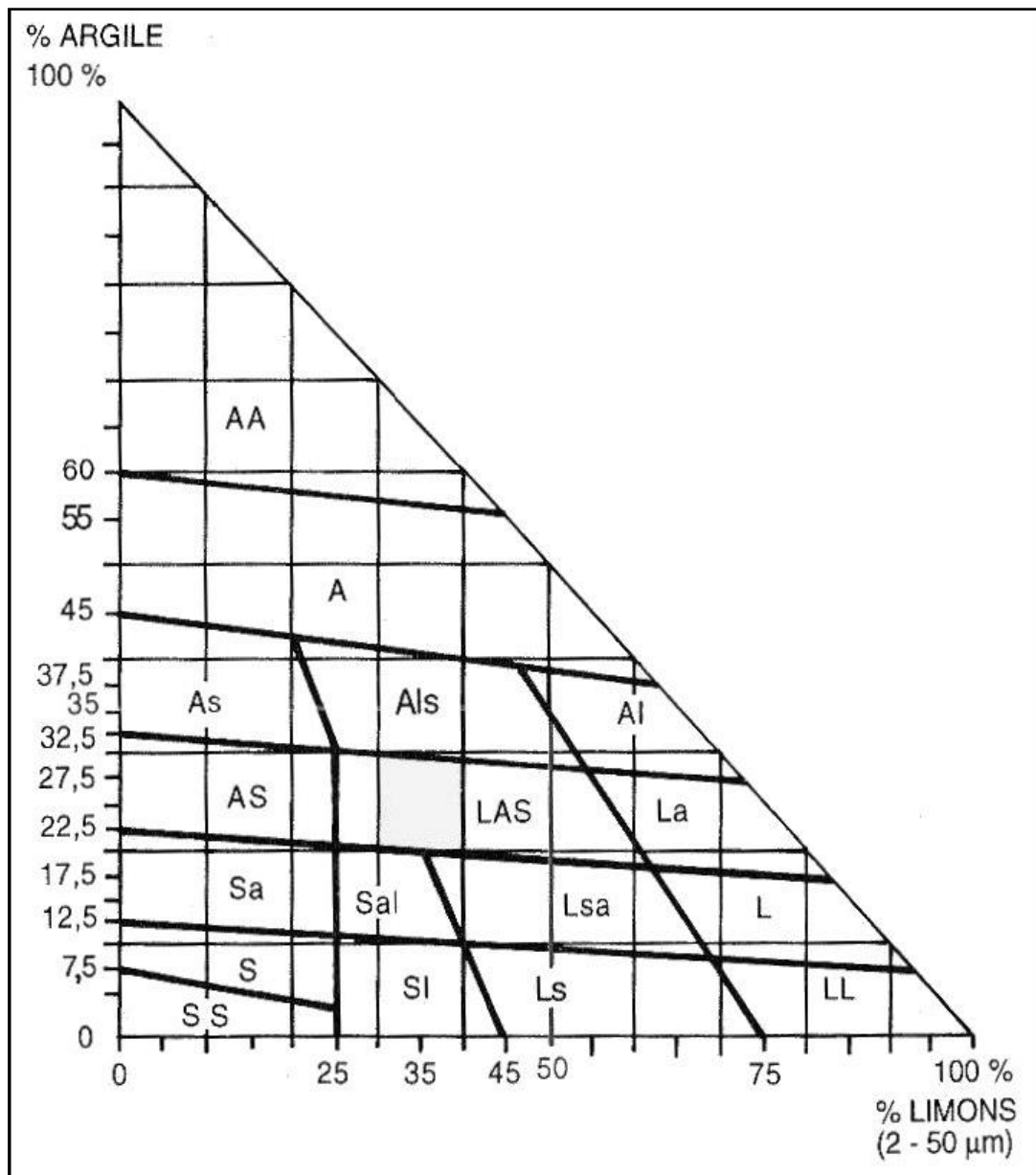
TAGAGTTTGGATCCTGGCTCAGAGTGAACGCTGGCGGTAGGCCTAACACATGCAAGTCGAACGGCAGCACA
GTAAGAGCTTGCTCTTATGGGTGGCGAGTGGCGGACGGGTGAGGAATACATCGGAATCTACCTTTTTCGTG
GGGGATAACGTAGGGAACTTACGCTAATACCGCATAACGACTTCGGGTGAAAGCAGGGGACCTTCGGGC
CTTGGCGGGATAGATGAGCCGATGTCCGATTAGCTAGTTGGCGGGGTAAAGGCCACCAAGGCGACGATC
CGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGAAGTACGACACGGTCCAGACTCCTACGGGAGGCAG
CAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATACCGCGTGGGTGAAGAAGGCCTTCGG
GTTGTAAAGCCCTTTTGTGGGAAAGAAAAGCAGTCGATTAATACTCGGTTGTTCTGACGGTACCCAAAG
AATAAGCACCGGCTAACTTCGTGCCAGCAGCCGCGTAATACGAAGGTTGCAAGCGTTACTCGGAATTAC
TGGGCGTAAAGCGTGCCTAGGTGGTGTGTTAAGTCTGTGTTGAAAGCCCTGGGCTCAACCTGGGAATTGC
AGTGATACTGGGCGACTAGAGTGTGGTAGAGGGTAGTGAATTCGGGTGTAGCAGTGAATGCGTAGA
GATCGGGAGGAACATCCATGGCGAAGGCAGCTACCTGGACCAACACTGACACTGAGGCACGAAAGCGTGG
GGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAAACGATGCGAACTGGATGTTGGGTGCAATT
TGGCACGCAGTATCGAAGCTAACGCGTAAAGTTCGCCGCTGGGGAGTACGGTCGCAAGACTGAAACTCA
AAGGAATTGACGGGGGCCCCGACAAGCGGTGGAGTATGTGGTTAATTCGATGCAACGCGAAGAACCCTTA
CCTGGTCTTGACATGTGAGAACTTTCCAGAGATGGATTGGTGCCTTCGGGAACTCGAACACAGGTGCTG
CATGGCTGTGCTGAGTTCGTGCTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCTTAA
GTTGCCAGCACGTAATGGTGGGAACTCTAAGGAGACCGCCGTTGACAAACCGGAGGAAGGTGGGGATGAC
GTCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTACTACAATGGTAGGGACAGAGGGCTGCAAA
CCCCGAGGGGCAAGCCAATCCCAGAAACCCTATCTCAGTCCGGATTGGAGTCTGCAACTCGACTCCATGA
AGTCCGAATCGCTAGTAATCGCAGATCAGCATTGCTGCGGTGAATACGTTCCCGGGCCTTGTACACACCG
CCCGTCACACCATGGGAGTTTGTGTCACCAGAAGCAGGTAGCTTAACCTTCGGGAGGGCGCTTGCCACGG
TGTGGCCGATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTA

Annexe B (suite):

>S9

```
CCGTAGAGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGGTAGC
ACAGAGAGCTTGCTCTCGGGTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAACTGCCTGATGGAGG
GGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCC
TCTTGCCATCAGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAATGGCTCACCTAGGCGACGATCC
CTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAGTACGACACGGTCCAGACTCCTACGGGAGGCAGC
AGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGG
TTGTAAAGTACTTTCAGCGGGGAGGAAGGTGTTGCGGTTAATAACCGCAGCAATTGACGTTACCCGCAGA
AGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACT
GGGCGTAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAAGTGA
TTCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAG
ATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAAGACTGACGCTCAGGTGCGAAAGCGTGGG
GAGCAAACAGGATTAGATAACCTGGTAGTCCACGCCGTAACGATGTCGACTTGGAGGTTGTTCCCTTGA
GGAGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCTGGGGAGTACGGCCGCAAGGTTAAAAC TCAA
TGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTAATTTCGATGCAACGCGAAGAACCTTACC
TACTCTTGACATCCAGAGAACTTTCAGAGATGGATTGGTGCCTTCGGGAACTCTGAGACAGGTGCTGCA
TGGCTGTCGTCAGCTCGTGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCC TATCCTTGT
TGCCAGCGGCTAGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTC
AAGTCATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTC
GCGAGAGCAAGCGGACCTCATAAAGTGCCTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGT
CGGAATCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCG
TCACACCATGGGAGTGGGTTGCAAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCAC TTTGTG
ATTCATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTA
```

Annexe C : Le triangle de texture GEPPA simplifié par Baize (1988)



المخلص:

إن التقنيات التقليدية للتعرف على البكتيريا ليست كافية لدراسة بكتيريا التربة. بالمقابل، التقنيات البيولوجية الجزيئية تتميز بسعة تمييز عالية و تمثل الأساليب الحالية لتحديد مختلف المتغيرات التصنيفية البكتيرية، مما يسمح باتباع نهج معمق و دقيق في دراسات التنوع. الاستكشاف الكامل لهذه الأخيرة دائما ما يشكل تحديا لعلماء الأحياء المجهرية، وهذا يرجع إلى تعقيد و عدم تجانس مصفوفة التربة (وجود البيئات الدقيقة) و وجود المجموعات الصغيرة عدديا، و نادرا ما اكتشفت من خلال التقنيات التقليدية.

في اطار هذه الأطروحة، استعملنا ثلاثة تقنيات جزيئية (MALDI-TOF MS و 16S و métagenomique) لتحقيق الأهداف التالية: إظهار مدى انطباق MALDI-TOF MS في مجال ميكروبيولوجيا التربة و قياس التنوع البكتيري في منطقة المريج (مدينة قسنطينة، الجزائر) من خلال métagenomique و دراسة تأثير الخصائص الفيزيائية على التنوع البيولوجي. في الجزء الأول، أكدنا على قابلية MALDI-TOF MS على تحديد جميع العزلات البكتيرية لدينا (9 سلالات) حتى مستوى الأنواع بنتيجة أكبر من 2, 000 وفي وقت قياسي مقارنة بالتقنية المظهرية و الوراثة (ADNr 16S) وفي الجزء الثاني، pyroséquençage 454 des métagénomomes لعينات التربة الثلاثة (بحيرة التربة، و التربة الغابات و التربة المزروعة) يكشف عن وجود 11 من الكائنات الحية و 20 عائلة و 13 جنسا. أظهر حساب مؤشرات التنوع (l'indice de Shannon-Weaver, l'inverse de l'indice de Simpson et l'indice Chao1) أن التربة في منطقة المريج يحتوي على تجانس كبير للبكتيريا و على توزيع عادل لثراء الأنواع التي يمكن ربطها بدرجة الحموضة القلوية لهذه المنطقة و المحتوى العضوي المرتفع.

يظهر عملنا بشكل واضح أنه من الممكن تحسين الوصول إلى مختلف لبكتيري التربة من خلال MALDI-TOF MS كنظام تحديد روتيني لمختبرات علم الأحياء الدقيقة في التربة و métagenomique لدراسات التنوع.

الكلمات المفتاحية: بكتيريا التربة، MALDI-TOF MS، ADNr 16S، métagenomique، التنوع البكتيري.

Abstract :

The traditional bacterial culture techniques are not sufficient to study the terrestrial bacterial populations. By against, molecular biology techniques have a power of high discrimination and they are the current methods for the identification of different bacterial taxonomic variants, allowing a finer approach in the diversity studies. The full exploration of bacterial diversity has always been a challenge; this is due to the complexity, to heterogeneity of the soil matrix and the existence of populations numerically small and rarely detected by conventional techniques.

In this thesis, we used three molecular techniques (MALDI-TOF MS, the 16S and metagenomics) to achieve the following objectives: (i) to demonstrate the applicability of the MALDI-TOF MS in the field of soil microbiology, (ii) measure the bacterial diversity of the region "El-Meridj" (city of Constantine, Algeria) through metagenomics and (iii) study the effect of physicochemical characteristics on this biodiversity. In the first part, we confirmed the susceptibility of proteomics technology MALDI-TOF MS to identify all our bacterial isolates (9 strains) in species level with a score greater than 2,000 and in record time compared to phenotypic techniques and genotypic techniques (amplification and sequencing of the 16S rDNA). In the second part, the 454 pyrosequencing of metagenomes isolated from three soil samples (lake soil, forest soil and cultivated soil) reveals the presence of 11 *phyla*, 20 families and 13 genera. Thus, the calculation of diversity indices (Shannon-Weaver index, Inverse Simpson index and Chao1 index) showed that the soil in the area "El-Meridj" contains significant heterogeneity of bacterial population and a fair distribution of species richness which can be correlated to the alkaline pH of this region and the high organic content.

Our work clearly shows that it is possible to improve access to different telluric bacterial taxonomic variations by the application of MALDI-TOF MS as routine identification system to soil microbiology laboratories and metagenomics for diversity studies.

Keywords: Soil bacteria, MALDI-TOF MS, 16S rDNA, metagenomics, bacterial diversity.

Nom : KHELILI	Diplôme : Doctorat 3 ^{ème} cycle LMD
Prénom : Kaoutar	Date de soutenance : 31 mai 2015

Titre : Application des outils biomoléculaires 16S, MALDI-TOF et métagénomique pour la détermination des variantes taxonomiques bactériennes telluriques.

Résumé :

Les techniques traditionnelles de culture bactérienne ne suffisent pas pour étudier les populations bactériennes telluriques. Par contre, les techniques moléculaires ont un pouvoir de discrimination élevé et constituent les méthodes actuelles de l'identification bactérienne. L'exploration complète de la diversité bactérienne représente un défi dû à la complexité, à l'hétérogénéité de la matrice sol et à l'existence de populations numériquement faibles et rarement détectées par les techniques conventionnelles.

Dans le cadre de cette thèse, nous avons fait appel à trois techniques moléculaires (le MALDI-TOF MS, le 16S et la métagénomique) afin d'atteindre les objectifs suivants : (i) mettre en évidence l'applicabilité du MALDI-TOF MS dans le domaine de la microbiologie du sol, (ii) mesurer la diversité bactérienne de la région « El-Meridj » (ville de Constantine, Algérie) par l'approche métagénomique et (iii) étudier l'effet des caractéristiques physicochimiques sur cette biodiversité. Dans la première partie, nous avons pu confirmer la susceptibilité de la technique protéomique MALDI-TOF MS à identifier tous nos isolats bactériens (9 souches) au niveau de l'espèce avec un score supérieur à 2,000 et dans un délai record par rapport aux techniques phénotypiques et génotypiques (amplification et séquençage de l'ADNr 16S). Dans la deuxième partie, le pyroséquençage 454 des métagénomomes issus des trois échantillons de sol (sol du lac, sol forestier et le sol cultivé) révèle la présence de 11 *phyla*, 20 familles et 13 genres. Le calcul des indices de diversité (l'indice de Shannon-Weaver, l'inverse de l'indice de Simpson et l'indice Chao1) a montré que le sol de la région " El-Meridj " renferme une forte hétérogénéité de la population bactérienne et une répartition équitable de la richesse spécifique qui peuvent être corrélées au pH alcalin de cette région et à la teneur élevée en matière organique.

Notre travail montre clairement qu'il est possible d'améliorer l'accès aux différentes variantes taxonomiques bactériennes telluriques par l'application du MALDI-TOF MS comme système d'identification de routine aux laboratoires de microbiologie du sol et la métagénomique pour des études de diversité.

Mots clés : Bactéries telluriques, MALDI-TOF MS, ADNr 16S, métagénomique, diversité bactérienne.