



**HAL**  
open science

## Diversité bactérienne des sols : accès aux populations à effectifs moniritaires “ the rare biosphere ”

Aurélie Faugier

► **To cite this version:**

Aurélie Faugier. Diversité bactérienne des sols : accès aux populations à effectifs moniritaires “ the rare biosphere ”. Sciences du Vivant [q-bio]. Ecole Centrale de Lyon, 2010. Français. NNT : . tel-00502105v1

**HAL Id: tel-00502105**

**<https://theses.hal.science/tel-00502105v1>**

Submitted on 13 Jul 2010 (v1), last revised 15 Feb 2011 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

présentée devant

## L'ÉCOLE CENTRALE DE LYON

pour obtenir le grade de

### DOCTEUR

Préparée au sein de

### L'ÉCOLE DOCTORALE ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE DE LYON

par

**Aurélie FAUGIER**

---

## Diversité bactérienne des sols : accès aux populations à effectifs monitaires « the rare biosphere »

---

*soutenue le 12 mars 2010 devant la commission d'examen*

### J U R Y

<b>Yves DESSAUX</b>	Directeur de recherche CNRS – Gif sur Yvette	<i>Président</i>
<b>Pascale BAUDA</b>	Professeur – Université de Metz	<i>Rapporteur</i>
<b>Pierre PEYRET</b>	Professeur – Université Blaise Pascal Clermont-Ferrand	<i>Rapporteur</i>
<b>Pascal SIMONET</b>	Directeur de recherche CNRS – Ecole centrale de Lyon	<i>Directeur de thèse</i>
<b>Simonetta GRIBALDO</b>	Chargée de recherche – Institut Pasteur – Paris	<i>Examineur</i>
<b>Lionel RANJARD</b>	Directeur de recherche INRA – Dijon	<i>Examineur</i>



## REMERCIEMENTS

*J'exprime mes profonds remerciements à mon directeur de thèse, Pascal Simonet, pour son encadrement durant ces trois années de thèse. Merci pour ses conseils avisés et pour ses encouragements. Merci de m'avoir donné la chance de travailler sur un sujet qui m'a passionnée.*

*Merci à la région Rhône Alpes qui a financé ma thèse, merci au directeur du laboratoire Ampère pour son accueil.*

*Je remercie les membres du jury qui ont bien voulu consacrer une partie de leur temps à ma thèse: Simonetta Gribaldo, Yves Dessaux et Lionel Ranjard. Je remercie en particulier ceux qui ont accepté la lourde charge d'être les rapporteurs de ce travail: Pascale Bauda et Pierre Peyret.*

*Un merci particulier à Maude, à qui je dois beaucoup tant sur le plan professionnel que personnel. Depuis 5 ans nous avons suivi le même chemin, sa présence va beaucoup me manquer. Merci à Tim pour ses discussions très intéressantes, Seb pour son investissement dans l'analyse des puces, Manu pour la formation R, Saliou pour ses très bons conseils et merci à Nathalie pour m'avoir initiée au clonage !! Merci aussi à Yoann, Cédric, Sibel, Kevin, Julie, Delina, Catherine et Nico, qui m'ont aidée d'une manière ou d'une autre durant ces trois années. Merci aussi pour les bons moments passés ensemble...*

*Merci à tous mes collègues de laboratoire : Sandrine, Jean-Michel, Sam, Tom, Alban, Jérémy, Laurine, Barbara, Marina, Mayssa, Monique, Isabelle, Richard. Merci aux personnes qui ont été impliquées à un moment ou un autre dans le projet « genefish » : Xiaojun, Céline, Laure, Sam...*

*Je remercie ma famille, qui m'a toujours aidée et soutenue. Plus particulièrement mes parents, Ludovic et Valère mes frères, et Marie ma sœur. Sans oublier Léa, Cassandra et Chléa. Et bien sur un énorme merci à Fred pour son soutien immense.*



# SOMMAIRE

<b>Introduction générale.....</b>	<b>11</b>
<b>CHAPITRE 1 : Synthèse bibliographique</b>	
<b>Evaluation et origine de la diversité bactérienne, plus particulièrement dans l'environnement sol.....</b>	<b>17</b>
<b>1. Introduction.....</b>	<b>19</b>
<b>2. Techniques d'étude de la diversité bactérienne .....</b>	<b>20</b>
<b>2.1. La culture bactérienne .....</b>	<b>20</b>
2.1.1. Approches traditionnelles .....	20
2.1.2. Classification et identification bactérienne appliquée aux isolats bactériens .....	20
<b>2.2. Identification bactérienne par analyse de l'ADN extrait de l'environnement (approche métagénomique) .....</b>	<b>23</b>
2.2.1. Caractérisation des communautés bactériennes.....	24
2.2.2. Apport des nouvelles technologies à haut débit.....	27
2.2.2.1. Principe et applications des puces à ADN.....	27
2.2.2.2. Le séquençage .....	30
<b>3. Plasticité génomique bactérienne .....</b>	<b>37</b>
<b>3.1. Processus d'évolution verticale.....</b>	<b>38</b>
3.1.1. Mutations spontanées.....	38
3.1.2. Réarrangements d'ADN.....	40
<b>3.2. Processus d'évolution horizontale.....</b>	<b>42</b>
3.2.1. Transformation naturelle.....	42
3.2.2. La transduction.....	43
3.2.3. La conjugaison .....	45
<b>4. Les bactéries dans le sol.....</b>	<b>46</b>
<b>4.1. Le sol et ses composants .....</b>	<b>46</b>
<b>4.2. Les bactéries dans le sol .....</b>	<b>49</b>
<b>4.3. Obtention du métagénome de sols.....</b>	<b>51</b>
<b>4.4. Facteurs influençant la structure des communautés bactériennes du sol .....</b>	<b>54</b>
4.4.1. Facteurs abiotiques.....	54
4.4.2. Facteurs biotiques .....	55
<b>5. Conclusions.....</b>	<b>57</b>

<b>CHAPITRE 2 : Combinaison d'une approche conceptuelle avec l'utilisation d'une technique sensible pour évaluer la diversité bactérienne des sols.....</b>	<b>59</b>
<b>PARTIE I : Evolution des communautés bactériennes inoculées dans des sols stérilisés au cours du temps.....</b>	<b>61</b>
<b>1. Introduction.....</b>	<b>61</b>
<b>2. Matériels &amp; méthodes.....</b>	<b>63</b>
<b>2.1. Préparation et conditionnement des sols .....</b>	<b>63</b>
<b>2.2. Extraction et inoculation des communautés bactériennes.....</b>	<b>65</b>
<b>2.3. Préparations des contrôles négatifs et positifs .....</b>	<b>66</b>
<b>2.4. Extraction de l'ADN et amplification de la région intergénique 16S-23S.....</b>	<b>67</b>
<b>2.5. Empreintes moléculaires RISA des communautés bactériennes extraites des échantillons de sol .....</b>	<b>68</b>
<b>2.6. Les puces à ADN taxonomiques .....</b>	<b>68</b>
<b>2.7. Amplification du gène 16S r DNA.....</b>	<b>68</b>
<b>2.8. Protocole d'hybridation sur puces à ADN .....</b>	<b>69</b>
<b>2.9. Lecture des lames et traitement des données .....</b>	<b>70</b>
<b>2.10. Reproductibilité des résultats .....</b>	<b>70</b>
<b>2.11. Analyses statistiques .....</b>	<b>70</b>
<b>3. Résultats.....</b>	<b>71</b>
<b>3.1. Validation de l'approche conceptuelle et criblage des premiers résultats .....</b>	<b>71</b>
3.1.1. Extraction et amplification des échantillons d'ADN .....	71
3.1.2. Contrôles négatifs et positifs.....	73
<b>3.2. Evaluation des structures des communautés bactériennes.....</b>	<b>75</b>
<b>3.3. Reproductibilité des résultats .....</b>	<b>88</b>
<b>4. Discussion .....</b>	<b>89</b>
<b>4.1. Criblage des premiers résultats.....</b>	<b>89</b>
<b>4.2. Contrôles négatifs et positifs de l'expérience .....</b>	<b>91</b>
<b>4.3. Reproductibilité de l'expérience.....</b>	<b>92</b>
<b>4.4. Evaluation des structures des communautés bactériennes.....</b>	<b>93</b>
<b>5. Conclusions.....</b>	<b>97</b>

PARTIE II : Modifications de communautés bactériennes induites par des changements de sols.....	99
<b>1. Introduction.....</b>	<b>99</b>
<b>2. Matériels &amp; Méthodes .....</b>	<b>100</b>
<b>2.1. Choix des échantillons d’ADN.....</b>	<b>100</b>
<b>2.2. Traitement des données obtenues par les puces à ADN.....</b>	<b>101</b>
<b>2.3. Identification taxonomique des sondes oligonucléotidiques .....</b>	<b>101</b>
<b>3. Résultats &amp; Discussion .....</b>	<b>102</b>
<b>3.1. Identification des taxons bactériens détectés dans les sols inoculés par chacune des communautés par rapport aux inocula de départ. ....</b>	<b>102</b>
<b>3.2. Nombre de genres différents cumulés obtenus par l’inoculation d’une communauté dans des sols stérilisés.....</b>	<b>121</b>
<b>4. Conclusions.....</b>	<b>125</b>
<b>CHAPITRE 3 : Développement d’une nouvelle approche métagénomique : capture in situ de l’ADN métagénomique.....</b>	<b>127</b>
<b>1. Introduction.....</b>	<b>129</b>
<b>2. Matériels &amp; Méthodes .....</b>	<b>131</b>
<b>2.1. Souches bactériennes et conditions de culture .....</b>	<b>131</b>
<b>2.2. Réactions enzymatiques .....</b>	<b>131</b>
<b>2.3. Clonage .....</b>	<b>133</b>
<b>2.4. Préparation des cellules pour l’électroporation.....</b>	<b>133</b>
<b>2.5. Transformation .....</b>	<b>134</b>
<b>2.6. Criblage Phénotypique.....</b>	<b>134</b>
<b>2.7. Test d’efficacité .....</b>	<b>135</b>
<b>3. Résultats &amp; discussion .....</b>	<b>135</b>
<b>3.1. Principe de l’approche métagénomique .....</b>	<b>136</b>
<b>3.2. Présentation des différents composants.....</b>	<b>140</b>
3.2.1. Les gènes suicides et leur régulation .....	140
3.2.2. Les zones de recombinaison .....	142
3.2.3. Définitions et modification des souches réceptrices .....	142
<b>3.3. Construction et efficacité du système suicide final .....</b>	<b>143</b>
3.3.1. Construction de la cassette toxique inductible à deux gènes .....	143



3.3.2. Efficacité du système toxique .....	145
3.3.3. Clonage de la cassette toxique dans pBAD342G1G2.....	147
3.3.4. Clonage de la cassette toxique dans le plasmide pBAD35 .....	148
<b>4. Conclusions &amp; perspectives.....</b>	<b>150</b>
<b>Conclusions générales.....</b>	<b>151</b>
<b>Références bibliographiques.....</b>	<b>157</b>

## LISTE DES ABREVIATIONS

ACP	Analyse en Composante Principale
ARDRA	Amplified Ribosomal DNA Restriction Analysis
(A) RISA	(Automated) Ribosomal Intergenic Spacer Analysis
AB	AntiBiotique
ADN	Acide DéoxyRiboNucléique
Ara	Arabinose
ARN(m)	Acide RiboNucléique (messenger)
BLAST	Basic Local Alignment Sequence Tool
Cm	Chloramphénicol
CP1	Composante Principale 1
CP2	Composante Principale 2
CS1	Communauté du Sol 1 (Côte Saint André)
CS2	Communauté du Sol 2 (Montrond)
CS3	Communauté du Sol 3 (Chinon)
CS5	Communauté du Sol 5 (Kenya)
CL	Communauté Lacustre
DGGE	Denaturing Gradient Gel Polymorphism
dNTP	désoxy Nucléotide Tri Phosphate
DO	Densité Optique
EDTA	Ethyl Diamine Tétra Acétate
EGM	Elément Génétique Mobile
G	Gramme
GM	Génétiquement Modifiée
H	Heure
IPTG	IsoPropyl $\beta$ -D-ThioGalactopyranoside
Kb	Kilobase
Kn	Kanamycine
L	Litre
LBm	Luria Bertani Modifié
Min	Minute
ml	Millilitre

MM	Millimètre
NCBI	National Center for Biotechnological Information
ORF	Open Reading Frame
Pb	Paire de Base
PCR	Polymerase Chain Reaction
PVPP	PolyVinylPyrrolidone
QSP	Quantité Suffisante Pour
RPM	Rotor Par Minute
S	Seconde
Sb	Simple brin
Sm	Streptomycine
SMC	Sites Multiples de Clonage
Sp	Spectinomycine
SSCP	Single Strand Conformation Polymorphism
TBE	Tris Borate EDTA
Tc	Tétracycline
THG	Transfert Horizontal de Gène
Tm	Temperature melting
TGGE	Temperature Gardient Gel Electrophoresis
T-RFLP	Terminal Restriction Fragment Length Polymorphism
Trp	Tryptophane
Ts	Thermosensible
U	Unité
UFC	Unités Formant Colonies
UV	Ultra Violet
µL	Microlitre

## Introduction générale

Les bactéries et archées présentes dans presque tous les types de biotopes rencontrés sur terre sont les organismes les plus ubiquitaires de notre planète. Elles peuvent être isolées du sol, des eaux douces, marines ou saumâtres, de l'air, des profondeurs océaniques, des déchets radioactifs (Fredrickson *et al.*, 2004), de la croûte terrestre, sur la peau et dans l'intestin des animaux mais elles peuvent également être rencontrées dans des environnements plus extrêmes tels que des lacs salés, des banquises, des sources d'eau chaude, etc... On estime que le nombre de cellules bactériennes dans la biosphère se situerait entre 4 et  $6.10^{30}$  (Whitman *et al.*, 1998), représentant une grande partie de la biomasse de notre planète.

Les bactéries ont une importance considérable dans les cycles biogéochimiques comme par exemple dans ceux du carbone ou de l'azote avec un rôle fondamental dans la fixation de l'azote atmosphérique, fonction qui a beaucoup été étudiée depuis plusieurs dizaines d'années (Vitousek, 1991). La plupart des bactéries sont inoffensives pour l'Homme, certaines étant même essentielles au bon fonctionnement de l'organisme. Il existe cependant de nombreuses espèces pathogènes à l'origine des maladies infectieuses comme le choléra, la syphilis, la peste, l'anthrax, la tuberculose (Kumar & Tuteja, 2009, Xu *et al.*, 2010). L'Homme a très tôt utilisé les propriétés des bactéries pour se nourrir, se soigner. Aujourd'hui les domaines d'application sont très variés. En agroalimentaire, les bactéries comme *Lactobacillus*, *Lactococcus* ou *Streptococcus*, combinées aux levures et moisissures interviennent dans l'élaboration d'aliments fermentés comme les fromages, les yaourts, la bière....(Hanniffy *et al.*, 2009).

La capacité des bactéries hétérotrophes à dégrader une large variété de composés organiques est exploitée pour le traitement des sols pollués dans des stratégies de bioremédiation ou pour le traitement des eaux usées (Furukawa, 2003). Des bactéries sont également utilisées dans les fosses septiques pour en assurer l'épuration. En agriculture, certaines bactéries peuvent être utilisées à la place de pesticides en lutte biologique pour combattre les parasites des plantes (Lucy *et al.*, 2004) (ex : *Bacillus thuringiensis*), d'autres bactéries vont avoir un effet bénéfique sur la croissance de plantes comme les PGPR (Lugtenberg & Kamilova, 2009).

Dans le domaine médical des isolats bactériens ont été sélectionnés sur leurs capacités de production de métabolites secondaires utilisables comme médicaments et le génie génétique a permis d'en améliorer les performances ou même de faire produire par des

bactéries des composés initialement issus d'autres organismes (production d'antibiotiques, d'insuline, de l'hormone de croissance, ...) (Wong, 1995, Demain, 1999, Zhou *et al.*, 2009). Si un millilitre d'eau douce contient environ 1 million de cellules bactériennes c'est au niveau du sol que se trouve le principal réservoir en termes de nombre de cellules et de diversité avec environ  $10^8$ - $10^9$  bactéries par gramme de sol.

Cependant, la majorité de ces bactéries ne sont encore ni identifiées ni caractérisées car elles demeurent récalcitrantes à la culture *in vitro* sur des milieux de culture qui ont pu leur être proposés (Rappe & Giovannoni, 2003). Ces limites imposées par la culture *in vitro* ont permis le développement d'approches s'en affranchissant. L'arrivée de la métagénomique dans les années 1990 a restreint l'étude des bactéries à leur ADN directement extrait de l'environnement (McHardy & 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.

L'exploration de la diversité bactérienne, notamment du sol, reste toutefois imparfaite et très incomplète (Aschby, 2007). Ces limitations sont d'ordre technique, les méthodes d'extraction de l'ADN ne permettant pas d'accéder à la totalité de l'information génétique bactérienne du fait principalement des caractéristiques physico-chimiques d'environnements complexes comme le sol (Saano, 1995). Comme nous le verrons ultérieurement, les cellules bactériennes ainsi que l'ADN libre peuvent être adsorbées irréversiblement sur des particules de sol, telles que les argiles (Fortin *et al.*, 2004). L'étape de la purification de l'ADN est aussi problématique, les interférences avec des substances humiques co-extraites conduisant également à une diminution du rendement d'extraction (Harry, 1999). Les étapes d'amplification de l'ADN par la technique de la PCR rajoutent un biais dans l'analyse des communautés bactériennes. Quant au séquençage d'échantillons environnementaux qui contiennent plusieurs micro-organismes en quantité variable, les nouvelles générations de séquençage ont également leurs limites liées aux erreurs de séquençage, au traitement des masses de données générées et à l'ampleur considérable de l'effort requis pour atteindre la totalité de l'ADN des espèces. Ces limites sont aussi conceptuelles, du fait de la colonisation par certaines populations de microenvironnements très protégés (Hattori, 1988) et surtout à

des différences de taille des populations. En effet, les analyses sur la distribution de l'abondance des bactéries indiquent que la diversité est divisée en 2 principales composantes. La première inclut les quelques espèces abondantes, qui représente la partie la plus étudiée des communautés, la deuxième nommée la « rare biosphere » (Sogin, 2006) comprend un nombre élevé d'espèces rares qui représentent la majorité de la diversité (Pedros-Alio, 2006) et qui a été démontrée en premier par pyroséquençage des bactéries marines dans les profondeurs de l'atlantique (Sogin, 2006).

Nous connaissons encore peu de choses sur leur rôle écologique et fonctionnel, même ces bactéries « rares » doivent jouer un rôle fondamental dans le fonctionnement biologique du sol. Certaines peuvent se développer quand les conditions du sol changent, et ainsi maintenir différentes fonctions bactériennes quand les bactéries les plus abondantes et actives sont affectées par les modifications environnementales. Certaines d'entre elles sur-expriment certains de leurs gènes impliqués dans des fonctions écologiques majeures pour compenser la faible taille de leur population (Taylor, 2002). Ces très nombreuses bactéries à très faibles effectifs constituent aussi un réservoir d'information génétique et notamment de gènes qui peuvent être transférés entre les différents membres de la communauté bactérienne par des mécanismes de transferts horizontaux qui sont un puissant moteur de l'évolution bactérienne, permettant notamment la très rapide adaptation de ces microorganismes aux changements biotiques ou abiotiques de leur environnement (Lawrence, 1999). Cette importante disparité au niveau de la taille des populations bactériennes liée à un niveau très élevé du nombre d'espèces rend l'inventaire des espèces présentes extrêmement ardu, sans même prendre en compte la diversité des fonctions réalisées, leur redondance, les interactions existant entre les bactéries et les autres composants biotiques et abiotiques de l'environnement. Différentes études sont basées sur l'estimation du nombre de bactéries par gramme de sol (Torsvik *et al.*, 1990, Gans *et al.*, 2005), malgré tout aucune estimation de cette diversité bactérienne n'est certaine, en grande partie du fait de l'existence de cette rare biosphere.

Nos objectifs dans le cadre de ce travail de thèse ont été de développer des approches tant conceptuelles que techniques qui permettent d'accéder à un niveau plus important de la diversité bactérienne présente dans les environnements complexes comme les sols.

La première approche est basée sur deux piliers expérimentaux, l'un conceptuel et l'autre technique. L'approche conceptuelle consiste à inoculer des sols stérilisés avec une même communauté bactérienne afin de provoquer des changements de sa structure pour permettre le développement de bactéries minoritaires. Cette approche expérimentale est combinée avec

l'utilisation d'une technique sensible de caractérisation et d'identification bactérienne : les puces à ADN de type taxonomique.

La deuxième approche consiste au développement d'un outil permettant de s'affranchir des étapes d'extraction de l'ADN du sol. Cet outil est basé sur l'utilisation d'une souche bactérienne susceptible d'acquérir une *in situ* des gènes de la microflore indigène du sol et sur la construction d'un système de contre sélection positive permettant de sélectionner les clones ayant intégrés des gènes d'intérêt.

La combinaison des deux approches a pour but de maximiser l'accès à ces bactéries « rares » source d'un potentiel génétique inestimable tant pour comprendre les mécanismes évolutifs et adaptatifs bactériens que pour l'exploiter industriellement.

Ce mémoire est constitué de trois parties entièrement dédiées à l'étude et la compréhension de la diversité des communautés bactériennes dans un environnement complexe tel que le sol.

Le chapitre I est une synthèse bibliographique portant sur l'évolution et la sensibilité des différentes techniques d'étude de la diversité bactérienne. Une partie est consacrée à la plasticité génomique des bactéries permettant d'expliquer l'origine de cet immense réservoir génétique. L'environnement complexe du sol est ensuite décrit, notamment ses propriétés physico-chimiques qui affectent la structure des communautés bactériennes.

Le chapitre II présente les résultats obtenus par l'approche conceptuelle d'inoculation de communautés bactériennes dans des sols stérilisés. L'objectif de cette partie était d'une part de montrer l'impact des sols, ayant des propriétés physico-chimiques différentes, sur une même communauté et d'autre part de mettre en évidence des bactéries minoritaires à partir d'une communauté bactérienne inoculée dans les sols stérilisés. Cette partie s'intègre dans le projet ANR : Ecomic-RMQS (Réseau de Mesure de la Qualité des Sols). Ce projet ECOMIC-RMQS a pour objectif de mieux définir et comprendre les processus qui génèrent et maintiennent la biodiversité microbienne des sols, notamment par une meilleure estimation et caractérisation de la diversité « beta » de ces communautés (changement de composition de la communauté à l'échelle du paysage). Pour cela, il s'appuie sur le RMQS pour caractériser les communautés bactériennes indigènes de sols échantillonnés à une grande échelle spatiale. La densité et la diversité/structure génétique des communautés bactériennes sont caractérisées grâce à l'utilisation d'outils moléculaires (PCR Quantitative, empreintes moléculaires, puces à ADN) directement sur l'ADN extrait des sols du RMQS. Ce réseau comprend 2200 sols échantillonnés et couvre tout le territoire français. Pour chacun des sols, les caractéristiques

physico-chimiques, le spectre proche infrarouge, les paramètres climatiques environnants, les compositions floristiques, l'utilisation des terres, les pratiques agricoles et événements antérieurs sont répertoriés. Des outils statistiques ont été développés pour confronter les données microbiologiques, pédoclimatiques et les distances géographiques afin i) de caractériser les profils biogéographiques de la diversité bactérienne des sols de France, ii) d'évaluer la contribution relative des paramètres pédoclimatiques et d'utilisation des terres sur la diversité des communautés bactériennes indigènes, iii) de calibrer l'information spectrale avec les paramètres chimiques et bactériens et iv) d'identifier des bio-indicateurs bactériens d'environnements et d'activités anthropiques spécifiques.

Le chapitre III présente le développement de la souche permettant la capture *in situ* de gènes ciblés. Notre travail s'inscrit dans les premières étapes du développement de cet outil. La stratégie de construction ainsi que les mécanismes cellulaires impliqués sont dans un premier temps décrits, puis les résultats des modifications génétiques sur la souche bactérienne utilisée pour capturer l'ADN indigène sont présentés, ainsi que les optimisations nécessaires pour son utilisation.

Une conclusion générale permet de conclure ce manuscrit et ouvre des problématiques nouvelles.





## **CHAPITRE I : SYNTHESE BIBLIOGRAPHIQUE**

Evaluation et origine de la diversité bactérienne, plus  
particulièrement dans l'environnement sol



## 1. Introduction

Comme nous l'avons vu dans l'introduction, les bactéries sont ubiquitaires. Elles ont un rôle écologique très important, mais sont aussi très utilisées par l'Homme dans différents domaines. Pourtant la majorité de ces bactéries ne sont pas connues, le monde des bactéries se caractérise par une diversité physiologique et génétique extrêmement importante à l'origine de leurs potentialités à coloniser les environnements les plus divers. La nature très dynamique de leur génome basée sur des mutations ponctuelles dont la fréquence d'apparition peut varier selon les conditions du milieu mais aussi sur des acquisitions, délétions ou réarrangements d'informations génétiques est en grande partie responsable de ce potentiel adaptatif. Cette extraordinaire plasticité génomique leur permet de s'adapter beaucoup plus rapidement et efficacement que les autres organismes de notre planète à des modifications même très importantes de leur milieu.

L'objectif de ce chapitre bibliographique est de présenter d'abord un état de l'art sur la caractérisation, l'identification et la classification des bactéries, étapes initiales indispensables pour comprendre comment évoluent et s'adaptent les bactéries mais qui sont rendues extrêmement ardues du fait de la très petite taille de ces organismes, du manque de critères simples permettant de les différencier et du caractère non cultivé de la grande majorité d'entre elles. Ce sont ces travaux parmi d'autres qui ont permis de révéler la nature dynamique du patrimoine génétique bactérien et la seconde partie de ce chapitre sera consacrée à dévoiler les différents processus évolutifs qui sont responsables de la nature mosaïque des génomes bactériens à l'origine de la diversité génétique et physiologique. Cependant, le développement des microorganismes se fait en interactions étroites avec leur milieu et la dernière partie de cette revue bibliographique sera consacrée aux interactions que les bactéries du sol entretiennent avec cette matrice, comment celle-ci conditionne la structure des populations bactériennes telluriques mais aussi perturbe l'analyse de la diversité naturelle et le recensement des gènes et fonctions et même l'exploitation de ces microorganismes.

## 2. Techniques d'étude de la diversité bactérienne

### 2.1. La culture bactérienne

#### 2.1.1. Approches traditionnelles

Les techniques classiques de microbiologie par étalement d'un échantillon issu de l'environnement sur ou dans un milieu de culture ont été utilisées pour favoriser la croissance cellulaire et permettre ainsi la multiplication clonale des différentes bactéries. Quand appliquée avec un milieu de culture solidifié cette approche permet de dénombrer les colonies résultant de cette multiplication clonale et ainsi estimer le nombre initial de bactéries « cultivables » présentes dans l'échantillon. A partir d'un gramme de sol, par exemple, la culture sur milieu solidifié peut révéler la présence de  $10^8$  cellules. Une grande diversité de biotopes ont ainsi été étudiés mais la véritable révolution a eu lieu en 1980, où le dénombrement des micro-organismes a pu être fait *in situ*, grâce à des sondes fluorescentes comme le DAPI (Caron, 1983). Ces techniques, indépendantes de la mise en culture, ont révélé l'existence d'une diversité jusque-là insoupçonnée et ont véritablement bouleversé notre vision du monde bactérien. Par exemple, un gramme de sol peut contenir plus de  $10^9$  cellules (Torsvik *et al.*, 1990). On a ainsi pu établir que 99% des bactéries de l'environnement ne sont pas cultivables dans les conditions du laboratoire. Depuis lors, le sol est considéré comme un réservoir de diversité bactérienne avec une impressionnante étendue de processus biochimiques et moléculaires susceptibles d'être mis en œuvre.

#### 2.1.2. Classification et identification bactérienne appliquée aux isolats bactériens

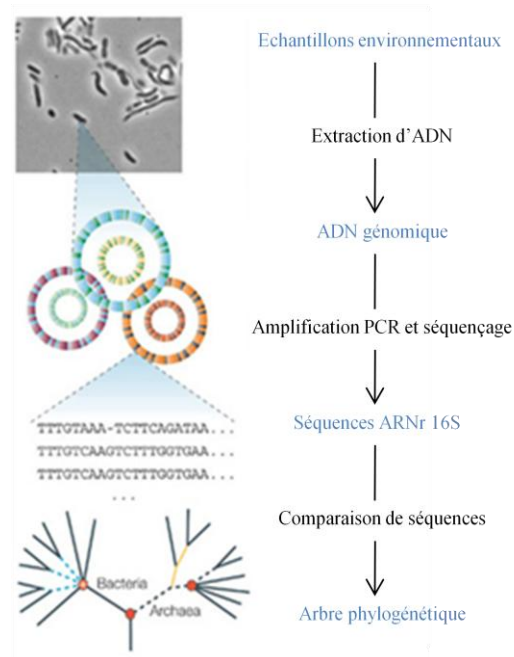
La taxonomie est l'ensemble des principes et théories qui permettent de classer puis de valider le classement des organismes (Fenchel & Finlay, 2006). Le besoin de procéder à des études de taxonomie en microbiologie existe, comme dans toutes les autres disciplines de la biologie depuis le tout début des travaux sur les microorganismes. Il faut décrire et classer les espèces en fonction d'un certain nombre de critères. Avant l'essor de la biologie moléculaire, les espèces pouvaient ainsi être classées et donc identifiées en fonction de leurs caractères : biochimiques (classification en biotypes ou biovars), antigéniques (classification en sérotypes ou sérovars), pathogéniques (classification en pathotypes ou pathovars), enzymatiques (classification en zymotypes ou zymovars), de sensibilité aux antibiotiques (classification en antibiotypes), de sensibilité aux bactériophages (classification en lysotypes ou lysovars).

D'autres critères se basent sur la composition de la paroi, discernable selon la coloration des cellules après application du protocole de Gram (coloration de Gram positive ou négative), la morphologie microscopique (bactéries de type coque, bacille, vibrion ; isolées, par deux, en chaînettes...), la morphologie macroscopique (taille, forme, couleur... des colonies sur milieux de culture gélosés), la mobilité (mobilité ou immobilité à une température donnée), la capacité à sporuler, la température de croissance, les besoins nutritionnels (nécessité de substances particulières pour le développement), le mode respiratoire (aérobie, anaérobie stricte, aéro-anaérobie facultative, microaérophilie...), la capacité de photosynthèse, l'utilisation des différentes sources de carbone ou d'azote (on parlera de biotypes ou biovars, de zymotypes ou zymovars). Cependant ces tests sont rendus imprécis par la variabilité des réponses d'une souche à l'autre ou pour une même souche, en fonction du temps et du milieu sur lequel les bactéries sont mises à cultiver pour effectuer les tests.

La notion d'espèce bactérienne ne peut être déterminée de manière précise seulement par l'utilisation de tels critères. Une étape importante qui a changé la classification bactérienne a été de prendre en compte les propriétés génomiques des bactéries qui sont principalement le rapport entre les bases guanine, cytosine et adénine, thymine, dans la séquence d'ADN (exprimé par le pourcentage moléculaire en G+C) et la similarité de séquences d'ADN obtenue par la technique d'hybridation ADN-ADN (Oren, 2004). Ces outils ont permis de proposer la notion d'espèce génomique bactérienne définie selon le niveau de similarité des génomes calculée d'après la méthode d'hybridation ADN-ADN. Des isolats dont les génomes présentent 70% ou plus de similarité entre leur génome (avec une variation de température de fusion ( $\Delta T_m$ ) inférieure à 5°C) (Wayne, 1987) pourront être classés au sein de la même espèce génomique.

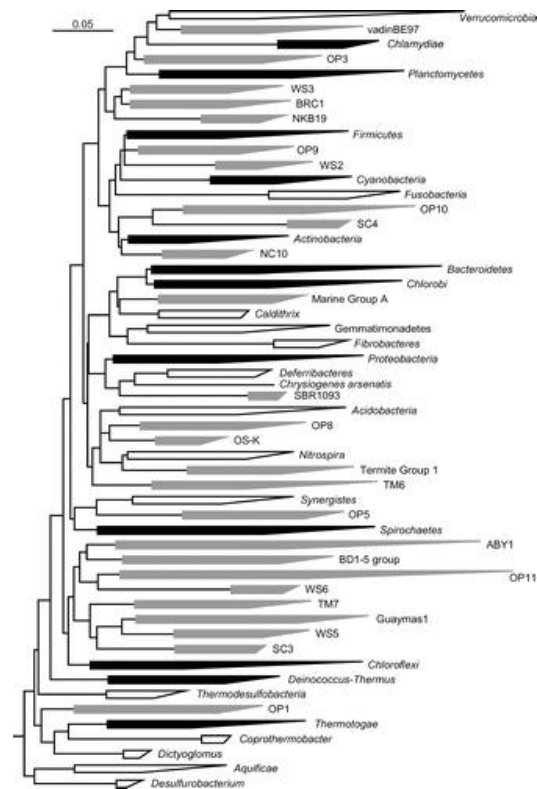
Ces études de similarité ADN-ADN sont cependant plus ou moins précises, longues et laborieuses, basées sur des comparaisons deux à deux, ne pouvant donc concerner qu'un nombre restreint d'isolats. Elles imposent de plus de travailler sur d'importantes quantités d'ADN, compliquant l'étude des souches à croissance lente et interdisant la possibilité d'inclure les bactéries non cultivées. Il a fallu attendre les progrès conjoints de la biologie moléculaire (amplification en chaîne d'ADN, séquençage à haut-débit...) et de la bioinformatique pour que se développe une taxonomie basée sur la phylogénie moléculaire. Il est alors devenu possible d'établir des arbres phylogénétiques de bactéries à partir de la comparaison de séquences d'un gène orthologue présent dans les espèces étudiées. Carl Woese a été le premier, en 1987 (Woese, 1987) à proposer l'utilisation du gène de l'ARN

ribosomique 16S comme marqueur de phylogénie. Il a été trouvé subséquemment que les niveaux de similarité des séquences du gène 16S rDNA entre isolats étaient représentatifs de ceux obtenus par hybridation ADN-ADN de la totalité des génomes, 70% de similarité ADN-ADN correspondant à 98% de similarité du gène 16S rADN (Oren, 2004). Ce gène de 1500 paires de bases répond à plusieurs critères essentiels pour la taxonomie : (Wayne, 1987) il est ubiquiste c'est-à-dire présent chez tous les procaryotes, où il remplit la même fonction (2) il évolue lentement (3) il peut être facilement séquencé car il possède des domaines hautement conservés avec peu de mutations fixées au cours de l'évolution et enfin, (4) il présente des régions de séquences plus variables qui sont d'autant plus différentes entre deux espèces que celles-ci sont phylogénétiquement éloignées. Ainsi, pour positionner dans un groupe phylogénétique une nouvelle espèce, on procède à l'amplification et au séquençage de ce gène puis à sa comparaison avec les séquences déjà connues disponibles dans les banques de données (**Figure 1**).



**Figure 1.** Stratégie utilisée pour l'analyse moléculaire de la diversité bactérienne.

La classification basée sur l'ADNr 16S a eu de profondes conséquences sur notre vision du monde bactérien. D'une part, elle est clairement différente de celle établie sur les critères physiologiques et a induit une refonte de l'arbre phylogénétique des bactéries cultivables. D'autre part, plus important encore, elle a permis de mettre en évidence des groupes phylétiques entiers n'ayant aucun représentant cultivable (**Figure 4**) (Rappe & Giovannoni, 2003).



**Figure 4.** Arbre phylogénétique représentant les principaux phylums des bactéries. Il illustre la présence de groupes entiers sans représentants cultivables (sur la figure: en noir et en blanc: groupes avec représentants cultivables; en gris: sans représentants cultivables) (Rappe & Giovannoni, 2003).

## 2.2. Identification bactérienne par analyse de l'ADN extrait de l'environnement (approche métagénomique)

Depuis le début des années 90, la microbiologie environnementale s'est dotée d'un nouveau concept en restreignant l'étude des bactéries à leur ADN directement extrait de l'environnement (McHardy & Rigoutsos, 2007). L'ADN métagénomique obtenu est exploité soit par amplification par la technique de réaction en chaîne de la polymérase (PCR) ciblée sur des gènes sélectionnés pour leur intérêt en analyse taxonomique (gène 16S rDNA) ou pour des fonctions écologiques particulières, soit par clonage dans un hôte cultivable avant sélection des clones d'intérêt de par la présence du gène recherché sur l'insert. Aujourd'hui, de nouvelles technologies facilitent l'étude de l'ADN métagénomique en proposant de séquencer directement l'ADN extrait s'affranchissant ainsi des étapes de clonage ou d'amplification (Shendure & Ji, 2008). L'émergence de la métagénomique présente l'une des plus grandes opportunités, peut être depuis l'invention du microscope, de révolutionner nos connaissances du monde microbien en permettant d'accéder à l'ensemble des bactéries d'un écosystème et non plus à la très faible minorité de celles cultivées *in vitro* et en permettant



l'analyse quasi simultanée de ces différents organismes (Streit & Schmitz, 2004). C'est donc l'ensemble d'une communauté microbienne qui en théorie devient accessible grâce à l'approche métagénomique, permettant ainsi de définir de possibles interactions entre les différentes populations, celles toujours associées comme celles s'excluant mutuellement, de définir la spécificité ou au contraire la redondance des fonctions potentielles, la diversité des gènes codant ces fonctions. A cause de l'écrasante majorité de bactéries non cultivables dans le sol, les premières recherches basées sur le métagénome résultent toujours en l'identification de taxa bactériens et de gènes jusqu'à présent inconnus, confirmant le potentiel que représente le sol comme réservoir quasi infini de diversité populationnelle et fonctionnelle (Rappe & Giovannoni, 2003).

### **2.2.1. Caractérisation des communautés bactériennes**

La caractérisation d'une communauté bactérienne permet de mettre en évidence des caractéristiques propres à cette communauté qui vont la différencier des autres. Les techniques d'empreintes moléculaires vont permettre la caractérisation de communautés bactériennes plus ou moins complexes, en obtenant des profils qui vont être dépendants des séquences d'ADN présentes dans la communauté bactérienne étudiée. Dans ces méthodes l'ADN métagénomique est extrait de l'échantillon environnemental et purifié. Un fragment d'ADN cible (16S, ITS...décrit ci-dessous) est amplifié par PCR avec des amorces spécifiques ou universelles et le résultat peut être séparé de différentes manières. Le principe de chaque méthode va être brièvement décrit, puis une synthèse sur les avantages et limites de ces techniques ainsi que leurs applications, sera présentée.

#### **DGGE (Denaturing Gradient Gel Electrophoresis / TGGE (Temperature Gradient Gel Electrophoresis)**

Ce sont deux méthodes similaires permettant l'étude des communautés bactériennes. Le gène 16S rDNA est composé de zones plus ou moins conservées. L'amplification des zones conservées cible toutes les bactéries, tandis que l'amplification de zones variables permet de cibler un niveau taxonomique précis. Dans ces techniques l'ADN est extrait des échantillons de sols et amplifié avec des amorces universelles ciblant la partie conservée de la séquence du gène 16S rDNA. L'extrémité 5' de l'amorce sens contient une queue composée d'une succession de GC de 35-40 paires de bases (=GC-clamp). L'ajout de ce GC-clamp va créer un domaine de grande stabilité, il ne s'ouvrira pas ou ne sera pas dénaturé dans le gel. Les

fragments d'ADN sont soumis à un gradient de concentration d'un agent dénaturant (formamide ou urée) ou à un gradient de température. Lorsque la concentration aboutissant à la dénaturation du fragment d'ADN ou la température de fusion ou  $T_m$  ( $T_m$ : melting temperature) du fragment dans le cas de la technique TGGE sont atteintes, le fragment d'ADN se dénature jusqu'au niveau du domaine GC-clamp, qui stoppe alors sa migration dans le gel. Cette dénaturation est séquence spécifique, les fragments vont donc migrer différemment dans le gel de polyacrylamide suivant leur séquence. La TGGE utilise le même principe que la DGGE, un gradient de température est alors utilisé à la place du gradient chimique dénaturant. Théoriquement la DGGE peut séparer des fragments d'ADN qui diffèrent par une seule paire de bases (Miller *et al.*, 1999). Les fragments d'ADN de différentes séquences peuvent avoir les mêmes caractéristiques de mobilité dans le gel polyacrylamide, donc une bande bien différenciée sur le gel ne représente pas forcément la présence dans la communauté d'une seule espèce bactérienne (Gelsomino *et al.*, 1999).

#### SSCP (Single Strand Conformation Polymorphism)

Dans cette technique, la région d'intérêt du gène 16S rDNA est amplifiée à partir d'une solution d'ADN extraite d'échantillons environnementaux à l'aide d'amorces marquées radioactivement. Les produits PCR sont ensuite soumis à dénaturation (Lee *et al.*, 1996).

Les ADN en conformation simple brin présentent une certaine mobilité dans un gel de polyacrylamide non dénaturant. Cette mobilité spécifique est due à la conformation spatiale de l'ADN simple brin qui résulte de la formation de liaisons hydrogènes et électrostatiques essentiellement au sein des structures primaires qui aboutissent à la formation des structures secondaires-tertiaires de l'ADN. Comme dans la technique précédente des profils de bandes d'ADN sont générés et permettent de se rendre compte de la diversité des gènes ribosomiques donc des bactéries présentes dans l'échantillon étudié. Ici encore des biais peuvent survenir liés au fait qu'un ADN simple brin peut avoir plus d'une conformation stable, de ce fait une même séquence d'ADN peut être représentée sur le gel d'agarose par plus d'une bande (Tiedje, 1999).

#### RFLP (restriction fragment length polymorphism) / ARDRA (amplified ribosomal DNA restriction analysis)

La RFLP aussi connue sous le nom d'ARDRA est une autre méthode reposant sur le polymorphisme de fragments d'ADN utilisée pour l'étude des structures des communautés

microbiennes. La région d'intérêt est amplifiée à l'aide d'amorces à partir de l'ADN extrait d'un échantillon, les fragments obtenus sont digérés par une enzyme de restriction appropriée. On obtient alors des fragments d'ADN de longueurs différentes selon les individus. Cette différence de taille est due à une différence du nombre de sites de restriction qui varie selon les individus, l'ADN à analyser étant alors fractionné en fragments de taille plus ou moins importante révélant alors le polymorphisme de longueur des fragments de restriction. Les fragments de restriction sont séparés selon leur taille par une électrophorèse en gel d'agarose (Liu *et al.*, 1997). Une espèce peut ainsi être caractérisée par 4 à 6 fragments de restriction dans la zone ciblée (Tiedje, 1999).

#### T-RFLP (Terminal-Restriction Fragment Length Polymorphism)

Cette technique est basée sur le même principe que la RFLP, mis à part le fait que l'une des amorces est marquée avec un colorant fluorescent. Ceci permet seulement la détection des fragments de restriction terminaux marqués (Liu *et al.*, 1997), simplifiant le profil des bandes. Cette méthode permet l'analyse de communautés complexes en fournissant des informations sur la diversité puisqu'une bande représente une seule OTU (Tiedje, 1999).

#### RISA (Ribosomal Intergenic Spacer Analysis) /ARISA (Automated Ribosomal Intergenic Spacer Analysis)

Cette technique est basée sur l'amplification par PCR de la région IGS (intergenic spacer) située entre les sous unités ribosomales 16SrDNA et 23SrDNA, région d'ADN présentant une variabilité de taille et de séquences entre individus supérieure à celle des gènes, pouvant alors différencier des espèces très proches. Les produits générés, de différentes tailles sont alors séparés sur gel de polyacrylamide sous conditions dénaturantes. (Fisher & Triplett, 1999).

Ces techniques d'empreintes moléculaires se sont révélées efficaces pour l'analyse comparative des structures des communautés microbiennes (Ramette, 2009). Elles présentent plusieurs avantages, étant indépendantes de la culture *in vitro*, hautement reproductibles, robustes et permettent d'étudier rapidement la structure des communautés bactériennes. Mais elles comportent également des biais, ces méthodes basées sur une étape d'amplification (von Wintzingerode *et al.*, 1997), sous-estiment la diversité, car seules les espèces numériquement dominantes sont détectées. Un problème majeur est donc le manque de sensibilité de ces techniques qui ne couvrent que les 100 à 200 bactéries les plus représentées. Ces techniques

ne répondent pas généralement directement aux questions concernant la richesse ou l'abondance relative des OTU (unité taxonomique opérationnelle) (Bent & Forney, 2008) à moins que les bandes ne soient extraites du gel et l'ADN séquencé. Ces méthodes d'empreintes moléculaires sont donc généralement inefficaces pour fournir l'identité taxonomique des bactéries impliquées (DeSantis *et al.*, 2007). La T-RFLP est celle parmi les méthodes d'empreintes moléculaires qui offre la meilleure résolution taxonomique, potentiellement capable de l'identification de classes. Comme déjà mentionné, l'association à la nomenclature taxonomique de chaque groupe peut être effectuée par séquençage, après une étape d'extraction et de purification de chaque bande, et n'est donc pas très adaptée pour des grandes études comparatives telles que le suivi environnemental (Dunbar *et al.*, 2002).

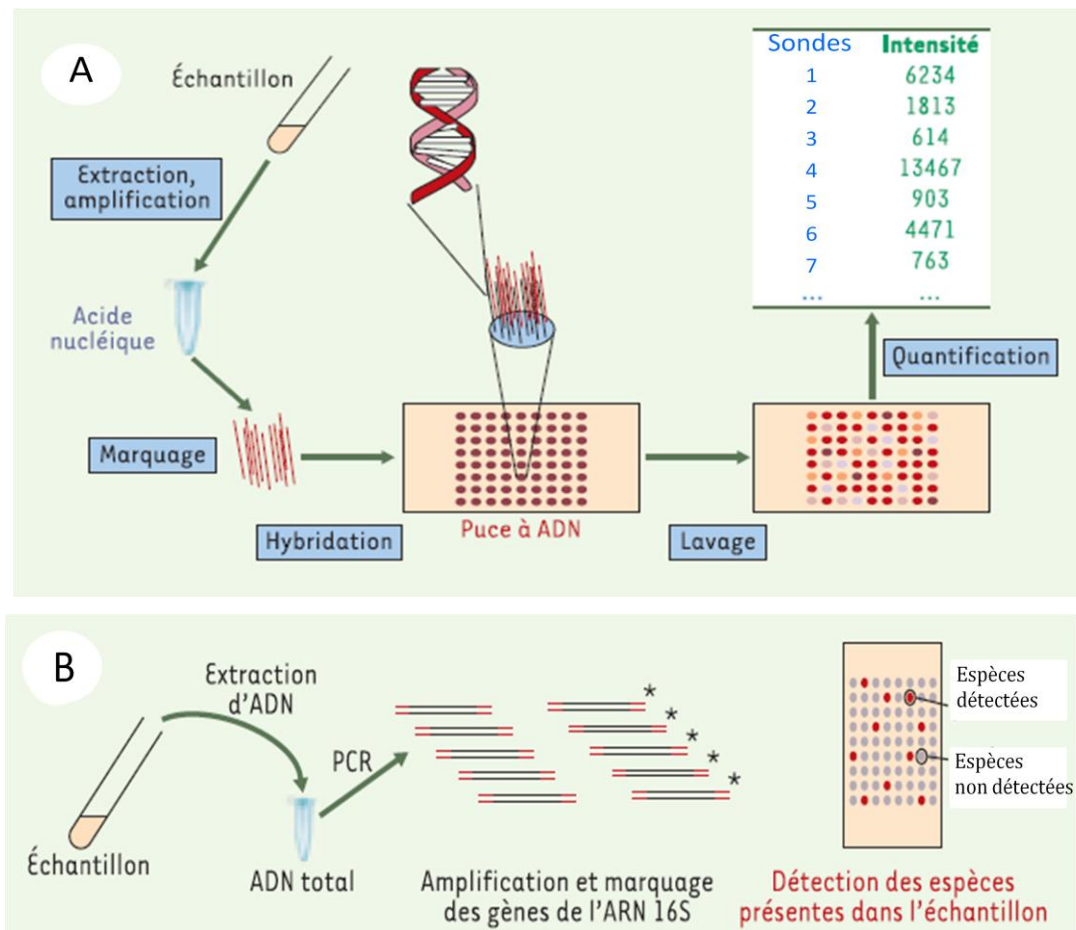
## 2.2.2. Apport des nouvelles technologies à haut débit.

### 2.2.2.1. Principe et applications des puces à ADN

Le principe d'une puce à ADN réside dans la reconnaissance, c'est-à-dire l'hybridation, entre deux molécules d'ADN simple brin complémentaires. L'échantillon (ADN ou ARN) à analyser, marqué de manière fluorescente, est mis en contact sur la puce avec les sondes dont le nombre peut dépasser plusieurs centaines de milliers qui sont des produits PCR ou des oligonucléotides synthétiques de séquence connue fixés ou synthétisés directement *in situ*. Après lavage de la lame pour ne maintenir que les duplex stables signes de la complémentarité de séquences la plus élevée possible entre sondes et cibles, le signal de chaque hybridation est quantifié pour être analysé. Sa valeur dépendra de la concentration en molécules marquées de l'échantillon complémentaire de la sonde et du degré de complémentarité (% d'identité) avec la sonde (Glaser, 2005) (**Figure 5.A**). La technologie des puces à ADN est l'un des développements clés de ces dernières années. Pour des applications biologiques les motifs sur la puce peuvent être de l'ADN, de l'ARN, des protéines, des polysaccharides, des lipides, des petits composés organiques, ou même des cellules entières (Hoheisel, 2006). Par conséquent la technologie des puces ADN a permis principalement l'estimation de l'abondance de cibles et la détection d'interactions biologiques au niveau moléculaire ou cellulaire. Parmi les différents types de motifs possibles, les puces à ADN sont les plus utilisées et les mieux développées. Initialement développées pour l'analyse de l'expression des gènes de la totalité du génome (transcriptome) (Schena *et al.*, 1996, Tao *et al.*, 1999) les puces à ADN ont commencé à être utilisées pour la détection et l'identification de souches microbiennes, d'espèces, de genres ou de niveaux taxonomiques supérieurs

(suivant la construction des sondes) présents dans un échantillon. Pour les raisons évoquées précédemment les cibles les plus couramment utilisées sont les gènes 16S rDNA et 23S rDNA (Small *et al.*, 2001 , Loy *et al.*, 2002) dont le potentiel pour l'identification bactérienne dans le cadre de l'approche « puces à ADN » a été initialement montré pour un nombre limité de sondes (jusqu'à 132 pour des puces à ADN 16S rDNA) (Bodrossy & Sessitsch, 2004). A l'inverse des gènes fonctionnels, le haut niveau de conservation de séquences du gène 16S rDNA et son ubiquité chez les procaryotes (Woese, 1977 , Woese, 1987) permettent d'accéder à la quasi totalité des bactéries (Bruce *et al.*, 1992), intérêt confirmé par la disponibilité de plusieurs dizaines de milliers de séquences dans les bases de données (Benson *et al.*, 2002 , Cole *et al.*, 2005). Par exemple, la base de données du gène 16S rDNA ARB contient des séquences (86 0158) de qualité correctement alignées permettant de définir des sondes oligonucléotidiques ciblant des bactéries à des niveaux taxonomiques différents, du phylum à l'espèce (Ludwig *et al.*, 2004).

Dans le cadre de l'identification des bactéries présentes dans un échantillon, la technologie « puces à ADN » qualifiée de taxonomique ou phylogénétique (phylochip) repose sur successivement, l'extraction de l'ADN total à partir d'un échantillon environnemental, l'amplification d'une partie plus ou moins longue du gène 16S rDNA grâce à des amorces définies selon le niveau de spécificité souhaité et l'hybridation des produits PCR contre des oligo-sondes préalablement définies, synthétisées *ex situ* puis greffées sur la lame ou directement synthétisées sur la puce comme le proposent aujourd'hui certaines compagnies privées (**Figure 5. B.**) (Glaser, 2005). Les puces à ADN sont utilisées pour détecter un produit de PCR et remplacent ainsi la migration sur gel en validant la spécificité de l'amplification par sa complémentarité avec la sonde. Pour l'analyse d'un échantillon simple cette procédure est peu compétitive par rapport à la PCR en temps réel ou l'électrophorèse capillaire. En revanche dans le cas d'un échantillon complexe, les puces à ADN taxonomiques permettent l'analyse simultanée de produits d'amplification complexes, permettant ainsi non seulement la caractérisation d'une communauté bactérienne mais aussi l'identification des bactéries présentes.



**Figure 5. A. Analyse d'acides nucléiques par puces à ADN.** L'ADN ou l'ARN est purifié et éventuellement amplifié à partir d'un échantillon biologique. Il est ensuite marqué de manière fluorescente et mis en contact avec les sondes portées par la puce. Lors de cette étape d'hybridation, les acides nucléiques marqués vont s'apparier avec les sondes ADN fixées sur le support. Une étape de lavage permet ensuite d'éliminer les acides nucléiques marqués fixés de manière non spécifique. Finalement, la fluorescence au niveau de chaque dépôt de sonde sera quantifiée au moyen de tubes photomultiplicateurs ou d'une caméra CCD (*charge-coupled device*). Les valeurs obtenues pour chaque sonde, comme indiquées sur le tableau, doivent ensuite être traitées aux moyens d'outils informatiques pour obtenir la caractérisation de l'échantillon. **B. Analyse d'une population bactérienne mixte.** Après extraction de l'ADN, les régions codant pour l'ARN ribosomique 16S sont amplifiées pour toutes les bactéries de l'échantillon en utilisant des amorces universelles correspondant à des régions conservées dans toutes les espèces. La puce porte des oligonucléotides de séquences spécifiques de chaque espèce. Après hybridation, la quantification du signal permet de détecter la présence de bactéries des différentes espèces représentées sur la puce ADN et d'évaluer leurs quantités relatives (Glaser, 2005).

Cependant comme toutes méthodes, les puces à ADN ont leurs limites. La plus importante est liée au risque d'hybridation croisée entre motifs différents présentant une forte similarité de séquences sans pour autant présenter une parfaite complémentarité (présence de bases non appariées) avec donc le maintien d'un duplex double brin (Shendure, 2008). Une autre limite

tient au fait qu'en dépit de la richesse en termes de séquences 16S rDNA dans les bases de données de nombreuses bactéries demeurent non identifiées. La séquence de leur gène 16S rDNA est non prise en compte pour la définition des sondes (Shiu & Borevitz, 2008). Ces lacunes constituent une autre limitation majeure de cette technologie qui ne pourra identifier dans un mélange d'ADN que les séquences préalablement répertoriées et ne permettra pas d'identifier de nouvelles bactéries. En dépit de cette limitation la technologie reste très prometteuse en comparaison des autres techniques, y compris le séquençage de clones d'une banque d'ADN métagénomique qui nécessite un très important investissement en termes de nombre de clones à produire pour obtenir un niveau d'accès à la diversité bactérienne comparable à celui obtenu par l'utilisation de puces à ADN taxonomiques disponibles aujourd'hui (DeSantis *et al.*, 2007). Il faut aussi considérer que l'acquisition de nouvelles séquences 16S rDNA se fait à un rythme important, soutenue par les progrès des technologies de séquençage. Le problème du faible nombre relatif de sondes disponibles en comparaison de la diversité réelle va tendre à se résorber progressivement dans les prochaines années. La technologie des puces à ADN reste en constante amélioration, avec donc une densité de sondes de plus en plus élevée, des coûts de fabrication plus faibles et le développement de programmes d'analyse performants.

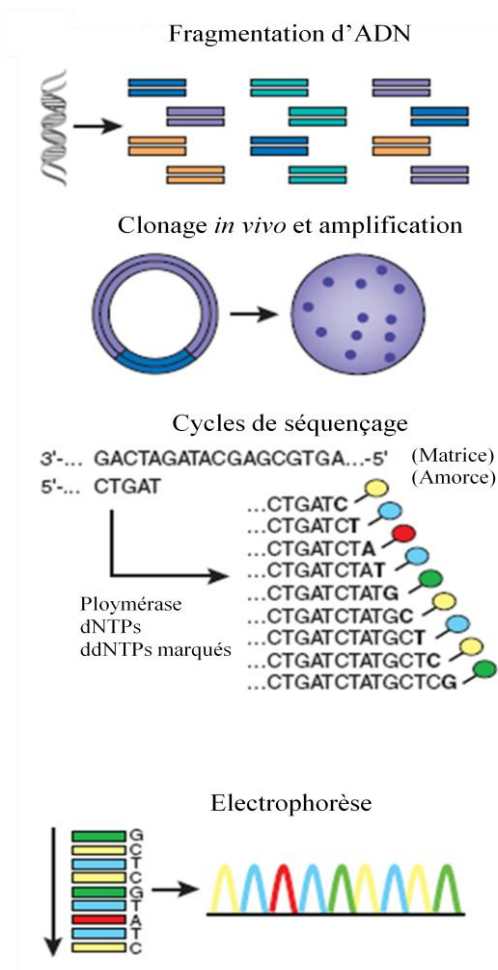
#### 2.2.2.2. *Le séquençage*

Traditionnellement, l'étude de gènes à partir d'un environnement naturel incluait le clonage d'un ADN dans un vecteur, l'insertion du couple vecteur-insert dans une bactérie hôte, le criblage et le séquençage de l'insert par la technique de Sanger. Les méthodes de séquençage par synthèse fournissent des méthodes rapides, et plus simples pour séquencer le métagénome en contournant les biais de clonage liés à la méthode Sanger. Les nouvelles technologies de séquençage diffèrent dans le type de sortie, la longueur de lecture, la vitesse et la distribution du taux d'erreur. L'objectif de cette partie est de décrire les différentes techniques de séquençage et de présenter une synthèse sur les applications et les impacts de ces techniques pour l'étude de la diversité bactérienne.

#### ***Principe de la méthode de Sanger et Nicklen 1977***

Cette méthode est basée sur l'interruption de la synthèse enzymatique d'un brin d'ADN complémentaire (arrêt d'élongation) (Sanger *et al.*, 1977). L'ADN génomique est fragmenté, puis cloné dans un vecteur plasmidique et transformé dans *E. coli*. Pour chaque réaction de

séquence une colonie bactérienne est utilisée et son ADN plasmidique isolé. Chaque cycle de séquençage a lieu à l'intérieur de micro-volume, générant plusieurs tailles de séquence d'ADN terminées par un ddNTP, qui sont séparées par électrophorèse à haute résolution à l'aide d'un séquenceur de 96 ou 384 capillaires. Les fragments de différentes tailles marqués par fluorescence passent devant un détecteur, les 4 canaux des spectres d'émissions sont utilisés pour obtenir les traces du séquençage (Shendure, 2008) (**Figure 6**). L'électrophorèse simultanée dans 96 ou 384 capillaires indépendants permet l'analyse simultanée de peu d'échantillons. Après beaucoup d'améliorations la technique de Sanger peut être appliquée pour déchiffrer des séquences jusqu'à environ 1 000 pb avec une précision par base de 99.999%.



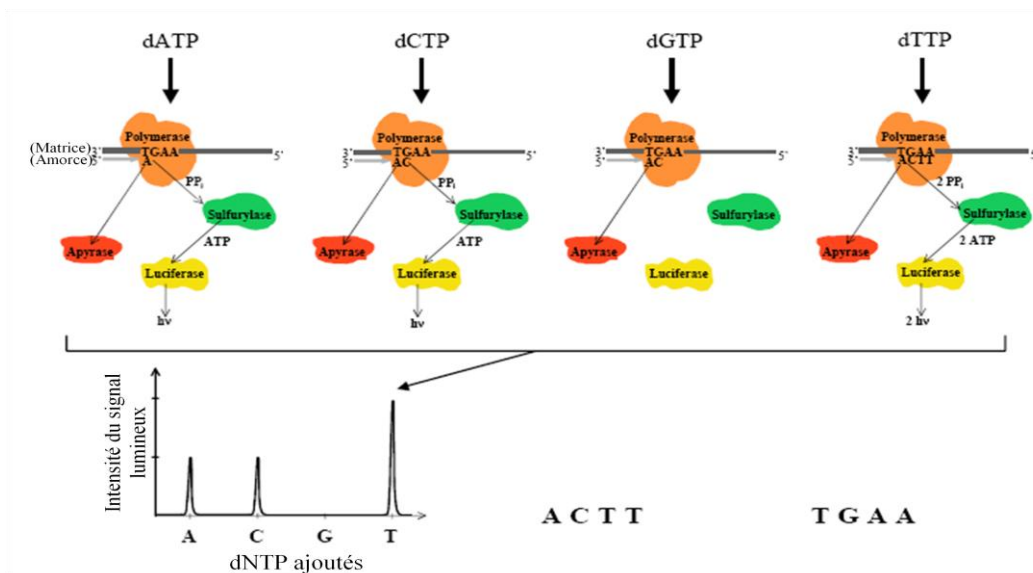
**Figure 6.** Etapes du séquençage Sanger (Shendure, 2008)

### *Le 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 (*sequencing by synthesis*, 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. On peut 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 7**).

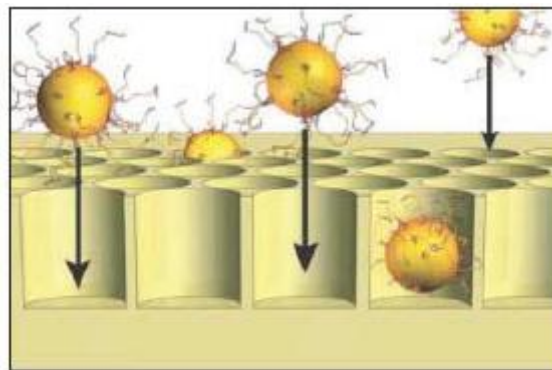


**Figure 7.** Etapes du pyroséquençage (Ahmadian *et al.*, 2006).

### ***Les nouvelles technologies de séquençage haut débit***

- **La 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 J. 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) (**Figure 8**) 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.



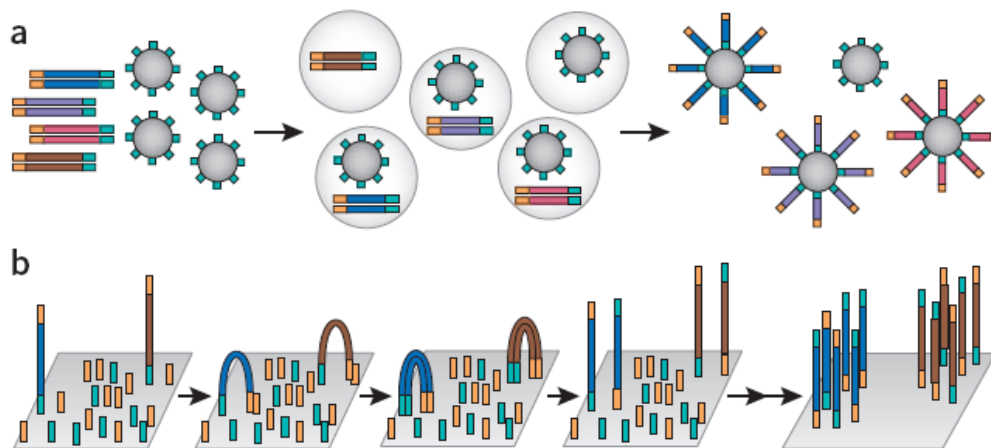
**Figure 8.** Représentation du dépôt d'une microbille dans un puits d'une plaque en fibre optique contenant 1,4 millions de puits (Shendure & Ji, 2008).

Etape 1: 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 (**Figure 8**).

Étape 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.

Étape 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 9.A**).



**Figure 9.** Amplification clonale des motifs de séquence. (a) Les plateformes 454, Polanotor, et SOLiD sont basées sur une PCR en émulsion pour amplifier les motifs à séquencer. (b) La technologie Solexa est basée sur des ponts PCR pour amplifier les motifs à séquencer (Shendure & Ji, 2008).

- **La technologie CRT Solexa/ illumina**

La société Solexa, récemment achetée par la firme Illumina (San Diego, USA), a développé une technologie de séquençage sur lames de type puces à ADN, fondée sur l'intégration de plusieurs techniques : les biopuces à ADN, la nanotechnologie, une variante de la technique de Sanger appelé CRT (*cyclic reversible termination*), ainsi que des technologies informatiques de pointe pour l'acquisition, le traitement et l'analyse des images (Fedurco *et al.*, 2006, Turcatti *et al.*, 2008). Les coûts de séquençage sont réduits de façon drastique en supprimant l'étape lente d'amplification de l'ADN, en travaillant dans un volume de réaction

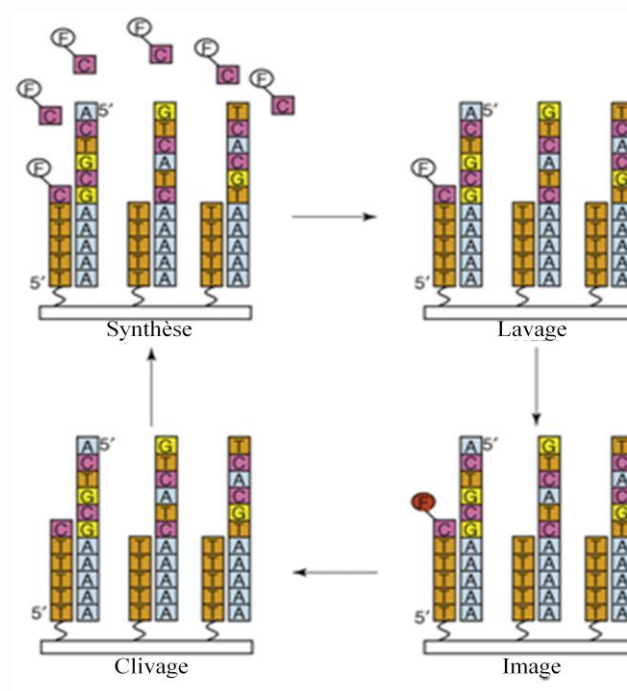
très réduit et avec une préparation minimale de l'échantillon et en éliminant l'étape d'électrophorèse. Le principe de séquençage est basé sur l'incorporation réversible de nucléotides fluorescents (CRT : *cyclic reversible termination*) et par lecture optique de la fluorescence. Comme pour la technique de Sanger il s'agit d'une terminaison de synthèse basée sur l'utilisation d'un terminateur réversible contenant un groupement de protection attaché au nucléotide qui termine la synthèse d'ADN (**Figure 9. b**). Il s'agit là également, d'un séquençage en temps réel, basé sur la détection de la fluorescence mais en présence des 4 nucléotides marqués (ce qui constitue un avantage par rapport à la technologie 454). La très haute densité de la puce à ADN (plus de 100 millions de molécules par centimètre carré) permet de séquencer environ 100 000 paires de bases par seconde. Des séquences de 25 paires de bases sont produites puis alignées, grâce à un logiciel, et comparées à une séquence de référence pour reconstituer la séquence.

- **Le système SOLiD**

La plateforme SOLiD du système décrit par Shendure *et al.* (Shendure *et al.*, 2005), emploie des processus de préparation et amplification d'échantillons similaires à ceux développés pour le 454. Le processus de séquençage SOLiD diffère considérablement des méthodes 454 et Solexa, dans le fait qu'il repose sur un séquençage par « ligation », où des oligonucléotides de 8 bases marqués par fluorescence sont liés de manière séquentielle sur l'amorce de séquençage. Chacun des quatre types de sonde porte un fluorophore distinct en 3', représentant la séquence de l'échantillon complémentaire aux quatrièmes et cinquièmes bases de la sonde oligonucléotidique. Le signal fluorescent généré par l'hybridation de la sonde est alors détecté et enregistré par un scanner laser et un processus d'acquisition de données similaires au système Solexa. La sonde liée est alors clivée entre la 5<sup>ème</sup> et 6<sup>ème</sup> base, enlevant la moitié fluorescente, et des séries répétées de la ligation de la sonde.

- **Helicos**

Sur la base des études de séquençage des molécules individuelles d'ADN (Braslavsky *et al.*, 2003), le séquenceur Helicos (Harris *et al.*, 2008) récemment commercialisé diffère des autres processus de séquençage décrits ci-dessus par le fait que l'échantillon d'ADN ne nécessite aucune amplification préliminaire. L'ADN est fragmenté de manière aléatoire, et flanqué d'une queue poly-A, puis capturé par hybridation à la surface de la lame où sont fixés des oligomères poly T. Les étapes du séquençage sont décrites **Figure 7** (Gupta, 2008).



**Figure 10.** Illustration schématique montrant les différentes étapes du séquençage Helicos (Bioscience) ; (Wayne) Incubation des séquences avec un nucléotide marqué avec un colorant et son incorporation ; (2) étape de lavage pour enlever les nucléotides qui n'ont pas été incorporés ; (3) capture d'image du nucléotide incorporé grâce au marquage Cy5 (excitation 645nm) ; (4) clivage au niveau du nucléotide marqué pour libérer le marqueur (Gupta, 2008).

### ***Applications et impact des différentes techniques de séquençage dans l'étude de la diversité microbienne des sols***

Les nouvelles générations de séquençage d'ADN permettent d'obtenir des données de séquences très importantes (ordre de grandeur : la gigabase) pour une variété d'applications. Nous nous intéressons dans cette synthèse bibliographique plus particulièrement aux applications du séquençage liées à l'étude des communautés microbiennes par l'analyse des gènes 16S rRNA. La plupart des études métagénomiques actuellement décrites sont basées soit sur la technique 454 (Roesch *et al.*, 2007) soit sur le séquençage conventionnel de type Sanger (Morales *et al.*, 2009), la raison principale étant la longueur de lecture, en effet plus grande sera la longueur du fragment d'ADN ribosomique séquençé plus informative sera son analyse et donc plus précis et juste sera le positionnement phylogénétique et taxonomique de l'organisme à l'origine du gène étudié (Wang *et al.*, 2007). Le sol est considéré comme le plus important réservoir de biodiversité bactérienne en termes de nombre de populations bactériennes différentes, suscitant un intérêt grandissant pour en estimer la richesse génomique. L'utilisation de la cinétique de réassociation de l'ADN appliquée à l'ADN extrait du sol a permis d'estimer le nombre d'espèces bactériennes dans un gramme de sol à environ

10 000 (Torsvik *et al.*, 1990). Récemment cette estimation a été réévaluée, ce qui a permis de prédire que près de  $10^7$  espèces microbiennes pouvaient coloniser un gramme de sol (Gans *et al.*, 2005), une valeur trop élevée pour être vérifiée par les technologies de séquençage actuelles (Gans *et al.*, 2005). D'autres études ont tenté d'estimer le nombre d'espèces bactériennes dans un gramme de sol par extrapolation basée sur l'amplification, le clonage puis le séquençage des gènes 16S rDNA. Schloss et Handelsman (Schloss & Handelsman, 2006) ont estimé ce nombre d'espèces compris entre 2000 et 5000 à partir d'une banque de clones 16S rDNA riche de 1033 et 600 clones, respectivement. En fait, aucune des estimations indiquées dans la littérature n'est basée sur un nombre suffisant de séquences pour fournir une extrapolation raisonnable, en effet le nombre d'espèces détectées dans un échantillon est fortement influencé par le nombre de séquences analysées (Schloss & Handelsman, 2005). Roesch *et al.* (2007) ont obtenu la plus importante banque de séquences 16Sr RNA (53 533), en utilisant la technique de pyroséquençage 454 (Roesch *et al.*, 2007). Le nombre d'OTU unique alors estimé par extrapolation est d'environ 6000 par gramme de sol. Incontestablement, les avancées des technologies de séquençage durant la dernière décennie ont des retombées directes sur l'étude de la diversité microbienne, la résolution étant de plus en plus fine. Ces technologies offrent ainsi la chance de répertorier, identifier, caractériser des bactéries appartenant à des phylums totalement inconnus même si ces études peuvent encore être limitées par des biais d'extraction de leur ADN et d'amplification, de clonage ou de séquençage.

### 3. Plasticité génomique bactérienne

Le très haut niveau de plasticité génomique des bactéries est un des arguments proposés pour expliquer l'importante diversité bactérienne. Les bactéries colonisent la terre depuis 3.8 milliards d'années, soit 2 milliards d'années avant l'apparition des premières cellules eucaryotes et c'est leur capacité à générer une très importante diversité génomique qui est proposée comme mécanisme à la base de leurs succès évolutifs (Arber, 2000). Les bactéries ont en effet acquis la capacité de survivre et même de se développer dans de nombreux milieux dont certains sont considérés comme extrêmes. Par exemple *Mycobacterium parascrofulaceum* colonise des geysers où la température atteint les 100°C (Santos *et al.*, 2007). Pourtant, les procaryotes sont dépourvus d'une véritable sexualité. La plasticité génomique bactérienne est assurée par deux types de mécanismes, qui permettent d'expliquer cette incroyable diversité du monde bactérien :

- Des modifications endogènes du patrimoine génétique au sein d'une population par accumulation successive de mutations (Drake, 1991) et de réarrangements de portions plus ou moins importantes du génome (Milkman, 1997).
- Des acquisitions de gènes ou de fragments d'ADN provenant d'autres espèces grâce à des mécanismes spécifiques relevant du transfert horizontal (Lawrence, 1999).

### 3.1. Processus d'évolution verticale

#### 3.1.1. Mutations spontanées

L'ADN polymérase réalise des erreurs lors de la réplication de l'ADN en insérant des bases incorrectes (insertion, répétition, délétion...). Il s'agit alors de mutations ponctuelles dont la fréquence d'apparition chez *E. coli* est d'environ une erreur tous les  $10^4$  nucléotides (Radman *et al.*, 2000). La structure primaire de l'ADN est un facteur important de l'existence et de la localisation des mutations (Drake, 1991). En effet, la présence de séquences répétées et/ou de séquences palindromiques, peut engendrer des hybridations incorrectes durant la réplication de l'ADN ce qui résulte en la production d'erreurs dans la séquence d'ADN fille par rapport à la séquence d'origine. Ces erreurs détectées par la présence de mésappariements sont immédiatement corrigées par l'activité 3'-5' exonucléase portée par la polymérase (Radman *et al.*, 2000). Il existe également des systèmes spécifiques de réparation des erreurs de réplication et des lésions de l'ADN, qui sont aussi impliqués dans la correction de ces erreurs. Chez *E. coli* le système MMS (methyl-directed mismatch repair), constitutivement exprimé, reconnaît les déformations physiques de l'ADN liées aux mésappariements. Les méthylases du système MMS agissent sur un ou deux nucléotides au niveau d'une séquence spécifique, par exemple GATC pour *E. coli*. La protéine MutS se fixe au niveau d'un mésappariement, recrute MutL, et ensemble activent MutH, qui est une endonucléase qui clive le brin non méthylé jusqu'au site spécifique hémiméthylé (Miller, 1996, Oliver *et al.*, 2002). Le système s'assure ainsi que le brin ciblé contient l'erreur puisqu'il n'est pas méthylé, et donc qu'il correspond au brin nouvellement synthétisé ou incorporé. Puis les protéines SSB, l'hélicase II et la polymérase III forment une fourche de réplication pour synthétiser un nouveau brin (Oliver *et al.*, 2002).

Des facteurs externes peuvent aussi agir directement sur le génome bactérien en induisant des mutations, on parle alors d'agents mutagènes (irradiation, accumulation de produits toxiques...). Lorsque l'ADN est très endommagé, le système SOS est alors induit. Ce système contrôle l'expression de plus de 40 gènes en réponse à des dommages de l'ADN (Courcelle *et*

*al.*, 2001), en conditions normales de croissances ces gènes sont exprimés à un niveau basal. La protéine LexA est un répresseur qui se lie à des séquences consensus opératrices : les SOS box, de plusieurs gènes impliqués dans la réponse SOS. Il demeure cependant, parmi ces gènes, une expression basale de la protéine RecA. L'apparition d'ADN simple brin constitue un signal d'altération pour la cellule. RecA s'y lie en formant un nucléofilament, et devient actif en tant que protéase : il clive LexA. Les gènes régulés par ce répresseur sont alors exprimés (Bridges, 2005). Le taux et la longueur de l'induction dépendent donc de la quantité et de la persistance de l'ADN simple brin, puisque lorsque la lésion est réparée, le système SOS est de nouveau réprimé. Le clivage de *lexA* induit l'expression du gène *umuCD*. Ce gène code deux sous-unités : la sous unité UmuC, et la sous unité UmuD qui subit l'action protéolytique de RecA pour être active. L'assemblage UmuC, 2 UmuD' forme la polymérase V, qui permet de passer les lésions et de répliquer l'ADN endommagé (Tang *et al.*, 1999). Ce système est largement distribué parmi les toutes bactéries (protéobactéries et bactéries à gram positif).

Tous ces mécanismes permettent de diminuer le taux de mutations, sans cependant totalement le réprimer. En fonction de la qualité et de la précision de l'activité exonucléasique de l'ADN polymérase, la fréquence des mutations spontanées sera différente. Il a été démontré qu'une mutation qui inactive la portion de gène codant pour cette activité exonucléasique engendre un mutant bactérien ayant ensuite un taux de mutations beaucoup plus élevé (Drake, 1991). Des défauts dans les gènes du système MMS, causant la perte des capacités de relecture de l'ADN durant la réplication, peuvent résulter en l'augmentation de plus de 100 fois du taux d'apparition de mutations génomiques (Nohmi, 2006).

La plupart des mutations spontanées sont soit silencieuses, soit délétères, mais très peu sont avantageuses, c'est-à-dire, favorables d'un point de vue adaptatif. En condition de croissance en présence d'un stress quelconque (chimique, physique etc), le taux de mutation dans les cellules bactériennes peut varier dans le temps de plusieurs ordres de grandeur, conduisant à la génération d'une importante population de mutants parmi lesquels certains présenteront une résistance ou une meilleure adaptation au stress impliqué. La stratégie adaptative impliquée est donc basée sur une véritable mutagénèse de la population, induite par un stress non léthal pour sélectionner des « mutations adaptatives » aboutissant à l'émergence de multiples mutants, pour sélectionner ceux présentant le phénotype de résistance (Rosenberg, 2001). Le système SOS intervient dans cette réponse transitoire au stress et fait intervenir des ADN polymérase peu fidèles, ayant une faible activité exonucléasique: polIV et polV, codées par



les gènes *dinB* et *umuDC* chez *E. coli*. Ces enzymes engendrent de nombreuses mutations par leur taux d'erreurs de réplication très élevé.

### ***Phénotype hypermutateur***

Les populations bactériennes naturelles comportent une proportion variable d'isolats présentant un phénotype hypermutateur, dont le taux de mutation peut être augmenté d'un facteur 100 par rapport aux isolats de référence, généralement suite à des mutations altérant le fonctionnement du système de réparation des mésappariements de base (MMR) (Martinez & Baquero, 2000). Ce phénotype est souvent dû à une inactivation (délétion) du gène *mutS*, mais une population ne peut rester sous sa forme hypermutatrice et doit donc récupérer un gène *mutS* fonctionnel ou transférer l'allèle muté à une cellule ayant un gène *mutS* fonctionnel. Le taux de mutation élevé de ces isolats, associé à une pression de sélection, apporté par exemple par la présence d'un antibiotique, peut aboutir à l'émergence de mutants résistants et à leur rapide prévalence au sein de la population bactérienne. Ainsi, les patients atteints de mucoviscidose présentent une très forte proportion d'isolats cliniques de *Pseudomonas aeruginosa* ayant un phénotype hypermutateur (Macia *et al.*, 2005). Des taux élevés de mutation jouent un rôle important dans l'évolution adaptative.

### **3.1.2. Réarrangements d'ADN**

La diversité génétique peut aussi résulter de remaniements endogènes des séquences d'ADN, c'est-à-dire au sein d'un génome. Le réarrangement d'ADN conduit à l'assemblage d'une nouvelle séquence d'ADN par déplacement de régions plus ou moins importantes du génome. Ces réarrangements peuvent conduire à des événements d'inversion, de duplication et de délétion dans la séquence d'ADN. Dans le cas de la délétion d'ADN, les séquences non essentielles peuvent être éliminées au cours de la sélection naturelle (Arber, 2000). La recombinaison permet un réarrangement de la molécule d'ADN par échange de brins homologues ou par recombinaison non homologue.

#### ***La recombinaison homologue***

C'est un mécanisme que la cellule met en place lorsque les systèmes de réparation de l'ADN par excision ne suffisent plus à rétablir la réplication. Une cassure au niveau de la séquence d'ADN double brin initie une recombinaison homologue chez les bactéries. Cette recombinaison homologue fait intervenir le complexe enzymatique RecBCD et la protéine

RecA (Dabert & Smith, 1997). La protéine RecA est notamment l'enzyme clé catalysant la réaction de recombinaison entre les séquences d'ADN homologues (Chaudhury & Smith, 1984). Les terminaisons au niveau de la cassure d'un ADN double brin servent de point d'entrée à l'enzyme multifonctionnelle RecBCD, qui se déplace jusqu'à la rencontre de séquences *chi* (5'-GCTGGTGG-3'). Ces séquences *chi* constituent un point chaud de recombinaison, favorisant les événements de la recombinaison homologue (Tracy *et al.*, 1997). La sous unité recD perd alors son affinité pour l'hétérodimère RecBC, qui devient déficient en activité hélicasique. Le complexe RecBCD dissocie alors les deux brins d'ADN et la protéine RecA s'associe alors à l'ADN simple brin nouvellement formé. La recombinaison est ensuite amorcée si une homologie de séquences suffisante est détectée (Matic *et al.*, 1996), une séquence minimale d'identité étant en effet requise (Shen & Huang, 1986). Des régions de similarité de 25 à 200pb suivant les espèces sont nécessaires pour initier la complémentation entre l'ADN entrant et l'ADN de la souche réceptrice (Thomas & Nielsen, 2005). La fréquence de recombinaison augmente avec la taille de la zone homologue.

***La recombinaison non homologue :*** est un terme générique qui regroupe les réarrangements génétiques avec comme particularité commune de ne pas nécessiter de grandes régions de parfaite identité entre deux séquences d'ADN. Ce mécanisme permet l'intégration de fragments d'ADN qui contiennent notamment des gènes codant des enzymes permettant leur propre recombinaison (Battista, 2004). Des recombinaisons non homologues peuvent se réaliser dans la cellule par le mouvement d'éléments mobiles tels que des séquences d'insertion (IS), des transposons ou intégrons. Dans le cas des transposons il s'agit de fragments d'ADN entourant un ou plusieurs gènes et codant une ou plusieurs transposases. Ces transposases assurent l'excision ou la réplication de la séquence d'ADN. L'intégration de la séquence est effectuée sur un site spécifiquement reconnu par ces mêmes enzymes. Les mouvements de ces éléments mobiles peuvent engendrer des délétions ou des insertions dans le génome (Battista, 2004). Ils peuvent aussi perturber l'expression des gènes en s'intégrant au niveau de la séquence codante. Ils peuvent influencer l'expression des gènes en apportant avec eux des modules de transcription tels que des promoteurs.

### 3.2. Processus d'évolution horizontale

Les transferts horizontaux de gènes entre bactéries et même entre bactéries et organismes eucaryotes sont dus à 3 processus majeurs : la transformation, la transduction et la conjugaison. Les éléments génétiques mobiles (EGM) tels que les plasmides, les bactériophages, les éléments conjugatifs intégratifs, les transposons, les éléments IS, les intégrons, les cassettes de gènes, et les îlots génomiques sont les principaux vecteurs intervenant dans les deux derniers processus.

#### 3.2.1. Transformation naturelle

La transformation naturelle est généralement définie comme l'acquisition d'ADN libre par des bactéries compétentes (Chen & Dubnau, 2004). Les pré-requis pour la transformation naturelle sont la disponibilité de l'ADN libre, le développement du stade de compétence, l'absorption et l'intégration de manière stable de l'ADN capturé dans le génome de la bactérie réceptrice. Différentes études ont montré qu'en dépit de la présence ubiquitaire des DNases, de l'ADN libre peut être détecté et sa persistance dans différents environnements a été démontrée (Paget, 1994, Gebhard, 1999). Pour que la transformation ait lieu, la bactérie réceptrice doit être dans un état physiologique particulier, appelé stade de compétence, permettant l'absorption efficace de l'ADN. Un tel état peut être obtenu artificiellement chez la plupart des bactéries cultivées, mais les espèces bactériennes connues pour être naturellement compétentes sont assez rares. De plus la transformabilité semble être une propriété qui n'est pas systématiquement partagée par tous les isolats appartenant à la même espèce, avec des fréquences de transformation pouvant différer jusqu'à 4 ordres de grandeur entre isolats transformables d'une même espèce (Sikorski *et al.*, 2002, Maamar & Dubnau, 2005). L'induction de la compétence est le plus souvent en relation avec le stade physiologique des cellules et/ou la diffusion d'un facteur de compétence dans le milieu (Chen & Dubnau, 2004). La transformation *in situ* dépend alors des conditions qui favorisent l'accumulation d'un facteur de compétence ou le développement du stade physiologique approprié, par exemple la bactérie du sol *Bacillus subtilis* répond à des stress environnementaux par le développement du stade de compétence. Chez *B. subtilis* le développement de la compétence est une partie de l'état physiologique, distinct de la croissance végétative ou de la sporulation, qui est développée par seulement 10-20% des cellules durant le dernier stade de croissance sous des conditions nutritionnelles spécifiques (Berka *et al.*, 2002). Récemment, l'absorption d'ADN par les bactéries *Pseudomonas sp.* soumises à des paramètres électriques simulant la décharge

de foudre a été décrite comme un mécanisme alternatif de transformation qui pourrait être particulièrement important pour les bactéries comme *Pseudomonas* qui ne possèdent pas le mécanisme moléculaire pour développer la compétence naturelle (Ceremonie *et al.*, 2004).

L'ADN absorbé peut servir comme source de nutriments, pour la réparation de l'ADN ou comme source d'innovation génétique (Dubnau, 1999). L'absorption de l'ADN peut être suivie par son intégration dans le génome bactérien grâce à l'action de mécanismes de recombinaison homologue, ou par recombinaison illégitime, celle-ci pouvant être facilitée par de très courtes zones de similarités de séquences (de Vries & Wackernagel, 2002). En plus des différences au niveau de l'absorption d'ADN, les bactéries ne montrent pas la même efficacité d'intégration de l'ADN par recombinaison homologue ou illégitime (Sikorski *et al.*, 2002). Une autre option pour la pérennisation de l'ADN pénétrant réside dans la formation d'un élément répliquatif autonome (plasmide).

La transformation, naturelle ou résultant de stress physico-chimiques est donc un mécanisme de transfert de gènes susceptible de concerner tous types de bactéries et tout type d'ADN présent dans l'environnement proche de la cellule. Si ce mécanisme permet aux bactéries de générer une forte variabilité génétique (Dubnau, 1999, Nielsen *et al.*, 2000) il n'en demeure pas moins fermement régulé, notamment pour la pérennisation de l'ADN qui a franchi les enveloppes cellulaires (Fall *et al.*, 2007).

### **3.2.2. La transduction**

La transduction est le processus de transfert d'ADN d'une bactérie à une autre via des virus bactériens, les bactériophages.

Il existe deux types de transduction : généralisée et spécialisée. Lorsqu'une cellule bactérienne est lysée par un phage, le chromosome bactérien est dégradé en petits fragments. Il arrive que les particules virales en formation incorporent par erreur un fragment d'ADN bactérien dans la tête phagique, à la place de l'ADN phagique. De tels virus peuvent se fixer à une cellule bactérienne et y injecter l'ADN qu'ils contiennent, en l'occurrence des gènes de la bactérie donatrice. Les gènes bactériens transduits peuvent alors être intégrés dans le chromosome par recombinaison. Puisque n'importe quel marqueur de la bactérie hôte peut être transduit on parle de transduction généralisée.

La transduction localisée est réalisée par des phages tempérés à sites d'intégration stricts. Ces phages tempérés peuvent demeurer dans un état quiescent en intégrant leur matériel génétique à l'ADN de la bactérie. On parle alors de provirus ou de prophage, c'est-à-dire un virus dont le

matériel génétique est intégré au génome de l'hôte. Ces phages endogènes, sont copiés à chaque division cellulaire avec l'ensemble de l'ADN de la bactérie, que l'on qualifie alors de lysogène. Dans certaines conditions, en particulier en cas de carence ou de stress, le prophage sort de son état quiescent et active son cycle répliatif. Il s'excise du génome de l'hôte et entre dans un cycle lytique. Normalement, l'ADN phagique est excisé dans son intégralité. Toutefois, à faible fréquence, l'excision est anormale aboutissant à l'excision d'une molécule d'ADN "hybride" constituée d'un fragment d'ADN phagique et d'un fragment d'ADN bactérien. Ce fragment d'ADN bactérien est alors adjacent à la zone d'intégration du prophage d'où les appellations de transduction spécialisée ou localisée. Elle ne correspond pas à une erreur d'encapsidation (transduction généralisée) mais à une excision anormale du prophage lorsqu'un cycle productif succède à une phase de lysogénie (donc lorsque le répresseur déterminant la lysogénie est inactivé).

La plupart des bactériophages étudiés à ce jour présentent un spectre d'hôtes étroit ce qui en théorie restreint la propension des gènes bactériens à être transférés à des hôtes phylogénétiquement éloignés. En contrepartie, les bactériophages libérés dans l'environnement sont plus résistants aux contraintes physico-chimiques que l'ADN extracellulaire et l'ADN encapsidé présente donc une persistance accrue (Wommack & Colwell, 2000). Par ailleurs, comme pour la transformation avec de l'ADN nu ce mécanisme de transfert de gènes ne nécessite pas de contact cellule à cellule. Le rôle de la transduction comme processus de transfert horizontal de gènes en conditions environnementales est lié à l'abondance des bactériophages dans différents milieux environnementaux (Ashelford *et al.*, 2003) et à la présence de séquences de prophages dans la plupart des génomes bactériens séquencés, preuve indirecte de l'importance de ce mécanisme (Canchaya *et al.*, 2003). Plusieurs déterminants de pathogénicité (toxines) ont été acquis via des phages comme par exemple la toxine Shiga produite par *E. coli* (Brussow *et al.*, 2004). Les îlots de pathogénicité sont de larges séquences génomiques qui portent un ou plusieurs gènes de virulence, qui proviennent souvent de bactériophages lysogènes et sont supposés être plus fréquents dans des souches pathogènes que non pathogènes (Dobrindt *et al.*, 2004). Cependant l'annotation complète des séquences de génomes a révélé que certaines souches non pathogènes peuvent aussi porter des îlots génomiques codant des caractères tels que des adhésines, des systèmes de capture du fer ou des protéases, qui contribuent à l'adaptabilité et à la compétitivité de ces microorganismes pourtant dépourvus des facteurs de virulence (Grozdanov *et al.*, 2004).

### 3.2.3. La conjugaison

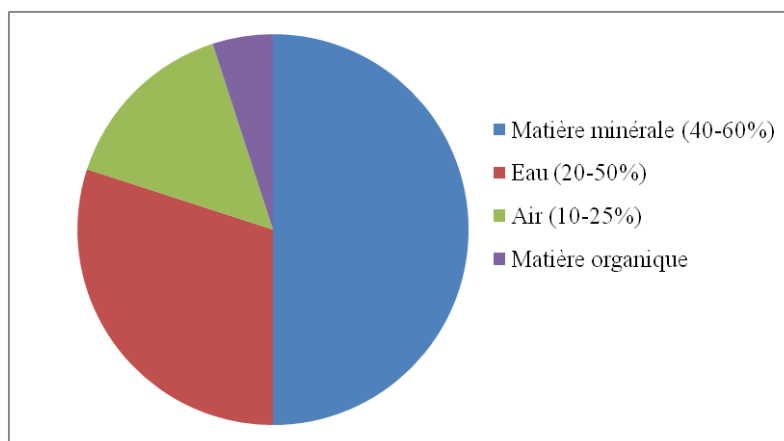
La conjugaison est le procédé par lequel une molécule d'ADN auto-répliquative (plasmide ou transposon conjugatif) est transférée entre cellule donatrice et réceptrice liée physiquement par l'appareil de conjugaison (Zechner, 2000). Bien que les grands principes soient communs entre la plupart des systèmes de conjugaison notamment la synthèse du pilus de conjugaison, il existe une diversité remarquable de systèmes de conjugaison notamment entre bactéries gram+ et protéobactéries. Les plasmides qui ne possèdent pas tous les gènes codant les protéines nécessaires à la conjugaison peuvent encore être transférés dans des cellules réceptrices grâce à l'action du système conjugatif d'autres plasmides de la cellule donatrice, après formation de co-intégrats et même par transduction via des bactériophages ou par transformation (Frey, 1989). Une propriété critique des plasmides est leur gamme d'hôtes qui peut être plus ou moins étendue suivant le type d'hôtes concernés (Heuer *et al.*, 2007). La proportion de plasmides portés par les populations bactériennes dépend donc de l'élément répliquatif lui-même, de l'espèce bactérienne concernée, de l'environnement étudié et de la pression de sélection (Johnsen & Kroer, 2007). Les plasmides possédant une gamme d'hôtes large ont souvent perdu des sites de reconnaissance pour des endonucléases de restriction spécifiques de ces souches hôtes et peuvent aussi porter des gènes codant des systèmes anti-restriction qui minimisent l'effet du clivage par des nucléases spécifiques, systèmes développés par ces organismes pour se protéger de l'invasion par de l'ADN étranger. Ces éléments génétiques mobiles augmentent les chances de survie des souches réceptrices en leur conférant des avantages sélectifs nouveaux par rapport à leurs voisins. Les communautés bactériennes et l'environnement changent continuellement, ce qui implique le transfert permanent de ces EGM permettant alors d'accroître la vitesse d'adaptation des souches réceptrices. Ces dernières possédant des EGM évoluent et survivent au détriment de celles qui n'en ont pas.

## 4. Les bactéries dans le sol

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. Dans le sol les bactéries sont soumises en permanence à des perturbations qui en modifient la structure des communautés. Pourtant, comme nous l'avons déjà indiqué, étudier cette diversité est extrêmement complexe, du fait du faible taux de cultivabilité des espèces bactériennes et des biais liés à l'extraction des cellules et de leur ADN hors de cette matrice complexe. Seule une meilleure connaissance des composants biotiques et abiotiques du sol, des interactions entre bactéries et ces environnements sera en mesure de permettre d'optimiser les techniques d'extraction de l'ADN et d'accéder ainsi à une fraction plus importante de la diversité génétique que recèle le sol.

### 4.1. Le sol et ses composants

Le sol est à l'interface entre la lithosphère, l'atmosphère et l'hydrosphère (Alexander, 1977) et sert de support à une partie de la biosphère (Gobat, 2003). Il est constitué de cinq composants majeurs : fraction minérale, matière organique, eau, air et organismes vivants (Alexander, 1977) (**Figure 1**). Les matières organiques et minérales s'organisent de manière à créer des vides alors occupés par l'air et l'eau (les pores).



**Figure 1.** Proportion des principaux composants du sol en volume (White, 2006).

*La fraction minérale* 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<sub>2</sub> 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 (>2mm) et la terre fine (<2mm) (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 électronégativement, 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).

***La 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, hemicellulose, 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 & 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, archeobacteria, 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.

**La 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 macroagrégats sont formés par des associations temporaires de microagrégats, de minéraux et de particules de matière organique principalement par enchevêtrement des hyphes fongiques et des racines de plantes (Oades, 1984). Les microagrégats 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 & Richaume, 2001).

**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 10µm, constitués entre les agrégats et à l'intérieur des macroagrégats. Une partie de l'eau est fortement retenue par capillarité à l'intérieur des micropores (<10 µm) (Robert, 1992) constitués à l'intérieur des microagrégats. 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 & 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, 1996).

La proportion relative de liquide et de gaz affectant les phénomènes aérobie et anaérobies, dépend de la saturation des sols et de la circulation à l'intérieur du réseau des pores (Robert, 1992).

#### **4.2. Les bactéries dans le sol**

L'importance de la variabilité spatiale en écologie microbienne du sol a depuis longtemps été reconnue et la distribution spatiale des microbes a été décrite à des échelles allant de quelques millimètres à plusieurs mètres (Grundmann & 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 microagrégats. Mummey *et al.* (2004) ont clairement démontré une stratification spatiale suivant la localisation à l'intérieur et à l'extérieur des microagrégats. Le fractionnement des microagrégats du sol par des irradiations UV permet de comparer les structures de communautés bactériennes au niveau écologique pertinent de la micro-échelle (Mummey, 2004).

A microéchelle, les organisations structurelles du sol fournissent par conséquent différents habitats dans lesquels les bactéries indigènes sont distribuées de manière hétérogène (Ranjard & 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 & Richaume, 2001). Le diamètre maximum des pores les plus fréquemment colonisés par les bactéries est estimé entre 2.5  $\mu\text{m}$  et 9  $\mu\text{m}$  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 microagrégats stables est préférentiellement colonisée, ce qui suggère une combinaison de conditions favorables pour la vie bactérienne dans ces environnements. Les microagrégats sont les habitats les plus favorables pour les bactéries dans la plupart des types de sol. En effet, les régions intérieures des microagrégats peuvent constituer un habitat relativement stable et isolé pour les microorganismes, caractérisé par une pression de prédations 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 microagrégats 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 microagrégats 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.

### 4.3. Obtention du métagénome de sols

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 (Volossiuk *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 & 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, 1998, Hurt, 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, 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<sub>3</sub> ou purification de l'ADN extrait par CaCl<sub>2</sub> (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 0.3µm) sont plus difficiles à lyser que les cellules plus volumineuses (2 à 10µm) (More *et al.*, 1994).

#### **4.4. Facteurs influençant la structure des communautés bactériennes du sol**

Beaucoup de paramètres biotiques et abiotiques interviennent pour moduler la structure des communautés bactériennes des sols tant en termes de diversité que d'effectifs (Hassink, 1993, Borneman *et al.*, 1996, Borneman & 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.

##### **4.4.1. Facteurs abiotiques**

Les structures des communautés microbiennes 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 & 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 microbiennes. Par exemple le pH influence fortement la composition des *Acidobacteria* et des *Actinobacteria*. Le pH est l'un des forts prédicteurs 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 & 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, 2001, Zhou *et al.*, 2002). Tarlera *et al.* 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, 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 (*ex* : 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).

#### 4.4.2. 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, 2008), le niveau de résolution phylogénétique avec lequel sont étudiées les communautés (*ex* : *Pseudomonas*) (Cho & 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, 1998). La présence de plantes résulte en une augmentation significative de l'humidité du sol, de la biomasse et de la respiration microbienne (Singh, 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'exudation 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, 1998 , Nusslein & Tiedje, 1999 , Tiedje, 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, 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, 2009). Des résultats précédents avaient déjà montré qu'un champignon *Glomus mossae* (Murty, 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.

## 5. Conclusions

Ce chapitre bibliographique a tenté de montrer toute la complexité d'un écosystème et plus particulièrement celui du sol, que ce soit au niveau du système biologique qu'au niveau du système abiotique.

La première partie nous a montré l'évolution des techniques permettant l'analyse de la diversité bactérienne, qui promettent d'offrir une étude de plus en plus exhaustive des communautés bactériennes qui se caractérisent par une très forte diversité génétique. Cette diversité s'explique par un mode d'évolution et d'adaptation très complexe qui rend d'autant plus difficile l'étude de ces populations.

Le sol est un environnement structuré, dans lequel les microorganismes sont distribués de manière hétérogène. Les propriétés physico-chimiques du sol ne permettent pas d'accéder à la totalité des bactéries, présentes dans un échantillon, soit à l'étape d'extraction ou bien lors de la purification de l'ADN.

Une étude exhaustive de la diversité bactérienne nécessiterait la combinaison d'une approche permettant d'accéder aux communautés bactériennes minoritaires, avec une approche évitant les biais liés à l'extraction de l'ADN bactérien.



## **CHAPITRE II**

Combinaison d'une approche conceptuelle avec l'utilisation  
d'une technique sensible pour évaluer la diversité  
bactérienne des sols.



# PARTIE I : Evolution des communautés bactériennes inoculées dans des sols stérilisés au cours du temps

## 1. Introduction

Il y a une apparente contradiction entre les observations empiriques attribuant à chaque environnement ses microorganismes propres et l'idée d'un cosmopolitisme beaucoup plus général (de Wit & Bouvier, 2006) selon le concept "Everything is everywhere, and the environment selects" formulé par Baas Becking (Baas Becking, 1934) basé sur ses propres études et fortement influencé par celles de son compatriote Beijerinck (Beijerinck, 1913). Le concept de Baas Becking repose sur un cosmopolitisme bactérien à l'échelle des différents écosystèmes telluriques de la planète mais avec une très forte hétérogénéité dans la répartition des espèces entre les différents environnements. Il est à noter que cette conception n'est pas sans rappeler la notion de génération spontanée qui agita le monde scientifique au XIX<sup>e</sup> siècle, avec comme différence fondamentale toutefois que la vie n'apparaît pas « spontanément » à partir de rien mais que le développement bactérien observable résulte d'une multiplication d'un faible nombre de cellules initiales suite à un changement des conditions environnementales.

Dans la synthèse bibliographique (chapitre I) nous avons indiqué que l'environnement et en particulier le sol comportait une diversité microbienne d'une très grande richesse et présenté les mécanismes moléculaires endogènes qui permettent aux bactéries de réguler la génération de la diversité génétique, mais aussi l'importance de l'environnement abiotique et biotique sur la structuration des communautés bactériennes par le biais des relations avec les autres organismes, notamment les plantes. L'exploration de la diversité bactérienne du sol reste toutefois imparfaite et majoritairement incomplète. S'il est pourtant un sujet scientifique qui passionne l'Homme depuis des millénaires c'est celui de dresser l'inventaire le plus exhaustif possible des espèces qu'il est amené à côtoyer plus ou moins régulièrement. Cela a déjà été largement réalisé pour beaucoup d'organismes de notre biosphère et continue de susciter un intérêt majeur quand les organismes considérés sont des animaux ou des plantes ne serait-ce que pour déterminer l'étendue de l'érosion de la biodiversité suite aux actions anthropiques. Comme nous l'avons vu précédemment la situation est bien plus complexe en ce qui concerne le monde bactérien pour lequel (presque) tout reste à faire. L'intérêt scientifique dépasse

toutefois celui lié à un inventaire aussi informatif puisse-t-il être car il concerne le fonctionnement de l'écosystème du fait des fonctions des microorganismes qui vont agir seuls ou en synergie. Nous avons vu toute l'importance que peuvent avoir les populations même les plus limitées en effectif du fait de leur capacité à sur-exprimer certains de leurs gènes, à survivre en état de latence jusqu'à ce que les conditions autorisent leur multiplication leur permettant ainsi de participer à la redondance fonctionnelle au sein de l'écosystème et à son potentiel de résilience et enfin à être un inépuisable réservoir de gènes transférables à d'autres individus de la communauté. Plus conceptuellement, cet inventaire populationnel est nécessaire pour comprendre les stratégies adaptatives et évolutives du monde procaryote qui seraient nécessairement diamétralement opposées s'il s'avérait que chaque gramme de sol héberge la quasi-totalité de la diversité bactérienne tellurique ou bien si chaque sol se caractérise par une spécificité forte de sa microflore. Les conséquences sociétales ne seraient pas non plus identiques puisqu'une spécialisation forte ferait redouter la perte de diversité en cas de stress important sur ces écosystèmes alors que le cosmopolitisme protège beaucoup mieux le maintien des ressources génétiques.

Notre travail se positionne dans cette problématique d'inventaire de la diversité microbienne des sols et pousse son ambition jusqu'à tenter de répondre au concept de cosmopolitisme proposé par Baas Becking (1934) pour le confirmer ou l'infirmer. Nos stratégies d'étude ont été basées sur deux piliers expérimentaux, l'un conceptuel, l'autre technique. L'approche conceptuelle se base sur les nombreuses données de la littérature montrant l'influence de l'environnement abiotique sur la structuration des communautés bactériennes. S'il est maintenant admis que chaque communauté bactérienne est la résultante d'interactions complexes entre les différentes populations et avec leur environnement abiotique il peut être postulé que soumettre cette communauté à un autre environnement devrait en modifier la structure et notamment les effectifs des différentes populations. Par ce biais il peut être envisagé que certaines des bactéries présentant initialement des effectifs extrêmement restreints donc indétectables, puissent suffisamment se multiplier pour être détectées. La répétition de ce type d'expérimentations d'inoculations d'une communauté bactérienne dans différents types de sol permettant d'étendre significativement l'accès à la biodiversité initiale. Nos objectifs ont été de mettre en place les protocoles d'extraction des cellules bactériennes à partir de différents types de sol, d'inoculer et d'incuber ces communautés bactériennes dans des échantillons de sol stérilisé avant d'extraire l'ADN de la microflore développée à partir de l'inoculum de départ pour l'analyser par des techniques moléculaires appropriées. Dans un

premier temps, cette approche a été validée, grâce à la technique RISA, et les échantillons d'ADN ont été criblés rapidement.

Dans un second temps cette approche conceptuelle a été combinée à l'utilisation d'une technique à fort potentiel qui est la technique des puces à ADNr 16S. Si cette technique permet de détecter des changements dans les communautés bactériennes et qui viendrait donc en complément de l'approche RISA, c'est surtout sa capacité à identifier avec un niveau élevé de spécificité et de sensibilité un nombre de plus en plus élevé de taxons bactériens. La technique des puces à ADN est alors utilisée non seulement pour analyser l'évolution de structure des communautés bactériennes mais surtout pour identifier les populations présentes et potentiellement détecter les bactéries initialement rares dans leurs écosystèmes originaux et dont le changement de milieu aurait permis la multiplication jusqu'à un niveau détectable.

## 2. Matériels & méthodes

### 2.1. Préparation et conditionnement des sols

Neuf sols provenant de France, d'Afrique, de Nouvelle Calédonie et de Martinique, ayant des caractéristiques physico-chimiques différentes et marquées, ont été sélectionnés. Les propriétés des sols sont présentées dans le **Tableau 1**. Les sols S8 et S9 sont des sols provenant du projet Ecomiq-RMQS. Les sols ont été séchés à température ambiante puis tamisés à 2mm. Tous les sols ont été stérilisés par irradiation gamma (25kGy), puis incubés durant 2 mois à température ambiante jusqu'à extinction complète des activités enzymatiques (Lensi, 1991). Les microcosmes ont été réalisés en plaçant 5g de sol stérilisé dans des récipients de 20 ml en verre dans le but d'une inoculation en 6 répliques de chaque type de sol par chacune des communautés bactériennes dans 54 microcosmes et donc un total de 486 pour les 9 sols. Pour parfaire la stérilisation les microcosmes ont été autoclavés avant inoculation, l'efficacité finale de la stérilisation étant vérifiée par suspension-dilution d'un peu de sol et étalement sur milieux nutritifs solidifiés.



**Tableau 1.** Caractéristiques des neufs sols.

Sols	Localisation	Sable (%)	Limon (%)	Argile (%)	pH eau	pHKCL	Calcaire total (g/kg)	Carbone organique (g/kg)	Matière organique (g/kg)	Azote total (g/kg)	C/N
S1	Côte Saint André (France)	42,9	43,6	13,5	7,24	6,66	10	17	29,2	1,7	9
S2	Montrond (France)	44,6	34,8	20,6	5,59	5	0	37,1	63,8	3,6	10
S3	Brévil (France)	58,3	25,2	16,5	4,44	3,82	0	59,4	102,2	2,7	21
S4	Chinon (France)	87	5	8	8,57	8,12	159	6,5	11,2	0,5	12
S5	Embu (Kenya)	11,6	65,6	22,7	6,78	5,78	0	24,9	42,9	2	11
S6	Black point (Congo)	88	2,2	9,8	4,79	4,37	0	5,1	8,9	0,4	12
S7	New caledonia	94,4	4,3	1,4	5,82	5,9	0	5,2	8,9	0	0
S8	Saint -Anne (Martinique)	18,8	26,6	54,7	7,33	5,54	5	21,3	36,6	2,1	10
S9	Talmon Saint Hilaire (France)	17,3	73	9,7	5,4	4,34	0	8,8	15,1	1	8

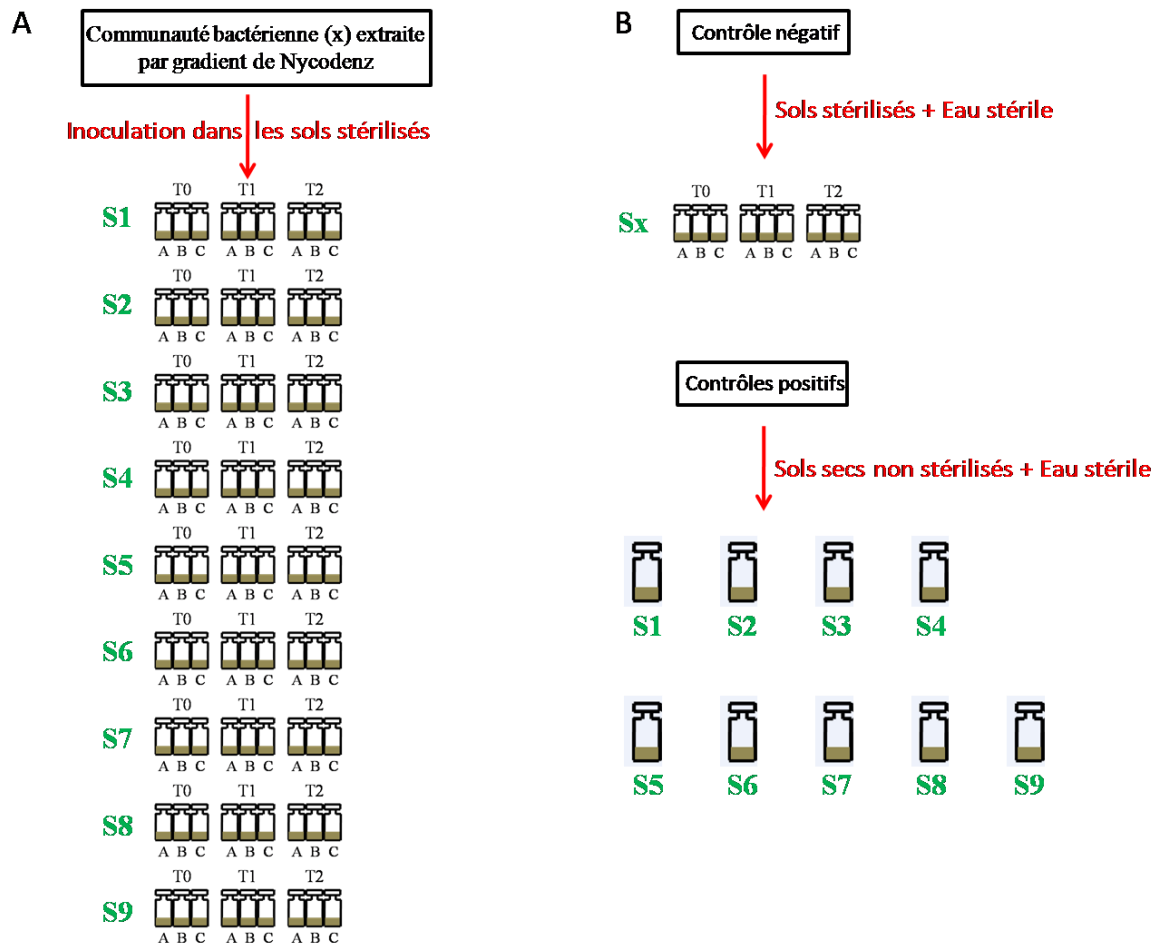
## 2.2. Extraction et inoculation des communautés bactériennes

Les communautés bactériennes des sols bruts S1, S2, S3 et S5 ont été extraites par centrifugation haute vitesse sur un gradient de densité Nycodenz (Axis-Shield PoC, Oslo, Norway) comme indiqué ci-dessous. Soixante grammes de sol ont été dilués dans 180ml d'hexamétaphosphate en présence d'une vingtaine de billes en verre de 2,5mm de diamètre et homogénéisés par agitation durant 1 heure. Les suspensions sont ensuite centrifugées pendant 1min à 300rpm et 10°C pour une récupération du surnageant, contenant la très grande majorité des cellules bactériennes, lui-même soumis à une seconde centrifugation de 20 min à 4000rpm et 10°C de façon à sédimenter les cellules. Le culot cellulaire est remis en suspension dans 20 ml de NaCl 0.8%. Cette suspension est additionnée de 11ml de Nycodenz (0.8g ml<sup>-1</sup> de densité), puis soumise à centrifugation pendant 40min à 4000rpm et 10°C. L'anneau composé des cellules bactériennes regroupées d'après leur densité au sommet du gradient formé par le polymère est prélevé soigneusement à l'aide d'une pipette.

La suspension cellulaire est ensuite purifiée (élimination du Nycodenz) par dilution dans de l'eau stérile (40ml) suivie d'une centrifugation (4000rpm, 15min, 10°C). Le culot cellulaire est resuspendu dans 1ml de NaCl 0.8% et 1ml de glycérol (50%) puis stocké à -80°C. Ce protocole a été appliqué autant de fois qu'il était nécessaire pour traiter 2,160kg, 1,360kg, 0,700kg et 0,800kg des sols S1, S2, S3 et S5, respectivement avant mise en commun de tous les échantillons d'un même sol.

L'efficacité de l'extraction cellulaire est réalisée par comptage des cellules colorées avec du BET et visionnée à l'aide d'un microscope à épi-fluorescence. Approximativement 30 surfaces sélectionnées au hasard ont été comptées par filtre, chaque surface contenait en général quelques centaines de cellules. La moyenne des cellules obtenues sur ces 30 surfaces a été calculée, puis la concentration en bactéries/ml de chaque communauté bactérienne extraite par Nycodenz a été calculée grâce aux surfaces connues de comptage et du filtre. Une dernière communauté bactérienne provenant d'eau de lac a été obtenue par ultrafiltration de 40l d'eau. La densité bactérienne finale est mesurée par la technique de cytométrie de flux et est de 7,1.10<sup>9</sup> bactéries/ml. Les communautés bactériennes des sols S1, S2, S3 et S5, nommées respectivement CS1, CS2, CS3 et CS5 sont inocuées indépendamment dans leur propre sol et dans les 8 autres sols stérilisés à une concentration égale à 10<sup>7</sup> bactéries /gramme de sol sec. La communauté bactérienne des eaux du lac nommée CL a été elle aussi inocuée dans les 9 sols stérilisés, dans les mêmes conditions. Après l'inoculation, tous les sols étaient à 50% d'humidité. Une communauté bactérienne a donc été inocuée dans 81 microcosmes

pour suivre l'évolution de cette communauté dans les 9 sols à 3 temps (T0=1jour, T1=2mois et T2=6mois) avec 3 répliques (**Figure1**). Tous les microcosmes ont été incubés à température constante de 20°C dans des boites hermétiques en présence d'eau stérile pour maintenir l'humidité stable.



**Figure 1.** A/ Schéma d'inoculation d'une communauté bactérienne dans les 9 sols stérilisés ; B/ Contrôle négatif pour une communauté inoculée, ainsi que les contrôles positifs valables pour toutes les communautés inoculées.

### 2.3. Préparation des contrôles négatifs et positifs

Dans un premier temps, une extraction ADN a été réalisée à partir de 0.3g des 9 sols stérilisés par rayon gamma puis conditionnés dans des pots en verre et enfin autoclavés, à l'aide du kit Mobio PowerSoil (Mobio, Carlsbad, USA). Ensuite pour chaque communauté inoculée dans les 9 sols stérilisés, 1 contrôle négatif contenant le sol stérilisé à partir duquel la communauté utilisée a été extraite, a été préparé en ajoutant de l'eau stérile pour se placer à 50% d'humidité, comme dans les autres microcosmes. Ce témoin négatif permet de valider les

conditions de stérilité durant les 3 temps de l'expérience avec 3 répliques (**Figure 1. A/**). Ces témoins négatifs ont été incubés dans les mêmes conditions que les microcosmes inoculés par les communautés bactériennes (ci-dessus).

En parallèle 5g des 9 sols secs non stérilisés ont été placés dans les récipients en verre de 20ml, stérilisés, puis de l'eau stérile a été ajoutée pour aboutir à un taux d'humidité de 50%. Les microcosmes sont alors incubés à 20°C, dans des boîtes hermétiques avec une humidité constante. Ces microcosmes permettent d'obtenir les communautés bactériennes des 9 sols dans les conditions identiques à celles des microcosmes contenant les sols stérilisés inoculés.

#### **2.4. Extraction de l'ADN et amplification de la région intergénique 16S-23S**

L'ADN a été extrait à partir de 0.3g de chacune des 3 répliques de chacun des 9 sols inoculés par chaque communauté bactérienne après un jour, 2 mois et 6 mois d'incubation en utilisant le kit Mobio PowerSoil suivant les recommandations du fournisseur. De plus l'ADN a été extrait directement à partir des cellules bactériennes récupérées sur le gradient Nycodenz des sols S1, S2, S3, et S5 à l'aide du kit MasterPure™ *Gram Positive DNA Purification* (Epicentre, Madison, USA), suivant les recommandations du fournisseur. Ces échantillons représentent les communautés bactériennes initiales utilisées comme inocula dans les sols stérilisés. Ces échantillons d'ADN sont nommés ENS1, ENS2, ENS3, ENS5, respectivement. La région intergénique (IGS) localisée entre la petite et la grande sous unité du gène rDNA a été amplifiée avec les amorces suivantes : RISA Rev (petite sous-unité du gène rDNA eubactérien, 5'-TGCGGCTGGATCCCCTCCTT-3') et Risa Fwd (grande sous unité du gène rDNA eubactérien, 5'-CCGGGTTTCCCCATTCGG-3') (Normand, 1996). Les séquences amplifiées contiennent l'IGS plus approximativement 150pb correspondant aux 20 nucléotides de l'amorce RISA Rev et environ 130pb du gène 23S rDNA. Les amplifications à partir des échantillons d'ADN sont réalisées à l'aide du kit Illustra Hot Start Mix RTG (GE Healthcare, Buckinghamshire, UK) qui combine une Taq ADN polymérase recombinante de haute qualité, une protéine activatrice recombinante Hot Start, et des nucléotides (sous forme d'une bille contenue dans un tube 0.2ml). Seules les amorces (à 1mM), de l'eau (qsp : 25µl) et la matrice d'ADN sont ajoutés pour compléter la solution réactionnelle. Les conditions PCR utilisées sont les suivantes : une dénaturation initiale de 10min à 95°C, suivi de 29 cycles : 94°C (45s), 55°C (45s), 72°C (60s), et une extension finale de 7min à 72°C.

## **2.5. Empreintes moléculaires RISA des communautés bactériennes extraites des échantillons de sol**

Les produits d'amplification ramenés à la même concentration sont déposés sur les lames à l'aide du kit Agilent DNA 7500 suivant les recommandations du fournisseur. Les profils RISA ont été obtenus à l'aide du Bioanalyser 2100 (Agilent, Santa Clara, CA). Le Bioanalyser détecte à l'aide de micro-capillaires la fluorescence des différentes bandes au cours du temps, basé sur des micro-flux. Le logiciel 2100 expert convertit la fluorescence détectée en électrophérogramme, les pics représentant les différentes tailles de fragment, et la hauteur de ces pics la proportion relative des fragments dans le produit total. Il est possible d'exporter un tableau constitué toujours des mêmes 1481 lignes représentant les temps (en seconde) et les valeurs de fluorescence enregistrées à ces différents temps. Les données RISA obtenues à l'aide du Bioanalyser ont été normalisées par la méthode des quantiles.

## **2.6. Les puces à ADN taxonomiques**

Le format de la puce à ADN utilisée dans ces expériences est celui d'Agilent Sureprint Technologies. La lame est constituée par 8 blocs de 15 000 emplacements chacun, sur un format standard de lame de verre de 1''x 3'' (25mm x 75mm). Chaque emplacement est réservé à la synthèse *in situ* de sondes de 20 oligonucléotides. Chaque sonde est présente en plusieurs exemplaires, minimum 3, à l'intérieur de chaque bloc. Tous les blocs sont identiques. Ce format est prévu pour l'hybridation conjointe de 8 échantillons différents (produits PCR) sur la même lame, nécessitant donc l'utilisation de plusieurs lames quand plus de 8 échantillons doivent être analysés. Les sondes ont été construites pour cibler le gène *rrs* et couvrir une large proportion de l'arbre phylogénétique des bactéries et des archées. Certaines sondes ont été conçues à l'aide du progiciel ARB, d'autres à l'aide du logiciel Phylarray (Peyret, 2007) dans le cadre d'une collaboration avec l'équipe Génomique Intégrée des Interactions Microbiennes (Clermont Ferrand). Les sondes ont des longueurs de 20 nucléotides, avec des températures de fusion variant de 65°C+/-5°C, un taux de mésappariement pondéré inférieur à 1.5, la conception incluant la caractérisation de sondes oligonucléotidiques ciblant les organismes à des niveaux taxonomiques différents.

## **2.7. Amplification du gène 16S r DNA**

Le gène 16S rDNA a été amplifié à partir des échantillons d'ADN en utilisant les amorces universelles (pour la plupart des bactéries et quelques archées) pA (5'-GTT TGA TCC TGG

CTC AG-3') and pH-T7 (5'-TAA TAC GAC TCA CTA TAG AAG GAG GTG ATC CAG CCG CA-3'). Les PCR ont été réalisées à l'aide du kit Hot Start RTG (GE Healthcare, Buckinghamshire, UK) comme ci-dessous, avec les réactions PCR des 3 répliques effectuées indépendamment, puis mises en commun avant analyse. Chaque mélange de réaction PCR provenant du kit contenait, dans un tube 0.2ml, une bille qui combine une Taq ADN polymérase recombinante de haute qualité, une protéine activatrice recombinante Hot Start, et des nucléotides. Seules les amorces (à 1mM), l'eau (qsp : 25µl) et la matrice d'ADN sont ajoutés (2µl). Les conditions PCR utilisées sont les suivantes : une dénaturation initiale de 5 min à 95°C, suivi de 29 cycles : 94°C (45s), 55°C (45s), 72°C (60s), et une extension finale de 7min à 72°C.

## 2.8. Protocole d'hybridation sur puces à ADN

Les produits PCR des trois répliques mis en commun pour chaque échantillon ont été purifiés à l'aide du kit Illustra GFX PCR DNA and Gel purification (GE Healthcare, Buckinghamshire, UK) suivant les recommandations du fournisseur.

Les produits PCR purifiés ont ensuite été marqués à l'aide d'UTP-Cy3 durant la transcription *in vitro* (GE Healthcare, Buckinghamshire, UK). Le Cy3 est un marqueur fluorescent émettant de la lumière à la longueur d'onde égale à 532 nm. Chaque mix (12µl) de transcription contenait : 4µl de tampon pour l'ARN-T7, 2µl de DDT (100mM), 1µl d'ATP, de CTP, de GTP (chacun à 10mM), 0.5µl d'UTP (10mM), 0.5µl d'RNAsine (40U/µl), 1µl d'ARN-T7 polymérase, de l'UTP-Cy3 (5mM). Huit µl de produit PCR ont été ajoutés aux 12µl du mélange de transcription avant incubation 37°C pendant 4 heures.

L'ARN marqué a été purifié à l'aide du minikit RNeasy (Qiagen, Hilden, Germany) en utilisant le protocole RNA Cleanup, suivant les recommandations du fournisseur. La concentration en ARN et en Cy3 a été mesurée à l'aide du nanophotomètre (IMPLEN, Wehrlestrasse, Germany). L'ARN a alors été fragmenté en fragments de 20pb par ajout de Tris CL (1mM) et ZnSO4 (100mM), le mélange étant incubé à 60°C pendant 30min avant blocage de la réaction par incubation des tubes dans la glace et en ajoutant de l'EDTA (à 500mM), et de l'RNAsin (40U/µl). Ensuite, l'ARN marqué et fragmenté de chaque échantillon a été déposé sur les lames construites par Agilent suivant un plan de disposition des oligosondes préalablement défini, et incubés à 60°C avec une légère rotation durant 16 heures. Le lavement des puces a été effectué suivant les protocoles standards Agilent.

## 2.9. Lecture des lames et traitement des données

Les puces ont été scannées à l'aide du scanner GenePix (Bucher Biotec, Viaduktstrasse, Switzerland), avec une taille de pixel de 5 $\mu$ m. La fluorescence a été transformée en image à l'aide du logiciel GenePix Pro, qui réduit les données à une valeur de signal pour chaque sonde. Les signaux de fluorescence des hybridations brutes ont été calculés en faisant la moyenne de la fluorescence de chaque pixel pour une sonde donnée. La valeur de bruit de fond a été calculée suivant la formule suivante : (moyenne des répliques du témoin négatif) + (3 x écart type du témoin négatif), formule standard IUPAC. La moyenne des différentes répliques de chaque sonde a été calculée en enlevant les valeurs aberrantes lorsque la valeur était supérieure à 2 écarts type de leur moyenne. Le bruit de fond a alors été soustrait à la moyenne corrigée de chaque sonde. Les résultats ont alors été normalisés par la méthode des quantiles. Les sondes ayant une valeur inférieure à zéro sont considérées comme non détectées et celles supérieures à zéro comme détectées.

## 2.10. Reproductibilité des résultats

La reproductibilité des expériences a été vérifiée tant pour la méthode RISA que pour la technique des puces à ADN.

Les échantillons d'ADN extraits des sols inoculés par la communauté de la Côte Saint André (CS1) à T1=2mois et T2=6mois ont été analysés par la méthode RISA. L'ADN a été extrait des 9 sols à partir de 0.3g des 3 répliques pour les deux durées d'incubation. Ces échantillons d'ADN ont alors été amplifiés séparément, puis les profils RISA des trois répliques de chaque échantillon de sol ont été comparés.

La reproductibilité de la méthode des puces à ADN a été contrôlée avec les échantillons d'ADN extraits du sol S1 inoculé par la communauté de la Côte Saint André à T2=6mois. L'ADN a été extrait à partir de 0.3g du sol S1 des 3 répliques au temps T2=6mois. Ces échantillons d'ADN ont été amplifiés séparément, puis analysés indépendamment à l'aide des puces à ADN. En parallèle les produits PCR des trois répliques ont été mis en commun (comme pour le reste des échantillons) puis analysés par hybridation sur puces à ADN.

## 2.11. Analyses statistiques

Les analyses en composante principale obtenues à partir des données normalisées RISA ou des puces à ADN, ont été utilisées pour évaluer les différences de structure des communautés bactériennes. Les ACP ont été réalisées à l'aide du logiciel ADE-4 software (Thioulouse,

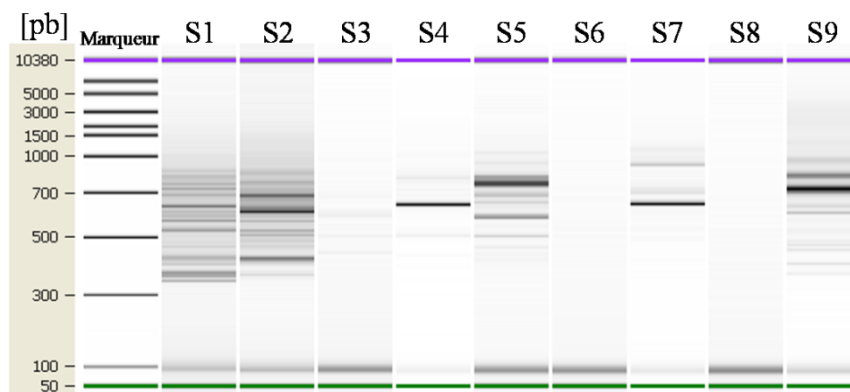
1997). Des ACP inter classes ont été réalisées à partir des données normalisées à l'aide du logiciel R.

### 3. Résultats

#### 3.1. Validation de l'approche conceptuelle et criblage des premiers résultats

##### 3.1.1. Extraction et amplification des échantillons d'ADN

L'amplification de l'espace intergénique bactérien (IGS) n'a pas été possible avec les ADN extraits des sols S3, S6 et S8 provenant de Brévil (France), Pointe Noire (Congo) et Saint-Anne (Martinique), quelle que soit la communauté inoculée, comme le montrent les profils RISA (**Figure 2**) obtenus à l'aide du bioanalyser à partir des échantillons d'ADN extraits des 9 sols inoculés avec la communauté CS2=Montrond, après 6 mois (T2) d'incubation (aucune bande présente sur le gel pour S3, S6 et S8 tandis que des profils plus ou moins complexes sont observés avec l'ADN extrait des autres sols).

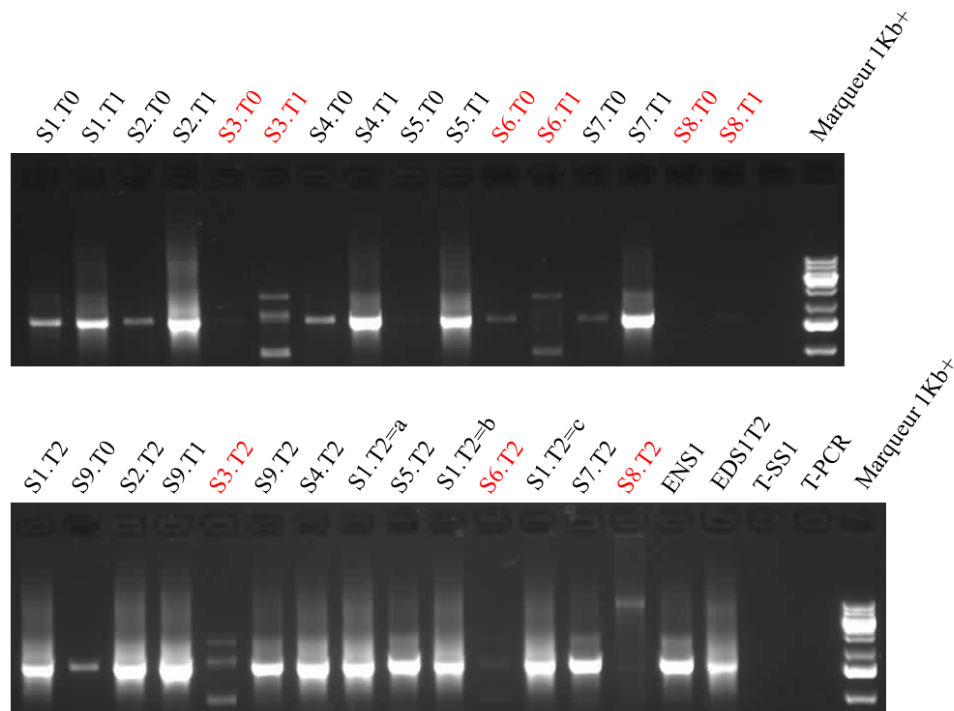


**Figure 2.** Profils RISA obtenus à partir des échantillons d'ADN extraits des 9 sols inoculés par la communauté CS2=Montrond, à T2=6mois.

L'amplification du gène 16S rDNA à partir des échantillons d'ADN de ces trois sols présente aussi des difficultés. La **Figure 3** représente les produits d'amplification du gène 16S rDNA à partir des échantillons d'ADN provenant des 9 sols inoculés avec la communauté CS1 et extraits aux temps T0, T1 et T2. Les produits PCR des sols S3, S6 aux temps T1 et T2, présentent des bandes aspécifiques. Les produits PCR du sol S3 à T1 et T2 présentent la bande attendue mais en très faible quantité par rapport aux autres bandes aspécifiques. Pour le sol S6 au temps T1 la bande attendue n'est pas présente, et au temps T2 une légère bande de la bonne taille est discernable. L'amplification du gène 16S rDNA pour le sol S8 aux temps T0 et T1 ne donne aucune bande et au temps T2 des bandes aspécifiques apparaissent. Pour



améliorer l'efficacité des amplifications du gène 16S rDNA des sols S3, S6 et S8, plusieurs stratégies ont été adoptées.



**Figure 3.** Amplification du gène 16S rDNA à partir des échantillons d'ADN extraits des 9 sols inoculés avec la communauté bactérienne de la Côte Saint André (CS1) aux temps T0=1jour, T1=2mois et T2=6mois.

Dans un premier temps une nouvelle tentative d'amplification a été réalisée à partir des échantillons d'ADN du sol S3, S6, et S8 inoculés par la communauté CS2 et extraits au temps T2 mais après dilution au  $1/10^{\text{ème}}$ , et au  $100^{\text{ème}}$  et au  $1/1000^{\text{ème}}$  des solutions obtenues.

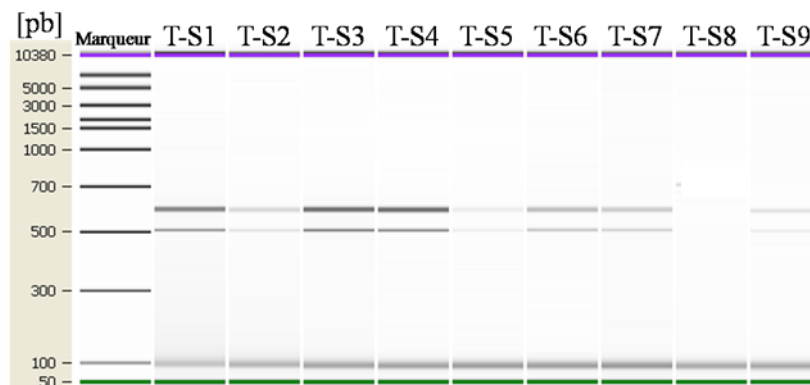
En effet, au temps T0 l'amplification du gène 16S rDNA à partir des ADN des sols S3 et S6 est positive indiquant que la concentration d'ADN (plus faible à T=0) peut influencer sur l'efficacité de l'amplification PCR. Cette hypothèse est toutefois infirmée, la dilution n'ayant eu aucun impact sur l'intensité des produits PCR aspécifiques.

Les échantillons d'ADN des sols S3, S6 et S8 inoculés par la communauté CS2 et extraits au temps T2=b et c ont été purifiés selon deux techniques alternatives, soit par filtration sur colonne d'affinité (kit GE Healthcare, GFX Microspin column), soit par combinaison de deux techniques (kits Nucleospin RNA II et RNA/DNA Buffer 7) (Macherey Nagel, Duren, Allemagne). Cette dernière technique a permis de limiter le nombre de bandes aspécifiques pour le sol S3 mais dans tous les cas la bande contenant l'ADN du gène 16S rDNA est en trop faible quantité pour, après purification sur gel, permettre son marquage. Pour l'ADN extrait

du sol S6 la bande attendue n'est pas présente et aucune bande n'a été détectée avec l'ADN extrait du sol S8.

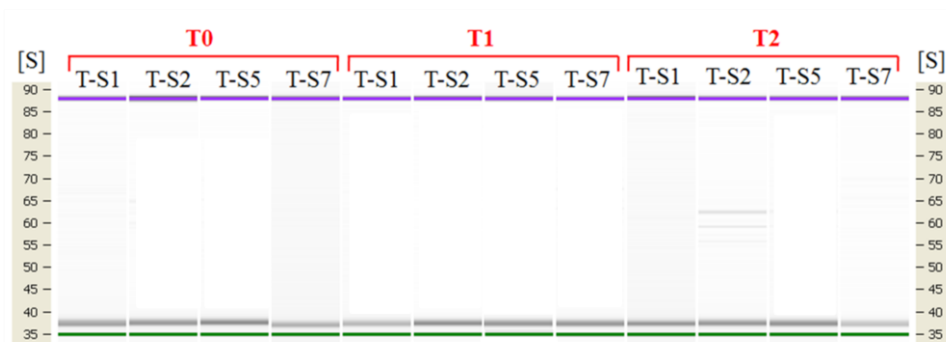
### 3.1.2. Contrôles négatifs et positifs

Une amplification de la région intergénique à partir des ADN extraits des 9 sols secs stérilisés par irradiation aux rayons gamma, puis conditionnés dans des pots en verre et enfin autoclavés, a été réalisée. Les profils RISA obtenus (**Figure 4.**) montrent la présence de 2 bandes identiques pour chaque échantillon.



**Figure 4.** Profils RISA obtenus à partir des échantillons d'ADN extraits des 9 sols secs après stérilisation par rayons gamma et par autoclavage.

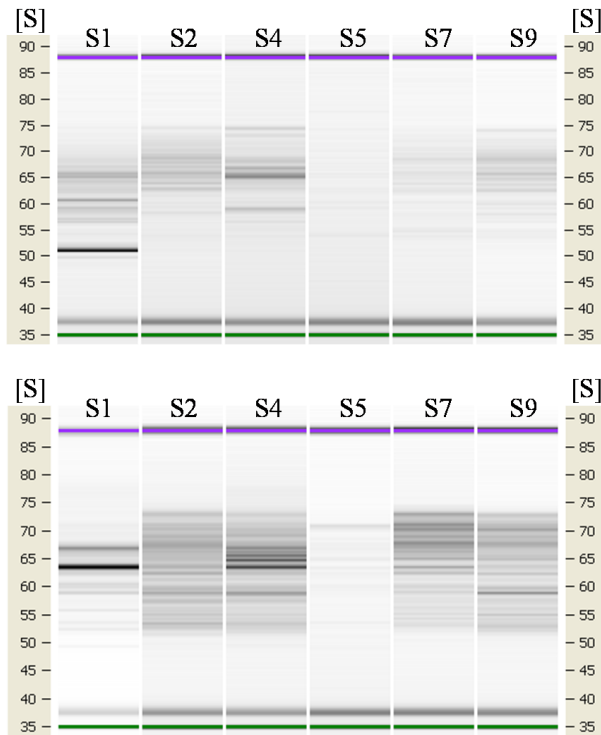
A partir de ces sols secs, une suspension-dilution a été réalisée puis étalée sur du bouillon nutritif gélosé en boîte de pétri. Après 24, 48h et 72h aucune colonie bactérienne n'a pu y être détectée. L'ADN a également été extrait des échantillons utilisés comme témoins négatifs d'inoculation: T-S1, T-S2 ; T-S5 ; T-S7, inoculés avec de l'eau stérile et incubés sur la même durée (T0=1 jour, T1=2 mois, T2=6mois) et dans les mêmes conditions que les sols inoculés. Les profils RISA obtenus après l'amplification de la région intergénique (**Figure 5.**) au cours du temps ne présentent aucune bande, sauf pour le T-S2 à T2=6mois où les deux bandes sont légèrement visibles.



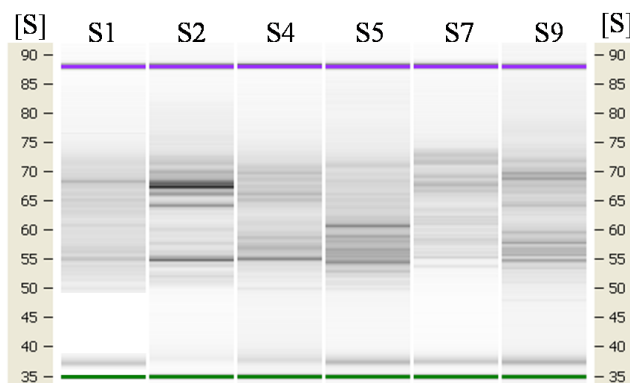
**Figure 5.** Profils RISA obtenus à partir des échantillons d'ADN extraits des contrôles négatifs inoculés avec de l'eau stériles et incubés dans les mêmes conditions que les sols stérilisés puis inoculés, au cours du temps : T0=1jour, T1=2mois et T2=6mois.

Une amplification du gène 16S rDNA a été réalisée à partir de l'ADN extrait des témoins négatifs T-S1, T-S2, T-S3, T-S5 et T-S7, aux temps T0, T1 et T2. Les produits PCR ont été visualisés sur gel d'agarose après électrophorèse aucune bande n'a alors été détectée.

Au niveau des profils RISA obtenus à partir des échantillons d'ADN extraits des sols inoculés par les 5 communautés, les deux bandes n'ont jamais été détectées, par exemple sur la **Figure 7** où sont représentés les profils RISA obtenus à partir des échantillons d'ADN des 6 sols inoculés par les communautés CS1 et CS2 au temps T0=1jour, aucune bande présente dans tous les échantillons clairement distincte n'est visualisée. En parallèle les profils RISA au temps T2=6mois, obtenus à partir des échantillons d'ADN extraits des 6 sols bruts inoculés avec de l'eau stérile et incubés dans les mêmes conditions que les sols stérilisés de l'expérience sont présentés **Figure 8**. Les profils RISA obtenus sont complexes et différents les uns des autres. Ces contrôles positifs nous permettent d'évaluer les communautés initiales de chacun des sols dans les conditions de l'expérience.



**Figure 7.** Profils RISA obtenus à partir des échantillons d'ADN extraits des 6 sols inoculés par les communautés : **A/** CS1=Côte Saint André ; **B/** CS2=Montrond, au temps T0=1jour

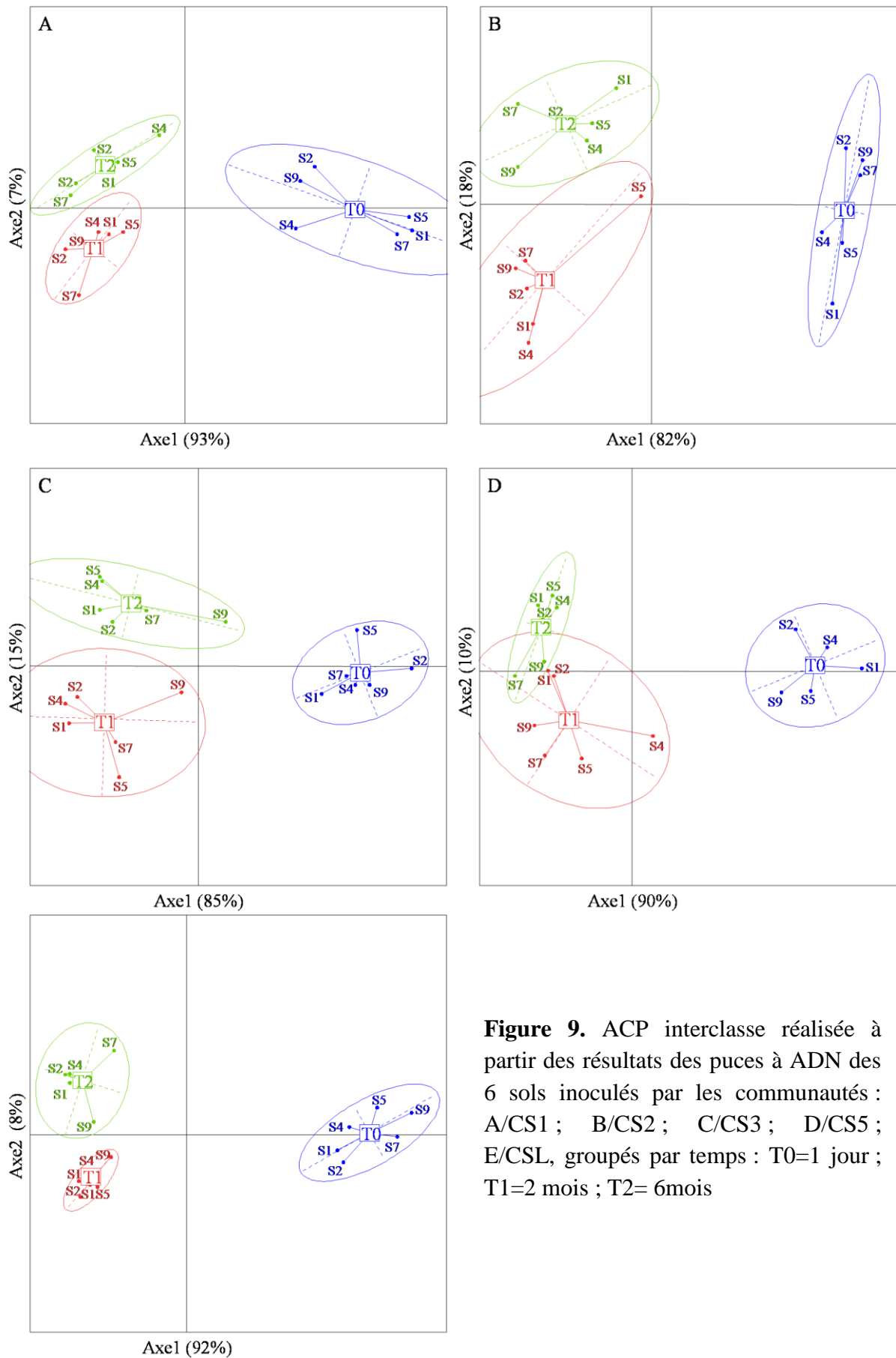


**Figure 7.** Profils RISA obtenus à partir des échantillons d'ADN extrait des contrôles positifs : sols bruts (non stériles) inoculés avec de l'eau stérile et incubés dans les mêmes conditions que les sols stérilisés puis inoculés.

### 3.2. Evaluation des structures des communautés bactériennes

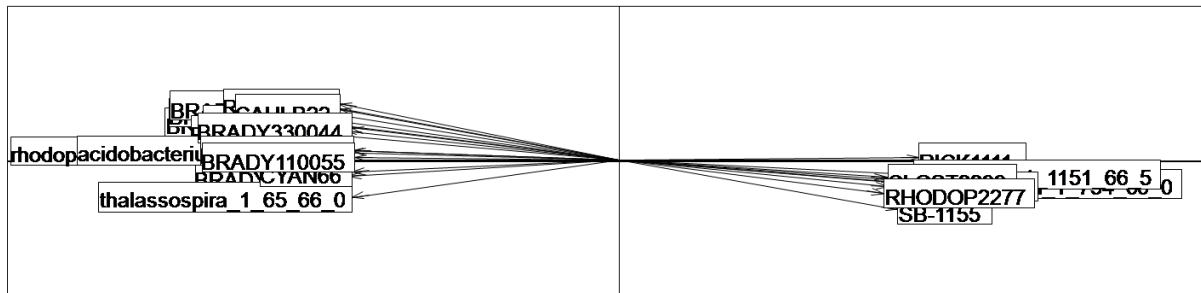
*Au cours du temps :* Lorsqu'une communauté bactérienne est inoculée dans les 6 sols stérilisés, puis extraite au temps T0=1jour, les profils RISA obtenus sont différents entre les 6 sols. Par exemple la **Figure 8** représente les profils RISA obtenus à partir des échantillons extraits des 6 sols inoculés à partir des communautés A/CS1 et B/CS2, indépendamment, au temps T0. Les échantillons d'ADN extraits des 6 sols inoculés par la communauté CS1 ont

également chacun des profils RISA différents de ceux obtenus lorsque les 6 sols ont été inoculés par la communauté CS2. Une ACP interclasse avec les résultats de l'hybridation sur puces à ADN a été réalisée regroupant en fonction des 3 temps les 6 sols inoculés par une communauté bactérienne. En considérant les durées d'incubation ( $T_0=1$ jour,  $T_1=2$ mois et  $T_2=6$ mois), une évolution de la structure de la communauté bactérienne inoculée dans chacun des 6 sols est observable au cours du temps. En effet, les groupes formés des 6 sols aux temps  $T_0$ ,  $T_1$  et  $T_2$ , inoculés par une même communauté bactérienne sont clairement distincts les uns des autres. Mais les groupes  $T_1$  et  $T_2$  sont très proches et sont séparés du groupe  $T_0$ , par l'axe 1 qui est le plus discriminant quelle que soit la communauté inoculée dans les 6 sols (CS1, CS2, CS3, CS5 et CSL) (**Figure 9. A/ ; B/ ; C/ ; D/ et E/** respectivement). Les différences observées au niveau des ACP sont basées sur une différence significative de fluorescence du signal pour chaque sonde oligonucléotidique présente sur la lame. Les sondes présentant la plus forte variation au niveau de leur intensité de fluorescence entre les différents échantillons sont celles les plus éloignées du barycentre de l'ACP au niveau de l'axe 1, il a été possible de les isoler pour chaque communauté inoculée pour les six sols (**Figures 10, 11, 12, 13 et 14, A/**). A partir de ces sondes les groupes taxonomiques présentant la plus forte évolution au cours du temps ont été identifiés (**Figures 10, 11, 12, 13 et 14, B/**). L'analyse en composante principale des résultats d'hybridation sur puces à ADN des 6 sols inoculés par une même communauté et considérés au temps  $T_2=6$ mois met en évidence deux sols qui favorisent une évolution très marquée de la structure de la communauté bactérienne en comparaison des autres sols. En effet, les sols 7 et 9, se séparent des autres sols par l'axe 1 (qui est le plus discriminant) quand ils ont été inoculés par les communautés bactériennes CS1, CS2, CS3 et CS5, (**Figure 15. A/ ; B/ ; C/ ; D/ ; E/** respectivement). Pour les communautés bactériennes CS2, CS3, CS5 et CS7 inoculées dans les 6 sols, l'axe 2 sépare aussi les sols S7 et S9. L'axe 2 différencie également les sols S1, S2, S4, S5 pour les communautés CS1 et CS3. Pour CS2, CS5 et CL ces 4 autres sols restent groupés, et ont donc des structures plus proches les uns des autres. Le sol S5 pour la communauté CL n'est pas représenté car les résultats ne sont pas disponibles.



**Figure 9.** ACP interclassée réalisée à partir des résultats des puces à ADN des 6 sols inoculés par les communautés : A/CS1 ; B/CS2 ; C/CS3 ; D/CS5 ; E/CSL, groupés par temps : T0=1 jour ; T1=2 mois ; T2= 6mois

A/



B/

## CS1 CP1&lt;0

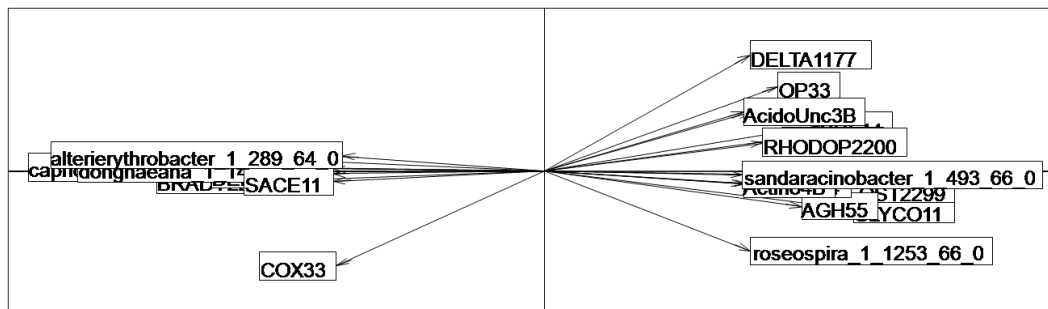
Phylum	Genre	Nombre
Acidobacteria	Edaphobacter	1
Bacteroidetes	Cytophaga	1
Cyanobacteria	Synechococcus	2
Proteobacteria	Burkholderia	47
	Dyella	16
	Escherichia	2
	Frateuria	4
	Gluconobacter	3
	Lebetimonas	1
	Luteibacter	2
	Methylobacillus	2
	Methylophilus	5
	Methylovorus	1
	Rhodanobacter	9
	Roseomonas	1
	Sulfurimonas	15
Xanthomonas	1	

## CS1 CP1&gt;0

Phylum	Genre	Nombre
Actinobacteria	Glycomyces	19
Firmicutes	Bacillus	21
	Oceanobacillus	26
	Ornithinibacillus	2
	Paucisalibacillus	1
	Pontibacillus	13
	Sediminibacillus	3
	Terribacillus	9
Virgibacillus	11	
Proteobacteria	Pseudomonas	2
	Sphingobium	11
	Sphingomonas	22

**Figure 10.** A/ Représentation des sondes expliquant le plus la distribution au niveau de l'axe 1 de l'ACP interclasse (Figure 9. A/), représentant la communauté CS1 inoculée dans les 6 sols stérilisés, au cours du temps (T0=1jour, T1=2mois et T2=6mois); B/ Identification taxonomique des sondes les plus discriminantes.

A/



B/

## CS2 CP1&lt;0

Phylum	Genre	Nombre
<b>Bacteroidetes</b>	Cytophaga	1
	Flavobacterium	1
	Persicivirga	1
<b>Proteobacteria</b>	Escherichia	2
	Novosphingobium	1
	Sphingomonas	1
	Sphingopyxis	1
	Sulfurospirillum	4

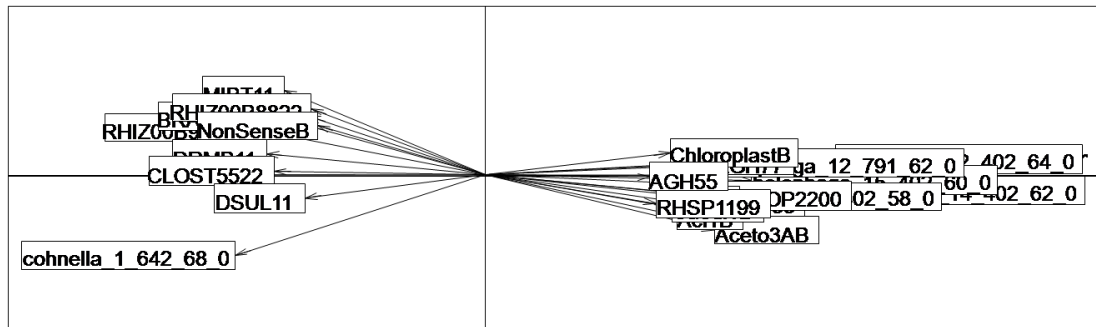
## CS2 CP1&gt;0

Phylum	Genre	Nombre
<b>Actinobacteria</b>	Glycomyces	19
<b>Chloroflexi</b>	Anaerolinea	1
<b>Firmicutes</b>	Bacillus	1
<b>Proteobacteria</b>	Desulfobotulus	2
	Sphingomonas	4

**Figure 11.** A/ Représentation des sondes expliquant le plus la distribution au niveau de l'axe 1 de l'ACP interclasse (Figure 9. B/), représentant la communauté CS2 inoculée dans les 6 sols stérilisés, au cours du temps (T0=1jour, T1=2mois et T2=6mois); B/ Identification taxonomique des sondes les plus discriminantes.



A/



B/

## CS3 CP1&lt;0

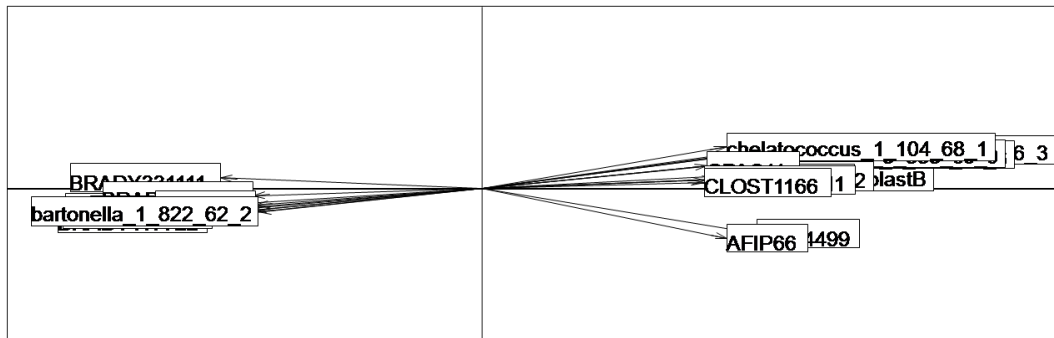
Phylum	Genre	Nombre
<b>Actinobacteria</b>	Acrocarpospora	5
	Arsenicococcus	2
	Brachybacterium	60
	Dermabacter	4
	Devriesea	1
	Herbidospora	4
	Microbacterium	1
	Microbispora	49
	Microtetraspora	29
	Planotetraspora	35
	Streptosporangium	3
	<b>Firmicutes</b>	Bacillus
Clostridium		1
Cohnella		11
Coprothermobacter		2
Paenibacillus		2
Sedimentibacter		12

## CS3 CP1&gt;0

Phylum	Genre	Nombre
<b>Chloroflexi</b>	Anaerolinea	1
<b>Proteobacteria</b>	Acidiphilium	2
	Roseococcus	1
	Roseomonas	4

**Figure 12.** A/ Représentation des sondes expliquant le plus la distribution au niveau de l'axe 1 de l'ACP interclasse (Figure 9. C/), représentant la communauté CS3 inoculée dans les 6 sols stérilisés, au cours du temps (T0=1jour, T1=2mois et T2=6mois); B/ Identification taxonomique des sondes les plus discriminantes.

A/



B/

## CS5 CP1&lt;0

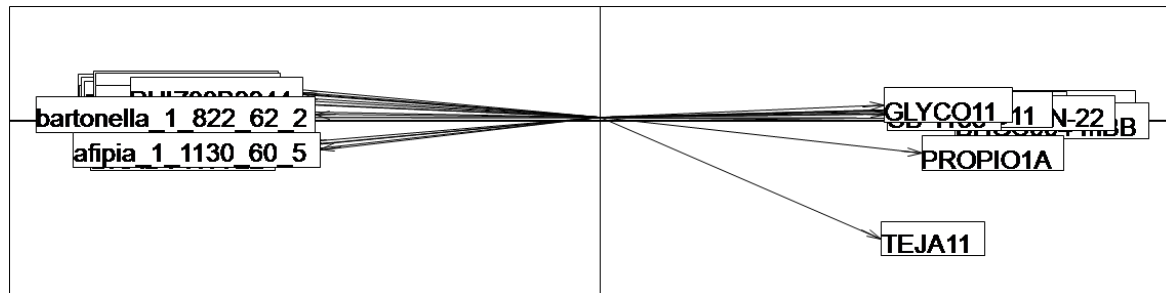
Phylum	Genre	Nombre
<b>Actinobacteria</b>	Arthrobacter	1
<b>Proteobacteria</b>	Burkholderia	23
	Mesorhizobium	1
	Pseudomonas	1
	Rhizobium	78
	Shinella	1

## CS5 CP1&gt;0

Phylum	Genre	Nombre
<b>Firmicutes</b>	Acetonema	2
	Clostridium	4
	Psychrosinus	1
	Sporomusa	8
	Sporotalea	1
	Staphylococcus	1
<b>Proteobacteria</b>	Afipia	1
	Brevundimonas	7
	Caulobacter	15
	Methylobacterium	6
	Sphingomonas	1
	Vibrio	1

**Figure 13.** A/ Représentation des sondes expliquant le plus la distribution au niveau de l'axe 1 de l'ACP interclasse (Figure 9. D/), représentant la communauté CS5 inoculée dans les 6 sols stérilisés, au cours du temps (T0=1jour, T1=2mois et T2=6mois); B/ Identification taxonomique des sondes les plus discriminantes.

A/



B/

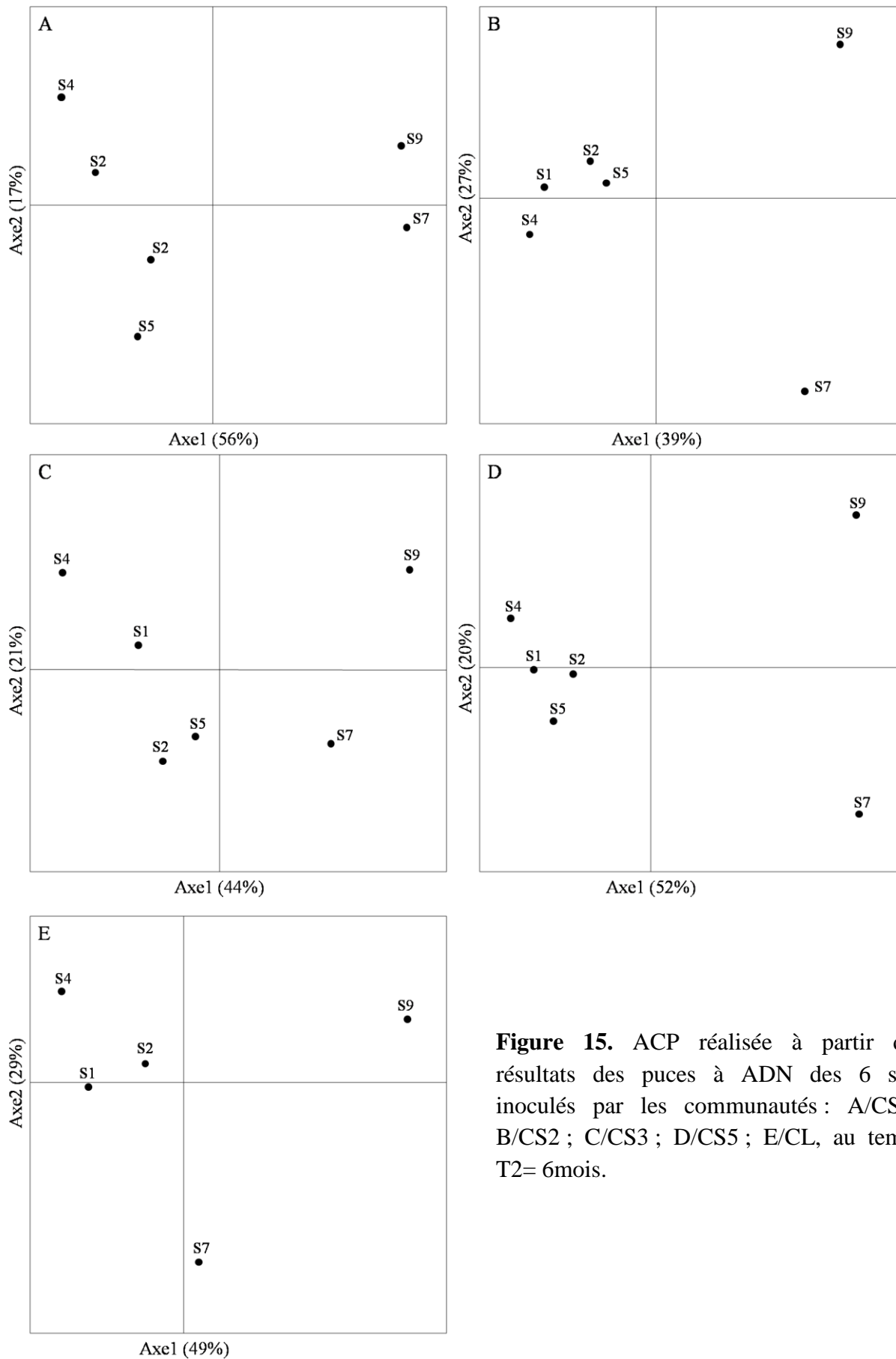
## CL CPI&lt;0

Phylum	Genre	Nombre
Proteobacteria	Brucella	6
	Burkholderia	23
	Pseudomonas	1
	Rhizobium	2
	Shinella	1
	Thermochromatium	1

## CL CPI&gt;0

Phylum	Genre	Nombre
Actinobacteria	Actinomadura	6
	Actinomyces	3
	Astrosporangium	1
	Catellatospora	1
	Cellulosimicrobium	6
	Curtobacterium	1
	Dietzia	1
	Glycomyces	19
	Gordonia	7
	Janibacter	25
	Jiangella	1
	Knoellia	3
	Nocardiosis	5
	Nonomurea	3
	Nostocoida type II	1
	Ornithinimicrobium	1
	Phycococcus	11
Propioniferax	1	
Pseudonocardia	2	
Rhodococcus	6	
Saccharomonospora	1	
Sanguibacter	5	
Streptomonospora	1	
Terrabacter	10	
Tetrasphaera	15	
Tsukamurella	1	
Firmicutes	Eubacterium	1

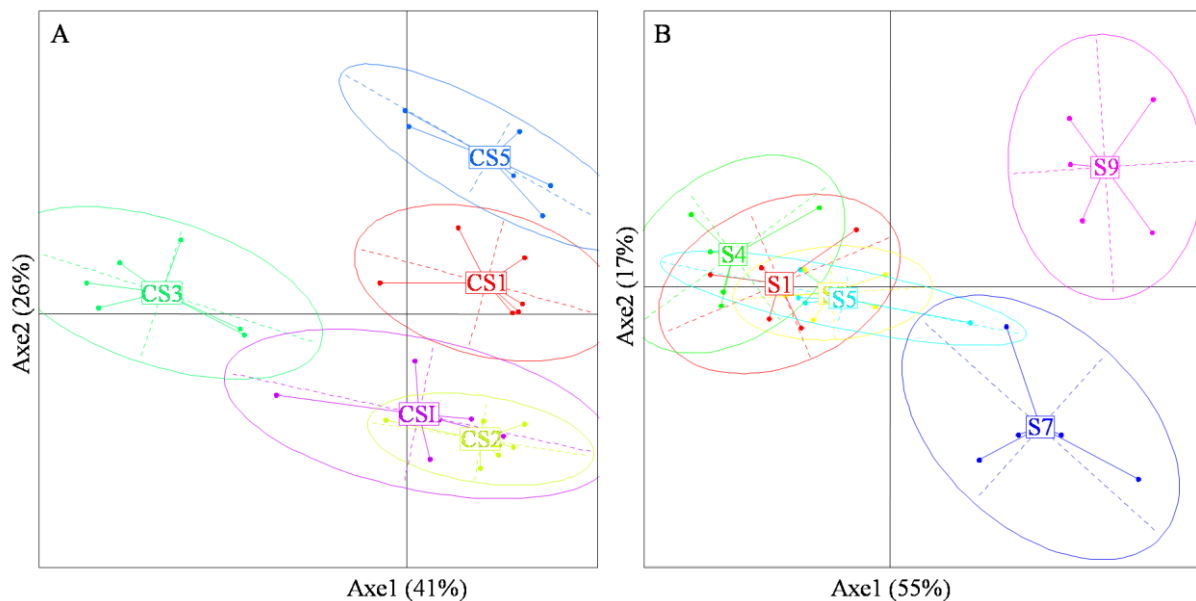
**Figure 14.** A/ Représentation des sondes expliquant le plus la distribution au niveau de l'axe 1 de l'ACP interclasse (Figure 9. E/), représentant la communauté CL inoculée dans les 6 sols stérilisés, au cours du temps (T0=1jour, T1=2mois et T2=6mois); B/ Identification taxonomique des sondes les plus discriminantes.



**Figure 15.** ACP réalisée à partir des résultats des puces à ADN des 6 sols inoculés par les communautés : A/CS1 ; B/CS2 ; C/CS3 ; D/CS5 ; E/CL, au temps T2= 6mois.

***Evaluation comparée des effets communautés inoculées et caractéristiques des sols sur la structure des communautés bactériennes dans les sols après incubation***

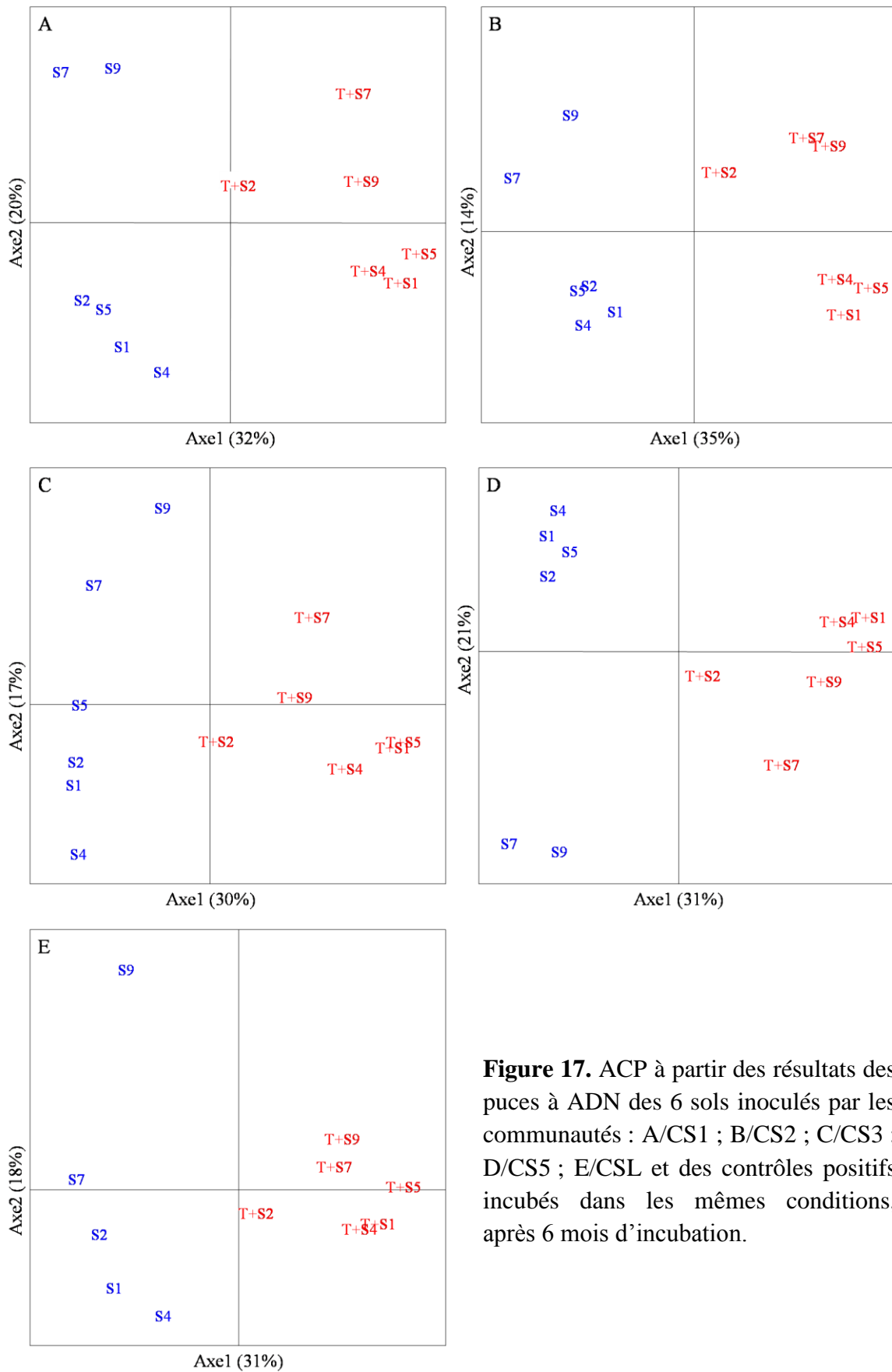
Une ACP interclasse a été réalisée à partir des résultats des puces à ADN, en groupant dans un premier temps les sols ayant été inoculés par la même communauté, incluant 5 groupes (pour 5 communautés inoculées) de 6 sols (S1, S2, S4, S5, S7 et S9) (Figure 16. A) Une autre ACP interclasse a été réalisée en groupant cette fois-ci chaque sol ayant reçu 5 communautés distinctes, incluant 6 groupes (car 6 sols inoculés) de 5 communautés (car 5 communautés bactériennes inoculées) (Figure 16. B). Dans la première configuration, les groupes CS1, CS3 et CS5 sont clairement distincts les uns des autres. Les groupes CL et CS2 ne sont pas différenciés entre eux mais distincts des autres groupes. L'axe 1 sépare clairement les groupes CS1, CS5, CS2 et CL du groupe CS3. L'axe 2 sépare les groupes CS5, CS1, CS2, CS2 et CL. Dans la deuxième configuration les groupes des sols S1, S2, S4 et S5 sont tous regroupés que ce soit au niveau de l'axe 1 que de l'axe 2. Au contraire les groupes des sols 7 et 9 sont clairement distincts des autres groupes au niveau de l'axe 1, de plus les groupes 7 et 9 sont distincts entre eux au niveau de l'axe 2.



**Figure 16.** **A/** ACP interclasse réalisée à partir des résultats des puces à ADN, représentant des groupes composés des 6 sols inoculés par la même communauté (5 groupes); **B/** ACP interclasse représentant des groupes composés d'un sol inoculé par les 5 communautés indépendamment (6 groupes).

***Comparaison des structures des communautés inoculées par rapport à celles des échantillons de sols « contrôles positifs »***

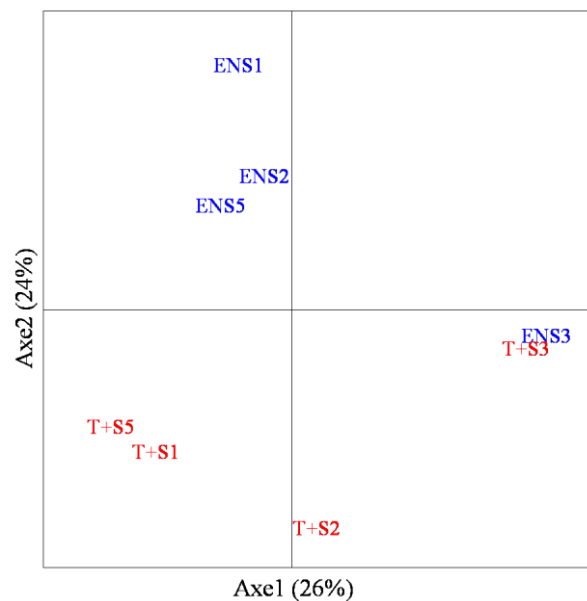
Une ACP a été réalisé à partir des résultats d'hybridation sur puces à ADN des produits PCR obtenus après amplification PCR de l'ADN des 6 sols inoculés par la même communauté extrait à T2=6mois et ceux des contrôles positifs, c'est-à-dire l'ADN les sols « bruts » incubés sur la même période (T2=6mois) et dans les mêmes conditions que les microcosmes inoculés. Les contrôles positifs sont clairement séparés des 6 sols inoculés par les communautés CS1, CS2, CS3, CS5 et CL (**Figure 17. A/; B/; C/; D/; E/** respectivement) au niveau de l'axe 1. De plus, l'ACP montre que les contrôles positifs sont groupés et distincts des sols inoculés, qui présentent plus de variabilités entre eux. Les sols 7 et 9 sont encore clairement isolés des autres, alors que dans les contrôles positifs, les sols 7 et 9 ne sont pas clairement éloignés des autres sols. Pour les échantillons inoculés, les sols 7 et 9 se séparent des autres sols inoculés par les communautés CS1 et CS5, le sol 9 étant isolé des autres sols pour les communautés CS3 et CSL, et le sol 7 distinct des autres sols pour la communauté CS2, au niveau de l'axe 2.



**Figure 17.** ACP à partir des résultats des puces à ADN des 6 sols inoculés par les communautés : A/CS1 ; B/CS2 ; C/CS3 ; D/CS5 ; E/CSL et des contrôles positifs incubés dans les mêmes conditions, après 6 mois d'incubation.

***Structure de la communauté bactérienne établie après extraction Nycodenz en comparaison de celle établie après extraction directe de l'ADN des échantillons de sols « contrôles positifs »***

Une ACP interclasse a été réalisée à partir des résultats obtenus après hybridation de l'ADN des communautés bactériennes extraites par Nycodenz des sols S1, S2, S3, S5 nommées ENS1, ENS2, ENS3, ENS5, et des contrôles positifs T+S1, T+S2, T+S3 et T+S5, qui sont les sols bruts incubés dans les mêmes conditions que les sols stérilisés (**Figure 18**). Seule la structure de la communauté « Nycodenz » du sol S3 est proche de celle de la communauté du sol contrôle T+S3. Les contrôles T+S1, T+S2, T+S5 présentent des structures se séparant significativement des communautés ENS1, ENS2 et ENS5. La communauté obtenue à partir d'eau de lac n'est pas intégrée dans cette analyse car elle est trop différente des autres et donc minimise les différences quand elle est intégrée dans l'analyse en composante principale. De plus le contrôle positif pour cette communauté avec laquelle pourrait être comparée la communauté « Nycodenz » n'a pas pu être réalisé.



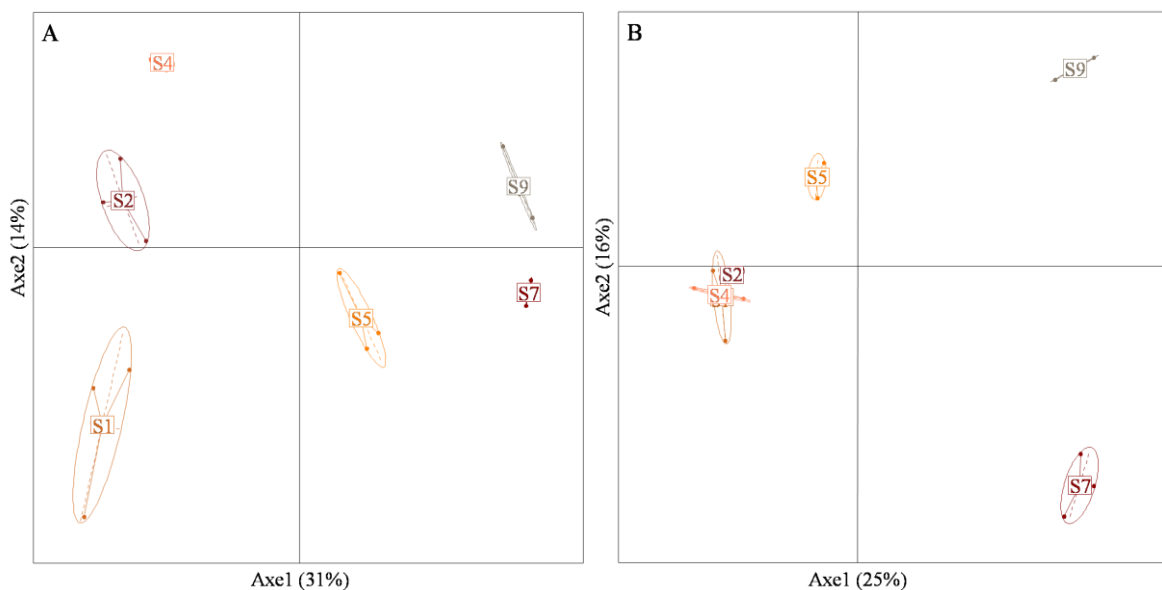
**Figure 18.** ACP réalisée à partir des résultats des puces à ADN des communautés bactériennes extraites à partir des Nycodenz du sol S1, S2, S3, S4, S5 nommées ENS1, ENS2, ENS3, ENS5 respectivement et des contrôles positifs (sols bruts incubés dans les mêmes conditions que les sols inoculés) T+S1, T+S2, T+S3, T+S5.



### 3.3. Reproductibilité des résultats

La reproductibilité de l'expérience a dans un premier temps été déterminée par la méthode RISA, qui a permis de valider l'approche avant d'analyser les échantillons d'ADN par la technique d'hybridation sur puces à ADN, qui présente une plus grande sensibilité, nécessaire pour l'analyse de communautés bactériennes complexes.

Application de la méthode RISA. Chaque microcosme contenant les différents sols stérilisés a été produit en 3 exemplaires. L'ADN a été extrait de chaque microcosme aléatoirement, puis amplifié séparément. Les ACP représentant les résultats RISA obtenus pour chaque répliques des 6 sols inoculés avec la communauté CS1=Côte-Saint-André au temps T1=2 mois (**Figure 19. A/**) et au temps T2=6mois (**Figure 19. B/**), montrent un bon regroupement des 3 répliques pour chacun des sols. Au temps T1 les groupes de répliques sont distincts les uns des autres au niveau de l'ACP, tandis qu'au temps T2 seuls les groupes des répliques des sols 5, 7 et 9 sont distincts.

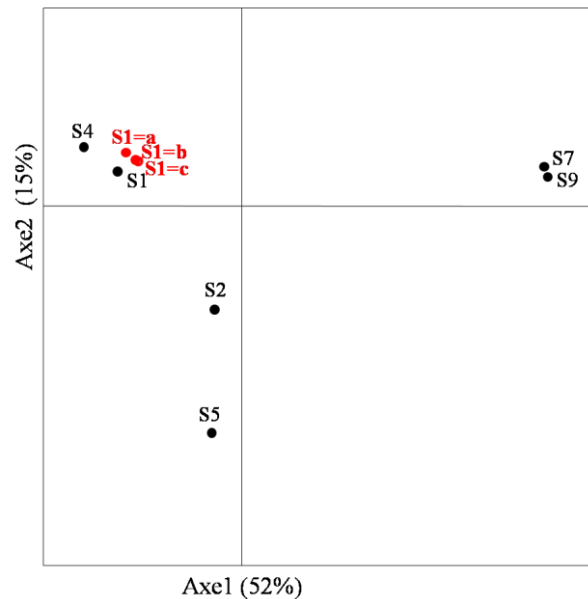


**Figure 19.** ACP à partir résultats RISA de chaque répliques des 6 sols inoculés par la communauté CS1 inoculée dans les 6 sols à : **A/** T1=2mois ; **B/** T2=6mois.

La reproductibilité a ensuite été déterminée par la technique des puces à ADN à partir des répliques du sol S1 inoculé par la communauté CS1 au temps T2=6mois : S1=a, S1=b, S1=c.

Une ACP a été réalisée avec les résultats des puces à ADN des 6 sols inoculés avec la communauté CS1=Côte-Saint-André au temps T2=6 mois (**Figure 20.**) plus ceux des répliques S1=a, S1=b, S1=c. Les répliques sont parfaitement regroupées. De plus les répliques

sont également très proches de S1 qui représente les résultats obtenus lorsque l'on a mis en commun les produits d'amplification 16S rDNA avant de les analyser par hybridation sur puces à ADN.



**Figure 20.** ACP à partir des résultats des 6 sols, inoculés par la communauté CS1, au temps T2=6mois. En rouge les répliques du sol 1 inoculées par CS1.

## 4. Discussion

### 4.1. Criblage des premiers résultats

L'analyse de la diversité bactérienne dans les sols a été abordée en combinant une approche conceptuelle novatrice basée sur la réinoculation de communautés bactériennes dans des environnements différents de ceux d'où elles ont été extraites et une approche technique de haute sensibilité. L'hybridation de produits PCR sur des lames de types puces à ADN taxonomiques constituait, à l'initiation de ce projet, la technique présentant le meilleur compromis entre sensibilité, capacité à traiter un nombre conséquent d'échantillons et coûts raisonnables. Cette technique a été couplée à une approche méthodologique (RISA) encore plus rapide pour la mise en œuvre et un coût modique mais de sensibilité moindre mais qui permettait de rapidement déterminer les modifications majeures affectant une communauté soumise à un nouvel environnement. L'étape initiale de ces approches et qui en constitue le facteur limitant est l'extraction de l'ADN bactérien à partir du sol et sa purification (Harry, 1999). Nos objectifs n'ont pas été de tenter d'optimiser l'efficacité de l'extraction de l'ADN mais d'utiliser des protocoles validés, certains commercialisés restant conscients toutefois que ces techniques présentent un certain nombre de limites qui vont entraîner des biais dans

l'évaluation de la diversité bactérienne. La purification de l'ADN extrait constitue un second défi à relever dont le résultat conditionne fortement l'emploi des outils de biologie moléculaire utilisés pour aborder les études d'inventaire de la biodiversité. Nos résultats montrent toute la difficulté que peut receler cette étape. C'est en effet un problème de purification de l'ADN extrait du sol qui limite l'amplification PCR de l'ADN extrait à partir de 3 échantillons de sol (sols 3, 6 et 8), que ce soit avec les amorces ciblant l'intergène 16S rDNA-23S rDNA ou avec celles ciblant une région interne au gène 16S rDNA.

L'analyse en composante principale (ACP) réalisée à partir des caractéristiques des 9 sols souligne une influence du sable pour séparer le sol 6 des autres sols. Le sol 6 présente en effet un taux élevé de sable (88%) mais ce paramètre peut difficilement être retenu comme responsable de l'inhibition de l'enzyme polymérase du fait qu'un autre sol (sol 7) qui affiche un taux de sable de 94.4% ne présente aucune limitation de l'amplification PCR de l'ADN extrait. Par contre l'ACP montre que ce sont le carbone organique, la matière organique et le rapport C/N, qui différencie le sol 3 des autres sols. Comme nous l'avons vu dans le chapitre bibliographique l'isolement de l'ADN bactérien du sol résulte en la co-extraction d'acides humiques, de polysaccharides ou même l'urée, dont les propriétés chimiques en particulier la solubilisation les rendent difficiles à séparer de l'ADN (Saano, 1995). Or ces composés, difficilement éliminés par les protocoles d'extraction et de purification sont connus pour fortement interférer avec la détection et la quantification de l'ADN (Zhou *et al.*, 1996) du fait de l'inhibition de la Taq DNA polymérase (Frostegard *et al.*, 1999, Robe, 2003). D'autres facteurs peuvent agir sur l'efficacité de la PCR, par exemple un pH acide inhibant fortement l'étape PCR (Burgmann *et al.*, 2001, Sagova-Mareckova *et al.*, 2008, Zhou *et al.*, 1996) ou bien le carbone organique du fait principalement de la forte concentration en substances humiques dans les sols très organiques (Sagova-Mareckova *et al.*, 2008; Zhou *et al.*, 1996). Le sol 3 se caractérise justement par un pH très acide (3.82), valeur la plus faible de tous les sols étudiés et le taux de carbone organique le plus élevé (59.4 g/kg), ce qui peut alors expliquer les problèmes d'amplification de l'ADN extrait de ce sol. Pourtant ce problème d'amplification pour le sol 3 ne se produit qu'avec les ADN extraits des sols inoculés par les 5 communautés. En effet, lorsque l'on extrait l'ADN à partir du sol brut S3, l'amplification de l'espace intergénique ou bien d'une région interne au gène 16S rDNA est possible avec respectivement un profil RISA complexe et une bande à la taille attendue.

L'incapacité à obtenir des produits d'amplification à partir des solutions extraites du sol 8 pourrait être liée au taux d'argile détecté dans ce sol (54,7%) qui est le plus élevé parmi les

sols étudiés. L'argile, de par sa capacité d'adsorption des cellules bactériennes (et de l'ADN) pourrait interférer négativement sur l'efficacité de l'extraction de l'ADN libéré dans le sol par les traitements de lyse. Cette hypothèse est confirmée par le fait que lorsque ce sol stérilisé est réensemencé par les communautés bactériennes des autres sols les quantités d'ADN réextraites demeurent extrêmement faibles, non détectables directement par visualisation sur gel d'agarose ni *a priori* après amplification PCR. Pour ce sol, les échecs des tentatives d'amplification seraient plus le fait du protocole d'extraction utilisé, dont le traitement lytique serait inefficace pour libérer l'ADN que de sa purification ultérieure.

Du fait des échecs de l'amplification PCR pour les sols 3, 6 et 8 soit par absence de molécules matrices (sol 8) ou par inhibition de l'enzyme polymérase (sols 3 et 6) les travaux ultérieurs ont été centrés sur 6 sols (1, 2, 4, 5, 7 et 9).

#### **4.2. Contrôles négatifs et positifs de l'expérience**

Seule la combinaison des deux méthodes de stérilisation (irradiation aux rayons gamma et autoclavage) permet d'obtenir un environnement tellurique totalement axénique comme le confirme l'absence de croissance de colonies bactériennes sur des milieux nutritifs solidifiés ensemencés avec des suspensions après dilutions sérielles des différents sols soumis à stérilisation par l'une, l'autre ou la combinaison des deux méthodes.

Afin de provoquer la mort cellulaire ce double traitement de stérilisation affecte la qualité de l'ADN extracellulaire initialement présent dans le sol ou libéré par la lyse cellulaire. Ces effets s'ajoutent à ceux qui affectent naturellement dans le sol l'ADN extracellulaire soumis à l'action enzymatique (nucléases), chimique et physique notamment par le déplacement des composants du sol et l'adsorption sur les surfaces minérales ou humiques (Nielsen *et al.*, 2007). L'exposition de l'ADN aux hautes températures pendant l'autoclavage va se traduire par la dénaturation de l'ADN et sa fragmentation que doit conforter l'irradiation des microcosmes par les rayons gamma (Nielsen *et al.*, 2007). Si l'objectif premier du double traitement de stérilisation était de rendre axénique l'échantillon de sol afin de permettre à une nouvelle communauté bactérienne de s'y implanter la dégradation de l'ADN qui résulte de ces traitements devait permettre de limiter, sinon totalement éliminer l'ADN résiduel susceptible de contaminer la nouvelle solution d'ADN extraite après développement de la communauté réensemencée dans ces échantillons de sol stérilisés. De plus la quasi-complète dégradation de l'ADN initial aboutissant à son inactivation biologique limite le risque de transfert horizontal

par transformation qui pourrait concerner les populations introduites et qui, du fait de leur multiplication intensive seraient susceptibles de développer un stade de compétence.

L'approche RISA montre cependant que tout l'ADN n'est pas totalement et immédiatement dégradé dans tous les sols par le double traitement puisque deux fragments sont visualisés à partir de l'ADN extrait des échantillons de sol stérilisés et non réensemencés. Par contre, cet ADN résiduel n'est pas mis en évidence dans la plupart des sols (échantillons T-S1, T-S2, T-S3, T-S5, T-S7) à T=0 et ces résultats demeurent négatifs après incubation de 2 et 6 mois, confirmant ainsi la stérilité des échantillons et notamment la mort de bactéries non cultivables qui auraient pu échapper à la détection par la technique d'étalement sur milieu nutritif. D'autre part, pour les échantillons dans lesquels l'ADN résiduel était détectable à T=0, l'incubation à 20°C et sous 50% d'humidité résulte dans la dégradation graduelle de l'ADN résiduel dont les 2 bandes RISA initialement détectées ne le sont plus même après la plus courte durée d'incubation.

Une question prévaut cependant sur l'impact que pourrait avoir l'ADN résiduel sur les résultats d'analyse des nouvelles communautés bactériennes dans les échantillons de sol stérilisés réensemencés. En fait, les deux fragments RISA identifiés dans quelques échantillons contrôles n'ont jamais été détectés dans les sols stérilisés après inoculation (**Figure 8**), la très faible concentration de l'ADN résiduel ne devant pas impacter les profils RISA qui représentent les populations bactériennes les plus fortement représentées en termes d'effectifs des communautés bactériennes en développement.

Enfin, la dernière série de microcosmes contrôles a été constituée d'échantillons des différents sols naturels qui n'ont pas été soumis aux traitements de stérilisation, qui ont été inoculés avec de l'eau stérile et incubés dans les mêmes conditions que les autres microcosmes. Les profils RISA à T2=6 mois des 6 sols étudiés montrent des profils complexes, très différents les uns des autres confirmant l'intérêt de ces sols aux caractéristiques physico-chimiques marquées pour aborder les objectifs de ces études qui sont de tenter de favoriser le développement des populations minoritaires en leur offrant un nouvel environnement afin de pouvoir les détecter.

### **4.3. Reproductibilité de l'expérience**

Dans un premier temps, la reproductibilité de l'expérience a été testée par la méthode RISA. Les amplifications de l'espace intergénique obtenues à partir de chaque réplique des échantillons d'ADN provenant de trois microcosmes indépendants pour chacun des 6 sols

inoculés par la communauté CS1, sont tous regroupés au niveau de l'ACP (**Figure 19**). Ce résultat valide plusieurs niveaux de l'expérience : i) la reproductibilité de l'extraction de l'ADN : résultats similaires obtenus à partir de 3 microcosmes répliqués ; ii) l'efficacité de l'amplification de l'ADN par PCR : résultats positifs et similaires obtenus à partir de répliques indépendantes) ; iii) Potentiel de la technique RISA. Ces résultats démontrent la sensibilité et la reproductibilité de cette méthode, donc son efficacité pour détecter des différences entre communautés bactériennes complexes extraites de différents sols (**Figure 19**).

La reproductibilité de l'expérience a aussi été validée par la technologie d'hybridation sur les puces à ADN pour les échantillons d'ADN extraits du sol S1 inoculé par la communauté CS1 au temps T2=6mois. Ce résultat a permis de valider les conditions expérimentales à plusieurs niveaux : i) l'extraction d'ADN (à partir de 3 microcosmes répliqués) ; ii) l'amplification du gène 16S rDNA, l'ADN (à partir des répliques indépendamment) ; iii) la technique des puces à ADN. Les répliques sont regroupées au niveau de l'ACP (**Figure 20**), elles sont également très proches de l'échantillon S1 inoculé par CS1 qui représente les produits PCR des répliques mis en commun, donc avec une structure de communauté similaire. Ce résultat valide ainsi le choix de regrouper les produits PCR des répliques pour chaque échantillon avant l'analyse par puces à ADN afin de limiter la perte d'informations.

#### **4.4. Evaluation des structures des communautés bactériennes**

##### ***Structure de la communauté bactérienne établie après extraction Nycodenz en comparaison de celle établie après extraction directe de l'ADN des échantillons de sols « contrôles positifs »***

Deux approches techniques ont été utilisées pour extraire l'ADN, la méthode de lyse directe *in situ* appliquée sur les échantillons inoculés et celle indirecte passant par l'extraction cellulaire puis la lyse bactérienne en parallèle à l'utilisation de ces suspensions comme inocula. La comparaison des résultats d'analyse de la diversité à T=0 (1 jour) permet de renseigner sur les biais liés à ces approches méthodologiques ou plus conceptuellement sur le changement de conditions environnementales que produit l'incubation du sol à températures et taux d'humidités déterminés. En fait, seule la communauté du sol 3 présente une certaine cohérence quelle que soit l'approche choisie, extraction indirecte (méthode Nycodenz) ou directe (contrôle T+S3).

Les communautés issues des sols contrôles T+ S1, T+S2 et T+S5 se différencient au niveau de l'axe 1 de l'ACP, des communautés établies d'après l'ADN extrait des échantillons ENS1, ENS2, et ENS5. L'hypothèse conceptuelle pour expliquer ce résultat est donc une évolution importante de la communauté de ces sols quand ceux-ci ont été placés dans d'autres conditions expérimentales que celles qui prévalaient auparavant.

### ***Evolution des structures des communautés bactériennes inoculées dans les sols stérilisés au cours du temps***

La première étape de cette analyse a été de déterminer l'évolution populationnelle successive affectant les communautés bactériennes pendant leur incubation dans des échantillons de sols stérilisés sur une durée allant jusqu'à 6 mois. Déjà, l'analyse RISA des solutions d'ADN immédiatement ré-extraites (T0=1jour) des 6 sols ensemencés par une même communauté bactérienne (extraite du même sol) montre des profils différents ce qui ne peut s'expliquer que par des biais d'extraction de l'ADN liés aux caractéristiques physicochimiques différentes des sols considérant que les communautés n'ont pu suffisamment évoluer en un temps si court pour aboutir à des modifications populationnelles détectables. Ces résultats confirment la difficulté à évaluer la structure populationnelle des communautés bactériennes des sols y compris par les approches basées sur l'étude de l'ADN bactérien qui, comme les approches culturales, sont limitées par des biais importants quoique différents.

Ces techniques permettent cependant de montrer que toutes les communautés bactériennes étudiées (CS1, CS2, CS3, CS5 et CL) subissent une adaptation graduelle aux nouveaux environnements que constituent les différents échantillons de sols stérilisés. Ces résultats peuvent être interprétés par le passage de la structure des communautés par une phase « d'adaptation » plus ou moins précoce et plus ou moins différenciée de la phase de stabilisation et qui peut être détectée à T1=2 mois pour tous les échantillons de sols. L'application de la technique d'hybridation sur puces à ADN taxonomiques dont les résultats sont analysés en ACP interclasse (**Figure 9**) permet d'identifier les taxons bactériens spécifiquement impliqués dans la phase temporaire d'adaptation de ceux responsables de la structuration plus aboutie des communautés bactériennes (T1= 2 mois et surtout T2= 6 mois). Il est à noter que les différents sols inoculés par une même communauté bactérienne se différencient systématiquement par les taxons bactériens impliqués dans les phases d'adaptation puis d'équilibre. Ce même résultat a été obtenu pour chacune des communautés bactériennes inoculées indiquant le fort potentiel structurant de la matrice tellurique. Très

clairement, ces résultats montrent la prédominance de l'effet matrice abiotique liée aux caractéristiques physico-chimiques du milieu dans l'évolution des communautés bactériennes ensemencées dans ce type de nouvel environnement que représente un échantillon de sol stérilisé.

Après 6 mois d'incubation (T2), les communautés développées dans chacun des 6 sols à partir d'un même inoculum présentent une structure de communautés spécifique mais avec cependant des différences entre les sols. Deux sols favorisent en effet des structures de communauté bactérienne nettement différentes des autres d'après les ACP obtenues pour chaque communauté. Ces deux sols (S7 et S9) sont non seulement intéressants par le fait qu'ils permettent un changement drastique de la communauté bactérienne initiale, mais aussi du fait que les deux structures populationnelles sont également très différentes l'une de l'autre. Cette structuration différenciée pourrait s'expliquer par les caractéristiques granulométriques de ces sols, le sol 7 (Nouvelle Calédonie) ayant le taux le plus élevé de sable (94.4%), le sol 9 (France : Thalmont Saint Hilaire) celui de limon (73%), l'ACP réalisée à partir des caractéristiques des 6 sols confirmant que ce sont ces deux paramètres qui différencient ces deux sols des autres sols.

Une question importante se pose quand on inocule une communauté bactérienne dans des sols stérilisés ayant des caractéristiques différentes : après 6 mois d'incubation observe-t-on plutôt un effet communauté inoculée ou bien un effet sol qui va orienter la structure des communautés bactériennes dans ces sols? Lorsque les 6 sols sont regroupés en fonction de la communauté par laquelle ils ont été inoculés, les groupes sont alors clairement séparés les uns des autres (**Figure 16. A/**). Il y a donc bien un effet communauté inoculée significatif sur la structure populationnelles dans les 6 sols, c'est-à-dire que les structures des communautés bactériennes des 6 sols inoculés par la même communauté de départ vont être plus similaires entre elles que les structures des communautés des 6 sols inoculés par une autre communauté de départ.

Lorsque les groupes sont formés par un sol ayant reçu les 5 communautés bactériennes indépendamment (**Figure 16. B/**), alors seuls les groupes des sols 7 et 9 sont distincts des autres groupes de sols et d'eux même, ce qui signifie que ces deux sols ont un poids important au niveau de la structure des communautés bactériennes inoculées. L'effet sol est moins important que l'effet communauté inoculée au niveau de la structure populationnelle, car seulement 2 groupes sur 6 sont distincts, mais il a un impact évident sur les communautés inoculées pour les sols 7 et 9. Les structures des 5 communautés inoculées dans les sols 7 et 9



ont tendance à évoluer de manière différente sous l'effet modulant des caractéristiques des deux sols au temps T2= 6 mois. L'ACP interclasse n'a pas pu mettre en évidence de différences de structure entre les groupes S1, S2, S4 et S5.

### ***Comparaison des structures des communautés inoculées par rapport à celles des échantillons de sols « contrôles positifs »***

Les résultats précédents indiquent clairement un effet communauté inoculée sur l'évolution de la structure des populations dans les sols. Une question fondamentale est alors de savoir si une communauté extraite d'un sol donné pour être inoculée dans un autre sol stérilisé allait effectivement se développer vers une structure similaire à celle originellement présente dans ce second sol, en accord alors avec l'hypothèse structurante de l'environnement abiotique. Cette hypothèse n'est cependant pas confirmée expérimentalement. Ce résultat pourrait être lié au fait qu'un échantillon de sol stérilisé est loin de produire des conditions similaires à celles qui ont prévalu à l'installation de la communauté originelle. La stérilisation, en particulier par autoclavage est connue pour déstructurer la matrice tellurique et donc modifier les caractéristiques physico-chimiques en comparaison du sol brut prélevé sur le terrain. L'environnement abiotique élimine aussi les effets liés à la prédation et la compétition entre les populations de l'inoculum s'établira sur des bases totalement différentes de celles existant dans les conditions naturelles. L'évolution séquentielle de la communauté bactérienne pendant l'incubation avec au moins 2 phases comme on l'a vu précédemment doit être analysée comme résultant de processus artificiels ne simulant que très imparfaitement l'environnement naturel. On peut malgré tout conclure que l'environnement abiotique est un facteur structurant la communauté bactérienne dans les conditions expérimentales utilisées dans nos études, qui va avoir un impact différent suivant les communautés inoculées. Cette hypothèse est confirmée par la détection pour chaque type de sols de communautés avec des structures différentes quand ces microcosmes stérilisés sont inoculés et incubés avec diverses communautés. Si l'environnement abiotique était le seul facteur structurant, un même sol devrait favoriser le développement des populations de façon à aboutir à terme à une structure de communauté bactérienne identique quel que soit l'inoculum utilisé. Nos résultats montrent que ce n'est pas le cas, même s'il n'est pas possible de conclure que la communauté ait atteint un état définitivement stable après 6 mois d'incubation. Il est intéressant aussi de constater que les communautés des sols naturels (témoins positifs) se différencient des sols inoculés quelles que soient les combinaisons prises en compte et l'ACP montre même plus de

différences entre les communautés des différents sols inoculés par un même inoculum qu'entre communautés naturelles (provenant des sols non traités et incubés dans les mêmes conditions que les microcosmes inoculés). Les sols 7 et 9 inoculés par les 5 communautés sont encore clairement isolés des autres sols, confirmant leur spécificité et leur pouvoir structurant fort en comparaison des autres sols même si leurs communautés naturelles (contrôles positifs) ne se distinguent pas particulièrement de celles provenant des autres sols.

## 5. Conclusions

Ces premiers résultats sur l'évolution des communautés bactériennes inoculées dans des sols stérilisés valident dans un premier temps notre approche conceptuelle, et montrent l'effet des propriétés des sols sur la structure des communautés bactériennes inoculées dans des sols stérilisés. Pourtant ces résultats soulèvent un point intéressant, l'approche conceptuelle aurait pu être basée sur d'autres facteurs engendrant des changements de structure d'une communauté d'un même sol, comme par exemple l'ajout de nutriments, des variations de température, etc...Mais comme nous l'avons vu dans la synthèse bibliographique les caractéristiques physico chimiques des sols sont les facteurs les plus importants dans le modelage de la structure des communautés microbiennes (Girvan *et al.*, 2003 , Singh, 2006). La technique des puces à ADN nous a permis dans cette partie d'évaluer les structures des communautés bactériennes dans des les 6 sols stérilisés au cours du temps, mais cette technique permet également d'identifier les bactéries correspondant aux sondes allumées, ce qui sera développé dans la deuxième partie.



## PARTIE II : Modifications de communautés bactériennes induites par des changements de sols

### 1. Introduction

Les travaux réalisés précédemment visaient à déterminer l'impact des paramètres physico-chimiques des sols sur les structures des communautés bactériennes telles que celles-ci pouvaient être déterminées par des approches techniques de profilage basées sur le gène ribosomique 16S rDNA. Pour atteindre ces objectifs les cellules bactériennes étaient extraites des sols par la technique de séparation sur gradient de Nycodenz, inoculées dans des microcosmes de sol stérilisé qui étaient ensuite incubés pendant des périodes de temps allant jusqu'à 6 mois pour permettre un équilibrage de la structure de la communauté. Les résultats acquis par analyse RISA après PCR de l'ADN extrait de ces échantillons de sol ont clairement confirmé que les caractéristiques physico-chimiques telluriques interviennent fortement dans la structuration de la communauté bactérienne, même si d'autres facteurs influent également. Deux sols, provenant l'un de Nouvelle Calédonie l'autre de Talmont Saint Hilaire (France) ont un impact particulièrement important sur la microflore tellurique comme l'attestent les différences très marquées notées entre les profils initiaux de chacune des communautés en comparaison de ceux obtenus après extraction de l'ADN des sols des microcosmes. Nos travaux montrent également que l'adaptation à un nouvel environnement se fait graduellement, les profils RISA observés indiquant systématiquement une structuration de la communauté différente après des temps d'incubation s'étageant entre 1 et 18 mois. Ces résultats peuvent être interprétés comme le reflet des différences adaptatives des bactéries, certaines se développant rapidement dans un écosystème dépourvu de compétition avant que d'autres, à croissance plus lente et/ou dont la fitness s'exprime plus tardivement ne les supplantent.

Ces résultats doivent être interprétés comme la première étape de nos objectifs qui visent à appréhender les populations minoritaires des sols afin de définir à terme le niveau de spécificité caractérisant chaque écosystème tant au niveau des populations que des fonctions potentiellement réalisables. Ces objectifs ont requis le développement ou l'application d'une double approche conceptuelle et méthodologique. D'un point de vue méthodologique nous avons montré tout l'intérêt pour la caractérisation de la structure des communautés

bactériennes, de la conjonction de la technologie RISA rapide et facile d'emploi et de la technique d'hybridation de produits PCR contre des oligosondes définies dans le gène 16S rDNA. Les sondes définies peuvent être directement synthétisées sur lames de type puces à ADN résultant en une technologie qui présente une bonne sensibilité et qui peut être appliquée à de nombreux échantillons à des coûts raisonnables. C'est toutefois conceptuellement par la ré-inoculation de communautés bactériennes dans des échantillons de sols stérilisés que nos travaux trouvent leur originalité et leur principal intérêt. Cette approche s'est systématiquement traduite par la mise en évidence après incubation de structures de communautés bactériennes divergeant significativement des structures originales laissant envisager que des populations initialement indétectables se sont suffisamment développées dans leur nouvel environnement pour pouvoir être comptabilisées. L'analyse dynamique du processus de colonisation a de plus permis d'identifier plusieurs étapes caractérisées par une succession de populations majoritaires accroissant ainsi la possibilité de détecter des taxons bactériens dont le niveau populationnel n'est élevé que dans des conditions biotiques et abiotiques très restreintes. Le fait enfin que les différents types de sol ré-inoculés par une même communauté bactérienne révèlent une structure spécifique (indiquant ainsi le rôle structurant de l'inoculum en complément des facteurs abiotiques) multiplie la possibilité de détecter des taxons initialement indétectables mais effectivement présents dans le sol original.

Les objectifs recherchés dans cette seconde partie du chapitre II sont d'identifier les taxons bactériens que l'approche conceptuelle de ré-inoculation permet de révéler en vue d'accroître l'inventaire des populations initialement présentes dans le sol original et que son analyse directe n'avait pas permis de détecter. Cet objectif sera réalisable grâce à la technologie de l'hybridation sur puces à ADN qui permet d'identifier les sondes responsables de la topologie de l'ACP et conséquemment les taxons bactériens produisant la nouvelle structure de la communauté bactérienne. Les résultats escomptés visent à confirmer ou infirmer l'hypothèse à la base de notre travail « Is everything everywhere ? »

## **2. Matériels & Méthodes**

### **2.1. Choix des échantillons d'ADN**

Dans cette partie nous nous intéresserons aux échantillons d'ADN extraits des sols S1, S2, S4, S5, S7, inoculés avec les communautés bactériennes CS1, CS2, CS3, CS5 et CL, aux temps T0=1jour, T1=2mois et T2=6mois, analysés par la technique des puces à ADN. Les solutions

d'ADN extraites à partir des suspensions bactériennes obtenues sur gradients de Nycodenz : ENS1, ENS2, ENS3, ENS5 (suspensions qui ont ensuite été inoculées dans les échantillons de sols stérilisés). Les protocoles d'inoculation, d'extraction et des puces à ADN sont décrits dans le Matériels et Méthodes de la PARTIE I : Evolution des communautés bactériennes inoculées dans des sols stérilisés.

## **2.2. Traitement des données obtenues par les puces à ADN**

Les puces ont été scannées à l'aide du scanner GenePix (Bucher Biotec, Viaduktstrasse, Switzerland) avec une taille de pixel de 5µm. La fluorescence a été transformée en image à l'aide du logiciel GenePix Pro, qui réduit les données à une valeur de signal pour chaque sonde. Les signaux de fluorescence des hybridations brutes ont été calculés en faisant la moyenne de la fluorescence de chaque pixel pour une sonde donnée. La valeur de bruit de fond a été calculée suivant la formule suivante : (moyenne des répliques du témoin négatif) + (3 x écart type du témoin négatif), formule standard IUPAC. La moyenne des différentes répliques de chaque sonde a été calculée en enlevant les valeurs aberrantes lorsque la valeur était supérieure à 2 écarts type de leur moyenne. Le bruit de fond a alors été soustrait à la moyenne corrigée de chaque sonde. Les résultats ont alors été normalisés par la méthode des quantiles. Les sondes dont le signal présente une valeur inférieure à zéro sont considérées comme n'ayant pas produit une hybridation significativement positive, alors que c'est l'inverse (hybridation positive) pour celles dont le signal est supérieur à zéro.

## **2.3. Identification taxonomique des sondes oligonucléotidiques**

Chaque sonde constituée de 20 oligonucléotides a été confrontée *in silico* aux données de la base de données NCBI en conditions d'acceptation de plus ou moins deux mésappariements (mismatches) et « gaps ». Dans ces conditions, chaque sonde présente une complémentarité avec plusieurs Hits correspondant à différentes OTU. La puce est constituée de telle sorte qu'il y ait un recouvrement au niveau des différents Hits identifiés par chaque sonde, c'est à dire que plusieurs sondes identifient le même Hit. Pour qu'un Hit soit déclaré positif après hybridation, c'est-à-dire l'OTU correspondant présent dans l'échantillon de sol étudié, il faut que toutes les sondes identifiant ce Hit présentent un signal positif. Un programme sous Matlab a permis d'obtenir pour chaque échantillon tous les Hits potentiellement présents.

### 3. Résultats & Discussion

#### 3.1. Identification des taxons bactériens détectés dans les sols inoculés par chacune des communautés par rapport aux inocula de départ.

L'analyse taxonomique obtenue par la technique d'hybridation des produits PCR 16S rDNA sur puces à ADN a été réalisée pour chaque communauté inoculée CS1, CS2, CS3, CS5 et CL. La puce taxonomique peut cibler potentiellement jusqu'à 32 phylums différents (**Tableau 1**). A l'aide du programme Matlab, il a été possible d'identifier tous les taxons potentiellement détectés pour chaque échantillon. Lorsque l'on compare les phylums détectés dans l'inoculum de départ et ceux détectés dans les six sols inoculés par ce dernier après six mois d'incubation, aucun nouveau phylum n'a pu être mis en évidence par notre approche d'inoculation pour les communautés CS1 et CS5 (**Tableaux 2 & 6**) par rapport aux inoculas références : ENS1 et ENS5. Par contre pour les communautés CS2, CS3 et CS7, des nouveaux phylums (en bleu) ont pu être détectés dans un ou plusieurs des six sols inoculés par rapport aux inoculas de départ ENS2, ENS3 et EDL (**Tableaux 3, 4 & 6**), respectivement 3, 1 et 12 phylums. L'augmentation de diversité à ce niveau taxonomique élevé est très intéressante mais ne s'applique pas à toutes les communautés inoculées étudiées. Notre étude s'est alors focalisée sur un niveau taxonomique inférieur, celui des genres qui est plus informatif car 1477 genres sont potentiellement ciblés par les différentes sondes nucléotidiques présentes au niveau de la puce taxonomique à ADN.

La première étape a été d'identifier les taxons bactériens non détectés dans les 5 inocula de départ soit ENS1, ENS2, ENS3, ENS5 et EDL parmi les taxons putatifs représentés par des oligosondes sur la puce à ADN ce qui a permis dans un second temps d'identifier parmi ces derniers, ceux détectables dans 1 ou plusieurs des 6 sols ré-inoculés indépendamment par chacune des 5 communautés bactériennes.

Ce type de résultats est présenté dans les tableaux avec comme exemple les genres identifiés comme absents dans l'inoculum ENS1 mais présents au temps T2=6mois, dans le sol 1 (**Tableau 7**), dans le sol 2 (**Tableau 8**), dans le sol 4 (**Tableau 9**), dans le sol 5 (**Tableau 10**), dans le sol 7 (**Tableau 11**) et dans le sol 9 (**Tableau 12**), tous inoculés par la même communauté ENS1 (CS1=communauté du sol 1). La même analyse a été réalisée pour chacune des communautés inoculées : CS1 comme présentée ci-dessous mais aussi CS2, CS3, CS5 et CL, en prenant comme référence ENS1, ENS2, ENS3, ENS5 et EDL respectivement. Les genres seulement détectés dans chacun des 6 sols inoculés par la même communauté ont pu être mis en évidence, ils sont représentés par la couleur rouge au niveau des tableaux. La

prise en compte des différents sols permet ainsi de définir une liste de taxons bactériens qui n'avaient pu être mis en évidence dans la communauté initiale de ce sol pour les raisons évoquées précédemment, notamment d'un niveau populationnel trop faible. L'approche de ré-inoculation dans un nouvel environnement a permis de détecter ces taxons démontrant la présence de ces bactéries correspondantes dans l'inoculum.

**Tableau 1.** Nombre de Hits représentant les différents phylums potentiels ciblés par les différentes sondes oligonucléotidiques présentes sur les puces à ADN.

Phylum	Puce
Acidobacteria	45
Actinobacteria	3637
Aquificae	12
Bacteroidetes	1715
Candidatus Poribacteria	1
Chlamydiae	72
Chlorobi	28
Chloroflexi	81
Chrysiogenetes	1
Crenarchaeota	92
Cyanobacteria	1588
Deferribacteres	14
Deinococcus-Thermus	165
Dictyoglomi	7
Elusimicrobia	1
Euryarchaeota	380
Fibrobacteres	6
Firmicutes	5205
Fusobacteria	67
Gemmatimonadetes	5
Korarchaeota	6
Lentisphaerae	9
Nitrospirae	14
Planctomycetes	170
Platyhelminthes	1
Proteobacteria	13701
Spirochaetes	319
Synergistetes	4
Tenericutes	363
Thermodesulfobacteria	19
Thermotogae	73
Verrucomicrobia	51



**Tableau 2.** Nombre de Hits correspondant aux phylums potentiellement détectés dans chacun des 6 sols inoculés par la communauté CS1, après 6 mois d'incubation. ENS1 correspond à l'inoculum de départ.

Phylum	S1_T2	S2_T2	S4_T2	S5_T2	S7_T2	S9_T2	ENS1
Acidobacteria	14	13	12	10	12	19	10
Actinobacteria	808	1018	545	1095	838	569	1108
Aquificae	2	1	1	1	2	1	2
Bacteroidetes	117	112	78	92	111	77	171
Candidatus-Poribacteria	1	1	1	1	1	1	1
Chlorobi	6	6	6	6	6	6	6
Chloroflexi	8	9	8	9	11	8	14
Crenarchaeota	20	20	7	20	20	7	90
Cyanobacteria	81	96	78	82	63	46	337
Deferribacteres	2	2	2	2	2	1	1
Deinococcus-Thermus	4	28	0	24	24	24	34
Euryarchaeota	21	21	21	21	21	3	120
Fibrobacteres	0	0	0	0	0	0	2
Firmicutes	656	608	483	499	612	457	854
Fusobacteria	1	25	1	25	25	25	3
Gemmatimonadetes	1	1	1	1	0	1	1
Nitrospirae	2	2	3	3	0	0	4
Planctomycetes	30	32	30	31	30	9	43
Proteobacteria	3627	3383	2363	2914	3213	2394	2805
Spirochaetes	112	112	110	112	115	115	199
Synergistetes	2	2	2	2	2	2	3
Tenericutes	48	1	28	21	33	19	37
Thermodesulfobacteria	17	0	17	0	0	0	17
Thermotogae	15	15	15	0	3	15	41
Verrucomicrobia	16	17	16	16	8	7	16

**Tableau 3.** Nombre de Hits correspondant aux phylums potentiellement détectés dans chacun des 6 sols inoculés par la communauté CS2, après 6 mois d'incubation. ENS2 correspond à l'inoculum de départ.

Phylum	S1_T2	S2_T2	S4_T2	S5_T2	S7_T2	S9_T2	ENS2
Acidobacteria	12	16	12	14	9	9	7
Actinobacteria	1109	1131	700	760	698	478	422
Aquificae	1	1	1	1	1	1	1
Bacteroidetes	134	134	100	80	98	60	49
Candidatus Poribacteria	1	1	1	0	1	1	1
Chlamydiae	0	0	0	0	0	0	1
Chlorobi	6	6	7	6	7	6	6
Chloroflexi	13	9	8	5	6	6	10
Crenarchaeota	20	20	7	20	7	7	7
Cyanobacteria	151	162	46	43	92	40	43
<b>Deferribacteres</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>5</b>	<b>0</b>	<b>0</b>
<b>Deinococcus-Thermus</b>	<b>22</b>	<b>24</b>	<b>4</b>	<b>0</b>	<b>20</b>	<b>0</b>	<b>0</b>
Euryarchaeota	21	3	21	3	60	3	2
Firmicutes	549	705	530	304	388	405	339
Fusobacteria	1	1	1	1	1	1	3
Gemmatimonadetes	3	3	1	1	1	1	1
Nitrospirae	3	2	2	2	2	0	1
Planctomycetes	45	45	43	30	46	7	46
Proteobacteria	3383	3150	3206	2007	2193	1905	1519
Spirochaetes	207	115	112	2	40	7	110
Synergistetes	2	2	2	2	2	2	2
<b>Tenericutes</b>	<b>76</b>	<b>20</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
Thermodesulfobacteria	0	0	0	0	0	0	16
Thermotogae	17	17	17	17	0	0	17
Verrucomicrobia	16	16	16	7	7	4	8

**Tableau 4.** Nombre de Hits correspondant aux phylums potentiellement détectés dans chacun des 6 sols inoculés par la communauté CS3, après 6 mois d'incubation. ENS3 correspond à l'inoculum de départ.

Phylum	S1_T2	S2_T2	S4_T2	S5_T2	S7_T2	S9_T2	ENS3
Acidobacteria	14	14	11	15	16	15	21
Actinobacteria	1707	1638	1427	1650	1127	450	1154
Aquificae	2	1	1	1	2	1	1
Bacteroidetes	113	136	77	108	186	53	93
Candidatus Poribacteria	1	1	1	1	1	1	1
Chlamydiae	0	0	0	0	0	28	1
Chlorobi	6	6	6	6	6	6	6
Chloroflexi	9	6	6	8	9	6	11
Crenarchaeota	7	7	7	20	7	7	20
Cyanobacteria	116	94	62	158	156	78	230
Deferribacteres	2	2	2	2	6	2	2
Deinococcus-Thermus	29	28	4	28	24	1	29
Elusimicrobia	0	0	0	0	0	0	1
Euryarchaeota	21	3	21	3	8	3	120
Firmicutes	1670	1320	956	904	784	496	880
Fusobacteria	26	26	26	26	25	1	25
Gemmatimonadetes	0	1	0	1	0	0	1
Nitrospirae	4	4	1	4	4	1	10
Planctomycetes	35	30	7	13	54	7	50
Proteobacteria	3549	2703	2681	2928	2962	1841	2844
Spirochaetes	151	196	112	115	244	115	115
Synergistetes	2	2	2	2	2	2	3
<b>Tenericutes</b>	<b>73</b>	<b>72</b>	<b>76</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>
Thermodesulfobacteria	16	16	16	0	0	0	17
Thermotogae	27	27	3	27	0	0	60
Verrucomicrobia	7	7	4	7	7	7	17

**Tableau 5.** Nombre de Hits correspondant aux phylums potentiellement détectés dans chacun des 6 sols inoculés par la communauté CS5, après 6 mois d'incubation. ENS5 correspond à l'inoculum de départ.

Phylum	S1_T2	S2_T2	S4_T2	S5_T2	S7_T2	S9_T2	ENS5
Acidobacteria	14	16	10	3	6	18	12
Actinobacteria	768	1443	373	1533	715	799	1239
Aquificae	1	1	1	1	1	1	2
Bacteroidetes	89	118	83	104	60	60	113
Candidatus Poribacteria	1	1	1	1	1	1	1
Chlamydiae	0	0	0	0	28	0	1
Chlorobi	6	6	6	6	6	6	6
Chloroflexi	9	9	8	8	6	6	14
Crenarchaeota	20	20	7	20	7	7	20
Cyanobacteria	46	56	46	43	43	41	199
Deferribacteres	2	2	2	2	1	2	2
Deinococcus-Thermus	20	44	24	4	24	0	44
Elusimicrobia	0	0	0	0	0	0	1
Euryarchaeota	21	21	3	21	3	2	16
Firmicutes	606	716	492	374	491	399	844
Fusobacteria	1	2	1	1	1	2	2
Gemmatimonadetes	1	1	1	1	0	0	1
Nitrospirae	0	0	0	0	0	1	1
Planctomycetes	30	32	7	7	7	7	45
Proteobacteria	2802	3282	2468	2470	2226	1728	3034
Spirochaetes	139	112	2	137	110	110	115
Synergistetes	2	2	2	2	2	2	2
Tenericutes	1	2	1	2	0	1	48
Thermodesulfobacteria	0	0	0	0	0	0	17
Thermotogae	39	24	2	26	3	0	44
Verrucomicrobia	17	7	16	16	4	4	10

**Tableau 6.** Nombre de Hits correspondant aux phylums potentiellement détectés dans chacun des 6 sols inoculés par la communauté CS7, après 6 mois d'incubation. EDL correspond à l'inoculum de départ.

Phylum	S1_T2	S2_T2	S4_T2	S7_T2	S9_T2	EDL
<b>Acidobacteria</b>	<b>15</b>	<b>13</b>	<b>9</b>	<b>12</b>	<b>10</b>	<b>0</b>
Actinobacteria	687	941	473	403	279	519
Aquificae	1	1	1	2	1	1
Bacteroidetes	135	109	64	71	29	41
Candidatus Poribacteria	1	1	1	1	0	1
<b>Chlamydiae</b>	<b>0</b>	<b>28</b>	<b>0</b>	<b>28</b>	<b>0</b>	<b>0</b>
Chlorobi	6	6	6	6	6	6
Chloroflexi	9	9	8	6	3	5
Crenarchaeota	20	34	7	7	7	7
Cyanobacteria	47	128	46	44	42	17
<b>Deferribacteres</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>
<b>Deinococcus-Thermus</b>	<b>0</b>	<b>24</b>	<b>0</b>	<b>24</b>	<b>0</b>	<b>0</b>
<b>Euryarchaeota</b>	<b>23</b>	<b>5</b>	<b>3</b>	<b>21</b>	<b>3</b>	<b>0</b>
Firmicutes	621	611	299	410	305	92
Fusobacteria	25	25	25	1	1	1
<b>Gemmatimonadetes</b>	<b>3</b>	<b>3</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Lentisphaerae</b>	<b>3</b>	<b>3</b>	<b>0</b>	<b>3</b>	<b>0</b>	<b>0</b>
<b>Nitrospirae</b>	<b>2</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>0</b>
Planctomycetes	32	32	30	7	7	3
Proteobacteria	4282	3143	2096	2495	1351	726
<b>Spirochaetes</b>	<b>115</b>	<b>115</b>	<b>105</b>	<b>115</b>	<b>5</b>	<b>0</b>
<b>Synergistetes</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>0</b>
Tenericutes	102	20	27	19	0	19
<b>Thermodesulfobacteria</b>	<b>16</b>	<b>0</b>	<b>16</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Thermotogae</b>	<b>17</b>	<b>17</b>	<b>15</b>	<b>2</b>	<b>0</b>	<b>0</b>
Verrucomicrobia	22	7	16	4	4	4

**Tableau 7.** Genres nouvellement détectés dans le sol 1 inoculé par CS1 au temps T2=6mois par rapport à l'inoculum ENS1.**S1.CS1/T2**

<b>Genre</b>	<b>Phylum</b>	<b>Genre</b>	<b>Phylum</b>
Acidothermus	Actinobacteria	Maritalea	Proteobacteria
<b>Advenella</b>	<b>Proteobacteria</b>	Massilia	Proteobacteria
Alkalibacterium	Firmicutes	Meganema	Proteobacteria
Altererythrobacter	Proteobacteria	<b>Mesoplasma</b>	<b>Tenericutes</b>
Aquamicrobium	Proteobacteria	Methylbium	Proteobacteria
<b>Azospira</b>	<b>Proteobacteria</b>	Naxibacter	Proteobacteria
Azotobacter	Proteobacteria	Nitrobacteria	Proteobacteria
Azovibrio	Proteobacteria	Nitrosomonas	Proteobacteria
Balneimonas	Proteobacteria	<b>Nostocoida type II</b>	<b>Actinobacteria</b>
<b>Bdellovibrio</b>	<b>Proteobacteria</b>	<b>Oceanobacillus</b>	<b>Firmicutes</b>
Bibersteinia	Proteobacteria	Ochrobactrum	Proteobacteria
<b>Bordetella</b>	<b>Proteobacteria</b>	Oligotropha	Proteobacteria
Buchnera	Proteobacteria	<b>Ornithinibacillus</b>	<b>Firmicutes</b>
Candidatus Endobugula	Proteobacteria	Oxalobacter	Proteobacteria
Candidatus Midichloria	Proteobacteria	<b>Paucisalibacillus</b>	<b>Firmicutes</b>
Candidatus Sulcia	Bacteroidetes	Petrobacter	Proteobacteria
Cardiobacterium	Proteobacteria	<b>Phycoccus</b>	<b>Actinobacteria</b>
Camobacterium	Firmicutes	<b>Pleomorphomonas</b>	<b>Proteobacteria</b>
<b>Castellaniella</b>	<b>Proteobacteria</b>	Polaromonas	Proteobacteria
Chelatococcus	Proteobacteria	<b>Pontibacillus</b>	<b>Firmicutes</b>
Chitinophaga	Bacteroidetes	Porphyrobacter	Proteobacteria
<b>Cyclobacterium</b>	<b>Bacteroidetes</b>	Prolixibacter	Bacteroidetes
Daeguia	Proteobacteria	Propionivibrio	Proteobacteria
<b>Dechloromonas</b>	<b>Proteobacteria</b>	<b>Pseudidiomarina</b>	<b>Proteobacteria</b>
Deferibacter	Deferibacteres	Pseudochrobactrum	Proteobacteria
Deinococcus	Deinococcus-Thermus	Pusillimonas	Proteobacteria
Desmospora	Firmicutes	Rhodocyclus	Proteobacteria
Desulfofaba	Proteobacteria	Salicola	Proteobacteria
Desulfotignum	Proteobacteria	<b>Salinibacter</b>	<b>Bacteroidetes</b>
Dichelobacter	Proteobacteria	<b>Sediminibacillus</b>	<b>Firmicutes</b>
Duganella	Proteobacteria	Spiroplasma	Tenericutes
Endozoicomonas	Proteobacteria	Sufflavibacter	Bacteroidetes
<b>Entomoplasma</b>	<b>Tenericutes</b>	Suttonella	Proteobacteria
Escherichia	Proteobacteria	Symbiobacterium	Firmicutes
Herbaspirillum	Proteobacteria	Tepidiphilus	Proteobacteria
Hermiimonas	Proteobacteria	<b>Terrabacter</b>	<b>Actinobacteria</b>
Hydrogenivirga	Aquificae	<b>Terribacillus</b>	<b>Firmicutes</b>
Hydrogenophilus	Proteobacteria	Thauera	Proteobacteria
<b>Idiomarina</b>	<b>Proteobacteria</b>	Thiofaba	Proteobacteria
Janthinobacterium	Proteobacteria	Thiothrix	Proteobacteria
Kaistobacter	Proteobacteria	Trigonala	Firmicutes
Kordiimonas	Proteobacteria	Undibacterium	Proteobacteria
Luteimonas	Proteobacteria	Variovorax	Proteobacteria
Maribacter	Bacteroidetes	Zhihengliuella	Actinobacteria

**Tableau 8.** Genres nouvellement détectés dans le sol 2 inoculé par CS1 au temps T2=6mois par rapport à l'inoculum ENS1.**S2.CS1/T2**

<b>Genre</b>	<b>Phylum</b>
Acidothermus	Actinobacteria
Alkalibacterium	Firmicutes
Azovibrio	Proteobacteria
Balneimonas	Proteobacteria
Bibersteinia	Proteobacteria
Brochothrix	Firmicutes
Buchnera	Proteobacteria
Candidatus Endobugula	Proteobacteria
Candidatus Sulcia	Bacteroidetes
Carboxydotherrmus	Firmicutes
Cardiobacterium	Proteobacteria
Carnobacterium	Firmicutes
<b>Caryophanon</b>	<b>Firmicutes</b>
Chelatococcus	Proteobacteria
Chitinophaga	Bacteroidetes
Daeguia	Proteobacteria
Deferribacter	Deferribacteres
Deinococcus	Deinococcus-Thermus
<b>Delftia</b>	<b>Proteobacteria</b>
Desmospora	Firmicutes
Desulfofaba	Proteobacteria
Desulfotignum	Proteobacteria
Dichelobacter	Proteobacteria
Endozoicomonas	Proteobacteria
Escherichia	Proteobacteria
<b>Fucophilus</b>	<b>Verrucomicrobia</b>
Herbaspirillum	Proteobacteria
Hydrogenophilus	Proteobacteria
Kaistobacter	Proteobacteria
Kitasatospora	Actinobacteria
Kordiimonas	Proteobacteria
Leptotrichia	Fusobacteria
Luteimonas	Proteobacteria
Mariibacter	Bacteroidetes
Maritalea	Proteobacteria
Meganema	Proteobacteria
<b>Niastella</b>	<b>Bacteroidetes</b>
Nitrobacteria	Proteobacteria
Ochrobactrum	Proteobacteria
Oligotropha	Proteobacteria
Oxalobacter	Proteobacteria
Pandoraea	Proteobacteria
<b>Pilimelia</b>	<b>Actinobacteria</b>
Polaromonas	Proteobacteria

<b>Genre</b>	<b>Phylum</b>
Porphyrobacter	Proteobacteria
Prolixibacter	Bacteroidetes
Propionivibrio	Proteobacteria
Pseudochrobactrum	Proteobacteria
Pusillimonas	Proteobacteria
Rhodocyclus	Proteobacteria
Salicola	Proteobacteria
<b>Solibacillus</b>	<b>Firmicutes</b>
Sufflavibacter	Bacteroidetes
Suttonella	Proteobacteria
Symbiobacterium	Firmicutes
Thauera	Proteobacteria
Thiobacter	Proteobacteria
Thiofaba	Proteobacteria
Thiothrix	Proteobacteria
Trigonala	Firmicutes
Variovorax	Proteobacteria
<b>Zobellia</b>	<b>Bacteroidetes</b>

**Tableau 9.** Genres nouvellement détectés dans le sol 4 inoculé par CS1 au temps T2=6mois par rapport à l'inoculum ENS1.**S4.CS1/T2**

<b>Genre</b>	<b>Phylum</b>
Acidothermus	Actinobacteria
Alkalibacterium	Firmicutes
Altererythrobacter	Proteobacteria
Azovibrio	Proteobacteria
Bibersteinia	Proteobacteria
Candidatus Endobugula	Proteobacteria
Cardiobacterium	Proteobacteria
Carnobacterium	Firmicutes
Chelatococcus	Proteobacteria
Chitinophaga	Bacteroidetes
Daeguia	Proteobacteria
Deferribacter	Deferribacteres
Desmospora	Firmicutes
Dichelobacter	Proteobacteria
Duganella	Proteobacteria
Endozoicomonas	Proteobacteria
Escherichia	Proteobacteria
<b>Finegoldia</b>	<b>Firmicutes</b>
Herbaspirillum	Proteobacteria
Herminiimonas	Proteobacteria
Janthinobacterium	Proteobacteria
Kaistobacter	Proteobacteria
Kordiimonas	Proteobacteria
Kozakia	Proteobacteria
Luteimonas	Proteobacteria
Maribacter	Bacteroidetes
Massilia	Proteobacteria
Meganema	Proteobacteria
Naxibacter	Proteobacteria
Nitrobacteria	Proteobacteria
Ochrobactrum	Proteobacteria
Oxalobacter	Proteobacteria
Petrobacter	Proteobacteria
Polaromonas	Proteobacteria
Porphyrobacter	Proteobacteria
Prolixibacter	Bacteroidetes
Pseudochrobactrum	Proteobacteria
Rhodocyclus	Proteobacteria
Salicola	Proteobacteria
Sufflavibacter	Bacteroidetes
Suttonella	Proteobacteria
Symbiobacterium	Firmicutes
Tepidiphilus	Proteobacteria
Thauera	Proteobacteria

<b>Genre</b>	<b>Phylum</b>
Thiofaba	Proteobacteria
Thiothrix	Proteobacteria
Trigonala	Firmicutes
Undibacterium	Proteobacteria
Variovorax	Proteobacteria
Zhihengliuella	Actinobacteria



**Tableau 10.** Genres nouvellement détectés dans le sol 5 inoculé par CS1 au temps T2=6mois par rapport à l'inoculum ENS1.**S5.CS1/T2**

<b>Genre</b>	<b>Phylum</b>	<b>Genre</b>	<b>Phylum</b>
Acidothermus	Actinobacteria	Suttonella	Proteobacteria
Acrocarpospora	Actinobacteria	Symbiobacterium	Firmicutes
Anaerobranca	Firmicutes	Tepidiphilus	Proteobacteria
Azovibrio	Proteobacteria	Thauera	Proteobacteria
Brochothrix	Firmicutes	<b>Thermobrachium</b>	<b>Firmicutes</b>
Buchnera	Proteobacteria	<b>Thermomonas</b>	<b>Proteobacteria</b>
<b>Caloramator</b>	<b>Firmicutes</b>	Thiofaba	Proteobacteria
Candidatus Endobugula	Proteobacteria	Thiothrix	Proteobacteria
Carboxydotherrmus	Firmicutes	<b>Tolumonas</b>	<b>Proteobacteria</b>
Cardiobacterium	Proteobacteria	Trigonala	Firmicutes
Chelatococcus	Proteobacteria	Variovorax	Proteobacteria
Chitinophaga	Bacteroidetes	Zhihengliuella	Actinobacteria
Daeguia	Proteobacteria		
Deferribacter	Deferribacteres		
Desmospora	Firmicutes		
Dichelobacter	Proteobacteria		
Endozoicomonas	Proteobacteria		
Escherichia	Proteobacteria		
Herbaspirillum	Proteobacteria		
Herbidospora	Actinobacteria		
Hydrogenophilus	Proteobacteria		
Kaistobacter	Proteobacteria		
Kitasatospora	Actinobacteria		
Kordiimonas	Proteobacteria		
Leptotrichia	Fusobacteria		
Luteimonas	Proteobacteria		
Maribacter	Bacteroidetes		
Marinilactibacillus	Firmicutes		
Maritalea	Proteobacteria		
Meganema	Proteobacteria		
Microtetraspora	Actinobacteria		
Nitrobacteria	Proteobacteria		
Nitrosomonas	Proteobacteria		
Ochrobactrum	Proteobacteria		
Oxalobacter	Proteobacteria		
Petrobacter	Proteobacteria		
Planotetraspora	Actinobacteria		
Polaromonas	Proteobacteria		
Porphyrobacter	Proteobacteria		
Prolixibacter	Bacteroidetes		
Pseudochrobactrum	Proteobacteria		
Rhodocyclus	Proteobacteria		
Salicola	Proteobacteria		
Sufflavibacter	Bacteroidetes		

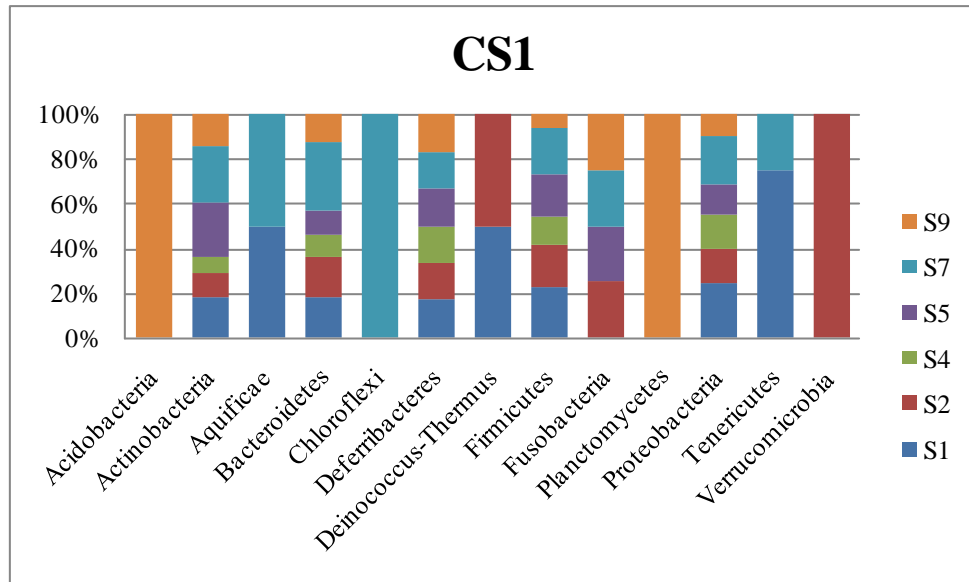
**Tableau 11.** Genres nouvellement détectés dans le sol 7 inoculé par CS1 au temps T2=6mois par rapport à l'inoculum ENS1.**S7.CS1/T2**

<b>Genre</b>	<b>Phylum</b>	<b>Genre</b>	<b>Phylum</b>
Acidothermus	Actinobacteria	Kordiiimonas	Proteobacteria
<b>Alistipes</b>	<b>Bacteroidetes</b>	Leptotrichia	Fusobacteria
Anaerobranca	Firmicutes	Lishizhenia	Bacteroidetes
<b>Anaerovorax</b>	<b>Firmicutes</b>	<b>Lonepinella</b>	<b>Proteobacteria</b>
<b>Anoxybacillus</b>	<b>Firmicutes</b>	Luteimonas	Proteobacteria
Aquamicrobium	Proteobacteria	Maribacter	Bacteroidetes
<b>Arsenicococcus</b>	<b>Actinobacteria</b>	Marinilactibacillus	Firmicutes
Azotobacter	Proteobacteria	<b>Marinomonas</b>	<b>Proteobacteria</b>
Azovibrio	Proteobacteria	Meganema	Proteobacteria
<b>Brachybacterium</b>	<b>Actinobacteria</b>	<b>Methylarcula</b>	<b>Proteobacteria</b>
Brochothrix	Firmicutes	Methylibium	Proteobacteria
Candidatus Endobugula	Proteobacteria	<b>Methylophaga</b>	<b>Proteobacteria</b>
Candidatus Midichloria	Proteobacteria	<b>Methylosulfonomonas</b>	<b>Proteobacteria</b>
<b>Candidatus Nitrotoga</b>	<b>Proteobacteria</b>	Mucilagibacter	Bacteroidetes
Candidatus Sulcia	Bacteroidetes	Nitrosomonas	Proteobacteria
Carboxydotherrmus	Firmicutes	Nubsella	Bacteroidetes
Cardiobacterium	Proteobacteria	Ochrobactrum	Proteobacteria
Chitinophaga	Bacteroidetes	Oligotropha	Proteobacteria
<b>Chloroflexus</b>	<b>Chloroflexi</b>	Pandoraea	Proteobacteria
Crocinitomix	Bacteroidetes	<b>Phocoenobacter</b>	<b>Proteobacteria</b>
Daeguia	Proteobacteria	<b>Planococcus</b>	<b>Firmicutes</b>
Deferribacter	Deferribacteres	Porphyrobacter	Proteobacteria
<b>Denitrobacterium</b>	<b>Actinobacteria</b>	Prolixibacter	Bacteroidetes
<b>Dermabacter</b>	<b>Actinobacteria</b>	Propionivibrio	Proteobacteria
Desmospora	Firmicutes	Pseudochrobactrum	Proteobacteria
Desulfofaba	Proteobacteria	Pusillimonas	Proteobacteria
Desulfotignum	Proteobacteria	<b>Rahnella</b>	<b>Proteobacteria</b>
<b>Devriesea</b>	<b>Actinobacteria</b>	Rhodocyclus	Proteobacteria
Dichelobacter	Proteobacteria	Ruegeria	Proteobacteria
Endozoicomonas	Proteobacteria	<b>Sideroxydans</b>	<b>Proteobacteria</b>
Escherichia	Proteobacteria	Spiroplasma	Tenericutes
<b>Ewingella</b>	<b>Proteobacteria</b>	<b>Starkeya</b>	<b>Proteobacteria</b>
<b>Flavococcus</b>	<b>Bacteroidetes</b>	Sufflavibacter	Bacteroidetes
<b>Gaetbulibacter</b>	<b>Bacteroidetes</b>	Suttonella	Proteobacteria
<b>Geobacillus</b>	<b>Firmicutes</b>	Thauera	Proteobacteria
<b>Hafnia</b>	<b>Proteobacteria</b>	Thiobacillus	Proteobacteria
Herbaspirillum	Proteobacteria	Thiobacter	Proteobacteria
Hydrogenivirga	Aquificae	Thiofaba	Proteobacteria
Hydrogenophilus	Proteobacteria	Thiothrix	Proteobacteria
Hyphomonas	Proteobacteria	<b>Tiedjeia</b>	<b>Proteobacteria</b>
<b>Ignatzschineria</b>	<b>Proteobacteria</b>	Trigonala	Firmicutes
Kaistobacter	Proteobacteria	Variovorax	Proteobacteria
Kitasatospora	Actinobacteria	<b>Wohlfahrtiimonas</b>	<b>Proteobacteria</b>
<b>Klebsiella</b>	<b>Proteobacteria</b>		

**Tableau 12.** Genres nouvellement détectés dans le sol 9 inoculé par CS1 au temps T2=6mois par rapport à l'inoculum ENS1.**S9.CS1**

<b>Genre</b>	<b>Phylum</b>
<b>Acidobacterium</b>	<b>Acidobacteria</b>
Acrocarpospora	Actinobacteria
Candidatus Endobugula	Proteobacteria
<b>Candidatus Solibacter</b>	<b>Acidobacteria</b>
Crocinitomix	Bacteroidetes
Dae guia	Proteobacteria
Deferribacter	Deferribacteres
Desmospora	Firmicutes
Desulfofaba	Proteobacteria
Endozoicomonas	Proteobacteria
Escherichia	Proteobacteria
Herbaspirillum	Proteobacteria
Herbidospora	Actinobacteria
Hyphomonas	Proteobacteria
Kaistobacter	Proteobacteria
Kordiimonas	Proteobacteria
Kozakia	Proteobacteria
Leptotrichia	Fusobacteria
Lishizhenia	Bacteroidetes
Luteimonas	Proteobacteria
Maribacter	Bacteroidetes
Marinilactibacillus	Firmicutes
<b>Marteella</b>	<b>Proteobacteria</b>
Microtetraspora	Actinobacteria
Mucilaginibacter	Bacteroidetes
Nitrobacteria	Proteobacteria
Nitrosomonas	Proteobacteria
Nubsella	Bacteroidetes
Ochrobactrum	Proteobacteria
Oligotropha	Proteobacteria
Planotetraspora	Actinobacteria
Porphyrobacter	Proteobacteria
Propionivibrio	Proteobacteria
Pseudochrobactrum	Proteobacteria
Pusillimonas	Proteobacteria
Ruegeria	Proteobacteria
<b>Singulisphaera</b>	<b>Planctomycetes</b>
<b>Sulfuricurvum</b>	<b>Proteobacteria</b>
Thiobacillus	Proteobacteria
Thiobacter	Proteobacteria
Trigonalia	Firmicutes
<b>Zoogloea</b>	<b>Proteobacteria</b>

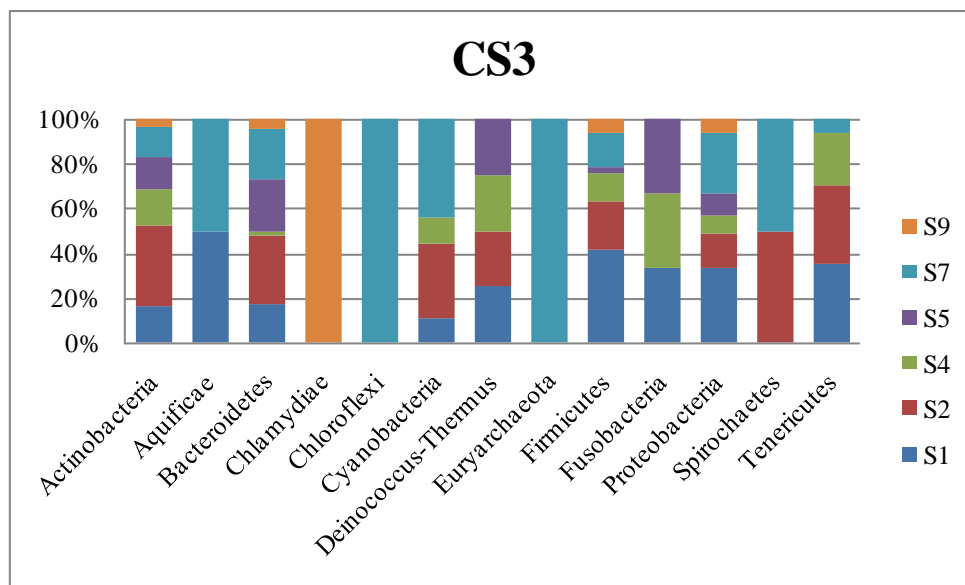
Les phylums correspondant aux genres détectés dans les 6 sols inoculés par la même communauté et non détectés dans l'inoculum de départ, au temps T2=6mois ont été comptabilisés, leur distribution dans chacun des six sols a pu alors être visualisée, pour chacune des 5 communautés inoculées : CS1, CS2, CS3, CS5 et CL (**Figure 2, 3, 4, 5, 6** respectivement).



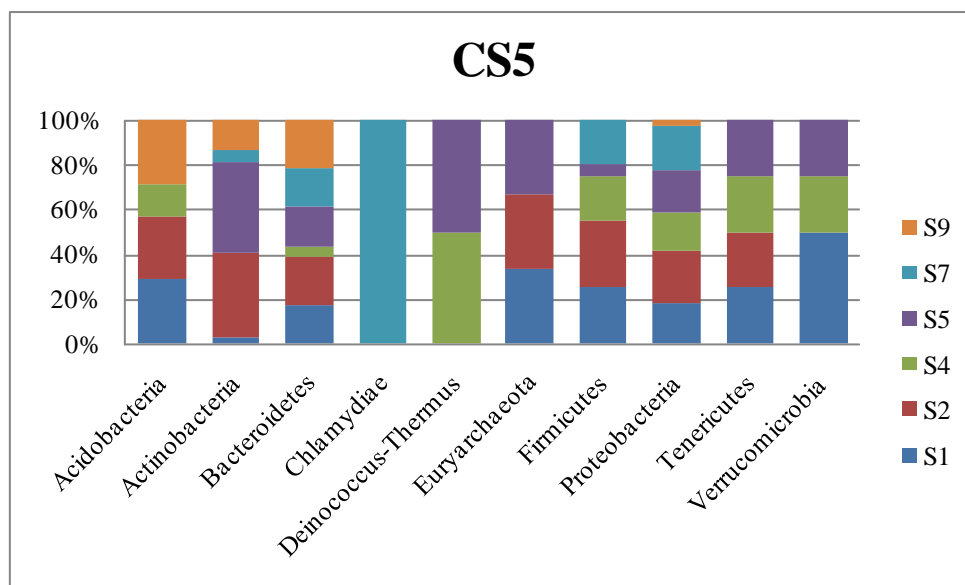
**Figure 2.** Taux de présence des phylums correspondant aux genres nouvellement détectés dans les 6 sols inoculés par la communauté CS1 (Côte Saint André), par rapport à l'inoculum ENS1.



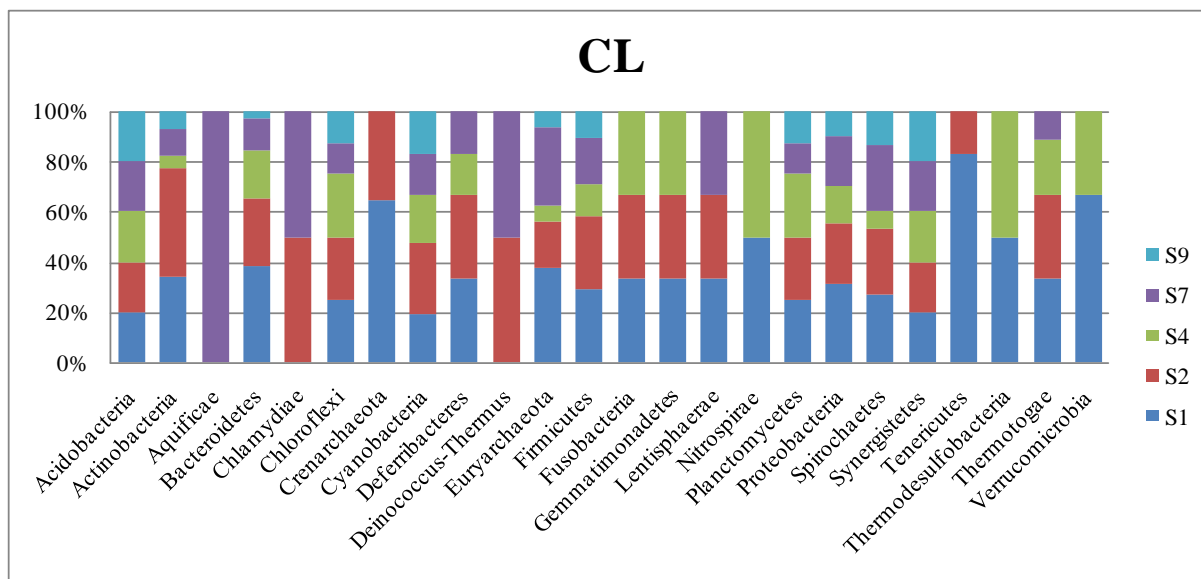
**Figure 3.** Taux de présence des phylums correspondant aux genres nouvellement détectés dans les 6 sols inoculés par la communauté CS2 (Montrond), par rapport à l'inoculum ENS2.



**Figure 4.** Taux de présence des phylums correspondant aux genres nouvellement détectés dans les 6 sols inoculés par la communauté CS3 (Brévil), par rapport à l'inoculum ENS3.



**Figure 5.** Taux de présence des phylums correspondant aux genres nouvellement détectés dans les 6 sols inoculés par la communauté CS5 (Kenya), par rapport à l'inoculum ENS5.



**Figure 6.** Taux de présence des phylums correspondant aux genres nouvellement détectés dans les 6 sols inoculés par la communauté CL (lacustre), par rapport à l'inoculum EDL.

Tous les taxons nouvellement identifiés appartenant au phylum Acidobacteria sont détectés dans le sol S9 inoculé par CS1 avec les genres *Acidobacterium*, *Candidatus Solibacter*, et dans les sols S5 et S2 inoculés par CS2 avec les genres *Candidatus* et *Solibacter*. Le genre *Acidobacterium* comprend un certain nombre d'espèces communément trouvées dans des réservoirs d'eau, de nombreux types de sols différents, des sédiments marins et d'eau douce... Ces bactéries sont impliquées dans la première étape de la destruction des molécules complexes produites par les organismes autotrophes (capables de synthétiser leurs propres éléments nutritifs). Le sol 9 présente l'acidité la plus forte, ce qui pourrait expliquer que ce sol favorise le développement de bactéries acidophiles.

Tous les taxons nouvellement identifiés appartenant au phylum Planctomycetes sont détectés dans le sol 9 inoculé par CS1 avec le genre *Singulisphaera*, et dans le sol S7 inoculé par CS2 avec le genre *Singulisphaera*. Les Planctomycetes forment un phylum de bactéries aquatiques, qui se trouvent dans des terrains saumâtres, marins ou d'eau douce. Ils se reproduisent par bourgeonnement (Glockner *et al.*, 2003).

Tous les taxons nouvellement identifiés appartenant au phylum Verrucomicrobia sont détectés dans le sol 2 inoculé par CS1 avec le genre *Fucophilus*. Le phylum Verrucomicrobia est un phylum divergeant à l'intérieur du domaine des bactéries incluant des membres de communautés microbiennes du sol et d'eaux douces et marines. Récemment des bactéries extrêmement acidophiles et capables d'oxyder le méthane ont été identifiées dans des sources chaudes (Lee *et al.*, 2009).

Tous les taxons nouvellement identifiés appartenant au phylum Chloroflexi sont détectés dans le sol 7 inoculé par CS1 avec le genre *Chloroflexus*, dans le S1 inoculé par CS2 avec le genre *Oscillochloris*, et dans le sol S7 inoculé par CS3 avec le genre *Caldilinea*. Le phylum Chloroflexi est constitué de bactéries photosynthétiques.

Tous les taxons nouvellement identifiés appartenant au phylum Tenericutes sont détectés dans les sols S7 et S1 inoculés par CS1, avec les genres *Spiroplasma* détectés dans les deux sols, *Mesoplasma* et *Entomoplasma* dans S1, et dans les sols S1, S2 inoculés par CL avec le genre *Acholeplasma* dans le sol S1, et les genres *Candidatus Phytoplasma*, *Entomoplasma*, *Mesoplasma*, *Spiroplasma*, *Candidatus Phytoplasma* dans les sols S1 et S2. Les Tenericutes sont des cellules dépourvues de paroi cellulaire, liées entre elles par une membrane plasmique. *Spiroplasma* est un genre de Mollicutes, un groupe de petites bactéries sans paroi cellulaire avec un métabolisme simple c'est-à-dire adapté à la vie parasitaire. La plupart des *Spiroplasmas* se développent soit dans l'intestin ou l'hémolymphe des insectes, soit dans le phloème des plantes (Jaenike *et al.*, 2009). Les *Spiroplasmas* sont des organismes sensibles, qui nécessitent un milieu de culture riche se développant bien à 30° C mais pas à 37° C.

Tous les taxons nouvellement identifiés appartenant au phylum Aquificae sont détectés dans les sols S1, S7 inoculés par CS1, dans les sols S7, S1 inoculés par CS3 et dans le sol S7 inoculé par CL, avec à chaque fois le genre *Hydrogenivirga*. Le phylum Aquificae recouvre une variété de bactéries qui vivent dans des environnements extrêmes. Ils ont été isolés au niveau des sources d'eau chaude avec une forte teneur en soufre, dans les fonds marins. Ce sont les membres dominants de la plupart des sources d'eau chaude terrestres, de pH neutre à alcalin et de températures supérieures à 60°C. Ce sont des bactéries autotrophes et sont les fixateurs primaires de carbone dans ces environnements (Nunoura *et al.*, 2008).

Tous les taxons nouvellement identifiés appartenant au phylum Deinococcus-Thermus ne sont détectés que dans les sols S1, S2 inoculés par CS1, dans les sols S4, S5 inoculés par CS5, avec le genre *Deinococcus* et dans les sols S2, S7 inoculés par CL avec le genre *Meiothermus*. *Deinococcus* est composé de plusieurs espèces résistantes aux radiations ionisantes (Albuquerque *et al.*, 2005). Ces organismes présentent des capacités à se développer au dépend de composés chimiques toxiques, notamment sur les déchets nucléaires, de survivre dans le vide de l'espace et à des températures extrêmes (Makarova *et al.*, 2007).

Tous les taxons nouvellement identifiés appartenant au phylum Chlorobi sont détectés dans les sols S4, S7 inoculés par CS2, avec les genres *Chloroherpeton* dans les deux sols et

*Chlorobaculum* dans le sol 4. Le phylum Chlorobi recouvre des bactéries du soufre anaérobies et photo autotrophes. Initialement apparentées aux Bacteroidetes dont elles se différencient, elles ont par la suite été affectées à un embranchement propre (Bryant & Frigaard, 2006). Ces bactéries sont photosynthétiques et non-mobiles, utilisent des ions sulfures, de l'hydrogène ou des ions ferreux comme donneurs d'électrons (réducteurs) et produisent du soufre élémentaire (minéral). Le soufre est déposé en dehors de la cellule où il peut être oxydé. Elles sont adaptées à des conditions extrêmes, réduites en ressources énergétiques. Leur niche écologique est souvent partagée avec des cyanobactéries: *prochlorococcus* et *synechococcus*. A titre d'exemple une bactérie du soufre a été trouvée à une profondeur de 2 500 mètres dans l'océan Pacifique au large du Mexique (Beatty *et al.*, 2005).

Tous les taxons nouvellement identifiés appartenant au phylum Chlamydiae sont détectés dans le sol S9 inoculé par CS3, avec le genre *Parachlamydia*, dans le sol S7 inoculé par CS5 et dans les sols S2, S7 inoculés par CL, avec les genres *Candidatus Clavochlamydi*, *Candidatus Fritschea*, *Candidatus Protochlamydia*, *Estrella*, *Neochlamydia*, *Parachlamydia*, *Chlamydia* et *Waddlia*. Le phylum Chlamydiae regroupe des bactéries, toutes parasites intracellulaires d'eucaryotes (Robertson *et al.*, 2009). Les espèces les plus connues infectent les mammifères et les oiseaux, mais certaines ont été retrouvées chez d'autres hôtes, comme des amibes. Ils sont apparentés à certains autres groupes bactériens, en particulier les Thermomicrobia mais constituent un phylum indépendant.

Tous les taxons nouvellement identifiés appartenant au phylum Euryarchaeota ne sont détectés que dans le sol S7 inoculé par CS3 avec les genres *Methanocalculus* et *Methanocorpusculum*. Le phylum des Euryarchaeota fait partie du domaine des Archées (Archaea). Les genres *Methanocalculus* et *Methanocorpusculum* sont des méthanogènes qui produisent du méthane et se rencontrent dans des environnements variés allant des sédiments marins aux intestins de nombreux animaux dont l'homme (Zhang *et al.*, 2003, Anderson *et al.*, 2009).

Tous les taxons nouvellement identifiés appartenant au phylum Spirochaetes sont détectés dans les sols S2, S7 inoculés par CS3, avec le genre *Spirocheta*. Les spirochètes (*Spirochaetes*) sont un phylum de bactéries distinctes à gram négatif (protéobactéries). Elles se distinguent par la présence de flagelles internes. Ces flagelles leur confèrent une mobilité et une rapidité de déplacement exceptionnellement élevées dans des milieux visqueux.



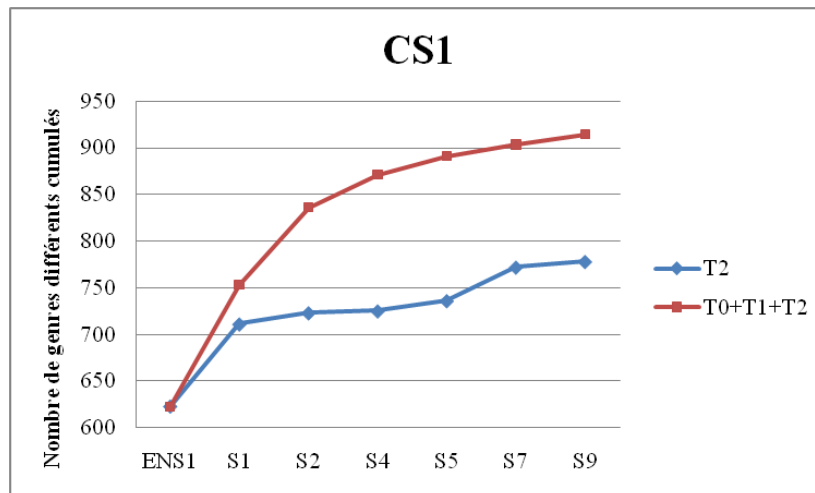
Tous les taxons nouvellement identifiés appartenant au phylum Crenarchaeota sont détectés dans les sols S1, S2 inoculés par CL, avec les genres *Acidilobus*, *Aeropyrum*, *Desulfurococcus*, *Geogemma*, *Hyperthermus*, *Ignicoccus*, *Pyrodictium*, *Staphylothermus* et *Thermofermentum* dans le sol S1 et les genres *Caldivirga*, *Pyrobaculum*, *Thermocladium*, *Thermoproteus* et *Vulcanisaeta* dans le sol S2. Les crénarchées appartiennent au domaine des archées (*Archaea*). Beaucoup d'espèces sont thermophiles (*ex* : *Thermoproteus*, *Pyrolobus*...). Leur environnement naturel est en général les sources d'eau chaude riche en soufre, comme c'est le cas dans le parc de Yellowstone aux Etats-Unis (Kozubal *et al.*, 2008).

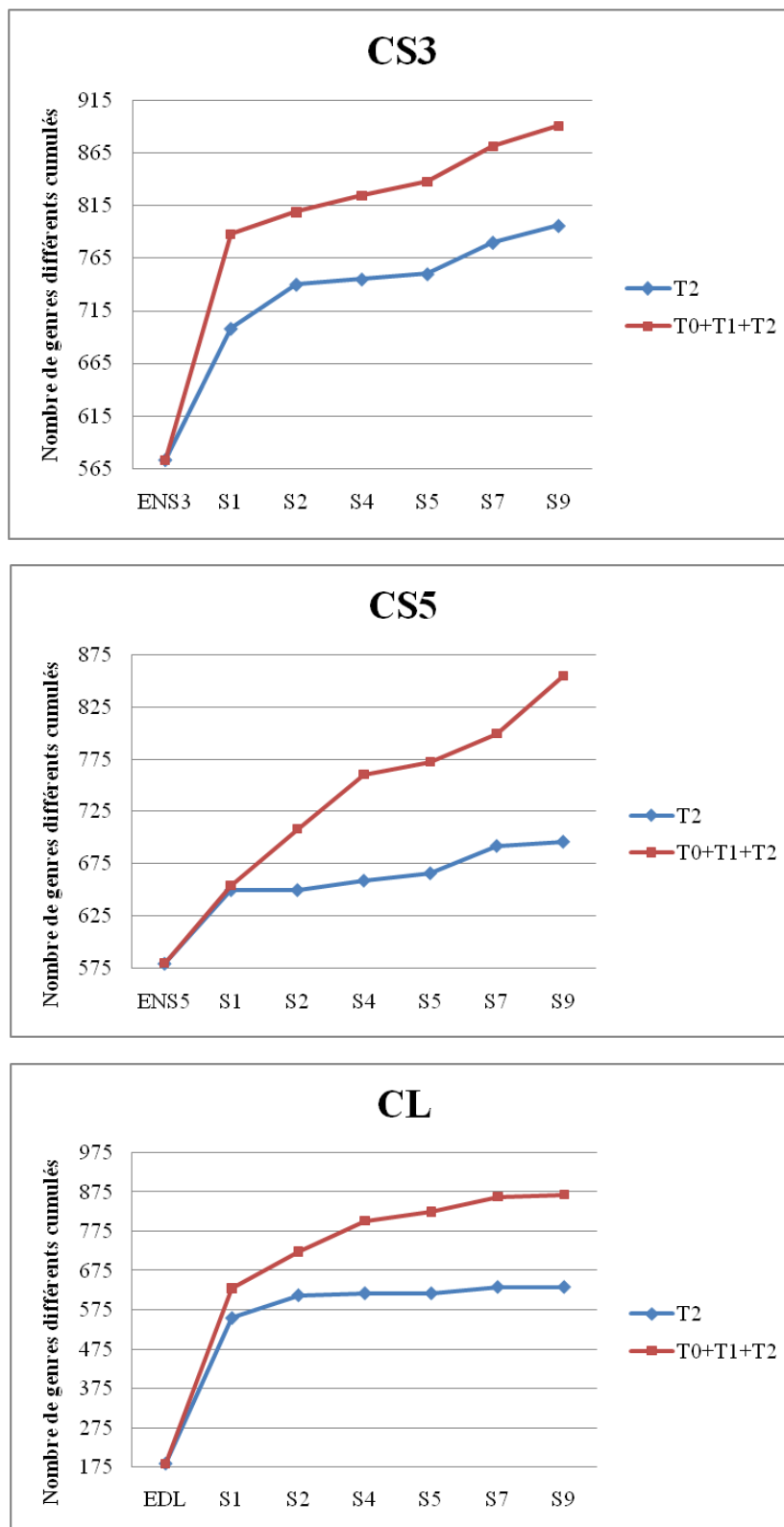
Et enfin les taxons nouvellement identifiés appartenant au phylum Nitrospirae sont détectés dans les sols S4, S1 inoculés par CL avec le genre *Thermodesulfovibrio* qui regroupe des bactéries sulfatoreductrices (Dillon *et al.*, 2007).

A T2=6mois, durée d'incubation pour laquelle nous avons considéré la stabilisation plus ou moins définitive de la nouvelle structure des communautés dans chacun des six sols, les résultats montrent que chaque sol ne favoriserait pas la croissance des mêmes bactéries quand il est inoculé par des communautés bactériennes différentes, ce qui devrait se réaliser si l'environnement abiotique était le seul facteur structurant de la communauté. Par exemple, le sol S2 est le seul sol sur les 6 inoculés par la communauté CS1 à mettre en évidence des taxons bactériens appartenant au phylum Verrucomicrobia, alors que ce sont les sols S1, S4 et S5 inoculés par la communauté CS5 qui ont mis en évidence des taxons appartenant à ce phylum. Par contre les genres bactériens appartenant au phylum Aquificae sont détectés à 3 reprises dans le sol S7 quand il est inoculé par les communautés CS1, CS2 et CL. Dans ce cas là le sol 7 pourrait avoir un impact important sur la croissance des genres appartenant à ce phylum. Il est difficile en règle générale, de corréler la présence de ces taxons appartenant à un phylum donné avec un sol particulier, et plus précisément avec des caractéristiques granulométriques ou physicochimiques. Ceci est confirmé par la détection dans des sols incubés à 20°C de genres sensés être inféodés à des milieux extrêmes (*ex* : le phylum Crenarchaeota qui regroupent des bactéries thermophiles) montrant qu'il est possible de détecter des microorganismes dans des environnements où leur présence n'est absolument pas attendue.

### 3.2. Nombre de genres différents cumulés obtenus par l'inoculation d'une communauté dans des sols stérilisés

Une communauté inoculée dans des sols stérilisés différents va-t-elle dévoiler des genres qui n'étaient pas détectés dans l'inoculum de départ, mais quel pourcentage d'augmentation est-il possible d'obtenir par rapport à l'inoculum de départ ?





**Figure 7.** Courbes des genres différents cumulés obtenus à partir des six sols inoculés par la même communauté bactérienne, en bleu : au temps T2=6mois, en rouge : durant les 3 temps T0=1jour + T1=2mois + T2=6mois. **A/** CS1; **B/** CS2 ; **C/**CS3 ; **D/**CS5 ; **E/**CL. EN et ED sont les communautés inoculées dans les 6 sols.

Les inocula initiaux (ENS1, ENS2, ENS3, ENS5 et EDL) correspondant aux communautés CS1, CS2, CS3, CS5 et CL, présentaient respectivement 623, 314, 573, 580, et 184 genres bactériens putatifs (**Figure 7. A/, B/, C/, D/, E/**). Après inoculation et incubation dans les 6 sols stérilisés pendant 6 mois (T2) qui permet la stabilisation de la nouvelle communauté bactérienne le nombre de genres bactériens détectables s'est accru significativement. La communauté bactérienne CS1 révèle 155 nouveaux genres, 331 pour CS2, 223 pour CS3, 116 pour CS5 et 450 pour CL. Cette augmentation est pour CS1 de 25%, pour CS2 de 105%, pour CS3 de 39%, pour CS5 de 20% et pour CL de 245%.

Le taux d'augmentation semble inversement corrélé au nombre de genres bactériens détectés dans l'inoculum de départ, plus grand étant le niveau de diversité et plus faible sera son accroissement. Ainsi l'inoculum ENS1 du sol 1 qui possède la plus grande diversité en nombre de genres bactériens détectés ne révélera que le plus faible nombre de nouveaux taxons dans les autres sols. Inversement l'inoculum « lacustre » présente une très faible diversité initiale (184 genres putatifs) mais montre la plus importante augmentation de diversité. Plusieurs hypothèses peuvent être proposées pour expliquer ce résultat : D'ordre méthodologique d'abord du fait que la puce à ADN taxonomique utilisée comporte une majorité d'oligosondes ciblant des taxons bactériens majoritairement telluriques. Certaines des populations majoritaires présentes dans cet environnement aquatique échapperaient donc à la détection du fait de ce biais. D'ordre écologique ensuite, le niveau de diversité dans les environnements aquatiques étant plus faible que dans les environnements telluriques. Il est toutefois intéressant de noter que l'inoculation dans les sols de cette communauté révèle un nombre très important de taxons connus pour leurs capacités à se développer dans le sol. Nos résultats confirment cependant l'intérêt de l'approche employée pour révéler des populations minoritaires mais surtout renseignent sur le niveau très élevé de cosmopolitisme des bactéries environnementales. Les environnements aquatiques hébergeraient donc une communauté bactérienne composée de très nombreuses populations minoritaires en termes d'effectifs dans les eaux mais capables de se développer dans les sols. Ce résultat pose à nouveau la question du niveau de spécificité de chaque habitat qui va bien au-delà des seuls environnements terrestres, le « everything is everywhere » pouvant être beaucoup plus large qu'initialement envisagé.

Il convient également de considérer ce concept à la lumière de la totalité des données générées par ces expérimentations qui montrent que le niveau de diversité est encore supérieur à celui révélé après 6 mois d'incubation. En effet, nos travaux sur l'évolution dans le temps de la

structure des communautés dans les sols stérilisés et inoculés montrent des différences significatives que nous avons analysées comme résultant d'une différence de valeurs adaptatives des différentes bactéries. Le fait que certains taxons détectés à T=0 (1 jour) ne le soient plus à T=1 (2 mois) et/ou à T=2 (6 mois) et que d'autres soient détectables uniquement à T=1 ou à T=2 indique que la méthodologie employée n'a pas atteint ses limites et que certainement beaucoup d'autres taxons restent à découvrir. Sur la **Figure 7** on peut clairement voir que la prise en compte des données obtenues à T=0 et T=1 augmente significativement le nombre de nouveaux taxons détectés par l'analyse par hybridation sur puces à ADN des solutions d'ADN provenant des échantillons de sol inoculés en comparaison des sols originaux.

C'est donc en fait 291 nouveaux genres bactériens détectés pour la communauté CS1, 539 pour la communauté CS2, 318 pour la communauté CS3, 274 pour la communauté CS5 et 685 pour la communauté CL correspondant à une augmentation de 47% pour CS1, 171% pour CS2, 55% pour CS3, 47% pour CS5, 372% pour CL par rapport à la diversité des inocula initiaux (ENS1, ENS2, ENS3, ENS5 et EDL respectivement).

Il est également intéressant de noter que chaque type de sols inoculé par un même inoculum a entraîné le développement d'une communauté présentant une structure spécifique (voir partie I) avec, pour chaque sol, la révélation de genres bactériens spécifiques. Des sols avec des propriétés physico-chimiques comparables favorisent le développement de communautés différenciées révélant chacune des taxons bactériens spécifiques qui dans chacun des cas n'avaient pas été détectés dans le sol d'origine. Quand par contre l'inoculation est réalisée dans des sols avec une texture clairement distincte comme c'est le cas avec le sol de Nouvelle Calédonie les structures des communautés bactériennes qui en résultent sont aussi très différenciées (voir partie I). Ces sols permettent de révéler un nombre significativement plus élevé de nouvelles populations bactériennes que les autres sols.

La question fondamentale qui se pose est de savoir quel niveau de diversité l'ajout d'autres sols dans les expérimentations permettrait d'atteindre, même si l'analyse statistique des résultats indique que la diversité révélée à partir d'une même communauté ait atteint un plateau. C'est en effet la conclusion qui peut être tirée des différentes courbes montrant le cumul des nouveaux genres détectés dans les 6 sols durant les 3 temps d'incubation par rapport à l'inoculum de départ, qui atteignent un nombre de genres cumulés très proches : 914 pour CS1, 853 pour CS2, 891 pour CS3, 854 pour CS5 et 869 pour CL (**Figure 7**. A/, B/, C/, D/, et E/ respectivement).

Cette limite peut s'expliquer par des biais techniques, de récents travaux visant à quantifier les biais liés aux méthodes d'extraction de l'ADN démontrent que l'emploi d'une seule méthode d'extraction de l'ADN bactérien à partir du sol ne permet la récupération que de moins de 10% de l'ADN total extractible. Des travaux futurs devraient permettre de définir si la diversité susceptible d'être révélée par l'emploi de techniques d'extraction optimisées se superpose ou s'additionne à celle liée à l'emploi de l'approche de ré-inoculation.

De plus l'approche utilisée peut souffrir d'une limite conceptuelle liée au relatif manque de diversité des environnements utilisés, qui tous étaient des environnements telluriques quoique de caractéristiques abiotiques différentes. Les résultats obtenus avec la communauté bactérienne aquatique inoculée dans des sols où le plus fort accroissement de diversité a pu être observé confirment l'intérêt qu'il y aurait à reproduire ces expérimentations avec un panel plus large d'environnements. Si la détection de nouveaux taxons est corrélée aux caractéristiques de la matrice, le nombre de taxons restant à découvrir pour chaque environnement pourrait être extrêmement élevé considérant le potentiel expérimental que représente la diversité des écosystèmes sur la planète.

De plus les puces à ADN utilisées dans cette expérience ciblaient un nombre de genre limité représentant les groupes dominants des écosystèmes telluriques. L'utilisation de puces à ADN taxonomiques assurant une couverture plus large des bactéries et archées présentes sur l'arbre phylogénétique du vivant ou bien le séquençage par les technologies haut débit actuellement disponibles qui ont été présentées précédemment (chapitre I) pourraient permettre de révéler une fraction encore plus importante de la diversité bactérienne.

#### **4. Conclusions**

L'inoculation de communautés bactériennes dans des sols stérilisés a permis de mettre en évidence des genres bactériens non détectés dans l'inoculum de départ. De nombreuses bactéries ne sont pas détectées dans les inocula de départ du fait de leur présence minoritaire, à des niveaux inférieurs aux limites de détection de la technique des puces à ADN. Nos travaux montrent que fournir à ces populations bactériennes de nouvelles conditions abiotiques, permet de stimuler la multiplication de certaines d'entre elles jusqu'à un niveau dépassant la limite de détection. Leur détection dans ce nouvel environnement révèle de fait leur présence dans la communauté bactérienne initiale confirmant ainsi la présence de ce que les anglo-saxons appellent la « rare biosphere » et l'intérêt de l'approche développée ici pour la mettre en évidence au moins en partie.

C'est donc une réponse au moins partielle qui est donnée à l'hypothèse de départ que « tout serait partout », les limites techniques (sensibilité des méthodes utilisées) et conceptuelles (faible nombre d'environnements différents explorés) ne permettent cependant pas de totalement confirmer ou infirmer. Les résultats obtenus par un réensemencement des sols avec des bactéries issues d'échantillons aquatiques qui ont permis de révéler de très nombreux taxons initialement indétectables et généralement plutôt inféodés aux environnements telluriques constituent des indices forts sur le cosmopolitisme des bactéries et l'importance en termes de réserve de biodiversité cachée que peut constituer la « rare biosphere ». Son exploration est cependant rendue particulièrement ardue quand des environnements aussi complexes et hétérogènes que les sols sont étudiés. L'étape d'extraction des cellules ou de leur ADN à partir de cette matrice en constitue une des limitations majeures. C'est fort de ce constat que nos efforts pour atteindre la biodiversité bactérienne tellurique se sont portés vers le développement d'une nouvelle approche indépendante de l'extraction des bactéries indigènes ou de leur ADN et que nous présentons dans le chapitre III.

## **CHAPITRE III**

Développement d'une nouvelle approche métagénomique :  
capture in situ de l'ADN métagénomique





## 1. Introduction

En contournant les biais liés à la mise en culture *in vitro* des microorganismes l'approche métagénomique est extrêmement prometteuse en vue de l'analyse (et de l'exploitation) de la diversité génétique bactérienne du sol tant au niveau populationnel que fonctionnel. Nous avons cependant vu dans les chapitres précédents que cette analyse demeure non exhaustive du fait de biais liés à l'extraction et à la purification des acides nucléiques du sol, mais aussi à la distribution extrêmement hétérogène des différentes populations au sein de la communauté bactérienne. De ce fait, les techniques d'investigation couramment employées, du fait de leur manque de sensibilité ne détecteront que très aléatoirement les très nombreuses bactéries à effectifs minoritaires révélant au contraire celles dont le niveau des effectifs est très élevé, mais qui ne constituent qu'une très faible proportion de la diversité. Nous avons proposé deux types d'approches pour tenter de contourner ces difficultés, l'une basée sur un artifice conceptuel (le changement de milieu) pour modifier le rapport entre populations majoritaires et minoritaires et l'autre l'utilisation de technologies à haute sensibilité comme les puces à ADN de type taxonomique. Nos résultats montrent tout l'intérêt de cette combinaison d'approches pour accroître le niveau de détection des bactéries du sol et aborder des questions essentielles sur leurs stratégies adaptatives dans des environnements aussi complexes que les sols de notre planète. Les objectifs des travaux présentés dans ce chapitre sont de compléter le registre des approches mises en places pour tenter d'atteindre l'exhaustivité de l'analyse tant populationnelle que fonctionnelle de la microflore tellurique. Une longue pratique des approches métagénomiques nous a conduits à identifier l'extraction des cellules bactériennes ou de leur ADN hors de la matrice tellurique comme une des étapes les plus limitante. Protégés dans des microsites du sol ou solidement adsorbés sur des composants comme l'argile, des populations bactériennes spécifiques et leurs acides nucléiques échappent aux techniques d'extraction et demeurent donc non détectés.

Le but de notre étude est de développer une technologie applicable *in situ* pour atteindre des populations bactériennes qui échappent aux méthodes conventionnelles que celles-ci soient dépendantes ou indépendantes de la mise en culture. Le principe de notre approche consiste à développer des bactéries domestiquées capables de remplir cette fonction quand elles sont inoculées dans le sol puis de rendre disponible à l'étude l'ADN ainsi piégé quand ces bactéries sont ré-isolées de l'échantillon de sol. Pour remplir ces fonctions de piégeage ces bactéries doivent répondre à plusieurs critères : Présenter une capacité de survie et possiblement de colonisation du sol pour pouvoir atteindre les populations indigènes

ciblées. Présenter une capacité d'intégration de l'ADN exogène, que celui-ci soit de l'ADN extracellulaire libéré par des cellules mortes ou de l'ADN intracellulaire. Cette capacité d'intégration de l'ADN peut être reliée au mécanisme de transformation génétique naturelle présent chez certains taxons bactériens ou à l'action de facteurs exogènes provoqués par l'expérimentateur afin de perméabiliser les enveloppes cellulaires et permettre l'entrée de l'ADN. Pouvoir pérenniser l'ADN ciblé après son entrée dans la cellule, notamment en l'incorporant au génome de la bactérie hôte pour qu'il soit répliqué concomitamment. Posséder l'arsenal moléculaire de façon à permettre à l'expérimentateur de ré-isoler spécifiquement du sol ces bactéries domestiquées et pouvoir spécifiquement reconnaître et isoler celles ayant accompli la tâche de piégeage de l'ADN des autres qui n'auront pas été transformées et qui ne présenteront donc aucun intérêt pour l'expérimentateur.

Notre travail s'inscrit dans les premières étapes du développement de cet outil, en utilisant une souche d' *E. coli* comme organisme receveur du fait des nombreux outils disponibles et en dépit du fait que cette bactérie dont l'écosystème naturel est le tube digestif des mammifères ne soit pas un organisme bien adapté à l'environnement tellurique. Quoique non transformable naturellement cette bactérie présente des fréquences de transformation élevées par électro-transformation. La machinerie cellulaire impliquée dans la recombinaison sera mise à profit pour pérenniser parmi l'ADN métagénomique internalisé les fragments d'intérêt requérant pour cela d'équiper la souche capture de séquences idoines sur lesquelles pourra s'amorcer la recombinaison. Enfin, l'isolement spécifique de la souche « capture » à partir du sol nécessitera la conjonction de plusieurs gènes marqueurs et nous montrerons par quel artifice basé sur l'utilisation de gènes létaux inductibles, cellules transformées et recombinées seront séparées du pool des cellules qui n'auront pas été modifiées génétiquement lors de leur passage dans le sol.

Dans un premier temps le principe général de l'outil ainsi que les différents éléments choisis seront expliqués. Enfin après différentes stratégies de clonage, les premiers résultats pourront être présentés. Ce travail se situe dans la continuité des travaux de Nathalie Lombard.

## 2. Matériels & Méthodes

### 2.1. Souches bactériennes et conditions de culture

Les souches d'*E. coli* utilisées sont répertoriées dans le **Tableau 1**. Les souches ont été conservées dans du glycérol (Euromedex, Souffelweyersheim, France) 25% en concentration finale et stockées à -80°C. Les étalements et cultures liquides ont été effectués dans le milieu nutritif Luria Bertani modifié (LBm), les milieux solides contenant de leur côté de l'agar à 10g.L<sup>-1</sup> et en présence de la pression de sélection adéquate. Les antibiotiques ont été ajoutés dans les milieux de culture adéquatement pour que les concentrations finales suivantes soient atteintes : ampicilline 100µg.ml<sup>-1</sup> (Euromedex), tétracycline 20µg.ml<sup>-1</sup> (Sigma Aldrich, Saint Quentin Fallavier, France), chloramphénicol 25µg.ml<sup>-1</sup> (Euromedex), Du L-arabinose (Carl Roth, Lautebourg, France) à 10g.L<sup>-1</sup> et de l'IPTG (Euromedex) à 500µg.ml<sup>-1</sup> ont été ajoutés dans le milieu quand précisé. Toutes les solutions dissoutes dans l'eau ont été stérilisées par passage sur filtre 0.2µm (Millipore Saint Quentin en Yvelines, France). Les plasmides ont été extraits avec les kits commerciaux Qiagen Plasmid Mini, Midi et Maxi (Courtaboeuf, France), Macherey-Nagel NucleoBond PC100 (Hoerdts, France), suivant les recommandations du fournisseur.

### 2.2. Réactions enzymatiques

Les mélanges réactionnels pour la restriction des ADN ont été réalisés avec 100 à 500 ng d'ADN suivant les cas, le tampon de digestion adapté (Fermentas, Mundolsheim, France), le ou les enzymes de restriction (Fermentas), le tout complété avec de l'eau stérile (Aguettant, Lyon, France) dans un volume final réactionnel de 25µl. Les mélanges ont été incubés à 37°C entre 1h et 2h ou plus suivant les enzymes utilisées et selon les recommandations du fournisseur. Les produits de digestion ont été vérifiés par migration sur gel d'agarose.

Les amplifications PCR ont impliqué l'enzyme Taq polymérase (0.025U.µL<sup>-1</sup>) de la société Invitrogen (Cergy Pontoise, France), la solution contenant les amorces sens (Fw) et antisens (Rev) 0.5µM et 0.1 à 0.5µg de l'échantillon d'ADN (plasmide) à amplifier en présence du tampon réactionnel, de 1.5mM de MgCl<sub>2</sub> et du mélange des dNTP (200µM concentration finale).

**Tableau 1.** Liste des souches *E. coli* utilisées

Souche <i>E. coli</i>	Génotype	Résistance
<b>DH5Z1</b>	F- $\phi$ 80lacZ $\Delta$ M15 $\Delta$ (lac-argF)U169 deoR recA1 endA1 hsdR15(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 $\lambda$ -spr lacI <sup>q</sup> tet <sup>R</sup>	Sp <sup>r</sup>
<b>TELI</b>	F- $\lambda$ -IN(rrnD-rrnE)1 $\Delta$ (Passalacqua <i>et al.</i> )X74 rpsL galK2 recD1903::Tn10d-Tet trpDC700::putA1303::[kn <sup>r</sup> -lacI <sup>q</sup> immE3-lac]	Kn <sup>r</sup> , Tc <sup>r</sup> , ImmE3 <sup>+</sup> , Cm <sup>s</sup>
<b>1661LI</b>	F- lac-3350, galK2, galT22, recD1014, IN(rrnD-rrnE)1, rpsL179 $\Delta$ trp::[Kn <sup>r</sup> -immE3-lacI <sup>q</sup> ]	Sm <sup>r</sup> , Kn <sup>r</sup> , Trp <sup>-</sup> , ImmE3 <sup>+</sup>
<b>1661LIAara</b>	F- lac-3350, galK2, galT22, recD1014, IN(rrnD-rrnE)1, rpsL179 $\Delta$ trp::[Kn <sup>r</sup> -immE3-lacI <sup>q</sup> ] $\Delta$ araBAD	Sm <sup>r</sup> , Kn <sup>r</sup> , Trp <sup>-</sup> , Ara <sup>-</sup> , ImmE3 <sup>+</sup>
<b>1661LIAara<math>\Delta</math>narG</b>	F- lac-3350, galK2, galT22, recD1014, IN(rrnD-rrnE)1, rpsL179 $\Delta$ trp::[Kn <sup>r</sup> -immE3-lacI <sup>q</sup> ] $\Delta$ araBAD $\Delta$ narGnarH	Sm <sup>r</sup> , Kn <sup>r</sup> , Trp <sup>-</sup> , Ara <sup>-</sup> , ImmE3 <sup>+</sup> , NarG <sup>-</sup>
<b>TELIAara<math>\Delta</math>narG</b>	F- $\lambda$ -IN(rrnD-rrnE)1 $\Delta$ (Passalacqua <i>et al.</i> )X74 rpsL galK2 recD1903::Tn10d-Tet trpDC700::putA1303::[knr-lacIqimmE3-lac] $\Delta$ araBAD $\Delta$ narGnarH	Kn <sup>r</sup> , Tc <sup>r</sup> , ImmE3 <sup>+</sup> , Cm <sup>s</sup> , Ara <sup>-</sup> , NarG <sup>-</sup>

Ara<sup>-</sup> : auxotrophe pour l'arabinose, Ara<sup>s</sup> : sensible à l'induction arabinose, Cm<sup>r</sup> : résistant au chloramphénicol, Cm<sup>s</sup> : sensible au chloramphénicol, ImmE3<sup>+</sup> : production d'antitoxine ImmE3, Kn<sup>r</sup> : résistant à la kanamycine, Sm<sup>r</sup> : résistant à la streptomycine, Tc<sup>r</sup> : résistant à la tétracycline, Trp<sup>-</sup> : Auxotrophe pour le tryptophane, NarG<sup>-</sup> : dépourvu des gènes *narG* et *narH*.

**Tableau 2.** Liste et caractéristiques des plasmides utilisés ou construits

Plasmide	Commentaires	Génotype	Phénotype
pB341	Vecteur de clonage	<i>cm<sup>r</sup></i> , ori p15A	Cm <sup>r</sup>
pSK360	Cible pour l'amplification de la cassette [P <sub>A1-O3/O4</sub> :: <i>relF-tcr</i> ]	<i>amp<sup>r</sup></i> , <i>tc<sup>r</sup></i> , 2 x (PA1-O3/O4):: <i>relF</i> , ori pMB1	Amp <sup>r</sup> , Tc <sup>r</sup> , IPTG <sup>s</sup>
pB33C3	Cible pour l'amplification de la cassette [ <i>araC</i> -P <sub>bad</sub> :: <i>colE3-rrnB</i> ]	<i>cm<sup>r</sup></i> , ori p15A, <i>araC</i> -P <sub>bad</sub> :: <i>colE3</i>	Cm <sup>r</sup> , Ara <sup>s</sup>
pB341C3	Clonage de la cassette [ <i>araC</i> -P <sub>bad</sub> :: <i>colE3-rrnB</i> ] dans pB341	<i>cm<sup>r</sup></i> , ori p15A, <i>araC</i> -P <sub>bad</sub> :: <i>colE3</i>	Cm <sup>r</sup> , Ara <sup>s</sup>
pB341RTc	Clonage de la cassette [P <sub>A1-O3/O4</sub> :: <i>relF-tcr</i> ] dans pB341	<i>cm<sup>r</sup></i> , ori p15A, P <sub>A1-O3/O4</sub> :: <i>relF-tcr</i>	Cm <sup>r</sup> , Tc <sup>r</sup> , IPTG <sup>s</sup>
pB341C3RTc1	Clonage de la cassette [P <sub>A1-O3/O4</sub> :: <i>relF-tcr</i> ] dans pB341C3	<i>cm<sup>r</sup></i> , ori p15A, <i>araC</i> -P <sub>bad</sub> :: <i>colE3</i> PA1-O3/O4:: <i>relF-tcr</i>	Cm <sup>r</sup> , Tc <sup>r</sup> , IPTG <sup>r</sup> , Ara <sup>s</sup>
pB341C3RTc2	Clonage de la cassette [P <sub>A1-O3/O4</sub> :: <i>relF-tcr</i> ] dans pB341C3	<i>cm<sup>r</sup></i> , ori p15A, <i>araC</i> -P <sub>bad</sub> :: <i>colE3</i> , PA1-O3/O4:: <i>relF-tcr</i>	Cm <sup>r</sup> , Tc <sup>r</sup> , IPTG <sup>s</sup> , Ara <sup>s</sup>
pBAD35K7	Clonage de la cassette [ <i>araC</i> -P <sub>bad</sub> :: <i>colE3-rrnB</i> ; P <sub>A1-O3/O4</sub> :: <i>relF-tcr</i> ]	<i>cm<sup>r</sup></i> , ori p15A, <i>araC</i> -P <sub>bad</sub> :: <i>colE3</i> , PA1-O3/O4:: <i>relF</i>	Cm <sup>r</sup> , IPTG <sup>s</sup> , Ara <sup>s</sup>
pB35K7G1G2	Clonage des zones de recombinaison narG1 et narG2	<i>cm<sup>r</sup></i> , ori p15A, <i>araC</i> -P <sub>bad</sub> :: <i>colE3</i> , PA1-O3/O4:: <i>relF</i>	Cm <sup>r</sup> , IPTG <sup>s</sup> , Ara <sup>s</sup>

Amp<sup>r</sup> : résistant à l'ampicilline, Ara<sup>r</sup> : résistant à l'induction par l'arabinose, Ara<sup>s</sup> : sensible à l'induction par l'arabinose, Cm<sup>r</sup> : résistant au chloramphénicol, IPTG<sup>r</sup> : résistant à l'induction par l'IPTG, IPTG<sup>s</sup> : sensible à l'induction par l'IPTG, Tc<sup>r</sup> : résistant à la tétracycline.

### **2.3. Clonage**

Les fragments d'ADN provenant de la restriction de l'ADN plasmidique ou de produits d'amplification soumis ou non à un traitement par des enzymes de restriction ont été utilisés pour les réactions de ligation. Préalablement, les produits PCR ont été purifiés à l'aide des kits commerciaux GFX PCR DNA and gel band purification (Amersham, Buckinghamshire, United Kingdom) ou Nucleospin Extract II (Macherey-Nagel, Hoerd, France) soit directement sur colonnes d'affinité soit suivant l'extraction à partir d'un gel d'agarose après électrophorèse. Lorsque précisé, des bouts francs ont été générés sur les extrémités cohésives des fragments d'ADN par l'utilisation de l'enzyme Klenow Fragments (Roche Diagnostics, Meylan, France). Le mélange réactionnel et les conditions d'incubation utilisés ont été ceux recommandés par le fabricant. La déphosphorylation des vecteurs a été effectuée par traitement avec l'enzyme phosphatase alcaline Shrimp (SAP) (Roche Diagnostics) suivant les recommandations du fournisseur. L'ADN a ensuite été purifié sur colonnes d'affinité commerciales. Les fragments d'ADN utilisés pour le clonage ont été purifiés à l'aide du kit Qubit (Turner Biosystem, Oregon, USA). Les fragments d'ADN, inserts et vecteurs ayant leurs deux extrémités franches ont été mélangés selon un ratio molaire insert/vecteur de 5 :1. Dans les autres cas inserts et vecteurs ont été mélangés selon un ratio molaire de 1 :1 quand leur taille était très similaire et de 3 :1 quand l'insert avait une taille inférieure au vecteur. Le mélange réactionnel d'un volume de 20µl était composé du mélange insert vecteur, de 2 µl de tampon adapté et d'au minimum 1µl de l'enzyme T4 ligase à 1U. µl<sup>-1</sup> (Roche, Meylan, France) le tout complété avec de l'eau stérile (Aguettant). Les conditions subséquentes de réaction incluaient une incubation d'une nuit à 16°C avant passage pour 10 min à 70°C et conservation sur glace avant transformation.

### **2.4. Préparation des cellules pour l'électroporation**

Les cellules d'*E. coli* ont été rendues électrocompétentes d'après le mode opératoire suivant : une colonie isolée fraîchement développée était inoculée en milieu liquide en présence de(s) pression(s) de sélection adéquate(s). La pré-culture a été incubée 1 nuit à 37°C sous agitation de 150 rpm, puis a été diluée au 1/100<sup>ème</sup> dans le milieu LBm contenant le(s) antibiotique(s) adéquat(s) pour être incubée une nuit à 37°C sous agitation de 150 rpm jusqu'à obtenir une densité optique (DO) à 600nm maximale de 0.6. A partir de cette étape, les cellules bactériennes ont été maintenues dans de la glace. La culture bactérienne a été centrifugée à

4000g pendant 10min à 4°C, le surnageant étant éliminé et le culot cellulaire lavé dans du glycérol 10% pré-refroidi à 4°C. Les étapes de centrifugation / lavage ont été effectuées successivement 4 fois, avec un volume de glycérol 10% froid de 1, 1/2, 1/5<sup>ème</sup> du volume initial de la culture bactérienne. Les cellules ont été reprises dans un volume final de glycérol 10% correspondant à 1/100<sup>ème</sup> du volume initial, puis utilisées directement ou bien réparties par aliquotes et stockées à -80°C.

## 2.5. Transformation

L'ADN recombinant a été introduit par électroporation dans les cellules bactériennes préparées selon le protocole détaillé ci-dessus soit directement soit à partir des stocks conservés à -80°C et après décongélation pendant 30min environ sur la glace. Dans une cuve d'électroporation de 2mm (BTX, MP Biomedicals, Illkirch, France) pré-refroidie sur glace, 2µl de solution d'ADN ont été mis en présence de 50µl de cellules et le volume a été ajusté avec 3µl d'eau stérile (Aguettant, Lyon, France). Un choc électrique de 2.5KV.cm<sup>-1</sup>, 200Ω et 25µF a été appliqué avec GenePulserLi (Bio-Rad, Marne La Coquette, France). Après choc électrique, du milieu LBm liquide a été ajouté en quantité suffisante pour (QSP) 1ml, puis les cellules ont été incubées de 1h à 2h à 37°C sous une agitation inférieure à 150 rpm. La suspension bactérienne a alors été diluée en série au 1/10<sup>ème</sup> dans une solution de chlorure de sodium (NaCl) à 8g.L<sup>-1</sup>. Les suspensions bactériennes aux dilutions 10<sup>0</sup>, 10<sup>-1</sup>, 10<sup>-2</sup>, ont été étalées sur milieux gélosé sélectif adéquat pour dénombrer les transformants les dilutions 10<sup>-4</sup>, 10<sup>-5</sup>, 10<sup>-6</sup>, étant étalées sur milieu gélosé non sélectif pour dénombrer les cellules réceptrices. Après incubation 1 nuit à 37°C, les colonies ont été dénombrées.

## 2.6. Criblage Phénotypique

Par induction en milieu liquide : L'insertion des gènes toxiques, leur expression et leur fonctionnalité ont été vérifiées en milieu liquide par criblage des transformants après induction chimique appropriée. Pour ce faire, les transformants ont été inoculés en microplaques 96 puits cellstar (Greiner Bio-one, Courtabeouf, France) contenant 100µl de milieu LBm et les antibiotiques adéquats. Les microplaques ont été incubées 4h à 37°C avec agitation puis les clones ont été transférés individuellement par réplique dans des microplaques contenant 100µl du milieu LBm additionné de l'antibiotique et de l'agent inducteur de l'expression toxique. L'expression des gènes toxiques placés sous contrôle du système régulateur dérivé de l'opéron lactose ont été induits par l'IPTG à été utilisé et le L(+)

arabinose a été utilisé pour induire l'expression des gènes toxiques placés sous contrôle du système régulateur dérivé de l'opéron arabinose. Les plasmides des transformants présentant le phénotype recherché ont été extraits à l'aide des kits commerciaux afin de contrôler d'après la taille des fragments d'ADN obtenus après restriction la validité de la construction génétique. .

## 2.7. Test d'efficacité

Une colonie fraîchement développée a été inoculée en milieu liquide contenant les antibiotiques adéquats. Après 12h d'incubation à 37°C sous agitation de 150rpm, la culture bactérienne a été diluée sériellement au 1/10<sup>ème</sup>, dans une solution de NaCl à 8g.L<sup>-1</sup> jusqu'à la dilution 10<sup>-6</sup>. Cent µl de chacune des dilutions ont été étalées sur milieu gélosé sélectif additionné ou non de l'agent inducteur (IPTG à 500µg.ml<sup>-1</sup> et/ou arabinose à 10g.L<sup>-1</sup>) avant incubation d'une nuit à 37°C puis dénombrement des colonies (UFC). Un taux d'échappement a été calculé en divisant le nombre d'UFC.ml<sup>-1</sup> se développant sur milieu avec inducteur par celui des UFC.mL<sup>-1</sup> se développant sur milieu sans inducteur.

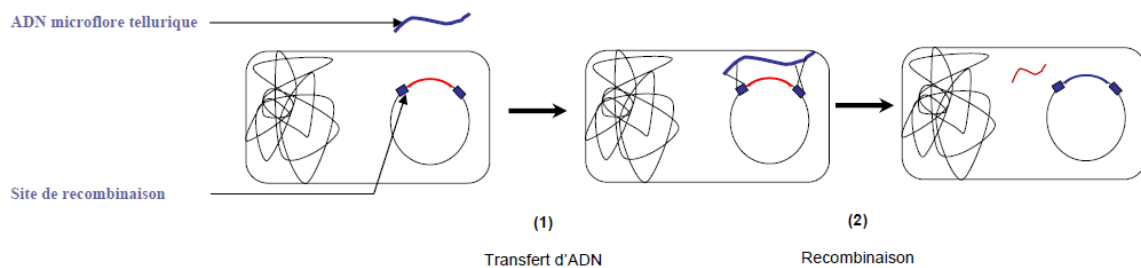
## 3. Résultats & discussion

En dépit d'avancées significatives apportées conjointement par la ré-inoculation de communautés bactériennes dans de nouveaux environnements et l'utilisation d'outils de grande sensibilité et spécificité, l'accès à la diversité bactérienne présente dans le sol demeure encore limité. Le biais est lié en grande partie à la difficulté d'atteindre dans le sol des cellules bactériennes appartenant à des taxons à faibles effectifs, protégées dans des microenvironnements, pouvant présenter une forte résistance aux traitements lytiques et dont l'ADN lorsqu'il peut être extrait s'adsorbe irréversiblement sur les composants du sol. C'est suite à ces constatations que fut mis en chantier le projet visant à utiliser une souche bactérienne qui inoculée dans le sol pourrait internaliser et pérenniser de l'ADN provenant de ces bactéries avant que son ré-isollement à partir du sol ne permette d'étudier et potentiellement d'exploiter cette information génétique « pêchée » *in situ*. L'objectif de nos travaux est de développer cet outil en vue d'en évaluer le potentiel en comparaison des différentes technologies conventionnelles de métagénomique et notamment de déterminer l'intérêt que représente cette nouvelle approche dans laquelle les cellules bactériennes indigènes ou leur ADN ne sont plus extraits des sols. Très clairement, cette technologie alternative n'a pas pour but de remplacer les approches conventionnelles de métagénomique,



du fait déjà du ciblage qui est fait sur des fractions de génome comme nous allons le voir mais de venir en complément ne serait-ce que pour aider à définir l'étendue de la diversité bactérienne tellurique. Du fait de la sophistication des constructions moléculaires nécessaires au développement de la souche et des différentes tâches de validation *in vitro* et *in situ* ce projet de grande ampleur ne peut résulter du travail réalisé dans le cadre d'une seule thèse de doctorat. Ce projet a été initié par une étudiante en thèse de l'équipe (NL) avant que mon arrivée dans le laboratoire me permette de m'impliquer dans la construction moléculaire de l'outil comme cela va être présenté ci-dessous. Les échéances d'un travail de doctorat m'ont contrainte à mon tour à passer le relais à d'autres scientifiques qui vont pouvoir maintenant déterminer les capacités de cette technologie.

### 3.1. Principe de l'approche métagénomique



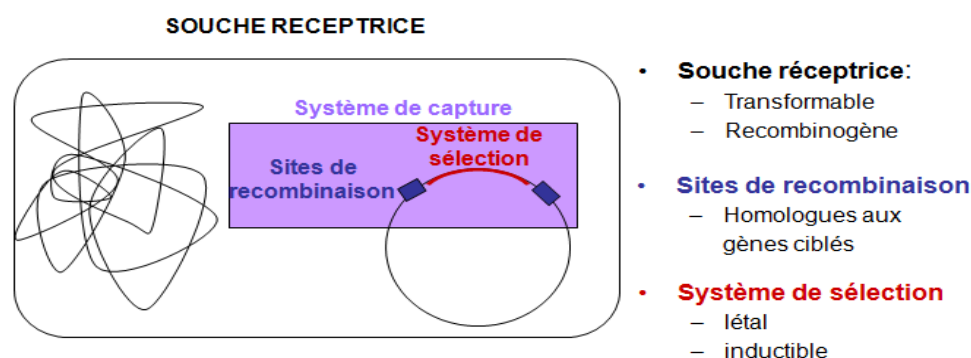
**Figure 1.** Mécanisme de piégeage de l'ADN de la microflore tellurique dans une souche réceptrice (Nathalie Lombard).

L'approche proposée est un processus en plusieurs étapes (**Figure 1**), au cours duquel l'ADN extracellulaire libre présent dans le sol serait internalisé par transformation (naturelle ou forcée) dans une souche bactérienne réceptrice modifiée génétiquement et inoculée dans le sol (Wayne, 1987). Une fois à l'intérieur de la cellule, cet ADN va être intégré au génome (plasmidique) de la souche réceptrice via un processus de recombinaison (2). Dans un second temps et grâce à un puissant système de sélection, les souches ayant recombiné l'ADN pourront être ré-isolées de l'environnement. La faisabilité de la première étape est appuyée par l'existence dans l'environnement de transferts naturels d'ADN entre espèces bactériennes différentes. Dans la synthèse bibliographique les différents processus responsables d'échanges génétiques ont été décrits, trois mécanismes principaux ayant été identifiés : la conjugaison, la transformation et la transduction auxquels il convient d'ajouter l'électro-transformation que

celle-ci découle d'événements climatiques comme la décharge de foudre ou de l'injection artificielle de courant dans des microcosmes de sol au laboratoire.

La faisabilité de la deuxième étape s'appuie sur l'existence de systèmes moléculaires bactériens de recombinaison permettant l'intégration dans le génome de l'ADN pénétrant. Le plus étudié est la voie de recombinaison RecBCD chez *E. coli*. Ce système intervient dans la réponse SOS pour la réparation de l'ADN endommagé. La recombinaison homologue fait intervenir le complexe enzymatique RecBCD et la protéine RecA (Dabert & Smith, 1997). La protéine RecA est l'enzyme clé catalysant la réaction de recombinaison entre l'ADN simple brin et l'ADN génomique (Chaudhury & Smith, 1984). L'ADN double brin entrant dans la cellule est dégradé par le complexe enzymatique RecBCD jusqu'à la rencontre de séquences *chi* (5'-GCTGGTGG-3'). La sous unité recD perd alors son affinité pour l'hétérodimère RecBC, qui devient déficient en activité hélicasique. Le complexe RecBCD dissocie alors les deux brins d'ADN et la protéine RecA s'associe alors à l'ADN simple brin nouvellement formé. L'intégration est ensuite amorcée si une homologie de séquence suffisante est détectée (Matic *et al.*, 1996), une séquence minimale d'identité étant en effet requise (Shen & Huang, 1986). Des régions de similarité de 25 à 200 pb suivant les espèces sont nécessaires pour initier la complémentation entre l'ADN entrant et l'ADN de la souche réceptrice (Thomas & Nielsen, 2005), la fréquence de recombinaison augmentant avec la taille de la zone de similarité.

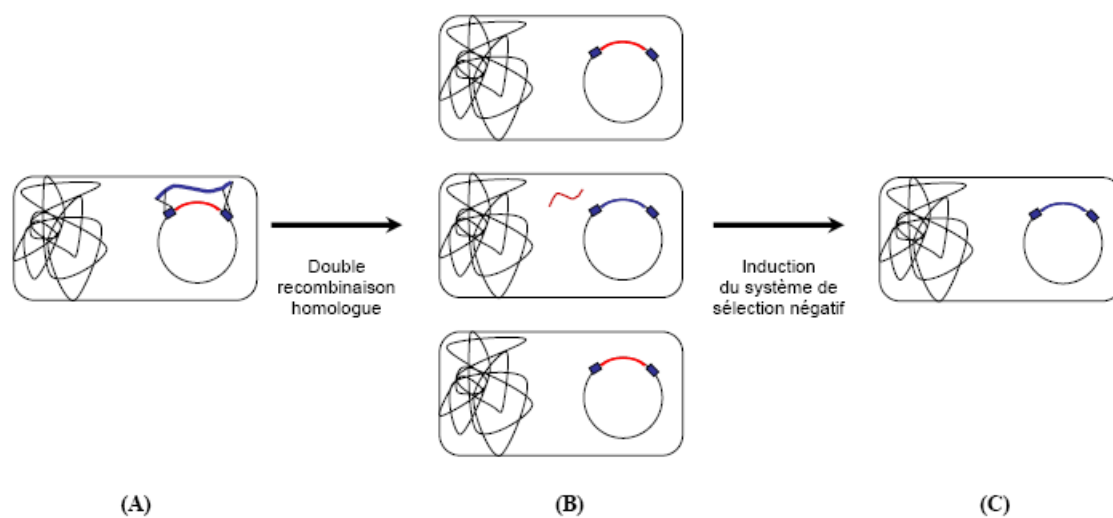
L'utilisation de ces mécanismes pour intégrer de l'ADN issu du sol dans la souche réceptrice nécessite la prise en compte de différents facteurs si la plus forte efficacité de recombinaison est recherchée (**Figure 2**).



**Figure 2.** Critères nécessaires à la construction pour la souche réceptrice les sites de recombinaison et le système de sélection.

L'étape critique de cette technologie repose sur la capacité pour l'expérimentateur d'identifier, sélectionner et isoler les cellules transformées (et recombinées) des autres

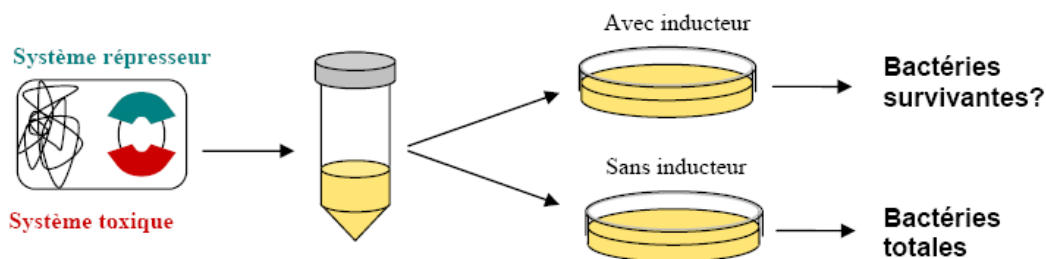
cellules de la souche réceptrice qui n'ont pas été modifiées génétiquement par la transformation. Le système de sélection développé repose sur un plasmide sur lequel est intégré une cassette comportant deux gènes toxiques, entourés de deux zones dévolues pour la recombinaison (**Figure 2**). Ces deux gènes étant sous le contrôle de promoteurs inducibles, sont donc réprimés en absence d'inducteur. L'intégration du fragment d'ADN ciblé par double recombinaison conduit à l'élimination du système de sélection. Sous conditions d'induction, les bactéries comportant le fragment d'intérêt peuvent être sélectionnées sur milieu de culture alors que les cellules non transformées seront tuées par les produits de l'un, de l'autre ou des 2 gènes létaux (**Figure 3**).



**Figure 3.** Principe de la contre sélection pour l'isolement des souches réceptrices transformées / recombinées. **(A)** La double recombinaison a lieu uniquement si une homologie de séquence suffisante est présente entre l'ADN entrant (bleu) et les sites de recombinaison (bleu) clonés dans la souche réceptrice. **(B)** Pour différencier les cellules réceptrices ayant intégré l'ADN entrant des autres cellules réceptrices, un système de sélection négatif inducible (en rouge), constitué de gènes exprimant une activité toxique pour la cellule en présence d'inducteur, a été ajouté entre les sites de recombinaison. La double recombinaison conduit à l'élimination du système de sélection (ou cassette toxique). **(C)** Suite à l'apport du substrat (inducteur) permettant l'expression des gènes de la cassette toxique, seules les souches ayant éliminé la cassette toxique survivront et pourront développer des colonies sur milieux de culture solidifiés (Nathalie Lombard).

La principale difficulté liée à la mise en place de ce système de sélection tient à sa nécessaire efficacité requise pour que la souche réceptrice puisse être utilisable dans les conditions décrites. Les taux de transformation escomptés, notamment en conditions *in situ* sont très faibles, entre  $10^{-5}$  et  $10^{-8}$ . Ceci est dû à la faible efficacité de la transformation *in situ*, à la dilution des gènes cibles dans le métagénome, aux divergences de séquences des cibles entre

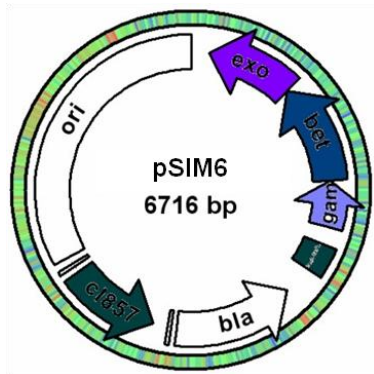
ADN donneur et régions clonées dans la souche réceptrice. Ces conditions nécessitent l'optimisation du système pour que les taux d'échappement soient inférieurs à  $10^{-8}$  voire plus bas de façon à ce que les seules colonies se développant sur milieux de culture soient celles relevant d'un événement de transfert-recombinaison et non pas celles qui, du fait d'une inactivation des gènes suicides ont échappé à la mort programmée. L'inactivation est liée aux mutations ponctuelles intervenant à chaque génération avec une fréquence d'environ une erreur tous les  $10^4$  nucléotides (Radman *et al.*, 2000) et sera donc fortement restreinte par l'utilisation de deux gènes placés sous deux systèmes régulateurs indépendants. Ce taux d'échappement peut être comptabilisé et correspond au nombre de colonies se développant sur milieu de culture solidifié en présence d'inducteur par rapport au nombre de colonies obtenues en absence d'inducteur (**Figure 4**).



**Figure 4.** Méthode d'évaluation de l'efficacité du système de sélection

Un effort important a également du être réalisé pour accroître l'efficacité de l'intégration par recombinaison de l'ADN entrant dans le génome de la souche réceptrice. La stratégie suivie a été d'utiliser le système lambda red gam constitué de gènes provenant du phage lambda, positionnés sur un plasmide et placés sous contrôle d'un promoteur inducible (Poteete *et al.*, 1988). Les gènes *exo* et *bet* (constituant le système red) du bactériophage permettent la recombinaison et l'intégration de l'ADN dans la souche hôte, le gène *exo* exprimant une exonucléase à activité 5'-3', le gène *bet* une protéine se fixant sur l'ADN simple brin, stimulant l'hybridation avec l'ADN génomique. L'invasion de l'ADN simple brin nécessite cependant l'action des protéines RecA (Delmas & Matic, 2005) et Gam (Murphy, 1998). Ces dernières se fixent sur le complexe enzymatique RecBCD, inhibant ainsi entre autres l'activité exonucléasique du complexe (Murphy, 1991). Chez *E. coli*, ces systèmes augmentent les fréquences de recombinaison d'ADN linéaire par rapport à la voie RecBCD (Murphy, 1998) et permettent l'amorçage de la recombinaison avec des séquences de seulement 30-50 nucléotides (Copeland *et al.*, 2001, Poteete, 2001). Les gènes *exo*, *bet* et *gam*, provenant du

phage lambda, ont été clonés sur des vecteurs pour leur emploi en biologie moléculaire (exemple plasmide pSIM6 **Figure 5**) (Datsenko & Wanner, 2000).



**Figure 5.** Représentation du plasmide pSIM6 obtenu à l'aide du logiciel pDRAW d'après sa séquence nucléotidique

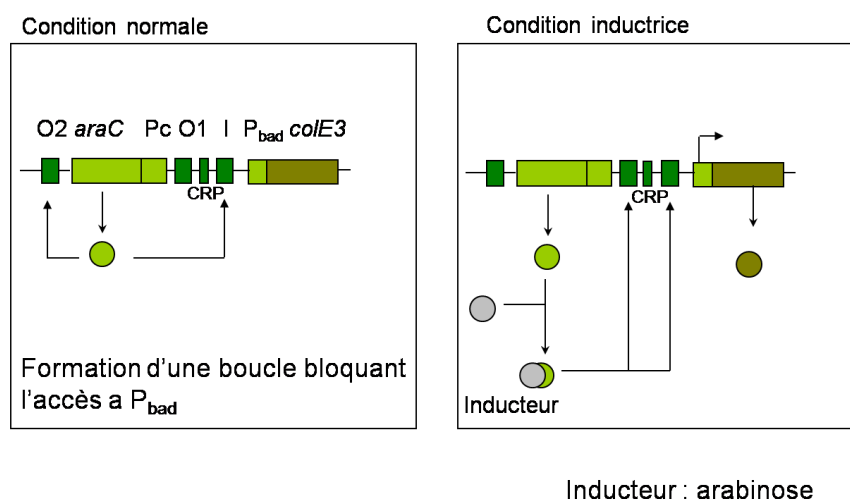
Ce système red gam est généralement placé sous le contrôle de promoteurs induits par l'IPTG ou l'arabinose pour activer de manière contrôlée la recombinaison (Chaveroche *et al.*, 2000, Datsenko & Wanner, 2000). Une série de plasmides récemment développés permet l'expression du système lambda red gam par augmentation de la température à 42°C (Datta *et al.*, 2006). Les gènes sont sous le contrôle du promoteur  $P_L$ , dont l'expression est réprimée par la protéine thermosensible (ts) CI857. La dénaturation de la molécule à 42°C permet ainsi une dérégulation du promoteur  $P_L$  et donc l'expression des gènes du système red gam. Les systèmes lambda red gam sont efficaces chez *E. coli* (Murphy, 1998). Leur fonctionnement a également été rapporté chez *Salmonella enterica* (Datta *et al.*, 2006).

### 3.2. Présentation des différents composants

#### 3.2.1. Les gènes suicides et leur régulation

Les gènes toxiques sont par définition des gènes dont l'expression entraîne des perturbations majeures chez la cellule hôte conduisant à sa mort. Ces gènes aux propriétés toxiques ont des origines et des fonctions variées. L'efficacité des gènes toxiques est aussi très variable d'un gène à l'autre. Pour une construction génétique dans une souche bactérienne donnée, une partie de la population est capable de survivre en dépit de l'expression toxique. Ce taux d'échappement peut varier de  $10^{-2}$  à  $10^{-5}$  lorsque le gène toxique est cloné sur un plasmide dans un système de régulation simple (Ahrenholtz *et al.*, 1994, Aizenman *et al.*, 1996). Ces bactéries échappent à l'action toxique suivant différents mécanismes dont la mutation du gène (Diaz *et al.*, 1994, Torres *et al.*, 2000) ou du système contrôlant son expression (Knudsen *et al.*, 1995, Torres *et al.*, 2003). Afin de diminuer cette fréquence d'échappement, des stratégies ont été développées consistant à cloner 2 gènes toxiques différents (Torres *et al.*, 2003).

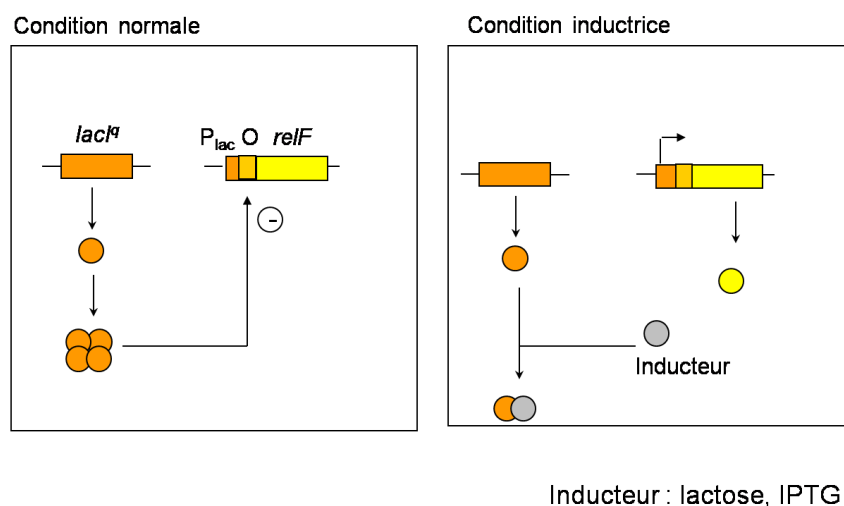
Le choix des gènes suicides s'est finalement porté sur *colE3* et *relF*, deux gènes différents ciblant deux composantes cellulaires distinctes et provoquant tous les deux la mort de la cellule. Le gène *colE3* code la toxine colicine E3, protéine à activité ribonucléasique. Cette toxine inhibe la synthèse protéique par clivage de la terminaison 3' de l'ARNr16S, le site de liaison à l'ARNm est alors éliminé, ce qui entraîne l'inhibition de la synthèse protéique aboutissant à la mort cellulaire (Senior & Holland, 1971). Le gène *colE3* est sous le contrôle d'un système régulateur dérivé de l'opéron arabinose, il est donc inductible en présence d'arabinose et réprimé en absence d'inducteur par le gène *araC* (**Figure 6**). Pour une efficacité optimale de la construction suicide, le système régulateur doit empêcher toute expression basale du système suicide en absence d'induction, une seule molécule de la toxine ColE3 suffisant à tuer la cellule. Pour contrer l'expression basale de la toxine ColE3 impossible à réprimer, le gène *immE3* codant une protéine labile à activité antitoxique immunité E3 qui neutralise l'activité RNase de *colE3* exprimé constitutivement après intégration dans le chromosome de la souche réceptrice.



**Figure 6.** Régulation du gène *colE3* sous le contrôle du promoteur P<sub>bad</sub>.

Le deuxième gène suicide utilisé, *relF* code la porine RelF, protéine membranaire agissant négativement sur la membrane par abolition du potentiel membranaire, ce qui a pour conséquence d'entraîner la mort de la cellule (Molin *et al.*, 1993). Ce gène est sous le contrôle d'un système régulateur dérivé de l'opéron lactose, donc induit en présence de lactose ou d'IPTG et réprimé en absence d'inducteur par le gène *lacI<sup>r</sup>*. L'opéron lactose est constitué d'un promoteur P<sub>lac</sub>, dont l'activité est contrôlée par le gène répresseur *lacI*. L'expression des gènes clonés sous P<sub>lac</sub> est permise par dérèpression en présence d'une molécule inductrice

l'IPTG (Griffiths, 1999) (**Figure 7**). Ce système est caractérisé par une très faible expression basale en absence d'inducteur, mais au contraire par une très forte expression en condition d'induction (Guzman *et al.*, 1995). Les promoteurs jouent un rôle important dans le contrôle de l'expression, d'une part dans l'efficacité de la répression par l'emplacement des séquences opératrices, d'autre part dans l'efficacité de la transcription par leur affinité avec l'ARN polymérase (Lanzer & Bujard, 1988).



**Figure 7.** Régulation du gène *relF* sous le contrôle du promoteur  $P_{lac}$

### 3.2.2. Les zones de recombinaison

Les zones de recombinaison peuvent être variables, en effet cet outil pouvant servir à différentes applications, les gènes ciblés peuvent alors aussi être variés. Dans un premier temps les zones de recombinaison nous permettront de valider l'outil moléculaire. Notre choix s'est porté sur les gènes *nar* intervenant dans la réduction des nitrates. Ces gènes sont relativement bien représentés dans les communautés bactériennes des sols, et pour la plupart bien conservés. De plus, ils présentent une organisation en opérons permettant de prendre en compte des fragments de taille variables pouvant atteindre plusieurs kb. C'est le gène *narG* codant pour la sous unité alpha du nitrate réductase qui a été sélectionné de par sa présence élevée chez les bactéries du sol (environ 30% des bactéries connues le possèdent).

### 3.2.3. Définitions et modification des souches réceptrices

Au début des travaux, l'option du système lambda red gam n'étaient pas au programme pour augmenter la recombinaison, c'est ainsi qu'une autre stratégie a été envisagée. Plusieurs souches ont été construites afin d'optimiser l'efficacité du plasmide suicide. L'intégration génomique de fragments d'ADN linéaires par recombinaison dans les souches recBCD est

stimulé par la présence de sites *chi*, qui diminuent l'activité exonucléasique de *recD* et favorisent la fixation de *recA* sur l'ADN simple brin. Ces sites *chi* ou des sites similaires (*chi*-like) sont présents chez de nombreux organismes (El Karoui *et al.*, 1999). La présence de 2 sites *chi* à chacune des extrémités du fragment linéaire et la restriction enzymatique à proximité de ces sites accroissent l'activité recombinogène (Dabert & Smith, 1997). Leur absence conduit à la diminution de la probabilité d'intégration dans les souches *recBCD*. La souche *E. coli* mutée *recD*, *E. coli* 1661, a été choisie pour son utilisation comme l'une des souches réceptrices. Ce gène *recD* code une protéine à activité exonucléasique, il intervient pour limiter l'intégration d'éléments génétiques étrangers à l'intérieur d'un génome bactérien. La mutation du gène *recD* entraîne une perte de l'activité exonucléasique tout en conservant l'activité recombinogène et la viabilité des cellules (Amundsen *et al.*, 1986). Sa mutation permet donc de faciliter la recombinaison de l'ADN avec la construction suicide puisque la souche peut internaliser l'ADN exogène sans le dégrader grâce à cette mutation du gène *recD*. Les mutants *recD* deviennent (hyper) recombinogènes en l'absence de sites *chi* (Dabert & Smith, 1997). La cassette [P3::immE3-kn<sup>r</sup>-lacI<sup>q</sup>] contenue dans *E. coli* TELI a été transduite dans le chromosome *E. coli* 1661 via le phage P1, pour ne pas augmenter la taille de la cassette toxique inductible. Cette souche nommée alors *E. coli* 1661LI possède alors les gènes *lacI<sup>q</sup>* et *immE3* au niveau chromosomique, qui assurent respectivement la répression du gène *relF* et la stabilisation du gène *colE3*. La quantité de protéines ImmE3 exprimée est suffisante pour permettre la stabilisation du plasmide contenant le gène *colE3*. Cette souche est capable de métaboliser l'arabinose, qui n'est donc plus disponible pour l'induction. D'où la nécessité de muter la souche *E. coli* 1661LI dans son opéron arabinose. L'absence de métabolisation de l'arabinose, liée à l'inactivation de l'opéron arabinose, améliore d'un ordre de grandeur l'efficacité tueuse liée à l'expression du gène *colE3* de la souche 1661LIΔara. Les zones de recombinaisons choisies sont celles permettant la recombinaison du gène *narG*. Cependant, *narG* et *narH* étant présents sur la souche réceptrice *E. coli* 1661LI, une étape supplémentaire de délétion de ces gènes a été réalisée. La souche *E. coli* après toutes ces modifications génétiques réalisées, est nommée 1661LIΔaraΔnarG.

### 3.3. Construction et efficacité du système suicide final

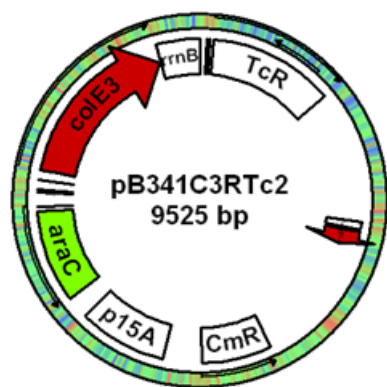
#### 3.3.1. Construction de la cassette toxique inductible à deux gènes

Après des essais de plusieurs stratégies, le clonage des deux gènes toxiques a été réalisé successivement sur le plasmide pB341, dérivé du plasmide pBAD33, qui possède les gènes



*cat* codant la résistance à l'antibiotique chloramphénicol. Dans un premier temps la cassette [*araC*-P<sub>bad</sub>::*colE3*-rrnB] a été amplifiée par le couple d'amorces Pbc<sub>colE3</sub>notIFw5 / Pbc<sub>colE3</sub>ndeIRv5 à partir du plasmide pB33C3. Le site de restriction *NotI* a été ajouté sur l'amorce sens (par modification nucléotidique) afin d'obtenir un clonage orienté des produits PCR dans le plasmide pB341. Le plasmide pB341C3 obtenu a alors été transformé dans la souche 1661LIΔara par électroporation, afin de tester l'efficacité du gène *colE3*. Ce plasmide montre une activité toxique en présence d'arabine, le gène *colE3* est donc fonctionnel. Dans un deuxième temps la cassette [P<sub>AI-03/04</sub>::*relF*-*tc*<sup>r</sup>] contenant les gènes *relF* et le gène de résistance à la tétracycline, a été amplifiée par le couple d'amorces ParelFTcAscIFw1 / ParelFTcascIRv2 à partir du plasmide pSK360. Le site de restriction a été ajouté sur l'amorce sens, puis le produit PCR a été cloné dans le plasmide pB341C3. Le plasmide pB341CRTc1 potentiellement obtenu a été transféré dans 1661LIΔara et mis en présence d'IPTG et de tétracycline afin de vérifier son phénotype. Aucun transformant n'a été obtenu. Aucun clone recombinant n'a cependant pu être détecté. Ceci peut être expliqué par le phénotype de la souche réceptrice utilisée 1661LIΔara. Cette dernière en effet n'est pas mutée dans le gène *hsdR*, dont les produits d'expression interviennent dans la restriction de l'ADN. Cette souche dégraderait donc l'ADN ne portant pas la méthylation adéquate, tels les produits PCR (Tock & Dryden, 2005).

La souche *E. coli* DH5αZ1 utilisée lors de cette dernière opération de clonage est quant à elle mutée dans son système de restriction. Les ADN non méthylés entrant dans la cellule ne sont donc pas dégradés, et des clones présentant les génotypes plasmidiques et les phénotypes attendus ont pu être isolés, confirmant ainsi la validité de l'hypothèse. Le système de modification de la souche *E. coli* DH5αZ1 n'est quant à lui pas muté. Il a donc fallu passer par le clonage de la cassette [PA1-O3/O4::*relF*-*tc*<sup>r</sup>] dans le plasmide pB341 puis transformer le plasmide obtenu dans la souche *E. coli* DH5αZ1. Le plasmide pB341RTc extrait de cette souche est donc méthylé, par conséquent la cassette [PA1-O3/O4::*relF*-*tc*<sup>r</sup>] issue de ce plasmide présente le même profil de méthylation. Le plasmide pB341RTc a été extrait de *E. coli* DH5αZ1 puis digéré par deux enzymes de restriction, *SgsI* et *NcoI* pour lesquelles 2 et un seul sites de restriction sont présents sur la séquence. Trois fragments d'ADN sont ainsi générés dont un contenant le gène d'intérêt *relF*. Le fragment contenant la cassette [ParelF-Tc] a alors été purifié sur gel puis cloné dans pB341C3, lui-même digéré par *SgsI* et déphosphorylé.



**Figure 8.** Représentation du plasmide pB341C3RTc2 obtenue à l'aide du logiciel pDRAW

Le plasmide pB341C3RTc2 (**Figure 8**) obtenu comporte une cassette mobilisable constituée de 2 gènes toxiques inductibles qui entraînent la mort des cellules hôtes après induction par l'IPTG, ainsi que par l'arabinose. Ces résultats montrent que les gènes *relF* et *colE3* fonctionnels ont pu être clonés avec succès sur un même plasmide. L'obtention de transformants dans la souche *E. coli* 1661LIΔara possédant pB341C3 malgré la possible restriction des produits PCR peut être expliquée par une saturation du système de restriction / modification via la pénétration d'une grande quantité d'ADN (Nielsen *et al.*, 1998).

### 3.3.2. Efficacité du système toxique

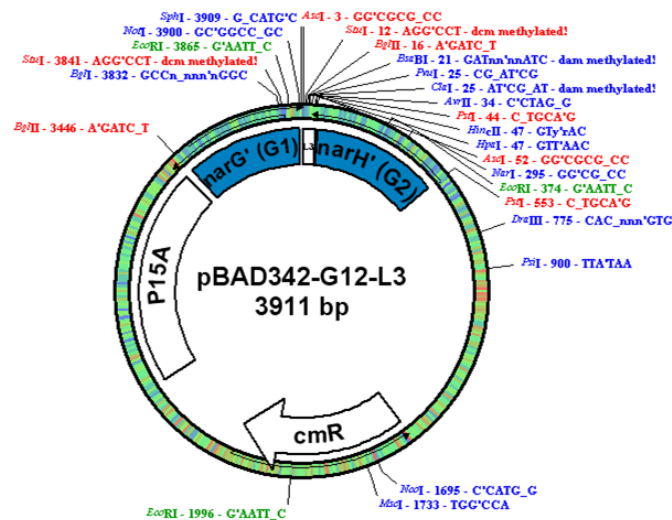
Le test d'efficacité de la construction toxique portée par le plasmide pB342C3RTc2 dans *E. coli* 1661LIΔara par induction simultanée de l'expression des deux gènes toxiques indique un taux d'échappement de  $10^{-7}$ . Le clonage de deux gènes toxiques sur le même plasmide a permis un gain d'efficacité, mais ce gain est inférieur à ce qui était attendu. La construction est seulement 10 fois plus efficace que la construction ne possédant qu'un seul gène toxique, alors qu'un taux d'échappement de  $10^{-8}$  à  $10^{-11}$  était attendu. Le test d'efficacité réalisé par induction de chacun des gènes toxiques de manière individuelle indique un taux d'échappement de  $10^{-6}$  pour le gène *colE3* et de seulement  $10^{-3}$  pour le gène *relF*. La cassette [PA1-O3/O4::*relF-tcr*] dans le plasmide pB341RTc présentait une efficacité de  $10^{-5}$ . Cette perte d'activité peut être liée soit au vecteur pB341C3, soit à la souche hôte utilisée. Cette perte d'activité serait liée au gène *lacI<sup>f</sup>* inséré dans le chromosome d'*E. coli* 1661LIΔara. Pour vérifier cette hypothèse l'efficacité du plasmide pB341RTc contenant seulement le gène *relF* a été testée dans *E. coli* 1661LIΔara d'une part et dans la souche *E. coli* DH5αZ1 contenant un gène *lacI<sup>f</sup>* chromosomique fonctionnel d'autre part. Le plasmide pB341RTc a donc été transformé dans les 2 souches d'*E. coli*. Le taux d'échappement en présence d'IPTG est de  $10^{-6}$  avec la souche DH5αZ1 et de  $10^{-2}$  avec la souche 1661LIΔara.

Au cours des étapes d'intégration chromosomique, de transduction, et de modification chromosomique de la souche réceptrice, des événements ont pu conduire à la mutation du gène *lacI<sup>q</sup>* inséré et consécutivement à la production d'une plus faible quantité de protéines que celle attendue. Le plasmide pB341RTc a donc été transféré dans la souche TELI qui a permis de construire la souche 1661LI à partir de laquelle la souche 1661LIΔara a été réalisée. Le taux d'échappement en présence d'IPTG est en moyenne de  $10^{-5}$ , pour les deux souches, ce qui vérifie l'hypothèse de mutations survenues au niveau du gène *lacI<sup>q</sup>* de la souche 1661LI/1661LIΔara. Après plusieurs essais infructueux pour améliorer l'efficacité du gène *relF*, nous avons sélectionné un nouveau clone de la souche 1661LI possédant un gène *lacI<sup>q</sup>* fonctionnel, qui a été à son tour muté au niveau de l'opéron arabinose pour devenir la souche 1661LIΔara avant que plasmide pB341C3RTc2 soit transféré dans cette nouvelle souche 1661LIΔara. Le test d'efficacité réalisé par induction de chacun des gènes toxiques de manière individuelle indique un taux d'échappement de  $10^{-6}$  pour le gène *colE3* et de seulement  $10^{-5}$  pour le gène *relF*. Le test d'efficacité de la construction toxique portée par pB341C3RTc2 dans *E. coli* 1661LIΔara par induction simultanée de l'expression des deux gènes toxiques indique un taux d'échappement de  $10^{-9}$  et peut même atteindre dans certains cas  $10^{-10}$ .

En 2007, une cassette toxique inductible contenant deux gènes toxiques différents a également été construite (Nielsen *et al.*, 2007).. Ces derniers ont utilisé le gène *sacB* et le gène *hok*. Le gène *hok* présente un mode d'action similaire au gène *relF* (Molin *et al.*, 1993). Le gène *sacB* exprime une enzyme responsable d'une accumulation de levane dans le périplasme de la cellule hôte, menant à la lyse cellulaire (Recorbet *et al.*, 1993). L'ajout de saccharose étant nécessaire à l'activité toxique de *sacB*, les problèmes d'expression basale sont évités. Cependant, un examen de la littérature indique une efficacité de *sacB* équivalente voire inférieure à celle de *relF* ou *colE3*. Dans la construction toxique réalisée par ces auteurs (Harms *et al.*, 2007), le gène *sacB* a un taux d'échappement de  $10^{-6}$ , et le gène *hok* de  $10^{-5}$ , le taux d'échappement de chacun des gènes sont équivalents à ceux obtenus pour *colE3* et *relF*, mais le taux d'échappement de leur construction sous induction simultanée des deux gènes est de  $10^{-8}$  (Harms *et al.*, 2007). Le plasmide pB341C3RTc2 permet d'obtenir une construction efficace que l'on pourra mobiliser aisément dans un plasmide contenant de part et d'autre des séquences dévolues à l'initiation de la recombinaison.

### 3.3.3. Clonage de la cassette toxique dans pBAD342G1G2

Dans cette étude les zones de recombinaison utilisées sont narG1 et narG2 ciblant les gènes intervenant dans la réduction des nitrates. Le plasmide pBAD342G12 contenant les zones de recombinaison narG1 et narG2 a été construit simultanément au développement du plasmide contenant la cassette suicide, afin de pouvoir y cloner la cassette toxique [*araC*-P<sub>bad</sub>::*colE3*, PA1-O3/O4::*relF*-*tc*r] ré-extraite du plasmide xxx. Un site multiple de clonage a donc été intégré, puis digéré par *Asc*I, pour être cloné dans le plasmide pBAD342G12 entre les deux zones de recombinaisons, lui-même digéré par *Asc*I et déphosphorylé.



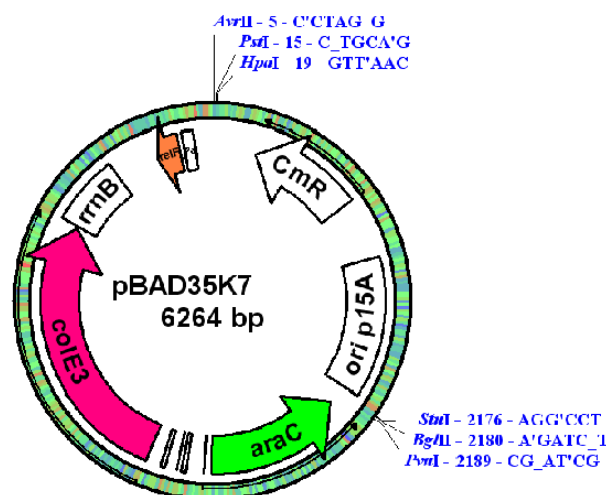
**Figure 9.** Représentation du plasmide pBAD342G12L3 contenant le site multiple de clonage entouré des zones de recombinaison narG1 et narG2.

Le plasmide obtenu pBAD342G12L3 (**Figure 9**) a été traité par les différentes enzymes de restriction présentes sur le site multiple de clonage inséré entre les zones dévolues à la recombinaison, les profils de digestion confirmant la présence de chaque site de restriction choisi. L'amplification de la cassette toxique [*araC*-P<sub>bad</sub>::*colE3*, PA1-O3/O4::*relF*-*tc*r] à partir du plasmide pB342C3RTc2 et à l'aide d'amorces contenant des sites de restriction pour permettre un clonage orienté n'a pas fonctionné. Ce qui peut notamment s'expliquer par la longueur importante du fragment à amplifier de 6616 pb. Pour réduire la taille de la cassette toxique finale une autre stratégie a été adoptée. Afin de minimiser la taille de la cassette toxique mobilisable le gène de résistance à la tétracycline ne sera pas amplifié et deux étapes d'amplification utilisées. Une première étape PCR permet l'amplification de la cassette [*araC*-P<sub>bad</sub>::*colE3*] puis une seconde celle de la cassette [PA1-O3/O4::*relF*] dépourvue de

gène de résistance à la tétracycline, les deux amplifications étant réalisées à partir de l'ADN du plasmide pB341C3RTc2 comme matrice. L'ajout de sites de restriction au niveau des amorces a permis le clonage des deux produits PCR au niveau du site multiple de clonage. Mais des difficultés au niveau de l'étape de PCR pouvant être due à l'ajout des sites de restriction sur l'extrémité 3' des amorces n'ont cependant pas permis l'obtention des produits PCR attendus.

### 3.3.4. Clonage de la cassette toxique dans le plasmide pBAD35

Les deux fragments [*araC*-P<sub>bad</sub>::*colE3*] et [PA1-O3/O4::*relF*] ont été soumis à amplification PCR avec les mêmes amorces dépourvues des sites de restriction (sans ajout par rapport à l'étape précédente) et à partir du plasmide pB342C3RTc comme matrice. Ces deux fragments ont pu alors être amplifiés et ont été clonés en bouts francs dans le plasmide pBAD35, pour donner le plasmide pBAD35K7 (Figure 10).

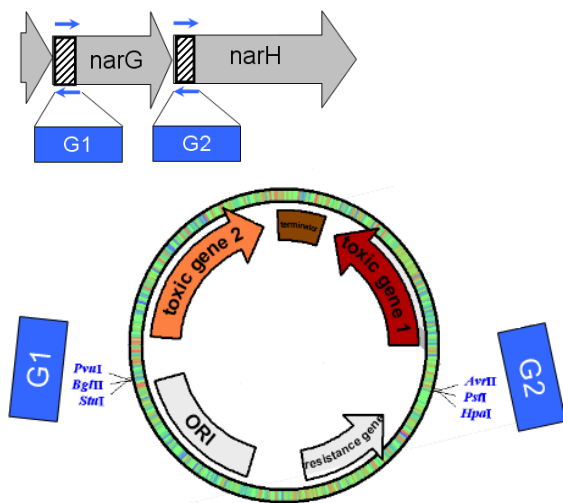


**Figure 10.** Représentation du plasmide pBAD35K7 obtenu à l'aide du logiciel pDRAW, d'après sa séquence nucléotidique

La cassette toxique est flanquée de deux sites multiples de clonage. La souche 1661LIΔaraΔnarG mutée *recD* est très recombino-gène et donc très instable au niveau de son contenu génétique. Pour palier à cette instabilité, toutes les souches utilisées dans les étapes décrites ci-dessous ont alors été choisies avec un génotype *recD*+. Le test d'efficacité de la construction toxique portée par pBAD35K7 dans *E. coli* 1661LIΔaraΔnarG par induction simultanée de l'expression des deux gènes toxiques indique un taux d'échappement variant de  $10^{-8}$  à  $10^{-9}$ . Le test d'efficacité réalisé par induction de chacun des gènes toxiques de manière

individuelle indique un taux d'échappement de  $10^{-6}/10^{-7}$  pour le gène *colE3* et de seulement  $10^{-5}$  pour le gène *relF*.

Nous nous sommes cependant rendus compte que la souche identifiée 1661LIΔaraΔnarG utilisée ne comportait pas la délétion des gènes *narG* et *narH*, leur présence entraînant des difficultés d'instabilité génique au moment du clonage des zones dévolues à la recombinaison narG1 et narG2 qui présentaient de fortes similarités de séquences. C'est la raison qui nous a conduits à utiliser la souche *E. coli* TELIΔaraΔnarG lors des étapes ultérieures de clonage. Les zones de recombinaisons narG1 et narG2 préalablement définies ont été amplifiées puis clonées au niveau des SMC du plasmide pBAD35K7, le plasmide obtenu étant nommé pB35K7G1G2 (**Figure 11**). Les taux d'échappement obtenus avec cette construction ont atteints  $10^{-7}$  pour le gène *colE3* et  $10^{-4}$  pour le gène *relF*. Le gène *relF* montre une perte d'efficacité d'un ordre de grandeur, due soit aux caractéristiques génétiques de la nouvelle souche TELIΔaraΔnarG soit à l'insertion des zones de recombinaison narG1 et narG2.



**Figure 11.** Clonages des zones de recombinaison G1 et G2 au niveau des SMC dans le plasmide pBAD35K7. Deux régions conservées ont été choisies : G1 se situant au début du gène *narG*, et G2 au début du gène *narH*.

## 4. Conclusions et perspectives

Un plasmide portant une cassette toxique constituée de deux gènes toxiques différents sous le contrôle de systèmes régulateurs indépendants, entourée de deux zones de recombinaisons a pu être construit, après plusieurs stratégies de clonage.

La cassette toxique comporte les gènes *colE3* et *relF*, régulés par les protéines LacI et AraC, respectivement. L'efficacité obtenue à partir de cette dernière construction, varie autour de  $10^{-8}$  à  $10^{-9}$ , lorsque les deux gènes sont activés simultanément. La combinaison des deux gènes toxiques a bien permis l'augmentation de l'efficacité de la cassette.

L'utilisation de cette cassette dans le système de contre sélection devrait ainsi permettre d'isoler des cellules bactériennes recombinées à une fréquence de  $10^{-8}$ .

Les perspectives quant à la suite des travaux, seront de valider l'outil en testant la souche « piègeuse » d'ADN, dans un premier temps *in vivo*, dans des conditions optimales en présence d'ADN cible connu, pour tester l'efficacité de transformation-recombinaison de la souche réceptrice et l'efficacité du système de sélection des souches ayant intégré l'ADN cible. Le plasmide lambda red gam pourra alors être intégré et activé dans la souche réceptrice, pour pouvoir augmenter ses capacités de recombinaison. Puis des tests identiques seront réalisés en présence d'un mélange d'ADN contenant notamment l'ADN cible (témoin positif), pour tester la limite de détection du système. La validation ultime se fera alors *in situ*, dans des conditions réelles, l'outil pourra alors être appliqué.

Ce projet innovant a mobilisé plusieurs personnes, il est dans la continuité des travaux de Nathalie Lombard (ancienne doctorante), ma contribution s'est située au niveau de la construction de la cassette toxique entourées des zones de recombinaison, d'autres personnes ont pris la suite des travaux, pour la partie validation de l'outil dans un premier temps *in vitro* puis *in situ*.

## Conclusions générales

La diversité microbienne représente un réservoir génétique immense. L'arrivée de la métagénomique a révolutionné la microbiologie permettant en théorie d'accéder directement à l'ADN de toutes les bactéries d'un échantillon, le métagénome. Malgré cette grande avancée, différents biais technologiques, au niveau des étapes d'extraction, d'amplification de l'ADN ou du fait des limites de sensibilité et de spécificité des différentes techniques analytiques limitent l'accès à la totalité du potentiel génétique présent dans un échantillon. De fait, ces biais limitent considérablement l'accès aux très nombreuses populations à effectifs minoritaires présentes dans les communautés bactériennes des sols. Les caractéristiques physicochimiques des sols contribuent à en faire des milieux extrêmement complexes, hétérogènes très favorables à l'éclosion d'un très haut niveau de diversité bactérienne mais aussi particulièrement délicats à étudier. C'est pourquoi, plus d'un demi-siècle après sa formulation le concept de Baas Becking sur le cosmopolitisme bactérien que résume le « Tout est partout ET l'environnement sélectionne » demeure ni confirmé ni infirmé. Depuis le début des années 90 l'application de différentes technologies basées sur l'ADN extrait de l'environnement a même sérieusement contribué à brouiller les estimations du nombre d'espèces colonisant les échantillons de sol étudiés puisque celles-ci varient de plus de 3 ordres de grandeur selon les méthodes utilisées. A ces incertitudes méthodologiques doivent être ajoutées celles liées à la notion d'espèce bactérienne et la diversité intraspécifique qui se traduit par des différences entre souches pouvant atteindre 30% du génome ce qui complique encore la situation d'inventoriage de la biodiversité bactérienne. Enfin, la nature extrêmement dynamique des génomes bactériens, continuellement remodelés par des modifications, des acquisitions, des délétions met en exergue que tout inventaire aussi exhaustif puisse-t-il être d'un échantillon environnemental ne révélera qu'un instantané d'une structure génomique des communautés en dehors même de toute prise en compte des variations dans la taille en effectifs des populations conditionnées par les facteurs biotiques et abiotiques des milieux.

Ces données soulignent la nécessité absolue qu'il y avait à l'initiation de ce travail de doctorat d'aborder cette récurrente question du niveau réel de la biodiversité bactérienne par des travaux novateurs et originaux. La nécessité en était d'abord d'ordre fondamental pour répondre au besoin naturel de l'homme de connaître et inventorier son milieu. Contrairement à l'exploration du cosmos et à sa cartographie en galaxies, étoiles, planètes et autres objets célestes l'étude des bactéries du sol ne se limite pas aux seules descriptions d'écosystèmes et



acquisitions de connaissances théoriques. Elle revêt au contraire un caractère mécanistique, les capacités d'évolution, d'adaptation des communautés bactériennes influant considérablement sur leurs activités qui, comme on l'a vu, sont essentielles à la vie sur notre planète. Cette diversité bactérienne est également un réservoir quasi-inépuisable de gènes et opérons codant des fonctions que l'homme exploite déjà largement pour son agriculture, son alimentation, pour réparer les dégâts qu'il cause à son environnement. C'est dans ce réservoir qu'il peut espérer trouver les nouveaux antibiotiques qui lui permettront de lutter contre les maladies infectieuses qui, du fait de la dispersion des résistances aux antibiotiques parmi les organismes pathogènes recommencent à menacer l'humanité.

L'analyse des approches utilisées depuis plus d'une vingtaine d'années pour répondre à ces objectifs, les écueils rencontrés que nous avons mentionnés plus haut mais aussi les perspectives qu'offrent des technologies de séquençage de plus en plus performantes et bon marché ont conforté le constat que nous avons dressé que pour tendre vers l'exhaustivité de l'inventoriage, l'utilisation des nouvelles technologies se devait d'être obligatoirement couplée à des approches conceptuelles novatrices.

Notre travail a donc consisté à développer de nouvelles stratégies pour tenter d'accéder à des populations récalcitrantes du fait de leurs faibles effectifs, de propriétés écologiques ou physiologiques caractérisées par une localisation dans des microenvironnements protégés ou une résistance particulièrement élevée aux traitements d'extraction ou de lyse. Ces stratégies ont été basées sur deux approches distinctes mais qui se veulent complémentaires pour en maximiser l'efficacité.

La première approche était basée sur la ré-inoculation de communautés bactériennes dans de nouveaux environnements rendus stériles qui nous a permis de constater dans un premier temps que les structures des communautés qui en résultaient dans les différents sols se différenciaient entre elles et de celles de leur environnement d'origine. Nos travaux ont ainsi pu mettre en évidence le rôle combiné dans la structuration de ces nouvelles communautés de l'environnement abiotique et de l'inoculum bactérien utilisé. Nos objectifs ne se limitaient cependant pas à ce constat mais voulaient, grâce à une technologie à la fois sensible et applicable à de nombreux échantillons identifier les taxons responsables des modifications de structure afin de déterminer si, parmi ceux-ci certains qui n'auraient pas été détectés initialement appartiendraient de fait à cette catégorie des « récalcitrants » et pourraient être des membres de cette « rare biosphere ».

L'analyse des genres bactériens dans les sols inoculés grâce à la technique d'hybridation sur puces à ADN taxonomique de produits PCR amplifiés à partir de l'ADN bactérien issu du sol a effectivement permis de mettre en évidence des genres qui n'avaient pas été détectés dans l'inoculum de départ. Ces résultats confirment l'intérêt de cette approche dans laquelle la modification des caractéristiques abiotiques à laquelle est soumise une communauté bactérienne brasse en profondeur les équilibres biotiques modifiant négativement pour certaines, positivement pour d'autres leur valeur adaptative.

Ces résultats confirment tout l'intérêt du concept formulé en son temps par Baas Becking « tout est partout », même s'ils ne permettent en l'état des recherches ni de définitivement le confirmer ni de l'infirmer. Ils ouvrent des pistes tant conceptuelles que technologiques pour poursuivre les recherches sur une telle base. Nos travaux se sont majoritairement restreints aux sols tant pour en extraire des inocula bactériens que pour s'en servir comme matrices axéniques pour les réensemencements ; les résultats ont été à la hauteur de l'important travail nécessaire pour réaliser ces expérimentations laborieuses et longues en confirmant la présence d'une « rare biosphere » qu'il est maintenant possible en partie de révéler et d'identifier. La seule expérience impliquant une communauté bactérienne issue d'un environnement aquatique a montré que l'utilisation d'un spectre beaucoup plus large d'environnements pour la préparation d'inocula et de matrices axéniques serait particulièrement pertinente pour continuer l'exploration de la diversité bactérienne. Le nombre des écosystèmes différents sur notre planète est suffisamment important, des différents environnements aquatiques aux types de sols les plus variés incluant ceux aux conditions extrêmes mais aussi tous ceux dans lesquels bactéries et archées vivent en étroite relation avec d'autres organismes, plantes et animaux pour que les conditions expérimentales ainsi créées soient en mesure de révéler toute la diversité cachée et en déterminer ainsi le niveau réel. De tels travaux devraient bénéficier en parallèle du développement des dernières technologies de séquençage, que celles-ci soient appliquées à l'ADN métagénomique total ou restreint aux séquences du gène 16S rDNA sur un principe comparable à celui utilisé dans nos travaux avec les puces à ADN taxonomiques. La sensibilité et la spécificité de ces techniques devraient ainsi permettre pour chaque échantillon étudié d'affiner le niveau de l'analyse tant quantitativement par le nombre de taxons détectés et identifiés que qualitativement en complétant le niveau populationnel de l'étude par son niveau fonctionnel.

Il convient cependant d'être conscient que le biais de l'extraction des cellules et de leur ADN peut limiter l'accès à cette rare biosphere en dépit des améliorations proposées ci-dessus.

C'est alors que pourrait intervenir la seconde stratégie abordée lors de cette thèse qui a pour but de contourner les biais liés à l'extraction en capturant directement l'ADN *in situ*. Nos efforts se sont portés sur le développement d'une souche bactérienne réceptrice capable de piéger dans le sol des fragments d'ADN ciblés provenant du génome des bactéries telluriques. Nous avons montré ici tous les développements réalisés pour construire cette souche bactérienne depuis la construction du système de contre sélection positive réalisée en collaboration forte avec d'autres chercheurs de l'équipe (thèse Nathalie Lombard). Ce système est constitué de deux gènes toxiques inductibles, flanqués de deux régions d'ADN, ici *narG1* et *narG2* dévolues à amorcer la recombinaison avec les séquences correspondantes dans l'ADN pénétrant. Dans cette étude préliminaire notre choix s'est en effet porté sur le gène *narG* intervenant dans la réduction des nitrates du fait de sa forte représentation dans les bactéries telluriques ce qui en facilitera sa détection dans les tests ultérieurs. Nous cherchons à ce que l'intégration du fragment d'ADN ciblé, conduise après un double événement de recombinaison à l'élimination du système de sélection, permettant sous conditions d'induction, que les bactéries comportant le fragment d'intérêt puissent être sélectionnées et développer des colonies sur milieu de culture. Après plusieurs essais, l'efficacité du système de sélection obtenu est conforme à nos attentes et permettra son utilisation dans les conditions visées pour son développement. Les travaux entrepris avec les autres chercheurs de l'équipe impliqués dans ce projet (Nathalie Lombard, Laure Franqueville, Samuel Jacquot...) visaient à mettre en place le système de recombinaison le plus efficace possible pour intégrer dans le génome de la souche capture les régions d'ADN du métagénome ciblées par les séquences dévolues à cet objectif, clonées de part et d'autre des gènes létiaux. Nous avons présenté dans le chapitre III tout l'intérêt du système basé sur le système lambda red gam, venant après des tentatives avec une souche dont le gène *recD* avait été inactivé. Les travaux en cours s'appliquent à définir le potentiel de ce mécanisme pour les applications de capture d'ADN qui ont motivé son développement. Les objectifs étant d'accéder à une diversité allélique pouvant s'étendre jusqu'aux zones dévolues à la recombinaison nos expérimentations visent à déterminer les conditions de recombinaison supportant les plus forts taux de divergence entre séquences cibles et ADN transformant ainsi que les tailles optimales et minimales de ces régions d'ADN sachant que pour beaucoup de gènes bactériens orthologues les zones consensus conservées entre bactéries sont de taille réduite. La série d'expérimentations suivante cherchera à déterminer le rapport entre fréquences de transformation-recombinaison et taille de la région d'ADN à intégrer, considérant que plus longue pourra être la région

d'ADN acquise plus informative en sera l'analyse de sa séquence. Enfin, le potentiel de l'outil à « pêcher » les régions d'ADN au sein d'une solution d'ADN de complexité croissante sera déterminé d'abord en conditions *in vitro* puis *in situ* afin de définir le seuil populationnel au sein d'une communauté pour que le ou les gènes cibles spécifiques de cette population puissent en théorie être capturés. Une fois tous ces contrôles réalisés et si le potentiel déterminé *in vitro* est conforme aux attentes en termes de spécificité et de sensibilité, la souche bactérienne construite sera utilisée *in situ*, équipée selon les objectifs de chaque projet des régions de capture adaptées comptant que le principe de contournement des biais liés à l'extraction de l'ADN permette d'atteindre les gènes de populations non accessibles par les approches métagénomiques classiques, y compris celle basée sur une ré-inoculation d'environnements axéniques par des inocula bactériens complexes. La possibilité de réaliser *in situ* l'électro-transformation de la souche bactérienne capturee inoculée dans des échantillons de sol grâce aux générateurs « haute tension » du laboratoire Ampère complète la gamme des outils disponibles pour la mise en pratique de cette nouvelle approche de métagénomique complémentaire de celles déjà développées considérant que seule leur combinaison permettra de répondre aux questions qui justifient ce travail de thèse : Quel est le niveau de la diversité bactérienne tellurique et est-ce que tout est vraiment partout ?



## Références bibliographiques

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

Ahrenholtz I, Lorenz MG & Wackernagel W. 1994. A conditional suicide system in *Escherichia coli* based on the intracellular degradation of DNA. *Appl Environ Microbiol*. **60**: 3746-3751.

Aizenman E, Engelberg-Kulka H & Glaser G. 1996. An *Escherichia coli* chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci U S A*. **93**: 6059-6063.

Albuquerque L, Simoes C, Nobre MF, *et al.* 2005. *Truepera radiovictrix* gen. nov., sp. nov., a new radiation resistant species and the proposal of Trueperaceae fam. nov. *FEMS Microbiol Lett*. **247**: 161-169.

Alexander M. 1977. Introduction to soil microbiology, second edn. *United States of America: John Wiley & Sons, Inc.*

Amundsen SK, Taylor AF, Chaudhury AM & Smith GR. 1986. recD: the gene for an essential third subunit of exonuclease V. *Proc Natl Acad Sci USA*. **83**: 5558-5562.

Anderson I, Ulrich LE, Lupa B, *et al.* 2009. Genomic characterization of methanomicrobiales reveals three classes of methanogens. *PLoS One*. **4**: 5797.

Arber W. 2000. Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev*. **24**: 1-7.

Artursson V, Finlay RD & Jansson JK. 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.

Ashby MN, Rine J, Mongodin EF, Nelson KE & Dimster-Denk D. 2007. Serial analyses of rRNA genes and the unexpected dominance of rare members of microbial communities. *Appl Environ Microbiol*. **73**: 4532-4542.

Ashelford KE, Day MJ & Fry JC. 2003. Elevated abundance of bacteriophage infecting bacteria in soil. *Appl Environ Microbiol*. **69**: 285-289.

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

Baas Becking LGM. 1934. Geobiologie of inleiding tot de milieukunde. *The Hague, the Netherlands: W. P. Van Stockum & Zoon (in Dutch)*.

Battista JM, Earl AM, Sandler SJ. 2004. Microbial evolution : gene establishment, survival and exchange. *ASM Press*.

Baver LD. 1961. Soil Physics. *John Wiley and Sons, New York*.

Beatty JT, Overmann J, Lince MT, *et al.* 2005. An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. *Proc Natl Acad Sci USA*. **102**: 9306-9310.

Beijerinck MW. 1913. De infusies en de ontdekking der bacterien. . *Jaarboek van de Koninklijke Akademie voor Wetenschappen. Amsterdam, the Netherlands: Muller*.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA & Wheeler DL. 2002. GenBank. *Nucleic Acids Res*. **30**: 17-20.

Bent SJ & Forney LJ. 2008. The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *Isme J*. **2**: 689-695.

Berka RM, Hahn J, Albano M, *et al.* 2002. Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Mol Microbiol*. **43**: 1331-1345.

Best M (1995) Classification of rocks and their abundances on the earth. In rocks physics and phase relation: a handbook of physical contents. 1-7.

Bodrossy L & Sessitsch A. 2004. Oligonucleotide microarrays in microbial diagnostics. *Curr Opin Microbiol*. **7**: 245-254.

Borneman J & Triplett EW. 1997. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl Environ Microbiol*. **63**: 2647-2653.

Borneman J, Skroch PW, O'Sullivan KM, *et al.* 1996. Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl Environ Microbiol*. **62**: 1935-1943.

Braslavsky I, Hebert B, Kartalov E & Quake SR. 2003. Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci U S A*. **100**: 3960-3964.

Bridges BA. 2005. Error-prone DNA repair and translesion DNA synthesis. II: The inducible SOS hypothesis. *DNA Repair (Amst)*. **4**: 725-726, 739.

Bruce KD, Hiorns WD, Hobman JL, Osborn AM, Strike P & Ritchie DA. 1992. Amplification of DNA from native populations of soil bacteria by using the polymerase chain reaction. *Appl Environ Microbiol*. **58**: 3413-3416.

Brussow H, Canchaya C & Hardt WD. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev*. **68**: 560-602.

Bryant DA & Frigaard NU. 2006. Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol*. **14**: 488-496.

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.

Canchaya C, Proux C, Fournous G, Bruttin A & Brussow H. 2003. Prophage genomics. *Microbiol Mol Biol Rev*. **67**: 238-276.

Caron DA. 1983. Technique for Enumeration of Heterotrophic and Phototrophic Nanoplankton, Using Epifluorescence Microscopy, and Comparison with Other Procedures. *Appl Environ Microbiol*. **46**: 491-498.

Ceremonie H, Buret F, Simonet P & Vogel TM. 2004. Isolation of lightning-competent soil bacteria. *Appl Environ Microbiol*. **70**: 6342-6346.

Chaudhury AM & Smith GR. 1984. A new class of Escherichia coli recBC mutants: implications for the role of RecBC enzyme in homologous recombination. *Proc Natl Acad Sci USA*. **81**: 7850-7854.

Chaveroche MK, Ghigo JM & d'Enfert C. 2000. A rapid method for efficient gene replacement in the filamentous fungus Aspergillus nidulans. *Nucleic Acids Res*. **28**: E97.

Chen I & Dubnau D. 2004. DNA uptake during bacterial transformation. *Nat Rev Microbiol*. **2**: 241-249.

Chenu C. 1993. Clay or sand polysaccharide associations as models for the interface between micro-organisms and soil: water related properties and microstructure. *Geoderma*. **56**.

Cho JC & Tiedje JM. 2000. Biogeography and degree of endemism of fluorescent Pseudomonas strains in soil. *Appl Environ Microbiol*. **66**: 5448-5456.

Cole JR, Chai B, Farris RJ, *et al.* 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res*. **33**: D294-296.

Copeland NG, Jenkins NA & Court DL. 2001. Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet*. **2**: 769-779.

Costa R, Gotz M, Mrozek 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.

Courcelle J, Khodursky A, Peter B, Brown PO & Hanawalt PC. 2001. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli. *Genetics*. **158**: 41-64.

Dabert P & Smith GR. 1997. Gene replacement with linear DNA fragments in wild-type Escherichia coli: enhancement by Chi sites. *Genetics*. **145**: 877-889.

Datsenko KA & Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA*. **97**: 6640-6645.



- Datta S, Costantino N & Court DL. 2006. A set of recombineering plasmids for gram-negative bacteria. *Gene*. **379**: 109-115.
- de Vries J & Wackernagel W. 2002. Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc Natl Acad Sci USA*. **99**: 2094-2099.
- de Wit R & Bouvier T. 2006. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environ Microbiol*. **8**: 755-758.
- Delmas S & Matic I. 2005. Cellular response to horizontally transferred DNA in *Escherichia coli* is tuned by DNA repair systems. *DNA Repair (Amst)*. **4**: 221-229.
- Demain AL. 1999. Pharmaceutically active secondary metabolites of microorganisms. *Appl Microbiol Biotechnol*. **52**: 455-463.
- DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM & Andersen GL. 2007. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol*. **53**: 371-383.
- Diaz E, Munthali M, de Lorenzo V & Timmis KN. 1994. Universal barrier to lateral spread of specific genes among microorganisms. *Mol Microbiol*. **13**: 855-861.
- Dillon JG, Fishbain S, Miller SR, Bebout BM, Habicht KS, Webb SM & Stahl DA. 2007. High rates of sulfate reduction in a low-sulfate hot spring microbial mat are driven by a low level of diversity of sulfate-respiring microorganisms. *Appl Environ Microbiol*. **73**: 5218-5226.
- Dobrindt U, Hochhut B, Hentschel U & Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol*. **2**: 414-424.
- Drake JW. 1991. Spontaneous mutation. *Annu Rev Genet*. **25**: 125-146.
- Dressman D, Yan H, Traverso G, Kinzler KW & Vogelstein B. 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci USA*. **100**: 8817-8822.
- Dubnau D. 1999. DNA uptake in bacteria. *Annu Rev Microbiol*. **53**: 217-244.
- Duchaufour P. 1977. Pédogénèse et classification. Paris: Masson S.A.
- Dunbar J, Barns SM, Ticknor LO & Kuske CR. 2002. Empirical and theoretical bacterial diversity in four Arizona soils. *Appl Environ Microbiol*. **68**: 3035-3045.
- El Karoui M, Biaudet V, Schbath S & Gruss A. 1999. Characteristics of Chi distribution on different bacterial genomes. *Res Microbiol*. **150**: 579-587.
- Fall S, Mercier A, Bertolla F, et al. 2007. Horizontal gene transfer regulation in bacteria as a "spandrel" of DNA repair mechanisms. *PLoS One*. **2**: 1055.

- Fedurco M, Romieu A, Williams S, Lawrence I & Turcatti G. 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* **34**: 22.
- Fenchel T & Finlay BJ. 2006. The diversity of microbes: resurgence of the phenotype. *Philos Trans R Soc Lond B Biol Sci.* **361**: 1965-1973.
- Fierer N, Bradford MA & Jackson RB. 2007. Toward an ecological classification of soil bacteria. *Ecology.* **88**: 1354-1364.
- Fisher MM & Triplett EW. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl Environ Microbiol.* **65**: 4630-4636.
- Fortin N, Beaumier D, Lee K & Greer CW. 2004. Soil washing improves the recovery of total community DNA from polluted and high organic content sediments. *J Microbiol Methods.* **56**: 181-191.
- Foster RC. 1988. Microenvironments of soil microorganisms *Biology and fertility of soils* **6**: 189-203.
- Fredrickson JK, Zachara JM, Balkwill DL, *et al.* 2004. Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the hanford site, washington state. *Appl Environ Microbiol.* **70**: 4230-4241.
- Frey J, Bagdasarian, M. 1989. The molecular biology of IncQ plasmids In Thomas CM, ed, Promiscuous plasmids of gram negative bacteria. In Thomas CM, ed, *Promiscuous plasmids of gram negative bacteria*, Academic Press, Orlando, USA. 79-94.
- Frostegard A, Courtois S, Ramisse V, *et al.* 1999. Quantification of bias related to the extraction of DNA directly from soils. *Appl Environ Microbiol.* **65**: 5409-5420.
- Furukawa K. 2003. 'Super bugs' for bioremediation. *Trends Biotechnol.* **21**: 187-190.
- Gans J, Wolinsky M & Dunbar J. 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science.* **309**: 1387-1390.
- Gebhard F & Smalla K. 1999. Monitoring field releases of genetically modified sugar beets for persistence of transgenic plant DNA and horizontal gene transfer. *FEMS Microbiol Ecol.* **28**: 261-272.
- Gelsomino A, Keijzer-Wolters AC, Cacco G & van Elsas JD. 1999. Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electrophoresis. *J Microbiol Methods.* **38**: 1-15.
- Girvan MS, Bullimore J, Pretty JN, Osborn AM & Ball AS. 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.

- Glaser P. 2005. Les puces à ADN vont-elles révolutionner l'identification des bactéries? *medecine sciences* **21**: 539-544
- Glockner FO, Kube M, Bauer M, *et al.* 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci U S A.* **100**: 8298-8303.
- Gobat J, Aragno M & Matthey W. 2003. Le sol vivant, second edn. *Lausanne: Presses polytechniques et universitaires romandes.*
- Grayston SJ, Wang SQ, Campbell CD, Edwards AC. 1998. Selective influence of plant species on microbial diversity in the rizhosphere. *Soil Biol Biochem.* **30**: 369-378.
- Griffiths AJF, Gelbart WM, Miller JH, Lewontin RC. 1999. Modern genetic analysis. *New York: W.H. Freeman and company.* **1**.
- Grozdanov L, Raasch C, Schulze J, Sonnenborn U, Gottschalk G, Hacker J & Dobrindt U. 2004. Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J Bacteriol.* **186**: 5432-5441.
- Grundmann GL. 2004. Spatial scales of soil bacteria diversity - the size of a clone. *FEMS Microbiol Ecol.* **48**: 119-127.
- Grundmann GL & Debouzie D. 2000. Geostatistical analysis of the distribution of NH<sub>4</sub>(+)  
and NO<sub>2</sub>(-)-oxidizing bacteria and serotypes at the millimeter scale along a soil transect. *FEMS Microbiol Ecol.* **34**: 57-62.
- Gupta PK. 2008. Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol.* **26**: 602-611.
- Guzman LM, Belin D, Carson MJ & Beckwith J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol.* **177**: 4121-4130.
- Hanniffy SB, Pelaez C, Martinez-Bartolome MA, Requena T & Martinez-Cuesta MC. 2009. Key enzymes involved in methionine catabolism by cheese lactic acid bacteria. *Int J Food Microbiol.* **135**: 223-230.
- Harms K, de Vries J & Wackernagel W. 2007. A double kill gene cassette for the positive selection of transforming non-selective DNA segments in *Acinetobacter baylyi* BD413. *J Microbiol Methods.* **69**: 107-115.
- Harris TD, Buzby PR, Babcock H, *et al.* 2008. Single-molecule DNA sequencing of a viral genome. *Science.* **320**: 106-109.
- 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 LA, Zwart KB & 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 *Biology and fertility of soils.* **6**: 189-203.
- Heuer H, Fox RE & Top EM. 2007. Frequent conjugative transfer accelerates adaptation of a broad-host-range plasmid to an unfavorable *Pseudomonas putida* host. *FEMS Microbiol Ecol.* **59**: 738-748.
- Hoheisel JD. 2006. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet.* **7**: 200-210.
- Holben WE, Jansson JK, Chelm BK & Tiedje JM. 1988. DNA Probe Method for the Detection of Specific Microorganisms in the Soil Bacterial Community. *Appl Environ Microbiol.* **54**: 703-711.
- Hurt R A, Qiu X, Wu L, Roth Y, Palumbo AV, Tiedje JM & Zhou J. 2001. Simultaneous recovery of RNA and DNA from soils and sediments. *Appl Environ Microbiol.* **67**: 4495-503
- Hyman ED. 1988. A new method of sequencing DNA. *Anal Biochem.* **174**: 423-436.
- Jaenike J, Stahlhut JK, Boelio LM & Unckless RL. 2009. Association between *Wolbachia* and *Spiroplasma* within *Drosophila neotestacea*: an emerging symbiotic mutualism? *Mol Ecol.*
- Janssen PH. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol.* **72**: 1719-1728.
- Johnsen AR & Kroer N. 2007. Effects of stress and other environmental factors on horizontal plasmid transfer assessed by direct quantification of discrete transfer events. *FEMS Microbiol Ecol.* **59**: 718-728.
- Kauffmann IM, Schmitt J & Schmid RD. 2004. DNA isolation from soil samples for cloning in different hosts. *Appl Microbiol Biotechnol.* **64**: 665-670.
- Knudsen S, Saadbye P, Hansen LH, Collier A, Jacobsen BL, Schlundt J & Karlstrom OH. 1995. Development and testing of improved suicide functions for biological containment of bacteria. *Appl Environ Microbiol.* **61**: 985-991.
- Kozubal M, Macur RE, Korf S, Taylor WP, Ackerman GG, Nagy A & Inskeep WP. 2008. Isolation and distribution of a novel iron-oxidizing crenarchaeon from acidic geothermal springs in Yellowstone National Park. *Appl Environ Microbiol.* **74**: 942-949.
- Kumar S & Tuteja U. 2009. Detection of Virulence-Associated Genes in Clinical Isolates of *Bacillus anthracis* by Multiplex PCR and DNA Probes. *J Microbiol Biotechnol.* **19**: 1475-1481.

- Kuske CR, Banton KL, Adorada DL, Stark PC, Hill K & Jackson PJ. 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 FM, Botha A & Prior BA. 2007. Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J Appl Microbiol.* **102**: 265-273.
- Lanzer M & Bujard H. 1988. Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A.* **85**: 8973-8977.
- Lauber CL, 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.*
- Lauber CL, Strickland MS, Bradford MA & 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.
- Lawrence JG. 1999. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr Opin Microbiol.* **2**: 519-523.
- Lee DH, Zo YG & Kim SJ. 1996. Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism. *Appl Environ Microbiol.* **62**: 3112-3120.
- Lee KC, Webb RI, Janssen PH, Sangwan P, Romeo T, Staley JT & Fuerst JA. 2009. Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. *BMC Microbiol.* **9**: 5.
- Lensi R, Lescure C, Steinberg J, Savoie M & Faurie G. 1991. Dynamics of residual enzyme activities, denitrification potential, and physico-chemical properties in a gamma-sterilized soil. *Soil Biol Biochem.* **23**: 367-373.
- Liu WT, Marsh TL, Cheng H & Forney LJ. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol.* **63**: 4516-4522.
- Loy A, Lehner A, Lee N, *et al.* 2002. Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol.* **68**: 5064-5081.
- Lucy M, Reed E & Glick BR. 2004. Applications of free living plant growth-promoting rhizobacteria. *Antonie Van Leeuwenhoek.* **86**: 1-25.
- Ludwig W, Strunk O, Westram R, *et al.* 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**: 1363-1371.

- Lugtenberg B & Kamilova F. 2009. Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol.* **63**: 541-556.
- Maamar H & Dubnau D. 2005. Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol Microbiol.* **56**: 615-624.
- Macia MD, Blanquer D, Togores B, Sauleda J, Perez JL & Oliver A. 2005. Hypermutation is a key factor in development of multiple-antimicrobial resistance in *Pseudomonas aeruginosa* strains causing chronic lung infections. *Antimicrob Agents Chemother.* **49**: 3382-3386.
- Makarova KS, Omelchenko MV, Gaidamakova EK, *et al.* 2007. *Deinococcus geothermalis*: the pool of extreme radiation resistance genes shrinks. *PLoS One.* **2**: 955.
- Margulies M, Egholm M, Altman WE, *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* **437**: 376-380.
- Martinez JL & Baquero F. 2000. Mutation frequencies and antibiotic resistance. *Antimicrob Agents Chemother.* **44**: 1771-1777.
- Martiny JB, Bohannan BJ, Brown JH, *et al.* 2006. Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol.* **4**: 102-112.
- Matic I, Taddei F & Radman M. 1996. Genetic barriers among bacteria. *Trends Microbiol.* **4**: 69-72.
- McHardy AC & Rigoutsos I. 2007. What's in the mix: phylogenetic classification of metagenome sequence samples. *Curr Opin Microbiol.* **10**: 499-503.
- Milkman R. 1997. Recombination and population structure in *Escherichia coli*. *Genetics.* **146**: 745-750.
- Miller DN. 2001. Evaluation of gel filtration resins for the removal of PCR-inhibitory substances from soils and sediments. *J Microbiol Methods.* **44**: 49-58.
- Miller DN, Bryant JE, Madsen EL & Ghiorse WC. 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol.* **65**: 4715-4724.
- Miller JH. 1996. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair. *Annu Rev Microbiol.* **50**: 625-643.
- Molin S, Boe L, Jensen LB, Kristensen CS, Givskov M, Ramos JL & Bej AK. 1993. Suicidal genetic elements and their use in biological containment of bacteria. *Annu Rev Microbiol.* **47**: 139-166.
- Morales SE, Cosart TF, Johnson JV & Holben WE. 2009. Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters. *Appl Environ Microbiol.* **75**: 668-675.

- More MI, Herrick JB, Silva MC, Ghiorse WC & Madsen EL. 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.
- Muller AK, Westergaard K, Christensen S & Sorensen SJ. 2002. The diversity and function of soil microbial communities exposed to different disturbances. *Microb Ecol.* **44**: 49-58.
- Mummey DL & Stahl PD. 2004. Analysis of soil whole and inner microaggregate bacterial communities. *Microb Ecol.* **48** 41-50.
- Murphy KC. 1998. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J Bacteriol.* **180**: 2063-2071.
- Murphy KC. 1991. Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme. *J Bacteriol.* **173**: 5808-5821.
- Murty D, Kirschbaum MFU, McMurtrie RE & McGilvray A. 2002. Does conversion of forest to agricultural land change soil carbon and nitrogen? *Global Change Biology.* **8**: 105-123.
- Nielsen KM, Smalla K & van Elsas JD. 2000. Natural transformation of Acinetobacter sp. strain BD413 with cell lysates of Acinetobacter sp., Pseudomonas fluorescens, and Burkholderia cepacia in soil microcosms. *Appl Environ Microbiol.* **66**: 206-212.
- Nielsen KM, Bones AM, Smalla K & van Elsas JD. 1998. Horizontal gene transfer from transgenic plants to terrestrial bacteria a rare event? *FEMS Microbiol Rev.* **22**: 79-103.
- Nielsen KM, Johnsen PJ, Bensasson D & Daffonchio D. 2007. Release and persistence of extracellular DNA in the environment. *Environ Biosafety Res.* **6**: 37-53.
- Noguez AM, Arita HT, Escalante AE, Forney LJ, Garcia-Oliva F & Souza V. 2005. Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Global Ecol. Biogeogr.* **14**: 241-248.
- Nohmi T. 2006. Environmental stress and lesion-bypass DNA polymerases. *Annu Rev Microbiol.* **60**: 231-253.
- Normand P, Ponsonnet C, Nesme X, Neyra M & Simonet P. 1996. ITS analysis of prokaryotes. In D. L. Akkermans, J. D. van Elsas and E. I. de Bruijn (ed.), *Molecular microbial ecology manual*. Kluwer Academic Publishers, Amsterdam, The Netherlands. 1-12.
- Nunan N, Wu K, Young IM, Crawford JW & Ritz K. 2002. In situ spatial patterns of soil bacterial populations, mapped at multiple scales, in an arable soil. *Microb Ecol.* **44**: 296-305.

- Nunan N, Daniell TJ, Singh BK, Papert A, McNicol JW & Prosser JI. 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 KJ, Crawford JW & Young IM. 2001. Quantification of the in situ distribution of soil bacteria by large scale imaging of thin sections of undisturbed soil. *FEMS Microbiol Ecol.* **37**: 67-77.
- Nunoura T, Miyazaki M, Suzuki Y, Takai K & Horikoshi K. 2008. Hydrogenivirga okinawensis sp. nov, a thermophilic sulfur-oxidizing chemolithoautotroph isolated from a deep-sea hydrothermal field, Southern Okinawa Trough. *Int J Syst Evol Microbiol.* **58**: 676-681.
- Nusslein K & Tiedje JM. 1999. Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. *Appl Environ Microbiol.* **65**: 3622-3626.
- Oades JM. 1984. Soil organic matter and structural stability: mechanisms and implications for management. *Plant Soil.* **76**: 319-337.
- Oliver A, Baquero F & Blazquez J. 2002. The mismatch repair system (mutS, mutL and uvrD genes) in Pseudomonas aeruginosa: molecular characterization of naturally occurring mutants. *Mol Microbiol.* **43**: 1641-1650.
- Oren A. 2004. Prokaryote diversity and taxonomy: current status and future challenges. *Philos Trans R Soc Lond B Biol Sci.* **359**: 623-638.
- Orsini M & Romano-Spica V. 2001. A micro-wave-based method for nucleic acid isolation from environmental samples *Lett Appl Microbiol.* **33**: 17-20.
- Paget E & Simonet P. 1994. On the track of natural transformation in soils. *FEMS Microbiol Ecol.* **15**: 109-118.
- Parent ME & Velegol D. 2004. E. coli adhesion to silica in the presence of humic acid. *Colloids Surf B Biointerfaces.* **39**: 45-51.
- Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME & Bergman NH. 2009. Structure and complexity of a bacterial transcriptome. *J Bacteriol.* **191**: 3203-3211.
- Pedros-Alio C. 2006. Marine microbial diversity: Can it be determined? *Trends Microbiol.* **14**: 257-263.
- Phillipot L, Renault P, Sierra J, Henault C, Clays-Josserand A, Chenu C, Chaussod R, Lensi R. 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.
- Postma J, van Veen JA & 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.



- Poteete AR. 2001. What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol Lett.* **201**: 9-14.
- Poteete AR, Fenton AC & Murphy KC. 1988. Modulation of Escherichia coli RecBCD activity by the bacteriophage lambda Gam and P22 Abc functions. *J Bacteriol.* **170**: 2012-2021.
- Radman M, Taddei F & Matic I. 2000. Evolution-driving genes. *Res Microbiol.* **151**: 91-95.
- 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 JM. 2007. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb Ecol.* **53**: 197-207.
- Ranjard L & Richaume A. 2001. Quantitative and qualitative microscale distribution of bacteria in soil. *Res Microbiol.* **152**: 707-716.
- Ranjard L, Richaume A, Jocteur Monorozier L & Nazaret S. 1997. Response of soil bacteria to Hg(II) in relation to soil characteristics and cell location. *FEMS Microbiol Ecol.* 321-331.
- Rappe MS & Giovannoni SJ. 2003. The uncultured microbial majority. *Annu Rev Microbiol.* **57**: 369-394.
- Recorbet G, Robert C, Givaudan A, Kudla B, Normand P & Faurie G. 1993. Conditional suicide system of Escherichia coli released into soil that uses the Bacillus subtilis sacB gene. *Appl Environ Microbiol.* **59**: 1361-1366.
- Ritz K, McNicol W, Nunan N, Grayston S, Millard P, Atkinson D, Gollotte A, Habeshaw D, Boag B, Clegg CD, Griffiths BS, Weatley RE, Glover LA, McCaig AE & Prosser JI. 2004. Spatial structure in soil chemical and microbiological properties in an upland grassland. *FEMS Microbiol Ecol.* **49**: 191-205.
- Robe P, Nalin R, Capellano C, Vogel T & Simonet P. 2003. Extraction of DNA from soil. *Eur. J. Soil Biol.* **39**: 183-190.
- Robert M, Chenu, C. 1992. Interactions between soil minerals and micro-organisms, in Stotzky, G., Bollag, J.M. (Eds) *Soil Biochemistry* 307-404.
- Robertson DK, Gu L, Rowe RK & Beatty WL. 2009. Inclusion biogenesis and reactivation of persistent chlamydia trachomatis requires host cell Sphingolipid Biosynthesis. *PLoS Pathog.* **5**: 1000664.
- Roesch LF, Fulthorpe RR, Riva A, et al. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *Isme J.* **1**: 283-290.

- Ronaghi M, Uhlen M & Nyren P. 1998. A sequencing method based on real-time pyrophosphate. *Science*. **281**: 363, 365.
- Ronaghi M, Pettersson B, Uhlen M & Nyren P. 1998. PCR-introduced loop structure as primer in DNA sequencing. *Biotechniques*. **25**: 876-878, 880-872, 884.
- 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.
- Rosenberg SM. 2001. Evolving responsively: adaptive mutation. *Nat Rev Genet*. **2**: 504-515.
- Rykiel EJ. 1985. Towards a definition of ecological disturbance. *Aus J Ecol*. **10**: 361-365
- Saano A & Lindstrom, K. 1995. Small-scale extraction of DNA from soil with spun column cleanup. *Molecular Microbial Ecology Manual*. 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 OE, Chapin FS, 3rd, Armesto JJ, et al. 2000. Global biodiversity scenarios for the year 2100. *Science*. **287**: 1770-1774.
- Sanger F, Nicklen S & Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*. **74**: 5463-5467.
- Santos R, Fernandes J, Fernandes N, Oliveira F & Cadete M. 2007. Mycobacterium parascrofulaceum in acidic hot springs in Yellowstone National Park. *Appl Environ Microbiol*. **73**: 5071-5073.
- Schena M, Shalon D, Heller R, Chai A, Brown PO & Davis RW. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA*. **93**: 10614-10619.
- Schloss PD & Handelsman J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol*. **71**: 1501-1506.
- Schloss PD & Handelsman J. 2006. Toward a census of bacteria in soil. *PLoS Comput Biol*. **2**: 92.
- Sebastianelli A, Sen T & Bruce IJ. 2008. Extraction of DNA from soil using nanoparticles by magnetic bioseparation. *Lett Appl Microbiol*. **46**: 488-491.
- Senior BW & Holland IB. 1971. Effect of colicin E3 upon the 30S ribosomal subunit of Escherichia coli. *Proc Natl Acad Sci U S A*. **68**: 959-963.
- Shen P & Huang HV. 1986. Homologous recombination in Escherichia coli: dependence on substrate length and homology. *Genetics*. **112**: 441-457.

- Shendure J. 2008. The beginning of the end for microarrays? *Nat Methods*. **5**: 585-587.
- Shendure J & Ji H. 2008. Next-generation DNA sequencing. *Nat Biotechnol*. **26**: 1135-1145.
- Shendure J, Porreca GJ, Reppas NB, *et al.* 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*. **309**: 1728-1732.
- Shiu SH & Borevitz JO. 2008. The next generation of microarray research: applications in evolutionary and ecological genomics. *Heredity*. **100**: 141-149.
- Sikorski J, Teschner N & Wackernagel W. 2002. Highly different levels of natural transformation are associated with genomic subgroups within a local population of *Pseudomonas stutzeri* from soil. *Appl Environ Microbiol*. **68**: 865-873.
- Singh BK, Millard P, Whiteley AS & Murrell JC. 2004. Unravelling rhizosphere-microbial interactions: opportunities and limitations. *Trends Microbiol*. **12**: 386-393.
- Singh BK, Tate KR, Kolipaka G, Hedley CB, Macdonald CA, Millard P & Murrell JC. 2007. Effect of afforestation and reforestation of pastures on the activity and population dynamics of methanotrophic bacteria. *Appl Environ Microbiol*. **73**: 5153-5161.
- Singh BK, Dawson LA, McDonald CA & Buckland SM. 2009. Impact of biotic and abiotic interaction on soil microbial communities and functions: A field study. *App. Soil Ecol*. **41**: 239-248.
- Singh BK, Munro S, Potts J & Millard P. 2007. Influence of grass species and soil type on rhizosphere microbial community structure in grassland soils. *Appl. Soil Ecol*. **36**: 147-155.
- Singh BK, Nazarie L, Munro S, Anderson I & Campbell CD. 2006. Use of multiplexterminal restriction fragment length polymorphism for rapid and simultaneous analysis of different components of the soil microbial community. *Appl Environ Microbiol*. **72**: 7278-7285.
- Six J, Elliot ET & 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 ET & Paustian K. 2000. Soil macroaggregate turnover and microaggregate formation: a mechanism for C sequestration under no-tillage agriculture. *Soil Biol Biochem* **32**: 2099-2103.
- Small J, Call DR, Brockman FJ, Straub TM & Chandler DP. 2001. Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl Environ Microbiol*. **67**: 4708-4716.
- SoginML, *et al.* 2006. Microbial diversity in the deep sea and the under explored "rare biosphere". *Proc Natl Acad Sci USA*. **103**: 12115-12120.
- Stach JE, Bathe S, Clapp JP & Burns RG. 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.

- Stevenson FJ. 1976. Stability constants of Cu, Pb and Cd complexes with humic acids. *Soil Sci Soc Am J.* **40**: 665-672.
- Streit WR & Schmitz RA. 2004. Metagenomics--the key to the uncultured microbes. *Curr Opin Microbiol.* **7**: 492-498.
- Tang M, Shen X, Frank EG, O'Donnell M, Woodgate R & Goodman MF. 1999. UmuD'(2)C is an error-prone DNA polymerase, Escherichia coli pol V. *Proc Natl Acad Sci USA.* **96**: 8919-8924.
- Tao H, Bausch C, Richmond C, Blattner FR & Conway T. 1999. Functional genomics: expression analysis of Escherichia coli growing on minimal and rich media. *J Bacteriol.* **181**: 6425-6440.
- Tarlera S, Jangid K, Ivester AH, Whitman WB & Williams MA. 2008. Microbial community succession and bacterial diversity in soils during 77,000 years of ecosystem development. *FEMS Microbiol Ecol.* **64**: 129-140.
- Taylor JP, Wilson, B, Mills MS & Burns RG. 2002. Comparison of microbial numbers and enzymatic activities in surface soils and subsoils using various techniques. *Soil Biology and Biochemistry.* **34**: 387-401.
- Thioulouse J, Chessel D, Dole'dec S & Olivier JM. 1997. ADE-4: a multivariate analysis and graphical display software. *Stat. Comput.* **7**: 75–83.
- Thomas CM & Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol.* **3**: 711-721.
- Tiedje JM, Assuming-Bremppong S, Nusslein K, Marsh TL & Flynn SJ. 1999. Opening the black box of soil microbial diversity. *Appl. Soil Ecol.* **13**: 109-122.
- Tisdall JM & Oades JM. 1982. Organic matter and water-stable aggregates in soils. *J Soil Sci.* **33**: 141-163
- Torres B, Jaenecke S, Timmis KN, Garcia JL & Diaz E. 2000. A gene containment strategy based on a restriction-modification system. *Environ Microbiol.* **2**: 555-563.
- Torres B, Jaenecke S, Timmis KN, Garcia JL & Diaz E. 2003. A dual lethal system to enhance containment of recombinant micro-organisms. *Microbiology.* **149**: 3595-3601.
- 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**: 776-781.
- Tracy RB, Chedin F & Kowalczykowski SC. 1997. The recombination hot spot chi is embedded within islands of preferred DNA pairing sequences in the E. coli genome. *Cell.* **90**: 205-206.

- Tsai YL & Olson BH. 1992. Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Appl Environ Microbiol.* **58**: 2292-2295.
- Turcatti G, Romieu A, Fedurco M & Tairi AP. 2008. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res.* **36**: e25.
- Verville JH, Hobbie SE, Chapin FS & Hooper DU. 1998. Response of tundra Ch<sub>4</sub> and Co<sub>2</sub> flux to manipulation of temperature and vegetation. *Biogeochemistry.* **41**: 215-235.
- Vitousek PM & Howarth RW. 1991. Nitrogen limitation on land and in the sea: How can it occur? *Biogeochemistry.* **13**: 87-115.
- Volossiouk T, Robb EJ & Nazar RN. 1995. Direct DNA extraction for PCR-mediated assays of soil organisms. *Appl Environ Microbiol.* **61**: 3972-3976.
- von Wintzingerode F, Gobel UB & Stackebrandt E. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev.* **21**: 213-229.
- Wang Q, Garrity GM, Tiedje JM & Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* **73**: 5261-5267.
- Wayne LG (and 11 others). 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* **37**: 463-464
- White RE. 2006. Principles and practice of soil science. The soil as a natural resource. . *Fourth edn: Blackwell Publishing.*
- Whitman WB, Coleman DC & Wiebe WJ. 1998. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A.* **95**: 6578-6583.
- Woese CR. 1987. Bacterial evolution. *Microbiol Rev.* **51**: 221-271.
- Woese CR & Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA.* **74**.
- Wommack KE & Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev.* **64**: 69-114.
- Wong SL. 1995. Advances in the use of *Bacillus subtilis* for the expression and secretion of heterologous proteins. *Curr Opin Biotechnol.* **6**: 517-522.
- Xu X, Stern A, Liu Z, Kan B & Zhu J. 2010. Virulence regulator AphB enhances toxR transcription in *Vibrio cholerae*. *BMC Microbiology.* **10**: 3.

Young CC, Burghoff RL, Keim LG, Minak-Bernero V, Lute JR & Hinton SM. 1993. Polyvinylpyrrolidone-Agarose Gel Electrophoresis Purification of Polymerase Chain Reaction-Amplifiable DNA from Soils. *Appl Environ Microbiol.* **59**: 1972-1974.

Zechner EL, de la Cruz F, Eisenbrandt R, Grahn AM, Koraimann G, Lanka E, Muth G, Pansegrau W, Thomas CM, Wilkins BM & Zatyka M. 2000. Conjugative-DNA transfer processes. In Thomas CM, ed, *The Horizontal gene pool: bacterial plasmids and gene spread. Harwood Scientif. Publ. UK.* 87-174.

Zhang T, Fang HH & Ko BC. 2003. Methanogen population in a marine biofilm corrosive to mild steel. *Appl Microbiol Biotechnol.* **63**: 101-106.

Zhou J, Bruns MA & Tiedje JM. 1996. DNA recovery from soils of diverse composition. *Appl Environ Microbiol.* **62**: 316-322.

Zhou J, Xia B, Treves DS, *et al.* 2002. Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol.* **68**: 326-334.

Zhou QF, Li MY & Li CW. 2009. Cloning and expression of a novel insulin-releasing peptide, brevinin-2GU from *Escherichia coli*. *J Biosci Bioeng.* **107**: 460-463.

**Titre: Diversité bactérienne des sols : accès aux populations à effectifs minoritaires  
« the rare biosphere »**

**Mots clés:** sols; biodiversité; adaptation; identification bactérienne ; puces à ADN

**Title:** *Bacterial diversity of soils : accessing the effective minority populations in the "rare biosphere"*

**Keywords:** *soil; biodiversity; adaptation; bacterial identification; microarray*

**Résumé:**

L'exploration complète de la diversité bactérienne reste incomplète, due à des limitations techniques (extraction d'ADN incomplète, limites des techniques de caractérisation...), à la complexité et l'hétérogénéité de la matrice sol (présence de microenvironnements...) et à l'existence de populations numériquement faibles. Bien que ces populations minoritaires jouent probablement un rôle important dans le fonctionnement de la communauté, elles sont rarement détectées par les techniques conventionnelles. Nos objectifs dans le cadre de ce travail de thèse ont été de développer des approches tant conceptuelles que techniques qui permettent d'accéder à un niveau plus important de la diversité bactérienne présente dans les environnements complexes comme les sols. La première approche a pour but de favoriser le développement de bactéries minoritaires grâce à l'inoculation de communautés bactériennes dans des sols stérilisés avec des propriétés physico-chimiques différentes, afin de confirmer ou d'infirmer le concept de Baas Becking « tout est partout et l'environnement sélectionne ». Les changements de structures des communautés bactériennes inoculées, sont analysés à l'aide de puces à ADN taxonomique nouvellement développées. Les résultats confirment clairement l'impact de l'environnement sur la structure des populations bactériennes. De plus l'analyse des puces à ADN révèle des bactéries précédemment non détectées, confirmant la présence de populations minoritaires et la possibilité d'augmenter leur abondance relative sous différentes conditions. La deuxième approche consiste au développement d'une souche bactérienne capable de capturer *in situ* des fragments d'ADN. Elle a pour but d'éviter l'étape d'extraction d'ADN. Cet outil est basé sur la construction d'un système de contre sélection positive permettant de sélectionner les clones ayant intégré des gènes d'intérêt. Nos travaux montrent clairement qu'il est possible d'améliorer l'accès à cette « rare biosphere » sans toutefois pouvoir répondre entièrement à la question : est-ce que tout est partout?

**Direction de Recherche :**

Directeur de recherche CNRS, SIMONET Pascal

[pascal.simonet@ec-lyon.fr](mailto:pascal.simonet@ec-lyon.fr)

**Laboratoire Ampère (CNRS UMR5005)**

<http://www.ampere-lab.fr>

**Ecole Centrale de Lyon**

36 avenue Guy de Collongue  
69134 Ecully cedex (France)

Version électronique de cette thèse :

<http://tel.archives-ouvertes.fr>