



HAL
open science

**Caractérisation des communautés procaryotiques
impliquées dans la bioremédiation d'un sol pollué par
des hydrocarbures et développement d'outils d'analyse à
haut débit**

Cécile Militon

► **To cite this version:**

Cécile Militon. Caractérisation des communautés procaryotiques impliquées dans la bioremédiation d'un sol pollué par des hydrocarbures et développement d'outils d'analyse à haut débit. Protistologie. Université Blaise Pascal - Clermont-Ferrand II; Université d'Auvergne - Clermont-Ferrand I, 2007. Français. NNT : 2007CLF21810 . tel-00718551

HAL Id: tel-00718551

<https://theses.hal.science/tel-00718551>

Submitted on 17 Jul 2012

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.

**ECOLE DOCTORALE
DES SCIENCES DE LA VIE ET DE LA SANTE
N° d'ordre 474**

THESE

Présentée à l'Université Blaise Pascal
pour l'obtention du grade de

DOCTEUR D'UNIVERSITE

Spécialité : GENOMIQUE ET ECOLOGIE MICROBIENNE

Présentée et soutenue publiquement par

Cécile MILITON

le 17 décembre 2007

**CARACTERISATION DES COMMUNAUTES PROCARYOTIQUES
IMPLIQUEES DANS LA BIOREMEDIATION D'UN SOL POLLUE
PAR DES HYDROCARBURES ET DEVELOPPEMENT D'OUTILS
D'ANALYSE A HAUT DEBIT: LES BIOPUCES ADN.**

JURY

Présidente : Arlette DARFEUILLE-MICHAUD (Pr., Université d'Auvergne, Clermont-Ferrand)
Rapporteurs : Pascale BAUDA (Pr., Université Paul Verlaine, Metz)
Pascal SIMONET (DR CNRS, Ecole Centrale, Ecully)
Examineur : Robert DURAN (Pr., Université de Pau et des Pays de l'Adour, Pau)
Directeurs : Pierre PEYRET (Pr., UMR CNRS 6023, Université d'Auvergne, Clermont-Ferrand)
Eric PEYRETAILLADE (Dr., UMR CNRS 6023, Université d'Auvergne, Clermont-Ferrand)

**Laboratoire Biologie des Protistes - UMR CNRS 6023 - Equipe Génomique Intégrée des
Interactions Microbiennes, Université Blaise Pascal, Clermont-Ferrand.**

REMERCIEMENTS

Je souhaite exprimer mes remerciements à Christian Amblard, directeur du laboratoire de Biologie des Protistes (UMR CNRS 6023), pour m'avoir accueilli au sein de son laboratoire et permis de réaliser cette thèse dans de bonnes conditions.

Mes remerciements vont ensuite à mon directeur de thèse, Pierre Peyret, pour avoir encadré mes travaux jusqu'à leurs aboutissements mais surtout pour m'avoir fait partager sa passion pour la recherche, son ouverture d'esprit et sa rigueur scientifique durant ces années passées à ses côtés. Merci de m'avoir donné une grande liberté et des conditions de travail idéales qui m'ont permis de m'épanouir scientifiquement. Merci enfin de m'avoir fait confiance en me permettant de présenter mes travaux au cours de différents congrès et de m'avoir soutenue jusqu'en Autriche (dont je garde de très bons souvenirs). Que vous trouviez ici l'expression de ma profonde gratitude pour tout ce que vous avez fait pour moi.

Mes remerciements vont aussi à Eric Peyretaillade pour l'intérêt qu'il a porté à ce travail en acceptant de co-diriger cette thèse, pour sa disponibilité, ses orientations et ses remarques fructueuses. Merci sincèrement Eric pour ton soutien scientifique mais aussi pour la bonne humeur que tu as le don de transmettre autour de toi.

Je souhaite exprimer mes remerciements les plus sincères à Pascale Bauda et à Pascal Simonet pour m'avoir fait l'honneur d'accepter d'être rapporteurs de cette thèse ainsi qu'à Arlette Darfeuille-Michaud et à Robert Duran pour avoir accepté d'examiner ce travail et de faire partie du jury.

Un grand merci à Anne Moné pour sa précieuse collaboration. Je tiens à lui rendre hommage pour son courage et pour le soutien qu'elle m'a offert depuis le premier jour. Anne, ce fut un véritable plaisir de travailler à tes côtés et de partager un neurone.

Merci aussi à Delphine Boucher et à Corinne Petit pour leur aide dans la rédaction de ce mémoire ainsi que pour toutes les connaissances qu'elles m'ont fait partager durant ces années. Merci à Sébastien Rimour et Geneviève Gagne pour leur contribution aux travaux bioinformatiques et biologiques ainsi que pour la rédaction des publications.

Merci à tous les membres de l'équipe GIIM qui ont été pour moi une seconde famille durant cette thèse. Un merci particulier à Brigitte Chebance pour son éternelle bonne humeur, sa disponibilité, son efficacité et sa gentillesse. Merci de m'avoir rendu plus douces ces années difficiles.

Je remercie Julien Troquet, directeur de Biobasic Environnement, et Cédric Vachelard pour le matériel biologique fourni et les données associées dans le cadre d'un projet commun.

Merci à Valérie et Abdel de m'avoir prise sous leur aile les premières années, de m'avoir offert leur précieuse amitié et pour toutes ces soirées passées ensemble. Un grand merci aussi à Cécile pour sa gentillesse et sa lucidité qu'elle m'a fait partager. Merci à Stéphanie et Isabelle, pour leur amitié sans faille. Merci aussi à Carine, Sylvie, Thomas, Antoine, Benoît, Maider, Marc-Antoine, Sigolène, Jean-Philippe, Olivier, Simon, Jean-Denis, Denis, Fabien et Ingrid pour leurs amitiés qui dure depuis le lycée (ou presque).

Pour terminer, j'aimerais dédier cette thèse à ma famille. A mon père, qui m'a permis d'être ce que je suis aujourd'hui, à ma mère pour m'avoir toujours soutenue et pour m'avoir permis de faire mes études dans les meilleures conditions, à Séverine, à Loïc, à mes neveux Lison, Eliot et Nine. Je n'oublie pas non plus ma belle-famille et Philippe pour leur soutien.

A Anthony...

**CARACTERISATION DES COMMUNAUTES PROCARYOTIQUES
IMPLIQUEES DANS LA BIOREMEDIATION D'UN SOL POLLUE PAR DES
HYDROCARBURES ET DEVELOPPEMENT D'OUTILS D'ANALYSE A HAUT DEBIT: LES
BIOPUCES ADN.**

Résumé :

Les activités humaines sont à l'origine de nombreuses pollutions par des hydrocarbures au niveau des écosystèmes et plus particulièrement au niveau des sols. Afin de préserver la santé humaine et environnementale, il est nécessaire de réduire les risques d'exposition et de contamination en dégradant les polluants en produit moins toxiques voire même inoffensifs. Dans ce but, les techniques de bioremédiation apparaissent aujourd'hui comme de réelles alternatives aux techniques classiques pouvant être très invasives et onéreuses. Cependant, la mise en place et/ou l'optimisation de tels procédés nécessitent une meilleure connaissance des communautés microbiennes impliquées dans la biodégradation de ces polluants. Ainsi, après avoir fait un état des lieux des outils développés pour isoler, identifier et caractériser ces populations, nous avons étudié la structure et la dynamique des communautés bactériennes au cours d'un procédé de biostimulation d'un sol pollué aux hydrocarbures aliphatiques. Outre la caractérisation d'un « core-set » bactérien présent tout au long du procédé de bioremédiation, cette étude nous a permis d'identifier des phylotypes potentiellement impliqués dans la biodégradation de ces polluants. Enfin, nous nous sommes attachés à mettre au point des outils moléculaires à haut débit, afin d'étudier plus globalement ces communautés bactériennes environnementales, et ce par le développement d'un logiciel de sélection de sondes sensibles, spécifiques et exploratoires pour biopuces ADN phylogénétiques (PhylArray).

**CHARACTERIZATION OF THE PROKARYOTIC COMMUNITIES IMPLIED IN
THE HYDROCARBON-POLLUTED SOIL BIOREMEDIATION AND
DEVELOPMENT OF HIGH-THROUGHPUT ANALYZE TOOLS :
THE DNA MICROARRAYS.**

Abstract :

Environmental ecosystems are sensitive to damage from human activities and there is an increasing need to develop better methods for removing pollutants from soils. The removal of pollutants or their transformation to less toxic products by bioremediation is a less invasive and expensive process than classical decontamination. However, the use and optimization of bioremediation treatments requires knowledge on the microbial communities directly and indirectly involved in the degradation of the pollutants. Therefore, after the creation of a synthesis describing the microbial diversity and the available tools for the isolation and the identification of microorganisms, we have studied the structure and the dynamic of the bacterial communities during the bioremediation of an aliphatic hydrocarbon-polluted soil. Beyond the characterization of a bacterial "core-set" present throughout the decontamination process, we have identified several phylotypes potentially involved in degradation. In order to globally monitor these bacterial communities, we also have developed a probe design software (PhylArray) generating efficient and explorative probes for high-throughput molecular tools: the DNA microarrays.

Discipline : Ecologie et Génomique Microbienne.

Mots-clés : Diversité bactérienne, ADNr 16S et ARNr 16S, Biopuces phylogénétique, Sélection de sondes, Sols pollués, Hydrocarbures aliphatiques, Bioremédiation.

Laboratoire : Biologie des Protistes, UMR CNRS 6023, 24 av. des Landais, 63177 Aubière.

SOMMAIRE

INTRODUCTION GENERALE	1
------------------------------------	----------

PARTIE I : SYNTHESE BIBLIOGRAPHIQUE	7
--	----------

<u>Chapitre I : Les sols pollués aux hydrocarbures et leur réhabilitation.....</u>	8
Introduction.....	8
1. Le système sol	9
1.1. La fraction minérale.....	9
1.2. La fraction organique	10
1.3. Les phases liquides et gazeuses.....	10
1.4. La fraction vivante	11
2. Les hydrocarbures aliphatiques.....	12
2.1. Caractéristiques physico-chimiques.....	12
2.2. Impact toxicologique	13
3. Comportement des polluants organiques au niveau du sol.	14
3.1. Géosorption aux différentes fractions du sol.....	15
3.2. Devenir des substances géosorbées.....	16
4. Réhabilitation non biologique des sols pollués par des hydrocarbures.....	16
4.1. Destruction/altération des polluants	16
4.2. Extraction des contaminants de la matrice sol	17
4.3. Techniques d'immobilisation	18
5. Réhabilitation biologique des sols pollués par des hydrocarbures.	19
5.1. La phytoremédiation	19
5.2. La bioremédiation	21
5.2.1. <i>L'atténuation naturelle</i>	21
5.2.2. <i>La biostimulation</i>	22
5.2.3. <i>La bioaugmentation</i>	22
Conclusions.	23
<u>Chapitre II : Identification des microorganismes : apport des techniques moléculaires</u>	24
Introduction	24
1. La révolution moléculaire en écologie microbienne.....	25
1.1. Méthodes basées sur l'amplification PCR	26

1.1.1. <i>Clonage/séquençage des produits PCR</i>	26
1.1.2. <i>Techniques d'empreintes génétiques</i>	27
1.2. Méthodes basées sur l'hybridation.....	28
2. Les nouvelles techniques à haut débit.	29
2.1. Métagénomique et métatranscriptomique.....	29
2.2. Le séquençage ultra-haut débit.....	31
Conclusions	32
<u>Chapitre III : Les biopuces ADN : outils à haut-débit.</u>	33
Introduction.	33
1. La technologie biopuce ADN.	33
1.1. Les sondes	34
1.2. Fixation des sondes	35
1.3. Préparation des cibles	36
1.4. L'hybridation	37
1.5. Acquisition et analyse des images de biopuces.	38
2. Les biopuces fonctionnelles et génomiques.	39
2.1. Les biopuces ciblant des gènes fonctionnels.....	39
2.2. Les biopuces ciblant des génomes.	41
2.3. Les biopuces génétiques.	42
3. Les biopuces oligonucléotidiques phylogénétiques.	43
3.1. Sélection des sondes oligonucléotidiques.....	43
3.2. Efficacité des sondes.....	45
3.3. Etudes de la diversité d'environnements complexes	46
Conclusions.	47
<u>Chapitre IV : Diversité bactérienne dans les écosystèmes sols.</u>	48
Introduction.	48
1. Diversité bactérienne des sols.	48
1.1. Structuration des communautés bactériennes	48
1.2. Diversité des communautés bactériennes	49
1.2.1. <i>Approches culturales</i>	49
1.2.1. <i>Approches moléculaires</i>	51
2. Diversité bactérienne des sols pollués.	52
2.1. Adaptations des populations suite à une pollution.....	53
2.2. Bactéries aérobies dégradant les polluants.....	54
2.3. Bactéries anaérobies dégradant les polluants.....	56

Conclusions.....	57
------------------	----

PARTIE II : NOUVEAUX OUTILS D'ISOLEMENT ET D'IDENTIFICATION DES MICRO-ORGANISMES.....	58
--	-----------

1. Contexte.....	59
2. Objectifs.....	59
ARTICLE 1 : New tools for isolation and identification of micro-organisms.....	61
1) Microbial Diversity and evolution.....	63
1.1) Phenotypic and genetic microbial classification.....	63
1.2) Microbial diversity in complex environments.....	67
1.3) Origin of the eukaryotic cell.....	70
2) Isolation and microbial cultivation methods.....	72
2.1) Standard methods for micro-organisms cultivation.....	72
2.2) New cultural strategies for micro-organism isolation.....	73
2.3) High-throughput micro-organism cultivation.....	74
2.4) Single cell isolation.....	75
3) PCR-based methods for microbial identification.....	75
3.1) Nucleic acids extraction.....	75
3.2) Extra cellular DNA and dead cells.....	77
3.3) PCR artefacts.....	78
3.4) rDNA markers and other biomarkers.....	78
3.5) Molecular fingerprinting methods.....	81
3.5.1) Amplified ribosomal DNA restriction analysis (ARDRA).....	81
3.5.2) Terminal-restriction fragment length polymorphism (T-RFLP) and Length heterogeneity polymerase chain reaction (LH-PCR).....	81
3.5.3) Denaturing gradient gel electrophoresis (DGGE) and Temperature gradient gel electrophoresis (TGGE).....	82
3.5.4) Ribosomal intergenic spacer analysis (RISA).....	83
3.5.5) Single-strand conformation polymorphism (SSCP).....	84
3.5.6) Randomly amplified polymorphic DNA (RAPD), Variable number of tandem repeats (VNTR), and Pulsed field gel electrophoresis (PFGE).....	84
4) Hybridization methods and Isotope probing.....	85
4.1) DNA reassociation.....	85

4.2) Fluorescence in situ hybridization (FISH).....	86
4.3) DNA microarrays.....	89
4.4) Stable isotope probing (SIP).....	93
5) Metagenomics.....	95
6) New high-throughput sequencing developments.....	98
Conclusions.....	99
Perspectives.....	100
References.....	101
3. Discussion.....	117

<p>PARTIE III : ÉTUDE DE LA DYNAMIQUE DES COMMUNAUTÉS BACTÉRIENNES AU COURS D'UN PROCÉDÉ DE BIOREMÉDIATION D'UN SOL POLLUÉ PAR DES HYDROCARBURES.118</p>
--

1. Contexte de l'étude.....	119
2. Objectifs de l'étude.....	119
3. Principaux résultats obtenus.....	120
ARTICLE 2 : Dynamics of the bacterial communities during a pilot-scale bioremediation process of a hydrocarbon-contaminated soil identified by SSU rDNA and SSU rRNA libraries.....	122
Summary.....	122
Introduction.....	123
Results.....	125
Discussion.....	129
Experimental procedures.....	133
References.....	137
4. Discussion.....	140

<p>PARTIE IV : DÉVELOPPEMENT D'UN ALGORITHME DE SELECTION DE SONDES OLIGONUCLÉOTIDIQUES POUR BIOPUCES PHYLOGÉNÉTIQUES144</p>

1. Contexte de l'étude.....	145
2. Objectifs de l'étude.....	145
3. Principaux résultats obtenus.....	146
ARTICLE 3 : PhylArray: Phylogenetic Probe Design Algorithm for MicroArray...147	

Abstract.....	147
1 Introduction.....	148
2 Probe design strategy.....	149
2.1 Targeted taxonomic groups and polymorphism questions.....	150
2.2 SSU rRNA database re-build.....	151
3 Design algorithm.....	151
4 Implementation.....	153
5 Experimental validation.....	154
5.1 Probe selection.....	154
5.2 Experimental procedure.....	154
5.3 Probe efficiency analysis.....	155
5.3.1. <i>Probe sensitivity</i>	156
5.3.2. <i>Probe specificity</i>	156
5.3.3. <i>Explorative power</i>	157
6 Discussion.....	157
7 Conclusion.....	160
References.....	161
Supplementary data.....	164
4. Discussion.....	171

PARTIE V : CONCLUSIONS ET PERSPECTIVES.	173
---	------------

1. Conclusions.....	174
2. Perspectives.....	177
2.1. Etude exhaustive des communautés procaryotiques.....	177
2.1.1. <i>Création d'une biopuce phylogénétique complète</i>	177
2.1.2. <i>Utilisation de biomarqueurs complémentaires</i>	177
2.2. De la structure à la fonction.....	178
2.2.1. <i>Activité biologique et structure</i>	178
2.2.2. <i>Voies métaboliques et structure</i>	181
2.3. Etude de l'ensemble du réseau trophique microbien des sols.....	183

REFERENCES.....	184
------------------------	------------

ANNEXE.....	206
--------------------	------------

LISTE DES FIGURES

PARTIE I : SYNTHESE BIBLIOGRAPHIQUE

<u>Figure 1.</u> Structure des 16 Hydrocarbures Aromatiques Polycycliques (HAP) prioritaires selon l'US EPA.....	9
<u>Figure 2.</u> Interactions entre les hydrocarbures aliphatiques et les différentes fractions du sol	15
<u>Figure 3 :</u> Principe de l'extraction sous vide <i>in situ</i> des polluants du sol.....	17
<u>Figure 4 :</u> Principe du lavage <i>in situ</i> des polluants du sol.....	18
<u>Figure 5 :</u> Principe de la vitrification <i>in situ</i> des polluants du sol.....	19
<u>Figure 6 :</u> Principe des barrières réactives passives (BRS).....	19
<u>Figure 7 :</u> Principe de la phytoremédiation.....	20
<u>Figure 8 :</u> Dégradation microbienne des polluants aliphatiques au niveau du sol.....	22
<u>Figure 9 :</u> Histogrammes représentant le nombre de séquences ADNr 16S soumis dans GenBank depuis 1993	26
<u>Figure 10 :</u> Principe de la technique SARST (Serial Analysis of Ribosomal Sequence Tags)	27
<u>Figure 11 :</u> Principe de la technique T-RFLP (Terminal Restriction Fragment Length Polymorphism).....	28
<u>Figure 12 :</u> Principe de la technique DGGE ou TGGE (Denaturing et Thermal Gradient Gel Electrophoresis)	28
<u>Figure 13 :</u> Cinétique de réassociation d'ADN de différentes communautés microbiennes du sol.....	29
<u>Figure 14 :</u> Hybridation <i>in situ</i> de cellules bactériennes et de cellules eucaryotes en biofilm	29
<u>Figure 15 :</u> L'analyse par métagénomique des communautés microbiennes	30
<u>Figure 16 :</u> Principe du pyroséquençage.....	32
<u>Figure 17 :</u> Principe des biopuces ADN	34
<u>Figure 18 :</u> Schéma d'une analyse transcriptomique par biopuce à ADN.....	34
<u>Figure 19 :</u> Fixation des sondes sur les biopuces ADN par un robot.....	35
<u>Figure 20 :</u> Synthèse <i>in situ</i> des sondes par la technique de la photolithographie.....	35
<u>Figure 21 :</u> Principe du marquage des cibles pour l'hybridation des biopuces ADN.....	36

<u>Figure 22</u> : Principe de la lecture des biopuces par un scanner à fluorescence.....	38
<u>Figure 23</u> : Carte d'accessibilité de l'ARNr 16S.....	44
<u>Figure 24</u> : Schéma de détermination de sondes pour POAs (Phylogenetic Oligonucleotides Arrays).....	45
<u>Figure 25</u> : Modèles évolutifs r/K	49
<u>Figure 26</u> : Box-plots représentant la contribution des différents phyla bactériens à la biodiversité des sols	51
<u>Figure 27</u> : Estimation du niveau de connaissance de différents phyla et des bactéries en général	52
<u>Figure 28</u> : Biodégradation aérobie des alcanes.....	54

PARTIE II : NOUVEAUX OUTILS D'ISOLEMENT ET D'IDENTIFICATION DES MICRO-ORGANISMES

<u>Figure 1</u> : Description of a possible scheme for microbial identification	67
<u>Figure 2</u> : Schematic representation of the different ways used for microbial identifications for an environmental sample.....	69
<u>Figure 3</u> : Schematic representation of the different strategies used for microbial identification by nucleic acids-based identification.	75
<u>Figure 4</u> : Schematic representation of DNA microarray experiment	89

PARTIE III : ÉTUDE DE LA DYNAMIQUE DES COMMUNAUTÉS BACTÉRIENNES AU COURS D'UN PROCÉDÉ DE BIOREMÉDIATION D'UN SOL POLLUÉ PAR DES HYDROCARBURES
--

<u>Figure 1</u> : Rarefaction analysis of the three SSU rDNA and the three SSU rRNA libraries generated before the bioremediation process, after 12 months and after 24 months	127
--	-----

PARTIE IV : DÉVELOPPEMENT D'UN ALGORITHME DE SELECTION DE SONDES OLIGONUCLÉOTIDIQUES POUR BIOPUCES PHYLOGÉNÉTIQUES

<u>Figure 1</u> : Schematic representation of probe design to target a group of species.	151
---	-----

<u>Figure 2</u> : Schematic illustration of the last step of the algorithm.	152
<u>Figure 3</u> : Sensitivity of designed <i>Staphylococcus</i> probes.	156
<u>Figure 4</u> : Specificity evaluation of the probe design.....	156
<u>Figure 5</u> : Potential explorative power of PhylArray probes.	157
<u>Figure 6</u> : Schéma de la stratégie de design GoArrays comparée celle utilisée classiquement pour le design de sondes oligonucléotidiques pour biopuces ADN.....	171

PARTIE V : CONCLUSIONS ET PERSPECTIVES

<u>Figure 1</u> : Images de MAR-FISH sur un biofilm montrant les groupes ayant incorporé le substrat radioactif (¹⁴ C méthanol)	179
<u>Figure 2</u> : Images d'Isotope Array sur une souche pure <i>Nitrosomonas eutropha</i>	180
<u>Figure 3</u> : Séparation des acides nucléiques par la méthode de marquage par isotopes stables (SIP : Stable Isotope Probing).....	180
<u>Figure 4</u> : Importance des communautés procaryotiques dans les réseaux trophiques des écosystèmes aquatiques.....	183

LISTE DES TABLEAUX

PARTIE I : SYNTHESE BIBLIOGRAPHIQUE

<u>Tableau 1.</u> Caractéristiques physico-chimiques d'hydrocarbures (HC) aliphatiques et aromatiques polycycliques (HAP)	13
<u>Tableau 2.</u> Synthèse exhaustive des projets de métagénomique.....	31
<u>Tableau 3.</u> Caractéristiques des différents formats de biopuces utilisées en écologie microbienne.....	47

PARTIE III : ÉTUDE DE LA DYNAMIQUE DES COMMUNAUTÉS BACTÉRIENNES AU COURS D'UN PROCÉDÉ DE BIOREMÉDIATION D'UN SOL POLLUÉ PAR DES HYDROCARBURES

<u>Table 1.</u> Phylogenetic groups representation in the libraries obtained from 16S rDNA and 16S rRNA throughout the pilot-scale bioremediation process of the hydrocarbon polluted soil and characterization of the phylogenetic groups common in the both types of libraries.....	126
<u>Table 2.</u> Dominant phylotypes present in the SSU rDNA, in the SSU rRNA or in both libraries generated before the bioremediation process, after 12 months and after 24 months	126
<u>Table 3.</u> Metabolically active bacterial core-set observed before and during the bioremediation process.....	129

LISTE DES ABREVIATIONS

%GC	Pourcentage en Guanine/Cytosine	MAR-FISH	Microautoradiography-Fluorescent <i>in situ</i> hybridization
16S	16 Svedberg	mCFU	Microcolony forming units
ADN/ARN	Acide désoxyribonucléique/ Acide ribonucléique	MGA	Metagenomic array
ADNc	Acide désoxyribonucléique complémentaire	MGGA	Multiple gene genealogical analysis
ADNr	Acide désoxyribonucléique ribosomique	MLSA	Multilocus sequence analysis
AhR	Aryl hydrocarbon receptor	MLST	Multilocus sequence typing
AIA	Alcool isoamylique	MSA	Multiple sequence alignment
ANI	Average nucleotide identity	NAST	Nearest alignment space termination
ATP	Adénosine triphosphate	NCBI	National Center for Biotechnology Information
ARDRA	Amplified ribosomal DNA restriction analysis	NHS	N-hydroxysuccinylimidyl
ARNm	Acide ribonucléique messenger	NLS	N-lauroyl sarcosine
ARNr	Acide ribonucléique ribosomique	nt	Nucléotide
BAC	Bacterial artificial chromosome	OGM	Organismes génétiquement modifiés
BLAST	Basic local alignment search tool	ORF	Open reading frame
BRS	Barrières réactives passives	OTU	Operational taxonomic unit
BSA	Bovine serum albumine	pb	Paire de bases
BTEX	Benzène toluène éthylbenzène xylène	PCE	Perchloroéthylène
CARD	Catalyzed reporter deposition	PCR	Polymerase chain reaction
CGA	Community genome array	PFGE	Pulsed field gel electrophoresis
CGA	Comparative genealogical analysis	PFLA	Phospholipid fatty acid
CsCl	Caesium chloride	PMA	Propidium monoazide
CsTFA	Caesium trifluoroacetate	PMT	Photomultiplicateur
CTAB	Bromure d'hexadécyltriméthylammonium	PNA	Peptide nucleic acid
Cy	Cyanine	POA	Phylogenetic oligonucleotide array
CYP	Cytochrome P450	ppm	Partie par million
dCTP	Désoxy cytosine tri-phosphate	PVPP	Polyvinylpyrrolidone
DGGE	Denaturing gradient gel electrophoresis	RAPD	Randomly amplified polymorphic DNA
DHPLC	Denaturing high-performance liquid chromatography	RDP	Ribosomal database project
DNase	Désoxyribonucléase	RING	Recognition of individual genes
dNTP	Désoxy nucléotide tri-phosphate	RISA	Ribosomal intergenic spacer analysis
DOTUR	Distance-based OTU and richness	RNase	Ribonucléase
DTT	Dithiothreitol	RSGP	Reverse sample genome probing
dUTP	Désoxy uracile tri-phosphate	RT	Reverse transcription
EDTA	Acide éthylène-diamine-tétraacétique	SARST	Serial analysis of ribosomal sequence tags
EMA	Ethidium monoazide	SDS	Sodium dodécyl sulfate
EMBL	European Molecular Biology Laboratory	SIGEX	Substrate-induced gene expression screening
FGA	Functional gene array	SIP	Stable isotope probing
FISH	Fluorescent <i>in situ</i> hybridization	SONS	Stands for shared OTUs and similarity
FU	Fluorescence Unit	SRP	Sulfate-reducing prokaryote
Gb	Gigabases	SSELO	Sequence-specific end labelling of oligonucleotides
GFP	Green fluorescent protein	SSC	Saline sodium citrate
GNS	Green Non Sulfur	SSCP	Single-strand conformation polymorphism
GOLD	Genome online database	S/S	Solidification/Stabilisation
HAP	Hydrocarbure aromatique polycyclique	SSU/LSU	Small subunit/Large subunit
HGT	Horizontal gene transfer	TGGE	Temperature gradient gelelectrophoresis
IRMS	Isotope-ratio mass spectrometry	TIV	Transcription <i>in vitro</i>
ITS	Intergenic transcribed sequence	T-RFLP	Terminal restriction fragment length polymorphism
IUPAC	International Union of Pure and Applied Chemistry	ULS	Universal linkage system
LGT	Lateral gene transfer	UniFrac	Unique fraction
LH-PCR	Length heterogeneity polymerase chain reaction	UV	Ultraviolet
LUCA	Last universal common ancestor	VNTR	Variable number of tandem repeats
		WGA	Whole genome array

INTRODUCTION GENERALE

Depuis plusieurs décennies, les activités humaines engendrent une production de plus en plus importante de produits chimiques, potentiellement toxiques, pouvant être déversés accidentellement ou non dans l'environnement. Ainsi, les risques pour la santé publique et le bon fonctionnement des écosystèmes sont de plus en plus importants et préoccupants. La base de données ARIA (Analyse, Recherche et Information sur les Accidents ; <http://aria.ecologie.gouv.fr/>), exploitée par le Ministère de l'Ecologie, du Développement et de l'Aménagement Durable, a recensé en France et à l'étranger les événements accidentels qui ont ou auraient pu porter atteinte à la santé, à la sécurité publique, à l'agriculture et à l'environnement. Depuis 1992, ces accidents, essentiellement dûs aux activités industrielles et agricoles, aux transports de matières dangereuses, auraient touchés plus de 32 000 sites (ARIA, 2006). Sur l'ensemble de ces accidents, 40 % d'entre-eux auraient entraîné la contamination de sites terrestres ou aquatiques. Notons que le recensement de ces accidents et incidents français ou étrangers ne peut être considéré comme exhaustif.

Afin de restaurer ces sites pollués, un arsenal de procédures a été développé. Ces techniques emploient des procédés mécaniques, physiques et/ou chimiques afin d'éliminer les contaminations. Grâce à ces méthodes, les polluants sont immobilisés, extraits et/ou détruits afin de réduire leurs impacts sur la santé et sur l'environnement (Khan *et al.*, 2004; Kalin, 2004). Cependant, ces procédés peuvent être très invasifs pour les écosystèmes traités car, d'une part, ils nécessitent souvent l'excavation des matériaux pollués ce qui bouleverse les environnements et, d'autre part, ils peuvent entraîner le rejet de composés toxiques pour les formes de vie. Il existe cependant une alternative à ces techniques invasives permettant de limiter ces effets néfastes : la remédiation biologique. Cette approche utilise les capacités naturelles de certains organismes à dégrader les polluants parfois même jusqu'à leur minéralisation (Serrano *et al.*, 2007). Ce type de décontamination peut impliquer les végétaux (phytoremédiation) et/ou les microorganismes (bioremédiation). Même si la phytoremédiation est potentiellement efficace (Tripathi *et al.*, 2007), elle semble toutefois limitée aux zones occupées par les racines (la rhizosphère) contrairement aux procédés de bioremédiation utilisant les microorganismes qui ne sont pas limités dans l'espace. De plus, cette dégradation microbienne de la pollution peut aussi être améliorée soit par adjonction d'agents stimulant les microorganismes endogènes (Perfumo *et al.*, 2007), soit par ajout de microorganismes épurateurs endogènes ou exogènes (Jacques *et al.*, 2007). Ces procédés sont nommés respectivement biostimulation et bioaugmentation.

Afin d'optimiser ces techniques de bioremédiation, il est nécessaire d'améliorer les connaissances sur les communautés microbiennes impliquées dans la dégradation des polluants et de surveiller leur évolution durant ces procédés. Cependant, dans la plupart des écosystèmes naturels, ces études se révèlent être de véritables challenges tant les communautés microbiennes sont denses et diversifiées (Curtis et Sloan, 2004). Wackett et Hershberger (2000) ont ainsi estimé qu'un gramme de terre pouvait compter jusqu'à 10^9 bactéries représentant 10 000 espèces différentes. Ainsi, de part leur diversité, leur fonctionnement en *consortia*, leur évolution rapide et leurs caractères culturels spécifiques, les microorganismes sont difficilement isolables et l'identification de souches présentant un intérêt en bioremédiation passe souvent par la mise en oeuvre de techniques moléculaires.

Durant la dernière décennie, l'utilisation de ces nouvelles techniques a connu un essor considérable, représentant une alternative puissante à l'identification des microorganismes par des approches culturales (Amann *et al.*, 1995). Ces approches moléculaires incluent des techniques de réassociation d'ADN, d'hybridations ADN-ADN ou ARNm-ADN, de clonage-séquençage et d'empreintes génétiques comme par exemple la Denaturing Gradient Gel Electrophoresis (DGGE), la Temperature Gradient Gel Electrophoresis (TGGE), la Ribosomal Intergenic Spacer Analysis (RISA) et la Denaturing High-performance Liquid Chromatography (DHPLC) (Kirk *et al.*, 2004, Nocker *et al.*, 2007). Les gènes codant pour les ARN ribosomiques sont généralement choisis comme biomarqueurs dans ces techniques d'identification car ils sont ubiquistes et possèdent en parallèle des régions conservées et variables. De plus, ils ne font l'objet apparemment que de très rares transferts horizontaux et subissent peu de recombinaisons (Miller *et al.*, 2005). Ces caractéristiques en font un bon marqueur moléculaire taxonomique (facilité d'isolement, discrimination évolutive de nombreuses espèces). Récemment, des techniques dites de métagénomique et de métatranscriptomique ont été développées afin d'accéder à l'immense réservoir génétique que forme l'ensemble des génomes et des transcriptomes de tous les microorganismes présents au niveau des environnements complexes. Ces méta « omics » peuvent permettre la découverte de nouveaux microorganismes et de nouvelles fonctions métaboliques grâce à la prise en considération globale de l'information génétique et non plus simplement de celle des biomarqueurs (Venter *et al.*, 2004). Toutes ces techniques ont permis de faire des avancées considérables dans l'identification de la diversité phylogénétique et métabolique des communautés microbiennes présentes au niveau des écosystèmes. Actuellement, un des derniers défis que se sont lancés les microbiologistes consiste à établir un lien entre diversité

et fonction afin d'accéder à l'écophysiologie des microorganismes. Afin d'atteindre cet objectif, différentes techniques, basées sur l'incorporation de marqueurs radioactifs (MAR-FISH, Isotope Array) ou isotopiques (Stable Isotope Probing) au niveau des macromolécules, ont été développées (Wagner *et al.*, 2006).

Une des dernières avancées en biologie moléculaire a été le développement des biopuces ADN. Apparues au milieu des années 90 (Schena *et al.*, 1995), elles ont été initialement mises au point pour l'étude globale de l'expression de tous les gènes d'un organisme (transcriptome). Cette technique, appliquée à l'étude des communautés microbiennes (DeSantis *et al.*, 2007 ; Moisander *et al.*, 2006 ; Peplies *et al.*, 2006 ; Sanguin *et al.*, 2006), représente un potentiel considérable car elle permet la détection simultanée de milliers d'espèces microbiennes en une seule manipulation (Wagner *et al.*, 2007). Ainsi, l'utilisation de cet outil à haut débit pourrait permettre d'accéder à l'extraordinaire diversité microbienne d'environnements complexes et, plus précisément, d'étudier la dynamique des communautés bactériennes et de mieux comprendre les mécanismes d'adaptation régissant le fonctionnement d'écosystèmes sains ou anthropisés.

Afin de participer à l'amélioration des connaissances sur les communautés microbiennes et plus particulièrement sur les bactéries impliquées dans les phénomènes de décontamination naturelle des sols pollués, les objectifs de cette thèse ont été les suivants. Tout d'abord, nous avons souhaité décrire la complexité de la diversité microbienne et réaliser une synthèse regroupant tous les outils culturels et moléculaires, récemment développés, permettant l'isolement et l'identification des microorganismes. En fonction des contraintes associées à notre projet, cet inventaire quasi exhaustif nous a servi de base dans le choix de techniques adaptées pour le suivi des communautés bactériennes au cours d'un procédé de bioremédiation d'un sol pollué par des hydrocarbures. Ces travaux de recherche ont été réalisés dans le but de caractériser la dynamique des populations bactériennes suite à la présence et à la disparition des polluants afin d'identifier les bactéries potentiellement impliquées dans ces phénomènes de biodégradation. Enfin, pour pouvoir suivre plus facilement l'extrême diversité des communautés microbiennes, nous avons mis en place une approche biopuces ADN. De part les limitations des logiciels classiquement utilisés pour la détermination de sondes, nous avons développé un logiciel permettant la sélection de sondes efficaces (sensibles, spécifiques et exploratoires) pour l'étude des microorganismes d'environnements complexes.

Afin de présenter le contexte de l'étude et les principaux résultats obtenus, le mémoire de cette thèse s'articulera autour de quatre parties.

La première partie constitue la synthèse bibliographique. Celle-ci est composée de quatre chapitres présentant un état de l'art allant de l'écosystème sol à l'identification des souches épuratrices. Lors du premier chapitre, consacré à l'étude des sols pollués, des polluants et des techniques de réhabilitation des écosystèmes, l'accent sera mis sur notre modèle d'étude : les hydrocarbures aliphatiques. Au cours du deuxième chapitre, nous nous employerons à décrire l'impact qu'a eu l'apparition des techniques moléculaires sur l'étude des communautés microbiennes. Dans le chapitre suivant, nous nous focaliserons, sur un de ces outils permettant une étude à haut débit de la diversité microbienne: les biopuces ADN avec notamment les biopuces phylogénétiques oligonucléotidiques. Enfin, le dernier chapitre dressera un état des lieux des connaissances sur la diversité des microorganismes présents dans les écosystèmes de sols sains ou contaminés par des hydrocarbures aliphatiques.

La partie II présentée sous forme d'un chapitre d'ouvrage fera une synthèse de la diversité microbienne et des outils culturels et moléculaires mis à la disposition des microbiologistes afin d'isoler, d'identifier et de caractériser les communautés microbiennes. Nous avons volontairement choisi de séparer ce travail de la partie synthèse bibliographique afin de pouvoir présenter plus en détails ces nouvelles techniques qui permettent d'accéder à la majorité non cultivée des microorganismes et ainsi d'ouvrir des perspectives stimulantes quant à l'amélioration de nos connaissances sur le monde microbien.

La partie III présentera les travaux de recherche portant sur la caractérisation de la structure et de la dynamique des communautés bactériennes au cours d'un procédé de bioremédiation d'un sol pollué aux hydrocarbures aliphatiques. Cette étude a eu pour objectifs de caractériser des groupes bactériens potentiellement impliqués dans la biodégradation du polluant et ainsi de définir des pistes quant à l'amélioration des techniques de décontamination naturelle.

Enfin, la partie IV exposera le développement d'outils biopuces ADN et plus particulièrement celui d'un nouveau logiciel de sélection de sondes nommé PhylArray. Nous démontrerons l'efficacité de ces nouvelles biopuces tant au niveau de la spécificité et de la sensibilité de détection des bactéries par rapport à d'autres logiciels de sélection de sondes. De plus, nous suggérerons l'aspect exploratoire de nos biopuces. Ainsi, ces nouvelles biopuces devraient permettre de donner une image plus réaliste des communautés bactériennes des écosystèmes et faciliter la compréhension de leur fonctionnement.

Une conclusion générale ainsi que les perspectives envisagées pour continuer les travaux initiés clôtureront ce mémoire.

PARTIE I

SYNTHESE BIBLIOGRAPHIQUE

*Des sols pollués à l'identification de souches
épuratrices : Apports des outils moléculaires*

Chapitre I

Les sols pollués aux hydrocarbures et leur réhabilitation.

INTRODUCTION

Les sols sont constitués de minéraux, de matières organiques, d'eau et d'air (Alexander, 1961). Ce mélange permet au sol d'être le support trophique de la production végétale et à ce titre d'être un des déterminants de la qualité et de la sécurité alimentaire. Les sols déterminent aussi la qualité des eaux souterraines et superficielles de part leurs rôles dans les processus de ruissellement et d'infiltration des eaux. De plus, les sols sont le support de la plus importante biomasse de la Terre : les microorganismes. Ceux-ci pourraient représenter jusqu'à 6.10^{30} bactéries au niveau des écosystèmes terrestres et aquatiques (Whitman *et al.*, 1998). De plus, cette biomasse apparaît très diversifiée. En effet, un gramme de terre contiendrait plus d'un million de génomes différents (Gans *et al.*, 2005), ceci laissant présager du potentiel génétique de cette biodiversité encore largement méconnue (Dunbar *et al.*, 1999). A ces différents titres et en tenant compte de leur caractère quasiment non renouvelable à l'échelle des générations humaines, les sols constituent un patrimoine qui doit impérativement être préservé. Leur gestion durable doit donc s'imposer dans les années à venir comme une préoccupation de plus en plus forte d'autant plus que ces sols sont des écosystèmes sensibles, très souvent touchés par la pollution.

L'apparition d'une pollution résulte le plus souvent d'accidents industriels, de dépôts ou du transport de matières dangereuses (Riser-Roberts, 1998). Selon les bases de données BASOL (2004) et BASIAS (2004), du Ministère de l'Ecologie, du Développement et de l'Aménagement Durable, il y aurait en France 3 900 sites et sols actifs, c'est-à-dire pollués, et 180 000 sites potentiels qui ont abrité, par le passé, une activité pouvant être à l'origine d'une pollution. Les principaux polluants de ces sols, constatés seuls ou en mélange, sont : les hydrocarbures, le plomb, le zinc, les solvants halogénés, le chrome, le cuivre, l'arsenic, le

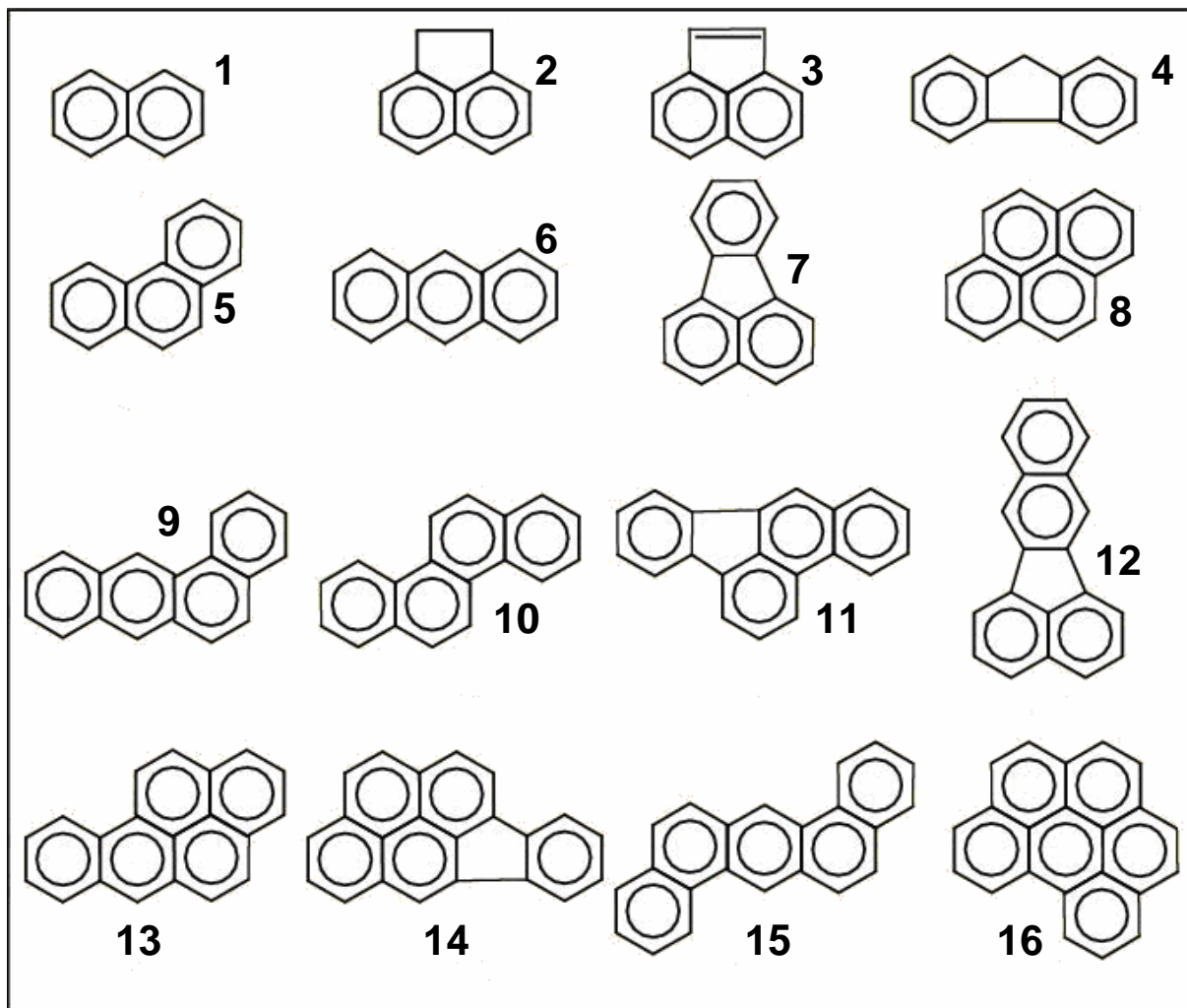


Figure 1. Structure des 16 Hydrocarbures Aromatiques Polycycliques (HAP) prioritaires selon l'US EPA. (1) Naphtalène, (2) Acénaphthène, (3) Acénaphthylène, (4) Fluorène, (5) Phénanthrène, (6) Anthracène, (7) Fluoranthène, (8) Pyrène, (9) Benzo[a]anthracène, (10) Chrysène, (11) Benzo[b]fluoranthène, (12) Benzo[k]fluoranthène, (13) Benzo[a]pyrène, (14) Indéno[1,2,3-cd]pyrène, (15) Dibenzo[a,h]anthracène et (16) Benzo[g,h,i]pérylène.

nickel et le cadmium. Les hydrocarbures font partie des polluants organiques majoritaires au niveau de ces écosystèmes. Dans les années 1970, les préoccupations envers ce type de polluants organiques devenant de plus en plus fortes, l'Agence de protection environnementale des Etats-Unis (US EPA) a élaboré une liste indiquant ceux qui devaient être prioritairement pris en compte. Cette première liste, mettant en évidence seize hydrocarbures polycycliques aromatiques (HAP) (figure 1), a été suivie par de nombreuses autres traitant d'autres types de polluants à considérer. Cependant, à l'heure actuelle, les HAP restent toujours au centre des recherches sur les polluants organiques alors que ce ne sont pas ceux qui sont les plus représentés au niveau des sites pollués (Block *et al.*, 1991). En effet, les plus importants, quantitativement, sont les hydrocarbures aliphatiques qui sont, au contraire, relativement peu étudiés.

Généralement, les origines des pollutions aux hydrocarbures sont des accidents résultant d'activités anthropiques ou de dépôts atmosphériques. Au niveau des sols, en fonction de leurs caractéristiques physico-chimiques, ces polluants peuvent avoir deux comportements majeurs : soit ils sont éliminés par biodégradation, par lessivage ou par volatilisation, soit ils sont retenus par les fractions du sol. Etant donné que ces polluants sont des composés pouvant avoir un impact très important sur la santé humaine mais aussi sur le fonctionnement des écosystèmes, il est nécessaire de développer des procédés de réhabilitation de ces sols pollués.

1. LE SYSTEME SOL.

Le sol est un système situé à l'interface de l'atmosphère, de la lithosphère et de l'hydrosphère. Il est constitué d'une fraction minérale, d'une fraction organique, d'eau, d'air et d'organismes vivants (Alexander, 1961). Certaines de ces fractions interagissent fortement avec les polluants organiques. Ainsi, il est nécessaire de connaître les constituants de cet écosystème afin de comprendre les mécanismes influençant le comportement de ces polluants au niveau du sol.

1.1. La fraction minérale.

La fraction minérale provient de la dégradation physique ou chimique de la lithosphère. La désagrégation physique consiste en une micro-division de la roche mère sans modification

de sa composition. Les facteurs impliqués dans cette altération sont les chocs thermiques et hydriques ainsi que l'action du gel. L'altération chimique implique majoritairement des procédés d'hydrolyse et plus marginalement des procédés d'oxydoréduction, d'hydratation et de dissolution. Le facteur impliqué est l'eau parfois associée à des acides minéraux (dérivant de minéraux inorganiques, par exemple l'acide chlorhydrique) ou organiques, à du dioxyde de carbone et à de l'oxygène (Gobat *et al.*, 2003). Cette dégradation entraîne la formation d'une terre fine, représentant un peu moins de 50 % du sol, constituée de minéraux hérités et de minéraux secondaires composés d'argiles et d'oxyhydroxydes (Duchaufour, 2001).

1.2. La fraction organique.

La fraction organique, représentant moins de 10 % du sol, provient de la dégradation de la litière composée de débris végétaux, animaux et microbiens. Cette transformation mécanique et enzymatique est réalisée par la microflore, la mésofaune et la macrofaune du sol (Alexander, 1961). Lorsque le milieu est biologiquement actif, certains produits de la litière, comme la cellulose ou l'hémicellulose, sont rapidement convertis en minéraux solubles ou gazeux lors de la minéralisation primaire. Les molécules n'ayant pas été minéralisées (lignine, tanins, lipides, résines, cires, etc.) sont alors utilisées en partie, lors de l'humification, pour la formation de molécules complexes de nature colloïdale : l'humus (Duchaufour, 2001). Celui-ci est constitué de sous-produits traditionnellement définis par leur solubilité : les humines (fraction insoluble quelle que soit la valeur du pH), les acides humiques (fraction insoluble à des pH acides inférieurs à 2.0) et les acides fulviques (fraction soluble quelque soit le pH) (Van Trump *et al.*, 2006). L'humus peut à son tour subir une minéralisation appelée secondaire.

1.3. Les phases liquides et gazeuses.

Occupant les espaces poreux du sol, l'eau et les mélanges gazeux représentent environ la moitié de son volume total (Alexander, 1961). Ces pores sont localisés au niveau de complexes organo-minéraux formés par l'agrégation des argiles, des oxyhydroxydes, des carbonates et des ciments organiques contenus dans le sol. Leur origine peut être biologique (formation par la microflore et structuration par la faune), chimique (précipitation et insolubilisation de composés comme les oxyhydroxydes ou la matière organique) ou physique

(création de failles dues à des changements de volume, notamment des argiles) (Duchaufour, 2001). Ces pores, pouvant être classés en fonction de leur taille, forment des environnements différents. Ainsi dans les macroporosités, on retrouve majoritairement l'atmosphère des sols qui est composée d'azote, d'oxygène, de dioxyde de carbone, de vapeur d'eau et de gaz divers (Gobat *et al.*, 2003). Ces gaz peuvent être à l'état libre ou dissous dans les solutions du sol. Dans les microporosités, on retrouve surtout l'eau qui est impliquée dans la nutrition, de façon directe ou indirecte par le transport d'éléments nutritifs, et dans la pédogenèse. Cette eau peut provenir des précipitations ou des nappes souterraines. Au niveau du sol, elle se retrouve sous forme d'eau de ruissellement (mouvements parallèles à la surface), d'eau de gravité (mouvements verticaux ou obliques) ou d'eau retenue (immobile car les forces capillarité et d'absorption sont supérieures aux forces de gravité) (Duchaufour, 2001).

1.4. La fraction vivante.

La fraction vivante représente 2 à 5 % de la matière organique totale. Elle est composée de la microflore et de la faune du sol (Gobat *et al.*, 2003). La microflore est majoritairement composée de bactéries et de mycètes mais, aussi, d'algues. Notons que, contrairement aux systèmes aquatiques, peu d'études ont été réalisées sur les virus des sols qui sont pourtant relativement abondants. En effet, Williamson *et al.* (2005) ont mis en évidence par l'analyse de plusieurs sols que leur nombre pouvait aller jusqu'à $4,17.10^9$ particules/g de sol. Ces virus sont assez diversifiés mais dominés par des bactériophages présentant des capsides de 20 à 160 nm de diamètre et de morphologies différentes. Les bactériophages, ubiquistes de nature, persistent dans le monde bactérien sous deux états distincts : en tant que phage virulent ou sous forme lysogène. Tout comme dans les réseaux aquatiques, les virus participent certainement activement aux transferts de gènes (transduction) mais aussi au maintien des différentes populations bactériennes du sol en décimant celles devenues trop dominantes (théorie « killing the winner »). Les bactéries, abondantes au niveau de la rhizosphère du fait entre autres de la production d'exsudats racinaires, sont impliquées dans la dégradation de la matière organique. Les bactéries hétérotrophes dégradent la cellulose et l'hémicellulose des parois végétales ainsi que les protéines provenant du protoplasme. Les bactéries autotrophes tirent leur énergie de l'oxydation de composés, tels le soufre ou encore l'ammoniac, et assimilent le carbone contenu dans le dioxyde de carbone (Duchaufour, 2001). Les bactéries du sol sont donc intimement liées au fonctionnement des grands cycles biogéochimiques. Le

rôle des mycètes, dans la dégradation de la matière organique, est multiple. Il semble qu'ils soient capables de dégrader la cellulose, tout comme les bactéries, mais aussi certains composés récalcitrants comme la lignine et les tanins. La faune du sol, très importante et très diversifiée, est également impliquée dans la dégradation et la structuration de la matière organique (Duchaufour, 2001). Un mètre cube de sol de prairie pourrait contenir jusqu'à 260 millions d'individus divisés, selon la taille, en quatre classes. La microfaune (< 0,2 mm) est constituée majoritairement de Protozoaires et de Nématodes mais aussi de Rotifères et de Tardigrades. La mésofaune (0,2 à 4 mm) est constituée de Microarthropodes tels les Acariens et les Collembolés, de Pseudoscorpions, de Protoures, de Diploures, de petits Myriapodes et de vers. La macrofaune (4 à 80 mm) est composée de vers de terre, d'insectes, de limaces, d'escargots, d'araignées et d'opilions. Enfin, la mégafaune (> 10 cm) regroupe les mammifères, les reptiles et les amphibiens. (Deprince, 2003).

Le fonctionnement des écosystèmes terrestres est donc très complexe et dépendant de multiples paramètres biotiques et abiotiques.


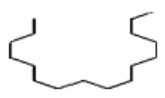

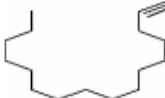


2. LES HYDROCARBURES ALIPHATIQUES.

Les hydrocarbures sont des polluants organiques majoritairement impliqués dans la pollution des écosystèmes précédemment décrits : les sols. Ce sont des molécules organiques formées d'atomes de carbone et d'hydrogène, sous des formes linéaires ou cycliques, qui peuvent présenter ou non des ramifications. Cette famille regroupe deux types de molécules : les aliphatiques et les aromatiques (Grignard, 1935). Dans ce paragraphe, nous traiterons uniquement des hydrocarbures aliphatiques qui, outre le fait d'être notre modèle d'étude, sont des polluants qui sont très largement retrouvés au niveau des sites pollués. Cependant, avant de décrire les comportements que peuvent adopter ces molécules au sein des systèmes sols, nous nous emploierons à décrire leurs caractéristiques physico-chimiques et leurs impacts sur l'Homme et les écosystèmes.

2.1. Caractéristiques physico-chimiques.

La famille des hydrocarbures aliphatiques est constituée de composés saturés et non saturés. Les hydrocarbures saturés, appelés alcanes, sont caractérisés par la présence de

Tableau 1. Caractéristiques physico-chimiques d'hydrocarbures (HC) aliphatiques et aromatiques polycycliques (HAP) (adapté de Stroud *et al.*, 2007). La structure des hydrocarbures influence leurs propriétés physico-chimiques (solubilité, état à température ambiante, etc.). *ND* : Non défini ; *Kow* : coefficient de partage octanol/eau.

Type d'HC	Nom	Formule	Poids moléculaire (g/mol)	Structure	Point de fusion (°C)	Point d'ébullition (°C)	Solubilité (mg/L)	Log Kow
Alcane	Tetradécane	C ₁₄ H ₃₀	198,38		5,5	253	0,000282	7,2
Alcane	Hexadécane	C ₁₆ H ₃₄	226,44		18	287	0,0009	9,1
Alcène	Hexadécène	C ₁₆ H ₃₂	224,43		3 à 5	274	ND	ND
Alcyne	Hexadécyne	C ₁₆ H ₃₀	222,42		15	148	ND	ND
Cyclane	Cyclohexane	C ₆ H ₁₂	84,16		6,55	80,74	ND	ND
HAP	Pyrène	C ₁₆ H ₁₀	202,6		156	404	0,135	5,19

liaisons simples entre les atomes de carbone (tableau 1). Leur formule brute est donc C_nH_{2n+2} . Ils forment deux grands groupes : les hydrocarbures à chaîne ouverte et les cyclanes dont la chaîne est fermée. Les hydrocarbures non saturés sont de deux types, les alcènes et les alcynes, caractérisés respectivement par la présence d'une double et d'une triple liaison entre deux atomes de carbone (tableau 1). Leurs formules brutes sont donc respectivement C_nH_{2n} et C_nH_{2n-2} . Les alcènes et les alcynes, possédant plus d'une liaison multiple, sont nommés respectivement alcadiènes, alcatriènes, [...] et alcadiynes, alcatriynes, [...].

Les hydrocarbures aliphatiques sont des composés non polaires et hydrophobes. Comparés aux molécules aromatiques polycycliques (Stroud *et al.*, 2007), ils semblent être beaucoup moins solubles dans l'eau de par leurs indices de solubilité moins élevés (tableau 1). Leurs températures de vaporisation peuvent être relativement importantes et sont directement influencées par la longueur de leur chaîne carbonée (tableau 1). En conséquence, ces caractéristiques font que ces polluants organiques sont peu lessivables et volatilisables (Stroud *et al.*, 2007). Il faut aussi noter que ces composés peuvent interagir avec certaines fractions du sol, entraînant potentiellement une disponibilité réduite envers les réactions bio-physico-chimiques permettant leur élimination. Les hydrocarbures aliphatiques saturés (alcanes) sont plus sensibles aux phénomènes de biodégradation que les hydrocarbures aliphatiques non saturés (alcènes, alcynes). Il en est de même pour les alcanes à longue chaîne ou non ramifiés comparés aux alcanes à chaîne plus courte ou ramifiés (Zhang et Bennett, 2005). Ainsi, certains hydrocarbures aliphatiques peuvent être considérés comme des composés récalcitrants à la dégradation et donc peuvent présenter une persistance au niveau des sols qui doit être prise en considération (Reid *et al.*, 2000).

2.2. Impact toxicologique.

Dans l'environnement, les contaminants peuvent être présents sous différentes formes augmentant ainsi les risques d'exposition. En effet, à température ambiante, les hydrocarbures aliphatiques sont gazeux jusqu'à C_4 , puis liquides et enfin solides à partir de C_{16} (excepté pour les alcynes qui se solidifient à partir de C_{15}). Ces formes induisent des modes de contamination différents. Par ordre d'importance, on notera l'inhalation, la pénétration à travers la peau puis l'ingestion (Baudouin *et al.*, 2002). Une fois dans l'organisme, ces composés hautement liposolubles rejoignent la circulation sanguine puis sont répartis vers un spectre très large de tissus avec une localisation préférentielle au niveau des tissus adipeux.

Du fait de leur caractère lipophile, ces polluants ont une tendance à la bioaccumulation *via* les réseaux trophiques (Fent, 2004).

De nombreux hydrocarbures aliphatiques présentent des risques toxicologiques, mutagènes voire même cancérigènes pour de nombreux organismes dont l'Homme (Spencer *et al.*, 2002). Il est cependant difficile de dissocier leurs effets propres car ils sont souvent présents en mélange avec d'autres hydrocarbures, tels les HAPs, qui sont largement plus étudiés. Les hydrocarbures aliphatiques chlorés (méthanes, éthanes et éthènes chlorés) sont les plus étudiés car ce sont des composés chimiques qui sont largement utilisés pour leurs propriétés de solvants et dont la toxicité est avérée (Ruder, 2006). En effet, le contact avec ces contaminants peut entraîner des dommages, pouvant aller jusqu'au développement de cancers, au niveau de nombreux organes (foie, reins, système cardio-vasculaire, système reproductif, système nerveux, etc.) (Ruder, 2006 ; Spencer *et al.*, 2002). Au niveau cellulaire, les mécanismes d'action des hydrocarbures aliphatiques sont assez mal connus comparés à ceux des hydrocarbures aromatiques polycycliques impliquant les enzymes telles les cytochromes P450 (CYP1 A1, CYP A2 et CYP1 B1), l'époxide hydrolase, la glutathione transférase, la UDP-glucuronosyltransferase, la sulfotransférase, la NAD(P)H quinone oxidoreductase 1 et l'aldo-keto reductase (Nebert *et al.*, 2004 ; Shimada, 2006). Ces enzymes participent à la conversion des polluants en métabolites hydrosolubles pouvant être éliminés. Cependant, ces réactions entraînent aussi la formation d'intermédiaires métaboliques hautement réactifs et potentiellement oncogéniques (PAH-diols, PAH diol-époxydes, PAH o-quinones, etc.).

3. COMPORTEMENT DES POLLUANTS ORGANIQUES AU NIVEAU DU SOL.

Certaines fractions du sol sont largement impliquées dans les comportements que peuvent adopter les hydrocarbures aliphatiques au niveau de ces écosystèmes. De plus, comme nous venons de le voir, du fait de leurs caractéristiques physico-chimiques, la majorité d'entre eux ne sont pas facilement volatilisables ou lessivables (Stroud *et al.*, 2007). Ainsi, les contaminants organiques se retrouvent, en grande partie, immobilisés au niveau des sols. Ce phénomène, nommé géosorption, est un procédé d'accumulation de molécules sur des surfaces solides (phénomènes d'adsorption) ou d'une phase du sol vers une autre (phénomène de partitionnement). Dans le cas des hydrocarbures, leur sorption se fait majoritairement sur

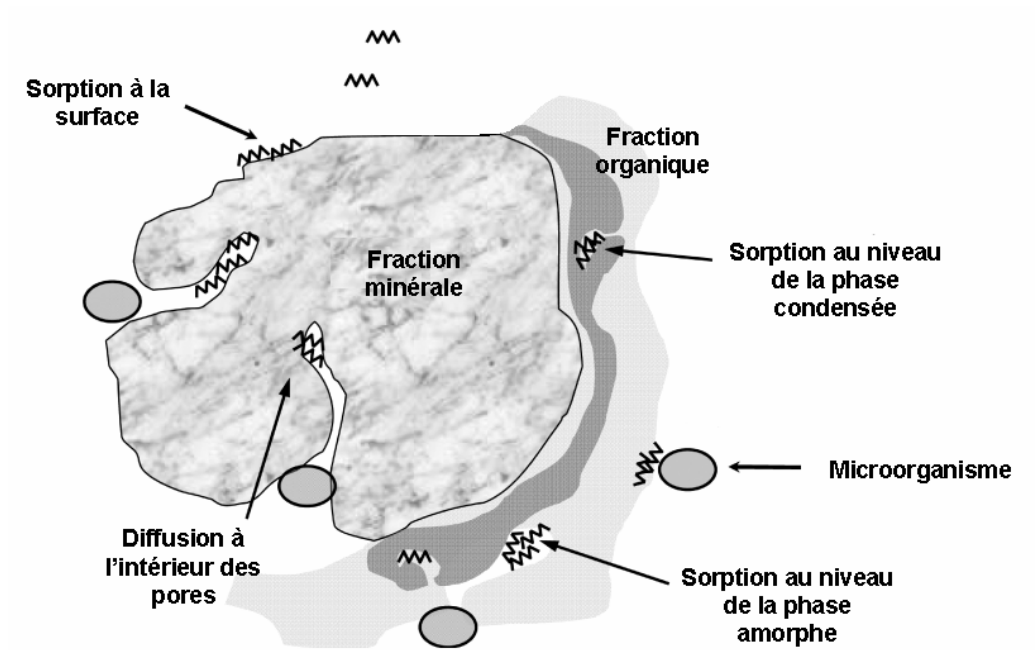


Figure 2. Interactions entre les hydrocarbures aliphatiques et les différentes fractions du sol (adapté de Stroud *et al.*, 2007). Les polluants organiques peuvent être géosorbés au niveau de la fraction minérale et des deux types de fraction organique (condensée et amorphe).

les fractions organique et minérale du sol. Ce phénomène est gouverné par différents paramètres : les propriétés intrinsèques de l'écosystème, les facteurs environnementaux et les propriétés physico-chimiques du polluant.

3.1. Géosorption aux différentes fractions du sol.

Lorsque le sol est riche en carbone organique ($> 0,1 \%$), la sorption se fait préférentiellement au niveau de la fraction organique (Ehlers et Loibner, 2006). Comme nous l'avons vu précédemment (cf. chapitre I ; § 1.2.), la fraction organique est composée de divers constituants. Dans ce contexte de géosorption, nous pouvons les subdiviser en deux catégories: la phase amorphe et la phase condensée. La phase amorphe, constituée par les substances humiques, les polysaccharides, les lipides, les protéines et la lignine, est majoritairement responsable du partitionnement des polluants organiques. La phase condensée, composée de substances diagénétiquement maturées comme le charbon, la suie et le kérogène, est, quant à elle, impliquée dans l'adsorption des hydrocarbures (Huang *et al.*, 2003). La séquestration des hydrocarbures est considérée comme réversible au niveau de la phase amorphe et irréversible au niveau de la phase condensée (figure 2). Dans le cas où le sol est pauvre en carbone organique ($< 0,1 \%$), la géosorption se fait préférentiellement au niveau de la fraction minérale (Schwarzenbach et Westall, 1981) (figure 2). Ce mécanisme peut prendre différentes formes : l'adsorption des polluants de l'air dans des conditions pauvres en eau, l'adsorption des polluants par des interactions donneurs/accepteurs d'électrons ou électrostatiques (Ehlers et Loibner, 2006).

En plus de leurs capacités propres à lier les polluants organiques, les fractions organiques et minérales du sol peuvent interagir l'une avec l'autre et ainsi engendrer un effet positif ou négatif sur cette fixation. Hur et Schlautman (2004) ont montré que les surfaces minérales pouvaient altérer la capacité de la fraction organique à lier les hydrocarbures. En effet, dans cette étude, la présence de kaolinite et d'hématite a induit la diminution des capacités des substances humiques à sorber le pyrène. Cependant, les interactions entre ces deux fractions peuvent aussi être à l'origine de structures stables. L'association de composés amorphes et condensés provenant de la fraction organique avec des composés minéraux peut entraîner une condensation des composés amorphes qui stabiliserait les liaisons avec les polluants organiques (Ehlers et Loibner, 2006).

3.2. Devenir des substances géosorbées.

Les polluants peuvent être relargués si les propriétés du géosorbant ou du polluant changent ou si les polluants migrent vers des sites permettant une désorption. Il semble que les polluants organiques tendent à être de moins en moins mobiles avec le temps (Alexander, 2000). Ces phénomènes de « vieillissement » affectent les mouvements et la réactivité des hydrocarbures au niveau des sols. Ainsi, cette disponibilité réduite a un impact direct sur les phénomènes de transport et la biodisponibilité des polluants pour les microorganismes (Huang *et al.*, 2003).

4. REHABILITATION NON BIOLOGIQUE DES SOLS POLLUES PAR DES HYDROCARBURES.

Comme nous l'avons vu précédemment, les hydrocarbures peuvent avoir des impacts très néfastes sur la santé humaine et celle des écosystèmes. Afin d'éviter les contacts entre les contaminants des sols et leurs cibles potentielles, il est nécessaire d'éliminer les polluants ou de les rendre moins toxiques. C'est pourquoi de nombreuses méthodes de réhabilitation ont été développées en fonction de la nature du site pollué, du type de polluant et du caractère *in-situ* ou *ex-situ* de la dépollution (Khan *et al.*, 2004). Celles-ci peuvent être classées selon trois types de stratégies utilisées seules ou en combinaison : la destruction ou l'altération des contaminants, l'extraction des contaminants du milieu environnemental et l'immobilisation des polluants (Kalin, 2004). Par souci de clarté, ne seront présentées ici que les méthodes de réhabilitation applicables à la décontamination des sols pollués par des hydrocarbures.

4.1. Destruction/altération des polluants.

La première façon de détruire ou d'altérer les polluants est de traiter thermiquement les sols contaminés. En présence d'oxygène, les sols peuvent être incinérés *ex-situ*, par des températures de 800 à 1 200° C, afin de transformer les hydrocarbures en carbone et en vapeur d'eau. Comme les propriétés bio-physico-chimiques de ces terres incinérées ne sont plus identiques à celles de terres non traitées, ces résidus de combustion ne peuvent plus être utilisés comme support de la végétation. Ils sont donc soit enfouis dans des centres spécialisés, soit réutilisés comme remblais (Delage et Schrefler, 2005). En absence

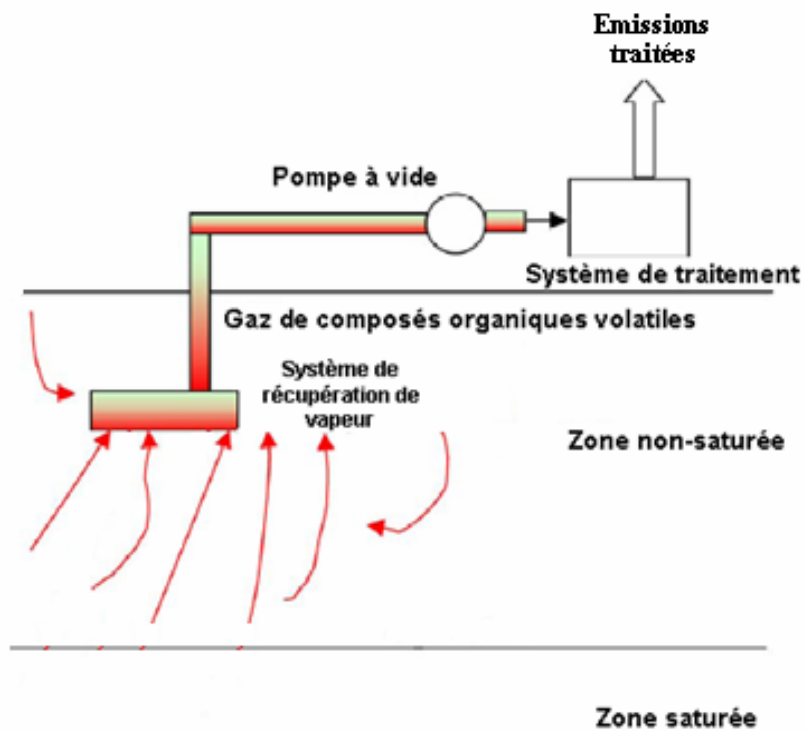


Figure 3 : Principe de l'extraction sous vide *in situ* des polluants du sol. La dépression créée par ce système de pompe à vide permet de collecter les vapeurs émises par les composés organiques volatiles. Ces vapeurs seront traitées avant d'être rejetées dans l'atmosphère (adapté de Khan *et al.*, 2006).

d'oxygène, le traitement est réalisé à des températures avoisinant 400° C (pyrolyse). Elle permet la transformation des polluants en coke et en gaz combustibles (monoxyde de carbone, hydrogène et méthane) (Van Deuren *et al.*, 2002).

L'oxydation chimique peut elle aussi permettre une altération des polluants organiques grâce à l'utilisation d'oxydants chimiques tels le peroxyde, l'ozone et le permanganate. Généralement, ces produits permettent d'obtenir de bonnes efficacités de traitement pour les composés aliphatiques non saturés et les HAPs (Van Deuren *et al.*, 2002). Bien que les alcanes soient moins réactifs, Yu *et al.* (2007) ont permis de dégrader, par ozonation *ex-situ*, jusqu'à 94 % des hydrocarbures saturés contenus dans un sol pollué à 1 485 ppm. Ces procédés sont très efficaces pour la remédiation des sols pollués ; cependant, ils ne sont pas sans conséquences sur les terres traitées. En effet, les oxydations affectent la mobilité des métaux, la structure de la matière organique et les microorganismes endogènes (Jung *et al.*, 2005).

4.2. Extraction des contaminants de la matrice sol.

Une autre façon de traiter les sols contaminés est d'en extraire les polluants. Ceci peut être réalisé par extraction sous-vide, par désorption thermique ou par lavage. L'extraction sous-vide est une technique efficace pour traiter les sols homogènes, non saturés, pollués par des composés volatiles ou semi-volatiles. Elle repose sur la pressurisation des sols, au travers de puits horizontaux et verticaux, afin de faciliter le procédé d'évaporation des polluants (figure 3). L'incorporation d'air est aussi souvent utilisée en combinaison afin de favoriser cette vaporisation. Les vapeurs ainsi extraites sont traitées avant d'être rejetées dans l'atmosphère (Delage et Schrefler, 2005). Cependant, l'hétérogénéité des sols peut causer de sérieuses limitations à l'utilisation de cette technique du fait de possibles pertes en air. Aussi, Hoier *et al.* (2007) ont développé un procédé alternatif, l'extraction sous vide pneumatique, qui permet de décontaminer ce type de sols. Grâce à un gradient de pression, dérivé d'un système de pompage pulsatile, un sol sablonneux a été débarrassé à 77 % de sa pollution au trichloroéthylène. L'efficacité de cette technique reste toutefois très limitée pour les sols saturés, possédant de forts pouvoirs de sorption.

La désorption thermique consiste à chauffer les sols pollués excavés jusqu'à la température d'ébullition du polluant, généralement de 100 à 600° C, afin de le vaporiser. Ce

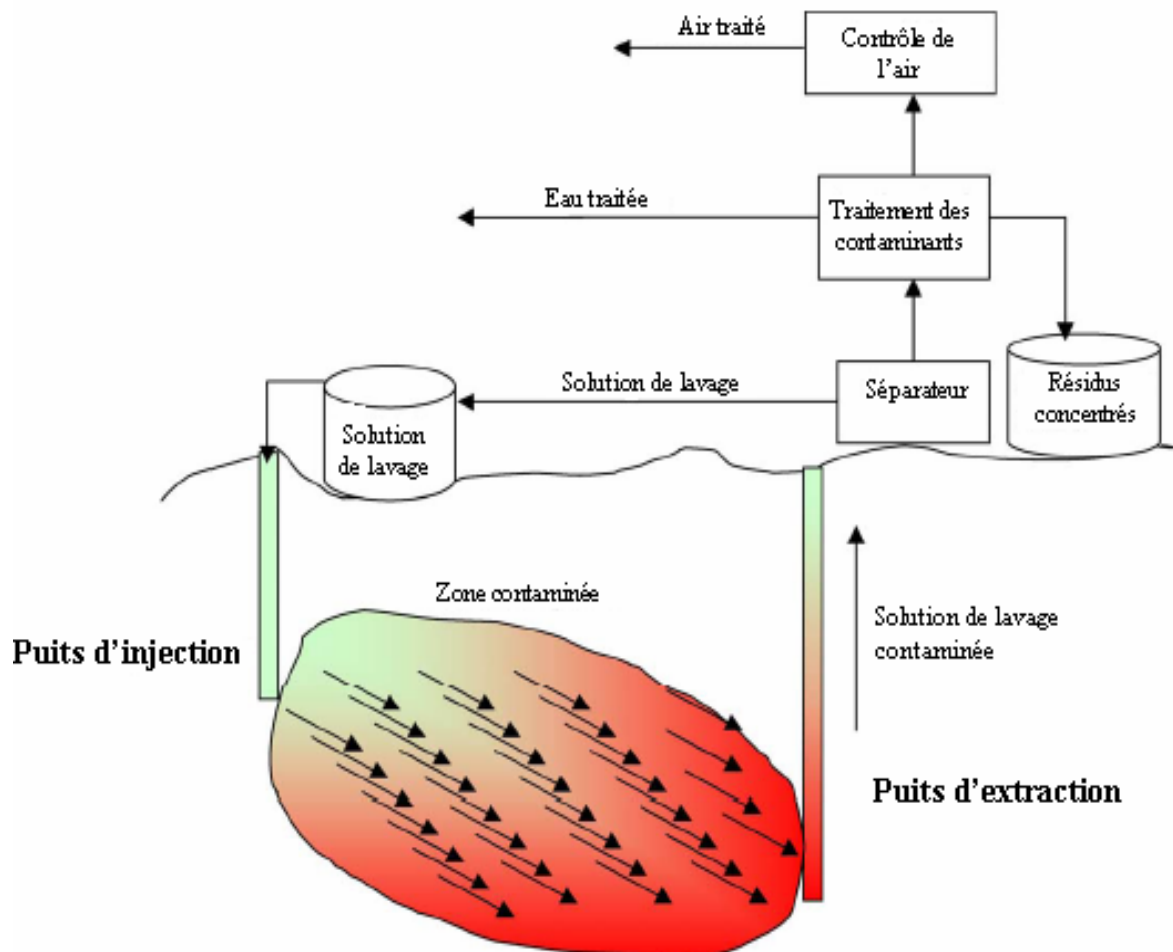


Figure 4 : Principe du lavage *in situ* des polluants du sol. Une solution de lavage est injectée au niveau de la zone polluée. Ce produit solubilise les contaminants et les extrait jusqu'à la surface où ils seront traités (adapté de Khan *et al.*, 2006).

changement d'état permet de séparer le contaminant de la matrice sol. Ces vapeurs sont ensuite collectées afin d'être traitées (Khan *et al.*, 2004). Merino et Bucala (2006) ont démontré *ex situ* l'efficacité de cette technique sur un sol organique pollué par des *n*-hexadécane. Des températures, de 150 à 800° C, ont ainsi permis la vaporisation de 80 à 88 % de la quantité initiale en alcanes.

Le lavage des sols utilise un mouvement de fluides afin d'éliminer les polluants contenus dans des sols non saturés. Ces lavages sont réalisés avec de l'eau mélangée à des agents tensioactifs afin d'augmenter la solubilisation des contaminants. Réalisée *in situ*, cette technique consiste à injecter une solution de lavage au niveau des sites pollués. Les fluides en résultant sont ensuite entraînés vers une zone où ils seront traités (figure 4). Ainsi, 79 % du pétrole brut contenu dans un sol provenant du gisement pétrolier Ekofisk en Norvège ont pu être solubilisés puis dégradés grâce à l'ajout de sodium dodécyl sulfate (SDS), de rhamnolipides et de saponines (Urum *et al.*, 2003). Cependant, comme les hydrocarbures s'adsorbent préférentiellement aux petites particules du sol (Riser-Roberts *et al.*, 1998), ce type de traitement ne permet pas une décontamination totale de l'ensemble des fractions du sol.

4.3. Techniques d'immobilisation.

L'immobilisation des polluants peut être réalisée par solidification/stabilisation (S/S). Cette méthode est basée, d'une part, sur l'emprisonnement des produits toxiques dans des structures de type monolithique afin de diminuer leur transfert au niveau des sols et, d'autre part, sur la transformation chimique des polluants en produits moins mobiles, moins toxiques et moins solubles (Delage et Schrefler, 2005). Il faut noter que ces procédés sont plus adaptés à la remédiation de sols contaminés avec des métaux lourds ou des produits inorganiques. En effet, Karamalidis et Voudrias (2007) ont démontré que des sols, à la fois pollués par des alcanes et des HAPs et ayant subi une S/S avec du ciment, présentaient une augmentation de la déstabilisation et du lessivage des polluants. Cependant, il existe des méthodes de S/S adaptées aux polluants organiques telles l'« incorporation à l'asphalte » ou la « vitrification ». L'incorporation à l'asphalte est une technique qui consiste à excaver les sols pollués par des hydrocarbures et à les substituer aux granulats dans l'asphalte (Khan *et al.*, 2004). Les mélanges inertes et imperméables alors obtenus, sont ensuite utilisés, par exemple, pour la construction des infrastructures de transport. La vitrification est une technique qui

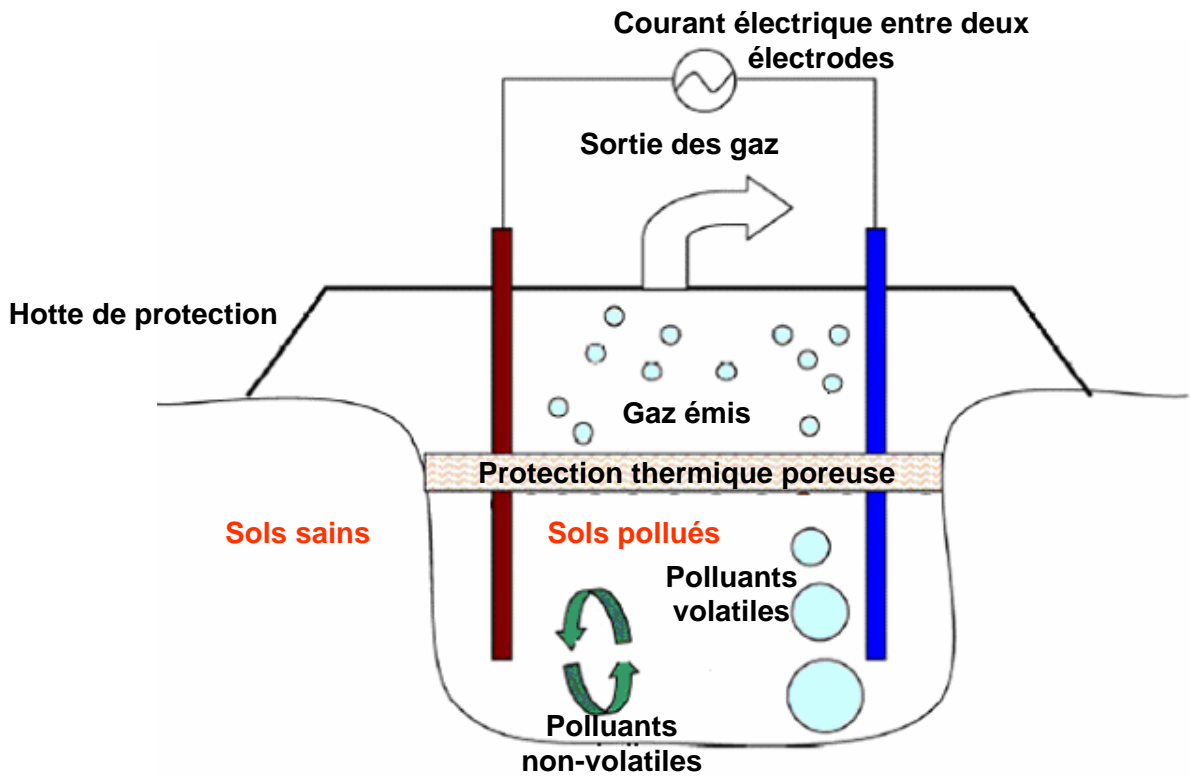


Figure 5 : Principe de la vitrification *in situ* des polluants du sol. La vitrification est une technique qui consiste à transformer, à hautes températures, les sols pollués en pâte de verre inerte. Cette fusion à hautes températures entraîne simultanément une dégradation des composés organiques et un piégeage des composés inorganiques (adapté de Khan *et al.*, 2006).

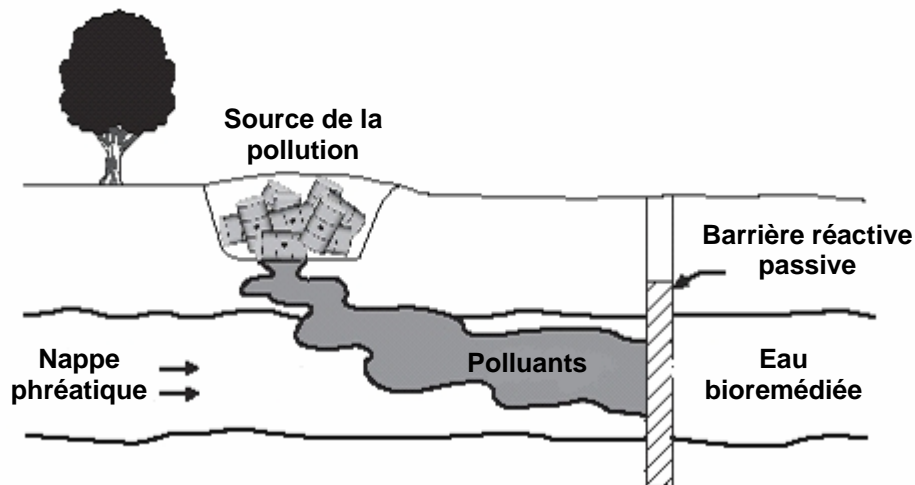


Figure 6 : Principe des barrières réactives passives (BRS). Les liquides qui traversent la zone polluée et se contaminent sont neutralisés à leur sortie en traversant la BRS.

consiste à transformer, à haute température, les sols pollués en pâte de verre inerte (figure 5). Cette fusion effectuée entre 1 600 et 2 000° C entraîne simultanément une dégradation des composés organiques et un piégeage des composés inorganiques (Acar et Alshawabkeh, 1993).

L'immobilisation des polluants peut aussi être réalisée grâce à un isolement physique des sols pollués. C'est le cas de l'encapsulation qui consiste à isoler la zone contaminée grâce à des systèmes de parois imperméables. Ainsi, les substances polluantes ne peuvent plus être entraînées par les eaux de ruissellement au niveau de zones saines ou de nappes phréatiques (Khan *et al.*, 2004). Contrairement à l'encapsulation, les barrières réactives passives (BRS) ne bloquent pas totalement le passage des fluides (figure 6). Leur principe est basé sur l'action de produits oxydo-réducteurs ou de sorption présents au niveau de ces parois. Ainsi, les liquides contaminés sont neutralisés lors de leur passage à travers la BRS (Kalin, 2004).

5. REHABILITATION BIOLOGIQUE DES SOLS POLLUES PAR DES HYDROCARBURES.

Les techniques d'extraction, de destruction et d'immobilisation, qui ont été mises en place afin de décontaminer les sites pollués, peuvent être onéreuses et très invasives pour les écosystèmes traités. En effet, réalisées *ex situ*, elles nécessitent l'excavation des terres à traiter et, réalisées *in situ*, elles requièrent la mise en place d'un complexe de traitement conséquent. Maier *et al.* (2000) ont estimé que la décontamination des terres polluées par des hydrocarbures, rien qu'aux Etats-Unis, pourrait entraîner une dépense supérieure à un milliard de dollars. Il faut aussi noter que ces techniques peuvent entraîner le rejet de constituants qui peuvent être plus toxiques que les polluants eux-mêmes. Aussi, depuis plusieurs années, des solutions alternatives de décontamination voient le jour. Elles utilisent les capacités naturelles des plantes et des microorganismes à dépolluer les sols.

5.1. La phytoremédiation

La phytoremédiation est l'ensemble des technologies utilisant les plantes pour réduire, dégrader ou immobiliser naturellement les polluants des écosystèmes terrestres, aquatiques ou aériens (figure 7). Bien que cette méthode soit principalement utilisée pour les métaux lourds

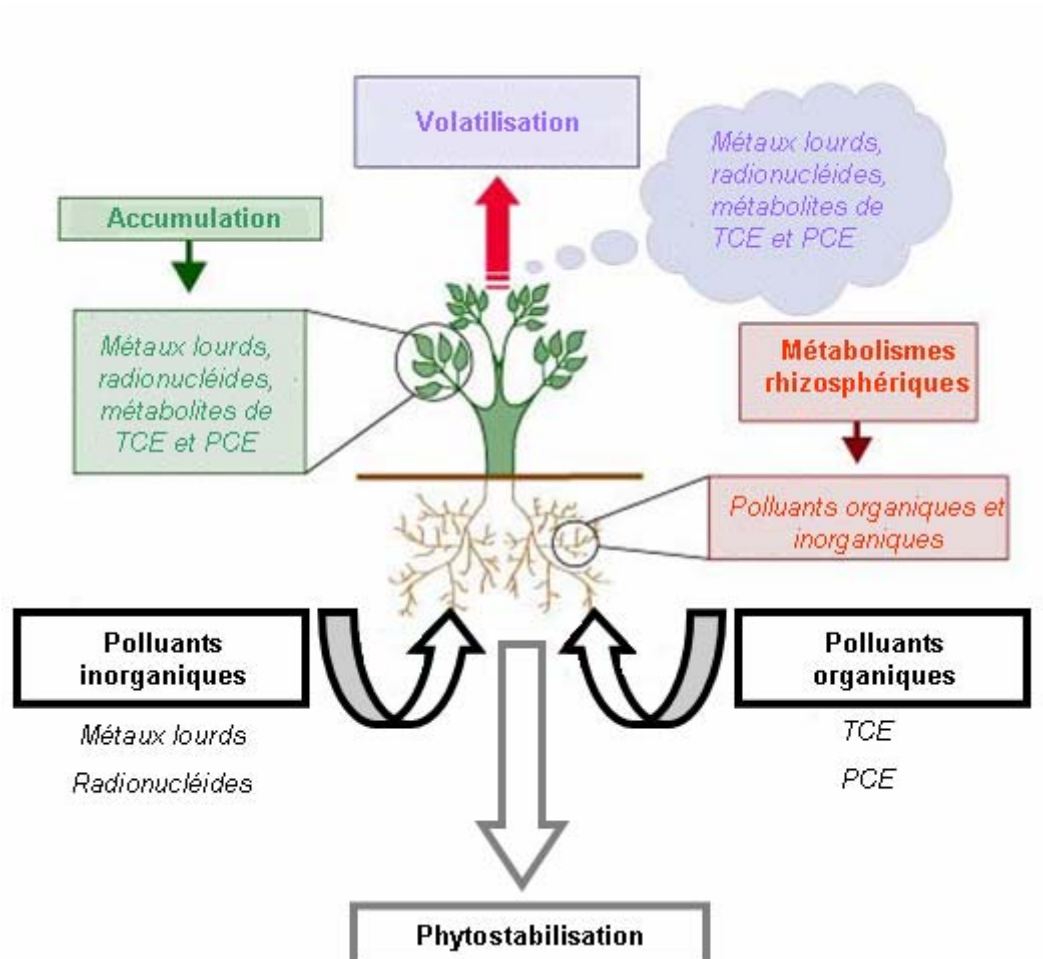


Figure 7 : Principe de la phytoremédiation. La bioremédiation par les plantes a lieu à travers des interactions complexes entre les végétaux, les microorganismes et le sol par des procédés d'accumulation, de volatilisation, de dégradation et de stabilisation. Bien que cette méthode soit majoritairement utilisée pour les métaux et les radionucléides, elle peut aussi s'appliquer au traitement des polluants organiques. *TCE* : trichloréthylène ; *PCE* : perchloréthylène.

(plomb, cadmium, arsenic, chrome, mercure, cuivre et zinc) et les radionucléides (uranium, strontium, césium et plutonium), elle peut également s'appliquer au traitement des polluants organiques (Salt *et al.*, 1998). Elle peut être réalisée selon deux procédés naturels différents : la phytostabilisation et la phytoextraction (Krämer, 2005). La phytostabilisation consiste à couvrir le site à décontaminer de plantes tolérantes aux polluants. Ainsi, ce couvert végétal permet de diminuer les processus d'érosion et d'immobiliser les polluants au niveau de la rhizosphère. La phytoextraction consiste, quant à elle, à cultiver sur les sites contaminés des végétaux qui peuvent puiser puis accumuler les polluants au niveau de leurs tissus aériens. Par exemple, Srivasta *et al.* (2006) ont mis en évidence les capacités hyperaccumulatrices de l'arsenic (As) de trois fougères appartenant au genre *Pteris* : *P. biaurita*, *P. quadriaurita* et *P. ryukyuensis*. Leur croissance sur des sols pollués par 100 mg d'As/kg a permis l'accumulation de 1 770 à 3 650 mg d'As/kg au niveau des frondes et de 182 à 507 mg d'As/kg au niveau des racines. Les plantes épuratrices sont ensuite récoltées puis incinérées afin de récupérer les matières résiduelles fortement concentrées en polluants (Van Nevel *et al.*, 2007). Cette dernière méthode semble plus efficace pour réhabiliter les sites contaminés. Cependant, elle peut être limitée par plusieurs facteurs comme la biodisponibilité des polluants, le prélèvement par les racines, la mobilité à travers la plante et l'efficacité du stockage et de la détoxification au niveau de ses parties aériennes (Clemens *et al.*, 2002). Des approches de génie génétique ont d'ailleurs été récemment développées afin d'améliorer les performances de ces plantes épuratrices. Bañuelos *et al.* (2007) ont notamment créé deux lignées transgéniques de *Brassica juncea* afin d'améliorer les capacités de ces plantes à phytoremédier les sols contaminés au sélénium. Grâce à la surexpression des gènes codant pour une sélénocysteine lyase et une sélénocysteine méthyltransférase, ces organismes génétiquement modifiés (OGM) ont permis d'accumuler jusqu'à deux fois plus de sélénium.

Même si ces dernières avancées technologiques montrent des résultats encourageants, l'utilisation de la phytoremédiation présente toutefois des inconvénients non négligeables. Le premier est la limitation de la décontamination à la zone racinaire des plantes qui est souvent peu étendue. L'utilisation d'arbres pourrait permettre d'améliorer ce point mais leur faibles capacités à accumuler les polluants nécessiterait la mise en place de modifications génétiques (Lyyra *et al.*, 2007). Le second inconvénient est une contamination potentielle de la chaîne alimentaire *via* une consommation accidentelle des plantes épuratrices par des herbivores au niveau des sites traités. Enfin, les polluants peuvent être réintroduits dans l'environnement du

fait de la chute, au niveau du sol, des organes aériens des plantes accumulatrices (Van Nevel *et al.*, 2007).

5.2. La bioremédiation.

Une autre technique, utilisant les capacités épuratrices naturellement présentes chez certains organismes, est la bioremédiation. Dans ce cas, ce sont les microorganismes, présents au niveau des environnements contaminés qui sont utilisés pour dégrader les polluants. Les bactéries prennent une part importante dans cette décontamination naturelle, du fait de leurs capacités à évoluer très rapidement en présence d'une pression de sélection. En effet, grâce aux mutations ponctuelles, aux réarrangements endogènes et aux transferts horizontaux, elles peuvent s'adapter à la présence de polluants en développant une machinerie enzymatique permettant de dégrader et/ou d'utiliser ce polluant pour leur survie et leur développement. Ainsi, quelque temps après l'apparition de molécules xénobiotiques ne possédant aucun équivalent dans la nature, on peut voir l'apparition de nouvelles voies métaboliques permettant la dégradation de ce composé. Un exemple est le cas du lindane, pesticide organochloré créé en 1946, dont la biodégradation anaérobie a été découverte dans les années 1960 (pour revue voir Phillips *et al.*, 2005). Boubakri *et al.* (2006) ont également simulé *in vitro* la création d'un gène impliqué dans la dégradation du lindane (*linA*) à partir de fragments métagénomiques extraits de sols sains par brassage de gènes (DNA shuffling). Ainsi, les microorganismes, grâce à leur réservoir génétique immense et à la sélection naturelle, peuvent s'adapter à un large spectre de polluants. Ils constituent, par conséquent, une voie intéressante dans le développement de techniques naturelles d'élimination des polluants. Cette bioremédiation peut être réalisée sous différentes formes : l'atténuation naturelle, la biostimulation et la bioaugmentation.

5.2.1. L'atténuation naturelle.

L'atténuation naturelle est un procédé qui utilise les capacités épuratrices des microorganismes présents au niveau des écosystèmes pollués. Même si cette technique de décontamination ne nécessite pas théoriquement l'intervention humaine, il est toutefois nécessaire d'éliminer ou de neutraliser la source de pollution et de surveiller en permanence le site jusqu'à la fin du traitement (Mulligan *et al.*, 2004). Ce type de bioremédiation est très peu onéreux car il ne réclame pas beaucoup de moyens mais il nécessite de longues durées de

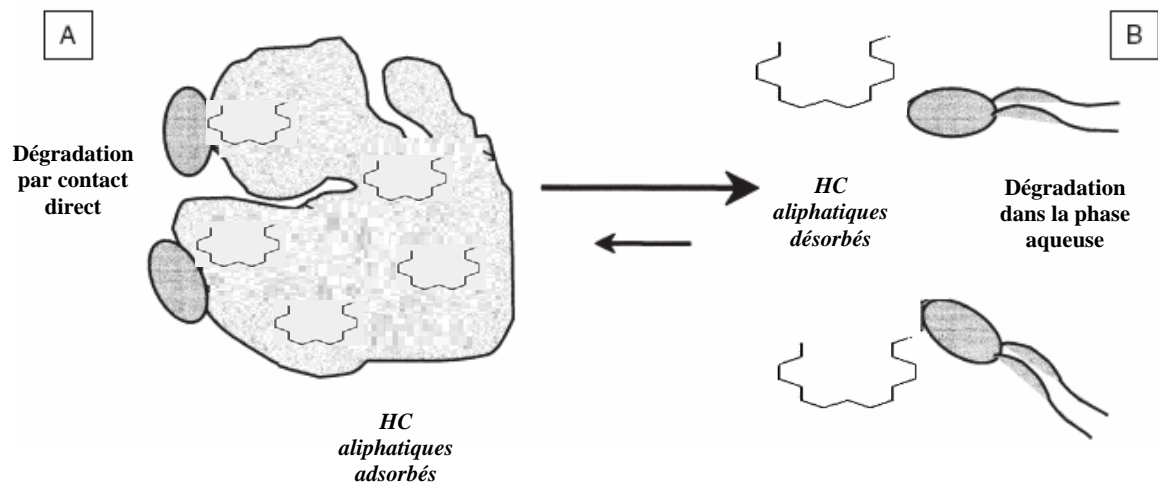


Figure 8 : Dégradation microbienne des polluants aliphatiques au niveau du sol. Les communautés microbiennes peuvent attaquer les polluants adsorbés aux fractions du sol (A) ou ceux biodisponibles au niveau des pores du sol (B) (adapté de Semple *et al.*, 2003).
 HC : Hydrocarbures.

traitement. Ainsi, Serrano *et al.* (2007) ont montré que ce procédé, réalisé en pilote de bioremédiation, permettait une restauration en 200 jours d'un sol pollué par 2 700 ppm d'hydrocarbures. Cette dégradation, comme dans tout procédé d'atténuation naturelle, ne peut cependant pas être attribuée uniquement aux microorganismes. En effet, des procédés abiotiques, tels que l'évaporation, la dissolution, la dispersion, l'émulsification, l'adsorption ou la photooxydation, peuvent aussi entrer en jeu.

5.2.2. La biostimulation.

Les capacités intrinsèques d'un environnement à pouvoir réaliser une atténuation naturelle sont corrélées à plusieurs paramètres. Le premier est la biodisponibilité du polluant (figure 8) qui peut être affectée par des liaisons aux fractions minérales et organiques du sol (cf. chapitre I ; § 3). Cette sorption peut être influencée par la quantité de matière organique, le pH, la minéralogie, la température et le type de polluant. Le second paramètre est l'environnement oxydo-réducteur de l'écosystème. En effet, la disponibilité en accepteurs ou donneurs d'électrons peut influencer les activités microbiennes. Enfin, la biodisponibilité des nutriments (azote et phosphore) et les associations microbiennes dans l'écosystème peuvent aussi entraîner un contrôle de l'atténuation naturelle (Roling et van Verseveld, 2002). La biostimulation va donc consister à améliorer les processus de biodégradation en modifiant les conditions environnementales par adjonction de surfactants, de nutriments, d'eau, de donneurs ou d'accepteurs d'électrons. Ainsi, Perfumo *et al.* (2007) ont étudié les conséquences de l'augmentation de la température, de l'ajout d'azote, de phosphate, de potassium inorganique et de surfactant sur la bioremédiation d'un sol pollué par des hexadécane. Après 40 jours de traitement, les terres biostimulées par l'ajout d'azote, de phosphore et de potassium inorganiques montrent 10 % de dégradation en plus comparées aux terres n'ayant subi qu'une atténuation naturelle.

5.2.3. La bioaugmentation.

Parfois, les populations microbiennes endogènes d'un écosystème pollué ne présentent pas tous les outils métaboliques permettant de réaliser une dégradation complète des polluants. La bioaugmentation permet de remédier à ce manque (El Fantroussi et Agathos, 2005). Elle consiste en l'incorporation, au niveau des écosystèmes contaminés, de microorganismes possédant des métabolismes adaptés. Cet apport peut se faire par l'ajout d'une souche seule

ou d'un groupe de souches appelé *consortia*. Ainsi, Da Silva et Alvarez (2004) ont démontré une amélioration de la dégradation anaérobie d'un mélange Benzène/Toluène/Ethylbenzène/Xylène (BTEX) et éthanol, contenu dans un aquifère, suite à l'apport d'un *consortia* méthanogène. Cependant, de nombreux facteurs, surtout biotiques (prédation, compétition), peuvent mettre en péril l'efficacité de telles techniques en conduisant au déclin des populations introduites (van Veen *et al.*, 1997). Afin de pallier ces contraintes écologiques, des systèmes de protection physique des microorganismes ensemencés, tel l'encapsulation en gellan gum, peuvent être mis en place (Moslemy *et al.*, 2002). Le génie génétique peut aussi permettre de résoudre ces problèmes de stabilité des souches exogènes. Dans ce contexte, des microorganismes endogènes peuvent être modifiés en leur apportant les gènes nécessaires à l'amélioration de leurs capacités épuratrices. Ainsi, Watanabe *et al.* (2002) ont montré une amélioration de la dégradation du phénol grâce à l'ajout d'une souche endogène de *Comamonas* génétiquement modifiée par introduction de gènes codant pour une phénol hydroxylase. Certaines études ont, de plus, démontré la possibilité d'améliorer l'efficacité de la bioremédiation en bioaugmentant les écosystèmes pollués par des gènes d'intérêt. Dans ce cas, les transferts de gènes entre les bactéries donneuses de plasmides cataboliques et les populations endogènes permettent la biodégradation des polluants (Bathe *et al.*, 2005).

CONCLUSIONS

Les procédés de bioremédiation sont envisagés comme d'intéressantes alternatives aux techniques classiques de dépollution. En effet, ils offrent la possibilité de réaliser une décontamination d'environnements pollués en diminuant les impacts sur les écosystèmes traités. Afin d'améliorer encore ces procédés de dépollution, il est cependant nécessaire de comprendre le mode de fonctionnement de ces communautés microbiennes vis à vis du polluant. De plus, l'utilisation du vivant nécessite une surveillance des microorganismes une fois la dépollution naturelle initiée, afin de s'assurer que le « biosystème » ne dévie pas de sa fonction initiale. De part leur organisation en *consortia* et leurs besoins spécifiques, les microorganismes sont cependant difficilement cultivables et leur étude par les techniques microbiologiques classiques reste par conséquent limitée. Ainsi, l'identification et le suivi des souches bactériennes présentant un intérêt en bioremédiation passe souvent par la mise en oeuvre de techniques moléculaires.

Chapitre II

Identification des microorganismes : apport des techniques moléculaires.

INTRODUCTION

Depuis le développement de l'ancêtre du microscope et les observations par Antonie van Leeuwenhoek en 1676, de nombreux chercheurs tendent à développer des techniques permettant de cultiver et d'identifier les microorganismes. Un des premiers fut Louis Pasteur qui, en 1861, développa les techniques de pasteurisation et de stérilisation qui permettront dès lors d'isoler les microorganismes et de les étudier en cultures pures. Cette avancée technique majeure lui permettra de réfuter la théorie de la génération spontanée. A la fin du XIXème siècle, Robert Koch et Hans Christian Gram ont mis au point des techniques de coloration permettant l'identification de microorganismes. R. Koch mettra ainsi en évidence le bacille responsable de la tuberculose (*Mycobacterium tuberculosis*). La technique développée par H.C. Gram permettra, quant à elle, d'élaborer une première classification entre les bactéries en fonction de la composition de leur paroi. A cette même période, J.R. Petri inventa la boîte de Petri facilitant permettant la culture en conditions stériles des microorganismes. Ces découvertes majeures vont être les fondements de la microbiologie moderne qui sera, dès le début du XXème siècle, révolutionnée par la caractérisation des activités enzymatiques des microorganismes. La découverte de métabolismes spécifiques de microorganismes permettra par la suite d'élaborer, dans les années 1960, des systèmes de classification des microorganismes en fonction de leurs caractères métaboliques plutôt que par rapport à leurs caractères phénotypiques (Pelczar *et al.*, 1993).

Jusqu'à récemment, la mise en culture était une étape obligatoire dans l'identification des microorganismes par l'étude de leur physiologie et de leurs capacités métaboliques. Or, les techniques de culture par enrichissement, utilisées classiquement, ne permettent d'isoler qu'une faible fraction des populations microbiennes. En effet, seules les bactéries capables de

pousser rapidement sur des milieux riches ont pu être identifiées (Leadbetter, 2003). Torsvik *et al.* (2002) ont notamment estimé, grâce à une comparaison du nombre de bactéries comptées sur boîtes et observées en microscopie à fluorescence, que seulement 0,1 à 1 % des bactéries du sol pouvaient être cultivés par des techniques classiques d'isolement. C'est pourquoi les microbiologistes ont développé de nouvelles méthodes de culture utilisant des milieux pauvres en nutriments, proposant une grande diversité de substrats ou simulant les conditions environnementales (Ferrari *et al.*, 2005). De nouveaux outils tels les micromanipulateurs, les « pinces à laser » et les capsules de cultures (Joseph *et al.*, 2003) ont également été mis au point afin de tenter d'accéder à l'ensemble de la diversité microbienne. Même si ces nouvelles techniques de culture représentent un réel potentiel et une avancée considérable permettant d'isoler de nouveaux microorganismes, elles s'avèrent cependant encore inadaptées pour appréhender la totalité de la biodiversité des écosystèmes. C'est avec le développement des techniques moléculaires que l'identification des microorganismes non cultivés va prendre un réel essor et ainsi permettre d'avoir une vision plus globale de la diversité microbienne.

1. LA REVOLUTION MOLECULAIRE EN ECOLOGIE MICROBIENNE.

La découverte de l'ADN par O. Avery, C. MacLeod et M. McCarthy en 1944 a été un tournant pour la microbiologie moderne. Au milieu du XX^{ème} siècle, de nouvelles techniques basées sur les informations contenues par cette molécule ont vu le jour. Ces nouvelles approches de phylogénie moléculaire reposent sur la comparaison de séquences génomiques, essentiellement celles codant pour les ARN ribosomiques (ARNr). Dès les années 70, le précurseur de la phylogénie moléculaire, Carl Woese, les a définies comme étant les « chronomètres moléculaires universels ». En effet, ces gènes présentent plusieurs avantages faisant d'eux de bons biomarqueurs. Ils sont présents chez tous les microorganismes, souvent en plusieurs copies dans le génome, et ne subissent que rarement des transferts horizontaux et des recombinaisons (Miller *et al.*, 2005). De plus, comme la synthèse des protéines est une fonction universelle chez tous les microorganismes, leurs séquences présentent à la fois des régions très conservées qui permettent de les isoler facilement et de faire des études de phylogénies longues distances et des régions hypervariables qui assurent une signature

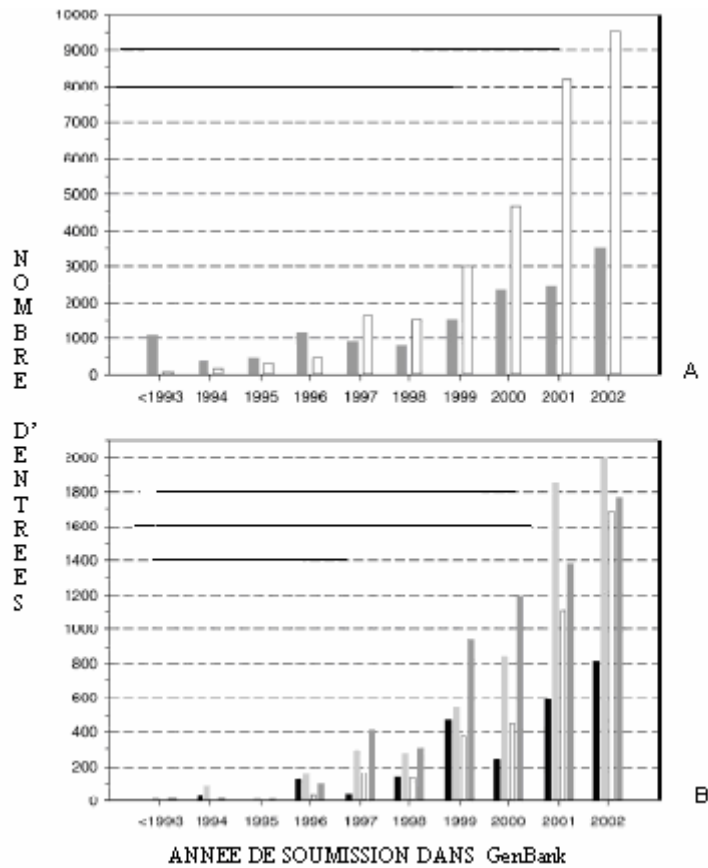


Figure 9 : Histogrammes représentant le nombre de séquences ADNr 16S soumis dans GenBank depuis 1993. Sur le premier histogramme (A) sont reportés en gris les séquences obtenues à partir de microorganismes cultivés et en blanc les séquences obtenues à partir de clones de l'environnement. Sur le deuxième histogramme (B) sont reportées les séquences obtenues à partir de clones de divers écosystèmes : en noir, des sédiments, en gris clair, le sol, en blanc, l'eau de mer, et en gris foncé, l'eau douce (d'après Rappé et Giovannoni, 2003).

spécifique d'espèces ou de genres microbiens permettant ainsi des études phylogénétiques courtes distances (Amann *et al.*, 1995).

Les premières analyses phylogéniques, basées sur ces séquences d'ARNr 16S, ont permis de mettre en évidence que les procaryotes pouvaient être classés en deux groupes : les *Eubacteria* et les *Archaeobacteria* (Woese et Fox, 1977). Suite à ce travail, une première classification moléculaire des microorganismes a été réalisée par Woese (1987). Elle répertorie la vie cellulaire en trois domaines : *Eukaryota*, *Bacteria* et *Archaea*. Concernant plus particulièrement les *Bacteria*, ces approches ont permis la caractérisation de douze divisions majeures. Olsen *et al.* (1986) introduisent l'idée que la fraction non cultivée des microorganismes peut aussi être identifiée par la caractérisation de ces séquences. C'est ainsi que la diversité microbienne des écosystèmes a commencé à être étudiée par une approche de clonage/séquençage de fragments d'ADN génomique environnementaux contenant ces gènes ribosomiaux. Cette méthode puissante est malgré tout très laborieuse. C'est pourquoi des techniques, visant à cibler directement les gènes d'intérêt, ont été développées.

Même si l'ARNr 16S (~1.500 nucléotides) ne contient environ que deux fois moins d'informations que l'ARNr 23S (~3.000 nucléotides), il est actuellement le biomarqueur phylogénétique bactérien de référence. D'une part, il est plus facile à cloner et à séquencer et d'autre part, les bases de données sur les ARNr 16S sont les plus denses et ne cessent d'augmenter (figure 9). Ainsi, la base de données du Ribosomal Database Project II (<http://rdp.cme.msu.edu/>) contient 440 891 séquences dans sa version 9.55 de Octobre 2007.

1.1. Méthodes basées sur l'amplification PCR.

Grâce à la réaction de polymérisation en chaîne (PCR), des gènes d'intérêts (également appelés biomarqueurs) peuvent être spécifiquement amplifiés à partir de l'ADN génomique des communautés microbiennes étudiées. Il en résulte cependant un mélange complexe de fragments provenant de l'ensemble des espèces présentes dans le milieu qu'il est nécessaire d'analyser plus finement.

1.1.1. Clonage/séquençage des produits PCR.

La méthode la plus directe est le clonage/séquençage de ces produits PCR afin d'identifier, notamment par méthode comparative ou phylogénie moléculaire, à quelles espèces appartiennent ces fragments (Amann *et al.*, 1995 ; Pace, 1997 ; Giovanonni *et al.*, 1990).

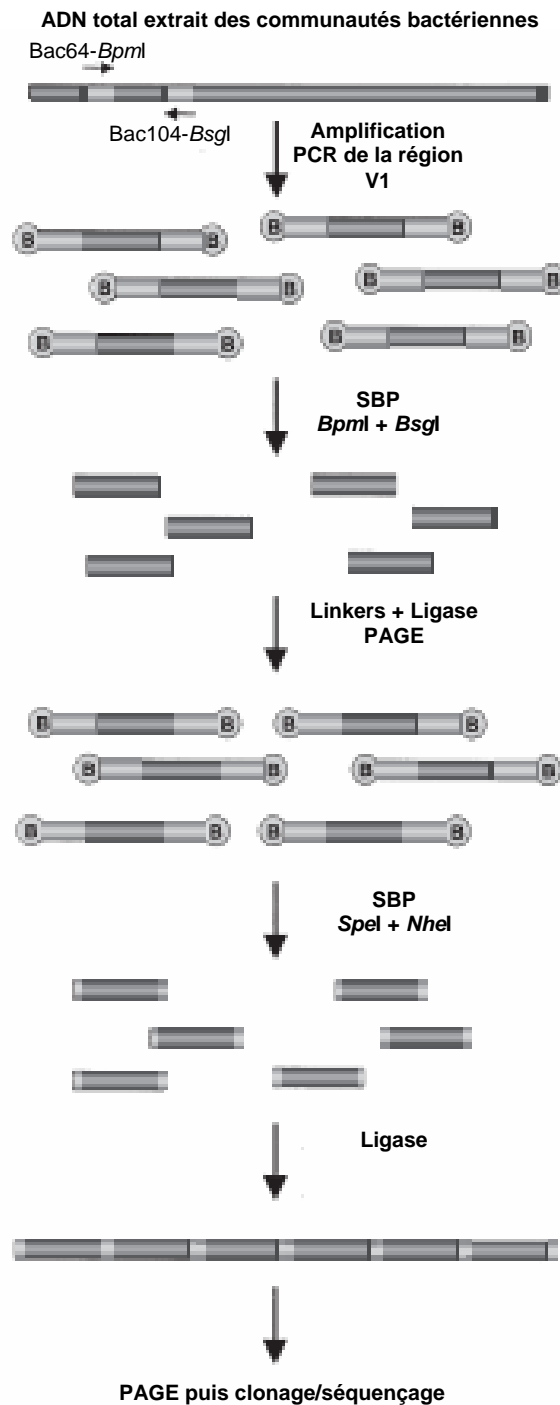


Figure 10 : Principe de la technique SARST (Serial Analysis of Ribosomal Sequence Tags). Les purifications en gel de polyacrylamide (PAGE) et les purifications par des billes de streptavidine (SBP) utilisant le marquage à la biotine (B) permettent la purification et la séparation des fragments qui seront concaténés pour être séquencés en série (adapté de Neufeld *et al.*, 2004).

Cependant, il est très difficile d'étudier l'ensemble de la biodiversité par cette technique car les communautés microbiennes sont très diversifiées. D'après Kemp *et al.* (2004), les études utilisant le clonage/séquençage des ADNr 16S ne permettent d'appréhender qu'environ 48 % de la diversité procaryotique des écosystèmes étudiés du fait d'un effort de séquençage généralement insuffisant. La technique Serial Analysis of Ribosomal Sequence Tags (SARST) pourrait permettre d'améliorer ces performances en terme de couverture de la biodiversité (figure 10). Elle consiste en la concaténation de courtes régions hypervariables de l'ADNr 16S (souvent V1 ou V6) pouvant provenir de différentes espèces. Ces séquences composites sont introduites dans un vecteur de clonage puis séquencées. Etant donné la taille de ces régions variables (17 à 55 pb), le séquençage de chaque clone permet l'identification de différentes espèces (Neufeld *et al.*, 2004).

1.1.2. Techniques d'empreintes génétiques.

Afin d'avoir une vue plus globale de la diversité des écosystèmes étudiés, des techniques dites d'empreintes génétiques ont été développées. Elles consistent en la séparation du mélange des produits PCR permettant une visualisation directe de la diversité des populations microbiennes. La plus simple d'entre-elles, du point de vue technique, est l'Amplified Ribosomal DNA Restriction Analysis (ARDRA). Elle repose sur une digestion des produits PCR par différentes enzymes de restriction puis sur une migration des fragments obtenus en gel d'agarose. Cette technique permet d'identifier des profils spécifiques d'espèces microbiennes (Wu *et al.*, 2006) mais son utilisation pour étudier des environnements complexes reste délicate en raison du nombre important de bandes obtenues (Dahllof, 2002).

Afin de diminuer la complexité des profils, la Terminal Restriction Fragment Length Polymorphism (T-RFLP) ne prend en compte que les fragments terminaux obtenus grâce à l'utilisation d'amorces marquées par un fluorochrome lors de l'étape d'amplification PCR (Liu *et al.*, 1997) (figure 11). La diminution de la complexité des profils peut aussi passer par des techniques permettant la migration différentielle des produits PCR de même taille en fonction de critères tels leur composition ou leur conformation. La DGGE et la TGGE permettent la séparation de produits PCR, en fonction de leur composition nucléotidique, grâce à une migration électrophorétique le long d'un gradient linéaire, respectivement d'agents dénaturants ou de température (Muyzer *et al.*, 1993) (figure 12). Dans la technique Single-Strand Conformation Polymorphism (SSCP), la séparation électrophorétique des produits PCR se fait grâce aux différentes structures tridimensionnelles que peuvent adopter

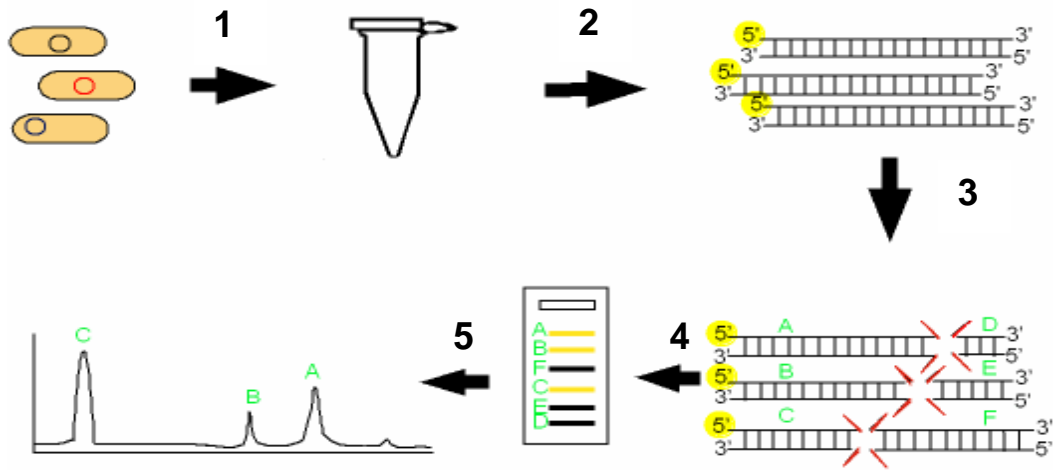


Figure 11 : Principe de la technique T-RFLP (Terminal Restriction Fragment Length Polymorphism). Les ADN des communautés microbiennes sont extraits des échantillons environnementaux (1). Les gènes d'intérêt sont ensuite amplifiés par PCR en utilisant une amorce marquée par fluorescence (2). Après purification, les produits PCR sont digérés avec des enzymes de restriction qui génèrent des fragments de tailles différentes (3). Ceux-ci sont séparés par électrophorèse (4) et les fragments présentant une fluorescence sont reconnus (5) (adapté de Grüntzig *et al.*, 2002 ; rdp8.cme.msu.edu).

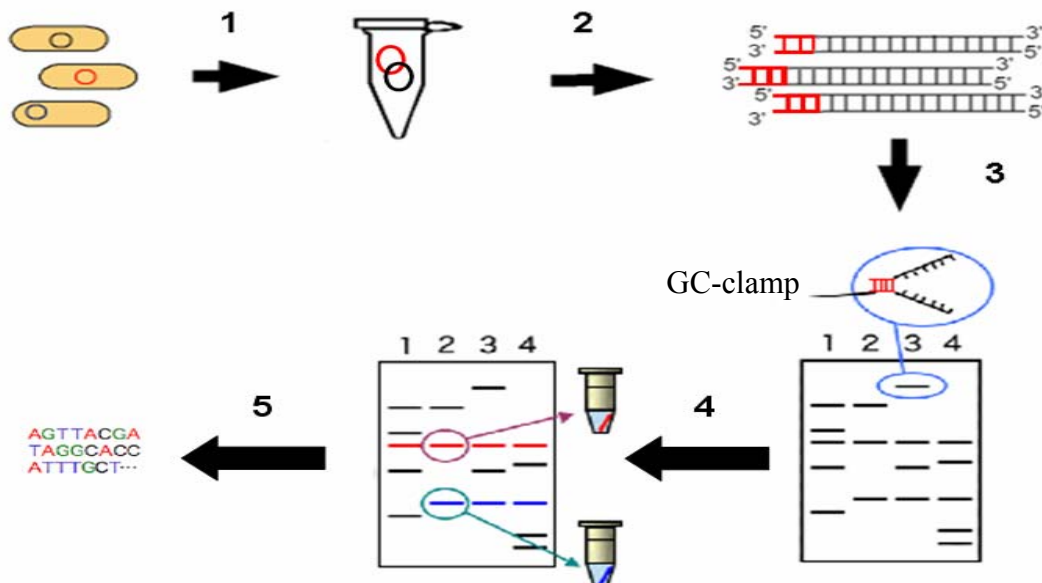


Figure 12 : Principe de la technique DGGE ou TGGE (Denaturing et Thermal Gradient Gel Electrophoresis). Les ADN des communautés microbiennes sont extraits des échantillons environnementaux (1). Les gènes d'intérêt sont ensuite amplifiés par PCR en utilisant une amorce couplée à un GC-clamp qui est une région riche en GC résistante à la dénaturation (2). Après purification, les produits PCR sont séparés par électrophorèse, en fonction de leur composition, selon un gradient de dénaturation chimique ou thermique (3). Les fragments, correspondant « théoriquement » à une espèce, peuvent être excisés (4) afin d'être séquencés (5).

les simples brins d'ADNr en conditions non dénaturantes. Ces variations de conformations sont induites par les différences de séquences des ADNr en fonction des espèces (Lee *et al.*, 1996). Enfin, la RISA permet une discrimination des fragments amplifiés, correspondant aux régions intergéniques (ITS), en fonction de leur taille (de 50 à 1 500 pb en fonction des espèces bactériennes) (Ranjard *et al.*, 2001).

1.2. Méthodes basées sur l'hybridation.

Les méthodes basées sur l'hybridation d'acides nucléiques permettent la caractérisation de microorganismes en s'affranchissant des biais liés à l'amplification PCR. La première d'entre-elles est la technique de réassociation de l'ADN qui permet de mesurer la complexité génétique d'une communauté. Dans une expérience de réassociation, l'ADN total extrait à partir d'échantillons environnementaux est dénaturé par la chaleur puis réassocié. Le taux de réassociation dépend alors de la similarité des séquences présentes dans le mélange (figure 13). Ainsi, lorsque la diversité des séquences d'ADN augmente, le taux de réassociation diminue (Torsvik *et al.*, 1990). Cependant, l'inconvénient de cette méthode réside dans le fait qu'elle manque de sensibilité et qu'elle ne permet pas de définir les espèces présentes dans l'échantillon étudié.

La méthode d'hybridation *in-situ* par fluorescence (FISH) permet, quant à elle, d'avoir une image qualitative et quantitative des populations microbiennes au sein de l'écosystème étudié (Amann *et al.*, 1990 ; DeLong *et al.*, 1989). Cette technique repose sur l'utilisation de sondes oligonucléotidiques marquées par un fluorochrome qui sont complémentaires de zones variables de l'ARNr des microorganismes (figure 14). Afin d'améliorer la sensibilité de la méthode, des systèmes d'amplification enzymatique du signal, comme la Catalyzed Reporter Deposition-Fluorescence In Situ Hybridization (CARD-FISH), ont été développés (Pernthaler *et al.*, 2002). Cette méthode reste cependant limitée par le nombre de fluorochromes utilisables. En effet, pas plus de sept populations microbiennes peuvent être révélées simultanément (Amann *et al.*, 1996). Ainsi, cette méthode ne permet pas de détecter simultanément de nombreux groupes bactériens.

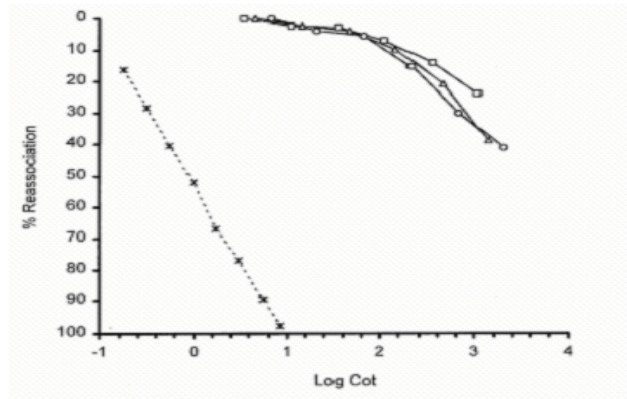


Figure 13 : Cinétique de réassociation d'ADN de différentes communautés microbiennes du sol. Ces cinétiques ont été établies à partir d'ADN extrait de microorganismes de sols de prairies différentes (◻ - ◊ - ◯) ou à partir d'ADN d'*Escherichia coli* (x). Chaque courbe représente la communauté microbienne d'un des sols. Le $Cot_{1/2}$, est un index de mesure du temps nécessaire pour que la moitié de l'ADN de chaque échantillon soit renaturé (d'après Clegg *et al.*, 1998).

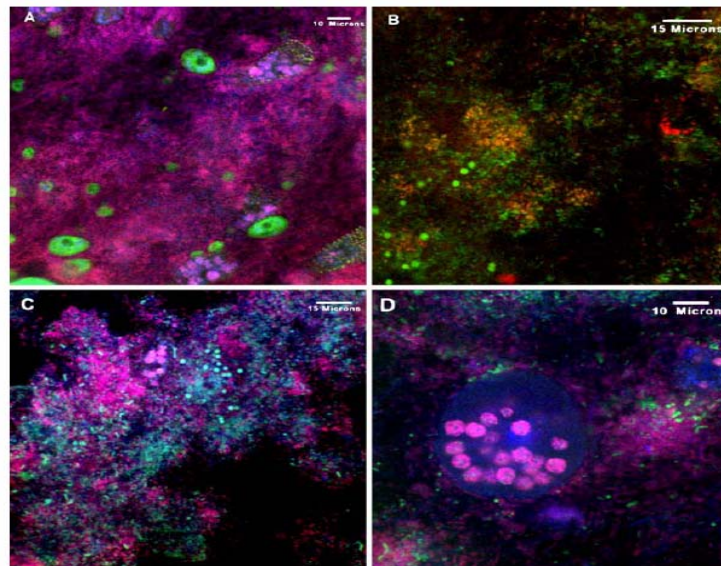


Figure 14 : Identification par hybridation *in situ* des populations bactériennes et eucaryotes d'un biofilm provenant d'un réacteur de traitement des eaux de mer. Cette technique permet de reconnaître et d'observer *in situ* plusieurs populations grâce à des sondes spécifiques présentant des fluorescences différentes. (A) Hybridation avec les sondes Euk516-FITC (vert), Gam42a-CY3 (rouge) et Eub-mix-CY5 (bleu). Les signaux magenta résultent de la fusion des signaux de Gam42a-CY3 et de Eub-mix-CY5. (B) Hybridation avec des sondes Alf1B-FITC (vert) et S*-OTU14-0129-a-A-20-CY3, correspondant aux *Hyphomicrobium* (rouge). (C) Hybridation avec les sondes Alf1B-FITC (vert), MPH-730-CY3 (rouge) et Eub-mix-CY5 (bleu). Les signaux cyan et magenta résultent de la fusion des sondes Eub-mix-CY5 avec Alf1B-FITC ou MPH-730-CY3. (D) Hybridations avec les sondes Alf1B-FITC (vert), MPH-730-CY3 (rouge) et Eub-mix-CY5 (blue). Les signaux magenta résultent de la fusion des signaux de MPH-730 et Eub-mix (adapté de Labbé *et al.*, 2007). Euk516 : plupart des eucaryotes ; Gamma42a : *Gamma Proteobacteria* ; Eub-mix : plupart des bactéries ; Alf1B : *Alpha*, quelques *Delta Proteobacteria* et *Spirochetes* ; MPH-730 : Genre *Methylophaga spp. except natronica*

2. LES NOUVELLES TECHNIQUES A HAUT-DEBIT.

La révolution moléculaire a ouvert des perspectives immenses pour l'étude de la fraction non cultivée des communautés microbiennes des écosystèmes complexes. Néanmoins, les limites liées aux techniques moléculaires dites « classiques » rendent nécessaires le développement d'outils permettant des analyses à haut débit. Parmi eux, on notera surtout le développement des techniques dites de méta«omiques», du pyroséquençage et des biopuces ADN. Ces dernières feront l'objet d'un chapitre à part entière (chapitre III).

2.1. Métagénomique et métatranscriptomique

La métagénomique est l'étude de l'ensemble des génomes contenus au niveau d'un environnement (figure 15). Cette approche consiste en une extraction directe ou indirecte de l'ADN génomique des microorganismes indépendamment de toute étape de mise en culture. En fonction de la taille des fragments générés, ceux-ci sont insérées dans des plasmides, des fosmidés (Lee *et al.*, 2004) ou des BACs (Bacterial Artificial Chromosome) (Rondon *et al.*, 2000). Une fois séquencés, par la méthode de Sanger ou par pyroséquençage, les fragments de génomes sont assemblés et assignés à une souche grâce à la présence potentielle de biomarqueurs, mais également par l'estimation du pourcentage en GC, de fréquences des di-, tri- et tétranucléotides du biais d'usage des codons de chaque séquence. Outre le fait d'identifier de nouvelles espèces microbiennes, la métagénomique permet l'étude de l'immense réservoir génétique microbien des écosystèmes et plus particulièrement de sa fraction non cultivée (Daniel, 2004).

Un des premiers projets de métagénomique a été le séquençage du biofilm des eaux de drainage de la mine de Richemond (Tyson *et al.*, 2004). La faible diversité, due aux conditions environnementales extrêmes, a permis une reconstruction de deux génomes dominants sur les cinq souches présentes dans cet environnement : ceux de *Leptospirillum* group II et de *Ferroplasma* type II. Des informations substantielles pour les autres membres de la communauté procaryotique (*Leptospirillum* group III, *Ferroplasma* type I, *Sulfobacillus thermosulfidooxydans*) et des Eucaryotes ont été obtenues mais n'ont pas permis de reconstituer l'intégralité des génomes. Ces reconstructions ont permis de mettre en évidence, entre-autres, que *Leptospirillum* group II n'était pas impliqué dans la fixation de l'azote,

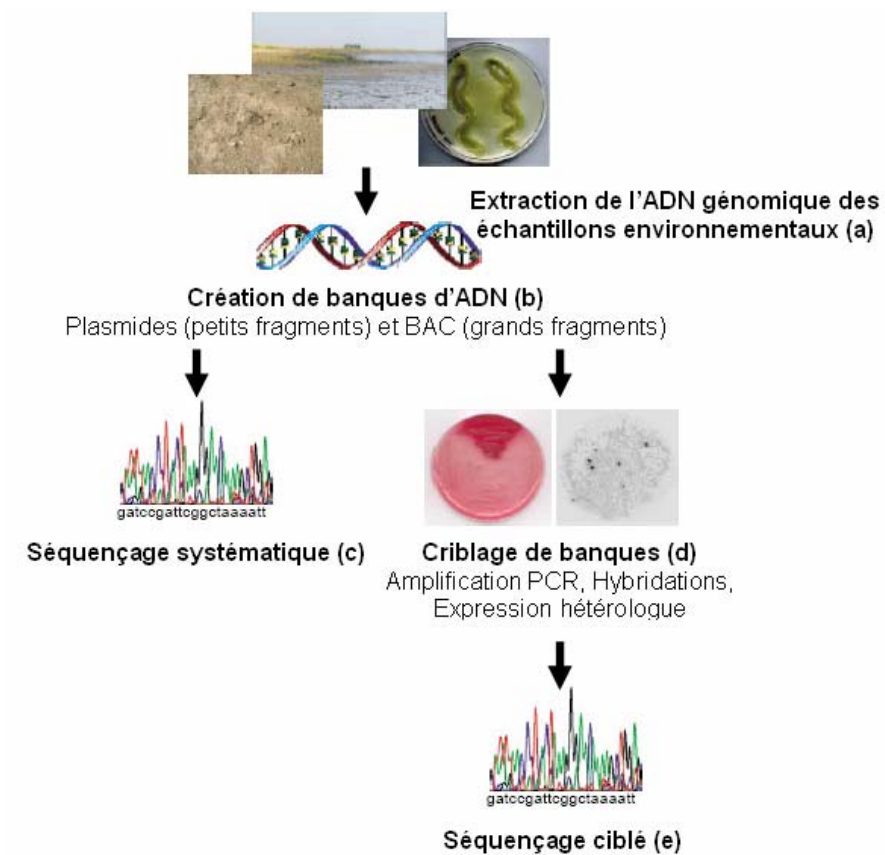


Figure 15 : L'analyse par métagénomique des communautés microbiennes. L'ADN génomique des communautés microbiennes est isolé des écosystèmes étudiés (a) : par exemple, des sols, des sédiments ou de cultures enrichies. Après fragmentation, les grands fragments d'ADN sont insérés dans des vecteurs de clonage (b) puis sont soit séquencés systématiquement (c) soit criblés (d) afin de ne séquencer que les fragments d'intérêt (e) (adapté de Streit et Schimtz, 2004).

comme certaines hypothèses le suggéraient, contrairement à *Leptospirillum* group III. La reconstruction de tous les génomes n'est actuellement pas envisageable dans le cas d'études d'environnements plus complexes. En effet, Tringe *et al.* (2005) ont estimé que 2 à 5 Gigabases (Gb) seraient nécessaires pour obtenir une couverture de génome de 8X et, cela, seulement pour l'espèce dominante d'un sol contenant 10^4 à 10^5 espèces microbiennes. Ainsi, le projet de séquençage de la mer des Sargasses (Venter *et al.*, 2004) qui a produit environ 1,36 Gb n'a permis de reconstituer que quelques génomes complets de procaryotes dont celui de l'espèce dominante appartenant au genre *Burkholderia*. Une des dernières études de métagénomique environnementale dont les résultats ont récemment été publiés par Rusch *et al.* (2007), est le « Global Ocean Sampling Expedition » de la fondation J. Craig Venter. Dans ce projet, 41 échantillons représentatifs d'habitats différents répartis sur 8 000 km entre l'Atlantique Nord et le Pacifique Sud, ont été prélevés et 6,3 Gb ont été séquencées. Une partie de cette étude a permis de mettre en évidence une importante diversité des individus appartenant aux clades majoritairement retrouvés au niveau de ces milieux marins (*Pelagibacter ubiquus*, *Prochlorococcus marinus* et *Synechococcus* WH8102). De plus, la caractérisation de la variation des séquences et des îlots génomiques a montré que jusqu'à 85 % des séquences assemblées étaient uniques.

A l'heure actuelle, de plus en plus d'études métagénomiques voient le jour. La base de données Genome OnLine Database (GOLD; <http://www.genomesonline.org>) recense pas moins de 115 projets de métagénomiques portant sur des sujets très différents, allant de l'étude des communautés fongiques de l'Arctique (Université Alaska Fairbanks) à celle des populations microbiennes contenues dans l'air de New-York avec le projet de l'Institut J Craig Venter (tableau 2). Même si elles ne permettent pas d'assembler tous les génomes, les études métagénomiques permettent de découvrir de nombreux nouveaux gènes. Afin de cibler le séquençage vers les fragments possédant des gènes d'intérêts, il est possible de cribler les banques métagénomiques sur la base de leur séquence ou de leur fonction. Le premier type de criblage est réalisé par amplification PCR ou hybridation avec des sondes spécifiques. Cette approche utilise les ADNr 16S ou encore des gènes conservés déjà connus. Ainsi, Liles *et al.* (2003) ont pu identifier une souche d'*Acidobacterium* non cultivée à partir d'une banque métagénomique de 24 400 clones sur la base d'un criblage par les ADNr 16S. Le second type de criblage est basé sur l'expression, en système hétérologue, de gènes codant pour une nouvelle fonction. Ce criblage de nouveaux gènes peut également s'effectuer sur un substrat d'intérêt qui active la transcription d'un opéron de gènes impliqués dans sa dégradation. Un

Tableau 2. Synthèse exhaustive des projets de métagénomique. A l'heure actuelle, 115 projets sont en cours ou ont été réalisés sur différents environnements (source GOLD : Genome OnLine Database ; <http://www.genomeonline.org>).

Acid Mine Drainage (ARMAN) euryarchaeal community from Richmond mine	Marine anaerobic ammonium oxidizing (anammox) community
Acid Mine Drainage microbial communities from Richmond mine	Marine and Aquatic ammonia oxidizing bacterial (AOB) communities
Air microbial communities from New York	Marine archaeal anaerobic methane oxidation (AOM) communities from Eel River sediments
Air microbial communities from Singapore	Marine Bacterioplankton communities from Antarctic
Alvinella pompejana microbiome from the East Pacific Rise hydrothermal vent	Marine Beggiatoxa community from coastal sediments
Aquatic archaeal communities from Nevada Hot Springs	Marine cyanobacterial communities from coastal water
Aquatic crenarchaeal communities from the Hishikari Goldmine Hot Springs	Marine microbial communities from Bras del Port saltern in Santa Pola Spain
Aquatic dechlorinating community (KB-1)	Marine microbial communities from deep sea hydrothermal vents
Aquatic hyperthermophilic communities from Arctic Springs	Marine microbial communities from English Channel site L4
Aquatic methylotrophic communities from Lake Washington sediments	Marine microbial communities from Nova Scotia coast surface waters
Aquatic microbial communities from Antarctic Lake	Marine microbial communities from Global Ocean Sampling (GOS)
Aquatic microbial communities from Cedars Ultrabasic Springs	Marine microbial communities from Guerrero Negro
Aquatic microbial communities from Drinking-water networks	Marine microbial communities from Monterey Bay
Aquatic microbial communities from Soudan Mine in Minnesota	Marine microbial communities from Near-Shore Anoxic Basin of Saanich Inlet of Vancouver
Aquatic microbial communities from South Africa Goldmine	Marine microbial communities from North Atlantic deep water
Aquatic microbial communities from Tilapia ponds	Marine microbial communities from planktonic and surface waters
Aquatic microbial communities from Yellowstone Bath Hot Springs	Marine microbial communities from Sargasso Sea
Aquatic microbial communities from Yellowstone Bison Hot Spring Pool	Marine microbial communities from sediment
Aquatic microbial community from Columbia River	Marine microbial communities from the Colne Estuary sediments
Aquatic microbial community from Contaminated Groundwater	Marine microbial community from Georgetown SC saltmarsh estuary
Aquatic microbial community from New Orleans Floodwaters	Marine microbial community from San Diego solar saltens
Aquatic microbial community from Yellowstone Hot Springs	Marine oligotrophic communities from Northwestern Sargasso Sea
Aquatic viral communities from Yellowstone Hot Springs	Marine phototrophic picocaryotic community
Benzene-Degrading Methanogenic communities	Marine Plankton communities from deep Mediterranean sea
Biogas-producing microbial community from Bioreactor	Marine planktonic communities from Hawaii Ocean Times Series Station (HOT/ALOHA)
Coral Porites astreoides microbiome	Marine RNA viral communities from coastal samples
Coral Porites compressa microbiome	Marine viral communities
Coral reefs microbiome from the Northern Line Islands	Marine viral communities from Chesapeake Bay
Daphnia pulex microbiome	Marine viral communities from deep-sea hydrothermal vents
Elephant Grass decomposer microbial community	Marine viral communities from ocean environments
Endophytic crenarchaeal communities of Plant roots	Marine viroplankton communities from surface and estuarine waters
Endophytic Frankia community of Plant roots	Mosquitos virome
Endophytic microbial communities of pepper bark tree	Olavius algarvensis (gutless worm) microbiome from Mediterranean sea
Endophytic microbiome from Rice	Oral microbiome of Human
Endophytic microbiomes from Poplar	Oral TM7 microbial communities of Human
Extreme microbial communities from Yellowstone National Park	Poplar biomass decaying microbial community
Fossil microbial community from Whale Fall	Sand microbial communities from Pacific Beach
Fossil microbiome from Mammothus primigenius	Sediment microbial communities from DOE Hanford Site
Fossil microbiome from Mammothus primigenius	Simulated microbial communities from 113 complete and draft JGI genomes
Fossil microbiome from Neanderthal	Soil Fungal communities from boreal forest
Fossil microbiome from Neanderthal	Soil microbial communities associated with greenhouse gases
Fossil microbiome from Pleistocene Cave Bears	Soil microbial communities from Alaska
Fungal communities from the Arctic	Soil microbial communities from contaminated sediments
Gut microbiome of Anoplophora glabripennis	Soil microbial communities from four geographic locations
Gut microbiome of Human	Soil microbial communities from Minnesota Farm
Gut microbiome of Human	Soil microbial communities from Wetland
Gut microbiome of Human	Soil microbial community from Tallgrass Prairie
Gut microbiome of Human	Soil PCBs-degrading microbial communities
Gut microbiome of Human	Sponge microbiome from Conch Reef, Key Largo Florida
Gut microbiome of Human	Stromatolites microbial communities
Gut microbiome of Macropus eugenii	Sulfur spring microbial communities from Zedletone Oklahoma
Gut microbiome of Mouse	Thermal and Saline microbial communities from China and UK
Gut microbiome of termites	Wastewater anaerobic ammonium oxidation (anammox) community from Bioreactor
Gut microbiome of Tipula abdominalis	Wastewater EBPR microbial communities from Bioreactor
Gut virome of Human	Wastewater Terephthalate-degrading communities from Bioreactor
Gut virome of Human	Wastewater treatment plant Plasnid pool
Lung microbiome of Human	Wastewater treatment type I Accumilbacter community from Bioreactor
Lung virome of Human	

gène rapporteur *gfp* (Green Fluorescent Protein), préalablement cloné à la suite de l'opéron, permet la mise en évidence des clones possédant un opéron actif (Uchiyama *et al.*, 2005). Ainsi, 62 clones régulés par la présence de benzoate et de naphthalène ont été identifiés au travers d'une banque de 152 000 clones. Il faut toutefois noter que cette approche peut être limitée par l'usage de bactéries hôtes pouvant ne pas posséder le matériel génétique nécessaire à transcription, à la traduction et à la maturation de la protéine d'intérêt (Handelsman, 2004).

A l'instar de la métagénomique, des études visant à identifier l'ensemble des activités exprimées *in-situ* par les microorganismes voient le jour. Cette approche originale, nommée métatranscriptomique, est actuellement encore peu utilisée. Son utilisation repose sur l'extraction des ARN totaux et sur le clonage des produits reverse-transcrits. Ainsi, Bailly *et al.* (2007) ont réalisé des études préliminaires sur le métatranscriptome des eucaryotes du sol qui ont permis l'identification de gènes impliqués dans les différentes grandes fonctions cellulaires dont la mobilisation des éléments nutritifs du sol.

2.2. Le séquençage ultra-haut débit.

La technique de séquençage, utilisée dans la plupart des techniques précédemment citées, est dérivée de la technique de Sanger (Sanger *et al.*, 1977). L'utilisation extensive de cette méthode, basée sur la migration électrophorétique de fragments désoxyribonucléotides triphosphates, est limitée par son coût élevé et son débit relativement faible. Actuellement, le séquençage d'un méga paire de bases, par une telle technique, est estimé à environ 6 000 € (Goldberg *et al.*, 2006). Le besoin de techniques plus efficaces et à moindre coût a entraîné le développement de nouvelles technologies de séquençage. Notons particulièrement la technique développée par Shendure *et al.* (2005) où les matrices ADN, attachées à des billes d'un micromètre, sont séquencées grâce à des réactions de polymérisation en chaîne en émulsion. D'autres techniques de séquençage permettent d'éviter les étapes de clonage mais également l'accès à des séquences qui ne peuvent être propagées dans un système bactérien. C'est le cas entre autres du séquençage sur biopuce et du séquençage au travers de nanopores.

Une des approches les plus prometteuses est la technique de pyroséquençage (Ronaghi *et al.*, 1998) selon le système 454 (454 Life Science, Branford, CT). Les échantillons d'ADN sont fixés sur des billes et placés dans des micropuits où s'effectue le pyroséquençage. Contrairement au séquençage classique, les nucléotides ne sont pas ajoutés tous ensemble mais les uns après les autres (figure 16). Le pyroséquençage permet ainsi d'obtenir de 200000

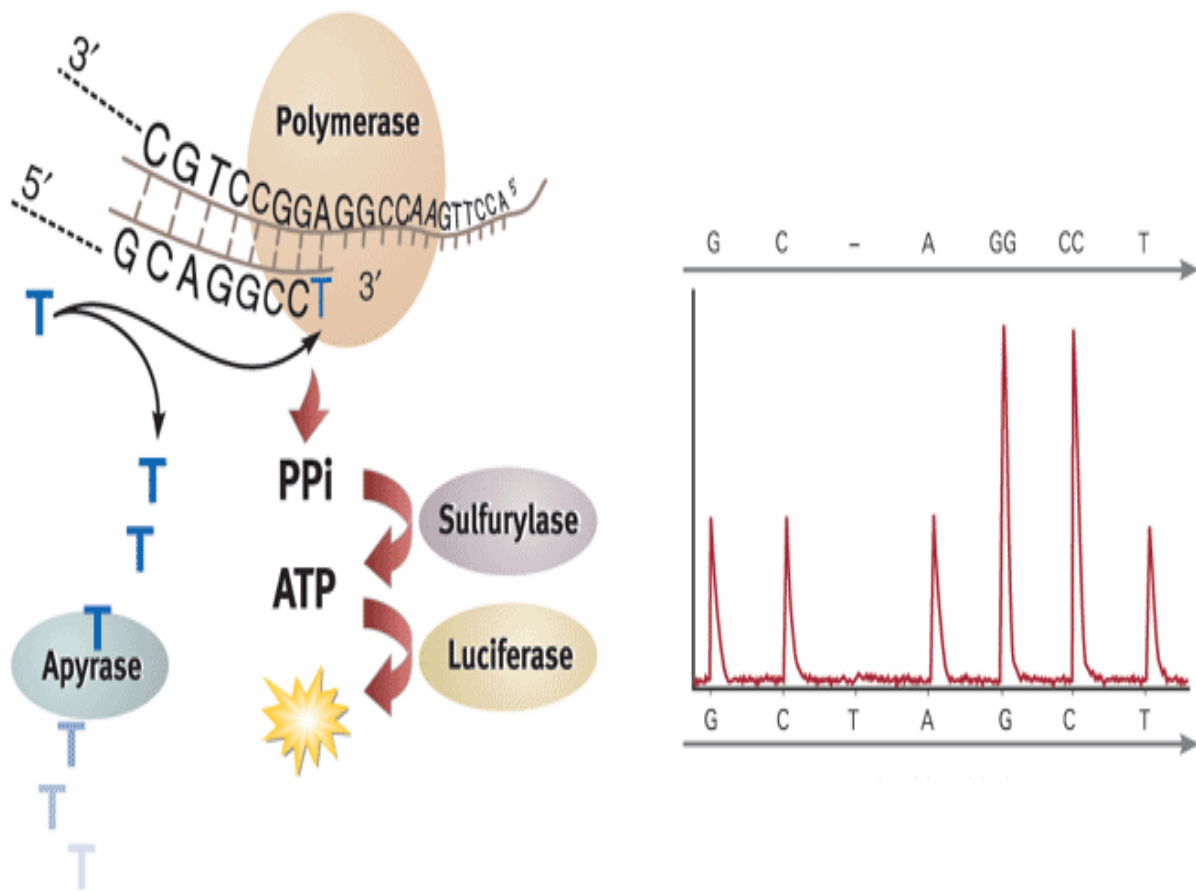


Figure 16 : Principe du pyroséquençage. Contrairement au séquençage classique, les nucléotides ne sont pas ajoutés tous ensemble mais les uns après les autres. Si le nucléotide ajouté est complémentaire de la matrice, l'ADN polymérase l'incorporera en libérant un pyrophosphate (PPi). Ce dernier va être transformé en ATP grâce à une ATP sulfurylase. L'ATP ainsi formé sera alors utilisée par une luciférase pour émettre un signal lumineux. Ce signal va être capté par le séquenceur et traduit en un pic qui sera plus ou moins important en fonction du nombre de nucléotides incorporés. Les nucléotides qui n'ont pas été incorporés par la polymérase sont dégradés par une apyrase avant qu'un nouveau nucléotide ne soit ajouté (adapté de England et Petterson, 2005).

à 400 000 pyrogrammes de 250 bases lors d'un seul run. Il faut toutefois noter que le nombre important de séquences obtenues par ce type de séquençage et surtout leurs tailles relativement réduites peuvent engendrer des difficultés d'assemblage. Même si le séquençage de 100 Megabases (Mb) par cette méthode s'élève à 15 000 €, ce coût reste toutefois inférieur à celui qu'aurait entraîné l'utilisation de la méthode de Sanger. Ce système principalement utilisé pour le séquençage de génome, a été testé sur des environnements complexes. Grâce à cette technique, Sogin *et al.* (2006) ont exploré la biodiversité d'écosystèmes marins profonds au travers de 118 000 régions hypervariables V6 de l'ADNr 16S. Ils ont ainsi pu mettre en évidence leur incroyable complexité, par rapport à d'autres environnements microbiens. Cette diversité serait attribuable à la présence de milliers de groupes bactériens peu abondants (biosphère rare) relativement anciens.

CONCLUSIONS

Actuellement, la majorité des espèces microbiennes ne sont connues qu'au travers de leurs biomarqueurs, essentiellement les gènes codants pour les ARNr 16S. La découverte de leurs fonctions biologiques permettrait d'élargir cette vision limitée. Cependant, il s'agit d'un véritable challenge car la majorité des espèces bactériennes n'est toujours pas cultivées. L'ensemble des outils moléculaires mis aujourd'hui à disposition des microbiologistes permet d'étudier de façon approfondie les communautés microbiennes, tant au niveau de leur structure que de leur fonction. Au cours de ces dernières décennies, de nouveaux microorganismes et de nouvelles voies métaboliques ont ainsi pu être découverts, permettant d'améliorer notre compréhension du monde microbien. Cependant, du fait de l'importante diversité des communautés microbiennes au niveau des environnements complexes, l'utilisation d'outils à hauts débits devient indispensable afin d'avoir une vue globale de tous les acteurs de ces écosystèmes. Un outil moléculaire semble particulièrement bien adapté à l'identification et au suivi simultané de nombreux groupes microbiens : les biopuces ADN.

Chapitre III

Les biopuces ADN : outils à haut-débit.

INTRODUCTION

Apparues au milieu des années 90, après le séquençage à grande échelle, les biopuces ADN ont été initialement mises au point pour l'étude globale de l'expression des gènes afin de découvrir leurs fonctions biologiques et d'étudier leurs réseaux de régulation (Skena *et al.*, 1995). Ces outils, dérivés des techniques de Southern blot (Southern, 1975) et de dot blot (Kafatos *et al.*, 1979), permettent d'étudier simultanément des milliers de gènes, assurant ainsi une meilleure compréhension des processus cellulaires complexes (Lander, 1999).

Utilisés en microbiologie, ces outils à haut débit ont permis d'apporter de nombreuses réponses aux questions posées sur les mécanismes de développement et d'adaptation des microorganismes à leur environnement. En effet, les biopuces ADN ont permis d'identifier, entre autres, de nouveaux gènes, des organisations originales de génomes, des réponses à des stress environnementaux ainsi que de nombreuses régulations transcriptionnelles (Dharmadi et Gonzalez, 2004).

Plus récemment, les biopuces ont été utilisées pour appréhender la diversité procaryotique et son fonctionnement au sein d'environnements complexes tels les systèmes aquatiques ou encore terrestres (DeSantis *et al.*, 2007 ; Moisander *et al.*, 2007 ; Peplies *et al.*, 2006 ; Sanguin *et al.*, 2006). Contrairement aux cultures pures, les écosystèmes naturels présentent une formidable diversité en microorganismes, augmentant ainsi les difficultés d'étude. Du fait de cette complexité, les biopuces sont des outils à haut débit de choix pour étudier cette diversité microbienne et son évolution spatio-temporelle.

1. LA TECHNOLOGIE BIOPUCE ADN.

La technique biopuce ADN est basée sur une hybridation dite inverse car, comparée aux techniques de Southern et de dot blot, ce sont les sondes et non les cibles qui sont

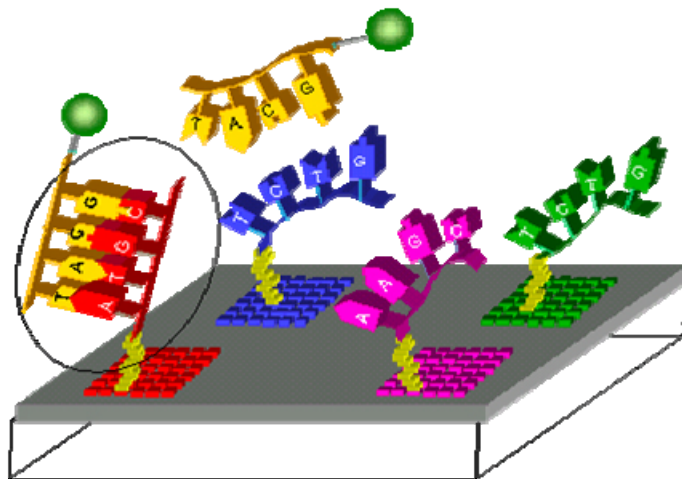


Figure 17 : Principe des biopuces ADN. Les cibles marquées par fluorescence s'hybrident spécifiquement avec les sondes complémentaires fixées sur un support solide tel que des lames de verre fonctionnalisées (laboratoire Ifos UMR CNRS 5621).

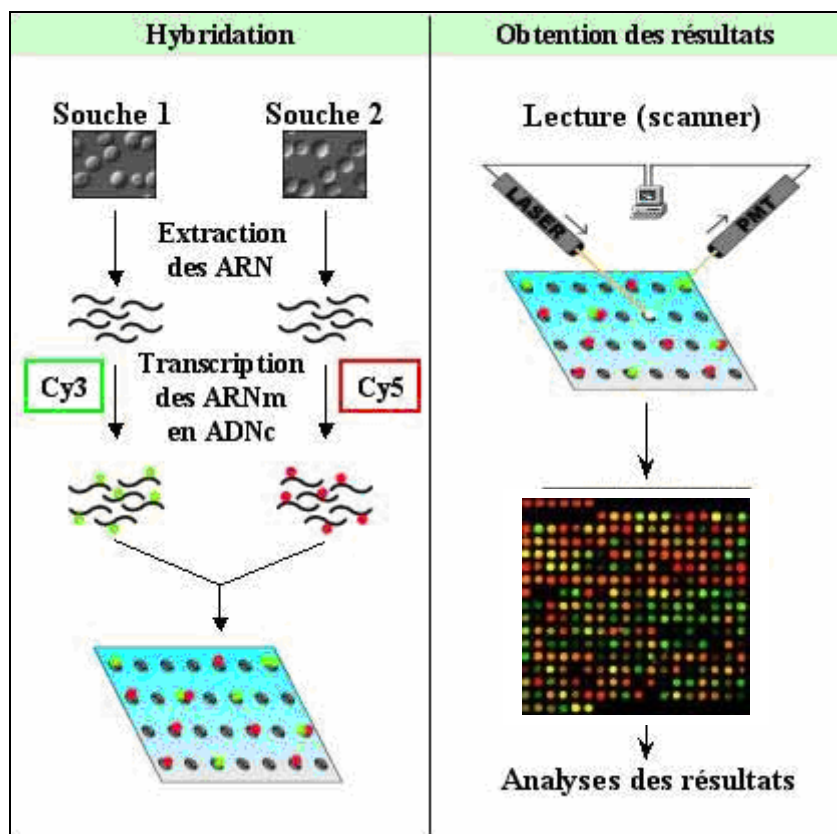


Figure 18 : Schéma d'une analyse transcriptomique par biopuce à ADN. Les ARN totaux sont extraits de deux souches bactériennes différentes. Les fractions ARNm sont rétrotranscrites puis marquées avec des fluorochromes (Cy3 et Cy5). Les cibles ainsi générées sont déposées sur la biopuce et s'apparient avec les sondes présentant des homologies de séquences. Ces hybridations sont mesurées à l'aide d'un scanner à fluorescence et les images générées sont analysées.

immobilisées sur un support solide. Des centaines voire même des dizaines de milliers de sondes différentes peuvent être fixées sur une seule biopuce, augmentant ainsi la détection simultanée d'une grande variété de gènes ou de microorganismes. Durant l'étape d'hybridation, jusqu'à deux populations de cibles, marquées différemment par des fluorochromes (généralement Cy3 et Cy5), peuvent être criblées par les sondes présentant une complémentarité de séquence. Les duplex ainsi formés sont résistants aux étapes de lavages nécessaires à l'élimination des cibles non ou mal appariées (figure 17). Après l'hybridation des sondes portées par la biopuce avec les cibles marquées, la fluorescence des duplex est détectée à l'aide d'un scanner (Ehrenreich, 2006).

1.1. Les sondes.

Initialement, les sondes désoxyribonuléotidiques étaient fixées sur des membranes de nylon poreuses (macroarray) et les cibles étaient marquées par radioactivité. Ce support, limité en terme de densité de sondes, a vite été abandonné au profit de lames de verre préactivées (microarray). Les sondes radioactives, quant à elles, ont été remplacées par des sondes fluorescentes en raison de leur facilité de manipulation. Ce format apporte ainsi de nombreux avantages tels la possibilité de fixation covalente des sondes sur le verre, la diminution des phénomènes de bruit de fond et la possibilité d'hybrider simultanément deux populations de cibles (Dharmadi et Gonzalez, 2004) (figure 18). Les sondes fixées sur ces lames de verre peuvent être des oligonucléotides ou des ADN doubles brins dénaturés (ADN complémentaires (ADNc), produits PCR ou ADN génomiques).

Les avantages des sondes ADN sont multiples. Ces biopuces peuvent être générées sans connaître à *priori* la séquence du gène ciblé (Schena *et al.*, 1995). De plus, elles possèdent une grande sensibilité car la taille des sondes (en moyenne 500 pb) augmente l'énergie de liaison du duplex sondes/cibles (Chou *et al.*, 2004). Cependant, cette grande taille représente aussi un inconvénient car elle augmente la possibilité de générer des hybridations croisées, c'est-à-dire de reconnaître des séquences non ciblées. En outre, il faut prendre en considération les efforts importants nécessaires afin de créer une biopuce avec ce type de sondes (coût de production important pour l'obtention des sondes et pour leur vérification) (Dharmadi et Gonzalez, 2004 ; Gibson, 2002).

Une alternative aux sondes ADN double brins (ADNdb) est l'utilisation de sondes oligonucléotidiques (~20- à 50-mers) qui sont faciles à synthétiser (Kreil *et al.*, 2006). Du fait de leur taille, ces sondes sont plus spécifiques que les sondes ADN mais peuvent présenter

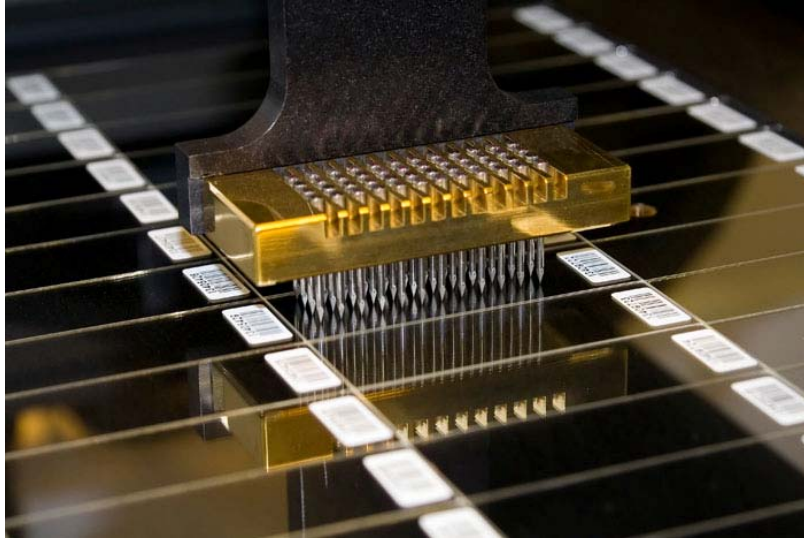


Figure 19 : Fixation des sondes sur les biopuces ADN par un robot. Les aiguilles du robot sont plongées dans la solution de sondes et ces dernières sont déposées par contact sur la surface de la biopuce.

des défauts de sensibilité. Contrairement aux sondes ADNdb, il est nécessaire de connaître la séquence du gène ciblé pour pouvoir créer une sonde adaptée. Il est aussi nécessaire de vérifier si celles-ci n'entraînent pas de croisements aspécifiques avec des séquences non ciblées. Ainsi, une analyse bioinformatique est nécessaire afin de choisir les sondes les plus efficaces (Dharmadi et Gonzalez, 2004). Il faut noter qu'à l'heure actuelle, les sondes oligonucléotidiques sont le plus souvent utilisées (Gibson, 2002).

1.2. Fixation des sondes.

En fonction de la nature des sondes, les biopuces peuvent être construites de deux façons. La première consiste en une synthèse indépendante des sondes suivie de leur fixation sur le support solide à l'aide d'un robot spotter. Dans ce cas, les aiguilles sont plongées dans la solution de sondes et ces dernières sont déposées par contact sur la surface de la biopuce (figure 19). Cette méthode peut être utilisée pour tout type de sondes mais, en fonction des sondes employées, les liaisons avec la surface ne se feront pas de la même façon (Stekel, 2003). Les liaisons peuvent être soit covalentes, impliquant des groupes fonctionnels carboxyliques, phosphate, aldéhyde ou amino, ou non covalentes, impliquant des liaisons électrostatiques. Pour les sondes oligonucléotidiques, de nombreuses chimies de surface des lames sont disponibles mais les surfaces les plus souvent utilisées sont recouvertes de groupements aminosilane (interaction électrostatique), époxy ou aldéhyde (fixations covalentes des sondes). Afin de réaliser la liaison covalente entre la lame et la sonde pré-synthétisée, cette dernière possède souvent une modification sur l'une de ses extrémités. Le plus souvent il s'agit d'un groupement amine situé en 5' qui réagit avec les groupements époxy ou aldéhyde. Parfois et cela afin d'améliorer la sensibilité, des espaceurs (par exemple plusieurs dizaines de nucléotides triphosphate identiques) sont utilisés entre la lame et la sonde (Guo *et al.*, 1994). En effet, l'éloignement des sondes de la surface de la biopuce semble limiter les contraintes spatiales lors de la reconnaissance sonde/cible.

La seconde façon de construire une biopuce consiste à réaliser directement la synthèse des sondes sur la lame. Cette synthèse *in situ*, uniquement utilisable pour les sondes oligonucléotidiques, peut être réalisée par plusieurs méthodes. Dans le premier cas, les nucléotides sont projetés contre la lame selon une technologie analogue à celle d'une imprimante jet d'encre (technologie Agilent). Cette technologie, bien que très flexible, ne permettrait pas la production de lames strictement identiques (Stekel, 2003). La seconde

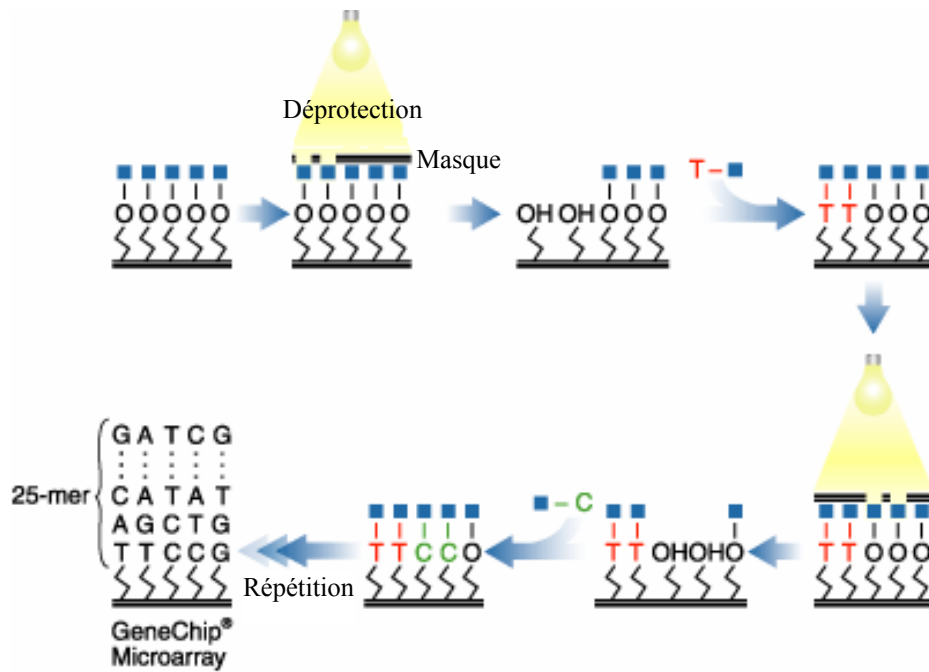


Figure 20 : Synthèse *in situ* des sondes par la technique de la photolithographie. Les biopuces sont recouvertes de composés chimiques sensibles à la lumière. Des masques photolithographiques sont utilisés pour bloquer ou laisser passer la lumière au niveau des zones à activer. La biopuce est ensuite recouverte de solutions contenant les nucléotides à ajouter (adénine, thymine, cytosine ou guanine). Ceux-ci se fixent seulement au niveau des zones qui ont été déprotégées par la lumière et, comme elles portent des groupements de protection sensibles à la lumière, le cycle peut être répété jusqu'à ce que l'oligonucléotide soit complètement synthétisé.

méthode de synthèse *in situ*, utilisée par le système Affymetrix GeneChip®, permet d'obtenir des biopuces oligonucléotidiques très denses contenant plusieurs centaines de milliers de sondes (DeSantis *et al.*, 2007) (figure 20). La méthode employée est la photolithographie. Des lames de quartz, recouvertes de composés chimiques sensibles à la lumière, sont utilisées comme support solide pour la synthèse *in situ* des sondes. Des masques photolithographiques sont utilisés pour bloquer ou laisser passer la lumière au niveau des zones à activer. La biopuce est ensuite recouverte de solutions contenant les nucléotides à ajouter. Ceux-ci se fixent seulement au niveau des zones qui ont été déprotégées par la lumière et, comme ils portent des groupements de protection sensibles à la lumière, le cycle peut être répété jusqu'à ce que l'oligonucléotide soit complètement synthétisé (généralement jusqu'à 25-mers). Chaque sonde Affymetrix est couplée à un contrôle négatif (même séquence mais avec un mésappariement central) et plusieurs couples de sondes sont généralement utilisés pour un même gène ciblé (Ehrenreich, 2006 ; Stekel, 2003).

1.3. Préparation des cibles.

La préparation des cibles consiste en un marquage permettant la reconnaissance par le scanner des duplex sondes/cibles formés lors de l'hybridation. Plusieurs fluorochromes peuvent être choisis mais les cyanines (Cy3 et Cy5) sont les plus souvent utilisées car, d'une part, elles présentent des spectres d'émission de fluorescence bien distincts (respectivement 580-645 nm et 670-735 nm) et une fluorescence importante et, d'autre part, elles entraînent peu de bruit de fond (Ehrenreich, 2006).

L'incorporation des fluorochromes peut s'effectuer de deux façons : par marquage direct ou indirect (figure 21). Le marquage direct consiste en une incorporation, au niveau de la cible, de nucléotides portant les fluorochromes (par exemple Cy5-dUTP ou Cy3-dCTP). Ce marquage est facile à réaliser car il ne nécessite qu'une étape. Cependant, il présente une incorporation différentielle selon la cyanine utilisée. En effet, la cyanine 5 présente un encombrement stérique plus important et, de ce fait, est moins bien incorporée que la cyanine 3 lors de la synthèse de la sonde. Afin d'éviter ce biais, le marquage peut être effectué indirectement. Cette méthode consiste en une incorporation préliminaire, au niveau des cibles, de nucléotides modifiés au niveau desquels seront greffés, dans un deuxième temps, les fluorochromes. Le système le plus souvent utilisé est la modification des nucléotides par un groupement amino-allyl (aa-dUTP ou aa-dCTP) combiné à des fluorochromes portant un

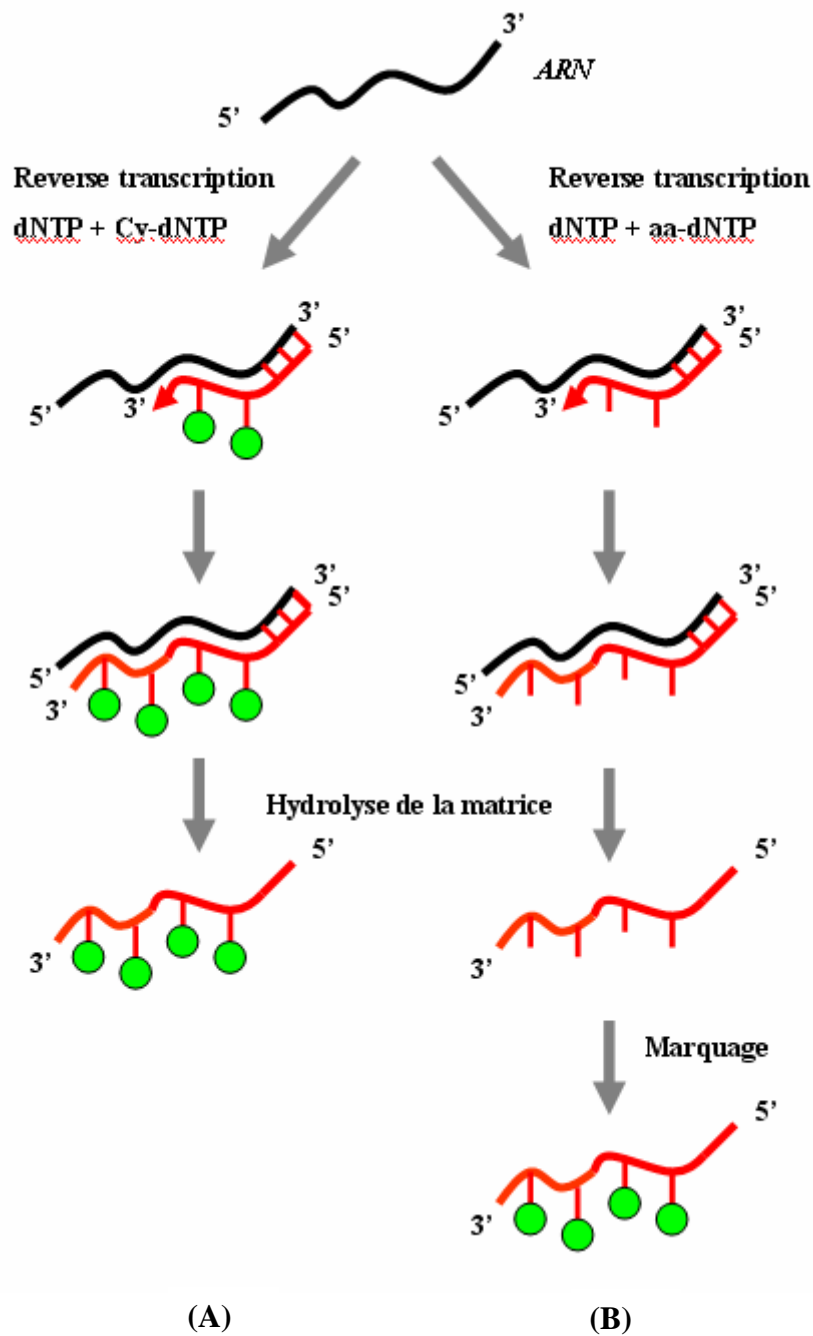


Figure 21: Principe du marquage des cibles pour l'hybridation des biopuces ADN. L'incorporation des fluorochromes peut s'effectuer de deux façons : par marquage direct avec incorporation d'un nucléotide fluorescent (A) ou indirect avec incorporation d'une base modifiée qui servira par la suite au couplage du fluorochrome (B).

groupement N-hydroxysuccinylimidyl ester (NHS ester). Enfin, le marquage des cibles peut aussi être réalisé directement sur les molécules d'ADN ou d'ARN grâce au système Universal Linkage System (ULS) de Kreatech. Cette technique implique une liaison entre la guanine des ADN/ARN cibles et un atome de platine contenu dans un haptène marqué par un fluorophore (Van Gijlswijk *et al.*, 2001).

1.4. L'hybridation.

L'hybridation est l'étape qui permet la formation des duplex cibles/sondes. Elle est basée sur le postulat que deux fragments d'ADN, possédant des séquences nucléiques complémentaires, s'apparient en formant des liaisons de faible énergie : les liaisons hydrogènes (Southern, 1975). En fonction de la stringence du milieu (température, concentration en ions sodium), cet appariement peut se rompre plus ou moins facilement. Ainsi, une température d'hybridation élevée et des concentrations faibles en Na^+ vont permettre des hybridations plus spécifiques car les duplex sondes/cibles possédant des mésappariements seront moins stables.

Les cibles marquées sont mélangées à un tampon d'hybridation contenant le plus souvent du Saline Sodium Citrate (SSC), une molécule masquant les sites de fixation aspécifique des cibles sur la biopuce et permettant ainsi de limiter le bruit de fond (par exemple, l'ADN de sperme de saumon) et des molécules permettant une meilleure reconnaissance des cibles à des températures plus basses (comme le formamide). Ce mélange est disposé sur la biopuce, au niveau de la zone spottée, puis est recouvert d'une lamelle afin d'éviter l'évaporation. Ce montage est placé pendant plusieurs heures dans une chambre d'hybridation où la température, calculée en fonction de la composition en bases des sondes, est maintenue constante. Une fois hybridée, la biopuce est lavée soit avec des solutions stringentes (généralement une solution faiblement concentrée en SSC contenant du SDS), soit à haute température afin d'éliminer les cibles mal ou non appariées (Ehrenreich, 2006). Ces lavages peuvent être réalisés à température constante (lavage isothermique) ou croissante (dissociation thermique non équilibrée). Notons toutefois que des analyses ont révélé que environ 20 % des sondes, et cela, indépendamment des caractéristiques spécifiques de ces séquences, ne sont pas éliminées correctement par cette dernière approche (Pozhitkov *et al.*, 2007).

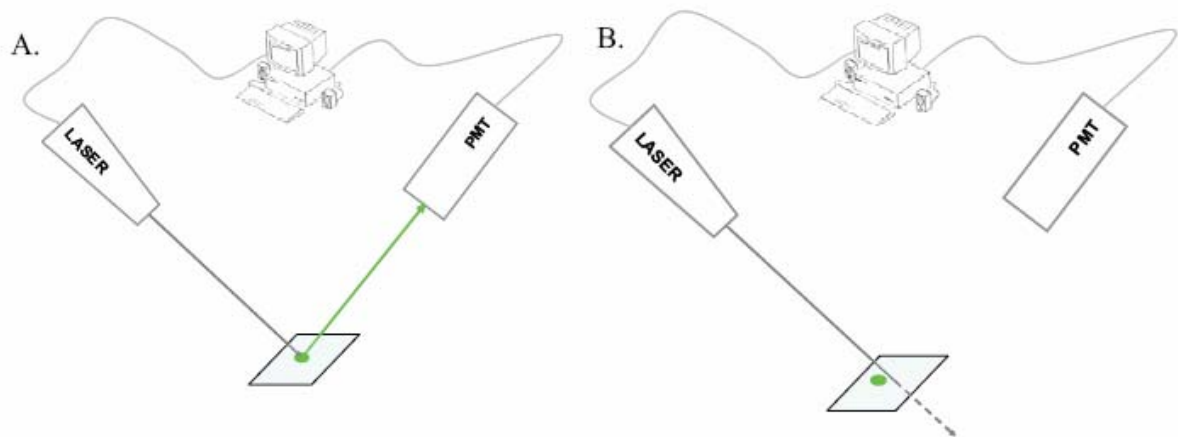


Figure 22 : Principe de la lecture des biopuces par un scanner à fluorescence. Le laser excite les fluorochromes présents au niveau des duplex sondes/cibles. Une fluorescence est émise et est collectée au niveau du photomultiplicateur (PMT) (A). Aucune émission n'est détectée lorsqu'il n'y a pas de duplex (B).

L'utilisation du système lame/lamelle ne permet pas une diffusion optimale des cibles sur toute la surface spottée ce qui peut entraîner des efficacités d'hybridations hétérogènes d'un spot à l'autre. Récemment, des avancées technologiques ont été réalisées afin d'améliorer cette étape d'hybridation en permettant une meilleure répartition des cibles. Ainsi, Yuen *et al.* (2003) ont amélioré de 2 à 5 fois l'efficacité de l'hybridation en utilisant un système de barreaux aimantés permettant de créer un chaos fluide à l'intérieur de la chambre d'hybridation. D'autres systèmes, basés sur le même principe, ont été développés comme, par exemple, un micromélangeur à advection chaotique ou des systèmes de rotation de la chambre durant l'hybridation (McQuain *et al.*, 2004).

1.5. Acquisition et analyse des images de biopuces.

Une fois hybridées, les biopuces vont être analysées à l'aide d'un scanner qui va collecter les niveaux de fluorescence émis par chaque zone de la biopuce. Les signaux fluorescents, témoignant de la reconnaissance de la cible par la sonde, sont obtenus après excitation laser des fluorochromes des cibles (figure 22). La lumière ainsi émise est collectée au niveau d'un photomultiplicateur (PMT) où elle est transformée en signaux électriques qui permettront de recréer une image en fausses couleurs (Stekel, 2003). Les données brutes de l'image sont stockées en format 16 bits. Cela signifie que les intensités de fluorescence peuvent s'étaler de 1 à 65 535 unités de fluorescence (FU). Etant donné que les biopuces montrent généralement un bruit de fond de fluorescence égal à 100 FU et que la saturation du signal apparaît vers 50000 FU, on peut considérer que cet outil offre une large dynamique de signal (Ehrenreich, 2006).

La première étape du traitement des images de biopuce est la reconnaissance des zones (spots) où les sondes ont été fixées. Cette étape se fait à l'aide d'une grille, souvent rectangulaire, que l'on positionne sur l'image de la biopuce. Les coordonnées de cette grille sont généralement générées automatiquement lors de la fabrication des biopuces (localisation du dépôt ou de la synthèse des sondes). Ensuite, l'étape de segmentation permet de tracer le contour des spots en définissant quels pixels entrent dans leur composition. Ceci est généralement réalisé par les méthodes des cercles fixes ou variables, des histogrammes ou des formes adaptatives (Stekel, 2003). Pour chacun des répliqués, la fluorescence de tous les pixels des spots est ensuite mesurée puis la fluorescence moyenne du bruit de fond en est extraite. Une normalisation à l'aide de spots contrôles, est ensuite réalisée afin de pallier d'éventuels

biais de fluorescence sur une même lame ou entre plusieurs lames (Dharmadi et Gonzalez, 2004).

2. LES BIOPUCES FONCTIONNELLES ET GENOMIQUES.

Il existe de nombreux types de biopuces utilisés en écologie microbienne. En 2006, Gentry *et al.* ont tenté de les classer en quatre formats selon les gènes ciblés : les biopuces ciblant les gènes fonctionnels (FGA : Functional Gene Arrays), les biopuces génomiques (CGA : Community Genome Arrays), les biopuces génétiques ciblant les cadres de lectures ouverts de génome complets (WGA : Whole Genome Arrays) et les biopuces phylogénétiques oligonucléotidiques (POA : Phylogenetic Oligonucleotide Arrays) ciblant les ARNr 16S. Ce dernier type de biopuces fera l'objet d'un paragraphe spécial dans ce chapitre (cf. chapitre III ; § 3).

2.1. Les biopuces ciblant les gènes fonctionnels.

Les biopuces ciblant les gènes fonctionnels ou « Functional Gene Arrays » (FGA) sont surtout utilisées pour l'étude des gènes impliqués dans les grands cycles biogéochimiques et dans les procédés de bioremédiation (Wu *et al.*, 2001 ; Bodrossy *et al.*, 2003 ; Rhee *et al.*, 2004). Les sondes utilisées peuvent être soit des produits PCR (~200 à 1 000 nucléotides), soit des oligonucléotides (~15 à 70 mers) (Sessitsch *et al.*, 2006). Cependant, ce sont ces derniers qui sont généralement favorisés car les produits PCR, du fait de leur taille, peuvent engendrer des croisements aspécifiques. Les sondes oligonucléotidiques longues (~40- à 70-mers) sont préférées aux sondes courtes car, d'une part, elles sont plus sensibles et, d'autre part, leur spécificité, moins importante que celle des oligonucléotides courts, est compensée par le fait que les gènes fonctionnels présentent suffisamment de variabilité. Les oligonucléotides de 50-mers permettent de différencier les cibles possédant en général moins de 88 % d'homologies avec leurs sondes si on se place dans des conditions d'hybridation assez stringentes (T= 50° C ; %Formamide= 50%) (Rhee *et al.*, 2004). Même si leur sensibilité est inférieure à celle des produits PCR, elles permettent tout de même de détecter leurs cibles dans 5 à 10 ng d'ADN génomique seul ou dans 50 à 100 ng d'ADN génomique en mélange (Rhee *et al.*, 2004). Wu *et al.* (2001) ont pu détecter, grâce à une FGA oligonucléotidique, la présence des gènes nitrate réductase (*nirS*) dans 1 ng d'ADN génomique pur et dans 25 ng d'ADN génomique en

mélange. Ces seuils de détection correspondraient à la mise évidence de cibles ADN provenant seulement de 10 cellules ou de 5 % des communautés bactériennes totales (Gentry *et al.*, 2006). De plus, Wu *et al.* (2001) ont noté que pour des valeurs de 1 à 100 ng d'ADN génomique (pure ou en mélange), il existait une relation linéaire entre l'intensité du signal et la concentration de la cible ce qui signifie qu'une analyse semi quantitative serait possible dans cette gamme.

Jusqu'à très récemment, les biopuces ciblant les gènes fonctionnels ne présentaient pas une forte densité car les gènes ciblés doivent répondre à des critères très précis. En effet, ils doivent : (1) coder pour une enzyme clé dans la voie métabolique d'intérêt, (2) présenter des régions suffisamment conservées pour permettre leur détection mais aussi des régions présentant suffisamment de variabilité pour permettre la sélection de sondes spécifiques de microorganismes, et (3) être représentés par un certain nombre de séquences dans les bases de données afin de disposer de suffisamment d'informations pour la détermination des sondes. C'est pourquoi, avant l'apparition de la GeoChip ciblant plus de 10 000 gènes impliqués dans les cycles du carbone, de l'azote, du soufre, du phosphore mais aussi dans la dégradation de contaminants organiques et dans la réduction et la résistance aux métaux (Yergeau *et al.*, 2007), la biopuce la plus dense publiée ne contenait que 1 662 sondes (Rhee *et al.*, 2004).

La majorité des études utilisant les FGA sont réalisées avec des cibles ADN permettant seulement de déterminer si les gènes des organismes de l'environnement étudié sont présents. Ainsi, Stralis-Pavese *et al.* (2004) ont étudié la structure des communautés méthanotrophes de lysimètres différemment végétalisés. Ils ont pu démontrer, grâce à une FGA contenant 68 sondes oligonucléotidiques ciblant les gènes particulate methane monooxygenase (*pmoA*), que la structure des communautés méthanotrophes était liée au type de végétation. Cependant, pour obtenir des informations sur les expressions des gènes d'intérêt, il est nécessaire d'utiliser des cibles ARN messagers (ARNm). Grâce à l'utilisation de telles molécules, Rhee *et al.* (2004) ont montré la présence et l'expression de gènes impliqués dans la dégradation du naphtalène. Il faut cependant noter que la plupart des études, utilisant les ARNm, sont réalisées sur des cultures pures, sur des systèmes simples ou sur des enrichissements en raison de la difficulté à extraire ce genre de molécules à partir d'environnements complexes (Gentry *et al.*, 2006).

2.2. Les biopuces ciblant les génomes.

La technologie des biopuces CGA dérive de la technique Reverse Sample Genome Probing (RSGP) qui est basée sur la fixation de génomes bactériens sur une membrane nylon ou nitrocellulose et sur l'utilisation de cibles environnementales radiomarquées. Dans le cas des biopuces CGA, les génomes de bactéries cultivables sont fixés sur une lame de verre (un génome par spot) et les cibles sont marquées par fluorescence (Sessitsch *et al.*, 2006).

Wu *et al.* (2004) ont ainsi développé une CGA ciblant 67 génomes bactériens. L'ADN génomique de ces microorganismes, appartenant aux α -, β - et γ -*Proteobacteria* et à des bactéries Gram positif, a été isolé et utilisé pour la fabrication de cette biopuce. Deux niveaux de différenciation ont pu être mis en évidence grâce à un ajustement des conditions d'hybridation. Une température d'hybridation à 55° C et un pourcentage de formamide à 50% a permis une différenciation au niveau de l'espèce alors qu'une augmentation de la stringence des conditions d'hybridation (T= 75° C ; %Formamide= 65 %) a permis une différenciation au niveau de la souche. La limite de détection de l'ADN génomique pur a été estimée à 0,2 ng. De plus, en conditions stringentes, il semble exister une relation linéaire entre l'intensité du signal et la concentration des cibles ADN (issues de cultures pures ou en mélanges). Ces résultats suggèrent que les biopuces CGA peuvent éventuellement devenir des outils quantitatifs pour détecter des microorganismes dans des échantillons environnementaux. Le caractère discriminant des biopuces CGA a été confirmé par Bae *et al.* (2005) grâce à l'utilisation de sondes issues de fragments génomiques obtenus par hybridation soustractive. La biopuce CGA ainsi construite, ciblant le genre *Salmonella*, a permis une discrimination au niveau de la souche.

Bien que l'utilisation de ce type de puce puisse potentiellement discriminer les communautés microbiennes, leur utilisation reste encore délicate car on ne peut exclure l'apparition de croisements aspécifiques, surtout si l'étude porte sur des environnements complexes. De plus, l'utilisation de ces biopuces est restreinte à l'étude des microorganismes cultivés du fait de la nécessité d'extraire leur génome pour la génération des sondes. Une alternative pourrait être l'utilisation de sondes correspondant aux grands fragments de génomes obtenus par des approches métagénomiques (Handelsman, 2004).

2.3. Les biopuces génétiques.

Du fait des transferts horizontaux de gènes entre les souches bactériennes, les microorganismes montrant une parenté phylogénétique (ADNr 16S proches) peuvent cependant présenter d'importantes différences (gènes accessoires et polymorphisme génétique). La comparaison de leurs génomes pourrait permettre de comprendre la base génétique de nombreuses différences phénotypiques mais, aussi, d'étudier les mécanismes à la base de ces évolutions. Cependant, seules quelques espèces bactériennes ont plusieurs de leurs représentants séquencés (par exemple *Escherichia coli*, *Campylobacter jejuni*, *Helicobacter pylori* ou encore *Staphylococcus aureus*). En effet, le séquençage systématique des génomes de toutes les espèces et souches ayant des ADNr 16S proches reste relativement long et coûteux. Le développement des techniques de séquençage à ultra haut débit, notamment le pyroséquençage, facilitera ces études dans un proche avenir. Actuellement, des biopuces ciblant les ORFs ont toutefois été développées pour évaluer les différences génétiques d'espèces et de souches. En effet, les séquences génétiques qui codent pour des protéines sont largement représentatives du génome des organismes (Dorell *et al.*, 2005).

Les sondes des biopuces WGA peuvent être générées par amplification PCR de chaque gène du génome et peuvent présenter une densité allant jusqu'à 50 000 produits PCR sur une lame de verre. Ainsi, une comparaison de 34 souches virulentes d'*Helicobacter pylori*, isolées de cas cliniques, a été réalisée à l'aide d'une biopuce WGA construite par amplification PCR des 1 636 ORFs de deux souches références (*Helicobacter pylori* 26 695 et J99). Les résultats montrent que le « core-set » (ensemble des gènes présents dans toutes les souches) d'*Helicobacter pylori* représenterait 66,7 % du génome. Les autres gènes « accessoires » semblent être spécifiques de l'une ou de l'autre des deux souches. (Han *et al.*, 2007). Ces souches cliniques semblent ainsi avoir augmenté leur diversité génétique afin de mieux s'adapter aux conditions variées que peuvent leur offrir leurs hôtes.

Il faut noter que l'utilisation des biopuces WGA nécessite de sélectionner des couples d'amorces pour chaque ORF et d'amplifier chaque gène. De plus, du fait de la taille de ces produits PCR des problèmes d'aspécificité peuvent apparaître. Une alternative est l'utilisation de sondes oligonucléotidiques spécifiques de chaque ORF dont la spécificité doit être vérifiée. Dans ce cas, il est nécessaire de s'assurer que les oligonucléotides ne reconnaissent pas aspécifiquement une zone non ciblée.

3. LES BIOPUCES OLIGONUCLEOTIDIQUES PHYLOGENETIQUES.

Les biopuces phylogénétiques oligonucléotidiques ou « Phylogenetic Oligonucleotide Arrays » (POA) sont actuellement le type de biopuces le plus largement utilisé pour étudier les communautés bactériennes. Grâce aux sondes oligonucléotidiques, ciblant les ARNr 16S, les POAs permettent la discrimination des groupes procaryotiques. En effet, comme nous l'avons vu précédemment (cf. chapitre II, § 2.1.), les ARNr 16S sont des biomarqueurs de choix pour étudier les relations phylogénétiques entre les différents microorganismes et pour l'analyse de la structure des communautés procaryotiques dans les environnements naturels. Les zones des ARNr 16S, ciblées par les sondes, sont choisies en fonction du degré de résolution attendu : les séquences hautement conservées sont utilisées pour une détermination à des niveaux taxonomiques supérieurs (famille, ordre, classe) alors que les séquences hypervariables peuvent différencier les espèces et les genres. Cependant, comme les séquences des ARNr restent très conservées, la détermination de sondes spécifiques de microorganismes peut se révéler être difficile (Zhou et Thompson, 2002).

3.1. Sélection des sondes oligonucléotidiques.

La pertinence de la caractérisation des communautés microbiennes, par une approche biopuce POA, repose sur l'efficacité des sondes sélectionnées. Celles-ci doivent être hautement sensibles (Peplies *et al.*, 2006) et reconnaître spécifiquement les groupes ciblés, même ceux faiblement présents dans l'échantillon étudié (Gentry *et al.*, 2006). Dans l'optique de sélectionner des sondes performantes, il est nécessaire de pouvoir se baser sur des critères prédisant l'efficacité de l'hybridation sondes/cibles (Pozhitkov *et al.*, 2007). Généralement, les logiciels de sélection de sondes oligonucléotidiques basent leurs choix sur un ou plusieurs paramètres tels l'identité entre les sondes et les cibles, la position des mésappariements, la présence de structures secondaires, les paramètres thermodynamiques ou bien encore la position de la zone d'hybridation par rapport à l'extrémité de la cible (He *et al.*, 2005).

Un des premiers critères pris en considération est la similarité entre les sondes et les séquences ciblées et non ciblées. En effet, Kane *et al.* (2000) ont montré que les sondes oligonucléotidiques de 50-mers et les séquences non ciblées ne devaient pas présenter plus de

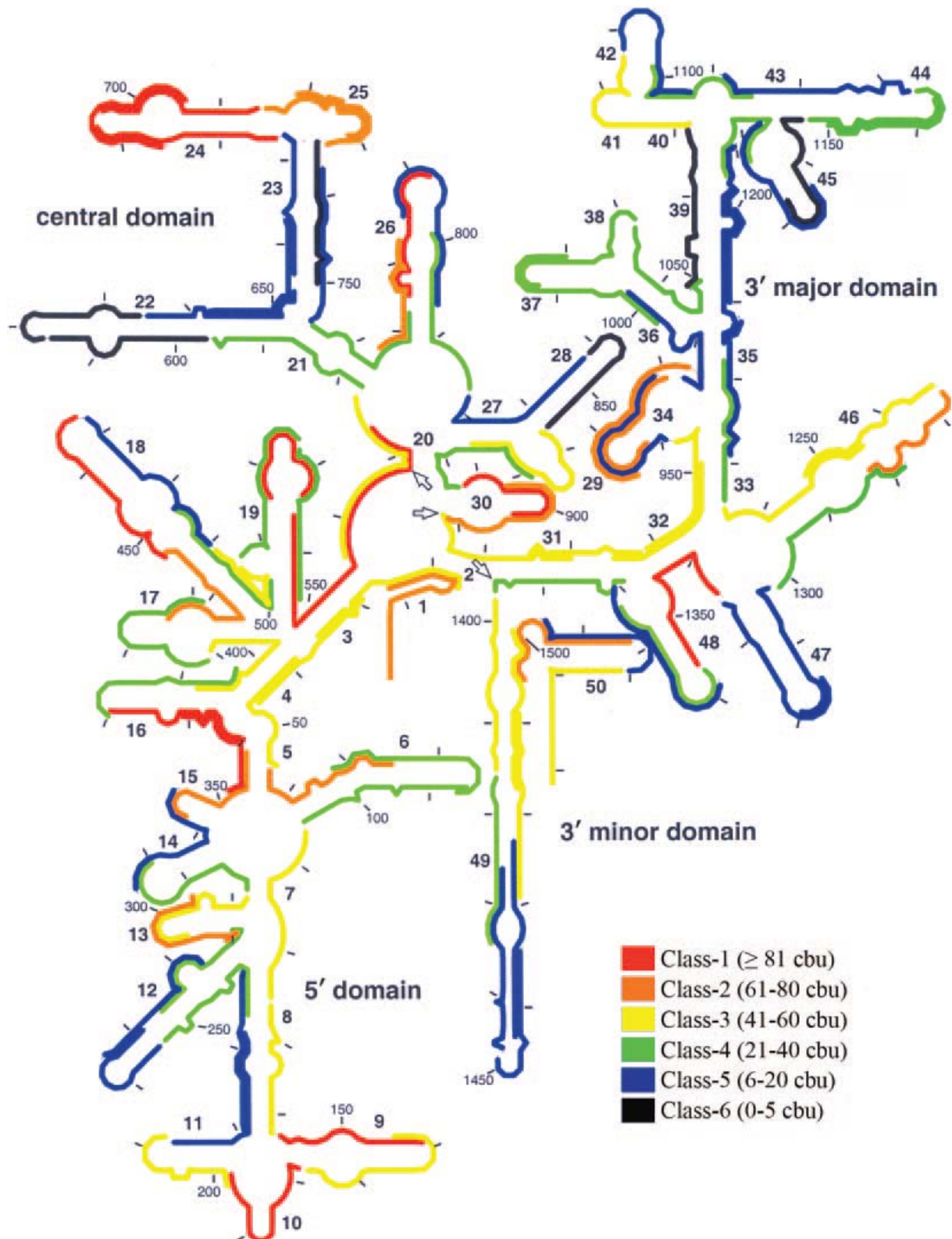


Figure 23 : Carte d'accessibilité de l'ARNr 16S. L'hybridation de centaines de sondes, tout au long de l'ARNr 16S, a permis de mettre en évidence des régions plus ou moins favorables aux hybridations : de la moins accessible (noir) à la plus accessible (rouge) (Yilmaz *et al.*, 2006).

75 % de similarité afin d'éviter une reconnaissance aspécifique. De plus, même si les sondes et les séquences non ciblées présentent une similarité inférieure à cette valeur seuil, elles ne doivent pas contenir de portions de séquences similaires de plus de 15 nucléotides successifs. Cependant, cette évaluation des hybridations croisées potentielles est difficile étant donné que la composition du mélange de cibles étudié est généralement inconnue. Afin de comprendre d'autres phénomènes impliqués dans la spécificité et la sensibilité des sondes, Urakawa *et al.* (2002) ont testé l'influence de la position et du type de mésappariements. Ainsi, un mésappariement localisé sur l'extrémité de la sonde déstabiliserait moins l'hybridation que ceux en position plus interne. Ces résultats vont dans le sens des travaux obtenus par Zhou *et al.* (2002) qui montrent qu'un seul mésappariement central déstabiliserait le duplex cibles/sondes au point de lui faire perdre de 15 à 25 % de fluorescence.

Les propriétés thermodynamiques des duplex sondes/cibles, largement étudiées en solution (SantaLucia *et al.*, 1996), ont aussi été utilisées pour sélectionner des sondes pour biopuces. Fotin *et al.* (1998) ont ainsi montré que les modèles définis pour les sondes en solution pouvaient être adaptés aux sondes attachées sur des substrats en gel de polyacrylamide. Yilmaz *et al.* (2006) ont proposé un modèle corrélant l'énergie libre globale de Gibbs ($\Delta G^{\circ}_{\text{globale}}$) et la fluorescence de sondes oligonucléotidiques utilisées pour les hybridations *in situ*. Ainsi, une carte représentant les zones de l'ARNr 16S plus ou moins favorables aux hybridations a été réalisée (figure 23). Toutefois, ces modèles ne peuvent s'appliquer aussi directement aux sondes fixées sur des lames de verre. En effet, d'autres paramètres tels la densité des sondes, la chimie de surface, la stabilité des duplex et les structures inter- et intra- molécules peuvent jouer sur l'efficacité des hybridations (Peterson *et al.*, 2001). Dans une étude de Pozhitkov *et al.* (2006), peu ou pas de corrélation a été obtenue entre les valeurs d'énergie libre de Gibbs et les intensités de signaux obtenues. Les auteurs suggèrent que les paramètres thermodynamiques ne peuvent pas être utilisés pour le design de sondes car leurs influences sur les efficacités d'hybridations sont encore loin d'être entièrement connues.

L'obtention de sondes spécifiques et sensibles est un réel défi mais la sélection de sondes exploratoires en est un encore plus grand. En effet, seule une petite partie des communautés microbiennes est connue à l'heure actuelle, ce qui implique que la majorité des séquences d'ADNr des microorganismes n'est pas disponible dans les bases de données. Or, la plupart des logiciels de sélection de sondes utilisent ces données incomplètes pour générer des sondes

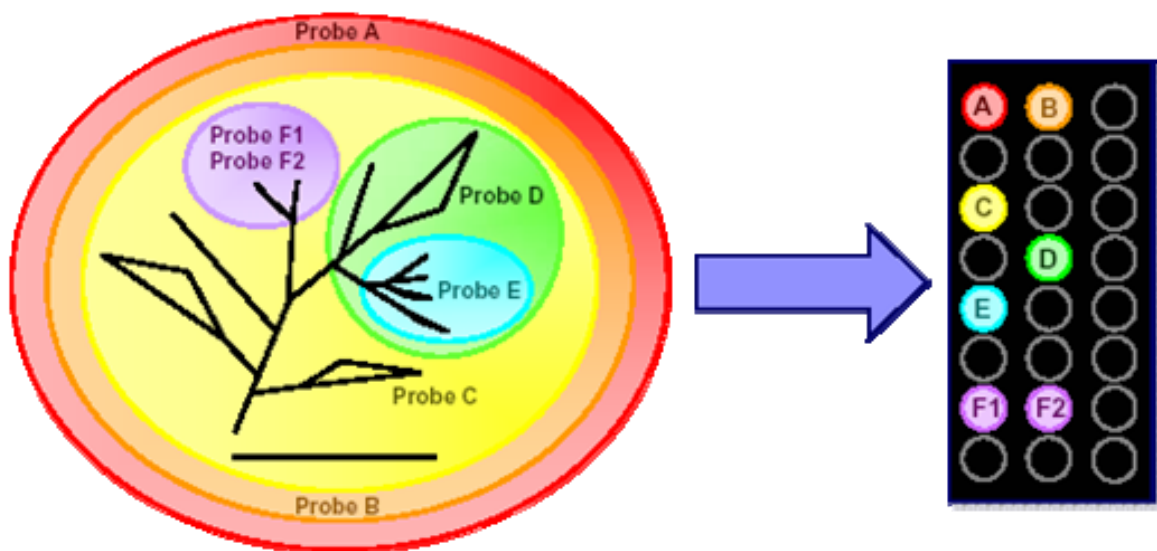


Figure 24: Schéma de détermination de sondes pour POAs (Phylogenetic Oligonucleotides Arrays). Les sondes ciblent plusieurs groupes microbiens à différents niveaux phylogénétiques. (Loy *et al.*, 2004).

« spécifiques » d'espèces. Ainsi, seule la fraction connue des communautés microbiennes peut être étudiée à l'aide de ces sondes. Toutefois, quelques logiciels de détermination de sondes tentent de diminuer ce biais grâce à la sélection de sondes ciblant des niveaux taxonomiques plus hauts que l'espèce (figure 24). Les logiciels Arb (Ludwig *et al.*, 2004) et Primrose (Ashelford *et al.*, 2002) peuvent générer ce type de sondes spécifiques de taxons. D'autres logiciels tentent de contourner ce problème en sélectionnant plusieurs sondes, plutôt qu'une seule, pour cibler un groupe microbien (Schliep and Rahmann, 2006).

3.2. Efficacité des sondes.

Sur les premières biopuces phylogénétiques, les sondes oligonucléotidiques étaient liées à des blocs de gel de polyacrylamide activés, eux-mêmes fixés à un support solide (Guschin *et al.*, 1997). L'utilisation de ce type de biopuce a permis les premières reconnaissances spécifiques d'acides nucléiques provenant de souches bactériennes (*Escherichia coli*, *Desulfovibrio vulgaris*, *Nitrosovibrio tenuis* et deux souches de *Nitrosomonas*). Les années suivantes, les biopuces phylogénétiques sur lame de verre, issues des avancées méthodologiques, confirmeront ces résultats. En effet, Small *et al.* (2001) ont reconnu spécifiquement les souches *Geobacter chapellei* et *Desulfovibrio desulfuricans* grâce à leurs sondes oligonucléotidiques. En ce qui concerne la sensibilité de ces biopuces, ils ont pu démontrer que la limite de détection des cibles de *Geobacter chapellei*, mélangées à des ARN extraits d'un sol, était de l'ordre de 500 ng d'ARN totaux, soit l'équivalent de $7,5 \cdot 10^6$ cellules. Les auteurs expliquent cette relativement faible sensibilité par l'analyse d'un extrait de sol non purifié et par l'utilisation d'un matériel de détection insuffisamment sensible.

La GeneChip® (Affymetrix) développée par le groupe de Gary L. Andersen, comportant 31 179 sondes oligonucléotidiques, a permis de reconnaître 15 espèces bactériennes sur les 17 testées. Cette biopuce s'est avérée, cependant, incapable de discriminer les souches *Rickettsia prowazekii* et *Mycoplasma pneumoniae* en mélange. Toutefois, les résultats obtenus sur des échantillons d'air ont montré des résultats concordants à ceux obtenus par une approche de clonage/séquençage. En effet, les deux approches ont mis en évidence la présence des mêmes groupes phylogénétiques. Seuls les nouveaux organismes, ne possédant pas de proches représentants dans les bases de données, n'ont pas pu être détectés par la biopuce phylogénétique (Wilson *et al.*, 2002). Ces résultats sont très encourageants pour l'utilisation des POAs dans les études d'environnements complexes mais doivent, toutefois, inciter les

chercheurs à rester prudents sur les données obtenues et à tenter de développer des sondes encore plus discriminantes et performantes.

3.3. Études de la diversité d'environnements complexes.

Ces dernières années, de nombreuses études utilisant les biopuces phylogénétiques oligonucléotidiques ont été réalisées pour explorer la diversité microbienne de différents environnements tels que les écosystèmes lacustres, les estuaires, des sols, les boues activées, les aérosols urbains, etc. (Koizumi *et al.*, 2002 ; Loy *et al.*, 2004 ; Castiglioni *et al.*, 2004 ; Sanguin *et al.*, 2006 ; Brodie *et al.*, 2007). Dans ce paragraphe, nous nous focaliserons sur les études menées sur les sols.

Afin d'étudier la diversité des communautés sulfato-réductrices (SRP), Loy *et al.* (2002) ont développé une puce phylogénétique, nommée « SRP-PhyloChip », composée de 132 sondes de 18-mers. Loy *et al.* (2004) ont utilisé cette biopuce pour caractériser les procaryotes impliqués dans la réduction du sulfate au niveau d'écosystèmes terrestres qui, au contraire de ceux procédant en milieux aquatiques anoxiques, sont très peu connus. Leurs modèles d'étude ont été les couches supérieures de deux marécages (Schlössnerbrunnen I et II) qui, bien que montrant de faibles concentrations en sulfate (20 à 200 μM) et une acidité assez importante (pH 4-5), présentent des activités sulfato-réductrices significatives. La SRP-PhyloChip a permis de caractériser une diversité assez stable des SRP entre ces deux marécages. Des membres de la famille des *Syntrophobacteraceae* ont été observés dans les deux échantillons et la présence du genre *Desulfomonile* n'a été démontrée que sur l'un des deux sites. Ces résultats de biopuces ont été confirmés par clonage/séquençage des gènes réductase dissimilatrice du sulfate (*dsrAB*). Cependant, cette approche a aussi mis en évidence que la SRP-PhyloChip n'a pas détecté la présence de SRP type *Desulfobacca*.

Dans le cadre d'une autre étude, Sanguin *et al.* (2006) ont étudié la composition des communautés bactériennes au niveau d'une rhizosphère de maïs et d'un sol non rhizosphérique grâce à une biopuce phylogénétique de 170 sondes ciblant essentiellement le phylum des *Proteobacteria*. Un effet significatif de la rhizosphère a été démontré à la fois par la présence plus importante de bactéries du genre *Agrobacterium* et par la faible représentation des phyla *Acidobacteria*, *Bacteroidetes*, *Verrucomicrobia* et *Planctomycetes*. Par contre, des bactéries appartenant aux groupes *Sphingomonas*, *Rhizobiaceae* et *Actinobacteria* ont été observées dans les deux environnements. Ces résultats ont été

Tableau 3. Caractéristiques des différents formats de biopuces utilisées en écologie microbienne (adapté de Gentry *et al.*, 2006). *POA* : *Phylogenetic Oligonucleotide Microarray* ; *FGA* : *Functionnal Gene Array* ; *CGA* : *Community Genome Array* ; *WGA* : *Whole-Genome Array* ; *ORF* : *Open Reading Frame* ; *nt* : *nucleotides*.

Propriétés	POA	FGA	CGA	WGA
Sondes	ADNr	Gènes fonctionnels	Génome entier	ORF
Longueur de sondes	~18-25 nt	~50-70 nt (oligos) ~200-1000 nt (PCR)	Génome entier	~25 nt (oligos) ~200-3000 nt (PCR)
Microorganismes ciblés	Cultivable et non cultivable	Cultivable et non cultivable	Cultivable	Cultivable
Informations fournies	Phylogénétique	Fonctionnelle	Phylogénétique	Fonctionnelle et Phylogénétique
Spécificité	Niveau spécifique ou différence d'un seul nt	<80-90% d'homologie	Niveau de l'espèce ou de la souche	Différence d'un seul nt
Sensibilité (ng d'ADN génomique pur)	~500	~1-8	~0,2	~0,1
Quantitatif	Variable	Oui	Oui	Variable

comparés à ceux obtenus par une approche clonage/séquençage. Bien que la concordance entre les données obtenues par ces deux types d'approches ne soit pas parfaite, du fait du faible nombre de clones analysés et des groupes bactériens non ciblés par la biopuce, les résultats sont tout de même très prometteurs.

Une des dernières études menées sur la diversité des communautés microbiennes utilisant des POAs est celle de DeSantis *et al.* (2006) qui ont utilisé une GeneChip® de 297 851 sondes ciblant les ARNr 16S de 842 sous-familles procaryotiques. Ainsi, la communauté procaryotique de trois types d'environnements différents a pu être étudiés : un sol contaminé, un aquifère contaminé biostimulé et des échantillons d'air. Afin de confirmer les données obtenues avec ces biopuces POAs, trois banques de clones, comprenant respectivement 417, 485 et 253 séquences, ont été construites à partir des échantillons d'air, de sol et d'eau. Cette étude montre que l'utilisation des biopuces permet de révéler une plus grande part de la diversité procaryotique que l'approche par clonage/séquençage. En effet, dans le cas du sol, la biopuce POA a permis la détection de 279 sous-familles dont 239 n'ont pas été révélées par la banque de clones. Par contre, les banques de clones ont permis quant à elles de mettre en évidence la présence de nouvelles sous-familles (14,4 % dans le cas du sol).

Ces différentes études confirment donc que les biopuces sont particulièrement adaptées pour l'étude globale des différents groupes microbiens connus.

CONCLUSIONS

Compte tenu de l'importance et de la diversité de la biomasse procaryotique au niveau des environnements naturels, les outils à haut débit, tels les biopuces, semblent désormais incontournables (tableau 3). Les biopuces phylogénétiques oligonucléotidiques qui sont les biopuces les plus abouties permettent de suivre simultanément des centaines voire des milliers de groupes microbiens simultanément. Bien qu'elles ne soient pas pour l'instant réellement quantitatives et exploratoires, de nombreuses études ont montré que leur utilisation était possible pour étudier des environnements complexes. Ainsi, leur application pour suivre les communautés bactériennes des sols et notamment des sols pollués semble particulièrement bien adaptée. Elles permettront ainsi d'obtenir de nombreuses informations qui compléteront celles déjà obtenues par d'autres approches moléculaires et aideront à la compréhension globale du fonctionnement des écosystèmes terrestres pollués, et à la caractérisation de souches présentant un intérêt en bioremédiation.

Chapitre IV

Diversité bactérienne dans les écosystèmes sols.

INTRODUCTION

Dans la plupart des écosystèmes naturels, les communautés microbiennes sont denses et diversifiées (Curtis et Sloan, 2004). Wackett et Hershberger (2000) ont ainsi estimé qu'un gramme de sol pouvait compter, en moyenne, 10^9 bactéries représentant plus de 10 000 espèces différentes dont certaines sont ubiquistes et d'autres endémiques. D'après de nombreuses études, la biodiversité de ces écosystèmes pourrait être directement influencée par les caractéristiques intrinsèques des sols (Sessitsch *et al.*, 2001). En effet, ce sont des écosystèmes complexes composés de nombreux microhabitats et présentant des conditions environnementales bio-physico-chimiques très variées (Ranjard et Richaume, 2001). Ainsi, ils représentent une mosaïque d'environnements qui sont autant de niches écologiques différentes pour les microorganismes.

Les approches dépendantes de la culture ont longtemps été utilisées pour l'investigation de la diversité bactérienne de différents écosystèmes. Cependant, avec l'apparition des techniques moléculaires, les techniques culturales sont apparues comme donnant une vision tronquée et donc potentiellement biaisée de la biodiversité. Aussi, les techniques indépendantes de la mise en culture sont aujourd'hui largement utilisées. Elles ont permis de faire des inventaires, non exhaustifs, des communautés bactériennes présentes dans différents types de sols mais également d'étudier leur répartition au niveau de ces écosystèmes.

1. DIVERSITE BACTERIENNE DES SOLS.

1.1. Structuration des communautés bactériennes.

Les bactéries ont une localisation préférentielle au niveau des limons fins (2 à 20 μm) et de la fraction fine ($< 2 \mu\text{m}$) des sols car elles y trouvent des conditions favorables à leur survie

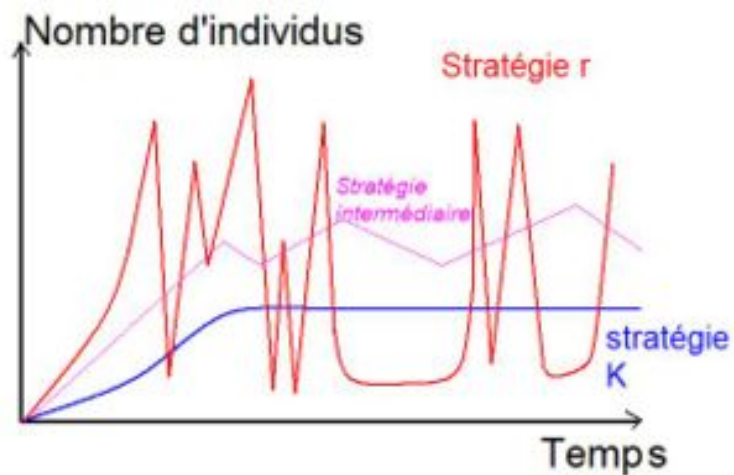


Figure 25 : Modèles évolutifs r/K. La survie de l'espèce devant être assurée constamment, face à des environnements différents, l'évolution a sélectionné des stratégies différentes. Les espèces vivant dans les milieux instables et imprévisibles adoptent une stratégie de la reproduction rapide et massive: c'est la stratégie r. Dans les milieux prévisibles, aux variations stables ou cycliques, la stratégie sélectionnée est celle de la reproduction moins rapide mais efficace, notamment par rapport aux quantités de ressources nécessaires. C'est la stratégie K (Maier *et al.*, 2000).

(Duchaufour, 2001). La porosité de ces fractions fines est constituée majoritairement de micropores où l'on retrouve environ 80 % des microorganismes. Ceci s'explique par plusieurs facteurs. En effet, les fractions fines du sol présentent de fortes quantités de matières organiques (50 à 80 %) qui sont les seules sources de carbone et d'énergie pour les bactéries hétérotrophes. De plus, comme les microporosités retiennent fortement l'eau par des phénomènes de capillarité (Duchaufour, 2001), ces espaces sont donc favorables aux microorganismes car ils les protègent contre la dessiccation et leur fournissent des nutriments. Par ailleurs, du fait de leur taille, les micropores protègent les microorganismes contre la prédation par les protozoaires (pour revue voir Ranjard et Richaume, 2001).

La localisation préférentielle des microorganismes au niveau des limons et de la fraction fine plutôt qu'au niveau des sables fins (50 à 200 μm) a été confirmée par Sessitsch *et al.*, (2001). Au cours de ce travail, les auteurs suggèrent que la présence de cette importante biomasse pourrait être la conséquence d'une plus grande diversité plutôt que d'une forte colonisation des espaces due à l'atténuation des pressions de prédation (protozoaires) et de la diminution de la compétition avec les mycètes. L'analyse par T-RFLP et par clonage/séquençage des ADNr 16S a également permis de mettre en évidence que, contrairement aux macropores qui sont dominés par des *Proteobacteria* de type α , les micropores comprendraient principalement des *Acidobacterium* du genre *Holophaga* et des *Prostheco bacter*. Cependant, selon Smit *et al.* (2001), cette répartition des bactéries serait plutôt à corrélérer avec la disponibilité des nutriments. En effet, une forte disponibilité des substrats favoriserait les espèces à stratégie r (favorisant une division des populations rapide et massive), telles les α - et γ -*Proteobacteria*, alors qu'une faible disponibilité favoriserait plutôt les espèces à stratégie K (favorisant une division moins rapide mais plus efficace) comme les *Acidobacterium*. Ces deux théories ne sont cependant pas si opposées si on prend en considération que les macropores sont des environnements plus instables du point de vue nutritif qui nécessitent que les microorganismes soient à adaptation rapide alors que les micropores sont des environnements plus stables qui nécessitent que les microorganismes soit compétitifs par rapport aux substrats (figure 25).

1.2. Diversité des communautés bactériennes.

1.2.1. Approches culturales.

Les approches culturales sont des méthodes qui ont été traditionnellement utilisées pour accéder à la diversité microbienne des sols. En 1988, une collection de souches bactériennes provenant de différents environnements terrestres a été établie : la Subsurface Microbial Culture Collection (SMCC). Celle-ci recense environ 8 000 souches différentes. A partir de 1993, des typages moléculaires ont été réalisés sur trois groupes bactériens de cette collection : les chimioautotrophes aérobies, les chimiohétérotrophes aérobies et les anaérobies. Ainsi, la présence au niveau des sols des phyla *Bacteroidetes*, *Actinobacteria* (Gram positif à haut GC%), *Firmicutes* (Gram positif à faible GC%) et des classes α -*Proteobacteria*, β -*Proteobacteria* et γ -*Proteobacteria*, a pu être mise en évidence (Balkwill *et al.*, 1997). La présence de ces divisions et sous-divisions bactériennes a aussi été démontrée dans des collections provenant de rizières et de prairies (Hattori *et al.*, 1997). Plus récemment, Joseph *et al.* (2003) ont réussi à cultiver 350 isolats bactériens à partir de terre provenant d'un pré pâturé. Une analyse phylogénétique des ADNr 16S a alors permis de mettre en évidence une dominance des *Actinobacteria* et des α -*Proteobacteria*. La présence de bactéries appartenant aux *Acidobacteria*, *Bacteroidetes*, *Deinococcus-Thermus*, *Firmicutes*, *Gemmatimonadetes*, *Planctomycetes*, β -*Proteobacteria*, γ -*Proteobacteria* et *Verrucomicrobia* a également été observée. A un niveau phylogénétique plus fin, Alexander (1961) a suggéré que les genres *Agrobacterium*, *Alcaligenes*, *Arthrobacter*, *Bacillus*, *Flavobacterium*, *Micromonospora*, *Nocardia*, *Pseudomonas* et *Streptomyces* étaient des membres fortement représentés dans les communautés bactériennes du sol.

Même si ces analyses permettent d'obtenir des informations sur les communautés bactériennes des sols, celles-ci sont encore largement incomplètes et biaisées. Torsvik *et al.* (2002) ont ainsi estimé que seulement 0,1 à 1 % des bactéries du sol pouvaient être isolés par les techniques de culture classiques. De plus, seule la moitié des phyla bactériens possèderaient des représentants cultivés (Rappé et Giovannoni, 2003). En effet, les techniques de culture sur des milieux riches en nutriment isolent préférentiellement les bactéries à stratégie r (Simu and Hagstrom, 2004). C'est pourquoi de nouvelles méthodes de culture ont été développées afin de pouvoir accéder aux bactéries présentant un modèle évolutif différent. Ainsi, Ferrari *et al.* (2005) ont développé une nouvelle méthode de culture simulant les conditions environnementales du sol qui a permis la culture d'une bactérie appartenant au phylum TM7 qui ne possédait dès lors aucun représentant cultivé.

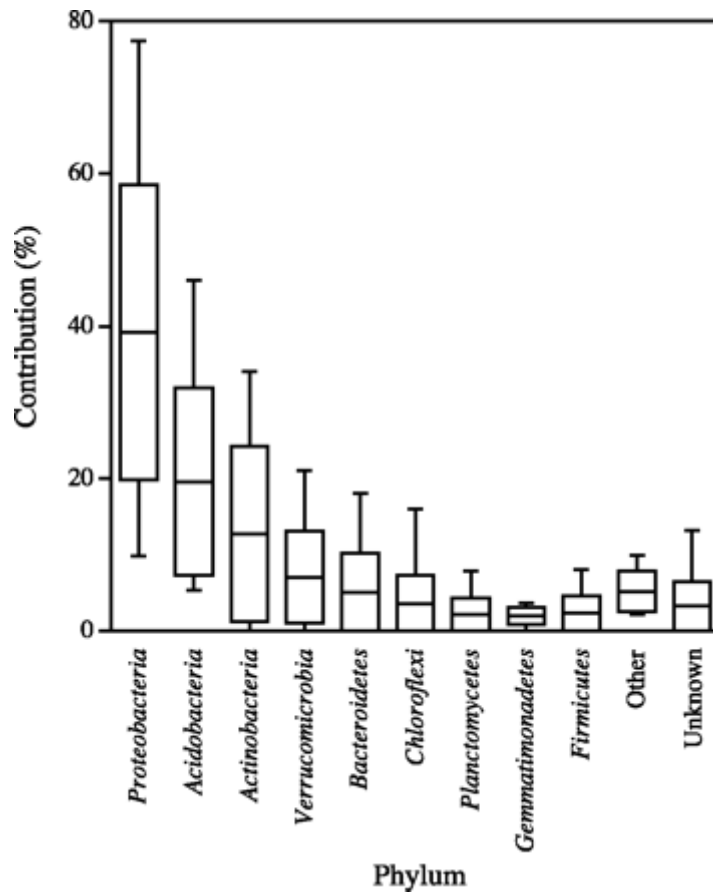


Figure 26 : Box-plots représentant la contribution des différents phyla bactériens à la biodiversité des sols. La ligne horizontale, au milieu de la boîte, indique le pourcentage moyen de la contribution de chaque phylum. Les extrémités de la boîte indiquent la déviation standard autour de la moyenne. Enfin, les lignes verticales indiquent les valeurs extrêmes des contributions (Janssen, 2006).

Bien que d'énormes progrès méthodologiques aient été réalisés en ce qui concerne les approches culturales, celles-ci ne permettent pas encore d'avoir une vision globale de la diversité procaryotique au niveau des sols. Il est donc nécessaire d'utiliser des approches indépendantes de la mise en culture afin d'étudier cette biodiversité.

1.2.2. Approches moléculaires.

Parmi les différentes techniques de biologie moléculaire, ce sont les approches de clonage/séquençage qui ont été les plus largement utilisées pour étudier la diversité bactérienne. Au travers de la réalisation et de l'affiliation de banques de clones, principalement d'ADNr 16S mais aussi d'ARNr 16S, elles ont permis d'accroître nos connaissances sur la composition des communautés bactériennes. Le caractère ubiquiste des *Proteobacteria*, des *Bacteroidetes*, des *Actinobacteria* et des *Firmicutes* a ainsi pu être mis en évidence. Ces quatre divisions représenteraient à elles seules, les $\frac{3}{4}$ des séquences environnementales et 90 % des bactéries cultivées (Hugenholtz *et al.*, 1998). Le clonage/séquençage a aussi permis de révéler l'importance de taxons ignorés ou sous-estimés par les approches culturales : les *Acidobacterium*, les *Verrucomicrobia*, les bactéries GNS (Green Non Sulfur) et les OP11. Cependant, outre certains groupes qui semblent être présents dans tous les écosystèmes (*Proteobacteria*, *Cytophagales*, *Firmicutes*), il existe aussi des phyla qui semblent occuper des niches écologiques particulières comme les OP11 au niveau des sols (Hugenholtz *et al.*, 1998).

Une synthèse des résultats réalisée par Janssen (2006), à partir de toutes les banques de clones ADNr 16S et ARNr 16S provenant de sols non rhizosphériques, a permis d'identifier les espèces bactériennes dominant ces environnements. Ainsi, il s'est avéré que les bactéries du sol étudié pouvaient être affiliées à 32 phyla dont les abondances varient en fonction des environnements. Par exemple, les *Proteobacteria* peuvent présenter des abondances allant de 10 à 80 % selon les sols étudiés (figure 26). Cependant, seuls 9 phyla semblent être dominants dans les sols. Ce sont essentiellement les *Proteobacteria* et les *Acidobacteria* mais aussi les *Actinobacteria*, les *Verrucomicrobia*, les *Bacteroidetes*, les *Chloroflexi*, les *Planctomycetes*, les *Gemmatimonadetes* et les *Firmicutes* (figure 26) qui représentent en moyenne 92 % des banques. Les autres taxons présents sont les *Chlamydiae*, *Chlorobi*, *Cyanobacteria*, *Deinococcus-Thermus*, *Fibrobacteres*, *Nitrospirae*, BRC1, NKB19, OP10, OP11, OS-K, SC3, SC4, termite groupe 1, TM6, TM7, WS2 et WS3. L'importance des *Firmicutes*, des *a-*

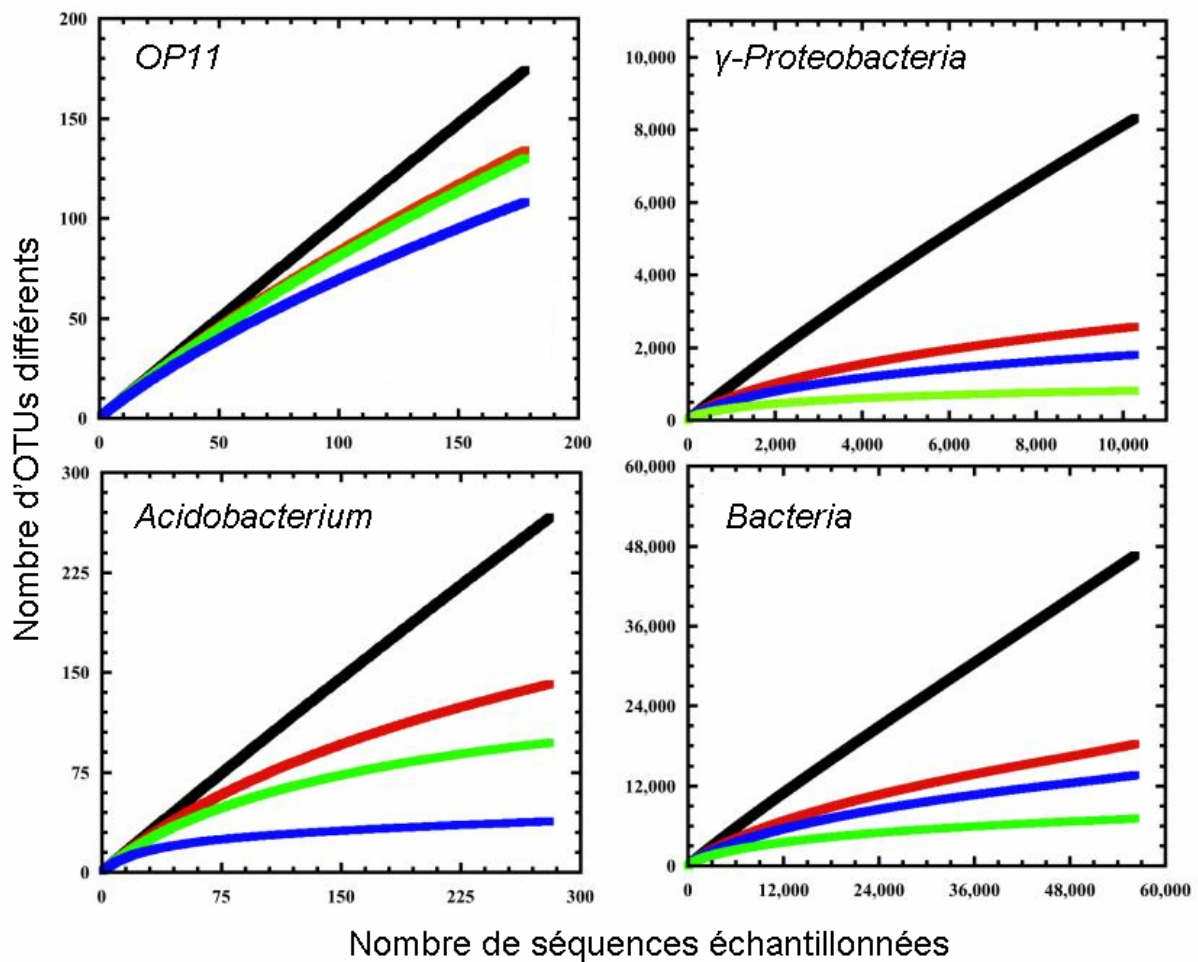


Figure 27 : Estimation du niveau de connaissance de différents phyla et des bactéries en général. Des courbes de raréfaction ont été réalisées, grâce aux séquences de la base de données ribosomique (RDP-II), à différents niveaux phylogénétiques : individu (noir), espèce (rouge), genre (vert) et famille/classe (bleu). Ces courbes montrent que le taux de récupération de nouvelles séquences bactériennes est important et que les asymptotes, correspondant à une diversité complètement connue, ne sont pas atteintes quelque soit le groupe. *OTU* : *Operational Taxonomic Unit*.

Proteobacteria, des *Holophaga-Acidobacterium* et des *Verrucomicrobia* a aussi été observée par Hackl *et al.* (2004), au niveau de différents sols forestiers, grâce à l'utilisation couplée du clonage/séquençage et de la technique T-RFLP. Les résultats obtenus à l'aide d'autres méthodes moléculaires n'impliquant pas l'analyse de banques, telle la technique FISH, ont aussi montré la forte proportion des *Proteobacteria* dans le sol. Ainsi, Song *et al.* (2004) ont mis en évidence que les communautés bactériennes d'un sol étaient composées de 13,7 % d' α -*Proteobacteria*, de 31,7 % de β -*Proteobacteria*, de 17,1 % de γ -*Proteobacteria*, de 16,8 % de *Cytophaga-Flavobacterium* et de 20,8 % d'autres groupes bactériens.

Idéalement, l'exploitation de ces séquences pourrait permettre d'avoir une vue globale de la diversité des communautés bactériennes au niveau des sols. Cependant, l'analyse de l'ensemble de ces séquences ribosomiques présentes au niveau du Ribosomal Database Project II (RDP-II) a montré que le recensement de la diversité bactérienne était loin d'être complet (Schloss et Handelsman, 2004) (figure 27). Une des raisons, hormis le fait que le nombre d'environnements différents étudiés n'est pas assez important, pourrait provenir de la taille des banques de clones. Kemp et Aller (2004) ont estimé statistiquement que les banques de clones jusqu'alors séquencées ne permettaient de couvrir qu'environ 48 % des communautés microbiennes des sols. Des banques de plus de 400 clones seraient nécessaires afin d'avoir une vision plus réaliste de la diversité bactérienne ou procaryotique de ces environnements. L'utilisation de techniques moléculaires à haut-débit pourrait donc permettre d'avoir une vue plus exhaustive des communautés bactériennes. Ce type d'outil ne remplace cependant pas pour l'instant les approches de clonage/séquençage car il ne permet pas d'avoir une vision exploratoire.

2. DIVERSITE BACTERIENNE DES SOLS POLLUES.

De part l'intérêt grandissant pour les techniques de bioremédiation, les écosystèmes pollués ont également été étudiés afin de découvrir des microorganismes épurateurs utilisables pour ces procédés alternatifs de décontamination. Les recherches concernant les sols pollués par des hydrocarbures se sont surtout focalisées autour des polluants de types HAP et BTEX. Comparativement, il existe relativement peu d'informations sur les écosystèmes pollués par des hydrocarbures aliphatiques. En effet, ceux-ci sont moins étudiés car ils sont considérés comme étant moins dangereux pour la santé (Gieg et Sulfita, 2002). Cependant, ces polluants

restent les plus importants quantitativement et leurs impacts sur la santé ne sont pas à négliger.

2.1. Adaptations des populations suite à une pollution.

Lors de l'apparition d'une pollution au niveau des sols, des pressions de sélection s'exercent sur les microorganismes endogènes induisant une restructuration des communautés bactériennes. En ce qui concerne les polluants organiques de type hydrocarbures, ceux-ci influent de façon passive et s'accumulent au niveau des membranes hydrophobes des bactéries (Sikkema *et al.*, 1995). Ce phénomène, nommé partitionnement, entraîne une modification de la structure des membres et peut altérer les fonctions des enzymes membranaires. Certaines populations sont alors favorisées par rapport à d'autres si elles présentent des réponses comportementales ou physiologiques et/ou des capacités métaboliques adaptées à la présence de ce polluant et à son utilisation. Les bactéries, pour lesquelles l'impact du polluant n'est pas léthal, peuvent mettre en place plusieurs types de réponse pour pallier ce stress (échappement par chimiotactisme, formation de biofilms, pompes à efflux, modification de l'hydrophobicité ou réparation des membranes) (Van Hamme *et al.*, 2003). Au contraire, pour les bactéries qui utilisent ces polluants organiques, ces mécanismes d'influx passifs sont bénéfiques car ils leur permettent d'acquérir les substrats de certaines de leurs voies métaboliques. Chez des bactéries de type *Rhodococcus erythropolis* ou *Sphingomonas paucimobilis*, Kim *et al.* (2002) et Story *et al.* (2000) ont même observé la présence de systèmes d'influx actifs des hydrocarbures. En fonction des réponses que les bactéries peuvent apporter à la présence de polluants, leurs abondances au sein de leurs écosystèmes vont varier. Ainsi, en schématisant, les populations bactériennes qui ne possèdent pas de mécanismes d'adaptation vont tendre à disparaître. Celles qui possèdent ces mécanismes mais qui ne métabolisent pas le polluant vont être relativement stables. Enfin, les bactéries qui possèdent des voies de dégradation du polluant pourront l'utiliser comme source de carbone et d'énergie et ainsi se développer.

Powell *et al.* (2006) ont suggéré qu'une fois les polluants utilisés, les populations bactériennes évoluent de nouveau en favorisant les bactéries pouvant utiliser les produits de dégradation des hydrocarbures ou d'autres sources de carbone et d'énergie. Cette adaptation des communautés bactériennes aux hydrocarbures est possible car les bactéries dégradant les composés organiques semblent ubiquistes au niveau des sols, or ces deux types de composés peuvent présenter des similarités structurales (Van Hamme *et al.*, 2003). Toutefois, il faut

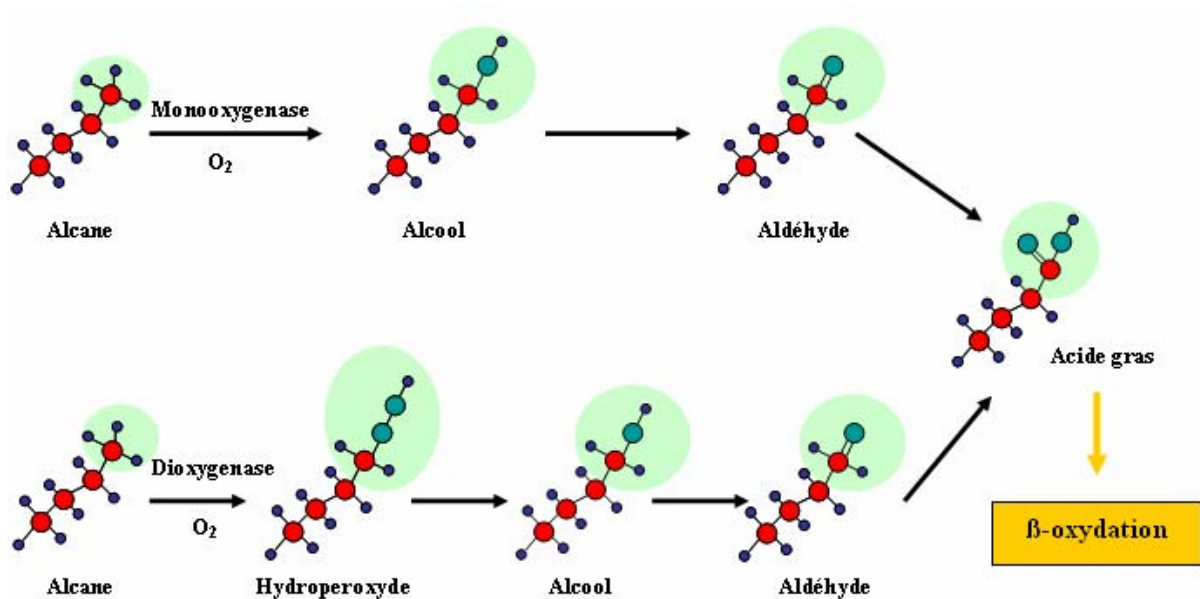


Figure 28 : Biodégradation aérobie des alcanes (adapté de Maier *et al.*, 2000).

noter que la restructuration des communautés microbiennes peut être aussi influencée par d'autres paramètres tels le type de polluants ou le type de sol. En effet, bien que les mélanges d'hydrocarbures comportent souvent les mêmes composés (alcane linéaires ou ramifiés, cycloalcanes, hydrocarbures phénoliques, HAP, etc.), ceux-ci ne sont généralement en même concentration. De plus, les propriétés bio-physico-chimiques des sols (microorganismes endogènes, température, pH, nutriments, etc.) varient d'un écosystème à l'autre. Ainsi, Hamamura *et al.* (2006) ont étudié l'influence d'un même mélange d'hydrocarbures, majoritairement composé d'hexadécane, sur l'évolution des communautés microbiennes de 7 sols pollués présentant des propriétés bio-physico-chimiques différentes. Les résultats montrent des abattements de polluants comparables au bout de 50 jours, mais les taux initiaux de dégradation des polluants et les types d'hydrocarbures préférentiellement dégradés sont variables en fonction des sols. Les différences de biodégradation sont ainsi imputables au développement de différentes populations bactériennes en fonction du type de sol. Cependant, certains groupes bactériens se révèlent être récurrents quelles que soient les caractéristiques du sol ce qui suggère leur importance dans la biodégradation des alcanes. Ces populations peuvent alors être de deux types : les bactéries aérobies dégradant les hydrocarbures en présence d'oxygène et les bactéries anaérobies ne nécessitant pas d'oxygène pour cette dégradation. Ces deux groupes sont présentés dans les paragraphes ci-dessous.

2.2. Bactéries aérobies dégradant les polluants.

Dans le cas des hydrocarbures aliphatiques les plus saturés, la première étape de la dégradation aérobie implique généralement une monooxygénase qui, en présence d'oxygène, transforme les alcanes en alcools (figure 28). Ceux-ci sont ensuite transformés en aldéhydes puis en acides gras qui seront ultérieurement utilisés dans la voie de la β -oxydation pour la biosynthèse des phospholipides membranaires et du matériel de stockage des lipides (Maier *et al.*, 2000). Ainsi, Powell *et al.* (2006) ont montré une corrélation positive entre la concentration en *n*-alcane, au niveau du sol, et la proportion de bactéries possédant le gène *alkB* codant pour une alcane monooxygénase. Il faut aussi noter que d'autres voies alternatives existent impliquant des enzymes telles la dioxygénase ou la cis-desaturase lors des premières étapes de la voie de dégradation (Van Hamme *et al.*, 2003). La biodégradation des hydrocarbures aliphatiques insaturés implique soit une oxydation monoterminale ou subterminale des alcènes ou une attaque de la double liaison entraînant la formation d'alcools

ou d'époxydes qui seront transformés en acides gras eux aussi utilisés dans la voie de la β -oxydation.

Comme nous l'avons vu précédemment, l'évolution des populations bactériennes suite à une pollution peut mettre en évidence les capacités épuratrices de certains microorganismes. En effet, l'augmentation de groupes bactériens au niveau d'un environnement pollué pourrait dénoter une capacité à dégrader le polluant. Ainsi, afin d'identifier les bactéries possédant ces aptitudes, des comparaisons de populations de sols pollués et non pollués ont été réalisées. Globalement, ces comparaisons semblent révéler que les divisions bactériennes majoritaires au niveau des sols sains le sont aussi au niveau des sols contaminés et que les variations de populations interviendraient à des niveaux phylogénétiques plus fins. L'étude parallèle, par DGGE et clonage/séquençage des ADNr 16S, des populations bactériennes de sols alpins sains et contaminés aux hydrocarbures a pu mettre en évidence que les espèces appartenant aux divisions des *Proteobacteria* et des *Actinobacteria* sont majoritaires qu'il y ait une pollution ou non (Labbé *et al.*, 2007). Par contre, les sols pollués et les sols sains présenteraient des abondances différentes au niveau des différentes sous-divisions des *Proteobacteria*. En effet, la présence des hydrocarbures semble induire une diminution des α -*Proteobacteria* (de 46 à 24 %) au profit des β -*Proteobacteria* et des γ -*Proteobacteria*. Cette évolution confirme les observations faites par Margesin *et al.* (2003) sur le même type de sol où suite à une pollution par des *n*-alcanes, les espèces appartenant aux γ -*Proteobacteria* ont été favorisées. Le développement des β - et des γ -*Proteobacteria*, lors d'une pollution aux hydrocarbures aliphatiques, a aussi été démontré dans des sols autres que alpins (Hamamura *et al.*, 2006).

L'identification de bactéries possédant des capacités épuratrices peut aussi être réalisée au cours de la bioremédiation des sols pollués. Ainsi, une étude sur l'évolution de la diversité d'un sol pollué aux alcanes et aux BTEX pendant un processus de bioremédiation a permis de mettre en évidence que le sol pollué est toujours dominé par des microorganismes de type γ -*Proteobacteria* (Popp *et al.*, 2006). Les bactéries appartenant aux sous-divisions α -*Proteobacteria* et β -*Proteobacteria* sont toutefois importantes tout au long de la décontamination. Par ailleurs, la plupart des souches bactériennes, caractérisées au cours de ce travail, avaient déjà été identifiées au niveau de sols contaminés par des hydrocarbures (Eriksson *et al.*, 2003 ; Thiel *et al.*, 2005). Elles appartiennent aux genres *Pseudomonas*, *Acinetobacter*, *Sphingomonas*, *Acidovorax* et *Thiobacillus* (Popp *et al.*, 2006). Ceci laisserait supposer que ces souches puissent être impliquées dans la dégradation de ces polluants. Popp

et al. (2006) suggèrent que la déviation de la diversité vers les γ -*Proteobacteria*, nommée « gamma-shift », interviendrait lors d'une augmentation de la richesse en nutriments ou lors de l'apparition d'une contamination. Ce gamma-shift pourrait être causé par le grand nombre de bactéries dégradant les hydrocarbures appartenant à cette sous-division. Les *Pseudomonales*, et surtout le genre *Pseudomonas*, sont très souvent dominants dans les sites contaminés aux hydrocarbures (Evans *et al.*, 2004). Ils semblent, en effet, être impliqués dans les premières phases des procédés de bioremédiation lorsque les polluants sont biodisponibles (Kaplan et Kitts, 2004). Les *Moraxellaceae*, avec le genre *Acinetobacter*, sont souvent présents lors des pollutions et ont montré des capacités à dégrader des alcanes linéaires en C₁₆ (Bogan *et al.*, 2003). La présence des genres *Xanthomonas*, *Stenotrophomonas* et *Hydrocarboniphaga* a aussi été reportée au niveau de sols contaminés (Popp *et al.*, 2006). Il faut cependant noter qu'il existe aussi des bactéries identifiées comme étant épuratrices et n'appartenant pas aux γ -*Proteobacteria* telles les *Actinobacteria*, les *Sphingomonas* et *Methylobacterium* qui appartiennent aux α -*Proteobacteria*, les *Rhodospirillum rubrum*, *Acidovorax* et *Thiobacillus* qui appartiennent aux β -*Proteobacteria*, et les *Flavobacterium* qui appartiennent aux *Flavobacteria* (Popp *et al.*, 2006).

2.3. Bactéries anaérobies dégradant les polluants.

En absence d'oxygène, la biodégradation des hydrocarbures est possible grâce à des *consortia* bactériens spécifiques qui utilisent des accepteurs d'électrons autres que l'O₂. En cultures pures, des bactéries méthanogènes ou des bactéries anaérobies réduisant les nitrates, les sulfates et le fer ont montré des capacités à dégrader les alcanes. En anaérobiose, l'activation des alcanes pourrait être similaire à celle du toluène impliquant le fumarate dans la formation d'un succinate substitué (Widdel and Rabus, 2001). Ces dégradations anaérobies sont des processus qui ont été largement étudiés au niveau des écosystèmes aquatiques mais qui sont encore assez méconnus au niveau terrestre (Kasai *et al.*, 2005).

Actuellement, il existe très peu d'études relatives à la biodégradation des hydrocarbures aliphatiques en anaérobiose. Au niveau des sols, la biodégradation des hydrocarbures aliphatiques la plus étudiée est celle des éthènes chlorés comme le tetrachloroéthène (PCE). Ceci est dû au fait que ces polluants sont largement répandus, cancérigènes et qu'ils ont longtemps été considérés comme résistants aux biodégradations aérobies. En anaérobiose, certains microorganismes peuvent déhalogéner ce PCE en trichloroéthène, dichloroéthène, chlorure de vinyle puis en éthène non toxique. Ce procédé respiratoire, nommé

halorespiration, peut être réalisé par certaines espèces bactériennes du genre *Dehalospirillum*, *Dehalobacter*, *Desulfitobacterium*, *Dehalococcoides ethenogenes* et *Sulfurospirillum* (Luijten *et al.*, 2003; Zhang et Bennet, 2005).

Les processus de biodégradation anaérobie sont donc importants et ne doivent pas être négligés et ce pour plusieurs raisons : (1) en effet, les sols peuvent présenter des anaérobioses lorsqu'ils sont saturés ou lorsque que l'oxygène a été consommé au niveau des microporosités ; (2) certains xénobiotiques ne peuvent être dégradés efficacement que par des bactéries anaérobies ; (3) enfin les processus anaérobies sont moins coûteux à mettre en place que les processus aérobie au cours d'un procédé de bioremédiation (Zhang et Bennett, 2005).

CONCLUSIONS.

Les études qui ont été menées sur la diversité microbienne permettent d'avoir une bonne idée des phyla bactériens présents au niveau des sols sains et des sols pollués. Les *Proteobacteria* et les *Acidobacteria* mais aussi les *Actinobacteria*, les *Verrucomicrobia*, les *Bacteroidetes*, les *Chloroflexi*, les *Planctomycetes*, les *Gemmatimonadetes* et les *Firmicutes* semblent jouer un rôle important dans le fonctionnement des écosystèmes sols du fait de leur caractère ubiquiste. En ce qui concerne les sols pollués par des hydrocarbures, certains groupes bactériens appartenant aux γ -*Proteobacteria*, aux *Actinobacteria*, aux α -*Proteobacteria*, aux β -*Proteobacteria* et aux *Flavobacteria* semblent prévalents et ainsi impliqués dans la dégradation naturelle des polluants. Cependant, la caractérisation des communautés bactériennes à un niveau phylogénétique plus fin est encore insuffisante comme l'ont montré Schloss et Handelsman (2004). Néanmoins, afin de comprendre les processus d'adaptation des communautés microbiennes à leur environnement ainsi qu'à l'apparition d'une pollution, il est nécessaire d'obtenir ce niveau de résolution. Les outils biopuces permettront certainement, dans un avenir proche, d'obtenir les informations nécessaires à la résolution de ces problèmes.

PARTIE II

NOUVEAUX OUTILS D'ISOLEMENT ET D'IDENTIFICATION DES MICRO- ORGANISMES

A. CONTEXTE.

Comme nous l'avons indiqué, Whitman *et al.* (1998) ont estimé que l'ensemble des procaryotes pourraient représenter jusqu'à 6.10^{30} cellules au niveau des écosystèmes terrestres et aquatiques faisant d'eux la plus importante biomasse portée par la Terre. En outre, ces microorganismes présentent une grande diversité (Curtis et Sloan, 2004) et une flexibilité métabolique et morphologique leur permettant de s'adapter à de nombreuses niches écologiques et aux variations environnementales naturelles ou accidentelles. Ces caractéristiques font qu'il est difficile de les étudier dans leurs milieux naturels. Cependant, leurs rôles dans le fonctionnement des écosystèmes et leur intérêt grandissant en biotechnologie nécessitent une meilleure connaissance de ces organismes.

Jusqu'à récemment, la mise en culture était une étape obligatoire dans l'identification et dans l'étude des microorganismes. Or, des travaux récents ont montré que seule une minorité de bactéries pouvait être isolée par les techniques de culture utilisées classiquement c'est-à-dire par les techniques d'enrichissement (Torsvik *et al.*, 2002). Afin d'éliminer ce biais, il est alors apparu nécessaire aux microbiologistes de développer de nouvelles technologies leur permettant d'accéder à la majorité non cultivée des communautés bactériennes. Deux voies principales ont été empruntées afin d'atteindre cet objectif. La première a porté sur l'amélioration des techniques culturales avec le développement de nouveaux milieux de cultures et outils cultureux (Ferrari *et al.*, 2005 ; Joseph *et al.*, 2003). Cependant, même si ces nouvelles techniques représentent un réel potentiel et une avancée considérable, elles s'avèrent encore inadaptées pour appréhender la totalité de la biodiversité des écosystèmes. La seconde a porté sur l'utilisation des approches moléculaires basées notamment sur l'analyse des séquences nucléiques des microorganismes. De plus, avec le développement de techniques moléculaires dites à haut-débit, la microbiologie exploratoire a pris un nouvel essor.

B. OBJECTIFS.

Afin d'avoir une vue générale des difficultés d'étude des microorganismes dans des environnements complexes et des outils développés pour les isoler, les identifier et les caractériser, nous avons réalisé une synthèse qui est présentée dans cette partie sous forme d'un chapitre de livre. Le travail présenté par la suite a volontairement été séparé de la partie

synthèse bibliographique afin de pouvoir présenter plus en détails ces nouvelles techniques qui permettent d'accéder à la majorité des microorganismes, et par conséquent, d'ouvrir des perspectives stimulantes quant à l'amélioration de nos connaissances sur le monde microbien.

New tools for isolation and identification of micro-organisms

« **Advances in Fermentation Technology** »

Asiatech Publishers, Inc., New Delhi

Cécile Militon, Corinne Biderre and Pierre Peyret

Génomique Intégrée des Interactions Microbiennes , Laboratoire de Biologie des Protistes,
UMR CNRS 6023, Blaise Pascal University, 24 avenue des Landais, Campus des Cézeaux,
63177 Aubière Cedex , France

pierre.peyret@univ-bpclermont.fr

Abstract

The microbial world contains a highly heterogeneous group of organisms sharing only one common characteristic, their small sizes. These organisms make up two (out of three) entire Domains of life on earth, the prokaryotic Bacteria and Archaea (Woese, 1987). Within the third Domain Eukarya, the majority of the phylogenetic diversity is contained within eukaryotic micro-organisms such as protozoa, algae, and fungi. Micro-organisms are estimated to make up more than a third of Earth's biomass (Whitman, et al., 1998).

Because culturing provided the platform for building the depth and detail of modern microbiological knowledge, for a long time microbiologists ignored the challenge to identify and characterize uncultured organisms. One of the indicators that cultured micro-organisms did not represent much of the microbial world was the oft-observed “great plate count anomaly” (Staley and Konopka, 1985). The discrepancy between the sizes of populations estimated by dilution plating and by microscopy is particularly dramatic in some aquatic environments, in which plate counts and viable cells estimated by acridine orange staining can differ by four to six orders of magnitude, and in soil, in which 0.1 to 1% of bacteria are readily culturable on common media under standard conditions (Torsvik and Ovreas, 2002). Understanding the extent and importance of the genetic and biochemical diversity among strains of the same or very closely related species is a cornerstone issue for many microbiological disciplines such as biotechnology, taxonomy, diagnosis, diversity surveys, etc. We are only at the beginning of the history of the discovery of the amazing microbial diversity.

- 1) Microbial Diversity and evolution
 - 1.1) Phenotypic and genetic microbial classification
 - 1.2) Microbial diversity in complex environments
 - 1.3) Origin of the eukaryotic cell
- 2) Isolation and microbial cultivation methods
 - 2.1) Standard methods for micro-organisms cultivation
 - 2.2) New cultural strategies for micro-organism isolation
 - 2.3) High-throughput micro-organism cultivation
 - 2.4) Single cell isolation
- 3) PCR-based methods for microbial identification
 - 3.1) Nucleic acids extraction
 - 3.2) Extra cellular DNA and dead cells
 - 3.3) PCR artefacts
 - 3.4) rDNA markers and other biomarkers
 - 3.5) Molecular fingerprinting methods
 - 3.5.1) Amplified ribosomal DNA restriction analysis (ARDRA)
 - 3.5.2) Terminal-restriction fragment length polymorphism (T-RFLP) and Length heterogeneity polymerase chain reaction (LH-PCR)
 - 3.5.3) Denaturing gradient gel electrophoresis (DGGE) and Temperature gradient gel electrophoresis (TGGE)
 - 3.5.4) Ribosomal intergenic spacer analysis (RISA)
 - 3.5.5) Single-strand conformation polymorphism (SSCP)
 - 3.5.6) Randomly amplified polymorphic DNA (RAPD), Variable number of tandem repeats (VNTR), and Pulsed field gel electrophoresis (PFGE)
- 4) Hybridization methods and Isotope probing
 - 4.1) DNA reassociation
 - 4.2) Fluorescence in situ hybridization (FISH)
 - 4.3) DNA microarrays
 - 4.4) Stable isotope probing (SIP)
- 5) Metagenomics
- 6) New high-throughput sequencing developments

1) Microbial diversity and evolution

The overwhelming majority of life on our planet is microbial, both in terms of phylogenetic diversity and sheer numbers of organisms. Virtually every conceivable environmental niche harbours micro-organisms capable of growing there. Urban aerosols from San Antonio and Austin U.S. cities contained at least 1,800 diverse bacterial types, a richness approaching that of some soil bacterial communities (Brodie, et al., 2006). The hot, reducing, gaseous water emanating from a fracture at 2.8 to 4.2 kms harboured a microbial community dominated by a single *Firmicutes* phylotype (Lin, et al., 2006). It is timely to reconsider the extent to which it is possible to provide closer marriage between an operational species “definition” and a theory-based species “concept” for those taxa that are currently poorly characterized or yet to be discovered (Gevers, et al., 2005). Species demarcation in prokaryotes is not defined by a theory-based concept and tends to be more arbitrary or rooted in practical necessity. In the case of prokaryotic pathogens, species are historically defined on the basis of the disease they cause, regardless of other ecological or evolutionary considerations. A species definition is a set of easily applied and stable rules by which to decide when two organisms are similar enough in their genomic and/or phenotypic properties to be given the same name.

1.1 Phenotypic and genetic microbial classification

The first credible approaches to the systematic classification of bacteria began in the latter part of the 19th century. These early studies separated groups of bacteria primarily on the basis of morphology, size, and motility. A pioneer investigator during this period was Ferdinand Cohn, who supported the concept of a diversity of micro-organisms and argued that, within species, varieties emerged and transmitted their characteristics to the next generation (Rossello-Mora and Amann, 2001). The subsequent development of agar-based media led to the *in vitro* isolation and propagation of pure cultures. However, because of the little number of biochemical and phenotypic tests available, characterization of proposed species was inadequate and imprecise. This problem often led to confusion and resulted in the discovery and rediscovery of the same bacterial species by different investigators who gave the same taxa new names based upon slightly different morphological, cultural, and phenotypic criteria. The Gram-negative bacterium *Klebsiella pneumoniae* provides an

example of this taxonomic quagmire. This member of the family *Enterobacteriaceae* was listed under six different genera, with seven different species epithets, between 1885 and 1928.

A major innovation in laboratory sciences occurred in the late 1960s, when manual miniaturized identification systems were first introduced into the clinical microbiology laboratory. These early systems included the nine-test Enterotube (Roche, Basel, Switzerland) and two-tube R-B systems for identifying members of *Enterobacteriaceae*. These kits were rapidly followed by the landmark API20E strip test (bioMérieux, Hazelwood, Mo.), which was a micromethod employing 20 different biochemical tests. The 20E strip generated a septyl (7-digit) code to 18 to 21h based upon biochemical reactions plus a screening oxidase test. The numeric code could then be located in a logbook that converted septyl codes into a final bacterial identification. The 1980s saw continued improvements in bacteriology with the introduction of automation into the laboratory. In 1983, MicroScan (Dade Behring MicroScan, West Sacramento, Calif.) released the AS-3/touchScan system, the first widely accepted automated system for microbial identification and drug susceptibility testing in the clinical laboratory. Other systems, such as the Vitek AutoMicrobic system, quickly followed (Stager and Davis, 1992). These automated systems transformed processes for bacterial identification by reducing the time required to identify rapidly growing bacteria to as little as 2 to 4h. However, phenotypic systems have several drawbacks. Biochemical properties do not accurately reflect the entire extent of the genomic complexity of a given species. Furthermore, phenotypic properties can be unstable at times and expression can be dependent upon changes in environmental conditions, e.g., growth substrate, temperature, and pH. A further problem with commercial systems is the construction of databases, not to mention which biochemical tests are included in panels. Molecular diagnoses were developed to obtain a more accurate identification.

The first major leap in molecular taxonomy and identification applicable to diagnostic microbiology occurred with the introduction of DNA-DNA hybridization studies pioneered by Don Brenner and his colleagues at the Walter Reed Army Institute of Research and later at the Centers for Disease Control and Prevention (Brenner, et al., 1977). The value of DNA hybridization was that it provided a quantitative definition of what is a species, ~70% or greater DNA-DNA relatedness with a ΔT_m of 5°C or lower. The downside of DNA hybridization is that it is an expensive, technically complex, and labour-intensive procedure. The world of molecular taxonomy was revolutionized in the mid-1980s with the advent of full sequence analysis of molecular chronometers such as rRNA. By the mid-1990s, sequencing of

the small subunit rDNA genes had become commonplace, considered a standard tool of microbial taxonomists not only for elucidating phylogenetic relatedness but also as a means of bacterial identification. Since ribosomal genes are highly conserved, sequence variation between strains belonging to different species from distinct genera is less evident with rDNA gene sequencing than with DNA pairing studies, where DNA relatedness values are used to compare strains. 16S rRNA subunit reflects a small portion of the genome complexity. It is generally accepted that an unidentified isolate whose rDNA gene sequence is <97% similar to those of the isolate's closest phylogenetic neighbours constitutes a new taxon.

Estimates of diversity frequently adopt a working definition of a species as a group that shares >97% 16S rRNA gene sequence identity and >70% DNA-DNA hybridizations (Stackebrandt and Fiedler, 1979). However, it is becoming clear that this measure greatly underestimates species number (DeLong and Pace, 2001). Recent studies defined a species match as strain exhibiting $\geq 99\%$ or 99.5% sequence relatedness to strains previously deposited in database. However, these arbitrary values may not always accurately reflect taxonomic relatedness (Janda and Abbott, 2002). Some *Edwardsiella* species revealed only 0.19% divergence for 16S rDNA sequence. Species are recognized as different species not only on the basis of DNA hybridization results but also because each group occupies a distinct ecologic niche.

Ideally, identification of any taxon is based upon a polyphasic approach that includes a combination of phenotypic testing methods (e.g. biochemical testing, cellular fatty acid analysis, and numerical analysis) and genotyping testing methods (e.g. DNA-DNA hybridizations, analysis of G+C content, Pulsed Field Gel Electrophoresis, and 16S rDNA gene sequencing). However, such methods are time consuming, expensive, and not easily adaptable to workflow in microbial laboratories. Janda and Abotte proposed in 2002 guidelines for identification of bacterial species for publication purposes that should be both technically and financially feasible and would help to reduce the number of publication with misidentifications.

The multilocus sequence typing (MLST) method has recently emerged as a method of choice for exploring and cataloguing intraspecies genetic diversity (Gevers, et al., 2005; Maiden, et al., 1998), thus setting the stage for linking the genetic diversity to the biochemical and functional diversities species. It has various other synonyms such as multiple gene genealogical analysis (MGGA) or comparative genealogical analysis (CGA). Typical applications of the MLST method employ the sequencing of six to eight genes (loci) and subsequent phylogenetic analysis of the concatenate sequence alignments to reveal exact

genetic relationships among the strains analyzed. Phylogenetic sequence analysis of multiple (six to eight) genes in the genome represents currently the most favourable approach for studying species diversity. Therefore, whereas the rRNA gene sequence might assign an isolate to a genus, MLST might assist in grouping isolates into the major genetic lineages within a species. The more general term multilocus sequence analysis (MLSA) could be used to be applied to the problem of species “definition”. This would invalidate the use of simple clustering procedures based on the number of allelic mismatches, and therefore phylogenetic procedures based on the nucleotide sequences of the alleles would be employed instead. Rapid and robust classification with MLSA could use a set of genes which would allow for the hierarchical classification of all prokaryotes. However, this approach might be impractical, as genes that are informative within a genus or family might not be useful or even present in more distantly related taxa. Also, genes that are conserved enough to be amplified by a set of primers common to all species might not evolve quickly enough to distinguish closely related taxa. MLSA should use genes that are ubiquitous (at least in the taxon under study) and in single copy, and should avoid those genes in which recombination might confer a selective advantage (virulence and antigen-encoding genes) or closely related genes. The simplest approach is to concatenate the sequences of the multiple genes and to use the concatenated sequences to construct a tree, which can identify deeply branching clusters that can help the divisions of the genus into species. Prokaryotic identification in this scenario is a two-step process: rRNA sequencing assigns an unknown strain to a group (genus or family), which defines which genes and primers are to be used for MLSA to assign a strain to a species. Ecotypes are defined as populations that are genetically cohesive and ecologically distinct. Defining species as ecotypes (ecovar) might remain reserved for those groups whose ecological distinctness is obvious and important (for example, pathogens), as well as species yet to be discovered. MLSA could not be applicable in some cases and several issues remain to be resolved. It might not be possible to delineate groups within a continuous spectrum of genotyping variation. Important phenotypes might not be conferred by stable chromosomal loci, resulting in marked differences among strains that are closely related. Recently, MLST has been applied to classify the Gram negative *Burkholderia* which comprise over 30 species (Mahenthiralingam, et al., 2006). The MLST strain identification scheme examines nucleotide polymorphisms in seven genomically disparate housekeeping genes (ATP synthase β chain, *atpD*; glutamate synthase large subunit, *gltB*; DNA gyrase B, *gyrB*; Recombinase A, *recA*; GTP binding protein, *lepA*; Acetoacetyl-CoA reductase, *phaC* and Tryptophan synthase, *trpB*). To date, 33 species-specific MLST databases are available online

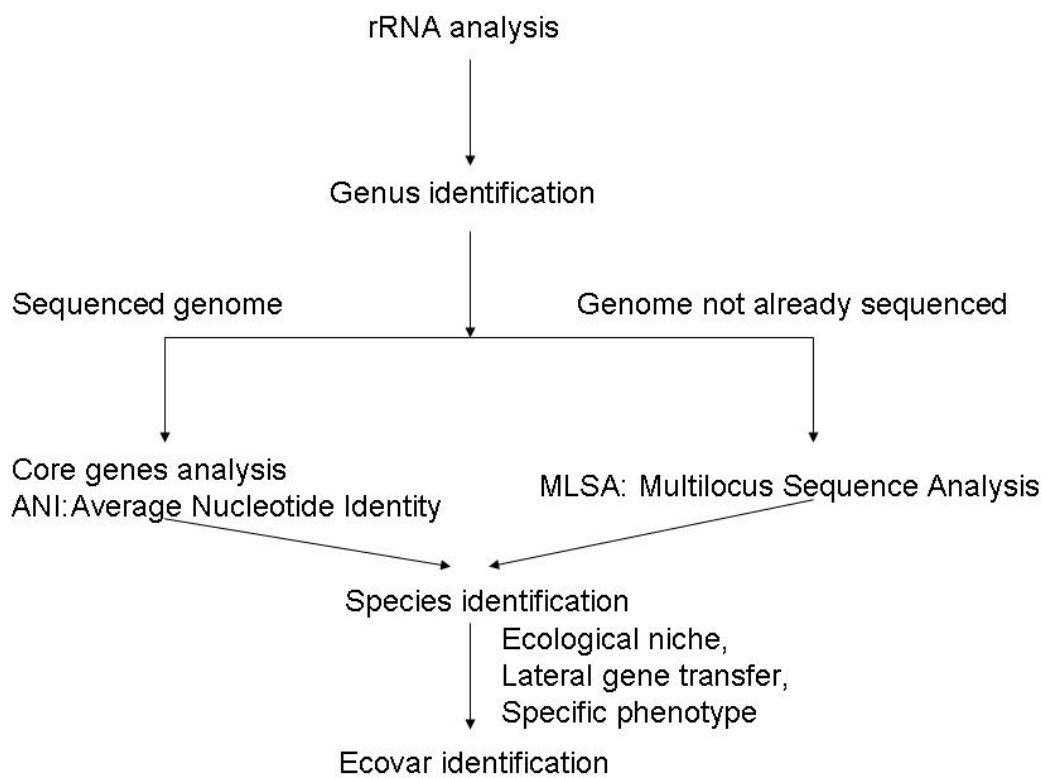


Figure 1: Description of a possible scheme for microbial identification.

(<http://www.mlst.net>). Recent report describes the development of an MLST pipeline that automates DNA sequence editing and analysis (Platt, et al., 2006).

The average nucleotide identity (ANI) of the shared genes between two strains were found to be a robust means to compare genetic relatedness among strains, and that ANI values ~94% corresponded to the traditional 70% DNA-DNA reassociation standard of the current species definition (Konstantinidis and Tiedje, 2005). Notably, a large fraction, e.g., up to 65%, of the differences in gene content within species is associated with bacteriophage and transposase elements, revealing an important role of these elements during bacterial speciation. Phylogenies of closely related organisms based on the average nucleotide identity (ANI) of their shared genes correspond accurately to phylogenies based of state-of-the-art analysis of their whole genome sequences (Konstantinidis, et al., 2006). The authors shown that almost all core genes, regardless of their functions and position in the genome, offer robust phylogenetic reconstruction among strains that show 80 to 95% ANI (16S rRNA identity, >98,5%). This precise measure is limited in its application to sequenced genomes. Lateral gene transfer (LGT), within-species genomic variability and homologous recombination all make it harder to imagine the evolution of species. Genomics tells us that bacteria often acquire evolutionary novelties from outside the ancestral population by LGT (Gogarten and Townsend, 2005; Lerat, et al., 2005). Many genes conferring a highly complex adaptation can be acquired in one event, instantly dividing a single population into two subpopulations that differ substantially in lifestyle but continue to share in a common gene pool. LGT radically uncouples the evolution of phenotype from the evolution of the bulk of the genome. For instance, *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis* show more than 94% ANI and are highly syntenic in chromosome structure. In fact, they are famously different in phenotype. Figure 1 represents the different levels of microbial identification.

1.2 Microbial diversity in complex environments

Estimates of the order of magnitude for the total number of microbial species on Earth range from 10^3 to 10^9 . Despite global dispersal of micro-organisms, this number is probably rather large (Pedros-Alio, 2006). Because micro-organisms are small and have many individuals, they are easily dispersed globally. This is an explanation for the classical dictum that “everything is everywhere, the environment selects” (Quispel, 1998). But the conflicting debate with the other hypothesis suggesting a biogeography distribution of micro-organisms

is not resolved (Horner-Devine, et al., 2004). The unknown diversity hides novel metabolisms that are the largest potential reservoir of useful genes for medicine and biotechnology.

Culture based methods suggest that a gram of soil contains fewer than 100 species, but these are undoubtedly underestimates because multiple lines of evidence indicate that fewer than 1% of the species in soil are presently culturable (Dunbar, et al., 1999). DNA reassociation experiments produce richness estimates ranging from 4,000 to 100,000,000 genome equivalents per 10 or 30 g of soil (Gans, et al., 2005). An alternative method relies on analysis of 16S rRNA gene sequences amplified from soil by PCR. Estimates of richness have been obtained through parametric or non parametric empirical models of species frequency distribution to produce richness estimates between 590 and 100,000 species per gram of soil (Lunn, et al., 2004). The simulated communities constructed based on the distribution of 16S rRNA gene sequences sampled from Alaskan and Minnesotan soils had a richness of 5,000 and 2,000 operational taxonomic units (OTUs), respectively, where an OTU represents a collection of sequences not more than 3% distant from each other (Schloss and Handelsman, 2006). To sample each of these OTUs in the Alaskan 16 rRNA gene library at least twice, 480,000 sequences would be required; however, to estimate the richness of the simulated communities using nonparametric richness estimators would require only 18,000 sequences. Among the members of a single OTU, there is undoubtedly considerable phenotypic and genomic diversity that is not reflected by 16S rRNA sequences.

Extensive shotgun sequencing of the Sargasso sea produced ~1.36 Gb of sequence data, estimated to derive from >1,000 genomic species (Venter, et al., 2004). Over 1.2 million gene products were identified and provided a basis for the estimation of diversity.

More recently, a set of 31 protein-coding marker genes, extracted from large-scale environmental shotgun sequencing data provided a more direct, quantitative and accurate picture of community composition (von Mering, et al., 2007). The authors find *Proteobacteria* to be the most dominant phylum of microbial life in both marine and soil environments. However, marked differences become apparent, for example relatives of the *Rickettsiales* (including the marine genus *Pelagibacter*) are mostly found in the surface water samples, whereas relatives of *Rhizobiales* or *Burkholderiales* are mostly found in the soil sample. Surprisingly, few endospore-forming organisms (*Bacilli* and *Clostridia*) were observed in the sequences community. Similarly, *Actinobacteria* (many of which have a spore stage) represent in soil sample a fraction of only 6.2%. Most microbial lineages remain associated with a certain environment for extended time periods, and successful competition in a new environment seems to be rare event. Existing communities would thus constantly be

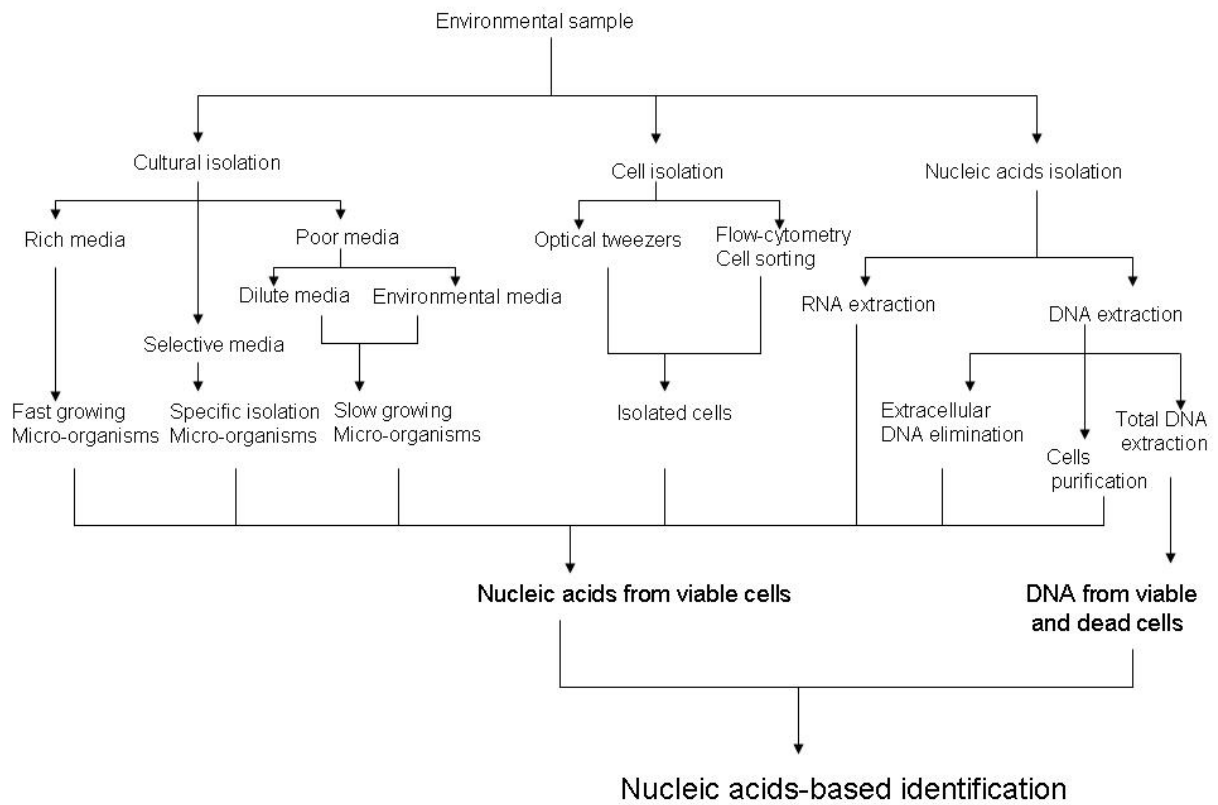


Figure 2: Schematic representation of the different ways used for microbial identifications for an environmental sample.

challenged by intruders from non-specialist phyla who may occasionally survive simply by chance, acquiring the necessary functionality through horizontal gene transfer. Figure 2 describes the different ways to characterize micro-organisms in an environmental sample.

An outstanding challenge in microbial ecology is to estimate species richness based on 16S rRNA sequences. Sequences with greater than 97% identity are typically assigned to the same species, those with >95% identities are typically assigned genus and those with >80% identities are assigned to the same phylum. Several methods recently developed address the challenge of assigning sequences to operational taxonomic units based on genetic distances between sequences. DOTUR (Distance-Based OTU and Richness; <http://www.plantpath.wisc.edu/fac/joh/dotur.html>) assigns sequences to OTUs by using either furthest, average, or nearest neighbour algorithm for each distance level (Schloss and Handelsman, 2005). DOTUR uses the frequency at which each OTU is observed to construct rarefaction and collector's curves for various measures of richness and diversity. Diversity indices and richness estimators are useful to compare the relative complexity of two or more communities and to estimate the completeness of sampling of a community. It is still impractical to exhaustively sample a whole microbial community. FastGroupII (<http://biome.sdsu.edu/fastgroup/>) is web-based platform for the dereplication of large 16S rDNA libraries and estimation of community composition and diversity (Yu, et al., 2006). One main feature of FastGroupII is the dereplication process, in which identical or nearly-identical sequences belonging to the same microbial ribotype are grouped together. The NAST (Nearest Alignment Space Termination; <http://greengenes.lbl.gov/NAST>) algorithm is used for creating multiple sequence alignment (MSA) with a web-interface (DeSantis, et al., 2006). Each user-submitted sequence is compared with Greengenes' "Core Set", comprising ~10 000 aligned non-chimeric sequences representative of the currently recognized diversity among bacteria and archaea. A current set of >80 000 16S rDNA genes gleaned from GenBank is maintained in aligned format on the Greengenes server (DeSantis, et al., 2006).

Comparisons between sequences from multiple environments are increasingly important because they can test whether microbial community composition changes in response to specific environmental variables. In this way, microbiologists could efficiently identify candidate micro-organisms involved in metabolic processes of interest. UniFrac (Unique Fraction; <http://bmf.colorado.edu/unifrac/index.psp>) or SONS (Stands for Shared OTUs and Similarity; <http://www.plantpath.wisc.edu/fac/joh/sons.html>) provide easy access to powerful multivariate techniques for comparing microbial communities in a phylogenetic

context (Lozupone, et al., 2006; Lozupone and Knight, 2005; Schloss and Handelsman, 2006).

1.3) Origin of the eukaryotic cell

Microbial diversity is the result of a complex evolution. The origin of the eukaryotic cell is still enigmatic (Dagan and Martin, 2006). Five distinctly different views try to explicit organism evolution.

The classical ribosomal RNA (rRNA) tree of life suggests that the universal ancestor of all life (the progenote) was a communal collection of information-storing and information-processing entities that were not yet organized as cells. Lateral gene transfer is seen as the main mode of genetic novelty at the early stages of evolution. The emerging cellular lineages of prokaryotes and eukaryotes from supramolecular aggregates become refractory to LGT. The classical tree assumed its current shape when anciently diverged protein-coding genes suggested that the root of the universal tree lies on the bacterial branch (Gogarten, et al., 1989; Iwabe, et al., 1989).

In the eukaryotes-first (intron-early) theory, tree emerged when Ford Doolittle suggested that the ancestral state of genes might be “split”, and that some introns in eukaryotic genes might thus be carryovers from the assembly primordial protein-coding regions (Doolittle and Sapienza, 1980). There is no prokaryote-to-eukaryote transition in the intron-early tree, because prokaryotic genome organization is seen as a very early derivative of eukaryotic gene organization. For some researchers the hypothetical fusion of an archaeon and a bacterium explains nothing about the special features of the modern eukaryote cell (Kurland, et al., 2006). A primitive eukaryote raptor with a larger, more complex cell structure than that of present-day prokaryotes is envisioned as the host of the ancestral endosymbiont.

The neomuran theory accounts mainly for cell biological characters, but not for sequence similarity among genes. The neomuran tree suggests that common ancestor of all cells was a free-living eubacterium (in the most recent version of the theory, a Chlorobium-like anoxygenic photosynthesizer) and that eubacteria were the only organisms on Earth until about 900 million years ago (Cavalier-Smith, 2006). A member of the eubacteria lost its murein-containing cell wall and the neomuran lineage underwent diversification into two lineages. One lineage invented isoprene ether lipid synthesis and gave rise to archaeobacteria. The other became phototrophic and gave rise to the eukaryotes.

On the other hand, the symbiotic tree proposes that the ancestor of eukaryotes originated by endosymbiosis of one prokaryote to another prokaryote giving rise to nucleated eukaryotic cells but mitochondrion lacking cell (an archezoon) followed by the origin of mitochondria (Cavalier-Smith, 1987).

The prokaryote host-tree proposes that endosymbiotic event that gave rise to a cell containing nucleus and mitochondria occurred in a prokaryotic host. This leads to a ring-like relationship between the ancestral organisms rather than a tree. Informational genes (genes involved in transcription, translation, and related processes) are most closely related to archaeal genes, whereas operational genes (genes involved in cellular metabolic processes such as amino acid synthesis, cell envelope and lipid synthesis, and so on) are most closely related to eubacterial genes (Rivera, et al., 1998). Genome fusions and horizontal (or lateral) gene transfer (HGT or LGT) convert phylogenetic trees into rings. All current theories about the origin of eukaryotes see them as chimeras arising from the coming together of different prokaryotic cellular lineages, both archaeal and bacterial, *via* some intermediate symbiotic association. One fusion partner branches from deep within an ancient photosynthetic clade, and the other is related to the archaeal prokaryotes. The eubacterial organism is either a proteobacterium, or a member of a larger photosynthetic clade that includes Cyanobacteria and the Proteobacteria (Rivera and Lake, 2004).

The origin of the nucleus is still debated. The most widespread and familiar model for the origin of the nucleus is the concept that the endomembrane system of eukaryotes, to which the nucleus and ER belong, is derived from invagination of the plasma membrane of a prokaryote (Martin, 2005). Martin and Koonin (2006) proposed that the nuclear envelope's incipient function was to allow mRNA splicing (Martin and Koonin, 2006), which is slow, to go to completion (0.005-0.01 intron per second; (Audibert, et al., 2002) so that translation, which is fast (10 amino acids in prokaryotes and 1 amino acids in modern eukaryotes), would occur only on mRNA intact reading frames. The rapid, fortuitous spread of introns following the origin of mitochondria is adduced as the selective pressure that forged nucleus-cytosol compartmentalization. Yet, exon shuffling and alternative splicing remain as paradigms of eukaryotic gene evolution. The evolutionary relationships of proteins specific to the nuclear envelope and nuclear pore complex reveal that this sizable protein set is a mix of proteins and domains of archaeobacterial and eubacterial origins, along with some eukaryotic innovations, suggesting that the nucleus arose in a cell that already contained a mitochondrial endosymbiont. The result of the inference of the authors is a cell bearing a mitochondrion, chimaeric chromosomes contained within a membrane-bounded splicing compartment (the

nucleus), and archaeobacterial ribosomes in the cytosol that translate mature mRNAs from intron-containing and intronless of both host (Archaeobacterial) and symbiont (α -proteobacterial).

2) Isolation and microbial cultivation methods

The design of media formulations involving the combination of non-traditional growth substrate is therefore a major driving force in cultivating and isolating new microbes. Cultivation of micro-organisms has been viewed as fundamental for understanding of microbial physiology and morphology (Pedros-Alio, 2007). It is also well known that conventional cultivation approaches access only a tiny subset of the wide diversity of micro-organisms inferred to be present in any given environment. Traditional microbiological methods of cultivation recover less than 1% of the total bacterial species, and the culturable portion of bacteria is not representative of the total phylogenetic diversity. Considering the extent of functional diversity described for microbes and the numerous applications of their secondary metabolites, the biotechnological potential hidden among the 99% of the bacteria that are nonculturable is immense. In fact, many prokaryotic phyla recognized so far consist solely of 16S rRNA gene sequences from as yet uncultured microbes. Most of the cultivated micro-organisms are those that grow quickly and are capable to growth in nutrient-rich media (Leadbetter, 2003).

2.1) Standard methods for micro-organisms cultivation

Classical strategies have traditionally supplied excessive nutrients to a system, resulting in the enrichment of fast-growing bacteria that are capable of colony forming or biofilm formation. Culture media often supply excessive amounts of nutrients and in some cases may be toxic to target bacteria. Oligophilic “k-selected” bacteria have been described as bacteria that are adapted to growth in nutrient-poor environments, and they may represent a high proportion of the missing culturable diversity (Simu and Hagstrom, 2004). The inability to form colonies or biofilm appears to be a strategy to explore dissolved organic matter in seawater as single cells.

New approaches to the cultivation of bacteria which rely on growth in dilute nutrient media and simulated environments are beginning to address this problem of selection.

2.2) New cultural strategies for micro-organism isolation

Methods for cultivation that are based on modified traditional approaches have resulted in the isolation of previously uncultured, phylogenetically distinct micro-organisms. New approaches to the cultivation of bacteria that rely on growth in dilute nutrient media or simulated natural environments are beginning to address (Green and Keller, 2006; Tyson and Banfield, 2005). In many of these studies traditional signs of bacterial growth, such as colony formation and turbidity, are poor indicators of cultivation as the majority of growth is microcolony CFU (mCFU) growth, which require microscopic identification (Kaeberlein, et al., 2002; Simu and Hagstrom, 2004). The majority of growth of marine bacteria (>99%) under environmental condition was growth to only microcolony stage, while up to 80% of soil bacteria grew to the microcolony stage only over a 63-day incubation period on dilute soil medium.

Changing various culturing conditions (oxygen concentration, nutrient level, addition of humics, specific wavelengths of light and signalling molecules) allowed to increase the culturability of micro-organisms in soil (Stevenson, et al., 2004). The SAR11 represents a marine α -proteobacterial “ribotype” believed to account for a staggering 50% of all cells encountered in many open-ocean systems. The organisms encoding the “ribotype” had remained elusive to all attempts to cultivation for over a decade since their discovery (Rappe, et al., 2002). The key to cultivating SAR11 appears to be related to the use of dilute growth media (i.e. nothing more than sterilized seawater amended with low amounts of phosphate and ammonium), the ability to recognize very significant (2-3 log) increases in population sizes at “optical density” thresholds far below the typical detection limit, and the patience to engage a microbe with low yields and slow growth rates (i.e. 30-40h doubling times). Janssen and co-workers have been making outstanding progress in the cultivation of organisms representing rich, phylogenetic diversity (Davis, et al., 2005; Joseph, et al., 2003; Sait, et al., 2002). A patient approach was developed for examining soil organisms that arise as abundant, but slow-to-appear, colonies on gellan-solidified, dilute nutrient broth. The collection includes the first isolate of a major subdivision within the *Verrucomicrobia*, previously known only through cloned environmental sequences. In a similar study using a heteropolysaccharide-based xylan, defined medium instead of diluted nutrient broth, by waiting 12 weeks for colonies to develop, a collection of isolates was obtained including the first isolates representing subdivision 3 of the phylum *Acidobacteria*.

Ferrari et al. (2005) described a novel microcultivation for soil bacteria that mimics natural conditions (Ferrari, et al., 2005). The slurry membrane system employed by this group combines the use of a polycarbonate membrane as a growth support and soil extract as the substrate. The result is abundant growth of uncharacterized bacteria as microcolonies. The majority of these microcolonies consisted of large numbers of cells that formed a single layer, in many cases with defined distances between individual cells, like a dispersal strategy for growth. In some instances more than one cell type was present in the same microcolony, suggesting that cells were interacting. By combining low nutrients enrichments and molecular methods, a high diversity of new amylase genes was detected in a neutral sulfid-rich hot spring in Iceland.

The selective fractionation of specific micro-organisms within a sample before cultivation can reduce the complexity of the microbial community. Miteva and Brenchley (2005) described a filtration-cultivation procedure combined with long incubations that enriched for novel ultra-small cell isolates (Miteva and Brenchley, 2005). The abundant microbial population in a 3,043-m-deep Greenland glacier ice core was dominated by ultra-small cells ($<0.1 \mu\text{m}^3$) that may represent intrinsically small organisms or starved, minute forms of normal-sized microbes. Initial studies found that more than 70% of the soil micro-organisms, named “dwarf”, had diameters of less than $0.3\mu\text{m}$, and only 0.2% of these were culturable. These small cells have been reported using different names, including the term “ultramicrobacteria”. Small microbial cell is considered to be advantageous for more efficient nutrient uptake in oligotrophic conditions due to a larger surface-to-volume ratio, protection against predators, and occupation of microenvironments.

2.3) High-throughput micro-organism cultivation

One of the most promising advances is the development of high-throughput methods to grow encapsulated single cells under environmentally relevant conditions. This method has been used for cultivation of marine micro-organisms, but could be applied directly to any environment (Zengler, et al., 2002; Zengler, et al., 2005). The approach involves the encapsulation of single cells (i.e. from a mixed microbial inoculum), within a “gel microdroplet”, after which the hardened beads are packed into a column through which dilute media is continuously flowed for long periods. Following long incubation periods of constant flow of low nutrients, flow cytometry was used to sort gel beads containing cells that had grown into small micro-colonies from those that not. Using this technique, increases in cell

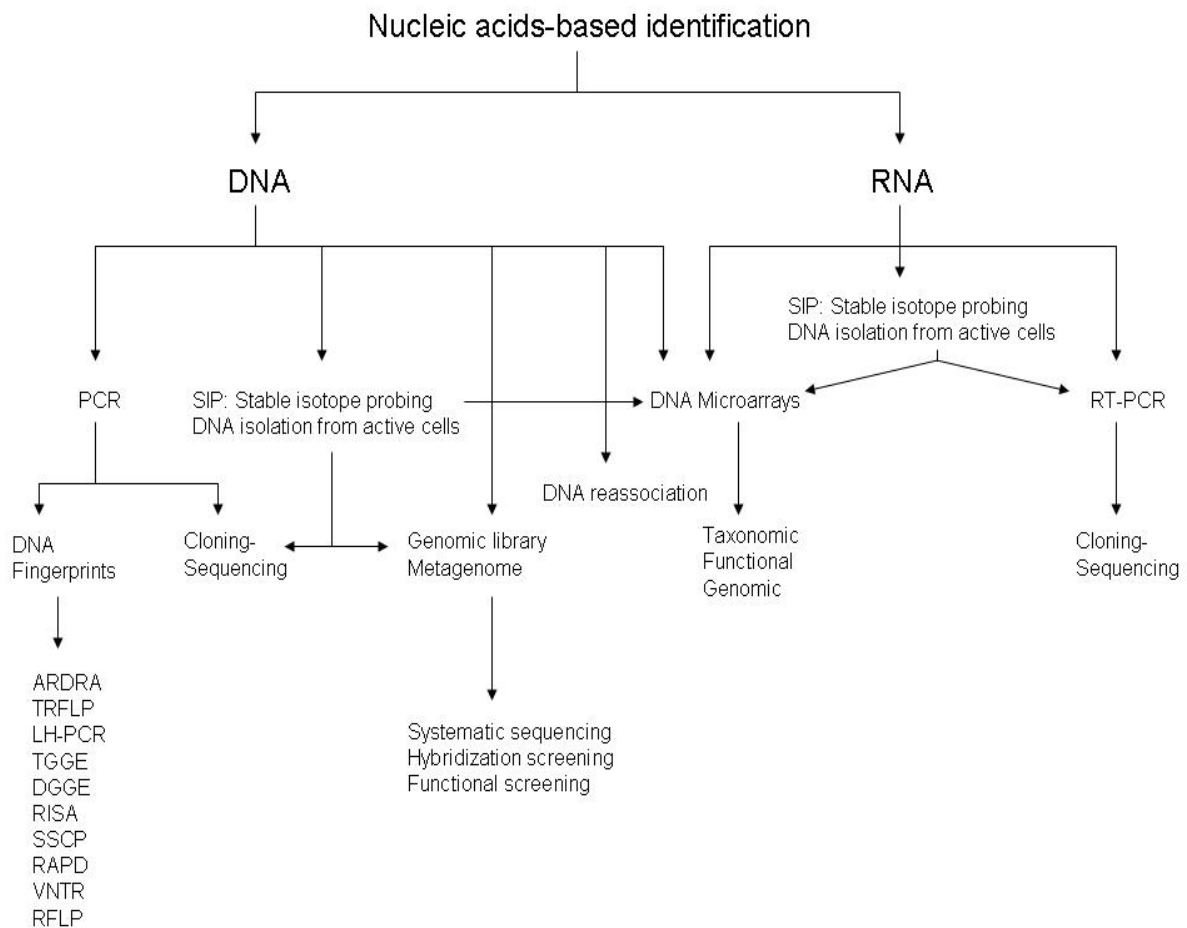


Figure 3: Schematic representation of the different strategies used for microbial identification by nucleic acids-based identification. For abbreviations and details of the techniques referred to the text.

number corresponding to as few as five to seven doubling events could be rapidly monitored in 1200 separate beads. Deep lineages of marine *Planctomyces* and *Bacteroidetes* were recovered, along with several other isolates scattered throughout the bacterial tree of life.

2.4) Single cell isolation

Single cells manipulation techniques such as optical tweezers and laser microdissection, also hold promise for targeted isolation of micro-organisms (Frohlich and Konig, 2000). Several targeted isolation approaches have used 16S rRNA-directed probes to track the progress of isolation attempts in serial dilutions without previous knowledge of the physiology of the organisms (Rappe, et al., 2002). Combining oligonucleotide probes with optical tweezers lead to the isolation of a novel thermophilic archaeon (Huber, et al., 1995).

Despite these advances, many cultivation attempts fail. We simply do not know how to provide the conditions necessary to enable growth in the laboratory. Micro-organisms might depend upon a specific combination of environmental conditions, require unsuspected growth factors, or be sensitive to unidentified compounds. In the environment, organisms grow as part of a community and it is likely that synergistic relationships develop where certain organisms carry out the necessary biosynthetic tasks and remove or change the chemical forms of toxic compounds. Figure 3 illustrates the following parts by describing the nucleic acids-based methods for microbial identification.

3) PCR-based methods for microbial identification

Nucleic-acid based methodology currently employed to characterize microbial species or microbial consortia is based largely, but not exclusively, on taxonomic markers such as genes encoding the small subunit of the ribosomal RNA.

3.1) Nucleic acids extraction

Assessing environmental microbial community structure by the use of molecular techniques requires a satisfactory sampling strategy that takes into account the high microbial diversity and the heterogeneous distribution of micro-organisms. The sampling procedure has

no influence on the assessment of the bacterial community structure. On the contrary the sampling aliquots of soil ≥ 1 g are required to obtain robust and reproducible fingerprinting analysis of the genetic structure of communities (Ranjard, et al., 2003). However, the smallest samples could be adequate for the detection of minor populations masked by dominant ones in larger samples.

The extraction of high quality DNA is central to the success of any sequencing project or molecular approach using DNA. In the case of environmental samples that contain mixed consortia of organism types, the objective is to obtain a quantitatively accurate representation of all community members during extraction and subsequent analysis. Realizing this importance, microbial ecologists have invested substantial effort in optimizing DNA and RNA extraction procedures for various environmental samples (Hurt, et al., 2001; Luna, et al., 2006; Yang, et al., 2006; Zhou, et al., 1996). Lysis of the microbial cells during DNA extraction represents a critical step in PCR-mediated approaches (von Wintzingerode, et al., 1997). Rigorous conditions needed to lyse Gram-positive cells may cause excessive shearing of nucleic acids of Gram-negative cells, potentially biasing the reported diversity as well as possibly creating artefacts and chimeric PCR products. Co-extracted components of soil, for example humic and fulvic acids, inhibit *Taq* polymerase and other modifying enzymes and could interfere with the analysis of the free DNA. Diluting out the sample to similarly dilute these undesired contaminants seems a viable option. To minimize loss of nucleic acids from small sample volumes, additives, such as bovine serum albumine (BSA) and T4 gene 32 (gp32), are used to reduce the inhibitory effect of contaminants (Kreader, 1996; McKeown, 1994). The use of affinity-binding column purification like polyvinylpyrrolidone (PVPP) that selectively binds humic acids is an alternative to eliminate inhibitory contaminants from DNA (Berthelet, et al., 1996).

Procedures extensively utilized for extraction of microbial cells could be employed to purify intracellular DNA (Holben, et al., 1988). Other methods like nuclease-based procedure eliminate extracellular DNA. Simultaneous recovery of extracellular and intracellular DNA suitable for molecular studies has been recently developed (Corinaldesi, et al., 2005). Assessment of the metabolically active fraction of the community should be done by analysis of RNA rather than DNA. However, extraction of RNA from environmental ecosystems requires special attention as RNAs are highly susceptible to degradation by RNases during the extraction procedures.

3.2) Extracellular DNA and dead cells

The differentiation between live and dead cells presents an important challenge in many microbiological applications. Due to the persistence of DNA in the environment after cells have lost viability, DNA-based methods cannot differentiate whether positive signals originate from live or dead bacterial targets. All aquatic sediments are characterized by high DNA concentrations (concentrations that are 3 to 4 order of magnitude greater than those found in the water column), mostly (up to 90%) due to extracellular DNA. Extracellular DNA may remain preserved in sedimentary deposits over geological time scales from 1,000 to more than 100,000 years (Willerslev, et al., 2004). Complex refractory organic molecules and/or inorganic particles are able to bind, adsorb, and stabilize free DNA in sediments. The free extracellular DNA fraction represented less than 5% of the total extracellular DNA pool. The adsorption of extracellular might reduce its degradability, and indeed only about 50% of this DNA can be hydrolyzed by nucleases. Extracellular DNA might have important implications for bacterial metabolism, providing a source of nitrogen and phosphorous and/or exogenous nucleotides, and may also contribute to horizontal gene transfer through natural transformation. The most important criterion for distinguishing between viable and irreversibly damaged cells is membrane integrity. Live cells with intact membranes are distinguished by their ability to exclude DNA-binding dyes that easily penetrate dead or membrane compromised cells. The most common membrane-impermeant dye is propidium iodide. In the last few years, EMA-PCR (Ethidium MonoAzide-Polymerase Chain Reaction) combines the use of a live-dead discriminating dye with the speed and sensitivity of real time PCR. Ethidium monoazide is a DNA-intercalating dye with the azide group allowing covalent binding of the chemical to DNA upon exposure to bright visible light. Photo-induced cross-linking was reported to inhibit PCR amplification of DNA from dead cells. Recently, it was shown that EMA cross-linking to DNA actually rendered the DNA insoluble and to its loss together with cells debris during genomic DNA extraction (Nocker and Camper, 2006). Though this technique is promising, the use of EMA prior DNA extraction was found to suffer a major drawback. Ethidium monoazide readily penetrates viable cells of some species resulting in partial DNA loss. The lack of selectivity and of overall applicability lead to developed alternative chemical: propidium monoazide (PMA). A higher impermeability was observed allowing an efficient removal of genomic DNA from cells with compromised cell membranes (Nocker, et al., 2006).

3.3) PCR artefacts

The genetic differences and diversity must be taken into account for the microbial community analysis. Both genome size and copy number differ from one organism to the other and could alter the number of 16S genes present in a single organism. In this situation PCR amplifications of 16S rRNA genes from a complex microbial community may not give an accurate quantification of the consortium (Farrelly, et al., 1995). After the proper primers are chosen, consideration must be given to the possible nucleotide composition of an unknown consortium. The varying %G+C content of the template DNA is suspected of causing differential amplification during the denaturation step of the PCR reaction. Sequences can also be less available for primers annealing due to selective priming or the formation of higher-order structures such as hairpin loops. Another pitfall of PCR is the production of minute errors by some *Taq* polymerase, which lacks the ability to proofread. Other limitation of the PCR technology is contaminating DNA introduced by unintentional tube-to-tube contamination or contaminated reagents. Each physical, chemical, and biological step involved in molecular analysis is a source of bias (von Wintzingerode, et al., 1997).

3.4) rDNA markers and other biomarkers

The key molecular player in microbial classification has been the RNA component of the small subunit of ribosomes (SSU rRNA, or 16S/18S rRNA), which Carl Woese picked in the early 1970s as a convenient and reliable « universal molecular chronometer ». In 1977, Carl Woese and George Fox announced that the prokaryotes could be divided in two distinct groups on the basis of the SSU rRNA sequence, the “eubacteria” and the “archeobacteria” (Woese and Fox, 1977). Later in 1980s some other differences have been revealed to differentiate these two groups. For example, Archaea RNA polymerase is closer to their eukaryotic counterparts than to Bacteria. According to a recent survey, 33 ribosomal proteins are shared uniquely among Archaea and eukaryotes (Lecompte, et al., 2002). In contrast, no ribosomal proteins are shared between Bacteria and eukaryotes that are not also found in Archaea. The Bacterial/Archaeal dichotomy has been widely accepted. But recent discoveries highlight that the physiological differences between Bacterial/Archaeal and eukaryotes are not so strict. *Planctomycetes* have (like Archaea) cell wall lacking peptidoglycan and (like eukaryotes) membrane-bounded nucleoids. Members of *Verruimicrobia* contain genes for the eukaryotic cytoskeletal protein tubulin and homologs of a few additional genes otherwise

considered restricted to eukaryotes. In Archaea, the surprise was the discovery of a new phylum, so far, represented by a single species, *Nanoarchaeum equitans* a parasite of the hyperthermophilic crenarchaeote *Ignococcus*. Archaea show so far only two or three groups: the Euryarchaeota, the Crenarchaeota and (possibly) the Korarchaeota. Among Bacteria, there are at least 31 phyla. Comparison of five different taxonomies uncovered surprisingly great disparity between expert curators (<http://greengenes.lbl.gov/TaxCompare>; RDP, Ludwig, NCBI, Hugenholtz and Pace show 31, 50, 68, 88 and, 88 phyla respectively). The main source of the disparity is the discordant naming of novel candidate phyla or the absence of names for candidate phyla. SSU rRNA sequencing provides the advantage of quick identification and the ability to define within-phylum phylogenetic relationships down to the level of “species” in a uniform way. Culture-independent approaches have revolutionized microbial ecology. Half the bacterial phyla are known only by cloning-sequencing.

Since protein synthesis is a conserved function in all organisms, the ribosomal RNA sequence possesses nucleotide-sequence motifs that are conserved, and other parts of the molecule can vary considerably, which can be used in the phylogenetic classification of micro-organisms. PCR is the simplest and currently widely used method to obtain 16S rRNA genes. The key to PCR is to use oligonucleotide primers designed to be complementary to the desired gene or genetic region. Given the varying degrees of conservation of the 16S gene, PCR primers can be designed to target the DNA of specific organisms or groups of organisms with any desired degree of specificity. rRNAs are found in ribosomes of prokaryotes and eukaryotes: the 16S, 23S and 5S rRNAs in prokaryotes and 18S, 28S, 5S and 5.8S analogues in eukaryotes. Individual 5.8S rRNA does not exist in prokaryotes and some eukaryotes like Microsporidia (Peyretailade et al., 2001). The terms Small-subunits (SSU) and Large-subunits (LSU) are used to refer to the 16S (18S) and 23S (28S) rRNAs, respectively. The 16S rRNA possesses approximately 1500 nucleotides, large enough to be useful for phylogenetic inference. The 120-nucleotide 5S rRNA is quite small and its low number of independently varying nucleotide positions limits its usefulness for phylogenetic studies. Though the 23S contains about twice as much information as the 16S, it is the SSU that has become the established phylogenetic reference because of its ease to clone and to sequence.

The Ribosomal Database Project II (<http://rdp.cme.msu.edu/>) contains 286,257 16S rRNAs in the release 9.45 of December 2006 and the Greengenes database (<http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>) in October 2006 archives 107,284 aligned 16S rDNA records >1250nt. These databases growth exponentially due to the facility of sequencing (in June 2005 RDP contained 155 708 sequences). Several “Universal primers”

have been identified based on conserved regions of 16S rRNA. probeBase (<http://www.microbial-ecology.net/probebase>) is a curated database of annotated rRNA-targeted oligonucleotide probes and supporting information (Loy, et al., 2006). probeBase contains 1258 probes and 8 microarrays from 266 references (status October 2006). Most probes in the database target highly variable sequence regions, e.g. around *Escherichia coli* position 450 and 650, while conserved regions, e.g. around position 350 and 1400 are typically occupied by only few probes that are specific for higher taxa, such as the bacterial probe EUB338 or the universal probe UNIV1390.

Comparative analysis of 16S small-subunit genes is commonly used to survey the constituents of microbial communities, to infer bacterial and archaeal evolution, and to design monitoring and analysis tools. 43% of full-length 16S small-subunit genes recorded in the GenBank database are amalgamated into pseudodivisions “environmental samples” and “unclassified” (DeSantis, et al., 2006).

While the principal phylogenetic marker is the SSU rRNA, it has at times been shown to have insufficient resolving power between closely related strains (Hill, et al., 2004), species, and even at the genus level (Forney, et al., 2004). Alternative phylogeny markers to the 16S rRNA are still presently used, such as the intergenic region between 16S and 23S (Spiegelman, et al., 2005), the coding sequence of eubacterial chaperone protein chaperonin 60 (*cpn 60*), RNA polymerase β -subunit (*rpoB*), and the DNA gyrase B protein (*gyrB*) coding for topoisomerase II. Authors have demonstrated that there is more phylogenetic information in the protein-encoding *cpn60* sequence relative to the structural RNA-encoding 16S rRNA gene (Brousseau, et al., 2001).

Completed genome sequences may offer the best and least biased measure of the dissimilarity and similarity of the two prokaryotes domains. It could be possible to determine “genomic signatures”. Ciccarelli et al. (2006) have developed an automatable procedure for reconstructing the tree of life with branch lengths comparable across all three domains (Ciccarelli, et al., 2006). The tree has its basis in a concatenation of 31 orthologs (all involved in translation) occurring in 191 species with sequenced genomes. Within Bacteria, the Firmicutes appear to comprise the earliest branching phylum, in agreement with a proposed Gram-positive ancestor for all Bacteria. Furthermore, the thermophilic Firmicute *Thermoanaerobacter tengcongensis* is the taxon with the shortest overall phylogenetic distance to the root of Bacteria. Together with the fact that slowest evolving, ancestral Archaea are also (hyper)thermophilic, this lends support to the hypothesis that the last

universal common ancestor (LUCA) was living at high temperatures. In a near future, the 31 “universal” orthologs could be probably used as biomarkers for microbial identification.

3.5) Molecular fingerprinting methods

PCR-based techniques can produce nothing more than a community fingerprint. However, gel bands in most cases can be cloned and sequenced for the precise identification of the micro-organisms.

3.5.1) Amplified ribosomal DNA restriction analysis (ARDRA)

Amplified ribosomal DNA restriction analysis (ARDRA) is a simple procedure in which a standard restriction digestion analysis is performed on PCR-amplified rDNA (SSU, ITS, LSU rRNA genes). The PCR products obtained using specific primers are then digested by various restriction enzymes or by a combination of restriction enzymes. The DNA restriction fragments are classically separated on an agarose gel producing a pattern of fragment sizes that is characteristic of the species of a microbial genus. Amplicons were digested with restriction enzyme *AluI* or *TaqI* to produce unique and distinguishable patterns to differentiate between 15 ATCC reference strains of *Bacillus* (Wu, et al., 2006). If the microbial community to characterize is complex, polyacrylamide gel is used to resolve the complex pattern.

3.5.2) Terminal-restriction fragment length polymorphism (T-RFLP) and Length heterogeneity polymerase chain reaction (LH-PCR)

Terminal-restriction fragment length polymorphism is a modified method from the ARDRA technique that uses fluorescent PCR primers and automated equipment to quantify DNA fragments. Patterns obtained by T-RFLP are more simple and precise than those obtained by ARDRA method. PCR is first performed on a community sample to amplify rDNA genes, except that one of the primers is labelled with a fluorescent dye. Restriction digestion and gel electrophoresis will always produce a single fragment that is fluorescently labelled (the terminal fragment). Automated sequencer is used to read both the size and the intensity of terminally labelled restriction fragments (T-RF). This method is useful both for creating a community fingerprint that can be compared with other communities and for

estimating species richness and diversity within a single community. Each T-RF can represent a single species, referred to in this context as a “ribotype”. rDNA database can be used to identify T-RFs predicted from all known sequences, for a given set of PCR primers and restriction enzymes (Kent, et al., 2003; Nakano, et al., 2006; Smith, et al., 2005). Care must be taken to select restriction enzymes that produce a broad community pattern. The actual limit of taxonomic resolution is often reduced to higher groups than species level. T-RFLP can provide a good view of the most abundant species and are typically limited to the 50 most abundant organisms. A limit sensitivity is estimated for member of the community which accounts for about 0.5% of the total rDNA amplified (Liu, et al., 1997). Like other PCR-based methods, biases could be considered to analysis results with caution. T-RFLP cannot effectively determine phylotype richness in extremely complex communities (Dunbar, et al., 2000). T-RF abundance or peak heights are most appropriate terms than species abundance. However, automated detection and digital output produced are clear advantages of T-RFLP. Length heterogeneity polymerase chain reaction (LH-PCR) also named amplicon length heterogeneity is a variant of the T-RFLP. Instead of identifying PCR amplicons based on restriction endonuclease sites, in LH-PCR the discrimination of amplicons is based on natural variation in the length of the 5' domain of the amplified 16S rDNA (Suzuki, et al., 1998). In LH-PCR the size is the only discriminating factor and by consequence is less accurate than T-RFLP to discriminate complex sample.

3.5.3) Denaturing gradient gel electrophoresis (DGGE) and Temperature gradient gel electrophoresis (TGGE)

Denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) are methods that separate PCR-amplified rDNA according to differences in sequence G-C content, based on differential mobility through DNA-denaturing conditions during gel electrophoresis separation.

DGGE and TGGE are the most commonly used methods of microbial community characterization. PCR-amplified DNA from rRNA genes is run on DNA-denaturing polyacrylamide gel containing urea and formamide for DGGE and under a gradient of temperature for TGGE. As DNA passes through a concentration gradient of denaturant it comes under increasing pressure to separate into single strands. The DNA is unable to denature completely because of the presence of a GC clamp, which is included in one of the primers used in the PCR reaction. The DNA comes to rest when it is almost completely

denatured. The position along the gradient at which this occurs is determined primarily by the relative proportions of G+C and A+T in a given amplicon, since G-C bonds are more difficult to denature than A-T bonds. Thus, differences in DNA composition lead to a differential migration of the PCR-products on the denaturing gel. DDGE is sensitive enough to detect even single base-pair differences between amplicons (Miller, et al., 1999). Careful calibration must be done to obtain reproducible results on microbial community analysis. A limitation of the DDGE method is the need to analysis DNA fragments below 500 bp to obtain efficient resolution during gel electrophoresis separation. Identification of differential DNA bands by cloning-sequencing is then limited for taxonomic identification due to size limitation. Large amount of DNA (500 ng) is required for effective resolution of DNA band patterns for complex environmental samples and could be an other limitation. DDGE is a useful method for visualizing the major members of a microbial community, the brightest DNA bands are often assumed to represent the dominant member of the community (Forney, et al., 2004). However, the biases associated with PCR could cause relative under- or over-representation of a given taxon. The limit of resolution of this method seems to be about 1% of the community population. In case of complex samples producing hundreds DNA bands, multiple co-migrating amplicons to the same position could occur and cannot be distinguished. On the other hand the existence of multiple copies of rRNA genes in a single species introduces ambiguity in the pattern.

3.5.4 Ribosomal intergenic spacer analysis (RISA)

Ribosomal intergenic spacer analysis (RISA) is a PCR-based method that amplifies the ribosomal intergenic region, creating community profile based on the species-specific length polymorphism in this region. Ribosomal intergenic spacer (IGS) region flanked by the 16S and 23S rRNA genes, varies in length between species from 50 to 1500 bp (Ranjard, et al., 2001). RISA can be used to distinguish between different strains and closely related species, where 16S and other rDNA-based analysis fail to do so (Shaver, et al., 2001). However, IGS database is not as large or comprehensive as other ribosomal databases reducing the potentiality to identify the species. RISA as other PCR-based techniques can be biased but in a disproportional level due to the broad range of sizes. PCR amplify preferentially shorter sequences. Biases imposed by secondary structures in the rDNA genes flanking the amplified region may also induce problem. An automated variant of RISA (ARISA) has been developed. Like T-RFLP and LH-PCR, ARISA involves the use of

universal, fluorescence-labelled PCR primers, resolved on polyacrylamide gels using automated DNA sequencers with laser detection (Ranjard, et al., 2001).

3.5.5) Single-strand conformation polymorphism (SSCP)

Single-strand conformation polymorphism is a method that separates single-stranded DNA from PCR-amplified ribosomal genes on the basis of variations in their 3D structure, which occur as a result of variations in their sequences. SSCP was originally developed to detect polymorphisms and mutation in human genes (Orita, et al., 1989). Since 16S rDNA genes are well conserved in size, these markers have been used to conduct SSCP analysis. Genes of interest are amplified by PCR and amplified products subjected to heat or chemical or enzymatic denaturation to obtain single-stranded DNA. The combination of lambda exonucleases to remove one strand of the amplicon by using one phosphorylated primer (the target for the 5'-3' exonuclease activity) and the other dephosphorylated primer for the PCR reaction and use of non-denaturing gel eliminates the artefacts of re-annealing of DNA strands after initial denaturation (Schwieger and Tebbe, 1998). Experimental evidence suggests that SSCP will produce a visible band for members of a community that comprise no lower than 1.5% of the total extracted DNA. Automation of SSCP using fluorescent primers could also be conducted on an automated DNA sequencer. Since no database exists for SSCP bands, the only way to get taxonomic information is to use other methods like cloning-sequencing of the resulting DNA bands on the pattern.

3.5.6) Randomly amplified polymorphic DNA (RAPD), Variable number of tandem repeats (VNTR), and Pulsed field gel electrophoresis

Randomly amplified polymorphic DNA is a sequence-independent PCR-based method that can be used to create a community profile based on its ability to generate a unique set of amplicons for each genome present in the sample. A single random PCR primer (usually 10 nt) will anneal to multiple sites throughout a single target DNA. Generally 5 to 15 amplicons could be generated (sites for annealing the primer could be sufficiently close to obtain PCR products). This method is usually used to differentiate between closely related species (Yost and Nattress, 2002). RAPD cannot be complemented by other methods to obtain phylogenetic affiliation due to the use of random amplification of DNA sequences but not of a common genetic marker like rRNA. Even if this technique is simple (no sequence information is

needed), rapid and of low cost, application of RAPD to a complex sample is limited by the power of resolution and of identification of the micro-organisms.

During recent years a typing method, which uses the information provided by whole genome sequencing of bacterial species, has gained increased attention. The variable number of tandem repeats (VNTRs), has proven to be suitable target for assessing genetic polymorphisms within bacterial species (Lindstedt, 2005). However, for most of the pathogenic bacteria, the method of macrorestriction of total agar-embedded DNA with a single rare cutting restriction endonuclease with subsequent fragment separation in agarose using PFGE (Pulsed Field Gel Electrophoresis) is still used today. This system for coordinated typing and surveillance has proven to be very successful in USA for foodborne disease-causing bacteria and was named PulseNet (<http://www.cdc.gov/pulsenet/>). PulseNet Europe is an initiative by a number of European countries for prevention and control of zoonoses (MED-VET-NET: <http://www.medvetnet.org>). The European “Salmgene Project” group (<http://www.salmgene.net>) is additionally promoting similar efforts to standardize typing capabilities for salmonellosis based on PFGE.

4) Hybridization methods and Isotope probing

DNA composition and DNA reassociation kinetics could also be used to characterize micro-organisms or microbial consortia.

4.1 DNA reassociation

DNA reassociation profiles use basic biochemical properties of the DNA molecule to measure the complexity of the DNA in a sample and thus estimate the diversity of the sample. The rate at which thermally denatured DNA reassociates at a fixed temperature below its melting point is a measure of the complexity of that DNA (Torsvik, et al., 1990). The absorbance of the reassociating DNA is measured spectrophotometrically over time. This value is considered to be proportionate to the complexity of the DNA in the sample. DNA reassociation only gives an estimate of total species number in a sample. This method must be completed by more specific methods to obtain phylogenetic information. % G-C profiles give information about the base composition of DNA, but cannot, on their own, be used to determine the presence or the abundance of the particular species. However, bisbenzimidazole-CsCl-gradient fractionation is a method of DNA fractionation based on

overall %G-C content, which produces a characteristic community profile of relative abundance of DNA vs. %G-C content (Nusslein and Tiedje, 1998). Bisbenzimidazole is fluorescent under long-wave UV illumination and preferentially binds to A+T regions of the DNA. This method could also be applied to enrich genomic DNA fraction based on their G-C content for subsequent molecular analysis.

4.2) Fluorescence in situ hybridization (FISH)

Fluorescence *in situ* hybridization (FISH) has become one of the most widely used tool in microbial ecology for the detection of individual cells in various environment (Amann, et al., 1990; DeLong, et al., 1989). The lack of morphological details in many bacteria and micro-organisms has usually hampered their identification, making the use of cultivation necessary for identification. This has proven difficult for many environmental or medically important micro-organisms. Therefore, in order to identify and quantify the micro-organisms occurring in a certain ecosystem, the development of molecular tools avoiding cultivation has become necessary. The FISH method is based on the hybridization of synthetic oligonucleotide probes to specific regions within the bacterial ribosome (rRNA targets) and does not require cultivation. Oligonucleotide probes are used to investigate the presence of various taxonomic groups in a sample, to measure the relative abundance of specific taxa, and to determine the spatial distribution of species or group of interest in natural microhabitat. A typical oligonucleotide probe is between 15 and 30 nucleotides in length. The accessibility of selected target sites for oligonucleotide probes can be increased by adding unlabelled probes that bind adjacent to the probe target site. The aim is to unfold the nucleic acid and thus facilitate probe hybridization. These so-called helper probes need to be designed carefully because of their specificity to the respective probe and must have a differentiation temperature at least as high as that of the probe to prevent dissociation of the helper at stringent temperature (Fuchs, et al., 2000). Under appropriate reaction conditions, complementary sequences in the probe and target cell anneal, and the site of probe hybridization is detected by fluorescence microscopy or flow cytometry. probeBase (<http://www.microbial-ecology.net/probebase>) is a curated database of annotated rRNA-targeted oligonucleotide probes and supporting information (Loy, et al., 2006).

Prior to hybridization, bacteria must be fixed and permeabilized in order to allow penetration of the fluorescent probes into cell and protect the RNA from degradation by endogenous RNAses (Moter and Gobel, 2000). In some cases, e.g. for Gram-positive, an

additional enzymatic treatment with lysozyme, lysostaphin or an enzyme mixture (Krimmer, et al., 1999) may be necessary to open the peptidoglycan layer. FISH originated in medicine and developmental biology for the localization of particular DNA sequences in mammalian chromosomes and subsequently has been applied in environmental microbiology (Bottari, et al., 2006). FISH combines the precision of molecular genetics with the visual information from microscopy, allowing visualization and identification of individual cells within their natural microhabitat or on membranes. In FISH, either the small-subunit (SSU) or large-subunit (LSU) rRNA is typically selected as the phylogenetic marker and probed in situ with fluorophore-labeled DNA oligonucleotides. rRNAs are the main target molecules because they are relatively stable, in large amount associated with ribosomes and they include both variable and highly conserved sequence domains. FISH targets rRNA, and the success of hybridization is affected by the cell size, growth rate, cell wall permeability, cellular ribosome content, number of rRNA operon copies per cell, and accessibility of target sites. When the ribosome content and permeability are not limiting, the lack of signal intensity can be attributed to quenching of the fluorophore or the inefficiency of hybridization. rRNA accessibility is a major sensitivity issue limiting the design of working probes for fluorescence *in situ* hybridization. The ability of probes to reach their target sites within the three-dimensional structure of the ribosome has been regarded as the major factor for hybridization efficiency (Amann, et al., 1995). The differential accessibility of target sites on the SSU and LSU rRNA of *Escherichia coli* were systematically studied (Fuchs, et al., 2001; Fuchs, et al., 1998), and the accessibility of target sites on the SSU rRNA of three other organisms (*Pirellula* sp strain 1: *Planctomycetes*, *Bacteri*; *Metallosphaera sedula* : *Crenarchaeota*, *Archaea*; *Saccharomyces cerevisiae* : *Eucarya*) was studied (Behrens, et al., 2003). These studies empirically highlighted the accessibility of target sites on rRNA maps by grouping probes into six classes according to their brightness levels. Although the accessibility maps have reportedly been used for selection of potential target sites, their predictive power regarding the hybridization efficiency of newly designed probes is limited, as probe brightness can vary remarkably with small shifts at the targeted region or when different organisms are targeted.

Calculations of thermodynamic properties of oligonucleotides could be done using the program OligoWalk (Mathews, et al., 1999) or the Excell macro "OligoAnal" (Matveeva, et al., 2003). A mechanistic model of FISH was proposed to provide a mathematical basis for the development of efficiency probes (Yilmaz and Noguera, 2004). The kinetic limitations imposed by structural restraints in the ribosome, which could be improved by increasing the

incubation period or using formamide, may be the reason for the dim signals obtained with some probes. It was recently proven that every nucleotide in the 16S rRNA of *Escherichia coli* could be targeted with a bright probe and, therefore, that there were no truly inaccessible target regions in the 16S rRNA (Yilmaz, et al., 2006). Accessibility of target site is thought to be function of secondary and tertiary rRNA-rRNA and protein-rRNA interactions. The presence or absence of the strongly denaturing detergent sodium dodecyl sulphate had a much more pronounced effect than a change of fixative from paraformaldehyde to ethanol (Behrens, et al., 2003). The authors recommend the following steps for rational design of FISH probes to achieve high fluorescence intensity. (i) Locate a target segment on rRNA that is specific to the organisms of interest. Kinetically accessible sites can be selected to previously published accessibility maps (Behrens, et al., 2003; Fuchs, et al., 2001; Fuchs, et al., 1998) which are already linked to the ARB software (Kumar, et al., 2005). (ii) Determine the length of the probe so that $\Delta G^0_{\text{overall}}$ has a value of ca. -13 kcal/mol. (iii) Starting with conventional periods (i.e., 2 to 5h), extend the hybridization time until the brightness reaches satisfactory levels. (iv) If step iii does not work with reasonable hybridization times, return to step ii, increase the magnitude of $\Delta G^0_{\text{overall}}$. (v) If steps ii to iv do not work, quenching might be the cause, so consider switching to a new dye (cy3 instead of fluorescein) or changing its position. Fluorescein is known to be severely quenched by nearby guanine nucleotides or G-C base pairs (Marras, et al., 2002). (vi) If step ii to iv require prolonged hybridizations or the design of high-affinity probes, consider using nonlabeled competitor oligonucleotides to maximize probe specificity. Since using probes with very high $\Delta G^0_{\text{overall}}$ values might enhance the affinity of the probe also to nontarget organisms with few mismatches, increasing $\Delta G^0_{\text{overall}}$ values far beyond the -13 kcal/mol threshold is not recommended (Yilmaz and Noguera, 2004). The theoretical value for melting point is -10 kcal/mol. Negative controls were routinely prepared using the complement to the EUB probe (nonEUB; 5'-ACTCCTACGGAGGCAGC-3'), which is not supposed to bind to rRNA (Wallner, et al., 1993).

Some improvements to the sensitivity of the standard fluorescence *in situ* hybridization procedure have come recently from the use of catalyzed reporter deposition (CARD-FISH). By using a novel tyramide substrate for the generation of fluorescent signal by an oligonucleotide-linked catalytic enzyme, researchers have been able to enhance fluorescence intensities and signal-to-noise ratios, increase the detection rates of particular taxa, and in some cases to detect taxonomic groups undetectable by standard FISH (Pernthaler, et al., 2002). To investigate the relationship between specific substrate utilization

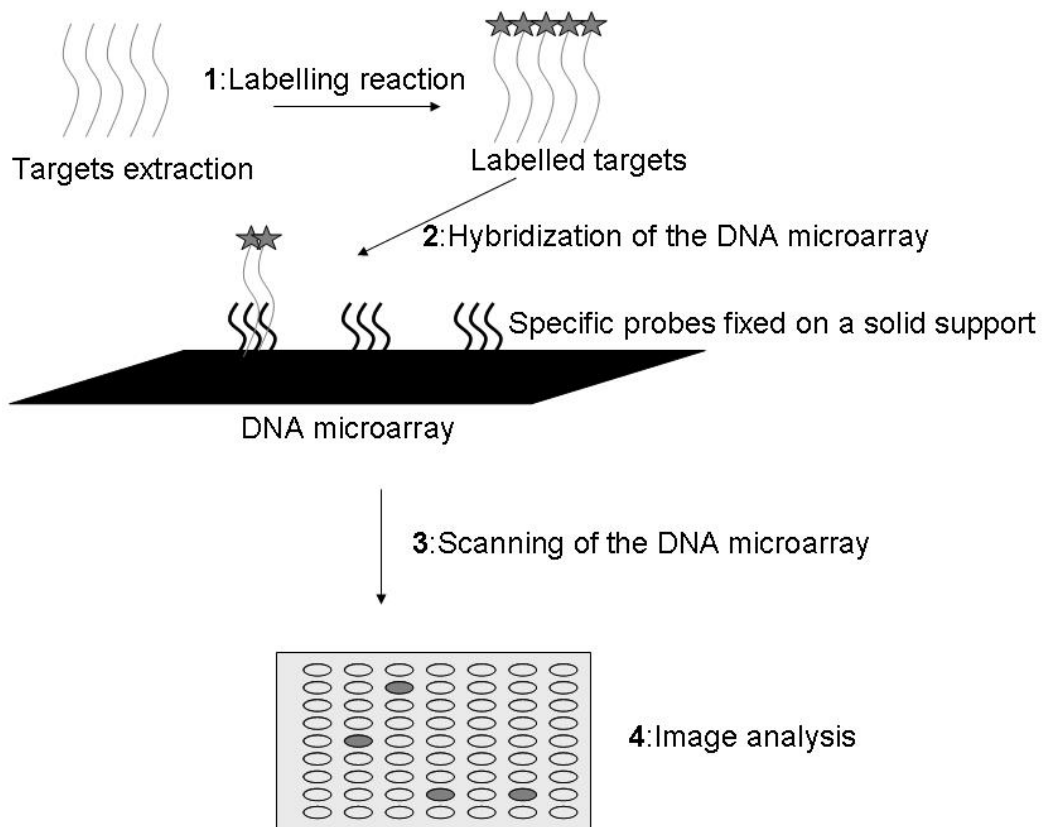


Figure 4: Schematic representation of DNA microarray experiment

and phylogeny, investigators have combined probe hybridization and microautoradiography MAR-FISH (Ouverney and Fuhrman, 1999). Alternatively, secondary ion mass spectrometry (SIMS) can be combined with *in situ* hybridization to reveal the relationship between phylogeny and naturally occurring variation in stable isotope ratios, indicative of particular metabolic processes such as anaerobic methane oxidation (Orphan, et al., 2001).

mRNAs could be used successfully for *in situ* visualisation (Pernthaler and Amann, 2004). The recognition of individual genes by FISH (RING-FISH) technology is based on target visualization mediated polynucleotide probe network formation (Zwirgmaier, et al., 2004).

DNA probes are often limited in their capability to distinguish single nucleotide changes. As a result, it is often difficult to design DNA probes that hybridize efficiently with a given stretch of the rRNA dictated by the nucleotide differences found between closely related species. The use of peptide nucleic acids (PNA) has been considered useful in overcoming the variable and sometimes insufficient penetration of probes into bacteria depending on their cell wall characteristics. PNAs are uncharged DNA analogues that bind to nucleic acids much more strongly than oligonucleotides because there is no electrical repulsion between the PNA probe and the negatively charged sugar-phosphate backbone of the target molecule (Ray and Norden, 2000). Moreover, the PNA probes hybridize virtually independently of the salt concentration, being ideal for targeting nucleic acids with high degree of secondary structure like rRNA. At this time, high prices and specificity problems are slowing the application of PNAs to FISH (Amann, et al., 2001).

4.3) DNA microarrays

DNA microarray (microchip, biochip, gene chip) technology is a powerful, high throughput experimental system that allows the simultaneous analysis of thousands to hundreds of thousands of genes at the same time (Figure 4). Originally developed for monitoring whole-genome gene expressions, micro-arrays have been used for other purposes such as the genome-wide mutational screening for single nucleotide polymorphisms and the distribution of species or strains in natural environments. The application of microarray technology for microbial diagnostics is a field in the stage of dynamic development (Loy and Bodrossy, 2006).

This is a very high-throughput technique in which a large number of taxonomically differentiated oligonucleotides or functional oligonucleotides are immobilized on a solid

surface and hybridized to labelled nucleic acids. Different platforms currently used in microbial diagnostics microarrays have been developed (Bodrossy and Sessitsch, 2004). Macroarrays (dot blots on nitrocellulose or nylon membranes) have the disadvantage of moderate throughput and uncontrolled binding of oligonucleotides. Planar glass microarrays have become the most widely used type of array. DNA micro-arrays are glass surfaces to which arrays of specific DNA fragments of various lengths have been attached or *in situ* synthesized (photolithography technology of Affymetrix : <http://www.affymetrix.com> or the ink jet technology of Agilent : <http://www.home.agilent.com>) at discrete locations. These fragments serve as probes for hybridization. Under conditions suitable for hybridization, the DNA spots on the chip are exposed to a solution containing a complex sample of fluorescent-labelled DNA. The use of oligonucleotide microarray technology requires a very detailed attention to the design of specific probes spotted on the solid phase (Militon, et al., 2007; Rimour, et al., 2005). The cross-hybridization is the major point that limits the determination of specific probes. Hybridization of rRNAs to microarray is a promising approach for prokaryotic and eukaryotic species identification. Recent results recommend against the application of probe design software tools that use thermodynamic parameters to assess probe quality (Pozhitkov, et al., 2006). It has become clear that the simple notion that short oligonucleotides with a mismatch should hybridize less efficiently than perfect match probes is not always applicable. The most optimal discrimination is provided with the mismatch in the middle of the duplex with short oligonucleotides (20 nt). The higher destabilization is obtained with GA and GC mismatches. On the other hand, lesser destabilization is seen with TC, TT and TG mismatches. Some factors affecting duplex formation on DNA microarrays include: probe density, microarray surface composition and the stabilities of oligonucleotides-target duplexes, intra- and inter- molecular self-structures and RNA secondary structures. Zhou et al. (2004) systematically tested a prototype POA (Phylogenetic Oligonucleotide Array) consisting of 19-mer probes for different bacterial 16S rRNA gene sequences with one to five mismatches in the mismatched probes (Zhou, et al., 2004). A single mismatch at the central position of the probes decreased the signal to only 15-25% the intensity of the perfectly matched probes.

BeadArray technology has been recently developed and is based on the random self-assembly of a bead pool onto a patterned substrate (Fan, et al., 2006). The density of a randomly assembled 300-nm-diameter silicon bead array is ~40,000 times higher than a typical spotted microarray. Currently, up to 1624 unique bead types are represented in each of the universal arrays, with an average 30-fold bead type redundancy in each array (~50,000

beads). Each bead is derivatized on its surface with approximately 700,000 covalently attached oligonucleotide probes of a unique sequence. The gene expression profiling assay called DASL (cDNA-mediated annealing, selection, extension, and ligation) multiplexes to over 1500 sequence targets, e.g., 500 genes at three probes per gene or 1500 genes at one probe per gene. To date, no microbial diagnostic has been made with these new arrays.

Recently, several types of micro-arrays have been developed and evaluated for bacterial detection and microbial community analysis (Gentry, et al., 2006; Loy and Bodrossy, 2006). The first proof of principle for the parallel detection of bacteria using a 16S rRNA-based DNA microarray was given by Guschin et al. (Guschin, et al., 1997). The 16S rRNA gene is the most widely used marker, however, it does not allow for resolution below the species level. Alternative universal marker genes with higher resolution include *rpoB*, *recA*, *gyrB*, *groEL*, *atpD* or the *tmRNA* gene. The application of functional genes as marker genes focuses array to a defined physiological group of microbes. A comprehensive thoroughly validated microarray covering the entire known diversity of methanotroph was conducted using *pmoA* genes as the functional marker (Bodrossy, et al., 2003). The array used in microbial evaluation can be divided into at least five categories based on the genes targeted by the array (Gentry, et al., 2006): (1) Phylogenetic oligonucleotide arrays (POAs) are designed based on a conserved marker such as the 16S ribosomal RNA (rRNA) gene, which is used to compare the relatedness of communities in different environments. The most comprehensive POA developed so far contained 31,179 perfectly matched hierarchical 20-mer probes (with a corresponding number of single mismatch probes as negative, mismatch controls) targeting 1945 prokaryotic and 431 eukaryotic sequences from the RDP (Wilson, et al., 2002). (2) Functional gene arrays (FGAs) are designed for key functional genes that encode for proteins involved in various metabolic processes. The largest FGA published to date contained 1662 probes for genes involved in carbon, nitrogen, and sulphur cycles, organic contaminant degradation, and metals resistance and reduction (Rhee, et al., 2004), but this FGA has recently been expanded to over 24,000 probes as indicated by Gentry et al. (Gentry, et al., 2006). (3) Community genome arrays (CGAs) contain the whole genomic DNA of cultured organisms and can describe a community based on its relationship to these cultivated organisms. A prototype array that contained the entire genomic DNA of 67 different bacteria (one species' genome per spot) including α -, β -, δ -*Proteobacteria* and Gram-positive bacteria was developed (Wu, et al., 2004). The CGA was used to compare the microbial populations in four marine sediments, three river sediments, and three soils. (4) Metagenomic arrays (MGA) are a potentially powerful technique because, unlike the other

arrays, they contain probes produced directly from environmental DNA itself and can be applied with no prior knowledge of the community. Sebat et al. (2003) generated an MGA using a cosmid library from groundwater enrichment (Sebat, et al., 2003). (5) Whole-genome open reading frame (ORF) arrays (WGA) contain probes for all of the ORFs in one or multiple genomes. Dong et al. (2001) used a WGA containing 96% of the annotated ORFs in *E. coli* K-12 to comparatively interrogate the genome of the closely related (97% based on 16S rRNA gene) *Klebsiella pneumonia* 342, which is a maize endophyte (Dong, et al., 2001).

An important aspect in DNA microarray analysis of environmental microbial communities is the sensitivity of the protocol applied, especially if an initial PCR amplification step in target molecule preparation is to be avoided. In this context, the 16S rRNA represents the phylogenetic marker of choice due its high copy number of up to 10^5 per cell (Amann, et al., 1995). Direct hybridizations of total RNA extracted from complex environment were successfully employed in several microarray studies (Chandler, et al., 2006; El Fantroussi, et al., 2003; Peplies, et al., 2006; Peplies, et al., 2004; Small, et al., 2001). Direct labelling of RNA could be done by the ULSTM technology (www.kreatech.com) that is part of the Ulysis labelling Kit from Roche and is based on the stable coordinative binding of a platinum dye complex to nucleic acids. The conjugate forms a coordinative bond, by binding to the N7 position of the guanine bases. The detection limit in terms of relative cellular abundance was found to be in the range of 2 to 5% after hybridization of 500 ng of environmental rRNA (Peplies, et al., 2006). In comparison, Loy et al., (2005) have hybridized 400 ng of PCR products of the 16S rRNA gene and reported the detection of members of the betaproteobacterial order *Rhodocyclades* with a relative cellular abundance of less than 1% only after applying primer pairs that were highly specific for this particular phylogenetic group (Loy, et al., 2005). Long (50-70mers) oligoprobes display fundamentally different hybridization properties (Kostic, et al., 2005). The sensitivity is several orders of magnitude above that of 1-5% found for short oligonucleotides (20-25mers). The disadvantage of long oligoprobes is their decreased specificity. GoArrays a new algorithm for probe design could resolve this problem in combining short oligonucleotides to obtain long ones (Milton, et al., 2007; Rimour, et al., 2005). Chandler et al. (2003) applied a two probe chaperone detection system, consisting of an immobilized, species-specific capture probe and a labelled proximal detector (chaperone) probe for the direct detection of 16S rRNA (Chandler, et al., 2003). Detection sensitivity in the range of 0.1% has been demonstrated using a microbial diagnostic microarray based on sequence-specific end labelling of oligonucleotides (SSELO) with *gyrB* as the marker gene for the identification of pathogenic bacteria in a background of

nonpathogen (Kostic, et al., 2007). However, this system seriously limits the options for probe design. Capture oligonucleotides are immobilized on the microarray. Reverse complements of the capture oligonucleotides are end labelled in a linear amplification reaction upon the availability of the corresponding target sequence. The entire mixture is then hybridized to the microarray to sort out the sequences that have been labelled.

The integration of isotope and microarray technologies produces one of the potentially most powerful combined approaches for microbial ecology research. Microarray analysis of DNA or RNA labelled with isotopes can differentiate between active and inactive organisms in a sample. Adamczyk et al. (2003) used ^{14}C labelled bicarbonate and a POA to study ammonia-oxidizing bacterial communities in two samples of nitrifying activated sludge (Adamczyk, et al., 2003).

4.4) Stable Isotope Probing (SIP)

One of the biggest challenges that microbiologists face is to identify which micro-organisms are carrying out a specific set of metabolic processes in the natural environment. FISH-microautoradiography (FISH-MAR) and the isotope array both use radioactive tracers to monitor the incorporation of substrate (Dumont and Murrell, 2005). The isotope array involves incubating an environmental sample with a ^{14}C -labelled substrate, after which the RNA is extracted from the sample, labelled with a fluorophore and analysed with an oligonucleotide array that targets 16S rRNA. The array is then scanned for fluorescence and incorporation of radioactive isotope to determine which community members have metabolized the substrate.

SIP was first applied in the analysis of phospholipid fatty acids (PFLA) that can be extracted from micro-organisms and analysed by isotope-ratio mass spectrometry (IRMS). Although PFLA analysis offers great sensitivity, the use of labelled nucleic acids as biomarkers has the potential to identify a wider range of bacteria with a greater degree of confidence. DNA-based SIP (DNA-SIP) is increasingly being used in attempts to link the identity of micro-organisms to their functions (Morris, et al., 2002; Radajewski, et al., 2000; Radajewski, et al., 2002). DNA-SIP is dependent on the commercial availability of compounds that are highly enriched in ^{13}C . This approach has been used to characterize bacteria metabolizing C1-compounds such as methane, methanol and methyl halides in various environment, as well as multi-carbon compounds, including acetate, glucose, caffeine, benzene, naphthalene, phenol, salicylate, phenanthrene, 2,2'-dichlorobiphenyl. The

incorporation of a high proportion of ^{13}C into DNA greatly enhances the density of labelled DNA compared with unlabelled (^{12}C) DNA. The DNA was isolated and subjected to caesium chloride (CsCl) buoyant density-gradient centrifugation with ethidium bromide. The heavy ^{13}C -DNA can be purified away from the light ^{12}C -DNA by needle collection and used as a template in PCR, with general primer sets that amplify rRNA genes of most known Bacteria, Archaea or Eukarya. It is possible to target “functional” genes as it was demonstrated for methanotroph bacteria (Cebron, et al., 2006). Methanotroph population diversity was studied by targeting 16S rRNA and *pmoA* genes encoding methane monooxygenase involved in the first step of CH_4 oxidation. However, unnatural concentrations of nutrients may artificially change bacterial diversity and activity. One major disadvantage of this incubation was that $^{13}\text{CH}_4$ was slow and some cross-feeding of ^{13}C occurred via trophic interactions, thereby leading to labelling of non-methanotroph micro-organisms. Recently, the potential to combine DNA-SIP with metagenomic analysis has been recognized and applied to identify a BAC (Bacterial Artificial Chromosome) clone containing an operon involved in one-carbon metabolism (Dumont, et al., 2006).

RNA-SIP provides a solid platform from which to scour genomic sequences for biotechnologically valuable and ecologically relevant information. RNA-SIP studies generally require shorter incubation time since RNA is labelled faster than DNA but recovery of pure RNA from soil is still difficult (Gallagher, et al., 2005; Kasai, et al., 2006; Manefield, et al., 2002; Whiteley, et al., 2006). RNA could serve as a more responsive biomarker than DNA for use in SIP, because in active cells, RNA synthesis occurs at a high rate and labelling can also occur without the need for DNA synthesis or replication of the organism. RNA-SIP cannot be performed in caesium chloride (CsCl) gradient media, because CsCl precipitates during centrifugation at the buoyant density for RNA banding ($\sim 2.0 \text{ g mL}^{-1}$). Caesium trifluoroacetate (CsTFA), however, allows an appropriate buoyant density for the banding of rRNA (36-60h runs are necessary), but has a reduced loading capacity. The complete fractionation of gradients is essential for RNA-SIP due to low relative RNA concentrations prevent visualization with nucleic acids stains (Lueders, et al., 2004). To verify label incorporation it is necessary to show that a given RNA template in “heavy” fractions is absent or less abundant in “light” ones and that this distribution in the gradient becomes apparent over time, as the labelled substrate is consumed. Even in pure culture, nucleic acids distribute across the length of density gradient such that they contaminate the more dense fractions. This could lead to the mistaken analysis of “light” nucleic acids that occur lower in a density gradient. Another concern is that high GC content can increase the buoyant density

approaching that of lower-GC content ^{13}C molecules. Obtaining a $\delta^{13}\text{C}$ value by isotope ratio mass spectrometry (IRMS) provides a useful confirmation that the retrieved nucleic acids are enriched in ^{13}C and to what extent. Singleton et al. (2005) used unlabeled DNA from *Escherichia coli* K12 as an unlabeled DNA in the high-density fractions (Singleton, et al., 2005). *E. coli* K12 was specifically targeted by PCR to determine the extent to which unlabeled DNA contaminated the high-density fractions. Theoretically, isotopically enriched mRNA could also be extracted and fractionated to retrieve genes from organisms actively incorporating labelled substrate. This has yet to be demonstrated and has thus far been limited by low yields of mRNA retrieved from environmental samples, the difficulty of cloning mRNA independent of background rRNA and potentially by technical difficulties in the separation of mRNA species by isotopic density centrifugation (Neufeld, et al., 2006).

5) Metagenomics

Metagenomics refers to the study of collective genomes in an environmental community. Tringe et al. (2005) estimated that more than 2×10^9 bp of sequence from 3×10^6 sequence reads would be necessary to sequence 8-fold coverage of the most abundant species in their soil sample assuming a genome size of 6 Mbp (Tringe, et al., 2005). Although not currently feasible, sequencing 8-fold coverage of every OTU_{0.03} in the Alaskan soil metagenome would require sequencing 950,000 genome equivalents or 6×10^{12} bp of DNA.

The analysis of genome sequence data that has been recovered from the environment is motivated by many objectives, which include the establishment of gene inventories and natural product discovery (Allen and Banfield, 2005; Handelsman, 2004). Genomics can resolve the genetic and metabolic potential of communities and established how functions are partitioned in and among populations, reveal how genetic diversity is created and maintained, and identify the primary drivers of genome evolution and speciation. Simple communities with low species diversity can be characterized thoroughly with moderate effort sequencing. The best example of such an analysis is the nearly complete sequencing of the metagenome of a community in acid drainage of the Richmond mine (California), which represents one of the most extreme environments on earth (Tyson, et al., 2004). The microbial community forms a pink biofilm that floats on the surface of the mine waters. The drainage water below the biofilm has a pH of between 0 and 1 and high levels of Fe, Zn, Cu, and As (317, 14, 4 and, 2 mM respectively). The solution around the biofilm water is 42°C and microaerophilic. There

is no source of carbon or nitrogen other than the gaseous forms in the air. Environmental sequence data (76 Mb) were obtained from a small-insert library from a single biofilm sample of the Richmond Mine at Iron Mountain (Tyson, et al., 2004). This metagenomic approach revealed that a relatively minor community component, *Leptospirillum* group III, possessed the sole complement of nitrogen fixation (*nif*) genes. This subsequently led to the design of a selective isolation strategy to successfully cultivate this organism using genome sequence data (Tyson and Banfield, 2005). All the genomes in the acid mine drainage are rich in genes associated with removing potentially toxic elements from the cell.

On the other hand, complex communities are more representative of most natural microbial assemblages, but their characterization presents myriad challenges that require special consideration. Once large scaffolds are generated, they can be assigned to the correct organism based on various parameters, which include: the phylogeny of conserved marker genes, such as 16 rRNA genes, *recA/radA*, *rpoB/C*, Hsp60/70, EF-Tu, and α IF2 β ; % GC content; depth of sequence coverage (number of reads per unit length of DNA); codon usage; di-, tri-, tetra-nucleotide frequencies. Extensive shotgun sequencing of the Sargasso sea produced ~1.36 Gb of sequence data, estimated to derive from >1,000 genomic species. Over 1.2 million gene products were identified and provided a basis for the estimation of diversity. Despite this massive effort, near-complete genome assembly was only possible for a few organisms, with the dominant organism sampled, *Burkholderia* spp., represented at ~21X coverage. The most significant obstacle in metagenomics is the microbial complexity in most communities (Schloss and Handelsman, 2005). The Minnesota farm soil probably contains more than 5,000 species and 10^4 - 10^5 strains, making it inevitable that the over 15,000 sequence reads could not be assigned into contigs (Tringe, et al., 2005). The US Joint genome Institute has essentially sequenced the metagenomes of the microbial communities associated with two extinct ancient cave bears, which contains less than 2 and 6% cave bear DNA, respectively (Noonan, et al., 2005). The J. Craig Venter Foundation is sequencing the metagenomes of samples taken along a path intended to simulate the voyage of Darwin's ship *The Beagle*, as well as samples of New York City's air. A future prospect is completing the human genome by sequencing the metagenome of 1012 microbial cells that are associated with the human body.

The advent of metagenomics was accompanied by promises of great biotechnological applications as well as fundamental discovery. The promise of metagenomics as a source of new technology has not been fully realized, primarily because of the challenges in screening for the desired activities. Screening of metagenomic libraries has traditionally followed two

paths: sequence-based and function-based screening. Some of metagenomic libraries have been screened by hybridization or by PCR to detect genes with homology to known genes. This approach has been fruitful, but is necessarily limited to discovery of genes in previously described families. The alternative is functional analysis, which enables researcher to discover entirely new classes of genes for useful functions but requires expression of genes from exotic organisms in tame bacteria such as *Escherichia coli*. The most significant requirement is that regulatory machinery that recognizes the promoter and the substrate must be present and functional in the host cell. Screen in other host species will overcome this barrier. High-throughput screen for catabolic pathways designated “substrate-induced gene expression screening” or SIGEX is based on the observation that genes encoding the enzymes for catabolic pathways are commonly arranged in operons that are induced by pathways’ substrates (Uchiyama, et al., 2005). The authors exploit this feature in an ‘operon trap” in which the metagenomic DNA is cloned upstream of the *gfp* gene (green fluorescent protein gene), thereby placing green fluorescent protein (GFP) expression under the control of promoters in the metagenomic DNA. Fluorescence-activated cell sorting is used to detect positive clones. From a library of 152,000 clones with inserts averaging 7 kb, 58 clones were identified that were regulated by benzoate and 4 clones by naphthalene. Small-molecule discovery by functional metagenomics has concentrated on antibiotics, which are of interest for their pharmaceutical applications as well as their roles in ecosystem function. Frequencies antibiotic-producing clones are detected at a frequency of approximately 1 producer per 10^4 clones (Gillespie, et al., 2002). The first polyketide synthases, enzymes involved, in synthesis of polyketides, broad class of antibiotics that includes erythromycin, epithilone, and rifamycin, were first cloned from soil with a PCR-based approach (Seow, et al., 1997). A surprising result from the Davies group indicated that subinhibitory concentrations of many antibiotics induce quorum sensing despite no resemblance in structure to acylated homoserine lactone that appear to be natural inducers (Goh, et al., 2002). An opportunity was investigated to capture molecules that are quorum sensing inducers as well as antibiotics by designing a high throughput screen (Andersen, et al., 2001). The potential for application of metagenomics to biotechnology seems endless.

Fingerprinting could be obtained with proteomics and metabolomics approaches but results are difficult to relate to a precise identification at a phylogenetic level. Databases must be constructed and related genomics, proteomics and metabolomics informations to obtain easily interpretable results for micro-organisms identification.

Currently, both the cost of sequencing and the challenges that are associated with the management of vast datasets precludes comprehensive genomic studies of highly complex communities.

6) New high-throughput sequencing developments

Micro-organisms are the first and most abundant species to be completely sequenced. Genomes Online Database is web resource for comprehensive access to information regarding complete and ongoing genome projects all around the world (<http://www.genomesonline.org/>; (Liolios, et al., 2006). The database currently (at the date of January, 20 2007) incorporates information on 2299 sequencing projects, of which 495 have been completely sequenced (418 Bacterial, 32 Archaeal, 45 Eukaryal among them 15 Protozoa and 14 Fungi). Sequencing programs in progress is largely dominated by prokaryotes (1034 Bacterial, 58 Archaeal, 640 Eukaryotic and 72 Metagenomes).

DNA sequencing can be used to gain important information on genes, genetic variation and gene function. Improvements of the Sanger sequencing technique (fluorescent labelling, automatic sequencing using capillary electrophoresis) lead to generation of long reads (1000 bases) unable for the success of complete genome sequencing. The vast majority of DNA sequences has been obtained by using this method, which is based on the electrophoretic separation of deoxyribonucleotide triphosphate fragments with single-base resolution. But cost and throughput remain the key limitations of sequencing.

Novel technologies are being developed to meet this need by generating massive amounts of sequence (Bentley, 2006; Shendure, et al., 2005). New sequencing technologies are expected to become the mainly of future research, and to make DNA sequencing centre stage as routine tool in genetic research in the coming years. A bacterial genome could be sequenced for €400, similar to the cost of some routine clinical pathology test. Further cost reductions of between 1000- and 10 000 fold are required. A 384-capillary instrument can be miniaturised conducting to a major reduction in reagent volumes (i.e. more than 1000 fold). Sequencing by hybridization using array of oligonucleotides is most widely used for genotyping because most repetitive sequences are excluded from this type of analysis. On the other hand, sequencing by synthesis on arrays seems highly efficient approaches. These systems avoid bacterial cloning steps and can potentially generate sequence data from DNA that cannot be propagated in bacteria. The 454 system (www.454.com) uses templates that are

immobilised and amplified on beads in aqueous-oil emulsion. Beads with DNA are purified and placed in individual microfabricated picolitre wells for pyrosequencing (chemoluminescent signal). Typically, 80 cycles of incorporation yield reads of approximately 100 bases. Exploitation of the massively parallel DNA sequencing of the 454 system was used to evaluate the microbial diversity in the deep sea (Sogin, et al., 2006). 118,000 PCR amplicons that spans the V6 hypervariable region of ribosomal RNAs were sequenced. In the Solexa system (www.solexa.com), single molecules are covalently attached to a planar surface and amplified *in situ*. Sequencing by synthesis is carried out by adding a mixture of four fluorescently labelled reversible chain terminators. Read lengths are 30-50 bases. Shendure et al. (2005) attach template DNA molecules to 1 µm beads that are embedded in polyacrylamide for amplification, forming “colonies” (polymerase colonies). Sequencing is performed using multiple cycles of ligation of labelled 9-mers from a start point determined by an anchor primer. Several groups have investigated the possibility of sequencing directly from single DNA molecules. In the nanopores technology, DNA molecule is driven by an electric field through a 1.5 nm pore at a rate of at least ~1000 bases per second. Conductance through the pore can be measured and provides a signal that is potentially characteristic of the identity of the base obstructing the pore. Other proposed arrangements use polymerase-based sequencing in real time, detecting base-specific incorporation as the polymerase moves along a single DNA molecule at a rate of ~100 bases per second (www.pacificbiosciences.com; www.visigenbio.com).

Conclusions :

Micro-organisms represent the largest diversity of organisms on Earth. Even with recent molecular-based methods we cannot estimate precisely the number of different species in complex environments like soils or aquatic environments. The major difficulties are to be able to find specific characters that distinguish easily and rapidly micro-organisms from each others. The small sizes, the large species number, the few visual phenotypic traits, the difficulty to isolate and cultivate micro-organisms reinforce these difficulties. Furthermore, the complex evolution of micro-organisms conducted by numerous lateral gene transfers create a confuse situation to define the “species” notion. We are at the beginning of discovery of the variety of micro-organisms and of their metabolic capacities. Micro-organisms are major players in biogeochemical processes, in plants and animals healthy but also for a small

part of them involved in severe diseases. Micro-organisms evolved continuously and are able to rapidly adapt to new environments. The microbial world is fascinating and should be a gold mine to understand the origin of life and to discover new valuable molecules for medicine and biotechnology.

Perspectives:

Considerable efforts must continue for the development of new cultivation approaches. We need a better evaluation of the physical and chemical composition of the different explored ecological niches to more precisely define new conditions for efficient cultivation. Due to a very slow growing of a large part of micro-organisms, technical developments for a more sensitive detection of the growth will be helpful for microbiologists. In the next years, the part of “uncultivable” (we should say “not already cultivate”) micro-organisms will progressively decrease. High throughput cultivation procedures will gain interest to isolate unique cells or consortia. Improvements of cultivation conditions will serve a better understanding of the microbial physiology.

In the same time, high throughput molecular-based methods will increase our knowledge on microbial genetic diversity. Given the current rate of progress in sequencing technology, we can expect in few years the sequencing of complete genomes at very low cost and in a reduced time. Microbial identification could benefit from these improvements.

References

- Adamczyk, J., Hesselsoe, M., Iversen, N., Horn, M., Lehner, A., Nielsen, P.H., Schloter, M., Roslev, P. and Wagner, M. (2003) The isotope array, a new tool that employs substrate-mediated labeling of rRNA for determination of microbial community structure and function, *Applied and environmental microbiology*, **69**, 6875-6887.
- Allen, E.E. and Banfield, J.F. (2005) Community genomics in microbial ecology and evolution, *Nature reviews*, **3**, 489-498.
- Amann, R., Fuchs, B.M. and Behrens, S. (2001) The identification of microorganisms by fluorescence in situ hybridisation, *Current opinion in biotechnology*, **12**, 231-236.
- Amann, R.I., Krumholz, L. and Stahl, D.A. (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology, *Journal of bacteriology*, **172**, 762-770.
- Amann, R.I., Ludwig, W. and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation, *Microbiological reviews*, **59**, 143-169.
- Andersen, J.B., Heydorn, A., Hentzer, M., Eberl, L., Geisenberger, O., Christensen, B.B., Molin, S. and Givskov, M. (2001) gfp-based N-acyl homoserine-lactone sensor systems for detection of bacterial communication, *Applied and environmental microbiology*, **67**, 575-585.
- Audibert, A., Weil, D. and Dautry, F. (2002) In vivo kinetics of mRNA splicing and transport in mammalian cells, *Molecular and cellular biology*, **22**, 6706-6718.
- Behrens, S., Ruhland, C., Inacio, J., Huber, H., Fonseca, A., Spencer-Martins, I., Fuchs, B.M. and Amann, R. (2003) In situ accessibility of small-subunit rRNA of members of the domains Bacteria, Archaea, and Eucarya to Cy3-labeled oligonucleotide probes, *Applied and environmental microbiology*, **69**, 1748-1758.
- Bentley, D.R. (2006) Whole-genome re-sequencing, *Current opinion in genetics & development*, **16**, 545-552.
- Berthelet, M., Whyte, L.G. and Greer, C.W. (1996) Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpyrrolidone spin columns, *FEMS microbiology letters*, **138**, 17-22.
- Bodrossy, L. and Sessitsch, A. (2004) Oligonucleotide microarrays in microbial diagnostics, *Current opinion in microbiology*, **7**, 245-254.

- Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A. and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs, *Environmental microbiology*, **5**, 566-582.
- Bottari, B., Ercolini, D., Gatti, M. and Neviani, E. (2006) Application of FISH technology for microbiological analysis: current state and prospects, *Appl Microbiol Biotechnol*, **73**, 485-494.
- Brenner, D.J., Fanning, G.R., Steigerwalt, A.G., Sodd, M.A. and Doctor, B.P. (1977) Conservation of transfer ribonucleic acid and 5S ribonucleic acid cistrons in Enterobacteriaceae, *Journal of bacteriology*, **129**, 1435-1439.
- Brodie, E.L., Desantis, T.Z., Parker, J.P., Zubietta, I.X., Piceno, Y.M. and Andersen, G.L. (2006) Urban aerosols harbor diverse and dynamic bacterial populations, *Proceedings of the National Academy of Sciences of the United States of America*.
- Brousseau, R., Hill, J.E., Prefontaine, G., Goh, S.H., Harel, J. and Hemmingsen, S.M. (2001) Streptococcus suis serotypes characterized by analysis of chaperonin 60 gene sequences, *Applied and environmental microbiology*, **67**, 4828-4833.
- Cavalier-Smith, T. (1987) Eukaryotes with no mitochondria, *Nature*, **326**, 332-333.
- Cavalier-Smith, T. (2006) Cell evolution and Earth history: stasis and revolution, *Philosophical transactions of the Royal Society of London*, **361**, 969-1006.
- Cebon, A., Bodrossy, L., Stralis-Pavese, N., Singer, A.C., Thompson, I.P., Prosser, J.I. and Murrell, J.C. (2006) Nutrient amendments in soil DNA stable isotope probing experiments reduce observed methanotroph diversity, *Applied and environmental microbiology*.
- Chandler, D.P., Jarrell, A.E., Roden, E.R., Golova, J., Chernov, B., Schipma, M.J., Peacock, A.D. and Long, P.E. (2006) Suspension array analysis of 16S rRNA from Fe- and SO(4)2-reducing bacteria in uranium-contaminated sediments undergoing bioremediation, *Applied and environmental microbiology*, **72**, 4672-4687.
- Chandler, D.P., Newton, G.J., Small, J.A. and Daly, D.S. (2003) Sequence versus structure for the direct detection of 16S rRNA on planar oligonucleotide microarrays, *Applied and environmental microbiology*, **69**, 2950-2958.
- Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B. and Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life, *Science*, **311**, 1283-1287.
- Corinaldesi, C., Danovaro, R. and Dell'Anno, A. (2005) Simultaneous recovery of extracellular and intracellular DNA suitable for molecular studies from marine sediments, *Applied and environmental microbiology*, **71**, 46-50.

- Dagan, T. and Martin, W. (2006) The tree of one percent, *Genome Biol*, **7**, 118.
- Davis, K.E., Joseph, S.J. and Janssen, P.H. (2005) Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria, *Applied and environmental microbiology*, **71**, 826-834.
- DeLong, E.F. and Pace, N.R. (2001) Environmental diversity of bacteria and archaea, *Systematic biology*, **50**, 470-478.
- DeLong, E.F., Wickham, G.S. and Pace, N.R. (1989) Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells, *Science*, **243**, 1360-1363.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, *Applied and environmental microbiology*, **72**, 5069-5072.
- DeSantis, T.Z., Jr., Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M., Phan, R. and Andersen, G.L. (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes, *Nucleic acids research*, **34**, W394-399.
- Dong, Y., Glasner, J.D., Blattner, F.R. and Triplett, E.W. (2001) Genomic interspecies microarray hybridization: rapid discovery of three thousand genes in the maize endophyte, *Klebsiella pneumoniae* 342, by microarray hybridization with *Escherichia coli* K-12 open reading frames, *Applied and environmental microbiology*, **67**, 1911-1921.
- Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution, *Nature*, **284**, 601-603.
- Dumont, M.G. and Murrell, J.C. (2005) Stable isotope probing - linking microbial identity to function, *Nature reviews*, **3**, 499-504.
- Dumont, M.G., Radajewski, S.M., Miguez, C.B., McDonald, I.R. and Murrell, J.C. (2006) Identification of a complete methane monooxygenase operon from soil by combining stable isotope probing and metagenomic analysis, *Environmental microbiology*, **8**, 1240-1250.
- Dunbar, J., Takala, S., Barns, S.M., Davis, J.A. and Kuske, C.R. (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning, *Applied and environmental microbiology*, **65**, 1662-1669.
- Dunbar, J., Ticknor, L.O. and Kuske, C.R. (2000) Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis, *Applied and environmental microbiology*, **66**, 2943-2950.

- El Fantroussi, S., Urakawa, H., Bernhard, A.E., Kelly, J.J., Noble, P.A., Smidt, H., Yershov, G.M. and Stahl, D.A. (2003) Direct profiling of environmental microbial populations by thermal dissociation analysis of native rRNAs hybridized to oligonucleotide microarrays, *Applied and environmental microbiology*, **69**, 2377-2382.
- Fan, J.B., Gunderson, K.L., Bibikova, M., Yeakley, J.M., Chen, J., Wickham Garcia, E., Lebruska, L.L., Laurent, M., Shen, R. and Barker, D. (2006) Illumina universal bead arrays, *Methods in enzymology*, **410**, 57-73.
- Farrelly, V., Rainey, F.A. and Stackebrandt, E. (1995) Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species, *Applied and environmental microbiology*, **61**, 2798-2801.
- Ferrari, B.C., Binnerup, S.J. and Gillings, M. (2005) Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria, *Applied and environmental microbiology*, **71**, 8714-8720.
- Forney, L.J., Zhou, X. and Brown, C.J. (2004) Molecular microbial ecology: land of the one-eyed king, *Current opinion in microbiology*, **7**, 210-220.
- Frohlich, J. and Konig, H. (2000) New techniques for isolation of single prokaryotic cells, *FEMS microbiology reviews*, **24**, 567-572.
- Fuchs, B.M., Glockner, F.O., Wulf, J. and Amann, R. (2000) Unlabeled helper oligonucleotides increase the in situ accessibility to 16S rRNA of fluorescently labeled oligonucleotide probes, *Applied and environmental microbiology*, **66**, 3603-3607.
- Fuchs, B.M., Syutsubo, K., Ludwig, W. and Amann, R. (2001) In situ accessibility of Escherichia coli 23S rRNA to fluorescently labeled oligonucleotide probes, *Applied and environmental microbiology*, **67**, 961-968.
- Fuchs, B.M., Wallner, G., Beisker, W., Schwippl, I., Ludwig, W. and Amann, R. (1998) Flow cytometric analysis of the in situ accessibility of Escherichia coli 16S rRNA for fluorescently labeled oligonucleotide probes, *Applied and environmental microbiology*, **64**, 4973-4982.
- Gallagher, E., McGuinness, L., Phelps, C., Young, L.Y. and Kerkhof, L.J. (2005) ¹³C-carrier DNA shortens the incubation time needed to detect benzoate-utilizing denitrifying bacteria by stable-isotope probing, *Applied and environmental microbiology*, **71**, 5192-5196.
- Gans, J., Wolinsky, M. and Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil, *Science*, **309**, 1387-1390.

- Gentry, T.J., Wickham, G.S., Schadt, C.W., He, Z. and Zhou, J. (2006) Microarray applications in microbial ecology research, *Microb Ecol*, **52**, 159-175.
- Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L. and Swings, J. (2005) Opinion: Re-evaluating prokaryotic species, *Nature reviews*, **3**, 733-739.
- Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman, R.M. and Handelsman, J. (2002) Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA, *Applied and environmental microbiology*, **68**, 4301-4306.
- Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T. and et al. (1989) Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes, *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 6661-6665.
- Gogarten, J.P. and Townsend, J.P. (2005) Horizontal gene transfer, genome innovation and evolution, *Nature reviews*, **3**, 679-687.
- Goh, E.B., Yim, G., Tsui, W., McClure, J., Surette, M.G. and Davies, J. (2002) Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics, *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 17025-17030.
- Green, B.D. and Keller, M. (2006) Capturing the uncultivated majority, *Current opinion in biotechnology*, **17**, 236-240.
- Guschin, D.Y., Mobarry, B.K., Proudnikov, D., Stahl, D.A., Rittmann, B.E. and Mirzabekov, A.D. (1997) Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology, *Applied and environmental microbiology*, **63**, 2397-2402.
- Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms, *Microbiol Mol Biol Rev*, **68**, 669-685.
- Hill, J.E., Penny, S.L., Crowell, K.G., Goh, S.H. and Hemmingsen, S.M. (2004) cpnDB: a chaperonin sequence database, *Genome research*, **14**, 1669-1675.
- Holben, W.E., Jansson, J.K., Chelm, B.K. and Tiedje, J.M. (1988) DNA Probe Method for the Detection of Specific Microorganisms in the Soil Bacterial Community, *Applied and environmental microbiology*, **54**, 703-711.
- Horner-Devine, M.C., Lage, M., Hughes, J.B. and Bohannan, B.J. (2004) A taxa-area relationship for bacteria, *Nature*, **432**, 750-753.

- Huber, R., Burggraf, S., Mayer, T., Barns, S.M., Rossnagel, P. and Stetter, K.O. (1995) Isolation of a hyperthermophilic archaeum predicted by in situ RNA analysis, *Nature*, **376**, 57-58.
- Hurt, R.A., Qiu, X., Wu, L., Roh, Y., Palumbo, A.V., Tiedje, J.M. and Zhou, J. (2001) Simultaneous recovery of RNA and DNA from soils and sediments, *Applied and environmental microbiology*, **67**, 4495-4503.
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. and Miyata, T. (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes, *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 9355-9359.
- Janda, J.M. and Abbott, S.L. (2002) Bacterial identification for publication: when is enough enough?, *Journal of clinical microbiology*, **40**, 1887-1891.
- Joseph, S.J., Hugenholtz, P., Sangwan, P., Osborne, C.A. and Janssen, P.H. (2003) Laboratory cultivation of widespread and previously uncultured soil bacteria, *Applied and environmental microbiology*, **69**, 7210-7215.
- Kaeberlein, T., Lewis, K. and Epstein, S.S. (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment, *Science*, **296**, 1127-1129.
- Kasai, Y., Takahata, Y., Manefield, M. and Watanabe, K. (2006) RNA-based stable isotope probing and isolation of anaerobic benzene-degrading bacteria from gasoline-contaminated groundwater, *Applied and environmental microbiology*, **72**, 3586-3592.
- Kent, A.D., Smith, D.J., Benson, B.J. and Triplett, E.W. (2003) Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities, *Applied and environmental microbiology*, **69**, 6768-6776.
- Konstantinidis, K.T., Ramette, A. and Tiedje, J.M. (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers, *Applied and environmental microbiology*, **72**, 7286-7293.
- Konstantinidis, K.T. and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes, *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 2567-2572.
- Kostic, T., Weilharter, A., Rubino, S., Delogu, G., Uzzau, S., Rudi, K., Sessitsch, A. and Bodrossy, L. (2007) A microbial diagnostic microarray technique for the sensitive detection and identification of pathogenic bacteria in a background of nonpathogens, *Analytical biochemistry*, **360**, 244-254.

- Kostic, T., Weilharter, A., Sessitsch, A. and Bodrossy, L. (2005) High-sensitivity, polymerase chain reaction-free detection of microorganisms and their functional genes using 70-mer oligonucleotide diagnostic microarray, *Analytical biochemistry*, **346**, 333-335.
- Kreader, C.A. (1996) Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein, *Applied and environmental microbiology*, **62**, 1102-1106.
- Krimmer, V., Merkert, H., von Eiff, C., Frosch, M., Eulert, J., Lohr, J.F., Hacker, J. and Ziebuhr, W. (1999) Detection of *Staphylococcus aureus* and *Staphylococcus epidermidis* in clinical samples by 16S rRNA-directed in situ hybridization, *Journal of clinical microbiology*, **37**, 2667-2673.
- Kumar, Y., Westram, R., Behrens, S., Fuchs, B., Glockner, F.O., Amann, R., Meier, H. and Ludwig, W. (2005) Graphical representation of ribosomal RNA probe accessibility data using ARB software package, *BMC bioinformatics*, **6**, 61.
- Kurland, C.G., Collins, L.J. and Penny, D. (2006) Genomics and the irreducible nature of eukaryote cells, *Science*, **312**, 1011-1014.
- Leadbetter, J.R. (2003) Cultivation of recalcitrant microbes: cells are alive, well and revealing their secrets in the 21st century laboratory, *Current opinion in microbiology*, **6**, 274-281.
- Lecompte, O., Ripp, R., Thierry, J.C., Moras, D. and Poch, O. (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale, *Nucleic acids research*, **30**, 5382-5390.
- Lerat, E., Daubin, V., Ochman, H. and Moran, N.A. (2005) Evolutionary origins of genomic repertoires in bacteria, *PLoS biology*, **3**, e130.
- Lin, L.H., Wang, P.L., Rumble, D., Lippmann-Pipke, J., Boice, E., Pratt, L.M., Sherwood Lollar, B., Brodie, E.L., Hazen, T.C., Andersen, G.L., DeSantis, T.Z., Moser, D.P., Kershaw, D. and Onstott, T.C. (2006) Long-term sustainability of a high-energy, low-diversity crustal biome, *Science*, **314**, 479-482.
- Lindstedt, B.A. (2005) Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria, *Electrophoresis*, **26**, 2567-2582.
- Liolios, K., Tavernarakis, N., Hugenholtz, P. and Kyrpides, N.C. (2006) The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide, *Nucleic acids research*, **34**, D332-334.
- Liu, W.T., Marsh, T.L., Cheng, H. and Forney, L.J. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA, *Applied and environmental microbiology*, **63**, 4516-4522.

- Loy, A. and Bodrossy, L. (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays, *Clinica chimica acta; international journal of clinical chemistry*, **363**, 106-119.
- Loy, A., Maixner, F., Wagner, M. and Horn, M. (2006) probeBase--an online resource for rRNA-targeted oligonucleotide probes: new features 2007, *Nucleic acids research*.
- Loy, A., Schulz, C., Lucker, S., Schopfer-Wendels, A., Stoecker, K., Baranyi, C., Lehner, A. and Wagner, M. (2005) 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales", *Applied and environmental microbiology*, **71**, 1373-1386.
- Lozupone, C., Hamady, M. and Knight, R. (2006) UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context, *BMC bioinformatics*, **7**, 371.
- Lozupone, C. and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities, *Applied and environmental microbiology*, **71**, 8228-8235.
- Lueders, T., Manefield, M. and Friedrich, M.W. (2004) Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients, *Environmental microbiology*, **6**, 73-78.
- Luna, G.M., Dell'Anno, A. and Danovaro, R. (2006) DNA extraction procedure: a critical issue for bacterial diversity assessment in marine sediments, *Environmental microbiology*, **8**, 308-320.
- Lunn, M., Sloan, W.T. and Curtis, T.P. (2004) Estimating bacterial diversity from clone libraries with flat rank abundance distributions, *Environmental microbiology*, **6**, 1081-1085.
- Mahenthiralingam, E., Baldwin, A., Drevinek, P., Vanlaere, E., Vandamme, P., Lipuma, J.J. and Dowson, C.G. (2006) Multilocus sequence typing breathes life into a microbial metagenome, *PLoS ONE*, **1**, e17.
- Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. and Spratt, B.G. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms, *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 3140-3145.
- Manefield, M., Whiteley, A.S., Griffiths, R.I. and Bailey, M.J. (2002) RNA stable isotope probing, a novel means of linking microbial community function to phylogeny, *Applied and environmental microbiology*, **68**, 5367-5373.

- Marras, S.A., Kramer, F.R. and Tyagi, S. (2002) Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes, *Nucleic acids research*, **30**, e122.
- Martin, W. (2005) Archaeobacteria (Archaea) and the origin of the eukaryotic nucleus, *Current opinion in microbiology*, **8**, 630-637.
- Martin, W. and Koonin, E.V. (2006) Introns and the origin of nucleus-cytosol compartmentalization, *Nature*, **440**, 41-45.
- Mathews, D.H., Burkard, M.E., Freier, S.M., Wyatt, J.R. and Turner, D.H. (1999) Predicting oligonucleotide affinity to nucleic acid targets, *RNA (New York, N.Y)*, **5**, 1458-1469.
- Matveeva, O.V., Shabalina, S.A., Nemtsov, V.A., Tsodikov, A.D., Gesteland, R.F. and Atkins, J.F. (2003) Thermodynamic calculations and statistical correlations for oligo-probes design, *Nucleic acids research*, **31**, 4211-4217.
- McKeown, B.J. (1994) An acetylated (nuclease-free) bovine serum albumin in a PCR buffer inhibits amplification, *BioTechniques*, **17**, 246-248.
- Militon, C., Rimour, S., Missaoui, M., Biderre, C., Barra, V., Hill, D., Mone, A., Gagne, G., Meier, H., Peyretailade, E. and Peyret, P. (2007) PhylArray: Phylogenetic Probe Design Algorithm for MicroArray, *Bioinformatics*.
- Miller, K.M., Ming, T.J., Schulze, A.D. and Withler, R.E. (1999) Denaturing gradient gel electrophoresis (DGGE): a rapid and sensitive technique to screen nucleotide sequence variation in populations, *BioTechniques*, **27**, 1016-1018, 1020-1012, 1024 passim.
- Miteva, V.I. and Brenchley, J.E. (2005) Detection and isolation of ultrasmall microorganisms from a 120,000-year-old Greenland glacier ice core, *Applied and environmental microbiology*, **71**, 7806-7818.
- Morris, S.A., Radajewski, S., Willison, T.W. and Murrell, J.C. (2002) Identification of the functionally active methanotroph population in a peat soil microcosm by stable-isotope probing, *Applied and environmental microbiology*, **68**, 1446-1453.
- Moter, A. and Gobel, U.B. (2000) Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms, *Journal of microbiological methods*, **41**, 85-112.
- Nakano, Y., Takeshita, T. and Yamashita, Y. (2006) TRFMA: a web-based tool for terminal restriction fragment length polymorphism analysis based on molecular weight, *Bioinformatics (Oxford, England)*, **22**, 1788-1789.
- Neufeld, J.D., Dumont, M.G., Vohra, J. and Murrell, J.C. (2006) Methodological Considerations for the Use of Stable Isotope Probing in Microbial Ecology, *Microb Ecol.*

- Nocker, A. and Camper, A.K. (2006) Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide, *Applied and environmental microbiology*, **72**, 1997-2004.
- Nocker, A., Cheung, C.Y. and Camper, A.K. (2006) Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells, *Journal of microbiological methods*, **67**, 310-320.
- Noonan, J.P., Hofreiter, M., Smith, D., Priest, J.R., Rohland, N., Rabeder, G., Krause, J., Deter, J.C., Paabo, S. and Rubin, E.M. (2005) Genomic sequencing of Pleistocene cave bears, *Science*, **309**, 597-599.
- Nusslein, K. and Tiedje, J.M. (1998) Characterization of the dominant and rare members of a young Hawaiian soil bacterial community with small-subunit ribosomal DNA amplified from DNA fractionated on the basis of its guanine and cytosine composition, *Applied and environmental microbiology*, **64**, 1283-1289.
- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms, *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 2766-2770.
- Orphan, V.J., House, C.H., Hinrichs, K.U., McKeegan, K.D. and DeLong, E.F. (2001) Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis, *Science*, **293**, 484-487.
- Ouverney, C.C. and Fuhrman, J.A. (1999) Combined microautoradiography-16S rRNA probe technique for determination of radioisotope uptake by specific microbial cell types in situ, *Applied and environmental microbiology*, **65**, 1746-1752.
- Pedros-Alio, C. (2006) Marine microbial diversity: can it be determined?, *Trends in microbiology*, **14**, 257-263.
- Pedros-Alio, C. (2007) Ecology. Dipping into the rare biosphere, *Science*, **315**, 192-193.
- Peplies, J., Lachmund, C., Glockner, F.O. and Manz, W. (2006) A DNA microarray platform based on direct detection of rRNA for characterization of freshwater sediment-related prokaryotic communities, *Applied and environmental microbiology*, **72**, 4829-4838.
- Peplies, J., Lau, S.C., Pernthaler, J., Amann, R. and Glockner, F.O. (2004) Application and validation of DNA microarrays for the 16S rRNA-based analysis of marine bacterioplankton, *Environmental microbiology*, **6**, 638-645.

- Pernthaler, A. and Amann, R. (2004) Simultaneous fluorescence in situ hybridization of mRNA and rRNA in environmental bacteria, *Applied and environmental microbiology*, **70**, 5426-5433.
- Pernthaler, A., Pernthaler, J. and Amann, R. (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria, *Applied and environmental microbiology*, **68**, 3094-3101.
- Platt, S., Pichon, B., George, R. and Green, J. (2006) A bioinformatics pipeline for high-throughput microbial multilocus sequence typing (MLST) analyses, *Clin Microbiol Infect*, **12**, 1144-1146.
- Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Nolte, A.W., Sonnenberg, R., Staehler, P., Beier, M. and Tautz, D. (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted, *Nucleic acids research*, **34**, e66.
- Quispel, A. (1998) Lourens G. M. Baas Becking (1895-1963). Inspirator for many (micro)biologists, *Int Microbiol*, **1**, 69-72.
- Radajewski, S., Ineson, P., Parekh, N.R. and Murrell, J.C. (2000) Stable-isotope probing as a tool in microbial ecology, *Nature*, **403**, 646-649.
- Radajewski, S., Webster, G., Reay, D.S., Morris, S.A., Ineson, P., Nedwell, D.B., Prosser, J.I. and Murrell, J.C. (2002) Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing, *Microbiology (Reading, England)*, **148**, 2331-2342.
- Ranjard, L., Lejon, D.P., Mougél, C., Schehrer, L., Merdinoglu, D. and Chaussod, R. (2003) Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities, *Environmental microbiology*, **5**, 1111-1120.
- Ranjard, L., Poly, F., Lata, J.C., Mougél, C., Thioulouse, J. and Nazaret, S. (2001) Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: biological and methodological variability, *Applied and environmental microbiology*, **67**, 4479-4487.
- Rappe, M.S., Connon, S.A., Vergin, K.L. and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade, *Nature*, **418**, 630-633.
- Ray, A. and Norden, B. (2000) Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future, *Faseb J*, **14**, 1041-1060.
- Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X. and Zhou, J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-

- mer oligonucleotide microarrays, *Applied and environmental microbiology*, **70**, 4303-4317.
- Rimour, S., Hill, D., Militon, C. and Peyret, P. (2005) GoArrays: highly dynamic and efficient microarray probe design, *Bioinformatics (Oxford, England)*, **21**, 1094-1103.
- Rivera, M.C., Jain, R., Moore, J.E. and Lake, J.A. (1998) Genomic evidence for two functionally distinct gene classes, *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 6239-6244.
- Rivera, M.C. and Lake, J.A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes, *Nature*, **431**, 152-155.
- Rossello-Mora, R. and Amann, R. (2001) The species concept for prokaryotes, *FEMS microbiology reviews*, **25**, 39-67.
- Sait, M., Hugenholtz, P. and Janssen, P.H. (2002) Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys, *Environmental microbiology*, **4**, 654-666.
- Schloss, P.D. and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness, *Applied and environmental microbiology*, **71**, 1501-1506.
- Schloss, P.D. and Handelsman, J. (2005) Metagenomics for studying unculturable microorganisms: cutting the Gordian knot, *Genome Biol*, **6**, 229.
- Schloss, P.D. and Handelsman, J. (2006) Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures, *Applied and environmental microbiology*, **72**, 6773-6779.
- Schloss, P.D. and Handelsman, J. (2006) Introducing TreeClimber, a test to compare microbial community structures, *Applied and environmental microbiology*, **72**, 2379-2384.
- Schwieger, F. and Tebbe, C.C. (1998) A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis, *Applied and environmental microbiology*, **64**, 4870-4876.
- Sebat, J.L., Colwell, F.S. and Crawford, R.L. (2003) Metagenomic profiling: microarray analysis of an environmental genomic library, *Applied and environmental microbiology*, **69**, 4927-4934.
- Seow, K.T., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C.R. and Davies, J. (1997) A study of iterative type II polyketide synthases, using bacterial genes cloned from

- soil DNA: a means to access and use genes from uncultured microorganisms, *Journal of bacteriology*, **179**, 7360-7368.
- Shaver, Y.J., Nagpal, M.L., Fox, K.F., Rudner, R. and Fox, A. (2001) Variation in 16S-23S rRNA intergenic spacer regions among *Bacillus subtilis* 168 isolates, *Molecular microbiology*, **42**, 101-109.
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome, *Science*, **309**, 1728-1732.
- Simu, K. and Hagstrom, A. (2004) Oligotrophic bacterioplankton with a novel single-cell life strategy, *Applied and environmental microbiology*, **70**, 2445-2451.
- Singleton, D.R., Powell, S.N., Sangaiah, R., Gold, A., Ball, L.M. and Aitken, M.D. (2005) Stable-isotope probing of bacteria capable of degrading salicylate, naphthalene, or phenanthrene in a bioreactor treating contaminated soil, *Applied and environmental microbiology*, **71**, 1202-1209.
- Small, J., Call, D.R., Brockman, F.J., Straub, T.M. and Chandler, D.P. (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays, *Applied and environmental microbiology*, **67**, 4708-4716.
- Smith, C.J., Danilowicz, B.S., Clear, A.K., Costello, F.J., Wilson, B. and Meijer, W.G. (2005) T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles, *FEMS microbiology ecology*, **54**, 375-380.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. and Herndl, G.J. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere", *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12115-12120.
- Spiegelman, D., Whissell, G. and Greer, C.W. (2005) A survey of the methods for the characterization of microbial consortia and communities, *Canadian journal of microbiology*, **51**, 355-386.
- Stackebrandt, E. and Fiedler, F. (1979) DNA-DNA homology studies among strains of *Arthrobacter* and *Brevibacterium*, *Archives of microbiology*, **120**, 289-295.
- Stager, C.E. and Davis, J.R. (1992) Automated systems for identification of microorganisms, *Clinical microbiology reviews*, **5**, 302-327.
- Staley, J.T. and Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats, *Annual review of microbiology*, **39**, 321-346.

- Stevenson, B.S., Eichorst, S.A., Wertz, J.T., Schmidt, T.M. and Breznak, J.A. (2004) New strategies for cultivation and detection of previously uncultured microbes, *Applied and environmental microbiology*, **70**, 4748-4755.
- Suzuki, M., Rappe, M.S. and Giovannoni, S.J. (1998) Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity, *Applied and environmental microbiology*, **64**, 4522-4529.
- Torsvik, V., Goksoyr, J. and Daae, F.L. (1990) High diversity in DNA of soil bacteria, *Applied and environmental microbiology*, **56**, 782-787.
- Torsvik, V. and Ovreas, L. (2002) Microbial diversity and function in soil: from genes to ecosystems, *Current opinion in microbiology*, **5**, 240-245.
- Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P. and Rubin, E.M. (2005) Comparative metagenomics of microbial communities, *Science*, **308**, 554-557.
- Tyson, G.W. and Banfield, J.F. (2005) Cultivating the uncultivated: a community genomics perspective, *Trends in microbiology*, **13**, 411-415.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. and Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment, *Nature*, **428**, 37-43.
- Uchiyama, T., Abe, T., Ikemura, T. and Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes, *Nature biotechnology*, **23**, 88-93.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H. and Smith, H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea, *Science*, **304**, 66-74.
- von Mering, C., Hugenholtz, P., Raes, J., Tringe, S.G., Doerks, T., Jensen, L.J., Ward, N. and Bork, P. (2007) Quantitative Phylogenetic Assessment of Microbial Communities in Diverse Environments, *Science*.
- von Wintzingerode, F., Gobel, U.B. and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis, *FEMS microbiology reviews*, **21**, 213-229.

- Wallner, G., Amann, R. and Beisker, W. (1993) Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms, *Cytometry*, **14**, 136-143.
- Whiteley, A.S., Manefield, M. and Lueders, T. (2006) Unlocking the 'microbial black box' using RNA-based stable isotope probing technologies, *Current opinion in biotechnology*, **17**, 67-71.
- Whitman, W.B., Coleman, D.C. and Wiebe, W.J. (1998) Prokaryotes: the unseen majority, *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 6578-6583.
- Willerslev, E., Hansen, A.J., Ronn, R., Brand, T.B., Barnes, I., Wiuf, C., Gilichinsky, D., Mitchell, D. and Cooper, A. (2004) Long-term persistence of bacterial DNA, *Curr Biol*, **14**, R9-10.
- Wilson, K.H., Wilson, W.J., Radosevich, J.L., DeSantis, T.Z., Viswanathan, V.S., Kuczmarski, T.A. and Andersen, G.L. (2002) High-density microarray of small-subunit ribosomal DNA probes, *Applied and environmental microbiology*, **68**, 2535-2541.
- Woese, C.R. (1987) Bacterial evolution, *Microbiological reviews*, **51**, 221-271.
- Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms, *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5088-5090.
- Wu, L., Thompson, D.K., Liu, X., Fields, M.W., Bagwell, C.E., Tiedje, J.M. and Zhou, J. (2004) Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications, *Environmental science & technology*, **38**, 6775-6782.
- Wu, X.Y., Walker, M.J., Hornitzky, M. and Chin, J. (2006) Development of a group-specific PCR combined with ARDRA for the identification of *Bacillus* species of environmental significance, *Journal of microbiological methods*, **64**, 107-119.
- Yang, Z.H., Xiao, Y., Zeng, G.M., Xu, Z.Y. and Liu, Y.S. (2006) Comparison of methods for total community DNA extraction and purification from compost, *Appl Microbiol Biotechnol*.
- Yilmaz, L.S. and Noguera, D.R. (2004) Mechanistic approach to the problem of hybridization efficiency in fluorescent in situ hybridization, *Applied and environmental microbiology*, **70**, 7126-7139.

- Yilmaz, L.S., Okten, H.E. and Noguera, D.R. (2006) Making all parts of the 16S rRNA of *Escherichia coli* accessible in situ to single DNA oligonucleotides, *Applied and environmental microbiology*, **72**, 733-744.
- Yost, C.K. and Nattress, F.M. (2002) Molecular typing techniques to characterize the development of a lactic acid bacteria community on vacuum-packaged beef, *International journal of food microbiology*, **72**, 97-105.
- Yu, Y., Breitbart, M., McNairnie, P. and Rohwer, F. (2006) FastGroupII: a web-based bioinformatics platform for analyses of large 16S rDNA libraries, *BMC bioinformatics*, **7**, 57.
- Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E.J., Short, J.M. and Keller, M. (2002) Cultivating the uncultured, *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 15681-15686.
- Zengler, K., Walcher, M., Clark, G., Haller, I., Toledo, G., Holland, T., Mathur, E.J., Woodnutt, G., Short, J.M. and Keller, M. (2005) High-throughput cultivation of microorganisms using microcapsules, *Methods in enzymology*, **397**, 124-130.
- Zhou, J., Bruns, M.A. and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition, *Applied and environmental microbiology*, **62**, 316-322.
- Zhou, X., Wu, L. and Zhou, J. (2004) Fabrication of DNA microarrays on nanoengineered polymeric ultrathin film prepared by self-assembly of polyelectrolyte multilayers, *Langmuir*, **20**, 8877-8885.
- Zwirgmaier, K., Ludwig, W. and Schleifer, K.H. (2004) Recognition of individual genes in a single bacterial cell by fluorescence in situ hybridization--RING-FISH, *Molecular microbiology*, **51**, 89-96.

C. DISCUSSION.

Le développement des techniques moléculaires, initié depuis plusieurs décennies, a permis des avancées considérables. Ces techniques ont permis de s'affranchir des étapes culturelles classiques qui limitaient notre vision du monde microbien en occultant une grande partie des microorganismes : la fraction non cultivée. Outre une meilleure caractérisation de la diversité microbienne au niveau des différents environnements complexes, elles ont permis d'identifier de nouvelles voies métaboliques et de nouvelles fonctions qui ont permis d'améliorer nos connaissances sur le fonctionnement de ces organismes.

Les dernières techniques développées permettent d'étudier encore plus largement les microorganismes d'une part en criblant à haut-débit la majorité des microorganismes (biopuces ADN), d'autre part en explorant la globalité de l'information génétique contenue au sein des communautés microbiennes (métagénomique, métatranscriptomique) et enfin en liant la structure des populations microbiennes à leurs fonctions (Isotope array, MAR-FISH, SIP). La puissance de ces nouvelles techniques offre donc des potentialités qui permettront sans aucun doute d'élargir encore plus nos connaissances sur ces organismes microscopiques. Cependant, même si ces nouvelles techniques sont enthousiasmantes, il n'existe pas de technique qui ne présente pas de limites ou d'inconvénients. Aussi, en fonction des objectifs de l'étude, de la complexité des communautés microbiennes et de l'environnement étudié, plusieurs techniques de cette « boîte à outils » moléculaire peuvent être utilisés de manière complémentaire afin d'obtenir les informations désirées. De plus, il est nécessaire, en complément de ces études moléculaires, de ne pas sous-estimer les approches culturelles qui permettront, à terme, d'accéder à des connaissances sur l'écophysiologie des microorganismes qui est nécessaire à la compréhension du fonctionnement des microorganismes dans leur contexte environnemental.

PARTIE III

ÉTUDE DE LA DYNAMIQUE DES COMMUNAUTÉS BACTÉRIENNES AU COURS D'UN PROCÉDÉ DE BIOREMÉDIATION D'UN SOL POLLUÉ PAR DES HYDROCARBURES

A. CONTEXTE DE L'ETUDE.

Comme nous l'avons vu précédemment dans la partie synthèse bibliographique, les activités humaines sont à l'origine de nombreuses pollutions au niveau des écosystèmes et plus particulièrement au niveau des sols. Dans ce type d'environnement, les principaux polluants décrits sont les hydrocarbures. Ceux-ci peuvent présenter des risques toxicologiques, mutagènes voire même cancérigènes pour de nombreux organismes dont l'Homme (Spencer *et al.*, 2002). Ainsi, afin de préserver la santé humaine et environnementale, il est nécessaire de réduire les risques d'exposition et de contamination en dégradant les polluants en produit moins toxiques voire même inoffensifs. Cette décontamination peut être effectuée à l'aide de techniques physiques et/ou chimiques (Khan *et al.*, 2004; Kalin, 2004). Classiquement utilisées, elles permettent l'extraction, la destruction et/ou l'immobilisation des polluants mais elles peuvent être très invasives pour les écosystèmes traités et onéreuses à mettre en place. Une alternative à ces techniques est la bioremediation qui utilise les capacités des microorganismes à dépolluer les sols de façon naturelle grâce à des machineries enzymatiques adaptées (Serrano *et al.*, 2007). Afin de mieux comprendre les mécanismes impliqués dans cette décontamination naturelle et d'améliorer les rendements d'épuration, il est nécessaire d'obtenir des données sur les communautés bactériennes impliquées dans ces procédés. Ces connaissances contribueront alors à une meilleure compréhension des mécanismes d'adaptation des microorganismes suite à des perturbations.

B. OBJECTIFS DE L'ETUDE.

Les objectifs de nos travaux de recherche ont par conséquent été de caractériser les microorganismes présents au niveau d'un sol alluvial pollué par des hydrocarbures aliphatiques durant le procédé de bioremédiation mis en place afin de décontaminer cet environnement. Ce procédé consiste en une biostimulation des communautés bactériennes du sol pollué par l'apport continu d'oxygène en réacteurs pilotes (ligne BIOrem, Biobasic Environnement[®]). La composition et la dynamique des communautés bactériennes de ce sol pollué ont été suivies au cours du procédé de dépollution (au début, après 12 mois et 24 mois de traitement). Afin d'atteindre ces objectifs, une approche de clonage/séquençage des séquences de la petite sous unité d'ARNr (16S) a été choisie. Pour chacun des points de la

cinétique, deux banques de clones ADNr 16S et ARNr 16S ont été générées. L'analyse combinée de ces deux types de banques a permis la caractérisation des communautés métaboliquement actives *versus* les communautés quiescentes durant le procédé de bioremédiation et ainsi de caractériser les microorganismes potentiellement impliqués dans la dégradation des hydrocarbures.

C. PRINCIPAUX RESULTATS OBTENUS.

La création de banques d'environ 400 séquences ribosomiques a permis d'obtenir une bonne image en terme de représentativité des communautés bactériennes présentes au niveau du sol pollué durant la cinétique de bioremédiation. En effet, 74,0 à 88,3 % de leur diversité serait couverte grâce à cet effort de clonage/séquençage. Le regroupement des séquences en unités taxonomiques (97% d'homologie) a montré la présence continue ou discontinue de 534 phylotypes au travers des différentes banques dont seuls 9,4 % sont communs aux banques ADNr 16S et ARNr 16S. Cette observation met en évidence une reconnaissance différentielle des groupes bactériens en fonction du type de molécules utilisé (ADN ou ARN). En effet, l'utilisation des matrices ADN pour l'amplification de l'ADNr 16S permet l'identification des microorganismes quiescents et des ADN extracellulaires en plus des microorganismes métaboliquement actifs. Cependant, le fait que la majorité des séquences appartiennent à ces quelques phylotypes communs et que la plupart des phylotypes dominants se retrouvent dans les deux types de banques indique qu'il existe tout de même une bonne correspondance entre-elles.

L'analyse phylogénétique des 534 phylotypes a montré leur appartenance à 19 divisions bactériennes dont la majorité sont typiques des écosystèmes sols. Les divisions les plus fortement représentées au sein des banques sont les *γ-Proteobacteria*, les *Actinobacteria* et les *α-Proteobacteria*. Au début du procédé de bioremédiation, les populations microbiennes du sol sont composées en majorité de *γ-Proteobacteria* et d'*α-Proteobacteria*. A une échelle phylogénétique plus fine, ce sont les microorganismes appartenant aux *Pseudomonadaceae/otu_2452*, aux *B91-12/otu_2175* et aux *Lysobacter sp./otu_2521* qui dominent ce qui suggère leur adaptation à la présence des hydrocarbures aliphatiques au niveau du sol pollué. Durant le procédé de bioremédiation, les communautés bactériennes semblent tendre vers une diminution des *Proteobacteria* contrebalancée par une augmentation des bactéries appartenant aux *Actinobacteria* : les *Microthrixineae* et ce particulièrement

après 24 mois de traitement. Outre l'évolution de plusieurs phylotypes bactériens, ces travaux ont aussi permis de mettre en évidence la présence d'un « core-set » bactérien.

Dynamics of the bacterial communities during a pilot-scale bioremediation process of a hydrocarbon-contaminated soil identified by SSU rDNA and SSU rRNA libraries.

Cécile Militon¹, Delphine Boucher¹, Julien Troquet², Eric Peyretailade¹ and Pierre Peyret¹*

¹ Génomique Intégrée des Interactions Microbiennes, Laboratoire de Biologie des Protistes, UMR CNRS 6023, Blaise Pascal University, 24 avenue des Landais, Campus des Cézeaux, 63177 Aubière Cedex, France ; ² Biobasic Environnement, Biopôle Clermont-Limagne, 63360 Saint-Beauzire. * For correspondence. E-mail: pierre.peyret@univ-bpclermont.fr fax : +33 (0) 4 7340 76 70 phone : +33 (0) 4 7340 76 04.

Summary

In order to identify and follow the dynamic of the microbial communities implied in a pilot-scale decontamination of a hydrocarbon-polluted soil, SSU rDNA and SSU rRNA libraries have been constructed during a two-year process. The analysis of 2,213 full-length sequences has allowed the getting of an interesting overview of the diversity covering from 74.0% to 88.3% of the bacterial populations. The clustering of all sequences has shown the presence of 534 phylotypes belonging to nineteen bacterial phyla. The comparison of six libraries has revealed that only 9.4% of these phylotypes which correspond to 58.5% of the overall sequences are common to both DNA and RNA libraries. The majority of the non-shared phylotypes are only present in the SSU rDNA libraries that show a greater phylogenetic diversity because of the presence of phylotypes other than those metabolically active. The following of the bioremediation process has shown a shift of the bacterial populations from the *γ-Proteobacteria* to the *Actinobacteria* between the polluted soils untreated and those at the end of treatment. Additionally, the sequence analysis revealed the presence of a bacterial “core-set” of 37 and 19 phylotypes (respectively identified by the SSU rDNA and the SSU rRNA) represented by nearly half of sequenced clones.

Introduction

Aliphatic hydrocarbons are widespread in soil subsurface as a result of industrial activity and accidental leakages during pollutants transport and storage. Their impacts on human and environmental health have necessitated the development of sites restoration processes. The technologies used conventionally include physical and chemical techniques such as solidification/stabilization, vitrification, incineration, oxidation/reduction, thermal or flushing extraction or excavation (Khan *et al.*, 2004). However, these techniques are expensive, labor-intensive and responsible for the emission of components more or less toxic. An alternative is the use of the bioremediation. This process uses natural capabilities of living microorganisms which can biodegrade pollutants into less toxic or no-toxic products. The bioremediation, identified as one of the most promising remediation technologies, often relies exclusively on the metabolic activity of these organisms (Paul *et al.*, 2005). Indeed, the environmental microorganisms can naturally acquire adaptation capacity to a wide range of xenobiotic pollutants. Thus, their abilities to degrade complex substrates, such as hydrocarbons, have been used in recent years to decontaminate various environments (Da Silva and Alvarez, 2004; de Carvalho *et al.*, 2005; Botton and Parsons, 2007; Perfumo *et al.*, 2007; Wang *et al.*, 2007). The improvement of bioremediation strategies requires an exploration of the microbial communities from the polluted soils and specifically those involved in biodegradation. However, the main limitation of these investigations is based on the difficulty of identifying all microbial populations and their dynamics in complex environments like soils. Actually, Gans *et al.* (2005) have estimated that 1 g of surface soil might contain more than one million distinct genomes. Moreover, only a minority of those microorganisms have hitherto been cultivated and characterized (Dunbar *et al.*, 1999). This cultural bias generates an underestimation of the bacterial biodiversity which has forced the development of more suitable investigation methods. During the last decade, cultivation-independent molecular

tools have been developed as an alternative in order to study environmental microbial communities more comprehensively (Amann *et al.*, 1995). These nucleic acid-based methods most often use the information contained in the genes encoding the small subunit ribosomal RNA (SSU rDNA). The small subunit ribosomal RNA (SSU rRNA) can also be used instead of SSU rDNA in order to reflect the metabolically active communities (Nogales *et al.*, 2001; Gremion *et al.*, 2003; Martinez *et al.*, 2006; Akob *et al.*, 2007). These ribosomal sequences are ubiquitous and are structurally and functionally conserved among all Prokaryotes. Actually, their evolution is relatively slow compared to other prokaryotic genes, making them good phylogenetic biomarkers. Moreover, these nucleic acids are composed, on one hand, by conserved regions allowing an easy isolation and, on the other hand, by variable regions allowing the differentiation of taxonomic units. So, these biological molecules are the targets of choice for the development of these kinds of studies (Olsen *et al.*, 1986). The polymerase chain reactions (PCR) combined to the cloning/sequencing approaches has often been used to study the environmental microbial communities (Amann *et al.*, 1995; Pace, 1997; Giovanonni *et al.*, 2000). These molecular techniques allow the generation of numerous sequences which reflect the microorganisms present in the studied environment. In addition, owing to statistical methods derived from indices used by the ecologists to study the diversity of macroorganisms, these data also allow an estimate of the microbial diversity. However, many collections are not large enough to really accurately represent the overall microbial biodiversity of studied environments (Kemp and Aller, 2004). In this paper we report the monitoring of the diversity of the soil microbial populations during a two years pilot-scale bioremediation process from an alluvial soil contaminated with various aliphatic hydrocarbons. The evolution of bacterial communities was followed by the construction of SSU rDNA as well as SSU rRNA clone libraries at each sampling date. The six clones libraries comprise 2,213 full-length ribosomal sequences and allow the characterization of

microorganisms present in the soil during this bioremediation process including those metabolically active which are potentially implied in the biodegradation of the contaminants such as aliphatic hydrocarbons.

Results

Composition of the SSU rDNA and SSU rRNA libraries

The evolution of the microbial communities in a hydrocarbon-polluted soil was analyzed before the beginning of its bioremediation procedure (t_0), after 12 months (t_{+12}) and after 24 months (t_{+24}), corresponding respectively to the middle and to the end of this process. For each date, the bacterial populations were characterized by the generation of a SSU rDNA library (DNA), and metabolically active microorganisms have been identified owing to the creation of a SSU rRNA library (RNA). Sum-total, the six libraries comprise 2,213 full-length sequences: 1,139 belonging to the DNA libraries and 1,074 to those RNA. These sequences have been subdivided in Operational Taxonomic Units (OUT). Each OTU grouping all sequences that share more than 97% sequence similarity. Consequently, in the clone libraries from SSU rDNA, 444 phylotypes were detected and in the one from rRNA, 265. The number of phylotypes was highest in the three clone libraries from rDNA with 152, 133 and 159 phylotypes, respectively, for the date t_0 , t_{+12} and t_{+24} , suggesting a greater bacterial diversity identified with DNA rather than with RNA. Regarding the three RNA libraries, they consist respectively of only 83, 101 and 81 phylotypes at the beginning, in the middle and at the end of the bioremediation process. Phylogenetic analysis distributed clone sequences into 19 bacterial divisions. Most sequences were affiliated with 3 phyla: *γ-Proteobacteria* (28.2 % of the sequences), *Actinobacteria* (27.6 %) and *α-Proteobacteria* (14.8 %).

The comparison of six different libraries has allowed to detect several bacterial divisions present in DNA libraries but not in those RNA: the *Bacteroidetes*, *Nitrospirae*,

Table 1. Phylogenetic groups representation in the libraries obtained from 16S rDNA (DNA) and 16S rRNA (RNA) throughout the pilot-scale bioremediation process of the hydrocarbon polluted soil (t_0 : before the process; t_{+12} : after 12 months; t_{+24} : after 24 months) and characterization of the phylogenetic groups common in the both types of libraries. The values mean to the number of observed OTU_{97%}/number of corresponding clones.

Bacterial division	DNA			RNA			Total	Total without redundancy	Found in both types of libraries
	t_0	t_{+12}	t_{+24}	t_0	t_{+12}	t_{+24}			
<i>Acidobacteria</i>	16/29	18/30	17/50	5/5	6/9	9/16	71/139	59/139	12/69
<i>Actinobacteria</i>	13/26	13/26	29/101	11/98	17/91	27/269	110/611	90/611	17/459
<i>Bacteroidetes</i>	6/17	7/25	-	-	7/15	-	20/57	20/57	7/36
<i>Chlorobi</i>	-	1/2	-	-	-	-	1/2	1/2	-
<i>Chloroflexi</i>	12/19	8/11	24/40	4/14	1/2	5/8	54/94	51/94	9/38
<i>Gemmatimonadetes</i>	12/16	11/18	5/10	2/4	4/5	2/2	36/55	36/55	2/21
<i>Nitrospirae</i>	1/1	-	-	-	-	-	1/1	1/1	-
OP10	-	-	1/1	-	-	-	1/1	1/1	-
<i>Planctomycetes</i>	10/13	6/6	12/17	-	2/2	2/2	32/40	32/40	1/3
α - <i>Proteobacteria</i>	34/68	22/54	33/75	27/42	28/56	22/33	166/328	166/328	31/204
β - <i>Proteobacteria</i>	10/18	10/15	10/24	7/13	4/6	3/8	44/84	44/84	7/48
γ - <i>Proteobacteria</i>	25/148	10/135	13/45	18/181	22/108	4/6	102/623	102/623	20/575
δ - <i>Proteobacteria</i>	7/9	7/18	9/29	7/26	6/31	2/6	38/119	38/119	7/88
SBR1093	1/1	3/7	2/3	-	-	1/2	7/13	7/13	1/7
SPAM	1/1	-	-	2/3	-	-	3/4	3/4	-
TM6	-	-	-	-	-	2/2	2/2	2/2	-
TM7	2/3	2/6	2/7	-	2/5	2/2	10/23	10/23	3/16
<i>Verrucomicrobia</i>	2/4	3/5	1/2	-	1/1	-	7/12	7/12	1/5
WS3	-	2/3	1/1	-	1/1	-	4/5	4/5	1/3
Total	152/373	133/361	159/405	83/386	101/332	81/356	709/2,213	709/2,213	119/1,572

Table 2. Dominant phylotypes present in the SSU rDNA (DNA), in the SSU rRNA (RNA) or in both (2) generated before the bioremediation process (t_0), after 12 months (t_{+12}) and after 24 months (t_{+24}).

Date	Bacterial division	Representative sequence	Length (nt)	Nb of clones	Closest relative according the Hugenholtz's taxonomy	% Identity
t_0	<i>Actinobacteria</i>	CMJA8 (2)	1344	82	Unclassified Microthrixinae; otu_257	93.8%
	α - <i>Proteobacteria</i>	AMMG12 (DNA)	1324	10	<i>Bradyrhizobium</i> sp.; otu_1930	99.3%
	δ - <i>Proteobacteria</i>	CMPF12 (RNA)	1380	16	Unclassified EB1021; otu_2606	88.1%
	γ - <i>Proteobacteria</i>	CMIF10 (2)	1356	96	Unclassified Pseudomonadaceae; otu_2452	93.8%
		CMIH3 (2)	1362	54	B91-12; otu_2175	98.7%
		AMKC1 (2)	1355	53	<i>Lysobacter</i> sp.; otu_2521	97.2%
		AMMD3 (2)	1374	28	KCM-B-13; otu_2380	99.0%
		AMNC8 (DNA)	1373	20	Ellin339; pLW 30; otu_2332	97.4%
		CMIC7 (2)	1356	15	<i>Microbulbifer</i> sp.; otu_2398	93.6%
		AMMC6 (2)	1373	10	MP1-2H; otu_2406	94.0%
t_{+12}	<i>Actinobacteria</i>	CM37G3 (RNA)	1344	70	Unclassified Microthrixinae; otu_257	94.3%
	α - <i>Proteobacteria</i>	CM37B11 (2)	1318	13	KCM-B-56; unclassified L-3; otu_2380	92.9%
		CM38F9 (2)	1319	10	Unclassified α -Proteobacteria; otu_1889	95.7%
	δ - <i>Proteobacteria</i>	CM36D3 (2)	1380	22	Unclassified EB1021; otu_2606	88.0%
		AMEE5 (2)	1396	11	Unclassified EB1021; otu_2607	88.1%
	γ - <i>Proteobacteria</i>	AMGH2 (2)	1373	65	B91-12; otu_2175	93.2%
		CM36H3 (2)	1362	18	B91-12; otu_2175	98.7%
		CM36C7 (2)	1357	41	KCM-B-13; otu_2380	99.1%
		AMPC5 (2)	1356	27	Ellin339; pLW 30; otu_2332	97.4%
		AMGC2 (2)	1372	17	<i>Microbulbifer</i> sp.; otu_2398	93.3%
	AMOC7 (2)	1365	24	<i>Lysobacter</i> sp.; otu_2521	97.3%	
t_{+24}	<i>Acidobacteria</i>	AMHG5 (DNA)	1393	12	Unclassified Acidobacteria-6; otu_200	98.6%
		AMHA10 (2)	1372	11	Unclassified Acidobacteria-6; iii1-15; iii3-15; otu_202	99.0%
	<i>Actinobacteria</i>	CM24F12 (2)	1344	155	Unclassified Microthrixinae; otu_257	95.0%
		CM1C08 (2)	1341	48	Unclassified Microbacteriaceae; otu_316	99.3%
		CM23H2 (2)	1342	44	Unclassified Microthrixinae; otu_257	94.4%
		CM17C12 (2)	1344	14	Acidimicrobidae; a13110; otu_239	94.8%
		CM24B8 (2)	1343	12	Unclassified Microthrixinae; otu_257	94.1%
		AMDB11 (DNA)	1353	10	<i>Agromyces</i> sp.; otu_317	97.9%
	α - <i>Proteobacteria</i>	AMDB5 (2)	1321	13	Unclassified Ellin314; otu_2009	94.1%
	δ - <i>Proteobacteria</i>	AMDD5 (DNA)	1377	16	Unclassified NB1-j; otu_2637	92.1%
γ - <i>Proteobacteria</i>	AMIE4 (DNA)	1356	19	B91-12; otu_2175	93.1%	

Planctomycetes, *Verrucomicrobia*, SBR1093 and TM7 divisions at t_0 , the *Chlorobi* and SBR1093 at t_{+12} , and the *Verrucomicrobia*, OP10 and WS3 divisions at t_{+24} (Table 1). However, the number of phylotypes and clones belonging to these DNA-specific bacterial divisions were relatively low. Regarding the SSU rRNA libraries, they are also characterized by a “RNA-library specific” bacterial group, the TM6 division (Table 1). On the 534 phylotypes, only 122 are present in both type of libraries but they correspond to 58.5% of the total sequences (Table 1). This global trend was also observed at a finer scale as the 32 shared phylotypes characterized at t_0 correspond to 58.4% of all sequences identified at this point, the 45 shared phylotypes characterized at t_{+12} correspond to 61.0% of all sequences identified at this second point and 44 shared phylotypes characterized at t_{+24} represent 53.7% of all sequences identified at this last point (Data not shown). These common taxonomic units included 73.3% of the dominant phylotypes belonging especially to the *γ-Proteobacteria* and *Actinobacteria* (Table 2).

Rarefaction analysis and indices describing the microbial communities

In order to evaluate the representativeness of the six libraries, rarefaction analyses were performed. For each assemblage, the resulting curve did not attain its asymptotic phase illustrating an incompletely described diversity of the studied microbial communities (Fig.1). Nevertheless, the Good's indices (C) calculation has shown that the size of the assemblages was sufficient to cover respectively 74.0% of the microbial communities at t_0 , 79.8% at t_{+12} and 77.8% at t_{+24} using the SSU rDNA, and 88.3% (t_0), 81.3% (t_{+12}) and 85.7% (t_{+24}) using the SSU rRNA. These estimates were corroborated by the rarefaction plots which showed higher expected numbers of phylotypes for the SSU rDNA assemblages (Fig.1). The estimated numbers of phylotypes (S_{Chao1}) in the libraries do forecasted the presence of 346 (t_0), 218 (t_{+12}) and 277 phylotypes (t_{+24}) in the SSU rDNA collections and 145 (t_0), 236 (t_{+12})

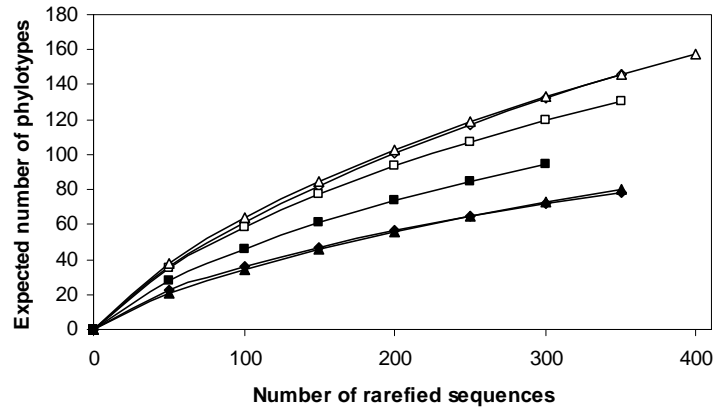


Fig. 1. Rarefaction analysis of the three SSU rDNA (straight shape) and the three SSU rRNA (dark shape) libraries generated before the bioremediation process (Δ), after 12 months (\square) and after 24 months (\diamond).

and 166 phylotypes (t_{+24}) in those of SSU rRNA. The calculation of the Shannon's diversity indices (H') has shown it was higher for the SSU rDNA libraries than for those of SSU rRNA: 4.476/3.234 at t_0 , 4.311/3.659 at t_{+12} and 4.552/2.819 at t_{+24} . Finally, equitability indices (J) were also higher for the SSU rDNA assemblages than for those of SSU rRNA: 0.524/0.376 at t_0 , 0.507/0.437 at t_{+12} and 0.526/0.333 at t_{+24} .

Diversity of microbial populations during the pilot-scale bioremediation process

The microbial communities of the hydrocarbon-polluted soil are dominated by the *Proteobacteria* which stand for 50.0% of phylotypes identified from DNA. The two largest subdivisions are the α -*Proteobacteria* and the γ -*Proteobacteria* representing respectively 22.4% and 16.4% of phylotypes. However, although the γ -type was less involved in the overall diversity, it represents the dominant bacterial subdivision with 39.7% of the clones. It comprises the dominant phylotypes present in this environment: the unclassified *Pseudomonadaceae*/otu_2452; B91-12/otu_2175 and *Lysobacter sp.*/otu_2521 (Table 2). These first observations were corroborated by the analysis of the RNA libraries showing that the *Proteobacteria* account for more than 3/4 of phylotypes of the metabolically active communities. The use of the SSU rRNA has also brought to light the place of *Actinobacteria* (25.4% of RNA clones), and especially the unclassified *Microthrixinae*/otu_257.

During the first part of the bioremediation process, the bacterial communities have changed, except for 21.9% of phylotypes which were still present in the DNA libraries after 12 months. These phylotypes represent the majority of the microorganisms (69.3% of clones). Concerning the metabolically active communities (SSU rRNA fraction), they seem to be more stable with 31.8% of the phylotypes common between two sampling points and which correspond to 74.4% of clones.

The *Proteobacteria* were still dominant at the second sampling point with 61.5% and 60.5% of the overall sequences obtained respectively from the DNA and RNA libraries. However, the number of phylotypes belonging to this phylum decreases down to 36.8% for assemblages from DNA library and to 59.4% for assemblages from RNA one (Table 1). This falling particularly concerns the “DNA-library specific” sequences belonging to the α - and γ -*Proteobacteria* suggesting a decrease of the diversity of the non-active parts of these both phyla.

The analysis of the SSU rRNA library has shown a more complex dynamic: the number of clones belonging to the α -*Proteobacteria* increased, whereas their number of phylotypes was quite stable; for the γ -type, the number of phylotypes decreased (Table 1). Thus, the unclassified *Pseudomonadaceae*/otu_2452, the B91-12/otu_2175 and the *Lysobacter sp.*/otu_2521, which were dominant before the bioremediation of the polluted soil, declined after 12 months of decontamination (Table 2). The analysis of SSU rRNA library has also shown that *Actinobacteria* take still an important place in the active microbial communities with the domination of the unclassified *Microthrixinae*/otu_257. At this sampling point, the active phylotypes belonging to the SPAM division identified before the beginning of the process began disappearing, while some active *Bacteroidetes*, *Planctomycetes*, TM7, *Verrucomicrobia* and WS3 appear.

During the second part of the bioremediation process, only 25.7% of the phylotypes representing 60.8% of the clones for DNA library and 18.7% of the phylotypes representing 66.3% of the clones for the RNA library have continued to evolve. The *Proteobacteria* were still dominant in the DNA library but the number of their clones largely decreased down to 42.7%. This rundown of the sequences belonging to the *Proteobacteria* concerns more particularly the γ -subdivision. This reduction is opposed to the increase of *Actinobacteria* and *Chloroflexi*. The importance of *Actinobacteria* inside the metabolically active communities

Table 3. Metabolically active bacterial core-set observed before and during the bioremediation process.

Bacterial division	Representative sequence	Length (nt)	Nb of clones	Closest relative according the Hugenholz's taxonomy	% Identity	
<i>Actinobacteria</i>	CM17C12	1344	21	Acidimicrobidae; a13110; otu_239	94.8%	
	CM23B2	1344	286	Microthrixinae; Unclassified; otu_257	94.4%	
	CM36G9	1343	7	Microthrixinae; Unclassified; otu_257	94.7%	
	CM23H2	1342	20	Microthrixinae; Unclassified; otu_257	94.4%	
<i>Gemmatimonadetes</i>	CMJH4	1339	6	Gemm-1b; Unclassified; otu_1656	94.5%	
<i>α-Proteobacteria</i>	CM38E10	1309	5	Ellin314; Bacterium_Ellin361; otu_2011	99.1%	
	CMJA9	1319	7	Ellin314; Unclassified; otu_2009	94.0%	
	CM1H03	1307	3	Kaistia; otu_2018	94.2%	
	CM37B11	1318	12	KCM-B-56; L-3; Unclassified; otu_2020	92.9%	
	CM1A07	1281	7	Rhodobacterales; H9; Unclassified; otu_2069	95.2%	
	CM36C11	1319	6	Unclassified; otu_1889	95.7%	
	<i>β-Proteobacteria</i>	CM41B9	1357	7	MND1; LO131; otu_2279	98.2%
	<i>γ-Proteobacteria</i>	CMOA12	1362	49	B91-12; otu_2175	98.7%
CMIC7		1356	22	Microbulbifer sp.; otu_2398	93.6%	
CMOF6		1357	6	Pseudomonadaceae; Unclassified; otu_2452	99.9%	
CMMG5		1356	95	Pseudomonadaceae; Unclassified; otu_2452	93.5%	
<i>δ-Proteobacteria</i>	CMMD11	1380	36	Unclassified EB1021; otu_2606	88.1%	
TM7	CM3D10	1301	4	TM7-1; Unclassified; otu_2810	92.6%	

was highlighted by the analysis of the rRNA library. Actually, the phylotypes belonging to this phylum represent 33.3% of the overall diversity and up to 75.6% of the total sequences. Moreover, three *Actinobacteria* saddled with the phylotypes which dominate at this sampling date: two unclassified *Microthrixinae* belonging to the otu_257 and one unclassified *Microbacteriaceae*/otu_316 (Table 2). On the contrary, the proportion of the proteobacterial phylotypes and clones respectively decreased down to 38.3% and 14.9% and only few *Proteobacteria* are still dominant at the end of the bioremediation process. At the last sampling date, the active phylotypes belonging to the *Bacteroidetes*, *Verrucomicrobia* and WS3 identified at t₊₁₂ disappeared whereas some active SBR1093 and TM6 appear.

In conclusion, the study of the bacterial communities during all the bioremediation process has shown the presence of a bacterial “core-set” composed by 37 phylotypes corresponding to 47.4% of the clones obtained from the three DNA libraries. This core-set was limited to 19 phylotypes for the metabolically active communities, corresponding to 55.8% of the clones obtained from the three RNA libraries (Table 3). The metabolically active phylotypes belonged to the *Actinobacteria*, *Gemmatimonadetes*, *α-Proteobacteria*, *β-Proteobacteria*, *δ-Proteobacteria*, *γ-Proteobacteria* and TM7 division.

Discussion

The improvement of the bioremediation technologies necessitates knowledges about the microorganisms involved in the pollutants degradation. Thereby it is necessary to characterize microbial communities inhabiting contaminated environments as well as their evolution during the decontamination processes. Thus, we have analyzed the bacterial populations present in a hydrocarbon-polluted soil before, at the middle and at the end of its treatment in a bioremediation pilot by the construction of ribosomal sequences libraries. In order to have a better overview of the metabolically active communities we have combined the study of SSU

rDNA and SSU rRNA sequences. Actually, as the detectability of the cells is strongly influenced by their physiological history and current physiological state (Oda *et al.*, 2000), the majority of microorganisms with metabolic and growth activities contain more SSU rRNA than the quiescent ones (Nomura *et al.*, 1984). However, due to the difficulties to isolate this ribosomal fraction from the complex environments, only few studies use the SSU rRNA to characterize the soil microbial communities (Felske *et al.*, 1997; Nogales *et al.*, 2001; Gremion *et al.*, 2003). For each library, nearly 400 full-length ribosomal sequences have been sequenced allowing to obtain important diversity coverage percentages (74-88.3% of the bacterial populations according to the Good's estimator). These results highlighted that, as the diversity is not completely described, the reached libraries sizes are sufficient to have a good overview of the soil microbial communities. Actually, Kemp and Aller (2004) have estimated that most of clones libraries obtained from soils only cover 48% (+/-10) of the studied environmental diversity due to their relatively weak size. This estimation has been increased up to 65% (+/-27) for the libraries judged to be large enough to have a stable estimation of the phylotypes richness. The lowest coverage indices obtained from the SSU rDNA sequences suggests a higher diversity in the DNA libraries. These observations were corroborated by higher observed and expected numbers of phylotypes, S_{Chao1} and Shannon indices for the DNA collections. This trend could be explained by the fact that the SSU rDNA does not recognize only the active microbial populations involved in the bioremediation process. Actually, the SSU rDNA libraries also collect the ribosomal sequences from the quiescent microorganisms and from the extracellular DNA which is persistent in the soils after the cell death (Recorbet *et al.*, 1993). The comparison of the libraries has shown that only 9.4% phylotypes were common in both types of libraries, illustrating the differential recognition between the both types of molecules. However, these shared phylotypes contain the majority of the sequences indicating, on one hand, a good correspondence between the both kinds of

collections and, on the other hand, that their differences came mostly from the rare phylotypes. Actually, the majority of the dominant phylotypes were present in the both types of libraries except for few of them. Moreover, the SSU rDNA libraries have shown the presence of several rare taxonomic units belonging to bacterial divisions which are not present in rRNA libraries (*Bacteroidetes*, *Nitrospirae*, *Planctomycetes*, *Verrucomicrobia*, SBR1093, TM7, *Chlorobi*, OP10 and WS3 divisions according to the sampling point) and conversely, the absence of the TM6 division present only in rRNA libraries. The identification of these “RNA-specific” sequences is surprising because the cells should contain both SSU rRNA and SSU rDNA molecules. These unexpected differences could be explained by a variability induced by methodological biases. Actually, the efficiency of the PCR and the RT-PCR used to generate the both types of libraries can be variable according different factors. First, the different amounts of the ribosomal templates in the reaction matrices could induce variable amplifications (Suzuki and Giovannoni, 1996). Moreover, the reverse transcription is known to potentially be abortive in some cases, as when the reverse primers used target the nearly terminal region of the SSU rRNA (Weller *et al.*, 1991). Despite of these potential biases, the different libraries have allowed to identify the active and the non-active parts of the microbial populations in the hydrocarbon-contaminated soil, and to follow their dynamic during the bioremediation process. Before and during the decontamination process, the soil has shown the continuous or discontinuous presence of 534 phylotypes belonging to 19 bacterial phyla. The found bacterial divisions of which the *Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Gemmatimonadetes*, *Planctomycetes* and *Verrucomicrobia* are often present in this kind of environment (Hugenholtz *et al.*, 1998; Jansen, 2006). More surprisingly, no sequences belonging to the *Firmicutes* were found, but their weak presence in a hydrocarbon-polluted soil was already observed (Popp *et al.*, 2006). At t_0 , the populations were dominated by the α - and the γ -*Proteobacteria*, and more especially by the

microorganisms belonging to *Pseudomonadaceae*, B91-12 and *Lysobacter*. These results are in accordance with most of the studies which have shown the importance of the *Proteobacteria*, and particularly the γ -division, in the hydrocarbon polluted-soil microbial communities (Margesin *et al.*, 2003; Hamamura *et al.*, 2006; Popp *et al.*, 2006). This γ -shift was explained by the presence of the pollutant representing a nutrient source favorable to these r-strategist bacteria. The noteworthy presence of the *Actinobacteria* (especially the *Microthrixinae*) before the bioremediation process was revealed by the RNA library. These observations suggest the adaptation and the potential implication of these both phyla in the first stage of the pollutant degradation. During the two-year bioremediation process, the microbial populations have evolved according the decrease of the pollution. The diminution of the contamination seems to generate an overall decrease of the *Proteobacteria* accompanied by an increase in *Actinobacteria*, particularly marked after 24 months of bioremediation. However, as the communities change, the comparison of the libraries has shown the presence of an important bacterial core-set throughout the decontamination process composed by nearly half of the sequences. It is constituted by 37 phylotypes for the DNA libraries and is restricted to 19 phylotypes for the metabolically active communities. These last phylotypes belonging to the *Actinobacteria*, *Gemmatimonadetes*, α -*Proteobacteria*, β -*Proteobacteria*, δ -*Proteobacteria*, γ -*Proteobacteria* and TM7 divisions could become potential targets to improve the degradation of the pollutants. In order to confirm their implications in the bioremediation mechanisms in the same way that the phylotypes dominating the bacterial communities, it will be necessary to obtain their isolation using cultivation techniques with the pollutants as sole carbon source. These isolated microorganisms could be used to improve the decontamination rates using the bioaugmentation techniques.

Experimental procedures

Nucleic acids recovery

The samples used for total nucleic acids recovery were taken from a bioremediation pilot (Biobasic environnement[®]) in which a contaminated alluvial soil (10,000 ppm of various aliphatic hydrocarbons) is biostimulated. These samples were removed before the beginning of the process and after 300 and 600 days of biostimulation. Both DNA and total RNA were extracted from each sample. DNA was extracted and purified from 5 g of biostimulation pilot soils using a protocol described previously by Zhou *et al.* (1996). The procedure used was slightly modified by addition of a phenol and a phenol/chloroform/AIA (25:24:1) purification step before the chloroform/AIA purification. The presence of DNAs was determined by agarose gel electrophoresis (1 % (wt/vol)) and staining with ethidium bromide.

Total RNA was extracted and purified from 10 g of biostimulation pilot soils after 30 min incubation at 37° C with 30 mL of NLS buffer (N-lauroyl sarcosine 1 % (wt/vol) ; CTAB 1 % (wt/vol) ; EDTA 0.5 M), 100 U of RNase inhibitor (Sigma) and 5 g of Proteinase K (Sigma). This mix was then grinded twice for 30 s at 30 Hz in the MM301 bead-beater (Retsch) with 15 g of glass and zirconium beads (106 µM, 2 mm and 1cm) before 10 min incubation at 60° C. The grinding solution was centrifuged for 10 min at 8,500 g at 4° C and the resulting supernatant was purified using phenol, phenol/chloroform (1:1) and chloroform extraction steps. The RNAs were pelleted, washed with ethanol 70 % and concentrated using the RNeasy Mini Kit as recommended by the manufacturer (Qiagen GmbH, Hilden, Germany). Residual DNA in RNA samples was eliminated using 2 U of RNase-free DNase I (AmpGrade, Invitrogen). RNA was analyzed on the 2100 Bioanalyzer using the RNA 6000 Nano and Pico kits as recommended by the manufacturer. The resultant DNA and RNA solutions were stored at -20° C for later used.

RT-PCR amplification of SSU rRNA and PCR amplification of SSU rDNA

Denatured SSU rRNA (125 ng) were reverse transcribed using the 1406r primer (5'-ACGGGCGGTGWGTRCAA-3') targeting most microorganisms belonging to the Eubacteria and the Archaea (Baker *et al.*, 2003). Reactions (20 µL) containing 0.625 µM of reverse primers, 100 U of Superscript III reverse transcriptase (Invitrogen), 1 X RT buffer, 0.1 M DTT, 0.25 mM of dNTP (Invitrogen), 1 U of RNasin+ (Promega) and 19 µg of T4 gene 32 Protein (Roche) were performed for 2 h at 42° C. RT products were analyzed on the 2100 Bioanalyzer using the RNA 6000 Nano and Pico kits after their hydrolysis with 0.25 M of NaOH for 15 min at 65° C.

PCR amplifications were performed using the primers 27f (5'-AGAGTTTGATCCTGGCTCAG-3') and 1406r described by Baker *et al.* (2003; Suzuki) and Giovanonni (1996) (Eurogentec). Negative controls without DNA were always included in PCR amplification experiments. The 50 µL reaction mixture contained 50 ng of soil DNA or RT products, 1 U of GoTaq DNA polymerase (Promega), 1 X PCR buffer containing MgCl₂ (concentration), 0.2 µM concentrations of each primer, and 250 µM concentrations of each deoxynucleotide. The reactions were carried in a iCycler thermocycler (Biorad) using a program composed by an initial denaturation step for 5 min at 95° C followed by 35 cycles of 3 steps (95° C for 1 min, 59.3° C for 30 s and 72° C for 1 min) and a final extension step for 7 min at 72° C. The presence and size of the amplification products were determined by agarose gel electrophoresis. Bands of the expected size were excised and purified with the Qiaquick Gel Elution kit as recommended by the manufacturer (Qiagen GmbH, Hilden, Germany).

Clone libraries construction, screening and sequencing

Agarose purified PCR products were ligated into the pCR®II-TOPO® vector (TOPO TA Cloning® kit Dual Promotor, Invitrogen) and transformed into *E. coli* One Shot® TOP10

cells (Invitrogen) following the manufacturer's instructions. Plasmid DNA was extracted using the Montage Plasmid Miniprep_{HTS} 96 kit (Millipore). Each clone was digested with *EcoRI* restriction enzyme, analyzed by electrophoresis on a 1 % (wt/vol) agarose gel and stained with ethidium bromide. Positive purified plasmids were sequenced by the company MWG Biotech (Ebersberg, Germany) using both SP6 and T7 vector primers.

Sequences treatment and analysis

The sequences treatment and joining were performed using the pregap4 and the gap4 tools of the Staden Package Program (Staden, 1996). The sequences were sense orientated using the OrientationChecker tool (Ashelford *et al.*, 2006) and aligned using the CLUSTAL W (1.83) Multiple Sequence Alignments (MSA) program (Thompson *et al.*, 1994). Sequences were checked for chimeric PCR artefacts using the Mallard software (Ashelford *et al.*, 2006) by inspection of sequences alignments. Potential composite sequences were excluded. The cleaned sequences were aligned using the NAST algorithm (DeSantis *et al.*, 2006) of the Greengenes web-interface (greengenes.lbl.gov/NAST). The Greengenes tools have also allowed the classification of the sequences according to the Hugenholtz taxonomy and the creation of distances matrix. The OTUs bringing together sequences sharing more than 97% homology were characterized using the furthest neighbor clustering algorithm of the DOTUR program (Schloss and Handelsman, 2005). Shannon diversity indexes (H') and the probable numbers of phylotypes (S_{Chao1}) along with the lower and the upper bounds of the confidence intervals (named respectively 95% lci and 95% uci) were also calculated using this tool. The Good's (C) and the equitability (J) indices were calculated using the following formulas: i) $C = [1 - (n_1/N)] * 100$ where n_1 is the number of singletons (OTUs with a single sequence) and N is the number of individuals sampled ; ii) $J = H'/H'_{\text{max}}$ where H' is the Shannon index and H'_{max} is the H' index when each OTU is represented by a single clone. Finally, the

rarefaction calculations were done using the Analytic Rarefaction (1.3) program (Raup, 1972) of the Stratigraphy Lab, University of Georgia (www.uga.edu/~strata/software/Software.html).

References

- Akob, D.M., Mills, H.J., and Kostka, J.E. (2007) Metabolically active microbial communities in uranium-contaminated subsurface sediments. *FEMS Microbiol Ecol* **59**: 95-107.
- Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143-169.
- Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. (2006) New screening software shows that most recent large SSU rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol* **72**: 5734-5741.
- Baker, G.C., Smith, J.J., and Cowan, D.A. (2003) Review and re-analysis of domain-specific SSU primers. *J Microbiol Methods* **55**: 541-555.
- Botton, S., and Parsons, J.R. (2007) Degradation of BTX by dissimilatory iron-reducing cultures. *Biodegradation* **18**: 371-381.
- Da Silva, M.L., and Alvarez, P.J. (2004) Enhanced anaerobic biodegradation of benzene-toluene-ethylbenzene-xylene-ethanol mixtures in bioaugmented aquifer columns. *Appl Environ Microbiol* **70**: 4720-4726.
- de Carvalho, C.C., and da Fonseca, M.M. (2005) Degradation of hydrocarbons and alcohols at different temperatures and salinities by *Rhodococcus erythropolis* DCL14. *FEMS Microbiol Ecol* **51**: 389-399.
- DeSantis, T.Z., Jr., Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M. et al. (2006) NAST: a multiple sequence alignment server for comparative analysis of SSU rRNA genes. *Nucleic Acids Res* **34**: W394-399.
- Dunbar, J., Takala, S., Barns, S.M., Davis, J.A., and Kuske, C.R. (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and SSU rRNA gene cloning. *Appl Environ Microbiol* **65**: 1662-1669.
- Felske, A., Rheims, H., Wolterink, A., Stackebrandt, E., and Akkermans, A.D. (1997) Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils. *Microbiology* **143** (Pt 9): 2983-2989.

- Gans, J., Wolinsky, M., and Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**: 1387-1390.
- Giovannoni, S.J., and Stingl, U. (2005) Molecular diversity and ecology of microbial plankton. *Nature* **437**: 343-348.
- Gremion, F., Chatzinotas, A., and Harms, H. (2003) Comparative SSU rDNA and SSU rRNA sequence analysis indicates that Actinobacteria might be a dominant part of the metabolically active bacteria in heavy metal-contaminated bulk and rhizosphere soil. *Environ Microbiol* **5**: 896-907.
- Hamamura, N., Olson, S.H., Ward, D.M., and Inskeep, W.P. (2006) Microbial population dynamics associated with crude-oil biodegradation in diverse soils. *Appl Environ Microbiol* **72**: 6316-6324.
- Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**: 4765-4774.
- Janssen, P.H. (2006) Identifying the dominant soil bacterial taxa in libraries of SSU rRNA and SSU rRNA genes. *Appl Environ Microbiol* **72**: 1719-1728.
- Kemp, P., and Aller, J. (2004) Bacterial diversity in aquatic and other environments: what SSU rDNA libraries can tell us. *FEMS Microbiol Ecol* **47**: 161-177.
- Khan, F.I., Husain, T., and Hejazi, R. (2004) An overview and analysis of site remediation technologies. *J Environ Manage* **71**: 95-122.
- Margesin, R., Labbe, D., Schinner, F., Greer, C.W., and Whyte, L.G. (2003) Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine Alpine soils. *Appl Environ Microbiol* **69**: 3085-3092.
- Martinez, R.J., Mills, H.J., Story, S., and Sobczyk, P.A. (2006) Prokaryotic diversity and metabolically active microbial populations in sediments from an active mud volcano in the Gulf of Mexico. *Environ Microbiol* **8**: 1783-1796.
- Nogales, B., Moore, E.R., Llobet-Brossa, E., Rossello-Mora, R., Amann, R., and Timmis, K.N. (2001) Combined use of SSU ribosomal DNA and SSU rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. *Appl Environ Microbiol* **67**: 1874-1884.
- Nomura, M., Gourse, R., and Baugham, G. (1984) Regulation of synthesis of ribosomes and ribosomal components. *Annu Rev Biochem* **53**: 75-117.
- Odaa, Y., Slagmana, S., Meijerb, W.G., Forneya, L.J., and Gottschala, J.C. (2000) Influence of growth rate and starvation on fluorescent in situ hybridization of *Rhodospseudomonas palustris*. *FEMS Microbiol Ecol* **32**: 205-213.

- Olsen, G.J., Pace, N.R., Nuell, M., Kaine, B.P., Gupta, R., and Woese, C.R. (1985) Sequence of the SSU rRNA gene from the thermoacidophilic archaebacterium *Sulfolobus solfataricus* and its evolutionary implications. *J Mol Evol* **22**: 301-307.
- Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276**: 734-740.
- Paul, D., Pandey, G., Pandey, J., and Jain, R.K. (2005) Accessing microbial diversity for bioremediation and environmental restoration. *Trends Biotechnol* **23**: 135-142.
- Perfumo, A., Banat, I.M., Marchant, R., and Vezzulli, L. (2007) Thermally enhanced approaches for bioremediation of hydrocarbon-contaminated soils. *Chemosphere* **66**: 179-184.
- Popp, N., Schlomann, M., and Mau, M. (2006) Bacterial diversity in the active stage of a bioremediation system for mineral oil hydrocarbon-contaminated soils. *Microbiology* **152**: 3291-3304.
- Raup, D.M. (1972) Taxonomic Diversity during the Phanerozoic. *Science* **177**: 1065-1071.
- Recorbet, G., Picard, C., Normand, P., and Simonet, P. (1993) Kinetics of the persistence of chromosomal DNA from genetically engineered *Escherichia coli* introduced into soil. *Appl Environ Microbiol* **59**: 4289-4294.
- Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501-1506.
- Staden, R. (1996) The Staden sequence analysis package. *Mol Biotechnol* **5**: 233-241.
- Suzuki, M.T., and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of SSU rRNA genes by PCR. *Appl Environ Microbiol* **62**: 625-630.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.
- Wang, F., Grundmann, S., Schmid, M., Dorfler, U., Roherer, S., Charles Munch, J. et al. (2007) Isolation and characterization of 1,2,4-trichlorobenzene mineralizing *Bordetella* sp. and its bioremediation potential in soil. *Chemosphere* **67**: 896-902.
- Weller, R., Weller, J.W., and Ward, D.M. (1991) SSU rRNA sequences of uncultivated hot spring cyanobacterial mat inhabitants retrieved as randomly primed cDNA. *Appl Environ Microbiol* **57**: 1146-1151.

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

D. DISCUSSION.

Cette étude cinétique a permis d'identifier des phylotypes bactériens potentiellement impliqués dans les processus de dégradation des hydrocarbures aliphatiques au cours de la bioremédiation du sol pollué. En effet, les phylotypes dominant les communautés bactériennes ou présentant une évolution positive durant le procédé de bioremédiation semblent posséder ou développer des adaptations leur permettant de survivre et même de croître sous des conditions défavorables induites par la présence du polluant.

Ainsi, nous avons pu identifier une implication potentielle de bactéries du type γ -*Proteobacteria* appartenant aux *Pseudomonadaceae*, aux B91-12 et au genre *Lysobacter* dans les premières phases de la bioremédiation. Ces observations vont dans le sens du phénomène nommé « gamma-shift » qui consiste en une déviation de la diversité vers les γ -*Proteobacteria* lors d'une augmentation de la richesse en nutriments ou lors de l'apparition d'une contamination (Popp *et al.*, 2006). D'après Evans *et al.* (2004), cette tendance pourrait être causée par le grand nombre de bactéries dégradant les hydrocarbures appartenant à cette classe. Il faut aussi noter que la famille des *Pseudomonadaceae* appartient à l'ordre des *Pseudomonales* qui est très souvent dominant dans les sites contaminés aux hydrocarbures (Evans *et al.*, 2004) ce qui va dans le sens de nos observations. L'abondance des *Microthrixineae*, appartenant aux *Actinobacteria*, semble aussi suggérer son adaptation à la présence des hydrocarbures aliphatiques voir même son implication lors des premières phases de la décontamination. Bien que certaines *Actinobacteria* aient été identifiées comme ayant des capacités de dégradation des hydrocarbures, les *Microthrixineae* n'ont jamais, à notre connaissance, été mise en évidence dans des sols pollués.

Durant le procédé de bioremédiation, nous pouvons observer une évolution des communautés bactériennes qui semble être corrélée à la diminution de la charge polluante de la terre. En effet, la dégradation des contaminants par certains microorganismes engendre la formation de nouvelles formes de polluants. Celles-ci produisent de nouvelles pressions de sélection sur les communautés microbiennes entraînant leur évolution : les bactéries adaptées persistent voire même croissent et celles ne possédant pas les mécanismes d'adaptations déclinent car elles ne trouvent plus dans cet environnement de niches écologiques adaptées à leur développement. Les observations durant la décontamination montrent que les *Proteobacteria* semblent diminuer au profit d'une augmentation des *Actinobacteria*. On observe que les *Pseudomonadaceae* disparaissent des phylotypes dominants après 12 mois de

traitement alors que les *Lysobacter* diminuent. Le phylotype B91-12 présent au début de la bioremédiation diminue au profit d'un autre phylotype appartenant au même groupe bactérien. Après 24 mois de traitement, les *Actinobacteria* dominent très largement les communautés bactériennes avec trois phylotypes appartenant aux *Microthrixineae* et aux *Microbacteriaceae*. Cette tendance a particulièrement été soulignée par l'analyse des banques ARN suggérant leur activité lors des derniers stades de la bioremédiation.

Bien que les communautés bactériennes présentent une évolution durant la décontamination du sol, la présence de certains phylotypes semble toutefois constante. Ce « core-set » bactérien est composé de 37 phylotypes dans le cas des banques ADN et est restreint à 17 pour les ARN. Le « core-set » métaboliquement actif est composé de phylotypes appartenant aux phyla *Actinobacteria*, *Gemmatimonadetes*, *α -Proteobacteria*, *β -Proteobacteria*, *δ -Proteobacteria*, *γ -Proteobacteria* et division TM7. Leur présence active durant les deux ans de bioremédiation semble aussi indiquer leur implication potentielle dans les procédés de dégradation.

Même si ce type d'étude permet d'explorer de façon approfondie la biodiversité des environnements complexes, il est cependant nécessaire de prendre en considération les limites liées aux différentes techniques utilisées. Par exemple, l'extraction des acides nucléiques peut être plus ou moins efficace en fonction des microorganismes. En effet, les méthodes d'extraction trop douces, impliquant des quantités d'enzymes insuffisantes, ne permettent pas de lyser efficacement toutes les bactéries, notamment les Gram positifs. Dans ce cas, leur biodiversité sera sous-estimée. Au contraire, les méthodes trop drastiques (broyage mécanique) conduisent à une fragmentation de l'ADN (von Wintzingerode *et al.*, 1997). Il faut aussi noter que les extractions classiques d'ADN ne font pas le distinguo entre l'ADN des cellules vivantes et l'ADN extracellulaire. Ainsi, l'image des populations bactériennes actives peut s'en trouver faussée. Récemment, l'utilisation de molécules permettant de faire une différenciation entre l'ADN intracellulaire et extracellulaire des bactéries a été décrite. En effet, le propidium monoazide (PMA), par une action intercalante au niveau des ADN extracellulaires ou de cellules mortes combinée à une photoinduction de son groupement azide, permettrait d'éliminer les ADN de la fraction bactérienne non vivante (Nocker *et al.*, 2007). L'extraction des acides nucléiques peut aussi co-entraîner des substances, présentes dans l'environnement, pouvant être inhibitrices des enzymes utilisées dans les réactions moléculaires. La purification des acides nucléiques peut également conduire à une perte de matériel susceptible de modifier l'analyse moléculaire de la diversité. Une autre étape très

sensible aux biais est la réaction d'amplification par PCR. En fonction du nombre de copies des ADNr, des amorces utilisées, du pourcentage en GC des gènes ciblés, de la qualité ou de la concentration de l'ADN matrice, chaque molécule pourra être amplifiée de façon différente (von Wintzingerode *et al.*, 1997).

Dans cette étude, certains microorganismes ont été identifiés comme étant potentiellement impliqués dans la dégradation des hydrocarbures aliphatiques. Afin de confirmer ces capacités, il serait nécessaire de mettre en place des expérimentations complémentaires. Dans un premier temps, des approches culturales pourraient être envisagées en utilisant des milieux minimums complétés en polluant afin d'isoler les souches l'utilisant comme seule source de carbone et d'énergie. Une fois isolées, ces souches pourraient être identifiées grâce à des analyses moléculaires notamment en identifiant leurs séquences ADNr 16S. Après confirmation de leurs capacités métaboliques (analyses moléculaires et biochimiques), ces souches pourront être utilisées pour améliorer les rendements de la bioremédiation des sols pollués aux hydrocarbures aliphatiques. Ceci pourrait passer soit par des aspects de bioaugmentation en inoculant ces souches au niveau du sol à traiter soit par des aspects de biostimulation. Dans ce cas, il serait nécessaire de mettre en place des tests afin de définir quels paramètres modifier afin de favoriser la dégradation du polluant par ces souches.

Notons cependant que les dynamiques des populations bactériennes au cours de ces processus de bioremédiation sont complexes et dépendantes de facteurs biotiques et abiotiques. L'identification des microorganismes impliqués dans ces phénomènes de dépollution pourra permettre à terme de caractériser plus facilement les interactions de ces *consortia* avec leur environnement et les différentes voies métaboliques impliquées dans la dégradation de polluants. Pour ce faire, il sera nécessaire de mettre en place des études de reconstruction de voies métaboliques par des approches d'inventaire exhaustif des gènes (séquençage) ou par des approches biochimiques (enzymologie, métabolome).

Finalement, même si l'étude globale du procédé de bioremédiation nous a permis de caractériser des souches potentiellement épuratrices, une analyse plus fine de l'évolution des microorganismes dans le temps nous auraient certainement permis de mieux comprendre les processus d'adaptations de ces communautés bactériennes. Cependant, du fait des efforts humains et financiers qu'impliquent la production et l'analyse de plus de 2 000 séquences complètes d'ADNr 16S et d'ARNr 16S, il ne nous a pas été possible d'envisager l'étude de ces communautés bactériennes à un pas expérimental plus fin. L'utilisation d'outils à haut-

débit, telles les biopuces à ADN, pourrait nous permettre d'envisager une exploration plus fine de l'évolution de ces microorganismes en réponse à la présence et à la dégradation du polluant. C'est pourquoi, le développement de ces outils à haut-débit a été initié.

PARTIE IV

DEVELOPPEMENT D'UN ALGORITHME DE SELECTION DE SONDES OLIGONUCLEOTIDIQUES POUR BIOPUCES PHYLOGENETIQUES

A. CONTEXTE DE L'ETUDE.

Comme nous l'avons vu précédemment, les techniques moléculaires classiquement utilisées pour étudier les communautés microbiennes peuvent s'avérer laborieuses voire même inadaptées pour analyser globalement la diversité bactérienne de nombreux échantillons complexes. Une technique alternative est l'utilisation de l'outil biopuce phylogénétique permettant d'identifier potentiellement la présence de milliers de phylotypes en une seule hybridation. Bien qu'elles ne soient pas encore utilisées de façon routinière au sein des laboratoires d'écologie microbienne, ces biopuces ont tout de même été utilisées dans plusieurs études afin d'appréhender la biodiversité microbienne et son fonctionnement au sein d'environnements complexes tels les systèmes aquatiques ou terrestres (Koizumi *et al.*, 2002 ; Castiglioni *et al.*, 2004 ; Loy *et al.*, 2004 ; DeSantis *et al.*, 2007 ; Sanguin *et al.*, 2006).

La pertinence de la caractérisation des communautés microbiennes par cette approche repose toutefois sur l'efficacité des sondes biopuces sélectionnées. Celles-ci doivent être hautement sensibles et reconnaître spécifiquement les groupes ciblés, même ceux présents en faibles abondances dans l'échantillon étudié (Gentry *et al.*, 2006 ; Peplies *et al.*, 2006). De plus, du fait du grand nombre de groupes bactériens encore méconnus, il est nécessaire que ces sondes possèdent un caractère exploratoire afin d'étudier l'ensemble des communautés bactériennes : connues et encore inconnues. Cependant, la plupart des logiciels de détermination de sondes utilisent les données incomplètes présentes dans les bases de données ribosomales publiques. Il en résulte donc une sélection de sondes uniquement dirigées vers les microorganismes répertoriés dans ces bases.

B. OBJECTIFS DE L'ETUDE.

L'objectif de ce travail de recherche a été le développement d'un algorithme permettant la sélection de sondes oligonucléotidiques pour biopuces phylogénétiques qui soient sensibles, spécifiques mais aussi exploratoires. Notre stratégie est basée sur la détection de groupes taxonomiques à des niveaux différents de l'espèce (tels le genre, la famille, etc.) et l'utilisation combinée de sondes dégénérées et non dégénérées afin d'étudier globalement les communautés bactériennes connues et inconnues. Cet algorithme a été implémenté dans un programme nommé PhylArray. Afin de tester l'efficacité de notre stratégie de design, des sondes ciblant les genres *Staphylococcus*, *Micrococcus*, *Aeromonas*, *Pseudomonas*,

Streptomyces, *Rhodococcus* et *Bifidobacterium* ont été sélectionnées avec le programme PhylArray mais aussi avec des outils classiquement utilisés pour le design de sondes phylogénétiques: le logiciel PRIMROSE (Ashelford *et al.*, 2002) et une base de sondes obtenues avec le logiciel ARB (Ludwig *et al.*, 2004). La stratégie de détermination de sondes GoArrays que nous avons également développée (Rimour *et al.*, 2005 ; voir Annexe) a aussi été utilisée afin d'améliorer la sensibilité et la spécificité des sondes courtes sélectionnées. Une biopuce prototype a été construite avec ces sondes et a été testée à l'aide des cibles extraites de souches bactériennes pures et d'environnements complexes.

C. PRINCIPAUX RESULTATS OBTENUS.

En terme de sensibilité, les sondes PhylArray ont montré des efficacités jusqu'à 20 fois supérieure par rapport aux sondes sélectionnées avec PRIMROSE ou dans la banque de sondes ARB. Nous avons aussi observé des différences de spécificité entre les différents types de sondes. En ce qui concerne les oligonucléotides longs (50 mers), les sondes PRIMROSE engendrent des signaux faussement-positifs dans 57 à 90% des cas alors que les sondes PhylArray présentent des pourcentages d'hybridations croisées inférieurs (0,05 à 46%). Les sondes courtes (20 mers) semblent globalement plus spécifiques avec de 0 à 42% de croisements pour les sondes ARB et de 3 à 28% de croisements pour les sondes PhylArray. De plus, certaines sondes PRIMROSE et ARB ont présenté des défauts de reconnaissance (absence de signal) contrairement aux sondes PhylArray qui ont toujours permis une reconnaissance des souches bactériennes testées. Il faut aussi noter que l'utilisation de la stratégie GoArrays sur les sondes courtes PhylArray et ARB a permis une amélioration de l'intensité du signal et une élimination des signaux faussement positifs.

En ce qui concerne le caractère exploratoire des sondes PhylArray, celui-ci a été testé en utilisant des cibles provenant des communautés bactériennes présentes au niveau du sol pollué par des hydrocarbures aliphatiques étudié dans la partie précédente (Partie III). La différence d'intensité de signaux entre les sondes dégénérées (SAerod2) et les sondes non dégénérées (SAero2.1, SAero2.2, SAero2.3 et SAero2.4) suggère la présence dans cette terre de microorganismes, appartenant ou proche du genre *Aeromonas*, non répertoriés dans les bases de données ribosomales.

Original papers: Sequence Analysis

PhylArray: Phylogenetic Probe Design Algorithm for MicroArray

Cécile Militon¹, Sébastien Rimour^{1,2}, Mohieddine Missaoui^{1,2}, Corinne Biderre¹, Vincent Barra², David Hill², Anne Moné¹, Geneviève Gagne¹, Harald Meier³, Eric Peyretailade¹ and Pierre Peyret^{1*}

¹Génomique Intégrée des Interactions Microbiennes, Laboratoire de Biologie des Protistes, UMR CNRS 6023, Blaise Pascal University, 24 avenue des Landais, Campus des Cézeaux, 63177 Aubière Cedex, France; ²LIMOS UMR CNRS 6158, Blaise Pascal University, Clermont-Ferrand II, BP 10125, 63177 Aubière Cedex, France; ³Lehrstuhl für Rechnerorganisation und Rechnerorganisation. Institut für Informatik. Technische Universität München. Germany. * For correspondence. E-mail: pierre.peyret@univ-bpclermont.fr fax : +33 (0) 4 7340 76 70 phone : +33 (0) 4 7340 76 04. The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors. Associate Editor: Prof. Thomas Lengauer

ABSTRACT

Motivation: Microbial diversity is still largely unknown in most environments, such as soils. In order to get access to this microbial “black-box,” the development of powerful tools such as microarrays are necessary. However, the reliability of this approach relies on probe efficiency, in particular sensitivity, specificity and explorative power, in order to obtain an image of the microbial communities that is close to reality. **Results:** We propose a new probe design algorithm that is able to select microarray probes targeting SSU rRNA at any phylogenetic level. This original approach, implemented in a program called “PhylArray”, designs a combination of degenerate and non-degenerate probes for each target taxon. Comparative experimental evaluations indicate that probes designed with PhylArray yield a higher sensitivity and specificity than those designed by conventional approaches. Applying the combined PhylArray/GoArrays strategy helps to optimize the hybridization performance of short probes. Finally, hybridizations with environmental targets have shown that the use of the PhylArray strategy can draw attention to even previously unknown bacteria. **Availability:** <http://fc.isima.fr/~rimour/phylarray/> **Contact:** pierre.peyret@univ-bpclermont.fr
Supplementary information: Supplementary data (Figures 1-2, Tables 1-2-3-4)

1 INTRODUCTION

Microorganisms are present in all environmental habitats, even the most extreme. Despite this extensive distribution, we have relatively little knowledge about these communities. Indeed, environmental studies are often limited by the difficulty to globally evaluate microbial populations and their dynamics in complex environments. For soils, Gans *et al.* (2005) have estimated that 1g of surface soil might contain more than one million distinct genomes. Moreover, a minority of those microorganisms have now been cultivated and characterized (Dunbar *et al.*, 1999). This underestimation of bacterial diversity, caused by the cultural bias, has forced the development of more suitable investigation methods. During the last decade, cultivation-independent molecular tools have been developed as an alternative in order to study environmental microbial communities more comprehensively (Amann *et al.*, 1995). These nucleic-acid based methodologies (PCR-based or hybridization methods) usually target the gene encoding the small subunit ribosomal RNA (SSU rDNA). However total nucleic acids from complex microbial environments are too rich in information to be easily analyzed by such molecular tools. Therefore SSU rDNA oligonucleotides microarrays have been developed (see for reviews Bodrossy and Sessitsch, 2004; Gentry *et al.*, 2006; Loy and Bodrossy, 2006). These high throughput molecular tools are able to detect up to several thousands of microbial phylotypes simultaneously in a single experiment using species-specific probes immobilized on a solid surface. The accuracy of such an approach, in terms of a comprehensive exploration of complex environments communities, relies on the efficiency of probes sets. They must be highly sensitive (Peplies *et al.*, 2006) and specifically recognize targeted groups even when the groups are present in low abundance (Gentry *et al.*, 2006). The majority of microarrays used for those studies are based on oligonucleotide probes, which present many design advantages (Ehrenreich, 2006). Cross-hybridization is the major point that limits the determination of specific probes. In order to evaluate the specificity of a given probe, it is necessary to have a reliable predictor for its hybridization performance. However, the dynamics of probe-target hybridization in a microarray experimental context are very complex and not yet fully understood. Recent work has demonstrated that the use of thermodynamics parameters to assess probe specificity does not show good results (Pozhitkov *et al.*, 2006). In our work, we decided to use sequence similarity to check probe specificity. A study made by Kane *et al.* (2000) on 50mer oligos shows that a probe must satisfy two conditions to be specific: (1) “The oligonucleotide sequence must not have more than 75% of similarity (among all sequences) with a non-targeted sequence present in the hybridization

pool”, (2) “The oligonucleotide sequence must not include a stretch of identical sequence greater than 15 contiguous bases”. Several probe design programs generalize these criteria to oligos of any length (for review, see Rimour *et al.*, 2005). In the remainder of this article, we use these criteria to check probe specificity. Values “75%” and “15 bases” are parameters of the algorithm and may be changed without modifying the bases of our method. In previous work, we also have proposed a new approach to design oligonucleotides that combines both the specificity and the sensitivity (Rimour *et al.*, 2005). In this strategy, named GoArrays, the oligonucleotide sequence is composed of two specific probe sequences (e.g. 25-mers) separated by a short random linker. As the oligonucleotide sequence is therefore quite long (e.g. 55-mers), it keeps the advantages of both short and long oligonucleotides. Obtaining specific and sensitive probes is a big challenge, but the design of explorative oligonucleotides is the greatest one. As described previously, the majority of microorganisms are still unidentified, and not present in public ribosomal databases. Most classical oligonucleotide design software uses therefore incomplete data sets to generate species specific probes. Thus only a small fraction of known bacterial communities can be studied with these probes. However a few design tools try to decrease this bias by allowing the selection of probes targeting higher bacterial taxa. ARB probe design tools (Ludwig *et al.*, 2004) and PRIMROSE software (Ashelford *et al.*, 2002) can generate this kind of taxon-specific probes. ARB is used in most of the biodiversity studies that use phylogenetic microarrays (Franke-Whittle *et al.*, 2005; Loy *et al.*, 2005; Sanguin *et al.*, 2006). Schliep and Rahmann (2006) use a statistical group-testing approach with non-unique probes to detect targets related by a phylogenetic tree. Their method can detect unknown targets, but it has been validated only on simulations of hybridization experiments.

In this paper, we describe a new algorithm, implemented in the program named PhylArray, allowing the generation of efficient probes. Our design strategy is based on the detection of high taxonomic groups and the use of a combination of degenerate and non-degenerate probes to globally monitor known and unknown bacterial communities.

2 PROBE DESIGN STRATEGY

Probe design for microarray experiments is not a trivial computational task. Parameters described previously have to be considered to obtain an efficient probe selection. Designing oligonucleotide probes for bacterial identification is basically the same problem as probe selection for classical gene expression experiments. The only difference is the specificity test.

In gene expression experiments, a probe identifying a given gene must be specific among all other gene sequences of the studied organism. When designing oligonucleotides from SSU rRNA, each probe must be specific among all SSU sequences that may be present in the sample during the hybridization step. If the mixture composition is totally unknown, the specificity can only be checked against all known SSU sequences. SSU rRNA sequences can be obtained from various sources: major primary databases (GenBank/EMBL/DDBJ) or curated secondary databases (Cole *et al.*, 2005; DeSantis *et al.*, 2006; Ludwig *et al.*, 2004; Wuyts *et al.*, 2004). The Ribosomal Database Project II provides aligned and annotated rRNA gene sequences, along with analysis services. It represents a widely used and good-quality data source for bacterial identification (Cole *et al.* 2005). Our aim is to develop a probe design algorithm for the selection of microarray oligonucleotides adapted to complex environments studies. The probes must be sensitive enough to detect all microbial community components, even in low abundance, and highly specific in order to recognize only the target groups. Moreover, as the majority of microorganisms from complex environments are still unidentified, we do not wish to use probes targeting only the known species, which would make the microarray unable to identify new species. Such constraints forced us to propose the following guidelines for developing our algorithm:

- The probes must target the genus as well as higher taxonomic units, in order to globally monitor known and unknown bacterial communities belonging to targeted taxonomic units. The polymorphism of the target group must be taken into account. To solve this problem, we decided to design a combination of degenerate and non-degenerate probes.
- The rRNA database must contain high-quality data in order to avoid cross-hybridizations that could be caused by sequences assigned to wrong species (false annotation). We decided to build our own secondary SSU rRNA database.

These points are discussed in the following paragraphs.

2.1 Targeted taxonomic groups and polymorphism questions.

It is very hard to identify and differentiate species of the same genus with a specific oligonucleotide microarray strategy because the SSU rRNA variability within some species is too low. Considering Kane's conditions, a probe that targets a species might cross-hybridize with another species of the same genus. This is why our algorithm selects probes that target at least a genus, or a higher taxonomic group (family, order...). In order to take into account the sequence polymorphism within the target group, the probe design software should generate a

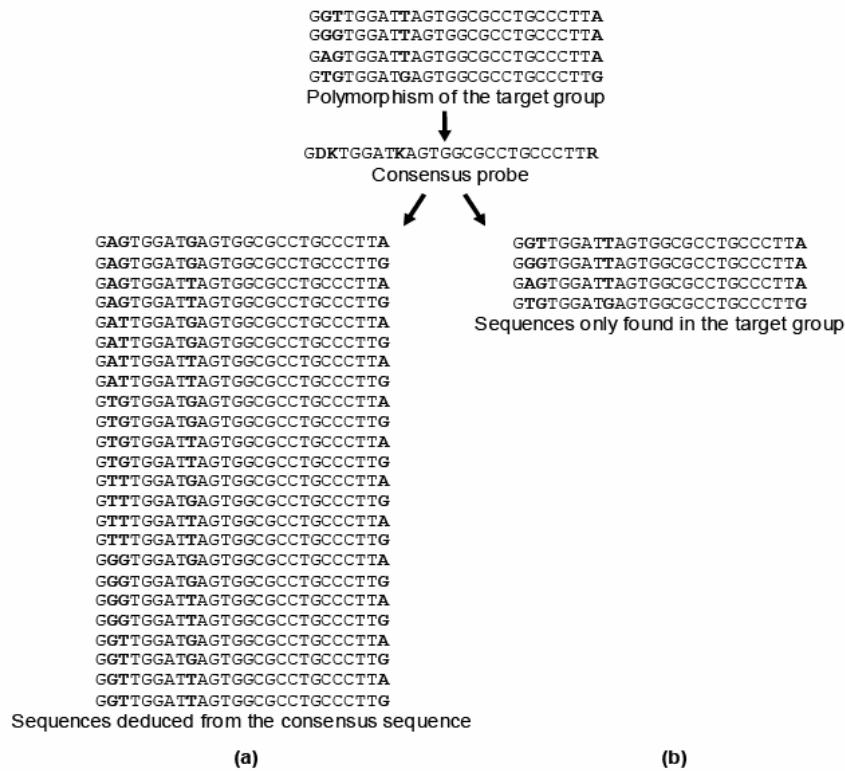


Fig. 1. Schematic representation of probe design to target a group of species. Sequences alignment allows the generation of the consensus sequence with degenerate positions. All sequences deduced from the consensus sequence (a) and real sequences used for the alignment step (b) are used in the search probe specificity.

consensus sequence for the group, using the IUPAC nomenclature. In a microarray experimental context, the spotted probe then is a mixture of all possible sequences which can be generated from the consensus sequence. By following this strategy some of these sequences targeted may not belong to real rRNAs (Fig. 1-a) and could lead to supplementary cross-hybridizations. In our method, the result given to the user contains the consensus sequence (degenerate oligonucleotide) and all the specific sequences derived from the single rRNA sequences that were used to build the consensus (Fig. 1-b). Thus, either the consensus sequence (with all sequence combinations) or each “real” sequence can be independently spotted on the microarray.

2.2 SSU rRNA database re-build.

All SSU sequences of taxonomic division PRO (prokaryotic sequences) were obtained from the EMBL database. The latter were classified by organism (OC line in EMBL entry). This classification is made only for the purpose of pre-filtering and might be modified by the next step of the database build. Candidate sequences were rejected from further analysis if:

- (1) The percentage of N (unknown base) in the sequence is $> 10\%$
 - (2) The sequence contains a stretch of 10 consecutive N
 - (3) The sequence is too short ($l_{max} < 400$, where l_{max} is the length of the longest sequence of the considered taxonomic unit)
- The next step is the database curation. It aims to reject sequences that are assigned to the wrong organism, or which present chimeric anomalies. We use the K-means algorithm. For each taxon, the sequences are first partitioned in two clusters ($K=2$). The distance between two sequences is derived from sequence similarity. When K-means has finished iterating for finding homogenous group of sequences, the well-annotated sequences are gathered in the same cluster. The other sequences are rejected.

3 DESIGN ALGORITHM

Input parameters for our algorithm are the name of the target taxon (T), the desired probe length (l), the specificity threshold (s) (Kane’s criteria), the maximum number of degenerate bases in the oligonucleotide sequence (xdeg), and the sequence database used for the specificity test. The target taxon could be a group located at any level of the phylogenetic tree, for example *Micrococcus* (Genus) or *Micrococcaceae* (Family). The specificity threshold is used to determine if the probe may hybridize with a non-target sequence. Thus, the user can modify Kane’s criteria described earlier in this paper (75% similarity, 15 identical contiguous

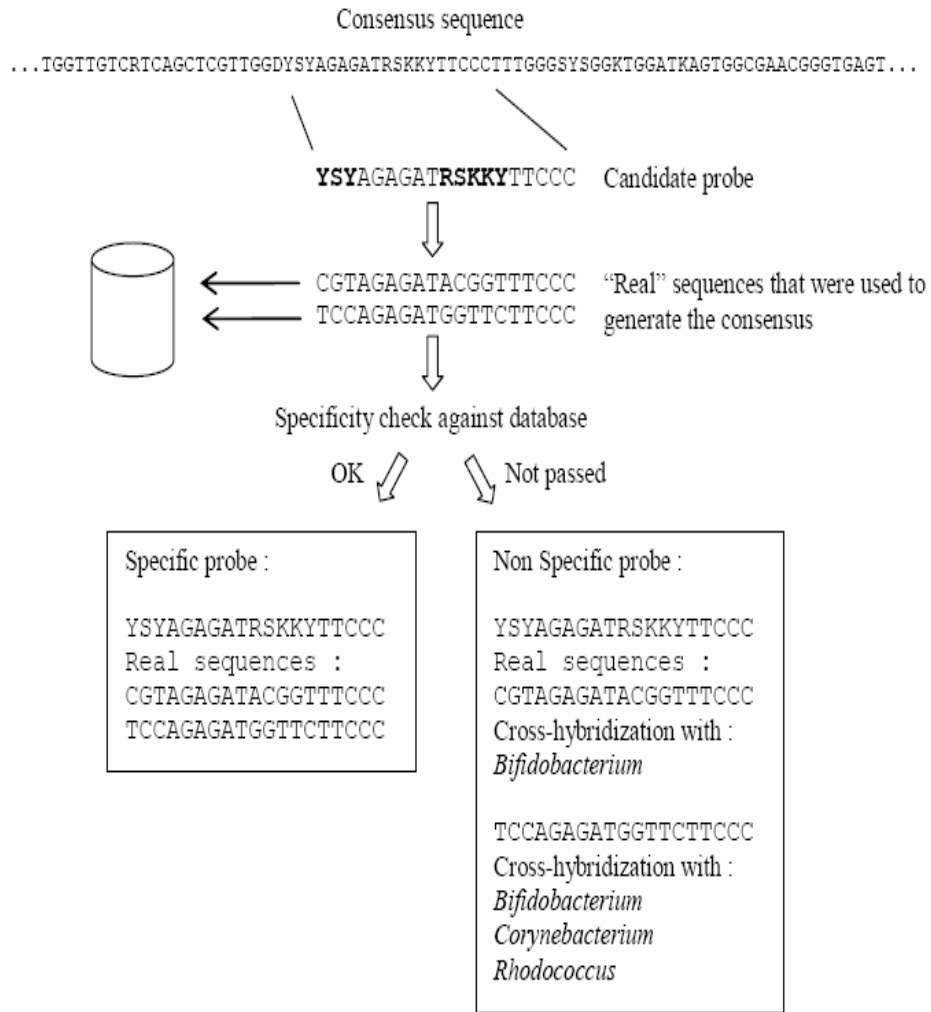


Fig. 2. Schematic illustration of the last step of the algorithm. A candidate probe is extracted from the consensus sequence. Only the sequences that were used to generate a consensus are tested for specificity. Even if the probes are not specific, the results are stored with possible cross-hybridizations.

bases) if it is too restrictive in the experimental conditions. The used database must contain all sequences which could be present in the sample during the hybridization step. The default database used has been described in the previous section with all SSU rRNA from the PRO division of EMBL database filtered using our algorithm. The obtained secondary database is composed of 25 110 sequences belonging to 1 900 genera. The design algorithm consists of four consecutive steps:

(1) *Extraction and filtering.* All sequences of the taxon T are extracted from the chosen rRNA database. We use NCBI Taxonomy (Wheeler *et al.*, 2000) stored in a relational database to facilitate the selection of sequences at different taxonomic levels.

(2) *Multiple sequence alignment.* The sequences obtained in step (1) are aligned using the ClustalW algorithm (Thompson *et al.*, 1994).

(3) *Search for a consensus sequence.* A “consensus sequence” using the IUPAC code is created from the alignment. The aim is not only to obtain a sequence that represents the group polymorphism, but also to remove possible sequencing errors. In each column of the alignment, the bases are replaced by a single consensus base. If numerous unspecified bases (N) or gaps (–) are observed in the alignment the following procedure is applied:

* If the number of “N” or “–“ characters in a column is greater or equal to 50% of the total number of characters, a “–“ is inserted in the consensus sequence.

* Otherwise, a consensus base is created, which corresponds to the bases present in the column. In this case, “N” and “–“ characters are assumed to be sequencing errors.

Figure 1 in supplementary data shows an example of this step.

(4) *Search for specific probes.* In the last step, specific probes are searched along the consensus sequence. The algorithm first tries to find a subsequence with less than xdeg degenerate bases (max. number of degenerate bases specified by the user) incrementing a window of length l along the consensus sequence. Then, the program checks the specificity of this subsequence, which is the critical step of our algorithm. In order to check the specificity of the potential probe, the program does not generate all base combinations from a consensus sequence, which would involve more checking than necessary. It gets only the sequences that were used to generate the consensus and checks their specificity. This process is illustrated on Figure 2.

If the tested sequences are specific, the probe is considered to be specific for the target taxon. The user result consists of the degenerate probe and the non-degenerate sequences that represent the taxon. Even if the probe is not specific, the results are stored including all targets

of potential cross-hybridizations. The user can specify a maximum number of cross-hybridizations for the probes to be stored in the result file, so that he can list all possible probes or only the most specific ones. The specificity of all possible sequences which can be generated from the consensus sequence is also tested, and the result can be added to the result file (program option). This gives information about the global specificity of the degenerate probe, including sequences that are not present in databases.

4 IMPLEMENTATION

The design algorithm is implemented in a program called PhylArray. The program is written in PERL, and it uses ClustalW for multiple sequence alignment and BLAST (with the following parameters: Word size $W=7$, Low-complexity Filtering $F=false$, Expectation value $E=1000$) for checking the specificity. A parallel implementation allows probe finding to be done in parallel on a computing cluster architecture. The parallelism is introduced in step (2) (multiple alignment) and step (4) (specificity checking) of the design algorithm. These are the most time-consuming tasks of the process. PhylArray has been tested on a cluster with 15 computing nodes: a “master” computer (management node) and 14 worker nodes. Each node is equipped with two processors (Xeon 2.67 GHz with hyperthreading) and 2 Gb of RAM. For the parallelization of multiple sequence alignment, we used ClustalW-MPI (Li, 2003). It is a distributed and parallel implementation of ClustalW, which uses the MPI library (Message Passing Interface) and runs on parallel architectures (a computing cluster in our case). The fourth step of our algorithm (search for a specific probe within the consensus sequence) has been parallelized using the fact that the specificity tests of the different probes extracted from the consensus sequence are totally independent. To partition the computation on p machines, we only need to split the consensus sequence in p parts. The databases used for specificity checking are sent to each computing node. For the implementation on the computing cluster, we use OpenPBS (<http://www.openpbs.org/>) to submit the jobs to the computing nodes. It is a flexible batch queuing system, which is easy to use when submitting independent jobs to a computing cluster. PhylArray is available via a Web interface at <http://fc.isima.fr/~rimour/phylarray/>. It is necessary to register to obtain a login and password. The user can access the PhylArray interface and submit jobs. This part of the application consists of a Web server (Apache) and is written in XHTML and PHP. Perl scripts are used to communicate with the master node of the cluster. All information concerning user

management and submitted jobs are stored in a relational database (MySQL). Figure 2 in supplementary data, presents a schematic overview of the PhylArray architecture.

5 EXPERIMENTAL VALIDATION

In order to evaluate the efficiency of probes selected with PhylArray, we compared them to oligonucleotides generated by another design program (PRIMROSE) and retrieved from a probe database (ARB Probe Library), both in common use for the design of probes (Freitag *et al.*, 2005). The GoArrays strategy, mentioned in the introduction of this paper, was used in order to increase the efficiency of short oligonucleotide probes. We focused on the design of oligonucleotides targeted to SSU rRNA of the following genera: *Staphylococcus*, *Micrococcus*, *Aeromonas*, *Pseudomonas*, *Streptomyces*, *Rhodococcus* and *Bifidobacterium*. 265 oligonucleotides targeting these groups were selected and spotted on a prototype microarray in order to evaluate them with nucleic acid extracts from pure bacterial cultures.

5.1 Probe selection

The selection of 10 probes was performed using PRIMROSE and the supplied SSU_Prok database including archaeal and bacterial sequences (Cole *et al.*, 2005). As advised, 15 to 20 SSU rRNA sequences belonging to the targeted genus have been autonomously selected in the database. The probe length was set to 40 or 50 nucleotides and the number of degenerate bases has been increased to 10 bases. 14 probes have been selected from the ARB Probe Library (20 bases long). PhylArray has been used to design 172 long probes (50-mers) and 34 short probes (25-mers). In order to increase the efficiency of small probes, the GoArrays strategy has been used on short probes. 31 long oligonucleotides (56-mers) were created by combining short probes proposed by PhylArray, and 12 oligonucleotides (46-mers) were designed by concatenating short probes selected from the ARB Probe Library. Oligonucleotide sequences and characteristics are available in the supplementary data (Tables 1-2-3).

5.2 Experimental procedure

For microarray production, oligonucleotides probes were synthesized with a 5' amino linker modification and spotted on Corning® GAPS II Coated Slides by Eurogentec. Each oligonucleotide was spotted in triplicate. Total RNA was extracted from pure cultures of *Staphylococcus xylosus* (DSM20266), *Enterococcus faecalis* (DSM20478), *Micrococcus*

antarcticus (JCM11467), *Micrococcus lylae* (DSM20315), *Nesterenkonia sandarakina* (DSM15664) and *Aeromonas sp.* (laboratory strain) using the RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. Total RNAs from a polluted soil were obtained using a modified protocol originally described by Fleming *et al.* (1998). Nucleic acid extraction is followed by a purification using phenol/chloroform and with the RNeasy Mini kit. To obtain labeled targets, the SSU rRNA fraction (125ng) was reverse transcribed (2h at 42°C) in a 20µl final volume using dNTPs from Invitrogen (0.25mM), RNasin+ from Promega (1U), SuperScriptIII (100U) and its associated buffer (1X) from Invitrogen, DTT (0.1M) and 0.625µM of the following primer (the bold part of the oligonucleotides allows the formation of a T7 promoter): R1406-T7: 5'-TAATACGACTCACTATAGGTACTACGGGCGGTGWGTRCAA-3' The second strand was created using all the neo-synthesized heteroduplex, dNTPs (0.3µM), Ribonuclease H (1U), E. coli DNA Polymerase I (20U) and its associated buffer (1X), E. coli Ligase (5U) (all these products are provided by Invitrogen) in 100µl final volume. In vitro transcription reaction allows the incorporation of amino allyl UTPs (final concentration 5mM) for the indirect labeling of the targets and was conducted using MEGAscript kit (Ambion) according to the manufacturer's instructions. Then, the labeling of the amplified SSU rRNA was done by coupling the amino modified aRNAs to the fluorescent dye Cy3 or Cy5 (0.5mM) by incubation of the aRNAs with the succinimidyl ester-derivitized reactive free dye (Cy3 or Cy5 mono-reactive NHS-ester) in a coupling buffer (1X) from Ambion. This reaction was performed in the obscurity for 1h at room temperature and was stopped with the addition of 1.3M of hydroxylamine. Hybridizations were carried out in a 25µl final volume (17µL of DigEasy buffer from Boehringer, 500ng of labeled aRNAs, 1.4µg of salmon sperm DNA and 6nM of doubly labeling-Cy5/Cy3-reference oligonucleotide) at 37°C for 2h in a TrayMixS1 hybridization chamber (Biotray; <http://www.biotray.fr/>). After hybridization, the microarrays were washed two times at room temperature during 5min with the following solutions (solution1: 0.2X SSC, 0.1% SDS; solution2: 0.2X SSC). The slides were then scanned on the Affymetrix 428 Array scanner to detect Cy3 and Cy5 fluorescence. Raw data analysis was carried out using a tool of the TM4 software suite: TIGR Spotfinder 3.1.0 (Saeed *et al.*, 2003). Spot segmentation was done with the Otsu method using 10 to 30px for the searched diameter. A median for each spot triplicate was calculated both in Cy5 and Cy3.

5.3 Probe efficiency analysis

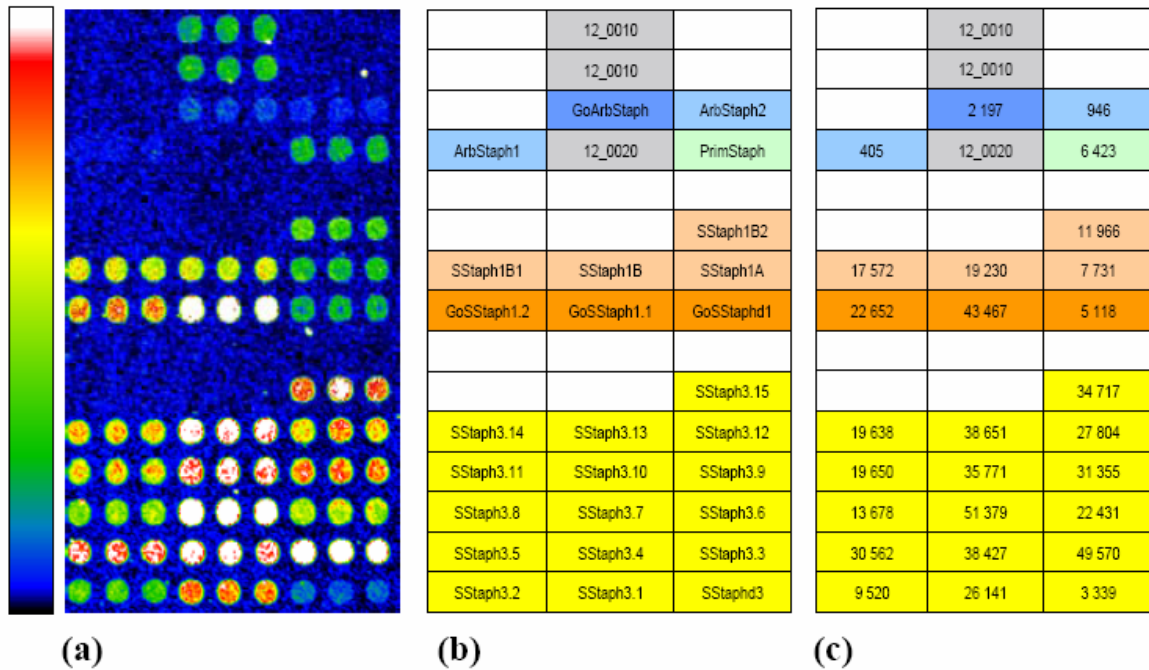


Fig. 3. Sensitivity of designed *Staphylococcus* probes. (a) : Microarray image obtained after hybridization of labeled 16S rRNAs of *Staphylococcus xylosus*. (b) : location of probes targeting *Staphylococcus* and control probes on the microarray. Probes were designed with the following software: PhylArray (50-mers: yellow and 25-mers: light orange), PRIMROSE (green), ARB probes (light blue) and the GoArrays strategy applied on Arb probes (dark blue) and PhylArray probes (dark orange). (c) : Measurements of signal intensities (Fluorescence Units) with the TIGR Spotfinder program. 12_0010 is the positive control oligonucleotide (the complementary oligonucleotide, doubly labeled, is added to the hybridization mix). 12_0020 is the negative control oligonucleotide. All probes are spotted in triplicate.

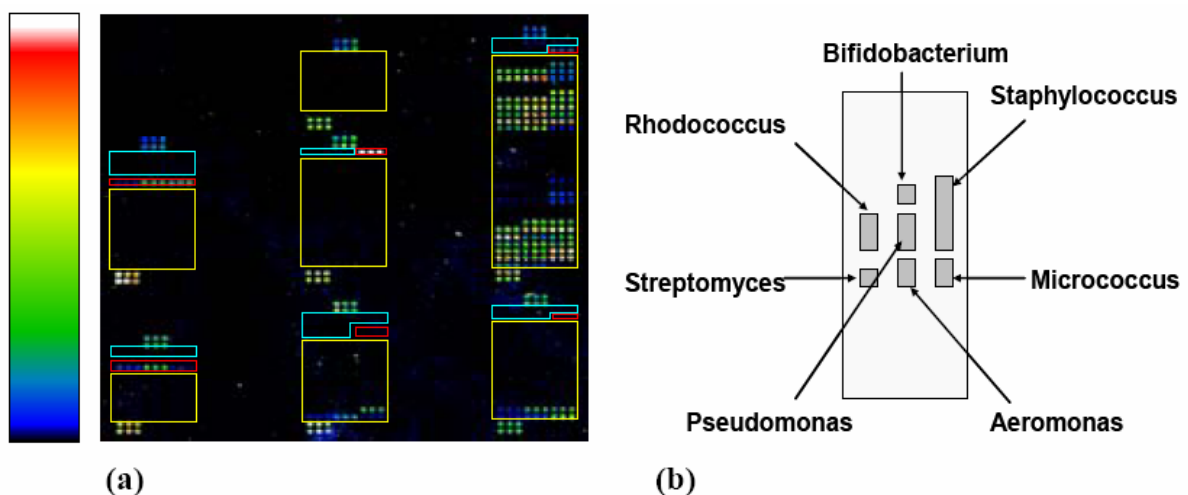


Fig. 4. Specificity evaluation of the probe design. Probes were designed with standard methods as PRIMROSE (red), ARB Probe Library (blue) and with a new method named PhylArray (yellow) (a). These oligonucleotides have been spotted on the microarray (b) and have been hybridized with 16S rRNA from *Staphylococcus xylosus* labeled with Cy5.

5.3.1. Probe sensitivity.

Figure 3 shows a section of the microarray with probes targeting *Staphylococcus* and the signals obtained after hybridization with labelled SSU rRNA extracted from the bacterial strain *Staphylococcus xylosus*. These targets hybridize with all oligonucleotides targeting this genus (Figure 3a). Nevertheless, significant differences in signal intensities can be observed according to the methods used for probes selection (Figure 3c). Probes designed with PhylArray are more sensitive than PRIMROSE and ARB ones (except, as expected for highly degenerate oligonucleotides like SStaphd3 showing a degeneracy of 1536). For long probes, signals of PhylArray oligonucleotides (from 9 520 to 51 379 FU (Fluorescence Unit)) are higher than the PRIMROSE ones (6 423 FU). This observation is the same for short PhylArray and ARB probes with respectively 7 731 to 17 572 FU and 405 to 946 FU. Moreover, hybridization analysis also shows that the use of the GoArrays strategy on short probes on average trebles their signal (for example GoArbStaph increase the signal by 2.3 to 5.4 fold).

5.3.2. Probe specificity.

Microarray analysis (Figure 4) shows that probes generated with PRIMROSE, the ARB Probe Library and PhylArray (50-mers probes) are not strictly specific. Actually, all PRIMROSE probes selected to target *Pseudomonas*, *Streptomyces* and *Rhodococcus* genera also recognize *Staphylococcus xylosus* targets. The SSU rRNA of *S. xylosus* does also cross-react with all probes from the ARB Probe Library targeting *Micrococcus* and *Aeromonas* and 50% of probes designed PhylArray (50-mers) that target *Micrococcus* and *Aeromonas*. The most important rate of cross-hybridization with *S. xylosus* targets is obtained for PRIMROSE probes (78%). Oligonucleotides selected with PhylArray (50-mers) and ARB have lower rates of false-positive hybridizations (respectively 8% and 1.5%). Moreover, false-positive intensities of ARB probes are clearly weaker and disappeared when the GoArrays strategy was used. PhylArray (25-mers) associated or not with the GoArrays strategy allowed the generation of more specific probes. Similar results have been obtained with targets extracted from others reference bacteria: *Micrococcus antarcticus*, *Enterococcus faecalis*, *Micrococcus lylae*, *Aeromonas sp.* and *Nesterenkonia sandarakina* (results not shown). PRIMROSE probes always show highest cross-hybridization rates (57 to 90%). PhylArray (50-mers) oligonucleotides have lower cross-hybridization percentages (0.05 to 46%) but quite higher compared to ARB probes (0 to 42%) and PhylArray short ones (3 to 28%). Depending on the tested bacterial strain, PhylArray short probes and ARB probes are more specific.

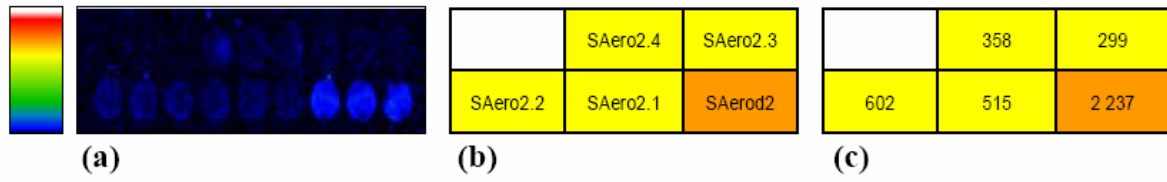


Fig. 5. Potential explorative power of PhylArray probes. Degenerate probes are designed by the PhylArray program allowing potential recognition of all species of the targeted genus, even the unknown fraction. (a) The microarray image shows hybridization of *Aeromonas*-targeting probes with bacterial targets extracted from a polluted soil. (b) Location of the degenerate probe SAerod2 and the specific probes SAero2.1-2.4. (c) Signal intensities (Fluorescence Units) are determined with TIGR Spotfinder 3.1.

Hybridization analyses have also shown that the use of the GoArrays strategy on short probes often allows the removal of false-positive signals (*Aeromonas sp.*, *Neisterenkonia sandarakina*, *Micrococcus antarcticus*, *Micrococcus lylae*). Furthermore, detection lacks for *M. lylae* have been observed using PRIMROSE probes. Oligonucleotides from the ARB probe library did not detect *M. lylae* and *Aeromonas sp.* PhylArray is the only design method allowing a sufficient recognition level for all the tested strains.

5.3.3. Explorative power.

Figure 5 shows an element of the microarray image (*Aeromonas* related part) obtained after hybridization with labelled SSU rRNA extracted from a polluted soil. The fluorescence of the degenerate probe (SAero2) is higher than non-degenerate probes fluorescence (SAero2.1, SAero2.2, SAero2.3 and SAero2.4). Thus unknown species belonging or close to the genus *Aeromonas* could be present in this complex environment.

6 DISCUSSION

In this paper, we have shown that conventional approaches to design microarray probes do not always allow the generation of efficient biosensors for studying complex environments accurately. The first critical point is the sensitivity of probes. In fact it is important to monitor all representatives of a bacterial community, even the low abundant ones, which also have an impact on the functionality of an ecosystem. In this study, the comparison of sensitivities between probes designed with PhylArray, PRIMROSE and those chosen in the ARB Probe Library has shown a better efficiency of longer probes. Fluorescence intensities of PhylArray probes are up to 20-fold higher than those obtained with PRIMROSE or selected from the ARB Probe Library (respectively compared to PhylArray 50- and 25-mers). It is well known that the presence and the positions of mismatches in the probe or target sequences can affect the signal intensity value (Urakawa *et al.*, 2003). In spite of the fact that ARB probes (20 mers) are smaller than PhylArray probes (25 and 50 mers), the sensitivity differences could also be somewhat explained by the localization of the hybridization sites on the targeted molecule (SSU rRNA). Fuchs *et al.* (1998) have demonstrated that probe accessibility is variable according to the binding region. If we localize probe hybridization sites on the drawn SSU rRNA accessibility map, we can observe that binding regions of PhylArray probes (SStaph3.x; position 50, SStaph1.x; position 650) could be more accessible compared to ARB (ArbStaph; positions 100 and 200) and PRIMROSE (PrimStaph; position 300) probes. Other

parameters could influence the hybridization efficiency. Several investigations have demonstrated the potential implication of thermodynamic properties of nucleic acids in the target-probe duplex formation and dissociation (e.g. secondary structures of SSU rRNA, intra- and inter-self structure of probes) that could be used to predict the probe efficiency (Gentry *et al.*, 2006). However, Pozhitkov *et al.* (2006) have recently shown that these thermodynamic parameters are only weakly correlated with probe efficiency. Our results demonstrated that probe sensitivity is influenced by the probe length, the perfect match between probes and targets, the location of the mismatches (Table 4 in supplementary data), the accessibility of the target and other complex thermodynamic parameters not yet well understood. The second critical point is the specificity. The design of probes only recognizing a defined taxon is a real challenge due to the high conservation of the SSU rDNA biomarker within members of the bacterial domain and the presence of unknown bacteria in the studied samples (Gentry *et al.*, 2006). We have observed specificity differences between different long probes (PRIMROSE vs. PhylArray 50-mers) as well as between short ones (ARB Probes Library vs. PhylArray 25-mers). Long oligonucleotide probes designed with PRIMROSE show a significantly high cross-hybridization rate, which is based on their high complementarity to non-targeted bacterial sequences. The sequences targeted by the PRIMROSE probes specific for *Streptomyces*, *Pseudomonas* and *Rhodococcus* and the sequence of *S. xylosum* SSU rRNA have a high similarity of up to 94%. This value exceeds by far Kane's criteria (Kane *et al.*, 2000) and is probably responsible for the observed cross-hybridization with *S. xylosum*. PhylArray 50-mer probes targeting *Streptomyces*, *Pseudomonas* and *Rhodococcus* are more specific because their target sequences are dissimilar enough to not be able to bind to *S. xylosum* SSU rRNA. The reason for not finding these specific regions with PRIMROSE could be in its degeneracy setting limitation of 10 degenerate nucleotides per probe. This constraint restricts the search for probes to more conserved regions which could create cross-hybridizations with others taxons. PhylArray identifies probes in less conserved regions that are more specific for a given taxon. The limitation of the degeneracy rate in PRIMROSE has been certainly done in order to reduce computing constraints. In fact, PRIMROSE generates all the combinations of oligonucleotides from the consensus sequence and checks the specificity of each one. Thus, this process is time and space-consuming if the consensus sequence is highly degenerated. The originality of our method lies in the specificity test of PhylArray probes which avoids problems due to a high degeneracy. In order to perform this test, the algorithm does not generate all possible sequences from the degenerate probe but only those used to create the consensus. Thus, the set of non-degenerate probes we obtain is

highly specific for the target taxon. Another restrictive issue is the homology threshold used to define a potential cross-hybridization with a non-targeted sequence. The PRIMROSE user can modify this setting while varying the number of tolerated mismatches (up to 7 mismatches). However, for long probes (50-mers), this threshold corresponding to 86% of similarity, is insufficient because cross-hybridization events can occur even at 75% as described by Kane (Kane *et al.*, 2000). PhylArray uses by default Kane's criteria: 75% similarity (1) and 15 identical contiguous bases (2), but if it is too restrictive the user can set individual parameter values. Compared to PRIMROSE and PhylArray long probes, short oligonucleotide probes designed with PhylArray or selected from the ARB Probe Library have shown smaller cross-hybridization rates (respectively 3 to 28% and 0 to 42%). These results are in accordance with Kane's criteria: the longer the oligonucleotide, the easier condition (1) is satisfied. However, to satisfy condition (2) the situation is more complex since as the length increases, the probability of finding a stretch of 15 identical bases might also increase. These observed cross-hybridizations are also explained by their similarities with non-targeted bacterial sequences. These false-positive signals seem to be unavoidable in a context of probes designed to target the SSU rRNA of closely related bacteria. However, we have shown a successful alternative by applying the GoArrays strategy. By concatenating suitable short oligonucleotide probes we could minimize or even avoid false-positive signals in most cases. PhylArray 25-mer probes as well as probes selected from the ARB Probe Library showed high specificity. This could be explained by greater destabilization of non-perfect probe-target duplexes due to the constraint of loop formation on the target during the hybridization with the probe. It is furthermore important to mention that the signal intensities increased significantly using this approach. It is of course important for probes to avoid cross-hybridizations but it is also essential to recognize the target taxon. Hybridization analyses have highlighted a recognition issue for probes designed by PRIMROSE or selected from the ARB Probe Library in order to target *Micrococcus*. These oligonucleotides did not detect strains of *M. lylae* and *M. antarcticus*. This lack could be due to the difficulty to design probes targeting higher phylogenetic levels than the species. This kind of design implies that the polymorphism of the targeted group is taken into account in order to recognize all its components. For polymorph taxa (e.g. *Micrococcus*), it is necessary to use highly degenerate oligonucleotides to cover all the group diversity. The ARB Probe Library and PRIMROSE do not allow the design of such probes because only 0 and 10 degenerate bases are permissible respectively. On the contrary, PhylArray allows the generation of highly degenerate probes in order to cover the polymorphism of all taxa. This could be the reason why no recognition

problem has been detected for PhylArray probes. The last critical point is the explorative challenge. Available design programs only generate probes for the known microbial fraction. It does not allow a global view of microbial communities and is mostly insufficient for comprehensive biological interpretations. Another specificity of our algorithm is to ensure an explorative process in the study of bacterial communities. The combined application of degenerate and non-degenerate probes can highlight the presence of bacteria which are not referenced in sequence databases if the spot composed of degenerate probes is the only one that give a hybridization signal. Hybridizations with environmental targets have suggested this potentiality (see results for SAerod2, SAero2.1, SAero2.2, SAero2.3 and SAero2.4). Unknown species belonging or close to the genus *Aeromonas* could be present in this complex environment. Biological experiments will help us to characterize these strains in order to validate this potential explorative power.

7 CONCLUSION

In summary, we present here a new probe design software tool called PhylArray, used to generate oligonucleotide probes for microarray analysis, allowing the targeting of the genus or higher taxonomic levels. PhylArray produces specific and sensitive probes which cover the polymorphism of the targeted taxon owing to the use of a high level of degeneracy. Moreover, the combined application of highly degenerate probes and associated non-degenerate probes even allows the exploration of the unknown part of bacterial communities. This exciting possibility could help us to create a better understanding of how microbial communities are functioning.

REFERENCES

- Amann, R.I., Ludwig, W. and Schleifer, K.H.** (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation, *Microbiol Rev*, 59, 143-169.
- Ashelford, K.E., Weightman, A.J. and Fry, J.C.** (2002) PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database, *Nucleic Acids Res*, 30, 3481-3489.
- Bodrossy, L. and Sessitsch, A.** (2004) Oligonucleotide microarrays in microbial diagnostics, *Curr Opin Microbiol*, 7, 245-254. Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., Garrity, G.M. and Tiedje, J.M. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis, *Nucleic Acids Res*, 33, D294-296.
- Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Bandela, A.M., Cardenas, E., Garrity, G.M. and Tiedje, J.M.** (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data, *Nucleic Acids Res*, 35, D169-172.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L.** (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, *Appl Environ Microbiol*, 72, 5069-5072.
- Dunbar, J., Takala, S., Barns, S.M., Davis, J.A. and Kuske, C.R.** (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning, *Appl Environ Microbiol*, 65, 1662-1669.
- Ehrenreich, A.** (2006) DNA microarray technology for the microbiologist: an overview, *Appl Microbiol Biotechnol*, 73, 255-273.
- Fleming, J.T., Yao, W.H. and Sayler, G.S.** (1998) Optimization of differential display of prokaryotic mRNA: application to pure culture and soil microcosms, *Appl Environ Microbiol*, 64, 3698-3706.
- Franke-Whittle, I.H., Klammer, S.H. and Insam, H.** (2005) Design and application of an oligonucleotide microarray for the investigation of compost microbial communities, *J Microbiol Methods*, 62, 37-56.
- Freitag, T.E., Chang, L., Clegg, C.D. and Prosser, J.I.** (2005) Influence of inorganic nitrogen management regime on the diversity of nitrite-oxidizing bacteria in agricultural grassland soils, *Appl Environ Microbiol*, 71, 8323-8334.
- Fuchs, B.M., Wallner, G., Beisker, W., Schwiapl, I., Ludwig, W. and Amann, R.** (1998) Flow cytometric analysis of the in situ accessibility of Escherichia coli 16S rRNA for fluorescently labeled oligonucleotide probes, *Appl Environ Microbiol*, 64, 4973-4982.
- Gans, J., Wolinsky, M. and Dunbar, J.** (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil, *Science*, 309, 1387-1390.

Gentry, T.J., Wickham, G.S., Schadt, C.W., He, Z. and Zhou, J. (2006) Microarray applications in microbial ecology research, *Microb Ecol*, 52, 159-175.

He, Z., Wu, L., Li, X., Fields, M.W. and Zhou, J. (2005) Empirical establishment of oligonucleotide probe design criteria, *Appl Environ Microbiol*, 71, 3753-3760.

Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D. and Madore, S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays, *Nucleic Acids Res*, 28, 4552-4557.

Li, K.B. (2003) ClustalW-MPI: ClustalW analysis using distributed and parallel computing, *Bioinformatics*, 19, 1585-1586.

Li, X., He, Z. and Zhou, J. (2005) Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation, *Nucleic Acids Res*, 33, 6114-6123.

Loy, A., Schulz, C., Lucker, S., Schopfer-Wendels, A., Stoecker, K., Baranyi, C., Lehner, A. and Wagner, M. (2005) 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales", *Appl Environ Microbiol*, 71, 1373-1386.

Loy, A. and Bodrossy, L. (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays, *Clin Chim Acta*, 363, 106-119.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. and Schleifer, K.H. (2004) ARB: a software environment for sequence data, *Nucleic Acids Res*, 32, 1363-1371.

Nordberg, E.K. (2005) YODA: selecting signature oligonucleotides, *Bioinformatics*, 21, 1365-1370.

Peplies, J., Lachmund, C., Glockner, F.O. and Manz, W. (2006) A DNA microarray platform based on direct detection of rRNA for characterization of freshwater sediment-related prokaryotic communities, *Appl Environ Microbiol*, 72, 4829-4838.

Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Nolte, A.W., Sonnenberg, R., Staehler, P., Beier, M. and Tautz, D. (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted, *Nucleic Acids Res*, 34, e66.

Reymond, N., Charles, H., Duret, L., Calevro, F., Beslon, G. and Fayard, J.M. (2004) ROSO: optimizing oligonucleotide probes for microarrays, *Bioinformatics*, 20, 271-273.

Rimour, S., Hill, D., Milton, C. and Peyret, P. (2005) GoArrays: highly dynamic and efficient microarray probe design, *Bioinformatics*, 21, 1094-1103.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. and Quackenbush, J. (2003) TM4: a free, open-source system for microarray data management and analysis, *Biotechniques*, 34, 374-378.

Sanguin, H., Herrera, A., Oger-Desfeux, C., Dechesne, A., Simonet, P., Navarro, E., Vogel, T.M., Moenne-Loccoz, Y., Nesme, X. and Grundmann, G.L. (2006) Development and validation of a prototype 16S rRNA-based taxonomic microarray for Alphaproteobacteria, *Environ Microbiol*, 8, 289-307.

Schliep, A. and Rahmann, S. (2006) Decoding non-unique oligonucleotide hybridization experiments of targets related by a phylogenetic tree, *Bioinformatics*, 22, e424-430.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res*, 22, 4673-4680.

Tiquia, S.M., Wu, L., Chong, S.C., Passovets, S., Xu, D., Xu, Y. and Zhou, J. (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples, *Biotechniques*, 36, 664-670, 672, 674-665.

Torsvik, V. and Ovreas, L. (2002) Microbial diversity and function in soil: from genes to ecosystems, *Curr Opin Microbiol*, 5, 240-245.

Urakawa, H., El Fantroussi, S., Smidt, H., Smoot, J.C., Tribou, E.H., Kelly, J.J., Noble, P.A. and Stahl, D.A. (2003) Optimization of single-base-pair mismatch discrimination in oligonucleotide microarrays, *Appl Environ Microbiol*, 69, 2848-2856.

Wuyts, J., Perriere, G. and Van De Peer, Y. (2004) The European ribosomal RNA database, *Nucleic Acids Res*, 32, D101-103.

SUPPLEMENTARY DATA

```

AF105024 ...CGTAGAGATACGGTTTCCCCTTTGGGGCGGG...
AF057289 ...CGTAGAGATACGGTTTCCCCTTTGGGGCGGG...
AB023371 ...CGTAGAGATACGGTTTCCC-TTTGGGGCGGG...
X86609 ...CGTAGAGATACGGTTTCC--TTTGGGGCGGG...
X86612 ...CGTAGAGATACGGTTTCC--TTTGGGGCGGG...
X80750 ...CGTAGAGATACGGTTTCCC-TTTGGGGCGGG...
AF196342 ...CGTAGAGATACGGTTTCCC-TTTGGGCCGGG...
AF057290 ...TCCAGAGATGGTTCTTCCCCTTTGGGGTCGG...
AF196343 ...TCCAGAGATGGTTCTTCCCCTTTGGGGTCGG...
          *****      ***      *****      **
Consensus...YSYAGAGATRSKKYTTCCC-TTTGGGSYSGG...
  
```

Fig. 1. Example of steps (1)-(2)-(3) of the design algorithm. The sequences of the target taxon are extracted from the rRNA database (represented by their accession number) and aligned with ClustalW (an extract of the alignment is represented on the figure). Then, a consensus sequence is generated, representative of the target taxon

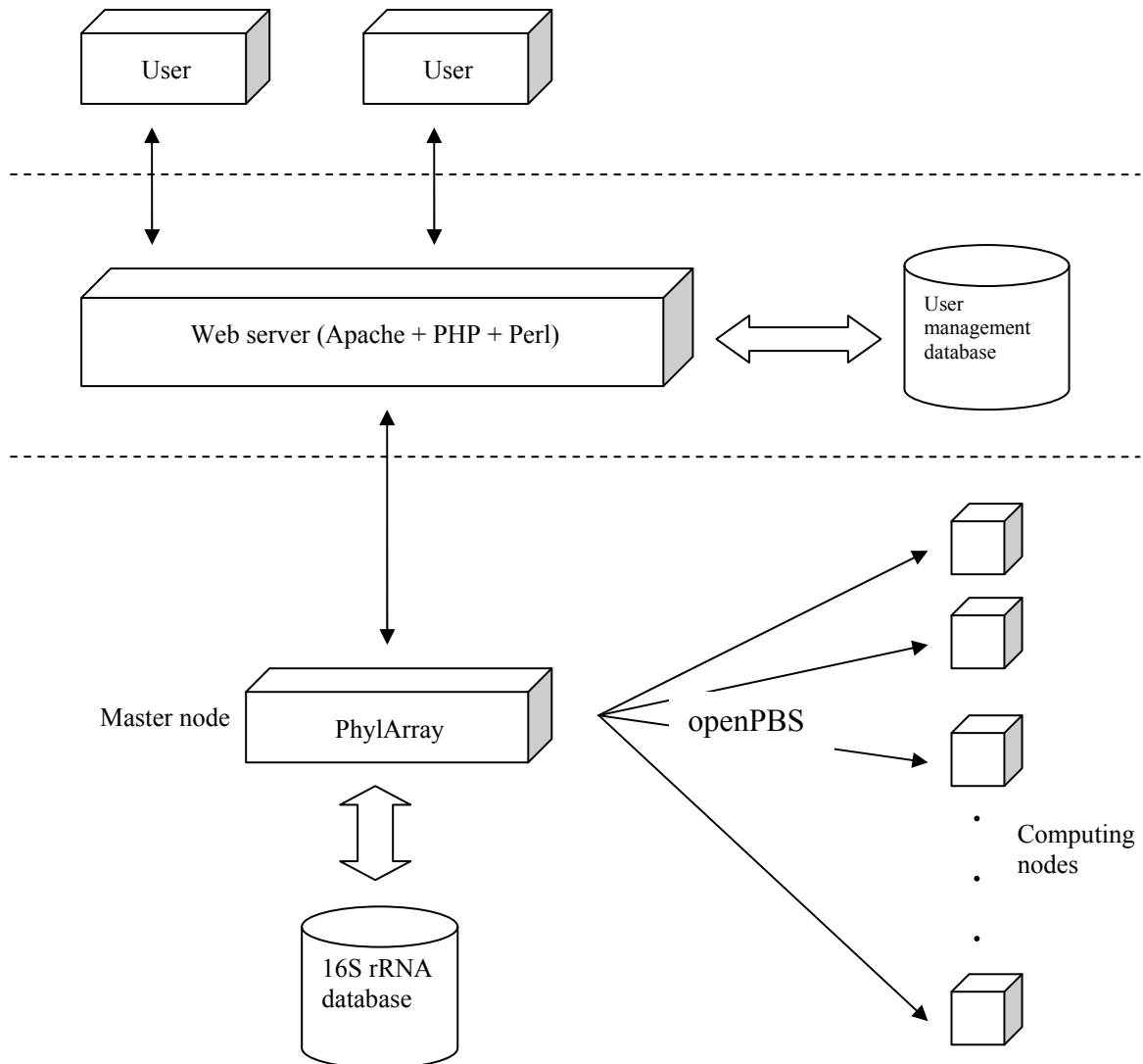


Fig. 2. Simplified schematic overview of the PhylArray software structure.

Table 1. Characteristics of 40-mers and 50-mers oligonucleotides, designed with Primrose, targeting 16S rRNA genes of *Micrococcus*, *Aeromonas*, *Streptomyces*, *Staphylococcus*, *Pseudomonas* and *Rhodococcus* genera.

Probe Name	Intended genus	Sequence 5'→3'	Position within 16S rDNA	Tm °C	Lg (mer)
PrimAero	<i>Aeromonas</i>	TTCGGGTTGTAAAGCACTTTCAGYGAGGAGGAAAGGTYRD	415	62	40
PrimMicro	<i>Micrococcus</i>	TCGGTYTTGGATGGACTCGCGGCCYATCAGCTTGTTGGTGAGGTAATGGC	200	67	50
PrimPseudo	<i>Pseudomonas</i>	GTGGGCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGAT	1130	71	50
PrimRhodo1	<i>Rhodococcus</i>	GTAGCRAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACCGGTGGGC	750	67	50
PrimRhodo2	<i>Rhodococcus</i>	TAGCRAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACCGGTGGGCR	740	66	50
PrimRhodo3	<i>Rhodococcus</i>	TTAGATACCCTGGTAGTCCACGCCGTAACCGGTGGGCRCTAGGTGTGGGT	760	67	50
PrimStaph	<i>Staphylococcus</i>	GGTCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGRCG	310	68	50
PrimSmyces1	<i>Streptomyces</i>	GGGACTCACRGGAGACNGCCGGGGTCAACTCGGAGGAAGGTGGGGACGA	1110	71	50
PrimSmyces2	<i>Streptomyces</i>	AAGGTGGGGACGACSTNAAGTCATCATGCCCTTATGTCTGGGCTGCAC	1150	67	50
PrimSmyces3	<i>Streptomyces</i>	CATGCCCTTATGTCTGGGCTGCACANGTGCTACAATGGCMGGTACAAT	1180	66	50

Table 2. Characteristics of 20-mers and 46-mers oligonucleotides, designed respectively with Arb and GoArrays strategy applied on Arb probes, targeting 16S rRNA genes of *Micrococcus*, *Aeromonas*, *Streptomyces*, *Staphylococcus*, *Pseudomonas* and *Rhodococcus* genera.

Probe Name	Intended genus	Sequence 5'→3'	Position within 16S rDNA	Tm °C	Lg (mer)
UNIV1390	Universal	GACGGGCGGTGTGTACAA	1390	66	18
EUB338	Eubacteria	GCTGCCTCCCGTAGGAGT	338	70	18
ArbAero1	<i>Aeromonas</i>	GAGCGAATCCCAAAAAGCGC	550	49	20
ArbAero2	<i>Aeromonas</i>	ACCTGGGAATTGCATTTAAA	190	41	20
ArbAero3	<i>Aeromonas</i>	GCGTACAGAGGGCTGCAAGC	1240	53	20
GoArbAero1	<i>Aeromonas</i>	ACCTGGGAATTGCATTTAAAGGATCCGCGTACAGAGGGCTGCAAGC	190-1240	67	46
GoArbAero2	<i>Aeromonas</i>	ACCTGGGAATTGCATTTAAAGGATCCGAGCGAATCCCAAAAAGCGC	190-550	65	46
GoArbAero3	<i>Aeromonas</i>	GCGTACAGAGGGCTGCAAGCGGATCCGAGCGAATCCCAAAAAGCGC	1240-550	70	46
ArbMicro1	<i>Micrococcus</i>	CCAAGGCTTGACATGTTCTC	950	47	20
ArbMicro2	<i>Micrococcus</i>	TCGTGAAAAGTCCGGGGCTTA	560	49	20
GoArbMicro	<i>Micrococcus</i>	TCGTGAAAAGTCCGGGGCTTAGGATCCCCAAGGCTTGACATGTTCTC	560-950	68	46
ArbPseudo	<i>Pseudomonas</i>	GATGAGCCTAGGTCCGATTA	215	47	20
ArbRhodo1	<i>Rhodococcus</i>	AAAACCCGGGGCTCAACTTC	570	49	20
ArbRhodo2	<i>Rhodococcus</i>	AAACAACCTGACGCTGAGGAA	710	45	20
ArbRhodo3	<i>Rhodococcus</i>	GGATCCGTGCCGTAGCTAAC	820	51	20
ArbRhodo4	<i>Rhodococcus</i>	GGCGGTTTGTCGCGTCGTTT	610	51	20
GoArbRhodo1	<i>Rhodococcus</i>	AAAACCCGGGGCTCAACTTCGGATCCAAACAACCTGACGCTGAGGAA	570-710	67	46
GoArbRhodo2	<i>Rhodococcus</i>	AAAACCCGGGGCTCAACTTCGGATCCGGATCCGTGCCGTAGCTAAC	570-820	69	46
GoArbRhodo3	<i>Rhodococcus</i>	GGCGGTTTGTCGCGTCGTTTGGATCCAAAACCCGGGGCTCAACTTC	610-570	69	46
GoArbRhodo4	<i>Rhodococcus</i>	GGCGGTTTGTCGCGTCGTTTGGATCCAAACAACCTGACGCTGAGGAA	610-710	68	46
GoArbRhodo5	<i>Rhodococcus</i>	GGCGGTTTGTCGCGTCGTTTGGATCCGGATCCGTGCCGTAGCTAAC	610-820	70	46
GoArbRhodo6	<i>Rhodococcus</i>	AAACAACCTGACGCTGAGGAAGGATCCGGATCCGTGCCGTAGCTAAC	710-820	68	46
ArbStaph1	<i>Staphylococcus</i>	ACGTGGATAACCTACCTATA	100	43	20
ArbStaph2	<i>Staphylococcus</i>	CTATCACTTATAGATGGACC	210	43	20
GoArbStaph	<i>Staphylococcus</i>	ACGTGGATAACCTACCTATAGGATCCCTATCACTTATAGATGGACC	100-210	63	46
ArbSmyces1	<i>Streptomyces</i>	GGGAGATCGGAATTCCTGGT	630	49	20
ArbSmyces2	<i>Streptomyces</i>	CAAGCCCTGGAAACGGGGTC	870	53	20
GoArbSmyces	<i>Streptomyces</i>	CAAGCCCTGGAAACGGGGTCGGATCCGGGAGATCGGAATTCCTGGT	630-870	70	46

Table 3. Characteristics of 25-mers and 56-mers oligonucleotides, designed respectively with PhylArray and GoArrays strategy applied on PhylArray probes, targeting 16S rRNA genes of *Micrococcus*, *Aeromonas*, *Streptomyces*, *Staphylococcus*, *Pseudomonas*, *Bifidobacterium* and *Rhodococcus* genera.

Name	Intended genus	Sequence 5'→3'	Position within 16S rDNA	Tm °C	Lg (mer)
SAerod1	<i>Aeromonas</i>	AATCGGRATTACTGGGCGTAAAGCGCACGCASGCGTTGGATAAGTTAGA	550	66	50
SAero1.1	<i>Aeromonas</i>	AATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGTTGGATAAGTTAGA	550	67	50
SAero1.2	<i>Aeromonas</i>	AATCGGAATTACTGGGCGTAAAGCGCACGCACGCGTTGGATAAGTTAGA	550	67	50
SAero1.3	<i>Aeromonas</i>	AATCGGGATTACTGGGCGTAAAGCGCACGCAGGCGTTGGATAAGTTAGA	550	68	50
SAerod2	<i>Aeromonas</i>	AAAGSAGKSGACCTTCGGGCCTTKCGCGATTGGATRWRCCAGGTGGGAT	190	68	50
SAero2.1	<i>Aeromonas</i>	AAAGGAGGGGACCTTCGGGCCTTTCGCGATTGGATGAACCCAGGTGGGAT	190	70	50
SAero2.2	<i>Aeromonas</i>	AAAGCAGGGGACCTTCGGGCCTTGC GCGATTGGATGAACCCAGGTGGGAT	190	71	50
SAero2.3	<i>Aeromonas</i>	AAAGCAGGGGACCTTCGGGCCTTGC GCGATTGGATATGCCCAGGTGGGAT	190	71	50
SAero2.4	<i>Aeromonas</i>	AAAGCAGTCGACCTTCGGGCCTTGC GCGATTGGATGAACCCAGGTGGGAT	190	70	50
SAero1-2A	<i>Aeromonas</i>	AAGCGCGTCGTAGTCCGGATYGGAG	1280	59	25
SAero1-2A1	<i>Aeromonas</i>	AAGCGCGTCGTAGTCCGGATCGGAG	1280	59	25
SAero1-2A2	<i>Aeromonas</i>	AAGCGCGTCGTAGTCCGGATTGGAG	1280	58	25
SAero1B	<i>Aeromonas</i>	CGTACAGAGGGCTGCAAGCTAGCGA	1240	58	25
SAero2B	<i>Aeromonas</i>	CCGTCACACCAYGGGAGTGGGTTGC	1400	59	25
SAero2B1	<i>Aeromonas</i>	CCGTCACACCATGGGAGTGGGTTGC	1400	59	25
SAero2B2	<i>Aeromonas</i>	CCGTCACACCACGGGAGTGGGTTGC	1400	61	25
GoSAerod1	<i>Aeromonas</i>	CGTACAGAGGGCTGCAAGCTAGCGAGGATCCAAGCGCGTCGTAGTCCGGATYGGAG	1240-1280	70	56
GoSAero1.1	<i>Aeromonas</i>	CGTACAGAGGGCTGCAAGCTAGCGAGGATCCAAGCGCGTCGTAGTCCGGATCGGAG	1240-1280	73	56
GoSAero1.2	<i>Aeromonas</i>	CGTACAGAGGGCTGCAAGCTAGCGAGGATCCAAGCGCGTCGTAGTCCGGATTGGAG	1240-1280	73	56
GoSAerod2	<i>Aeromonas</i>	AAGCGCGTCGTAGTCCGGATYGGAGGATCCCCGTCACACCAYGGGAGTGGGTTGC	1280-1400	71	56
GoSAero2.1	<i>Aeromonas</i>	AAGCGCGTCGTAGTCCGGATCGGAGGATCCCCGTCACACCATGGGAGTGGGTTGC	1280-1400	74	56
GoSAero2.2	<i>Aeromonas</i>	AAGCGCGTCGTAGTCCGGATCGGAGGATCCCCGTCACACCACGGGAGTGGGTTGC	1280-1400	75	56
GoSAero2.3	<i>Aeromonas</i>	AAGCGCGTCGTAGTCCGGATTGGAGGATCCCCGTCACACCATGGGAGTGGGTTGC	1280-1400	73	56
GoSAero2.4	<i>Aeromonas</i>	AAGCGCGTCGTAGTCCGGATTGGAGGATCCCCGTCACACCACGGGAGTGGGTTGC	1280-1400	74	56
SBifidoA	<i>Bifidobacterium</i>	GTCAAGYCATGAAAGTGGGYAGCAC	1380	56	25
SBifidoA1	<i>Bifidobacterium</i>	GTCAAGCCATGAAAGTGGGCAGCAC	1380	56	25
SBifidoA2	<i>Bifidobacterium</i>	GTCAAGTCATGAAAGTGGGCAGCAC	1380	54	25
SBifidoA3	<i>Bifidobacterium</i>	GTCAAGCCATGAAAGTGGGTAGCAC	1380	54	25
SBifidoA4	<i>Bifidobacterium</i>	GTCAAGTCATGAAAGTGGGTAGCAC	1380	53	25
SBifidoB	<i>Bifidobacterium</i>	ATCAGCAACGYYGCGGTGAATGCGT	1330	57	25
SBifidoB1	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGT	1330	56	25
SBifidoB2	<i>Bifidobacterium</i>	ATCAGCAACGCCGCGGTGAATGCGT	1330	58	25
SBifidoB3	<i>Bifidobacterium</i>	ATCAGCAACGCTGCGGTGAATGCGT	1330	57	25
GoSBifidod1	<i>Bifidobacterium</i>	ATCAGCAACGYYGCGGTGAATGCGTGGATCCGTCAGYCATGAAAGTGGGYAGCAC	1330-1380	68	56
GoSBifido1.1	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGTGGATCCGTCAGCCATGAAAGTGGGCAGCAC	1330-1380	71	56
GoSBifido1.2	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGTGGATCCGTCAGTCATGAAAGTGGGCAGCAC	1330-1380	71	56
GoSBifido1.3	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGTGGATCCGTCAGCCATGAAAGTGGGTAGCAC	1330-1380	71	56
GoSBifido1.4	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGTGGATCCGTCAGTCATGAAAGTGGGTAGCAC	1330-1380	70	56
GoSBifido1.5	<i>Bifidobacterium</i>	ATCAGCAACGCCGCGGTGAATGCGTGGATCCGTCAGCCATGAAAGTGGGCAGCAC	1330-1380	72	56
GoSBifido1.6	<i>Bifidobacterium</i>	ATCAGCAACGCCGCGGTGAATGCGTGGATCCGTCAGTCATGAAAGTGGGCAGCAC	1330-1380	71	56
GoSBifido1.7	<i>Bifidobacterium</i>	ATCAGCAACGCCGCGGTGAATGCGTGGATCCGTCAGCCATGAAAGTGGGTAGCAC	1330-1380	71	56
GoSBifido1.8	<i>Bifidobacterium</i>	ATCAGCAACGCCGCGGTGAATGCGTGGATCCGTCAGTCATGAAAGTGGGTAGCAC	1330-1380	71	56
GoSBifido1.9	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGTGGATCCGTCAGCCATGAAAGTGGGCAGCAC	1330-1380	71	56
GoSBifido1.10	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGTGGATCCGTCAGTCATGAAAGTGGGCAGCAC	1330-1380	71	56
GoSBifido1.11	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGTGGATCCGTCAGCCATGAAAGTGGGTAGCAC	1330-1380	71	56
GoSBifido1.12	<i>Bifidobacterium</i>	ATCAGCAACGTCGCGGTGAATGCGTGGATCCGTCAGTCATGAAAGTGGGTAGCAC	1330-1380	70	56
SMicro1d	<i>Micrococcus</i>	TCGYYSYAGAGATRSKKYTTCCCTTTGGGSSYSGGTWBACAGGTGGTGCA	970	68	50

SMicro1.1	<i>Micrococcus</i>	TCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTACACAGGTGGTGCA	970	71	50
SMicro1.2	<i>Micrococcus</i>	TCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTACACAGGTGGTGCA	970	71	50
SMicro1.3	<i>Micrococcus</i>	TCGTTCCAGAGATGTTTCTCCCCTTTGGGGTCGGTATACAGGTGGTGCA	970	69	50
SMicro1.4	<i>Micrococcus</i>	TCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTGACAGGTGGTGCA	970	71	50
SMicro1.5	<i>Micrococcus</i>	TCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTACACAGGTGGTGCA	970	71	50
SMicro2d	<i>Micrococcus</i>	RTWCYSGWTCGYYSYAGAGATRSKKYTTCCCCTTTGGGSYSGGTWBACAG	970	67	50
SMicro2.1	<i>Micrococcus</i>	GTTCTCGATCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTACACAG	970	70	50
SMicro2.2	<i>Micrococcus</i>	GTTCTCGATCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTACACAG	970	70	50
SMicro2.3	<i>Micrococcus</i>	GTTCTCGTTCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTACACAG	970	70	50
SMicro2.4	<i>Micrococcus</i>	ATACCGGATCGTTCAGAGATGGTCTTCCCCTTTGGGGTCGGTATACAG	970	68	50
SMicro2.5	<i>Micrococcus</i>	GTTCTCGATCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTGACAG	970	70	50
SMicro2.6	<i>Micrococcus</i>	GTTCTCGATCGCCGTAGAGATACGGTTTCCCCTTTGGGGCCGGGTTACACAG	970	70	50
SMicroA	<i>Micrococcus</i>	GATRSKKYTTCCCCTTTGGGSYSGG	980	59	25
SMicroA1	<i>Micrococcus</i>	GATACGGTTTCCCCTTTGGGGCCGGG	980	59	25
SMicroA2	<i>Micrococcus</i>	GATACGGTTTCCCCTTTGGGGCCGGG	980	59	25
SMicroA3	<i>Micrococcus</i>	GATGGTCTTCCCCTTTGGGGTCGG	980	58	25
SMicroB	<i>Micrococcus</i>	AAGGCTTGACATRTWCYSGWTCGY	950	54	25
SMicroB1	<i>Micrococcus</i>	AAGGCTTGACATGTTCTCGATCGCC	950	54	25
SMicroB2	<i>Micrococcus</i>	AAGGCTTGACATGTTCTCGTTCGCC	950	54	25
SMicroB3	<i>Micrococcus</i>	AAGGCTTGACATATACCGATCGTT	950	51	25
GoSMicrod	<i>Micrococcus</i>	AAGGCTTGACATRTWCYSGWTCGYGGATCCGATRSKKYTTCCCCTTTGGGSYSGG	950-980	69	56
GoSMicro1	<i>Micrococcus</i>	AAGGCTTGACATGTTCTCGATCGCCGGATCCGATACGGTTTCCCCTTTGGGGCCGGG	950-980	72	56
GoSMicro2	<i>Micrococcus</i>	AAGGCTTGACATGTTCTCGATCGCCGGATCCGATACGGTTTCCCCTTTGGGGCCGGG	950-980	72	56
GoSMicro3	<i>Micrococcus</i>	AAGGCTTGACATGTTCTCGATCGCCGGATCCGATGGTCTTCCCCTTTGGGGTCGG	950-980	71	56
GoSMicro4	<i>Micrococcus</i>	AAGGCTTGACATGTTCTCGTTCGCCGGATCCGATACGGTTTCCCCTTTGGGGCCGGG	950-980	72	56
GoSMicro5	<i>Micrococcus</i>	AAGGCTTGACATGTTCTCGTTCGCCGGATCCGATACGGTTTCCCCTTTGGGGCCGGG	950-980	72	56
GoSMicro6	<i>Micrococcus</i>	AAGGCTTGACATGTTCTCGTTCGCCGGATCCGATGGTCTTCCCCTTTGGGGTCGG	950-980	71	56
GoSMicro7	<i>Micrococcus</i>	AAGGCTTGACATATACCGATCGTTGGATCCGATACGGTTTCCCCTTTGGGGCCGGG	950-980	71	56
GoSMicro8	<i>Micrococcus</i>	AAGGCTTGACATATACCGATCGTTGGATCCGATACGGTTTCCCCTTTGGGGCCGGG	950-980	71	56
GoSMicro9	<i>Micrococcus</i>	AAGGCTTGACATATACCGATCGTTGGATCCGATGGTCTTCCCCTTTGGGGTCGG	950-980	70	56
SPseudo1d	<i>Pseudomonas</i>	ACBYDDBDRBBWTGACGKTACCVAMAGAATAAGCACCGSCTAACTYYGTG	460	65	50
SPseudo1.1	<i>Pseudomonas</i>	ACGTGATTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	65	50
SPseudo1.2	<i>Pseudomonas</i>	ACGTTAGTGTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	65	50
SPseudo1.3	<i>Pseudomonas</i>	ACGTTGGTGTCTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	67	50
SPseudo1.4	<i>Pseudomonas</i>	ACCTTGCTGTTTTGACGTTACCAACAGAATAAGCACCGGCTAACTTCGTG	460	65	50
SPseudo1.5	<i>Pseudomonas</i>	ACCTTGAGTGTGTTTTGACGTTACCAACAGAATAAGCACCGGCTAACTTCGTG	460	64	50
SPseudo1.6	<i>Pseudomonas</i>	ACGTATCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	65	50
SPseudo1.7	<i>Pseudomonas</i>	ACGTTAGTGTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTTCGTG	460	65	50
SPseudo1.8	<i>Pseudomonas</i>	ACTCTGCAATTTTACGTTACCGACAGAATAAGCACCGCTAACTCTGTG	460	65	50
SPseudo1.9	<i>Pseudomonas</i>	ACCTTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo1.10	<i>Pseudomonas</i>	ACGTTGGTGTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo1.11	<i>Pseudomonas</i>	ACGTGATGGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo1.12	<i>Pseudomonas</i>	ACTCTGCAATTTTACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	65	50
SPseudo1.13	<i>Pseudomonas</i>	ACTCTGCAGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo1.14	<i>Pseudomonas</i>	ACTCTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo1.15	<i>Pseudomonas</i>	ACTCTGCCGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	67	50
SPseudo1.16	<i>Pseudomonas</i>	ACTCTGCGGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	67	50
SPseudo1.17	<i>Pseudomonas</i>	ACTTTGCAGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	65	50
SPseudo1.18	<i>Pseudomonas</i>	ACTTTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	65	50
SPseudo1.19	<i>Pseudomonas</i>	ACTTTGCCGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo1.20	<i>Pseudomonas</i>	ACTTTGCGGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo1.21	<i>Pseudomonas</i>	ACCTTGCTGTTTTGACGTTACCGACAGAATAAGCACCGGCTAACTTCGTG	460	66	50

SPseudo1.22	<i>Pseudomonas</i>	ACGTAAGTGTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	65	50
SPseudo1.23	<i>Pseudomonas</i>	ACCCGGTTGGGATGACGGTACCCAAAGAATAAGCACCGGCTAACTTCGTG	460	69	50
SPseudo1.24	<i>Pseudomonas</i>	ACGCTGCAACTTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	65	50
SPseudo1.25	<i>Pseudomonas</i>	ACGCTGCAACTTTGACGTTACCAACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo1.26	<i>Pseudomonas</i>	ACGTGATTGTCTTGACGTTACCGACAGAATAAGCACCGGCTAACTCTGTG	460	66	50
SPseudo2d	<i>Pseudomonas</i>	TRCMTMGAAAYCTRCYYDDTMTGTTTTGGGGGAYAACGYNBSGAAASKDRGCT	110	66	50
SPseudo2.1	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTGGTAGTGGGGGATAACGTTTCGAAACGGACGCT	110	69	50
SPseudo2.2	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGGAACGCT	110	60	50
SPseudo2.3	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTGGTAGTGGGGGATAACGTTTCGAAACGGACGCT	110	70	50
SPseudo2.4	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTGGTAGTGGGGGACAACGTTCCGAAAGGGACGCT	110	70	50
SPseudo2.5	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTGGTAGTGGGGGATAACGTTCCGAAACGGGGCGCT	110	71	50
SPseudo2.6	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTGATAGTGGGGGACAACGTTTCGAAAGGAACGCT	110	68	50
SPseudo2.7	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGGGACGCT	110	70	50
SPseudo2.8	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTATTAGTGGGGGACAACGTTTCGAAAGGAACGCT	110	67	50
SPseudo2.9	<i>Pseudomonas</i>	TGCCTAGGAACCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGGAACGCT	110	69	50
SPseudo2.10	<i>Pseudomonas</i>	TACATCGGAATCTACTTTTTCTGTTTTGGGGGATAACGTAGGGAAACTTACGCT	110	64	50
SPseudo2.11	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCCCGTAGTGGGGGATAACGTTTCGAAACGGACGCT	110	70	50
SPseudo2.12	<i>Pseudomonas</i>	TGCCTAGGAATCTACCTGGTAGTGGGGGATAACGTTTCGAAACGGACGCT	110	69	50
SPseudo2.13	<i>Pseudomonas</i>	TGCCTAGGAATCTGCCTGATAGTGGGGGATAACGTTTCGAAACGGACGCT	110	69	50
SRhodo1d	<i>Rhodococcus</i>	GCAACCCYTRTCYYRTGTTGCCAGCRMRTDVGKBBGGKACTCRYRRGAG	1065	68	50
SRhodo1.1	<i>Rhodococcus</i>	GCAACCCCTGTCTGTGTTGCCAGCGCGTAATGGCGGGGACTCGCAGGAG	1065	73	50
SRhodo1.2	<i>Rhodococcus</i>	GCAACCCCTATCTTATGTTGCCAGCACGTTATGGTGGGGACTCGTAAGAG	1065	68	50
SRhodo1.3	<i>Rhodococcus</i>	GCAACCCCTGTCTTATGTTGCCAGCGCGTTATGGCGGGGACTCGTAAGAG	1065	69	50
SRhodo1.4	<i>Rhodococcus</i>	GCAACCCCTGTCTTATGTTGCCAGCACGTAATGGTGGGGACTCGTAAGAG	1065	68	50
SRhodo1.5	<i>Rhodococcus</i>	GCAACCCCTGTCTGTGTTGCCAGCACGTTATGGTGGGGACTCGCAGGAG	1065	72	50
SRhodo1.6	<i>Rhodococcus</i>	GCAACCCCTGTCTGTGTTGCCAGCAAGTAATGTTGGGGACTCGCAGGAG	1065	69	50
SRhodo1.7	<i>Rhodococcus</i>	GCAACCCCTGTCTCATGTTGCCAGCACGTTATGGTGGGGACTCGTGAGAG	1065	70	50
SRhodo1.8	<i>Rhodococcus</i>	GCAACCCCTGTCTTATGTTGCCAGCACATTGTTGGTGGGTACTCATGAGAG	1065	67	50
SRhodo1.9	<i>Rhodococcus</i>	GCAACCCCTGTCTGTGTTGCCAGCAAAATTCGGTTGGGGACTCGCAGGAG	1065	70	50
SRhodo1.10	<i>Rhodococcus</i>	GCAACCCCTGTCTGTGTTGCCAGCACATTTCGGTTGGGGACTCGCAGGAG	1065	71	50
SRhodo1.11	<i>Rhodococcus</i>	GCAACCCCTGTCTGTGTTGCCAGCACGTAATGGTGGGGACTCGCAGGAG	1065	71	50
SRhodo1.12	<i>Rhodococcus</i>	GCAACCCCTGTCTTATGTTGCCAGCACGTAATGGGGGGGACTCGTAAGAG	1065	69	50
SRhodo1.13	<i>Rhodococcus</i>	GCAACCCCTGTCTGTGTTGCCAGCAAGTAAAGTTGGGGACTCGCAGGAG	1065	69	50
Srhodo2d	<i>Rhodococcus</i>	CTRGYYTTGACATRBACMRGANVRSBBYAGAGATRBKSKYYCCCTTGTGG	945	66	50
Srhodo2.1	<i>Rhodococcus</i>	CTGGGTTTGACATATAACCGAAAAGCCGTAGAGATACGGCCCCCTTGTGG	945	69	50
Srhodo2.2	<i>Rhodococcus</i>	CTGGGTTTGACATATAACCGAAAAGCTGCAGAGATGTGGCCCCCTTGTGG	945	69	50
Srhodo2.3	<i>Rhodococcus</i>	CTGGGTTTGACATACACCGAAAACCGTAGAGATACGGTCCCCCTTGTGG	945	69	50
Srhodo2.4	<i>Rhodococcus</i>	CTGGGTTTGACATATAACCGAAAAGCCGTAGAGATACGGCCCCCTTGTGG	945	69	50
Srhodo2.5	<i>Rhodococcus</i>	CTGGGTTTGACATATAACCGAAAAGCTGCAGAGATGTGGCCCCCTTGTGG	945	69	50
Srhodo2.6	<i>Rhodococcus</i>	CTGGGTTTGACATGTACCGGACGACTGCAGAGATGTGGTTCCCTTGTGG	945	69	50
Srhodo2.7	<i>Rhodococcus</i>	CTGGGTTTGACATATAACCGGATCGCTCAGAGATGGGGTTCCCTTGTGG	945	69	50
Srhodo2.8	<i>Rhodococcus</i>	CTAGGCTTGACATATAACAGGACGACGGCAGAGATGTGGTTCCCTTGTGG	945	68	50
Srhodo2.9	<i>Rhodococcus</i>	CTGGGTTTGACATATAACCGGACGACTGCAGAGATGTGGTTCCCTTGTGG	945	68	50
Srhodo2.10	<i>Rhodococcus</i>	CTGGGCTTGACATGGACCGGATCGCGGTAGAGATACGTTTCCCTTGTGG	945	69	50
Srhodo2.11	<i>Rhodococcus</i>	CTGGGTTTGACATACACCGGATCGCTCAGAGATGGGGTTCCCTTGTGG	945	69	50
Srhodo2.12	<i>Rhodococcus</i>	CTGGGTTTGACATACACCGGACCGCCCCAGAGATGGGGTTCCCTTGTGG	945	71	50
Srhodo2.13	<i>Rhodococcus</i>	CTGGGTTTGACATACACAGATCGCCGTAGAGATACGGTTTCCCTTGTGG	945	68	50
SStaphd1	<i>Staphylococcus</i>	NRBDYGTWRGTMACRWRNRYVYCTTGACGGYACCTRAYCAGAAAGCCAC	450	65	50
SStaph1.1	<i>Staphylococcus</i>	AATTTGTTAGTAACTGAACAAGTCTTGACGGTACCTAACAGAAAGCCAC	450	63	50
SStaph1.2	<i>Staphylococcus</i>	AGGATGTAAGTAACTATGCATCCCTTGACGGTACCTAATCAGAAAGCCAC	450	64	50
SStaph1.3	<i>Staphylococcus</i>	AATGTGTAAGTAACTGTGCACATCTTGACGGTACCTAATCAGAAAGCCAC	450	64	50
SStaph1.4	<i>Staphylococcus</i>	TACGTGTAAGTAACTATGCACGCTTGACGGTACCTAATCAGAAAGCCAC	450	64	50

SStaph1.5	<i>Staphylococcus</i>	AACGTGTAAGTAACTGTGCACGCTTTGACGGTACCTAATCAGAAAGCCAC	450	65	50
SStaph1.6	<i>Staphylococcus</i>	TATGTGTAAGTAACTGTGCACATCTTGACGGTACCTAATCAGAAAGCCAC	450	64	50
SStaph1.7	<i>Staphylococcus</i>	AACGTGTAAGTAACTGTGCACGCTTTGACGGTACCTAACCAGAAAGCCAC	450	66	50
SStaph1.8	<i>Staphylococcus</i>	AATGTGTAAGTAACTATGCACGCTTTGACGGTACCTAATCAGAAAGCCAC	450	64	50
SStaph1.9	<i>Staphylococcus</i>	AATGTGTAAGTAACTATGCACGCTTTGACGGTACCTAATCAGAAAGCCAC	450	64	50
SStaph1.10	<i>Staphylococcus</i>	AATGTGTAAGTAACTGTGCACATCTTGACGGTACCTGATCAGAAAGCCAC	450	64	50
SStaph1.11	<i>Staphylococcus</i>	CATGTGTAAGTAACTATGCACATCTTGACGGTACCTAACCAGAAAGCCAC	450	65	50
SStaph1.12	<i>Staphylococcus</i>	AATGCGTAAGTAACTGTGCGCGTCTTGACGGTACCTAATCAGAAAGCCAC	450	66	50
SStaph1.13	<i>Staphylococcus</i>	AGTGCGTAGGTAAGTATGCACCTTGACGGTACCTAATCAGAAAGCCAC	450	67	50
SStaph1.14	<i>Staphylococcus</i>	AATGTGTAAGTAACTGTGCACATCTTGACGGTACCTAACCAGAAAGCCAC	450	64	50
SStaph1.15	<i>Staphylococcus</i>	AATGTGTAAGTAACTGTGCACGCTTTGACGGTACCTAATCAGAAAGCCAC	450	64	50
SStaph1.16	<i>Staphylococcus</i>	TATGTGTAAGTAACTGTGCACATCTTGACGGTACCTAACCAGAAAGCCAC	450	64	50
SStaph1.17	<i>Staphylococcus</i>	AGTGCGTAAGTAACTGTGCGCACCTTGACGGTACCTAATCAGAAAGCCAC	450	67	50
SStaph1.18	<i>Staphylococcus</i>	AACGTGTAAGTAACTGTGCACGCTTTGACGGTACCTGATCAGAAAGCCAC	450	66	50
SStaph1.19	<i>Staphylococcus</i>	AATGTGTAAGTAACTGTGCACGCTTTGACGGTACCTAACCAGAAAGCCAC	450	65	50
SStaphd2	<i>Staphylococcus</i>	CGRRGCTAATRCDDGATAABAYDYNRAACCGCATGGTTYNRNRDTRAAAG	150	63	50
SStaph2.1	<i>Staphylococcus</i>	CGGGGCTAATACCTGATAATATTTTGAACCGCATGGTTCAATAGTAAAAG	150	63	50
SStaph2.2	<i>Staphylococcus</i>	CGGGGCTAATACCGGATAACACATGAAACCGCATGGTTTCATGATGAAAAG	150	65	50
SStaph2.3	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAACATTAGAACCGCATGGTTCTAAAAGTAAAAG	150	64	50
SStaph2.4	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAATATTTTGAACCGCATGGTTTCGATAGTAAAAG	150	64	50
SStaph2.5	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAATATTTTGAACCGCATGGTTCAAAAAGTAAAAG	150	63	50
SStaph2.6	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAACATTGGAACCGCATGGTTCTAAAAGTAAAAG	150	64	50
SStaph2.7	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAACATGTTGAACCGCATGGTTCAACAGTAAAAG	150	65	50
SStaph2.8	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAACATGTTGAACCGCATGGTTCAACAGTAAAAG	150	64	50
SStaph2.9	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAACATATTTGAACCGCATGGTTCAATAGTAAAAG	150	64	50
SStaph2.10	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAATATTTTGAACCGCATGGTTCAATAGTAAAAG	150	63	50
SStaph2.11	<i>Staphylococcus</i>	CGGGGCTAATACCGGATAATATTTTGAACCGCATGGTTCAATAGTAAAAG	150	64	50
SStaph2.12	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAATATTTAGAACCGCATGGTTTCGATAGTAAAAG	150	64	50
SStaph2.13	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAATATTTTGAACCGCATGGTTCAAAAAGTAAAAG	150	62	50
SStaph2.14	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAATATGTTGAACCGCATGGTTCAACAGTAAAAG	150	66	50
SStaph2.15	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAACATGTTGAACCGCATGGTTCTACAGTAAAAG	150	67	50
SStaph2.16	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAATATGCGGAACCGCATGGTTCTGCAATGAAAAG	150	68	50
SStaph2.17	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAATATGCGGAACCGCATGGTTCCGCAATGAAAAG	150	69	50
SStaph2.18	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAATATGCTGAACCGCATGGTTCCGCAATGAAAAG	150	68	50
SStaph2.19	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAACATATCGAACCGCATGGTTTCGATAGTAAAAG	150	67	50
SStaph2.20	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAACATTTAGAACCGCATGGTTCTAAAAGTAAAAG	150	65	50
SStaph2.21	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAATATGCGAACCGCATGGTTCTGCAATGAAAAG	150	67	50
SStaph2.22	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAACATATAGAACCGCATGGTTCTATAGTAAAAG	150	65	50
SStaph2.23	<i>Staphylococcus</i>	CGGAGCTAATGCCGGATAACATTTGGAACCGCATGGTTCTAAAAGTAAAAG	150	64	50
SStaph2.24	<i>Staphylococcus</i>	CGGGGCTAATGCCGGATAACATGTTGAACCGCATGGTTCAACAGTAAAAG	150	67	50
SStaph2.25	<i>Staphylococcus</i>	CGAAGCTAATGCCGGATAACATGTTAGAACCGCATGGTTCTACAGTAAAAG	150	65	50
SStaph2.26	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAATATGTTGAACCGCATGGTTCAATAGTAAAAG	150	64	50
SStaph2.27	<i>Staphylococcus</i>	CGGGGCTAATACCGGATAATATATTTGAACCGCATGGTTCAATGTTGAAAAG	150	64	50
SStaph2.28	<i>Staphylococcus</i>	CGGAGCTAATACCGGATAAGATTTTGAACCGCATGGTTCAATAGTAAAAG	150	64	50
SStaphd3	<i>Staphylococcus</i>	GCGAACDGANRAGRAGCTTKCTYCTYTGANGTAGCGGCGGACGGGTGAG	50	68	50
SStaph3.1	<i>Staphylococcus</i>	GCGAACAGATGAGAAGCTTGCTTCTCTGATGTTAGCGGCGGACGGGTGAG	50	69	50
SStaph3.2	<i>Staphylococcus</i>	GCGAACAGACGAGGAGCTTGCTCCTCTGACGTTAGCGGCGGACGGGTGAG	50	73	50
SStaph3.3	<i>Staphylococcus</i>	GCGAACAGATAAGGAGCTTGCTCCTTTGAAGTTAGCGGCGGACGGGTGAG	50	69	50
SStaph3.4	<i>Staphylococcus</i>	GCGAACAGACAAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	50	71	50
SStaph3.5	<i>Staphylococcus</i>	GCGAACAGACGAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	50	72	50
SStaph3.6	<i>Staphylococcus</i>	GCGAACGAGCAGAGAAGCTTGCTTCTCTGATGTTAGCGGCGGACGGGTGAG	50	71	50
SStaph3.7	<i>Staphylococcus</i>	GCGAACGGATAAGGAGCTTGCTCCTTTGAAGTTAGCGGCGGACGGGTGAG	50	70	50

SStaph3.8	<i>Staphylococcus</i>	GCGAACGGACGAGAAGCTTCTCTCTGATGTTAGCGGCGGACGGGTGAG	50	70	50
SStaph3.9	<i>Staphylococcus</i>	GCGAACAGATAAAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	50	70	50
SStaph3.10	<i>Staphylococcus</i>	GCGAACAGATGAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	50	71	50
SStaph3.11	<i>Staphylococcus</i>	GCGAACAGATGAGGAGCTTGCTCCTCTGACGTTAGCGGCGGACGGGTGAG	50	72	50
SStaph3.12	<i>Staphylococcus</i>	GCGAACTGACGAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	50	72	50
SStaph3.13	<i>Staphylococcus</i>	GCGAACAGATGAGGAGCTTGCTCCTCTGATGTTAGCGGCGGACGGGTGAG	50	71	50
SStaph3.14	<i>Staphylococcus</i>	GCGAACTGAAGAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	50	71	50
SStaph3.15	<i>Staphylococcus</i>	GCGAACGGACAAGGAGCTTGCTCCTTTGAAGTTAGCGGCGGACGGGTGAG	50	71	50
SStaph1A	<i>Staphylococcus</i>	TGTAGCGGTGAAATGCGCAGAGATA	660	53	25
SStaph1B	<i>Staphylococcus</i>	TGAKGTGCGAAAGCGTGGGGATCAA	750	57	25
SStaph1B1	<i>Staphylococcus</i>	TGATGTGCGAAAGCGTGGGGATCAA	750	54	25
SStaph1B2	<i>Staphylococcus</i>	TGAGGTGCGAAAGCGTGGGGATCAA	750	56	25
GoSStaphd1	<i>Staphylococcus</i>	TGTAGCGGTGAAATGCGCAGAGATAGGATCCTGAKGTGCGAAAGCGTGGGGATCAA	660-750	67	56
GoSStaph1.1	<i>Staphylococcus</i>	TGTAGCGGTGAAATGCGCAGAGATAGGATCCTGATGTGCGAAAGCGTGGGGATCAA	660-750	69	56
GoSStaph1.2	<i>Staphylococcus</i>	TGTAGCGGTGAAATGCGCAGAGATAGGATCCTGAGGTGCGAAAGCGTGGGGATCAA	660-750	70	56
SSmycesd17	<i>Streptomyces</i>	CMACGCCGTAACGKTGGGMACTARGTGTGRCRRRCATTCCACGTYGTYS	770	67	50
SSmyces17.1	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGAACTAGGTGTTGGCGACATTCCACGTCGTCG	770	71	50
SSmyces17.2	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGAACTAGGTGTTGGCGACATTCCACGTCGTCG	770	70	50
SSmyces17.3	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGCACTAGGTGTGGGCAACATTCCACGTTGTCC	770	71	50
SSmyces17.4	<i>Streptomyces</i>	CAACGCCGTAACGTTGGGCACTAGGTGTGGACAGCATTCCACGTTGTCC	770	70	50
SSmyces17.5	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGCACTAGGTGTGGGCAACATTCCACGTTGTCC	770	70	50
SSmyces17.6	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGCACTAGGTGTGGGCAACATTCCACGTTGTCC	770	72	50
SSmyces17.7	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGAACTAAGTGTGGCGACATTCCACGTCGTCG	770	69	50
SSmyces17.8	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGCACTAAGTGTGGGCAACATTCCACGTTGTCC	770	70	50
SSmyces17.9	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGAACTAGGTGTGGGCGACATTCCACGTCGTCG	770	71	50
SSmyces17.10	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGCACTAGGTGTTGGCGACATTCCACGTCGTCG	770	72	50
SSmyces17.11	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGCACTAGGTGTGGGCGACATTCCACGTCGTCG	770	73	50
SSmyces17.12	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGAACTAGGTGTTGGCGACATTCCACGTCGTTG	770	69	50
SSmyces17.13	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGCACTAGGTGTGGGCGGCATTCCACGTCGTCG	770	73	50
SSmyces17.14	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGCACTAGGTGTGGGCGGCATTCCACGTCGTCG	770	73	50
SSmyces17.15	<i>Streptomyces</i>	CCACGCCGTAACGTTGGGAACTAGGTGTGGGCGACATTCCACGTTGTCC	770	70	50

Table 4. Influence of mismatches number and location on sensitivity detection of targets. Mismatches between probe and target are indicated in blue colour.

Probe Name	Probe Sequence	Fluorescence Units
SStaph3.3	GCGAACAGATAAAGGAGCTTGCTCCTTTGAAGTTAGCGGCGGACGGGTGAG	49.570
SStaph3.7	GCGAACGGATAAAGGAGCTTGCTCCTTTGAAGTTAGCGGCGGACGGGTGAG	51.379
SStaph3.13	GCGAACAGATGAGGAGCTTGCTCCTCTGATGTTAGCGGCGGACGGGTGAG	38.651
SStaph3.4	GCGAACAGACAAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	38.427
SStaph3.10	GCGAACAGATGAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	35.771
SStaph3.15	GCGAACGGACAAGGAGCTTGCTCCTTTGAAGTTAGCGGCGGACGGGTGAG	34.717
SStaph3.9	GCGAACAGATAAAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	31.355
SStaph3.5	GCGAACAGACGAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	30.562
SStaph3.12	GCGAACTGACGAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	27.804
SStaph3.1	GCGAACAGATGAGAAGCTTGCTTCTCTGATGTTAGCGGCGGACGGGTGAG	26.141
SStaph3.6	GCGAACGGACGAGAGCTTGCTTCTCTGATGTTAGCGGCGGACGGGTGAG	22.431
SStaph3.11	GCGAACAGATGAGGAGCTTGCTCCTCTGACGTTAGCGGCGGACGGGTGAG	19.650
SStaph3.14	GCGAACTGAAGAGGAGCTTGCTCCTTTGACGTTAGCGGCGGACGGGTGAG	19.638
SStaph3.8	GCGAACGGACGAGAGCTTCTCTCTGATGTTAGCGGCGGACGGGTGAG	13.678
SStaph3.2	GCGAACAGACGAGGAGCTTGCTCCTCTGACGTTAGCGGCGGACGGGTGAG	9.520

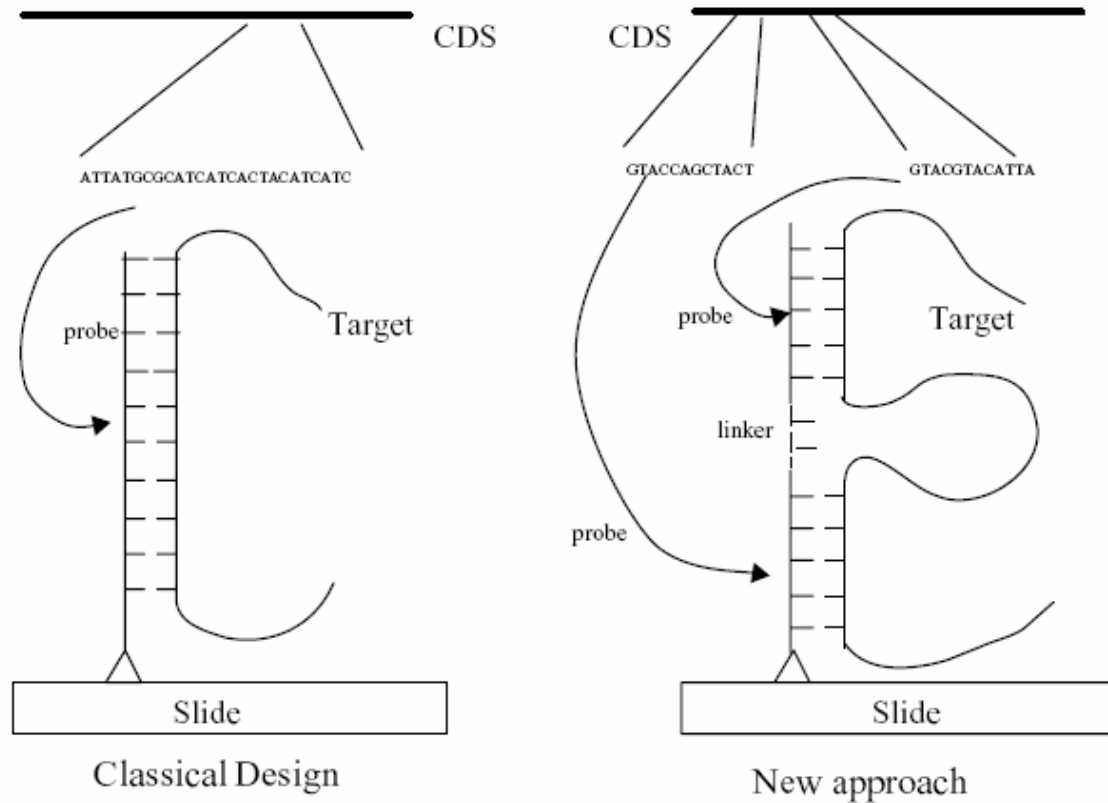


Figure 6: Schéma de la stratégie de design GoArrays comparée celle utilisée classiquement pour le design de sondes oligonucléotidiques pour biopuces ADN. Dans l'approche commune, l'oligonucléotide est spécifique d'une sous-séquence de sa cible (dans cet exemple d'un CDS). Dans la stratégie GoArrays, l'oligonucléotide est composé de deux sous-séquences, ciblant deux zones non continues de la cible, séparées par un espaceur choisi aléatoirement. Une hybridation stable entre la sonde composite et la cible entraîne la formation d'une boucle

D. DISCUSSION.

Nos résultats ont montré que les approches conventionnelles de sélection de sondes ne permettaient pas toujours d'obtenir des biocapteurs suffisamment efficaces pour identifier correctement les communautés bactériennes des environnements complexes.

En terme de sensibilité, la comparaison des différents types de sondes nous a montré que les sondes PhylArray semblaient plus sensibles que celles sélectionnées avec PRIMROSE et dans la banque de sondes ARB. Ainsi, elles pourraient potentiellement permettre d'accéder plus facilement aux phylotypes faiblement abondants présents dans les environnements étudiés. Ces résultats semblent être expliqués en partie par plusieurs paramètres tels la taille des sondes, la présence/position de mésappariements et l'accessibilité variable des sondes selon leur position sur la molécule cible. L'influence de ces paramètres, décrite par Urakawa *et al.* (2003) et Fuchs *et al.* (1998), ne suffit cependant pas à expliquer toute la variabilité d'intensité entre les différentes sondes. Ainsi, d'autres facteurs thermodynamiques non encore élucidés, semblent influencer les hybridations sonde-cible sur un support solide comme l'avait suggéré Pozhitkov *et al.* en 2006.

En terme de spécificité, les sondes courtes PhylArray et ARB apparaissent les plus efficaces en engendrant moins de croisements aspécifiques que les sondes longues PRIMROSE et PhylArray. Ces résultats sont en accord avec ceux obtenus par Kane *et al.* (2000) suggérant que plus la sonde oligonucléotidique est courte moins elle a de chance de présenter des critères qui la rendraient aspécifique (critères de Kane). Ces reconnaissances faussement positives s'expliquent par la similarité des sondes avec des séquences non ciblées. Ceci peut être induit d'une part par des critères de sélection insuffisamment contraignants du fait des caractéristiques propres aux logiciels et d'autre part par l'impossibilité de vérifier la spécificité des sondes envers tous les microorganismes car la plupart sont inconnus. Ainsi, les hybridations croisées semblent inévitables dans ce contexte. Cependant, nos résultats ont démontré que l'utilisation de la stratégie GoArrays pouvait inhiber ces signaux faussement positifs probablement grâce à une déstabilisation du duplex sonde/séquence non ciblée due à une contrainte imposée par la présence de la boucle (figure 6). Dans des expériences utilisant les biopuces, il est important d'éviter les reconnaissances aspécifiques mais il est tout autant nécessaire de reconnaître les microorganismes qui sont ciblés. Ainsi, nous avons pu mettre en évidence des défauts de reconnaissance des sondes PRIMROSE et ARB pour certaines souches du genre *Micrococcus*. Les sondes PhylArray ont quant à elles reconnu les souches

de ce taxon très polymorphe certainement grâce à l'utilisation de sondes reconnaissant des régions plus variables.

Enfin, les sondes sélectionnées avec le programme PhylArray semblent permettre une vision exploratoire des communautés bactériennes étudiées contrairement aux autres outils de design qui ne génèrent que des sondes ciblant les fractions microbiennes connues. La caractérisation des souches dont les cibles ont été hybridées aux sondes SAerod2, SAero2.1, SAero2.2, SAero2.3 et SAero2.4 ciblant le genre *Aeromonas* permettrait de valider ce caractère exploratoire.

PARTIE V

CONCLUSIONS ET PERSPECTIVES

1. CONCLUSIONS.

De part leur rôle dans les grands cycles biogéochimiques, les bactéries sont des acteurs extrêmement importants dans l'équilibre et le fonctionnement des écosystèmes et plus particulièrement dans celui des sols. Leur diversité et leur capacité évolutive, conséquences entre autres des échanges d'informations génétiques entre microorganismes, leurs permettent de s'adapter à des conditions environnementales variées. Ainsi elles peuvent occuper de nombreuses niches écologiques dont les plus extrêmes (Lin *et al.*, 2006). Ces aptitudes à se développer et à coloniser de « nouveaux » milieux ont notamment été mises en évidence lors de l'apparition de nouvelles molécules xénobiotiques qui ne possèdent aucun équivalent dans la nature, tel le lindane (Phillips *et al.*, 2005). Après une phase d'adaptation plus ou moins longue à de telles molécules, les bactéries arrivent à acquérir de nouvelles voies métaboliques leur permettant de résister voire même de les utiliser à leur avantage. Ainsi, l'idée de recourir à ces microorganismes pour dégrader naturellement des polluants a germé d'autant plus que l'on pourrait alors s'affranchir des techniques classiques de dépollution onéreuses et bien souvent invasives pour les écosystèmes traités. A l'heure actuelle, ces techniques de bioremédiation ont montré leur efficacité à atténuer naturellement la charge polluante de différents environnements mais dans certains cas il est nécessaire d'améliorer encore les rendements de décontamination. Pour cela, il est indispensable de pouvoir identifier les bactéries composant cette « boîte noire » épuratrice et de comprendre leurs mécanismes d'adaptation à la présence d'une pollution. Les informations ainsi obtenues pourraient permettre de guider la mise en place de procédés stimulant l'activité des bactéries impliquées dans la dégradation du polluant ou permettant leur implantation massive dans un environnement à traiter.

Les différents travaux présentés au cours de cette thèse nous ont permis d'approfondir les connaissances sur les communautés bactériennes impliquées dans les phénomènes de biodégradation de polluants. Jusqu'à récemment, l'étude des microorganismes nécessitait obligatoirement le passage par des étapes de mise en culture afin d'accéder aux critères phénotypiques et métaboliques permettant leur identification. Cependant, l'usage de milieux dits d'enrichissement a rapidement montré ses limites en raison, d'une part, de la présence massive de nutriments et, d'autre part, de l'absence de certains éléments qui favorisaient certaines espèces au détriment d'autres moins bien adaptées. C'est pourquoi des milieux

dilués en éléments nutritifs ou mimant les conditions environnementales ont été récemment développés. Couplés à des approches de cultures à haut débit, ils permettent d'augmenter fortement le nombre de microorganismes cultivés.

Malgré ces progrès impressionnants, la majorité des microorganismes n'est toujours pas cultivée à ce jour. C'est pourquoi des approches alternatives ont été mises en place afin de pouvoir identifier les microorganismes sans nécessiter leur mise en culture. Ces approches dites moléculaires sont basées sur l'étude des informations phylogénétiques apportées par les acides nucléiques des microorganismes et plus particulièrement par leurs biomarqueurs ribosomiques. Le recueil de ces informations passe le plus souvent par des techniques d'amplification de ces gènes suivies de techniques dites d'empreintes génétiques permettant une discrimination des différentes populations microbiennes. L'étude des microorganismes peut également être réalisée par hybridation *in situ* à l'aide de sondes spécifiques de phylotypes ou de groupes microbiens. Récemment, ces techniques moléculaires ont connu d'énormes avancées du point de vue méthodologique. Avec l'apparition de techniques permettant de cribler à haut débit les populations microbiennes et d'accéder à l'ensemble de l'information génomique des microorganismes, une nouvelle ère voit le jour. Elle permettra sans doute de mettre en lumière l'incroyable diversité des communautés microbiennes environnementales au sein de leurs écosystèmes, qu'ils soient sains ou perturbés.

De manière à mieux appréhender les mécanismes d'adaptation des communautés bactériennes à une pollution de type hydrocarbures aliphatiques, une partie du travail réalisé lors de cette thèse a consisté à caractériser les bactéries présentes au niveau d'un sol pollué au cours de sa réhabilitation par un procédé de bioremédiation. La création de banques d'ADNr 16S et d'ARNr 16S a permis de mettre en évidence l'évolution de la structure des communautés bactériennes au fur et à mesure de l'abattement de la charge polluante. Il s'est avéré que les bactéries appartenant au phylum *γ-Proteobacteria* étaient dominantes lors des premiers stades de bioremédiation, au moment où la pollution est maximale. L'analyse des communautés métaboliquement actives montre également l'importance des *Actinobacteria* au stade initial du processus de dépollution. Au cours du traitement, une tendance nette à la diminution des *Proteobacteria* en faveur des *Actinobacteria* a également été observée et ce particulièrement après 24 mois de traitement. Ce phylum des *Actinobacteria* semble donc jouer un rôle très important dans la dégradation des polluants lors des stades tardifs du traitement. Un autre résultat important a été la mise en évidence de la présence d'un « core-set » bactérien métaboliquement actif durant toute la cinétique. Celui-ci est composé de 17

phylotypes appartenant aux *Actinobacteria*, *Gemmatimonadetes*, *α -Proteobacteria*, *β -Proteobacteria*, *δ -Proteobacteria*, *γ -Proteobacteria* et division TM7. Leurs proportions au sein des banques de clones ARNr 16S suggèrent leur importance dans le procédé de bioremédiation mais aussi dans le fonctionnement global de cet écosystème perturbé.

Même si la création de larges banques de clones permet de couvrir la majorité de la diversité bactérienne, il a été montré que les données obtenues par les techniques de clonage/séquençage pouvaient être avantageusement complétées par l'utilisation de biopuces ADN. En effet, ces dernières permettent de révéler une plus grande part de la diversité bactérienne que les banques de clones pour un même environnement étudié et cela uniquement en quelques manipulations.

Cependant, comme les sondes de biopuces sont générées à partir d'informations provenant de banques de données où ne sont répertoriées que les séquences ribosomiques des bactéries connues, elles ne permettent pas l'exploration de la fraction bactérienne encore inconnue. Ainsi, afin de pouvoir combiner le haut débit des biopuces avec la possibilité de cibler des bactéries connues et inconnues, nous avons développé un logiciel de design de sondes oligonucléotidiques pour biopuces phylogénétiques. Ce logiciel, nommé PhylArray, repose sur une stratégie de sélection de sondes ciblant des groupes taxonomiques d'un niveau supérieur à l'espèce et sur l'utilisation combinée de sondes dégénérées et non dégénérées. La stratégie GoArrays, que nous avons développé au sein de l'équipe de recherche, a elle aussi été utilisée pour la création des sondes (Rimour *et al.*, 2005 ; voir Annexe). Comme la pertinence de la caractérisation des communautés bactériennes par une approche biopuce repose aussi sur l'efficacité des sondes sélectionnées (spécificité et sensibilité), nous nous sommes attachés à vérifier les performances de nos sondes comparées à celles de sondes sélectionnées avec d'autres logiciels (PRIMROSE, ARB). L'étude du comportement des différentes sondes, lors de l'hybridation de nos biopuces prototypes avec différentes cibles extraites de souches bactériennes de référence, a montré l'efficacité de nos approches (PhylArray et GoArrays) comparée aux approches classiquement utilisées (PRIMROSE, ARB). En effet, elles semblent permettre la sélection de sondes qui apparaissent suffisamment efficaces, en terme de sensibilité et de spécificité, pour étudier correctement les communautés bactériennes des environnements complexes. De plus, les résultats d'hybridation obtenus sur notre sol pollué aux hydrocarbures semblent suggérer que l'utilisation combinée des sondes dégénérées et non dégénérées est une stratégie intéressante permettant de donner un caractère exploratoire aux biopuces phylogénétiques.

2. PERSPECTIVES.

1. Etude exhaustive des communautés procaryotiques.

1.1. Création d'une biopuce phylogénétique.

Nos stratégies de détermination de sondes ayant été validées par la conception et l'hybridation d'une biopuce prototype ciblant 7 genres bactériens, une biopuce phylogénétique complète est actuellement en cours de réalisation. Celle-ci permettra l'étude simultanée de 80 genres appartenant au domaine des *Archaea* et de 1 335 genres appartenant au domaine des *Bacteria*. Ce nouvel outil nous permettra sans aucun doute d'analyser plus finement la composition et la dynamique des communautés bactériennes dans des écosystèmes complexes. Grâce à un pas d'échantillonnage plus court, cette biopuce nous permettra aussi d'étudier les mécanismes d'adaptation de façon plus intégrée.

De plus, l'utilisation de cette biopuce couplée à d'autres approches moléculaires permettra de valider ou d'infirmer définitivement le caractère exploratoire des sondes designées en utilisant la stratégie PhylArray.

1.2. Utilisation de biomarqueurs complémentaires.

Bien que présentant de nombreux avantages, certains inconvénients sont inhérents à l'utilisation des séquences d'ADNr 16S. En effet, les différentes copies des ADNr 16S, présentes au sein d'une même espèce, peuvent présenter une hétérogénéité pouvant entraîner des erreurs d'interprétation de la diversité (Shimizu *et al.*, 2001). De plus, ces gènes peuvent montrer des faiblesses pour différencier certaines espèces appartenant au même genre (Hill *et al.*, 2005). C'est pour ces différentes raisons que d'autres marqueurs phylogénétiques tendent à être utilisés. Ainsi, de plus en plus d'études utilisent les régions hypervariables Internal Transcribed Spacer (ITS) (Spiegelman *et al.*, 2005) mais aussi les gènes *cpn60* (chaperone protein chaperonin 60), *rpoB* (RNA polymerase β -subunit) et *gyrB* (DNA gyrase B protein) qui possèdent autant, voire plus d'informations que les ADNr 16S (Brousseau *et al.*, 2001). Cependant, le faible nombre de séquences disponibles pour ces biomarqueurs est actuellement encore assez limitant. L'utilisation de gènes fonctionnels pour des groupes de bactéries, par exemple le gène *pmoA* (particulate methane monooxygenase) pour les méthanotrophes, peut

aussi être envisagée (Horz *et al.*, 2001). Il est néanmoins nécessaire de s'assurer de leur capacité résolutive et de leur homogénéité au sein d'une même espèce (Dahllof, 2002). Récemment, afin d'améliorer la résolution des études phylogénétiques, plusieurs gènes bactériens, généralement des gènes de ménage, ont pu être étudiés simultanément. En effet, cette technique prometteuse, nommée MultiLocus Sequence Typing (MLST), est de plus en plus utilisée pour l'identification microbienne (Ciccarelli *et al.*, 2006 ; Whatmore *et al.*, 2007). Ainsi, une approche biopuces ADN utilisant d'autres marqueurs phylogénétiques pourrait être envisagée afin de compléter les informations apportées par les séquences ribosomiques. L'accumulation des données de séquençage de génomes bactériens complets apportera sans aucun doute de nouvelles informations sur la pertinence d'utilisation de nouveaux biomarqueurs permettant de discriminer les « espèces » bactériennes.

2. De la structure à la fonction.

Comme nous l'avons vu précédemment, l'évolution des communautés bactériennes au cours d'un processus de dépollution permet de suggérer l'implication de certains phylotypes dans les phénomènes de biodégradation des polluants. Cependant, il est primordial d'effectuer des analyses complémentaires permettant de faire un lien plus direct entre la structure de ces populations bactériennes et leurs fonctions (activité et voies métaboliques).

2.1. Activité biologique et structure.

Dans un premier temps, l'identification des phylotypes bactériens impliqués dans la bioremédiation de l'écosystème pollué doit être réalisée. Dès lors, il est nécessaire de révéler non seulement les bactéries activant la dégradation des contaminants mais aussi celles étant impliquées dans les étapes entraînant leur minéralisation. Afin de mettre en évidence ces biotransformations, l'utilisation de molécules traceuses (radioactives ou isotopiques) peut être envisagée.

Une technique permettant de relier, au niveau cellulaire, l'identité phylogénétique d'un microorganisme à son activité est l'utilisation combinée de la microautoradiographie (MAR) aux techniques FISH ou CARD-FISH (Lee *et al.*, 1999 ; Ouverney *et al.*, 1999 ; Sekar *et al.*, 2003). Dans cette technique, les échantillons des environnements étudiés sont initialement incubés avec un substrat d'intérêt marqué radioactivement (^{14}C , ^3H ou ^{32}P). Celui-ci est assimilé par certains microorganismes et la radioactivité est incorporée au niveau de leurs

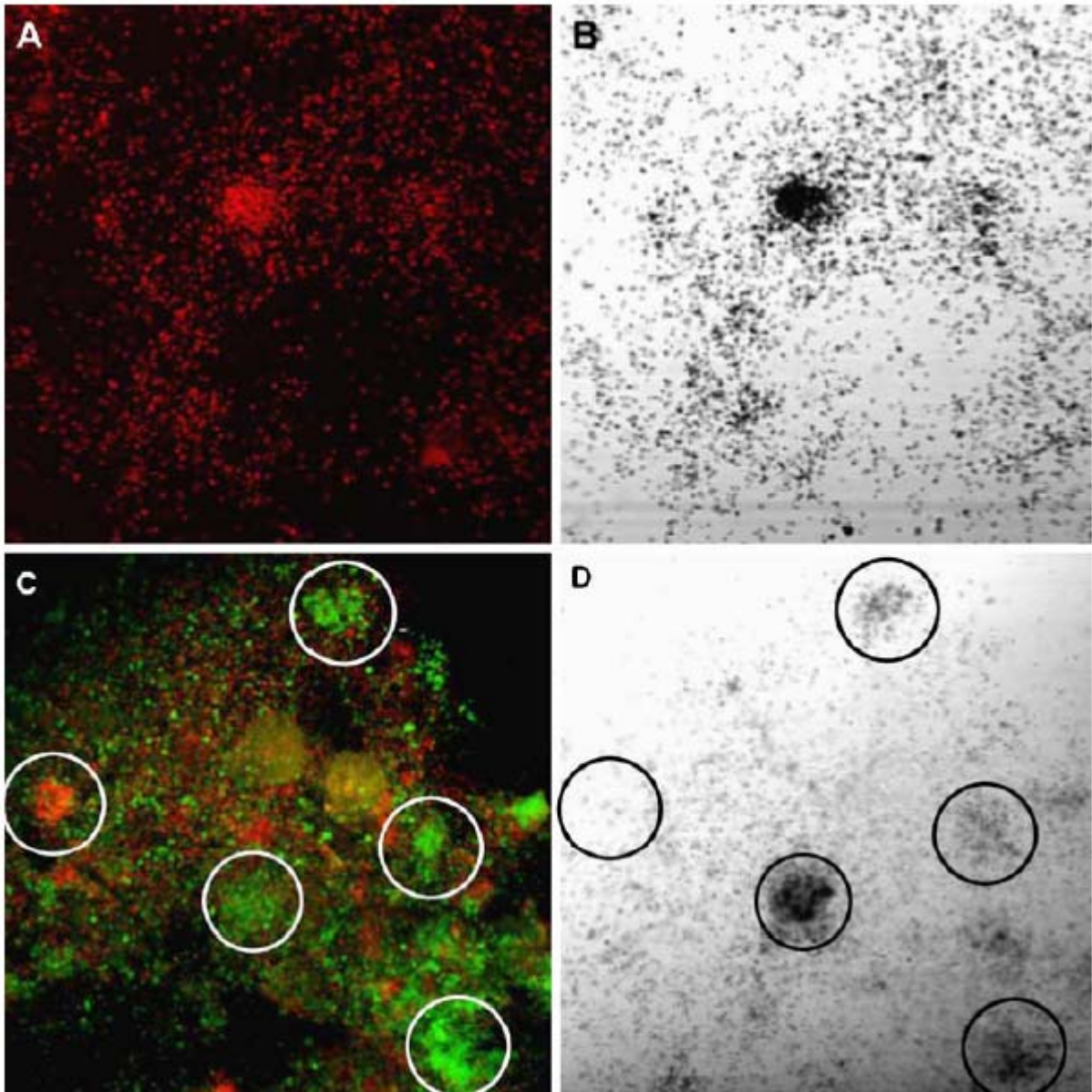


Figure 1 : Images de MAR-FISH sur un biofilm montrant les groupes ayant incorporé le substrat radioactif (^{14}C méthanol). La biofilm a été cultivé en présence de substrat radioactif (^{14}C méthanol) en conditions oxiqes (A et B) et en conditions anoxiques avec du nitrate (C et D). Hybridation *in-situ* par fluorescence (FISH) avec la sonde Gamma42a-CY3 (A) et microautoradiographie (MAR) sur le même champ (B). FISH avec les sondes Alf1b-FITC (vert) et MPH-730-CY3 (rouge) (C) et MAR sur le même champ (B) (Labbé *et al.*, 2007). Gamma42a : *Gamma Proteobacteria* ; Alf1b : *Alpha Proteobacteria* ; MPH-730 : Genus *Methylophaga* spp. except *natronica*.

macromolécules. Les communautés procaryotiques sont ensuite identifiées à l'aide de sondes FISH, spécifiques de groupes bactériens ; celles ayant consommé le substrat sont reconnues par la radioactivité qu'elles émettent (Wagner *et al.*, 2006) (figure 1). Ainsi, cette technique a permis de mettre en évidence les traits écophysiologiques de nombreux microorganismes impliqués dans le fonctionnement d'écosystèmes aquatiques (Nielsen *et al.*, 2006 ; Herndl *et al.*, 2005; Sintès et Herndl, 2006) ou bien encore dans le traitement des eaux usées (Thomsen *et al.*, 2007 ; Wong et Liu, 2007). Labbé *et al.* (2007) ont identifié de la sorte des populations méthylophiles au sein de biofilms provenant de réacteurs de traitement des eaux de mer. Leurs résultats montrent que les *α -Proteobacteria* utilisent le méthanol marqué en ^{14}C en présence de nitrate ce qui suggère leur implication dans la dénitrification. Outre le fait de mettre en évidence certaines fonctions au niveau de phyla bactériens, cette technique peut aussi permettre de souligner des différences en terme de consommation et d'assimilation de substrats entre plusieurs espèces d'un même genre. C'est ce qui a été mis en évidence par Kragelund *et al.* (2007) au sein du genre *Mycolata*. Il faut toutefois noter que cette technique ne semble, à l'heure actuelle, pas encore applicable à tous les environnements. Bien adaptée à l'étude des environnements aquatiques, elle semble beaucoup plus difficile à utiliser pour étudier les communautés bactériennes du sol du fait, entre autre, de l'autofluorescence de certaines de ses particules. Cependant, des améliorations technologiques récentes ont été développées. Elles portent notamment sur l'utilisation du Nycodenz pour séparer les microorganismes des particules du sol (Cérémonie *et al.*, 2004 ; Bertaux *et al.*, 2007). Enfin, une autre limitation à l'utilisation de cette technique provient du nombre limité de groupes microbiens détectables simultanément (Amann *et al.*, 1996). Ainsi, l'utilisation de cette technique pourrait être envisagée pour vérifier les fonctions d'un ou plusieurs groupes bactériens mais non pour l'étude globale de la structure et la fonction des communautés bactériennes de notre sol pollué.

Deux alternatives sont alors envisageables afin d'avoir une vue d'ensemble des communautés bactériennes impliquées dans la biodégradation des polluants. Celles-ci passent toutes les deux par l'utilisation de l'outil biopuce qui permet de cribler à haut débit les taxons bactériens et par l'emploi de molécules traceuses afin d'identifier les bactéries qui dégradent le polluant.

Le premier moyen de relier la structure des communautés bactériennes aux activités de biodégradation passe par l'utilisation de biopuces à isotopes (Isotopes Array). Celles-ci sont basées sur le même principe d'incorporation de substrats radioactifs que la méthode MAR-FISH mais, grâce à leur format, permettent l'identification simultanée de centaines de groupes

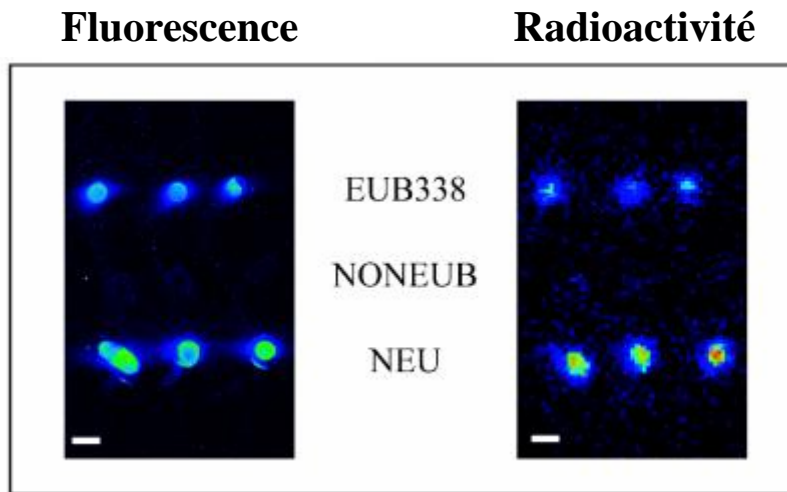


Figure 2 : Images d'Isotope Array sur une souche pure *Nitrosomonas eutropha*. Les ARNr ont été extraits de la souche bactérienne après incubation avec du [^{14}C]bicarbonate. Après marquage des ARNr avec du Cy5, une biopuce prototype a été hybridée. Les lectures de fluorescences et de radioactivité ont ensuite été effectuées. (Adapté de Adamczyk *et al.*, 2003). *EUB338* : sondes ciblant les Eubactéries, *NONEUB* : sondes contrôle ne ciblant pas les Eubactéries et *NEU* : sondes ciblant la plupart des nitrosomonadales halophiles et halotolérants.

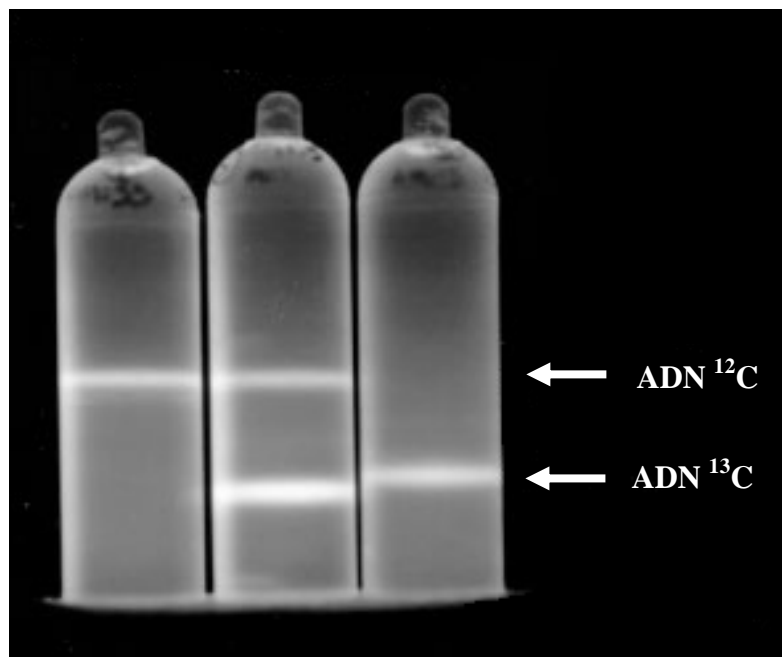


Figure 3 : Séparation des acides nucléiques par la méthode de marquage par isotopes stables (SIP : Stable Isotope Probing). La centrifugation par gradient isopycnique permet de séparer les acides nucléiques des cellules ayant incorporé le substrat d'intérêt de ceux des microorganismes ne l'ayant pas métabolisé. L'ADN ^{13}C provient des microorganismes capables de métaboliser le substrat radioactif. L'ADN ^{12}C appartient aux microorganismes présents dans l'environnement étudié ne pouvant pas métaboliser le substrat radioactif.

bactériens (Wagner *et al.*, 2006). Cette technique consiste, elle aussi, en un radiomarquage préliminaire des communautés bactériennes ayant assimilé le substrat radioactif et ce, notamment au niveau des acides nucléiques. Par la suite, les acides nucléiques de tous les microorganismes de l'écosystème étudié seront extraits, marqués par fluorescence et hybridés sur une biopuce phylogénétique. Lors de l'excitation de cette biopuce par un scanner, les duplex sondes/cibles émettent une fluorescence permettant l'identification des microorganismes présents dans l'échantillon. La lecture de la biopuce par un β imager permet, quant à elle, de cribler ceux ayant incorporé le substrat radioactif (Wagner *et al.*, 2006) (figure 2). L'efficacité d'une biopuce à isotopes ciblant les bactéries oxydant l'ammoniaque a ainsi été démontrée par la mise en évidence de l'incorporation de bicarbonate marqué au ^{14}C par une souche pure de *Nitrosomonas eutropha* (Adamczyk *et al.*, 2003). Cette biopuce à isotopes a permis l'étude de l'évolution de la composition des communautés bactériennes et de leurs activités fixatrices de CO_2 dans deux échantillons de boues activées nitrifiantes (Adamczyk *et al.*, 2003). Toutefois, il faut noter que très peu d'études utilisent cette technique à l'heure actuelle pour étudier le lien entre la structure des populations procaryotiques et leurs fonctions.

La seconde approche permettant d'attribuer la biodégradation des polluants à des phylotypes bactérien particuliers est celle utilisant le marquage par des isotopes stables (SIP : Stable Isotope Probing). Cette méthode consiste à incuber les échantillons étudiés avec un substrat d'intérêt marqué par des isotopes stables (^{13}C , ^{15}N ou ^2H). Celui-ci est assimilé par certains microorganismes et incorporé au niveau des acides nucléiques grâce aux réactions d'anabolisme. Le neutron supplémentaire de l'isotope stable rend les constituants cellulaires plus lourds que ceux ayant incorporé l'atome naturel à la place. Ainsi, une fois extraits, les acides nucléiques des microorganismes ayant utilisé le substrat marqué (fraction lourde) peuvent être séparés de ceux des microorganismes ne l'ayant pas consommé (fraction légère) (figure 3). Les acides nucléiques ainsi récupérés peuvent être utilisés comme matrice dans des techniques moléculaires afin d'identifier les microorganismes ayant métabolisé le substrat (Polz *et al.*, 2003 ; Lueders *et al.*, 2004). Cette technique a très souvent été utilisée afin d'identifier les microorganismes impliqués dans la dégradation de polluants tels le pyrène, le benzène, le naphthalène, le phénanthrène et les PolyChlorinated Biphenyl (PCB) (Singleton *et al.*, 2005 ; Tillmann *et al.*, 2005 ; Yu et Chu, 2005 ; Kasai *et al.*, 2006 ; Madsen, 2006 ; Singleton *et al.*, 2005). Par exemple, l'utilisation de ^{13}C naphthalène et de ^{13}C phénanthrène pour étudier un sol pollué aux HAP et bioremédié en bioréacteur a permis à Singleton *et al.* (2005) d'identifier des souches de *Pseudomonas* et de *Ralstonia* dégradant le naphthalène et

des souches d'*Acidovorax* dégradant le phénanthrène. A notre connaissance, hormis les études portant sur les métabolismes associés au méthane, aucune étude utilisant cette technique n'a été réalisée sur la dégradation des hydrocarbures aliphatiques. Notons aussi que la technique SIP combinée à l'utilisation de biopuces phylogénétiques n'a été employée, à notre connaissance, qu'au cours d'une seule étude. Dans celle-ci, Cebren *et al.* (2007) ont, d'une part, suivi les activités méthanotrophes des communautés bactériennes du sol et, d'autre part, mis en évidence que la stimulation de l'incorporation des isotopes stables pouvait éviter les phénomènes de marquage aspécifiques. Même s'il est évident que ces techniques nécessiteraient des optimisations pour permettre leur utilisation sur nos modèles d'études, les informations qu'elles pourraient apporter sur la structure et la fonction des communautés bactériennes environnementales apparaissent très prometteuses. Notons que les techniques de SIP-ARN (séparation des ARNr marqués) devraient permettre de réduire les temps d'incubation avec les substrats facilitant ainsi la mise en évidence des bactéries métaboliquement actives (Whiteley *et al.*, 2006).

2.2. Voies métaboliques et structure.

Une fois les phylotypes bactériens épurateurs identifiés, il est nécessaire de caractériser les voies métaboliques et les enzymes leur permettant d'effectuer cette biodégradation. Il faut garder à l'esprit que la minéralisation d'un polluant peut impliquer plusieurs étapes enzymatiques réalisées par une ou plusieurs espèces microbiennes (*consortium*). En effet, les membres d'un même consortium peuvent posséder des voies métaboliques séquentielles et complémentaires pour la dégradation d'un même constituant, l'un produisant des métabolites alors dégradés par un autre membre. De même, la co-métabolisation d'un produit par un microorganisme nécessite la présence d'un substrat de croissance pouvant être apporté par un autre membre du consortium (Wackett et Herseberg, 2000). Cette coopération entre les microorganismes peut aussi passer par la production de molécules facilitatrices : les biosurfactants. Ces molécules tensio-actives peuvent améliorer la biodisponibilité des polluants organique hydrophobe et ainsi de favoriser leur biodégradation par d'autres microorganismes (Volkering *et al.*, 1998).

Dans un premier temps, il est nécessaire d'identifier toutes les enzymes et les voies métaboliques microbiennes impliquées dans la dégradation du polluant ayant déjà été décrites dans la littérature ou dans les bases de données. Une fois identifiées, leur suivi au niveau de l'écosystème étudié sera réalisé à l'aide de biopuces fonctionnelles. Leur but est de

caractériser les gènes codant pour les enzymes des voies métaboliques d'intérêts. Dans notre cas, de celles impliquées dans les voies de dégradation des hydrocarbures aliphatiques. Ces enzymes, intervenant dans les étapes de transformation d'un substrat en un produit, exercent leurs activités grâce à un ou plusieurs domaines fonctionnels. Soumis à une forte contrainte fonctionnelle, ces domaines protéiques ont des vitesses d'évolution très lentes. Il existe donc une forte conservation de leurs séquences primaires entre les différentes enzymes isolées chez différents organismes. Ces régions hautement conservées vont permettre la détermination de sondes oligonucléotidiques pour chaque enzyme ou famille d'enzymes et ceci en tenant compte de la dégénérescence du code génétique. Cette sélection de sondes sera réalisée à l'aide d'un logiciel, nommé MetabolicDesign, développé au sein de notre équipe de recherche.

Le suivi parallèle des voies métaboliques (biopuces fonctionnelles) et de la structure des communautés bactérienne (biopuces phylogénétiques) permettra alors de suggérer des liens entre les phylotypes présents et les fonctions identifiées au sein de l'environnement étudié. Cependant, bien que la stratégie de cette biopuce fonctionnelle permette de cibler des enzymes non encore décrites chez certains microorganismes, elle ne possède toutefois pas la capacité de découvrir de gènes codant pour de nouvelles enzymes *stricto sensus*.

Afin d'avoir une vision exploratoire des communautés microbiennes, des approches de métagénomique peuvent être envisagées. Elles pourraient permettre d'accéder non seulement à l'identité des microorganismes présents dans l'environnement étudiés mais également à leurs fonctions. Même si l'assemblage de tous les génomes n'est certainement pas envisageable, ces approches pourraient permettre toutefois de découvrir de nouveaux gènes. Afin de cibler le séquençage vers les fragments possédant des gènes d'intérêts, c'est-à-dire les gènes impliqués dans les voies de dégradation des hydrocarbures aliphatiques, il serait intéressant de mettre en place un crible des banques métagénomiques grâce à la technique d'expression en système hétérologue sur notre substrat d'intérêt. Notons que les informations obtenues sur les gènes d'intérêt pourront permettre la détermination de nouvelles sondes utilisables pour les approches biopuces fonctionnelles. L'utilisation de nouveaux systèmes hétérologues devrait aussi faciliter l'identification de nouveaux gènes d'intérêts.

Outre le fait d'accéder à une meilleure compréhension du fonctionnement des voies de dégradation, ces connaissances pourrait permettre la mise en place des systèmes de sur-expression des enzymes d'intérêt afin d'améliorer les rendement de bioremédiation.

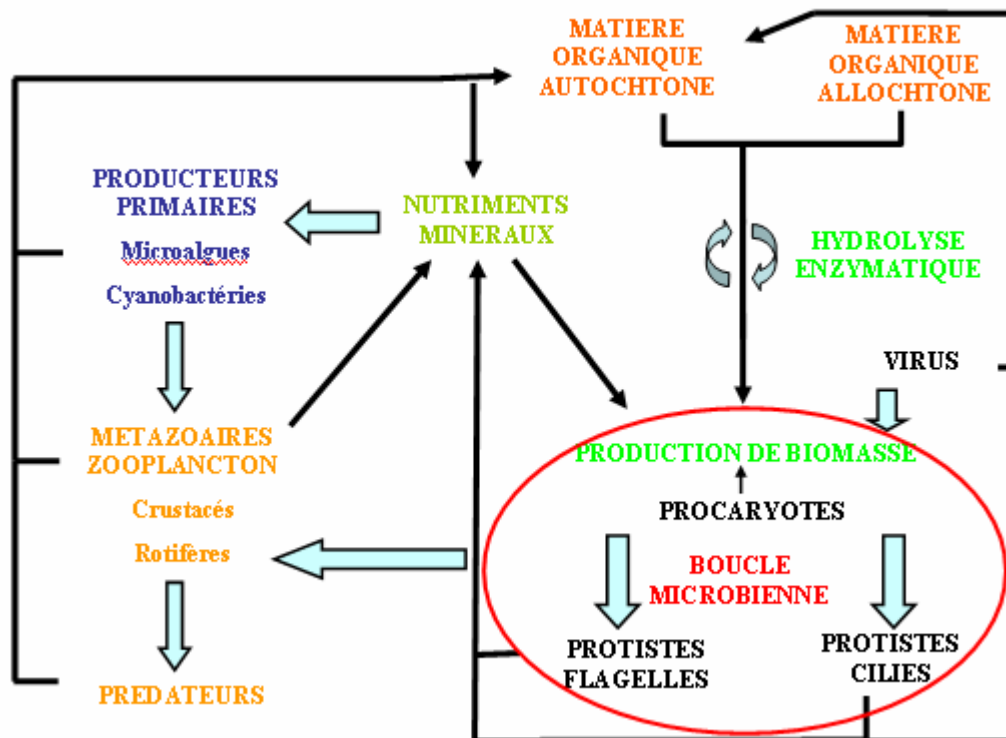


Figure 4 : Importance des communautés procaryotiques dans les réseaux trophiques des écosystèmes aquatiques.

3. Etude de l'ensemble du réseau trophique microbien des sols.

Les différents travaux réalisés au cours de cette thèse se sont focalisés sur les communautés bactériennes des environnements pollués. Cependant, même si ce sont les principaux acteurs de la dégradation de la matière organique, il ne faut pas négliger d'autres microorganismes, tels les champignons, qui interviennent également dans son recyclage. De plus, cette communauté fongique représente en terme d'abondance au niveau de différents types de sols une fraction tout aussi importante que la communauté bactérienne (Fierer *et al.*, 2007) et a montré des capacités à dégrader certains polluants aliphatiques et aromatiques légers (April *et al.*, 2000).

Il est important de garder à l'esprit que ces populations bactériennes et fongiques, qui jouent un rôle fondamental dans la régulation des flux de matière et d'énergie, font partie de la boucle microbienne au même titre que les protistes et les virus. Ce réseau trophique, qui a été essentiellement décrit dans les environnements aquatiques (Pomeroy, 1974 ; Azam, 1983 ; Porter *et al.*, 1985), a permis de mettre en évidence des processus particulièrement importants dans le fonctionnement de ce type d'écosystème (figure 4). Ce même concept a été observé dans les sols par Clarholm (1985). Parmi ces processus citons le fait que les communautés bactériennes sont soumises à différents facteurs de régulation dits « top down » et « bottom up ». Ainsi, elles vont dépendre des ressources (ici notamment le polluant) mais également de la mortalité induite notamment par les virus et par la prédation des protistes. Il serait donc intéressant de prendre en compte ces organismes qui, en jouant sur les communautés bactériennes (Weinbauer et Rassoulzadegan, 2004 ; Wardle, 2006), vont avoir un effet indirect sur l'efficacité de dégradation des polluants.

Ainsi, nos résultats et les approches que nous envisageons ouvrent des perspectives excitantes quant à l'étude des microorganismes environnementaux impliqués dans les processus de biodégradation mais aussi plus largement sur l'étude du fonctionnement global des communautés microbiennes au sein de différents écosystèmes.

REFERENCES BIBLIOGRAPHIQUES

A

- Acar, Y.B. and Alshawabkeh, A.N. (1993) Principles of electrokinetic remediation, *Environ Sci Technol*, **27**, 2638-2647.
- Adamczyk, J., Hesselsoe, M., Iversen, N., Horn, M., Lehner, A., Nielsen, P.H., Schloter, M., Roslev, P. and Wagner, M. (2003) The isotope array, a new tool that employs substrate-mediated labeling of rRNA for determination of microbial community structure and function, *Appl Environ Microbiol*, **69**, 6875-6887.
- Alexander, M. (1961) Introduction to soil microbiology. In United States of America: John Wiley & Sons, I. (ed).
- Alexander, M. (2000) Aging, bioavailability, and overestimation of risk from environmental pollutants, *Environ Sci Technol*, **34**, 4259-4265.
- Amann, R.I., Krumholz, L. and Stahl, D.A. (1990) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology, *J Bacteriol*, **172**, 762-770.
- Amann, R.I., Ludwig, W. and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation, *Microbiol Rev*, **59**, 143-169.
- Amann, R.I., Snaidr, J., Wagner, M., Ludwig, W. and Schleifer, K.H. (1996) In situ visualization of high genetic diversity in a natural microbial community, *J Bacteriol*, **178**, 3496-3500.
- April, T.M., Foght, J.M. and Currah, R.S. (2000) Hydrocarbon-degrading filamentous fungi isolated from flare pit soils in northern and western Canada, *Can J Microbiol*, **46(1)**, 38-49.
- ARIA (2006) Bilan des accidents technologiques 1992-2005. <http://aria.ecologie.gouv.fr/>
- Ashelford, K.E., Weightman, A.J. and Fry, J.C. (2002) PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database, *Nucleic Acids Res*, **30**, 3481-3489.
- Azam, F., Febchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A. and Thingstad, F. (1983) The ecological role of water column microbes in the sea, *Mar Ecol Prog Ser*, **10**, 257-263.

B

- Bae, J.W., Rhee, S.K., Nam, Y.D. and Park, Y.H. (2005) Generation of subspecies level-specific microbial diagnostic microarrays using genes amplified from subtractive suppression hybridization as microarray probes, *Nucleic Acids Res*, **33**, e113.

- Bailly, J., Lehembre, F., Fraissinet-Tachet, L., Verner, M.C., M., L., Wesolowski-Louvel, M., Debaud, J.C. and Marmeisse, R. (2007) Analyse du "métatranscriptome" eucaryote des sols. VIIème Congrès National de la SFM, Nantes, France.
- Balkwill, D.L., Reeves, R.H., Drake, G.R., Reeves, J.Y., Crocker, F.H., King, M.B. and Boone, D.R. (1997) Phylogenetic characterization of bacteria in the subsurface microbial culture collection, *FEMS Microbiol Rev*, **20**, 201-216.
- Banuelos, G., LeDuc, D.L., Pilon-Smits, E.A. and Terry, N. (2007) Transgenic Indian mustard overexpressing selenocysteine lyase or selenocysteine methyltransferase exhibit enhanced potential for selenium phytoremediation under field conditions, *Environ Sci Technol*, **41**, 599-605.
- BASIAS (2004) <http://basias.brgm.fr/>.
- BASOL (2004) <http://basol.environnement.gouv.fr/>.
- Bathe, S., Schwarzenbeck, N. and Hausner, M. (2005) Plasmid-mediated bioaugmentation of activated sludge bacteria in a sequencing batch moving bed reactor using pNB2, *Lett Appl Microbiol*, **41**, 242-247.
- Baudouin, C., Charveron, M., Tarroux, R. and Gall, Y. (2002) Environmental pollutants and skin cancer, *Cell Biol Toxicol*, **18**, 341-348.
- Bertaux, J., Gloger, U., Schmid, M., Hartmann, A. and Scheu, S. (2007) Routine fluorescence in situ hybridization in soil, *J Microbiol Methods*, **69**, 451-460.
- Block, R., Allworth, N. and Bishop, M. (1991) Assessment of diesel contamination in soils. In Calabrese, E.a.K., P. (ed), Hydrocarbon contaminated soils. Lewis Publishers, Chelsea, 135-148.
- Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A. and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs, *Environ Microbiol*, **5**, 566-582.
- Bogan, B.W., Sullivan, W.R., Kayser, K.J., Derr, K.D., Aldrich, H.C. and Paterek, J.R. (2003) *Alkanindiges illinoisensis* gen. nov., sp. nov., an obligately hydrocarbonoclastic, aerobic squalane-degrading bacterium isolated from oilfield soils, *Int J Syst Evol Microbiol*, **53**, 1389-1395.
- Boubakri, H., Beuf, M., Simonet, P. and Vogel, T.M. (2006) Development of metagenomic DNA shuffling for the construction of a xenobiotic gene, *Gene*, **375**, 87-94.
- Brodie, E.L., DeSantis, T.Z., Parker, J.P., Zubietta, I.X., Piceno, Y.M. and Andersen, G.L. (2007) Urban aerosols harbor diverse and dynamic bacterial populations, *Proc Natl Acad Sci U S A*, **104**, 299-304.

Brousseau, R., Hill, J.E., Prefontaine, G., Goh, S.H., Harel, J. and Hemmingsen, S.M. (2001) Streptococcus suis serotypes characterized by analysis of chaperonin 60 gene sequences, *Appl Environ Microbiol*, **67**, 4828-4833.

C

Castiglioni, B., Rizzi, E., Frosini, A., Sivonen, K., Rajaniemi, P., Rantala, A., Mugnai, M.A., Ventura, S., Wilmotte, A., Boutte, C., Grubisic, S., Balthasart, P., Consolandi, C., Bordoni, R., Mezzelani, A., Battaglia, C. and De Bellis, G. (2004) Development of a universal microarray based on the ligation detection reaction and 16S rna gene polymorphism to target diversity of cyanobacteria, *Appl Environ Microbiol*, **70**, 7161-7172.

Chou, C.C., Chen, C.H., Lee, T.T. and Peck, K. (2004) Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression, *Nucleic Acids Res*, **32**, e99.

Cebon, A., Bodrossy, L., Stralis-Pavese, N., Singer, A.C., Thompson, I.P., Prosser, J.I. and Murrell, J.C. (2007) Nutrient amendments in soil DNA stable isotope probing experiments reduce the observed methanotroph diversity, *Appl Environ Microbiol*, **73**, 798-807.

Ceremonie, H., Buret, F., Simonet, P. and Vogel, T.M. (2004) Isolation of lightning-competent soil bacteria, *Appl Environ Microbiol*, **70**, 6342-6346.

Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B. and Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life, *Science*, **311**, 1283-1287.

Clarholm, M. (1985) Possible roles for roots, bacteria, protozoa and fungi in supplying nitrogen to plants. In: Fitter, AH (ed): Ecological interactions in soil. Blackwell, Oxford, pp 355-365.

Clemens, S., Palmgren, M.G. and Kramer, U. (2002) A long way ahead: understanding and engineering plant metal accumulation, *Trends Plant Sci*, **7**, 309-315.

Curtis, T.P. and Sloan, W.T. (2004) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology, *Curr Opin Microbiol*, **7**, 221-226.

D

- Da Silva, M.L. and Alvarez, P.J. (2004) Enhanced anaerobic biodegradation of benzene-toluene-ethylbenzene-xylene-ethanol mixtures in bioaugmented aquifer columns, *Appl Environ Microbiol*, **70**, 4720-4726.
- Dahllof, I. (2002) Molecular community analysis of microbial diversity, *Curr Opin Biotechnol*, **13**, 213-217.
- Daniel, R. (2004) The soil metagenome--a rich resource for the discovery of novel natural products, *Curr Opin Biotechnol*, **15**, 199-204.
- Delage, P. and Schrefler, B. (2005) Géomécanique environnementale : sols pollués et déchets. Hermès Science publications.
- DeLong, E.F., Wickham, G.S. and Pace, N.R. (1989) Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells, *Science*, **243**, 1360-1363.
- Deprince, A. (2003) La faune du sol. Le courrier de l'environnement de l'INRA. <http://www.inra.fr/dpenv/>.
- DeSantis, T.Z., Brodie, E.L., Moberg, J.P., Zubieta, I.X., Piceno, Y.M. and Andersen, G.L. (2007) High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment, *Microb Ecol*, **53**, 371-383.
- Dharmadi, Y. and Gonzalez, R. (2004) DNA microarrays: experimental issues, data analysis, and application to bacterial systems, *Biotechnol Prog*, **20**, 1309-1324.
- Dorrell, N., Hinchliffe, S.J. and Wren, B.W. (2005) Comparative phylogenomics of pathogenic bacteria by microarray analysis, *Curr Opin Microbiol*, **8**, 620-626.
- Duchaufour, P. (2001) Introduction à la Science du Sol. Sol, végétation, environnement. DUNOD, Paris.
- Dunbar, J., Takala, S., Barns, S.M., Davis, J.A. and Kuske, C.R. (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning, *Appl Environ Microbiol*, **65**, 1662-1669.

E

- Ehlers, G.A. and Loibner, A.P. (2006) Linking organic pollutant (bio)availability with geosorbent properties and biomimetic methodology: a review of geosorbent characterisation and (bio)availability prediction, *Environ Pollut*, **141**, 494-512.
- Ehrenreich, A. (2006) DNA microarray technology for the microbiologist: an overview, *Appl Microbiol Biotechnol*, **73**, 255-273.

- El Fantroussi, S. and Agathos, S.N. (2005) Is bioaugmentation a feasible strategy for pollutant removal and site remediation?, *Curr Opin Microbiol*, **8**, 268-275.
- Eriksson, M., Sodersten, E., Yu, Z., Dalhammar, G. and Mohn, W. W (2003) Degradation of polycyclic aromatic hydrocarbons at low temperature under aerobic and nitrate-reducing conditions in enrichment cultures from northern soils, *Appl Environ Microbiol*, **69**, 275–284.
- Evans, F.F., Seldin, L., Sebastian, G.V., Kjelleberg, S., Holmstrom, C. and Rosado, A.S. (2004) Influence of petroleum contamination and biostimulation treatment on the diversity of *Pseudomonas* spp. in soil microcosms as evaluated by 16S rRNA based-PCR and DGGE, *Lett Appl Microbiol*, **38**, 93-98.

F

- Fent, K. (2004) Ecotoxicological effects at contaminated sites, *Toxicology*, **205**, 223-240.
- Ferrari, B.C., Binnerup, S.J. and Gillings, M. (2005) Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria, *Appl Environ Microbiol*, **71**, 8714-8720.
- Fotin, A.V., Drobyshev, A.L., Proudnikov, D.Y., Perov, A.N. and Mirzabekov, A.D. (1998) Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips, *Nucleic Acids Res*, **26**, 1515-1521.
- Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., Rayhawk, S., Knight, R., Rohwer, F. and Jackson, R.B. (2007) Metagenomic and small-subunit rRNA analyses of the genetic diversity of bacteria, archaea, fungi, and viruses in soil, *Appl Environ Microbiol*, **7**, [Epub ahead of print].
- Fuchs, B.M., Wallner, G., Beisker, W., Schwippl, I., Ludwig, W. and Amann, R. (1998) Flow cytometric analysis of the in situ accessibility of *Escherichia coli* 16S rRNA for fluorescently labeled oligonucleotide probes, *Appl Environ Microbiol*, **64**, 4973-4982.

G

- Gans, J., Wolinsky, M. and Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil, *Science*, **309**, 1387-1390.
- Gentry, T.J., Wickham, G.S., Schadt, C.W., He, Z. and Zhou, J. (2006) Microarray applications in microbial ecology research, *Microb Ecol*, **52**, 159-175.
- Gibson, G. (2002) Microarrays in ecology and evolution: a preview, *Mol Ecol*, **11**, 17-24.

- Gieg, L.M. and Suflita, J.M. (2002) Detection of anaerobic metabolites of saturated and aromatic hydrocarbons in petroleum-contaminated aquifers, *Environ Sci Technol*, **36**, 3755-3762.
- Giovannoni, S.J., Britschgi, T.B., Moyer, C.L. and Field, K.G. (1990) Genetic diversity in Sargasso sea bacterioplankton, *Nature*, **345**, 60-63.
- Gobat, J.-M., Aragno, M. and Matthey, W. (2003) *Le Sol vivant. Bases de pédologie, biologie des sols*. Presses polytechniques et universitaires romandes.
- GOLD: Genomes Online Database (2007) www.genomesonline.org.
- Goldberg, S.M., Johnson, J., Busam, D., Feldblyum, T., Ferriera, S., Friedman, R., Halpern, A., Khouri, H., Kravitz, S.A., Lauro, F.M., Li, K., Rogers, Y.H., Strausberg, R., Sutton, G., Tallon, L., Thomas, T., Venter, E., Frazier, M. and Venter, J.C. (2006) A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes, *Proc Natl Acad Sci U S A*, **103**, 11240-11245.
- Grignard, V. (1935) *Traité de chimie organique. Tome III.*, France.
- Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R. and Smith, L.M. (1994) Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports, *Nucleic Acids Res*, **22**, 5456-5465.
- Guschin, D.Y., Mobarry, B.K., Proudnikov, D., Stahl, D.A., Rittmann, B.E. and Mirzabekov, A.D. (1997) Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology, *Appl Environ Microbiol*, **63**, 2397-2402.

H

- Hackl, E., Zechmeister-Boltenstern, S., Bodrossy, L. and Sessitsch, A. (2004) Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils, *Appl Environ Microbiol*, **70**, 5057-5065.
- Hamamura, N., Olson, S.H., Ward, D.M. and Inskeep, W.P. (2006) Microbial population dynamics associated with crude-oil biodegradation in diverse soils, *Appl Environ Microbiol*, **72**, 6316-6324.
- Han, Y.H., Liu, W.Z., Shi, Y.Z., Lu, L.Q., Xiao, S., Zhang, Q.H. and Zhao, G.P. (2007) Comparative genomics profiling of clinical isolates of *Helicobacter pylori* in Chinese populations using DNA microarray, *J Microbiol*, **45**, 21-28.
- Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms, *Microbiol Mol Biol Rev*, **68**, 669-685.

- Hattori, T., Mitsui, H., Haga, H., Wakao, N., Shikano, S., Gorlach, K., Kasahara, Y., el-Beltagy, A. and Hattori, R. (1997) Advances in soil microbial ecology and the biodiversity, *Antonie Van Leeuwenhoek*, **72**, 21-28.
- He, Z., Wu, L., Li, X., Fields, M.W. and Zhou, J. (2005) Empirical establishment of oligonucleotide probe design criteria, *Appl Environ Microbiol*, **71**, 3753-3760.
- Herndl, G.J., Reinthaler, T., Teira, E., van Aken, H., Veth, C., Pernthaler, A. and Pernthaler, J. (2005) Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean, *Appl Environ Microbiol*, **71(5)**, 2303-2309.
- Hill, J.E., Gottschalk, M., Brousseau, R., Harel, J., Hemmingsen, S.M. and Goh, S.H. (2005) Biochemical analysis, cpn60 and 16S rDNA sequence data indicate that *Streptococcus suis* serotypes 32 and 34, isolated from pigs, are *Streptococcus orisratti*, *Vet Microbiol*, **107**, 63-69.
- Hoier, C.K., Sonnenborg, T.O., Jensen, K.H., Kortegaard, C. and Nasser, M.M. (2007) Experimental investigation of pneumatic soil vapor extraction, *J Contam Hydrol*, **89**, 29-47.
- Horz, H.P., Yimga, M.T. and Liesack, W. (2001) Detection of methanotroph diversity on roots of submerged rice plants by molecular retrieval of pmoA, mmoX, mxaF, and 16S rRNA and ribosomal DNA, including pmoA-based terminal restriction fragment length polymorphism profiling, *Appl Environ Microbiol*, **67**, 4177-4185.
- Huang, W., Peng, P., Yu, Z. and FU, J. (2003) Effects of organic matter heterogeneity on sorption and desorption of organic contaminants by soils and sediments, *Appl Geochem*, **18**, 955-972.
- Hugenholtz, P., Goebel, B.M. and Pace, N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity, *J Bacteriol*, **180**, 4765-4774.
- Hur, J. and Schlautman, M.A. (2004) Effects of mineral surfaces on pyrene partitioning to well-characterized humic substances, *J Environ Qual*, **33**, 1733-1742.

J

- Jacques, R.J., Okeke, B.C., Bento, F.M., Teixeira, A.S., Peralba, M.C. and Camargo, F.A. (2007) Microbial consortium bioaugmentation of a polycyclic aromatic hydrocarbons contaminated soil, *Bioresour Technol*.
- Janssen, P.H. (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes, *Appl Environ Microbiol*, **72**, 1719-1728.

- Joseph, S.J., Hugenholtz, P., Sangwan, P., Osborne, C.A. and Janssen, P.H. (2003) Laboratory cultivation of widespread and previously uncultured soil bacteria, *Appl Environ Microbiol*, **69**, 7210-7215.
- Jung, H., Ahn, Y., Choi, H. and Kim, I.S. (2005) Effects of in-situ ozonation on indigenous microorganisms in diesel contaminated soil: survival and regrowth, *Chemosphere*, **61**, 923-932.

K

- Kafatos, F.C., Jones, C.W. and Efstratiadis, A. (1979) Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure, *Nucleic Acids Res*, **7**, 1541-1552.
- Kalin, R.M. (2004) Engineered passive bioreactive barriers: risk-managing the legacy of industrial soil and groundwater pollution, *Curr Opin Microbiol*, **7**, 227-238.
- Kane, M.D., Jatkoa, T.A., Stumpf, C.R., Lu, J., Thomas, J.D. and Madore, S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays, *Nucleic Acids Res*, **28**, 4552-4557.
- Kaplan, C.W. and Kitts, C.L. (2004) Bacterial succession in a petroleum land treatment unit, *Appl Environ Microbiol*, **70**, 1777-1786.
- Karamalidis, A.K. and Voudrias, E.A. (2007) Cement-based stabilization/solidification of oil refinery sludge: Leaching behavior of alkanes and PAHs, *J Hazard Mater*, **148**, 122-135.
- Kasai, Y., Takahata, Y., Manefield, M. and Watanabe, K. (2006) RNA-based stable isotope probing and isolation of anaerobic benzene-degrading bacteria from gasoline-contaminated groundwater, *Appl Environ Microbiol*, **72**, 3586-3592.
- Kemp, P.F. and Aller, J.Y. (2004) Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us, *FEMS Microbial Ecol*, **47**, 161-177.
- Khan, F.I., Husain, T. and Hejazi, R. (2004) An overview and analysis of site remediation technologies, *J Environ Manage*, **71**, 95-122.
- Kim, I.S., Foght, J.M. and Gray, M.R. (2002) Selective transport and accumulation of alkanes by *Rhodococcus erythropolis* S+14He, *Biotechnol Bioeng*, **80**, 650-659.
- Kirk, J.L., Beaudette, L.A., Hart, M., Moutoglis, P., Klironomos, J.N., Lee, H. and Trevors, J.T. (2004) Methods of studying soil microbial diversity, *J Microbiol Methods*, **58**, 169-188.

- Koizumi, Y., Kelly, J.J., Nakagawa, T., Urakawa, H., El-Fantroussi, S., Al-Muzaini, S., Fukui, M., Urushigawa, Y. and Stahl, D.A. (2002) Parallel characterization of anaerobic toluene- and ethylbenzene-degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA-DNA membrane hybridization, and DNA microarray technology, *Appl Environ Microbiol*, **68**, 3215-3225.
- Kragelund, C., Remesova, Z., Nielsen, J.L., Thomsen, T.R., Eales, K., Seviour, R., Wanner, J. and Nielsen, P.H. Ecophysiology of mycolic acid-containing Actinobacteria (Mycolata) in activated sludge foams, *FEMS Microbiol Ecol*, **61(1)**, 174-184.
- Kramer, U. (2005) Phytoremediation: novel approaches to cleaning up polluted soils, *Curr Opin Biotechnol*, **16**, 133-141.
- Kreil, D.P., Russell, R.R. and Russell, S. (2006) Microarray oligonucleotide probes, *Methods Enzymol*, **410**, 73-98.

L

- Labbe, N., Laurin, V., Juteau, P., Parent, S. and Villemur, R. (2007) Microbiological community structure of the biofilm of a methanol-fed, marine denitrification system, and identification of the methanol-utilizing microorganisms, *Microb Ecol*, **53**, 621-630.
- Labbe, D., Margesin, R., Schinner, F., Whyte, L.G. and Greer, C.W. (2007) Comparative phylogenetic analysis of microbial communities in pristine and hydrocarbon-contaminated Alpine soils, *FEMS Microbiol Ecol*, **59**, 466-475.
- Lander, E.S. (1999) Array of hope, *Nat Genet*, **21**, 3-4.
- Leadbetter, J.R. (2003) Cultivation of recalcitrant microbes: cells are alive, well and revealing their secrets in the 21st century laboratory, *Curr Opin Microbiol*, **6**, 274-281.
- Lee, D.H., Zo, Y.G. and Kim, S.J. (1996) Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism, *Appl Environ Microbiol*, **62**, 3112-3120.
- Lee, N., Nielsen, P.H., Andreasen, K.H., Juretschko, S., Nielsen, J.L., Schleifer, K.H. and Wagner, M. (1999) Combination of fluorescent in situ hybridization and microautoradiography-a new tool for structure-function analyses in microbial ecology, *Appl Environ Microbiol*, **65**, 1289-1297.
- Lee, S.W., Won, K., Lim, H.K., Kim, J.C., Choi, G.J. and Cho, K.Y. (2004) Screening for novel lipolytic enzymes from uncultured soil microorganisms, *Appl Microbiol Biotechnol*, **65**, 720-726.

- Liles, M.R., Manske, B.F., Bintrim, S.B., Handelsman, J. and Goodman, R.M. (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library, *Appl Environ Microbiol*, **69**, 2684-2691.
- Lin, L.H., Wang, P.L., Rumble, D., Lippmann-Pipke, J., Boice, E., Pratt, L.M., Sherwood Lollar, B., Brodie, E.L., Hazen, T.C., Andersen, G.L., DeSantis, T.Z., Moser, D.P., Kershaw, D. and Onstott, T.C. (2006) Long-term sustainability of a high-energy, low-diversity crustal biome, *Science*, **314**, 479-482.
- Liu, W.T., Marsh, T.L., Cheng, H. and Forney, L.J. (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., Adamczyk, J., Meier, H., Ernst, J., Schleifer, K.H. and Wagner, M. (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.
- Loy, A., Kusel, K., Lehner, A., Drake, H.L. and Wagner, M. (2004) Microarray and functional gene analyses of sulfate-reducing prokaryotes in low-sulfate, acidic fens reveal cooccurrence of recognized genera and novel lineages, *Appl Environ Microbiol*, **70**, 6998-7009.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. and Schleifer, K.H. (2004) ARB: a software environment for sequence data, *Nucleic Acids Res*, **32**, 1363-1371.
- Lueders, T., Pommerenke, B. and Friedrich, M.W. (2004) Stable-isotope probing of microorganisms thriving at thermodynamic limits: syntrophic propionate oxidation in flooded soil, *Appl Environ Microbiol*, **70**, 5778-5786.
- Luijten, M.L., de Weert, J., Smidt, H., Boschker, H.T., de Vos, W.M., Schraa, G. and Stams, A.J. (2003) Description of *Sulfurospirillum halorespirans* sp. nov., an anaerobic, tetrachloroethene-respiring bacterium, and transfer of *Dehalospirillum multivorans* to the genus *Sulfurospirillum* as *Sulfurospirillum multivorans* comb. nov, *Int J Syst Evol Microbiol*, **53**, 787-793.

Lyyra, S., Meagher, R.B., Kim, T., Heaton, A., Montello, P., Balish, R.S. and Merkle, S.A. (2007) Coupling two mercury resistance genes in Eastern cottonwood enhances the processing of organomercury, *Plant Biotechnol J*, **5**, 254-262.

M

Madsen, E.L. (2006) The use of stable isotope probing techniques in bioreactor and field studies on bioremediation, *Curr Opin Biotechnol*, **17**, 92-97.

Maier, R.M., Pepper, I.L. and Gerba, C.P. (2000) Environmental microbiology. San Diego: Academic Press.

Margesin, R., Labbe, D., Schinner, F., Greer, C.W. and Whyte, L.G. (2003) Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine Alpine soils, *Appl Environ Microbiol*, **69**, 3085-3092.

McQuain, M.K., Seale, K., Peek, J., Fisher, T.S., Levy, S., Stremler, M.A. and Haselton, F.R. (2004) Chaotic mixer improves microarray hybridization, *Anal Biochem*, **325**, 215-226.

Merino, J. and Bucala, V. (2007) Effect of temperature on the release of hexadecane from soil by thermal treatment, *J Hazard Mater*, **143**, 455-461.

Miller, S.R., Augustine, S., Olson, T.L., Blankenship, R.E., Selker, J. and Wood, A.M. (2005) Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene, *Proc Natl Acad Sci U S A*, **102**, 850-855.

Moisander, P.H., Shiue, L., Steward, G.F., Jenkins, B.D., Bebout, B.M. and Zehr, J.P. (2006) Application of a nifH oligonucleotide microarray for profiling diversity of N₂-fixing microorganisms in marine microbial mats, *Environ Microbiol*, **8**, 1721-1735.

Moisander, P.H., Morrison, A.E., Ward, B.B., Jenkins, B.D. and Zehr, J.P. (2007) Spatial-temporal variability in diazotroph assemblages in Chesapeake Bay using an oligonucleotide nifH microarray, *Environ Microbiol*, **9**, 1823-1835.

Moslemy, P., Neufeld, R.J. and Guiot, S.R. (2002) Biodegradation of gasoline by gellan gum-encapsulated bacterial cells, *Biotechnol Bioeng*, **80**, 175-184.

Mulligan, C.N. and Yong, R.N. (2004) Natural attenuation of contaminated soils, *Environ Int*, **30**, 587-601.

Muyzer, G., de Waal, E.C. and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA, *Appl Environ Microbiol*, **59**, 695-700.

N

- Nebert, D.W., Dalton, T.P., Okey, A.B. and Gonzalez, F.J. (2004) Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer, *J Biol Chem*, **279**, 23847-23850.
- Neufeld, J.D., Yu, Z., Lam, W. and Mohn, W.W. (2004) Serial analysis of ribosomal sequence tags (SARST): a high-throughput method for profiling complex microbial communities, *Environ Microbiol*, **6**, 131-144.
- Nielsen, J.L., Klausen, C., Nielsen, P.H., Burford, M. and Jorgensen, N.O. (2006) Detection of activity among uncultured Actinobacteria in a drinking water reservoir, *FEMS Microbiol Ecol*, **55(3)**, 432-438.
- Nocker, A., Burr, M. and Camper, A.K. (2007) Genotypic microbial community profiling: a critical technical review, *Microb Ecol*, **54**, 276-289.
- Nocker, A., Sossa, K.E., Camper, A.K. (2007) Molecular monitoring of disinfection efficacy using propidium monoazide in combination with quantitative PCR, *J Microbiol Methods*, **70(2)**, 252-260.

O

- Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R. and Stahl, D.A. (1986) Microbial ecology and evolution: a ribosomal RNA approach, *Annu Rev Microbiol*, **40**, 337-365.
- Ouverney, C.C. and Fuhrman, J.A. (1999) Combined microautoradiography-16S rRNA probe technique for determination of radioisotope uptake by specific microbial cell types in situ, *Appl Environ Microbiol*, **65**, 1746-1752.

P

- Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere, *Science*, **276**, 734-740.
- Pelczar, M.J., Chan, E.C.S. and Krieg, N.R. (1993) Prologue: discovering the microbial world. In Prancan, K.M. and Gordon, H. (eds), *Microbiology: concepts and applications*. McGraw-Hill, New-York, 2-20.
- Peplies, J., Lachmund, C., Glockner, F.O. and Manz, W. (2006) A DNA microarray platform based on direct detection of rRNA for characterization of freshwater sediment-related prokaryotic communities, *Appl Environ Microbiol*, **72**, 4829-4838.

- Perfumo, A., Banat, I.M., Marchant, R. and Vezzulli, L. (2007) Thermally enhanced approaches for bioremediation of hydrocarbon-contaminated soils, *Chemosphere*, **66**, 179-184.
- Pernthaler, A., Pernthaler, J. and Amann, R. (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria, *Appl Environ Microbiol*, **68**, 3094-3101.
- Peterson, A.W., Heaton, R.J. and Georgiadis, R.M. (2001) The effect of surface probe density on DNA hybridization, *Nucleic Acids Res*, **29**, 5163-5168.
- Phillips, T.M., Seech, A.G., Lee, H. and Trevors, J.T. (2005) Biodegradation of hexachlorocyclohexane (HCH) by microorganisms, *Biodegradation*, **16**, 363-392.
- Polz, M.F., Bertilsson, S., Acinas, S.G. and Hunt, D. (2003) A(r)Ray of hope in analysis of the function and diversity of microbial communities, *Biol Bull*, **204**, 196-199.
- Pomeroy, L.R. (1974) The ocean's food web, a changing paradigm, *Bioscience*, **24**, 499-504.
- Popp, N., Schlomann, M. and Mau, M. (2006) Bacterial diversity in the active stage of a bioremediation system for mineral oil hydrocarbon-contaminated soils, *Microbiology*, **152**, 3291-3304.
- Porter, K.G., Sherr, E.B., Sherr, B.F., Pace, M. and Sanders, R.W. (1985) Protozoa in planktonic food webs, *J Protozool*, **32**, 409-415.
- Powell, S.M., Ferguson, S.H., Bowman, J.P. and Snape, I. (2006) Using real-time PCR to assess changes in the hydrocarbon-degrading microbial community in antarctic soil during bioremediation, *Microb Ecol*, **52**, 523-532.
- Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Nolte, A.W., Sonnenberg, R., Staehler, P., Beier, M. and Tautz, D. (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted, *Nucleic Acids Res*, **34**, e66.
- Pozhitkov, A.E., Stedtfeld, R.D., Hashsham, S.A. and Noble, P.A. (2007) Revision of the nonequilibrium thermal dissociation and stringent washing approaches for identification of mixed nucleic acid targets by microarrays, *Nucleic Acids Res*, **35**, e70.

R

- Ranjard, L., Poly, F., Lata, J.C., Mougél, C., Thioulouse, J. and Nazaret, S. (2001) Characterization of bacterial and fungal soil communities by automated ribosomal

- intergenic spacer analysis fingerprints: biological and methodological variability, *Appl Environ Microbiol*, **67**, 4479-4487.
- Ranjard, L. and Richaume, A. (2001) Quantitative and qualitative microscale distribution of bacteria in soil, *Res Microbiol*, **152**, 707-716.
- Rappe, M.S. and Giovannoni, S.J. (2003) The uncultured microbial majority, *Annu Rev Microbiol*, **57**, 369-394.
- Reid, B.J., Jones, K.C. and Semple, K.T. (2000) Bioavailability of persistent organic pollutants in soils and sediments--a perspective on mechanisms, consequences and assessment, *Environ Pollut*, **108**, 103-112.
- Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X. and Zhou, J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays, *Appl Environ Microbiol*, **70**, 4303-4317.
- Ribosomal Database Project II (2007) <http://rdp.cme.msu.edu/>.
- Rimour, S., Hill, D., Milton, C. and Peyret, P. (2005) GoArrays: highly dynamic and efficient microarray probe design, *Bioinformatics*, **21**, 1094-1103.
- Riser-Roberts, E. (1998) Remediation of Petroleum Contaminated Soil: Biological, Physical, and Chemical Processes. Lewis Publishers.
- Roling, W.F. and van Verseveld, H.W. (2002) Natural attenuation: what does the subsurface have in store?, *Biodegradation*, **13**, 53-64.
- Ronaghi, M., Uhlen, M. and Nyren, P. (1998) A sequencing method based on real-time pyrophosphate, *Science*, **281**, 363, 365.
- Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J. and Goodman, R.M. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms, *Appl Environ Microbiol*, **66**, 2541-2547.
- Ruder, A.M. (2006) Potential health effects of occupational chlorinated solvent exposure, *Ann N Y Acad Sci*, **1076**, 207-227.
- Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.H., Falcon, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Birmingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Neilson, K.,

Friedman, R., Frazier, M. and Venter, J.C. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific, *PLoS Biol*, **5**, e77.

S

Salt, D.E., Smith, R.D. and Raskin, I. (1998) Phytoremediation, *Annu Rev Plant Physiol Plant Mol Biol*, **49**, 643-668.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc Natl Acad Sci U S A*, **74**, 5463-5467.

Sanguin, H., Remenant, B., Dechesne, A., Thioulouse, J., Vogel, T.M., Nesme, X., Moenne-Loccoz, Y. and Grundmann, G.L. (2006) Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. and bacterial communities, *Appl Environ Microbiol*, **72**, 4302-4312.

SantaLucia, J., Jr., Allawi, H.T. and Seneviratne, P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability, *Biochemistry*, **35**, 3555-3562.

Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science*, **270**, 467-470.

Schliep, A. and Rahmann, S. (2006) Decoding non-unique oligonucleotide hybridization experiments of targets related by a phylogenetic tree, *Bioinformatics*, **22**, e424-430.

Schloss, P.D. and Handelsman, J. (2004) Status of the microbial census, *Microbiol Mol Biol Rev*, **68**, 686-691.

Schwarzenbach, R.P. and Westall, J. (1981) Transport of nonpolar organic compounds from surface water to groundwater. Laboratory sorption studies., *Environ Sci Technol*, **15**, 1360-1367.

Sekar, R., Pernthaler, A., Pernthaler, J., Warnecke, F., Posch, T. and Amann, R. (2003) An improved protocol for quantification of freshwater Actinobacteria by fluorescence in situ hybridization, *Appl Environ Microbiol*, **69**(5), 2928-35.

Serrano, A., Gallego, M., Gonzalez, J.L. and Tejada, M. (2007) Natural attenuation of diesel aliphatic hydrocarbons in contaminated agricultural soil, *Environ Pollut*.

Sessitsch, A., Hackl, E., Wenzl, P., Kilian, A., Kostic, T., Stralis-Pavese, N., Sandjong, B.T. and Bodrossy, L. (2006) Diagnostic microbial microarrays in soil ecology, *New Phytol*, **171**, 719-735.

- Sessitsch, A., Weilharter, A., Gerzabek, M.H., Kirchmann, H. and Kandeler, E. (2001) Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment, *Appl Environ Microbiol*, **67**, 4215-4224.
- Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. and Church, G.M. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome, *Science*, **309**, 1728-1732.
- Shimada, T. (2006) Xenobiotic-metabolizing enzymes involved in activation and detoxification of carcinogenic polycyclic aromatic hydrocarbons, *Drug Metab Pharmacokinet*, **21**, 257-276.
- Shimizu, T., Ohshima, S., Ohtani, K., Hoshino, K., Honjo, K. and Hayashi, H. (2001) Sequence heterogeneity of the ten rRNA operons in *Clostridium perfringens*, *Syst Appl Microbiol*, **24**, 149-156.
- Sikkema, J., de Bont, J.A. and Poolman, B. (1995) Mechanisms of membrane toxicity of hydrocarbons, *Microbiol Rev*, **59**, 201-222.
- Simu, K. and Hagstrom, A. (2004) Oligotrophic bacterioplankton with a novel single-cell life strategy, *Applied and environmental microbiology*, **70**, 2445-2451.
- Singleton, D.R., Powell, S.N., Sangaiah, R., Gold, A., Ball, L.M. and Aitken, M.D. (2005) Stable-isotope probing of bacteria capable of degrading salicylate, naphthalene, or phenanthrene in a bioreactor treating contaminated soil, *Appl Environ Microbiol*, **71**, 1202-1209.
- Sintes, E. and Herndl, G.J. (2006) Quantifying substrate uptake by individual cells of marine bacterioplankton by catalyzed reporter deposition fluorescence in situ hybridization combined with microautoradiography, *Appl Environ Microbiol*, **72(11)**, 7022-7028.
- Small, J., Call, D.R., Brockman, F.J., Straub, T.M. and Chandler, D.P. (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays, *Appl Environ Microbiol*, **67**, 4708-4716.
- Smit, E., Leeftang, P., Gommans, S., van den Broek, J., van Mil, S. and Wernars, K. (2001) Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods, *Appl Environ Microbiol*, **67**, 2284-2291.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M. and Herndl, G.J. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere", *Proc Natl Acad Sci U S A*, **103**, 12115-12120.

- Song, H.G., Kim, O.S., Yoo, J.J., Jeon, S.O., Hong, S.H., Lee, D.H. and Ahn, T.S. (2004) Monitoring of soil bacterial community and some inoculated bacteria after prescribed fire in microcosm. *J Microbiol*, **42**, 285-291.
- Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J Mol Biol*, **98**, 503-517.
- Spencer, P.S., Kim, M.S. and Sabri, M.I. (2002) Aromatic as well as aliphatic hydrocarbon solvent axonopathy, *Int J Hyg Environ Health*, **205**, 131-136.
- Spiegelman, D., Whissell, G. and Greer, C.W. (2005) A survey of the methods for the characterization of microbial consortia and communities, *Can J Microbiol*, **51**, 355-386.
- Srivastava, M., Ma, L.Q. and Santos, J.A. (2006) Three new arsenic hyperaccumulating ferns, *Sci Total Environ*, **364**, 24-31.
- Stekel, D. (2003) Microarray bioinformatics. Cambridge university press.
- Story, S.P., Parker, S.H., Kline, J.D., Tzeng, T.R., Mueller, J.G. and Kline, E.L. (2000) Identification of four structural genes and two putative promoters necessary for utilization of naphthalene, phenanthrene, fluoranthene by *Sphingomonas paucimobilis* var. EPA505, *Gene*, **260**, 155-169.
- Stralis-Pavese, N., Sessitsch, A., Weilharter, A., Reichenauer, T., Riesing, J., Csontos, J., Murrell, J.C. and Bodrossy, L. (2004) Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers, *Environ Microbiol*, **6**, 347-363.
- Stroud, J.L., Paton, G.I. and Semple, K.T. (2007) Microbe-aliphatic hydrocarbon interactions in soil: implications for biodegradation and bioremediation, *J Appl Microbiol*, **102**, 1239-1253.

T

- Thiel, M., Kaschabek, S. R., Gröning, J., Mau, M. and Schlömann, M. (2005) Two unusual chlorocatechol catabolic gene clusters in *Sphingomonas* sp. TFD44, *Arch Microbiol*, **183**, 80-94.
- Thomsen, T.R., Kong, Y. and Nielsen, P.H. (2007) Ecophysiology of abundant denitrifying bacteria in activated sludge, *FEMS Microbiol Ecol*, **60(3)**, 370-382.
- Tillmann, S., Strompl, C., Timmis, K.N. and Abraham, W.R. (2005) Stable isotope probing reveals the dominant role of Burkholderia species in aerobic degradation of PCBs, *FEMS Microbiol Ecol*, **52**, 207-217.

- Torsvik, V., Salte, K., Sorheim, R. and Goksoyr, J. (1990) Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria, *Appl Environ Microbiol*, **56**, 776-781.
- Torsvik, V. and Ovreas, L. (2002) Microbial diversity and function in soil: from genes to ecosystems, *Curr Opin Microbiol*, **5**, 240-245.
- Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P. and Rubin, E.M. (2005) Comparative metagenomics of microbial communities, *Science*, **308**, 554-557.
- Tripathi, R.D., Srivastava, S., Mishra, S., Singh, N., Tuli, R., Gupta, D.K. and Maathuis, F.J. (2007) Arsenic hazards: strategies for tolerance and remediation by plants, *Trends Biotechnol*, **25**, 158-165.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. and Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment, *Nature*, **428**, 37-43.

U

- Uchiyama, T., Abe, T., Ikemura, T. and Watanabe, K. (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes, *Nat Biotechnol*, **23**, 88-93.
- Urakawa, H., Noble, P.A., El Fantroussi, S., Kelly, J.J. and Stahl, D.A. (2002) Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses, *Appl Environ Microbiol*, **68**, 235-244.
- Urum, K., Pekdemir, T. and Gopur, M. (2003) Optimum Conditions for Washing of Crude Oil-Contaminated Soil with Biosurfactant Solutions., *Process Saf Enviro*, **81**.

V

- Van Deuren, J., Lloyd, T., Chhetry, S., Liou, R. and Peck, J. (2002) Remediation Technologies Screening Matrix and Reference Guide, 4th Edition.
- Van Gijlswijk, R.P., Talman, E.G., Janssen, P.J., Snoeijers, S.S., Killian, J., Tanke, H.J. and Heetebrij, R.J. (2001) Universal Linkage System: versatile nucleic acid labeling technique, *Expert Rev Mol Diagn*, **1**, 81-91.

- Van Hamme, J.D., Singh, A. and Ward, O.P. (2003) Recent advances in petroleum microbiology, *Microbiol Mol Biol Rev*, **67**, 503-549.
- Van Nevel, L., Mertens, J., Oorts, K. and Verheyen, K. (2007) Phytoextraction of metals from soils: How far from practice?, *Environ Pollut*.
- Van Trump, J.I., Sun, Y. and Coates, J.D. (2006) Microbial interactions with humic substances, *Adv Appl Microbiol*, **60**, 55-96.
- Van Veen, J.A., van Overbeek, L.S. and van Elsas, J.D. (1997) Fate and activity of microorganisms introduced into soil, *Microbiol Mol Biol Rev*, **61**, 121-135.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H. and Smith, H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea, *Science*, **304**, 66-74.
- Volkering, F., Breure, A.M. and Rulkens, W.H. (1998) Microbiological aspects of surfactant use for biological soil remediation, *Biodegradation*, **8(6)**, 401-417.
- Von Wintzingerode, F., Gobel, U.B. and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis, *FEMS Microbiol Rev*, **21**, 213-229.

W

- Wackett, L.P. and Hershberger, D.C. (2000) Biocatalysis and biodegradation. Microbial transformation of organic compounds. ASM Press Washington, DC.
- Wagner, M., Nielsen, P.H., Loy, A., Nielsen, J.L. and Daims, H. (2006) Linking microbial community structure with function: fluorescence in situ hybridization-microautoradiography and isotope arrays, *Curr Opin Biotechnol*, **17**, 83-91.
- Wagner, M., Smidt, H., Loy, A. and Zhou, J. (2007) Unravelling microbial communities with DNA-microarrays: challenges and future directions, *Microb Ecol*, **53**, 498-506.
- Watanabe, K., Teramoto, M. and Harayama, S. (2002) Stable augmentation of activated sludge with foreign catabolic genes harboured by an indigenous dominant bacterium, *Environ Microbiol*, **4**, 577-583.
- Weinbauer, M.G. (2004) Ecology of prokaryotic viruses, *FEMS Microbiol Rev*, **28**, 127-181.
- Weinbauer, M.G. and Rassoulzadegan, F. (2004) Are viruses driving microbial diversification and diversity?, *Environ Microbiol*, **6**, 1-11.

- Whatmore, A.M., Perrett, L.L. and MacMillan, A.P. (2007) Characterisation of the genetic diversity of *Brucella* by multilocus sequencing, *BMC Microbiol*, **7**, 34.
- Whiteley, A.S., Manefield, M. and Lueders, T. (2006) Unlocking the 'microbial black box' using RNA-based stable isotope probing technologies, *Curr Opin Biotechnol*, **17**, 67-71.
- Whitman, W.B., Coleman, D.C. and Wiebe, W.J. (1998) Prokaryotes: the unseen majority, *Proc Natl Acad Sci U S A*, **95**, 6578-6583.
- Widdel, F. and Rabus, R. (2001) Anaerobic biodegradation of saturated and aromatic hydrocarbons, *Curr Opin Biotechnol*, **12**, 259-276.
- Williamson, K.E., Radosevich, M. and Wommack, K.E. (2005) Abundance and diversity of viruses in six Delaware soils, *Appl Environ Microbiol*, **71**, 3119-3125.
- Wilson, K.H., Wilson, W.J., Radosevich, J.L., DeSantis, T.Z., Viswanathan, V.S., Kuczmarski, T.A. and Andersen, G.L. (2002) High-density microarray of small-subunit ribosomal DNA probes, *Appl Environ Microbiol*, **68**, 2535-2541.
- Woese, C.R. (1987) Bacterial evolution, *Microbiol Rev*, **51**, 221-271.
- Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms, *Proc Natl Acad Sci U S A*, **74**, 5088-5090.
- Wong, M.T. and Liu, W.T. (2007) Ecophysiology of *DeFluviicoccus*-related tetrad-forming organisms in an anaerobic-aerobic activated sludge process, *Environ Microbiol*, **9(6)**, 1485-1496
- Wu, L., Thompson, D.K., Li, G., Hurt, R.A., Tiedje, J.M. and Zhou, J. (2001) Development and evaluation of functional gene arrays for detection of selected genes in the environment, *Appl Environ Microbiol*, **67**, 5780-5790.
- Wu, L., Thompson, D.K., Liu, X., Fields, M.W., Bagwell, C.E., Tiedje, J.M. and Zhou, J. (2004) Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications, *Environ Sci Technol*, **38**, 6775-6782.
- Wu, X.Y., Walker, M.J., Hornitzky, M. and Chin, J. (2006) Development of a group-specific PCR combined with ARDRA for the identification of *Bacillus* species of environmental significance, *J Microbiol Methods*, **64**, 107-119.

Y

- Yergeau, E., Kang, S., He, Z., Zhou, J. and Kowalchuk, G.A. (2007) Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect, *ISME Journal*, **1**, 163-179.
- Yilmaz, L.S., Okten, H.E. and Noguera, D.R. (2006) Making all parts of the 16S rRNA of *Escherichia coli* accessible in situ to single DNA oligonucleotides, *Appl Environ Microbiol*, **72**, 733-744.
- Yu, C.P. and Chu, K.H. (2005) A quantitative assay for linking microbial community function and structure of a naphthalene-degrading microbial consortium, *Environ Sci Technol*, **39**, 9611-9619.
- Yu, D.Y., Kang, N., Bae, W. and Banks, M.K. (2007) Characteristics in oxidative degradation by ozone for saturated hydrocarbons in soil contaminated with diesel fuel, *Chemosphere*, **66**, 799-807.
- Yuen, P.K., Li, G., Bao, Y. and Muller, U.R. (2003) Microfluidic devices for fluidic circulation and mixing improve hybridization signal intensity on DNA arrays, *Lab Chip*, **3**, 46-50.

Z

- Zhang, C. and Bennett, G.N. (2005) Biodegradation of xenobiotics by anaerobic bacteria, *Appl Microbiol Biotechnol*, **67**, 600-618.
- Zhou, J. and Thompson, D.K. (2002) Challenges in applying microarrays to environmental studies, *Curr Opin Biotechnol*, **13**, 204-207.

ANNEXE

**CARACTERISATION DES COMMUNAUTES PROCARYOTIQUES
IMPLIQUEES DANS LA BIOREMEDIATION D'UN SOL POLLUE PAR DES
HYDROCARBURES ET DEVELOPPEMENT D'OUTILS D'ANALYSE A HAUT DEBIT: LES
BIOPUCES ADN.**

Résumé :

Les activités humaines sont à l'origine de nombreuses pollutions par des hydrocarbures au niveau des écosystèmes et plus particulièrement au niveau des sols. Afin de préserver la santé humaine et environnementale, il est nécessaire de réduire les risques d'exposition et de contamination en dégradant les polluants en produit moins toxiques voire même inoffensifs. Dans ce but, les techniques de bioremédiation apparaissent aujourd'hui comme de réelles alternatives aux techniques classiques pouvant être très invasives et onéreuses. Cependant, la mise en place et/ou l'optimisation de tels procédés nécessitent une meilleure connaissance des communautés microbiennes impliquées dans la biodégradation de ces polluants. Ainsi, après avoir fait un état des lieux des outils développés pour isoler, identifier et caractériser ces populations, nous avons étudié la structure et la dynamique des communautés bactériennes au cours d'un procédé de biostimulation d'un sol pollué aux hydrocarbures aliphatiques. Outre la caractérisation d'un « core-set » bactérien présent tout au long du procédé de bioremédiation, cette étude nous a permis d'identifier des phylotypes potentiellement impliqués dans la biodégradation de ces polluants. Enfin, nous nous sommes attachés à mettre au point des outils moléculaires à haut débit, afin d'étudier plus globalement ces communautés bactériennes environnementales, et ce par le développement d'un logiciel de sélection de sondes sensibles, spécifiques et exploratoires pour biopuces ADN phylogénétiques (PhylArray).

**CHARACTERIZATION OF THE PROKARYOTIC COMMUNITIES IMPLIED IN
THE HYDROCARBON-POLLUTED SOIL BIOREMEDIATION AND
DEVELOPMENT OF HIGH-THROUGHPUT ANALYZE TOOLS :
THE DNA MICROARRAYS.**

Abstract :

Environmental ecosystems are sensitive to damage from human activities and there is an increasing need to develop better methods for removing pollutants from soils. The removal of pollutants or their transformation to less toxic products by bioremediation is a less invasive and expensive process than classical decontamination. However, the use and optimization of bioremediation treatments requires knowledge on the microbial communities directly and indirectly involved in the degradation of the pollutants. Therefore, after the creation of a synthesis describing the microbial diversity and the available tools for the isolation and the identification of microorganisms, we have studied the structure and the dynamic of the bacterial communities during the bioremediation of an aliphatic hydrocarbon-polluted soil. Beyond the characterization of a bacterial "core-set" present throughout the decontamination process, we have identified several phylotypes potentially involved in degradation. In order to globally monitor these bacterial communities, we also have developed a probe design software (PhylArray) generating efficient and explorative probes for high-throughput molecular tools: the DNA microarrays.

Discipline : Ecologie et Génomique Microbienne.

Mots-clés : Diversité bactérienne, ADNr 16S et ARNr 16S, Biopuces phylogénétique, Sélection de sondes, Sols pollués, Hydrocarbures aliphatiques, Bioremédiation.

Laboratoire : Biologie des Protistes, UMR CNRS 6023, 24 av. des Landais, 63177 Aubière.