

# TD Evolution et Biodiversité Microbienne

## Les outils moléculaires permettant l'étude de la biodiversité :

### Problème 1 : La PCR-DGGE

Des chercheurs veulent mettre au point un protocole de PCR-DGGE permettant d'évaluer la diversité des espèces appartenant au genre *Burkholderia* directement dans des échantillons environnementaux. Les bactéries du genre *Burkholderia* représentent en effet une des populations bactériennes les plus importantes des communautés bactériennes du sol. Les différentes espèces composant ce genre bactérien ont des implications très diverses dans le fonctionnement des sols. Certaines espèces ont même des effets opposés puisque quelques une sont connues pour leurs capacités phytopathogéniques alors que d'autres semblent procurer à certaines plantes une protection contre d'autres bactéries phytopathogènes. Actuellement les *Burkholderia* comprennent 21 espèces.

Pour déterminer le jeu d'amorce nécessaire à l'amplification spécifique de l'ADNr 16S du genre, un alignement de 19 séquences ADNr 16S *Burkholderia* et de 19 séquences ADNr 16S non *Burkholderia* a été réalisé.

Une partie du résultat de l'alignement est donné ci-dessous :

Species	155 ↓ Sequence	169 ↓ Sequence	Numbering Position	646 ↓ Sequence	663 ↓ Sequence
	5'CGGCGAAAGCCGGAT	3'	⇐	5'GCRRGCTAGAGTATGGCA	3'
<i>Burkholderia pseudomallei</i> AJ131790	ATAGCC	-----	TAATAC	TGACTG	-----
<i>Burkholderia thailandensis</i> U91838	-----	-----	-----	-----	-----
<i>Burkholderia glathei</i> Y17052	-----	-----	-----	-----	-----
<i>Burkholderia ambifaria</i> AY02844	-----	-----	-----	-----	-----
<i>Burkholderia cepacia</i> AF311970	-----	-----	-----	-----	-----
<i>Burkholderia caryophylli</i> AB021423	-----	-----	-----	-----	-----
<i>Burkholderia vandii</i> U96932	-----	-----	-----	-----	-----
<i>Burkholderia cepacia</i> U96927	-----	-----	-----	-----	-----
<i>Burkholderia vietnamiensis</i> U96929	-----	-----	-----	-----	-----
<i>Burkholderia graminis</i> U96941	-----	-----	-----	-----	-----
<i>Burkholderia phenazinium</i> U96936	-----	-----	-----	-----	-----
<i>Burkholderia caribiensis</i> Y17010	-----	-----	-----	-----	-----
<i>Burkholderia andropogonis</i> X67937	-----	-----NNN--	-----	-----	-----
<i>Ralstonia solanacearum</i> X67036	--A-T	A-T-----A-T-GC	-----	-----CAC-----G--T--	-----A
<i>Ralstonia eutropha</i> M32021	--A-T	A-T-----ATTAGC	-----	-----CA-----T--	-----
<i>Pandoraea norimbergensis</i> Y09879	-----T	-----	-----	-----CA-----	-----
<i>Leptothrix discophora</i> L33974	-----	-----	-----	-----CACA-----C--T--	-----
<i>Thiothrix sp</i> AB042544	-----	T--A---T--A---	-----	AT-----T--A---A-G---G-	-----A

**Question 1 :** Quelles amorces doit-on choisir pour l'amplification par PCR. ? Quel sera la longueur du produit obtenu ? est-ce compatible avec la DGGE ? Que doit-on ajouter à ces amorces pour les utiliser en DGGE ?

La spécificité des amorces déterminées ci-dessus a ensuite été testée sur les ADN génomiques des espèces figurant ci-dessus. Avec une température d'hybridation de 50°C une forte amplification a été obtenue pour les *Burkholderia*, une très faible amplification est observée pour *Pandoraea* et *Leptothrix*, et rien pour les autres.

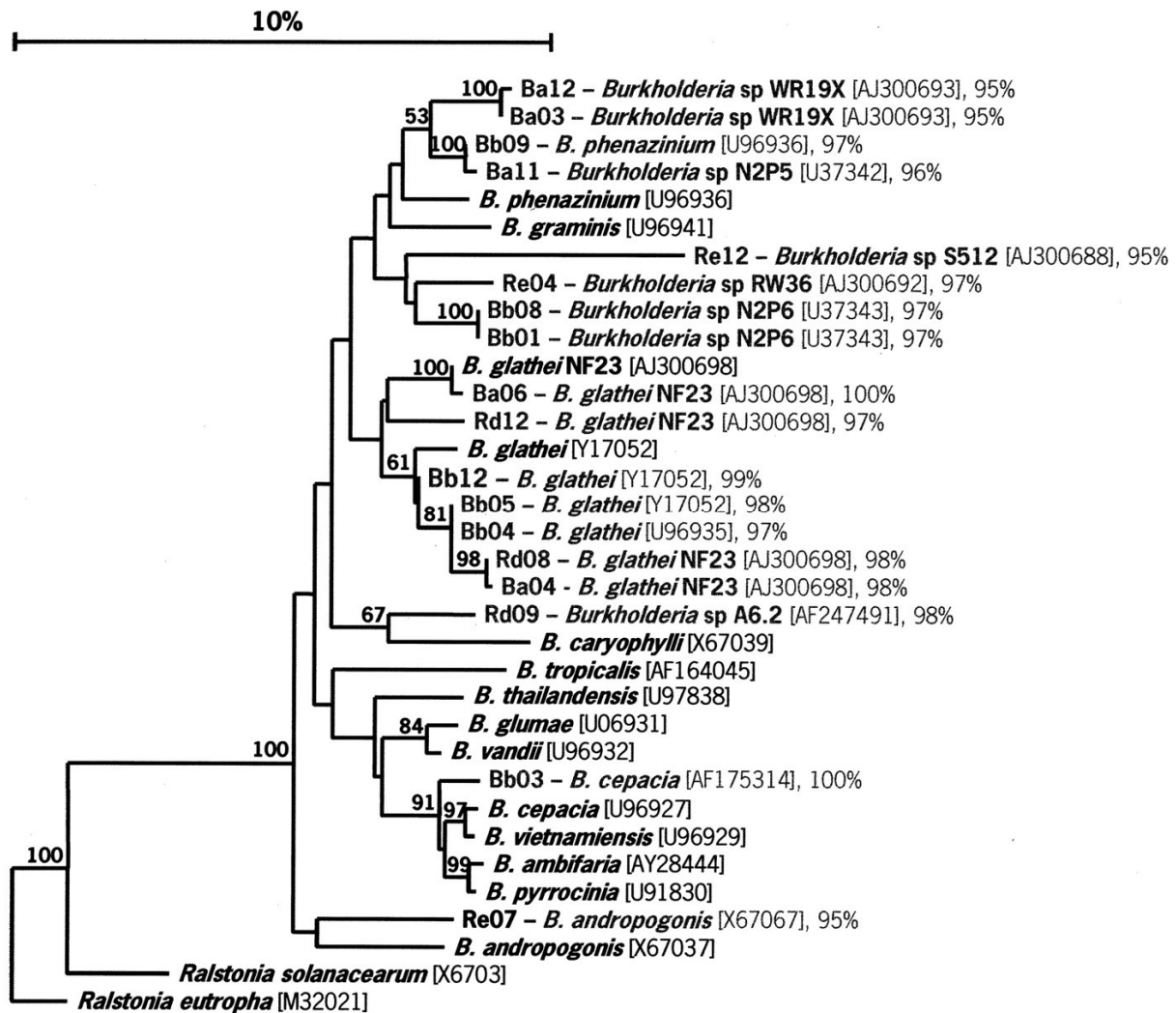
**Question 2 :** Expliquer les résultats observés ?  
En choisissant une température d'hybridation de 4 degrés plus élevée, la spécificité des amorces a été re-testée sur l'ADN génomique de 20 espèces bactériennes appartenant aux β-

Proteobacteria (dont 14 *Burkholderia*), 5 espèces appartenant aux  $\alpha$ -Proteobacteria, 13 espèces appartenant aux  $\gamma$ -Proteobacteria, 2 espèces appartenant aux Actinomycètes et 4 espèces Gram positives.

Cette fois, seuls les ADN extraits des *Burkholderia* ont été amplifiés.

Question 3 : Qu'en concluez-vous quant à la spécificité des amorces ?

Les chercheurs ont voulu confirmer la spécificité de leurs amorces en réalisant le séquençage de plusieurs produits PCR obtenus à partir d'ADN extrait du sol. Le résultat de ces séquençages est donné dans l'arbre phylogénétique ci-dessous :

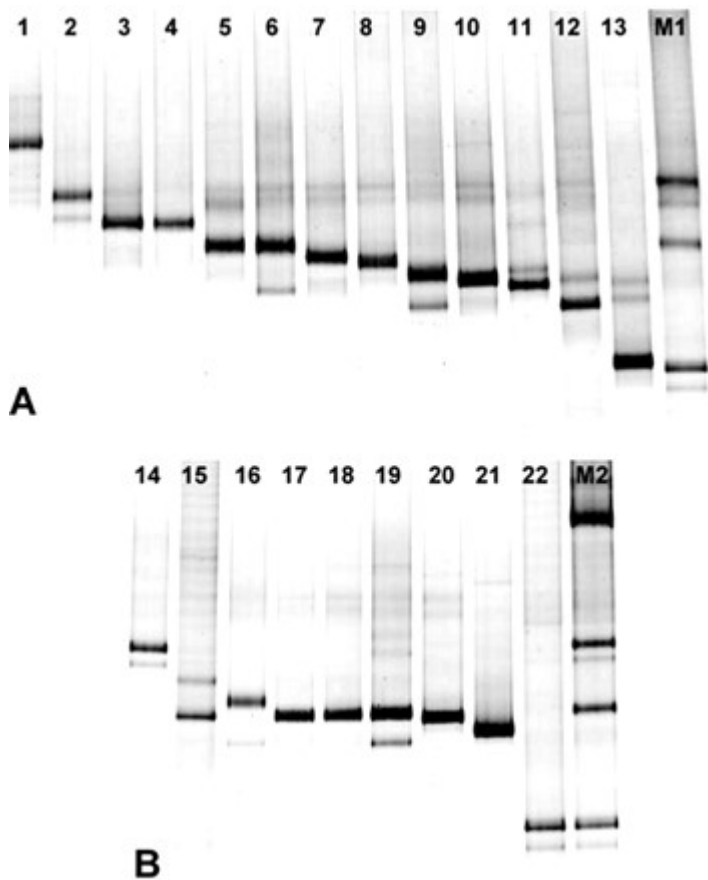


Question 4 : Interpréter les résultats obtenus ?

Pour déterminer cette fois la sensibilité des amorces choisies, des expériences d'incorporation dans le sol de différentes concentrations d'ADN de 4 espèces de *Burkholderia* ont permis de déterminer la limite de détection à  $5.10^5$  cellules/g de sol.

Question 5 : Pouvez-vous imaginer un moyen d'améliorer la limite de détection ?

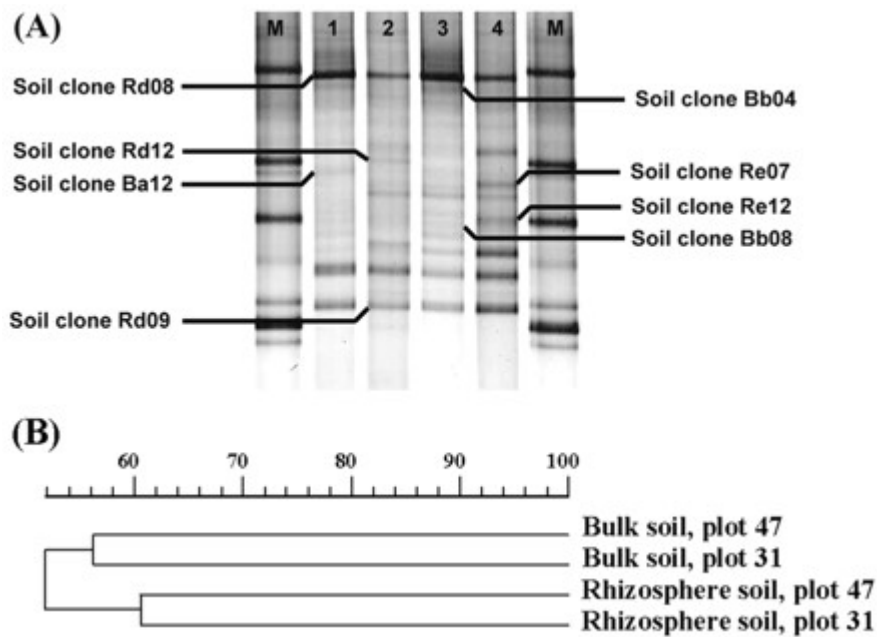
Afin de tester cette fois l'efficacité de la PCR-DGGE pour l'identification des différentes espèces, l'ADN de différentes espèces de *Burkholderia* a été utilisé. Le résultat est exposé ci-dessous :



DGGE patterns of 16S rDNA fragments of *Burkholderia* species (A) and members of the *B. cepacia* complex (B) generated by PCR with *Burkholderia*-specific primers in a 50 to 60% denaturing gradient. Lane 1, *B. glathei* WD1; lanes 2 and 14, *B. multivorans* LMG13010; lane 3, *B. plantarii* NCPPB3590; lane 4, *B. gladioli* ATCC 33664; lane 5, *B. pyrrocinia* ATCC 15958; lanes 6 and 16, *B. stabilis* LMG14294; lanes 7 and 18, *B. vietnamiensis* LMG10929; lane 8, *B. phenazinium* LMG2247; lane 9, *B. caribensis* WD3; lane 10, *B. glumae* NCPPB3708; lane 11, *B. graminis* WD2; lane 12, *B. caribensis* LMG18531; lane 13, *B. caryophylli* NCPPB353; lane 15, *B. cepacia* LMG16656 (genomovar III); lane 17, *B. cepacia* ATCC 25416 (genomovar I); lane 19, *B. cepacia* NCPPB945; lane 20, *B. cepacia* P2; lane 21, *B. cepacia* IPO1718; lane 22, *B. cepacia* LMG18941 (genomovar VI); lane M1, *Burkholderia* marker containing (from top to bottom) *B. multivorans* LMG13010, *B. cepacia* ATCC 25416, and *B. cepacia* LMG18941; lane M2, *Burkholderia* marker containing (from top to bottom) *B. andropogonis* LMG6872, *B. multivorans* LMG13010, *B. cepacia* ATCC 25416, and *B. cepacia* LMG18941.

**Question 6:** Analyser les résultats. La mobilité électrophorétique des souches de la même espèce est-elle toujours identique ? La mobilité électrophorétique des différentes espèces est-elle toujours différente ? Comment expliquer la présence de bandes minoritaires ?

La méthode décrite ici a ensuite été utilisée sur 4 échantillons environnementaux différents. 2 proviennent de la rizosphère d'une prairie (localisés à deux endroits différents) et les 2 autres d'échantillons de sols nus de la même prairie. Les résultats obtenus sont exposés dans la figure ci-dessous.



Comparison of DGGE patterns for bulk and rhizosphere soil communities in a grassland field. Samples were taken in two different plots at the same location.

(A) DGGE patterns for bulk soil from plot 47 (lane 1), rhizosphere soil from plot 47 (lane 2), bulk soil from plot 31 (lane 3), and rhizosphere soil from plot 31 (lane 4). Lane M contained *Burkholderia* markers (from top to bottom, *B. andropogonis* LMG6872, *B. multivorans* LMG13010, *B. cepacia* ATCC 25416, and *B. cepacia* LMG18941).

(B) Clustering by the unweighted pair group method with mathematical averages, showing the levels of similarity of the microbial communities

obtained by using the *Burkholderia*-specific PCR-DGGE system.

### Question 7:

La présence de plante influence-t-elle la population des *Burkholderia* ?

## Problème 2 : La PCR-SSCP

Des chercheurs se proposent d'utiliser la technique de PCR-SSCP afin de suivre la succession des communautés microbiennes qui interviennent au cours d'un processus de compostage.

Pour cela ils utilisent les différents jeux d'amorces suivants :

TABLE 1. Characterization of primers used for SSCP-based microbial-community analyses

Primer name	Primer sequence (5'-3')	Primer-target (positions)	Targeted variable region(s) of the small-subunit rRNA genes	Primer annealing temp during PCR (°C)
Com1	CAG CAG CCG CGG TAA TAC	Bacteria (519-536)	V4 and V5	50
Com2	CCG TCA ATT CCT TTG AGT TT	Bacteria (907-926)		
F243	GGA TGA GCC CGC GGC CTA	Actinomycetes (226-243)	V3	60
R531	CGG CCG CGG CTG CTG GCA CGT A	Actinomycetes (512-534)		
NS7	GAG GCA ATA ACA GGT CTG TGA TGC	Fungi (1185-1207)	V8 and V9	50
NS8	TCC GCA GGT TCA CCT ACG GA	Fungi (1508-1527)		

Voici les résultats des différents profils SSCP obtenus :

Figure 1 :

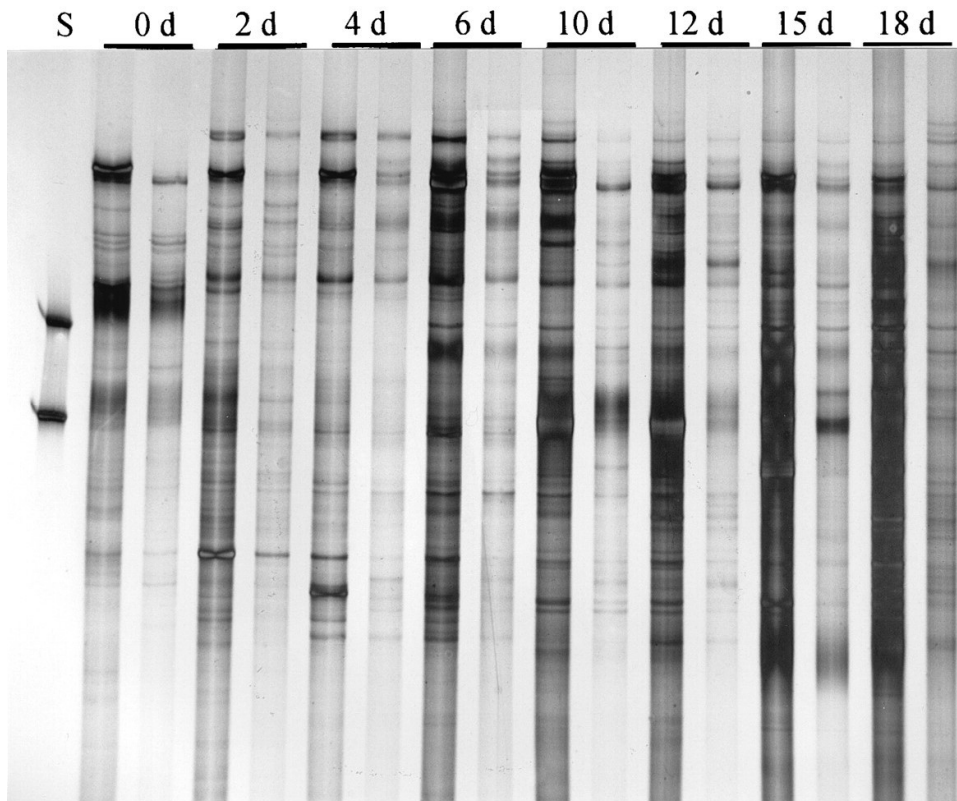


Fig 1 : Succession of PCR-amplified products during a composting process as detected by SSCP on a polyacrylamide gel. PCR primers were designed to amplify the hypervariable regions V4 and V5 of eubacterial 16S rRNA genes from directly extracted compost DNA (S, standard DNA). For each day (d), results obtained from two separate composting windows are shown.

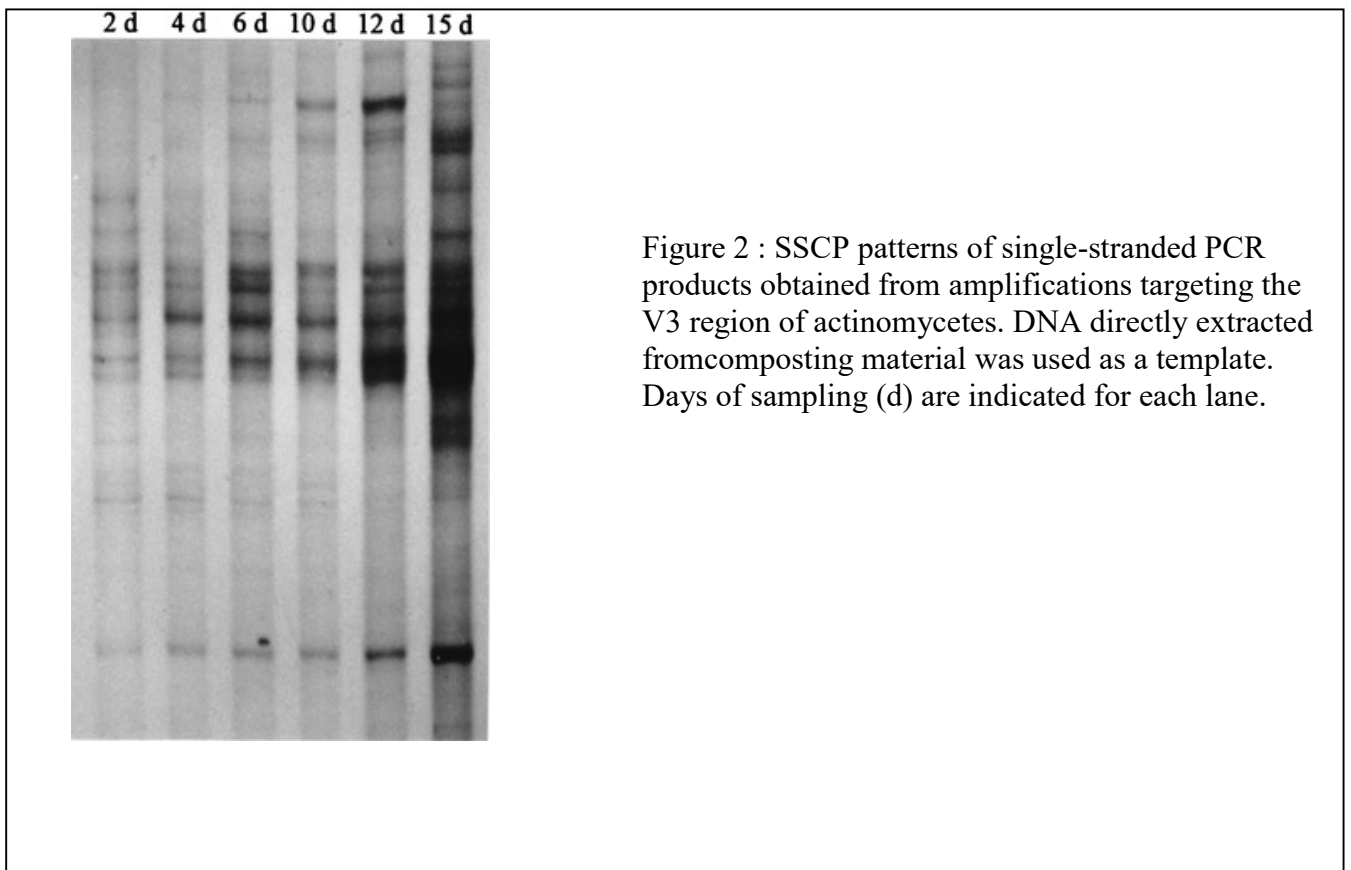


Figure 2 : SSCP patterns of single-stranded PCR products obtained from amplifications targeting the V3 region of actinomycetes. DNA directly extracted from composting material was used as a template. Days of sampling (d) are indicated for each lane.

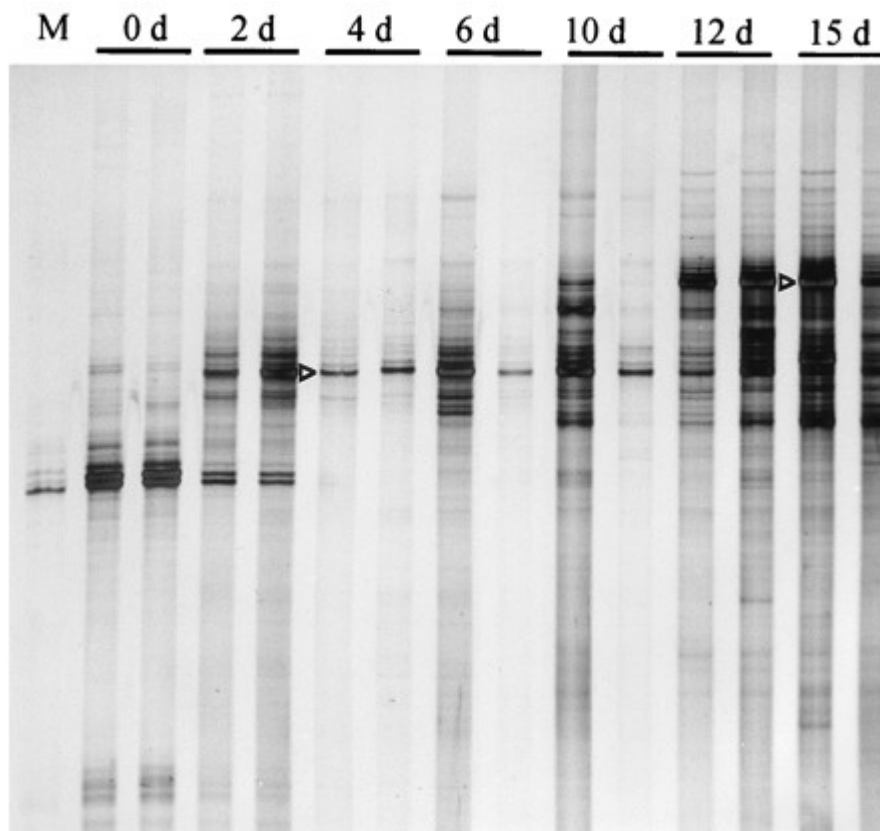


Figure 3 : Succession of SSCP patterns obtained from single-stranded DNA products amplified by PCR from directly extracted compost DNA. Primers targeted conserved regions to amplify the V8 and V9 regions of the eukaryotic 18S rRNA gene sequence (M, products obtained from DNA extracted from leaves of maize plants). Days of sampling (d) are given above lanes. For each day, parallel samples were obtained from two separate composting windows.

Question 1: Analyser chaque gel, que peut-on dire sur la variation des communautés microbiennes durant le processus de compostage ?

Les profils obtenus sont-ils toujours reproductibles ?

Certains microorganismes sont-ils régulièrement présents tout au long du processus ?

Certains microorganismes subissent-ils un phénomène d'enrichissement au terme du processus ? Comment expliquer ces variations ?

Afin de vérifier l'efficacité de la technique, certaines bandes (ADN simple brin, entourées sur le gel ci-dessous) ont été extraites des gels, réamplifiées par PCR et réanalysées sur un gel SSCP.

Les résultats sont exposés ci-dessous

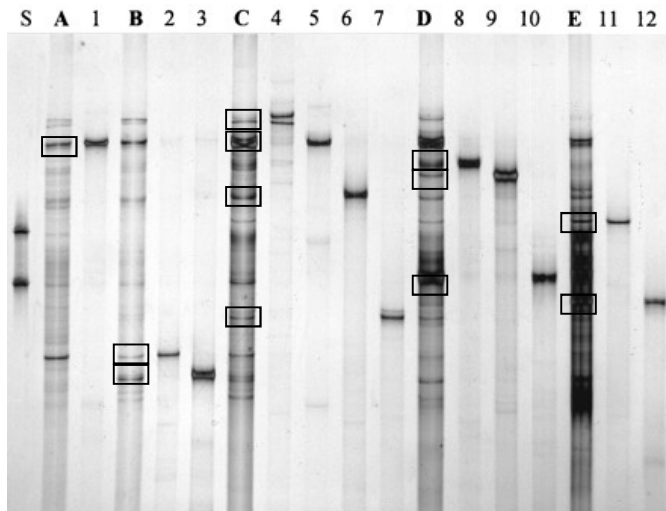
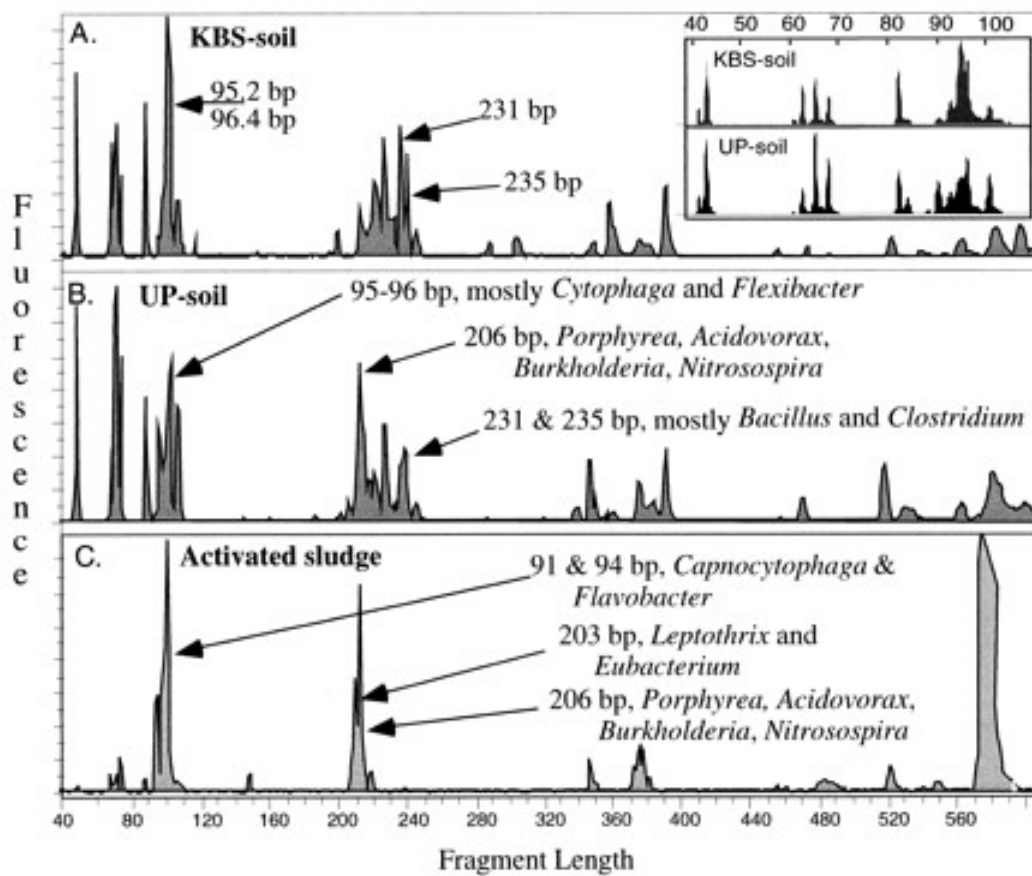


Figure 4 : Comparison of PCR-SSCP community patterns with single products isolated from profiles and reamplified by PCR.

Question 2 : Quelles conclusions peut-on tirer de ces résultats ?

### Problème 3 : La T-RFLP

Une expérience de T-RFLP a été entreprise pour caractériser la communauté microbienne de trois échantillons de sols différents :





T-RFLP profile of three bacterial communities. Terminal fragments were generated from an HhaI digestion of 16S rDNAs amplified from total community DNA using a fluorescently tagged 27F primer and an unlabeled 1392R primer. The terminal fragments were sized on an ABI 373 automated sequencer using internal size standards (ABI TAMRA 2500). (A) Agricultural soil from Kellogg's Biological Station (KBS), Michigan State University; (B) uncultivated soil from the upper peninsula (UP) of Michigan; (C) activated sludge. The insert presents an expanded view of the soil profile for fragments from 40 to 110 bp long.

**Question 1:** Quel est le sol le plus riche en termes de diversité microbienne ?

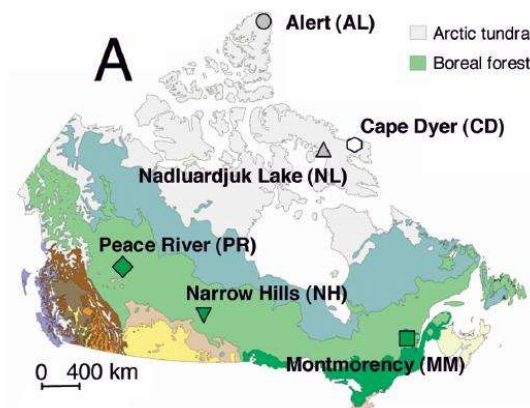
Dans chacun des trois cas étudiés, quel est la taille du fragment de restriction majoritaire ?

Comment faire à l'aide de ces résultats pour identifier l'espèce majoritaire ?

Quelle expérience pourrait-on proposer pour vérifier in situ la présence des trois espèces majoritaires ?

### **Problème 4 : la technique SARST**

Cette technique a été récemment employée pour étudier et comparer la diversité microbienne des sols de la toundra arctique et des forêts boréales du nord canadien :



A : carte localisant les lieux de prélèvements

TABLE 1. Composite soil sample characteristics and associated information

Soil sample	Site	Sample date <sup>a</sup>	Latitude	Longitude	Location	H <sub>2</sub> O (%)	Soil type	pH	C/N ratio	DNA (µg/g)	Organic (%)	RSTs sequenced
Alert	Tundra	980804	82°30'N	62°19'W	Nunavut	19.0	Silty loam	7.6	31.4	30.0	9.2	2,117
Nadluardjuk Lake	Tundra	020815	68°37'N	73°12'W	Nunavut	2.5	Sandy loam	5.3	16.0	8.5	1.4	2,562
Cape Dyer	Fuel pad	021104	66°36'N	61°34'W	Nunavut	11.0	Loamy sand	6.0	8.7	0.4	0.3	1,785
Montmorency	Balsam fir	030909	47°19'N	71°06'W	Quebec	24.6	Sandy loam	4.7	25.3	26.0	6.1	2,659
Narrow Hills	Jack pine	030624	53°54'N	104°41'W	Saskatchewan	7.9	Sand	5.3	37.1	19.2	3.2	2,240
Peace River	Mixed wood	030613	56°46'N	118°22'W	Alberta	27.2	Silty loam	4.5	24.4	19.3	6.3	1,487

<sup>a</sup> Year, month, day.

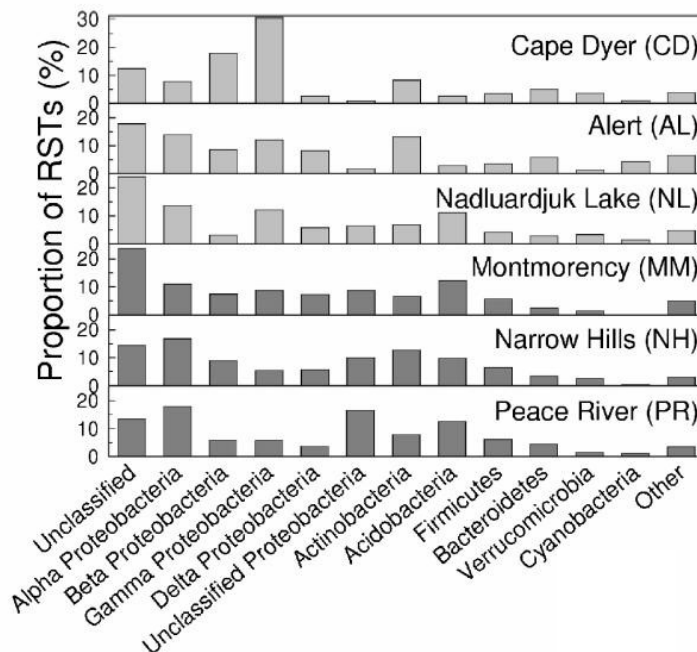
Tableau 1 : Caractéristiques des sols prélevés et nombre de RSTs obtenus.

Les auteurs de cette étude, afin de comparer la richesse de chacun de ces sols, ont déterminé le nombre de RSTs redondants au fur et à mesure du séquençage.

Analyser le contenu du tableau ci-dessous : quel est le biotope le plus riche en termes d'espèces ?

	Number of sequenced RSTs					
	250	500	1000	1500	1750	2000
Redundant RSTs at CD	150	330	750	1180	1350	ND
Redundant RSTs at AL	180	220	500	800	1000	1150
Redundant RSTs at NL	170	250	550	880	1070	1250
Redundant RSTs at PR	160	300	680	ND	ND	ND
Redundant RSTs at NH	165	280	700	1000	1200	1400
Redundant RSTs at MM	160	300	680	1080	1250	1450

La distribution par phylum a ensuite été déterminée:



Question 2 : Ce résultat est-il en accord avec le précédent ?

## Métagénomique :

### Exercice 1 (cas fictif).

Le laboratoire public UDUP géré par le ministère de l'environnement intitulé Unité de Dépollution des Usines Polluantes a pour mission principale de développer des processus de décontamination. Or l'agence nationale pour l'environnement vient d'annoncer publiquement que tous les sols occupés par les usines de l'industrie automobile française sont gravement contaminés par des composés chlorés tels que **chlorophénol** et **chloroéthènes**. Ces composés sont des cancérigènes sévères qui s'infiltrent facilement jusqu'aux nappes phréatiques. Malheureusement ils sont aussi difficilement biodégradés, seules quelques bactéries anaérobies strictes cultivables sont connues comme étant capables de les dégrader pour les rendre inoffensifs. C'est pourquoi une approche métagénomique est entreprise par ce laboratoire afin d'étudier les populations bactériennes susceptibles d'être impliquées dans la dégradation de ces composés et afin de découvrir de nouveaux gènes éventuellement utilisables pour la décontamination de ces sites. Dans ce but deux méthodes différentes sont expérimentées :

- 1) Extraction de l'ADN total d'un enrichissement (culture micro-anaérobie dans un milieu utilisé habituellement pour *E. coli* de 10 g de sol contaminé en présence de chlorophénol)
- 2) Extraction directe de l'ADN total de 10 g du même sol contaminé

Les séquences d'ARN ribosomal (ARNr) 16S sont très fréquemment utilisées pour l'identification au niveau spécifique car il s'agit de molécules ubiquitaires et possédant une structure conservée. Afin de déterminer la diversité bactérienne de ces deux échantillons, une expérience de PCR-DGGE (*denaturing gradient gel electrophoresis*) sur l'ADNr 16S est réalisée. Cette technique d'électrophorèse permet de séparer des molécules d'ADN double brin de même taille selon leur séquence sur un gel présentant un gradient dénaturant croissant. La dénaturation partielle des molécules d'ADNdb le long du gel permet de séparer les bandes. Dans le cas présenté ci-dessous, elle permet de distinguer les différents ADNr 16S présents dans un échantillon.

1 2

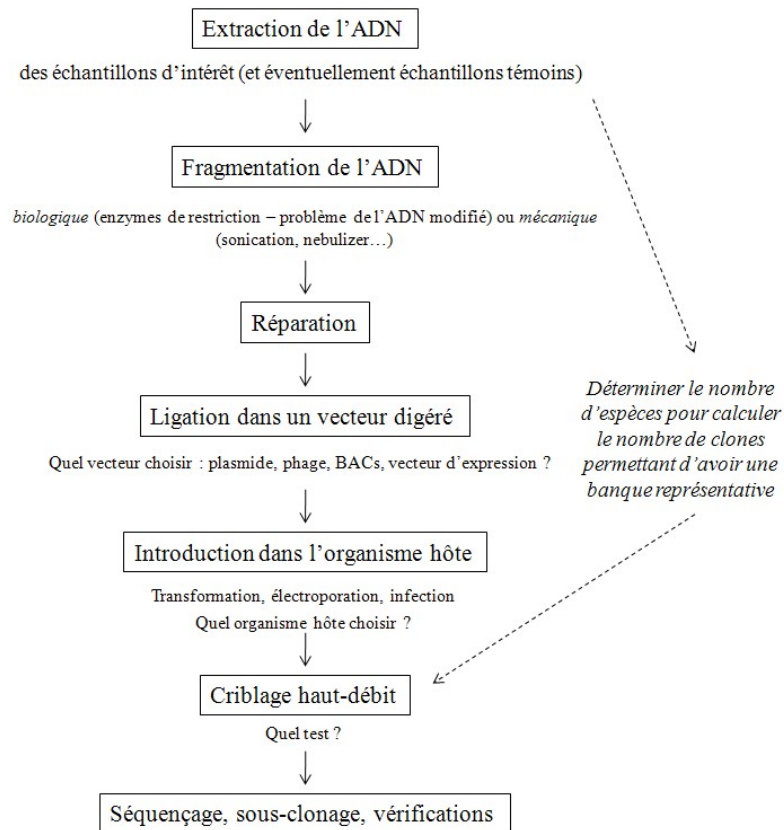


- 1 : PCR-DGGE sur l'ADN total directement extrait du sol contaminé  
2 : PCR-DGGE sur l'ADN provenant de l'enrichissement

Question 1 : Quel est le nombre minimal de génomes bactériens différents présents dans chaque échantillon ? Comment expliquer cette différence ?

En parallèle, une banque d'ADN génomique a été créée à partir de ces deux échantillons. Pour cela l'ADN a été soumis à des chocs mécaniques afin d'obtenir des fragments de taille moyenne

de 5 Kb. Après réparation des extrémités, les fragments obtenus ont été insérés dans un vecteur de clonage plasmidique et transformés dans la souche *E. coli* AE15 qui possède la propriété de facilement croître en microanaérobiose. 80 000 clones ont été obtenus pour l'ADN extrait directement du sol, 30 000 pour l'enrichissement. Un schéma présentant les principales étapes d'une démarche métagénomique visant à isoler un gène codant des propriétés d'intérêt est présenté ci-dessous :



La taille minimale requise d'une banque génomique peut être estimée par la formule suivante :

$$N_{\min} = \ln(1-p) / \ln[1 - (L-X)/(n \cdot G)]$$

où p correspond à la probabilité de trouver chaque gène au moins une fois, n correspond au nombre d'espèces, L correspond à la taille moyenne des inserts, X correspond à la taille moyenne d'un gène, et G est la taille moyenne des génomes

D'après la base de données du TIGR, G (pour les procaryotes) = 3100 Kpb et X = 0.9 Kpb.

**Question 2 :** Pourquoi les auteurs utilisent-ils une souche hôte ayant la capacité de croître en micro-anaérobiose ?

**Question 3 :** Que signifie l'étape de réparation des extrémités ?

**Question 4 :** En choisissant une probabilité = 0.9, déterminer si les banques obtenues sont représentatives ?

Après séquençage de chaque clone, le pourcentage en GC ainsi que la fréquence des tetranucléotides de chaque insert ont pu être déterminés. De cette manière, 16 « catégories d'ADN différents » ont été obtenues pour l'enrichissement alors que 55 ont été obtenues pour la banque génomique directement réalisée à partir du sol.

Question 5 : Ces données sont-elles en accord avec les résultats précédemment exposés ? Dans la négative, quelles hypothèses peut-on formuler?

Afin de découvrir de nouveaux gènes capables de dégrader le chlorophénol, les clones des deux banques génomiques ont été cultivés en micro-anaérobiose en présence de chlorophénol. La dégradation de ce composé chloré a ensuite été mesurée grâce à un *chloro-counter* (mesure des chlorures libres). La banque génomique réalisée à partir de l'enrichissement a permis de découvrir 12 nouveaux gènes différents. La banque génomique directement réalisée à partir du sol a permis de découvrir seulement 5 nouveaux gènes.

Question 6 : Comment expliquer les résultats obtenus ?

Les auteurs ont ainsi pu caractériser une nouvelle enzyme capable de dégrader le chlorophénol.

Question 7 : Comment faire pour connaître l'organisme responsable de l'activité recherchée ?

Question 8 : Quelle méthode allez-vous utiliser pour vérifier que le gène d'intérêt est exprimé *in situ* ?

Question 9 : Pouvez-vous déterminer grâce à votre stratégie la distribution géographique et biologique du gène responsable de l'activité recherchée ?

Question 10 : Donner un moyen de quantifier la présence et l'expression du gène de l'activité recherchée ?