

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉVALUATION DES CLASSIFICATIONS PHYLOGÉNÉTIQUES DES
BACILLACEAE BASÉES SUR LES GÈNES DE L'OPÉRON *rrn* ET DE
GÈNES DE MÉNAGE

MÉMOIRE
PRESENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN BIOLOGIE

PAR
MOUNA CHEIKH ROUHOU

DÉCEMBRE 2006

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens à exprimer ma gratitude à mon directeur de recherche Monsieur Jean-Charles Côté, Ph.D., directeur du laboratoire de microbiologie moléculaire et appliquée du Centre de Recherche et de Développement en Horticulture d'Agriculture et Agroalimentaire Canada à Saint-Jean-Sur-Richelieu, et professeur associé au Département des sciences biologiques de l'Université du Québec à Montréal, pour ses précieux conseils et directives, ainsi que pour la confiance qu'il m'a accordée et les moyens qu'il a mis à ma disposition.

Je tiens à exprimer ma reconnaissance et ma considération au Professeure Lucie Lamontagne, Ph.D., ma co-directrice de recherche, pour son assistance et son suivi.

Mes remerciements s'adressent également aux membres de mon comité d'orientation, les professeurs Anne Bergeron, Ph.D. et Vladimir Makarenkov, Ph.D., pour l'aide précieuse qu'ils ont apportée à la réalisation de ce travail.

Mes remerciements vont également aux membres de mon comité d'évaluation de ce mémoire de maîtrise, les professeurs Guy Charpentier, Ph.D. et Yves Hurtubise, Ph.D.

J'ai apprécié tout particulièrement l'accueil chaleureux qui m'a été réservé au Centre de Recherche et de Développement en Horticulture d'Agriculture et Agroalimentaire Canada à Saint-Jean-Sur-Richelieu, ainsi que le soutien de mes collègues de travail et amis : Hoon-Yong Lee, Brahim Soufiane, Dong Xu, Ph.D., Sabarimatou Yakoubou et Suzanne Fréchette. L'expression de ma reconnaissance s'adresse aussi aux membres du service informatique pour le support technique apporté durant ce travail.

Enfin, je ne saurais oublier le soutien et les encouragements de ma famille et de mon mari.

TABLE DES MATIÈRES

LISTE DES FIGURES	v
LISTE DES TABLEAUX	vii
LISTE DES TABLEAUX DE L'ANNEXE.....	viii
RÉSUMÉ.....	ix
CHAPITRE I	
INTRODUCTION.....	1
1.1. La bioinformatique :.....	1
1.2. Les <i>Bacillaceae</i> :.....	2
1.2.1. <i>Bacillus subtilis</i> :.....	3
1.2.2. <i>Bacillus licheniformis</i> :.....	4
1.2.3. <i>Bacillus halodurans</i> :.....	4
1.2.4. <i>Bacillus cereus</i> :.....	5
1.2.5. <i>Bacillus anthracis</i> :.....	5
1.2.6. <i>Bacillus thuringiensis</i> :.....	6
1.3. L'historique de la classification des <i>Bacillaceae</i> :.....	6
1.4. L'opéron ribosomal :.....	7
1.4.1. Le gène 16S :.....	9
1.4.2. Le gène 23S :.....	9
1.4.3. Le gène 5S :.....	10
1.4.4. Les régions inter géniques (ITS) :.....	10
1.4.5. Les ARN de transfert (ARNt) :.....	11
1.5. Les gènes domestiques ou gènes de ménage :.....	11
1.5.1. L'adénylate kinase (<i>adk</i>) :.....	12
1.5.2. La shikimate 5-déshydrogénase (<i>aroE</i>) :.....	12
1.5.3. La sous-unité B de l'ADN gyrase (<i>gyrB</i>) :.....	13

CHAPITRE II	
MATÉRIEL ET MÉTHODES	14
2.1. Les banques de données et la récupération des séquences :	14
2.2. Le format FASTA :	16
2.3. Le programme BLAST :	16
2.3.1. Le "score" :	17
2.3.2. Le "E-Value" :	18
2.4. Le programme CLUSTALW :	18
2.5. Le programme TreeView :	19
2.6. La phylogénie moléculaire :	19
2.7. Les arbres phylogénétiques :	20
2.8. Les méthodes de reconstruction des arbres phylogénétiques :	21
2.8.1. Les méthodes de distances ou méthodes phénétiques :	21
2.8.1.1. La méthode "Neighbor-Joining" (NJ) :	21
2.8.1.2. Le modèle de Kimura :	22
2.8.1.3. Les avantages des méthodes phénétiques :	24
2.8.1.4. Les inconvénients des méthodes phénétiques :	24
2.8.2. Les méthodes de caractères ou méthodes cladistiques :	24
2.8.2.1. La parcimonie :	25
2.8.2.2. La méthode de vraisemblance :	25
2.8.2.3. Les avantages des méthodes cladistiques :	25
2.8.2.4. Les inconvénients des méthodes cladistiques :	26
2.8.3. Le "Bootstrap" :	26
2.9. Enracinement des arbres phylogénétiques :	26
2.10. Récapitulatif des différentes étapes de la classification phylogénétique :	27
PROBLÉMATIQUE	29
CHAPITRE III	
CLASSIFICATION PHYLOGÉNÉTIQUE DES <i>BACILLACEAE</i> BASÉE SUR LES GÈNES 16S, 23S ET 5S DE L'OPÉRON RIBOSOMAL	31
3.1. Caractéristiques des allèles des gènes 16S, 23S, 5S et des deux régions inter géniques :	32
3.2. Analyse phylogénétique des <i>Bacillaceae</i> basée sur l'opéron ribosomal :	33
3.2.1. Classification phylogénétique basée sur les gènes 16S, 23S et 5S :	33

3.2.2.	Classification phylogénétique basée sur les régions ITS1 et ITS2 :.....	42
3.3.	Conclusion :.....	47
CHAPITRE IV		
CLASSIFICATION PHYLOGÉNÉTIQUE DES <i>BACILLACEAE</i> BASÉE SUR LES GÈNES DE MÉNAGE <i>adk</i> , <i>aroE</i> ET <i>gyrB</i>		
4.1.	Caractéristiques des gènes de ménage (<i>adk</i> , <i>aroE</i> et <i>gyrB</i>) :.....	49
4.2.	Analyse phylogénétique basée sur les séquences nucléotidiques des gènes de ménage <i>adk</i> , <i>aroE</i> et <i>gyrB</i> :.....	52
4.3.	Analyse phylogénétique basée sur les séquences en acides aminés des protéines codées par les gènes de ménage <i>adk</i> , <i>aroE</i> et <i>gyrB</i> :.....	58
4.4.	Conclusion :.....	64
CHAPITRE V		
HÉTÉROGÉNÉITÉ DES ALLÈLES DES GÈNES DE L'OPÉRON RIBOSOMAL ET SON EFFET SUR LA CLASSIFICATION PHYLOGÉNÉTIQUE.....		
5.1.	Hétérogénéité des allèles des gènes de l'opéron ribosomal (16S, 23S et 5S) :.....	67
5.2.	Effet de l'hétérogénéité des allèles sur la classification phylogénétique :.....	69
5.3.	Conclusion :.....	70
CHAPITRE VI		
DISTRIBUTION DES MUTATIONS.....		
6.1.	Distribution des mutations au sein des gènes 16S, 23S et 5S pour les <i>Bacillaceae</i> :.....	72
6.2.	Distribution des mutations au sein du gène 16S pour les γ -protéobactéries :.....	76
6.3.	Spécificités des régions hypervariables du gène 16S :.....	78
6.4.	Classification phylogénétique basée sur les régions hypervariables du gène 16S : ...	78
6.5.	Conclusion :.....	82
CHAPITRE VII		
CONCLUSION.....		
ANNEXE.....		
RÉFÉRENCES BIBLIOGRAPHIQUES.....		
		103

LISTE DES FIGURES

Figure 1.1 : Caractéristiques morphologiques des <i>Bacillus</i>	3
Figure 1.2 : Structure du ribosome.	8
Figure 1.3 : Organisation de l'opéron ribosomal.....	8
Figure 2.1 : Les différents types de substitutions	23
Figure 2.2 : Organigramme des différentes étapes de l'analyse phylogénétique des dix <i>Bacillus</i>	28
Figure 3.1 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences du gène 16S.	39
Figure 3.2 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences du gène 23S.	40
Figure 3.3 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences du gène 5S.	41
Figure 3.4 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les régions inter génique ITS1 (16S-23S).....	44
Figure 3.5 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les régions inter génique ITS2 (23S-5S).....	46
Figure 4.1 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences nucléotidiques du gène adénylate kinase (<i>adk</i>).	53
Figure 4.2 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences nucléotidiques du gène shikimate 5-déshydrogénase (<i>aroE</i>).	54
Figure 4.3 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences nucléotidiques du gène de la sous-unité B de l'ADN gyrase (<i>gyrB</i>).	55
Figure 4.4 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur la concaténation des séquences nucléotidiques des gènes adénylate kinase (<i>adk</i>), shikimate 5-déshydrogénase (<i>aroE</i>) et de la sous-unité B de l'ADN gyrase (<i>gyrB</i>).	56
Figure 4.5 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences en acides aminés de la protéine codée par le gène adénylate kinase (<i>adk</i>).	59
Figure 4.6 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences en acides aminés de la protéine codée par le gène shikimate 5-déshydrogénase (<i>aroE</i>).	60
Figure 4.7 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les séquences en acides aminés de la protéine codée par le gène de la sous-unité B de l'ADN gyrase (<i>gyrB</i>).	61
Figure 4.8 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur la concaténation des séquences en acides aminés des protéines codées par les gènes adénylate kinase (<i>adk</i>),	

shikimate 5-déshydrogénase (<i>aroE</i>) et de la sous-unité B de l'ADN gyrase (<i>gyrB</i>).....	62
Figure 6.1 : Histogramme de distribution des mutations au sein du gène 16S pour les <i>Bacillaceae</i>	73
Figure 6.2 : Histogramme de distribution des mutations au sein du gène 23S pour les <i>Bacillaceae</i>	74
Figure 6.3 : Histogramme de distribution des mutations au sein du gène 5S pour les <i>Bacillaceae</i>	75
Figure 6.4 : Histogramme de distribution des mutations au sein du gène 16S pour 28 souches appartenant à la sous division des γ -protéobactéries.....	77
Figure 6.5 : Histogramme de distribution des mutations au sein du gène 16S pour les <i>Bacillaceae</i> et les γ -protéobactéries.....	79
Figure 6.6 : Arbre phylogénétique des <i>Bacillaceae</i> basé sur les régions hypervariables du gène 16S.....	80

LISTE DES TABLEAUX

Tableau 2.1 : Liste et caractéristiques des <i>Bacillus</i> étudiées.....	15
Tableau 2.2 : Liste et définition des cinq programmes de BLAST	17
Tableau 2.3 : Comparaison entre le nombre de substitutions réelles et détectées.....	23
Tableau 3.1 : Caractéristiques des différents allèles des gènes de l'opéron ribosomal.....	34
Tableau 3.2 : Taux de similarité au sein du groupe <i>cereus</i> au sens large des gènes 16S, 23S et 5S.....	42
Tableau 4.1 : Caractéristiques des séquences nucléotidiques des différents gènes de ménage (<i>adk</i> , <i>aroE</i> et <i>gyrB</i>) des <i>Bacillus</i>	51
Tableau 4.2 : Comparaison des taux de similarité entre les séquences nucléotidiques et les séquences en acides aminés pour les gènes de ménage (<i>adk</i> , <i>aroE</i> et <i>gyrB</i>) et les protéines correspondantes.....	64
Tableau 5.1 : Taux de divergence intra-allélique au sein de chaque gène de l'opéron ribosomal	68

LISTE DES TABLEAUX DE L'ANNEXE

Tableau A.1 : Caractéristiques des <i>Bacillus</i> utilisés dans l'étude de la distribution des mutations au sein du gène 16S	87
Tableau A.2 : Caractéristiques des <i>Bacillus</i> utilisés dans l'étude de la distribution des mutations au sein du gène 23S	90
Tableau A.3 : Caractéristiques des <i>Bacillus</i> utilisé dans l'étude de la distribution des mutations au sein du gène 5S	92
Tableau A.4 : Caractéristiques des γ -protéobactéries utilisées dans l'étude de la distribution des mutations au sein du gène 16S	94
Tableau A.5 : Caractéristiques des régions hypervariables du gène 16S pour les dix <i>Bacillus</i> étudiés ainsi que pour "l'out group"	99

RÉSUMÉ

Plusieurs approches ont, jusqu'à présent, été utilisées pour l'identification de nouvelles souches ainsi que pour la caractérisation et la classification des différentes espèces au sein des *Bacillaceae*. Cependant, la plupart de ces méthodologies présentent certaines insuffisances.

L'objectif de cette étude est d'établir et d'évaluer des classifications phylogénétiques des *Bacillaceae* basées sur divers gènes, d'analyser l'effet de l'hétérogénéité allélique sur celles-ci, ainsi que d'étudier la distribution des mutations. Seuls les *Bacillaceae* dont le génome est complètement séquencé ont été considérés. Dans ce travail, tous les allèles des trois gènes de l'opéron ribosomal ont été utilisés. Trois gènes codant pour des protéines ont également été choisis pour la construction de phylogénies et la comparaison avec celles basées sur l'opéron ribosomal.

L'analyse phylogénétique de dix souches de *Bacillaceae* a été réalisée suivant la méthode de Neighbour-Joining (NJ) en utilisant les paramètres de Kimura. La fiabilité et la robustesse de la classification ont été évaluées à l'aide de la méthode "Bootstrap".

L'analyse basée sur tous les allèles des trois gènes de l'opéron ribosomal (16S, 23S et 5S) montre que les arbres obtenus présentent la même topologie et répartissent les *Bacillaceae* en trois groupes et deux sous-groupes. Aucun des gènes n'a permis la discrimination des espèces ou des souches au sein du groupe *cereus* au sens large. L'analyse phylogénétique des *Bacillaceae* pourrait donc se limiter à l'un des gènes de l'opéron ribosomal puisque aucune information additionnelle n'est apportée par l'utilisation d'un gène en particulier. Les régions inter géniques ITS1 (16S-23S) et ITS2 (23S-5S) n'ont pas permis, non plus, de discriminer au sein du groupe *cereus* au sens large. Mais elles ont, néanmoins, mis en évidence l'hétérogénéité allélique.

L'approche basée sur les séquences nucléotidiques des gènes de ménage : adénylate kinase (*adk*), shikimate 5-déshydrogénase (*aroE*) et la sous-unité B de l'ADN gyrase (*gyrB*), ainsi que de leurs séquences concaténées, a permis une meilleure discrimination notamment au sein du groupe *cereus* au sens large. Les *Bacillaceae* sont répartis en trois groupes et sept sous-groupes dont cinq au sein du groupe *cereus* au sens large. L'utilisation des séquences en acides aminés des protéines codées par les gènes *adk*, *aroE* et *gyrB* et de leurs séquences concaténées a généré, quant à elle, des classifications incongruentes. De plus, elle n'a pas

permis une meilleure discrimination des *Bacillaceae* au sein du groupe *cereus* au sens large. Plusieurs espèces se retrouvent dans un même faisceau secondaire. Un tel regroupement trouve son explication dans la dégénérescence du code génétique. L'utilisation des séquences nucléotidiques des gènes de ménage serait donc une meilleure approche pour la classification phylogénétique des *Bacillaceae*.

L'analyse de tous les allèles des différents gènes de l'opéron ribosomal (16S, 23S et 5S) montre qu'ils sont hétérogènes et que cette hétérogénéité n'a pas d'effet notable sur la classification phylogénétique. Il est donc justifié de se limiter à un seul allèle de l'un des trois gènes de l'opéron ribosomal lors de l'étude phylogénétique des *Bacillaceae*.

L'étude de la distribution des mutations au sein des trois gènes de l'opéron ribosomal (16S, 23S et 5S) pour les *Bacillaceae* montre qu'elle n'est pas uniforme et qu'il existe au sein de ces gènes des régions hypervariables dans lesquels se concentrent les mutations. La comparaison de la distribution des mutations au sein du gène 16S pour les *Bacillaceae* et les γ -protéobactéries montre que ces régions hypervariables sont spécifiques pour chaque groupe de bactéries. L'étude phylogénétique basée sur la concaténation des régions hypervariables du gène 16S pour les *Bacillaceae* a révélé que la topologie de l'arbre découlerait de la structure de ces régions. L'étude phylogénétique des *Bacillaceae* basée le gène 16S pourrait se limiter à l'analyse des régions hypervariables de ce gène.

Mots clés : adénylate kinase (*adk*), ARN 16S, ARN 23S, ARN 5S, *Bacillaceae*, classification phylogénétique, distribution des mutations, gènes de ménage, hétérogénéité alléliques, régions hypervariables, γ -protéobactéries, shikimate5-déshydrogénase (*aroE*) et sous-unité B de l'ADN gyrase (*gyrB*).

CHAPITRE I

INTRODUCTION

1.1. La bioinformatique :

Les techniques de séquençage génomique ont connu au cours des dernières décennies une évolution rapide. Face à l'énorme quantité de données de séquençage qui en a résulté et qui ne cessait d'augmenter chaque jour, le développement d'une discipline permettant de comparer et d'analyser ces données d'ordre moléculaire est devenu indispensable. C'est ainsi, au début des années 80, que la bioinformatique a fait son apparition avec la naissance des banques de séquences de biomolécules telles que EMBL (European Molecular Biology Laboratory, Heidelberg, Germany, <http://www.ebi.ac.uk>), TIGR (Institute for Genomic Research, Rockville, MD, USA, <http://www.tigr.org>) et GenBank (National Center for Biotechnology Information, Bethesda, MD, USA, <http://www.ncbi.nlm.nih.gov>).

C'est une discipline fondée sur des acquis de la biologie, de l'informatique et des mathématiques. La bioinformatique est donc une branche théorique de la biologie dite biologie "in silico". Elle propose des modèles, des logiciels et des techniques mathématiques permettant d'analyser, de gérer, d'organiser, d'explorer et de faire la synthèse des séquences nucléotidiques et protéiques, des structures tridimensionnelles des protéines et des interactions de protéines-protéines. La bioinformatique permet également d'interpréter biologiquement l'information sous forme génétique et d'énoncer des hypothèses de prédiction fonctionnelle.

1.2. Les *Bacillaceae* :

Le genre *Bacillus* appartient à la famille des *Bacillaceae* (Bradley et Galil, 1977). Les bactéries de cette famille sont de grands bacilles en forme de bâtonnets à Gram positif, groupées en chaînettes, aérobie à anaérobie facultatif, formant des spores et mobiles sauf *Bacillus anthracis* (voir figure 1.1) (Claus et Berkeley, 1986). La plupart sont des saprophytes du sol, de l'eau, de l'air et des plantes : ce sont des microorganismes ubiquitaires. Le genre *Bacillus* apparaît extrêmement hétérogène tant sur le plan génétique que phénotypique notamment au niveau respiratoire, métabolisme, ainsi que dans sa distribution dans l'environnement (Claus et Berkeley, 1986; Gordon *et al.*, 1973).

Les *Bacillus* présentent plusieurs intérêts d'ordre médical, agronomique, pharmaceutique et industriel (Carr, 1983; Gordon *et al.*, 1973; Hara et Veda, 1982; Parry *et al.*, 1983). Certaines souches des *Bacillaceae* sont non pathogènes pour l'homme, d'autres provoquent des infections et peuvent parfois causer la mort (Parry *et al.*, 1983). *Bacillus cereus*, par exemple, peut se multiplier dans les aliments et y produire une entérotoxine qui provoque des diarrhées par un mécanisme similaire à celui de l'entérotoxine d'*E. coli* (Drobniewski, 1993; Jackson *et al.*, 1995). *Bacillus anthracis* est, sans doute, le pathogène le plus virulent du genre *Bacillus* (Turnbull *et al.*, 1992). D'autres souches de ce genre ne sont pas pathogènes pour le genre humain, mais le sont pour certaines catégories d'insectes ou de parasites : elles sont utilisées par le secteur agronomique dans la lutte biologique (Krieg *et al.*, 1981).

Au sein de la famille des *Bacillaceae*, on retrouve, entre autres, *B. subtilis*, *B. licheniformis*, *B. halodurans*, *B. cereus* (Drobniewski *et al.*, 1993), *B. anthracis* (Turnbull *et al.*, 1992), *B. thuringiensis* (Turnbull, 1992), *B. mycoides* (Petersen *et al.*, 1995), *B. pseudomycoides* et *B. weihenstephanensis* (Mayer *et al.*, 1999).

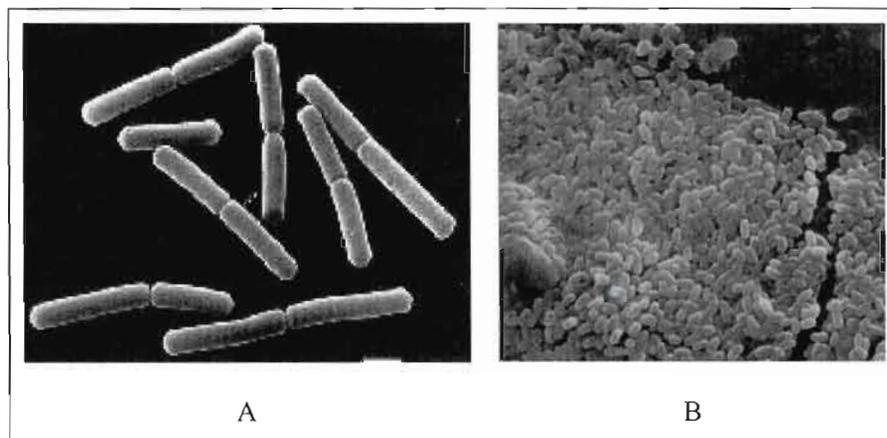


Figure 1.1 : Caractéristiques morphologiques des *Bacillus*
(A : cellules végétatives, B : spores).

1.2.1. *Bacillus subtilis* :

Bacillus subtilis et *Escherichia coli* sont parmi les espèces les mieux connues du point de vue génétique et biochimique. *B. subtilis* est une bactérie qui forme des spores capables de survivre dans des conditions extrêmes. C'est une bactérie inféodée aux plantes et retrouvée à la surface du sol. Elle n'est pas considérée comme pathogène pour l'homme, mais elle peut contaminer des aliments et peut exceptionnellement provoquer une intoxication alimentaire. (Kunst *et al.*, 1997).

B. subtilis constitue un excellent modèle pour l'étude de nombreuses bactéries pathogènes telles que *Bacillus anthracis* responsable de la maladie du charbon, *Listeria monocytogenes* causant de graves infections alimentaires et *Streptococcus pneumoniae* à l'origine d'infections nosocomiales sévères. Plusieurs secteurs industriels s'intéressent à *Bacillus subtilis*. Cette bactérie est une source d'enzymes industriels tel que les amylases utilisés dans l'industrie du pain, les protéases et les cellulases employés dans les industries agroalimentaires et des détergents. L'industrie pharmaceutique exploite la capacité de *Bacillus subtilis* à produire des antibiotiques, comme la bacitracine, pour fabriquer des médicaments pour les maux de gorge (Kunst *et al.*, 1997).

1.2.2. *Bacillus licheniformis* :

B. licheniformis se retrouve principalement dans le sol et les plantes (Sneath *et al.*, 1986). Cette espèce peut causer occasionnellement des toxi-infections alimentaires et des infections oculaires. Certaines souches de *B. licheniformis* possèdent un pouvoir dénitrifiant (élimination des nitrates). Cette propriété, très intéressante, est utilisée dans la dépollution de l'environnement (Alexander, 1977).

B. licheniformis présente un intérêt pour le secteur industriel où elle est utilisée pour la production à grande échelle de coenzymes (Schallmeyer *et al.*, 2004) et particulièrement de protéases pour les industries du textile et des détergents (Erickson, 1976). Le domaine agricole utilise cette bactérie dans la lutte contre certains microorganismes fongiques pathogènes attaquant les récoltes de maïs. (Neyra *et al.*, 1996). Finalement, *B. licheniformis* suscite l'intérêt du monde médical grâce à sa capacité de produire de la proticine et de l'acide poly- γ -glutamique (Ghera *et al.*, 1989)

1.2.3. *Bacillus halodurans* :

Une bactérie alcaline, isolée en 1975, a été identifiée comme appartenant au genre *Bacillus*. Ce n'est que récemment qu'elle a été nommée *B. halodurans*. Jusqu'à présent, aucune étude n'a montré que cette espèce est pathogène pour le monde animal ou végétal.

Bacillus halodurans se démarque par sa production de β -galactosidase (Ikura *et al.*, 1979) et du xylanase (Honda *et al.*, 1985a; Honda *et al.*, 1985b). De plus, elle produit beaucoup d'enzymes alcalins utiles pour plusieurs industries tel que la protéase, la cellulase et l'amylase (enzymes dégradant respectivement les protéines, la cellulose et l'amidon).

1.2.4. *Bacillus cereus* :

B. cereus est une bactérie qui appartient au groupe *cereus* au sens large (sensu lato). Les membres de ce groupe présentent à la fois de grandes similarités génétiques et une grande diversité de virulence et de pathogénicité (Helgason *et al.*, 2000; Patra *et al.*, 2002). *B. cereus* est un micro-organisme qui est généralement retrouvé dans la nourriture telle que le lait, les céréales et le riz (Drobniewski *et al.*, 1993; Jackson *et al.*, 1995). De plus en plus d'études montrent que cette bactérie produit une toxine émétique et des entérotoxines qui causent plusieurs types d'infections chez l'homme dont la pneumonie et la méningite (Drobniewski, 1993; Logan et Turnbull, 1999).

1.2.5. *Bacillus anthracis* :

B. anthracis est un autre membre du groupe *cereus* au sens large. Elle a été identifiée comme non hémolytique, non mobile, sporulante, aérobie, sensible à la pénicilline et encapsulée. Elle est l'agent étiologique de la maladie du charbon ou anthrax. Cette maladie est une zoonose (maladie animale contagieuse) touchant principalement les mammifères (y compris l'homme).

Certaines souches de *B. anthracis* (Ames et Vollum) sont très virulentes et sont considérées comme des armes biologiques (Jernigan *et al.*, 2001; Jernigan *et al.*, 2002; Mock & Fouet, 2001). Les spores de *B. anthracis* sont hautement résistantes à la sécheresse, la chaleur, les rayons ultra-violets, les rayons gamma et à de nombreuses substances désinfectantes (Turnbull, 2002). La virulence de *B. anthracis* provient de la synthèse de toxines composées de trois protéines distinctes et de la formation d'une capsule qui lui permet d'échapper à la phagocytose.

1.2.6. *Bacillus thuringiensis* :

Bt est l'abréviation de *B. thuringiensis*. Cette bactérie fait partie du groupe *cereus* au sens large (Joung et Côté, 2001). *Bacillus thuringiensis* a été isolé en 1901 à partir du ver à soie. On le retrouve dans l'air, l'eau, le feuillage des végétaux et dans pratiquement tous les sols.

Il se distingue des autres *Bacillus* du groupe *cereus* par sa capacité à synthétiser des cristaux formés de multiples protéines. Ces dernières présentent certaines propriétés insecticides pour certaines lépidoptères, coléoptères et diptères (Lecadet *et al.*, 1999). Ces protéines agissent en détruisant les cellules de l'intestin moyen de la larve de l'insecte (Garcia - Martinez *et al.*, 1999; Joung et Côté, 2000).

1.3. L'historique de la classification des *Bacillaceae* :

L'intérêt des secteurs médical, industriel et agronomique face aux *Bacillaceae* a vite révélé un énorme besoin de classification pour distinguer et identifier les différentes espèces. Dans un premier temps, les *Bacillaceae* ont été classifiés en utilisant des approches phénétiques telles que les caractères morphologiques, physiologiques (composition des cellules), biochimiques (test API) et nutritionnels (Berkeley *et al.*, 1984; Gordon *et al.*, 1973; Logan et Berkeley 1984; Smith *et al.*, 1952; Stackebrandt *et al.*, 1987). Puis, dans un deuxième temps, d'autres approches moléculaires basées sur l'hybridation ADN-ADN (Priest *et al.*, 1981) et la composition d'ADN (Fahmy *et al.*, 1985) ont aussi été utilisées.

Cependant, ces classifications basées sur les approches phénétiques ou moléculaires se sont avérées lentes, coûteuses, peu fiables et peu concluantes en raison de la très grande hétérogénéité des espèces à l'intérieur du genre, notamment aux niveaux physiologique et écologique (Claus & Berkeley, 1986; Priest *et al.*, 1988; Turnball et Kramer, 1991).

Pour surmonter ces difficultés, l'intérêt s'est alors orienté vers la classification des *Bacillaceae* en se basant sur les séquences du gène 16S (Ash *et al.*, 1991b; Rössler *et al.*, 1991). On a construit des arbres phylogénétiques des *Bacillaceae* à partir de la séquence d'une seule copie du gène 16S par espèce. Néanmoins, cette approche, n'a pas permis de révéler les hétérogénéités au niveau intra-spécifique. De plus, de récentes études ont montré que les différents allèles intra-spécifiques du gène 16S peuvent présenter un taux élevé de divergence (Clayton *et al.*, 1995). C'est le cas, par exemple, de *Thermoanaerobacter tengcongensis*, où ce taux peut atteindre 11.6% (Acinas *et al.*, 2004).

Il apparaît dès lors important de s'interroger sur la robustesse de la classification actuelle des *Bacillaceae* basée sur la séquence d'un seul allèle du gène 16S. Ne pas prendre en considération la possible hétérogénéité des différents allèles du gène 16S pourrait mener à la construction d'arbres phylogénétiques non fiables (Crosby et Criddle, 2003); différents allèles pouvant générer différents arbres.

Ceci nous amène à se poser la question suivante : quelle allèle faut-il choisir et combien d'allèles sont nécessaires pour élucider, s'il y a lieu, l'hétérogénéité intra-spécifique au sein des *Bacillaceae*? Le présent travail portera sur toutes les allèles de chaque gène formant l'opéron ribosomal : 16S, 23S et 5S. Ce qui permettrait d'apporter des éléments de réponses à cette question.

1.4. L'opéron ribosomal :

Les ribosomes sont des organites cellulaires résultant d'un assemblage complexe de diverses protéines ribosomiques avec des ARN ribosomiaux (ARNr).

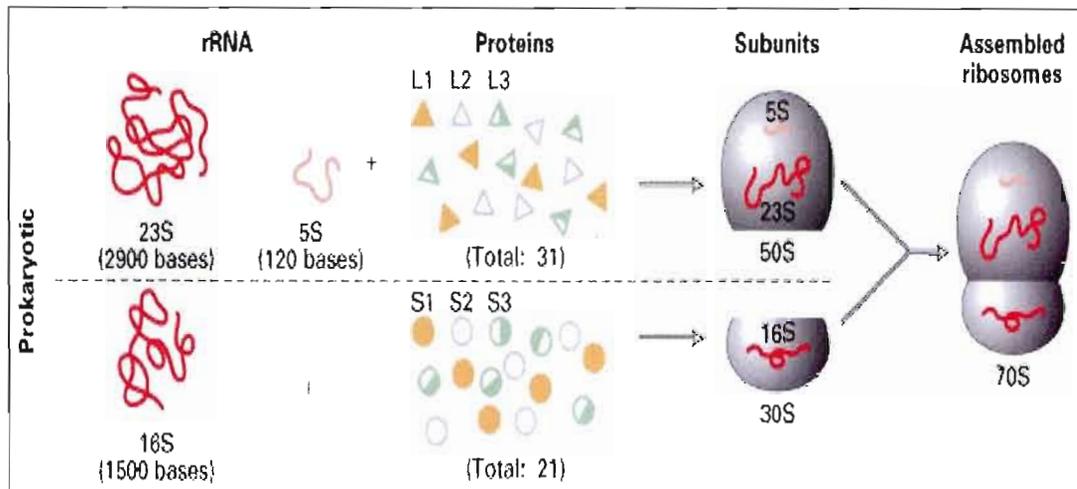


Figure 1.2 : Structure du ribosome.

(http://phs.prs.k12.nj.us/Science_Dept/APBio/Ribosome.html)

Le ribosome comprend deux sous-unités de tailles différentes (voir figure 1.2) (Mears *et al.*, 2002) :

- La grande sous-unité 50S constituée des deux types de molécules : l'ARN 23S composé de 2900 nucléotides et l'ARN 5S composé de 120 nucléotides.
- La petite sous-unité 30S constituée par la molécule d'ARN 16S composée, en moyenne, de 1500 nucléotides.

Généralement, chez les bactéries, l'opéron ribosomal est organisé dans l'ordre suivant (Bricker, 2000, Ueda *et al.*, 1999) : le gène 16S (*rrs*, ARN 16S), une région inter génique (ITS) contenant parfois des gènes d'ARN de transfert (ARNt), le gène 23S (*rrl*, ARN 23S), une seconde région inter génique (ITS) et le gène 5S (*rrf*, ARN 5S) (voir figure 1.3).

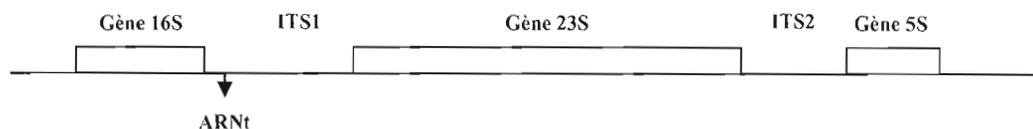


Figure 1.3 : Organisation de l'opéron ribosomal.

Chez les bactéries, on retrouve plusieurs copies de l'opéron ribosomal dont le nombre est proportionnel à la taille et à la vitesse de croissance de la bactérie. Chez la famille des *Bacillaceae*, ce nombre varie entre 8 et 13 copies (Johansen *et al.*, 1996; Lechner *et al.*, 1998; Priest *et al.*, 1994). L'opéron ribosomal est un bon marqueur moléculaire, fréquemment utilisé pour le traçage des rapports génétiques et même l'identification des nouvelles souches (Amann *et al.*, 1995).

1.4.1. Le gène 16S :

Pour l'étude des rapports phylogénétiques et la comparaison des diverses souches bactériennes, on s'est basé généralement sur la comparaison des séquences du gène d'ARNr 16S. Ce dernier est très abondant dans les cellules bactériennes ce qui facilite son extraction et sa purification (Amann *et al.*, 1995; Stahl *et al.*, 1991; Woese, 1987). De plus, il évolue lentement ce qui permet de comparer des espèces assez éloignées.

Cependant, ce critère est insuffisant pour certaines espèces (Bavykin *et al.*, 2004). En effet, la classification des espèces bactériennes étroitement proches peut être très difficile puisque les gènes ARNr 16S sont hautement conservés. Par exemple, au sein de la famille des *Bacillaceae*, ils présentent une similitude de l'ordre de 99% (Ash *et al.*, 1991a; Ash *et al.*, 1991b). D'autant plus que le gène ARNr 16S peut subir un transfert ou une recombinaison latérale occasionnelle (Rintaro *et al.*, 2000). Pour remédier à ces insuffisances, cette étude phylogénétique prendra en considération tous les gènes de l'opéron ribosomal.

1.4.2. Le gène 23S :

Le gène 23S est composé de régions présentant une évolution rapide qui renferment un taux de mutations très élevé. Il est utilisé, lui aussi, dans la distinction et la différenciation des espèces

notamment les plus proches pour lesquelles l'étude du gène 16S ne permet pas d'en déduire des conclusions quant à leur classification phylogénétique.

1.4.3. Le gène 5S :

L'ARN 5S est un composant intégral de la grande sous-unité du ribosome. Sa petite taille et son association avec les protéines ribosomales font de lui une molécule modèle pour les études des interactions ARN-protéines.

1.4.4. Les régions inter géniques (ITS) :

Les ITS sont des régions qui séparent les différents gènes de l'opéron ribosomal : 16S-23S et 23S-5S. Ce sont les régions les plus variables indépendamment des substitutions, des suppressions et des insertions qui peuvent exister au sein de l'opéron ribosomal. La taille des ITS est, elle aussi, très variable notamment d'une espèce à une autre. Chez les bactéries il existe plusieurs types d'ITS qui diffèrent selon la présence ou non de gènes d'ARNt et selon l'identité de l'ARNt encodé (Casanova *et al.*, 2001; Condon *et al.*, 1992; Maeda *et al.*, 2000; Vold, 1985).

Les ITS (16S-23S) représentent un bon marqueur moléculaire pour analyser les rapports phylogénétiques et la diversité des procaryotes contrairement aux ITS (23S-5S) qui sont beaucoup moins utilisées par manque de données sur les séquences du gène 5S (Garcia - Martinez *et al.*, 1999; Gürtler *et al.*, 1999; Jensen *et al.*, 1993).

1.4.5. Les ARN de transfert (ARNt) :

Ces molécules qui font partie de l'opéron ribosomal, sont formées d'environ 80 nucléotides et participent à la traduction de l'information entre les nucléotides et les acides aminés.

1.5. Les gènes domestiques ou gènes de ménage :

Les gènes domestiques, ou gènes de ménage ("house-keeping genes"), assurent des fonctions indispensables à la vie de tous les types de cellules. Ils existent en un seul exemplaire chez pratiquement tous les microorganismes et présentent une grande diversité intra-spécifique. Cette caractéristique a fait que ces gènes sont considérés comme un outil de choix dans les expériences de typage moléculaire (Multi-Locus Sequence Typing : MLST) (Maeda *et al.*, 2000). Ils permettent d'obtenir un résultat rapide, précis et fiable dans l'identification des souches très proches. Ceci présente une grande importance dans le domaine thérapeutique (identification des souches pathogène appartenant à la même espèce) (Maeda *et al.*, 2000).

Les gènes de ménage sont généralement de petite taille quoique celle-ci est très variable, notamment d'une espèce à une autre. Par exemple, pour le *pdhC* (codant pour la pyruvate déshydrogénase), alors que la taille de celui-ci est de 1239 paire de base (pb) au sein de *Staphylococcus aureus* (Forsyth *et al.*, 2002) et de 1329 pb chez *Bacillus subtilis* (Kunst *et al.*, 1997), elle n'est que de 480 pb chez *Neisseria meningitidis* (Maeda *et al.*, 2000).

Dans la présente étude on s'intéresse aux trois gènes suivants : adénylate kinase (*adk*), shikimate 5-déshydrogénase (*aroE*) et la sous-unité B de l'ADN gyrase (*gyrB*).

1.5.1. L'adénylate kinase (*adk*) :

Le gène *adk* code pour l'adénylate kinase (appelé aussi myokinase). Cette protéine est présente dans toutes les cellules des différents organismes vivants (procaryotes et eucaryotes). Elle catalyse réversiblement le transfert du troisième phosphate et de sa liaison, riche en énergie, d'un nucléoside triphosphate vers un nucléoside mono phosphate en présence de Mg^{2+} ($ATP + (d) NDP \rightleftharpoons ADP + (d) NTP$). Plusieurs études suggèrent que cet enzyme pourrait jouer un rôle essentiel dans la croissance de certaines bactéries (Munoz-Dorado *et al.*, 1990). Il a été souvent employé comme marqueur pour la détection et l'énumération rapide des bactéries Gram-négatives.

La structure d'*adk* est fortement conservée d'une espèce à une autre. Par exemple, son taux de similarité entre l'être humain (eucaryote) et la bactérie *E. coli* (procaryote) est de 42% (De La Rosa *et al.*, 1995; Hama *et al.*, 1991; Ray et Mathews 1992). Cette forte homologie entre les différents organismes fait d'*adk* une molécule modèle pour l'étude phylogénétique d'espèces très éloignées.

1.5.2. La shikimate 5-déshydrogénase (*aroE*) :

La shikimate 5-déshydrogénase est un gène codant pour un enzyme jouant un rôle principal dans la voie de biosynthèse d'acides aminés aromatiques. En effet, cet enzyme catalyse la quatrième étape de la voie biosynthétique de shikimate. En plus il est responsable de la réduction du 3-dehydroshikimate en shikimate en présence de NADPH (Anton et Coggins, 1988; Padyana et Burley, 2003). La présence de shikimate 5-déshydrogénase a été confirmée au sein de la plupart des génomes bactériens complètement séquencés. Cet enzyme est fortement conservé chez tous les organismes.

1.5.3. La sous-unité B de l'ADN gyrase (*gyrB*) :

Le *gyrB* est un gène qui code pour la protéine de la sous-unité B de l'ADN gyrase (ou la topoisomérase II). C'est un enzyme qui est responsable de la présence des supercoils négatifs dans les chromosomes bactériens et qui joue un rôle crucial dans leur réplication (Watt et Higgins, 1994). La sous-unité B de l'ADN gyrase est essentielle pour la survie des cellules bactériennes. Le transfert horizontal de ce gène est rare ce qui explique son utilisation dans la classification taxonomique des bactéries (Fukushima *et al.*, 2002; Sawada *et al.*, 1999; Watanabe *et al.*, 2001; Yanez *et al.*, 2003).

CHAPITRE II

MATÉRIEL ET MÉTHODES

2.1. Les banques de données et la récupération des séquences :

L'évolution rapide qu'ont connu les techniques de séquençage a été à l'origine de la disponibilité d'importantes quantités de données moléculaires. La gestion d'un tel volume de données rendait indispensable une approche structurée et efficace. C'est ainsi, qu'au début des années 80, les premières banques de données génomiques ont fait leur apparition. Ces banques de données moléculaires ont rapidement connu une expansion fulgurante si bien qu'elles sont devenues indispensables à la communauté scientifique.

Parmi ces banques de données on retrouve :

- "EMBL" (European Molecular Biology Laboratory) : Elle a été créée en 1980 et était financée par l'EMBO (European Molecular Biology Organisation). Elle est aujourd'hui diffusée par l'EBI (European Bioinformatics Institute, Cambridge, UK).
- "DDBJ" (DNA Data Bank of Japan) : Elle a été créée en 1986 et est diffusée par le NIG (National Institute of Genetics, Japon).
- "GenBank" : Elle a été créée en 1982 par la société IG (IntelliGenetics) et est diffusée maintenant par le NCBI (National Center for Biotechnology Information, Los Alamos, USA).

La principale mission de ces banques de données est de rendre publique les données de séquençage soumises par les laboratoires à travers le monde. En 1990, une convention a été conclue entre trois banques de données ("EMBL", "Genbank" et "DDBJ") pour permettre l'échange interbancaire des séquences. Aujourd'hui, les résultats d'une requête devraient être identiques quelle que soit la banque utilisée.

Les séquences de l'opéron ribosomal ainsi que des gènes de ménage des *Bacillus*, qui font l'objet de cette étude, ont été téléchargées à partir de "GenBank" (National Center for Biotechnology Information, Bethesda, MD, USA) (Benson *et al.*, 1997). Cette base de données est accessible sur Internet à l'adresse suivante (<http://www.ncbi.nlm.nih.gov>). Le tableau 2.1 donne la liste et les caractéristiques des *Bacillaceae* de cette étude.

Tableau 2.1 : Liste et caractéristiques des *Bacillus* étudiées

Nom de la bactérie	Numéro d'accèsion		Date	Taille du génome	N°ARN _r
<i>Bacillus anthracis</i> str. Ames	AE016879	NC_003997	2 Mai 2003	5227293 pb	11
<i>Bacillus anthracis</i> str. Ames 0581 ou Ancestor	AE017334	NC_007530		5227419 pb	11
<i>Bacillus anthracis</i> str. Sterne	AE017225	NC_005945		5228663 pb	11
<i>Bacillus cereus</i> ATCC 10987	AE017194	NC_003909	13 Fév 2004	5224283 pb	12
<i>Bacillus cereus</i> ATCC 14579	AE016877	NC_004722	2 Mai 2003	5411809 pb	13
<i>Bacillus cereus</i> EL33	CP000001	NC_006274		5300915 pb	13
<i>Bacillus halodurans</i> C-125	BA000004	NC_002570	23 Mar 1999	4202352 pb	8
<i>Bacillus licheniformis</i> ATCC14580	CP000002	NC_006270	6 Oct 2004	4222336 pb	7
<i>Bacillus subtilis</i> str. 168	AL009126	NC_000964	31 Déc 1997	4214630 pb	10
<i>Bacillus thuringiensis</i> serovar <i>konkukian</i> str. 97-27	AE017355	NC_005957		5237682 pb	14

Légende :

- NC : Le numéro d'accèsion des ARNr dans "GenBank".
- AE : Le numéro d'accèsion du génome complet dans "GenBank".
- Date : La date de la mise en place du génome entièrement séquencé dans "GenBank".
- N° ARN : Le nombre de copies de l'opéron ribosomal.

2.2. Le format FASTA :

Pour l'analyse des séquences nucléotidiques ou d'acides aminés par les différents programmes d'alignement, celles-ci doivent être écrites selon le format FASTA. Une séquence au format FASTA commence par une ligne de titre (nom, définition, ...), suivie par la séquence répartie sur plusieurs lignes. La ligne de titre est précédée par le symbole ">" et il est recommandé de limiter la longueur des lignes de la séquence à 80 caractères. Ci-dessous un exemple d'une séquence écrite suivant le format FASTA :

```
>rrnAB14579
TTATTGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACA
TGCAGTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGG
```

2.3. Le programme BLAST :

BLAST (Basic Local Alignment Search Tool) est un algorithme utilisé par une famille de cinq programmes qui permettent l'alignement d'une nouvelle séquence par rapport à une base de données. Ceci consiste à faire la comparaison de cette séquence requête avec les génomes existants dans les banques de données pour y déceler les séquences homologues. Des tests statistiques permettent de décider si l'alignement obtenu est significatif et les résultats fournis sont classés par ordre de fiabilité.

Le programme BLAST utilise l'algorithme développé par Altschul *et al.*, (1990) pour rechercher les segments similaires entre une séquence requête (ou séquence "query") et l'ensemble des séquences présentes dans la banque nucléique ou protéique. Les séquences sont classées en fonction d'un "score" qui dépend de l'homologie avec la séquence requête, de la taille de la banque et de la valeur du "E-Value". Plus celui-ci est petit, plus l'homologie entre la séquence requête et celle de la banque est grande.

BLAST regroupe plusieurs sous-programmes qui sont utilisés selon la nature de la séquence requête et de la banque (voir tableau 2.2).

2.3.1. Le "score" :

Le "score" global est un paramètre calculé par BLAST qui permet de quantifier la similitude entre la séquence requête et celle de la banque de données. Il résulte de la somme des "scores" élémentaires calculés pour chacune des positions des deux séquences dans leur appariement optimal. Il s'agit donc du nombre total de bons appariements duquel est retranché le nombre de mésappariements. Un mésappariement peut être soit :

- La substitution d'un caractère par un autre (une mutation).
- L'introduction d'un "gap", permettant d'optimiser l'alignement entre les deux séquences. Le "gap" matérialise une insertion (ou délétion).

Tableau 2.2 : Liste et définition des cinq programmes de BLAST

Sous-programme	Type de la séquence requête ("query")	Type de banque	Définition
BlastP	Protéique	Protéique	Compare la séquence protéique requête avec toutes celles présentes dans la banque de données protéiques
BlastN	Nucléique	Nucléique	Compare la séquence nucléique requête avec toutes celles présentes dans la banque de données nucléiques
TblastN	Protéique	Nucléique	Compare la séquence protéique requête traduite dans les 6 phases de lecture avec la banque de données
BlastX	Nucléique	Protéique	Compare la séquence nucléique requête traduite dans les 6 phases de lecture avec la banque de données
TblastX	Nucléique	Nucléique	La séquence requête et les séquences de la banque sont traduites dans les 6 cadres de lecture avant la recherche de segments homologues.

La valeur du score élémentaire est de 1 lorsqu'il s'agit d'un bon appariement et 0 pour un mésappariement.

2.3.2. Le "E-Value" :

Pour évaluer la fiabilité d'un alignement ayant un "score" "s", le programme BLAST effectue, dans d'autres banques de données, une recherche des alignements qui ont un score supérieur à celui obtenu. Le "E-Value" est le nombre total des alignements retrouvés dans les autres banques. Plus la valeur de ce paramètre est petite plus l'alignement est fiable.

2.4. Le programme CLUSTALW :

CLUSTALW est un programme d'alignements multiples des séquences d'acides nucléiques ou protéiques. Ces séquences sont fournies par l'utilisateur ou proviennent des banques de données. Ce type d'alignement est très utile puisqu'il permet de comparer des séquences provenant de différentes sources et d'étudier les différences et les similitudes entre les séquences de même type mais appartenant à des groupes différents. Ceci permet d'identifier les régions qui sont conservées d'une espèce à une autre (Thompson *et al.*, 1994).

Les principales étapes d'un alignement multiple avec CLUSTALW sont :

- Chaque paire de séquences est alignée séparément selon une méthode approximative globale et rapide (Wilbur et Lipman, 1983). Une matrice de distances traduisant la divergence de chaque paire de séquences est ainsi calculée.
- Un dendrogramme "guide tree" est ensuite construit à partir de la matrice des distances en utilisant la méthode de Neighbor-Joining (Higgins et Sharp, 1988). Cet arbre phylogénétique sert de guide qui détermine l'ordre selon lequel l'alignement de l'étape suivante sera réalisé.

- L'alignement de la paire de séquence qui présente le plus de similarités conduit à une séquence consensus avec laquelle la séquence suivante sera alignée. Le processus continue ainsi, selon l'ordre établi dans le dendrogramme jusqu'à l'alignement multiple final.

2.5. Le programme TreeView :

TreeView (Page, 1996) est un programme utilisé pour visualiser les arbres phylogénétiques à partir des fichiers de résultats donnés par CLUSTALW.

2.6. La phylogénie moléculaire :

C'est dans les années 1960, avec l'apparition des premières séquences de protéines homologues spécifiques à chaque espèce, qu'est venue l'idée d'utiliser les séquences des gènes pour retracer l'histoire évolutive des organismes vivants (Zuckerandl et Pauling, 1965). Cette approche a complètement révolutionné la systématique des procaryotes (Doolittle et Brown, 1994; Olsen *et al.*, 1994; Woese, 1994). En effet, la comparaison entre les morphologies, les comportements et les répartitions géographiques étaient les seuls moyens disponibles pour la classification des espèces. Toutefois, cette approche a rapidement montré ses limites dès qu'il s'agissait de comparer des organismes très divergents, notamment les microorganismes.

Au sein des bactéries, l'utilisation des répartitions géographiques et des caractéristiques morphologiques et physiologiques tels que la forme de la cellule ou de la colonie, la mobilité et la présence d'activités enzymatiques ou métaboliques ne permet de dégager que des conclusions assez limitées. Le plus souvent, ces caractères présentent trop de similarités ou

de divergences ou sont difficiles à interpréter pour constituer des marqueurs évolutifs convenables.

Face à ces difficultés, l'utilisation des caractéristiques moléculaires s'est imposée comme une alternative intéressante qui permet une étude cohérente de la phylogénie au sein des procaryotes. Cependant, l'accumulation des mutations au cours du temps peut affecter la fiabilité de cette approche. Cette insuffisance a été adressée par Zuckerland et Pauling (1965) qui ont développé la théorie de l'horloge moléculaire.

La phylogénie moléculaire consiste donc à reconstruire les relations de parenté entre les séquences de nucléotides ou d'acides aminés. Elle permet d'étudier l'évolution des séquences et la relation de parenté entre espèces. La nature quantitative des relations, ainsi établies, a permis le développement de méthodes et de règles rigoureuses pour la construction des arbres phylogénétiques.

2.7. Les arbres phylogénétiques :

Un arbre phylogénétique est une représentation graphique des relations de parenté entre des espèces, supposés ayant un ancêtre commun. Chaque nœud de l'arbre représente l'ancêtre commun de ses descendants. En général, un arbre phylogénétique renferme :

- Des feuilles qui représentent les "espèces" étudiées ou les taxons.
- Des nœuds internes qui représentent les ancêtres virtuels ayant divergé.
- Des arrêtes qui illustrent les liens de filiation.

Il existe plusieurs techniques de construction des arbres phylogénétiques. Ceux-ci peuvent être enracinés ou non.

2.8. Les méthodes de reconstruction des arbres phylogénétiques :

Il existe deux grands types de méthodes de reconstruction des arbres phylogénétiques moléculaires :

- Les méthodes de distances : Elles se basent sur la mesure des distances entre les séquences, c'est à dire le nombre de substitutions de nucléotides ou d'acides aminés entre chaque paire de séquences.
- Les méthodes de caractères : Il existe deux méthodes dans cette catégorie : la parcimonie et la vraisemblance. Elles se basent toutes les deux sur le nombre de mutations (substitutions, insertions, délétions) qui affectent chacune des séquences étudiées.

2.8.1. Les méthodes de distances ou méthodes phénétiques :

Elles se proposent de reconstruire des arbres sans racine en se basant sur les ressemblances observées entre chaque paire d'unités évolutives. Pour ces méthodes, plus la ressemblance entre deux unités est importante, plus leurs liens de parenté sont étroits. Parmi ces méthodes, on distingue la méthode Neighbor-Joining ; celle-ci est la seule qui sera utilisée dans cette étude. En effet, la conservation très élevée du gène 16S au sein de la famille des *Bacillaceae* ne permet pas l'utilisation d'autres méthodes de reconstruction phylogénétique.

2.8.1.1. La méthode "Neighbor-Joining" (NJ) :

C'est la méthode la plus utilisée, elle a été développée par Saitou et Nei (1987). Elle tente de remédier à l'un des défauts de la méthode "Unweight Pair Group Method with Arithmetic mean" (UPGMA) en permettant un taux de substitution variable dans les différentes branches. La méthode "Neighbor-Joining" (NJ) consiste à construire des arbres

phylogénétiques sans racine à partir d'un indice d'écart (par exemple distance ou dissimilarités entre les séquences). Elle est basée sur la recherche d'une paire de "Operational Taxonomic Units" (OTU) qui minimise la longueur totale des branches de l'arbre et ceci à chaque étape de regroupement. Lorsque deux séquences sont liées, le nœud représentant l'ancêtre commun est ajouté à l'arbre et les deux feuilles sont enlevées. Ce processus convertit l'ancêtre en un nœud terminal dans un arbre de taille réduite. Cependant, plus la divergence entre deux séquences est grande, plus la probabilité qu'une mutation intermédiaire ait eu lieu est importante. Par conséquent, il est probable que le simple comptage des différences entre deux séquences n'est pas le reflet exact de la réalité mais sous-estime le nombre d'évènements mutationnels.

Pour remédier à cette déficience, des hypothèses ont été émises sur la façon dont la substitution des bases ou acides aminés a lieu à un locus donné. Les premiers à avoir proposé une solution à ce problème sont Jukes et Cantor (1969). Il s'agit d'un modèle très simple, ne comportant qu'un seul paramètre : le taux de substitution d'une base d'ADN vers une autre. Ce modèle suppose la même probabilité de substitution quelle que soit la nature de la mutation. La fréquence de substitution utilisée par ce modèle est de 25% pour les quatre bases de l'ADN. Cependant, les données moléculaires ne sont pas cohérentes avec l'hypothèse d'une probabilité uniforme pour tous les types de substitutions. Ainsi on a assisté à la naissance du modèle de Kimura qui a essayé de remédier à l'insuffisance du modèle proposé par Jukes et Cantor.

2.8.1.2. Le modèle de Kimura :

Des études antérieures ont confirmé que la probabilité des transitions est plus élevée que celle des transversions. En effet, une transition correspond au remplacement d'une purine (A ou G) ou d'une pyrimidine (C ou T) par une autre. Alors que dans une transversion, une purine est substituée par une pyrimidine ou inversement (voir figure 2.1).

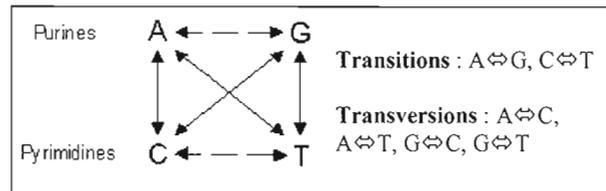


Figure 2.1: Les différents types de substitutions.

Kimura (1980) a proposé un modèle à deux paramètres où la probabilité des transitions est différente de celle des transversions, ce qui permet d'estimer les distances évolutives en termes de nombre de substitutions de nucléotides. Le modèle de Kimura fut rapidement adopté par la communauté scientifique. La complexité additionnelle apportée par l'ajout d'un paramètre supplémentaire est largement compensée par une nette amélioration de la cohérence entre les modèles de substitutions et les données réelles.

Le tableau 2.3 montre l'ordre et les conséquences des substitutions ainsi qu'une comparaison entre les substitutions réelles et celles qui peuvent être détectées.

Tableau 2.3 : Comparaison entre le nombre de substitutions réelles et détectées

	Séquence 1	Séquence 2	Nombre de substitutions observées	Nombre réel de substitutions
Substitution unique	C	C → A	1	1
Substitutions multiples	A	A → C → T	1	2
Substitutions coïncidentes au même site	C → A	C → G	1	2
Substitutions parallèles	T → A	T → A	0	1
Substitutions convergentes	C → T → A	C → A	0	3
Substitutions inverses	C → T → C	C	0	1

2.8.1.3. Les avantages des méthodes phénétiques :

Ces méthodes sont les seules capables d'analyser certains types de données tel que les distances immunologiques et d'hybridation d'acides nucléiques. Elles sont rapides et permettent d'analyser de grandes bases de données et de tester un grand nombre d'hypothèses alternatives. Elles permettent aussi d'intégrer des modèles de changements évolutifs (modèle de Kimura) qui ne sont pas intégrables dans d'autres méthodes.

2.8.1.4. Les inconvénients des méthodes phénétiques :

Les méthodes de distances réduisent la matrice de caractères à une matrice de distance ce qui induit la perte d'une certaine quantité d'information. De plus, ces méthodes ne permettent pas de combiner dans une même matrice des caractères de natures différentes tel que des caractères morphologiques et des séquences d'ADN.

2.8.2. Les méthodes de caractères ou méthodes cladistiques :

Le principe de ces méthodes a été élaboré par l'entomologiste Willi Hennig (1940). Ce sont des méthodes de classification dans lesquelles les espèces appartenant au même taxon partagent un ancêtre commun plus récent que celui qu'elles ont en commun avec d'autres espèces. Dans de telles classifications, tous les taxons doivent être monophylétiques. Ces méthodes sont lentes et imprécises.

2.8.2.1. La parcimonie :

La méthode de parcimonie est la plus simple sur le plan théorique (Fitch, 1971). Elle consiste à minimiser le nombre de mutations (ou de substitutions) nécessaires pour passer d'une séquence à une autre dans une topologie de l'arbre. Elle recherche parmi tous les arbres possibles celui qui nécessite le moins de substitutions d'acides nucléiques ou d'acides aminés pour expliquer les différences entre les séquences. Cette méthode s'appuie sur les hypothèses suivantes :

- Les sites évoluent indépendamment les uns des autres.
- La vitesse d'évolution est lente et constante au cours du temps.

2.8.2.2. La méthode de vraisemblance :

Elle a été introduite en phylogénie moléculaire par Felsenstein (1981). Cette méthode de reconstruction phylogénétique évalue, en terme de probabilités, l'ordre des branchements et la longueur des branches d'un arbre selon un modèle évolutif donné. Elle s'appuie sur les taux de substitutions, par analogie aux matrices de substitutions utilisées dans la construction des alignements multiples.

2.8.2.3. Les avantages des méthodes cladistiques :

Contrairement aux méthodes des distances, les séquences sont traitées sous forme de caractères, ce qui permet d'éviter toute perte d'information. Les méthodes cladistiques essayent d'identifier les séquences ancestrales en évaluant différents arbres.

2.8.2.4. Les inconvénients des méthodes cladistiques :

Les méthodes cladistiques sont très lentes, ne donnent aucune information sur la longueur des branches, n'utilisent pas toute l'information disponible et ne font pas de corrections pour les substitutions multiples.

2.8.3. Le "Bootstrap" :

Le "Bootstrap" (Efron *et al.*, 1993) est une méthode souvent utilisée pour tester la fiabilité des branches internes des arbres phylogénétiques. Elle consiste à effectuer un tirage au hasard avec remise des sites. Dans chaque réplique de "Bootstrap", certains sites peuvent être présents plusieurs fois, tandis que d'autres peuvent être absents.

Chaque réplique produit un nouveau alignement "artificiel" qui est utilisé pour construire un arbre "artificiel". Pour chaque branche interne on calcule le pourcentage des arbres "artificiels" contenant cette branche. On considère généralement que les branches définies par une valeur de "Bootstrap" supérieure à 95% sont fiables. Il faut 1000 répliques de "Bootstrap" pour que cette méthode soit statistiquement valable.

2.9. Enracinement des arbres phylogénétiques :

L'utilisation de la méthode "Bootstrap" permet d'aboutir à un arbre consensus construit à partir des noeuds les plus fréquemment rencontrés dans l'ensemble des arbres possibles. Toutefois, s'il est possible sans trop de difficultés de calculer la similarité entre toutes les paires de séquences d'un groupe et de positionner les points de branchements relatifs, il est beaucoup plus difficile de déterminer la séquence du groupe qui pourrait être l'ancêtre

commun. L'approche la plus souvent adoptée est d'enraciner l'arbre phylogénétique. Cette technique consiste à ajouter une séquence connue comme étant beaucoup plus ancienne que toutes les autres. Cependant, il ne faut pas que la séquence choisie soit ni trop éloignée ni trop proche des autres. L'utilisation de cette séquence "out group" améliore la fiabilité de la classification.

2.10. Récapitulatif des différentes étapes de la classification phylogénétique :

Les différentes étapes de la classification phylogénétiques des dix *Bacillaceae* sont illustrées dans l'organigramme de la figure 2.2.

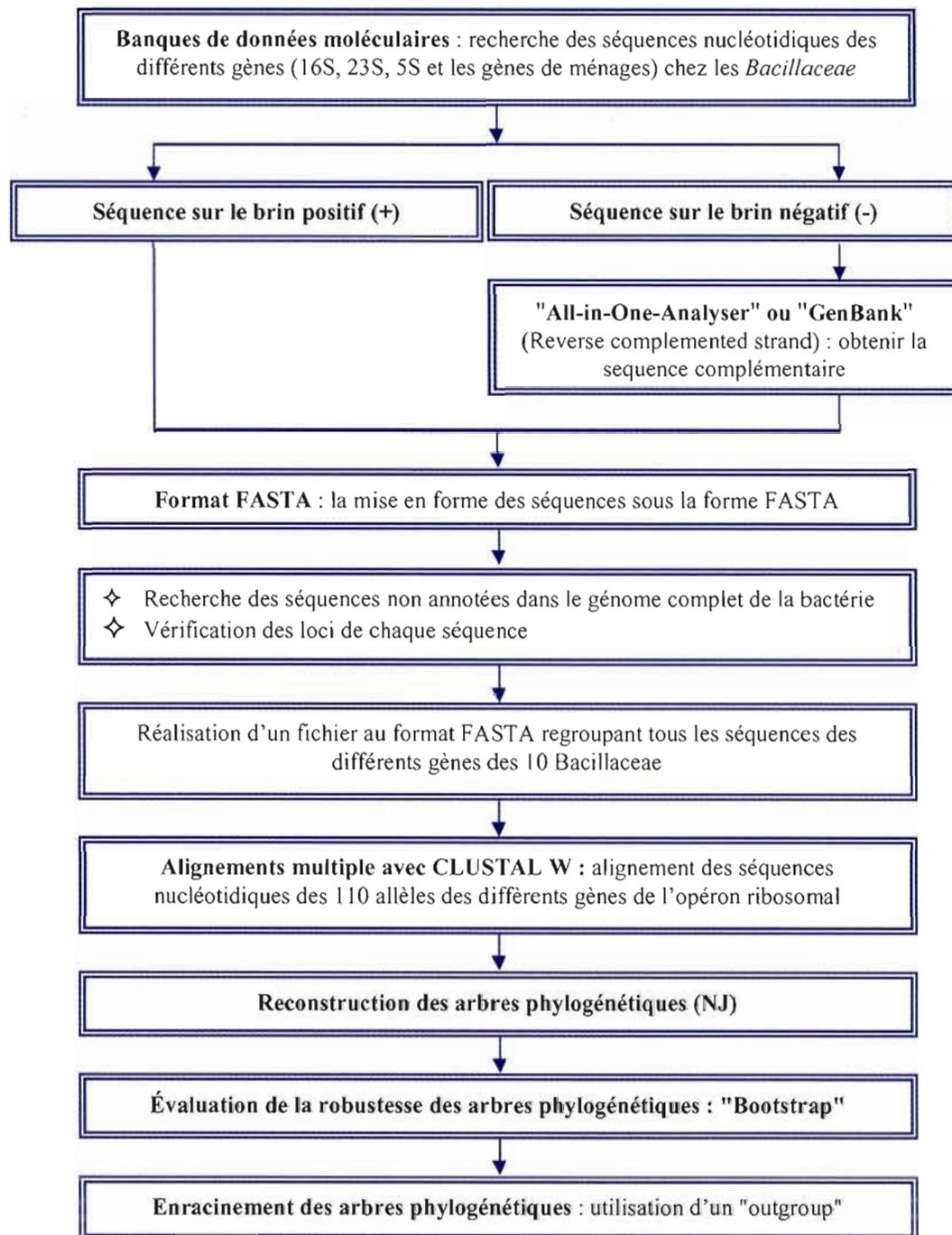


Figure 2.2: Organigramme des différentes étapes de l'analyse phylogénétique des dix *Bacillus*.

PROBLÉMATIQUE

La famille des *Bacillaceae* est très répandue dans la nature. C'est une famille très hétérogène, aussi bien sur le plan phénétique que génétique. L'intérêt que présente les *Bacillaceae* pour les secteurs industriel, médical et agricole a suscité plusieurs travaux d'identification et de classification.

Les premières classifications phylogénétiques au sein de cette famille étaient basées sur les propriétés morphologiques et physiologiques. Puis, l'approche basée sur l'analyse d'une seule copie du gène ARN 16S a fait son apparition. Elle demeure jusqu'à présent l'une des approches d'identification et de classification les plus utilisées. Cependant, certaines insuffisances inhérentes à cette méthode nécessitent d'être soulevées (Stackbrandt et Goebel, 1994). En effet, le gène 16S est fortement conservé au sein de la famille des *Bacillaceae* (Gürtler et Stanisich, 1996, Kolbert et Persing, 1999). De plus, plusieurs études ont mis évidence l'hétérogénéité des allèles, ce qui rend l'utilisation d'une seule copie du gène 16S discutable.

Cette étude vise à déterminer si l'utilisation d'un seul allèle du gène 16S est justifiée, ainsi que d'explorer si les autres gènes (ARN 23S et ARN 5S) de l'opéron ribosomal ou certains gènes de ménage offrent une meilleure alternative pour la classification phylogénétique des *Bacillaceae*. La distribution des mutations, la recherche de régions hypervariables au sein de l'opéron ribosomal et éventuellement leur utilisation pour établir une classification phylogénétique des *Bacillaceae* figurent aussi parmi les objectifs fixés pour cette étude.

Dans le chapitre III des classifications phylogénétiques des *Bacillaceae* basées sur tous les allèles de tous les gènes de l'opéron ribosomal (16S, 23S et 5S) ainsi que sur les deux régions

inter géniques (16S-23S) et (23S-5S) sont établies. Ces phylogénies sont ensuite comparées entre elles pour déterminer si l'utilisation d'un seul gène de l'opéron ribosomal est suffisante et justifiée.

L'approche d'une classification phylogénétique basée sur les séquences nucléotidiques des trois gènes de ménage suivants : adénylate kinase (*adk*), shikimate 5-déshydrogénase (*aroE*) et la sous-unité B de l'ADN gyrase (*gyrB*) ainsi que sur les séquences en acides aminés des protéines codées par ces trois gènes est explorée au chapitre IV.

Le chapitre V analyse l'hétérogénéité des allèles des trois gènes (16S, 23S et 5S) de l'opéron ribosomal et évalue son effet sur la classification phylogénétique des *Bacillaceae*.

Le dernier chapitre est consacré à l'étude de la distribution des mutations au sein des gènes de l'opéron ribosomal (16S, 23S, 5S) pour les dix souches de *Bacillaceae* complètement séquencées ainsi que dix autres dont le génome ne l'est que partiellement. L'étude de la distribution des mutations au sein du gène 16S est élargie pour inclure 28 souches de γ -protéobactéries. Finalement, une recherche de régions hypervariables est effectuée et l'efficacité d'une classification phylogénétique basée sur ces zones est évaluée.

CHAPITRE III

CLASSIFICATION PHYLOGÉNÉTIQUE DES *BACILLACEAE* BASÉE SUR LES GÈNES 16S, 23S ET 5S DE L'OPÉRON RIBOSOMAL

Au cours de la dernière décennie, plusieurs études ont établi des classifications phylogénétiques des *Bacillaceae* en se basant uniquement sur le gène 16S et en utilisant une seule copie de celui-ci. L'hypothèse d'homogénéité des allèles a souvent été invoquée dans ces études pour justifier cette approche. Cependant, récemment, les travaux de Clayton *et al.*, (1995) et Acinas *et al.*, (2004) ont démontré que les différents allèles sont hétérogènes et que les taux de divergence peuvent atteindre dans certains cas des valeurs significatives. Ceci met en doute la robustesse des classifications ainsi établies. En effet, différents allèles génèrent différents arbres, donc différentes classifications.

Dans cette étude on se propose d'établir une classification phylogénétique consensus des *Bacillaceae* en utilisant tous les allèles des trois gènes 16S, 23S et 5S ainsi que des régions inter géniques ITS1 et ITS2 de l'opéron ribosomal. Cette approche permettrait de valider si l'utilisation d'une seule partie de l'opéron ribosomal (par exemple le gène 16S) est suffisante pour établir une classification phylogénétique fiable. La nécessité d'utiliser tous les allèles des gènes de l'opéron ribosomal sera discutée au chapitre V qui étudie l'effet de l'hétérogénéité des allèles sur la classification des *Bacillaceae*.

3.1. Caractéristiques des allèles des gènes 16S, 23S, 5S et des deux régions inter géniques :

Le groupe de bactéries, faisant l'objet de cette étude, est formé d'un *B. thuringiensis*, trois *B. cereus*, trois *B. anthracis*, un *B. halodurans*, un *B. licheniformis* et un *B. subtilis*. La bactérie *Clostridium tetani* E88 a été choisie comme "out group" pour enraciner les arbres phylogénétiques. La liste complète des dix souches de *Bacillus* qui ont été analysées est donnée dans le tableau 3.1. Il s'agit de bactéries qui sont déjà complètement séquencées. Les séquences de celles-ci ont été récupérées à partir de "GenBank" (NCBI, <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>, 17 janvier 2006).

Les numéros d'accèsion des toutes les bactéries ainsi que les localisations des gènes 16S, 23S et 5S, tel qu'annotés dans "GenBank", figurent également au tableau 3.1. En absence d'annotation, le programme Blast a été utilisé pour identifier les différents allèles orthologues. Le programme Blast a également été l'outil employé, dans certains cas, pour confirmer les annotations et vérifier l'existence de tous les allèles. Lorsque requis, un caractère (A, B, C, ...) a été ajouté au nom du gène donnée dans "GenBank" pour permettre de distinguer les différents allèles.

Un premier alignement local, suivi d'un deuxième alignement global, ont montré que les annotations de certaines gènes, tel que fournis par "GenBank", nécessitent une correction touchant uniquement le début et la fin des séquences.

La taille des gènes 16S, 23S et 5S est de 1554, 2923 et 116 nucléotides respectivement. Celle des régions inter géniques 16S-23S (ITS1) et 23S-5S (ITS2) est assez variable. En effet, les tailles des régions ITS1 et ITS2 varient de 166 et 50 respectivement chez *B. subtilis* à 999 et 409 nucléotides chez *B. halodurans*.

3.2. Analyse phylogénétique des *Bacillaceae* basée sur l'opéron ribosomal :

Les allèles des trois gènes (16S, 23S et 5S) ainsi que des deux régions inter géniques (ITS1 et ITS2) de chacune des souches ont été téléchargés sous format FASTA à partir de "GenBank". Un alignement multiple des séquences nucléotidiques des 109 allèles orthologues du 16S, des 109 du 23S, des 105 du 5S, des 109 des ITS1 et des 105 des ITS2 a été réalisé en utilisant le programme ClustalW. La phylogénie a été établie suivant la méthode de Neighbor-Joining (Saitou et Nei, 1987) par la suite informatique "Molecular Evolutionary Genetics Analysis" version 3 (MEGA3, Kumar *et al.*, 2004; <http://www.megasoftware.net>) en utilisant les paramètres de Kimura des substitutions nucléotidiques multiples (Kimura, 1980). Les arbres phylogénétiques obtenus ont été illustrés à l'aide du programme MEGA3 TreeExplorer et leur fiabilité a été évaluée par le programme "Bootstrap". Six autres allèles de la bactérie *Clostridium tetani* E88, utilisée comme "out group", ont également été récupérés. Celle-ci a été choisie comme bactérie "out group" parce qu'elle est assez proche de la famille des *Bacillus* pour que les allèles orthologues 16S, 23S et 5S partagent une certaine homologie sans pour autant appartenir au même groupe phylogénétique.

3.2.1. Classification phylogénétique basée sur les gènes 16S, 23S et 5S :

Les résultats d'un premier alignement multiple des différents gènes de l'opéron ribosomal (16S, 23S et 5S) ont montré que le début et la fin des allèles, tel qu'annotés dans "GenBank", sont incorrects et ont besoin d'être corrigés. La nécessité d'apporter des corrections similaires à des séquences provenant de "GenBank" a déjà été signalée dans d'autres travaux (Olivier *et al.*, 2005). Pour la correction des annotations, *B. subtilis* a été choisi comme référence. En effet, *B. subtilis* ainsi que *E. coli* sont les deux espèces les mieux étudiées et connues du point de vue génomique. De plus, le génome de cette bactérie, ainsi que les annotations de ses gènes, ont été vérifiés par plusieurs travaux antérieurs. Les modifications apportées aux annotations des gènes sont présentées au tableau 3.1.

Tableau 3.1 : Caractéristiques des différents allèles des gènes de l'opéron ribosomal

Ce tableau donne le numéro d'accèsion pour le génome de chaque souche, le nom du gène pour 16S, 23S et 5S, le nom suggéré pour différencier entre les allèles et la localisation de début et de fin pour chaque gène tels qu'indiqués dans "GenBank" et leurs localisations après les corrections s'il y a lieu. Les signes (+) et (-) indiquent le brin d'ADN porteur du gène.

Souches	Numéro d'accèsion	Nom du gène dans "GenBank"	Nom suggéré	Gène 16S		Gène 23S		Gène 5S	
				localisation dans "GenBank"	localisation corrigée	localisation dans "GenBank"	localisation corrigée	localisation dans "GenBank"	localisation corrigée
<i>B. anthracis</i> Ames	NC_003997	<i>rRNA</i>		9335 10841(+)	9307 10861(+)	11247 14154(+)	11233 14154(+)	14203 14318(+)	
		<i>rRNAB</i>		29129 30635(+)	29101 30655(+)	31041 33948(+)	31027 33948(+)	33997 34112(+)	
		<i>rRNAC</i>		82451 83957(+)	82423 83977(+)	84136 87043(+)	84122 87043(+)	87093 87208(+)	
		<i>rRNAD</i>		145515 147021(+)	145487 147041(+)	147200 150107(+)	147186 150107(+)	150194 150309(+)	
		<i>rRNAE</i>		246235 247741(+)	246207 247761(+)	247920 250827(+)	247906 250827(+)	250877 250992(+)	
		<i>rRNAF</i>		266814 268320(+)	266786 268340(+)	268499 271406(+)	268485 271406(+)	271456 271571(+)	
		<i>rRNAG</i>		280112 281618(+)	280084 281638(+)	281797 284704(+)	281783 284704(+)	284755 284870(+)	
		<i>rRNAH</i>		290085 291591(+)	290057 291611(+)	291770 294677(+)	291756 294677(+)	294728 294843(+)	
		<i>rRNAI</i>		532463 533968(+)	532435 533988(+)	534140 537047(+)	534126 537047(+)	537148 537263 (+)	
		<i>rRNAJ</i>		741564 743070(+)	741536 743090(+)	743249 746156(+)	743235 746156 (+)	746258 746373(+)	
		<i>rRNAK</i>		4655211 4656717(-)	4655191 4656745(-)	4652125 4655032(-)	4652125 4655046(-)	4651908 4652023 (-)	
<i>B. anthracis</i> Ames 0581 "Ancestor"	NC_007530	<i>rRNA</i>		9335 10841(+)	9307 10861(+)	11247 14154(+)	11233 14154(+)	14203 14318(+)	
		<i>rRNAB</i>		29129 30635(+)	29101 30655(+)	31041 33948(+)	31027 33948(+)	33997 34112(+)	
		<i>rRNAC</i>		82451 83957(+)	82423 83977(+)	84136 87043(+)	84122 87043(+)	87093 87208(+)	
		<i>rRNAD</i>		145515 147021(+)	145487 147041(+)	147200 150107(+)	147186 150107(+)	150194 150309(+)	
		<i>rRNAE</i>		246235 247741(+)	246207 247761(+)	24792 250827(+)	247906 250827(+)	250877 250992(+)	
		<i>rRNAF</i>		266814 268320(+)	266786 268340(+)	268499 271406(+)	268485 271406(+)	271456 271571(+)	
		<i>rRNAG</i>		280112 281618(+)	280084 281638(+)	281797 284704(+)	281783 284704(+)	284755 284870(+)	
		<i>rRNAH</i>		290085 291591(+)	290057 291611(+)	291770 294677(+)	291756 294677(+)	294728 294843(+)	
		<i>rRNAI</i>		532463 533968(+)	532435 533988(+)	534140 537047(+)	534126 537047(+)	537148 537263 (+)	
		<i>rRNAI</i>		741565 743071(+)	741537 743091(+)	743250 746157(+)	743236 746157(+)	746259 746374 (+)	
		<i>rRNAK</i>		4655337 4656843(-)	4655317 4656871(-)	4652251 4655158(-)	4652251 4655172(-)	4652034 4652149(-)	
<i>B. anthracis</i> Sterne	NC_005945	<i>rRNA1</i>		9336 10845(+)	9308 10862(+)	11248 14158(+)	11234 14155(+)	14204 14322(+)	14204 14319(+)
		<i>rRNA2</i>		29130 30639(+)	29102 30656(+)	31042 33952(+)	31028 33949(+)	33998 34116(+)	33998 34113(+)
		<i>rRNA3</i>		82452 83961(+)	82424 83978(+)	84197 87047(+)	84123 87044(+)	87094 87212(+)	87094 87209(+)
		<i>rRNA4</i>		145516 147025(+)	145488 147042(+)	147201 150111(+)	147187 150108(+)	150195 150313(+)	150195 150310(+)
		<i>rRNA5</i>		246248 247757(+)	246220 247774(+)	247933 250843(+)	247919 250840(+)	250890 251008(+)	250890 251005(+)
		<i>rRNA6</i>		266827 268336(+)	266799 268353(+)	268512 271422(+)	268498 271419(+)	271469 271587(+)	271469 271584(+)
		<i>rRNA7</i>		280125 281634(+)	280097 281651(+)	281810 284720(+)	281796 284717(+)	284768 284886 (+)	284768 284883 (+)
		<i>rRNA8</i>		290098 291607(+)	290070 291624(+)	291783 294693(+)	291769 294690(+)	294741 294859(+)	294741 294856(+)
		<i>rRNA9</i>		532506 534014(+)	532478 534031(+)	534183 537093(+)	534169 537090(+)	537191 537309 (+)	537191 537306 (+)
		<i>rRNA10</i>		741460 742969(+)	741432 742986(+)	743145 746055(+)	743131 746052(+)	746154 746272 (+)	746154 746269 (+)
		<i>rRNA11</i>		4656430 4657939(-)	4656413 4657967(-)	4653344 4656254(-)	4653347 4656268(-)	4653133 4653241(-)	4653130 4653245(-)

Souches	Numéro d'accension	Nom du gène dans "GenBank"	Nom suggéré	Gène 16S		Gène 23S		Gène 5S		
				localisation dans "GenBank"	localisation corrigée	localisation dans "GenBank"	localisation corrigée	localisation dans "GenBank"	localisation corrigée	
<i>B. cereus</i> ATCC 10987	NC_003909	<i>Bc A</i>		9335 10842(+)	9307 10860(+)	11246 14153(+)	11232 14153(+)	14203 14318(+)	150200 150315(+)	
		<i>Bc B</i>		29109 30616(+)	29081 30634(+)	31020 33927 (+)	31006 33927(+)	33977 34092(+)		
		<i>Bc C</i>		82453 83960(+)	82425 83978(+)	84136 87044(+)	84122 87044(+)	87094 87209 (+)		
		<i>Bc D</i>		145521 47029(+)	145493 147047(+)	147206 150113(+)	147192 150113(+)	150202 150312(+)		
		<i>Bc E</i>		279626 281133(+)	279598 281151(+)	281309 284217(+)	281295 284217(+)	284267 284382(+)		
		<i>Bc F</i>		300676 302184(+)	300648 302202(+)	302361 305269 (+)	302347 305269(+)	305319 305434(+)		
		<i>Bc G</i>		313887 315394(+)	313859 315412(+)	315570 318477(+)	315556 318477(+)	318527 318642 (+)		
		<i>Bc H</i>		322986 324494(+)	322958 324512(+)	324671 327578(+)	324657 327578(+)	327628 327743 (+)		
		<i>Bc I</i>		330804 332312(+)	330776 332330 (+)	332489 335396(+)	332475 335396(+)	335446 335561(+)		
		<i>Bc J</i>		604958 606465(+)	604930 606483 (+)	606642 609549(+)	606628 609549(+)	609653 609763(+)		609651 609766(+)
		<i>Bc K</i>		818165 819672(+)	818137 819690 (+)	819849 822756(+)	819835 822756(+)	822861 822971(+)		822859 822974 (+)
		<i>Bc L</i>		4654163 4655724(-)	4654170 4655723 (-)	4651104 4654012(+)	4651104 4654026(-)	4650883 4651001(-)		4650886 4651001(-)
		<i>B. cereus</i> ATCC 14579	NC_004722	<i>rRNAA</i>		9188 10690(+)	9187 10741(+)	11108 13891(+)		11111 14033(+)
<i>rRNAB</i>				28929 30440(+)	28928 30482(+)	30849 33632(+)	30852 33774(+)	33823 33937 (+)	33823 33938 (+)	
<i>rRNAC</i>				82216 83727(+)	82215 83769(+)	83910 86693(+)	83913 86835(+)	92591 92705(+)	92591 92706(+)	
<i>rRNAD</i>				87924 89435(+)	87923 89477(+)	89618 92401(+)	89621 925431(+)	249846 249960 (+)	249846 249961 (+)	
<i>rRNAE</i>				151138 152649(+)	151137 152691(+)	152832 155615(+)	152835 155757(+)	270412 270526(+)	270412 270527(+)	
<i>rRNAF</i>				24517 246690(+)	245178 246732(+)	246872 249655 (+)	246875 249757 (+)	277685 277799(+)	277685 277800(+)	
<i>rRNAG</i>				265744 267255 (+)	265743 267297(+)	267438 270221(+)	267441 270363(+)	290981 291095(+)	290981 291096(+)	
<i>rRNAH</i>				273017 274528(+)	273016 274570(+)	274711 277494 (+)	274714 277636(+)	298088 298202 (+)	298088 298203 (+)	
<i>rRNAI</i>				286314 287825(+)	286313 287867(+)	288008 290791 (+)	288011 290933 (+)			
<i>rRNAJ</i>				293421 294932(+)	293420 294974(+)	295115 297898(+)	295118 298040(+)			
<i>rRNAK</i>				510018 511529(+)	510017 511571(+)	511711 514494(+)	511714 514636(+)			
<i>rRNAL</i>				718441 719952(+)	718440 719994(+)	720134 722917(+)	720137 723059(+)			
<i>rRNAM</i>				4807135 4808646(-)	4807094 4808648(-)	4804170 4806953(-)	4804029 4806951(-)			
<i>B. cereus</i> EL33	NC_006274			<i>rRNA1</i>		9337 10846(+)	9309 10863(+)	11248 14158(+)	11234 14155(+)	14206 14323(+)
		<i>rRNA2</i>		29135 30644(+)	29107 30661(+)	31046 33956(+)	31032 33953(+)	34004 34121(+)	34003 34118(+)	
		<i>rRNA3</i>		82456 83958(+)	82428 83982(+)	84141 87051(+)	84127 87048(+)	87099 87216(+)	87098 87213(+)	
		<i>rRNA4</i>		145514 147016(+)	145486 14704(+)	147199 150109(+)	147185 150106(+)	150195 150308(+)	150193 150308(+)	
		<i>rRNA5</i>		249459 250968(+)	249431 250985(+)	251144 254054(+)	251130 254051(+)	254103 254220(+)	254102 254217(+)	
		<i>rRNA6</i>		272048 273557(+)	272020 273574(+)	273733 276643(+)	273719 276640(+)	276691 276808(+)	276690 276805(+)	
		<i>rRNA7</i>		279241 280750(+)	279213 280767(+)	280925 283837(+)	280911 283834(+)	283885 284002(+)	283884 283999(+)	
		<i>rRNA8</i>		286608 288117(+)	286580 288134(+)	288293 291203(+)	288279 291200(+)	291251 291368(+)	291250 291365(+)	
		<i>rRNA9</i>		296724 298233 (+)	296696 298250(+)	298408 301319(+)	298394 301316(+)	301367 301484(+)	301366 301481(+)	
		<i>rRNA10</i>		303857 305366(+)	303829 305383(+)	305542 308452(+)	305528 308449(+)	308500 308617(+)	308499 308614(+)	
		<i>rRNA11</i>		514673 516175(+)	514645 516199(+)	516358 519268(+)	516344 519265(+)	519368 519481(+)	519366 519481(+)	
		<i>rRNA12</i>		731439 732948(+)	731411 732965(+)	733124 736034(+)	733110 736031(+)	736136 736249(+)	736134 736249(+)	
		<i>rRNA13</i>		4726430 4727932(-)	4726406 4727960(-)	4723337 4726247(-)	4723340 4726261(-)	4723122 4723237(-)	4723123 4723238(-)	

Souches	Numéro d'accension	Nom du gène dans "GenBank"	Nom suggéré	Gène 16S		Gène 23S		Gène 5S	
				localisation dans "GenBank"	localisation corrigée	localisation dans "GenBank"	localisation corrigée	localisation dans "GenBank"	localisation corrigée
<i>B. halodurans</i> C-125	NC_002570	<i>rrnA</i> <i>rrnB</i> <i>rrnC</i> <i>rrnD</i> <i>rrnE</i> <i>rrnF</i> <i>rrnI</i> <i>rrnG</i> <i>rrnH</i>		22819 24370(+) 50588 52139(+) 117374 118925(+) 124393 125945(+) 261929 263480(+) 559243 560794(+) 656197 657748(+) 1017921 1019473(+)		25054 27992(+) 52991 5929(+) 119924 122862(+) 126753 129691(+) 264163 267101(+) 561522 564460(+) 658720 661658(+) 1020325 1023263(+)		28401 28516(+) 56235 56350 (+) 123273 123388 (+) 129997 130112 (+) 267406 267521(+) 564869 564984(+) 577146 577261(+) 662067 662182 (+) 10236721023787(+)	
<i>B. licheniformis</i> ATCC14580	NC_006270	<i>rRNA</i> <i>rRNA</i> <i>rRNA</i> <i>rRNA</i> <i>rRNA</i> <i>rRNA</i> <i>rRNA</i>	A B C D E F G	9908 11452(+) 34611 36155(+) 95347 96892(+) 158300 159845(+) 612013 613557(+) 920492 922037(+) 3121432 3122977(-)	9900 11452(+) 34603 36155 (+) 95339 96892(+) 158293 159846(+) 612005 613557(+) 920484 922037(+) 3121433 3122986(-)	11787 14719(+) 36490 39420(+) 97065 99995(+) 160018 162951(+) 613730 616662(+) 922210 925140(+) 3118329 3121262(-)	11788 14719(+) 160019 162952(+) 3118330 3121263(-)	14851 14966(+) 39553 39668(+) 100128 100243(+) 163084 163119(+) 616795 616910(+) 925324 925416(+) 3118082 3118197(-)	163085 163200(+) 925311 925426(+) 3118083 3118198(-)
<i>B. subtilis</i> 168	NC_000964	<i>rrnO</i> <i>rrnA</i> <i>rrnJ</i> <i>rrnW</i> <i>rrnI</i> <i>rrnH</i> <i>rrnG</i> <i>rrnE</i> <i>rrnD</i> <i>rrnB</i>		9809 11361(+) 30277 31829(+) 90533 92085(+) 96389 97941(+) 160892 162443(+) 166498 168051(+) 171497 173046(+) 634981 636533(+) 946030 947583(+) 3176132 3177681(-)	 166499 168051(+) 171497 173047(+) 946031 947582(+)	11707 14634(+) 32175 35102 (+) 92251 95178(+) 98107 101034(+) 162609 165534(+) 168217 171140(+) 173213 176140(+) 636702 639629(+) 947751 950679(+) 3173037 3175963(-)	 165590 165705(+) 171196 171307(+) 176196 176307(+) 639685 639800(+) 950791 950906(+) 3172866 3172981(-)	14690 14801(+) 35235 35346(+) 95234 95349(+) 101090 101205(+) 165590 165705(+) 171196 171307(+) 176196 176307(+) 639685 639800(+) 950791 950906(+) 3172866 3172981(-)	14690 14805(+) 35235 35350(+) 171196 171311(+) 176196 176311(+)
<i>B. thuringiensis</i> serovar <i>konkukian</i> 97-27	NC_005957	<i>rRNA1</i> <i>rRNA2</i> <i>rRNA3</i> <i>rRNA4</i> <i>rRNA5</i> <i>rRNA6</i> <i>rRNA7</i> <i>rRNA8</i> <i>rRNA9</i> <i>rRNA10</i> <i>rRNA11</i> <i>rRNA12</i> <i>rRNA13</i>		9337 10846(+) 29138 30647(+) 82466 83975(+) 88179 89688(+) 151247 152755(+) 252230 253739(+) 272749 274258(+) 279760 281269(+) 287106 288614(+) 297075 298584(+) 521476 522985(+) 732426 733935 (+) 4660586 4662095(-)	9309 10863(+) 29110 30664(+) 82438 83992 (+) 88151 89705(+) 151219 152772(+) 252202 253756(+) 272721 274275(+) 279732 281286(+) 287078 288631(+) 297047 298601(+) 521448 523002(+) 732398 733952(+) 4660569 4662123(-)	11249 14160(+) 31050 33960(+) 84151 87061(+) 89864 92774(+) 152931 155841(+) 253915 256825(+) 274434 277344(+) 281445 284355(+) 298760 301671(+) 305935 308846(+) 523160 526070(+) 734110 737020(+) 465750 4660410(-)	11235 14157(+) 31036 33957(+) 84137 87058(+) 89850 92771(+) 152917 155838(+) 253901 256822(+) 274420 277341(+) 281431 284352 (+) 298746 301668(+) 305921 308843(+) 523146 526067(+) 734096 737017(+) 4657503 4660424(-)	14209 14326(+) 34009 34126(+) 87162 87275(+) 92822 92939(+) 155927 156040(+) 256874 256991(+) 277392 277509(+) 284402 284519(+) 291748 291865(+) 308895 309012(+) 526171 526284(+) 737121 737234(+) 4657286 4657399(-)	14208 14323(+) 34008 34123(+) 87160 87275(+) 92821 92936(+) 155925 156040(+) 256873 256988(+) 277391 277506(+) 284401 284516(+) 291747 291862(+) 308894 309009(+) 526169 526284(+) 737119 737234(+) 4657286 4657401(-)

Souches	Numéro d'accèsion	Nom du gène dans "GenBank"	Nom suggéré	Gène 16S		Gène 23S		Gène 5S	
				localisation dans "GenBank"	localisation corrigée	localisation dans "GenBank"	localisation corrigée	localisation dans "GenBank"	localisation corrigée
Out group <i>Clostridium tetani</i> E88	NC_004557	<i>rRNA1</i>	A	8715 10223(-)		5606 8450(-)		5340 5427(-)	
		<i>rRNA2</i>	B	41801 43309(-)		38692 41536(-)		38426 38513(+)	
		<i>rRNA3</i>	C	176113 177621(+)		177886 180730(+)		180909 180996(+)	
		<i>rRNA4</i>	D	565000 566508 (+)		567056 569901(+)		570080 570165(-)	
		<i>rRNA5</i>	E	2791982 2793490(-)		2788789 2791633(-)		2788540 2788627(-)	
		<i>rRNA6</i>	F	2797299 2798807(-)		2794106 2796950(-)		2793840 2793927(-)	

La taille des matrices générées par l'alignement multiple des séquences nucléotidiques est de 109 allèles x 1554 nucléotides pour le gène 16S, de 109 allèles x 2929 nucléotides pour le 23S et de 105 allèles x 116 nucléotides pour le 5S.

Les figures 3.1, 3.2 et 3.3 montrent les arbres phylogénétiques obtenus à partir des séquences nucléotidiques des trois gènes 16S, 23S et 5S suivant la méthode de NJ. La figure 3.1, montrant l'arbre phylogénétique basé sur les séquences nucléotidiques du gène 16S, permet de subdiviser les *Bacillus* en trois groupes bien distincts (I, II et III). Le groupe I renferme *B. anthracis* Ames, *B. anthracis* Ancestor, *B. anthracis* Sterne, *B. cereus* 10987, *B. cereus* 14579, *B. cereus* EL33 et *B. thuringiensis* serovar *konkukian* 97-27. Le groupe II englobe *B. licheniformis* 14580 et *B. subtilis* 168. Finalement, le groupe III est formé par *B. halodurans* C-125. Il est intéressant de noter que le groupe I est formé uniquement par les bactéries *B. anthracis*, *B. cereus* et *B. thuringiensis* et qu'aucun allèle de celles-ci ne se retrouve ailleurs dans un autre faisceau. Ceci montre que ces souches sont étroitement liées. Ce résultat est conforme aux conclusions de certaines études antérieures basées sur des données d'électrophorèses de multilocus enzymatique ainsi que sur la présence de la protéine S-layer à la surface des cellules et l'hybridation d'ADN. Ces travaux ont montré que *B. anthracis*, *B. cereus* et *B. thuringiensis* sont très proches et ont même suggéré qu'ils appartiennent probablement à la même espèce (Helgason, 2000). Néanmoins, la topologie de l'arbre découlant de l'analyse du gène 16S ne permet pas de discriminer entre les espèces du groupe I c'est à dire du groupe *cereus* au sens large.

Un examen plus approfondi du groupe II permet de déceler l'existence de deux sous-groupes. Au sein du premier on retrouve les allèles de *B. licheniformis* 14580, alors que le second renferme ceux de *B. subtilis* 168. Ce résultat supporte que *B. licheniformis* 14580 est très proche de *B. subtilis* 168. Ceci est en accord avec les travaux de Lapidus *et al.*, (2002) dans lesquels ils ont montré que le génome de *B. licheniformis* 14580 renferme des régions identiques à ceux retrouvées au sein de celui de *B. subtilis* 168. D'autres études antérieures arrivaient à des conclusions similaires affirmant que *B. licheniformis* 14580 et *B. subtilis* 168 sont étroitement liée (Rey *et al.*, 2004; Xu et Côté, 2003).

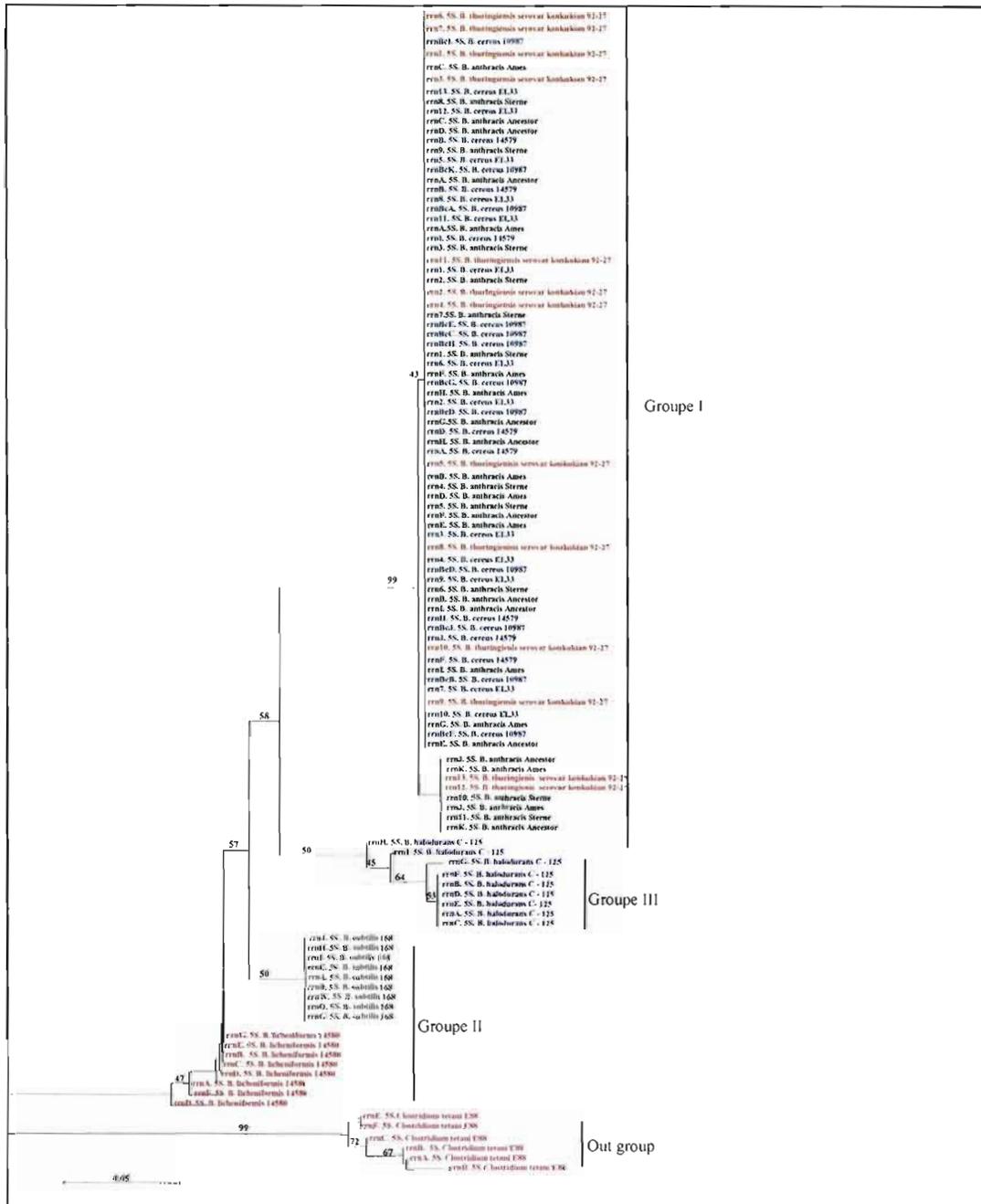


Figure 3.3 : Arbre phylogénétique des *Bacillaceae* basé sur les séquences du gène 5S.
 Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura et leur fiabilité a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 répliquations sont ramenées en pourcentages et sont indiquées au niveau des nœuds.
 ■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

Tableau 3.2 : Taux de similarité au sein du groupe *cereus* au sens large des gènes 16S, 23S et 5S

Gènes	Taux de similarité au sein du groupe <i>cereus</i> au sens large
16S (84 allèles)	99.91%
23S (84 allèles)	99.94%
5S (79 allèles)	98.82%

Les figures 3.2 et 3.3 montrent respectivement les arbres phylogénétiques reconstruits en utilisant les séquences nucléotidiques des gènes 23S et 5S. Ces arbres partagent entre eux la même topologie qui est elle-même similaire à celle découlant de l'analyse du gène 16S. On y retrouve notamment la même répartition des *Bacillus* en trois groupes ainsi que la subdivision du groupe II en deux sous-groupes.

Aucune donnée supplémentaire n'a été apportée par l'analyse basée sur les gènes 23S et 5S concernant la discrimination au sein du groupe *cereus* au sens large. Ceci s'explique par la forte homologie entre les différents allèles des trois gènes qu'illustrés au tableau 3.2.

L'étude phylogénétique des *Bacillus* pourrait, donc, être limité à l'un des trois gènes (16S, 23S ou 5S). Aucun avantage n'est apporté par l'utilisation d'un gène en particulier. Cette sélection pourrait être basée sur d'autres critères tels que la disponibilité ou la facilité de générer des données et la taille du gène.

3.2.2. Classification phylogénétique basée sur les régions ITS1 et ITS2 :

Bien que l'analyse phylogénétique basée sur les trois gènes de l'opéron ribosomal (16S, 23S et 5S) pour les dix souches de *Bacillus* a permis de subdiviser cette famille en trois groupes et deux sous-groupes au niveau de l'espèce, elle a été incapable de discriminer entre les différents membres du groupe *cereus* au sens large. L'utilisation des régions inter géniques pourrait être l'approche qui permettrait de surmonter cette difficulté.

En effet, ces régions sont variables aussi bien au niveau de leur taille que dans leurs séquences nucléotidiques (Jenson *et al.*, 1993; Leblond-Bourget *et al.*, 1996). Plusieurs travaux antérieurs ont affirmé que l'utilisation des régions inter géniques, approche relativement nouvelle, permettrait de différencier entre des espèces bactériennes étroitement liées (Gürtler et Stanisich, 1996; Garcia-Martinez *et al.*, 1999).

La première partie de cette section est consacrée à l'analyse basée sur la région inter génique ITS1 (16S-23S). La taille de la matrice générée par l'alignement multiple des séquences nucléotidiques est de 109 allèles x 999 nucléotides. Les grandes variabilités de la taille des régions inter géniques (de 160 pb à 999 pb) ainsi que de leurs séquences nucléotidiques ont été un obstacle pour la suite informatique MEGA3 lors de l'établissement de la classification phylogénétique. Pour contourner cette difficulté, l'interface Internet du programme DDBJ a été utilisée pour établir et illustrer cette classification.

La figure 3.4 illustre l'arbre phylogénétique établi à partir des séquences nucléotidiques des régions inter géniques ITS1 (16S-23S) suivant la méthode de NJ. Il est possible d'y distinguer deux classes. La première renferme les allèles de petites tailles (160 à 220 pb), alors que la deuxième regroupe les allèles de grandes tailles (400 pb à 999 pb). Aucune information supplémentaire, par rapport à l'analyse basée sur les gènes de l'opéron ribosomal, n'est apportée par l'utilisation de la région inter génique ITS1. En particulier concernant la discrimination au sein du groupe *cereus* au sens large. En effet, les allèles de tailles similaires du groupe *cereus* au sens large se retrouvent regroupés ensemble.

À l'exception de allèles de *B. halodurans* C-125 qui ne sont présent que dans la classe II, tous les autres allèles des autres souches se retrouvent réparties entre les deux classes. La première de ces classes est formée par les allèles correspondant aux ITS1 sans ARN de transfert, alors que la deuxième regroupe les allèles avec ARN de transfert. Les résultats obtenus montrent que ce sont les deux premières copies (A et B) qui contiennent les ARNt sauf pour *B. halodurans* C-125. Pour cette bactérie, ce sont les copies C et G qui renferment les ARNt.

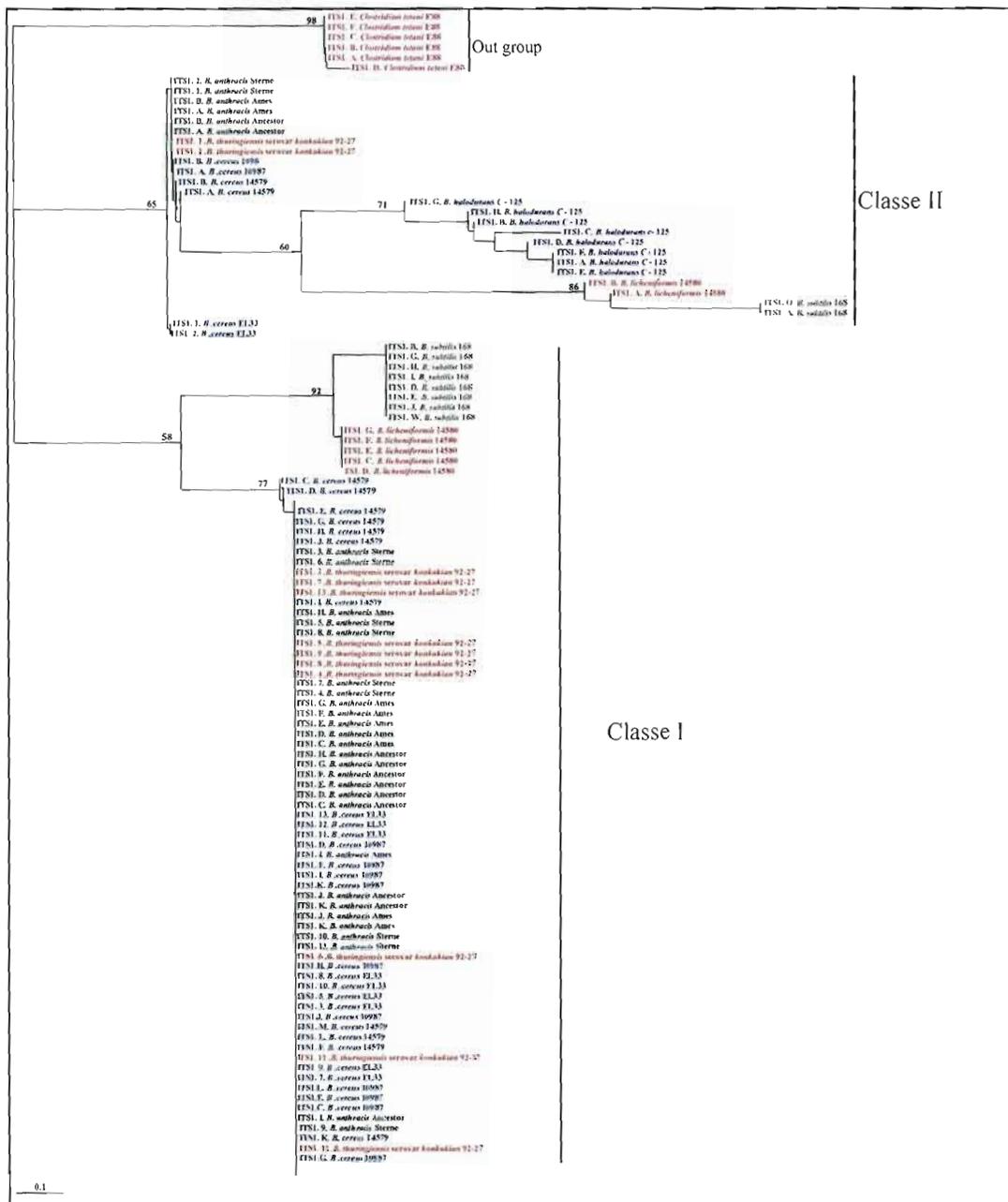


Figure 3.4 : Arbre phylogénétique des *Bacillaceae* basé sur les régions inter génique ITS1 (16S-23S).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par l'interface internet DDBJ en utilisant les paramètres de Kimura et leur fiabilité a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 répliquions sont ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

Deux types de ARNt ont été identifiés; le ARNt^{Ala} (ARN de transfert d'alanine) et le ARNt^{Ile} (ARN de transfert d'isoleucine). Ceci est conforme aux résultats d'autres recherches antérieurs (Ojaimi *et al.*, 1994). Ces résultats permettent d'affirmer que l'utilisation des régions inter géniques ITS1 pour la discrimination entre les espèces est limitée par les grandes variabilités de la taille et des séquences nucléotidiques de ces régions.

La deuxième partie de cette section est consacrée à l'analyse basée sur les régions inter géniques ITS2 (23S-5S). La taille de la matrice générée par l'alignement multiple des séquences nucléotidiques est de 105 allèles x 409 nucléotides.

La figure 3.5 montre l'arbre phylogénétique obtenu à partir des séquences nucléotidiques des régions inter géniques ITS2 (23S-5S). On y retrouve une répartition des *Bacillus* en trois groupes et deux sous-groupes similaire à celle obtenue lors de l'analyse phylogénétique basée sur les gènes de l'opéron ribosomal.

Un examen plus approfondi de la topologie de l'arbre de la figure 3.5 permet de déceler l'existence de plusieurs classes regroupant les allèles suivant leurs homologies et leurs tailles :

- La première classe renferme les huit allèles de *B. cereus* 14579, les copies A, C, E, F, G et H de chacune des *anthracis*, les copies A, B, C, E, G, H et I de *B. cereus* 10987, les copies 2, 3, 5, 6, 7, 8, 9 et 10 de *B. cereus* EL33 et les copies de 1, 2, 4, 6, 7, 8 et 9 de *B. thuringiensis* serovar *konkukian* 97-27.
- La deuxième regroupe la copie D de tous les *anthracis*, de *B. cereus* 10987 et de EL33 ainsi que la copie 5 de *B. thuringiensis* serovar *konkukian* 97-27.
- La troisième classe est formée par la copie I de tous les *anthracis*, les copies J et K de *B. cereus* 10987, les copies 11 et 12 de *B. cereus* EL33 et la copie 10 de *B. thuringiensis* serovar *konkukian* 97-27.
- Les derniers allèles de chaque souche du groupe I à l'exception de *Bt.* forment la quatrième classe. Il s'agit donc des copies J et K de tous les *anthracis*, la copie L de

B. cereus 10987, la copie 13 de *B. cereus* EL33 et les copies 3, 11 et 12 de *B. thuringiensis* serovar *konkukian* 97-27.

- La cinquième classe est représentée par tous les allèles de la souche de *B. halodurans* C-125.
- La sixième comporte l'allèle D de *B. subtilis* 168 et l'allèle F de *B. licheniformis* 14580.
- La septième classe renferme le reste d'allèles de *B. subtilis* 168.
- La dernière regroupe le reste des allèles de *B. licheniformis* 14580.

La faible valeur de "Bootstrap" relative à la classe VII (24) ne remet pas en cause les positions des classes I, II, III, IV, V et VI mais plutôt la position de la classe VII par rapport aux autres.

Ces résultats permettent d'affirmer que l'utilisation des régions inter géniques ITS2 ne permet pas de tirer des conclusions supplémentaires concernant la classification des *Bacillaceae* ou la discrimination au sein des souches appartenant au groupe *cereus* au sens large. En contre partie, cette approche a permis de répartir les allèles en huit classes distinctes.

3.3. Conclusion :

Les résultats de ce chapitre montrent que quelque soit le gène de l'opéron ribosomal utilisé dans l'étude phylogénétique, les arbres obtenus convergent vers la même topologie. Les dix souches des *Bacillus* sont réparties en trois groupes et deux sous-groupes. Toutefois, cette approche s'est révélée incapable de discriminer au sein du groupe *cereus* au sens large.

L'étude phylogénétique des *Bacillus* pourrait donc se limiter à l'un des gènes de l'opéron ribosomal puisqu'aucune information additionnelle n'est apportée par l'utilisation d'un gène en particulier.

L'analyse basée sur les régions inter géniques ITS1 et ITS2 a montré que ces régions ne permettent pas d'extraire des informations supplémentaires concernant la classification phylogénétique des *Bacillus* ni de distinguer au sein du groupe *cereus* au sens large. Néanmoins, cette approche a permis de révéler les hétérogénéités alléliques de ces régions et les subdiviser en plusieurs classes. Ceci pourrait suggérer qu'une étude phylogénétique basée sur une sélection d'allèles, à raison d'un par espèce mais appartenant chacun à une classe différente, pourrait compléter une première analyse basée sur tous les allèles. Cette approche pourrait apporter peut être, une meilleure discrimination notamment pour des espèces étroitement liées.

L'objectif du prochain chapitre est d'apporter plus de robustesse à la classification phylogénétique établie ainsi que d'essayer de discriminer au sein du groupe *cereus* au sens large. L'approche qui sera adoptée dans le chapitre suivant est basée sur l'utilisation des gènes de ménage.

CHAPITRE IV

CLASSIFICATION PHYLOGÉNÉTIQUE DES *BACILLACEAE* BASÉE SUR LES GÈNES DE MÉNAGE *adk*, *aroE* ET *gyrB*

Bien que l'étude phylogénétique basée sur l'opéron ribosomal a permis la subdivision des *Bacillaceae* en trois groupes, cette approche n'a pas permis d'établir des relations phylogénétiques entre les membres formant le groupe *cereus* au sens large. Ce chapitre explore l'approche basée sur les gènes de ménage ("house-keeping genes") dans le but de surmonter cette difficulté; c'est-à-dire de discriminer au sein du groupe *cereus* au sens large et d'offrir plus de robustesse et de fiabilité à l'analyse phylogénétique des *Bacillaceae*.

Les gènes de ménage assurent des fonctions vitales dans tous les types de cellules. Ils existent en un seul exemplaire chez tous les microorganismes et présentent une grande diversité intraspécifique. Ils permettent souvent de donner un résultat rapide, précis et fiable dans l'identification des souches très proches. Les gènes de ménage sont généralement de petite taille. Dans la présente étude on s'intéresse aux trois gènes de ménage suivants : adénylate kinase (*adk*), shikimate 5-déshydrogénase (*aroE*) et la sous-unité B de l'ADN gyrase (*gyrB*).

4.1. Caractéristiques des gènes de ménage (*adk*, *aroE* et *gyrB*) :

Les numéros d'accèsion des toutes les bactéries, les tailles des gènes ainsi que les localisations des séquences nucléotidiques des gènes de ménage, tel qu'annotés dans "GenBank", sont donnés dans le tableau 4.1. Leurs séquences nucléotidiques ainsi que

protéiques ont été téléchargées sous format FASTA à partir de "GenBank". Les gènes de ménage qui sont considérés dans cette étude sont : l'adénylate kinase (*adk*), la shikimate 5-déshydrogénase (*aroE*) et la sous-unité B de l'ADN gyrase (*gyrB*).

La taille des séquences nucléotidiques de ces gènes est variable. La taille d'*adk* est de 651 pb pour les membres du groupe *cereus* au sens large et de 654 pb pour *B. subtilis* 168, *B. halodurans* C-125 et *B. licheniformis* 14580. Celle du shikimate 5-déshydrogénase est de 927 pb pour *B. anthracis* Ancestor et Ames, de 834 pb pour les trois *B. cereus* et *B. anthracis* Sterne, de 837 pb pour *B. halodurans* C-125 et *B. licheniformis* 14580 et de 843 pb pour *B. subtilis* 168 et *B. thuringiensis* serovar *konkukian* 97-27. La taille du gène *gyrB* est de 1923 pb pour les trois *B. anthracis*, *B. cereus* 10987, *B. cereus* EL 33 et *B. thuringiensis* serovar *konkukian* 97-27, de 1929 pb pour *B. cereus* 14579, de 1914 pb pour *B. halodurans* C-125, de 1920 pb pour *B. licheniformis* 14580 et de 1917 pb pour *B. subtilis* 168.

L'alignement multiple des séquences nucléotidiques des gènes de ménage a été réalisé en utilisant le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining (Saitou et Nei, 1987) en utilisant la suite informatique "Molecular Evolutionary Genetics Analysis" version 3 (MEGA3, Kumar *et al.*, 2004, <http://www.megasoftware.net>), en utilisant les paramètres de Kimura pour les substitutions nucléotidiques multiples (Kimura, 1980).

Les arbres phylogénétiques obtenus ont été illustrés à l'aide du programme MEGA3 TreeExplorer et leur fiabilité a été évaluée par le programme "Bootstrap". La même bactérie utilisée précédemment au chapitre III, c'est-à-dire, *Clostridium tetani* E88, a été sélectionnée comme "out group".

Tableau 4.1 : Caractéristiques des séquences nucléotidiques des différents gènes de ménage (*adk*, *aroE* et *gyrB*) des *Bacillus*

Ce tableau donne le nom du gène, la taille de chaque gène (pb) et la localisation de début et de fin pour chaque gène tel qu'indiqué dans "GenBank". Les signes (+) et (-) indiquent le brin d'ADN porteur du gène.

Espèce	Gènes								
	adénylate kinase (<i>adk</i>)			shikimate 5-déshydrogénase (<i>aroE</i>)			sous-unité B de l'ADN gyrase (<i>gyrB</i>)		
	nom du gène dans "GenBank"	taille (pb)	localisation	nom du gène dans "GenBank"	taille (pb)	localisation	nom du gène dans "GenBank"	taille (pb)	localisation
<i>B. anthracis</i> Ames	<i>adk</i>	651	131497 132147(+)	-----	927	4146706 4145780(-)	<i>gyrB</i>	1923	4584 6506(+)
<i>B. anthracis</i> Ancestor	<i>adk</i>	651	131497 132147(+)	<i>aroE</i>	927	4146579 4145653(-)	<i>gyrB</i>	1923	4584 6506(+)
<i>B. anthracis</i> Sterne	<i>adk</i>	651	131498 132148(+)	<i>aroE</i>	834	4147081 4146248(-)	-----	1923	4585 6507(+)
<i>B. cereus</i> ATCC 10987	<i>adk</i>	651	131500 132150(+)	-----	834	4095218 4094385(-)	<i>gyrB</i>	1923	4584 6506(+)
<i>B. cereus</i> ATCC145798	<i>adk</i>	651	137017 137667(+)	<i>aroE</i>	834	4273096 4272263(-)	-----	1929	4458 6386(+)
<i>B. cereus</i> EI33	<i>adk</i>	651	131492 132142(+)	-----	834	4188748 4187915(-)	<i>gyrB</i>	1923	4585 6507(+)
<i>B. halodurans</i> C-125	<i>adk</i>	651	179331 179984(+)	<i>aroE</i>	837	1411430 1412266(+)	<i>gyrB</i>	1914	5130 7043(+)
<i>B. licheniformis</i> ATCC 14580	<i>adk</i>	651	143982 144635(+)	<i>aroD</i>	837	2652879 2652043(-)	<i>gyrB</i>	1920	4985 6904(+)
<i>B. subtilis</i> 168	<i>adk</i>	651	145875 146528(+)	<i>aroD</i>	843	2644695 2643853(-)	<i>gyrB</i>	1917	4865 6782(+)
<i>B. thuringiensis</i> serovar <i>konkukian</i> 97-27	<i>adk</i>	651	137229 137879(+)	<i>aroD</i>	843	4134521 4133688(-)	<i>gyrB</i>	1923	4586 6508(+)
<i>Clostridium tetani</i> E88	-----	651	2736846 2737496(-)	<i>aroE</i>	771	1185379 1186149(+)	-----	1914	46198 48111(+)

4.2. Analyse phylogénétique basée sur les séquences nucléotidiques des gènes de ménage *adk*, *aroE* et *gyrB* :

Un total de 10 allèles pour chacun des trois gènes de ménage suivants : adénylate kinase (*adk*), shikimate 5-déshydrogénase (*aroE*) et la sous-unité B de l'ADN gyrase (*gyrB*) ont été téléchargés à partir de NCBI. Les allèles de ces trois gènes ont aussi été téléchargés pour *Clostridium tetani* E88 utilisé comme "out group".

Les trois matrices résultant de l'alignement des séquences nucléotidiques des gènes de ménage sont composées de 10 allèles x 654 nucléotides pour *adk*, de 10 allèles x 927 nucléotides pour *aroE* et de 10 allèles x 1929 nucléotides pour *gyrB* sans compter l'allèle de "l'out group". À ces matrices, s'ajoute une autre résultant de l'alignement de la concaténation des séquences nucléotidiques des trois gènes de ménage. Elle est formée de 10 allèles x 3510 nucléotides.

Les figures 4.1, 4.2, 4.3 et 4.4 représentent les arbres phylogénétiques, obtenus à partir des séquences nucléotidiques des gènes de ménage (*adk*, *aroE* et *gyrB*) ainsi que de la séquence résultant de leur concaténation, suivant la méthode de NJ.

La figure 4.1, basée sur les séquences nucléotidiques d'*adk*, permet de distinguer trois groupes et sept sous-groupes. Le premier groupe englobe les trois *B. anthracis*, les trois *B. cereus* et *B. thuringiensis*. Le deuxième est formé par *B. subtilis* et *B. licheniformis*, alors que le dernier est formé par *B. halodurans*. L'analyse de cet arbre montre que les groupes I et II se subdivisent respectivement en cinq et deux sous-groupes. Le premier sous-groupe renferme les trois *B. anthracis*, le deuxième est formé par *B. thuringiensis* serovar *konkukian* 97-27, le troisième est représenté par *B. cereus* EL33, le quatrième renferme *B. cereus* 10987 et le cinquième est constitué par *B. cereus* 14579.

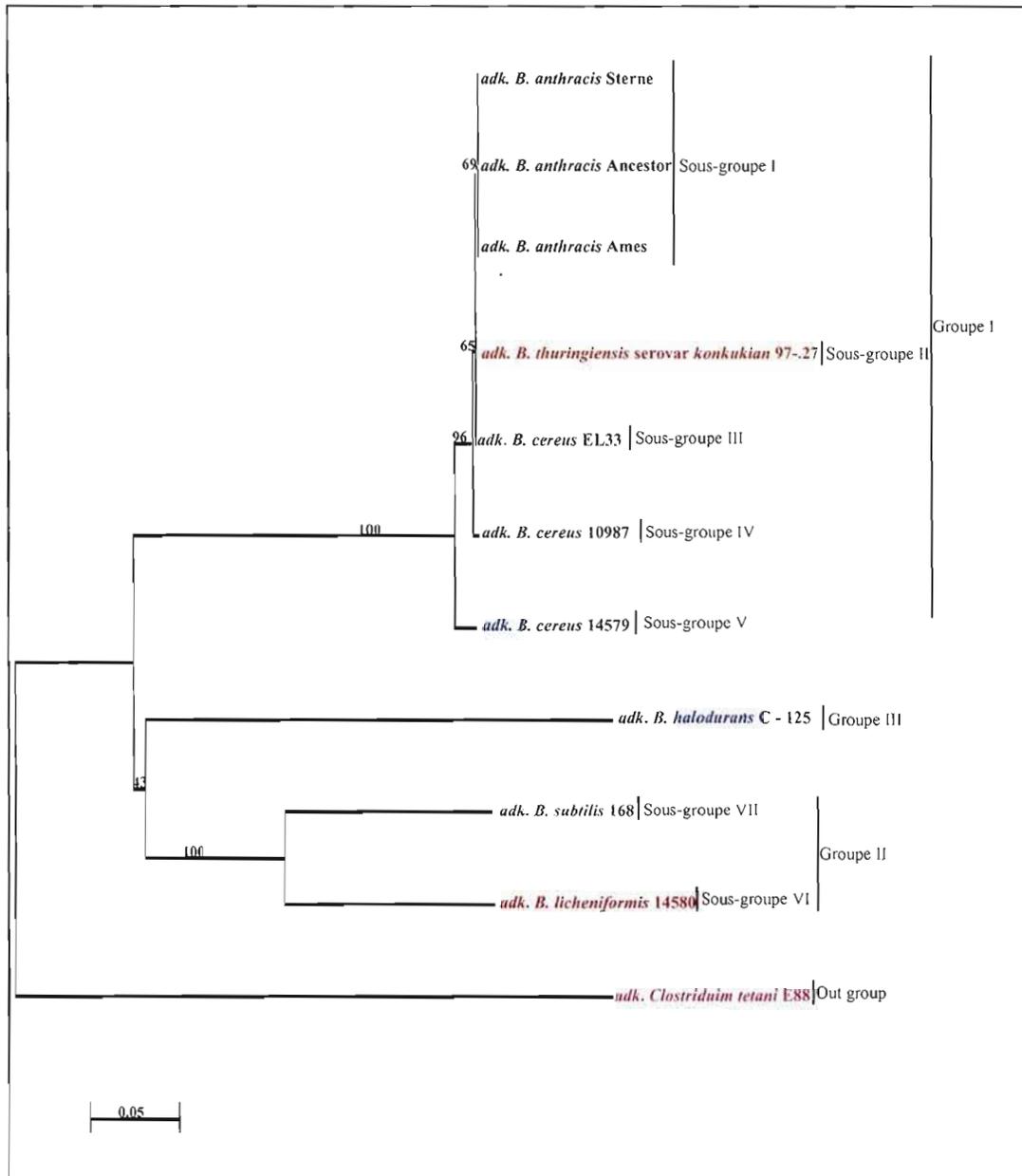


Figure 4.1 : Arbre phylogénétique des *Bacillaceae* basé sur les séquences nucléotidiques du gène adénylate kinase (*adk*).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura. La fiabilité de la classification a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 répliques sont ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

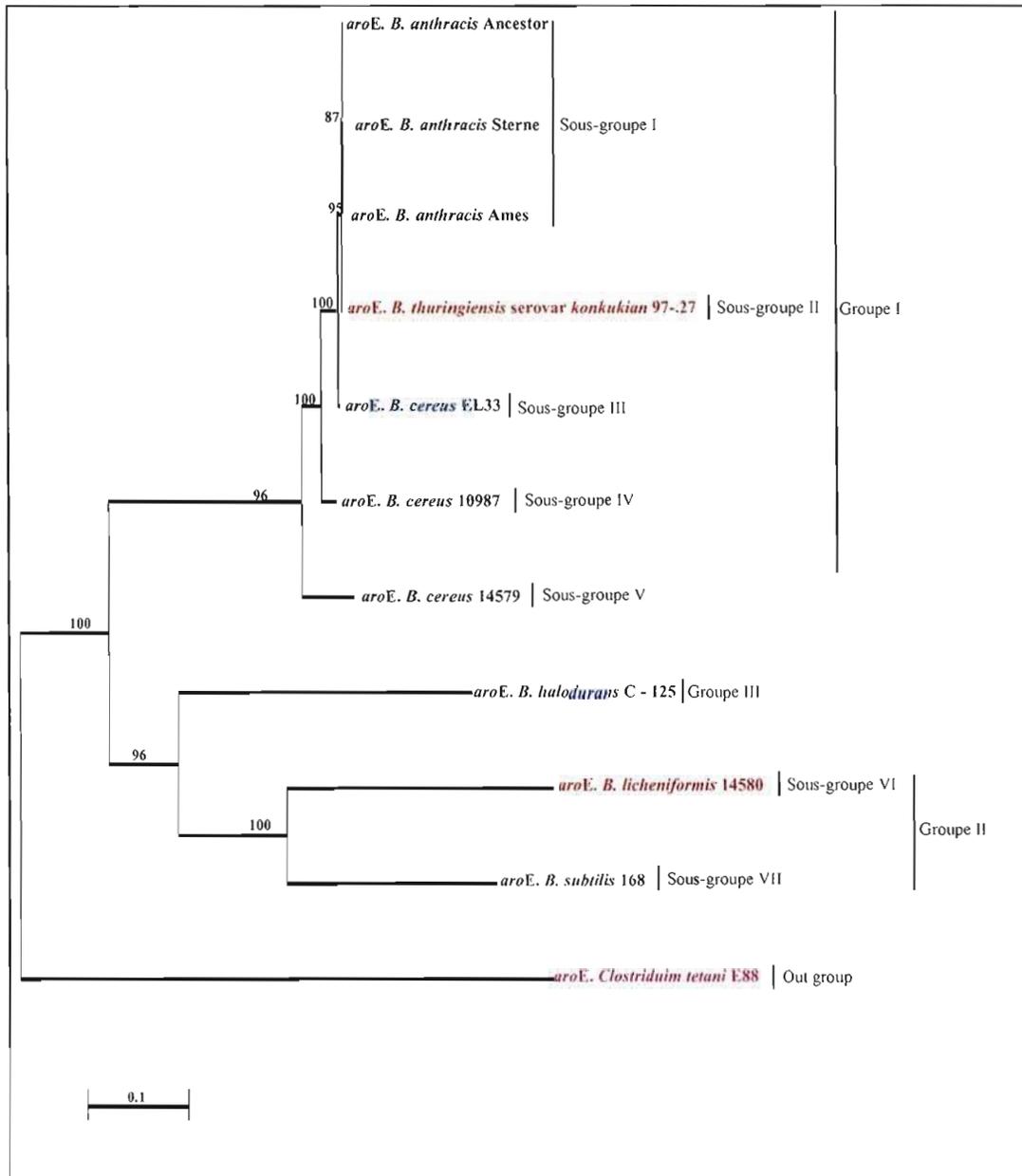


Figure 4.2 : Arbre phylogénétique des *Bacillaceae* basé sur les séquences nucléotidiques du gène shikimate 5-déshydrogénase (*aroE*).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura. La fiabilité de la classification a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 répliquions sont ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

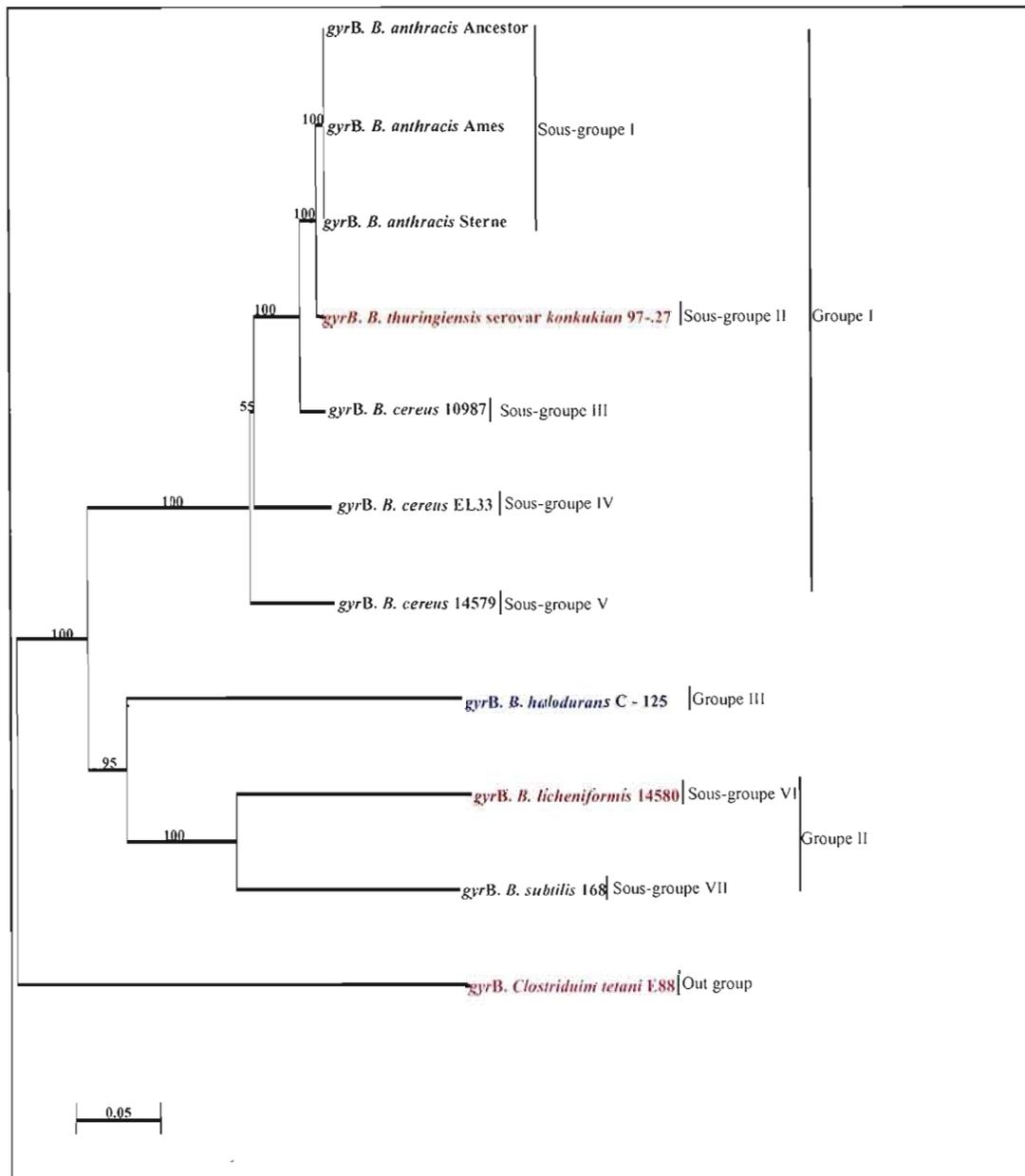


Figure 4.3 : Arbre phylogénétique des *Bacillaceae* basé sur les séquences nucléotidiques du gène de la sous-unité B de l'ADN gyrase (*gyrB*).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura. La fiabilité de la classification a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 répliques sont ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

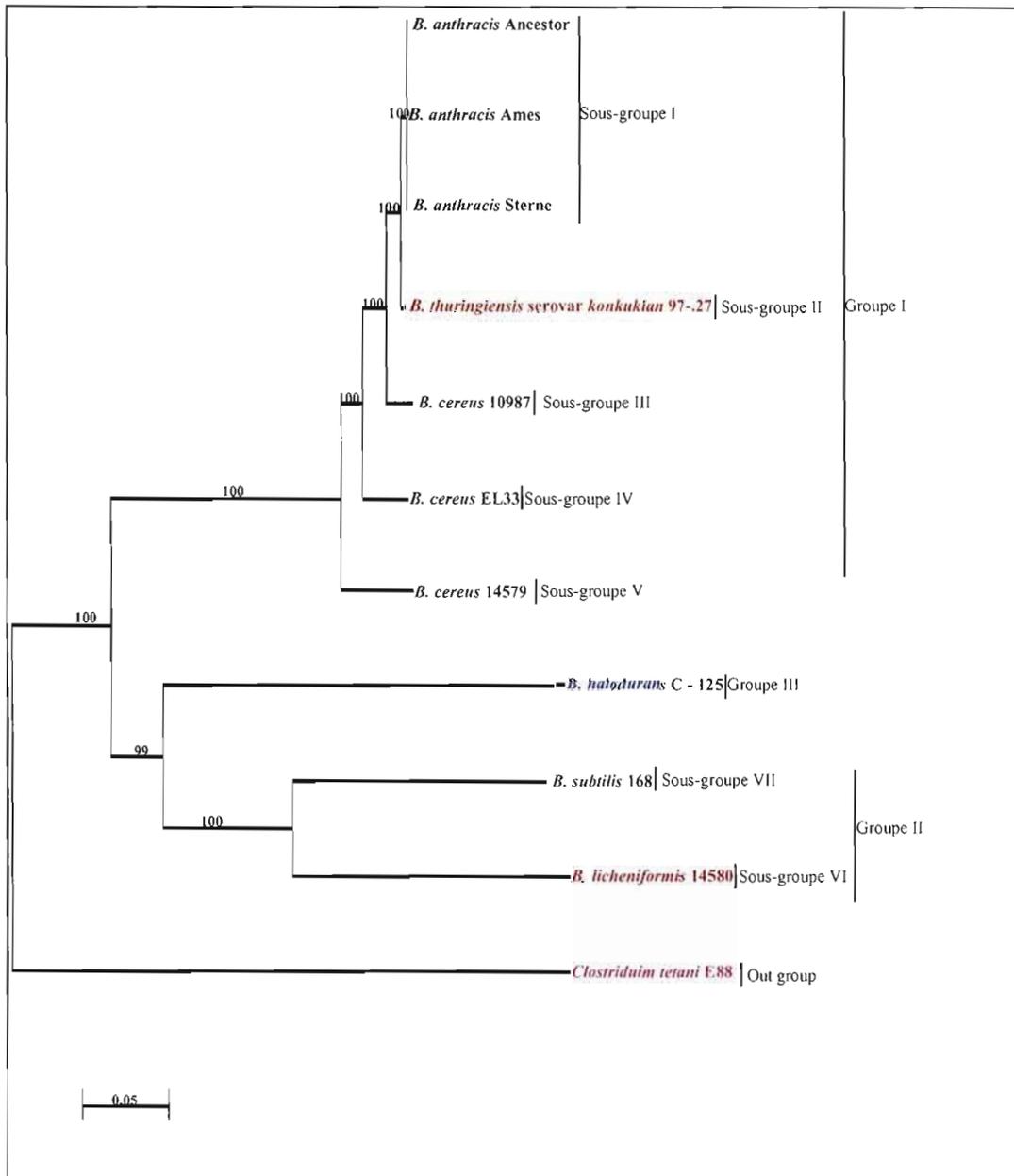


Figure 4.4 : Arbre phylogénétique des *Bacillaceae* basé sur la concaténation des séquences nucléotidiques des gènes adénylate kinase (*adk*), shikimate 5-déshydrégénase (*aroE*) et de la sous-unité B de l'ADN gyrase (*gyrB*).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura. La fiabilité de la classification a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 réplifications sont ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

Il est intéressant de noter que les souches du sous-groupe I, c'est à dire les trois *B. anthracis*, se retrouvent dans le même faisceau secondaire sans qu'elles soient séparées des autres membres du groupe *cereus* au sens large (groupe I). Ceci signifie que bien que *B. anthracis* apparaît comme étant une espèce étroitement liée à *B. cereus* et à *B. thuringiensis*, elle constitue une lignée phylogénétique indépendante qui a probablement divergé des *B. cereus* (Cherif *et al.*, 2003). Les travaux de Kwan *et al.*, (2004) concluent même que les *B. anthracis* proviennent d'une seule souche qui a subi des mutations au cours du temps. Ceci pourrait expliquer l'incapacité de l'approche basée sur l'analyse du gène de ménage (*adk*) à discriminer entre les trois *B. anthracis*.

Le faisceau du sous-groupe II, formé par *B. thuringiensis*, se retrouve entre les *B. anthracis* (sous-groupe I) et les *B. cereus* (sous-groupes III, IV et V). *B. thuringiensis* apparaît donc comme une espèce intermédiaire entre ces deux dernières. Ce résultat rejoint les conclusions des travaux de Cherif *et al.* (2003).

La figure 4.1 révèle aussi la subdivision du groupe II en deux sous-groupes. L'un est représenté par *B. subtilis* 168 et l'autre par *B. licheniformis* 14580. Ce résultat confirme les conclusions de l'analyse basée sur les différents gènes de l'opéron ribosomal (16S, 23S et 5S) et il est en accord avec plusieurs études antérieures qui montrent que *B. licheniformis* 14580 est très proche de *B. subtilis* 168 (Lapidus *et al.*, 2002; Rey *et al.*, 2004; Xu et Côté, 2003). Les positions relatives des groupes II et III ne sont pas très fiables telles qu'indiquées par la valeur du "Bootstrap" 43 (voir figure 4.1).

Les arbres basés sur les séquences nucléotidiques des autres gènes de ménage (*aroE*, *gyrB*) et sur la séquence concaténée des trois gènes (*adk*, *aroE*, *gyrB*) (figures 4.2, 4.3 et 4.4), présentent la même topologie que celui obtenu lors de l'analyse basée sur le gène *adk*. Les conclusions qui peuvent être dégagées à partir de ces arbres sont similaires à celles qui ont été présentées précédemment. Tous ces résultats concordent bien avec ceux provenant de l'analyse basée sur les gènes de l'opéron ribosomal (16S, 23S et 5S) (chapitre III).

L'étude des taux de similarité des séquences nucléotidiques des gènes de ménage au sein des trois *B. anthracis* révèle que celui-ci est de 100% pour les gènes *adk* et *gyrB* et de 89.96% pour *aroE*. Pour ce dernier, les séquences nucléotidiques des trois souches sont identiques jusqu'à la position 819. À la position 820, *B. anthracis* Ames et Ancestor présentent une délétion unique touchant le nucléotide G. Ceci engendre une longueur supplémentaire dans les séquences de ces deux *B. anthracis*. Ce résultat renforce l'hypothèse que tous les *B. anthracis* ont pour origine une même souche clonale qui a subi des mutations au cours du temps.

Les résultats présentés ici montrent que l'utilisation des séquences nucléotidiques des gènes de ménage permet une meilleure discrimination au sein du groupe *cereus* au sens large que l'utilisation des séquences nucléotidiques des gènes de l'opéron ribosomal (16S, 23S et 5S). Toutefois, cette approche demeure incapable d'établir un ordre phylogénétique au sein des *B. anthracis*.

4.3. Analyse phylogénétique basée sur les séquences en acides aminés des protéines codées par les gènes de ménage *adk*, *aroE* et *gyrB* :

Les matrices résultant de l'alignement des séquences en acides aminés sont formées de 10 allèles x 218 acides aminés pour la protéine codée par le gène *adk*, de 10 allèles x 309 acides aminés pour *aroE* et de 10 allèles x 643 acides aminés pour *gyrB*. À ces matrices, s'ajoute une autre résultant de l'alignement de la concaténation des séquences en acides aminés des protéines codées par ces trois gènes de ménage. Elle est formée de 10 allèles x 1170 acides aminés. Ici, également, *Clostridium tetani* E88 a été utilisé comme "out group".

Les figures 4.5, 4.6, 4.7 et 4.8 représentent les arbres phylogénétiques obtenus à partir des séquences en acides aminés des protéines codées par les gènes de ménage (*adk*, *aroE* et *gyrB*) ainsi que des séquences résultant de leur concaténation, suivant la méthode de NJ.

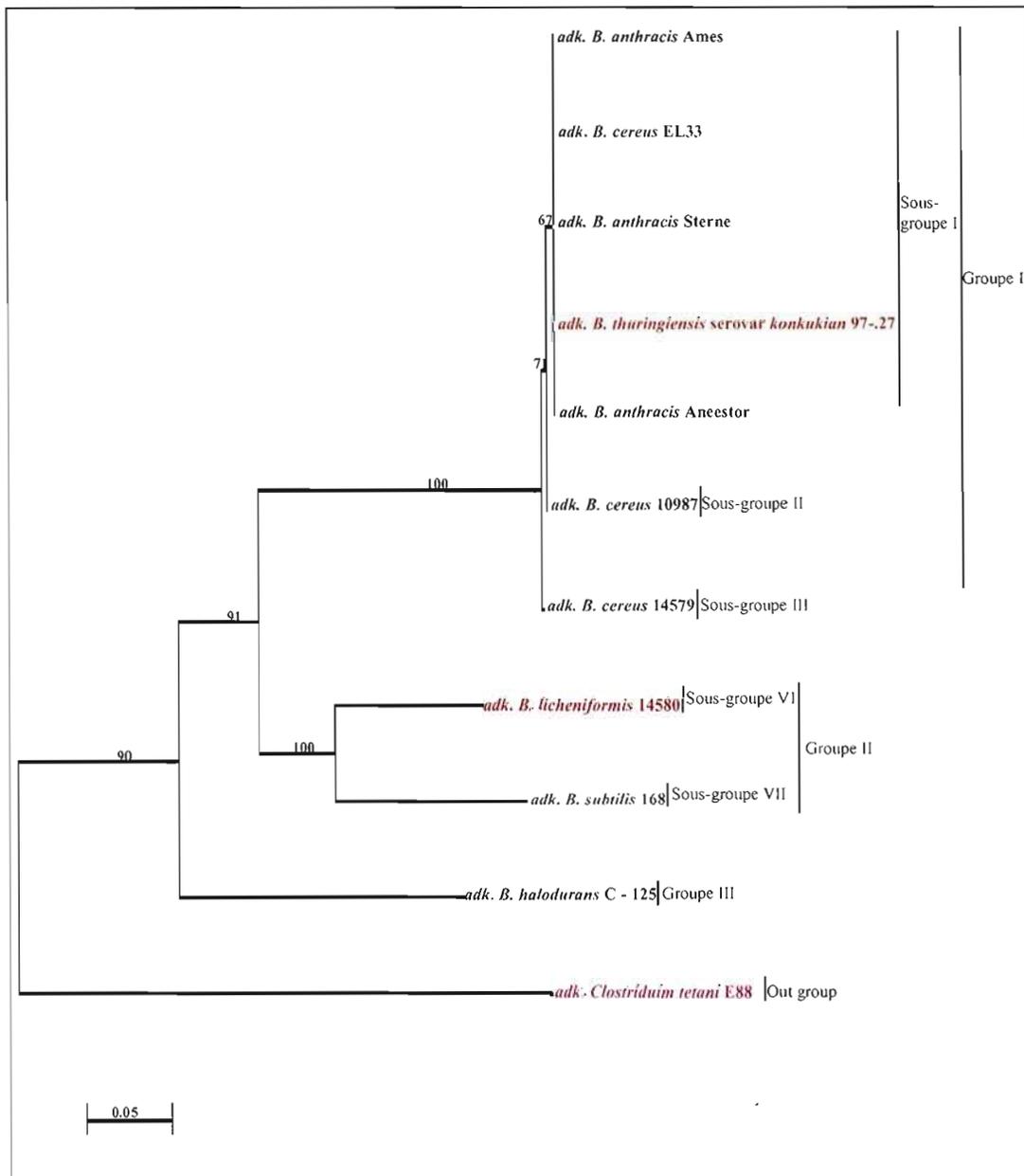


Figure 4.5 : Arbre phylogénétique des *Bacillaceae* basé sur les séquences en acides aminés de la protéine codée par le gène adénylate kinase (*adk*).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura. La fiabilité de la classification a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 répliques ont été ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

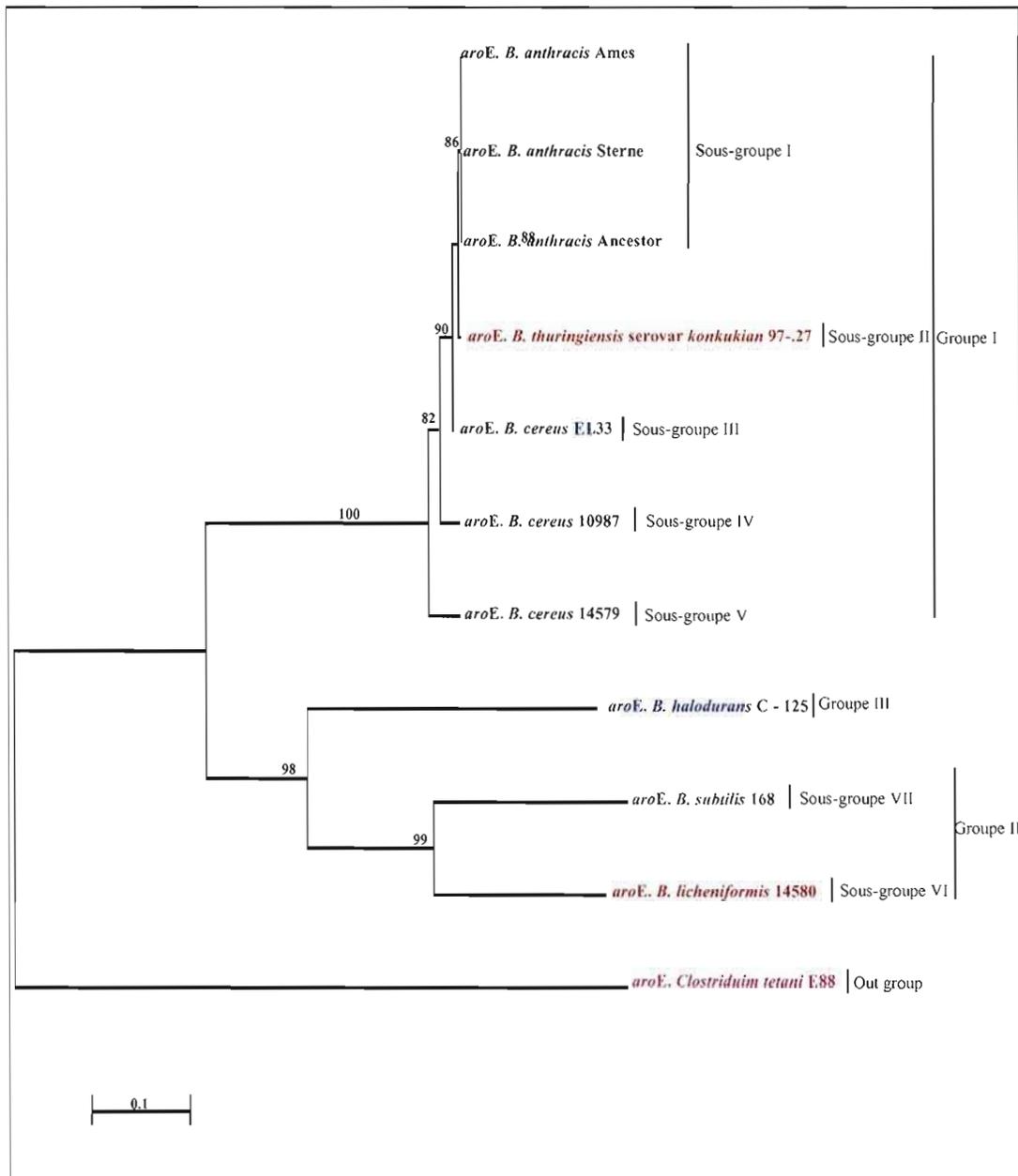


Figure 4.6 : Arbre phylogénétique des *Bacillaceae* basé sur les séquences en acides aminés de la protéine codée par le gène shikimate 5-déshydrogénase (*aroE*).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura. La fiabilité de la classification a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 répliques sont ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

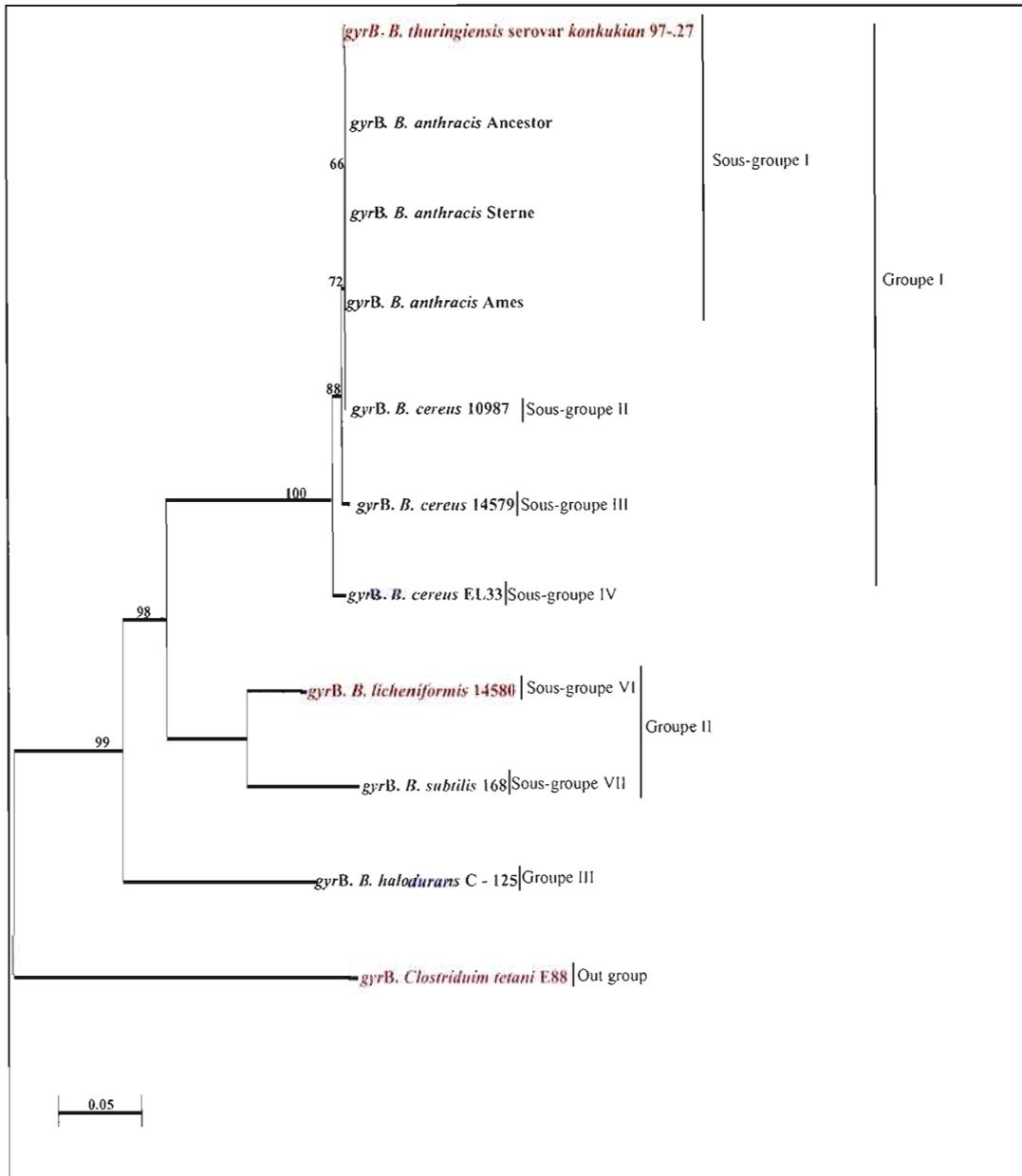


Figure 4.7 : Arbre phylogénétique des *Bacillaceae* basé sur les séquences en acides aminés de la protéine codée par le gène de la sous-unité B de l'ADN gyrase (*gyrB*).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura. La fiabilité de la classification a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 réplifications sont ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

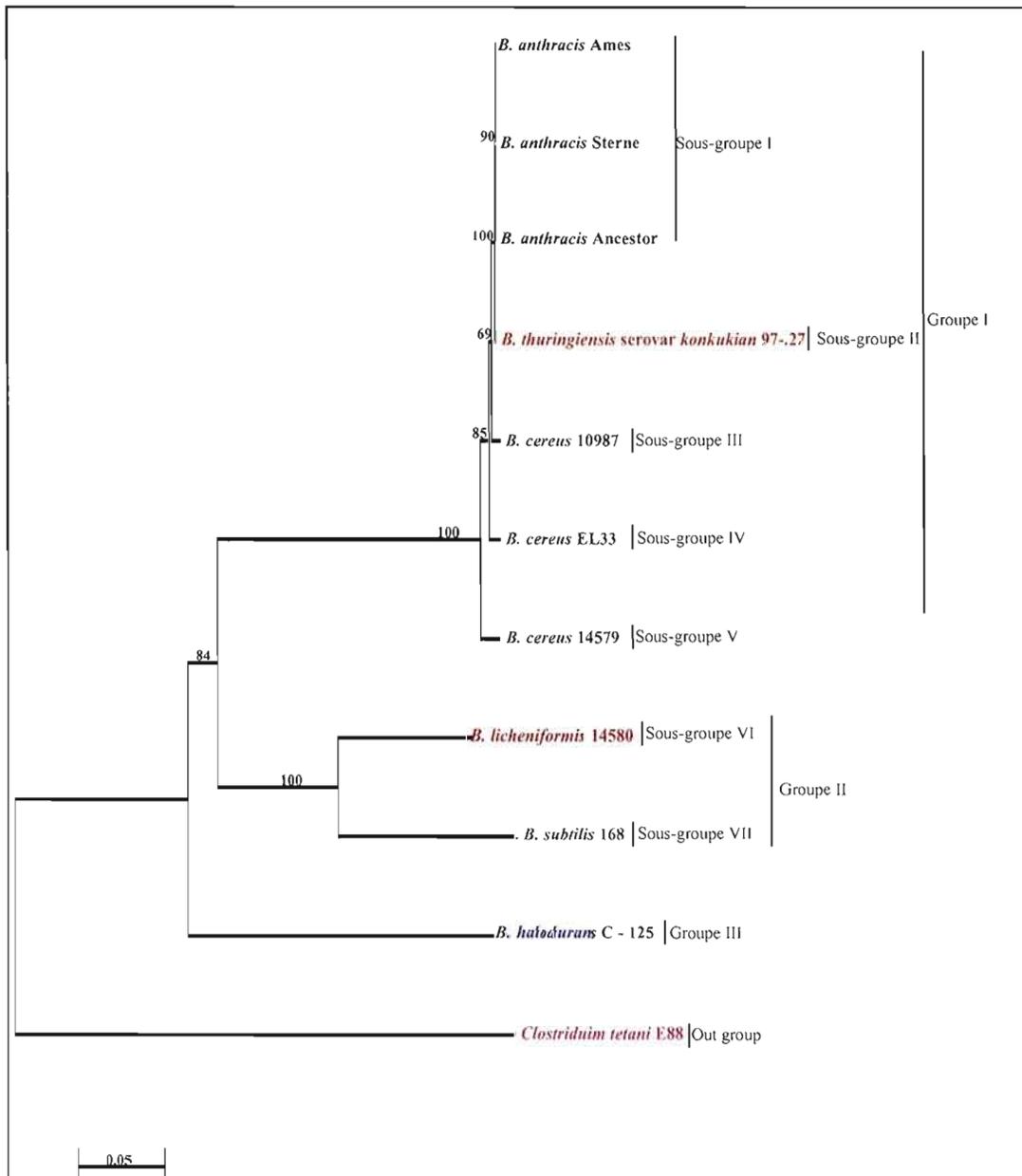


Figure 4.8 : Arbre phylogénétique des *Bacillaceae* basé sur la concaténation des séquences en acides aminés des protéines codées par les gènes adénylate kinase (*adk*), shikimate 5-déshydrégénase (*aroE*) et de la sous-unité B de l'ADN gyrase (*gyrB*).

Ces séquences ont été alignées avec le programme ClustalW et la phylogénie a été établie suivant la méthode de Neighbor-Joining par la suite informatique MEGA3 en utilisant les paramètres de Kimura. La fiabilité de la classification a été évaluée par le programme "Bootstrap". Les valeurs de "Bootstrap" pour 1000 répliquions sont ramenées en pourcentages et sont indiquées au niveau des nœuds.

■ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. licheniformis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *Clostridium tetani*.

Seuls les deux arbres basés sur la séquences en acides aminés de la protéine codée par le gène *aroE* (figure 4.6) et celle résultant de la concaténation des séquences en acides aminés des protéines codées par les trois gènes de ménage (*adk*, *aroE* et *gyrB*) (figure 4.8) présentent une topologie similaire à celle des arbres phylogénétiques obtenus en se basant sur les séquences nucléotidiques des trois gènes de ménage.

La comparaison des arbres reconstruits en se basant sur les séquences en acides aminés des protéines codées par les gènes *adk* et *gyrB* (figures 4.5 et 4.7) montre qu'ils convergent vers la même topologie et subdivisent les *Bacillus* en trois groupes principaux. Ceci est en accord avec les résultats des analyses phylogénétiques basées sur l'opéron ribosomal et les séquences nucléotidiques des gènes de ménage.

Cependant ces deux arbres sont incongruents dans leurs subdivisions du groupe I en sous-groupes. En effet, l'arbre relatif à la protéine codée par le gène *adk* subdivise le groupe I en trois sous-groupes alors que celui de la protéine codée par le gène *gyrB* montre une subdivision en quatre sous-groupes.

Le premier sous-groupe, révélé par l'analyse basée sur la protéine codé par le gène *adk*, renferme les trois *B. anthracis*, *B. cereus* EL33 et *B. thuringiensis* serovar *konkukian* 97-27. Le deuxième est composé par *B. cereus* 10987 et le dernier est constitué par *B. cereus* 14579. Le premier des quatre sous-groupes, découlant de l'analyse basée sur la protéine codée par le gène *gyrB*, est composé des trois *B. anthracis* et de *B. thuringiensis* serovar *konkukian* 97-27. Le deuxième est représenté par *B. cereus* 10987. Le troisième renferme *B. cereus* 14579. Finalement, c'est *B. cereus* EL33 qui se retrouve au quatrième sous-groupe. Toutefois, ces deux subdivisions présentent certaines similarités. En effet, plusieurs espèces se retrouvent au même sous groupe aussi bien avec la protéine codée par le gène *adk* qu'avec la protéine codée par le gène *gyrB*.

Tableau 4.2 : Comparaison des taux de similarité entre les séquences nucléotidiques et les séquences en acides aminés pour les gènes de ménage (*adk*, *aroE* et *gyrB*) et les protéines correspondantes

Nom du gène	Taux de similarités	
	séquences nucléotidiques	séquences en acides aminés
Adénylate Kinase (<i>adk</i>)	49.38%	58.25%
Shikimate 5-déshydrogénase (<i>aroE</i>)	34.11%	33%
Sous unité B d'ADN gyrase (<i>gyrB</i>)	53.86%	68.58%

Contrairement aux arbres générés à partir des séquences nucléotidiques des gènes de ménage, dans lesquels chaque souche à l'exception des *B. anthracis* se retrouve seule dans un faisceau, ceux établis à partir des séquences en acides aminés montrent plutôt un regroupement de plusieurs espèces au sein du même faisceau. Ceci s'explique par un taux d'homologie moins élevé entre les séquences nucléotidiques comparées aux séquences en acides aminés (voir tableau 4.2). Ce taux plus faible est dû à la dégénérescence du code génétique, dans lequel un acide aminé peut être codé par plusieurs codons.

L'utilisation des séquences en acides aminés des protéines codées par les trois gènes de ménage (*adk*, *aroE* et *gyrB*) n'apporte donc aucun élément supplémentaire à l'analyse phylogénétique des *Bacillaceae* par rapport aux séquences nucléotidiques.

4.4. Conclusion :

Les résultats de ce chapitre montrent que les topologies des arbres reconstruits en se basant sur les séquences nucléotidiques des trois gènes de ménage ainsi que leurs séquences concaténées convergent vers la même topologie. Les *Bacillaceae* sont subdivisés en trois groupes et sept sous-groupes. Cette approche a permis de discriminer les souches au sein du groupe *cereus* au sens large. L'analyse phylogénétique suivant cette approche pourrait être limitée à l'un des trois gènes de ménage. Les conclusions qui peuvent en être tirées étant indépendantes du choix du gène.

L'utilisation des séquences en acides aminés n'a pas permis d'avoir une meilleure résolution de la classification phylogénétique notamment au sein du groupe *cereus* au sens large. Cette approche a même été incapable de discriminer entre *B. thuringiensis*, *B. cereus* et les *B. anthracis*. Ceci est dû à la dégénérescence du code génétique qui se manifeste en un taux d'homologie plus élevé pour les séquences en acide aminés.

L'utilisation des séquences en acides aminés des protéines codées par les trois gènes de ménage (*adk*, *aroE* et *gyrB*) ne serait donc une meilleure approche que pour l'étude phylogénétique des espèces éloignées.

De manière globale, les gènes de ménage ont établi une meilleure classification que l'opéron ribosomal. L'utilisation des gènes de ménage serait donc suggérée.

CHAPITRE V

HÉTÉROGÉNÉITÉ DES ALLÈLES DES GÈNES DE L'OPÉRON RIBOSOMAL ET SON EFFET SUR LA CLASSIFICATION PHYLOGÉNÉTIQUE

Jusqu'à récemment, les classifications phylogénétiques étaient le plus souvent basées sur les séquences nucléotidiques du gène d'ARNr 16S (Lane *et al.*, 1985; Garrity et Holt, 2001; Stackebrandt et Goebel, 1994; Vandamme *et al.*, 1996; Woese, 1987; Woese *et al.*, 1990). De plus, la plupart des études n'ont utilisé qu'une seule copie de ce gène; l'hypothèse de l'homogénéité des allèles a été souvent invoquée pour justifier cette approche (Hillis *et al.*, 1991). Cependant, certains travaux récents montrent que des variations intra-spécifiques peuvent exister entre les différents allèles et que cette hétérogénéité pourrait atteindre des taux aussi élevés que 11.6% (Acinas *et al.*, 2004; Cilia *et al.*, 1996; Clayton *et al.*, 1995; Martinez-Murcia *et al.*, 1999).

Cette diversité intra-spécifique trouve ses origines dans l'organisation de ce gène en une famille de multi-gènes et dans le nombre d'allèles qui varie au sein de chaque famille. Par exemple chez les *Bacillaceae* ce nombre peut varier entre 7 et 13. La présence de plusieurs allèles reflète la stratégie de l'organisme à se défendre contre les changements des conditions du milieu extérieur (Klappenbach *et al.*, 2000). En général, les différents allèles de ces familles multi-gènes coévoluent en même temps (Ohta, 1991). Cependant, le degré de polymorphisme au sein de chaque gène dépend de la fréquence des interactions moléculaires (Hillis *et al.*, 1991). Jusqu'à présent, peu de travaux se sont intéressés à l'étude de la

variabilité intra allélique des différents gènes de l'opéron ribosomal et son effet sur la classification phylogénétique.

Le chapitre V a pour objectif d'évaluer l'hétérogénéité intra-allélique des séquences nucléotidiques des différents gènes de l'opéron ribosomal (16S, 23S et 5S) des dix souches de *Bacillaceae* et son effet sur la classification phylogénétique.

5.1. Hétérogénéité des allèles des gènes de l'opéron ribosomal (16S, 23S et 5S) :

Le tableau 5.1 donne pour chaque gène d'ARNr 16S, 23S et 5S le nombre d'allèles, le nombre de mutations entre les différentes copies et les taux d'hétérogénéité intra-allélique. Le nombre d'allèles pour chacun des trois gènes étudiés (16S, 23S et 5S) pour les dix souches de *Bacillaceae* varie entre 7 et 13. Le nombre de différences ayant été observées, en considérant toutes les mutations (substitution, insertion et délétion), varie de 6 chez *B. halodurans* C-125 à 49 chez *B. subtilis* 168. Les plus petit et plus grand nombres de mutations se retrouvent au sein du gène 23S.

L'étude du taux de divergence au sein des différents allèles des trois gènes d'ARNr 16S, 23S et 5S de l'opéron ribosomal révèle qu'il est variable. En effet, pour le gène 16S le taux de divergence est de 0.45% chez *B. cereus* 14579 et de 2.44% chez *B. subtilis* 168. Pour le gène 23S il est de 0.2% au sein *B. halodurans* C-125 et de 1.67% au sein de *B. subtilis* 168. Finalement, le taux de divergence pour le gène 5S est de 0% chez *B. cereus* 14579 et de 18.96% chez *B. thuringiensis* serovar *konkukian* 97-27. Les valeurs des taux de divergence données précédemment représentent les extrêmes observés au sein de chaque gène. C'est le gène 5S qui possède les taux de divergence le plus grand et le plus petit de l'opéron ribosomal.

Tableau 5.1 : Taux de divergence intra-allélique au sein de chaque gène de l'opéron ribosomal

Ce tableau donne pour chaque souche et gène le nombre d'allèles, le nombre de mutations et le taux d'hétérogénéité intra-allélique.

Souches	Gènes								
	16S			23S			5S		
	N° ARN	nombre de mutations	taux de divergence	N° ARN	nombre de mutations	taux de divergence	N° ARN	nombre de mutations	taux de divergence
<i>B. anthracis</i> Ames	11	18	1.1%	11	19	0.6%	11	18	15.5%
<i>B. anthracis</i> Ancestor	11	18	1.1%	11	18	0.6%	11	18	15.5%
<i>B. anthracis</i> Sterne	11	18	1.1%	11	20	0.7%	11	18	15.5%
<i>B. cereus</i> 10987	12	10	0.6%	12	17	0.6%	12	16	13.8%
<i>B. cereus</i> 14579	13	7	0.4%	13	27	0.9%	8	0	0%
<i>B. cereus</i> EI33	13	17	1.1%	13	18	0.6%	13	16	13.8%
<i>B. licheniformis</i> 14580	7	19	1.2%	7	24	0.8%	7	8	7%
<i>B. halodurans</i> c-125	8	17	1.1%	8	6	0.2%	9	8	7%
<i>B. subtilis</i> 168	10	38	2.4%	10	49	1.7%	10	8	7%
<i>B. thuringiensis</i> serovar <i>konkukian</i> 97-27	13	14	0.9%	13	16	0.5%	13	22	19%

(*) Les valeurs maximales et minimales du nombre des mutations et du taux de divergences sont indiquées en caractères gras.

Il est donc clair que le taux d'hétérogénéité intra-allélique est variable aussi bien au niveau des souches que des gènes. Le taux d'hétérogénéité intra-allélique dépend non seulement du nombre de mutations mais aussi de la longueur du gène. Ainsi au nombre de mutations le plus grand peut ne pas correspondre le taux d'hétérogénéité le plus élevé comme c'est le cas chez *B. subtilis* 168.

5.2. Effet de l'hétérogénéité des allèles sur la classification phylogénétique :

La majorité des travaux antérieurs n'ont considéré qu'une seule copie du gène 16S pour établir la classification phylogénétique des *Bacillaceae*. L'hypothèse d'homogénéité des allèles a souvent été invoquée pour justifier cette approche. Cependant les résultats du paragraphe précédent montrent clairement que les différentes copies des trois gènes de l'opéron ribosomal ne sont pas homogènes et que les taux de divergences sont variables aussi bien en fonction du gène et de la souche (Klappenbach *et al.*, 2000). Ces taux de divergence atteignent dans certains cas des valeurs aussi élevées que 19%.

La comparaison des arbres reconstruits en se basant sur tous les allèles des trois gènes (16S, 23S et 5S) (voir les figures 3.1, 3.2 et 3.3 du chapitre III, aux pages 45, 46 et 47) montre qu'ils convergent vers la même topologie. Les *Bacillaceae* sont répartis en trois groupes avec la subdivision du groupe II en deux sous-groupes. Seules des différences mineures ont été décelées. En effet, il a été constaté que la répartition des allèles de certaines souches est affectée par le taux de divergence, notamment au sein du groupe I. Plus ce taux est élevé, plus les allèles ont tendance à se disperser le long de l'arbre.

Il est donc possible de conclure que bien que les taux de divergence puissent atteindre des valeurs significatives (tel que le taux de 19% constaté chez *B. thuringiensis* serovar *konkukian* 97-27 pour les allèles 3, 5, 11, 12 et 13 du gène 5S), cette hétérogénéité n'a pas d'effet notable sur la classification phylogénétique des *Bacillaceae*.

5.3. Conclusion :

Les résultats du chapitre V montrent que les allèles des trois gènes de l'opéron ribosomal ne sont pas homogènes et que le taux de divergence dépend de la souche ainsi que de la taille du gène. Ce taux peut atteindre des valeurs élevées, mais cette hétérogénéité intra-allélique n'a pas d'effet notable sur la classification phylogénétique des *Bacillaceae*.

Il est donc justifié de se limiter à l'analyse d'un seul allèle de l'un des trois gènes de l'opéron ribosomal pour l'étude phylogénétique.

Le prochain chapitre sera consacré à l'analyse de la distribution des mutations au sein des trois gènes (16S, 23S et 5S) de l'opéron ribosomal ainsi qu'à la recherche de régions hypervariables qui seraient qualifiées de "hot-spots". Ces régions, si elles existent, permettraient peut être une meilleure discrimination des espèces, *anthracis*, *thuringiensis*, *cereus*, au sein du groupe *cereus* au sens large grâce à leurs faibles taux de conservation.

CHAPITRE VI

DISTRIBUTION DES MUTATIONS

La majorité des travaux antérieurs n'ont pas pris en considération l'hétérogénéité possible des allèles des gènes de l'opéron ribosomal et son effet sur la classification phylogénétique. Les résultats du chapitre précédent montrent que les taux d'hétérogénéité des allèles peuvent atteindre des valeurs significatives (tel que le taux de 19% constaté chez *B. thuringiensis* serovar *konkukian* 97-27). Néanmoins, cette hétérogénéité n'affecte pas de manière notable les conclusions de l'analyse phylogénétique.

Ce chapitre a pour objectif d'étudier la distribution des mutations au sein de chacun des trois gènes de l'opéron ribosomal (16S, 23S et 5S) pour la famille des *Bacillaceae*, ainsi qu'au sein du gène 16S pour la sous division des γ -protéobactéries. Cette étude permettra d'élucider s'il existe des régions hypervariables ("hot-spots"). Si elles existent, sont-elles spécifiques à chaque groupe de bactéries ? L'utilisation des séquences concaténées de telles régions lors d'une analyse phylogénétique pourrait offrir une approche intéressante pour discriminer entre des espèces étroitement liées.

6.1. Distribution des mutations au sein des gènes 16S, 23S et 5S pour les *Bacillaceae* :

Cette partie considère tous les allèles d'une vingtaine de souches qui appartiennent à la famille des *Bacillaceae*. Le génome de certaines souches est complètement séquencé alors que celui d'autres ne l'est que partiellement. Le tableau 3.1 du chapitre III aux pages 39 à 42 montre les numéros d'accession, le nombre d'allèles ainsi que les localisations des gènes 16S, 23S et 5S pour les souches complètement séquencées. Celles des souches partiellement séquencées sont données aux tableaux A.1, A.2 et A.3 de l'annexe. Dans ces tableaux figurent aussi les numéros de contiguïté. La recherche des différentes copies pour les trois gènes a été réalisée par le programme Blast. Un total de 188 allèles de 21 souches pour le gène 16S, 171 allèles de 21 souches pour le gène 23S et 163 allèles de 20 souches ont été alignés en utilisant le programme ClustalW. Au sein de chaque souche le nombre de copies varie entre 2 et 13.

Les figures 6.1, 6.2 et 6.3 représentent les histogrammes de distribution des mutations au sein des trois gènes de l'opéron ribosomal (16S, 23S et 5S). La figure 6.1 révèle l'existence de cinq régions hypervariables ayant une concentration élevée de mutations au sein du gène 16S. Ces zones s'étendent respectivement entre 0 et 10 pb, 60 et 100 pb, 180 et 230 pb, 1130 et 1160 pb et finalement 1450 et 1480 pb. Ceci est conforme aux conclusions du travail de Coenye et Vandamme (2003) dans lequel ils ont montré, pour 55 bactéries appartenant à des familles différentes, que le gène 16S présente des régions hypervariables dans lesquelles se concentrent les mutations.

Les figures 6.2 et 6.3 montrant la distribution des mutations au sein des gènes 23 S et 5S de l'opéron ribosomal, permettent d'en déduire les mêmes résultats quant à l'existence des régions hypervariables. En effet, ces histogrammes révèlent la présence de quatre régions hypervariables au sein du gène 23S et de trois régions au sein du gène 5S.

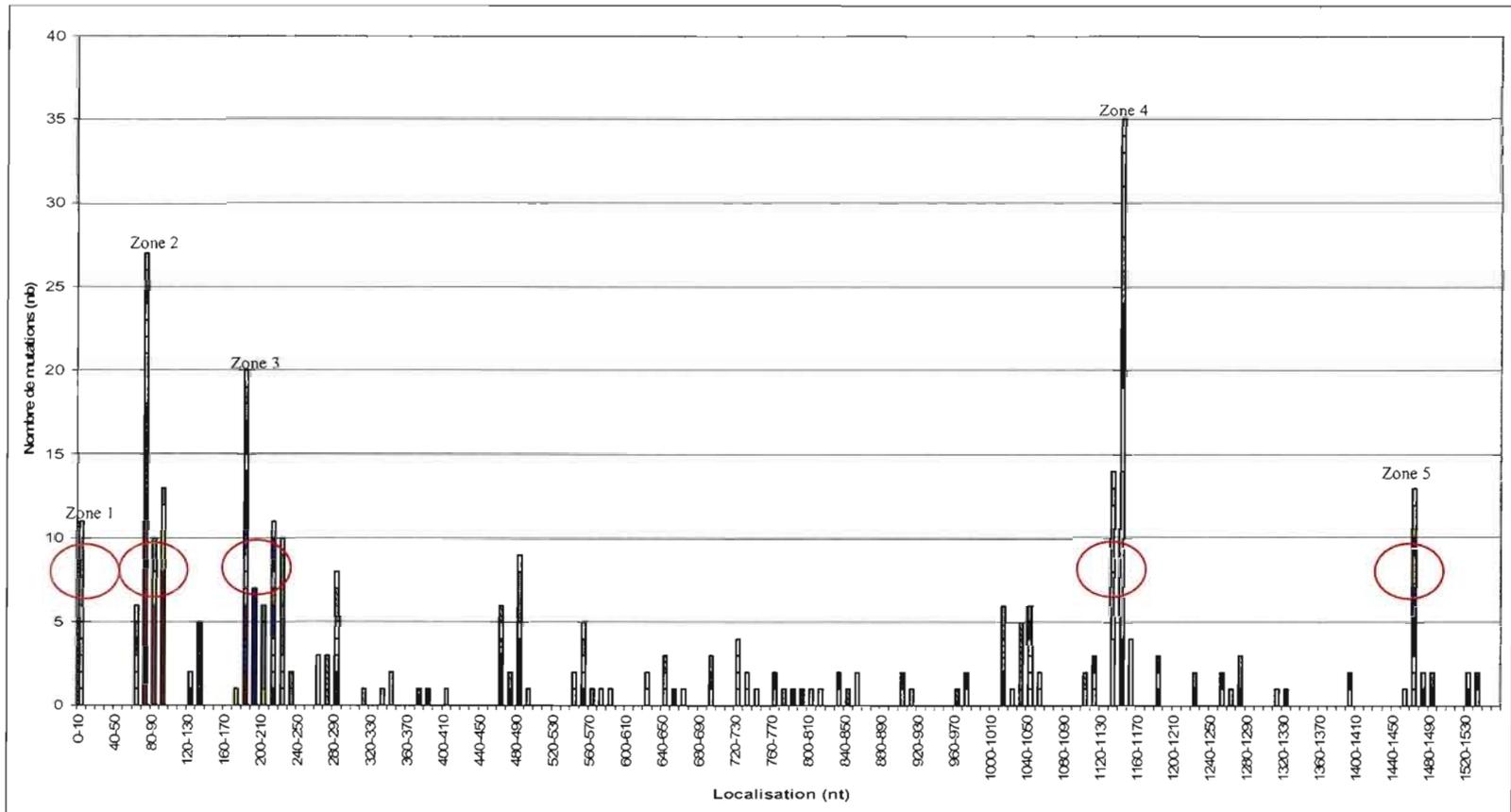


Figure 6.1 : Histogramme de distribution des mutations au sein du gène 16S pour les *Bacillaceae*.

L'abscisse donne la localisation par nombre de paire de base et l'ordonnée le nombre de mutations constatées.

□ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *B. licheniformis*, ■ *B. weihenstephanensis*, ■ *B. clausii*, ■ *B. sp*

○ Zone hypervariable.

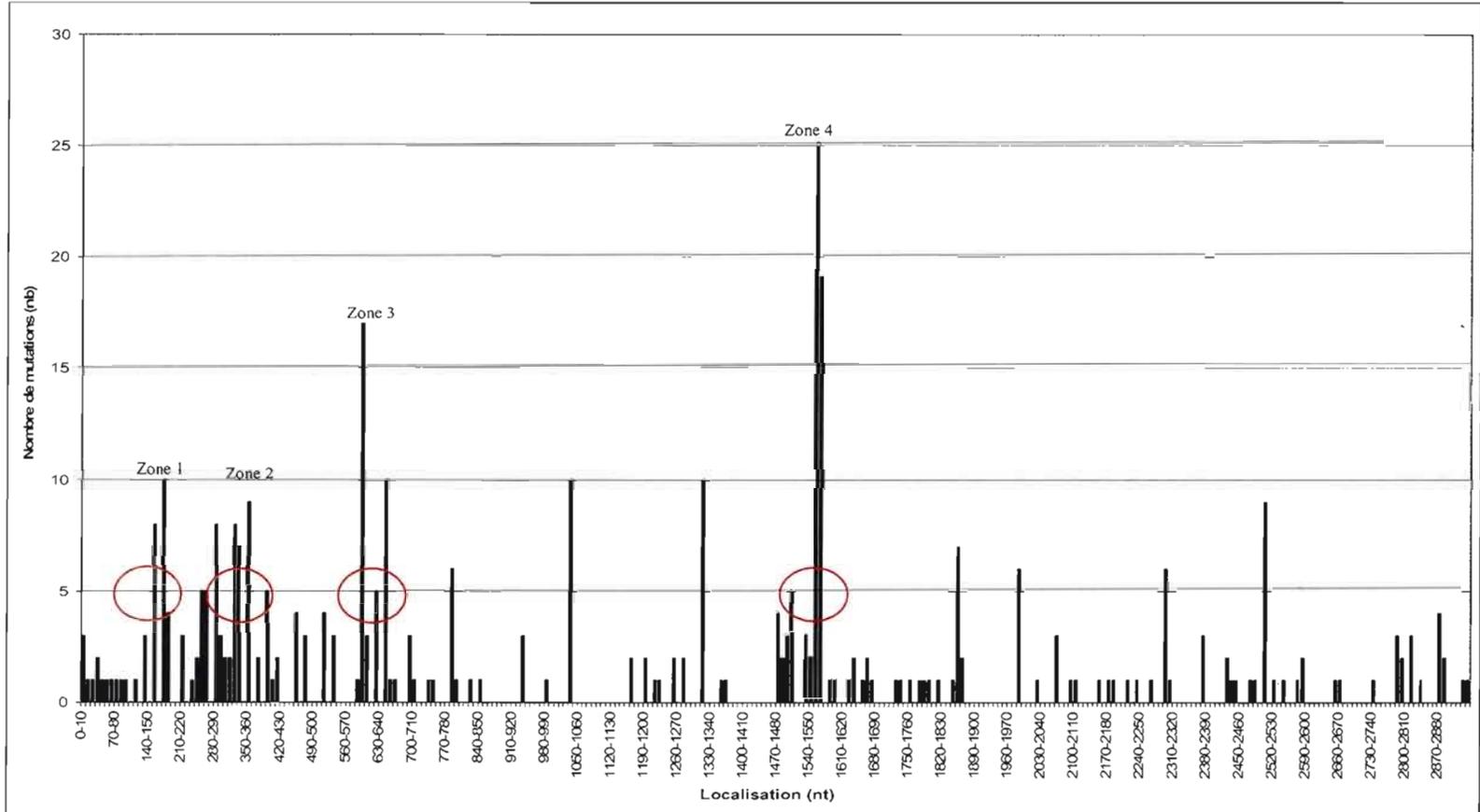


Figure 6.2 : Histogramme de distribution des mutations au sein du gène 23S pour les *Bacillaceae*.

L'abscisse donne la localisation par nombre de paire de base et l'ordonnée le nombre de mutations constatées.

□ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *B. licheniformis*, ■ *B. clausii*, ■ *B. sp.* ○ Zone hypervariable.

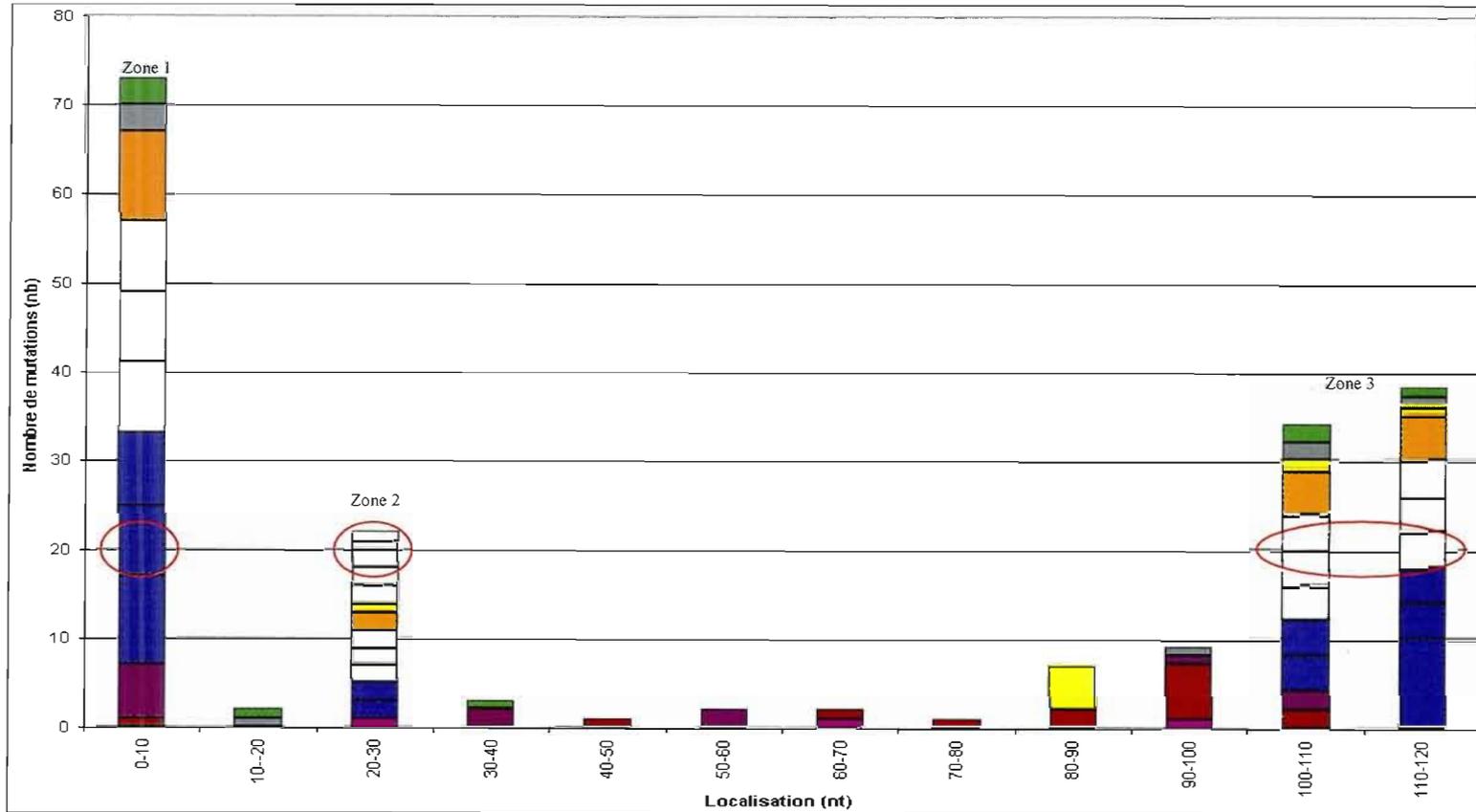


Figure 6.3 : Histogramme de distribution des mutations au sein du gène 5S pour les *Bacillaceae*.

L'abscisse donne la localisation par nombre de paire de base et l'ordonnée le nombre de mutations constatées.

□ *B. anthracis*, ■ *B. cereus*, ■ *B. thuringiensis*, ■ *B. subtilis*, ■ *B. halodurans*, ■ *B. licheniformis*, ■ *B. weihenstephanensis*, ■ *B. clausii*, ■ *B. methanolicus*.

○ Zone hypervariable.

Ces zones s'étendent entre 140 et 160 pb, 210 et 390 pb, 560 et 640 pb et finalement entre 1510 et 1590 pb pour le gène 23S. Pour le 5S, ces régions se situent entre 0 et 10 pb, 20 et 30 pb et 100 et 120 pb.

La distribution des mutations n'est donc pas uniforme au sein de chacun des gènes de l'opéron ribosomal des *Bacillaceae*. Il existe au sein de chaque gène des zones hypervariables ou "hot-spots" dans lesquelles se concentrent les mutations.

6.2. Distribution des mutations au sein du gène 16S pour les γ -protéobactéries :

Dans cette section, 169 allèles du gène 16S de 28 souches, appartenant à la sous division des γ -protéobactéries (Gupta, 2000), ont été considérés. Les caractéristiques de ces souches telles que le nom, le numéro d'accèsion et le nombre d'allèles pour le gène 16S sont listés dans le tableau A.4 de l'annexe. Toutes les souches considérées ont vu leur génome être complètement séquencé. Le nombre de copies du gène 16S varie de 2 à 9 au sein des γ -protéobactéries.

La figure 6.4 représente la distribution des mutations au sein du gène 16S pour les γ -protéobactéries étudiées. Cet histogramme révèle que la distribution des mutations n'est pas uniforme. Quatre régions peuvent être distinguées, au sein du gène 16S, dans lesquelles se concentrent les mutations. Ces zones sont situées entre 60 pb et 110 pb, 250 pb et 280 pb, 450 pb et 480 pb et finalement entre 1000 pb et 1060 pb.

Aussi bien, au sein de la famille des *Bacillaceae* que chez la sous division des γ -protéobactéries, la distribution des mutations n'est pas uniforme. De plus, il a été possible de mettre en évidence l'existence au sein du gène 16S de régions hypervariables avec une concentration élevée de mutations.

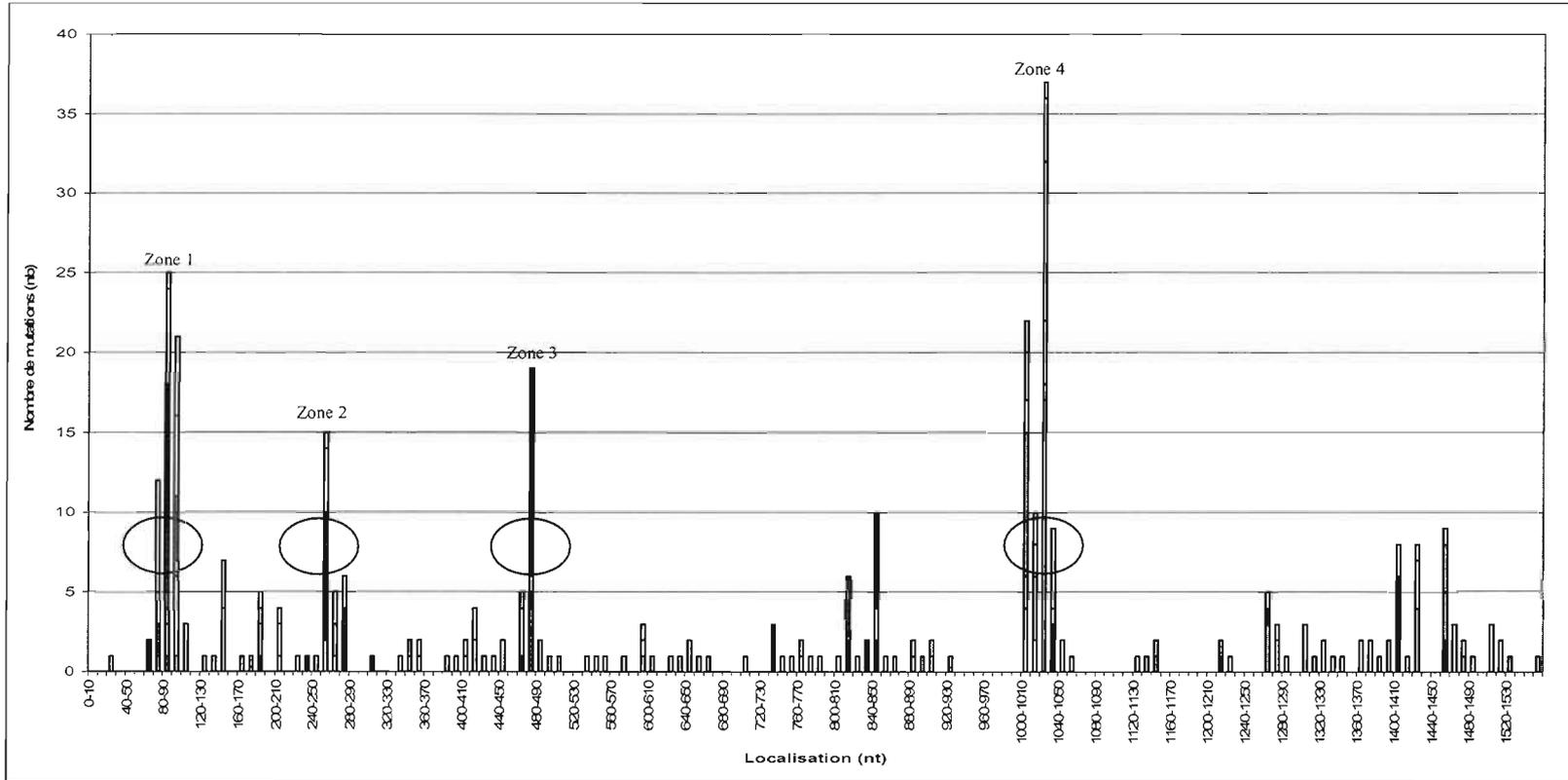


Figure 6.4 : Histogramme de distribution des mutations au sein du gène 16S pour 28 souches appartenant à la sous division des γ -protéobactéries.

L'abscisse donne la localisation par nombre de paire de base et l'ordonnée le nombre de mutations constatées.

Escherichia coli,
 shigella,
 Shewanella,
 Vibrio,
 Salmonella,
 Xylella,
 Xanthomonas,
 Haemophilus,
 Yersinia,
 Pasteurella,
 Pseudomonas,
 Wiggleworthia,
 Photorhabdus.
 Zone hypervariable.

Il serait intéressant d'analyser la conservation des ces zones hypervariables au sein des deux groupes (*Bacillaceae* et γ -protéobactéries) ainsi que chez d'autres groupes de bactéries.

6.3. Spécificités des régions hypervariables du gène 16S :

Dans le but d'apporter des éléments de réponse concernant la conservation des régions hypervariables, ce paragraphe est consacré à la comparaison entre les distributions des mutations au sein du gène 16S pour les *Bacillaceae* et les γ -protéobactéries.

La figure 6.5 représente l'histogramme de distribution des mutations au sein du gène 16S pour les 21 souches de *Bacillaceae* et 28 de γ -protéobactéries. Le nombre total d'allèles pour chaque groupe est de 188 et de 169 respectivement. Cet histogramme permet d'identifier huit régions dans lesquelles se concentrent les mutations. Ces zones sont situés entre 0 et 10 pb, 60 et 110 pb, 180 et 230 pb, 250 et 280 pb, 450 et 480 pb, 1000 et 1060 pb, 1130 et 1160 et finalement 1450 et 1480 pb. Seule la région située entre 60 et 100 pb semble être commune aux deux groupes. Toutes les autres sont spécifiques et coïncident avec celles qui ont été identifiées lors de l'analyse séparée de chaque groupe : les *Bacillaceae* et les γ -protéobactéries. Les régions hypervariables sont donc spécifiques à chaque groupe de bactéries.

6.4. Classification phylogénétique basée sur les régions hypervariables du gène 16S :

Les sections précédentes ont montré l'existence de régions hypervariables au sein du gène 16S qui sont spécifiques à chaque groupe de bactéries. Ces résultats suggèrent que l'utilisation de séquences concaténées des régions hypervariables des différents allèles du gène 16S pour

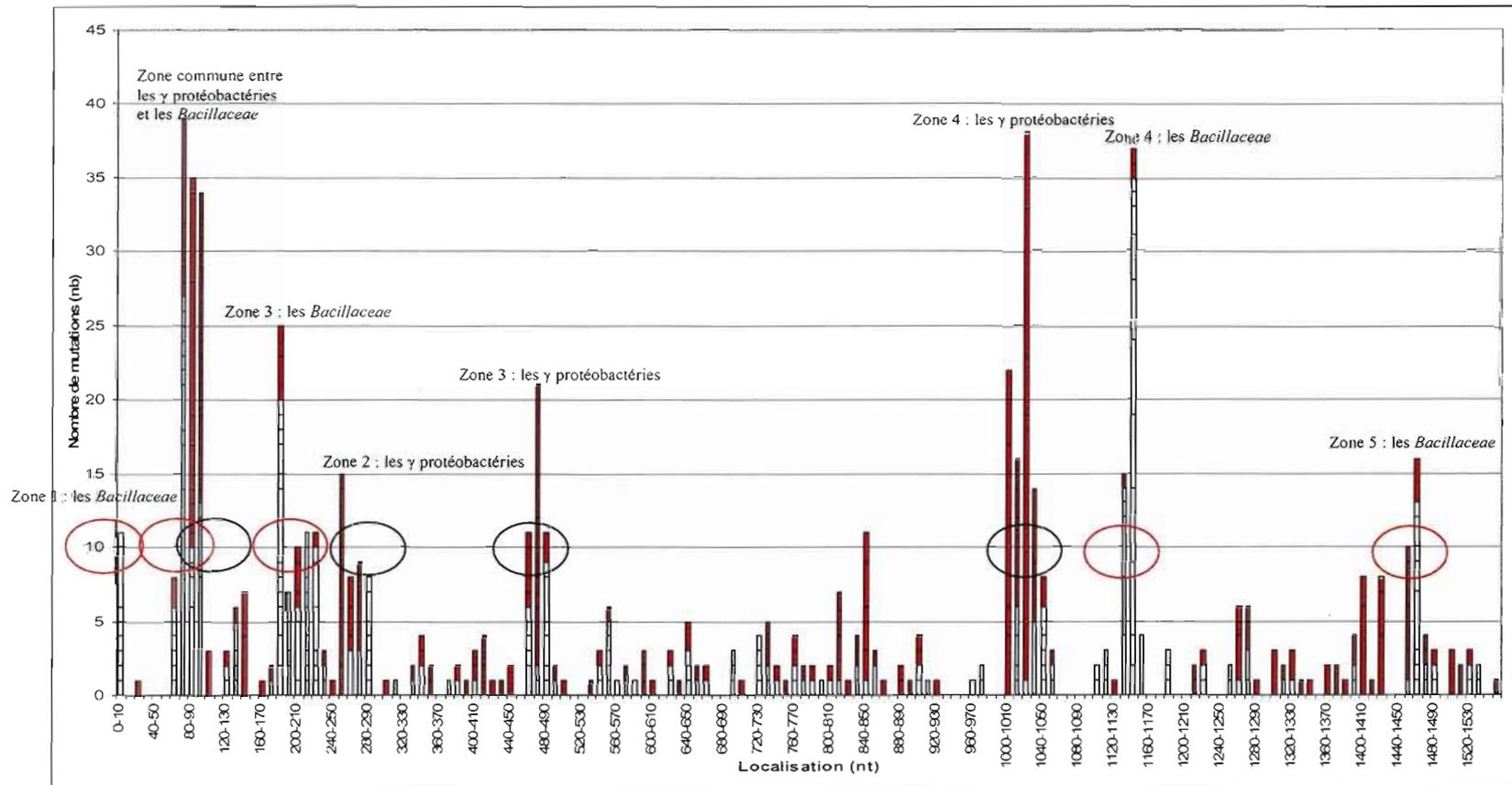


Figure 6.5 : Histogramme de distribution des mutations au sein du gène 16S pour les *Bacillaceae* et les γ -protéobactéries.

L'abscisse donne la localisation par nombre de paire de base et l'ordonnée le nombre de mutations constatées.

□ Les *Bacillaceae*, ■ Les Gamma protéobactéries; ○ Zone hypervariable des γ -protéobactéries ○ Zone hypervariable des *Bacillaceae*

établir une classification phylogénétique des *Bacillaceae* pourrait être une approche intéressante.

Les allèles des séquences concaténées des régions hypervariables du gène 16S de chacune des souches ont été téléchargés sous format FASTA à partir de "GenBank". L'alignement multiple de ces séquences nucléotidiques concaténées a été réalisé en utilisant le programme ClustalW. La phylogénie a été établie suivant la méthode de Neighbor-Joining (NJ, Saitou et Nei, 1987) par la suite informatique "Molecular Evolutionary Genetics Analysis" version 3 (MEGA3, Kumar *et al.*, 2004, <http://www.megasoftware.net/>) en utilisant les paramètres de Kimura des substitutions nucléotidiques multiples (Kimura 1980). L'arbre phylogénétique obtenu a été illustré à l'aide du programme MEGA3 TreeExplorer et sa fiabilité a été évaluée par le programme "Bootstrap". *Clostridium tetani* E88, a été utilisée comme "out group" comme dans les sections précédentes.

Les annotations des cinq régions hypervariables ont été calculées et sont données dans le tableau A.5 de l'annexe. La taille de la matrice générée par l'alignement multiple des séquences nucléotidiques concaténées du gène 16S est de 109 allèles x 160 nucléotides.

La figure 6.6 représente l'arbre phylogénétique des *Bacillaceae* obtenu à partir de la concaténation des séquences nucléotidiques des cinq régions hypervariables suivant la méthode NJ. Cette classification phylogénétique montre la répartition des *Bacillaceae* en trois groupes avec la subdivision du groupe II en deux sous-groupes.

Le groupe I englobe *B. anthracis* Ames, *B. anthracis* Ancestor, *B. anthracis* Sterne, *B. cereus* 10987, *B. cereus* 14579, *B. cereus* EL33 et *B. thuringiensis* serovar *konkukian* 97-27. Le groupe II renferme *B. licheniformis* 14580 et *B. subtilis* 168. Finalement le groupe III est formé par *B. halodurans* C-125.

Il est important de noter que la topologie de cet arbre est similaire à celle obtenue avec le gène 16S au complet. Ce résultat indique que ce sont les régions hypervariables qui sont responsables de la topologie des arbres phylogénétiques. Néanmoins, cette approche a été incapable d'apporter une meilleure discrimination entre les sept souches du groupe I c'est-à-dire le groupe *cereus* au sens large.

L'utilisation du gène 16S au complet pour établir une classification phylogénétique des *Bacillaceae*, pourrait être remplacée par l'analyse des régions hypervariables de ce gène.

6.5. Conclusion :

Les résultats de ce chapitre montrent que la distribution des mutations n'est pas uniforme au sein des trois gènes de l'opéron ribosomal (16S, 23S et 5S) pour les *Bacillaceae*. L'existence de régions hypervariables au sein de chacun de ces gènes est mise en évidence et la comparaison avec les γ -protéobactéries a permis de confirmer que ces zones sont spécifiques à chaque groupe.

L'analyse basée sur les séquences nucléotidiques concaténées des zones hypervariables du gène 16S a révélé que ce sont ces régions qui sont vraisemblablement responsables de la topologie de l'arbre phylogénétique. L'utilisation des séquences concaténées des zones hypervariables est suffisante pour l'analyse phylogénétique des *Bacillaceae*. Toutefois, cette approche n'est pas capable d'offrir une meilleure discrimination des espèces au sein du groupe *cereus* au sens large.

CHAPITRE VII

CONCLUSION

L'objectif de cette étude a été de déterminer si l'utilisation d'un seul allèle du gène 16S pour la classification phylogénétique des *Bacillaceae* est justifiée, ainsi que d'explorer si les autres gènes (ARN23S et ARN5S) de l'opéron ribosomal ou certains gènes de ménage représentent un meilleur choix de gènes. La distribution des mutations, la recherche de régions hypervariables au sein de l'opéron ribosomal et leur utilisation pour établir une classification phylogénétique des *Bacillaceae* figurent aussi parmi les objectifs fixés pour ce travail.

Lors de cette étude, un total de 109 allèles pour le gène 16S, 109 allèles pour le gène 23S, 105 allèles pour le gène 5S, ainsi que 109 allèles pour la région inter génique ITS1, 105 allèles pour la région ITS2, dix allèles pour chacun des trois gènes de ménage adénylate kinase (*adk*), shikimate 5-déshydrogénase (*aroE*) et la sous-unité B de l'ADN gyrase (*gyrB*) ont été téléchargés à partir de "GenBank" pour 10 souches de *Bacillaceae*. L'analyse phylogénétique de dix souches de *Bacillaceae* a été réalisée suivant la méthode de Neighbour-Joining (NJ) en utilisant les paramètres de Kimura. La fiabilité et la robustesse de la classification ont été évaluées à l'aide de la méthode "Bootstrap".

La comparaison des classifications phylogénétiques établies à partir de tous les allèles des gènes 16S, 23S et 5S de l'opéron ribosomal a montré que l'analyse phylogénétique des *Bacillaceae* pourrait être limitée à l'un des trois gènes de l'opéron. Aucune information supplémentaire n'est apportée par l'utilisation d'un gène en particulier. Les résultats obtenus répartissent les *Bacillaceae* en trois groupes avec la subdivision du groupe II en deux sous-groupes. L'utilisation des régions inter géniques (ITS1 et ITS2) n'a pas permis d'apporter des

informations supplémentaires ni de discriminer au sein du groupe *cereus* au sens large. Néanmoins, elle a permis de mettre en évidence les hétérogénéités alléliques.

Dans la deuxième partie de cette étude, une approche basée sur l'utilisation des séquences nucléotidiques des trois gènes de ménage suivants : adénylate kinase (*adk*), shikimate 5-déshydrogénase (*aroE*) et la sous-unité B de l'ADN gyrase (*gyrB*), ainsi que sur leurs séquences concaténées a été adoptée. Les phylogénies ainsi établies se sont avérées similaires entre elles et en accord avec la classification découlant de l'utilisation de l'opéron ribosomal. Toutefois, cette approche s'est révélée plus efficace et capable de discriminer au sein du groupe *cereus* au sens large. En effet, l'utilisation des séquences nucléotidiques des gènes de ménage a permis la répartition des *Bacillaceae* en trois groupes et la subdivision des groupes I et II en cinq et deux sous-groupes respectivement. L'analyse basée sur les séquences en acides aminés des protéines codant par les trois gènes de ménage ainsi que leurs séquences concaténées s'est avérée, quant à elle, une approche moins efficace. En effet, la discrimination s'est limitée à la répartition des *Bacillaceae* en trois groupes. La dégénérescence du code génétique est à l'origine de l'incapacité de cette approche de discriminer au sein du groupe *cereus* au sens large. Ceci s'est manifesté par un regroupement de plusieurs espèces au sein du même faisceau. L'utilisation des séquences nucléotidiques est donc préférable aux séquences en acides aminés pour l'étude d'espèces étroitement liées. Aucun avantage n'est apporté par le choix d'un gène en particulier.

L'hétérogénéité des allèles des trois gènes de l'opéron ribosomal (16S, 23S et 5S) a été analysée dans la dernière section de cette étude. Les résultats montrent que l'hétérogénéité dépend de la souche et de la taille du gène et peut atteindre dans certains cas des taux significatifs. Néanmoins, elle n'affecte pas de façon notable la classification phylogénétique des *Bacillaceae*. Une étude de la distribution des mutations au sein des gènes de l'opéron ribosomal a été réalisée pour un ensemble de 20 souches de la famille des *Bacillaceae* (aux dix souches initiales ont été rajoutés dix autres dont les génomes sont partiellement séquencés) et de 28 souches de la sous division des γ -protéobactéries. Les résultats montrent que cette distribution n'est pas uniforme au sein des différents gènes de l'opéron ribosomal

(16S, 23S et 5S) et qu'il existe des régions hypervariables ("hot-spots") dans lesquelles se concentrent les mutations. La comparaison avec les γ -protéobactéries a révélé que ces régions sont spécifiques à chaque groupe de bactéries. Les conclusions découlant de la classification phylogénétique des *Bacillaceae* basée les régions hypervariables 16S sont similaires à ceux de l'analyse basée sur les séquences nucléotidiques du gène au complet. La topologie des arbres semble découler de la structure de ces zones hypervariables. Il est donc suggéré de se limiter à l'utilisation des zones hypervariables du gène 16S au lieu du gène au complet.

ANNEXE

Tableau A.1 : Caractéristiques des *Bacillus* utilisés dans l'étude de la distribution des mutations au sein du gène 16S

Ce tableau donne le numéro d'accèsion pour chaque contiguë, le nom du gène, le nom suggéré pour différencier les allèles, le numéro de contiguë et la localisation du début et de la fin pour chaque gène tels qu'indiqués dans "GenBank". Les signes (+) et (-) indiquent le brin d'ADN porteur du gène.

Nom de la souche	Numéro d'accèsion	Nom du gène dans "GenBank"	Nom suggéré	Numéro de contiguë	Localisation dans "GenBank"	
<i>B. anthracis</i> Australia 94	NZ_AAES01000046	rRNA 16S	A	cont 57	1359	2913 (-)
	NZ_AAES01000045	rRNA 16S	B	cont 56	204394	205948 (+)
	NZ_AAES01000041	rRNA 16S	C	cont 52	11664	13218(+)
	NZ_AAES01000036	rRNA 16S	D	cont 47	1095	2650 (-)
	NZ_AAES01000043	rRNA 16S	E	cont 54	238974	240528(+)
	NZ_AAES01000037	rRNA 16S	F	cont48	2832	1278(-)
<i>B. anthracis</i> CNEVA-9066	NZ_AAEN01000023	rRNA 16S	A	cont 2232	21165	22719(+)
		rRNA 16S	B		31140	32694(+)
		rRNA 16S	C		273929	275484(+)
		rRNA 16S	D		584	2138(+)
	NZ_AAEN01000020	rRNA 16S	E	cont 2229	19682	21236(+)
	NZ_AAEN01000026	rRNA 16S	F	cont 2235	135	1692(-)
	NZ_AAEN01000021	rRNA 16S	G	cont 2230	20323	21876(+)
<i>B. anthracis</i> A1055	NZ_AAEO01000026	rRNA 16S	A	cont 241	144521	146075(+)
		rRNA 16S	B		164316	165870(+)
		rRNA 16S	C		217627	219181(+)
	NZ_AAEO01000038	rRNA 16S	D	cont 253	1468	3022 (-)
		rRNA 16S	E		cont 245	21275
	NZ_AAEO01000030	rRNA 16S	F		34575	36130(+)
		rRNA 16S	G		700	2255(+)
		rRNA 16S	H		44548	46103(+)
	NZ_AAEO01000032	rRNA 16S	I	cont 247	3157	1603 (-)
	NZ_AAEO01000034	rRNA 16S	J	cont 249	2073	3629 (-)
	NZ_AAEO01000035	rRNA 16S	K	cont 250	77822	79376(+)
<i>B. anthracis</i> Western North America USA6153	NZ_AAER01000036	rRNA 16S	A	cont 1971	12506	14061(+)
		rRNA 16S	B		22480	24034(+)
	NZ_AAER01000031	rRNA 16S	C	cont 1966	32040	33594(+)
	NZ_AAER01000029	rRNA 16S	D	cont 1964	112718	114272(+)
	NZ_AAER01000026	rRNA 16S	E	cont 1961	36349	37903(+)

Nom de la souche	Numéro d'accension	Nom du gène dans "GenBank"	Nom suggéré	Numéro de contiguité	Localisation dans "GenBank"	
<i>B. anthracis</i> Vollum	NZ_AAEP01000040	rRNA 16S	A	cont 49	185964	187518(+)
	NZ_AAEP01000047	rRNA 16S	B	cont 56	19746	21300(-)
	NZ_AAEP01000045	rRNA 16S	C	cont 54	6542	8096(+)
	NZ_AAEP01000043	rRNA 16S	D	cont 52	17147	18701(+)
	NZ_AAEP01000038	rRNA 16S	E	cont 47	1785	3339(-)
	NZ_AAEP01000036	rRNA 16S	F	cont 45	2702	4256(-)
	NZ_AAEP01000046	rRNA 16S	G	cont 55	239401	240955(+)
	NZ_AAEP01000044	rRNA 16S	H	cont 53	9887	11442(+)
	NZ_AAEP01000039	rRNA 16S	I	cont 48	534	2087(+)
<i>B. cereus</i> cytotoxis NVH 391-98	NZ_AALL01000081	rRNA 16S	A	cont 113	6671	8224(+)
	NZ_AALL01000002	rRNA 16S	B	cont 166	176013	177566(-)
	NZ_AALL01000007	rRNA 16S	C	cont 162	100331	101883(+)
	NZ_AALL01000033	rRNA 16S	D	cont 142	35011	36564(+)
	NZ_AALL01000003	rRNA 16S	E	cont 165	158910	160463(+)
<i>B. sp.</i> NRRL B-14911	NZ_AAOX01000008	rRNA 16S	A	cont 53141	74722	76275(+)
	NZ_AAOX01000029	rRNA 16S	B	cont 53144	65	1618(-)
<i>B. cereus</i> BGSC 6A5	AY224379	rRNA 16S	A		974	2527(+)
	AY224380	rRNA 16S	B		974	25727(+)
	AY224381	rRNA 16S	C		974	2527(+)
	AY224382	rRNA 16S	D		6682	8235(+)
	AY224383	rRNA 16S	E		974	2527(+)
	AY224384	rRNA 16S	F		2255	3808(+)
		rRNA 16S	G		974	2527(+)
	AY224385	rRNA 16S	H		8247	9800(+)
		rRNA 16S	I		974	2527(+)
		rRNA 16S	J		8081	9634(+)
	AY224386	rRNA 16S	K		974	2527(+)
	AY224387	rRNA 16S	L		974	2527(+)
	AY224388	rRNA 16S	M		973	2526(+)
	<i>B. cereus</i> G9241	NZ_AA EK01000138	rRNA 16S	A	cont1702	43
NZ_AA EK01000078		rRNA 16S	B	cont1902	63	1617(-)
NZ_AA EK01000017		rRNA 16S	C	cont1896	92582	94136(+)
		rRNA 16S	D	cont1896	72057	73611(+)
NZ_AA EK01000059		rRNA 16S	E	cont1870	5020	6574(+)
NZ_AA EK01000035		rRNA 16S	F	cont1897	49433	50988(+)
NZ_AA EK01000032		rRNA 16S	G	cont1889	38461	40015(+)
NZ_AA EK01000013		rRNA 16S	H	cont1894	10302	11856(+)

Nom de la souche	Numéro d'accèsion	Nom du gène dans "GenBank"	Nom suggéré	Numéro de contiguité	Localisation dans "GenBank"	
<i>B. clausii</i> KSM-K16	NC_006582	rRNA 16S A rRNA 16S B rRNA 16S C rRNA 16S D rRNA 16S E rRNA 16S F rRNA 16S G			11675 52364 136407 142939 218075 933638 1440581	13227 53916 137957 144489 219625 935188 1442131
<i>B. weihenstephanensis</i> KBAB4	NZ_AAOY01000004 NZ_AAOY01000016 NZ_AAOY01000012	rRNA 16S rRNA 16S rRNA 16S	A B C	cont 264 cont 258 cont 262	278 82716 92669	1832(-) 84270(+) 94223(+)

Tableau A.2 : Caractéristiques des *Bacillus* utilisés dans l'étude de la distribution des mutations au sein du gène 23S

Ce tableau donne le numéro d'accèsion pour chaque contiguë, le nom du gène, le nom suggéré pour différencier les allèles, le numéro de contiguë et la localisation du début et de la fin pour chaque gène tels qu'indiqués dans "GenBank". Les signes (+) et (-) indiquent le brin d'ADN porteur du gène.

Nom de la souche	Numéro d'accèsion	Nom du gène dans "GenBank"	Nom suggéré	Numéro de contiguë	Localisation dans "GenBank"	
<i>B. anthracis</i> Australia 94	NZ_AAES01000040.1	rRNA 23S	A	cont 51	65	2986 (+)
	NZ_AAES01000041.1	rRNA 23S	B	cont 52	65	2986 (+)
<i>B. anthracis</i> CNEVA-9066	NZ_AAEN01000023.1	rRNA 23S	A	cont 2232	2283	5204 (+)
		rRNA 23S	B		32839	35760(+)
		rRNA 23S	C		22865	25786(+)
		rRNA 23S	D		75	2996 (+)
		rRNA 23S	E		22022	24943 (+)
		rRNA 23S	F		1815	4736 (+)
<i>B. anthracis</i> A1055	NZ_AAEO01000026.1	rRNA 23S	A	cont 241	146447	149368(+)
		rRNA 23S	B		166242	169164(+)
	NZ_AAEO01000030.1	rRNA 23S	C	cont 245	219326	222247(+)
		rRNA 23S	D		46248	49170(+)
		rRNA 23S	E		36275	39195(+)
		rRNA 23S	F		2400	5321(+)
		rRNA 23S	G		22974	25895(+)
	NZ_AAEO01000035.1	rRNA 23S	H	cont 250	7952	82442(+)
	<i>B. anthracis</i> Western North America USA6153	NZ_AAER01000029.1	rRNA 23S	A	cont1964	11441
rRNA 23S			B	cont1971	14206	17127(+)
NZ_AAER01000036.1		rRNA 23S	C		24181	27103 (+)
		rRNA 23S	D		904	3827 (+)
		rRNA 23S	E	cont1966	33740	36661(+)
<i>B. anthracis</i> Vollum	NZ_AAEP01000047.	rRNA 23S	A	cont56	72696	69775 (-)
		rRNA 23S	B		19601	16680(-)
<i>B. cereus</i> BGSC 6A5	AY224379	rRNA 23S	A		2894	5819
	AY224380	rRNA 23S	B		2894	5819
	AY224381	rRNA 23S	C		2668	5593
		rRNA 23S	D		8376	11301
	AY224382	rRNA 23S	E		2668	5593
	AY224383	rRNA 23S	F		3948	6873

Nom de la souche	Numéro d'accesion	Nom du gène dans "GenBank"	Nom suggéré	Numéro de contiguïté	Localisation dans "GenBank"	
<i>B. cereus</i> BGSC 6A5	AY224384	rRNA 23S	G		2668	5593
	AY224385	rRNA 23S	H		9941	12866
		rRNA 23S	I		2668	5593
	AY224386	rRNA 23S	J		9775	12700
		rRNA 23S	K		2667	5592
	AY224387	rRNA 23S	L		2667	5592
AY224388	rRNA 23S	M		2666	5591	
<i>B. cereus</i> G9241	NZ_AAEK01000059	rRNA 23S	A	cont1870	1006	3927(+)
	NZ_AAEK01000017	rRNA 23S	B		6718	9639(+)
		rRNA 23S	C	cont1896	73756	76677(+)
	NZ_AAEK01000002	rRNA 23S	D		94280	97201(+)
		rRNA 23S	E	cont1905	1006	3927(+)
	NZ_AAEK01000013	rRNA 23S	F	cont1894	12000	14921(+)
NZ_AAEK01000032	rRNA 23S	G	cont1889	40385	43306(+)	
<i>B. clausii</i> KSM-K16	NC_006582	rRNA 23S A			13559	16361(+)
		rRNA 23S B			54372	57174(+)
		rRNA 23S C			138412	141215(+)
		rRNA 23S D			144821	147624(+)
		rRNA 23S E			220080	222882(+)
		rRNA 23S F			935921	938723(+)
		rRNA 23S G			1442780	1445583(+)
<i>B. cereus</i> cytotoxis NVH 391-98	NZ_AALL01000007		A	cont 62	102129	105049(-)
	NZ_AALL01000002		B	cont 66	175767	172846(+)
<i>B. sp.</i> NRRL B-14911	NZ_AAOX01000040		A	53102	932	3875(+)
	NZ_AAOX01000052		B	53099	10732	13673(-)
	NZ_AAOX01000021		C	53143	88331	91274(-)
	NZ_AAOX01000068		D	53134	932	3876(+)
	NZ_AAOX01000008		E	53141	76565	79498(+)

Tableau A.3 : Caractéristiques des *Bacillus* utilisé dans l'étude de la distribution des mutations au sein du gène 5S

Ce tableau donne le numéro d'accèsion pour chaque contiguë, le nom du gène, le nom suggéré pour différencier les allèles, le numéro de contiguë et la localisation de début et de fin pour chaque gène tels qu'indiqués dans "GenBank". Les signes (+) et (-) indiquent le brin d'ADN porteur du gène.

Souche	Numéro d'accèsion	Nom du gène dans "GenBank"	Nom suggéré	Numéro de contiguë	Localisation dans "GenBank"	
<i>B. anthracis</i> Australia 94	NZ_AAES01000019	rRNA 5S	A	cont 30	596	711(+)
	NZ_AAES01000021	rRNA 5S	B	cont 32	32	147(+)
	NZ_AAES01000022	rRNA 5S	C	cont 33	633	748(+)
<i>B. anthracis</i> CNEVA-9066	NZ_AAEN01000019	rRNA 5S	A	cont 2228	2018	2133(+)
	NZ_AAEN01000003	rRNA 5S	B	cont 2202	384	499(+)
	NZ_AAEN01000022	rRNA 5S	C	cont 2231	3098	3213(+)
	NZ_AAEN01000027	rRNA 5S	D	cont 2236	263612	263727(-)
<i>B. anthracis</i> A1055	NZ_AAEO01000024	rRNA 5S	A	cont 239	2589	2704(+)
	NZ_AAEO01000023	rRNA 5S	B	cont 238	95639	95754(-)
	NZ_AAEO01000031	rRNA 5S	C	cont 246	263223	263338(-)
	NZ_AAEO01000033	rRNA 5S	D	cont 248	446912	447027(-)
<i>B. weihenstephanensis</i> KBAB4	NZ_AAOY01000015	rRNA 5S	A	cont 255	88730	88845 (-)
	NZ_AAOY01000012	rRNA 5S	B	cont 262	679	794 (+)
	NZ_AAOY01000129	rRNA 5S	C	cont 165	7538	7423 (-)
	NZ_AAOY01000002	rRNA 5S	D	cont 265	243384	243269 (-)
	NZ_AAOY01000031	rRNA 5S	E	cont 238	54203	54318(-)
	NZ_AAOY01000182	rRNA 5S	F	cont 85	594	709(+)
<i>B. anthracis</i> Western North America USA6153	NZ_AAER01000041	rRNA 5S	A	cont 1976	151452	151567 (-)
	NZ_AAER01000035	rRNA 5S	B	cont 1970	1166062	1166177(+)
	NZ_AAER01000025	rRNA 5S	C	cont 1960	20	135(+)
<i>B. anthracis</i> Kruger B	NZ_AAEQ01000037	rRNA 5S	A	cont 1008	1902	2017(+)
	NZ_AAEQ01000020	rRNA 5S	B	cont 991	94659	94544(-)
	NZ_AAEQ01000031	rRNA 5S	C	cont 1002	263028	263143(-)
	NZ_AAEQ01000022	rRNA 5S	D	cont 993	516	631(+)
	NZ_AAEQ01000005	rRNA 5S	E	cont 951	2	117(-)
<i>B. cereus</i> G9241	NZ_AAEK01000169	rRNA 5S	A	cont 1757	85	200(-)
	NZ_AAEK01000164	rRNA 5S	B	cont 1735	38	153(-)
	NZ_AAEK01000123	rRNA 5S	C	cont 1901	860	975(-)
	NZ_AAEK01000059	rRNA 5S	D	cont 1870	4029	4144(+)
	NZ_AAEK01000013	rRNA 5S	E	cont 1894	15022	15137(+)

Souche	Numéro d'accension	Nom du gène dans "GenBank"	Nom suggéré	Numéro de contiguïté	Localisation dans "GenBank"	
<i>B. cereus</i> BGSC 6A5	AY224379	rRNA 5S	A		5868	5983(+)
	AY224380	rRNA 5S	B		5868	5983(+)
	AY224381	rRNA 5S	C		5694	5809(+)
		rRNA 5S	D		11351	11461(+)
	AY224382	rRNA 5S	E		5684	5799(+)
	AY224383	rRNA 5S	F		6922	7037(+)
	AY224384	rRNA 5S	G		5642	5757(+)
		rRNA 5S	H		12915	13030(+)
	AY224385	rRNA 5S	I		5641	5756(+)
		rRNA 5S	J		12748	12863(+)
	AY224386	rRNA 5S	K		5692	5807(+)
	AY224387	rRNA 5S	L		5694	5809(+)
AY224388	rRNA 5S	M		5692	5809(+)	
<i>B. methanolicus</i>	Z11826	rRNA 5S	A		1	116(+)
	Z11821	rRNA 5S	B		1	116(+)
	Z11818	rRNA 5S	C		1	116(+)
	Z11827	rRNA 5S	D		1	116(+)
	Z11816	rRNA 5S	E		1	116(+)
	Z11817	rRNA 5S	F		1	116(+)
	Z11819	rRNA 5S	G		1	116(+)
<i>B. clausii</i> KSM-K16	NC_006582		A		16733	16848(+)
			B		57448	57563(+)
			C		141700	141815(+)
			D		148117	148232(+)
			E	O	223378	223493(+)
			F		263227	263342(+)
			G		939096	939211(+)
				1446077	1446192(+)	

Tableau A.4 : Caractéristiques des γ -protéobactéries utilisées dans l'étude de la distribution des mutations au sein du gène 16S

Ce tableau donne le numéro d'accèsion pour le génome complet, le nom du gène dans "GenBank", le nom suggéré pour différencier les allèles et la localisation du début et de la fin pour chaque gène tels qu'indiqués dans "GenBank", ainsi que les corrections lorsque nécessaire. Les signes (+) et (-) indiquent le brin d'ADN porteur du gène.

Souches	Numéro d'accèsion	Nom du gène	Nom du gène dans "GenBank"	Nom suggéré	Localisation dans "GenBank"		Localisation corrigée	
<i>Aleromonadaceae</i> <i>Shewanella oneidensis</i> MR-1	NC_004347	16S rRNA	<i>Sp16SA</i>		269814	271342 (+)	269805	271347
		16S rRNA	<i>Sp16SB</i>		216788	218316 (+)	216779	218321
		16S rRNA	<i>Sp16SC</i>		46116	47644 (+)	46107	47649
		16S rRNA	<i>Sp16SD</i>		4945901	4947429 (-)	4945896	4947438(-)
		16S rRNA	<i>Sp16SE</i>		4691979	4693507 (-)	4691974	4693516 (-)
		16S rRNA	<i>Sp16SF</i>		4682869	4684397 (-)	4682864	4684406 (-)
		16S rRNA	<i>Sp16SG</i>		4380831	4382359 (-)	4380826	4382368 (-)
		16S rRNA	<i>Sp16SH</i>		3738378	3739906 (-)	3738373	3739915 (-)
		16S rRNA	<i>Sp16SI</i>		2973049	2974577 (-)	2973044	2974586 (-)
<i>Enterobacteriales</i> <i>Escherichia coli</i> K12	NC_000913	16S rRNA	<i>rrsA</i>		4033120	4034661(+)		
		16S rRNA	<i>rrsB</i>		4164238	4165779(+)		
		16S rRNA	<i>rrsC</i>		3939431	3940971(+)		
		16S rRNA	<i>rrsD</i>		3424858	3426399(-)		
		16S rRNA	<i>rrsE</i>		4205725	4207266(+)		
		16S rRNA	<i>rrsG</i>		2727636	2729178(-)		
<i>Escherichia coli</i> O157:H7 EDL933	NC_002655	16S rRNA	<i>rrsA</i>		4900418	4901959(+)		
		16S rRNA	<i>rrsB</i>		5044690	5046231(+)		
		16S rRNA	<i>rrsC</i>		4804233	4805773(+)		
		16S rRNA	<i>rrsD</i>		4229456	4230997(-)		
		16S rRNA	<i>rrsE</i>		5085699	5087240(+)		
		16S rRNA	<i>rrsG</i>		3519578	3521124 (-)		
		16S rRNA	<i>rrsH</i>		227103	228644 (+)		
<i>Escherichia coli</i> O157:H7 VT2-Sakai	NC_002695	16S rRNA	<i>rrsA</i>		4831654	4833195(+)		
		16S rRNA	<i>rrsB</i>		4975927	4977468 (+)		
		16S rRNA	<i>rrsC</i>		4735252	4736793(+)		
		16S rRNA	<i>rrsD</i>		4162234	4163775(-)		
		16S rRNA	<i>rrsE</i>		5016953	5018494 (+)		
		16S rRNA	<i>rrsG</i>		3449735	3451276 (-)		
		16S rRNA	<i>rrsH</i>		227102	228643 (+)		
<i>Escherichia coli</i> CFT073	NC_004431	16S rRNA	<i>rrsA</i>		4561194	4562736 +)		
		16S rRNA	<i>rrsB</i>		4699063	4700613(+)		

Souches	Numéro d'accèsion	Nom du gène	Nom du gène dans "GenBank"	Nom suggéré	Localisation dans "GenBank"		Localisation corrigée	
<i>Escherichia coli</i> CFT073	NC_004431	16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA	<i>rrsC</i> <i>rrsD</i> <i>rrsE</i> <i>rrsG</i> <i>rrsH</i>		4442840 3858590 4739329 2992309 235186	4444382 (+) 3860131 (-) 4740870 (+) 2993859 (-) 236727 (+)		
<i>Photobacterium luminescens</i> TTO1	NC_005126	16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA	16s_rRNA 16s_rRNA 16s_rRNA 16s_rRNA 16s_rRNA 16s_rRNA 16s_rRNA	1 2 3 4 5 6 7	58438 536233 801957 1472782 5143181 5473370 5509393	59982 (+) 537777 (+) 803501 (+) 1474326 (+) 5144725 (-) 5474914 (-) 5510937 (-)		
<i>Salmonella typhimurium</i> LT2 SGSC1412	NC_003197	16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA	<i>rrsA</i> <i>rrsB</i> <i>rrsC</i> <i>rrsD</i> <i>rrsE</i> <i>rrsG</i> <i>rrsH</i>		4196059 4351130 4100132 3570463 4394675 2800118 289189	4197600 (+) 4352673 (+) 4101675 (+) 3572006 (-) 4396219 (+) 2801660 (-) 290732 (+)	4351130 4100132 3570466 4394675 2800119 289179	4352671(+) 4101673(+) 3572007(-) 4396217(+) 2801660(-) 290720(+)
<i>Salmonella enterica</i> serovar Typhi CT18	NC_003198	16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA	16s_rRNA 16s_rRNA 16s_rRNA 16s_rRNA 16s_rRNA 16s_rRNA 16s_rRNA	A B C D E G H	3598527 3747528 3556311 3421900 4257492 2715999 287479	3600068 (-) 3749069 (-) 3557853 (-) 3423441 (-) 4259033 (+) 2717540 (-) 289020 (+)		
<i>Salmonella enterica</i> serovar Typhi Ty2	NC_004631	16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA	<i>rrsA</i> <i>rrsB</i> <i>rrsC</i> <i>rrsD</i> <i>rrsE</i> <i>rrsG</i> <i>rrsH</i>	1 2 3 4 5 6 7	287477 2691223 3407559 3541972 3584188 3733020 4242149	289010 (+) 2692757 (-) 3409092 (-) 3543505 (-) 3585721 (-) 3734553 (-) 4243682 (+)	287470 2691222 2407558 3541971 3584187 3733019 4242142	289011(+) 2692764(-) 3409099(-) 3543512(-) 3585728(-) 3734560(-) 4243683(+)
<i>Shigella flexneri</i> 2a 2457T	NC_004741	16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA 16S rRNA	<i>rrsA</i> <i>rrsB</i> <i>rrsC</i> <i>rrsD</i> <i>rrsE</i> <i>rrsG</i> <i>rrsH</i>		3723289 3585007 3820462 3400184 3544571 2720026 214156	3724830 (-) 3586548 (-) 3822003 (-) 3401730 (-) 3546112 (-) 2721566 (-) 215696 (+)		

Souches	Numéro d'accèsion	Nom du gène	Nom du gène dans "GenBank"	Nom suggéré	Localisation dans "GenBank"		Localisation corrigée	
<i>Shigella flexneri</i> 2a .301	NC_004337	16S rRNA	<i>rrsA</i>		4048482	4050023 (+)		
		16S rRNA	<i>rrsB</i>		4186671	4188212 (+)		
		16S rRNA	<i>rrsC</i>		3951291	3952832 (+)		
		16S rRNA	<i>rrsD</i>		3410015	3411556 (-)		
		16S rRNA	<i>rrsE</i>		4227109	4228650 (+)		
		16S rRNA	<i>rrsG</i>		2726674	2728214 (-)		
<i>Wigglesworthia glossinidia</i> <i>brevipalpis</i>	NC_004344	16S rRNA	<i>rrsH</i>	<i>rrsH+</i>	136582	138132 (+)		
		16S rRNA	<i>rrsH</i>	<i>rrsH-</i>	684192	685742 (-)		
<i>Yersinia pestis</i> CO92	NC_003143	16S rRNA	<i>16s_rRNA</i>	001	12292	13763 (+)	12265	12807
		16S rRNA	<i>16s_rRNA</i>	002	1217505	1218993 (+)	1217478	1219020
		16S rRNA	<i>16s_rRNA</i>	003	3652660	3654148 (-)	3652633	3654175
		16S rRNA	<i>16s_rRNA</i>	004	4178944	4180432 (-)	4178917	4180459
		16S rRNA	<i>16s_rRNA</i>	005	4221808	4223296 (-)	4221781	4223323
		16S rRNA	<i>16s_rRNA</i>	006	4388245	4389733 (-)	4388218	4389760
<i>Yersinia pestis</i> KIM	NC_004088	16S rRNA		<i>yr001</i>	12016	13600 (+)	12015	13557
		16S rRNA		<i>yr005</i>	355930	357514 (+)	355929	357471
		16S rRNA		<i>yr008</i>	522393	523977 (+)	522392	523934
		16S rRNA		<i>yr011</i>	565257	566841 (+)	565256	566798
		16S rRNA		<i>yr014</i>	1024389	1025973 (+)	1024388	1025930
		16S rRNA		<i>yr020</i>	3415745	3417329 (-)	3415788	3417330
<i>Pasteurellaceae</i>	NC_000907	16S rRNA	<i>HlrrnA16S</i>		623825	625364 (+)		
		16S rRNA	<i>HlrrnB16S</i>		657107	658646 (+)		
		16S rRNA	<i>HlrrnC16S</i>		771212	772750 (+)	771210	772749
		16S rRNA	<i>HlrrnD16S</i>		1820465	1822003 (-)	1820460	1821999
		16S rRNA	<i>HlrrnE16S</i>		127176	128715 (-)		
		16S rRNA	<i>HlrrnF16S</i>		246013	247552 (-)		
<i>Haemophilus ducreyi</i> 35000HP	NC_002940	16S rRNA		1	10643	12179 (+)	10642	12182
		16S rRNA		2	240156	241692 (+)	240155	241695
		16S rRNA		3	489319	490855 (+)	489318	490858
		16S rRNA		4	616741	618277 (+)	616740	618280
		16S rRNA		5	1539720	1541256 (-)	1539717	1541257
		16S rRNA		6	1604540	1606076 (-)	1604537	1606077
<i>Pasteurella multocida</i> PM70	NC_002663	16S rRNA		<i>Pm1</i>	341429	342970 (+)		
		16S rRNA		<i>Pm2</i>	541958	543499 (+)		
		16S rRNA		<i>Pm3</i>	1080340	1081882 (-)		
		16S rRNA		<i>Pm4</i>	1690576	1692118 (+)		
		16S rRNA		<i>Pm5</i>	1761353	1762894(+)		
		16S rRNA		<i>Pm6</i>	1941231	1942772 (-)		

Souches	Numéro d'accèsion	Nom du gène	Nom du gène dans "GenBank"	Nom suggéré	Localisation dans "GenBank"		Localisation corrigée	
<i>Pseudomonadaceae</i>								
<i>Pseudomonas aeruginosa</i> PAO1	NC_002516	16S rRNA		1	722096	723631 (+)		
		16S rRNA		2	4792196	4793731 (-)		
		16S rRNA		3	5267724	5269259 (-)		
		16S rRNA		4	6043208	6044743 (-)		
<i>Pseudomonas syringae</i> DC3000	NC_004578	16S rRNA	<i>Ps16SA</i>		666741	668258 (+)	666727	668265
		16S rRNA	<i>Ps16SB</i>		827555	829072 (+)	827541	829079
		16S rRNA	<i>Ps16SC</i>		1072092	1073609 (+)	1072078	1073616
		16S rRNA	<i>Ps16SD</i>		3873151	3874668 (-)	3873144	3874682
		16S rRNA	<i>Ps16SE</i>		6217622	6219139 (-)	6217615	6219153
<i>Vibrionales</i>								
<i>Vibrio cholerae</i> El Tor N16961	NC_002505	16S rRNA	<i>16Sa</i>		53823	55357 (+)	53816	55358
		16S rRNA	<i>16Sb</i>		151059	152593 (+)	15105	152594
		16S rRNA	<i>16Sc</i>		324147	325181 (+)	324140	325682
		16S rRNA	<i>16Sd</i>		401751	403286 (+)	401744	403287
		16S rRNA	<i>16Se</i>		762775	764309 (+)	762768	764310
		16S rRNA	<i>16Sf</i>		2679884	2681418 (-)	2679883	2681425
		16S rRNA	<i>16Sg</i>		2931745	2933279 (-)	2931744	2933286
		16S rRNA	<i>16Sh</i>		2937466	2939001 (-)	2937465	2939008
		<i>Vibrio parahaemolyticus</i> RIMD 2210633	NC_004603	16S rRNA	<i>16Sa</i>		33633	35103 (+)
16S rRNA	<i>16Sb</i>				581626	583096 (+)	581624	583176
16S rRNA	<i>16Sc</i>				2779817	2781287 (-)	2779737	2781289
16S rRNA	<i>16Sd</i>				2885630	2887100 (-)	2885550	2887102
16S rRNA	<i>16Se</i>				3066302	3067772 (-)	3066222	3067774
16S rRNA	<i>16Sg</i>				3071548	3073018 (-)	3071468	3073020
16S rRNA	<i>16Sh</i>				3131721	3133191 (-)	3131641	3133193
16S rRNA	<i>16Si</i>				3136967	3138437 (-)	3136887	3138439
16S rRNA	<i>16Sj</i>				3196944	3198414 (-)	3196865	3198416
16S rRNA	<i>16Sk</i>				3238281	3239751 (-)	3238201	3239753
<i>Vibrio vulnificus</i> CMCP6	NC_004459			16S rRNA		<i>16Sa</i>	475699	477241 (-)
		16S rRNA		<i>16Sb</i>	932194	933736 (-)		
		16S rRNA		<i>16Sc</i>	937697	939239 (-)		
		16S rRNA		<i>16Sd</i>	978401	979943 (-)		
				<i>16Se</i>	1060773	1062317 (+)		
				<i>16Sf</i>	1170287	1171829 (+)		
				<i>16Sg</i>	1389748	1391290 (+)		
				<i>16Sh</i>	1478333	1479875 (+)		

Souches	Numéro d'accèsion	Nom du gène	Nom du gène dans "GenBank"	Nom suggéré	Localisation dans "GenBank"		Localisation corrigée	
<i>Vibrio vulnificus</i> YJ016	NC_004459	16S rRNA	<i>rRNA-16Sa</i>		32728	34262 (+)	32721	34263
		16S rRNA	<i>rRNA-16Sb</i>		177895	179429 (+)	177888	179430
		16S rRNA	<i>rRNA-16Sc</i>		717240	718774 (+)	717233	718775
		16S rRNA	<i>rRNA-16Sd</i>		2946932	2948466 (-)	2946931	2948473
		16S rRNA	<i>rRNA-16Se</i>		2952670	2954200 (-)	2952669	2954207
		16S rRNA	<i>rRNA-16Sg</i>		3041406	3042940 (-)	3041405	3042947
		16S rRNA	<i>rRNA-16Sh</i>		3261628	3263162 (-)	3261627	3263169
		16S rRNA	<i>rRNA-16Si</i>		3303433	3304967 (-)	3303432	3304974
<i>Xanthomonadales</i>								
<i>Xylella fastidiosa</i> Temecula 1	NC_004556	16S rRNA		<i>PD0048</i>	66041	67585 (+)		
		16S rRNA		<i>PD0133</i>	171124	172668 (+)		
<i>Xylella fastidiosa</i> 9a5c	NC_002488	16S rRNA	<i>XFrnaA-1</i>		66558	68102 (+)		
		16S rRNA	<i>XFrnaA-2</i>		172274	173818 (+)		
<i>Xanthomonas</i> group								
<i>Xanthomonas campestris</i> pv. <i>campestris</i> ATCC33913	NC_003902	16S rRNA		<i>SSU1</i>	4561295	4562841 (-)		
		16S rRNA		<i>SSU2</i>	4949163	4950709 (-)		
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> 306	NC_003919	16S rRNA		<i>SSU1</i>	4580055	4581601 (-)		
		16S rRNA		<i>SSU2</i>	5069297	5070843 (-)		

Tableau A.5 : Caractéristiques des régions hypervariables du gène 16S pour les dix *Bacillus* étudiés ainsi que pour "l'out group"

Ce tableau donne le nom de la souche, le numéro d'accèsion pour chaque souche, le nombre d'allèles de l'ARNr16S et les annotations de chaque zone. Les signes (+) et (-) indiquent le brin d'ADN porteur du gène.

Souches	Numéro d'accèsion	Copies ARNr16S		Annotation zone 1 (10 pb)	Annotation zone 2 (30 pb)	Annotation zone 3 (50 pb)	Annotation zone 4 (30 pb)	Annotation zone 5 (30 pb)
<i>B. anthracis</i> Ames	NC_003997	<i>16S rRNA A</i>	(+)	9307 9317	9367 9407	9487 9537	10437 10467	10757 10787
		<i>16S rRNA B</i>	(+)	29101 29111	29161 29201	29281 29331	30231 30261	30551 30581
		<i>16S rRNA C</i>	(+)	82423 82433	82483 82523	82603 82653	83553 83583	83873 83903
		<i>16S rRNA D</i>	(+)	145487 145497	145547 145587	145667 145717	146617 146647	146937 146967
		<i>16S rRNA E</i>	(+)	246207 246217	246267 246307	246387 246437	247337 247367	247657 247687
		<i>16S rRNA F</i>	(+)	266786 266796	266846 266886	266966 267016	267916 267946	268236 268266
		<i>16S rRNA G</i>	(+)	280084 280094	280144 280184	280264 280314	281214 281244	281534 281564
		<i>16S rRNA H</i>	(+)	290057 290067	290117 290157	290237 290287	291187 291217	291507 291537
		<i>16S rRNA I</i>	(+)	532435 532445	532495 532535	532615 532665	533565 533595	533885 533915
		<i>16S rRNA J</i>	(+)	741536 741546	741596 741636	741716 741766	742666 742696	742986 743016
		<i>16S rRNA K</i>	(-)	4656745 4656735	4656685 4656645	4656565 4656515	4655295 4655265	4655295 4655265
<i>B. anthracis</i> Ames 0581 "Ancestor"	NC_007530	<i>16S rRNA A</i>	(+)	9307 9317	9367 9407	9487 9537	10437 10467	10757 10787
		<i>16S rRNA B</i>	(+)	29101 29111	29161 29201	29281 29331	30231 30261	30551 30581
		<i>16S rRNA C</i>	(+)	82423 82433	82483 82523	82603 82653	83553 83583	83873 83903
		<i>16S rRNA D</i>	(+)	145487 145497	145547 145587	145667 145717	146617 146647	146937 146967
		<i>16S rRNA E</i>	(+)	246207 246217	246267 246307	246387 246437	247337 247367	247657 247687
		<i>16S rRNA F</i>	(+)	266786 266796	266846 266886	266966 267016	267916 267946	268236 268266
		<i>16S rRNA G</i>	(+)	280084 280094	280144 280184	280264 280314	281214 281244	281534 281564
		<i>16S rRNA H</i>	(+)	290057 290067	290117 290157	290237 290287	291187 291217	291507 291537
		<i>16S rRNA I</i>	(+)	532435 532445	532495 532535	532615 532665	533565 533595	533885 533915
		<i>16S rRNA J</i>	(+)	741537 741547	741597 741637	741717 741767	742667 742697	742987 743017
		<i>16S rRNA K</i>	(-)	4656871 4656861	4656811 4656771	4656691 4656641	4655741 4655711	4655421 4655391
<i>B. anthracis</i> Sterne	NC_005945	<i>16S rRNA 1</i>	(+)	9308 9318	9368 9408	9488 9538	10438 10468	10758 10788
		<i>16S rRNA 2</i>	(+)	29102 29112	29162 29202	29282 29332	30232 30262	30552 30582
		<i>16S rRNA 3</i>	(+)	82424 82434	82484 82524	82604 82654	83554 83584	83874 83904
		<i>16S rRNA 4</i>	(+)	145488 145498	145548 145588	145668 145718	146618 146648	146938 146968
		<i>16S rRNA 5</i>	(+)	246220 246230	246280 246320	246400 246450	247350 247380	247670 247700
		<i>16S rRNA 6</i>	(+)	266799 266809	266859 266899	266979 267029	267929 267959	268249 268279
		<i>16S rRNA 7</i>	(+)	280097 280107	280157 280197	280277 280327	281227 281257	281547 281577
		<i>16S rRNA 8</i>	(+)	290070 290080	290130 290170	290250 290300	291200 291230	291520 291550
		<i>16S rRNA 9</i>	(+)	532478 532488	532538 532578	532658 532708	533608 533638	533928 533958
		<i>16S rRNA 10</i>	(+)	741432 741442	741492 741532	741612 741662	742562 742592	742882 742912
		<i>16S rRNA 11</i>	(-)	4657967 4657957	4657907 4657867	4657787 4657737	4656837 4656807	4656517 4656487

Souches	Numéro d'accension	Copies ARNr16S		Annotation zone 1 (10 pb)	Annotation zone 2 (30 pb)	Annotation zone 3 (50 pb)	Annotation zone 4 (30 pb)	Annotation zone 5 (30 pb)
<i>B. cereus</i> ATCC 10987	NC_003909	<i>16S rRNAA</i>	(+)	9307 9317	9367 9407	9487 9537	10437 10467	10757 10787
		<i>16S rRNAB</i>	(+)	29081 29091	29141 29181	29261 29311	30211 30241	30531 30561
		<i>16S rRNAC</i>	(+)	82425 82435	82485 82525	82605 82655	83555 83585	83875 83905
		<i>16S rRNAD</i>	(+)	145493 145503	145553 145593	145673 145723	146623 146653	146943 146973
		<i>16S rRNAE</i>	(+)	279598 279608	279658 279698	279778 279828	280728 280758	281048 281078
		<i>16S rRNAF</i>	(+)	300648 300658	300708 300748	300828 300878	301778 301808	302098 302128
		<i>16S rRNAG</i>	(+)	313859 313869	313919 313959	314039 314089	314989 315019	315309 315339
		<i>16S rRNAH</i>	(+)	322958 322968	323018 323058	323138 323188	324088 324118	324408 324438
		<i>16S rRNAI</i>	(+)	330776 330786	330836 330876	330956 331006	331906 331936	332226 332256
		<i>16S rRNAJ</i>	(+)	604930 604940	604990 605030	605110 605160	606060 606090	606380 606410
		<i>16S rRNAK</i>	(+)	818137 818147	818197 818237	818317 818367	819267 819297	819587 819617
		<i>16S rRNAL</i>	(-)	4655723 4655713	4655663 4655623	4655543 4655493	4654593 4654563	4654273 4654243
		<i>B. cereus</i> ATCC 14579	NC_004722	<i>16S rRNAA</i>	(-)	9187 9197	9247 9287	9367 9417
<i>16S rRNAB</i>	(+)			28928 28938	28988 29028	29108 29158	30058 30088	30378 30408
<i>16S rRNAC</i>	(+)			82215 82225	82275 82315	82395 82445	83345 83375	83665 83695
<i>16S rRNAD</i>	(+)			87923 87933	87983 88023	88103 88153	89053 89083	89373 89403
<i>16S rRNAE</i>	(+)			151137 151147	151197 151237	151317 151367	152267 152297	152587 152617
<i>16S rRNAF</i>	(+)			245178 245188	245238 245278	245358 245408	246308 246338	246628 246658
<i>16S rRNAG</i>	(+)			265743 265753	265803 265843	265923 265973	266873 266903	267193 267223
<i>16S rRNAH</i>	(+)			273016 273026	273076 273116	273196 273246	274146 274176	274466 274496
<i>16S rRNAI</i>	(+)			286313 286323	286373 286413	286493 286543	287443 287473	287763 287793
<i>16S rRNAJ</i>	(+)			293420 293430	293480 293520	293600 293650	294550 294580	294870 294900
<i>16S rRNAK</i>	(+)			510017 510027	510077 510117	510197 510247	511147 511177	511467 511497
<i>16S rRNAL</i>	(+)			718440 718450	718500 718540	718620 718670	719570 719600	719890 719920
<i>16S rRNAM</i>	(-)			4808648 4808638	4808588 4808548	4808468 4808418	4807518 4807488	4807198 4807168
<i>B. cereus</i> EL33	NC_006274	<i>16S rRNA1</i>	(+)	9309 9319	9369 9409	9489 9539	10439 10469	10759 10789
		<i>16S rRNA2</i>	(+)	29107 29117	29167 29207	29287 29337	30237 30267	30557 30587
		<i>16S rRNA3</i>	(+)	82428 82438	82488 82528	82608 82658	83558 83588	83878 83908
		<i>16S rRNA4</i>	(+)	145486 145496	145546 145586	145666 145716	146616 146646	146936 146966
		<i>16S rRNA5</i>	(+)	249431 249441	249491 249531	249611 249661	250561 250591	250881 250911
		<i>16S rRNA6</i>	(+)	272020 272030	272080 272120	272200 272250	273150 273180	273470 273500
		<i>16S rRNA7</i>	(+)	279213 279223	279273 279313	279393 279443	280343 280373	280663 280693
		<i>16S rRNA8</i>	(+)	286580 286590	286640 286680	286760 286810	287710 287740	288030 288060
		<i>16S rRNA9</i>	(+)	296696 296706	296756 296796	296876 296926	297826 297856	298146 298176
		<i>16S rRNA10</i>	(+)	303829 303839	303889 303929	304009 304059	304959 304989	305279 305309
		<i>16S rRNA11</i>	(+)	514645 514655	514705 514745	514825 514875	515775 515805	516095 516125
		<i>16S rRNA12</i>	(+)	731411 731421	731471 731511	731591 731641	732541 732571	732861 732891
		<i>16S rRNA13</i>	(-)	4727960 4727950	4727900 4727860	4727780 4727730	4726830 4726800	4726510 4726480

Souches	Numéro d'accèsion	Copies ARNr16S		Annotation zone 1 (10 pb)		Annotation zone 2 (30 pb)		Annotation zone 3 (50 pb)		Annotation zone 4 (30 pb)		Annotation zone 5 (30 pb)			
<i>B. halodurans</i> C 125	NC_002570	<i>16S rRNA</i>	(+)	22819	22829	22879	22919	22999	23049	23949	23979	24269	24299		
		<i>16S rRNAB</i>	(+)	50588	50598	50648	50688	50768	50818	51718	51748	52038	52068		
		<i>16S rRNAC</i>	(+)	117374	117384	117434	117474	117554	117604	118504	118534	118824	118854		
		<i>16S rRNAD</i>	(+)	124393	124403	124453	124493	124573	124623	125523	125553	125843	125873		
		<i>16S rRNAE</i>	(+)	261929	261939	261989	262029	262109	262159	263059	263089	263379	263409		
		<i>16S rRNAF</i>	(+)	559243	559253	559303	559343	559423	559473	560373	560403	560693	560723		
		<i>16S rRNAG</i>	(+)	656197	656207	656257	656297	656377	656427	657327	657357	657647	657677		
		<i>16S rRNAH</i>	(+)	1017921	1017931	1017981	1018021	1018101	1018151	1019051	1019081	1019371	1019401		
		<i>B. licheniformis</i> ATCC14580	NC_006270	<i>16S rRNA</i>	(+)	9900	9910	9960	10000	10080	10130	11030	11060	11350	11380
<i>16S rRNAB</i>	(+)			34603	34613	34663	34703	34783	34833	35733	35763	36053	36083		
<i>16S rRNAC</i>	(+)			95339	95349	95399	95439	95519	95569	96469	96499	96789	96819		
<i>16S rRNAD</i>	(+)			158293	158303	158353	158393	158473	158523	159423	159453	159743	159773		
<i>16S rRNAE</i>	(+)			612005	612015	612065	612105	612185	612235	613135	613165	613455	613485		
<i>16S rRNAF</i>	(+)			920484	920494	920544	920584	920664	920714	921614	921644	921934	921964		
<i>16S rRNAG</i>	(-)			3122986	3122976	3122926	3122886	3122806	3122756	3121856	3121826	3121536	3121506		
<i>B. subtilis</i> 168	NC_000964			<i>rrnO</i>	(+)	9809	9819	9869	9909	9989	10039	10939	10969	11259	11289
				<i>rrnA</i>	(+)	30277	30287	30337	30377	30457	30507	31407	31437	31727	31757
		<i>rrnJ</i>	(+)	90533	90543	90593	90633	90713	90763	91663	91693	91983	92013		
		<i>rrnW</i>	(+)	96389	96399	96449	96489	96569	96619	97519	97549	97839	97869		
		<i>rrnI</i>	(+)	160892	160902	160952	160992	161072	161122	162022	162052	162342	162372		
		<i>rrnH</i>	(+)	166499	166509	166559	166599	166679	166729	167629	167659	167949	167979		
		<i>rrnG</i>	(+)	171497	171507	171557	171597	171677	171727	172627	172657	172947	172977		
		<i>rrnE</i>	(+)	634981	634991	635041	635081	635161	635211	636111	636141	636431	636461		
		<i>rrnD</i>	(+)	946031	946041	946091	946131	946211	946261	947161	947191	947481	947511		
<i>rrnB</i>	(-)	3177681	3177671	3177621	3177581	3177501	3177451	3176551	3176521	3176231	3176201				
<i>B. thuringiensis</i> serovar Konkukian 97 27	NC_005957	<i>16S rRNA1</i>	(+)	9309	9319	9369	9409	9489	9539	10439	10469	10759	10789		
		<i>16S rRNA2</i>	(+)	29110	29120	29170	29210	29290	29340	30240	30270	30560	30590		
		<i>16S rRNA3</i>	(+)	82438	82448	82498	82538	82618	82668	83568	83598	83888	83918		
		<i>16S rRNA4</i>	(+)	88151	88161	88211	88251	88331	88381	89281	89311	89601	89631		
		<i>16S rRNA5</i>	(+)	151219	151229	151279	151319	151399	151449	152349	152379	152669	152699		
		<i>16S rRNA6</i>	(+)	252202	252212	252262	252302	252382	252432	253332	253362	253652	253682		
		<i>16S rRNA7</i>	(+)	272721	272731	272781	272821	272901	272951	273851	273881	274171	274201		
		<i>16S rRNA8</i>	(+)	279732	279742	279792	279832	279912	279962	280862	280892	281182	281212		
		<i>16S rRNA9</i>	(+)	287078	287088	287138	287178	287258	287308	288208	288238	288528	288558		
		<i>16S rRNA10</i>	(+)	297047	297057	297107	297147	297227	297277	298177	298207	298497	298527		
		<i>16S rRNA11</i>	(+)	521448	521458	521508	521548	521628	521678	522578	522608	522898	522928		
		<i>16S rRNA12</i>	(+)	732398	732408	732458	732498	732578	732628	733528	733558	733848	733878		
		<i>16S rRNA13</i>	(-)	4662123	4662113	4662063	4662023	4661943	4661893	4660993	4660963	4660673	4660643		

Souches	Numéro d'accèsion	Copies ARNr16S		Annotation zone 1 (10 pb)		Annotation zone 2 (30 pb)		Annotation zone 3 (50 pb)		Annotation zone 4 (30 pb)		Annotation zone 5 (30 pb)	
<i>Clostridium tetani</i> E88	NC_004557	<i>16S rRNAA</i>	(-)	10223	10213	10163	10123	10043	9993	9093	9063	8773	8743
		<i>16S rRNAB</i>	(-)	43309	43299	43249	43209	43129	43079	42179	42149	41859	41829
		<i>16S rRNAC</i>	(+)	176113	176123	176173	176213	176293	176343	177243	177273	177563	177593
		<i>16S rRNAD</i>	(+)	565000	565010	565060	565100	565180	565230	566130	566160	566450	566480
		<i>16S rRNAE</i>	(-)	2793490	2793480	2793430	2793390	2793310	2793260	2792360	2792330	2792040	2792010
		<i>16S rRNAF</i>	(-)	2798807	2798797	2798747	2798707	2798627	2798577	2797677	2797647	2797357	2797327

RÉFÉRENCES BIBLIOGRAPHIQUES

- Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. et Polz, M. F. (2004).** Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J. Bacteriol.* **9**: 2629 - 2635.
- Alexander, M. (1977).** Introduction to soil microbiology. New York: John Wiley. pp: 467.
- Amann, R. L., Ludwig, W. et Schleifer, K. - H. (1995).** Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**: 143 - 169.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. et Lipman, D. J. (1990).** Basic local alignment search tool. *J. Mol. Biol.* **215**: 403 - 410.
- Ash, C., Farrow, J. A. E., Dorsch, M., Stackebrandt, E. et Collins, M. D. (1991a).** Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. *Int. J. Syst. Bacteriol.* **41**: 343 - 346.
- Ash, C., Farrow, J. A. E., Wallbanks, S. et Collins, M. D. (1991b).** Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit ribosomal RNA sequence. *Lett. Appl. Microbiol.* **13**: 202 - 206.
- Bavykin, S. G., Lysov, Y. P., Zakhariyev, V., Kelly, J. J., Jackman, J., Stahl, D. A. et Cherni, A. (2004).** Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *J. Clin. Microbiol.* **42**: 3711- 3730.
- Benson, D. A., Boguski, M. S., Lipman, D. J. and Ostell, J. (1997).** GenBank. *Nucleic Acids Res.* **25**: 1- 6.

- Berkeley, R. C. W., Logan, N. A., Shute, L. A., Capey, A. G. (1984).** Identification of *Bacillus* species. In *Met. Microbiol.* 16 ed. Bergan, T. London: Academic Press. pp. 291- 328.
- Bradley, J. C., Galil, B. (1977).** The taxonomic arrangement of the Phasmatodea with keys to the subfamilies and tribes. *Proc. Entomol. Soc. Washington DC.* 2: 176 - 208.
- Bricker, B. J. (2000).** Characterization of the three ribosomal RNA operons *rrnA*, *rrnB*, and *rrnC*, from *Brucella melitensis*. *Gene* 255: 117 - 126.
- Carr, J. G. (1983).** Microbes I have known: a study of those associated with fermented products. *J. Appl. Bacteriol.* 55: 383 - 402.
- Casanova, A., Obreque, J. C., Sandino, A. M. G. et Jashés, M. (2001).** tRNA genes were found in *Pisicrickettsia salmonis* 16S-23S rDNA spacer region (ITS). *FEMS Microbiol. Lett.* 197: 19 - 22.
- Cherif, A., Borin, S., Rizzi, A., Ouzari, H., Boudabous, A. et Daffonchio, D. (2003).** *Bacillus anthracis* diverges from related clades of the *Bacillus cereus* group in 16S-23S ribosomal DNA intergenic transcribed spacers containing tRNA Genes. *Appl. Environ. Microbiol.* 69: 33 - 40.
- Claus, D. et Berkeley, R. C. W. (1986).** Genus *Bacillus* Cohn 1872. In *Bergey's Manual of Systematic Bacteriology*. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore: Williams & Wilkins. vol 2: 1105 - 1139.
- Clayton, R. A., Sutton, G., Hinkle, P. S. Jr., Bult, C. et Fields, C. (1995).** Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. *Int. J. Syst. Bacteriol.* 45: 595 - 599.
- Coenye, T. et Vandamme, P. (2003).** Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol. Lett.* 228: 45- 49.
- Condon, C., Philips, J., Squires, Z. Fu. C., and Squires, C. L. (1992).** Comparison of the expression of seven ribosomal RNA operons in *Escherichia coli*. *EMBO J.* 11: 4175 - 4185.

- Crosby, L. D. et Criddle, C. S. (2003).** Understanding bias in microbial community analysis techniques due to *rrn* operon copy number heterogeneity. *Biotech.* **34** :790 - 802.
- De La Rosa, A., Williams, R. L. et Steeg, P. S. (1995).** Nm23/nucleoside diphosphate kinase: toward a structural and biochemical understanding of its biological functions. *Bioessays.* **17**: 53 - 62.
- Doolittle, W. F. et Brown, J. R. (1994).** Tempo, mode, the progenote and the universal root. *Proc. Natl. Acad. Sci. USA.* **91**: 6721 - 6728.
- Drobniewski, F. A. (1993).** *Bacillus cereus* and related species. *Clin. Microbiol. Rev.* **6**: 324 - 338.
- Efron, B. et Tibshirani, R. J. (1993).** An introduction to the Bootstrap. Chapman and Hall, New York. p:178-201.
- Erickson, R. J. (1976).** Industrial applications of the bacilli: a review and prospectus. In *Microbiol.* Edited by Schlesinger D. Washington: American Society for microbiology. 406 - 419.
- Fahmy, F., Flossdorf, J. et Claus, D. (1985).** The DNA composition of the type strains of the genus *Bacillus*. *Syst. App. Microbiol.* **6**: 60 - 65.
- Felsenstein, J. (1981).** Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**: 368 - 376.
- Fitch, W. M. (1971).** Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20**: 406 - 416.
- Forsyth, R. A., Haselbeck, R. J., Ohlsen, K. L., Yamamoto, R. T., Xu, H., Trawick, J. D., Wall, D., Wang, L., Brown-Driver, V., Froelich, J. M., Kedar, G. C., King, P., McCarthy, M., Malone, C., Misiner, B., Robbins, D., Tan, Z., Zhu, Z. - Y., Carr, G., Mosca, D. A., Zamudio, C., Foulkes, J. G. et Zyskind, J. W. (2002).** A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol. Microbiol.* **43**: 1387 - 1400.

- Fukushima, M., Kakinuma, K. et Kawaguchi, R. (2002).** Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *J. Clin. Microbiol.* **40**: 2779 - 2785.
- Garcia - Martinez, J., Acinas, S. G., Anton, A. I. et Rodriguez - Valera, F. (1999).** Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *J. Microbiol. Meth.* **36**: 55 - 64.
- Garrity, G. M. et Holt, J. G. (2001)** The road map to the manual. In: Bergey's Manual of Systematic Bacteriology, 2nd edn. Springer, New York. **1**: 119 - 141.
- Gherna, R., Pienta, P. et Cote, R. (1989).** American type culture collection catalogue of bacteria and phages. Rockville: American Type Culture Collection.
- Gordon, R. E., Hayens, W. C & Pang, C. H. N. (1973).** The genus *Bacillus*. In *Bacillus*. Edited by Colin R. Harwood. New - York & London: Plenum Press. p: 293 - 320.
- Gupta, R. S. (2000).** The phylogeny of Proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol. Lett.* **24** : 367- 402.
- Gürtler, V., Rao, Y., Pearson, S. R., Bates, S. M. et Mayall, B. C. (1999).** DNA sequences heterogeneity in three copies of the long 16S-23S rDNA spacer of *Enterococcus faecalis* isolates. *Microbiology* **145**: 1785 - 1796.
- Gürtler, V. & Stanisich, V. A. (1996).** New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology* **142**: 3 - 16.
- Hama, H., Almaula, N., Lerner, C. G., Inouye, S. & Inouye, M. (1991).** Nucleoside diphosphate kinase from *Escherichia coli*; its overproduction and sequence comparison with eukaryotic enzymes. *Gene* **105**: 31 - 36.
- Hara, T. & Veda, S. (1982).** Regulation of polyglutamate production in *Bacillus subtilis* (natto): Transformation of high PGA productivity. *Agric. Biol. Chem.* **29**: 45 - 72.

- Helgason, E., Okstad, O. A., Caugant, D. A. et Johansen, H. A. (2000).** *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis* - one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* **66**: 2627 - 2630.
- Higgins, D. G. & Sharp, P. M. (1988).** CLUSTALW: a package for performing multiple sequence alignments on a microcomputer. *Gene* **73**: 237 - 244.
- Hillis, D. M. Moritz, C. Porter, C. A. et Baker, R. J. (1991).** Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* **251**: 308 - 310.
- Honda, H., Kudo, T. et Horikoshi, K. (1985a).** Molecular cloning and expression of the xylanase gene of alkaliphilic *Bacillus* sp. strain C-125 in *E. coli*. *J. Bacteriol.* **161**: 784 - 785.
- Honda, H., Kudo, T., Ikura, Y. et Horikoshi, K. (1985b).** Two types of xylanases of alkaliphilic *Bacillus* sp. no. C-125. *Can. J. Microbiol.* **31**: 538 - 542.
- Ikura, Y. et Horikoshi, K. (1979).** Isolation and some properties of β -galactosidase producing bacteria. *Agric. Biol. Chem.* **43**: 85 - 88.
- Jackson, S. G., Goodbrand, R. B., Ahmed, R. et Kasatiya, S. (1995).** *Bacillus cereus* and *Bacillus thuringiensis* isolated in a gastroenteritis outbreak investigation. *Lett. App. Microbiol.* **21**: 103 - 105.
- Jensen, M. A., Webster, J. A. et Straus, N. (1993).** Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA polymorphism. *App. Environ. Microbiol.* **59**: 945 - 952.
- Jernigan, D. B., Raghunathan, P. L., Bell, B. P., Brechner, R., Bresnitz, E. A., Butler, J. C., Cetron, M., Cohen, M., Doyle, T., Fischer, M., Greene, C., Griffith, K. S., Guarner, J., Hadler, J. L., Hayslett, J. A., Meyer, R., Petersen, L. R., Phillips, M., Pinner, R., Popovic, T., Quinn, C.P., Reefhuis, J., Reissman, D., Rosenstein, N., Schuchat, A., Shieh, W.J., Siegal, L., Swardlow, D. L., Tenover, F. C., Traeger, M., Ward, J. W., Weisfuse, I., Wiersma, S., Yeskey, K., Zaki, S., Ashford, D. A., Perkins, B. A., Ostroff, S., Hughes, J., Fleming, D., Koplan, J. P. et Gerberding, J. L. (2002).** Investigation of bioterrorism-related anthrax, United States, 2001: Epidemiologic Findings. *Emerg. Infect. Dis.* **8**: 1019 - 1028.

- Jernigan, J. A., Stephens, D. S., Ashford, D. A., Omenaca, C., Topiel, M. S., Galbraith, M., Tapper, M., Fisk, T. L., Zaki, S., Popovic, T., Meyer, R. F., Quinn, C. P., Harper, S. A., Fridkin, S. K., Sejvar, J. J., Shepard, C. W., McConnell, M., Guarner, J., Shieh, W. J., Malecki, J. M., Gerberding, J. L., Hughes, J. M. et Perkins, B.A. (2001).** Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. *Emerg. Infect. Dis.* **7**: 933 - 944.
- Johansen, T., Carlson, C. R. et Kolsto, A. - B. (1996).** Variable numbers of rRNA operons in *Bacillus cereus* strains. *FEMS Microbiol. Lett.* **136**: 325 - 328.
- Joung, K. - B. & Côté, J. - C. (2001).** A phylogenetic analysis of *Bacillus thuringiensis* serovars by RFLP- based ribotyping. *J. Appl. Microbiol.* **91**: 279 - 289.
- Joung, K. - B. & Côté, J. - C. (2004).** Ecological impact of *Bacillus thuringiensis*-based insecticides. *Microbiology* **1**:1 - 14.
- Jukers, T. H. et Cantor, C. (1969).** Mammalian protein metabolism. Evolution of protein molecules, Academic Press. New York. p: 21 - 132.
- Klappenbach, J. A., Dunbar, J. M. et Schmidt, T. M. (2000).** rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66**: 1328 - 1333.
- Ko, K. S., Kim, J. - W., Kim, J. - M., Kim, W., Chung, S. - I., Kim, I. J. et Kook, Y. - H. (2004).** Population structure of the *Bacillus cereus* group as determined by sequence analysis of six housekeeping genes and the *plcR* gene. *Infect. Immun.* **72**: 5253 - 5261.
- Kolbert, C. P. et Persing, D. H. (1999).** Ribosomal DNA sequencing as a tool for identification of bacterial pathogens. *Curr. Opin. Microbiol.* **2**: 299 - 305.
- Krieg, A. (1981).** In The genus *Bacillus*: insect pathogen in the prokaryotes, a handbook on habitats, isolation, identification of bacteria. Starr, M.P., Stolp, H., Trüper, H.G., Balows, A. and Schlegel, H.G. New York: Springer-Verlag. **2**: 1743 - 1755.
- Kumar, S., Tamura, K. & Nei, M. (2004).** MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Brief Bioinformatics* **5**: 150 - 163.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Dusterhoft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B. S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Amakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H. et Danchin, A. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**: 249 - 256.

Lapidus, A., Galleron, N., Andersen, J. T., Jørgensen, P. L., Ehrlich, S. D. et Sorokin, A. (2002). Co-linear scaffold of the *Bacillus licheniformis* and *Bacillus subtilis* genomes and its use to compare their competence genes. *FEMS Microbiol. Lett.* **209**: 23 - 30.

Leblond-Bourget, N., Philippe, H., Mangin, I. et Decaris, B. (1996). 16S rRNA and 16S to 23S internal transcribed spacer sequence analysis reveal inter- and intraspecific *Bifidobacterium* phylogeny. *Int. J. Syst. Bacteriol.* **46**: 102 - 111.

Lecadet, M. M., Frachon, E., Dumanoir, V. C., Ripouteau, H., Hamon, S., Laurent, P. et Thiery, I. (1999). Updating the hantigen classification of *Bacillus thuringiensis*. *J. App. Microbiol.* **86**: 660 - 672.

- Lechner, S., Mayr, R., Francis, K. P., Pruss, B. M., Kaplan, T., Wiessner-Gunkel, E., Stewart, G. S. et Scherer, S. (1998).** *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int. J. Syst. Bacteriol.* **48**: 1373 - 1382.
- Logan, N. A. & Berkeley, R. C. W. (1984).** Identification of *Bacillus* strains using API system. *J. Gen. Microbiol.* **130**: 1871 - 1882.
- Logan, N. A. & Turnbull, P. C. (1999).** Manual of Clinical Microbiology (Murray, P.R., Ed.). American Society for Microbiology. Washington, DC. pp: 357 - 369.
- Maeda, T., Takada, N., Furushita, M. et Shiba, T. (2000).** Structural variation in the 16S-23S rRNA intergenic spacers of *Vibrio parahaemolyticus*. *FEMS Microbiol. Lett.* **192**: 73 - 77.
- Mears, J. A., Cannone, J. J., Stagg, S. M., Gutell, R. R., Agrawal, R. K. & Havery, S. C. (2002).** Modeling a minimal ribosome based on comparative sequence analysis. *J. Mol. Biol.* **321**: 215 - 234.
- Mock, M. et Fouet, A. (2001).** Anthrax. *Annu. Rev. Microbiol.* **55**: 647 - 671.
- Muñoz-Dorado, J., Inouye, M. & Inouye, S. (1990).** Nucleoside diphosphate kinase from *Myxococcus xanthus*. I. Cloning and sequencing of the gene. *J. Biol. Chem.* **265**: 2702 - 2706.
- Neyra, C., Atkinson, L. A., Olubayi, O., Sadasivan, L., Zaurov, D. et Zappi, E. (1996).** Novel microbial technologies for the enhancement of plant growth and biocontrol of fungal diseases in crops. *Ca. Opt. Méd.* **31**: 447 - 456.
- Ohta, T. (1991).** Multigene families and the evolution of complexity. *J. Mol. Evol.* **33**: 34 - 41.
- Ojaimi, C., Davidson, B. E., Saint Girons, I. et Old, I. G. (1994).** Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of the Lyme disease spirochaetes *Borrelia burgdorferi*, *B. garinii* and *B. afzelii*. *Microbiology* **140**: 2931 - 2940.

- Olivier, A., Lee, H. - Y. et Côté, J.- C. (2005).** Study of the variability of 16S rRNA genes in γ -proteobacteria: implications for phylogenetic analysis. *J. Gen. Appl. Microbiol.* **51**: 395 - 405.
- Olsen, G. J., Woese, C. R. et Overbeek, R. (1994).** The winds of (evolutionary) change: breathing new life into microbiology. *J. Bacteriol.* **176**: 1 - 6.
- Padyana, A. K. et Burley, S. K. (2003).** Crystal structure of shikimate 5- dehydrogenase (SDH) bound to NADP: insights into function and evolution. *Structure* **11**: 1005 - 1013.
- Page, R. D. M. (1996).** TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**: 357 - 358.
- Parry, J. M., Turnbull, P. C. B. et Gibson, J. R. (1983).** A Color atlas of *Bacillus* species. Wolfe Medical Atlases, **19**. London: Wolfe Medical Publications
- Patra, G., Fouet, A., Vaissaire, J., Guesdon, L. L. et Mock, M. (2002).** Variation in rRNA operon number as revealed by ribotyping of *B. anthracis* strains. *Res. Microbiol.* **153**: 139 - 148.
- Priest, F. G., Goodfellow, M. & Todd, C. (1981).** The genus *Bacillus*: a numerical analysis. In the aerobic endospore-forming bacteria: classification and identification ed. Berkeley, R.C.W. and Goodfellow, M. New York: Academic Press. p: 91-103.
- Priest, F. G, Goodfellow, M. & Todd, C. (1988).** A numerical classification of the genus *Bacillus*. *J. Gen. Microbiol.* **134**: 1847 - 1882.
- Priest, F. G., Kaji, D. A., Rosato, Y. B. et Canhos, V. P. (1994).** Characterization of *Bacillus thuringiensis* and related bacteria by ribosomal RNA gene restriction fragment polymorphisms. *Microbiology* **140**: 1015 - 1022.
- Ray, N. B. & Mathews, C. K. (1992).** Nucleoside diphosphokinase: a functional link between intermediary metabolism and nucleic acid synthesis. *Curr. Top. Cell Regul.* **33**: 343 - 357.

- Rey, M. W., Ramaiya, P., Nelson, B. A., Brody-Karpin, S. D., Zaretsky, E. J., Tang, M., De Leon, A. L., Xiang, H., Gusti, V., Clausen, I. G., Olsen, P. B., Rasmussen, M. D., Andersen, J. T., Jørgensen, P. L., Larsen, T. S., Sorokin, A., Bolotin, A., Lapidus, A., Galleron, N., Ehrlich, S. D. et Berka, R. M. (2004).** Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biol.* **5**: R77.
- Rintaro, S., Yousuke, O., Naoki, K. & Masaru, T. (2000).** Computer analysis of potential stem structures of rRNA operons in various prokaryote genomes. *Gene* **259**: 217 - 222.
- Rössler, D., Ludwig, W., Schleifer, K. H., Lin, C., McGill, T. J., Wisotzkey, J. D., Jurtshuk, P. Jr. & Fox, G. E. (1991).** Phylogenetic diversity in the genus *Bacillus* as seen by 16S rRNA sequencing studies. *Syst. Appl. Microbiol.* **14**: 266 - 269.
- Saitou, N. & Nei, M. (1987).** The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol.* **4**: 406 - 425.
- Sawada, H., Suzuki, F., Matsuda, I. et Saitou, N. (1999).** Phylogenetic analysis of *Pseudomonas syringae* pathovars suggests the horizontal gene transfer of argK and the evolutionary stability of hrp gene cluster. *J. Mol. Evol.* **49**: 627 - 644.
- Schallmeyer, M., Singh, A. et Ward, O. P. (2004).** Developments in the use of *Bacillus* species for industrial production. *Can. J. Microbiol.* **50**: 1 - 17.
- Schouls, L. M., Schot, C. S. et Jacobs, J. A. (2003).** Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. *J. Bacteriol.* **185**: 7241-7246.
- Sneath, P. H. A., Mair, N. S., Sharpe, M. E. et Holt, J. G. (1986).** Bergey's Manual of Systematic Bacteriology, vol 2. Baltimore, Williams & Wilkins. pp. 1105 - 1139.
- Stackebrandt, E. et Goebel, B. M. (1994).** Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44**: 846 - 849.

- Stackebrandt, E., Ludwig, W. et Weizenegger, M. (1987).** Comparative 16S rRNA oligonucleotide analyses and murein types of round-spore-forming bacilli and non-spore-forming relatives. *J. Gen. Microbiol.* **133**: 2523 - 2529.
- Stahl, D. A. et Amann, R. (1991).** Development and application of nucleic acid probes in bacterial systematics. In *E. Stackebrandt and M. Goodfellow (ed)*. p: 205 - 248.
- Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hiramata, C., Nakamura, Y., Ogasawara, N., Kuhara, S. et Horikoshi, K. (2000).** Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res.* **21**: 4317 - 4331.
- Thompson, J. D., Higgins, D. G. et Gibson, T. J. (1994).** CLUSTALW: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673 - 4680.
- Turnbull, P. C. B. (2002).** Introduction: anthrax history, disease and ecology. *Anthrax* **271**: 1- 19.
- Turnbull, P. C. B. et Kramer, J. M. (1991).** *Bacillus*. In *Manual of Clinical Microbiology*, 5th edn. Edited by A. Balows, W. J. Hausler, Jr, K. L. p: 296 - 303.
- Turnbull, P. C., Huston, R. A., Ward, M. J., Jones, M. N., Quinn, C. P., Finnie, N. J., Duggleby, C. J., Kramer, J. M. et Melling, J. (1992).** *Bacillus anthracis* but not always anthrax. *J. Appl. Bacteriol.* **72**: 21 - 28.
- Ueda, K., Seki, T., Kudo, T., Yoshida, T. & Kataoka, M. (1999).** Two distinct mechanisms cause heterogeneity of 16S rRNA. *J. Bacteriol.* **181**: 78 - 82.
- Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K. et Swings, J. (1996).** Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* **60**: 407 - 438.
- Vold, B. S. (1985).** Structure and organization of genes for transfer ribonucleic acid in *Bacillus subtilis*. *Microbiol. Rev.* **49**: 71 - 80.

- Watanabe, K., Nelson, J. S., Harayama, S. et Kasai, H. (2001).** ICB database: the *gyrB* database for identification and classification of bacteria. *Nucleic Acids Res.* **29**: 344 - 345.
- Watt, P. M. et Higgins, D. (1994).** Structure and function of type II DNA topoisomerases. *J. Biochem.* **303**: 681 - 695.
- Wilbur, W. J. et Lipman, D. J. (1983).** Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA.* **80**: 726 - 730.
- Woese, C. R. (1987).** Bacterial evolution. *Microbiol. Rev.* **51**: 221- 271.
- Woese, C. R. (1994).** There must be a prokaryote somewhere: microbiology's search for itself. *Microbiol. Rev.* **58**: 1 - 9.
- Xu, D. et Côté, J. - C. (2003).** Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3'end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences. *Int. J. Syst. Evol. Microbiol.* **53**:695 - 704.
- Yanez, M. A., Catalan, V., Apraiz, D., Figueras, M. J. et Martinez - Murcia, A. J. (2003).** Phylogenetic analysis of members of the genus *Aeromonas* based on *gyrB* gene sequences. *Int. J. Syst. Evol. Microbiol.* **53**: 875 - 883.
- Zuckerland, E. et Pauling, L. (1965).** Molecules as documents of evolutionary history. *J. Theoret. Biol.* **8**: 357 - 366.