

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

**ÉVALUATION D'UN MARQUEUR D'ADN UNIVERSEL
POUR LA RECONSTRUCTION PHYLOGÉNÉTIQUE
DE TAXA BACTÉRIENS**

**THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN BIOLOGIE**

PAR

SABARIMATOU YAKOUBOU

Avril 2011

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse 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

J'adresse mes plus grands et sincères remerciements à mon directeur de recherche, Dr. Jean-Charles Côté. Je lui témoigne ma gratitude pour m'avoir accueillie dans son laboratoire et m'avoir offert un encadrement d'une grande qualité scientifique, son esprit scientifique rigoureux, sa grande disponibilité, les nombreuses discussions scientifiques, et parfois philosophiques, les très nombreuses heures passées à écrire ensemble nos articles scientifiques, les nombreux suivis, sa grande patience, etc. Ses conseils et critiques pertinents lors de la rédaction de cette thèse m'ont été d'une aide précieuse. Sa confiance, ses encouragements et son soutien permanent m'ont permis de tenir et d'aller au bout de ces études doctorales. Merci de ne m'avoir jamais abandonnée durant toutes ces années. Mille mercis Dr. Côté.

Je remercie ma co-directrice, la professeure Lucie Lamontagne, pour sa grande disponibilité et ses conseils à chaque fois que nécessaire.

J'aimerais remercier les professeurs : Tatiana Scorza, Guy Charpentier et Marc Sirois pour avoir accepté de faire partie du comité d'évaluation de cette thèse.

Ce travail a été réalisé dans le laboratoire de microbiologie moléculaire et appliquée du Centre de Recherche et de Développement en Horticulture d'Agriculture et Agro-Alimentaire Canada.

Je tiens à remercier sincèrement les membres présents et passés du laboratoire 257, Suzanne, Dong et Brahim pour votre aide technique lors de mes travaux de labo, vos conseils et vos soutiens. Merci également à Elisabeth, Véronique, Kwang-Bo, Audrey, Hoon-Yong et Mouna, ex-étudiantes et stagiaires du labo 257 pour leur amitié, leur soutien moral et pour nos diverses discussions constructives.

J'aimerais remercier tout le personnel du Centre de recherche et de développement en horticulture qui, par sa bienveillance, m'a permis d'évoluer dans un cadre de travail très agréable. Un grand merci aux personnels des services d'informatique et de création pour publications scientifiques.

Je rends hommage à mes parents. N'eût-ce été l'enfance très heureuse que vous m'avez offerte et la notion de persévérance que vous avez su m'inculquer, je n'aurais pas pu passer à travers les moments difficiles rencontrés durant la réalisation de ce travail. Merci pour votre soutien indéfectible sur tous les plans.

Je remercie sincèrement ma tante Amama Touré qui a suivi mes premiers pas à l'école et qui continue de me soutenir moralement.

Je remercie mes frères et sœurs et ma cousine Nadjibatou pour leur soutien moral et leur écoute. Merci d'avoir toujours été là pour moi.

Je remercie de tout cœur mon époux. Merci pour ton amour, les sacrifices consentis, ton soutien moral et informatique et ta patience qui m'ont permis d'arriver au bout de ce travail.

Je remercie mes enfants. Il n'existe pas de mot assez fort pour exprimer ce que vous représentez pour moi. Merci pour les sacrifices endurés.

Je remercie chaleureusement mon groupe d'amis du Canada et leurs familles pour m'avoir soutenu moralement tout le long de ce travail.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
TABLE DES MATIÈRES.....	iv
LISTE DES TABLEAUX	iv
LISTE DES FIGURES	v
LISTE DES ABRÉVIATIONS	vi
RÉSUMÉ.....	iv
INTRODUCTION	1
CHAPITRE I.....	1
PHYLOGENY OF THE ORDER <i>BACILLALES</i> INFERRRED FROM 3' 16S rDNA AND 5' 16S-23S ITS NUCLEOTIDE SEQUENCES	1
Abstract	29
Introduction.....	30
Materials and Methods.....	31
Results.....	38
Discussion.....	44
CHAPITRE II.....	2
ASSESSMENT OF A SHORT PHYLOGENETIC MARKER BASED ON COMPARISONS OF 3' END 16S rDNA and 5' END 16S-23S ITS NUCLEOTIDE SEQUENCES OF THE <i>BACILLUS CEREUS</i> GROUP	2
Abstract	3
Introduction.....	50
Materials and Methods.....	52
Results and Discussion	56
Conclusion.....	58
CHAPITRE III	59
PHYLOGENY OF Γ -PROTEOBACTERIA INFERRRED FROM COMPARISONS OF 3' END 16S rRNA GENE AND 5' END 16S-23S ITS NUCLEOTIDE SEQUENCES	59
Abstract	60

Materials and Methods.....	64
Results and discussion.....	68
Conclusion.....	78
CHAPITRE IV	80
ASSESSMENT OF A SHORT PHYLOGENETIC MARKER BASED ON COMPARISONS OF 3' END 16S rDNA AND 5' END 16S-23S ITS NUCLEOTIDE SEQUENCES ON THE GENUS <i>XANTHOMONAS</i>	80
Abstract	81
Introduction.....	82
Materials and Methods.....	83
Results and Discussion	86
Conclusion.....	90
CONCLUSION	91
RÉFÉRENCES	100

LISTE DES TABLEAUX

Tableau	Page
1-1. <i>Bacillale</i> families, genera, species and strains used in this study.....	32
2-1. Bacterial species used in this study.....	53
3-1. γ -proteobacteria strains used in this study.....	65
4-1. <i>Xanthomonadaceae</i> species used in this study	85

LISTE DES FIGURES

Figure	Page
1. Arbre phylogénétique universel adapté de Woese <i>et al.</i> , 1990.....	1
2. Représentation schématique de l'opéron <i>rrn</i> des Prokaryotes	8
3. Régions conservées de l'extrémité 3' du gène 16S de l'ARN ribosomal des bactéries (Adapté de Görtler et Stanisich, 1996)	19
4. Représentation schématique du marqueur universel d'ADN.....	24
1-1. Schematic representation of the 16S and 23S rRNA genes separated by an Internal transcribed spacer (ITS).	38
1-2. Agarose gel electrophoresis of the amplification products in selected species in the Order <i>Bacillales</i> using the L516SF/L523SR primer pair.	39
1-3. Bootstrapped neighbor-joining tree of 72 <i>Bacillales</i> species and strains inferred from the alignment of the 220 bp marker.....	41
2-1. Bootstrapped neighbor-joining tree inferred from comparative alignment of the 220 bp marker from the 129 alleles from the 12 bacterial species and strains under study.	54
3-1. Bootstrapped neighbor-joining tree of γ -proteobacteria species inferred from the alignment of a 157 nucleotide sequence at the 3' end of the 16S rRNA gene.....	70
3-2. Schematic representation of allelic 16S-23S rDNA Internal Transcribed Spacer of γ -proteobacteria.	73
3-3. Bootstrapped neighbor-joining tree of γ -proteobacteria species inferred from the alignment of 232 nucleotide sequence marker	75
4-1. Bootstrapped neighbor-joining tree of all 30 bacterial strains inferred from the alignment of the 224 bp marker.....	87

LISTE DES ABRÉVIATIONS

<i>Aci</i>	<i>Acinetobacter</i>
<i>Act</i>	<i>Actinobacillus</i>
ADNr	Acide desoxyribonucléique ribosomal
<i>Ae</i>	<i>Aeromonas</i>
AFLP	Amplified fragment length polymorphism
<i>Al</i>	<i>Alicyclobacillus</i>
<i>Alc</i>	<i>Alcanivorax</i>
<i>Alk</i>	<i>Alkaliminicola</i>
<i>Am</i>	<i>Amphibacillus</i>
<i>Ane</i>	<i>Aneurinibacillus</i>
<i>Ano</i>	<i>Anoxybacillus</i>
ARNr	Acide ribonucléique ribosomal
<i>B.</i>	<i>Bacillus</i>
bp	Base pair
<i>Br</i>	<i>Brevibacillus</i>
<i>Bu</i>	<i>Buchnera</i>
<i>Chr</i>	<i>Chromohalobacter</i>
<i>Cit</i>	<i>Citrobacter</i>
<i>Col</i>	<i>Colwellia</i>
<i>Cox</i>	<i>Coxiella</i>
<i>D</i>	<i>Dichelobacter</i>
<i>E</i>	<i>Escherichia</i>
<i>Ent</i>	<i>Enterobacter</i>
<i>F</i>	<i>Filobacillus</i>
<i>Fr</i>	<i>Francisella</i>
<i>Ge</i>	<i>Geobacillus</i>
<i>Gr</i>	<i>Gracilibacillus</i>

RÉSUMÉ

Différents marqueurs moléculaires ont été utilisés au cours des deux dernières décennies pour des études systématiques des espèces bactériennes à différents niveaux taxinomiques. L'apparition de nouvelles méthodes d'identification et de classification des espèces bactériennes telles que la taxonomie numérique et la taxonomie phylogénétique ont amenées à la révision de plusieurs taxa bactériens. Dans l'Ordre des *Bacillales* et dans la Classe des γ -protéobactéries par exemple, ces méthodes ont aboutit à la révision de la famille des *Bacillaceae* et du genre *Xanthomonas*, respectivement. Compte tenu de l'importance des bactéries des points de vue économique, environnemental et médical, le travail d'identification et de classification des bactéries demeure un enjeu majeur pour les bactériologistes.

Nous proposons dans cette thèse d'évaluer un marqueur d'ADN basé sur une courte séquence nucléotidique située à l'extrémité 3' du gène 16S ARNr et l'extrémité 5' de la région inter-génique 16S-23S (ITS) pour l'identification et la classification des bactéries à Gram-positif et à Gram-négatif à différents niveaux taxonomiques: Classe, Ordre, famille et genre.

Dans le premier chapitre, la phylogénie des bactéries de l'ordre des *Bacillales* a été construite à l'aide d'un marqueur d'ADN de 220 paires de bases (pb). Il s'agit d'une combinaison des 150 dernières pb de l'extrémité 3' du gène 16S de l'ARNr et des 70 premières paires de bases de l'extrémité 5' de la région inter-génique 16S-23S (ITS). Au total, 72 espèces et souches de l'Ordre des *Bacillales*, réparties en huit familles et 21 genres ont été analysées. Un arbre phylogénétique a été construit et sa robustesse a été statistiquement établie par une analyse bootstrap. Huit groupes ont été révélés et chaque groupe a été subdivisé en sous-groupes et branches. L'arbre phylogénétique ainsi construit à l'aide du marqueur de 220 pb est en général conforme à l'arbre phylogénétique obtenu à partir du gène 16S de l'ARNr et basé sur les données phénotypiques et moléculaires. Nos résultats indiquent que le marqueur présente un pouvoir résolution supérieur à celui du gène 16S de l'ARNr.

Dans le deuxième chapitre, le pouvoir de résolution du marqueur de 220 pb a été évalué à un niveau taxonomique très inférieur à l'Ordre des *Bacillales*: le groupe *Bacillus cereus*. Au cours de cette étude, 129 allèles de huit espèces bactériennes dont entre autres quatre espèces du groupe *B. cereus* ont été analysés. L'arbre phylogénétique construit a révélé quatre groupes majeurs et des sous-groupes. Le marqueur de 220 pb a permis une discrimination au niveau du genre mais n'a pas permis de distinguer les quatre espèces phylogénétiquement très proches du groupe *B.*

cereus sous étude: *B. anthracis*, *B. cereus*, *B. thuringiensis* et *B. weihenstephanensis*. Nous avons atteint la limite du pouvoir de résolution du marqueur de 220 pb.

Dans le troisième chapitre, un marqueur similaire à celui des bactéries Gram-positif a été développé pour l'établissement de la phylogénie chez les espèces de la Classe des γ -protéobactéries. Une analyse comparative de l'extrémité 3' du gène 16S de l'ARNr a permis de déterminer la portion la plus conservée à ce niveau. De même, une analyse comparée de la région inter-génique 16S-23S (ITS) des différents allèles d'une même souche bactérienne a permis de déterminer la partie la plus conservée de cette extrémité 5'. Les deux parties les plus conservées ont été combinées pour former un marqueur d'ADN de 232 pb: 157 pb de l'extrémité 3' du gène 16S de l'ARNr et 75 premières paires de bases de l'extrémité 5' de la région ITS. Un arbre phylogénétique dont la précision a été statistiquement établie par une analyse bootstrap a été construit. Quatre groupes majeurs ont été révélés et des sous-groupes et des clusters ont été formés. Cet arbre est dans l'ensemble conforme à celui basé sur le gène 16S de l'ARNr. Il est intéressant de noter que le marqueur de 232 pb a permis une meilleure discrimination entre les espèces très proches que le gène 16S de l'ARNr en raison d'une part de la conservation de l'extrémité 3' du gène 16S de l'ARNr entre les allèles d'une même souche et d'autre part, de la conservation des 75 pb de région inter-génique 16S-23S (ITS) entre les allèles d'une même souche.

Dans le quatrième chapitre, le pouvoir de résolution d'un marqueur d'ADN de 224 pb est évalué à un niveau taxonomique de beaucoup inférieur à la Classe des γ -protéobactéries: le genre *Xanthomonas* de la classe des γ -protéobactéries. Le marqueur est constitué d'une combinaison des 157 pb à l'extrémité 3' du gène 16S ARNr et des 67 premières paires de bases de l'extrémité 5' région inter-génique 16S-23S (ITS). Un total de 23 espèces du genre *Xanthomonas* ont été analysées. Les espèces éloignées phylogénétiquement et certaines espèces proches du genre *Xanthomonas* sont discriminées par le marqueur de 224 pb. Les espèces très proches phylogénétiquement et les pathovars par contre ne sont pas discriminées par le marqueur. Ceci constitue la limite du pouvoir de résolution de ce marqueur.

Le marqueur d'ADN universel évalué à différents niveaux hiérarchiques des bactéries à Gram-positif et Gram-négatif est un marqueur PCR (Polymerase Chain Reaction) facile d'utilisation du point de vue technique. Il est constitué des dernières paires de bases conservées de l'extrémité 3' du gène 16S de l'ARNr et des premières paires de bases conservées de la région inter-génique 16S-23S. Or le gène 16S de l'ARNr est très conservé entre les espèces d'un même genre tandis que la région inter-génique est hyper-variable à l'intérieur d'une même espèce. Ainsi, ce seul marqueur pourrait servir à la fois aux études systématiques intra et inter-spécifiques aboutissant à l'identification des espèces bactériennes, à leur classification, voire même à la révision de certaines classifications.

INTRODUCTION

Le monde vivant a été divisé en trois domaines sur la base des travaux de phylogénie moléculaire réalisés par Carl Woese et collaborateurs : les Eubactéries, les Archæabactéries et les Eucaryotes (Woese *et al.*, 1990) (Fig. 1).

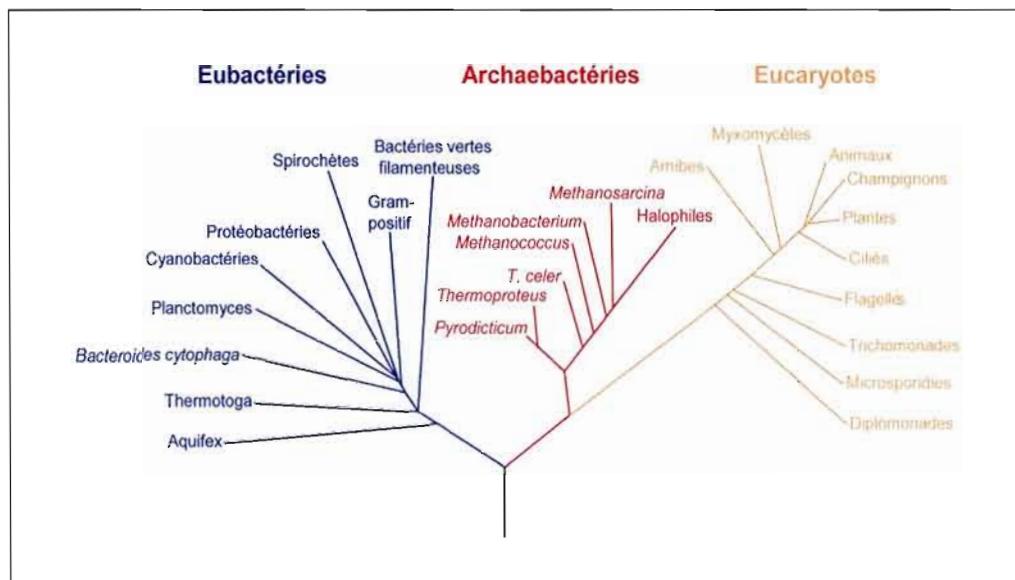


Fig. 1. Arbre phylogénétique universel adapté de Woese *et al.*, 1990.

Les Archæabactéries et les Eubactéries forment les procaryotes. Les procaryotes sont des organismes vivants unicellulaires de petites tailles en général et caractérisés par une absence de noyau et d'organites intracellulaires. Les Eucaryotes sont des organismes vivants plus complexes que les procaryotes. Les principales caractéristiques qui permettent de distinguer les Eucaryotes des Procaryotes sont : la présence d'un noyau renfermant l'information génétique, la présence d'organites intracellulaires pourvus de membranes (les mitochondries, les chloroplastes, les appareils de Golgi, le réticulum endoplasmique) et la présence d'un cytosquelette. Les Archæabactéries se distinguent des Eubactéries par l'absence d'acide muramique.

dans leur paroi cellulaire (Kandler et Konig, 1993). Sur l'arbre phylogénétique universel décrit par Woese *et al.*, 1990, seuls quelques organismes représentatifs de chaque domaine sont présentés. La topographie de l'arbre phylogénétique universel nous révèle que ces trois domaines tirent leur origine d'un ancêtre commun à partir duquel ils auraient divergé. Ce dernier ancêtre commun généralement appelé Luca (The last universal common ancestor) fait l'objet de plusieurs études (Forterre *et al.*, 2005). La racine des trois lignées de l'arbre universel serait dans le domaine des Eubactéries (Iwabe *et al.*, 1989, Woese *et al.*, 1990). L'arbre phylogénétique universel a permis également de révéler que bien que les Archæabactéries soient des Prokaryotes, elles sont phylogénétiquement plus proches des Eucaryotes que des Eubactéries. Les Archæabactéries présentent certains caractères similaires à ceux des Eucaryotes tels que : la présence de plusieurs ARN polymerase complexes, les gènes en mosaïque et les protéines impliquées dans les systèmes de réPLICATION d'ADN, la transcription et la traduction (Olsen et Woese, 1997; Forterre 1999; Kelman, 2000). Selon plusieurs études, les Archæabactéries et Eucaryotes auraient un ancêtre commun à partir duquel l'ancêtre des Eucaryotes aurait divergé plus tôt (Woese, 1981, Woese et Gupta, 1981, Woese 1998, Gupta, 2000; Yutin *et al.*, 2008).

Le domaine des Eubactéries est très diversifié et regroupe les bactéries très variées quant à leur morphologie, leur distribution dans l'environnement, leurs niches, leurs besoins nutritifs et leur métabolisme. Un total de 24 phyla est reconnu dans la dernière édition du Bergey's Manual of Systematic Bacteriology. Citons, les protéobactéries, les bactéries Gram-positif, les cyanobactéries, les spirochètes, etc. Les protéobactéries constituent le phylum le plus important. On y retrouve les bactéries de très grande importance des points de vue économique, médical, scientifique et environnemental connues à ce jour. L'identification moléculaire des bactéries a permis la caractérisation de bactéries non-cultivables des systèmes environnementaux (Ward *et al.*, 1990; Amann *et al.*, 1995; Dauga *et al.*, 2005). Un

total de 3289 taxa bactériens sont présentement décrits sur le site web du "National Center for Biotechnology Information" (NCBI), (<http://www.ncbi.nlm.nih.gov/guide/taxonomy/>).

Compte tenu de la grande diversité des Eubactéries, nous avons volontairement limité notre étude aux niveaux taxinomiques suivants : l'Ordre des *Bacillales* chez les bactéries Gram-positif et la Classe des γ -protéobactéries chez les bactéries Gram-négatif.

Importance des bactéries de l'Ordre des *Bacillales*

Les bactéries ont une grande importance des points de vue économique, environnemental et médical (Gordon *et al.* 1973; Hara et Veda 1982; Carr 1983). L'Ordre des *Bacillales* renferme entre autres des bactéries qui produisent les trois types de substances à grande valeur industrielle : les enzymes, les antibiotiques et les insecticides (Parry *et al.* 1983; Harwood, 1989; Devine, 1995).

Dans l'industrie des enzymes, les bactéries du genre *Bacillus* constituent la source majeure d'enzymes hydrolytiques utilisées surtout dans le domaine alimentaire et dans celui des détergents (Parry *et al.* 1983; Harwood, 1989). Plus des deux tiers des enzymes produits industriellement par des bactéries, le sont par des espèces de *Bacillus* (Devine, 1995). *Bacillus subtilis* est utilisé pour produire des protéases. La subtilisine, protéase alcaline qu'elle produit représentent 60% de la totalité de la production d'enzymes à des fins industrielles. Ce marché est évalué à \$ 26 millions par année rien qu'aux USA (Devine, 1995). Cette enzyme est utilisée dans les lessives, dans les tanneries pour attendrir les cuirs et dans les conserves alimentaires comme additifs (Towalski et Rothman, 1986; Girard, 1997). *Bacillus licheniformis* est le principal producteur d'alpha-amylase, une enzyme très utilisée en agro-

alimentaire pour le traitement de l'amidon. L'alpha-amylase est également ajoutée dans certains savons liquides (Devine, 1995; Girard, 1997). Plusieurs espèces du genre *Geobacillus* produisent des enzymes thermostables. Les lipases et les amylases sont produits respectivement par *Geobacillus stearothermophilus* et *Geobacillus thermoleovorans* (Sifour *et al.*, 2010; Uma Maheswar Rao et Satyanarayana 2007). Dans le genre *Alicyclobacillus*, plusieurs souches d'*Alicyclobacillus acidocaldarius* produisent des amylases (Schwermann *et al.*, 1994) et des mannanases, enzymes qui jouent une rôle important dans l'hydrolyse des hémicelluloses (Zhang *et al.*, 2008).

Dans le domaine médical et vétérinaire, plusieurs genres de différentes familles de l'Ordre des *Bacillales* sont des pathogènes pour l'humain et certains animaux. Chez les humains, *Listeria monocytogenes* est responsable de plusieurs pathologies chez les humains et les animaux telles que la gastroentérite, la méningite, la septicémie et l'avortement (Doumith *et al.*, 2004). *Staphylococcus aureus* est responsable des infections nosocomiales, des intoxications alimentaires, d'infections localisées suppurées (Klutmans *et al.*, 1997). Une souche de *St. aureus* résistante à la méticilline (SARM) est responsable des cas de septicémies graves à fort taux de mortalité (Cosgrove *et al.*, 2003; Klevens *et al.*, 2006). Chez les animaux, *Paenibacillus larvae* est l'agent causal de la loque américaine, infection qui fait des ravages dans les ruches de par le monde (Rieg *et al.*, 2010). Chez les bovins et autres ruminants, *St. aureus* est la cause principale des mastites (Herron *et al.*, 2002).

En biotechnologie, certaines souches du genre *Bacillus* servent à la production d'antibiotiques. *Brevibacillus brevis*, *B. polymyxa* et *B. subtilis* produisent respectivement la gramicidine S, la polymyxine B et la bacitracine (Girard, 1997). Dans le genre *Geobacillus*, deux souches de *Geobacillus thermoleovorans* ont des activités bactériocines sur certaines bactéries d'importance médicale dont *Salmonella typhimurium* (Novotny et Perry, 1992). Récemment, le développement, des systèmes hôte/vecteur et des outils de génétique moléculaire tels que la technologie de l'ADN

recombinant ont mis en exergue l'importance des membres du genre *Bacillus* plus précisément *Bacillus subtilis*. En effet, ils sont utilisés comme hôtes biologiques lors de l'expression de l'ADN recombinant. (Priest, 1981; Harwood, 1989).

En agriculture, certains pesticides biologiques sont produits par des espèces du genre *Bacillus* telles que *B. thuringiensis*, *B. sphaericus* et *B. popilliae*. *B. thuringiensis* synthétise au moment de la sporulation une inclusion parasporale, le cristal (Höfte et Whiteley, 1989; Martin, 1994; Devine, 1995; Joung et Côté, 2000; Wellman-Desbiens et Côté, 2004). Il est pathogène chez les larves de plusieurs ordres d'insectes tels que les lépidoptères, les coléoptères, les diptères, etc. En environnement, des études ont montré l'utilisation de *Geobacillus pallidius* en biorémédiation (Bustard *et al.*, 2002).

Importance des bactéries de la Classe des γ -protéobactéries

La Classe des γ -protéobactéries est constituée de bactéries ayant une grande importance du point de vue médical, scientifique et industriel. On distingue de nombreux genres pathogènes pour les humains et les animaux dans cette Classe.

Dans le domaine médical et vétérinaire, la classe des γ -protéobactéries renferme de nombreux genres pathogènes pour l'humain et certains animaux. Chez les humains, citons quelques genres tels que : *Escherichia coli*, responsable de gastro-entérites, *Klebsiella*, responsable d'infections urinaires et pulmonaires, *Salmonella*, responsable de gastro-entérite, de fièvres typhoïdes et d'intoxications alimentaires, *Shigella*, responsable de la dysenterie bacillaire, *Vibrio*, responsable du cholera, *Yersinia*, responsable de la peste (Kerster *et al.*, 2006). Chez les animaux, citons : *Actinobacillus*, responsable de la shigellose du poulin (Blackall *et al.*, 1998), *Pasteurella* responsable de plusieurs affections chez les mammifères et les oiseaux dont le cholera des poules et la septicémie hémorragique (Boyce *et al.*, 2002).

En biotechnologie, certaines souches d'*Escherichia coli*, organisme modèle en physiologie et biochimie sont utilisées en biologie moléculaire comme hôtes biologiques lors de l'expression de l'ADN recombinant (Harwood, 1989).

En agriculture, certains genres de la Classe des γ -protéobactéries pathogènes de plantes causent des ravages dans les cultures ayant une grande importance économique : *Pseudomonas*, *Xanthomonas* et *Xylella* sont pathogènes pour les plantes (Kerster *et al.*, 2006). En environnement, certaines espèces du genre *Pseudomonas* (*Pseudomonas halodenitrificans*) et le genre *Methylococcus* sont utilisés pour la dépollution des terrains contaminés, la biorémediation (Smith et Murell, 2009).

Considérant leur grande importance, le travail d'identification et de classification des bactéries demeure un enjeu majeur pour les bactériologistes.

Principales techniques de classification et d'identification des bactéries

Au début du 19^{eme} siècle, la taxonomie bactérienne était essentiellement basée sur des critères phénotypiques. Les premières classifications bactériennes étaient basées donc sur des caractères morphologiques, physiologiques et biochimiques (Smith *et al.*, 1952; Wolf et Barker, 1968; Gordon *et al.*, 1973; Gibson et Gordon, 1974; Hobbs et Cross, 1983; Harwood, 1989). La taxonomie microbienne étant un domaine dynamique, les méthodes de classification ont évolué au rythme des découvertes scientifiques. La découverte de la structure de l'ADN par Watson et Crick en 1953 a permis l'avènement des techniques moléculaires d'identification des bactéries (Towner et Cockayne, 1993). Le développement de la méthode de séquençage des protéines par Sanger en 1955 et la publication de l'article “Molecules as documents of evolutionary history” par Zuckerkandl et Pauling en 1965 qui démontre que les séquences des protéines et des acides nucléiques pouvaient

servir de chronomètres moléculaires ouvrent l'ère des phylogénies moléculaires. Ainsi de nouvelles méthodes d'identification et de classification des espèces font leur apparition amenant la révision de différents taxa bactériens. C'est le cas du genre *Bacillus* avec de la taxonomie numérique (Priest, 1981) et du genre *Xanthomonas* avec la taxonomie phylogénétique (Vauterin *et al.*, 1990).

Les techniques de classification et d'identification des bactéries peuvent être regroupées en fonction des critères utilisés, en deux catégories : les techniques de classification phénotypique et génotypique. Ces deux techniques de classification se combinent pour constituer la taxonomie polyphasique qui est l'approche utilisée depuis plus d'une vingtaine d'années pour l'identification des taxas bactériens à tous les niveaux (Vandamme *et al.*, 1996).

L'ARN ribosomal des bactéries.

Chez les bactéries le ribosome est formé de deux sous-unités: la sous-unité 30S et la sous-unité 50S. La sous-unité 30, ou petite sous-unité, est composée de l'ARN 16S et de 21 protéines ribosomales. La sous-unité 50S, ou grande sous-unité, est composée respectivement des ARN 23S et 5S et de 31 protéines ribosomales (Mears *et al.*, 2002).

Les gènes codant pour les ARN ribosomaux (ARNr) des bactéries sont organisés en opéron, appelé opéron *rrn*, localisé sur le chromosome bactérien. Dans l'ordre 5' 3', l'opéron *rrn* est constitué comme suit : le gène 16S, une région inter-génique, aucun, un ou des gènes d'ARN de transfert (ARNr), une région inter-génique, le gène 23S, une région inter-génique, et le gène 5S (Watson *et al.*, 1987; Gürtler et Stanisich, 1996; Garcia-Martinez *et al.*, 1996) (Fig.2). Sa structure est semblable chez toutes les bactéries. De une à 11 copies de l'opéron sont présentes dans les différentes espèces bactériennes (Gürtler, 1993; Garcia-Martinez *et al.*,

1996). *Enococcus œni* possède 1 copie (Le Jeune et Lonvaud-Funel, 1997), *Escherichia coli* en possède 7 (Hill et Harnish, 1981; Ellwood et Nomura, 1982) et *B. subtilis* en possède de 9 à 10 (Loughney *et al.*, 1982).

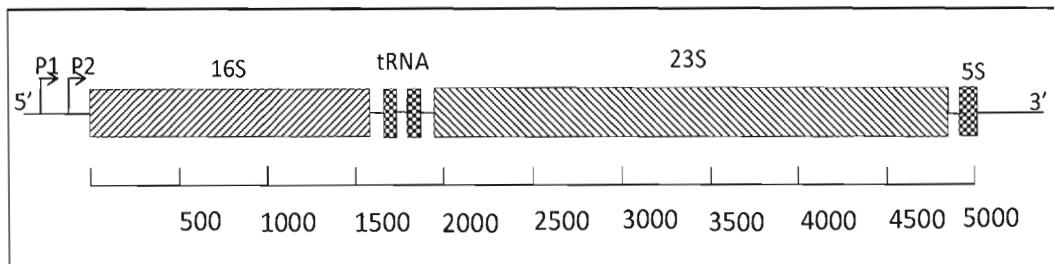


Fig. 2. Représentation schématique de l'opéron *rrn* des Prokaryotes

Les techniques de classification phénotypiques.

Il s'agit de toutes les techniques de classification n'utilisant pas directement l'ADN et l'ARN mais uniquement les caractères phénotypiques (Vandamme *et al.*, 1996). On y retrouve les caractères morphologiques, physiologiques, biochimiques et chimiques.

Les techniques phénotypiques classiques.

On regroupe sous ce terme les techniques se basant sur les caractéristiques morphologiques, physiologiques et biochimiques. Ces techniques constituent la base de la systématique bactérienne et sont nécessaires à la description des bactéries à tous les niveaux taxonomiques, de la sous espèce à l'espèce et du genre à la famille (Vandamme *et al.*, 1996).

- L'étude des caractères morphologiques se fait à la fois au niveau de la cellule et au niveau de la colonie. On étudie la forme, la taille, l'aspect et la couleur des colonies, l'aspect des cultures liquides, la forme et la taille de la cellule, la

motilité, la nature, le nombre et la disposition des flagelles, les réactions de coloration; la présence de structures intracellulaires particulières.

- Comme caractéristiques physiologiques, on étudie : la température optimale de croissance, le pH optimal de croissance, la tolérance à la salinité et à la pression osmotique etc.

- Les caractères biochimiques étudiés sont : la nature des sources de carbone, d'azote et de soufre, le potentiel de fermentation, l'oxydation des hydrates de carbone, la présence et l'activité des enzymes, l'effet des antibiotiques, les produits terminaux du métabolisme etc.

La technique de classification numérique.

La classification numérique a pris son essor dans les années 50 avec le développement des ordinateurs. Elle est basée sur le degré de ressemblance phénotypique entre les organismes vivants (Sneath, 1984; Régnault, 1990). Elle consiste à étudier un grand nombre de caractères phénotypiques pour chaque souche d'un grand échantillon de souches bactériennes et à attribuer le même poids à chacun des caractères qui seront codés 1 (présence du caractère) ou 0 (absence du caractère) et enfin à calculer le degré de ressemblance entre ces souches (Priest *et al.*, 1988; Vandamme *et al.*, 1996). Ce dernier est exprimé par un coefficient de similarité qui correspond au pourcentage de caractères partagés par deux organismes ou par deux groupes d'organismes. Enfin, on construit des dendrogrammes qui permettent d'établir les relations entre les différents groupes étudiés sur la base de leur pourcentage de similarité révélant ainsi le degré de parenté des espèces et des genres (Priest *et al.*, 1988.; Régnault, 1990; Vandamme *et al.*, 1996).

Les techniques chimiques de classification.

Les caractéristiques chimiques constituent de bons outils de taxonomie bactérienne (Claus et Fritze, 1989). La chimintaxonomie ou technique de classification chimique est basée principalement sur l'analyse de la composition chimique de la paroi cellulaire bactérienne, des acides gras présents dans la cellule et des quinones isoprenoïdes, etc...

La composition chimique des parois cellulaires.

La composition chimique de la paroi cellulaire est un important critère d'identification des bactéries (Pelczar *et al.*, 1982). Le composant principal de la paroi, le peptidoglycane est un polymère constitué de différentes substances, N-acétyl glucosamine, l'acide N-acétyl muramique et un peptide. Ce dernier est composé de quatre ou cinq acides aminés, la L- et la D-alanine, l'acide D-glutamique et selon les espèces, la lysine ou l'acide diaminopimélique (Pelczar *et al.*, 1982). La composition et la structure du peptidoglycane varient d'une espèce à l'autre chez les bactéries Gram-positif tandis qu'elles sont fixes chez les Gram- négatif ce qui en fait un bon outil de systématique (Schleifer, 1985; Schleifer et Seidl, 1985). Le peptidoglycane est responsable de la rigidité de la paroi. Environ 100 différents types de peptidoglycane ont été décrits (Schleifer et Kandler, 1972; Schleifer et Seidl, 1985; Stackebrandt *et al.*, 1987).

Les acides gras de la cellule.

Les acides gras sont les constituants essentiels des phospholipides et des lipopolysaccharides de la membrane cytoplasmique cellulaire (Embley et Wait, 1994; Vandamme *et al.*, 1996). Ils diffèrent par leur longueur, par le nombre et la position de leur liaison double (Darnel *et al.*, 1993). Une molécule d'acide gras est

constituée d'une longue chaîne hydrocarbonée reliée à un groupe carboxyle de propriété acide. Ce sont ces critères qui servent à la caractérisation des bactéries (Suzuki *et al.*, 1993). Plus de 300 différentes structures chimiques d'acide gras ont été identifiées (Vandamme *et al.*, 1996).

C'est en 1963 que Able *et al.* et Kaneda ont chacun démontré que l'analyse des acides gras par chromatographie en phase gazeuse permettait l'identification des bactéries (Able *et al.*; 1963; Kaneda, 1963; Siegel *et al.*, 2000). Dans l'Ordre des *Bacillales*, cette méthode a été utilisée dans le genre *Bacillus*, chez les espèces entomopathogènes (Kaneda, 1968; Frachon *et al.*, 1991; Schenkel *et al.*, 1992; Stahly et Klein, 1992; Kampfer, 1994; Siegel *et al.*, 1997; Siegel *et al.*, 2000). Dans la Classe des γ -protéobactéries, la méthode a été utilisée chez les genres *Aeromonas* (Canonica et Pisano, 1988), *Pseudomonas* (Moss et Dees. 1975; Peladan et Monteil, 1984) et *Legionella* (Moss *et al.*, 1983).

Les quinones isoprénoides de la chaîne respiratoire des bactéries.

Les quinones isoprénoides sont les constituants de la membrane cytoplasmique des bactéries qui jouent un rôle important dans le transport des électrons et dans la phosphorylation oxydative (Collins et Jones; 1981; Busse *et al.*, 1996; Vandamme *et al.*, 1996). Ils sont utilisés comme critères taxinomiques à cause de leur structure qui varie en fonction des espèces et qui est constante au sein d'une même espèce (Collins, 1985; Vandamme *et al.*, 1996; Hiraishi, 1999). On distingue deux types principaux de quinones isoprénoides: les benzoquinones et les naphtoquinones qui eux sont subdivisés en phylloquinones et menaquinones présentes chez les bactéries (Vandamme *et al.*, 1996).

L'analyse des protéines totales de la cellule bactérienne.

L'analyse comparative des protéines totales par migration électrophorétique sur un gel de polyacrylamide en présence de dodécylsulfate de sodium (SDS PAGE) permet d'établir les relations systématiques au niveau du genre et de l'espèce (Kersters *et al.*, 1994; Pot *et al.*, 1994; Busse *et al.*, 1996). Le degré de similarité entre les différents patrons protéiques permet d'établir la parenté, la relation, entre les souches étudiées. C'est une technique très utilisée en systématique bactérienne.

Les techniques de classification génotypique.

Ce sont des techniques de classification basées directement sur l'analyse de l'ADN et de l'ARN bactérien. Elles ont fait leur apparition suite aux progrès réalisés en biologie moléculaire. Différentes études basées sur l'analyse de l'ADN ont permis l'identification et la classification de plusieurs souches bactériennes ainsi que l'établissement des relations phylogénétiques entre les taxons (Gürtler et Mayall, 2001). Il existe plusieurs méthodes moléculaires de classification : l'hybridation ADN-ADN et ADN-ARN, la détermination des pourcentages molaires des bases puriques et pyrimidiques, l'amplification aléatoire d'ADN polymorphe (Random amplification of polymorphic DNA ; RAPD) (Williams *et al.*, 1990), le polymorphisme de longueur des fragments de restriction de l'ADN (Restriction fragment length polymorphism; RFLP), le ribotypage, l'électrophorèse en champ pulsé (Pulse-field gel electrophoresis; PFGE, Tenover *et al.*, 1995), le polymorphisme de longueur des fragments amplifiés de l'ADN (Amplified fragment length polymorphism ; AFLP), (De Vos *et al.* , 1995); le séquençage de l'ARN ribosomal, etc.

La détermination des bases de l'ADN (G+C mol%).

La détermination des pourcentages molaires de guanine et de cytosine (G+C mol%) contenus dans l'ADN est une méthode génotypique classique de description des espèces bactériennes (Vandamme *et al.*, 1996; Busse *et al.*, 1996). Ce pourcentage varie entre 24 et 76% chez les bactéries (Vandamme *et al.*, 1996). Chez deux souches appartenant à la même espèce, la plage de variation du pourcentage doit être inférieure ou égale à 3, tandis que chez des espèces appartenant au même genre, elle doit être inférieure ou égal à 10 (Stackebrandt et Liesack, 1993). L'analyse du pourcentage molaire de guanine et de cytosine (G+C mol%) peut se faire par la méthode de dénaturation thermique (De Ley *et al.*, 1970) ou par chromatographie liquide haute performance (High Performance Liquid Chromatography; HPLC, Mesbah *et al.*, 1989).

L'hybridation ADN-ADN.

L'hybridation ADN-ADN est la technique la plus indiquée pour la classification de nouvelles souches au niveau de l'espèce. Il s'agit d'une ré-association des chaînes complémentaires d'un acide nucléique. Elle est basée sur la propriété que possède une molécule d'ADN monobrin de s'associer spontanément et de façon spécifique et réversible à une autre molécule monobrin si celle-ci lui est complémentaire (Wayne *et al.*, 1987; Darnel *et al.*, 1993).

Différentes méthodes d'hybridation ADN-ADN ont été mises au point telles que la méthode de l'hydroxyapatite, la méthode de la membrane filtrante et la méthode à la nuclease S1. Cette dernière est la plus utilisée car contrairement aux autres, elle n'est pas basée sur le marquage radioactif de l'ADN (Legault-Démare *et al.*, 1967; Brenner *et al.*, 1969; De Ley *et al.*, 1970; Crosa *et al.*, 1973).

L'hybridation moléculaire peut se produire entre brins complémentaires d'ADN ou d'ARN mais également entre un brin d'ARN et son brin d'ADN complémentaire (De Vos et De Ley, 1983; Darnel *et al.*, 1993). L'hybridation ADN-ARN a permis de démontrer l'hétérogénéité du genre *Bacillus* (Priest, 1981). Elle a permis également d'établir la phylogénie des *Proteobacteria* et de démontrer l'hétérogénéité des genres *Pseudomonas*, *Xanthomonas* par exemple (De Vos et De Ley, 1983; De Ley *et al.*, 1986; De Vos *et al.*, 1985a, b). Ainsi au sein d'une même espèce bactérienne, le taux d'homologie de l'ADN est de 70% (Johnson, 1973, 1985).

Les marqueurs moléculaires d'identification et de classification des bactéries.

Les méthodes phénotypiques d'identification et de classification des bactéries souffrent de plusieurs limitations telles que l'expression variable et la reproductibilité entre autres (Olive et Bean, 1999). Afin de pallier à ces limites, plusieurs techniques moléculaires basées sur la caractérisation directe de l'ADN et de l'ARN microbien ont été mises au point au cours de la dernière décennie (Vandamme *et al.*, 1996; Olive et Bean, 1999). L'avènement de la réaction de polymérase en chaîne (Mullis *et al.*, 1986) qui permet de dupliquer en très grande quantité les régions de l'ADN que l'on souhaite étudier, ainsi que la découverte d'enzymes de restriction bactériennes ont permis la mise au point de différents marqueurs moléculaires pour la caractérisation de la diversité bactérienne. Ceux-ci permettent le typage des souches bactériennes et l'établissement des relations phylogénétiques entre les micro-organismes. Ils ont l'avantage d'être faciles à réaliser, rapides, reproductibles et permettent l'établissement de bases de données interchangeables entre laboratoires (Vandamme *et al.*, 1996; Olive et Bean, 1999).

La technique d'électrophorèse en champ pulsé (Pulsed-field gel electrophoresis; PFGE)

Cette technique est utilisée pour l'analyse de grands fragments d'ADN (10-800 kb). Il s'agit dans un premier temps d'extraire l'ADN génomique total, puis de le faire digérer avec des enzymes de restriction et enfin de séparer et de visualiser les différents fragments d'ADN obtenus par électrophorèse sur un gel d'agarose (Vandamme *et al.*, 1996; Olive et Bean, 1999). Elle a été utilisée lors de différentes études taxinomiques notamment pour des classifications intra-spécifique et inter-spécifique. Cependant, elle est plus indiquée pour des études épidémiologiques (Perales et Audicana, 1988; Olsen *et al.*, 1994; Baggesen *et al.*, 1997; Punia *et al.*, 1998; Baggesen *et al.*, 2000).

La technique de polymorphisme dans la longueur des fragments de restriction de l'ADN (Restriction fragment length polymorphism; RFLP)

Cette technique est souvent utilisée pour la différenciation des espèces d'un même genre ou pour le typage des souches d'une même espèce (Lazo *et al.*, 1987; Bouvet *et al.*, 1991; Jacquet *et al.*, 1992; Blanc *et al.*, 1993). Elle permet la détection des réarrangements au sein du génome microbien et celle des mutations apparues au niveau des sites de reconnaissance des enzymes de restriction utilisées pour la digestion de l'ADN (Busse *et al.*, 1996).

Le principe de base de ce marqueur moléculaire est l'extraction de l'ADN suivie de sa digestion par les enzymes de restriction; les différents fragments d'ADN ainsi obtenus seront séparés en fonction de leur taille par migration sur un gel d'agarose, puis transférés soit sur une membrane de nylon soit sur un filtre de nitrocellulose par la technique de Southern Blot (Southern, 1975); Ils seront visualisés par l'intermédiaire de sondes radioactives ou de molécules fluorescentes

(Busse *et al.*, 1996; Olive et Bean, 1999). C'est la haute spécificité des enzymes de restriction qui est exploitée pour la mise en évidence du polymorphisme. Les positions des sites de reconnaissance des enzymes de restriction au niveau d'un locus peuvent varier d'une souche à l'autre donnant ainsi des bandes de tailles différentes (Olive et Bean, 1999).

Le ribotypage.

Cette technique est basée sur le même principe que la technique RFLP, mais dans ce cas-ci, on analyse les gènes codant pour les ARN ribosomaux 16S et 23S, et la région inter-génique (ITS) de l'ARN ribosomal (rARN) bactérien (Olive et Bean, 1999; de Caesar *et al.*, 2001).

Le ribotypage permet d'obtenir un nombre réduit de patrons à analyser et par conséquent une meilleure interprétation des résultats. Il a été utilisé avec succès dans de nombreuses études taxinomiques particulièrement lors de la différenciation des souches bactériennes présentant une forte hétérogénéité au niveau de l'opéron ARNr (Gerner-Smidt, 1992; Joung et Côté, 2000; de Caesar *et al.*, 2001; Shaver *et al.*, 2002).

Un système de ribotypage automatisé a été mis au point et est connu sous le nom de "Ribo Printer" (Andersson *et al.*, 1999).

La technique d'amplification aléatoire d'ADN polymorphe (Random amplified polymorphism DNA; RAPD).

Cette technique de classification repose sur la mise en évidence du polymorphisme moléculaire généré par l'amplification aléatoire de fragments d'ADN grâce à des amorces dont les séquences ont été définies arbitrairement (Williams *et al.*, 1990; Welsh et McClelland, 1990). Les amorces utilisées doivent être de courtes longueurs (10 nucléotides) et l'amplification est suivie de la séparation des fragments

par électrophorèse sur gel d'agarose. L'analyse comparative des patrons obtenus permet d'évaluer la similarité des souches étudiées. Elle permet de détecter les différences entre de petites séquences d'ADN et ainsi de différencier les souches entre elles (Stephan, 1996; Ronimus *et al.*, 1997). C'est une technique surtout utilisée pour des études en épidémiologie clinique et en agro-alimentaire pour l'identification des espèces ou sous espèces bactériennes telles que d'une part *Listeria* (Mazurier et Wernars, 1992; Levett *et al.*, 1993; Czajka et Batt, 1994; Faber et Addison, 1994), *Bacillus licheniformis* (Stephan *et al.*, 1994), *B. cereus* (Stephan, 1996) et *B. thuringiensis* (Brousseau *et al.*, 1993) dans l'Ordre des *Bacillales* et d'autre part, *E.coli* et *Yersinia* dans la Classe des γ -protéobactéries (Cavé *et al.*, 1994; Rasmussen *et al.*, 1994).

Une variante de cette technique, l'amplification en chaîne par polymérisation à partir d'amorces aléatoires (arbitrarily primed PCR; AP-PCR) a été utilisée entre autres pour l'identification de souches bactériennes pathogènes ou non (Welsh et McClelland, 1990; Williams *et al.*, 1993).

La technique de polymorphisme de longueur des fragments amplifiés de l'ADN (Amplified fragment length polymorphism; AFLP).

C'est une nouvelle approche de la mise en évidence du polymorphisme moléculaire des micro-organismes. Mise au point en 1995 par De Vos *et al.*, l'AFLP est basée sur une amplification sélective de fragments de restriction obtenus suite à la digestion de l'ADN génomique (Vandamme *et al.*, 1996; Meijer *et al.*, 1999; Lan et Reeves, 2002). Elle se fait en trois phases successives : une digestion de l'ADN génomique, la fixation d'adaptateurs aux extrémités cohésives de l'ADN à l'aide d'une ADN ligase, une amplification pré-sélective suivie d'une amplification sélective qui permettent de réduire le grand nombre de fragments engendrés par la digestion (Lan et Reeves, 2002). Cette technique a été utilisée avec succès pour

l'étude de la diversité génétique de plusieurs organismes y compris les bactéries telles que *Bacillus anthracis* (Keim *et al.*, 1997) et *Xanthomonas vasicola* (Janssen *et al.*, 1996).

Le séquençage de l'ADN.

Le séquençage de l'ADN fait partie des grands progrès réalisés en biologie moléculaire. Cette technique qui consiste à déterminer l'ordre des nucléotides sur la molécule d'ADN a été mis au point dans les années 1970 (Maxam et Gilbert, 1977; Sanger *et al.*, 1977). Les premières méthodes de séquençage sont la méthode chimique de Maxam et Gilbert qui est une technique de séquençage par coupure chimique et la méthode de synthèse enzymatique de Sanger (Sanger dideoxy sequencing). Les avancées technologiques réalisées au début du 21^{ème} siècle ont permis de mettre au point de nouvelles techniques de séquençage permettant de réduire le coût et le temps nécessaire au séquençage. On distingue, la technique de séquençage automatique (automated sequencing) (Hillis *et al.*, 1996), la technique de pyroséquençage (Ronaghi *et al.*, 1998), les technologies de séquençage de haut débit, la technologie 454 (Margulies *et al.*, 2005) et la technologie CRT Solexa/illumina (Illumina, San Diego, USA).

Le séquençage du gène 16S de l'ARNr.

L'ARN ribosomal s'associe aux protéines pour former le ribosome, siège de la synthèse protéique. Chez les bactéries, le ribosome est formé de deux sous-unités, la sous-unité 30S, ou petite sous-unité composée de l'ARN 16S et de 21 protéines, et la sous-unité 50S ou grande sous-unité composée des ARN 23S et 5S et de 31 protéines (Mears *et al.*, 2002). Le gène 16S de l'ARNr est un gène universel qui a conservé une fonction identique durant l'évolution. Toute mutation pouvant être léthale, il subit de fortes pressions de sélection. Ce gène présente des régions très conservées entre les

espèces d'un même genre, régions qui jouent un rôle biologique critique, et de régions variables (Woese, 1987). Il constitue de ce fait un excellent marqueur moléculaire pour des études phylogénétiques entre les différents groupes de bactéries et pour l'identification de nouvelles espèces bactériennes (Woese *et al.*, 1990; Amann *et al.*, 1995; Cilia *et al.*, 1996; Goto *et al.*, 2000; Sacchi *et al.*, 2002) (Fig. 3).

Les premières études du gène 16S de l'ARNr étaient basées sur l'établissement des catalogues d'oligonucléotides. Le gène 16S était digéré par la ribonucléase T1 et les oligonucléotides ainsi obtenus étaient entièrement séquençés afin de constituer les catalogues (Fox *et al.*, 1977; Stackebrandt *et al.*, 1985; Fox et Stackebrandt, 1987).

La découverte de la PCR ayant permis l'amélioration de la technique de séquençage des gènes, la première méthode d'analyse du gène 16S fut remplacée par le séquençage direct du gène codant pour la sous-unité 16S de l'ARN ribosomal (Busse *et al.*, 1996; Cilia *et al.*, 1996). On assiste alors à l'éclosion de bases de données de séquences du gène 16S de bactéries usuelles mais également de bactéries difficiles à cultiver (Ward *et al.*, 1990; Amann *et al.*, 1995). Parallèlement à cela, de nombreux programmes informatiques nécessaires à l'analyse des données et à la construction des arbres phylogéniques ont été mis au point.

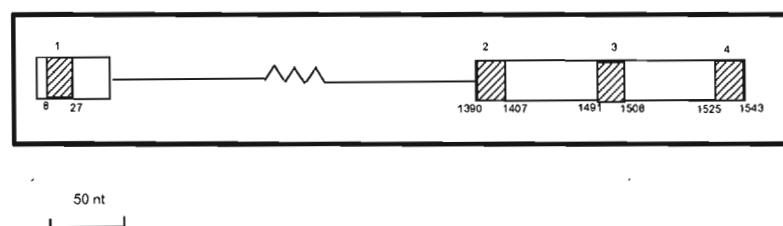


Fig. 3. Régions conservées de l'extrémité 3' du gène 16S de l'ARN ribosomal des bactéries (Adapté de Gürtler et Stanisich, 1996)

Le séquençage de la région inter-génique 16S-23S de l'ARNr.

La région inter-génique 16S-23S ITS, à cause de sa variabilité, a servi à l'étude de la diversité bactérienne au niveau de l'espèce et de la sous-espèce (Gürtler et Stanisich, 1996; Garcia-Martinez *et al.*, 1999; Gonçalves et Rosato, 2002; Wang *et al.*, 2008). La taille de l'ITS varie non seulement en fonction des espèces mais également entre les différents opérons d'une même cellule dans le cas d'opérons multiples (Condon *et al.*, 1995). Le typage par PCR de cette région a permis de démontrer l'existence d'un polymorphisme de longueur entre les espèces de *Bacillus* (Jensen *et al.*, 1993 ; Wunschel *et al.*, 1994; Gürtler et Stanisich, 1996; Daffonchio *et al.*, 1998). De même une partie de l'ITS a été amplifiée et utilisée comme sonde pour l'identification et la phylogénie des espèces de *Bacillus* (de Silva *et al.*, 1998).

La région inter-génique 23S-5S est moins utilisée comme marqueur moléculaire d'identification en raison non seulement de sa grande taille mais aussi parce qu'on ne dispose pas d'autant de données sur les séquences du gène 5S que sur ceux du gène 16S (Yoon *et al.*, 1997).

Les limites des études de systématique bactérienne basées sur l'analyse des séquences des ARN ribosomaux

Les ARN ribosomaux plus particulièrement les gènes 16S et 23S ont fait leur preuve en tant que marqueur moléculaire pour révéler des relations phylogénétiques et pour identifier des espèces bactériennes. Cependant, leur utilisation à ces fins taxinomiques présente des limites (Garcia-Martinez *et al.*, 1999; Kolbert et Persing, 1999).

Le gène 16S.

L'analyse de la séquence du gène 16S communément utilisée pour les études phylogénétiques et celles de la diversité bactérienne présente également des limites. Les gènes codant pour la sous-unité 16S de l'ARNr ont une taille très constante ce qui fait qu'il est techniquement difficile de séparer plusieurs gènes en fonction de leur taille. Ceci limite par conséquent l'analyse comparative de grands nombres d'échantillons (Garcia-Martinez *et al.*, 1999). En plus de cela, malgré sa structure composée d'une alternance de régions variables et très conservées, le nombre de substitutions nucléotidiques est trop faible pour permettre une analyse phylogénétique entre des espèces très proches (Garcia-Martinez *et al.*, 1999 ; Kolbert et Persing, 1999). Ceci a été mis en évidence lors de l'analyse phylogénétique d'espèces appartenant à un même genre, tel que le genre *Bacillus* (Ash *et al.*, 1991a; Rössler *et al.*, 1991). On peut pallier à cette insuffisance en amplifiant en plus du gène 16S, la région inter-génique 16S-23S, région hyper-variable par excellence dont l'analyse révèle un polymorphisme de longueur entre les espèces d'un même genre (Gürtler et Stanisich, 1996).

Le gène 23S.

La sous-unité 23S de l'ARNr est plus indiquée pour la distinction des espèces très proches non seulement à cause de sa taille (environ 2.5 kb) mais aussi parce qu'elle renferme des domaines ayant eu une évolution rapide et contenant de forts taux de mutations (Garcia-Martinez *et al.*, 1999 ; Kolbert et Persing, 1999). Elle est toutefois très peu utilisée comme méthode d'identification des espèces bactériennes.

La bio-informatique

La modernisation de la technique de séquençage a eu pour conséquence la génération d'énormes quantités de données qui sont utilisés pour l'établissement des phylogénies. Le besoin de stocker et de traiter ces données a amené à la création d'une nouvelle discipline qui est la bio-informatique. La bio-informatique est l'ensemble des concepts et techniques permettant l'interprétation de l'information génétique dans le cas des séquences et de l'information structurale tel que le repliement 3D des protéines. Il s'agit ici de la biologie "in silico", par opposition à la biologie *in vivo* ou *in vitro*. Elle a pour objectifs d'interpréter les informations génétique (séquences) et structurale (Repliement 3D des protéines), d'énoncer des hypothèses et de formuler les prédictions telles que la fonction d'un gène.

Résumé de la problématique

Les Eubactéries sont constituées d'espèces très variées quant à leur distribution dans l'environnement, leurs niches, leurs besoins nutritifs et leur métabolisme.

La taxonomie microbienne étant un domaine dynamique, les méthodes de classification évoluent au rythme des progrès technologiques. De nouvelles méthodes d'identification et de classification des espèces bactériennes qui en découlent telles que la phylogénie moléculaire par exemple (Woese, 1987) ont amené à la révision de plusieurs taxa bactériens.

Dans l'Ordre des *Bacillales*, il y a une dizaine d'années, les espèces bactériennes de la famille des *Bacillaceae* étaient regroupées essentiellement dans le genre *Bacillus*. Au cours de la dernière décennie, l'accumulation de données,

essentiellement moléculaires, a amené la création d'une dizaine de nouveaux genres. Ainsi une classification numérique basée sur l'utilisation de caractères phénétiques a permis de classer 368 souches de *Bacillus* en 79 groupes (Priest *et al.*, 1988). Au début des années 2000, une partie de la séquence du gène 16S rDNA (Goto *et al.*, 2000) ainsi que des patrons de restriction du gène 16S (Joung et Côté, 2002) ont été utilisés respectivement pour une identification rapide et une classification des espèces de *Bacillaceae* et des genres proches. Contrairement à la région codante du gène 16S RNAr qui est très conservée, la région inter-génique 16S-23S (ITS) est hyper-variable ce qui fait qu'elle a servi à l'étude de la diversité des procaryotes (espèces et sous-espèces) (Gürtler et Stanisich, 1996; García-Martínez *et al.*, 1999).

Dans la Classe des γ -protéobactéries, il y a une quinzaine d'année, la classification des espèces du genre *Xanthomonas* était basée sur un critère unique qui était la spécificité de l'hôte (Vauterin *et al.*, 1990). Des études basées sur l'hybridation ADN-ADN ont amené à la réorganisation de ce genre en révélant 20 espèces génomiques (Vauterin *et al.*, 1995).

Un marqueur d'ADN basé sur une courte séquence nucléotidique située dans la région 3' du gène 16S ARNr et 5' de la région régionale inter-génique (ITS) 16S-23S pour la reconstruction de la phylogénie de bactéries Gram-positif du genre *Bacillus* a été mis au point par Xu et Côté en 2003. Il s'agit d'un marqueur PCR facile d'utilisation du point de vue technique qui était composé de 220 paires de bases (pb) à savoir les 150 dernières pb en 3' du gène 16S ARNr et en 5' des 70 premières paires de bases de la région inter-génique 16S-23S (ITS). Au cours de cette étude, un total de 40 espèces du genre *Bacillus* étaient analysées. La phylogénie de ce genre a été établie à l'aide du marqueur de 220 pb. L'arbre phylogénétique construit est similaire celui du gène 16S ARNr qui est considéré comme arbre consensus et utilisé comme référence. Le gène 16S est très conservé entre les espèces d'un même genre tandis que la région inter-génique est une région hyper-variable par excellence dont

l'analyse a révélé un polymorphisme de longueur entre les espèces d'un même genre (Gürtler et Stanisich, 1996). L'extrémité 5' de la région inter-génique 16S-23S (ITS) en revanche est conservée à l'intérieur d'une même espèce (Xu et Côté, 2003). Le marqueur ainsi constitué pourrait servir aux études systématiques intra et inter-spécifiques.

Compte tenu de la grande diversité bactérienne en particulier à l'intérieur de l'Ordre des *Bacillales* et de la Classe des γ -protéobactéries, nonobstant toutes les techniques développées pour classifier ces bactéries, la classification des bactéries est en perpétuelle réorganisation.

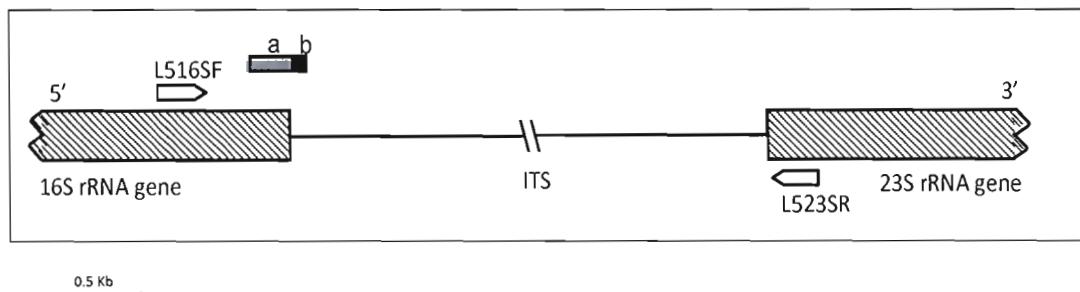


Fig. 4. Représentation schématique du marqueur universel d'ADN.

Objectifs

L'objectif général de cette thèse est d'évaluer un marqueur d'ADN basé sur une courte séquence nucléotidique située dans l'extrémité 3' du gène 16S ARNr et l'extrémité 5' de la région inter-génique 16S-23S (ITS) pour l'identification et la classification des bactéries Gram-positif de l'Ordre des *Bacillales* et Gram-négatif de la Classe des γ -protéobactéries à différents niveaux taxinomiques (Fig. 4).

Les quatre chapitres de la thèse ont les objectifs concrets suivants :

- 1- Établir la phylogénie des bactéries de l'Ordre des *Bacillales* à l'aide d'un marqueur d'ADN de 220 pb. Il s'agit d'une combinaison des 150 pb de l'extrémité 3' du gène 16S de l'ARNr et des 70 premières paires de bases de l'extrémité 5' de la région inter-génique 16S-23S (ITS). Un total de 72 espèces et souches de l'Ordre des *Bacillales* réparties en huit familles et 21 genres seront analysées.
- 2- Évaluer le pouvoir de résolution du marqueur de 220 pb à un niveau hiérarchique inférieur de l'Ordre des *Bacillales* : le groupe *Bacillus cereus* : *Bacillus anthracis*, *B. cereus*, *B. mycoides* et *B. thuringiensis*.
- 3- Développer un marqueur similaire à celui des bactéries Gram-positif pour les bactéries Gram-négatif. La phylogénie de la Classe des γ -protéobactéries est établi à l'aide d'un marqueur d'ADN de 232 pb : 157 pb de l'extrémité 3' du gène 16S de l'ARNr et 75 premières paires de bases de l'extrémité 5' du de la région inter-génique 16S-23S (ITS). Un total de 64 γ -protéobactéries réparties en 13 Ordres, 22 Familles, 40 genres et 59 espèces seront analysés.
- 4- Évaluer le pouvoir de résolution d'un marqueur d'ADN de 224 pb à un niveau taxonomique inférieur de bactéries Gram-négatif de la classe des γ -protéobactéries : le genre *Xanthomonas*.

Hypothèses

Hypothèse générale.

L'amélioration des techniques d'identification et de classification des espèces bactériennes a permis de révéler l'hétérogénéité phénotypique et/ou génotypique des plusieurs espèces bactériennes tels que la famille des *Bacillaceae* et le genre *Xanthomonas* par exemple. Il s'en est suivi la révision de plusieurs taxa bactériens.

L'opéron *rrn* a une structure semblable chez toutes les bactéries et il existe plusieurs copies de l'allèle de cet opéron chez une même espèce. Le gène 16S est très conservé entre les espèces d'un même genre tandis que la région inter-génique est hyper-variable à l'intérieur d'une même espèce. La région inter-génique du fait de sa variabilité a été utilisée pour l'étude de la diversité des procaryotes au niveau espèce et sous espèce. Une courte séquence d'ADN composée des dernières paires de bases de l'extrémité 3' du gène 16S de l'ARNr et des premières paires de bases de l'extrémité 5' de la région inter-génique 16S-23S (ITS) développée comme marqueur moléculaire peut servir à l'identification des espèces bactériennes.

À cause de l'importance des bactéries sur les plans économique, médical et environnemental, il est indispensable de développer un tel marqueur d'ADN pour l'identification rapide et la classification des espèces bactériennes à différents niveaux taxonomiques.

Hypothèse chapitre 1

Il est possible d'étudier les relations phylogénétiques des espèces de l'Ordre des *Bacillales* grâce à l'élaboration de dendrogrammes basés sur l'analyse comparative des séquences de notre marqueur moléculaire. De plus en plus d'ADN génomiques bactériens sont connus et les séquences des gènes codant pour l'ARNr 16S sont déterminées et conservées dans des banques de données telles que GenBank, EMBL, etc. Un court marqueur d'ADN composé des dernières paires de bases de l'extrémité 3' du gène 16S de l'ARNr et des premières paires de bases de l'extrémité 5' de la région inter-génique 16S-23S (ITS) peut servir à l'identification rapide et la classification des bactéries de l'Ordre des *Bacillales*.

Hypothèse chapitre 2

Certaines espèces bactériennes de par leur position sur l'arbre phylogénétique des *Bacillus* ont besoin de plus de caractérisation par des analyses complémentaires. C'est le cas de *Bacillus anthracis*, *Bacillus cereus*, *Bacillus thuringiensis* et *Bacillus weihenstephanensis* qui constituent le groupe I du genre *Bacillus* ou groupe de *Bacillus cereus*. Une analyse comparative des séquences du marqueur de 220 pb de plusieurs souches de chacune des espèces précitées permettrait de les distinguer.

Hypothèse chapitre 3

Si le marqueur développé pour l'Ordre des *Bacillales* s'avérait robuste, il deviendrait possible d'évaluer un marqueur similaire pour l'identification rapide et la classification des bactéries Gram-négatif de la Classe des γ -protéobactéries.

Hypothèse chapitre 4

Une fois le marqueur évalué au niveau de la Classe des γ -protéobactéries, il serait possible de l'évaluer à un niveau taxonomique inférieur : le genre *Xanthomonas*.

CHAPITRE I

Phylogeny of the Order *Bacillales* inferred from 3'
16S rDNA and 5' 16S-23S ITS nucleotide
sequences

Sabarimatou Yakoubou, Dong Xu et Jean-Charles Côté
Natural Science, 2(9): 990-997.

Abstract

A short 220 bp sequence was used to study the taxonomic organization of the bacterial Order *Bacillales*. The nucleotide sequences of the 3' end of the 16S rDNA and the 16S-23S Internal transcribed spacer (ITS) were determined for 32 *Bacillales* species and strains. The data for 40 additional *Bacillales* species and strains were retrieved directly from Genbank. Together, these 72 *Bacillales* species and strains encompassed eight families and 21 genera. The 220 bp sequence used here covers a conserved 150 bp sequence located at the 3' end of the 16S rDNA and a conserved 70 bp sequence located at the 5' end of the 16S-23S ITS.

A neighbor-joining phylogenetic tree was inferred from comparative analyses of all 72 nucleotide sequences. Eight major Groups were revealed. Each Group was sub-divided into sub-groups and branches. In general, the neighbor-joining tree presented here is in agreement with the currently accepted phylogeny of the Order *Bacillales* based on phenotypic and genotypic data. The use of this 220 bp sequence for phylogenetic analyses presents several advantages over the use of the entire 16S rRNA genes or the generation of extensive phenotypic and genotypic data. This 220 bp sequence contains 150 bp at the 3' end of the 16S rDNA which allows discrimination among distantly related species and 70 bp at the 5' end of the 16S-23S ITS which, owing to its higher percentage of nucleotide sequence divergence, adds discriminating power among closely related species from same genus and closely related genera from same family. The method is simple, rapid, suited to large screening programs and easily accessible to most laboratories.

Introduction

In the 1st Edition of Bergey's Manual of Systematic Bacteriology (1986), the genus *Bacillus*, a member of the Class *Bacilli*, encompassed the Gram-positive, rod-shaped, endospore-forming, either obligate or facultative aerobe bacteria (Claus and Berkeley 1986). A total of 34 species and 26 additional *species incertae sedis* were described. The genus was highly heterogeneous, exhibiting a wide range of nutritional requirements, physiological and metabolic diversity and DNA base composition. In the following two decades, numerical classifications (Priest *et al.*, 1988) 16S rDNA sequence alignments (Ash *et al.*, 1991a), characterizations at the genotypic and phenotypic levels of selected *Bacillus* species, all led to the creation of several new genera (briefly reviewed in Xu and Côté 2003).

In 2003, we developed a method for the identification, classification and phylogenetic analyses of *Bacillus* species and species from closely related genera (Xu and Côté 2003). The method was simple, rapid, suited to large screening programmes and easily accessible to most laboratories. In summary, the method relied on comparisons of a 220 bp nucleotide sequence marker among *Bacillus* species. This 220 bp marker was a combination of a 150 bp sequence at the 3' end of the 16S rRNA gene and a 70 bp sequence at the 5' end of the 16S-23S ITS sequence. In our original study, a total of 40 species was analyzed. We showed that the phylogeny inferred from the 220 bp marker was in agreement with then current classifications based on phenetic and molecular data, with exceptions. It revealed species and genera which appeared misassigned and for which additional characterization appeared warranted.

In the 2nd Edition of Bergey's Manual of Systematic Bacteriology (Ludwig *et al.*, 2009), a new taxonomy of the Class *Bacilli*, is presented. It is the result of phylogenetic and principal-component analyses of comprehensive datasets of 16S

rDNA sequences (Ludwig *et al.*, 2009; Garrity *et al.*, 2005). The *Bacillus* genus belongs to the Order *Bacillales*. The Order *Bacillales* contains nine families: *Alicyclobacillaceae*, *Bacillaceae*, *Listeriaceae*, *Paenibacillaceae*, *Pasteuriaceae*, *Planococcaceae*, *Sporolactobacillaceae*, *Staphylococcaceae* and *Thermoactinomycetaceae*. All nine families contain a total of 51 genera.

Today, we further assess the usefulness of the 220 bp marker by extending its analyses beyond the genus *Bacillus*, to a higher taxa level, the Order *Bacillales*. Whereas our first study focused on species from the genus *Bacillus* and species from closely related genera (Xu and Côté 2003), we report here the phylogenetic analyses of 72 species and strains from eight *Bacillales* families and 21 genera.

Materials and Methods

Bacterial strains and culture conditions.

A total of 72 strains of the Order *Bacillales* were used in this study. These encompassed eight families, 21 genera and 67 species. A total of 31 *Bacillales* strains were obtained from the "Deutsche Sammlung von Mikroorganismen und Zellkulturen" (DSMZ) GmbH, Braunschweig, Germany and were grown according to the DSMZ guidelines ([http://www.dsmz.de/microorganisms /media_list.php](http://www.dsmz.de/microorganisms/media_list.php)). *Jeotgalibacillus alimentarius* was obtained from the "Czech Collection of Microorganisms" (CCM), Masaryk University, Brno, Czech Republic and grown according to the CCM guidelines (<http://www.sci.muni.cz/ccm/index.html>). The nucleotide sequences of the 40 remaining strains were retrieved directly from GenBank. All bacterial strains and their sources are listed in Table 1.

Table 1-1. *Bacillale* families, genera, species and strains used in this study.

Familles, genres et espèces	Origine et Souches	No accession GenBank Étude*	GenBank **	Groupes	Sous-groupes
<i>Alicyclobacillaceae</i>					
<i>Alicyclobacillus</i>					
<i>acidocaldarius</i> subsp. <i>acidocaldarius</i>	DSM 446	EU723605	V		
<i>acidocaldarius</i> subsp. <i>rittmannii</i>	DSM 11297	EU723607	V		
<i>acidoterrestris</i>	DSM 3922	EU723608	V		
<i>cycloheptanicus</i>	DSM 4006	EU723610	V		
<i>herbarius</i>	DSM 13609	EU723613	V		
<i>hesperidum</i>	DSM 12489	EU723615	V		
<i>Bacillaceae</i>					
<i>Amphibacillus</i>					
<i>fermentum</i>	DSM 13869	VIII	Non groupées		
<i>tropicus</i>	DSM 13870	VIII	iii		
<i>xylanus</i>	DSM 6626	VIII	iii		
<i>Anoxybacillus</i>					
<i>flavithermus</i>	WK1	VIII	Non groupées		
<i>Bacillus</i>					
<i>clausii</i>		VI	i		
<i>subtilis</i>	168	VI	i		
<i>thuringiensis</i>		VI	ii		
<i>thuringiensis</i> var. <i>konkukian</i>	97-27	VI	ii		
<i>weiheinstphanensis</i>	KBAB4	VI	ii		
<i>Filobacillus</i>					
<i>milensis</i>	DSM 13259		Non groupées		
<i>Geobacillus</i>					
<i>caldoxylolyticus</i>	DSM 12041	EU723625	I		
<i>kaustophilus</i>	HTA426	BA000043	I		

* Séquences générées au cours de cette étude. ** Séquences obtenues sur GenBank

Table 1-1. *Bacillale* families, genera, species and strains used in this study.

Familles, genres et espèces	Origine et Souches	Étude*	No accession GenBank GenBank **	Groupes	Sous-groupes
<i>Bacillaceae</i>					
<i>Geobacillus</i>					
<i>kaustophilus</i>	DSM 7263		EU723629	I	
<i>stearothermophilus</i>	DSM 22		EU723631	I	
<i>subterraneus</i>	DSM 113552		EU723636	I	
<i>thermocatenulatus</i>	DSM 730		EU723638	I	
<i>thermodenitrificans</i>	NG80-2		CP000557	I	
<i>thermodenitrificans</i>	DSM 465		EU723643	I	
<i>thermoglucosidasius</i>	DSM 2542		EU723644	I	
<i>thermoleovorans</i>	DSM 5366		EU723648	I	
<i>uzenensis</i>	DSM 13551		EU723651	I	
<i>Gracilibacillus</i>					
<i>diplosauri</i>	DSM 11125		EU723655	VIII	
<i>halotolerans</i>	DSM 11805		EU723657	VIII	ii
<i>Halobacillus</i>					
<i>sp.</i>	SA-Hb6				
<i>Oceanobacillus</i>					
<i>iheyensis</i>	HTE831		AB367166	VII	i
<i>Virgibacillus</i>					
<i>marismortui</i>	DSM 12325		EU723664	VIII	ii
<i>pantothenticus</i>	DSM 26		EU723672	VIII	ii
<i>proomii</i>	DSM 13055		EU723675	VIII	ii
<i>saderigens</i>	DSM 11483		EU723666	VIII	ii

* Séquences générées au cours de cette étude. ** Séquences obtenues sur GenBank

Table 1-1. Bacillale familles, genera, species and strains used in this study

Families, genera and species	Source/Strains	No accession GenBank This study*	GenBank Retrieved **	Groups	Sub-groups
<i>Listeriaceae</i>					
<i>Listeria</i>					
<i>inoccua</i>	Clip11262			VI	iii
<i>monocytogenes</i>	EGD-e	AL592022	VI	iii	iii
<i>welshimeri</i>	SLCC5334	AL591824	VI	iii	iii
<i>Paenibacillaceae</i>					
<i>Aneurinibacillus</i>					
<i>aneariniticus</i>	DSM 5562	EU723616			
<i>migulanus</i>	DSM 2895	EU723617			
<i>thermoaerophilus</i>	DSM 10154	EU723618			
<i>Brevibacillus</i>					
<i>agri</i>	ATCC 51360		AF478091		
<i>borstelensis</i>	ATCC 51668		AF478093		
<i>brevis</i>	ATCC8246		AY478094		
<i>brevis</i>	NBRC10059		AP008955		
<i>chosinensis</i>	ATCC 51359		AF478095		
<i>parabrevis</i>	ATCC 8186		AF478097		
<i>formosus</i>	ATCC51669		AF478096		
<i>Paenibacillus</i>					
<i>alginolyticus</i>	ATCC 51185		AF478104		
<i>alvei</i>	ATCC 6344		AF478098		
<i>chondroitinus</i>	ATCC 51184		AF478105		
<i>larvae</i>	ATCC 9545		AF487106		
<i>lautus</i>	ATCC 43898		AF487100		
<i>lenitimorbus</i>	ATCC 8244		AY763503		
<i>macerans</i>	ATCC 43899		AF478101		
<i>pabuli</i>	ATCC 14707		AF478102		

* Séquences générées au cours de cette étude. ** Séquences obtenues sur GenBank

Table 1-1. *Bacillale* families, genera, species and strains used in this study.

Families, genera and species	Source/Strains	No accession GenBank This study*	Retrieved **	Groups	Sub-groups
<i>Paenibacillaceae</i>					
<i>Paenibacillus</i>	KLN 3	CP001656			
<i>popilliae</i>	JDR-2	DQ062684			
<i>sp.</i>	ATCC 43897	AF478103			
<i>Pasteuriaceae</i>					
<i>Pasteuria</i>					
<i>ramosa</i>	PS	AY762091	II		
<i>penetrans</i>	CJ-1	AY123968	II		
<i>Planococcaceae</i>					
<i>Jeotgalibacillus</i>					
<i>alimentarius</i>	CCM 7134	EU723660	VI		
<i>Marinibacillus</i>					
<i>marinus</i>	DSM 1297	EU723661	VI		
<i>Ureibacillus</i>					
<i>terrenus</i>	DSM 12654	EU723667	VI		
<i>thermosphaericus</i>	DSM 10633	EU723670	VI		
<i>Sporolactobacillaceae</i>					
<i>Sporolactobacillus</i>					
<i>terrae</i>	M-1116	D16289.1	Non groupées		
<i>Staphylococcaceae</i>					
<i>Staphylococcus</i>					
<i>aureus</i>	RF122	AJ938182	VII		
<i>armosus</i>	TM300	AM295250	VII		
<i>epidermidis</i>	RP62A	CP000029	VII		
<i>haemolyticus</i>	JCSCL1435	AP006716	VII		
<i>aprophyticus</i>	ATCC 15305	AP008934	VII		
<i>Macrococcus</i>					
<i>caseolyticus</i>	JCSCL5402	AP009484	VII		

* Séquences générées au cours de cette étude. ** Séquences obtenues sur GenBank

Escherichia coli strain TOP10 (Invitrogen Inc., Burlington, ON, Canada) was used for cloning PCR fragments. Strain TOP10 was cultured on Luria-Bertani (LB) agar plates to select transformants or in LB broth to multiply the cells, with shaking at 180–200 rpm at 37°C, 1 h.

DNA extraction.

Bacterial cells were washed with TESS buffer [10 mM Tris/HCl (pH 8.0), 1 mM Na₂EDTA, 0.1 M NaCl and 0.1 % Sarkosyl (*N*-lauroylsarcosine)] and resuspended in TE buffer [10 mM Tris/HCl (pH 8.0), 1 mM Na₂EDTA]. Cells were lysed with 10 mg/ml lysozyme and 0.1 % SDS. The subsequent phenol/chloroform extractions and ethanol precipitation were carried out as described by Sambrook and Russel (Sambrook and Russel, 2001).

Recombinant plasmid from *E. coli* strain TOP10 was isolated using the alkaline-lysis method (Stephen *et al.*, 1990) with some modifications as described elsewhere (Xu and Côté 2003).

Amplification of the 3' end 16S rDNA and the 16S–23S ITS region.

A pair of primers: L516SF (5'-TCGCTAGTAATCGCGGATCAGC-3') and L523SR (5'-GCATATCGGTGTTAGTCCCGTCC-3'), (Xu and Côté 2003) was used for the amplification of the 3' end of 16S rDNA, the 16S–23S ITS region and the 5' end of 23S rDNA. Amplification was performed in a Thermal Cycler 9600 (Perkin Elmer, Waltham, MA, USA) and the reaction mixtures contained 50 ng template DNA, 0.25 µM each primer, 200 µM dNTP, 1.5 mM MgCl₂ and 1.25 U *Taq* DNA polymerase (QIAGEN Inc. Mississauga, ON, Canada) in a final volume of 50 µl. PCR was performed under the following conditions: 45 s at 95°C and then 30 cycles

of 15 s at 94°C, 30 s at 53°C and 90 s at 72°C. Amplification products were visualized on agarose gels.

Cloning and sequencing methods.

The amplified DNAs were cloned into a pCRII-TOPO cloning vector using the TOPO TA cloning kit (Invitrogen, Inc.), following the manufacturer's instructions. *Escherichia coli* strain TOP10 transformants were selected on LB agar plates containing kanamycin (50 µg/ml), 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal) (40 µg/ml). Multiple clones were submitted for further analyses for each *Bacillales* species. The recombinant plasmids were isolated using the modified alkaline-lysis method, digested with *Eco*RI and visualized on agarose gels to confirm the presence of an inserted fragment.

The nucleotide sequences of cloned fragments were determined by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977), using a capillary array automated DNA sequencer (ABI 3730xl DNA analyzer; Applied Biosystems, Foster City, CA, USA). The sequences of both strands were determined.

Sequence analysis.

The 3' end of the 16S rDNA and the 16S–23S ITS of the 32 *Bacillales* species and strains sequenced in this study were used for analysis. Forty sequences publicly available from GenBank were added for comparison purposes to cover a wider range of *Bacillales* families and genera, for a total of 72 *Bacillales* species and strains. A neighbor-joining tree was constructed (Saitou and Nei, 1987) based on the alignment of the 72 3' end of the 16S rDNA and the 5' end of the 16S–23S ITS sequences. The tree was bootstrapped using 1,000 random samples of sites from the alignment, all using CLUSTAL W software (Thompson *et al.*, 1994) at the DNA Data Bank of

Japan (DDBJ) (<http://clustalw.ddbj.nig.ac.jp/top-e.html>), with the Kimura's parameter method (Kimura, 1980). The neighbor-joining phylogenetic tree was drawn using TreeView (version 1.6.6) (Page 1996; Page 2000).

Results

Two primers, one located about 200 nt upstream from the 3' end of the 16S rRNA gene, the other about 80 nt downstream from the 5' end of the 23S rRNA gene (Fig.1), were used to amplify the last 200 bp of the 16S rRNA gene and the entire 16S-23S Internal transcribed spacer (ITS) region from 32 *Bacillales* species and strains (Table 1). The amplicons ranged in length from 450 to 1,200 bp. The number of amplicons per strain ranged from 1 to 6. A subset of these results is shown in Fig. 2. Each amplicon was cloned and its nucleotide sequence determined. The homologous DNA sequences from 40 more *Bacillales* species and strains were retrieved directly from GenBank and added in the study (Table 1). Together, these 72 *Bacillales* species and strains belong to eight *Bacillales* families and 21 genera.

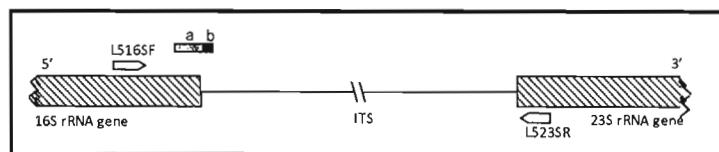


Fig. 1-1. Schematic representation of the 16S and 23S rRNA genes separated by an Internal transcribed spacer (ITS). Orientations and positions of the primers used for amplification, L516SF and L523SR, are shown. The contiguous small grey and black boxes, indicated by the letters "a" and "b" correspond to the last 150 bp at the 3' end of the 16S rRNA gene and the first 70 bp at the 5' end of the ITS, respectively. Together, these boxes correspond to the 220 bp marker used in this study.

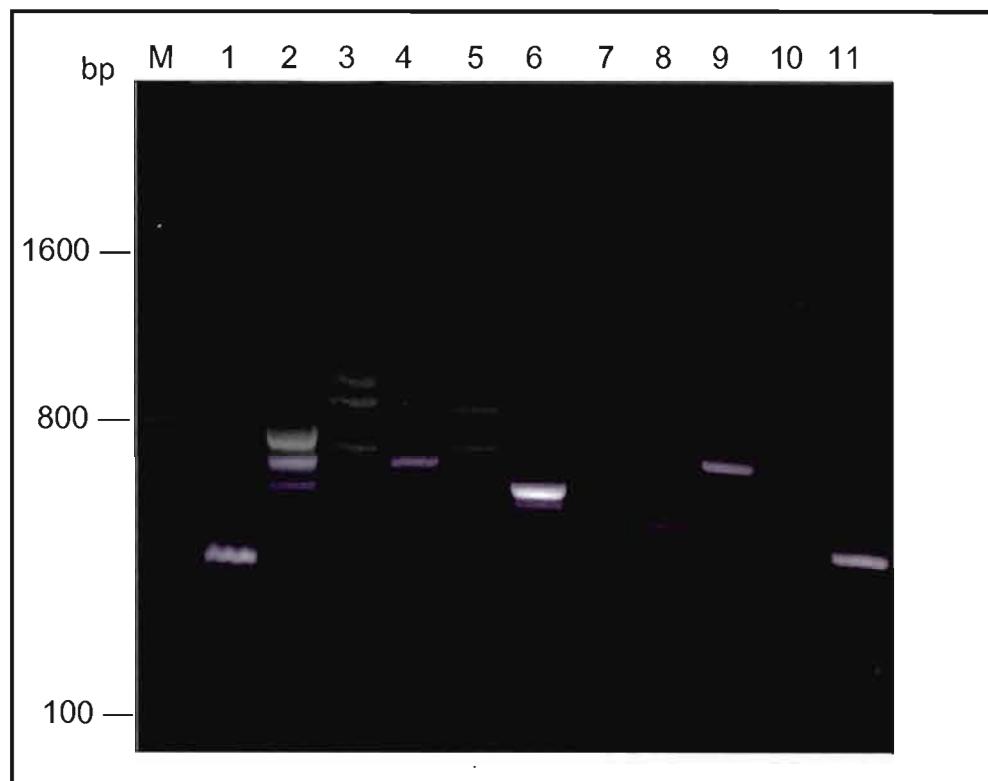


Fig. 1-2. Agarose gel electrophoresis of the amplification products in selected species in the Order *Bacillales* using the L516SF/L523SR primer pair.

Lane M, 100 bp DNA marker; lane 1, *Alicyclobacillus acidocaldarius* subsp. *acidocaldarius*; lane 2, *Alicyclobacillus herbarius*; lane 3, *Geobacillus uzenensis*; lane 4, *Gracilibacillus halodurans*; lane 5, *Geobacillus kaustophilus*; lane 6, *Amphibacillus tropicus*; lane 7, *Virgibacillus puumii*; lane 8, *Virgibacillus salexigens*; lane 9, *Marinibacillus marinus*; lane 10, *Aneurinibacillus aneurinilyticus*; lane 11, *Filobacillus milensis*.

The last 150 bp located at the 3' end of 16S rDNA and the first 70 bp located at the 5' end of 16S-23S ITS, were combined into a 220 bp sequence. A multiple alignment of these nucleotide sequences from the 72 *Bacillale* species and strains was performed (supplementary data) and a bootstrapped neighbor-joining tree was constructed (Fig. 3).

At first sight, eight major Groups are revealed at the 80% nucleotide sequence identities, Group I to VIII. Group I contains all eleven *Geobacillus* species and strains. *Alicyclobacillus herbarius* forms a single branch between Groups I and II. Group II contains both *Pasteuriaceae* species. Group III contains ten species and strains from two genera of the *Paenibacillaceae* family. Two sub-groups can be revealed. Sub-group i encompasses the three *Aneurinibacillus* species. They share at least 82% nucleotide sequence identities. Sub-group ii encompasses the seven *Brevibacillus* species and strains. They share at least 80% nucleotide sequence identities. Group IV contains all 11 *Paenibacillus* species. Group V contains five of the six species and strains in the *Alicyclobacillaceae* family. *Alicyclobacillus herbarius* is the sixth species, and forms a single branch as explained above, because it shares less than 80% nucleotide sequence identities with members of Group V. Group VI is more heterogenous. It contains 12 species and strains from five genera from three families: the *Planococcaceae*, *Bacillaceae* and *Listeriaceae*. Four sub-groups and two branches are revealed. *Jeotgalibacillus alimentarus* forms the first branch of Group VI, and belongs to the *Planococcaceae* family. Sub-group i contains two *Bacillus* species of the *Bacillaceae* family, *B. clausii* and *B. subtilis*. *Marinibacillus marinus* forms the second branch of Group VI. It is the second genus of the *Planococcaceae* family. Sub-group ii contains three highly related *Bacillus* species and strains, the two *B. thuringiensis* strains and *B. weihenstephanensis*. Sub-group iii contains all three *Listeria* species of the *Listeriaceae* family. Sub-group iv contains both species of the third *Planococcaceae* genus, *Ureibacillus*. Next, a branch is formed by *Sporolactobacillus terrae*. This genus is the only one known in the *Sporolactobacillaceae* family. This is followed by a branch formed by *Filobacillus milensis*, a member of the *Bacillaceae* family. Group VII contains five *Staphylococcus* and one *Macrococcus* species. Both genera belong to the *Staphylococcaceae* family. Group VIII contains 12 species from six genera, all in the *Bacillaceae* family. Three subgroups and thee branches can be revealed. Sub-group i contains *Oceanobacillus* and *Halobacillus*.

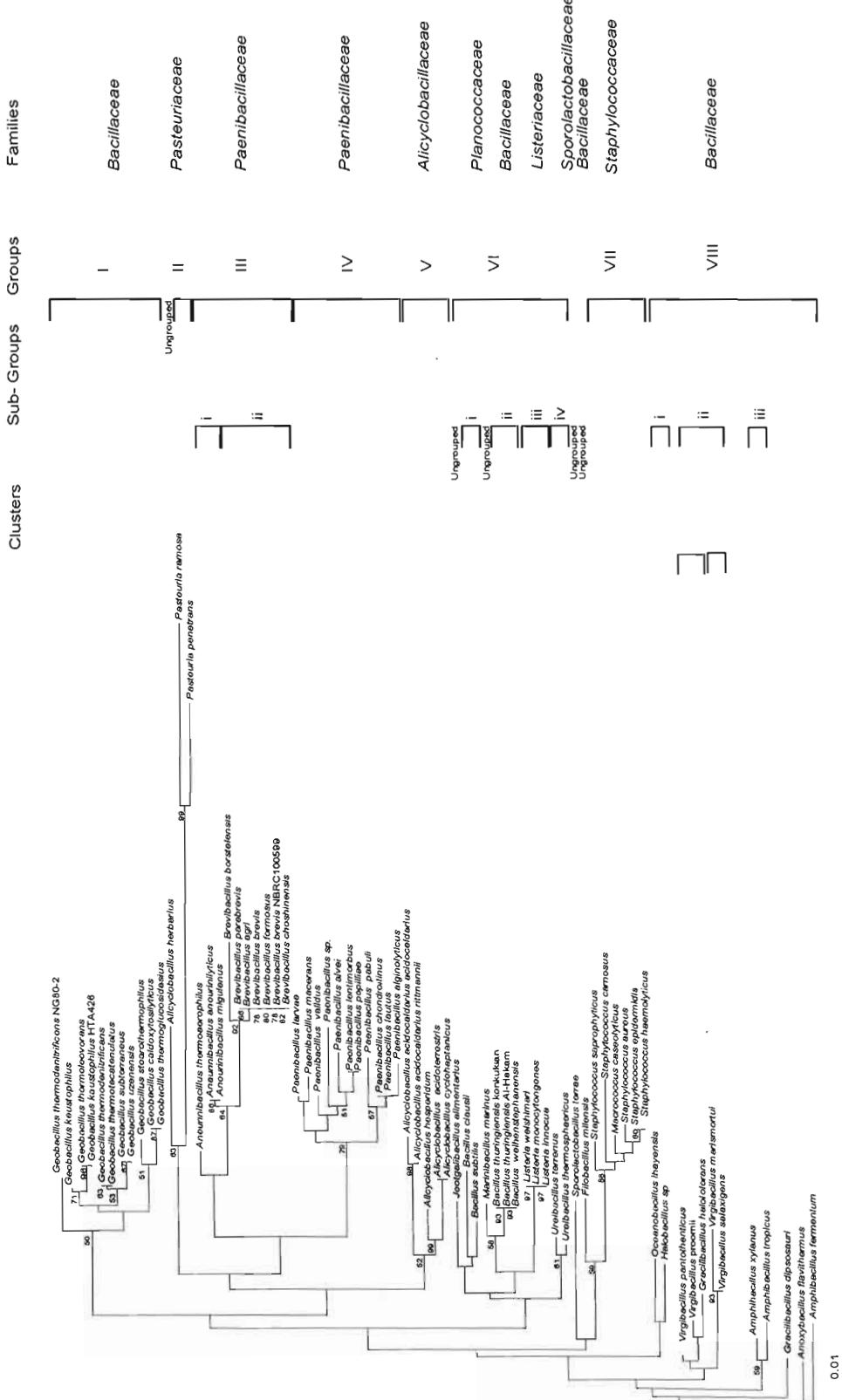


Fig. 1-3. Bootstrapped neighbor-joining tree of 72 *Bacillales* species and strains inferred from the alignment of the 220 bp marker.

Major Groups are indicated in capital roman numerals. Sub-groups are indicated in lower case roman numerals. Bootstrap values higher than 50% are indicated (expressed as percentage of 1000 replication). The horizontal bar represents 1% nt difference.

They share 87% nucleotide sequence identities. The grouping of *Oceanobacillus* and *Halobacillus* species in sub-group i is in agreement with the work of Lu *et al.*, based on morphology, physiological and biochemical properties, and genetic data (Lu *et al.*, 2001). Sub-group ii contains five species from two genera. It can be further divided into two clusters. The first one contains *Virgibacillus (Vir.) pantothenicus*, *Vir. pرومii* and *Gracilibacillus halotolerans*, the second cluster contains *Vir. marismortui* and *V. salexigens*. The subdivision of all four *Virgibacillus* species into two sub-groups is in agreement with the work of Heyrman *et al.*, based on genetic, chemotaxonomic and phenotypic data (Heyrman *et al.*, 2003). Sub-group iii contains two of the three *Amphibacillus (Am.)* species, *Am. xylinus* and *Am. tropicus*. Group VIII is completed by three branches: *Gracilibacillus dipsosauri*, *Anoxybacillus flavithermus* and *Amphibacillus fermentum*. All members of Group VIII are halophilic and alkaliphilic bacilli. This grouping is in agreement with the one proposed by Zhilina *et al.*, based on physiology and genetic data (Zhilina *et al.*, 2001).

When each of the eight major Groups of this neighbor-joining tree are analyzed separately, strains from same species, species from same genus and genera from same family are grouped together. Each Group corresponds to a single *Bacillales* family, exclusive of other families, with one exception, Group VI which contains three families. In some cases, Groups were divided into sub-groups which corresponded to genera, with one exception, Group VIII where sub-groups encompassed different genera. The family *Sporolactobacillaceae* forms a branch between Groups VI and VII. Three *Bacillales* families: *Bacillaceae*, *Paenibacillaceae* and *Alicyclobacillaceae*, show some level of heterogeneity. Genera of the *Bacillaceae* family are found into three Groups, Groups I, VI and VIII, and in a branch between Groups VI and VII. Genera of the *Paenibacillaceae* family are found into two Groups, Groups III and IV. Analysis of the *Alicyclobacillaceae* family reveals a different story. Five of the six species and strains are clustered together in Group V.

The sixth, *Alycyclobacillus herbarius*, is more distant. It forms a branch between Groups I and II. This is in agreement with the work of Goto *et al.*, (Goto *et al.*, 2000). Although distinct from all other *Alycyclobacillaceae* based on genomic data, including 16S rDNA sequences, *Alycyclobacillus herbarius* was grouped with the family based on the presence of ω -cycloheptane fatty acids Goto *et al.* (Goto *et al.*, 2000).

Discussion

In a 2003 study (Xu and Côté 2003), on *Bacillus* and closely-related genera, a multiple alignment of the 3' end of the 16S rRNA sequence showed that the last 157 nucleotides shared extensive identities among closely related species from same genus. This 157 nucleotide sequence, however, was not conserved among species from different genera. In the same study, a multiple alignment of the 16S-23S Internal transcribed spacer (ITS) sequence showed that the first 70 nucleotides were conserved between alleles of the same strain and between alleles of different strains from same species. This sequence, however, was not conserved among alleles of different species of the same genus. These two sequences, the last 150 bp at the 3' end of the 16S rRNA gene and the first 70 bp at the 5' end of the 16S-23S rRNA ITS, were combined into a single 220 bp marker. This marker was used to infer the phylogeny of *Bacillus* species and species from closely related genera. It could cluster *Bacillus* species and species from closely related genera into taxa akin to genera and could also distinguish closely related species. In this 2003 study, a total of 40 species was analyzed.

In the current study, we further assessed the usefulness of the 220 bp marker at a higher taxonomic level, the Order *Bacillales*. A total of 72 *Bacillales* species and strains from eight *Bacillales* families and 21 genera were covered. The number of *Bacillus* species included in this current study on *Bacillales* was deliberately kept low since this genus had already been covered extensively in our earlier study (Xu et Côté 2003). The neighbor-joining tree presented here was compared with the revised road map of the Order *Bacillales* shown in the 2nd Edition of Bergey's Manual of Systematic Bacteriology (Ludwig *et al.*, 2009). This revised road map is a consensus phylogenetic tree of the Order *Bacillales* (Ludwig *et al.*, 2009). It is the consensus tree inferred from numerous phylogenetic and principal-component analyses of comprehensive datasets of 16S rDNA sequences (Ludwig *et al.*, 2009; Garrity *et al.*, 2005). Our phylogenetic tree presented here is in agreement with the currently accepted phylogeny of the Order *Bacillales*, based on phenotypic and genotypic data. It is, in general, in agreement with the revised road map of the Order *Bacillales* (Ludwig *et al.*, 2009). In addition, some bacterial species that were not grouped at the genus level in our neighbor-joining tree, exemplified by *Alicyclobacillus herbarius*, were also confirmed by others to be different based on phenotypic and genotypic data (Goto *et al.*, 2000).

The main discrepancy between our results, obtained with the 220 bp marker, and the revised road map shown in the 2nd Edition of Bergey's Manual of Systematic Bacteriology (Ludwig *et al.*, 2009), rests on the grouping of the *Bacillaceae* family. In our study, members of the *Bacillaceae* family are present in three of the eight Groups. In the revised road map (Ludwig *et al.*, 2009), two major *Bacillaceae* groups are presented. It is recognized, however, that some species have been misassigned to the *Bacillaceae* family. The revised road map is constructed based on 16S rDNA sequences (Ludwig *et al.*, 2009). Our 220 bp marker contains 150 bp from 16S rDNA and 70 bp from ITS. Owing to its higher rate of nucleotide substitutions, this 70 bp

adds discriminating power among species from same genera and genera from same family. As indicated by Ludwig *et al.*, and as shown here, the reorganization of the *Bacillaceae* family is still a work in progress (Ludwig *et al.*, 2009).

The use of this 220 bp marker presents several advantages over the use of the entire 16S rRNA gene or the generation of extensive phenotypic and genotypic data in phylogenetic analyses. As shown in an earlier study (Xu and Côté 2003), the method is simple, rapid, suited to large screening programmes and easily accessible to most laboratories. We have shown here that it can group *Bacillales* families and genera in accordance with established phylogenies. Because the 220 bp marker shows a higher percentage of nucleotide sequence divergence than the 16S rRNA gene, it can better discriminate among closely related *Bacillales* species. It can also reveal *Bacillales* species which may appear misassigned and for which additional characterization appear warranted.

In conclusion, in an earlier study (Xu and Côté 2003), a 220 bp marker, based on 3' end of 16S rRDA and 5' end of 16S-23S rRNA ITS, was developed and used to classify species in the *Bacillus* genus and in closely related genera. Here, we showed that this 220 bp marker could be used to reconstruct the phylogeny at a higher taxa level: the Order *Bacillales*. We are planning to follow-up this work by assessing the resolving power of this marker in reconstructing the phylogeny at a lower taxa level: the *Bacillus cereus* group, sensu lato.

Recently, in parallel, a similar maker was tested and shown to be able to reconstruct the phylogeny of the Class γ -proteobacteria (Yakoubou and Côté 2010).

CHAPITRE II

Assessment of a short phylogenetic marker based
on comparisons of 3' end 16S rDNA and
5' end 16S-23S ITS nucleotide sequences of the
Bacillus cereus group

Sabarimatou Yakoubou et Jean-Charles Côté
Natural Science, 2 (10):1113-1118.

Abstract

A short phylogenetic marker previously used in the reconstruction of the Order *Bacillales* and the genus *Bacillus* was assessed here at a lower taxa level: species in the *Bacillus cereus* group: *B. anthracis*, *B. cereus*, *B. thuringiensis* and *B. weihenstephanensis*. This marker is 220 bp in length. It is a combination of 150 bp at the 3' end of the 16S rDNA and 70 bp at the 5' end of the 16S-23S ITS sequence. Three additional *Bacillus* species, *B. halodurans*, *B. licheniformis* and *B. subtilis*, and *Clostridium tetani* were included for comparison purposes. A total of eight bacterial species and 12 strains were analyzed.

A bootstrapped neighbor-joining tree was inferred from comparative analyses of all allelic sequences of the bacterial species and strains under study. Based on its topology, four major Groups were revealed at the 90% nucleotide sequence identities, Group I to IV. Group I contains all alleles of the *Bacillus cereus* group. Group II contains all alleles of *B. halodurans*. Group III contains all alleles of *B. licheniformis* and *B. subtilis*. Group IV contains all alleles of *Clostridium tetani*.

The 220 bp phylogenetic marker used here could resolve different species from different genera. At the genus level, distant species could be distinguished. Very closely-related species, however, were undistinguishable. Species in the *B. cereus* group, most notably *B. cereus*, *B. anthracis* and *B. thuringiensis*, could not be distinguished. After successfully inferring the phylogenies of the Order *Bacillales* and the genus *Bacillus*, we have met the resolving limit of this short phylogenetic marker: *B. cereus*, *B. anthracis* and *B. thuringiensis*.

Introduction

The *Bacillus cereus* group comprises six genetically highly related species: *B. cereus* sensu stricto, *B. anthracis*, *B. thuringiensis*, *B. weihenstephanensis*, *B. mycoides* (Claus and Berkeley, 1986) and *B. pseudomycoides* (Nakamura, 1998). They are Gram-positive, rod-shaped, endospore-forming, either obligate or facultative aerobic bacteria (Claus and Berkeley, 1986).

Bacillus cereus is an ubiquitous soil bacterium. It can be a contaminant of a variety of foods: meats, vegetables and dairy products (Drobniewski, 1993; Schoeni and Wong, 2005). It can cause diarrheal, and emetic food poisoning syndromes (Kramer and Gilbert, 1989). It can also be the etiologic agent of some opportunistic infections (Das *et al.*, 2001; Le Scanff *et al.*, 2006). *Bacillus anthracis* is the etiologic agent of anthrax, an acute disease in herbivorous mammals, transmissible to other animals, including humans (Logan and De Vos, 2009). This species has been studied and developed as a biological weapon (Inglesby *et al.*, 2002). Virulent strains of *B. anthracis* carry two plasmids, pXO1 (181 kb) and pXO2 (96 kb) which may be transmitted to others members of *Bacillus cereus* group (Turnbull, 2002). *Bacillus thuringiensis* is an insect pathogen. It is characterized by the synthesis upon sporulation of a parasporal inclusion body. This inclusion body is made of proteins, the δ -endotoxins, which are toxic to several insect larvae (Höfte and Whiteley, 1989; Garcia-Robles *et al.*, 2001) and other invertebrates (Feitelson, 1993). *B. thuringiensis* formulations have been developed for the control of insect pests in agriculture and forestry (Schneppf *et al.*, 1998; Otvos *et al.*, 2005; Bravo *et al.*, 2007) and for the control of insect vectors of human diseases such as malaria, yellow fever, onchocerciasis, etc (Guillet *et al.*, 1997). *Bacillus weihenstephanensis* is a psychotolerant species characterized by the ability to grow at 7°C and the absence of growth at 43°C. It is also characterized by the presence of specific signature

sequences on the 16S rRNA gene (small subunit ribosomal RNA gene) and the *cspA* gene (gene encoding the major cold shock protein) (Lechner *et al.*, 1998). *B. mycoides* is characterized by the formation of rhizoid colonies and the absence of motility (Nakamura and Jackson, 1995). *B. pseudomycoides* is phenotypically similar to *B. mycoides* and is distinguished by DNA relatedness and fatty acid composition (Nakamura, 1998).

The 16S rDNA is the macromolecule of choice in the reconstruction of bacterial phylogenies (Woese *et al.*, 1990; Amann *et al.*, 1995; Cilia *et al.*, 1996; Goto *et al.*, 2000; Sacchi *et al.*, 2002). The 16S rDNA, however, cannot distinguish among species in the *Bacillus cereus* group (Ash *et al.*, 1991a; Ash *et al.*, 1991b). Genomic approaches have been used in an attempt to elucidate the genetic diversity of three highly closely related species in the *B. cereus* group: *B. cereus*, *B. anthracis* and *B. thuringiensis*. They appear as a single species on the basis of genetic evidence (Vilas-Boa *et al.*, 2007).

In a previous study, a 220 bp marker was developed and used to infer the phylogeny of species in the genus *Bacillus* and closely-related genera (Xu and Côté, 2003). This marker was a combination of the last 150 bp at the 3' end of the 16S rDNA and the first 70 bp at the 5' end of the 16S-23S rDNA internal transcribed spacer (ITS). More recently, we assessed the usefulness of the 220 bp marker at a higher taxonomic level, the Order *Bacillales* (Yakoubou *et al.*, 2010). This marker showed several advantages over the use of 16S rDNA sequences or the generation of extensive phenotypic and genotypic data in phylogenetic analyses. First, the 150 bp at the 3' end of the 16S rDNA allowed discrimination among distantly related species. Owing to its higher rate of nucleotide substitutions, the 70 bp at the 5' end of the 16S-23S rDNA (ITS) added discriminating power among closely related species from same genus and closely related genera from same family. Because of its higher percentage of nucleotide sequence divergence than the 16S rDNA, the 220 bp marker

could better discriminate among closely related *Bacillus* (Xu and Côté, 2003) and *Bacillales* species (Yakoubou *et al.*, 2010). Second, the method was simple, rapid, suited to large screening programs and easily accessible to most laboratories. Third, the marker also revealed species which appeared misassigned and for which additional characterization appeared warranted.

In the current study, we further analyze the resolving power of this short marker in inferring phylogenies at a much lower taxa level: the *Bacillus cereus* group.

Materials and Methods

Bacterial species and strains.

Four species in the *Bacillus cereus* group: *B. anthracis*, *B. cereus*, *B. thuringiensis* and *B. weihenstephanensis* were analyzed. Three additional *Bacillus* species, *B. halodurans*, *B. licheniformis* and *B. subtilis*, and *Clostridium tetani* were included for comparison purposes. A total of eight bacterial species and 12 strains were analyzed (Table 1). They were selected on the basis that their complete genome sequences were freely available in GenBank at the National Center for Biotechnology Information (NCBI) completed microbial genomes database (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>, August 2009). *Bacillus mycoides* and *B. pseudomycoides* were not included because their complete genome sequences have not been determined.

Sequences

The 16S rDNA and 16S–23S ITS for the 12 bacterial species and strains were retrieved from GenBank, for a total of 129 allelic sequences. The last 150 bp at the 3' end of 16S rDNA and the first 70 bp at the 5' end of 16S–23S ITS were merged into a

Table 2-1. Bacterial species used in this study.

Genera	Species	Strains	GenBank accession no.
<i>Bacillus</i>	<i>anthracis</i>	Ames	AE016879
		Ames Ancestor	AE017334.2
		Sterne	AE017225.1
	<i>cereus</i>	ATCC 14579	AE016877.1
		ATCC 10987	AE017194.1
		E33L	CP000001.1
	<i>thuringiensis</i> serovar konukian	97-27	AE017355.1
	<i>weihenstephanensis</i>	KBAB4	NC_010184.1
	<i>halodurans</i>	C-125	BA000004.3
	<i>licheniformis</i>	ATCC 14580	AE017333.1
	<i>subtilis</i> subsp. <i>subtilis</i>	168	AL009126.3
	<i>Clostridium tetani</i>	E88	AE015927.1

single 220 bp sequence for each of the 129 alleles under study as described before (Xu and Côté, 2003). This 220 bp sequence will be used as a phylogenetic marker for the 12 bacterial species and strains under study.

Phylogenetic analyses

All 129 allelic sequences were aligned using ClustalW (Thompson *et al.*, 1994) (data not shown). A neighbor-joining tree was constructed (Saitou and Nei, 1987), based on the alignment of the 129 alleles of the 220 bp sequence (Fig. 1). The tree was bootstrapped using 1,000 random samples. The neighbor-joining tree was

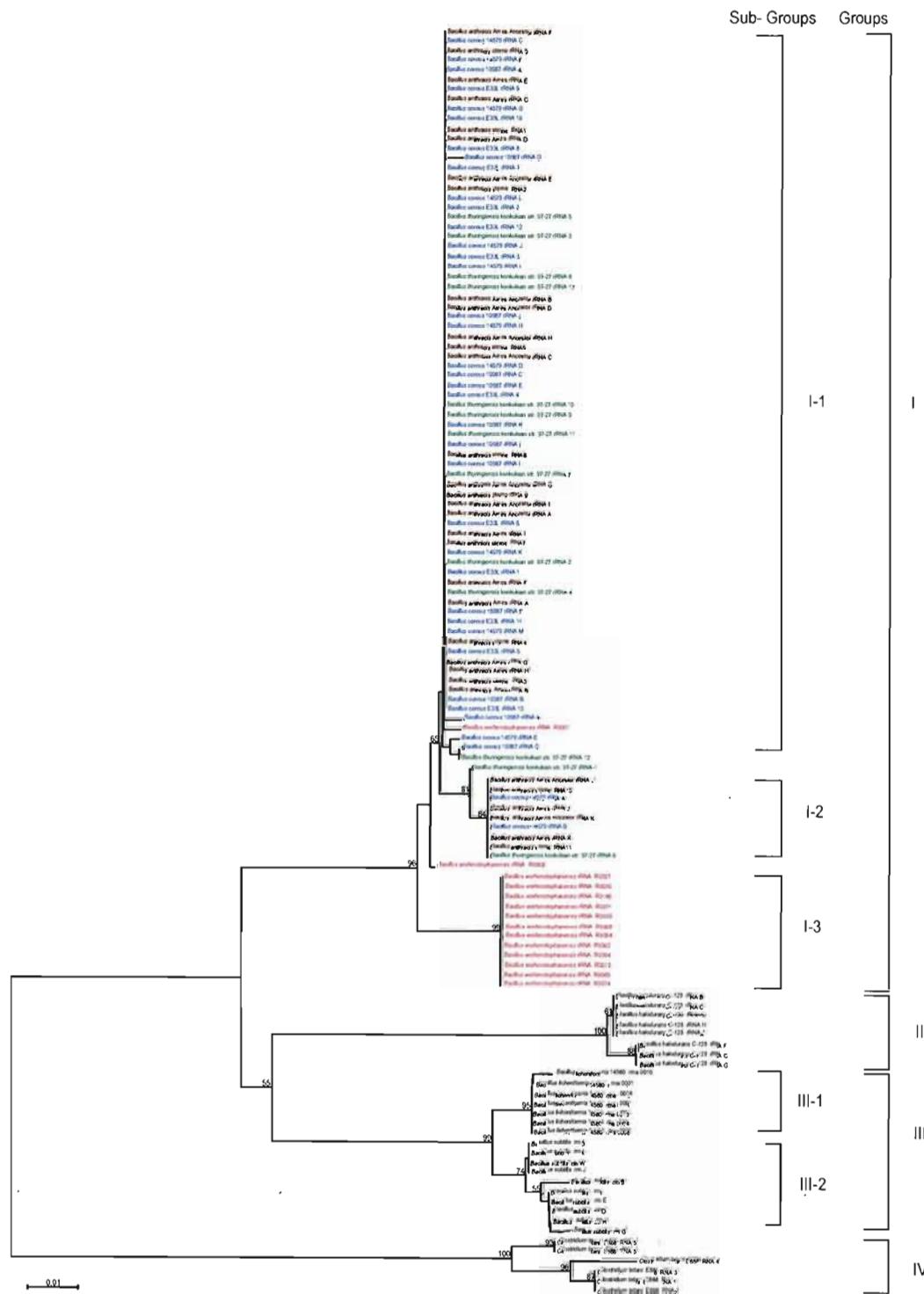


Fig.2-1. Bootstrapped neighbor-joining tree inferred from comparative alignment of the 220 bp marker from the 129 alleles from the 12 bacterial species and strains under study.

Major Groups are indicated in capital roman numerals. Sub-groups are indicated in arabic numerals. Bootstrap values higher than 50% are indicated (expressed as percentage of 1000 replication). The horizontal bar represents 1% nt difference.

Bacillus anthracis's strains and alleles; *Bacillus cereus*'s strains and alleles; *Bacillus thuringiensis*'s alleles; *Bacillus weihenstephanensis*'s alleles

drawn and printed with Tree explorer, all components of the Molecular Evolutionary Genetics Analysis (MEGA, version 3.1) software package (Kumar *et al.*, 2004).

Results and Discussion

In a previous study, a 220 bp sequence was developed as a DNA marker and used to infer the phylogeny of species in the Gram-positive genus *Bacillus* and closely-related genera (Xu and Côté, 2003). This marker was a combination of the last 150 bp at the 3' end of the 16S rDNA and the first 70 bp at the 5' end of the 16S-23S rDNA internal transcribed spacer (ITS). More recently, we assessed the usefulness of the 220 bp marker by extending its analyses at a higher taxonomic level, the Gram-positive Order *Bacillales* (Yakoubou *et al.*, 2010). In parallel, a similar marker was used to infer the phylogeny of the Gram-negative Class γ -proteobacteria (Yakoubou and Côté, 2010). In the current study, we further analyze the resolving power of this marker in inferring the phylogeny at a much lower taxa level: the *Bacillus cereus* group.

A bootstrapped neighbor-joining phylogenetic tree was inferred from comparative analyses of the 220 bp marker from the 129 alleles from the bacterial species and strains under study (Fig. 1). Four major Groups were revealed based on the topology of the neighbor-joining tree at the 90% nucleotide sequence identities, Group I to IV. Group I contains all alleles of the species in the *Bacillus cereus* group. Group II contains all alleles of *B. halodurans*. Group III contains all alleles of *B. licheniformis* and *B. subtilis*. Group IV contains all alleles of *Clostridium tetani*. Based on nucleotide sequence identities, sub-groups and branches can be revealed. Group I can be sub-divided into three sub-groups at the 95% nucleotide sequence identities. Sub-group I-1 encompasses 27 alleles from the *B. anthracis* strains, 36

alleles from the *B. cereus* strains, 11 alleles from *B. thuringiensis* and one allele from *B. weihenstephanensis*. Sub-group I-2 encompasses six alleles from *B. anthracis*, two alleles from *B. cereus* and one allele from *B. thuringiensis*. Sub-group I-3 encompasses 12 alleles from *B. weihenstephanensis*. A branch corresponding to an allele from *B. weihenstephanensis* is present between sub-groups I-2 and I-3. Group II contains all eight alleles from *B. halodurans*. They show at least 20% nucleotide sequence divergences with alleles from the other *Bacillus* species. Group III contains all alleles from *B. licheniformis* and *B. subtilis*. It can be sub-divided into two sub-groups at the 95% nucleotide sequence identities. Sub-group III-1 encompasses all seven alleles from *B. licheniformis*. Sub-group III-2 encompasses all ten alleles from *B. subtilis*. Group IV contains all six alleles from *Clostridium tetani*. These alleles show at least 26% nucleotide sequence divergence with alleles from species and strains in the genus *Bacillus*.

In accordance with our previous work on the Order *Bacillales*, the 220 bp sequence used as a phylogenetic marker was able to group alleles from same species for *B. halodurans*, *B. licheniformis*, *B. subtilis*, and *Clostridium tetani*, respectively. However, this 220 bp sequence could not group most alleles from same species, exclusive of alleles from others, for the *B. cereus* group. Sub-group I-1 is heterogeneous. It contains alleles from all four species from the *B. cereus* group. The close proximity of *B. cereus*, *B. anthracis* and *B. thuringiensis* is in agreement with previous works based on whole-genome DNA hybridization (Kaneko *et al.*, 1978), pulsed-field gel electrophoresis (PFGE) (Carlson *et al.*, 1994), multilocus enzyme electrophoresis (MEE) (Helgason *et al.*, 2000), amplified fragment length polymorphism (AFLP) fingerprinting (Ticknor *et al.*, 2001) and multilocus sequence typing (MLST) (Helgason *et al.*, 2004; Olsen *et al.*, 2007), which showed that all three species are genetically highly related. They appear as a single species on the basis of genetic evidence (Vilas-Boas *et al.*, 2007). Sub-groups I-2 is more

homogeneous. It mostly contains alleles from *B. anthracis*. Sub-groups I-3 is homogeneous. It only contains alleles from *B. weihenstephanensis*.

As shown earlier, on the genus *Bacillus* (Xu and Côté, 2003) and the Order *Bacillales* (Yakoubou *et al.*, 2010), this 220 bp sequence contains 150 bp at the 3' end of 16S rDNA which allowed discrimination among distantly related species and 70 bp at the 5' end of 16S-23S ITS which, owing to its higher percentage of nucleotide sequence divergence, added resolving power among closely related species. Here, species in the *B. cereus* group, most notably *B. cereus*, *B. anthracis* and *B. thuringiensis*, are too closely related to be discriminated with the 220 bp sequence previously used as a phylogenetic marker. Our work, however, has shown that the alleles in sub-group I-3 could distinguish *B. weihenstephanensis* from all other species.

Conclusion

Previous genetic analyses have shown that *B. cereus*, *B. anthracis* and *B. thuringiensis* should be regarded as a single species. We have shown here that a 220 bp marker, used to reconstruct the phylogeny of the Order *Bacillales* and the family *Bacillaceae*, was unable to discriminate between these three highly-related species. We have reached the limit of the resolving power of the 220 bp sequence as a phylogenetic marker: *B. cereus*, *B. anthracis* and *B. thuringiensis*.

CHAPITRE III

Phylogeny of γ -proteobacteria inferred
from comparisons of 3' end 16S rRNA gene and
5' end 16S-23S ITS nucleotide sequences

Sabarimatou Yakoubou et Jean-Charles Côté
Natural Science, 2 (6), 535-543.

Abstract

The phylogeny of γ -proteobacteria was inferred from nucleotide sequence comparisons of a short 232 nucleotide sequence marker. A total of 64 γ -proteobacterial strains from 13 Orders, 22 families, 40 genera and 59 species were analyzed. The short 232 nucleotide sequence marker used here was a combination of a 157 nucleotide sequence at the 3' end of the 16S rRNA gene and a 75 nucleotide sequence at the 5' end of the 16S-23S Internal Transcribed Spacer (ITS) sequence.

Comparative analyses of the 3' end of the 16S rRNA gene nucleotide sequence showed that the last 157 bp were conserved among strains from same species and less conserved in more distantly related species. This 157 bp sequence was selected as the first part in the construction of our nucleotide sequence marker. A bootstrapped neighbor-joining tree based on the alignment of this 157 bp was constructed. This 157 bp could distinguish γ -proteobacterial species from different genera from same family. Closely related species could not be distinguished.

Next, an alignment of the 16S-23S ITS nucleotide sequences of alleles from same bacterial strain was performed. The first 75 bp at the 5' end of the 16S-23S ITS was highly conserved at the intra-strain level. It was selected as the second part in the construction of our nucleotide sequence marker.

Finally, a bootstrapped neighbor-joining tree based on the alignment of this 232 bp sequence was constructed. Based on the topology of the neighbor-joining tree, four major Groups, Group I to IV, were revealed with several sub-groups and clusters. Our results, based on the 232 bp sequence were, in general, in agreement with the phylogeny of γ -proteobacteria based on the 16S rRNA gene. The use of this 232 bp sequence as a phylogenetic marker presents several advantages over the use of

the entire 16S rRNA gene or the generation of extensive phenotypic and genotypic data in phylogenetic analyses. First, this marker is not allele-dependant. Second, this 232 bp marker contains 157 bp from the 3' end of the 16S rRNA gene and 75 bp from the 5' end of the 16S-23S ITS. The 157 bp allows discrimination among distantly related species. Owing to its higher rate of nucleotide substitutions, the 75 bp adds discriminating power among closely related species from same genus and closely related genera from same family. Because of its higher percentage of nucleotide sequence divergence than the 16S rRNA gene, the 232 bp marker can better discriminate among closely related γ -proteobacterial species. Third, the method is simple, rapid, suited to large screening programs and easily accessible to most laboratories. Fourth, this marker can also reveal γ -proteobacterial species which may appear misassigned and for which additional characterization appear warranted.

Keywords: γ -proteobacteria; 16S rRNA; 16S-23S ITS; Phylogeny.

Introduction

The phylum proteobacteria or “purple bacteria and their relatives” encompasses bacteria with a wide variety of phenotype and physiological attributes and habitats (Stackebrandt *et al.*, 1988; Gupta, 2000). Proteobacteria have been classified based on the homology of 16S ribosomal RNA or by hybridization of ribosomal DNA with 16S and 23S ribosomal RNA (Fox *et al.*, 1980; Woese *et al.*, 1985a; Woese, 1987; De Ley, 1992). They have been subdivided in five major classes: α , β -, γ -, δ - and ε - (Woese *et al.*, 1984a et b; Woese *et al.*, 1985a et b).

Most γ -proteobacteria are Gram-negative. This class comprises 14 Orders and more than 200 genera. The γ -proteobacteria exhibit a wide range of metabolic diversity. Most are chemo-organotrophs, some are phototrophs or chemolithotrophs (Stackebrandt *et al.*, 1988; Gupta, 2000; Brenner *et al.*, 2005; Kersters *et al.*, 2006). This class include several medically and scientifically important bacteria. Some genera are human (*Klebsiella*, *Shigella*, *Salmonella*, *Yersina*, *Vibrio*), animal (*Pasteurella*) or plant pathogens (*Pseudomonas*, *Xanthomonas*, *Xylella*). Others are obligate endosymbionts (*Buchnera*, *Sodalis*, *Wigglesworthia* and *Coxiella*) (Brenner *et al.*, 2005; Kersters *et al.*, 2006; Belda *et al.*, 2005). Because of their biological importance, γ -proteobacteria are extensively studied.

The 16S ribosomal RNA (rRNA) gene has been established as the macromolecule of choice for phylogenetic analyses (Woese, 1987; Woese *et al.*, 1990). The current phylogeny of γ -proteobacteria is based on the homology of 16S rDNA nucleotide sequences (Fox *et al.*, 1980; Woese *et al.*, 1985a; Woese, 1987; De Ley, 1992; Kersters *et al.*, 2006; Ludwig et Klenk 2005).

The 16S-23S internal transcribed spacer (ITS) region is more variable than the 16S rRNA gene. It has been used, among others, in the study of specific γ -proteobacterial diversity at the species level, including *Escherichia*, *Haemophilus*, *Xanthomonas*, *Klebsiella* and *Pseudomonas* (Anton *et al.*, 1998; Giannino *et al.*, 2001; Gonçalves et Rosato, 2002; Wang *et al.*, 2008; Tambong *et al.*, 2009).

Additional approaches, based on different genes, have been used for the study of γ -proteobacterial phylogeny (Belda *et al.*, 2005; Kunisawa, 2001; Lerat *et al.*, 2003); Brown and Volker, 2004; Ciccardelli *et al.*, 2006; Mrazek *et al.*, 2006; Lee and Côté, 2006). Very recently, Gao *et al.*, (2009) have used a combination of phylogenomic and comparative genomic approaches to reconstruct the phylogeny of γ -proteobacteria.

In an earlier work on the bacterial Gram-positive *Bacillus* genus and related genera (Xu and Côté, 2003), a short 220 bp nucleotide sequence "marker" was used to reconstruct their phylogeny. This 220 bp marker was a combination of a 150 bp sequence at the 3' end of the 16S rRNA gene and a 70 bp sequence at the 5' end of the 16S-23S ITS sequence. Owing to its higher rate of nucleotide substitution, the 70 bp sequence at the 5' end of the 16S-23S ITS sequence added a greater discriminatory power among closely related species than 16S rRNA gene nucleotide sequences alone. They showed that the phylogeny inferred from the 220 bp marker was in agreement with then current classifications based on phenetic and molecular data. The marker also identified species which appeared mis-assigned. It also created new clusters suggesting the creation of new taxa levels. In a very recent study, we have tested whether or not this marker could reconstruct the phylogeny of the bacterial Gram-positive Order of the *Bacillales* (Yakoubou *et al.*, 2010).

In the current study, we further assess the usefulness of a similar marker among 13 of the 14 γ -proteobacterial Orders. The last 157 bp at the 3' end of the 16S rRNA

gene was combined with the first 75 bp at the 5' end of the 16S-23S Internal Transcribed Spacer (ITS) to yield a single 232 bp DNA marker. This marker was used to reconstruct the phylogeny of γ -proteobacteria. A total of 64 γ -proteobacteria from 13 Orders, 22 families, 40 genera and 59 species was analyzed.

Materials and Methods

Bacterial species and strains

A total of 64 γ -proteobacterial species and strains were analyzed. They were selected on the basis that their complete genome sequences were freely available in GenBank, at the National Center for Biotechnology Information (NCBI) completed microbial genomes database (<http://www.ncbi.nlm.nih.gov/genomes/MICROBES/Complete.html>). They encompassed 13 Orders, 22 families, 40 genera and 59 species. These 13 γ -proteobacterial Orders included six *Aeromonadales* families, one *Cardiobacteriales* family, two *Chromatiales* families, one *Enterobacteriales* family, two *Legionellales* families, one *Methylococcales* family, two *Oceanospirillales* families, one *Pasteurellales* family, two *Pseudomonadales* families, two *Thiotrichales* families, one *Vibrionales* family and one *Xanthomonadales* family. All bacterial species and strains and the GenBank accession number for their fully sequenced genome are listed in Table 1.

Sequences analysis

The 16S rRNA gene nucleotide sequences were retrieved from GenBank (Table 1) for the 64 γ -proteobacteria species and strains under study. First, all 64 sequences were aligned using ClustalW (Thompson *et al.*, 1994) (data not shown).

Table 3-1. γ -proteobacteria strains used in this study.

Orders, Families, Genera, Species	Source/Strain	GenBank accession no.
<i>Aeromonadales</i>		
<i>Aeromonadaceae</i>		
<i>Aeromonas hydrophyla</i>	ATCC 7966	NC_008570.1
<i>Aeromonas salmonida</i>	A449	NC_009348.1
<i>Alteromonadales</i>		
<i>Alteromonadaceae</i>		
<i>Marinobacter aquaeolei</i>	VT8	NC_008740.1
<i>Saccharophagus degradans</i>	2-40	NC_007912.1
<i>Colwelliaceae</i>		
<i>Colwellia psychrerythraea</i>	34H	NC_003910.7
<i>Idiomarinaceae</i>		
<i>Idiomarina ihoihensis</i>	L2TR	NC_006512.1
<i>Pseudoalteromonadaceae</i>		
<i>Pseudoalteromonas atlantica</i>	T6c	NC_008228.1
<i>Pseudoalteromonas haloplanktis</i>	TAC125	NC_007481.1
<i>Shewanellaceae</i>		
<i>Shewanella amazonensis</i>	SB2B	NC_008700.1
<i>Shewanella denitrificans</i>	OS217	NC_007954.1
<i>Shewanella frigidimarina</i>	NCIMB 400	NC_008345.1
<i>Shewanella oneidensis</i>	MR-1	NC_004347.1
<i>Cardiobacteriales</i>		
<i>Cardiobacteriaceae</i>		
<i>Dichelobacter nodosus</i>	VCS1703A	NC_009446.1
<i>Chromatiales</i>		
<i>Chromatiaceae</i>		
<i>Nitrosococcus oceanii</i>	ATCC 19707	NC_007484.1
<i>Ectothiorhodospiraceae</i>		
<i>Alkalilimnicola ehrlichei</i>	MLHE-1	NC_008453.1
<i>Halorhodospira halophila</i>	SL1	NC_008789.1
<i>Enterobacteriales</i>		
<i>Enterobacteriaceae</i>		
<i>Buchnera aphidicola</i>	APS	NC_002528.1
<i>Citrobacter koseri</i>	ATCC BAA-895	NC_009792.1
<i>Enterobacter sakazakii</i>	ATCC BAA-894	NC_009778.1
<i>Enterobacter sp</i>	638	NC_009436.1
<i>Escherichia coli</i>	CFT073	NC_004431.1
<i>Escherichia coli</i>	O157:H7 Sakai	NC_002695.1
<i>Klebsiella Pneumoniae</i>	342	NC_011283.1
<i>Photobacterium luminescens</i>	NTUH-K2044	NC_012731.1
<i>Salmonella enterica</i>	TT01	NC_005126.1
<i>Salmonella enteric</i>	Ty2	NC_004631.1
<i>Salmonella enteric</i>	arizonae	NC_010067.1
<i>Shigella boydii</i>	Sb227	NC_007613.1
<i>Shigella dysenteriae</i>	Sd197	NC_007606.1
<i>Shigella flexineri</i>	2457T	NC_004741.1
<i>Sodalis glossinidius</i>	morsitans	NC_007712.1
<i>Wigglesworthia glossinidiae</i>		NC_004344.2

Table 3-1. γ -proteobacteria strains used in this study.

Orders, Families, Genera, Species	Source/Strain	GenBank accession no.
<i>Enterobacteriales</i>		
<i>Enterobacteriaceae</i>		
<i>Yersinia enterocolitica</i>	8081	NC_008800.1
<i>Yersinia pestis</i>	CO92	NC_003143.1
<i>Yersinia pseudotuberculosis</i>	IP31758	NC_009708.1
<i>Legionellales</i>		
<i>Coxiellaceae</i>		
<i>Coxiella burnetii</i>	Dugway 5j108-111	NC_009727.1
<i>Coxiella burnetii</i>	RSA 493	NC_002971.3
<i>Legionellaceae</i>		
<i>Legionella pneumophila</i>	Lens	NC_006369.1
<i>Legionella pneumophila</i>	Corby	NC_009494.1
<i>Methylococcales</i>		
<i>Methylococcaceae</i>		
<i>Methylococcus capsulatus</i>	Bath	NC_002977.6
<i>Oceanospirillales</i>		
<i>Alcanivoraceae</i>		
<i>Alcanivorax borkumensis</i>	SK2	NC_008260.1
<i>Halomonadaceae</i>		
<i>Chromohalobacter salexigens</i>	DSM 3043	NC_007963.1
<i>Pasteurellales</i>		
<i>Pasteurellaceae</i>		
<i>Actinobacillus pleuropneumoniae</i>	L20	NC_0090053.1
<i>Actinobacillus succinogenes</i>	130Z	NC_009655.1
<i>Haemophilus ducreyi</i>	35000HP	NC_002940.2
<i>Haemophilus influenzae</i>	Rd KW20	NC_000907.1
<i>Mannheimia succiniciproducens</i>	MBEL55E	NC_006300.1
<i>Pasteurella multocida</i>	Pm70	NC_002663.1
<i>Pseudomonadales</i>		
<i>Moraxellaceae</i>		
<i>Acinetobacter baumannii</i>	ATCC 17978	NC_009085.1
<i>Acinetobacter sp.</i>	ADP1	NC_005966
<i>Psychrobacter arcticus</i>	273-4	NC_007204.1
<i>Psychrobacter cryohalolentis</i>	K5	NC_007969.1
<i>Pseudomonadaceae</i>		
<i>Pseudomonas fluorescens</i>	Pf5	NC_004129.6
<i>Pseudomonas syringae</i>	DC3000	NC_004578.1
<i>Thiotrichales</i>		
<i>Francisellaceae</i>		
<i>Francisella philomiragia</i>	<i>philomiragia</i>	NC_010336.1
<i>Francisella tularensis</i>	<i>horlatica</i>	NC_007880.1
<i>Piscirickettsiaceae</i>		
<i>Thiomicrospira crunogena</i>	XCL-2	NC_007520.2
<i>Vibrionales</i>		
<i>Vibrionaceae</i>		
<i>Photobacterium profundum</i>	SS9	NC_006370.1
<i>Vibrio cholerae</i>	N16961	NC_002505.1

Table 3-1. γ -proteobacteria strains used in this study.

Orders, Families, Genera, Species	Source/Strain	GenBank accession no.
<i>Vibrionales</i>		
<i>Vibrionaceae</i>		
<i>Vibrio parahaemolyticus</i>	RIMD 2210633	NC_004603.1
<i>Vibrio vulnificus</i>	CMCP6	NC_004459.2
<i>Xanthomonadales</i>		
<i>Xanthomonadaceae</i>		
<i>Xanthomonas axonopodis</i>	306	NC_003919.1
<i>Xanthomonas campestris</i>	8004	NC_007086.1
<i>Xylella fastidiosa</i>	9a5c	NC_002488.3

Next, the 3' end of the 16S rRNA gene nucleotide sequences of alleles from same bacterial strain, of alleles from different strains from same species, and of alleles from different species from same genus (data not shown) were aligned using ClustalW (Thompson *et al.*, 1994). The length of the nucleotide sequence most conserved was determined at 157 bp. Likewise, the 16S-23S Internal Transcribed Spacer (ITS) nucleotide sequences of alleles from same bacterial strain were also aligned using ClustalW. The length of the nucleotide sequence most conserved was determined at 75 bp. These two most conserved nucleotide sequences, the 157 bp at the 3' end of 16S, and the 75 bp at the 5' end of 16S-23S ITS were combined into a single 232 bp sequence for each bacterial species and strain under study. This 232 bp sequence will be used here as a phylogenetic marker for the γ -proteobacteria under study.

Phylogenetic trees

Two neighbor-joining trees were constructed (Saitou and Nei, 1987), a first one based on the alignment of the last 157 bp at the 3' end of the 16S rRNA gene described above, a second one based on the alignment of the 232 bp sequence also described above. Both trees were bootstrapped using 1,000 random samples of sites from the alignment, all using CLUSTAL W (Thompson *et al.*, 1994) at the DNA Data

Bank of Japan (DDBJ) (<http://clustalw.ddbj.nig.ac.jp/top-e.html>), with the Kimura's parameter method (Kimura, 1983). The neighbor-joining tree was drawn using TreeView (version 1.6.6) (Page, 1996; Page, 2000).

Results and discussion

In a previous study (Xu and Côté, 2003), on the bacterial genus *Bacillus* and closely-related genera, we reported the development of a short DNA marker that could be used to reconstruct their phylogeny. This marker was a combination of the last 150 bp at the 3' end of the 16S rRNA gene and the first 70 bp at the 5' end of the 16S-23S rRNA internal transcribed spacer (ITS) into a single 220 bp "marker". It could cluster *Bacillus* species and species from closely related genera into taxa akin to genera and could also distinguish closely related species. The 3' end of the 16S rRNA gene contained three regions that were known to be highly conserved among bacteria (Gürtler and Stanisich, 1996). The 5' end of the 16S-23S rRNA ITS was conserved among alleles from same strains (Xu and Côté, 2003).

In the current study on γ -proteobacteria, we further assessed the usefulness of this marker. The sizes of the 3' end of the 16S rRNA gene and the 5' end of the 16S-23S rRNA ITS retained here for the construction of our phylogenetic marker for γ -proteobacteria are slightly different at 157 and 75 bp, respectively, for a total marker size of 232 bp. These sizes were selected as follows: first, an alignment of the 16S rRNA gene nucleotide sequence of alleles from same strain showed that these sequences were highly conserved. The intra-strain alleles shared 99% nucleotide sequence identities. Alleles from species from same genus, however, covered a wider spectrum of nucleotide sequence identities. Whereas alleles from *Pseudomonas* (*Ps.*) *fluorescens* Pf-5 and from *Ps. syringae* pv. tomato strain DC3000 share 98% nucleotide sequence identities, alleles from *Vibrio* (*V.*) *cholerae* and from *V.*

parahaemolyticus RIMD 2210633 share 92% nucleotide sequence identities. Comparative analyses of the 3' end of the 16S rRNA gene nucleotide sequence showed that the last 157 bp were in many cases highly conserved among strains from same species. This is exemplified by *Salmonella (Sal.) enterica* arizonae and *Sal. enterica* Ty2 which share 99% nucleotide sequence identities over the last 157 bp at the 3' end of the 16S rRNA gene. Species from same genus share lower nucleotide sequence identities. This is exemplified by *Shewanella (She) amazonensis* and *She. denitrificans* which share 94% nucleotide sequence identities over the last 157 bp at the 3' end of the 16S rRNA gene.

A bootstrapped neighbor-joining tree based on the alignment of this 157 bp located at the 3' end of the 16S rRNA gene was constructed (Fig.. 1). Although, in most cases, this 157 bp could distinguish species from different genera from same family, in some cases, closely related species from different genera from same family appeared undistinguishable. This is the case for *Mannheimia succiniciproducens* and both *Actinobacillus (Act)*, species, *Act. pleuropneumonia* and *Act. succinogenes*. This is also true for *Xylella (Xy.) fastidiosa* and both *Xanthomonas (X.)* species, *X. axonopodis* pv. *citri* str.306 and *X. campestris* pv. *campestris* str.8004. And this is true for the human and animal pathogenic *Enterobacteriaceae*: the *Yersinia*, *Shigella*, *Klebsiella*, *Escherichia*,

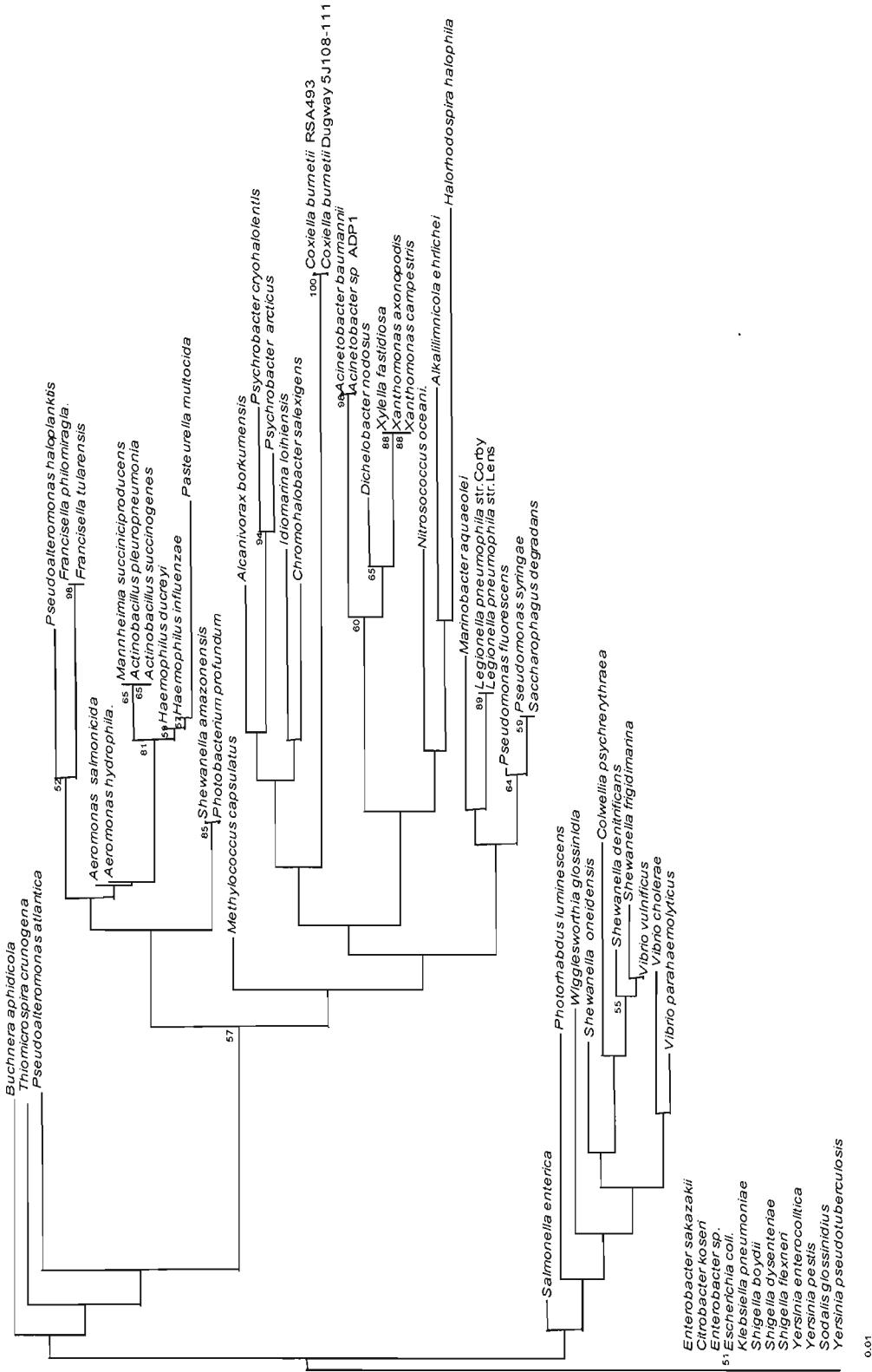


Fig. 3-1. Bootstrapped neighbor-joining tree of γ -proteobacteria species inferred from the alignment of a 157 nucleotide sequence at the 3' end of the 16S rRNA gene. Bootstrap values higher than 50% are indicated (expressed as percentage of 1000 replication). The horizontal bar represents 1% nucleotide difference.

Enterobacter and *Citrobacter* species. In some cases, this 157 bp could distinguish species from same genus as exemplified by the *Psychrobacter*, *Pseudomonas*, *Shewanella* and *Vibrio* species. In other cases, closely related species from same genus could not be distinguished as exemplified by the *Francisella*, *Aeromonas*, *Actinobacillus*, *Haemophilus*, *Acinetobacter* and *Xanthomonas* species.

In all cases, this 157 bp could distinguish species from different families, with one exception: *She. amazonensis* and *Photobacterium (Ph.) profundum*, members of the *Shewanellacea* and *Vibrionacea* family, respectively. Both appear undistinguishable. Clearly, this 157 bp sequence cannot distinguish closely-related species. An additional DNA sequence appears necessary to better distinguish closely-related species.

Next, an alignment of the 16S-23S ITS nucleotide sequences of alleles from same bacterial strain with a subset of the bacteria under study: *Xanthomonas campestris* pv. *campestris* str. 8004, *Ps. syringae* pv. *tomato* str. DC3000, *Actinobacillus succinogenes* 130Z, *E. coli* K12, *V. parahaemolyticus* RIMD 2210633 and *Shigella flexneri* 2a str. 301 (Figure 2). The total number of alleles vary from two to ten for *X. campestris* pv. *campestris* str. 8004 and *V. parahaemolyticus* RIMD 2210633, respectively. The allelic sequences were highly homologous for some species and highly hererologous for others. *Xanthomonas campestris* pv. *campestris* str. 8004 and *Ps. syringae* pv. *tomato* str. DC3000 carry two and five identical alleles, respectively. *Actinobacillus succinogenes* 130Z, *E. coli* K12, *V. parahaemolyticus* RIMD 2210633 and *Shigella flexneri* 2a str. 301 carry five, eight, ten and six alleles respectively, with varying level of heterogeneity, where highly homologous alleles are grouped together and can be distinguished from different alleles in same strain (Figure 2). Alleles carry from zero to four tRNA genes. An alignment of the nucleotide sequences among alleles at the intra-strain level required the introduction of several gaps.

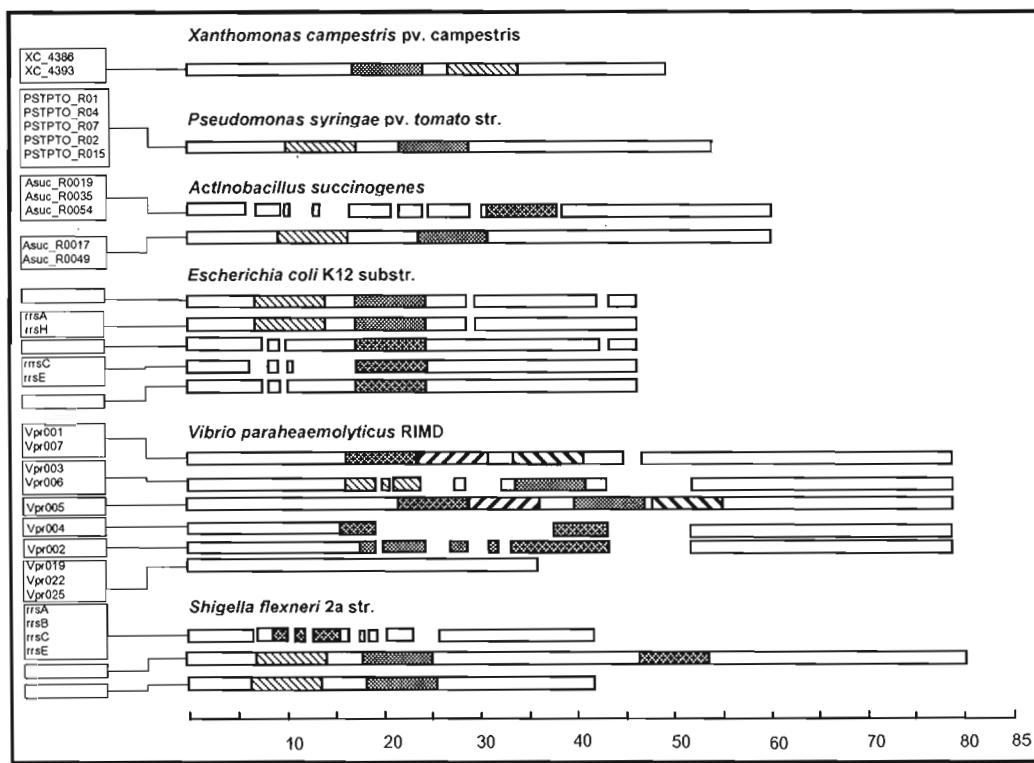


Fig. 3-2. Schematic representation of allelic 16S-23S rDNA Internal Transcribed Spacer of γ -proteobacteria.

The non filed boxes represent regions of homologous nucleotide sequences between allelic ITS of the same bacteria. Filed boxes represent tRNA. The blank spaces between boxes represent non conservation regions between allelic ITS of the same bacteria.

tRNA Ala tRNA Asp tRNA Glu tRNA Lys tRNA Val

The first 75 bp at the 5' end of the 16S-23S ITS, however, was highly conserved at the intra-strain level. It was retained here for the construction of our phylogenetic marker. The two conserved nucleotide sequences identified above, the 157 bp at the 3' end of 16S rRNA gene and the 75 bp at the 5' end of 16S-23S ITS, were combined into a single 232 bp sequence. This will be used here as a phylogenetic marker for the γ -proteobacteria under study.

A bootstrapped neighbor-joining tree based on the alignment of a 232 bp sequence was constructed (Figure 3). Based on the topology of the neighbor-joining tree, four major Groups, Group I to IV are revealed (Figure. 3). Based on nucleotide sequence identities, sub-groups and clusters can be formed. Group I contains seven Orders and nine families. Group II contains eight Orders and eleven families. Group III contains three Orders and five families. Group IV contains one Order and one family. Of the 13 Orders under study, four are present in more than one Group. The *Thiotrichales* are present in Groups I and III. The *Alteromonadales* are present in Groups I, II and III. The *Legionellales* are present in Groups I and II. The *Vibrionales* are present in Groups I and III. All other nine Orders are present in a single Group. Of the 22 families under study, four are present in more than one Group. The *Alteromonadaceae* are present in Groups I and II. The *Pseudoalteromonadaceae* (*Pse*) are present in Groups I and III. The *Vibrionaceae* are present in Groups I and III. The *Shewanellaceae* are present in Groups I and III. All other 18 families are present in a single Group. All species from same genus are present in same Group with the exception of *Pse. haloplanktis* and *Pse. atlantica* present in Group I and III, respectively.

Group I can be sub-divided into six sub-groups, sub-group I-1 to I-6. Sub-group I-1 contains both species of the *Francisellaceae* family. Sub-group I-2 contains three species, members of two families: *Pseudomonadaceae* and *Alteromonadaceae*. Sub-group I-3 comprises *Pse. haloplanktis* and both *Legionella pneumophila* strains. Although in the same sub-group, *Pse. haloplanktis* shows 20% nucleotide sequence divergence with the two *Legionella* strains. Both *Legionella* strains are tightly grouped together and form cluster I-3-1. Sub-group I-4 contains *Ph. profundum* and *She. amazonensis*, two species from two different families, *Vibrionaceae* and *Shewanellaceae*, respectively. Although they appear very similar on the neighbor-joining tree, both sequences show 15% nucleotide divergence. Sub-group I-5 contains both *Aeromonas* species, tightly grouped together. Sub-group I-6 contains all six

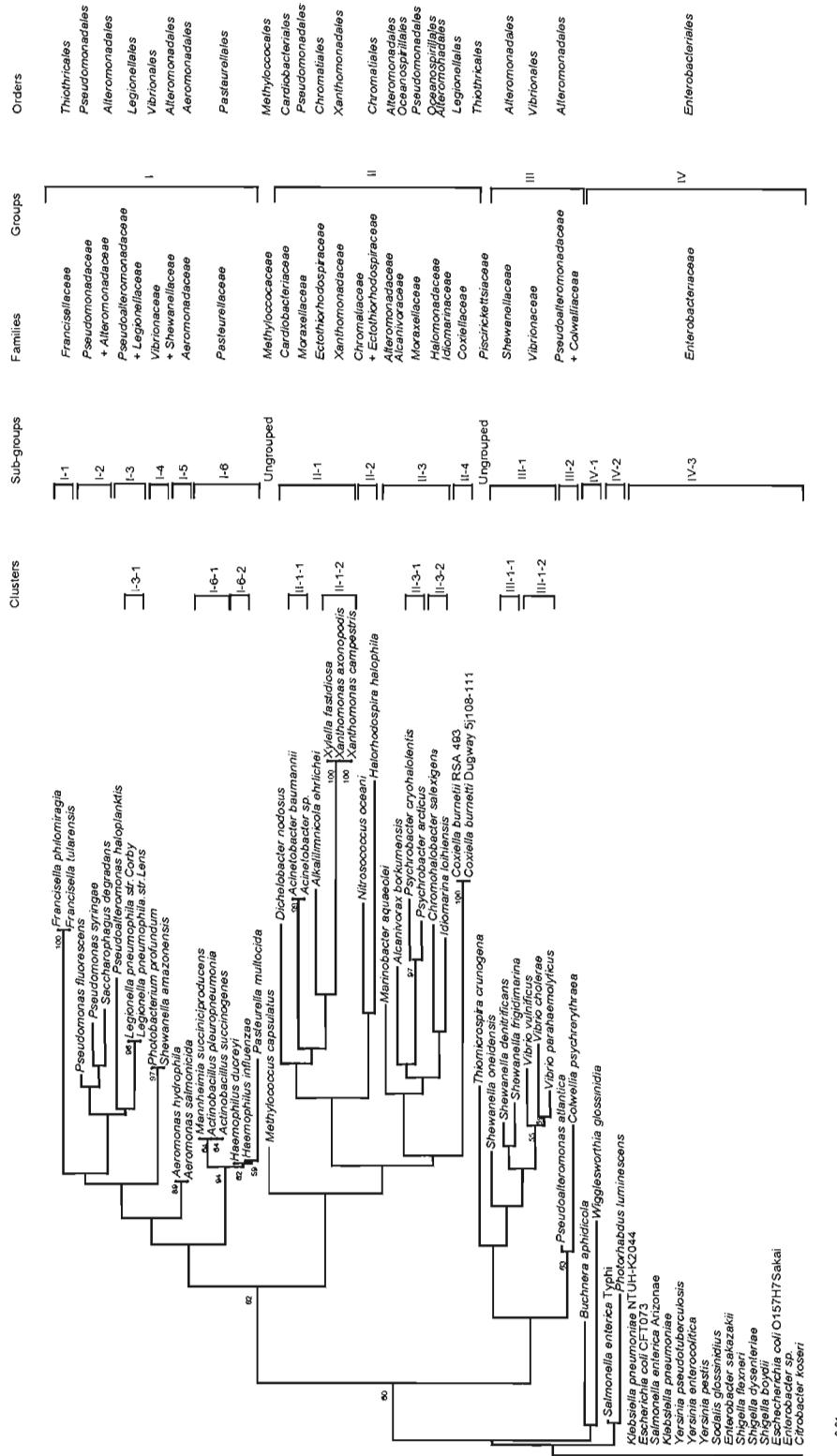


Fig. 3-3. Bootstrapped neighbor-joining tree of γ -proteobacteria species inferred from the alignment of 232 nucleotide sequence marker.

This 232 nucleotide sequence marker is a combination of a 157 nucleotide sequence at the 3' end of the 16S rRNA gene and a 75 nucleotide sequence at the 5' end of the 16S-23S Internal Transcribed Spacer (ITS) sequence. Major Groups are indicated in capital roman numerals. Sub-groups and clusters are indicated in arabic numbers. Bootstrap values higher than 50% are indicated (expressed as percentage of 1000 replication). The horizontal bar represents 1% nt difference.

Pasteurellaceae species. The *Mannheimia* species is tightly grouped with the two *Actinobacillus* species and form cluster I-6-1. Both *Haemophilus* species form cluster I-6-2. In Group I, clusters comprised species from same genus or closely related species from different genera from same family. All other branches corresponded to families.

Group II can be sub-divided into four sub-groups, sub-group II-1 to II-4, and one ungrouped species, *Methylococcus capsulatus*. Sub-group II-1 comprises members of four families from four Orders. Members of sub-group II-1 share up to 30% nucleotide sequence divergences. Closely related species can be further grouped together. This is the case for the two *Acinetobacter* species and all three *Xanthomonadaceae* species which form cluster II-1-1 and II-1-2, respectively. Sub-group II-2 contains *Nitrosoccus oceanii* and *Halorhodospira halophila*, member of the *Chromatiaceae* and *Ectothiorhodospiraceae* family respectively. Both species show 42% nucleotide sequence divergences. Both families belong to the Chromatiales Order. Sub-group II-3 contains six species, from five families and three Orders. Members of sub-group II-3 share up to 24% nucleotide sequence divergences. Two closely related species, *Psychrobacter (Psy) cryohalolentis* and *Psy. arcticus* form cluster II-3-1. Sub-group II-4 comprises both strains of *Coxiella burnetti*, members of the *Coxelliaceae* family. In Group II, clusters comprised species from same genus or closely related species from different genera from same family. All other branches corresponded to families.

Group III can be sub-divided into two sub-groups, sub-groups III-1 and III-2, and one ungrouped species, *Thiomicrospira crunogena*. Sub-group III-1 contains three *Shewallenaceae* and three *Vibrionaceae* species. *She. denitrificans* and *She. frigidimarina*, and the three *Vibrio* species form two clusters, III-1-1 and III-1-2, respectively. Sub-group III-2 contains two genera from two families of the same Order. Both species, on two separate branches, show 23% nucleotide sequence

divergence. In Group III, both clusters comprised species from same genus. All other branches corresponded to families.

Group IV contains all the *Enterobacteriaceae* species under study. Three sub-groups can be revealed: sub-group IV-1 to IV-3. Sub-group IV-1 contains two insect obligate endosymbionts. Sub-group IV-2 contains *Salmonella enterica* Ty2 and *Photobacterium luminescens*. All other *Enterobacteriaceae* species are in sub-group IV-3. The latter are closely related to each others.

Our results, based on the 232 bp phylogenetic marker described here are, in general, in agreement with the phylogeny of γ -proteobacteria based on the 16S rRNA gene with some exceptions. In the neighbor-joining tree, clusters comprised species from same genus or closely related species from different genera from same family. All other branches corresponded to families. As indicated above, of the 22 families under study, 18 are present in a single Group and four are present in more than one Group. These latter four families encompass the marine bacteria (Suen *et al.*, 2007). They are *Vibrionaceae*, in Groups I and III; and the *Alteromonas*-related protobacteria (Ivanova *et al.*, 2004): *Alteromonadaceae*, in Groups I and II; *Pseudoalteromonadaceae*, in Groups I and III; and *Shewanellaceae*, in Groups I and III. Interestingly, within a Group, these marine bacteria are found in close proximity of one another. It reflects the varying level of heterogeneity among *Alteromonas*-related protobacteria. The grouping showed here is based on a 232 bp marker. The biological significance of this grouping is unknown. Clearly, however, the phylogenetic analyses of these related marine heterotrophic bacteria is a work in progress (Ivanova *et al.*, 2004).

Conclusion

In conclusion, the use of this 232 bp marker presents several advantages over the use of the entire 16S rRNA gene or the generation of extensive phenotypic and

genotypic data in phylogenetic analyses. First, this marker is not allele-dependant. The 3' end of the 16S rRNA gene is highly conserved at the intra-strain level. We have shown here that although the 16S-23S ITS allelic sequences can be very heterogeneous within a strain, the first 75 bp, however, are conserved among alleles from same strain in γ -proteobacteria. Clearly, any allele would generate the same results. Second, this 232 bp marker contains 157 bp from the 3' end of the 16S rRNA gene and 75 bp from the 5' end of the 16S-23S ITS. The 157 bp is highly conserved among closely related species. Owing to its higher rate of nucleotide substitutions, the 75 bp adds discriminating power among closely related species from same genus and closely related genera from same family. Because of its higher percentage of nucleotide sequence divergence than the 16S rRNA gene, the 232 bp marker can better discriminate among closely related γ -proteobacteria species. Third, the method is simple, rapid, suited to large screening programmes and easily accessible to most laboratories. More importantly, however, this 232 bp marker can group γ -proteobacteria families and genera in accordance with established phylogenies, with the exceptions indicated above. It can also reveal γ -proteobacteria species which may appear misassigned and for which additional characterization appear warranted.

CHAPITRE IV

Assessment of a short phylogenetic marker
based on comparisons of 3' end 16S rDNA and
5' end 16S-23S ITS nucleotide sequences
on the genus *Xanthomonas*

Sabarimatou Yakoubou et Jean-Charles Côté
Natural Science, 2 (12):1369-1374.

Abstract

A short phylogenetic marker previously used in the reconstruction of the class γ -proteobacteria was assessed here at a lower taxa level, the genus *Xanthomonas*. This marker is 224 nucleotides in length. It is a combination of a 157 nucleotide sequence at the 3' end of the 16S rRNA gene and a 67 nucleotide sequence at the 5' end of the 16S-23S ITS sequence. A total of 23 *Xanthomonas* species were analyzed. Species from the phylogenetically related genera *Xylella* and *Stenotrophomonas* were included for comparison purposes.

A bootstrapped neighbor-joining phylogenetic tree was inferred from comparative analyses of the 224 bp nucleotide sequence of all 30 bacterial strains under study. Four major Groups were revealed based on the topology of the neighbor-joining tree, Group I to IV. Group I and II contained the *Stenotrophomonas* and *Xylella*, respectively. Group III included five *Xanthomonas* species: *X. theicola*, *X. sacchari*, *X. albineans*, *X. transluscens* and *X. hyacinthi*. This group of *Xanthomonas* species is often referred to as the *hyacinthi* group. Group IV contained the other 18 *Xanthomonas* species.

The overall topology of the neighbor-joining tree was in agreement with currently accepted phylogenetic. The short phylogenetic marker used here could resolve species from three different *Xanthomonadacea* genera: *Stenotrophomonas*, *Xylella* and *Xanthomonas*. At the level of the *Xanthomonas* genus, distant species could be distinguished, and whereas some closely-related species could be distinguished, others were undistinguishable. Pathovars could not be distinguished. We have met the resolving limit of this marker: pathovars and very closely related species from same genus.

Introduction

The genus *Xanthomonas* comprises 27 species. These species are primarily characterized by the production of xanthomonadins, a water-insoluble yellow pigment, and the production of an exo-polysaccharide, the xanthan gum, which is used as a thickening, stabilizing and gelling agent in food, pharmaceutics, cosmetics and oil industries (McNeely Kang 1973; Kennedy Bradsrlaw1984). Most *Xanthomonas* species are plant pathogens (Vauterin *et al.*, 1990). They cause diseases on several economically important plants including crucifers, Solanaceae, citrus, cotton, cereals, ornamentals, fruit and nut trees (Vauterin *et al.*, 1990; Hayward, 1993). It is estimated that at least 124 monocotyledons and 268 dicotyledons are infected by *Xanthomonas* species (Hayward, 1993; Leyns *et al.*, 1984; Vauterin *et al.*, 1995).

Up to the mid-90's, the classification of *Xanthomonas* species and isolates was based on phenotypic data. The main criteria for the creation of new species rested on host specificity. The taxa "pathovar" was added to distinguish *Xanthomonas* species at the infrasubspecific level. Some species, e.g. *X. axonopodis*, *X. transluscaens* or *X. campestris*, comprised more than ten, 40 and 125 pathovars, respectively (Vauterin *et al.*, 1990). Several methods were used in an attempt to classify *Xanthomonas* species and pathovars: restriction fragment-length polymorphism (RFLP) (Lazo *et al.*, 1987; Lazo et Gabriel, 1987), protein profiles (Vauterin *et al.*, 1991) and fatty acid methyl ester profiles (Chase *et al.*, 1992; Yang *et al.*, 1993). Vauterin *et al.* reorganized the classification of the genus *Xanthomonas* based on DNA-DNA hybridization (Vauterin *et al.*, 1995). They revealed 20 DNA homology groups which they considered genomic species. Since then, other approaches based on different nucleotide sequences have been used to study the phylogeny of *Xanthomonas*: the 16S rRNA gene (Hauben *et al.*, 1997), a multilocus sequence typing (MLST),

(Maiden *et al.*, 1998), the 16S-23S intergenic spacer (Gonçalves and Rosáto, 2002), the repetitive palindromic-based polymerase chain reaction fingerprinting (Rep-PCR) (Rademaker *et al.*, 2005), the *gyrB* gene (Parkinson *et al.*, 2007) and a multilocus sequence analysis (MLSA) (Young *et al.*, 2008). Seven additional *Xanthomonas* species have now been described (Trébaol *et al.*, 2000; Jones *et al.*, 2004; Euzéby, 2006; Euzéby, 2007; Schaad *et al.*, 2005; Schaad *et al.*, 2006; Schaad *et al.*, 2007).

In a recent study (Yakoubou and Côté, 2010), a short 232 bp nucleotide sequence "marker" was used to reconstruct the phylogeny of the Class γ -proteobacteria. This 232 bp marker was a combination of the last 157 bp at the 3' end of the 16S rRNA gene and the first 75 bp at the 5' end of the 16S-23S rRNA internal transcribed spacer (ITS). We showed that the 157 bp sequence was highly conserved among closely related species. Owing to its higher rate of nucleotide substitutions, the 75 bp added discriminating power among species from same genus and closely related genera from same family. This marker could reconstruct the phylogeny of the species, genera, families and Orders within the Class γ -proteobacteria in accordance with the accepted classification.

In the current study, we further assess the resolving power of this marker at a much lower taxa level: species within the genus *Xanthomonas*.

Materials and Methods

Bacterial species and strains.

A total of 25 *Xanthomonas* strains from 23 species were analyzed. Four *Xylella fastidiosa* strains and one *Stenotrophomonas maltophilia* stain were added for comparison purposes. They were selected on the basis that their complete genome

sequences were freely available in GenBank, at the National Center for Biotechnology Information (NCBI) completed microbial genomes database (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>, August 2009). All bacterial strains and their GenBank accession number are listed in Table 1.

Sequences analysis

First, the 16S rRNA and the 16S–23S ITS sequences of the 30 bacterial strains under study were retrieved from GenBank. Second, the 16S rRNA gene nucleotide sequences were aligned using ClustalW (Thompson *et al.*, 1994) (data not shown). The length of the nucleotide sequence most conserved was determined at 157 bp. Third, the 16S-23S ITS were aligned using ClustalW (data not shown). The length of the nucleotide sequence most conserved was determined at 67 bp. These two most conserved nucleotide sequences, the 157 bp at the 3' end of 16S, and the 67 bp at the 5' end of 16S-23S ITS, were combined into a single 224 bp sequence for each bacterial species and strain under study. This 224 bp sequence will be used here as a phylogenetic marker for the *Xanthomonas* species and related genera under study.

Phylogenetic trees

A neighbor-joining tree was constructed (Saitou and Nei, 1987) based on the alignment of the 224 bp sequence of the 30 bacterial strains under study. The tree was bootstrapped using 1,000 random samples of sites from the alignment, all using CLUSTAL W software (Thompson *et al.*, 1994) at the DNA Data Bank of Japan (DDBJ) (<http://clustalw.ddbj.nig.ac.jp/top-e.html>), with the Kimura's parameter method (Kimura, 1983). The neighbor-joining tree was drawn using TreeView (version 1.6.6) (Page, 1996; Page 2000).

Table 4-1. *Xanthomonadaceae* species used in this study

Genera	Species	Pathovars/Strain	GenBank accession no.
<i>Stenotrophomonas</i>			
	<i>maltophilia</i>	R551-3	NC_011071
<i>Xanthomonas</i>			
	<i>albilineans</i>	LMG 494 ^T	X95918
	<i>alfalfaе</i>	F1	AF442741
	<i>arboricola</i>	pv. juglandis LMG 747 ^T	Y10757
	<i>axonopodis</i>	LMG 538 ^T	X95919
	<i>bromi</i>	LMG 947 ^T	AF209754
	<i>campestris</i>	pv. campestris ATCC 33913 ^T pv. campestris B100 pv. campestris 8004	NC_003902 NC_010688 NC_007086
	<i>cassavae</i>	LMG 673 ^T	AF209756
	<i>codiaeи</i>	LMG 8678 ^T	Y10765
	<i>cucurbitae</i>	LMG 690 ^T	Y10760
	<i>euvesicatoria</i>	85-10	NC_007508
	<i>fragariae</i>	LMG 708 ^T	X95920
	<i>gardneri</i>	CNPH496	AY288083
	<i>hortorum</i>	LMG 733 ^T	Y10759
	<i>hyacinthi</i>	LMG 739 ^T	Y10754
	<i>melonis</i>	LMG 8670 ^T	Y10756
	<i>oryzae</i>	pv. oryzae MAFF 311018	NC_007705
	<i>perforans</i>	CNPH411	AY288081
	<i>pisi</i>	LMG 847 ^T	Y10758

Table 4-1. *Xanthomonadaceae* species used in this study (Continued)

Genera	Species	Pathovars/Strain	GenBank accession no.
	<i>Sacchari</i>	LMG 471 ^T	Y10766
	<i>theicola</i>	LMG 8684 ^T	Y10763
	<i>transluscens</i>	pv. graminis	AY247064
	<i>vasicola</i>	LMG 736 ^T	Y10755
	<i>vesicatoria</i>	LMG 911 ^T	Y10761
<i>Xylella</i>			
	<i>fastidiosa</i>	9a5c	NC_002488
		M12	NC_010513
		M23	NC_010577
		Temecula 1	NC_004556

Results and Discussion

A bootstrapped neighbor-joining tree based was inferred from the alignment of the 224 bp sequence of all 25 *Xanthomonas* species and pathovars, four *Xylella fastidiosa* strains and *Stenotrophomonas maltophilia* under study (Fig. 1). Four Groups, Group I to IV, were revealed at the 95% nucleotide sequence identities. Group I contains *Stenotrophomonas maltophilia*. Group II encompasses all four *Xylella fastidiosa* strains. They share 99% nucleotide sequence identities. Group III includes five *Xanthomonas* species: *X. theicola*, *X. sacchari*, *X. albineans*, *X. transluscens* and *X. hyacinthi*. They share at least 96% nucleotide sequence identities. This group of *Xanthomonas* species is often referred to as the *hyacinthi* group (Hauben *et al.*, 1997; Gonçalves et Rosato, 2002). Our results are in agreement with the first identification of the *hyacinthi* group based on the homology of their 16S rRNA

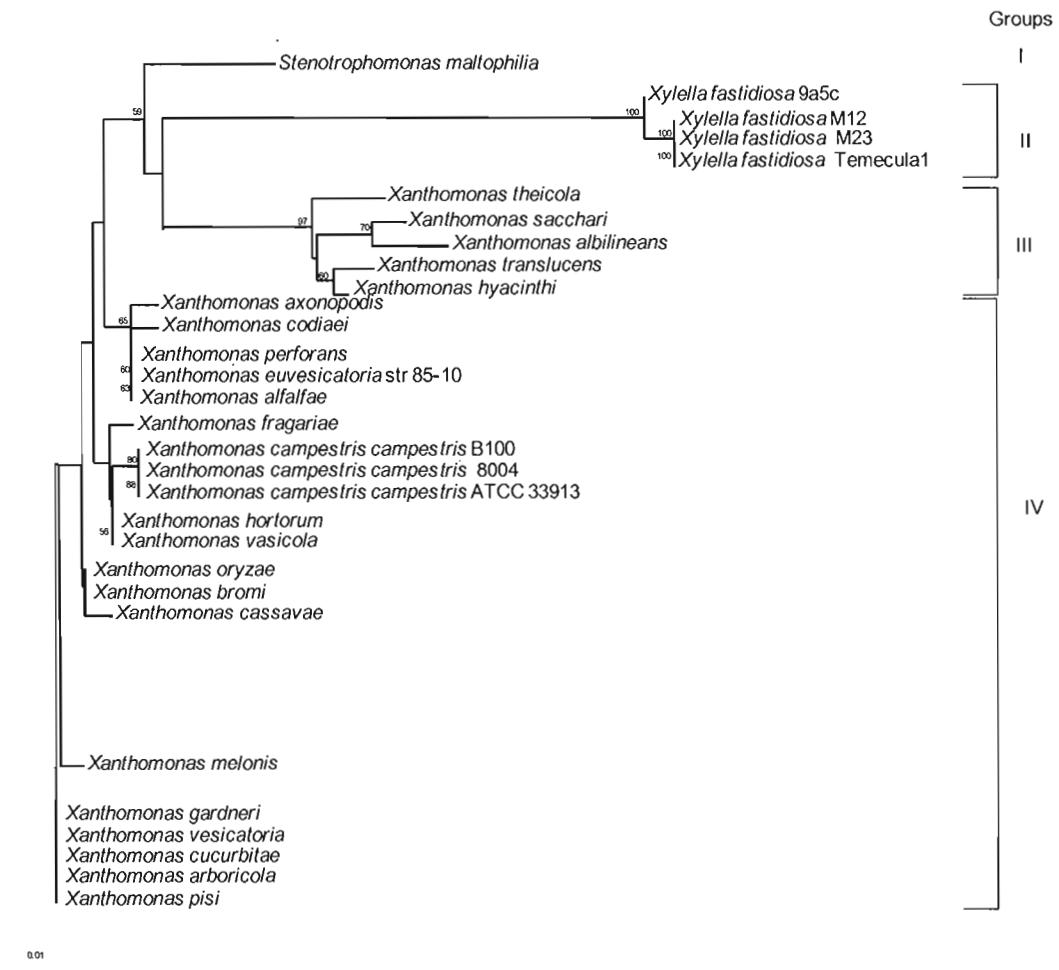


Fig.4-1. Bootstrapped neighbor-joining tree of all 30 bacterial strains inferred from the alignment of the 224 bp marker. Major Groups are indicated in capital roman numerals. Bootstrap values higher than 50% are indicated (expressed as percentage of 1000 replication). The horizontal bar represents 1% nt difference.

(Hauben *et al.*, 1997), 16S-23S ITS (Gonçalves and Rosato, 2002) and *gyrB* nucleotide sequences (Parkinson *et al.*, 2007) and MLSA (Young *et al.*, 2008). Group IV contains 18 *Xanthomonas* species. These species share at least 95% nucleotide sequence identities. Six species can be distinguished: *X. axonopodis*, *X. codiae*, *X. fragariae*, *X. campestris*, *X. cassavae* and *X. melonis*. *Xanthomonas perforans*, *X. euvesicatoria* and *X. alfalfa*e are grouped together and appear undistinguishable. These species share 100% nucleotide sequence identities. The grouping of these six species is in agreement with the work of Parkinson *et al.* (Parkinson *et al.* 2007) based on comparison of *gyrase B* gene sequences. Furthermore, the grouping of *X. perforans*, *X. euvesicatoria* and *X. alfalfa*e is in agreement with the work of Young *et al.* (Young *et al.* 2008) based on MLSA. *Xanthomonas hortorum* and *X. vasicola*, and *X. oryzae* and *X. bromi* are grouped together, respectively, and appear undistinguishable. Both pair of species share 99% and 100% nucleotide sequence identities, respectively. Five other *Xanthomonas* species, *X. gardneri*, *X. vesicatoria*, *X. cucurbitae*, *X. arboricola* and *X. pisi* are grouped together and appear undistinguishable. These species share 100% nucleotide sequence identities. The three *X. campestris* strains appear undistiguishable. They share 100% nucleotide sequence identities.

Of the 23 *Xanthomonas* species under study, 15 species or group of species could be distinguished by the 224 bp sequence used as marker. Very closely related species, such as those in Group IV could not be distinguished. Pathovars could not be distinguished, as exemplified by the three *X. campestris* pathovars. The overall topology of the neighbor-joining tree was, however, in agreement with phylogenetic trees based on the 16S rRNA (Hauben *et al.*, 1997) and the 16S-23S ITS (Gonçalves and Rosato, 2002).

In previous studies, we showed that a DNA sequence from 3' end 16S rRNA gene and 5' end 16S-23S ITS could be used as a marker in the reconstruction of phylogenies in the Gram-positive genus *Bacillus* and closely-related genera (Xu and Côté, 2003), the Gram-positive Order *Bacillales* (Yakoubou *et al.*, 2010), and the Gram-negative Class γ -proteobacteria (Yakoubou et Côté, 2010). This marker ranged in size from 220 bp to 232 bp. It contained 150-157 bp from the 3' end of the 16S rRNA gene and 67-75 bp from the 5' end of the 16S-23S ITS. The 150-157 bp from the 3' end of the 16S rRNA gene was often sufficient to distinguish bacterial Orders, families, and species from different genera. This sequence was, however, highly conserved among closely related species. Owing to its higher rate of nucleotide substitutions, the 67-75 bp from the 5' end of the 16S-23S ITS added resolving power among closely related species from same genus. This marker had proven useful in reconstructing the phylogenies of the genus *Bacillus* and closely-related genera (Xu and Côté, 2003), the Order *Bacillales* (Yakoubou *et al.*, 2010) and the Class γ -proteobacteria (Yakoubou and Côté, 2010) in accordance with accepted phylogenies inferred from much more comprehensive datasets. This marker presented several advantages over the use of the entire 16S rRNA gene or the 16S-23S ITS or the generation of extensive phenotypic and genotypic data in phylogenetic analyses. We showed that the method was simple, rapid, suited to large screening programmes and easily accessible to most laboratories. It also proved very valuable in revealing bacterial species which appeared misassigned and for which additional characterization appeared warranted. The resolving power of this marker has been further analyzed here in a much deeper branch of the Class γ -proteobacteria: the genus *Xanthomonas*. As expected, we have shown here that this marker could resolve species from three different *Xanthomonadaceae* genera: *Stenotrophomonas*, *Xylella* and *Xanthomonas*. At the level of the *Xanthomonas* genus, distant species could be distinguished. However, although some closely-related species could be distinguished, others were grouped together and some were undistinguishable.

Clearly, pathovars could not be distinguished. We have met the resolving limit of this marker: pathovars or very closely-related species.

Conclusion

We have shown here that a short DNA marker based on 3' end 16S rDNA and 5' end ITS, previously used to reconstruct the phylogeny of the Class γ -proteobacteria, was unable to discriminate between *Xanthomonas* very closely-related species and pathovars. We have reached the limit of the resolving power of the 224 bp sequence as a phylogenetic marker.

CONCLUSION

Compte tenu de l'importance économique, industrielle, médicale et scientifique des bactéries, leur identification et leur classification demeurent l'un des enjeux majeurs des bactériologistes. Différentes techniques d'identification et de classification ont été utilisées au fil des avancées technologiques. Elles sont principalement regroupées en deux catégories: les techniques de classification phénotypique et génotypique.

La phylogénétique est l'établissement de relation génétique entre organismes. Elle a connu son essor avec les travaux de phylogénie moléculaire réalisés par Carl Woese en 1987 et qui ont permis de révéler le monde vivant en 3 grands domaines (Woese, 1987). Grâce aux travaux de Woese, le gène 16S de l'ARNr s'est imposé comme la macromolécule de choix pour la réalisation des études phylogénétiques. Les études de phylogénie chez les bactéries ont permis la re-classification de taxa bactériens et la caractérisation de bactéries non-cultivables (Ward *et al.*, 1990; Amann *et al.*, 1995). Outre le gène 16S de l'ARNr, de nombreux autres marqueurs moléculaires ont été développés et utilisés en phylogénie bactérienne. Aujourd'hui pourtant, la phylogénie de nombreux taxa bactériens, espèces, genres, familles, Ordres, Classes de grande importance, demeure un travail en perpétuel changement.

L'objectif général de cette thèse était d'évaluer un marqueur d'ADN pour la classification de bactéries Gram-positif et Gram-négatif à différents niveaux taxonomiques. Ce marqueur d'ADN était basé sur une courte séquence nucléotidique située aux extrémités 3' du gène 16S de l'ARNr et 5' de la région inter-génique 16S-23S (ITS). Les niveaux taxinomiques étudiés ici correspondaient:

- 1) à des bactéries Gram-positif de l'Ordre des *Bacillales*,

- 3) et à des bactéries Gram-négatif de la Classe des γ -protéobactéries,
- 4) et du genre *Xanthomonas*.

Chaque niveau taxonomique étudié correspond à un chapitre de la thèse. Les conclusions majeures tirées des résultats obtenus dans chacun des quatre chapitres suivent.

Phylogénie de l'Ordre des *Bacillales* basée sur les séquences nucléotidiques des extrémités 3' 16S rDNA et 5' 16S-23S ITS

Phylogeny of the Order *Bacillales* inferred from 3' 16S rDNA and 5' 16S-23S ITS nucleotide sequences

L'objectif de cette première étude consistait à évaluer le marqueur phylogénétique à un niveau taxonomique élevé de bactéries Gram-positif : l'Ordre des *Bacillales*. Ce marqueur d'ADN est composé de 220 pb: les 150 dernières pb à l'extrémité 3' du gène 16S de l'ARNr et des 70 premières pb à l'extrémité 5' de la région inter-génique 16S-23S. Les résultats obtenus dans ce premier chapitre nous permettent de confirmer l'hypothèse émise: ce court marqueur d'ADN permet l'identification rapide et la classification des espèces bactériennes de l'Ordre des *Bacillales*.

Notre étude a porté sur huit des neuf familles que compte l'Ordre des *Bacillales*. Les bactéries de ces huit familles sont réparties en 21 genres et 72 espèces. Notre arbre phylogénétique construit à l'aide du marqueur est similaire à celui considéré comme l'arbre consensus utilisé comme référence (Ludwig *et al.*, 2009). Ce dernier, très robuste, a été construit à partir de données exhaustives de séquences du gène 16S de l'ARNr . Dans notre étude, huit groupes ont été révélés. Chaque groupe

correspond à une famille de l'Ordre des *Bacillales* à une exception près. Sur notre arbre phylogénétique, certaines familles apparaissent comme hétérogènes. C'est le cas des *Bacillaceae*, *Paenibacillaceae* et *Alicyclobacillaceae*. Dans notre étude, les membres de la famille des *Bacillaceae* sont répartis en quatre groupes. Ces membres sont répartis en deux groupes sur l'arbre consensus utilisé comme référence. Cette différence est due à la présence sur notre marqueur de la séquence des 70 premières pb de la région inter-génique qui augmente le pouvoir de discrimination entre les espèces. Nous en concluons que le marqueur phylogénétique analysé ici présente un pouvoir de résolution supérieur à celui du gène 16S de l'ARNr. Il permet aussi de révéler les espèces requérant une caractérisation additionnelle telle que *Alicyclobacillus herbarius*.

Évaluation d'un court marqueur phylogénétique basé sur la comparaison des séquences nucléotides des extrémités 3' 16S rDNA et 5' 16S-23S ITS du groupe *Bacillus cereus*

Assessment of a short phylogenetic marker based on comparison of 3' end 16S rDNA and 5' end 16S-23S ITS of the *Bacillus cereus* group

L'objectif de cette deuxième étude consistait à évaluer le pouvoir de résolution du marqueur phylogénétique de 220 pb à un niveau taxonomique inférieur de l'Ordre des *Bacillales* : le groupe *Bacillus cereus*. Ce marqueur d'ADN, strictement identique à celui utilisé pour l'Ordre des *Bacillales*, est composé de 220 pb: les 150 dernières pb à l'extrémité 3' du gène 16S de l'ARNr et des 70 premières pb à l'extrémité 5' de la région inter-génique 16S-23S. Les résultats obtenus dans ce deuxième chapitre ne nous permettent pas de confirmer l'hypothèse émise: ce court marqueur d'ADN permet l'identification rapide et la classification un niveau taxonomique inférieur de bactéries Gram-positif: le groupe *Bacillus cereus*.

Cependant, ils nous ont permis de déterminer la limite du pouvoir de résolution du marqueur phylogénétique de 220 pb au niveau de l'Ordre des *Bacillales*: les trois espèces phylogénétiquement très proches: *B. cereus*, *B. anthracis* et *B. thuringiensis*.

Notre étude a porté sur 129 allèles de 12 souches bactériennes réparties en deux genres et huit espèces. Notre arbre phylogénétique construit à l'aide du marqueur a permis de révéler quatre groupes. Chaque groupe correspond à l'ensemble des allèles d'une même espèce à une exception près, le groupe I. Le groupe I a été subdivisé en trois sous-groupes. Les sous-groupes I-1 et I-2 sont hétérogènes. Le sous-groupe I-1, le plus hétérogène, contient des allèles des quatre espèces du groupe *B. cereus*. La grande proximité des allèles de *B. cereus*, *B. anthracis* et *B. thuringiensis* observée dans ce sous-groupe est en accord avec plusieurs études antérieures qui ont démontré que ces trois espèces étaient très proches. De nombreux auteurs ont suggéré de considérer ces trois espèces comme une seule et même espèce (Vilas-Boas *et al.*, 2007). D'autre part, le sous-groupe I-3 est homogène. Il contient uniquement des allèles de *B. weihenstephanensis*.

Ici, le marqueur phylogénétique n'a pu discriminer les trois espèces très proches du groupe *B. cereus*: *B. cereus*, *B. anthracis* et *B. thuringiensis*. Certaines allèles spécifiques ont permis de distinguer *B. weihenstephanensis* de *B. cereus*, *B. anthracis* et *B. thuringiensis*. Nous concluons que cette étude nous a permis de révéler la limite du pouvoir de résolution du marqueur phylogénétique de 220 pb au niveau de l'Ordre des *Bacillales* : les trois espèces très proches du groupe *Bacillus cereus*.

Phylogénie des γ -protéobactéries basée sur la comparaison des séquences nucléotidiques des extrémités 3' du gène 16S rRNA et 5' de l'ITS 16S-23S.

Phylogeny of γ -proteobacteria inferred from comparisons of 3' end 16S rRNA gene and 5' end 16S-23S ITS nucleotide sequences

L'objectif de cette troisième étude consistait à évaluer le marqueur phylogénétique à un niveau taxonomique élevé de bactéries Gram-négatif : la Classe des γ -protéobactéries. Ce marqueur d'ADN est composé de 232 pb: les 157 dernières pb à l'extrémité 3' du gène 16S de l'ARNr et des 75 premières pb à l'extrémité 5' de la région inter-génique 16S-23S. Les résultats obtenus dans ce troisième chapitre nous permettent de confirmer l'hypothèse émise: ce court marqueur d'ADN permet l'identification rapide et la classification des espèces bactériennes de la Classe des γ -protéobactéries.

Notre étude a porté sur 64 espèces et souches de γ -protéobactéries réparties en 13 ordres, 22 familles, 40 genres et 59 espèces. Notre arbre phylogénétique construit à l'aide du marqueur est similaire à celui considéré comme l'arbre consensus utilisé comme référence (Ludwig et Klenk, 2005). Ce dernier, très robuste, a été construit à partir de données exhaustives de séquences du gène 16S de l'ARNr . Dans notre étude, les 22 familles sont classées dans quatre groupes majeurs. Sur les 22 familles étudiées, 18 sont homogènes. Les quatre autres familles se retrouvent dans plus d'un groupe et apparaissent comme hétérogènes. C'est le cas des *Alteromonadaceae*, *Pseudoalteromonadaceae*, *Vibrionaceae* et *Shewanellaceae*. Ces quatre familles sont constituées d'espèces de bactéries marines chez lesquelles des études antérieures ont révélé une hétérogénéité. Leur phylogénie est en cours de révision (Ivanova *et al.*, 2004). Dans notre étude, le marqueur de 232 pb a permis de distinguer les espèces proches d'un même genre ainsi que les genres proches d'une même famille. Cette

discrimination est due à la présence sur notre marqueur de la séquence des 75 premières pb de la région inter-génique qui augmente le pouvoir de résolution entre les espèces et genres proches, respectivement.

Nous en concluons que le marqueur phylogénétique analysé ici présente un pouvoir de résolution supérieur à celui du gène 16S de l'ARNr. Il permet aussi de révéler les espèces requérant une caractérisation additionnelle.

Évaluation d'un court marqueur phylogénétique basé sur la comparaison des séquences nucléotides des extrémités 3' 16S rDNA et 5' 16S-23S de l'ITS du genre *Xanthomonas*.

Assessment of a short phylogenetic marker based on comparisons of 3' end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences on the genus *Xanthomonas*

L'objectif de cette quatrième étude consistait à évaluer le pouvoir de résolution du marqueur phylogénétique de 224 pb à un niveau taxonomique inférieur de la Classe des γ -protéobactéries : le genre *Xanthomonas*. Ce marqueur d'ADN est composé de 224 pb: les 157 dernières pb à l'extrémité 3' du gène 16S de l'ARNr et des 67 premières pb à l'extrémité 5' de la région inter-génique 16S-23S. Les résultats obtenus dans ce quatrième chapitre nous permettent de confirmer l'hypothèse émise: ce court marqueur d'ADN permet l'identification rapide et la classification à un niveau taxonomique inférieur de bactéries Gram-négatif : le genre *Xanthomonas*.

Notre étude a porté sur 23 espèces du genre *Xanthomonas*. Deux genres additionnels de la famille des *Xanthomonadaceae*, *Stenotrophomonas* et *Xylella*, ont été rajoutées pour comparaison. Dans notre étude, quatre groupes ont été révélés. Les

genres *Stenotrophomonas* et *Xylella* de retrouvent dans les groupes I et II, respectivement. Le genre *Xanthomonas* est plus hétérogène. Les espèces se retrouvent dans les groupes III et IV. Le marqueur phylogénétique a discriminé 11 espèces sur les 23 espèces de *Xanthomonas* analysées. Les espèces très proches et les pathovars n'ont pas été discriminés.

Nous en concluons que cette étude nous a permis de révéler la limite du pouvoir de résolution du marqueur phylogénétique de 224 pb au niveau de la Classe des γ -protéobactéries : les espèces très proches et les pathovars du genre *Xanthomonas*.

Cette thèse a permis de construire quatre arbres phylogénétiques différents basés sur un court marqueur d'ADN dont la taille varie entre 220 pb et 232 pb. Le marqueur d'ADN est composé : des 150 à 157 dernières pb à l'extrémité 3' du gène 16S de l'ARNr et des 67 à 75 premières pb à l'extrémité 5' de la région inter-génique 16S-23S. Les quatre arbres phylogénétiques sont similaires aux arbres consensus construits à partir de données exhaustives de séquences du gène 16S de l'ARNr et utilisés comme référence dans chaque niveau taxonomique étudié. Les résultats obtenus lors de ces quatre études permettent de confirmer l'hypothèse générale de cette thèse qui stipule que le court marqueur d'ADN composé des dernières paires de bases de l'extrémité 3' du gène 16S de l'ARNr et des premières paires de bases de l'extrémité 5' de la région inter-génique 16S-23S permet l'identification rapide et la classification de bactéries à Gram-positif et à Gram-négatif à différents niveaux taxonomiques.

La limite du pouvoir de résolution du marqueur phylogénétique a été révélée au niveau de la Classe des γ -protéobactéries chez les espèces très proches et les

pathovars du genre *Xanthomonas* et au niveau de l'Ordre des *Bacillales* chez les trois espèces très proches du groupe *Bacillus cereus*. Ces niveaux taxinomiques pourraient être discriminés soit par l'ajout de séquences additionnelles de la région variable 16S-23S ITS, soit par une étude basée sur les gènes de ménage. Ces gènes codent pour des protéines indispensables au fonctionnement cellulaire de base. Mentionnons entre autres, la sous-unité B de l'ADN gyrase (*gyrB*), la protéine RecA (*recA*), la sous-unité β de l'ARN polymérase (*rpoB*), etc. Ils sont présents chez la plupart des bactéries et ont un taux d'évolution, donc un taux de substitutions nucléotidiques supérieur à celui du gène 16S et ne pas transmis horizontalement.

L'originalité de cette thèse réside dans la nature même du court marqueur d'ADN. La séquence de 150-157 pb a permis de distinguer certaines espèces de genres différents, au niveau de l'Ordre, de la Famille et des espèces. Cette séquence très conservée au niveau des espèces très proches n'arrive pas à discriminer ces dernières. La séquence des 67 à 75 premières pb à l'extrémité 5' de la région intergénique 16S-23S, à cause de son taux de substitution de nucléotides élevé, augmente le pouvoir de résolution du marqueur d'ADN au niveau des espèces très proches d'un même genre. Il s'agit d'une méthode simple d'utilisation qui peut être utilisée dans les programmes rapides de criblages de plusieurs centaines de souches bactériennes et accessible à tous les laboratoires. Le court marqueur d'ADN peut servir à la fois aux études systématiques intra et inter-spécifiques aboutissant à l'identification des espèces bactériennes et à la révision de certaines classifications. Par la même occasion, cette thèse permettra d'apporter une contribution à la définition de la notion d'espèce au niveau bactérien.

À moyen terme, hors du cadre des présentes études doctorales, il pourrait devenir possible de développer une puce biologique (Phylo-BioChip) pour la détection, et l'identification des espèces bactériennes. Une puce biologique, ou biopuce, est un support solide de petite taille qui peut être en verre ou en silicium et

qui contient plusieurs milliers de courtes séquences d'ADN simple brin marquées avec une molécule fluorescente. Celles-ci servent de sondes pour l'hybridation spécifique avec des séquences d'ADN complémentaires de bactéries provenant d'échantillons à tester. Ces derniers peuvent être d'origine alimentaire, environnementale, médicale, industrielle etc. Elle permet l'analyse simultanée de plusieurs milliers de séquences d'ADN différentes.

RÉFÉRENCES

- Able, K., De Schmertzing H. et Peterson, J. J. 1963. Classification of microorganisms of chemical composition I. Feasibility of utilizing gas chromatography. *J. Bacteriol.* 85: 1039-1044.
- Amann, R., Ludwig, W. et Schleifer, K. H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59: 143-169.
- Andersson, A., Svensson, B., Christiansson, A. et Rönner, U. 1999. Comparison between automatic ribotyping and random amplified polymorphic DNA analysis of *Bacillus cereus* isolates from the dairy industry. *Int. J. Food Microbiol.* 47: 147-151.
- Anton, A. I., Martinez-Murcia, A. J. et Rodriguez-Valera, F. 1998. Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the *Escherichia coli* ECOR collection. *J. Mol. Evol.* 47 (1): 62-72.
- Ash, C., Farrow, A. E., Wallbanks, S. et Collins, M. D. 1991a. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. *Lett. Appl. Microbiol.* 13: 202-206.
- Ash, C., Farrow, J. A. Dorsch, M. Stackebrandt, E. et Collins, M. D. 1991b. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. *Int. J. Syst. Evol. Microbiol.* 41: 343-346.
- Baggesen, D. L., Sandvang, D. et Aarestrup, F. M. 2000. Characterization of *Salmonella enterica* serovar Typhimurium DT104 isolated from Denmark and comparison with isolates from Europe and the United States. *J. Clin. Microbiol.* 38: 1581-1586.
- Baggesen, D. L., Wegener, H. C. et Madsen, M. 1997. Correlation of conversion of *Salmonella enterica* serovar Enteriditidis phage type 1, 4, or 6 to phage type 7 with loss of polysaccharide. *J. Clin. Microbiol.* 35: 330-333

- Blackall., P. J., Christensen, .P. et Bisgaard, M. 1998. Diversity among isolates of *Actinobacillus equuli* and related organisms as revealed by ribotyping. *Aust. Vet. J.* 76: 423-425.
- Blanc, D.S., Siegrist., H. H., Sahli, R. et Francioli, P. 1993. Ribotyping of *Pseudomonas aeruginosa*: discriminatory power and usefulness as a tool for epidemiological studies. *J. Clin. Microbiol.* 31: 71-77.
- Bouvet, A., Grimont, F. et Grimont, P. A. D. 1991. Intraspecies variations in nutritionally variant streptococci: rRNA gene restriction patterns of *Streptococcus defectus* and *Streptococcus adjacens*. *Int. J. Syst. Bacteriol.* 41: 483-486.
- Boyce, J. D., Wilkie, I., Harper, M., Paustian, M. L., Kapur, V. et Adler, B. 2002. Genomic Scale Analysis of *Pasteurella multocida* Gene Expression during Growth within the natural chicken host. *Infect. Immun.* 70: 6871 - 6879.
- Bravo, A., Gill, S. S. et Soberon, M. 2007. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon*, 49 (4): 423-435.
- Brenner D. J., Krieg N. R., Staley J. T. et Garrity G. M. 2005. Eds. Bergey's Manual of Systematic Bacteriology, 2nd Ed, vol. 2 (The *Proteobacteria*), part C (The *Alpha-, Beta-, Delta-, and Epsilonproteobacteria*), pp.1256. Springer, New York,
- Brenner, D. J., Fanning, G. R., Johnson, K. E., Citarella, R. V. et Falkow, S. 1969. Polynucleotide sequence relationships among members of *Enterobacteriaceae*. *J. Bacteriol.* 98: 637-650.
- Brousseau, R., Saint-Onge, A., Préfontaine, G., Masson, L. et Cabana, J. 1993. Arbitrary primer polymerase chain reaction, a powerful method to identify *Bacillus thuringiensis* serovars and strains. *Appl. Environ. Microbiol.* 59: 114-119.
- Brown, J. R. and Volker, C. 2004. Phylogeny of γ -proteobacteria: resolution of one branch of the universal tree? *Bioessays*, 26: 463-468.
- Busse, H. J., Denner, E. B. M. et Werner L. 1996. Classification and identification of bacteria: current approach to an old problem. Overview of methods used in bacterial systematics. *J. Biotechnol.* 47: 3-38.

- Bustard, M. T., Whiting, S., Cowan, D. A. et Wright, P. C. 2002. Biodegradation of high-concentration isopropanol by a solventtolerant thermophile, *Bacillus pallidus*. *Extremophiles* 6: 319–323.
- Canonica F. P. et Pisano M. A. 1988. Gas-liquid chromatographic analysis of fatty acid methyl esters of *Aeromonas hydrophila*, *Aeromonas sobria* and *Aeromonas caviae*. *Clin. Microbiol.* 26(4): 681–685.
- Carlson, C.R., Caugant, D.A. et Kolsto, A.-B. 1994. Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. *Appl. Environ. Microbiol.* 60: 1719-1725.
- Carr, J. G. 1983. Microbes I have known: A study of those associated with fermented products. In *Bacillus* pp 293-320. Harwood, C. R. (éd.) Plenum Press, New-York.
- Cavé, H., Bingen, E. E. J. et Denamur, E. 1994. Differentiation of *Escherichia coli* strains using randomly amplified polymorphic DNA analysis. *Res. Microbiol.* 145: 141-150.
- Chase, A.R.; Stall, R.E.; Hodge, N.C. et Jones, J.B. 1992. Characterization of *Xanthomonas campestris* strains from aroids using physiological, pathological, and fatty acid analyses. *Phytopathology* 82:754-759.
- Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B. and Bork, P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* 311: 1283-1287.
- Cilia, V., Lafay, B. et Christen, R. 1996. Sequence heterogeneities among 16S ribosomal RNA sequences and their effect on phylogenetic analyses at species level. *Mol. Biol. Ecol.* 13: 451-461.
- Claus, D. et Berkeley, R.C.W. 1986. Genus *Bacillus* Cohn, 1872. In *Bergey's Manual of Systematic Bacteriology*, vol. 2, pp. 1105-1139. Sneath, P. H. A., Mair, N. S., Sharpe, M. E. and J. G. Holt. (éd.) The Williams et Wilkins Co., Baltimore.
- Claus, D. et Fritze, D. 1989. Taxonomy of *Bacillus*. In *Bacillus* pp 5-26. Harwood, C. R. (éd.) Plenum Press, New-York.
- Collins, M. D. 1985. Analysis of isoprenoid quinone. In *Methods in Microbiology* Vol.18, pp. 329-366. Gottschalk, G. (éd.) Academic Press, London.

- Collins, M. D. et Jones, D. 1981. Distribution of isoprenoid quinone structural type in bacteria and their taxonomic implication. *Microbiol. Rev.* 45: 316-354.
- Condon, C., Squires, C. et Squires, C. L. 1995. Control of rRNA transcription in *Escherichia coli*. *Microbiol. Rev.* 59: 623-645.
- Cosgrove S. E., Sakoulas G., Perencevich E. N. et als. 2003 Comparison of mortality associated with methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* bacteremia: a meta-analysis *Clin. Infect. Dis.* 36:53-59.
- Crosa, J. H., Brenner, D. J. et Falkow, S. 1973. Use of a single-strand specific nuclease for analysis of bacterial and plasmid deoxyribonucleic acid homo- and heteroduplexes. *J. Bacteriol.* 115: 904-911.
- Czajka, J. et Batt, C. A. 1994. Verification of causal relationships between *L. monocytogene* isolates by RAPD DNA patterns. *J. Clin. Microbiol.* 32: 1280-1287.
- Daffonchio, D., Borin, S., Consolandi, A., Mora, D., Manachini, P. L. et Sorlini, C. 1998. 16S-23S rRNA internal transcribed spacers as molecular markers for the species of the 16S rRNA group I of the genus *Bacillus*. *FEMS Microbiol. Lett.* 163 : 229-236.
- Darnell, J., Lodish, H., Baltimore, D. et François, C. 1993. *Biologie moléculaire et cellulaire*. De Boeck-Wesmael (éd.) De Boeck-Wesmael s.a., Bruxelles.
- Das, T., Choudhury, K., Sharma, S., Jalali S., Nuthethi R. et the Endophthalmitis Research Group 2001. Clinical profile and outcome in *Bacillus* endophthalmitis. *Ophthalmology* 108: 1819-1825.
- Dauga, C. Doré, J. et Sghir, A. 2005. La diversité insoupsonnée du monde microbien. *Med. Sci.* 21 (3): 290-296.
- De Cesare, A., Bruce, J. L., Dambaugh, T. R., Guerzoni, M. E. et Wiedmann, M. 2001. Automated ribotyping using different enzymes to improve discrimination of *Listeria monocytogenes* isolates, with a particular focus on serotype 4b strains. *J. Clin. Microbiol.* 39 : 3002-3005.
- De Ley, J. 1992. The proteobacteria: ribosomal RNA cistron similarities and bacterial taxonomy. In *The prokaryotes*, 2nd ed., pp. 2111-2140. Balows, A., Trüper, H. G., Dworkin, M., Harder, W. et Schleifer, K. H., (éds.) Springer Verlag, Germany.

- De Ley, J., Cattoir, H. et Reynaerts, A. 1970. The quantitative measurement of DNA hybridization from renaturation rates. *Eur. J. Biochem.* 12: 133-142.
- De Ley, J., Segers, P., Kersters, K. Mannheim, W. et Lievens, A. 1986. Intra-and intergeneric similarities of the *Bordetella* ribosomal ribonucleic acid cistrons: Proposal for a new family, *Alcaligenaceae*. *Int. J. Syst. Bacteriol.* 36 : 405-414.
- de Silva, S., Petterson, B., de Muro, M. A. et Priest, F. G. 1998. A DNA probe for the detection and identification of *Bacillus sporothermodurans* using the 16S-23S rDNA spacer region and phylogenetic analysis of some field isolates of *Bacillus* which form highly heat resistant spores. *Syst. Appl. Microbiol.* 21: 398-407.
- Devine, K. M. 1995. The *Bacillus subtilis* genome project : aims and progress. *Tibtech.* 13 : 210-216.
- De Vos, P. et De Ley, J. 1983. Intra and intergeneric similarities of *Pseudomonas* and *Xanthomonas* ribosomal ribonucleic cistron. *Int. J. Syst. Bacteriol.* 33: 487-509.
- De Vos, P., Goor, M., Gillis, M. et De Ley, J. 1985a. Ribosomal ribonucleic acid cistron of phytopathogenic *Pseudomonas* species. *Int. J. Syst. Bacteriol.* 35: 169-184.
- De Vos, P., Kersters, K., Falsen, E., Pot, B., Gillis, M., Sergers, P. et De Ley, J. 1985b. *Comamonas* Davis and Park 1962 gen. nov., nom. Rev. emend. and *Comamonas terrigena* Hugh 1962 sp. nov., nom. rev. *Int. J. Syst. Bacteriol.* 35: 443-453.
- De Vos, P., Rogers, R., Bleeker M., Reijans M., van der Lee, T., Hornes, M., Freijters, A., Pot, J., Peleman, J. Kulper M. et Zabeau M. 1995. AFLP: a new method for DNA fingerprinting. *Nucleic Acids Res.* 23: 4407-4414.
- Doumith, M., Buchrieser, C., Glaser, P., Jacquet, C. et Martin, P. 2004. Differentiation of the major *Listeria monocytogenes* serovars by multiplex PCR. *J. Clin. Microbiol.* 42: 3819 - 3822.
- Drobniewski, F. A. 1993. *Bacillus cereus* and related species. *J. Clin. Microbiol.* 6: 324-338.
- Ellwood, M. et Nomura, M. 1982. Chromosomal locations of the genes for rRNA in *Escherichia coli* K-12. *J. Bacteriol.* 143: 1077-1080.

- Embley, T. M. et Wait. R. 1994. Structural lipids of eubacterial. In *Modern microbial methods: chemical methods in prokaryotic systematics*. pp. 121-163. Goodfellow M. et O'Donnell A.G. (éds.) John Wiley et Sons, England.
- Euzéby, J. 2006. In Validation of the publication of new names and new combinations previously effectively published outside the IJSB, List 109. *Int. J. Syst. Evol. Microbiol.* 56: 925-927.
- Euzéby, J. 2007. In List of New names and new combinations previously effectively, but not validly published list 115. *Int. J. Syst. Evol. Microbiol.* 57:893-897.
- Farber, J. M. et Addison, C. J. 1994. RAPD typing for distinguishing species and strains in the genus *Listeria*. *J. Appl. Bacteriol.* 77: 242-250.
- Feitelson, J. S. 1993. The *Bacillus thuringiensis* family tree. In *Advanced Engineered Pesticides*. pp. 63-71. L. Kim (éd.) Marcel Dekker Inc., New York.
- Forterre, P., Gribaldo, S. et Brochier, C. 2005. LUCA: the last universal common ancestor. *Med Sci.* 21 (10): 860-865.
- Forterre P. 1999. Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. *Mol. Microbiol.* 33: 457-65.
- Fox, G. E. et Stackebrandt, E. 1987. The application of 16S rRNA cataloguing and 5S rRNA sequencing in bacterial systematics. *Int. J. Syst. Bacteriol.* 27: 406-458.
- Fox, G. E., Pechman, K. R. et Woese, C. R. 1977. Comparative cataloguing of 16S ribosomal ribonucleic acid: molecular approach to prokaryotic systematics. *Int. J. Syst. Bacteriol.* 27: 44-57.
- Fox, G. E., Stackebrandt, E., Hespell R. B., Gibson J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrs, K. R., Chen, K. N., et Woese, C. R. 1980. The phylogeny of prokaryotes. *Science* 209: 457- 463.
- Frachon, E. S., Nicholas, I. et de Barjac, H. 1991. Cellular fatty acid analysis as a potential tool for predicting mosquitocidal activity of *Bacillus sphaericus* strains. *Appl. Environ. Microbiol.* 57: 3394-3398.

- Gao, B., Mohan, R. et Gupta, R. S. 2009. Phylogenomics and protein signatures elucidating the evolutionary relationships among the *Gammaproteobacteria* *Int. J. Syst. Evol. Microbiol.* 59: 234-247.
- García-Martínez, J., Acinas, S. G., Antón, A. I. et Rodríguez-Valera, F. 1999. Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *J. Microbiol. Methods* 36: 55-64.
- García-Martínez, J., Martínez-Murcia, A., Antón, A. I. et Rodríguez-Valera, F. 1996. Comparison of the small 16S to 23S Intergenic Spacer Region (ISR) of the rRNA operons of some *Escherichia coli* Strains of the ECOR Collection and *E. coli* K-12. *J. Bacteriol.* 178: 6374-6377.
- Garcia-Robles, I., Sánchez, J., Gruppe, A., Martínez-Ramírez, A. C., Rausell, C., Real, M. D. et Bravo, A. 2001. Mode of action of *Bacillus thuringiensis* PS86Q3 strain in hymenopteran forest pests. *Insect Biochem. Mol. Biol.* 31: 849–856.
- Garrity, G. M., Bell, J. A. et Lilburn, T. 2005. The Revised Road Map to the Manual. In *Bergey's Manual of Systematic Bacteriology*, 2nd ed., vol. 2, *The Proteobacteria, Part A, Introductory Essays* pp. 159–220. Brenner, Krieg, Staley et Garrity (éds.) Springer, New York.
- Gerner-Smidt, P. 1992. Ribotyping for *Acinetobacter calcoaceticus* – *Acinetobacter baumannii* complex. *J. Clin. Microbiol.* 30: 2580-2585.
- Giannino, V., Rappazzo, G., Scuto, A., Di Marco, O., Privitera, A., Santagati, M. et Stefani, S. 2001. rrn operons in *Haemophilus parainfluenzae* and mosaicism of conserved and species-specific sequences in the 16S-23S rDNA long spacer. *Res. Microbiol.* 152 (5): 461-468.
- Gibson, T. et Gordon, R. E. 1974. *Bacillus*. In *Bergey's Manual of Determinative Bacteriology*. pp 529-550. Buchanan R. E. et Gibbons N. E. (éds.) Williams et Wilkins, Baltimore.
- Girard, F. 1997. Utilisation industrielle de certaines souches de *Bacillus*. http://membres.lycos.fr/microbio/systematique/Bacillus_doc/bacilUI.html.
- Goncalves, E. R. et Rosato, Y. B. 2002. Phylogenetic analysis of *Xanthomonas* species based upon 16S-23S rDNA intergenic spacer sequences. *Int. J. Syst. Evol. Microbiol.* 52: 355-361.

- Gordon, R. E., Haynes, W. C. et Pang, C. H. N. 1973. The genus *Bacillus*. In *Bacillus* pp 293-320. Harwood C. R. (éd.) Plenum Press, New-York.
- Goto, K., Omura, T., Hara, Y. et Sadaie, Y. 2000. Application of the partial 16S rDNA sequence as an index for rapid identification of species in the genus *Bacillus*. *J. Gen. Appl. Microbiol.* 46: 1-8.
- Guillet, P., Chandre F. et Mouchet, J. 1997. L'utilisation des insecticides en santé publique: état et perspectives. *Med. Mal. Infect.* 27 (5): 552-557.
- Gupta R. S. 2000. The natural evolutionary relationships among Prokaryotes. *Crit. Rev. Microbiol.* 26 (2):111-31.
- Gupta, R. S. 2000. The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol. Rev.* 24: 367-402.
- Gürtler, V. 1993. Typing of *Clostridium difficile* strains by PCR-amplification of variable lenght 16S-23S rDNA spacer regions. *J. Gen. Microbiol.* 139: 3089-3097.
- Gürtler, V. et Mayall, B. C. 2001. Genomic approaches to typing, taxonomy and evolution of bacterial isolates. *Int. J. Syst. Evol. Microbiol.* 51: 3-16.
- Gürtler, V. et Stanisich, V. A. 1996. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology*, 142: 3-16.
- Hara, T. et Veda, S. 1982. Regulation of polyglutamate production in *Bacillus subtilis* (natto): transformation of high PGA productivity. *Agric. Biol. Chem.* 29: 45-72.
- Harwood, C. R. 1989. Introduction to the biotechnology of *Bacillus*. In *Biotechnology Handbooks*, vol 2 *Bacillus* pp 1-4. Harwood C. R. (éd.). Plenum Press, New-York.
- Hauben, L., Vauterin, L., Swings, J. et Moore, E. 1997. Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int. J. Syst. Bacteriol.* 47: 328-335.
- Hayward A.C. 1993. The host of *Xanthomonas*. In: *Xanthomonas*; pp. 51-54. J. G. Swings et E. L. Civerolo (éds.). Chapman et Hall, London, United Kingdom.
- Helgason, E., Caugant, D. A. Olsen, I. et Kolsto, A.-B. 2000a. Genetic structure of population of *Bacillus cereus* and *Bacillus thuringiensis* isolates associated with periodontitis and other human infections. *J. Clin. Microbiol.* 38: 1615-1622..

- Helgason, E., Tourasse, N. J., Meisal, R., Caugant, D. A. et Kolsto, A. B. 2004. Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. *Appl. Environ. Microbiol.* 70: 191–201.
- Herron, L. L., Chakravarty, R., Dwan, C., Fitzgerald, J. R., Musser, J. M., Retzel, E. et Kapur, V. 2002. Genome sequence survey identifies unique sequences and key virulence genes with unusual rates of amino acid substitution in Bovine *Staphylococcus aureus*. *Infect. Immun.* 70: 3978-3981.
- Heyman, J., Logan, N.A., Busse, H. J., Balcaen, A., Lebbe, L. et al. 2003. *Virgibacillus carmonensis* sp. nov., *Virgibacillus necropolis* sp. nov. and *Virgibacillus picturae* sp. nov., three novel species isolated from deteriorated mural paintings, transfer of the species of the genus *Salibacillus* to *Virgibacillus*, as *Virgibacillus marismortui* comb. nov. and *Virgibacillus salexigens* comb. nov., and emended description of the genus *Virgibacillus*. *Int. J. Syst. Evol. Microbiol.* 53: 501–511.
- Hill, C. Wet Harnish, B. W. 1981. Inversion between ribosomal RNA genes of *Escherichia coli*. *Proc. Natl. Acad. Sci USA*. 78: 7069-7072.
- Hillis, M., Mable, B. K., Larson, A., Davis, S. K et Zimmer, E. A. 1996. Nucleic acids IV : sequencing and cloning. In *Molecular systematics*. Hillis, M. D., Moritz, C. et Mable B. K. (éds.) Sinauer Associates, Inc, Massachussets.
- Hiraishi, A. 1999. Isoprenoid quinone as biomarkers of microbial populations in the environment. *J. Biosci. Bioeng.* 88 (5): 449-460.
- Hobbs, G. et Cross, T. 1983. Identification of endospore-forming bacteria. In *The Bacterial Spore*, vol. 2, pp. 50-78. A. Hurst et G. W. Gould (éds.) Academic Press, London.
- Höfte H. et Whiteley, H. R. 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.*, 53: 242–255.
- Inglesby, T. V., O'Toole, T., Henderson, D. A., Bartlett, J. G., Ascher, M. S., Eitzen, E., Friedlander, A. M., Gerberding, J., Hauer, J., Hughes, J., McDade, J., Osterholm, M. T., Parker, G., Perl, T. M., Russell, P. K. et Tonat, K. 2002. Anthrax as a biological weapon: updated recommendations for management. *JAMA* 287 (17): 2236-52.
- Ivanova, E. P., Flavier, S. et Christen, R. 2004. Phylogenetic relationships among marine *Alteromonas*-like proteobacteria: emended description of the family

- Alteromonadaceae* and proposal of *Pseudoalteromonadaceae* fam. nov., *Colwelliaceae* fam. nov., *Shewanellaceae* fam. nov., *Moritellaceae* fam. nov., *Ferrimonadaceae* fam. nov., *Idiomarinaceae* fam. nov. and *Psychromonadaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* 54: 1773-1788.
- Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. et Miyata, T. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci. USA* 86: 9355-9359.
- Jacquet, C., Aubert, S. El Solh, N. et Rocourt, J. 1992. Use of rRNA gene restriction patterns for the identification of *Listeria* species. *Syst. Appl. Microbiol.* 15: 42-46.
- Janssen, P., Coopman, R., Huys, G. Swings, J., Bleeker, M., Vos, P., Zabeau, M. et Kersters, K. 1996. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiol.* 142: 1881-1893.
- Jensen, M. A., Webster, J. A. et Straus, N. 1993. Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Appl. Environ. Microbiol.* 59: 945-952.
- Johnson, J. L. (1985). Determination of DNA base composition In *Methods in Microbiology*, vol 18. pp. 33-74. Gottschalk, G. (éd.) Academic Press, London.
- Johnson, J. L. 1973. Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria. *Int. J. Syst. Bacteriol.* 23: 308-315.
- Jones, J. B., Lacy, G. H., Bouzar, H., Stall, R. E. et Schaad, N. W. 2004. Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Syst. Appl. Microbiol.* 27: 755-762.
- Joung, K.-B. et Côté, J.-C. 2000. Ecological impacts of *Bacillus thuringiensis*-based insecticides. *Curr. Trends Microbiol.* 1: 1-15.
- Kämpfer, P. 1994. Limits and possibilities of total fatty acids analysis for classification and identification of *Bacillus* species. *Syst. Appl. Microbiol.* 17: 86-98.
- Kaneda, T. 1968. Fatty acids in the genus *Bacillus* II. Similarity in the fatty acid compositions of *Bacillus thuringiensis*, *Bacillus anthracis* and *Bacillus cereus*. *J. Bacteriol.* 95: 2210-2216.

- Kandler O. and König H. 1993. Cell envelopes of archaea:structure and chemistry. In: *The Biochemistry of Archaea (Archaeabacteria)*. pp. 223–259. Kates M., Kushner D. J. et Matheson A. T. (éds), Elsevier, Amsterdam.
- Kaneda, T., 1963. Biosynthesis of branched-chain fatty acids. I. Isolation and identification of fatty acids from *Bacillus subtilis* (ATTC 7059). *J. Biol. Chem.* 238: 1222-1228.
- Kaneko, T., Nozaki, R. et Aizawa, K. 1978. Deoxyribonucleic acid relatedness between *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis*. *Microbiol. Immunol.* 22: 639–641.
- Keim, P., Kalif, A., Schupp, J., Hill, K., Travis, S. E., Richmond, K., Adair, D. M., Hugh-Jones, M., Kuske, C. R. et Jackson, P. 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. *J. Bacteriol.* 179: 818-824.
- Kelman, Z. 2000. DNA replication in the third domain (of life). *Curr. Protein. Pept. Sci.* 1: 139-154.
- Kennedy J. F. et Bradshaw I .J. 1984. Production properties and applications of xanthan. *Progr.Industr.Microbiol.* 19: 319-371.
- Kersters, K., De Vos, P., Gillis, M., Swings, J., Vandamme, P. et Stackebrandt, E. 2006. Introduction to the *Proteobacteria*. In *The Prokaryotes: A Handbook on the Biology of Bacteria*. pp. 3-37. Dworkin M., Falkow, S. Rosenberg, E. Schleifer K. H. et Stackebrandt E. (éds.), Springer, New York.
- Kersters, K., Pot, B., Dewettinck, D., Torck, U., Vancanneyt M., Vauterin, l. et Vandamme P. (1994). Identification and typing of bacteria by protein electrophoresis. In *Bacterial diversity and systematics*. pp. 51-66. Priest, F. G., Ramos-Cormenzana, A. et Tyndall, B. (éds.) Plenum Press, New-York.
- Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111-120
- Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Klevens R. M., Edwards J. R., Tenover F. C. et als. 2006 Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* in intensive care units in U.S. hospitals, *Clin. Infect. Dis.* 42:389-391.

- Kluytmans, J., van Belkum, A. et Verbrugh, H. 1997. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks *Clin. Microbiol. Rev.* 10: 505 - 520.
- Kolbert, C. P. et Persing, D. H. 1999. Ribosomal DNA sequencing as tool for identification of bacterial pathogens. *Microbiology* 2: 299-305.
- Kramer, J. M. et Gilbert R. J., 1989. *Bacillus cereus* and other *Bacillus* species. In: *Foodborne bacterial pathogens*, pp 21-50. Doyle, M. P. (éd.), Marcel Dekker Inc., New York.
- Kumar, S., Tamura, K. and Nei, M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 5: 150–163.
- Kunisawa, T. 2001. Gene arrangements and phylogeny in the class *Proteobacteria*. *J. Theor. Biol.* 213: 9-19.
- Lan, R. et Reeves, P. 2002. Pandemic spread of cholera: genetic diversity and relationships within the seventh pandemic clone of *Vibrio cholerae* determined by Amplified Fragment Length Polymorphism. *J. Clin. Microbiol.* 40: 172-181.
- Lazo, G. R., et Gabriel, D. W. 1987. Conservation of plasmid DNA sequences and pathovar identification of strains of *Xanthomonas campestris*. *Phytopathology* 77: 448453.
- Lazo, G.R., Roffey, R. et Gabriel, D. W. 1987. Pathovars of *Xanthomonas campestris* are distinguishable by restriction fragment-length polymorphism. *Int. J. Syst. Bacteriol.* 37 (2): 144-221.
- Le Scanff, J., Mohammedi, J. I., Thiebaut, A., Martin, O., Argaud, L. et Robert, D. 2006. Necrotizing gastroenteritis due to *Bacillus cereus* in an immunocompromised patient. *Infection.* 34: 98-99.
- Lechner, S., Mayr, R., Francis, K. P., Prüß, B. M., Kaplan, T., Wießner-Gunkel E., Stewart Gordon, S.A.B. et Scherer, S. 1998. *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int. J. Syst. Bacterial.* 48: 1373 –1382.
- Lee, H. Y. et Côté, J.-C. 2006. Phylogenetic analysis of γ -proteobacteria inferred from nucleotide sequence comparisons of the house-keeping genes *adk*, *aroE*

- and *gdh*: comparisons with phylogeny inferred from 16S rRNA gene sequences. *J. Gen. Appl. Microbiol.* 52: 147-158.
- Legault-Démare, J., Desseaux, B., Heyman, T., Séror, S. et Ress, G. P. 1967. Studies on hybrid molecules of nucleic acids. 1. DNA-DNA hybrids on nitrocellulose filters. *Biochem. Biophys. Res. Commun.* 28: 550-557.
- Le Jeune, C. et Lonvaud-Funel, A. 1997. Sequence of DNA 16S/23S spacer region of *Leuconostoc oenos* (*Oenococcus oeni*): application to strain differentiation. *Res. Microbiol.* 148:79-86.
- Lerat, E., Daubin, V. et Moran, N. A. 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the γ -proteobacteria. *PLOS* 1, E19.
- Levett, P. N., Benett., P., Odonaghue, K., Bowker, K., Reeves, D. et Mac Gowan, A. 1993. Relapsed infection due to *L. monocytogenes* confirmed by RAPD analysis. *J. Infect.* 27: 205-207.
- Leyns, F., De Cleene, M., Swing, J. and De Ley, J. 1984. The host range of the genus *Xanthomonas*. *Bot. Rev.* 50 : 308-356.
- Logan N. A. et De Vos P. 2009. Genus I. *Bacillus* Cohn 1872, 174^{AL}. In: *Bergey's Manual of Systematic Bacteriology*, 2nd ed., vol. 3, *The Firmicutes*. pp. 21-128. De Vos, P., Garrity, G.M., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A., Schleifer, K. H. et Whitman, W. B. (éds.), Springer, New York.
- Loughney, K., Lund, E. et Dahlberg, J. E. 1982. tRNA genes are found between 16S and 23S rRNA genes in *Bacillus subtilis*. *Nucl. Acids Res.* 10: 1607-1624.
- Lu, J., Nogi, Y. et Takami, H. 2001. *Oceanobacillus iheyensis* gen. nov., sp. nov., a deep-sea extremely halotolerant and alkaliphilic species isolated from a depth of 1050 m on the Iheya ridge. *FEMS Microbiol. Lett.* 205:291-297.
- Ludwig, W. et Klenk, H. P. 2005. Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In *Bergey's Manual of Systematic Bacteriology*. pp. 49-65. Brenner, D. J. Krieg, N. R. Staley J. T. et Garrity, G. M. (éds.), Springer-Verlag, Berlin.
- Ludwig, W., Schleifer, K.-H. et Whitman, W. B. 2009. Revised road map to the phylum Firmicutes. In: *Bergey's Manual of Systematic Bacteriology*, 2nd ed., vol. 3 The Firmicutes. pp. 1-17. P. De Vos et al. (éds.) Springer-Verlag, New York.

- Maiden, M. C. J., Bygraves, J. A., Feil, E. J., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zurth, K., Caugant, D., Feavers, I. M., Achtman, M., et Spratt, B. G. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* 95:3140-3145.
- Margulies, M, Egholm, M, Altman, W. E., Attiya, S., Bader, J. S., et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
- Martin, P. A. W. 1994. An iconoclastic view of *Bacillus thuringiensis* ecology. *Am. Entomol.* 40:85-90.
- Maxam, A. M. et Gilbert, W. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA*. 74: 560-564.
- Mazurier, S.I. et Wernars, K. 1992. Typing of *Listeria* strains by random amplification of polymorphic DNA. *Res. Microbiol.* 143: 499-505.
- McNeely, W. H. et Kang, K. S. 1973. Xanthan and some other biosynthetic gums, In: *Industrial Gums*, 2nd Ed., p. 473–497. R. L. Whistler et J. N. BeMiller (éds.) Academic Press, New York.
- Mears, J. A., Cannnone, J. J., Stagg, S. M., Gutell, R. R., Agrawal, R. K. et Harvey, S. C. 2002. Modeling a minimal ribosome based on comparative sequence analysis. *J. Mol. Biol.* 321: 215-234.
- Meijer, A., Morré, S. A., Van Den Brule, A. J. C., Savelkoul, P. H. M. et Ossewaarde, J. M. 1999. Genomic relatedness of *Chlamydia* isolates determined by Amplified fragment length polymorphism analysis. *J. Bacteriol.* 181: 4469-4475.
- Mesbah, M., Premachandran, U. et Whitman, W. B. 1989. Precise measurement of the G+C content of deoxyribonucleic acid by high-performance liquid chromatography. *Int. J. Syst. Bacteriol.* 39: 159-167.
- Moss, C. W. et Dees, S. B. 1975. Identification of microorganisms by gas chromatographic-mass spectrometric analysis of cellular fatty acids. *J. Chromatogr.* 112:595-604.
- Moss, C. W., W. F. Bibb, D. E. Karr, G. O. Guerrant, et M. A. Lambert. 1983. Cellular fatty acid composition and ubiquinone content of *Legionella feeleii* sp. nov. *J. Clin. Microbiol.* 18:917-919.

- Mrazek, J., Spormann, A. M. et Karlin, S. 2006. Genomic comparisons among γ -proteobacteria. *Environ. Microbiol.* 8: 273-288.
- Mullis, K. Faloona, F. Scharf, F., Saiki, R., Horn, G. et Erlich, H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. CSH Symp. Quant. Biol. 51: 263-273.
- Nakamura L.K. 1998. *Bacillus pseudomycoides* sp. nov. *Int. J. Syst. Bacteriol.* 48: 1031–1035.
- Nakamura, L. K. et Jackson, M. A. 1995. Clarification of the taxonomy of *Bacillus mycoides*. *Int. J. Syst. Bacteriol.* 45: 46 - 49.
- Novotny, J. F. Jr et Perry, J. J. 1992. Characterization of bacteriocins from two strains of *Bacillus thermoleovorans*, a thermophilic hydrocarbon-utilizing species. *Appl. Environ. Microbiol.* 58: 2393-2396.
- Olive, D. M. et Bean, P. 1999. Principles and application of methods for DNA-based typing of microbial organisms. *J. Clin. Microbiol.* 37: 1661-1669.
- Olsen, G. J. et Woese, C. R. 1996. Lessons from an Archaeal genome: what are we learning from *Methanococcus jannaschii*. *Trends Genet.* 12 (10): 377-9.
- Olsen, G. J. et Woese, C. R. 1997. Archaeal genomics: an overview. *Cell* 89: 991–994.
- Olsen, J. E., Skov, M. N., Threlfall, E. J. et Brown, D. J. 1994. Clonal lines of *Salmonella enterica* serotype Enteritidis documented by IS200-, ribo-, pulsed-field gel electrophoresis and RFLP typing. *J. Med. Microbiol.* 40: 15-22.
- Olsen, J., Skogan, G., Fykse, E., Rawlinson, E., Tomaso, H., Granum, P. et Blatny, J. 2007. Genetic distribution of 295 *Bacillus cereus* group members based on adk-screening in combination with MLST (Multilocus Sequence Typing) used for validating a primer targeting a chromosomal locus in *B. anthracis*. *J. Microbiol. Meth.* 71 (3): 265-274.
- Otvos, I. S. Armstrong, H. et Conder, N. 2005. Safety of *Bacillus thuringiensis* var. *kurstaki*, Applications for Insect Control to Humans and Large Mammals. In *Proceedings of the 6th Pacific Rim Conference on the Biotechnology of Bacillus thuringiensis and its Environmental Impact*. pp 45-59. Côté, J.-C., Otvos, I. S., Schwartz, J.-L., et Vincent, C. (éds.) Erudit, Montréal, Canada.

- Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp. Appl. Biosci.* 12: 357-358.
- Page, R. D. M. 2000. TreeView—tree drawing software for Apple Macintosh and Windows. [HUhttp://taxonomy.zoology.gla.ac.uk/rod/treeview.htmlUH](http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).
- Parkinson, N., Aritua, V., Heeney, J., Cowie, C., Bew, J. et Stead, D. 2007. Phylogenetic analysis of *Xanthomonas* species by comparison of partial *gyrase B* gene sequences. *Int. J. Syst. Evol. Microbiol.* 57: 2881-2887.
- Parry, J. M., Turnbull, P. C. B. et Gibson J. R. 1983. A colour Atlas of *Bacillus* species, London: Wolfe Medical Publications Ltd.
- Peladan, F. et Monteil, H. 1984. Intérêt taxonomique de la caractérisation des *Pseudomonas* par l'analyse des acides gras volatils produits en culture, *Ann. Microbiol.* 135A : 411-425.
- Pelczar, M. J. Jr., Chan E. C. S. et Fontaine, J. 1982. Éléments de microbiologie. Montréal : Les Éditions HRW Ltée.
- Perales, I. et Audicana, A. 1988. *Salmonella enteridis* and eggs. *Lancet* ii: 1133
- Pot, B., Vandamme, P. et Kersters, K. 1994. Analysis of electrophoretic whole-organism protein fingerprints. In *Modern microbial methods. Chemical methods in prokaryotic systematics*. pp. 463-521. Goodfellow M. et O'Donnell A.G., Chichester (éds.) John Wiley et Sons.
- Priest, F. G. 1981. DNA homology in the genus *Bacillus*. In *The aerobic endospore-forming bacteria*. pp. 33-57. Berkeley R. C. W. et Goodfellow. (éds.) Academic Press, London.
- Priest, F.G., Goodfellow, M. et Todd, C. 1988. A numerical classification of the genus *Bacillus*. *J. Gen. Microbiol.* 134: 1847-1882.
- Punia, P., Hampton, M. D., Ridley A. M. Ward, L.R., Rowe, B. et Threfall, E. J. 1998. Pulsed-field eletrophoretic fingerprinting of *Salmonella india* and its epidemiological applicability. *J. Appl. Microbiol.* 84 103-107.
- Rademaker, J. L. W., Louws, F. J., Schultz, M. H., Rossbach, U., Vauterin, L., Swings, J. et de Bruijn, F. J. 2005. A comprehensive species to strain taxonomic framework for *Xanthomonas*. *Phytopathology* 95: 1098-1111.

- Rasmussen, H. N., Olsen, J. E. et Tasmussen, O. F. 1994. RAPD analysis of *Yersina enterocolitica*. *Lett. Appl. Microbiol.* 19: 359-362.
- Régnault, J. P. 1990. Microbiologie générale. Montréal: Décarie Éditeur inc.
- Rieg, S., Bauer, T. M., Peyerl-Hoffmann, G., Held, J., Ritter, W. Wagner, D., Vinzenz Kern, W. E. et Serr, A. 2010. *Paenibacillus larvae* bacteremia in injection drug users. *Emerg. Infect. Dis.* 1-7.
- Ronaghi, M., Uhlen, M. et al. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281(5375): 363-365.
- Ronimus, R. S., Parker, L. E. et Morgan, H. W. 1997. The utilisation of RAPD-PCR for identifying thermophilic and mesophilic *Bacillus* species. *Microbiol. Lett.* 147: 75-79.
- Rössler, D., Ludwig, W., Schleifer, K. H., Lin, C., McGill, T. J., Wisotzkey, J. D., Jurtschuk Jr, P. et Fox, G. E. 1991. Phylogenetic diversity in the genus *Bacillus* as seen by 16S rRNA sequencing studies. *Syst. Appl. Microbiol.* 14: 266-269.
- Sacchi, C. T., Whitney, A. M., Mayer, L. W., Morey, R., Steigerwalt, A., Boras, A., Weyant, R. S. et Popovic, T. 2002. Sequencing of 16S rRNA gene: A rapid tool for identification of *Bacillus anthracis*. *Emerg. Infect. Dis.* 8 (10): 1117-1123.
- Saitou, N. et Nei, M. 1987. The neighbour-joining method: a new method of constructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.
- Sambrook, J. et Russell, D.W. 2001. Molecular Cloning: A Laboratory Manual. (3rded.) Cold Spring Harbor Laboratory Press, N.Y., USA.
- Sanger, F., Nicklen, S. et Coulson, A.R. 1977. DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Schaad, N. W., Postnikova, E., Lacy, G. H., Sechler, A., Agarkova, I., Stromberg, P. E., Stromberg, V. K. et Vidaver, A. K. 2005. Reclassification of *Xanthomonas campestris* pv. *citri* (ex Hasse 1915) Dye 1978 forms A, B/C/D, and E as *X. smithii* subsp. *citri* (ex Hasse) sp. nov. nom. rev. comb. nov., *X. fuscans* subsp. *aurantifoliae* (ex Gabriel 1989) sp. nov. nom. rev. comb. nov., and *X. alfalfae* subsp. *citrumelo* (ex Riker and Jones) Gabriel et al., 1989 sp. nov. nom. rev. comb. nov.; *X. campestris* pv. *malvacearum* (ex Smith 1901) Dye 1978 as *X. smithii* subsp. *smithii* nov. comb. nov. nom. nov.; *X. campestris* pv. *alfalfae* (ex Riker and Jones, 1935) Dye 1978 as *X. alfalfae* subsp. *alfalfae* (ex Riker et al., 1935) sp. nov. nom. rev.; and “var. *fuscans*” of *X. campestris* pv. *phaseoli* (ex

- Smith, 1987) Dye 1978 as *X. fuscans* subsp. *fuscans* sp. nov. *Syst. Appl. Microbiol.* 28: 494–518.
- Schaad, N. W., Postnikova, E., Lacy, G., Sechler, A., Agarkova, I., Stromberg, P. E., Stromberg, V. K. et Vidaver, A. K. 2007. *Xanthomonas alfalfae* sp. nov., *Xanthomonas citri* sp. nov. and *Xanthomonas fuscans* sp. nov. In List of new names and new combinations previously effectively, but not validly, published, validation list no. 115. *Int. J. Syst. Evol. Microbiol.* 57: 893–897.
- Schaad, N. W., Postnikova, E., Lacy, G., Sechler, A., Agarkova, I., Stromberg, P. E., Stromberg, V. K. et Vidaver, A. K. 2006. (Erratum) Emended classification of *Xanthomonad* pathogens on citrus. *Syst. Appl. Microbiol.* 29: 690–695.
- Schenkel, R. G. M., Nicholas, L. Frachon, E. et Hamon S. 1992. Characterization and toxicity to mosquito larvae for four *Bacillus sphaericus* strains isolates from brazilian soils. *J. Invertebr. Pathol.* 60:10-14.
- Schleifer, K. H. 1985. Analysis of the chemical composition and primary structure of mureine. *Meth. Microbiol.* 18: 123-156.
- Schleifer, K. H. et Seidl, H. P. 1985. Chemical composition and structure of murein, In *Chemical methods in bacterial systematics*. pp. 201-219. Goodfellow, M. et Minnkin, D. E. (éds.) Academic Press, London.
- Schleifer, K. H., et Kandler, O. 1972. Pentidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* 36: 407-477.
- Schnepf, H. E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D. R. et Dean, D. H. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol. Molecular Biol. Rev.* (62):775–806.
- Schoeni, J. L. et Wong, A. C. L. 2005. *Bacillus cereus* food poisoning and its toxins. *J. Food. Protect.* 68 (3): 636–648.
- Schwermann, B., Pfau, K., Liliensiek, B., Schleyer, M., Fischer, T. et Bakker, E. P. 1994. Purification, properties and structural aspects of a thermoacidophilic α -amylase from *Alicyclobacillus acidocaldarius* ATCC 27009 · insight into acidostability of proteins. *Eur. J. Biochem.* 226: 981-991.
- Shaver, Y. J., Nagpal, M. L., Rudner, R., Nakamura, L. K., Fox, K. F. et Fox, A. 2002. Restriction fragment length of rRNA operons for discrimination and intergenic spacer sequences for cataloguing of *Bacillus subtilis* sub-groups. *J. Microbiol. Methods* 50:215-223.

- Siegel, J. P., Smith, A. R. et Novak, R. J. 1997. Comparison of the cellular fatty acid composition of a bacterium isolated from a human and alleged to be *Bacillus sphaericus* with that of *Bacillus sphaericus* isolated from mosquito larvicide. *Appl. Environ. Microbiol.* 63: 1006-1010.
- Siegel, J. P., Smith, A. R. et Novak, R. J. 2000. Cellular fatty acid analysis of isolates of *Bacillus thuringiensis* serovar *kurstaki*, strain HD-1. *Biological Control*. 17: 82-91.
- Sifour, M., Saeed, H. M., Zaghloul, T. I., Berekaa, M. M. et Abdel-Fattah, Y. R. 2010. Isolation of lipase gene of the thermophilic *Geobacillus stearothermophilus* strain-5. *Biotechnology*, 9: 55-60.
- Smith, N. R., Gordon, R. E., et Clark, F. E. 1952. Aerobic sporeforming Bacteria, U.S.D.A. Monogr. No 16, Washington, DC.
- Smith, T. J. et Murrell, J. C. 2009. Methanotrophy/methane oxidation. In *Encyclopedia of Microbiology* vol.3 pp. 293-298. Schaechter, M. (éd.) Elsevier.
- Sneath, P. H. A. 1984. Numerical taxonomy. In *Bergey's Manual of Systematic Bacteriology*, vol I. pp 5-7. Krieg, N. R. et Holt, J. G. (éds.) Williams et Wilkins Co. Baltimore.
- Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503-517.
- Stackebrandt, E., Ludwig, W. et Fox, G. E. 1985. 16S ribosomal RNA oligonucleotide cataloguing. *Methods Microbiol.* 18: 75-107.
- Stackebrandt, E., Murray, R. G. E., et Trüper, H. G. 1988. Proteobacteria classis nov., a name for the phylogenetic taxon that includes the "purple bacteria and their relatives". *Int. J. Syst. Bacteriol.* 38: 321-325.
- Stackebrandt, E., Weizenegger, Ludwig, W. M., Dorn, S., McGill, T. J., Fox, G. E., Woese, C. R., Schubert, W. et Schleifer, K. H. 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.
- Stackebrandt, E. et Liesack, W. 1993. Nucleic acids and classification. In *Handbook of new bacterial systematics*. pp 151-194. Goodfellow M. and O'Donnell A. G. (éds.) Academic Press Ltd, London.

- Stahly, D. P. et Klein, M. 1992. Problems with in vitro production of spores of *Bacillus popilliae* for the use in biological control of the japanese beetle. *J. Invertebr. Pathol.* 60: 282-291.
- Stephan R. 1996. Randomly amplified polymorphic DNA (RAPD) assay for genomic fingerprinting of *Bacillus cereus* isolates. *Int. J. Food. Microbiol.* 31: 311-316.
- Stephan, R., Schraft, H. et Untermann, F. 1994. Charecterization of *Bacillus licheniformis* with the RAPD technique (randomly amplified polymorphic DNA). *Lett. Appl. Microbiol.* 18: 260-263.
- Stephen, D., Jones, C. et Schofield, J. P. 1990. A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucl. Acids Res.* 18:7463-7464.
- Suen G., Goldman, B. S. et Welch, R. D. 2007. Predicting prokaryotic ecological niches using genome sequence analysis. *PLOS ONE*, 2 (8): e743.
- Suzuki, K., Goodfellow, M. et O'Donnell, A. G. 1993. Cell envelopes and classification. In *Handbook of new bacterial systematics*. pp. 195-250. Goodfellow M. and O'Donnell A. G. (éds.) Academic Press Ltd, London.
- Tambong, J. T., Xu, R. et Bromfield, E. S. P. 2009. Intercistronic heterogeneity of the 16S-23S rRNA spacer region among *Pseudomonas* strains isolated from subterranean seeds of hog peanut (*Amphicarpa bracteata*). *Microbiology* 155: 2630-2640.
- Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H. et Swaminathan, B. 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacteria strain typing. *J. Clin. Microbiol.* 33: 2233-2239.
- Thompson, J. D., Higgins, D. G. et Gibson, T. J. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22: 4673-4680.
- Ticknor, L. O., Kolsto, A. B., Hill, K. K., Keim, P., Laker, M. T., Tonks, M., et Jackson, P. J. 2001. Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates, *Appl. Environ. Microbiol.* 67:4863–4873.

- Towalski, Z. et Rothman, H. 1986. Enzyme technology. In *The Biotechnological challenge*. pp 37-76. Jacobsson A. J. et Rothman H. (éds.) Cambridge University Press, Cambridge.
- Towner, K. J. et Cockayne, A. 1993. Molecular methods for microbial identification and typing. Chapman et Hall (éds.) London.
- Trébaol, G., Gardan, C., Manceau, J., Tanguy, Y., Trilly, Y. et Boury, S. 2000. Genomic and phenotypic characterisation of *Xanthomonas cynarae*; a new species causing bacterial bract spot of artichoke (*Cynara scolymus L.*). *Int. J. Syst. Evol. Microbiol.* 50: 1471–1478.
- Turnbull, P. C. B. 2002. Introduction: Anthrax history, disease and ecology. In: *Anthrax*, pp. 1–20. Koehler T. (Ed.) Springer-Verlag, Berlin.
- Uma Maheswar Rao, J. L. et Satyanarayana, T. 2007. Purification and characterization of a hyperthermophilic and high maltogenic α -amylase of an extreme thermophile *Geobacillus thermoleovorans*. *Appl. Biochem. Microbiol.* 142 (2): 179-193.
- Vandamme, P., Pot, B., Gillis, P., De vos, K. et Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60: 407-438.
- Vauterin, L., Yang, P., Hoste, B., Vancanneyt, M., Civerolo, E. L., Swings, J. et Kersters, K., 1991. Differentiation of *Xanthomonas campestris* pv. Citri strains by sodium dodecyl sulphate-polyacrylamide gel electrophoresis of proteins, fatty acid analysis, and DNA-DNA hybridization. *Int. J. Syst. Bacteriol.* 41: 535-542.
- Vauterin, L., Hoste, B., Kersters, K. et Swings, J. 1995. Reclassification of *Xanthomonas*. *Int. J. Syst. Bacteriol.* 45: 472–489.
- Vauterin, L., Swings, J., Kersters, K., Gillis, M., Mew, T. W., Schroth, M. N., Palleroni, N. J., Hildebrand, D. C., Stead, D. E. et other authors 1990. Towards an improved taxonomy of *Xanthomonas*. *Int. J. Syst. Bacteriol.* 40: 312–316.
- Vilas-Boas, G. T., Peruca, A. P. et Arantes, O. M. 2007. Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis*, and *Bacillus thuringiensis*. *Can. J. Microbiol.* 53:673–687.

- Wang, M., Cao, B., Yu, Q., Liu, L., Gao, Q., Wang, L. et Feng, L. 2008. Analysis of the 16S-23S rRNA gene internal transcribed spacer region in *Klebsiella* species. *J. Clin. Microbiol.* 46: 3555-3563.
- Ward, D. M., Weller, R. et Bateson, M. M. 1990. 16S RNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345: 63-65.
- Watson, J. et Crick, F. 1953. Molecular structure of nucleic acids: A structure for deoxyribonucleic acids. *Nature* 171: 737-740.
- Watson, J., Hopkins, N., Roberts, J., Steitz, J. et Weiner, A. M. 1987. Molecular biology of the gene, fourth edition. The Benjamin/Cummings Publishing Company Inc., Menlo Park, CA.
- Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont P .A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrabdt, E., Starr, M. P. et Trüper, H. G. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37: 463-464.
- Wellman-Desbiens, É. et J.-C. Côté. 2004. Screening of the insecticidal activity of *Bacillus thuringiensis* strains against *Lygus hesperus* Knight (Hemiptera: Miridae) nymphal population. *J. Econ. Entomol.* 97: 251-258.
- Welsh, H. et McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.* 18: 7213-7218.
- Williams, J. G. K., Hanafey, M. K., Rafalski, J. A. et Tingey, S. V. 1993. Genetics analysis using random amplified polymorphic DNA markers. *Methods Enzymol.* 218: 704-740.
- Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. et Tingey, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18: 6531-6535.
- Woese, C. R., 1981. Archaebacteria. *Scientific American* (Jun): 98-122.
- Woese, C. R. 1987. Bacterial evolution. *FEMS Microbiol. Rev.* 51: 221-271.
- Woese C. 1998. The universal ancestor. *Proc. Natl. Acad. Sci. USA*, 95 (12) 6854-9.

- Woese, C. R., Kandler, O. et Wheelis, M. L. 1990. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria* and *Eucarya*. *Proc. Natl. Acad. Sci. USA*, 87: 4576-4579.
- Woese, C. R., Stackebrandt, E., Macke, T. J., et Fox, G. E. 1985a. A phylogenetic definition of the major eubacterial taxa. *Syst. Appl. Microbiol.* 6:143-151.
- Woese, C. R., Stackebrandt, E., Weisburg, W. G., Paster, B. J., Madigan, M. T., Fowler, V. J., Hahn, C. M., Blanz, P., Gupta, R., Nealson, K. H., et Fox, G. E. 1984a. The phylogeny of purple bacteria: the alpha subdivision *Syst. Appl. Microbiol.* 5: 315-326.
- Woese, C. R., Weisburg, W. G., Hahn, C. M., Paster, B. J., Zablen, L. B., Lewis, B. J., Macke, T. J., Ludwig, W., et Stackebrandt, E. 1985b. The phylogeny of purple bacteria: the gamma subdivision. *Syst. Appl. Microbiol.* 6:25-33.
- Woese, C. R., Weisburg, W. G., Paster, B. J., Hahn, C. M., Tanner, R. S., Krieg, N. R., Koops, H.-P., Harms, H., et Stackebrandt, E. 1984b. The phylogeny of purple bacteria: the beta subdivision. *Syst. Appl. Microbiol.* 5:327-336.
- Woese, C.R. et Gupta R. 1981. Are Archaebacteria merely derived 'Prokaryotes'? *Nature*. 289 (5793):95-6.
- Wolf, J. et Barker, A. N. 1968. The genus *Bacillus*:Aids to the identification of its species. In: *Identification methods for microbiologist*. pp. 93-109. Academic Press, New York.
- Wunschel, D., Fox, K. F., Black, G. E. et Fox, A. 1994. Discriminating among the *B. cereus* group, in comparison to *B. subtilis* by structural carbohydrate profiles and ribosomal RNA spacer region PCR *Syst. Appl. Microbiol.* 17: 625-635.
- 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.
- Yakoubou, S. et Côté, J.-C. 2010. Assessment of a short phylogenetic marker based on comparisons of 3' end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences of the *Bacillus cereus* group. *Nat. Sci.* In Press.
- Yakoubou, S. et Côté, J-C. 2010. Phylogeny of γ -proteobacteria inferred from comparisons of 3' end 16S rRNA gene and 5' end 16S-23S ITS nucleotide sequences. *Nat. Sci.* 2 (6): 535-543.

- Yakoubou, S., Xu, D. et Côté J.-C. 2010. Phylogeny of the Order *Bacillales* inferred from 3' 16S rDNA and 5' 16S-23S ITS nucleotide sequences. *Nat. Sci.* In Press.
- Yang. P., Vauterin L., Vancanneyt, M., Swings J., et Kersters K. 1993. Application of fatty acid methyl esters for the taxonomic analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 16: 47-71.
- Yoon, J.-H., Lee, S. T., Kim, S.-B., Goodfellow, M. et Park, Y.-H. 1997. Inter- and intraspecific genetic analysis of the genus *Saccharomonospora* with 16S to 23S ribosomal DNA (rDNA) and 23S to 5S rDNA internally transcribed spacer sequences. *Int. J. Syst. Bacteriol.* 47:661–669.
- Young, J. M., Park, D.-C., Shearman, H. M. et Fargier, E. 2008. A multilocus sequence analysis of the genus *Xanthomonas*. *Syst. Appl. Microbiol.* 31: 366–377.
- Yutin, N., Makarova, K. S., Mekhedov, S. L., Wolf, Y. I., Koonin, E. V. 2008. The deep archaeal roots of eukaryotes. *Mol. Biol. Evol.*, 25: 1619-1630.
- Zhang, Y., Ju, J., Peng, H., Gao, F., Zhou, C., Zeng, Y., Xue, Y., Li, Y., Henrissat, B., Gao, G. F. et Ma, Y. 2008. Biochemical and structural characterization of the intracellular mannanase AaManA of *Alicyclobacillus acidocaldarius* reveals a novel glycoside hydrolase family belonging to clan GH-A. *J. Biol. Chem.* 283: 31551-31558.
- Zhilina, T. N., Garnova, E. S., Tourova, T. P., Kostrikina, N. A. et Zavarzin, G. A. 2001. *Amphibacillus fermentum* sp. nov. and *Amphibacillus tropicus* sp. nov., new alkaliphilic, facultatively anaerobic, saccharolytic bacilli from Lake Magadi. *Mikrobiologiya*. 70: 825-837 (in Russian). English translation: *Microbiology* 70: 711-722.