

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE Abdelhamid Ibn Badis

Faculté des Sciences de la Nature et de la Vie

Département de biologie

Spécialité: génétique fondamentale et appliquée

Mémoire de fin d'études

Présentée par : Bahnes Fethi

Komichi Salah

En vue de l'obtention du Diplôme de Master

Génétique fondamentale et appliquée

Thème

**Approche bioinformatique pour l'analyse
et l'indentification moleculaire d'une bactérie du genre *bacillus***

Président : CHIBANI Abdelwaheb Professeur

Université de Mostaganem

Examineur : Guedouar Youcef M.C.B

Université de Mostaganem

Rapporteur : Dahmani Chahinez Amira M.A.B

Université de Mostaganem

Année universitaire : 2020-2021

REMERCIEMENT

En premier lieu, nous tenons à remercier le DIEU notre créateur, pour nous avoir donné la force, le courage et la patience d'accomplir ce travail.

Nous adressons nos vifs remerciements à :

La président : Monsieur **Chibani Abdelwaheb**

Professeur au département de Génétique Moléculaire Appliquée (université mosaganem) qui nous a fait un grand honneur de présider notre jury. Qu'il trouve, ici, notre témoignage de notre gratitude et de notre profond respect.

L'examineur : Monsieur **Guedouar Youcef**

Doctorant à l'université de Mostaganem qui nous a fait l'honneur de bien vouloir examiner ce travail. Qu'il soit assurée de notre vive reconnaissance et de notre profond respect.

Maitre conférences à l'université Es-Sania Oran qui nous a fait l'honneur de bien vouloir examiner ce travail. Qu'elle soit assurée de notre vive reconnaissance et de notre profond respect.

Notre encadreur : Mme **Dahmani Chahinez Amira**

Doctorante au département de GMA à (USTO). Nous vous remercions d'avoir consacré autant de temps à lire et corriger notre mémoire, pour votre compréhension, patience et vos remarques qui ont été précieuses. Qu'elle soit assurée de notre profonde reconnaissance, notre respect et toute notre gratitude.

Nos plus sincères remerciements vont également aux l'ensemble du personnel département de Génétique Moléculaire Appliquée.

Ainsi qu'à tous ceux qui nous aidés de près ou de loin.

Dédicaces

Je dédie ce modeste travail à :

A mes parents

Aucun hommage ne pourrait être à la hauteur de l'amour Dont ils ne cessent de me combler.

Que dieu leur procure bonne santé et longue vie.

A Mes beaux parents **AFIF** et **MALIKA**.

A mes chers frères **Mohamed** et **AMINE**.

A ma seule sœur que j'aime beaucoup **IMANE**.

A mes amis **HICHAM** et **SIDALI** et **ABDELHAK** et **HAMED** et **SIDALI** et **Sofiane**

A toute la promotion de Biologie et contrôle de qualité.

A mon binôme **SALAH** et toute la famille **KOMICHI**

Dédicaces

Je dédie ce travail à tous ceux qui m'ont soutenu de près et de loin,

A ma mère, et mon père

A mes sœurs et

A mes proches

A tous mes amis

Résumé

L'amidon est le polymère glucidique le plus abondant dans la nature après la cellulose et la chitine et est utilisé par les plantes comme source de carbone et d'énergie. En plus des intérêts que présente ce bio-polymère, les enzymes responsables de sa dégradation ont suscité un grand intérêt. Ces enzymes sont les amylases, notamment les α -amylases qui reçoivent un intérêt grandissant vu leur large champ d'application dans le traitement de l'amidon dans différentes industries. La source industrielle la plus utilisée pour la production des α -amylases parmi les genres bactériens est sans conteste le genre *Bacillus*.

Le séquençage du gène codant pour l'ARN ribosomal 16S (ARNr 16S) permet l'étude de la phylogénie et de la taxonomie bactériennes. Ce gène de 1500 pb environ est présent dans la plupart des bactéries et sa fonction n'a pas changé au cours de l'évolution. Par ailleurs, sa taille est suffisante pour une analyse informatique puissante.

Ce mémoire a eu pour objectif de réaliser la dernière étape de l'identification moléculaire d'une bactérie du genre *Bacillus* ayant pour rôle la production d'enzymes de type α -amylases par l'étude du gène codant l'ARNr 16S. Ces souches bactériennes du genre *Bacillus* ont été isolées et étudiées préalablement dans la thèse de Mr. CHOUBANE Slimane. En effet, nous avons utilisé la base de données BLAST et le logiciel MEGA X version 10,2,6 afin de réaliser l'approche bio-informatique.

La constitution de l'arbre phylogénétique de *Bacillus Mojavensis* nous a permis de la classer et la positionner comme étant non pathogène mais productrice d'enzymes principalement une souche identifiée comme appartenant à l'espèce *Bacillus subtilis* et est apparenté génétiquement aux espèces *Bacillus spp.*, utilisées pour la production d'enzymes industrielles.

Mots clés : ARNr 16S, bio-informatique, séquençage, taxonomie, arbre phylogénétique.

Abstract

Starch is the most abundant carbohydrate polymer in nature after cellulose and chitin and is used by plants as a source of carbon and energy. In addition to the interests of this biopolymer, the enzymes responsible for its degradation have aroused great interest. These enzymes are the amylases, in particular the α -amylases which are receiving increasing interest due to their wide field of application in the processing of starch in different industries. The most widely used industrial source for the production of α -amylases among bacterial genera is undoubtedly the genus *Bacillus*.

Sequencing of the gene encoding 16S ribosomal RNA (16S rRNA) allows the study of bacterial phylogeny and taxonomy. This around 1500 bp gene is present in most bacteria and its function has not changed over time. Moreover, its size is sufficient for a powerful computer analysis.

The objective of this thesis was to carry out the last step in the molecular identification of a bacterium of the genus *Bacillus* whose role is the production of α -amylase-type enzymes by studying the gene encoding 16S rRNA. These bacterial strains of the genus *Bacillus* were isolated and studied beforehand in the thesis of Mr. CHOUBANE Slimane. Indeed, we used the BLAST database and MEGA X software version 10,2,6 in order to carry out the bioinformatics approach.

The constitution of the phylogenetic tree of *Bacillus Mojavensis* allowed us to classify and position it as being non-pathogenic but producing enzymes mainly a strain identified as belonging to the species *Bacillus subtilis* and is genetically related to the species *Bacillus* spp., used for the production of industrial enzymes.

Keywords : 16S rRNA, bioinformatics, sequencing, taxonomy, phylogenetic tree.

ملخص

النشا هو بوليمر الكربوهيدرات الأكثر وفرة في الطبيعة بعد السليلوز والكتين وتستخدمه النباتات كمصدر للكربون والطاقة. بالإضافة إلى مصالحي هذا البوليمر الحيوي ، أثارت الإنزيمات المسؤولة عن تحليلها اهتمامًا كبيرًا. هذه الإنزيمات هي الأميلاز ، ولا سيما الأميلاز ألفا والتي تحظى باهتمام متزايد بسبب مجال تطبيقها الواسع في معالجة النشا في الصناعات المختلفة. المصدر الصناعي الأكثر استخدامًا لإنتاج الأميلاز ألفا بين الأجناس البكتيرية هو بلا شك جنس

Bacillus.

يسمح تسلسل الجين المشفر للحمض النووي الريبوزي 16 بدراسة السلالات البكتيرية والتصنيف.

هذا الجين حوالي 1500 زوج قاعدي موجود في معظم البكتيريا ولم تتغير وظيفته بمرور الوقت. علاوة على ذلك، حجمه كافٍ لتحليل كمبيوتر قوي.

كان الهدف من هذه الرسالة هو تنفيذ الخطوة الأخيرة في التعريف الجزيئي لبكتيريا من جنس العصوية التي يتمثل دورها في إنتاج إنزيمات من نوع أميلاز ألفا عن طريق دراسة الجين المشفر. تم عزل هذه السلالات البكتيرية من جنس (*Bacillus*) ودرستها مسبقًا في أطروحة السيد شوبان سليمان في الواقع ، استخدمنا قاعدة بيانات وإصدار برنامج

MEGA X 10,2,6. من أجل تنفيذ نهج المعلوماتية الحيوية .

سمح لنا تكوين شجرة النشوء والتطور في *Bacillus Mojavensis*

بتصنيفها ووضعها على أنها غير مسببة للأمراض ولكنها تنتج الإنزيمات بشكل أساسي وهي سلالة تم تحديدها على أنها تنتمي إلى الأنواع المستخدمة في إنتاج الإنزيمات الصناعية وترتبط وراثيًا بالأنواع

(*Bacillus* spp)

الكلمات الرئيسية: شجرة النشوء والتطور ، المعلوماتية الحيوية ، التسلسل ، التصنيف ،

.RNAr16s

Sommaire

Liste des figures

Liste des tableaux

Liste des abréviations

REVUE BIBLIOGRAPHIQUE

I. Biologie de l'amidon.....	p (1)
I.1. Définition de l'amidon.....	p (1)
I.2. Les enzymes amylases.....	p (2)
II. Le genre Bacillus.....	p (2)
II.1. Définition du groupe Bacillus.....	p (2)
II.2. Définition d'une espèce bactérienne.....	p (3)
II.3. Taxonomie du Bacillus.....	p (5)
III. Le ribosome bactérien.....	p (8)
III.1. Définition.....	p (8)
III.2. L'ARNr 16S.....	p (9)
III.2,1. Définition et Origines.....	p (9)
III.2,2. Utilisation	p (9)
IV. Place de la PCR universelle 16S dans l'identification bactérienne...	p (13)
IV.1. La technique PCR.....	p (13)
IV.2. L'approche PCR 16S.....	p (14)
V. Le séquençage 16S.....	p (16)
V.1. Les types de séquençages.....	p (18)

V.1,1. Méthode Sanger.....	p (18)
V.1,2. Méthode de Maxam et Gilbert.....	p (20)
V.1,3. La technique shotgun.....	p (22)
V.2. Le principe de séquençage de grands génomes.....	p (22)
VI. L'identification bactérienne par séquençage du gène ARN 16S.....	p (25)
VII. L'analyse bio-informatique.....	p (27)
VII.1 Alignement de séquence	p (27)
VII.1,1. Alignement local ou global.....	p (28)
VII.1,2. Alignement par paire ou multiple.....	P (29)
VII.2. Etude phylogénétique.....	p (29)
VII.2,2. Arbre phylogénétique.....	p (30)

OBJECTIFS DU MEMOIRE

MATERIEL ET METHODE

I. Isolement des colonies et extraction d'ADN.....	p (35)
II. Amplification du gène ribosomique ARNr 16S.....	p (35)
III. Séquençage du gène ARNr 16S.....	p (36)
IV. Analyse phylogénique du gène ARNr 16S.....	p (36)

RESULTATS ET DISCUSSION

I. Lecture et alignement de la séquence du gène ARNr 16S.....	p (40)
II. Construction de l'arbre phylogénétique de la souche <i>Bacillus</i>	p (41)

CONCLUSION

REFERENCES BIBLIOGRAPHIQUES

Liste des figures

- Figure 01** : Principe de l'hybridation ADN-ADN entre deux espèces bactériennes. Plus les espèces sont proches, moins de mésappariements surviennent pendant l'hybridation.....p (4)
- Figure 02** : Phylogénie de *Bacillus* à partir des séquences de gènes ARNr 16S.....p (7)
- Figure 03** : L'ARN ribosomique 16S : un constituant de la petite sous-unité ribosomale des procaryotes.....p (8)
- Figure 04** : Représentation schématique du gène codant pour l'ARNr 16S. Les cercles représentent les régions conservées utilisées pour le design d'amorces nécessaires à l'amplification et au séquençage.....p (10)
- Figure 05** : Structure secondaire de l'ARN ribosomique 16S d'*Escherichia coli*.....p (11)
- Figure 06** : Aperçu schématique de la réaction de PCR.p (14)
- Figure 07** : Figure représentative de la séquence de ARNr 16S et des sites de liaison d'amorce.....p (16)
- Figure 08** : Principes du séquençage selon la méthode de Sanger.....p (20)
- Figure 09** : Technique de Maxam-Gilbert.....p (21)
- Figure 10** : Types d'alignements.....p (28)
- Figure 11** : alignement de deux séquences et alignement multiple.....p (29)
- Figure 12** : Exemple de représentation de la parenté entre les grands groupes d'êtres vivants, à l'aide de cet arbre phylogénétique simplifié et buissonnant.....p (31)
- Figure 13** : Photo de l'interface de la base de données BLAST sur NCBI.....p (37)
- Figure 14** : Photo représentant l'interface du logiciel Mega X version 10,2,6.....p (38)
- Figure 15** : Photo qui représente l'alignement de séquences du gène ARNr 16S de *Bacillus Mojavensis* (PF2_TPU1) avec d'autres séquences d'espèces connues par le logiciel Mega X.....p (41)
- Figure 16** : Arbre phylogénétique de la souche bactérienne étudiée dans notre mémoire *Bacillus Mojavensis* (PF2_TPU1) réalisé par le logiciel Mega X.....p (42)

Liste des tableaux

Tableau 1 : Résumé des différentes amorces utilisées pour la PCR16S dans un contexte de détection bactérienne à partir des prélèvements cliniques.....p (12)

Tableau 2 : Exemples d'oligonucléotides standard utilisés dans l'amplification des gènes 16S ARNr.....p (15)

Tableau 3 : Recommandations pour l'identification bactérienne par séquençage de l'ARNr 16S.....p (27)

Liste des abréviations

ADN : acide désoxyribonucléique

ARN : acide ribonucléique

ARNr ou rRNA : acide ribonucléique ribosomique

ARNr16S ou 16S rRNA : constituant ARN de la petite sous-unité Ribosomale du 30S des procaryotes

A : Adénine

C : Cytosine

T : Thymine

G : Guanine

ddntp : Deoxyribonucleotide triphosphate. A generic term referring to the four deoxyribonucleotides: **dATP**, **dCTP**, **dGTP** and **dTTP**.

16SrRNA gene : le gène codant les ARN ribosomaux 16S réaction en chaîne par polymérase

S : Le Svedberg, de symbole S , est une unité de mesure du taux de sédimentation.

STS : sequence tagged sites

B : Bactérie

BAC : Bacterial Artificial Chromosome

DMS : Dimethyl Sulfide

F : Forward

P : Phage Artificial Chromosome

R : Reverse

RCA : Rolling-circle Amplification

RFLP : Restriction Fragment Length Polymorphism

PCR : Polymrase Chain Reaction

REVUE BIBLIOGRAPHIQUE

I. Biologie d'amidon

I.1. Définition de l'amidon

L'amidon est le polymère glucidique le plus abondant dans la nature après la cellulose et la chitine et est utilisé par les plantes comme source de carbone et d'énergie (*Taggart et Eliasson, 2004*). Aussi, l'amidon est le polysaccharide alimentaire le plus important, le plus abondant et le plus digestible, et représente ainsi une source majeure d'énergie dans notre régime alimentaire (*Peris-Tortajada et Eliasson, 2004*). En effet l'importance de l'amidon dans l'alimentation peut être vérifiée par le fait qu'il représente 30% de la moyenne du régime alimentaire, et plus de 25% de l'énergie fournie (*Butler et al. 2004*).

L'amidon est la réserve principale de sucres chez toutes les plantes supérieures. Dans certains cas il représente 70% du poids sec du matériel végétal. Il se présente sous la forme de granules insolubles dans l'eau. L'amidon est un polysaccharide hétérogène composé de deux entités à poids moléculaire élevé appelées : amylose et amylopectine. Ces deux polymères ont des structures et des propriétés physiques différentes (*Aiyer 2005*).

Les sources de l'amidon sont diverses à savoir les céréales, les tubercules et certains fruits ; cependant la majorité de l'amidon utilisé à travers le monde vient d'un nombre réduit de cultures telles que le maïs, la pomme de terre, le blé et le tapioca (*Wang et al. 1998*). Ce polysaccharide est utilisé dans différentes applications telles que : l'industrie alimentaire, pharmaceutique, médicale, biomatériaux, bio-alcool etc.

En plus des intérêts que présente ce bio-polymère, les enzymes responsables de sa dégradation ont suscité un grand intérêt. Ces enzymes sont les amylases, appartenant majoritairement aux glycosides hydrolases.

I.2. Les enzymes amylases

Les amylases, notamment les α -amylases, reçoivent un intérêt grandissant vu leur large champ d'application dans le traitement de l'amidon dans différentes industries (alimentaire, textile, détergents, pharmaceutique, bio-alcools etc.).

Les amylases, sont distribuées dans les mondes animal, végétal et microbien. Cependant, les amylases à partir des sources microbiennes ont dominé les applications industrielles grâce aux grandes quantités d'enzymes produites et leurs propriétés physicochimiques intéressantes (Hussain *et al.* 2013).

Les α -Amylases appartiennent au groupe des endoamylases. Ce sont des calcium-métalloenzymes, complètement incapable de fonctionner en l'absence de calcium.

Les α -amylases agissent à des sites aléatoires tout au long de la chaîne d'amidon, libérant le maltotriose et le maltose à partir de l'amylose, ou maltose, glucose et limit-dextrines à partir de l'amylopectine. Les α -Amylases tendent à être plus rapide que les β -amylases parce qu'elles peuvent agir n'importe où sur le substrat (Singh *et al.* 2011).

La source industrielle la plus utilisée pour la production des α -amylases parmi les genres bactériens est sans conteste le genre *Bacillus* notamment les espèces *Bacillus subtilis*, *Bacillus stearothermophilus*, *Bacillus licheniformis* et *Bacillus amyloliquefaciens* qui sont connues pour être de bonnes productrices d'alpha-amylases thermostables (Sundarram et Murthy 2014).

II. Le genre Bacillus

II.1. Définition du groupe Bacillus

Le genre bactérien *Bacillus* a une histoire longue et riche dans les annales de la microbiologie. Dans les années 1870, Ferdinand Cohn travaillant à l'université de Breslau (Pologne), a isolé une bactérie petite, motile et aérobie à partir d'une infusion de foin bouillie. Il l'a nommé *Bacillus subtilis* (bâtonnet fin). Cohn a fourni la première description précise du cycle de vie de *Bacillus* (Daniel et John 2009). La création de ce genre en 1872 par Ferdinand

Cohn, venait de changer le nom donné par Ehrenberg en 1835 à une des plus anciennes bactéries décrites à savoir *Vibrio subtilis* (Slepecky et Hemphill 2006).

Le genre décrit par Cohn fait partie de la famille des *Bacillacées* dont le caractère spécifique est la formation d'endospores (structures ovales, circulaires ou cylindriques très réfractiles qui se forment à l'intérieur de la bactérie).

Le genre *Bacillus* représente un ensemble très large et diversifié de bactéries qui ont cependant une caractéristique commune: la capacité de former des endospores dormants en aérobie quand elles sont confrontées à des conditions de croissance défavorables (Zeigler et Perkins 2009).

Ce genre est l'un des groupes de microorganismes les plus utiles et les plus diversifiés. Il a deux attributs majeurs qui le rendent intéressant : les endospores robustes et la production d'une grande variété d'enzymes. Grace à la résistance des endospores à l'altération par la chaleur, le froid et les produits chimiques ; les industriels peuvent inclure ces bactéries dans des formules de produits avec la certitude que celles-ci restent en vie.

Les représentants du genre sont largement distribués dans le sol, l'air et l'eau. La diversité métabolique des *Bacillus spp* ainsi que la faible incidence de pathogénicité rapportée, a mené à l'utilisation des membres de ce genre dans un large champ d'applications industrielles. Les souches *Bacillus spp* sont utilisées pour la production de quatre types essentiels de produits à savoir : les enzymes, les antibiotiques, les produits de chimie fine et les insecticides (Harwood, 1989).

II.2. Définition d'une espèce bactérienne

La définition du concept d'espèce bactérienne est un point délicat ; la correspondance entre une souche et une espèce déjà décrite est basée sur une similarité phénotypique et génétique (Gevers et al. 2005). Les méthodes actuelles utilisées pour définir les espèces procaryotes ne permettent pas de couvrir la diversité trouvée dans la nature et la taxonomie bactérienne est influencée par les avancées dans la génétique des populations,

l'écologie, la génomique et par la facilité avec laquelle les données peuvent être obtenues (Gevers et al. 2005).

Traditionnellement, une espèce bactérienne est définie à partir de l'hybridation ADN-ADN, qui est une technique lourde, complexe, onéreuse et de moins en moins utilisée (Figure 1) (Stackebrandt et al. 2002). Si on se base sur la cinétique de réassociation ADN-ADN, la définition génétique d'une espèce est quantifiable. Une espèce est alors définie génétiquement comme le rassemblement de souches qui ont des relations ADN-ADN se traduisant par (i) un pourcentage d'hybridation ADN-ADN 70% et (ii) une stabilité thermique des hybrides 5°C (Janda and Abbott, 2007).

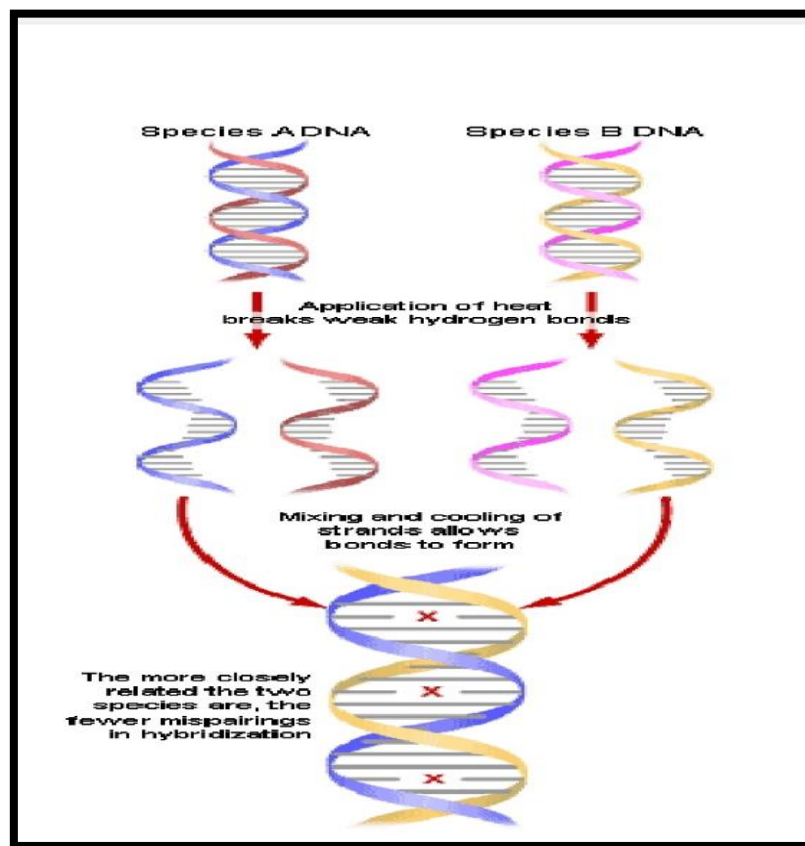


Figure 1 : Principe de l'hybridation ADN-ADN entre deux espèces bactériennes. Plus les espèces sont proches, moins de mésappariements surviennent pendant l'hybridation (Renvoisé A 2012).

II.3. Taxonomie du *Bacillus*

Les premières tentatives de classification des espèces *Bacillus* ont été basées sur deux caractéristiques: la croissance aérobie et la formation d'endospores.

Ces tentatives ont résulté en le groupement d'un grand nombre de bactéries qui possèdent différents traits physiologiques et génétiques. Cependant, depuis l'avènement des technologies de séquençage, la classification phylogénétique des espèces bactériennes a été rendue possible (O'Sullivan 2005).

Smith *et al.* (1946) a rassemblé un ensemble méticuleux et rigoureux de données morphologiques et physiologiques à partir d'une collection de 1134 souches *Bacillus spp.* et a réussi de les placer dans 19 espèces définies (Smith *et al.* 1946). Après cela, Smith et Gordon ont refait encore le même travail sur 206 souches thermophiles qui ont été placées dans deux espèces additionnelles (Gordon et Smith 1949).

En 1973, dans la monographie désormais classique *The Genus Bacillus* (Gordon *et al.* 1973), Ruth Gordon et ses collègues ont fourni une description détaillée et précise de 18 espèces de *Bacillus*, avec une batterie de 33 tests morphologiques et physiologiques et une clé d'identification, qui peut sans ambiguïté assigner la plupart des isolats connus à l'une de ces espèces. La taxonomie de *Bacillus* est devenue une science logique, ordonnée et lucide (Zeigler et Perkins 2009). En 1989, Ruth Gordon a été l'auteur d'un chapitre dans le livre *Practical Handbook of Microbiology* (CRC Press) qui a passé en revue la systématique du genre *Bacillus* en utilisant les données les plus extensives à ce moment-là (Gordon *et al.* 1989).

Ensuite, les différentes espèces de *Bacillus* ont été divisées en trois groupes selon le phénotype des sporanges (ex. gonflées vs. Non-gonflées) et des spores matures (ex. sphérique, cylindrique ou ellipsoïde) et selon des tests biochimiques (ex. catalase, hydrolyse d'amidon) et les propriétés de croissance (ex. 7% NaCl). Cependant les espèces au sein de ces groupes étaient très hétérogènes. Depuis ce temps, des reclassifications majeures ont été faites grâce à des

méthodes d'analyse plus sophistiquées qui s'appuient notamment sur la comparaison des génomes (Zeigler et Perkins 2009). La majorité des espèces *Bacillus* avec des noms valablement publiés, sont phylogénétiquement groupées en sous-groupes à l'intérieur du genre. Toutefois, quelques espèces de *Bacillus* validement nommés ne sont pas reliés à l'espèce type *B. subtilis*, et sont plus proche phylogénétiquement d'autres genres (Ludwig *et al.* 2009).

Les données récentes montrent clairement que dans le groupe de *B. subtilis*, à l'intérieur duquel la délimitation des espèces est très difficile, des gènes tels que *gyrB* permettent une différenciation sur une base génétique (Wang *et al.* 2007).

Quand les nouveaux outils de taxonomie moléculaire (tel que le séquençage de l'ARNr 16S) ont été appliqués au genre *Bacillus*, les données obtenues ont mis en cause le paradigme existant, que le genre était composé d'un petit nombre d'espèces rigoureusement défini identifiable par certain caractéristiques phénotypiques (Zeigler et Perkins 2009). Ash *et al.* (1991) ont utilisé cette nouvelle technologie pour analyser minutieusement la majorité des espèces connues de *Bacillus* et ont conclu qu'elles tombent dans cinq groupes distincts selon la similarité des séquences ARNr 16S, et ont conclu que ces groupes correspondent à cinq nouveaux genres.

Les espèces *B. subtilis*, *B. licheniformis*, *B. amyloliquefaciens*, *B. megaterium*, *B. cereus*, *B. anthracis*, *B. thuringiensis*, et la plupart des autres espèces connues et étudiées ont été placées dans le même groupe.

L'arbre phylogénétique *sensu lato* se basant sur les séquences ARNr 16S est présenté dans la Figure (14). Cependant il est connu que les séquences de l'ARNr 16S ne permettent pas toujours aux espèces d'être identifiées, et que pour cela d'autres techniques sont nécessaires telle que l'hybridation ADN-ADN (Logan et Vos 2009).

Avec les avancées et le développement des technologies de séquençage de nouvelle génération (NGS), un grand nombre de données de séquençage a été généré dans les dernières

années. L'accumulation rapide de données NGS de génomes complets d'espèces *Bacillus* dans la base de données *Genbank*, a rendu possible la comparaison des différences génomiques sur l'ensemble du génome qui ne peuvent être identifiées dans les analyses des séquences de gènes uniques (Wang et Ash 2015).

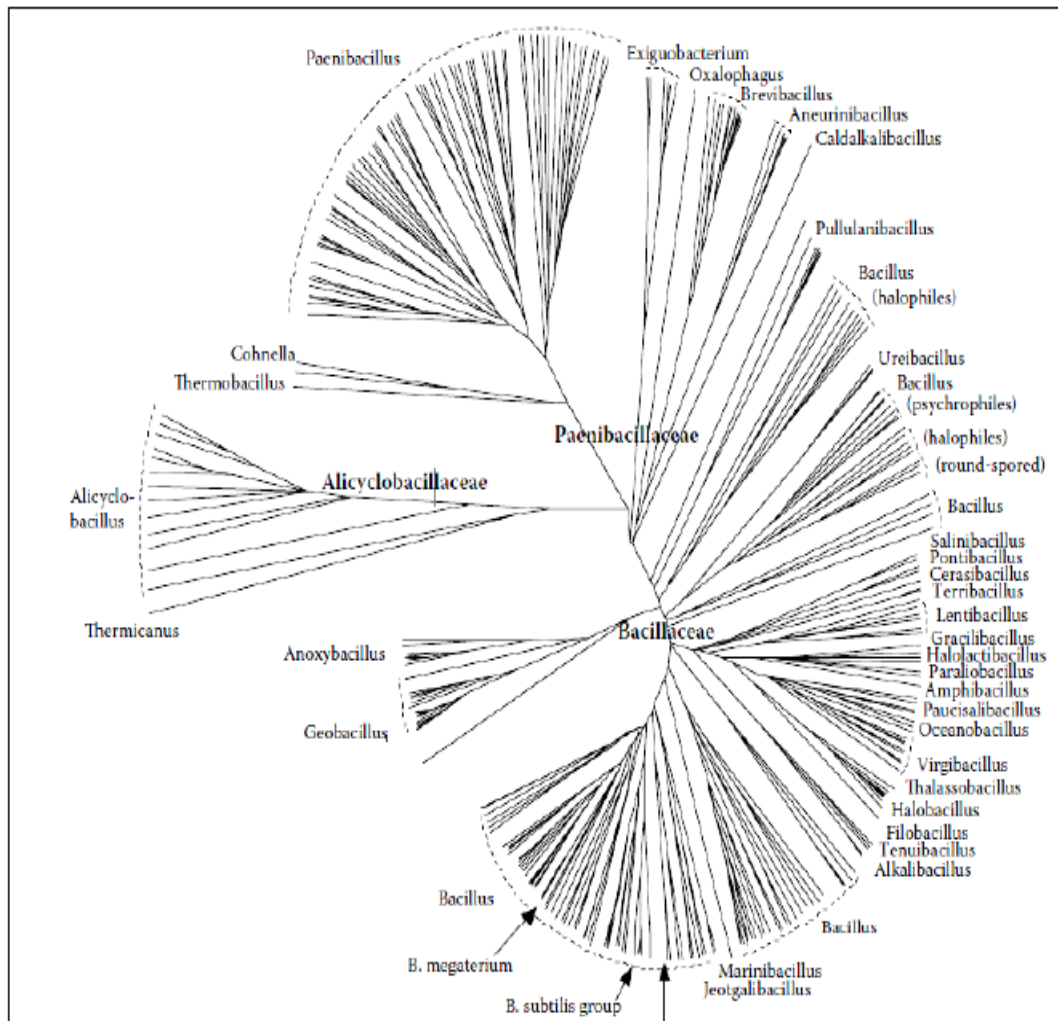


Figure 02 : Phylogénie de *Bacillus* à partir des séquences de gènes ARNr 16S (Ziegler et Perkins, 2009).

III. Le ribosome bactérien

III.1. Définition

Le ribosome est un complexe ribonucléoprotéique (c'est-à-dire composé de protéines et d'ARN) ; il permet la synthèse des protéines en utilisant l'ARN messager comme source d'information et les ARN de transfert associés aux acides aminés comme substrats. Cette synthèse protéique, encore appelée traduction est divisée en cinq phases principales : la liaison de l'ARN messager au ribosome, l'initiation, l'élongation, la terminaison et le recyclage des

Sous-unités ribosomales. Chez les bactéries, le ribosome est composé d'une grande sous unité (50S) et d'une petite sous-unité (30S) (**Figure 3**).

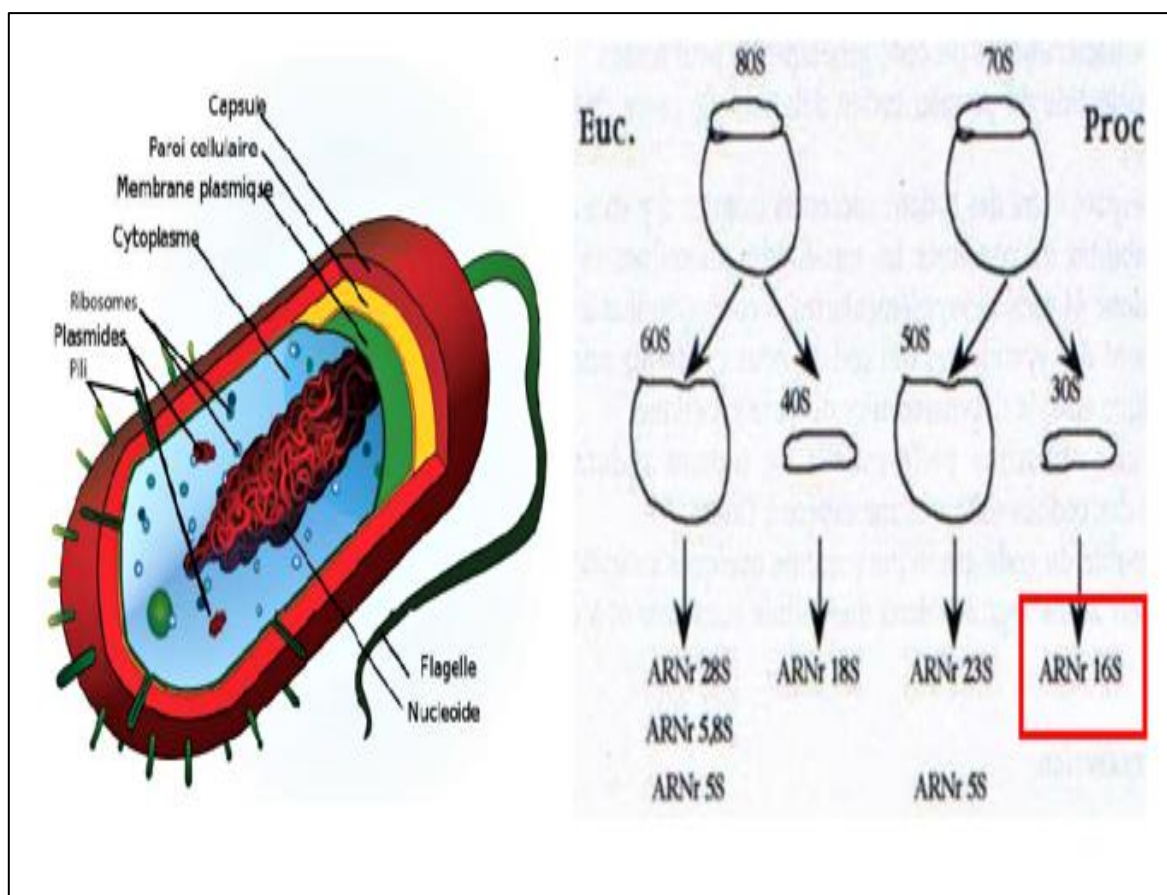


Figure 03 : L'ARN ribosomique 16S : un constituant de la petite sous-unité ribosomale des procaryotes (Aurélien RENVOISÉ 2012).

Au total, le ribosome fonctionnel (composé des deux sous-unités réunies) a une masse moléculaire de 2 5-megadalton et un coefficient de sédimentation de 70S (Schmeing et Ramakrishnan, 2009). La petite sous-unité est composée de l'ARN ribosomal 16S (Figure 2) (codé par un gène d'environ 1500 nucléotides) et d'environ 20 protéines ; elle permet la « lecture de l'ARN messenger ». La grande sous-unité est composée de l'ARN ribosomal 23S (codé par un gène d'environ 2900 nucléotides), de l'ARN ribosomal 5S (codé par un gène d'environ 120 nucléotides) et d'environ 30 protéines ; elle permet la synthèse de la protéine correspondant à l'ARN messenger «lu » par la petite sous-unité. De plus, divers facteurs protéiques agissent sur le ribosome aux différentes étapes de la traduction (Schmeing et Ramakrishnan,2009).

III.2. L'ARNr 16S

III.2,1. Définition et origine :

L'ARN ribosomal (ARNr) 16S est le constituant ARN de la petite sous- unité ribosomale du 30S des procaryotes (**Figure 2**). Le gène codant pour cet ARNr est le «le gène16S ARNr » (Clarridge, 2004), présent dans l'ensemble des espèces bactériennes en un nombre variable de copies (Woese, 1987 ;Petti, 2007).

Il est composé d'environ 1500 nucléotides et est constitué de sept régions conservées et de neuf régions hypervariables (Chakravorty et al. 2007). L'ensemble permet donc théoriquement d'utiliser ce gène pour identifier et détecter toute espèce bactérienne.

III.2,2. Utilisation

Dans les années 1980, il a été montré que les relations phylogéniques entre les êtres vivants pouvaient être déterminées en comparant leurs séquences nucléiques (Clarridge, 2004).

Des travaux préliminaires ont pu montrer que le gène 16S ARNr code pour un ARN ribosomal de fonction constante dans l'évolution ; ce gène peut donc servir d'horloge moléculaire pour suivre les changements dans l'évolution bactérienne. Les séquences du gène 16S ont ainsi joué un rôle majeur dans l'étude de la phylogénie et de la taxonomie bactérienne (Olsen and Woese, 1993 ; Janda et Abbott, 2007).

Ce gène a d'abord été utilisé à la fin des années 1980 par Carl Woese comme outil d'étude de l'évolution bactérienne (Woese, 1987). En 1991, Weisburg et al, ont décrit des amorces dites « universelles » permettant d'amplifier l'intégralité du gène codant pour l'ARNr 16S de la plupart des bactéries (**Figure 5**) (Weisburg et al. 1991).

Actuellement de multiples amorces universelles ont été décrites dans la littérature. Il n'existe malheureusement aucune revue ni aucune base de données qui répertorierait les dizaines d'amorces décrites depuis 20 ans (**Figure 4**). Dans le domaine de la détection bactérienne, Sontakke et al, ont proposé en 2009 une synthèse des diverses amorces utilisées. De plus la nomenclature des amorces reste non consensuelle ; souvent les auteurs nomment une amorce avec un nombre faisant référence à la position de la première base sur l'ADNr 16S d'*Escherichia coli*, suivi de F ou R pour forward ou reverse. Mais cela n'est pas toujours le cas, comme on peut noter pour les amorces fréquemment utilisées fD1 et rP2 (Weisburg et al).

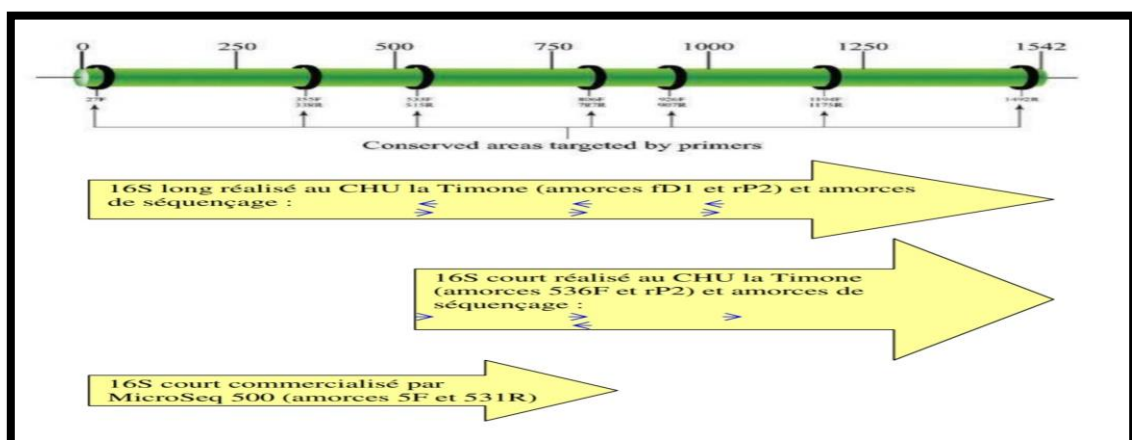


Figure 04 : Représentation schématique du gène codant pour l'ARNr 16S et quelques exemples de choix d'amorces sont donnés (Petti, 2007).

Les cercles représentent les régions conservées utilisées pour le design d'amorces nécessaires à l'amplification et au séquençage.

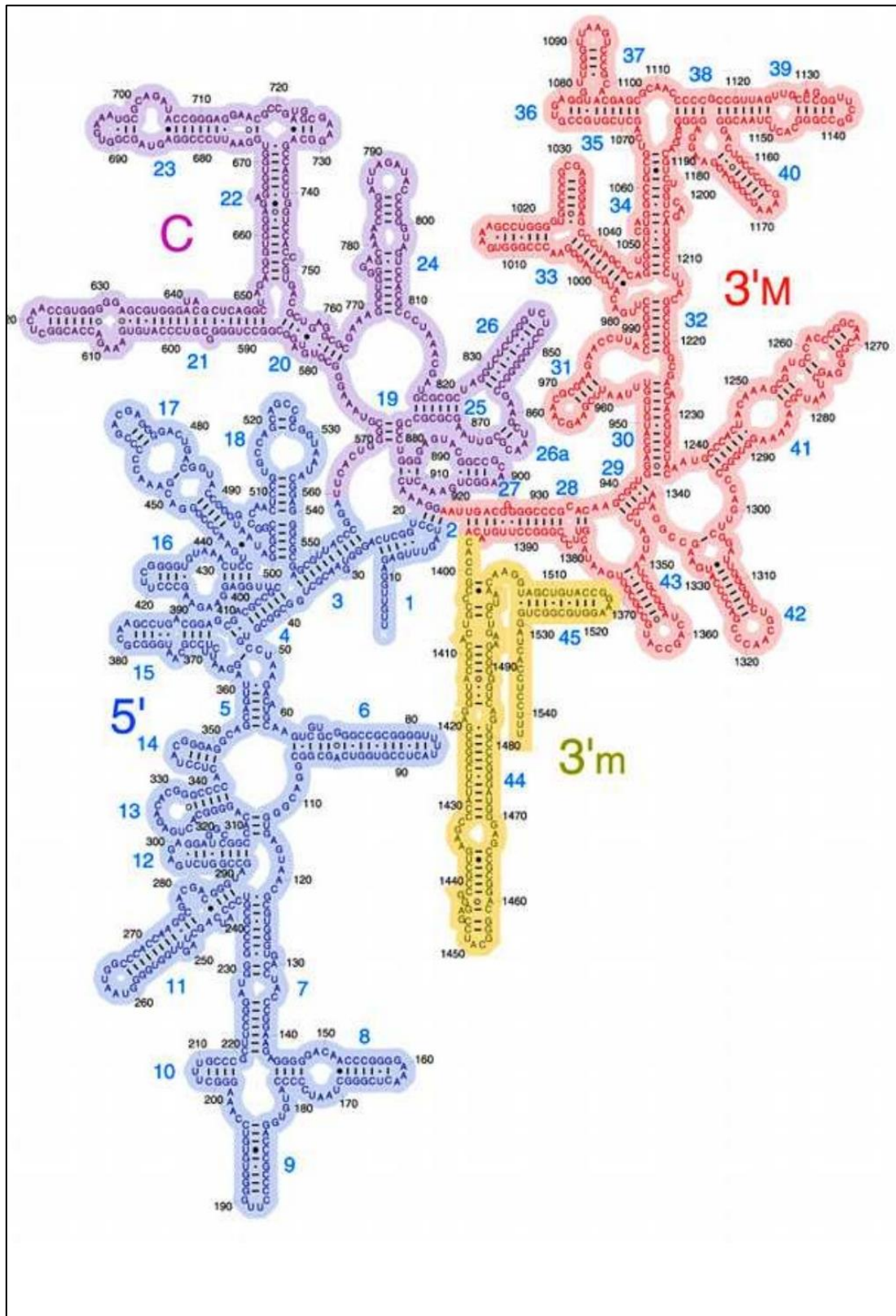


Figure 05 : Structure secondaire de l'ARN ribosomique 16S d'*Escherichia coli* (Renvoisé A, 2012).

Tableau 1 : Résumé des différentes amorces utilisées pour la PCR16S dans un contexte de détection bactérienne à partir des prélèvements cliniques (Sontakkeet et al.2009)

Blood and CSF	PSL	UN1	630 (783-1412)	Meningitis	Xu et al., 2003
	PSL	P13P	607 (783-1389)		
	PSL	PSR	312 (783-1094)		
	P11P	P13P	216 (1174-1389)		
CSF	518-537	1513-1491	996	Meningitis	Pandit et al., 2005
Amniotic fluid	fd1mod	rP2	1500	Subclinical intra amniotic infections	Jalava et al., 1996
Valve	fd1	rP2		Endocarditis	Greub et al., 2005
	536f	1050r			
CSF	518-537	1513-1491	996	Meningitis	Lu et al., 2000
Tissue, blood, pus and CSF	16SFa and 16SFb	16SR	320	Routine clinical microbiology service and diagnosis of bacterial infections	Harris and Hartley 2003
CSF	27F	1492R	1500	Meningitis	Shuurman et al., 2004
Pleural fluid	16SFa and 16Fb	16SR		Empyema	Saglani et al., 2005
Heart valves	8-27	907-936	918	Infectious	Reinken et al., 2003

Sample	Published location of primers ^b		Amplicon size ^a (location ^b)	Applications	References
	Forward	Reverse			
Blood, serum, pleural fluid and CSF	8-27	575-556		Unexplained life-threatening infections	Nikkari et al., 2002
	515-533	1390-1371			
Heart valves	8-27	806-787	800	Bacterial endocarditis	Goldenberger et al., 1997
		533-515	500		
Heart valves	91E, 13 BS		492	Bacterial endocarditis	Gauduchon et al., 2003
Synovial fluid, synovial tissue and peripheral blood	8-27	556-575	550	Septic arthritis	van der Heijden et al., 1999
CSF	c	c	241	Meningitis	Saravolatz et al., 2003
Blood	786-805	1184-1203	481	Infective endocarditis	Rothman et al., 2002
Aortic tissue	fd1 mod, 8-27	rP2, 1512-1492	1408	Bartonella infective endocarditis	Jalava et al., 1995
Blood	d	d	365 (1175-1540)	Septicemia	Cursons et al., 1999
Pus from intracerebral	8-27	1510-1492	1500 370	Brain abscesses	Tsai et al., 2004

Sample	Published location of primers ^b		Amplicon size ^a (location ^b)	Applications	References
	Forward	Reverse			
Synovial fluid, synovial tissue and peripheral blood	8-27	556-575	550	Reactive arthritis, osteoarthritis,	Wilbrink et al., 1998a,b
20 mg of the lesion	16SFa and 16SFb	16SR	320	Osteomyelitis, hard tender mass on right thigh	Harris et al., 2002
Sternal and femoral pus	fd1	rP2	1500	Osteomyelitis with hypogammaglobulinemia	La Scola et al., 1997

Aortic, mitral and prosthetic valves, Blood			(10-806)	Infectious endocarditis	Bosshard et al., 2003
Arterial wall	PA	PD		Aortic aneurisms	Marques da Silva et al., 2006
Intracranial and spinal pus or tissue				Epidural abscess, brain abscess or subdural empyema	Kupila et al., 2003
Heart valves	RW01	DG74	370	Infective Endocarditis	Grijalva et al., 2003
CSF and blood			630 (783-1412) 607 (783-1389) 312 (783-1094) 216 (1174-1412)	Meningitis	Xu et al., 2005

IV. Place de la PCR universelle 16S dans l'identification bactérienne

L'identification bactérienne se fait traditionnellement à partir des caractères phénotypiques de la bactérie : coloration (de Gram par exemple), morphologie, aptitude à croître sur certains milieux de culture, caractères biochimiques détectés par diverses techniques commerciales (galeries API, systèmes Vitek, Phoenix, Biolog...).

Toutefois certaines bactéries sont mal identifiées phénotypiquement pour diverses raisons. En effet, certaines bactéries expriment peu de caractères phénotypiques. Pour d'autres espèces, une situation de stress peut avoir altéré les caractères phénotypiques. Une identification basée exclusivement sur des caractères phénotypiques peut donc mener à un résultat erroné. Par ailleurs, les bactéries rares ne sont pas répertoriées dans les bases de données des systèmes commercialisés et en conséquence, les techniques basées sur les caractères phénotypiques ne permettront pas de les identifier. Enfin, pour certaines bactéries à croissance difficile les caractères phénotypiques sont difficiles à déceler ; la biologie moléculaire simplifie dans ce cas l'identification (Petti et al. 2005). Dans ces situations, l'amplification et le séquençage du gène 16S ARNr suivi de la comparaison de la séquence obtenue à des bases de données ont démontré leur efficacité pour l'identification bactérienne. Il s'agit d'une méthode universelle, précise et objective, dont la variabilité inter- opérateur est limitée par rapport aux techniques conventionnelles (Petti,2007).

IV.1. La technique PCR :

La PCR universelle ciblant le gène codant pour l'ARNr 16S à l'aide d'amorces universelles, a d'abord été développée pour des études phylogénétiques. En effet ce gène est universellement retrouvé chez les bactéries et sa fonction est conservée. Ainsi, il peut servir d'«horloge moléculaire» pour mesurer les distances phylogénétiques entre les différentes

espèces bactériennes. La PCR universelle a ensuite été appliquée en microbiologie clinique dans deux domaines distincts : la détection et l'identification bactériennes (Renvoisé A, 2012).

IV.2. L'approche PCR 16S

La PCR est une méthode de biologie moléculaire utilisée pour amplifier des fragments spécifiques d'ADN à travers une série de cycles qui comprennent:

- 1) Dénaturation d'un modèle d'ADN à double brin
- 2) Hybridation des amorces (oligonucléotides courts) qui sont complémentaires au modèle
- 3) Extension des amorces par l'enzyme de polymérase d'ADN, qui synthétise un nouveau brin d'ADN.

Un aperçu schématique de la méthode est montré dans la **Figure 6** :

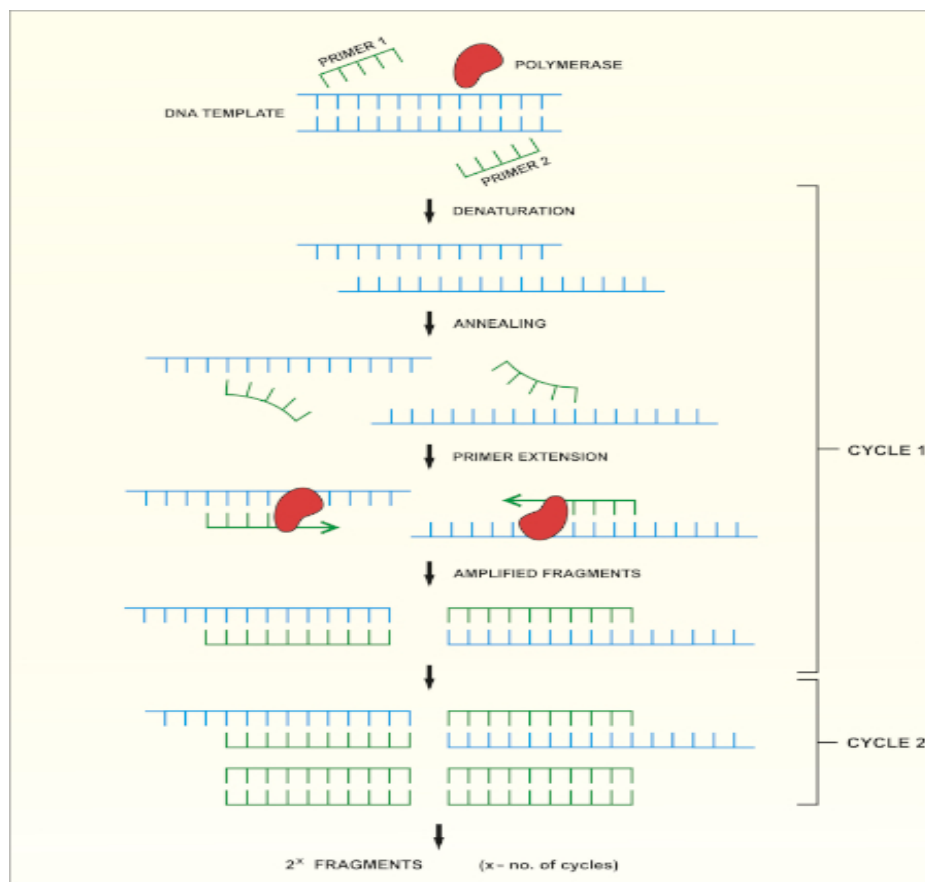


Figure 06 : Aperçu schématique de la réaction de PCR (Renvoisé A, 2012).

Il y a plusieurs facteurs qui sont importants pour une réaction réussie de PCR, dont l'un est la qualité du modèle d'ADN. L'isolement de l'ADN chromosomique des bactéries peut être effectué à l'aide de protocoles standard ou de kits commerciaux. Une attention particulière doit être prise pour obtenir de l'ADN exempt de contaminants qui peuvent inhiber la réaction de PCR.

Les régions conservées du gène 16S ARNr permettent la conception de paires d'apprêt universels (un avant et un revers) qui peuvent se lier à la région cible et l'amplifier dans n'importe quelle espèce bactérienne. La région cible peut varier en taille. Alors que certaines paires d'amorces peuvent amplifier la plupart du gène 16S ARNr, d'autres n'amplifient que des parties de celui-ci. Des exemples d'amorces couramment utilisées sont présentés dans le **Tableau 2** et leurs sites de liaison sont représentés à la **Figure 6**.

Tableau 2 : Exemples d'oligonucléotides standard utilisés dans l'amplification des gènes 16S ARNr (shendure J. et al ,2017).

Nom d'amorce	Séquence (5'-3')	Avant/revers	référence
8F (b)	AGAGTTTGATCCTGGCTCAG	en avant	-1
27F (en)	AGAGTTTGATCMTGGCTCAG	en avant	-10
515F (en)	GTGCCAGCMGCCGCGGTAA	en avant	-11
911R	GCCCCCGTCAATTCMTTTGA	revenir sur	-12
en 1391R	GACGCGGTGTGTRCA	revenir sur	-11
1492R	GGTTACCTTGTTACGACTT GGTTACCTTGTTACGACTT GGTTACCTTGTTACGACTT GGTTA	revenir sur	-11

Les longueurs prévues du produit PCR générées à l'aide des différentes combinaisons d'amorces peuvent être estimées en calculant la distance entre les sites de liaison pour l'avant et l'amorce inverse (**Figure 7**), par exemple la taille du PCR produit utilisant la paire d'apprêt 8F-1492R est de 1500 bp, et pour la paire d'apprêt 27F-911R 900 bp.

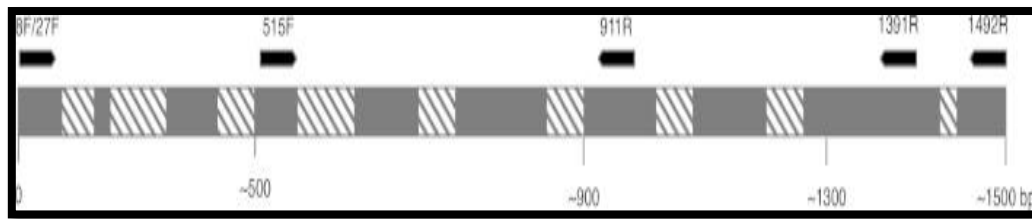


Figure 07: Figure représentative de la séquence de l'ARNr 16S et des sites de liaison d'amorce. (shendure J. et al ,2017)

Les régions conservées sont colorées en gris et les régions variables sont remplies de lignes diagonales. Pour permettre la plus haute résolution, l'amorce 8F et 1492R (nom basé sur l'emplacement sur la séquence d'ARNr) sont utilisées pour amplifier la séquence entière, permettant le séquençage de plusieurs régions variables du gène.

Les conditions de cyclisme pour la PCR (c'est-à-dire la température et le temps requis pour que l'ADN soit dénaturé, annealed avec des amorces et synthétisé) dépendent du type de polymérase qui est utilisé et des propriétés des amorces. Il est recommandé de suivre les directives du fabricant pour une polymérase particulière.

V. Le séquençage 16S

La planète Terre est un habitat pour des millions d'espèces bactériennes, dont chacune a des caractéristiques spécifiques. L'identification des espèces bactériennes est largement utilisée dans l'écologie microbienne pour déterminer la biodiversité des échantillons environnementaux et la microbiologie médicale pour diagnostiquer les patients infectés. Les bactéries peuvent être classées à l'aide de méthodes de microbiologie conventionnelles, telles que la microscopie, la croissance sur des médias spécifiques, les tests biochimiques et sérologiques, et les tests de sensibilité aux antibiotiques.

Au cours des dernières décennies, les méthodes de microbiologie moléculaire ont révolutionné l'identification bactérienne. Une méthode populaire est le séquençage du gène 16S

ribosomal ARN (ARNr). Cette méthode est non seulement plus rapide et plus précise que les méthodes conventionnelles, mais permet également d'identifier les souches qui sont difficiles à cultiver en laboratoire. En outre, la différenciation des souches au niveau moléculaire permet la discrimination entre les bactéries phénotypiquement identiques (Weisburg, 1991).

L'ARNr 16S se joint à un complexe de 19 protéines pour former une sous-unité 30S du ribosome bactérien. Il est encodé par le gène 16S ARNr, qui est présent et fortement conservé dans toutes les bactéries en raison de sa fonction essentielle dans l'assemblage de ribosome ; cependant, il contient également des régions variables qui peuvent servir d'empreintes digitales pour des espèces particulières. Ces caractéristiques ont fait du gène 16S ARNr un fragment génétique idéal à utiliser dans l'identification, la comparaison et la classification phylogénétique des bactéries (Woese, 1987).

Le séquençage du gène ARNr 16S est basé sur la réaction en chaîne de polymérase (PCR) suivie du séquençage de l'ADN. Une fois le programme PCR terminé, les produits sont analysés par électrophorèses de gel d'agarose. Une PCR réussie donne une seule bande de taille prévue. Le produit doit être purifié avant le séquençage pour enlever les amorces résiduelles, les désoxyribonucléotides, la polymérase et le tampon qui étaient présents dans la réaction PCR. Les fragments d'ADN purifiés sont habituellement envoyés pour le séquençage aux services commerciaux de séquençage ; cependant, certaines institutions effectuent le séquençage de l'ADN dans leurs propres installations de base.

La séquence d'ADN est automatiquement générée à partir d'un chromatogramme d'ADN par un ordinateur et doit être soigneusement vérifiée pour la qualité, car l'édition manuelle est parfois nécessaire. Après cette étape, la séquence génétique est comparée aux séquences déposées dans la base de données ARNr 16S. Les régions de similitude sont identifiées, et les séquences les plus similaires sont livrées (Shendure, J et al 2017).

V.1. Les types de séquençage

V.1.1. Méthode Sanger

La diffusion de la méthode de Sanger, la commercialisation d'automates utilisant des fluorophores quatre couleurs ainsi que le déploiement de la PCR dans les laboratoires ont considérablement amélioré les procédures de séquençage. La méthode de Sanger a en effet rapidement dépassé la méthode de Maxam-Gilbert pour la remplacer et reste à ce jour la principale méthode de séquençage utilisée dans les laboratoires. Son principe est le suivant. Dans un premier temps, il est nécessaire d'amplifier l'ADN cible par PCR, puis de le dénaturer afin d'obtenir un ADN simple brin. À l'aide d'une amorce spécifique et complémentaire du brin étudié (sens ou antisens), identique ou différente de celle utilisée pour la PCR, une ADN polymérase effectue alors la synthèse de l'ADN complémentaire à partir de cette amorce. De l'extrémité 5' vers l'extrémité 3', cette enzyme ajoute les désoxyribonucléotides-triphosphates (dNTP) complémentaires et de manière aléatoire et inconstante des didéoxyribonucléotides triphosphates (ddNTP), par exemple un ddGTP sera parfois ajouté à la place d'un dGTP. La réaction se faisant dans un seul tube, les ddNTP (ddATP, ddGTP, ddCTP et ddTTP) sont marqués à l'aide de fluorophores différents pour chaque ddNTP (fluorophores « quatre couleurs »). Lorsqu'un ddNTP est incorporé à la place d'un dNTP, l'ADN polymérase ne peut plus continuer sa polymérisation. La réaction d'extension s'arrête (en effet, le didéoxynucléotide ne possède pas de groupe 3'-hydroxyle indispensable à la réaction de polymérisation de l'enzyme).

Statistiquement, au cours de la réaction, pour chaque « base » de l'ADN cible, au moins une fois, un ddNTP complémentaire sera incorporé à la place d'un dNTP. Par conséquent, à la fin de la réaction, nous obtiendrons des fragments de taille différente. L'analyse de la réaction est ensuite effectuée. Différentes méthodes d'analyse sont possibles. Aujourd'hui, l'électrophorèse capillaire réalisée sur un automate de séquençage est la méthode de choix. Lors de la migration, chaque fragment (contenant un ddNTP marqué par un fluorophore) sera excité

par un laser et le signal obtenu analysé par un logiciel spécifique. L'analyse informatique des signaux permet d'obtenir la séquence étudiée, par exemple, sous forme d'un électrophorégramme, de lecture manuelle aisée mais souvent fastidieuse. Des logiciels d'analyse des séquences peuvent être utilisés. Dans tous les cas, l'analyse d'un fragment d'ADN après PCR se fait toujours à l'aide d'une amorce sens et antisens afin de confirmer la séquence (et une éventuelle anomalie de séquence). En général, cette technique permet d'obtenir des séquences de longueur comprise entre 400 et 850 pb. Comme déjà indiqué, (Hall N, 2007) cette technique, décrite pour la première fois en 1977, reste la plus utilisée dans les laboratoires, notamment en milieu hospitalier.

À titre d'exemple, de nombreux laboratoires hospitaliers utilisent un séquenceur commercialisé par la société Applied Biosystems permettant l'analyse de séquences en plaques de 96 ou 384 puits par électrophorèse capillaire (analyse multicapillaire en parallèle). Ce séquenceur contient un, quatre, huit, 16, 48 ou 96 capillaires selon le modèle. Ainsi, le modèle ABI3130Xl (96 puits, 16 capillaires) permet le séquençage d'environ 400 pb/puits en trois heures (soit 28 8 kb pour la plaque entière). En sachant que cette machine permet de lire environ 18 bases par seconde (pour des 96 capillaires), un an serait nécessaire pour séquencer un génome humain à l'aide de 100 machines utilisées en parallèle, en recouvrant cinq fois le génome (équivalent à cinq séquençages du génome), minimum nécessaire pour s'assurer de l'absence d'erreurs et en supposant que le temps de préparation de ces machines et des échantillons soient négligeable. D'autre développement de la méthode de Sanger sont néanmoins en cours et notamment la miniaturisation de la technique. À titre d'exemple, récemment des auteurs ont réussi à séquencer 600 pb en 6 5 minutes à l'aide d'une puce constituée d'une micro-fluidique permettant une électrophorèse avec un capillaire de 7 5 cm de long constitué d'un polymère spécifique (Fredlake C.P et al, 2008). D'autres technologies ont

donc été développées pour améliorer le rendement, la rapidité et le coût du séquençage (Sanger F., Nicklen S, 1977).

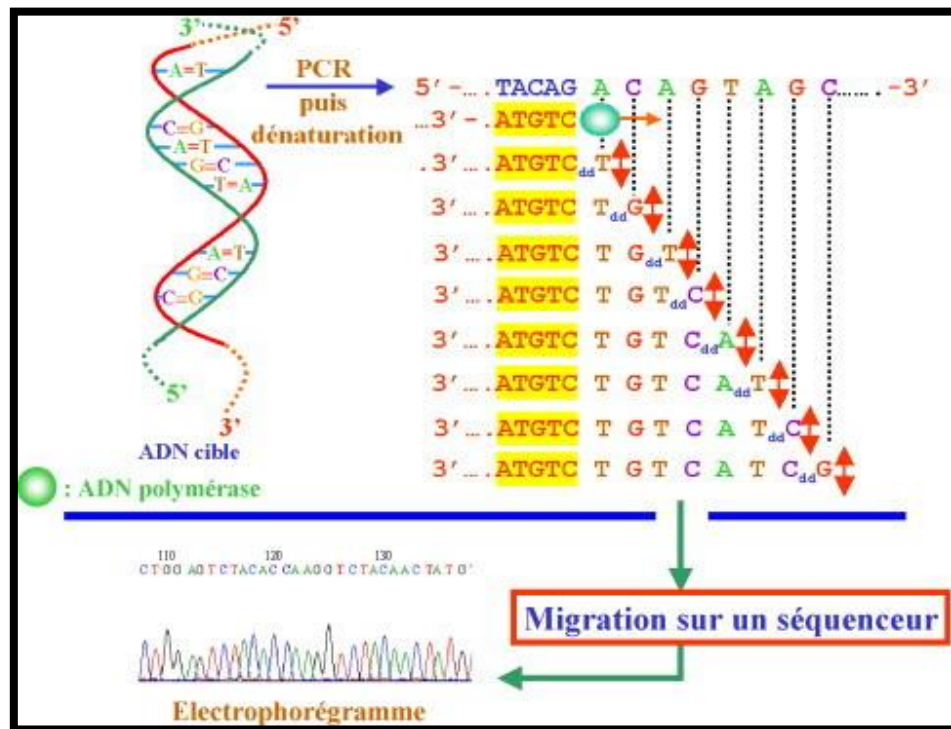


Figure 08 : Principes du séquençage selon la méthode de Sanger.(Sanger F,1977)

V.1.2. Méthode de Maxam et Gilbert

Cette technique est pratiquement abandonnée de nos jours. Nous la décrivons brièvement pour des raisons historiques. Cette méthode, publiée parallèlement à celle de Sanger en 1977, par son caractère révolutionnaire a grandement contribué à l'histoire de la biologie moléculaire. Il s'agit d'une méthode chimique de séquençage. Les réactifs clivent spécifiquement après chacune des bases A, C, G, [A + G], [C + T]. Cette technique est basée sur la propriété de certains agents chimiques, l'hydrazine, le diméthyl sulfate (DMS) et l'acide formique, de modifier les bases de l'ADN. Dans un second temps, la pipéridine est ajoutée et « casse » les brins d'ADN au niveau des bases modifiées. Les agents chimiques sont utilisés dans des conditions telles qu'ils n'agissent qu'avec un faible pourcentage des bases de l'ADN étudié. Le DMS agit au niveau des bases « G ». L'acide formique agit au niveau des bases « A + G ».

L'hydrazine agit au niveau des bases « C + T » (en milieu alcalin, l'hydrazine agit uniquement au niveau des « C »). L'ADN à séquencer est marqué à une extrémité. Le plus souvent, il s'agit d'un marqueur radioactif. Le produit de séquence est déposé sur un gel d'acrylamide, puis la séquence lue après autoradiographie (Figure 9). L'ADN étudié peut être simple ou double brin. Cette technique permettait d'analyser des fragments allant jusqu'à 500 pb. (Gilbert W., Maxam A, 1973)

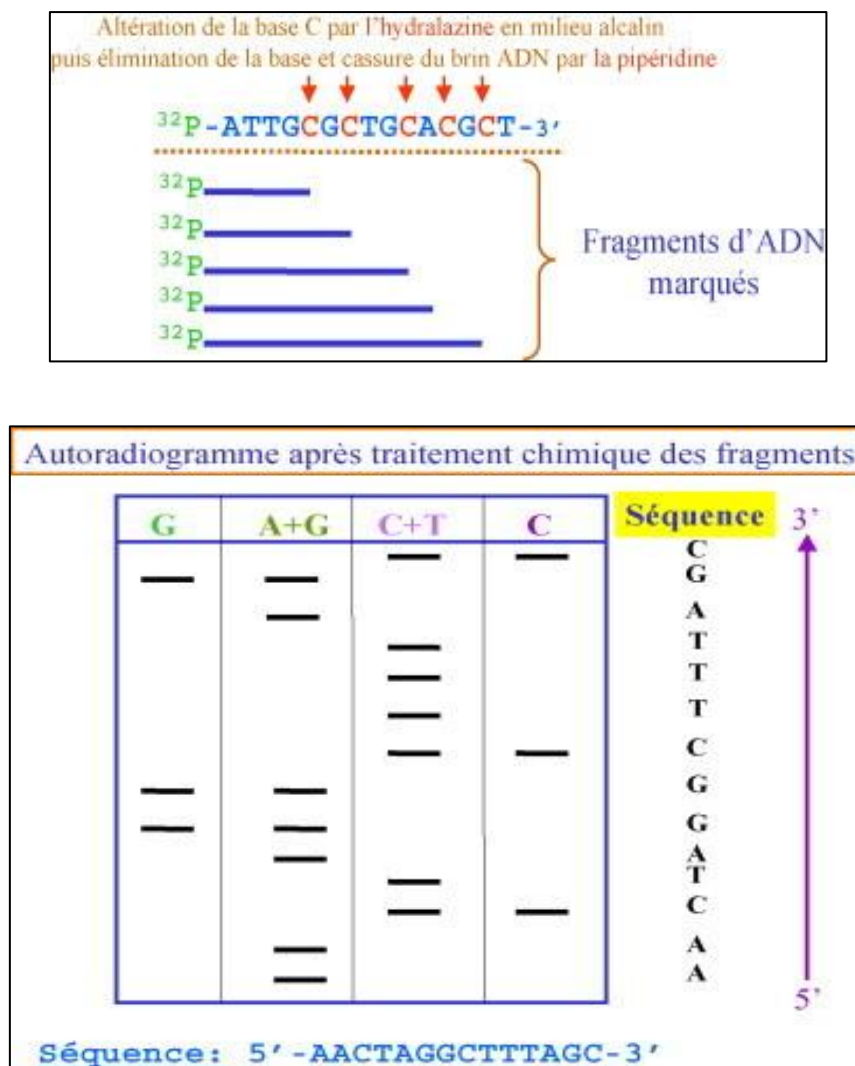


Figure 09 : Technique de Maxam-Gilbert. (Gilbert W., Maxam A, 1973)

V.1,3. La technique shotgun (séquençage aléatoire globale)

Elle a été utilisée massivement pour le séquençage d'un grand nombre de génomes notamment bactériens (séquençage de novo). Schématiquement, cette méthode consiste à fragmenter le génome entier à étudier en petits fragments d'ADN à l'aide de moyens mécaniques. Sur chaque fragment, une réaction de ligation permet de fixer de courtes séquences d'ADN appelées adaptateurs, ces derniers servant d'amorce pour la PCR. Ces fragments sont ensuite intégrés dans des plasmides et constituent une bibliothèque (library) de fragments aléatoires d'ADN simple brin. Ils sont ensuite amplifiés par PCR, par exemple, puis séquencés à l'aide de la méthode de Sanger.

À l'aide de logiciels informatiques, les séquences sont ensuite alignées et recoupées par chevauchement. Ces séquences représentent plusieurs blocs de séquences continus : on parle alors de « contig ». Certaines séquences manquent : ce sont des trous de séquences (gap). Ceux-ci sont comblés par séquençage à partir des séquences déjà déterminées. À partir de ces dernières, sont dessinées des amorces de PCR servant pour l'amplification en direction de ces trous. Cette technique nécessite de séquencer le génome cible plusieurs fois (minimum cinq fois), tant pour éviter les erreurs de séquençage que pour s'assurer du maximum de chevauchement entre les séquences et faciliter l'ordonnancement des séquences. Cette technique est essentiellement utilisée pour des génomes de petite taille comme ceux des bactéries.

V.2. Le principe de séquençage de grands génomes :

L'automatisation à l'aide de robots a permis de séquencer de nombreux génomes dont le génome humain. Le travail a été considérable et n'a pu être effectuée que grâce à une robotisation et une informatisation poussée dans des instituts dédiés à ces séquençages à grande échelle. À ce jour, plus de 280 génomes procaryotes ont ainsi été totalement séquencés. Pour

réaliser le séquençage d'un grand génome (eucaryote ou procaryote, par exemple), une carte physique du génome est d'abord constituée. Il s'agit d'établir des repères sur le génome à l'aide de marqueurs spécifiques de chaque chromosome. Schématiquement, deux types de marqueurs sont utilisés :

- **les polymorphismes** (avec étude de la liaison génétique entre ceux-ci). Ce type de marqueurs a été étudié soit à l'aide d'enzymes de restriction (on parle alors de restricted fragment length polymorphisms [RFLP]), soit à l'aide de microsatellites (séquences répétées dont le nombre de répétition est le plus souvent variable [polymorphe] au même locus) ;
- **les séquences uniques polymorphes** ou non appelées aussi sequence tagged sites (STS).

La carte physique permet de limiter la zone à étudier et ainsi de faciliter la reconstitution du génome à séquencer. Plusieurs méthodes existent. Nous en décrivons une de manière simplifiée. Pour permettre le séquençage du génome, il est nécessaire de le fragmenter. Pour cela, l'ADN est cassé mécaniquement. Les nombreux fragments obtenus sont ensuite insérés au hasard dans des plasmides (ces fragments sont alors appelés inserts). Ils peuvent être aussi insérés dans des cosmides ou des fosmides (vecteurs particuliers dérivés de plasmides, utilisés pour des inserts de plus grande taille). (Ameziane N, et al 2005). L'ensemble est introduit dans des bactéries (*Escherichia coli*) par transformation. Les fragments insérés ont une taille d'environ 2 à 4 kb. On obtient alors des clones dont on peut étudier les marqueurs et les STS. Pour cela, il est nécessaire de sélectionner les bactéries ayant incorporé le plasmide/insert (la transformation ne présente pas un rendement de 100 %). Des marqueurs de sélection sont utilisés (par exemple, résistance à des antibiotiques et système de couleurs des colonies s'assurant du succès de la transformation). Étant donné le nombre important de bactéries à sélectionner, plusieurs robots sont utilisés. Ils reconnaissent et sélectionnent les bactéries ayant incorporé le plasmide avec l'insert. La connaissance de la carte génétique et des STS facilite ensuite la sélection des clones à séquencer de manière ordonnée.

D'autres robots permettent de récupérer l'ensemble plasmide/insert après culture des bactéries et d'extraire l'ADN de l'ensemble plasmide/insert. Cet ADN est alors amplifié non par PCR mais par rolling-circle amplification (RCA). L'ADN est ensuite séquencé à l'aide d'hexamères aléatoires par la technique de Sanger. Cette réaction est également réalisée à l'aide de robots. Les séquences sont ensuite lues par électrophorèse capillaire à l'aide d'automates de séquençage de 384 capillaires. De nombreux fragments sont obtenus et l'analyse informatique regroupe les séquences chevauchantes pour obtenir une séquence continue. Cette séquence, une fois validée, sera la séquence de référence pour d'autres études sur le génome cible. Dans le cas du génome humain, plutôt que des plasmides, ce sont des chromosomes bactériens artificiels appelés bacterial artificial chromosomes (BAC, pouvant contenir des inserts d'une taille allant jusqu'à 300 kb) et parfois des phage artificial chromosomes (PAC, de même capacité) qui ont été utilisés. Ces clones ont ensuite été ordonnés (positionnés les uns par rapport aux autres et le long des chromosomes humains).

Pour cela, l'utilisation des profils de restriction, la connaissance de la carte génétique et des STS ainsi que l'étude par hybridation des clones entre eux ont été nécessaires. Le séquençage a alors été réalisé sur les clones chevauchant pour obtenir la séquence humaine de référence. Depuis le séquençage initial du génome à l'aide de la technique shotgun précédemment décrite, d'autres approches similaires ont été développées pour permettre un séquençage plus rapide. Par exemple et schématiquement, après fragmentation de l'ADN et addition d'adaptateurs aux extrémités par ligation, les différents fragments ainsi préparés sont fusionnés ensemble au hasard et forment ainsi une bibliothèque (library) de fragments aléatoires simple brin. Chaque fragment est ensuite immobilisé sur une bille, chaque bille ne contenant qu'un seul fragment aléatoire. Une PCR émulsion est réalisée (cette technique est décrite dans la partie II de l'article) (Dressman D et al ,2003) , (Holmberg K,2003), (Hori M,2003). Chaque amplification est clonale et permet d'obtenir des millions de copies d'un même ADN par puits.

Un automate de séquence peut alors démarrer le séquençage dans chaque puits (par pyroséquençage, par exemple) et la séquence sera déduite après analyse informatique. Cette dernière méthode est plus rapide et moins chère que la méthode utilisant le séquençage par la méthode de Sanger. (Ziebolz B., Droege M,2007)

VI. L'identification bactérienne par séquençage du gène ARN 16S)

Le séquençage du gène codant pour l'ARN ribosomal 16S (ARNr 16S) permet l'étude de la phylogénie et de la taxonomie bactériennes. Ce gène de 1500 pb environ est présent dans la plupart des bactéries (sous forme d'opéron ou de famille multigènes) et sa fonction n'a pas changé au cours de l'évolution. Par ailleurs, sa taille est suffisante pour une analyse informatique puissante. Au 9 juillet 2008, 9263 espèces de bactéries ont pu être classées grâce au séquençage de l'ARNr 16S.

Le séquençage du gène de l'ARNr 16S par sa simplicité a rendu la classification des bactéries plus aisée même si la référence absolue (gold standard) pour l'identification de nouvelles espèces et leur classification taxonomique reste l'hybridation ADN/ADN de réalisation plus lourde et plus difficile. C'est en 1994 que des chercheurs ont proposé et établi l'utilisation potentielle du séquençage de l'ARNr 16S pour la définition d'espèces en microbiologie (Goebel B.M. 1994).

En pratique, le séquençage de l'ARNr 16S permet l'identification du genre et de l'espèce d'un isolat. Cette méthode est intéressante notamment lorsque les classiques méthodes biochimiques ne permettent pas d'identifier clairement une bactérie (par méthode commerciale ou non). Le séquençage de l'ARNr 16S permet l'identification du genre bactérien dans plus de 90 % des cas, de l'espèce bactérienne dans 65–85 %. Selon les études, 1 à 14 % des isolats étudiés demeurent non identifiées après utilisation de cette méthode. (Janda JM, Abbott SL J 2007)

Dans tous les cas, l'identification par séquençage de l'ARNr 16S reste supérieure aux méthodes bactériologiques conventionnelles. Ce séquençage présente néanmoins des limites. À titre d'exemple, dans le genre *Bacillus*, deux souches — *Bacillus globisporus* et *Bacillus psychrophilus* — partagent plus de 95 % de séquences identiques sur leurs gènes ARNr 16S alors que la technique de référence d'hybridation ADN/ADN montre 23–50 % d'identité.

Certaines études ont établi une liste de genres bactériens posant des difficultés d'identification à l'aide de cette technique (à titre d'exemple, le genre *Campylobacter* pour les espèces « non » *Jejuni coli*, les mycobactéries à croissance rapide, le genre *Actinomyces*). Malgré ces inconvénients, le séquençage de l'ARNr 16S joue un rôle important dans l'identification bactérienne, que ce soit pour l'identification de bactéries inconnues ou pour celles au profil biochimique ambigu. Pour aider à mieux interpréter le séquençage de l'ARNr 16S, des recommandations ont été publiées (**Tableau 3**). Bien évidemment, les autres techniques de séquençage, par exemple, le pyroséquençage sont également utilisées pour réaliser de telles études.

Tableau 3 : Recommandations pour l'identification bactérienne par séquençage de l'ARNr 16S.

Catégorie	Recommandations
Souche à séquencer	En cas de difficulté d'identification phénotypique ou de difficulté identifiée et connue pour l'analyse par le gène ARNr 16S, il est nécessaire de séquencer en parallèle un autre gène commun (par exemple, le gène rpoB).
Séquençage de l'ARNr 16S	Séquencer au minimum 500–525 pb Pour obtenir moins de 1 % d'ambiguïté, séquencer 1300–1500 pb
Critères d'identification d'espèce	Au moins > 99 % d'identité de séquence (idéalement > 99,5 %) par rapport à la souche référence dont les caractères ont été définis par les études d'hybridation ADN/ADN En cas de différence < 0,5 % par rapport à l'espèce la plus proche, il est nécessaire d'évaluer les autres propriétés de la souche, notamment au niveau phénotypique pour permettre l'identification définitive.

VII. L'analyse bio-informatique

VII.1 Alignement de séquence :

L'alignement séquentiel est une manière de représenter deux ou plusieurs séquences de macromolécules biologiques (ADN, ARN ou protéines) les unes sous les autres, de manière à en faire ressortir les régions homologues ou similaires. L'objectif de l'alignement est de disposer les composants (nucléotides ou acides aminés) pour identifier les zones de concordance. Ces alignements sont réalisés par des programmes informatiques dont l'objectif est de maximiser le nombre de coïncidences entre nucléotides ou acides aminés dans les différentes séquences. Ceci

nécessite en général l'introduction de « trous » à certaines positions dans les séquences, de manière à aligner les caractères communs sur des colonnes successives. Ces trous correspondent à des insertions ou des délétions (appelés indel) de nucléotides ou d'acides aminés dans les séquences biologiques. Le résultat final est traditionnellement représenté comme des lignes d'une matrice.

En bio-informatique, l'opération d'alignement vise à identifier des zones communes a un groupe de k séquences. Des zones qui se ressemblent sont dites similaires ou homologues si elles dérivent d'un ancêtre commun (Iwata et al., Nucl.Acids Res. 2012)

VII.1.1. Alignement local ou global

Global : on tente d'identifier des similarités sur la longueur totale des séquences (→ pb si séquences de longueur différentes)

Local : on tente d'identifier des similarités entre une séquence et une sous-séquence :

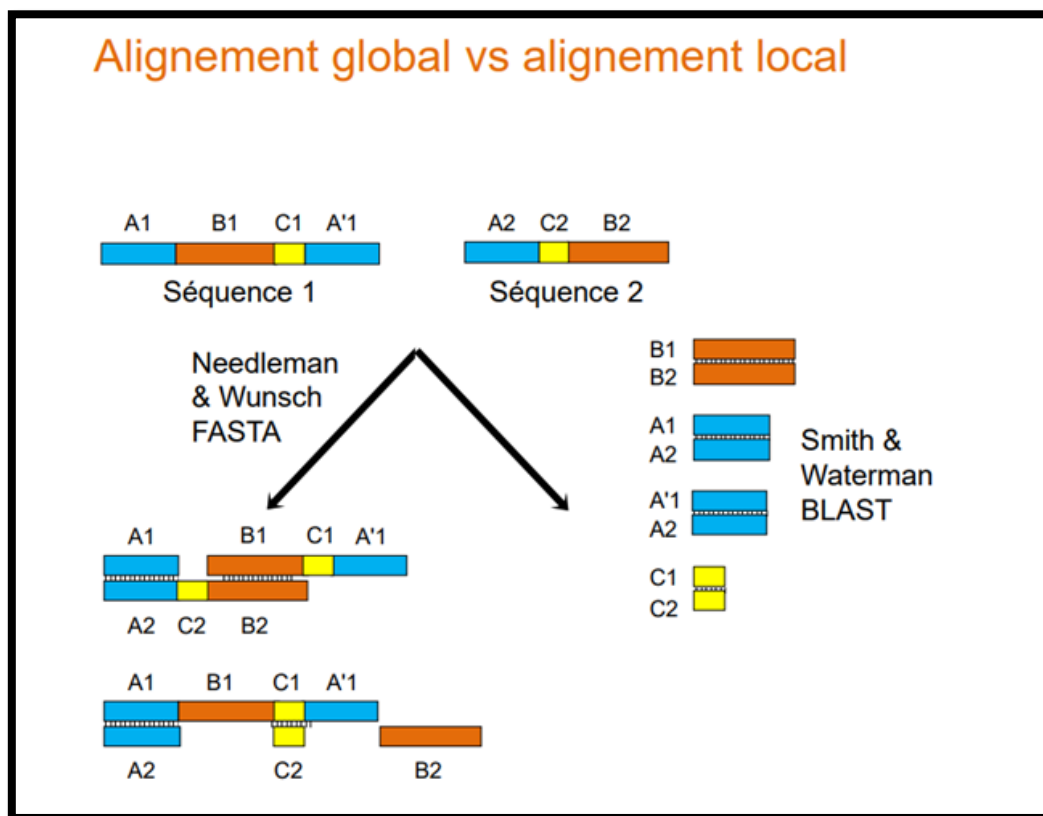


Figure 10 : Types d'alignements (Iwata et al., Nucl.Acids Res. (2012))

VII.1,2. Alignement par paire ou multiple

- par paires : on aligne 2 séquences
- multiple : on aligne plus de 2 séquences

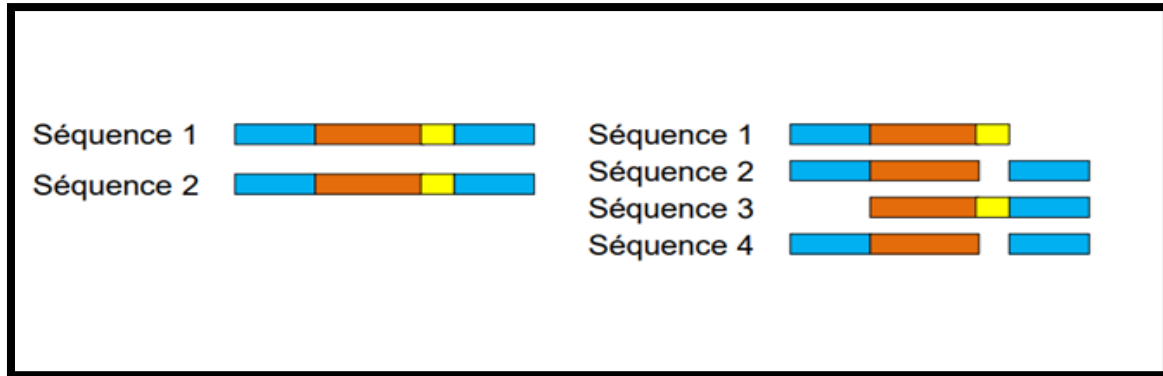


Figure 11 : alignement de deux séquences et alignement multiple (Iwata et al., Nucl. Acids Res. 2012)

VII.2. Etude phylogénétique :

La phylogénie moléculaire est basée sur l'analyse comparative des séquences d'ADN ou de protéines. Elle complète les méthodes traditionnelles de classification basées sur l'observation des caractères morphologiques et anatomiques.

La phylogénie moléculaire utilise les gènes des organismes vivants pour élaborer des arbres phylogénétiques en comparant les caractères moléculaires (séquences nucléotidiques ou protéiques). La phylogénie moléculaire essaie de retracer l'accumulation des mutations dans les génomes au cours de l'évolution des espèces. Elle permet de reconstruire l'histoire évolutive des organismes. Les espèces ont des génomes d'autant plus proches qu'elles ont divergé récemment depuis leur ancêtre commun.

La phylogénie moléculaire a pour but de reconstruire les relations de parenté entre des séquences de nucléotides ou d'acides aminés. On peut ainsi étudier les relations de parenté entre les espèces qui les portent mais, aussi, l'évolution du génome. En particulier, pour chaque

famille multigénique, on peut déterminer l'importance relative des événements de duplications et de transferts horizontaux de gènes.

La fiabilité des méthodes de reconstruction phylogénétique repose sur la compréhension des mécanismes d'évolution des séquences, un domaine qui a beaucoup progressé ces dernières années. Cela a abouti à une vision sans cesse plus correcte de l'arbre universel du vivant. L'étude des contraintes fonctionnelles agissant sur les protéines bénéficie de ces avancées. En particulier, la détection, dans une protéine, des positions qui sont soumises à une sélection darwinienne est devenue assez performante, permettant de prédire les substitutions à l'origine d'un changement de fonction et donc de guider les études expérimentales. (Castresana J. 2000).

VII.2,2. Arbre phylogénétique :

Un arbre phylogénétique est une représentation schématique et buissonnante, permettant de mettre en avant une parenté entre espèces ou groupes d'espèces. Les êtres les plus proches ne divergent qu'au fil des dernières branches de l'arbre, tandis qu'une ancestralité commune éloignée est marquée par un positionnement sur des branches très divergentes. Il s'établit normalement à l'aide d'une matrice taxons-caractères et s'accompagne souvent de chacune des innovations apparues successivement, permettant de distinguer les espèces entre elles (**Figure 12**) (Chaw SM et al .2000) .

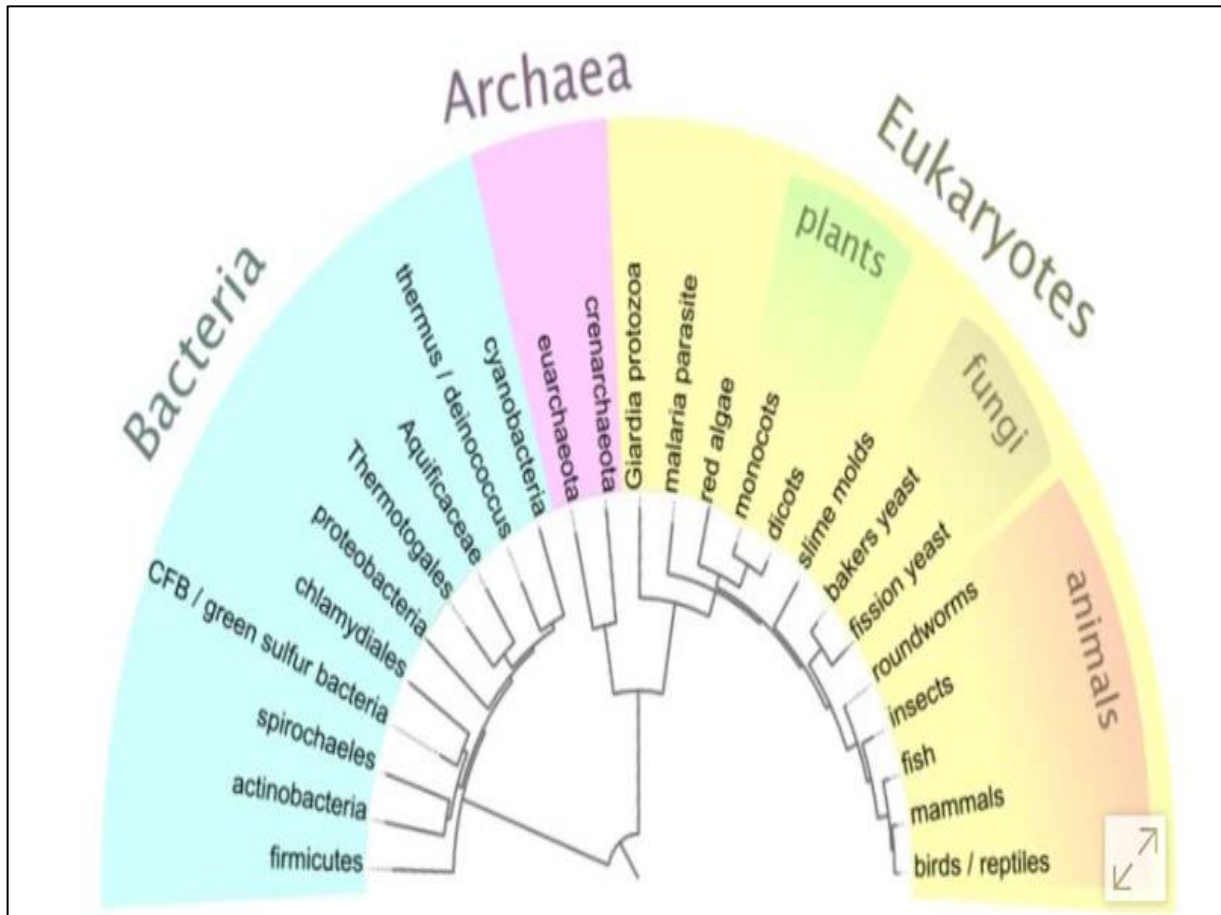


Figure 12 : Exemple de représentation de la parenté entre les grands groupes d'êtres vivants, à l'aide de cet arbre phylogénétique simplifié et buissonnant. (Chaw SM et al .2000)

OBJECTIFS DU MEMOIRE

Les α -amylases agissent à des sites aléatoires tout au long de la chaîne d'amidon, libérant le maltotriose et le maltose à partir de l'amylose, ou maltose, glucose et limit-dextrines à partir de l'amylopectine (Singh *et al.* 2011).

La source industrielle la plus utilisée pour la production des α -amylases parmi les genres bactériens est sans conteste le genre *Bacillus* qui sont des bactéries connues pour être de bonnes productrices d'alpha-amylases thermostables (Sundarram et Murthy 2014).

L'ARN ribosomal (ARNr) 16S est le constituant ARN de la petite sous-unité ribosomale du 30S des procaryotes. Le gène codant pour cet ARNr est le «le gène16S ARNr présent dans l'ensemble des espèces bactériennes en un nombre variable de copies (Woese, 1987 ;Petti, 2007).

Ce mémoire a eu pour objectif de réaliser la dernière étape de l'identification moléculaire d'une bactérie du genre *Bacillus* ayant pour rôle la production d'enzymes de type α -amylases par l'étude du gène codant l'ARNr 16S. Ces souches bactériennes du genre *Bacillus* ont été isolées et étudiées préalablement dans la thèse de Mr. CHOUBANE Slimane (Maitre de conférences à l'école supérieure de biologie d'Oran ESSBO).

Nous avons eu l'opportunité de réaliser l'analyse bio-informatique du séquençage qui a été réalisé par Dr. CHOUBANE concernant la séquence du gène ARNr 16S. Nous avons d'abord, réalisé la lecture et conversion de la séquence du gène de l'ARNr 16S de la souche d'intérêt *Bacillus* en utilisant l'outil Mega X. Ensuite, nous avons pu faire un alignement de séquences de gènes de l'ARNr 16S dans la base de données BLAST-NCBI. Ainsi, une comparaison de notre séquence a été réalisée avec toutes les séquences du même gène enregistrées dans BLAST-*nucleotide*. Finalement, l'arbre phylogénétique de la souche d'intérêt *Bacillus* a été exploré et analysé par le logiciel Mega X.

MATERIEL ET METHODES

Nous avons réalisé dans ce mémoire la dernière et importante partie de l'identification moléculaire de la bactérie *Bacillus* par une approche bio-informatique. Ces souches bactériennes *Bacillus* ont été isolées et étudiées préalablement dans la thèse de Mr. CHOUBANE Slimane (Maitre de conférences à l'école supérieure de biologie d'Oran ESSBO).

I. Isolement des colonies et extraction d'ADN :

Des échantillons de sol ont été prélevés par Dr CHOUBANE.S dans différentes zones, principalement de l'Ouest Algérien pour obtenir des souches *Bacillus spp.*, qui possèdent une activité amylolytique.

La caractérisation morphologique, biochimique et physiologique d'un isolat spécifique a été sélectionnée par Dr. CHOUBANE Slimane comme organisme performant hyper-producteur d'amylase avec des caractéristiques idéales pour la suite du travail (méthode moléculaire). Cette sélection a eu lieu en se basant sur l'activité amylolytique et autres activités enzymatiques des souches bactériennes.

Des jeunes colonies ont été inoculées dans 5 ml de bouillon LB et incubées au Laboratoire de Génétique Moléculaire et Cellulaire (LGMC, USTO-MB).

La culture a été ensuite transférée dans un microtube Eppendorf et centrifugée. L'ADN chromosomique a été alors extrait en utilisant le kit GenElute Bacterial Genomic DNA kit (Sigma, USA) (Choubane S, 2017).

II. Amplification du gène ribosomique ARNr 16S :

Le gène ribosomique ARNr 16S a été amplifié par réaction en chaîne par polymérase (PCR) en utilisant des amorces standards pour amplifier la totale longueur du gène (1500 pb).

L'amplification du gène a été réalisée avec le kit Q5® High-fidelity DNA polymerase kit (New England, USA) (Choubane S, 2017).

III. Séquençage du gène ARNr 16S :

Le produit de la PCR purifié et analysé par électrophorèse sur gel d'agarose à 1% a été soumis au séquençage en utilisant la méthode de Sanger (Sanger *et al.* 1977).

Le séquençage a été réalisé par Dr. CHOUBANE Slimane à Genevision TA Geneius (Northumberland, Royaume-Unis).

IV. Analyse phylogénique du gène ARNr 16S :

Nous avons eu l'opportunité de réaliser l'analyse bio-informatique du séquençage qui a été réalisé par Dr. CHOUBANE Slimane concernant la séquence du gène ARNr 16S. Ce travail a fait l'objet de notre mémoire.

La séquence du gène ARNr 16S reçue a été comparée à la base de données des nucléotides BLASTn de NCBI. Selon les résultats obtenus et le pourcentage d'homologie avec les séquences présentes déjà sur la base de données, la séquence a été enregistrée dans la base de données Genbank et un numéro d'accession a été attribué à Dr Choubane Slimane (comme auteur de la souche).

Dans ce travail de mémoire, la séquence du gène ARNr 16S a été d'abord alignée en utilisant l'algorithme BLAST et nous avons éliminé les séquences de basse qualité.

Par la suite, les séquences ont été alignées en utilisant l'outil d'alignements multiples ClustalW, avec d'autres séquences de gènes ARNr 16S d'autres souches de différentes espèces du genre *Bacillus*. Nous avons sélectionné 16 souches, parmi lesquelles des souches pathogènes et productrices d'enzymes.

Par ailleurs, nous avons construit un arbre phylogénétique avec toutes ces séquences ainsi que les séquences d'autres espèces en utilisant le programme MEGA X version 10,2,4.

La base de données BLAST :

C'est l'abréviation de « Basic Local Alignment Search Tool » ou, en français, L'outil de recherche basique d'alignement local. Il se trouve sur le site de NCBI (Figure).

Le BLAST est une méthode de recherche heuristique utilisée en bio-informatique. Il permet de trouver les régions similaires entre deux ou plusieurs séquences de nucléotides ou d'acides aminés, et de réaliser un alignement de ces régions homologues en comparant avec toutes les séquences enregistrées.

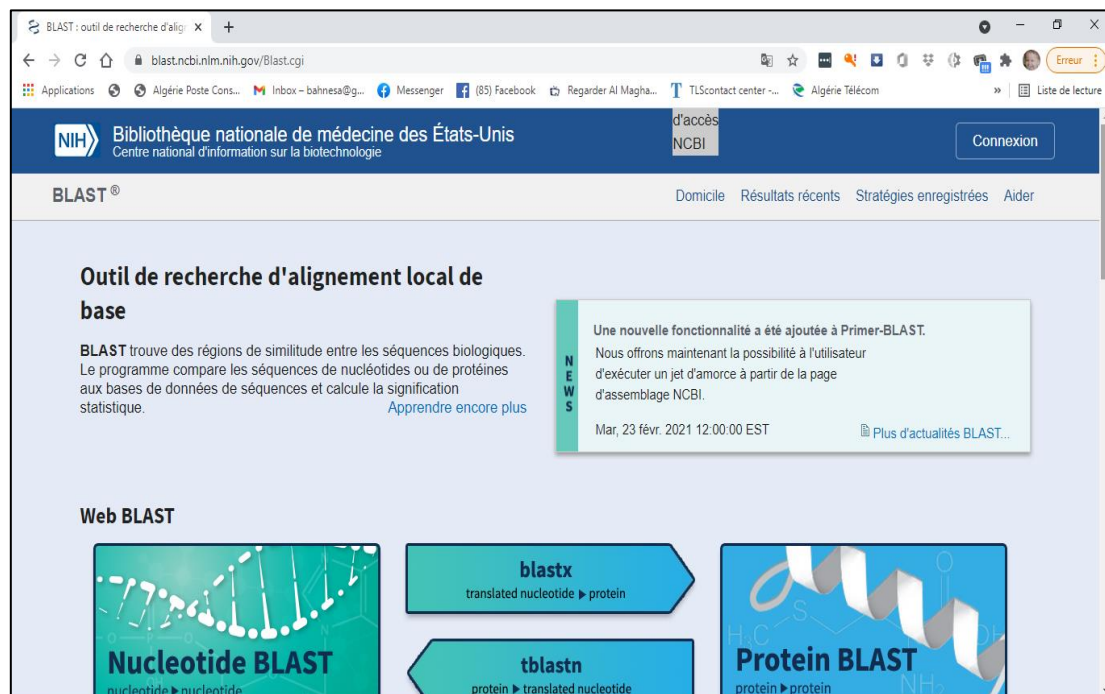


Figure 13 : Photo de l'interface de la base de données BLAST sur NCBI.

Les étapes de l'alignement par BLAST :

- Ouvrir la séquence en format AB1 (obtenue par séquençage du gène ARNr 16S) avec *Mega X* et éliminer les piques de basse qualité
- Copier la séquence nucléotidique du chromatogramme et la mettre en format texte
- Accéder à l'outil d'alignement en ligne BLAST et cliquer sur BLAST nucléotide
- Introduire la séquence du gène sous format FASTA et cliquer sur BLAST
- Dans le tableau affiché relever les scores, la couverture et le degré de pourcentage

Le logiciel MEGA X :

Le logiciel informatique Mega X « Molecular Evolutionary Genetics Analysis » permet d'effectuer des analyses statistiques de l'évolution moléculaire et de construire des arbres phylogénétiques. Il comprend de nombreuses méthodes et outils sophistiqués pour la phylogénomique et la phylomédecine.

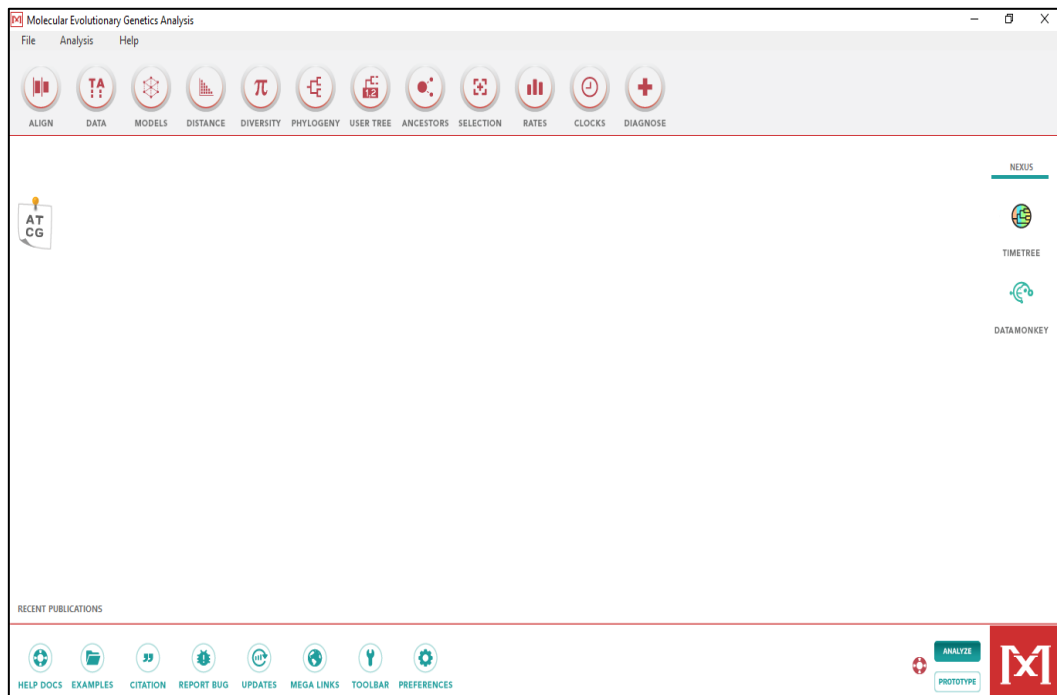


Figure 14 : Photo représentant l'interface du logiciel Mega X version 10,2,6.

Les étapes de la construction d'un arbre phylogénétique par Mega X sont comme suite :

- Mettre la séquence du gène avec d'autres séquences ARNr 16S (téléchargés de Genbank) dans un fichier texte et convertir au format FASTA avec BIOEDIT.
- Ouvrir le fichier dans MEGA X puis sélectionner les séquences et lancer l'alignement dans Align ; ClustalW
- Exporter la séquence de l'alignement sous format MEGA
- Ouvrir MEGA X et sélectionner « phylogeny » ; construct *Neighbour joining Tree* et sélectionner le fichier d'alignement puis générer l'arbre

Résultats et discussions

Dans ce mémoire, nous avons procédé à une approche bio-informatique afin d'analyser les résultats du séquençage du gène ARNr 16S pour une meilleure identification moléculaire de la bactérie d'intérêt.

Tout d'abord, les résultats du séquençage nous ont permis d'accéder à un fichier en format ABI qui contient la succession nucléotidique du gène ARNr 16S. Nous avons utilisé la plateforme BLAST et le logiciel MEGA X version 10,2,4 pour la lecture et l'alignement de la séquence de la souche *Bacillus*.

I. Lecture et alignement de la séquence du gène ARNr 16S :

Premièrement, nous avons éliminé les parties de séquences de basse qualité. Ceci a été fait là où les pics des quatre nucléotides n'étaient pas pointus. Ensuite, nous avons réussi à convertir la séquence du gène ARNr 16S en format FASTA afin de produire l'alignement par BLASTn et la phylogénie par MEGA X.

La séquence du gène ARNr 16S en Format FASTA générée par BLAST nucléotide :

```
TATGCTGCACATGCAGTCGTAACGGGGTGAATCGTGAGCTTGCTCCCTGATGTTAGCGGCGGACGGGTGAGTAACACGTGGG
TAACCTGCCTGTAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATGGTTGTTTGAACCGCATGGTTCAAACATA
AAAGGTGGCTTCGGCTACCACTTACAGATGGACCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGA
TGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGA
ATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTAG
GGAAGAACAAGTACCGTTTCAATAGGGCGGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGCAGGCGGTTTCTTAAGTCTGATGTGAA
AGCCCCCGGCTCAACCGGGAGGGTCATTGAAAAGTGGGGAACCTGAGTGCAAGAGGAGAGTGGAATTCCACGTGTAGC
GGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTCTGGTCTGTAAGTACGCTGAGGAGCGAAAGC
GTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGGGGGTTCCGCCCCCTT
AGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTGCGAAGACTGAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCC
```

L'alignement de la séquence à identifier par BLAST nous a révélé que la bactérie s'aligne parfaitement avec la souche *Bacillus Mojavensis* avec un pourcentage d'identification de 100%. En effet, la recherche d'homologie pour la séquence nucléotidique de l'ARNr 16S en utilisant l'outil BLASTn a révélé 99% de ressemblance avec la souche *Bacillus subtilis* PVR05. Cette souche est une bactérie endophyte résistante à l'arsenic (Zhu *et al.* 2014).

Nos résultats sont concordants avec ceux rapportés par Mr CHOUBANE Slimane dans sa thèse de doctorat (Choubane S, 2017). Cette souche a été nommée *Bacillus Mojavensis*.

II. Construction de l'arbre phylogénétique de la souche *Bacillus* :

Nous avons utilisé le logiciel Mega X afin de constituer l'arbre phylogénétique de notre souche bactérienne. En effet, nous avons comparé notre séquence du gène ARNr 16S avec celles de différentes souches bactériennes (pathologiques, productrices d'enzyme...). La constitution de l'arbre phylogénétique de notre souche nous a permis de la classer et la positionner par rapport à un groupe de 16 séquences de différentes souches connues.

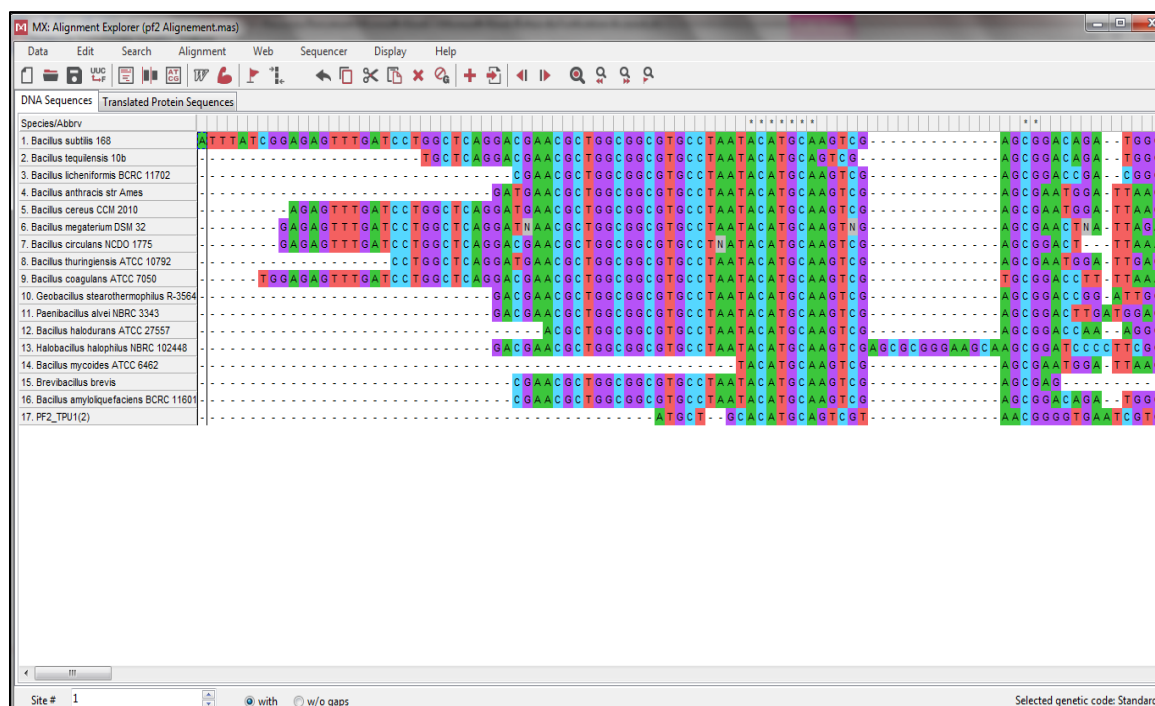


Figure 15 : Photo qui représente l'alignement de séquences du gène ARNr 16S de *Bacillus Mojavensis* (PF2_TPU1) avec d'autres séquences d'espèces connues par le logiciel Mega X.

Trois espèces du genre *Bacillus* sont spécialement connues pour leur pathogénicité :

***Bacillus cereus*, un agent d'intoxication alimentaire** qui cause deux types d'intoxications alimentaires (les syndromes émétique et diarrhéique), et une variété d'infections locales et systémiques telles que : l'endophtalmie, l'endocardite, la méningite, la parodontite, l'ostéomyélite, les infections de plaies et la septicémie (Drobniewski 1993).

Le groupe *Bacillus cereus* est composé de six espèces bactériennes dont les trois plus connues en raison de leur pathogénicité sont *Bacillus anthracis*, *Bacillus thuringiensis* et *Bacillus mycoides*.

Les espèces *Bacillus* autres que les membres du groupe *B. cereus* sont une cause rare de maladies d'origine alimentaire. Dans de cas pareil, l'aliment contiendrait un grand nombre (entre 10^5 et 10^9 UFC/g) de la bactérie *Bacillus sp* suspectée (Kramer et Gilbert 1989).

L'analyse phylogénétique (**Figure16**) a montré que notre souche est proche des espèces *Bacillus subtilis*, *Bacillus tequilensis*, *Bacillus amyloliquefaciens* et *Bacillus licheniformis*, ce qui confirme les résultats de l'identification conventionnelle. Les valeurs Bootstrap (%) pour 500 répétitions sont affichées au niveau des noeuds.

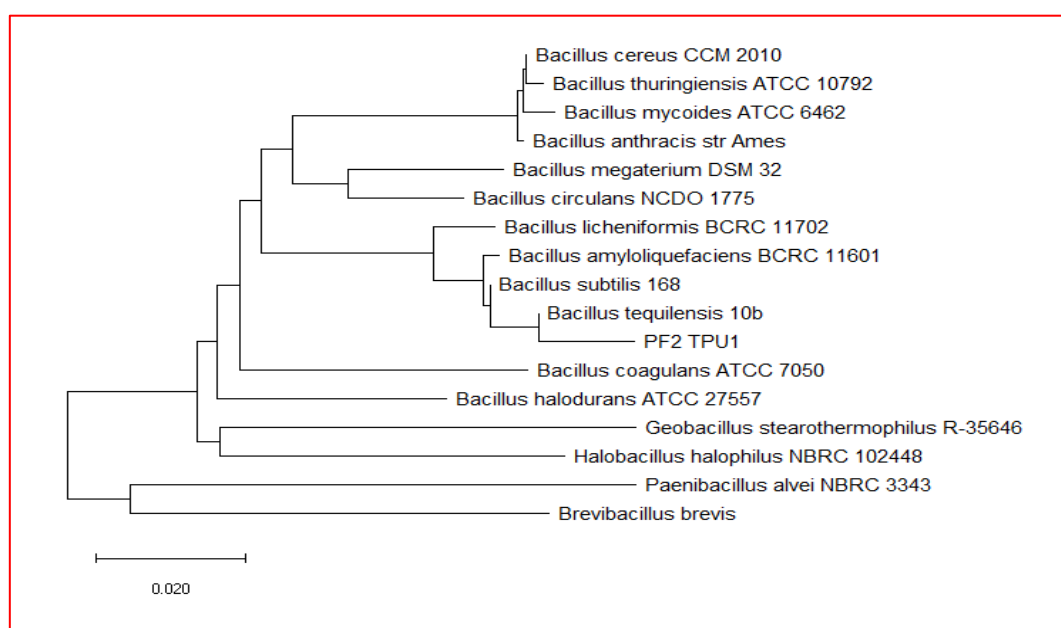


Figure 16 : Arbre phylogénétique de la souche bactérienne étudiée dans notre mémoire *Bacillus Mojavensis* (PF2_TPU1) réalisé par le logiciel Mega X.

La proximité avec ces espèces fait de cette souche un candidat potentiel pour l'utilisation en industrie sachant que *B. amyloliquefaciens* et *B. licheniformis* sont parmi les microorganismes les plus utilisés pour des applications commerciales (Pandey *et al.* 2000).

Ces espèces sont utilisées dans l'industrie pour la production d'enzymes grâce à leur capacité d'endurer des conditions rudes et de produire des quantités élevées d'enzymes.

Cependant, l'arbre a montré aussi que la souche *Bacillus Mojavensis* (PF2_TPU1) a été phylogénétiquement éloignée des espèces pathogènes telles que *B. anthracis* et *B. cereus*, ce qui représente un atout de sécurité.

Selon la thèse de Mr CHOUBANE Slimane, les bactéries *Bacillus subtilis* sont certifiées sûres par le statut GRAS (FDA 2014).

Dans cette étude, nous avons pu réaliser l'identification moléculaire de la souche étudiée par une approche bio-informatique. Nous avons aussi confirmé que *Bacillus Mojavensis* (PF2_TPU1) était proche phylogénétiquement des enzymes productrices d'enzymes notamment : *Bacillus subtilis* 168, *Bacillus licheniformis* BCRC 11702 et *Bacillus amyloliquefaciens* BCRC 11601 (Figure).

En conclusion, les résultats de l'étude ont indiqué la possibilité d'utiliser la souche isolée *Bacillus Mojavensis* pour la production de l'enzyme amylase. La souche identifiée comme appartenant à l'espèce *Bacillus subtilis* est apparentée génétiquement aux espèces *Bacillus spp.* utilisées pour la production d'enzymes industrielles.

CONCLUSION

Les espèces *Bacillus* ont été très importantes en industrie avec des rôles dans la microbiologie appliquée qui datent de plus d'un millier d'années en arrière, depuis la production du *natto* par fermentation à l'état solide du soja en utilisant *Bacillus subtilis* (Schallmey *et al.* 2004).

Les bactéries *Bacillus* sont utilisées pour la production de quatre types principaux de produits : enzymes, antibiotiques, produits chimiques et insecticides. Commercialement, c'est le marché des enzymes qui est le plus important et ce genre est en particulier la source majeure d'enzymes hydrolysantes pour le marché alimentaire et des détergents (Harwood 1989).

Bacillus a deux attributs qui le rendent attractif dans la microbiologie industrielle : les endospores robustes et la production d'une variété d'enzymes. Grace à la résistance des spores, les producteurs les incluent dans les processus de production avec la certitude que la bactérie reste en vie (Maczulak 2011).

Dans ce mémoire, nous avons eu le plaisir de travailler sur les résultats de Dr CHOUBANE Slimane en contribution avec notre encadreur Dr DAHMANI C.A afin de réaliser une partie bioinformatique concernant l'identification moléculaire d'une souche du genre *Bacillus*.

Nous avons réussi d'abord, à identifier la souche d'intérêt par l'utilisation des données du séquençage du gène ARNr 16S réalisé par Dr CHOUBANE. Ensuite, nous avons confirmé qu'il s'agit bel et bien de la bactérie *Bacillus Mojavensis* qui partage 99% de ressemblance avec la souche *Bacillus subtilis* PVR05. Cette souche est une bactérie endophyte résistante à l'arsenic. Aussi, nous avons construit l'arbre phylogénétique de cette souche en comparant notre séquence du gène ARNr 16S avec celles de différentes souches bactériennes pathologiques et productrices d'enzyme.

Par ailleurs, la constitution de l'arbre phylogénétique de *Bacillus Mojavensis* nous a permis de la classer et la positionner comme étant non pathogène mais productrice d'enzymes

principalement une souche identifiée comme appartenant à l'espèce *Bacillus subtilis* et est apparenté génétiquement aux espèces *Bacillus spp.* utilisées pour la production d'enzymes industrielles.

L'arbre a montré aussi que la souche *Bacillus Mojavensis* (PF2_TPU1) a été phylogénétiquement éloignées des espèces pathogènes telles que *Bacillus anthracis* et *Bacillus cereus*, ce qui représente un atout de sécurité.

REFERENCES BIBLIOGRAPHIQUES

- Aiyer, P. V. (2005). "Amylases and their applications." African Journal of Biotechnology.
- Ameziane N, Bogard M, Lamoril J. (2005). Principe de biologie moléculaire en biologie clinique. Collection Campus Référence Elsevier
- Ash, C., J. Farrow, S. Wallbanks et M. Collins (1991). "Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences." Letters in Applied Microbiology .
- Butler, D., M. Van der Maarel et P. Steeneken (2004). "Starch-acting enzymes." Eliasson A.C. Starch in food: structure, function and applications.
- Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol; 17 : 540–52.
- Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods.
- Chaw SM, Parkinson CL, Cheng Y, Vincent TM, Palmer JD. (2000) Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc Natl Acad Sci USA
- Choubane Slimane. Amylase de *Bacillus sp.* : Production, étude biochimique et moléculaire et application biotechnologique. 2016-2017. Thèse soutenue à Université des sciences et de la technologie d'Oran (USTO-MB).
- Clarridge, J. E., III (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. Clin Microbiol Rev 17.
- Daniel, Z. et P. John (2009). The genus *Bacillus*. Practical Handbook of MICROBIOLOGY. G. Emanuel et G. Lorrence, CRC Press

- Dressman D., Yan H., Traverso G., Kinzler K.W., Vogelstein B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A*.100:8817–8822.
- Gevers,D.,Cohan,F.M.,Lawrence,J.G.,Spratt,B.G.,Coenye,T.,Feil,E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F. L. & Swings, J. (2005). Re-evaluating prokaryotic species. *Nat. Rev. Microbiol*.
- Gilbert W., Maxam A. (1973). The nucleotide sequence of the lac operator. *Proc Natl Acad Sci U S A*.;70:3581–3584.
- Gordon, R. E. etN. R. Smith (1949). "Aerobic sporeforming bacteria capable of growth at high temperatures." *J Bacteriol* 58(3).
- Goebel B.M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol*. 44:846–849.
- Gordon, R. E., W. C. Haynes, C. H.-N. Pang etN. R. Smith (1973). "The genus *Bacillus*."
- Harwood, C. R. (1989). Introduction to the biotechnology of *Bacillus*. *Bacillus*, Springer.
- Harwood, C. R. (1989). Introduction to the biotechnology of *Bacillus*. *Bacillus*, Springer.
- Holmberg K. (2003) Organic reactions in microemulsions. *Curr Opin Colloid Interface Sci*. 8:187–196.
- Hori M., Fukano H., Suzuki Y. (2007) Uniform amplification of multiple DNAs by emulsion PCR. *Biochem Biophys Res Commun*. 352:323–328.

- Hussain, I., F. Siddique, M. S. Mahmood et S. I. Ahmed (2013). "A Review of the Microbiological Aspect of α -amylase Production." International Journal of Agriculture & Biology 15.
- Iwata et al., Nucl. Acids Res. (2012) : Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features.
- Janda, J. M. & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J. Clin. Microbiol.
- Logan, N. A. et P. D. Vos (2009). "Bacillus." Bergey's Manual of Systematics of Archaea and Bacteria.
- Ludwig, W., K.-H. Schleifer et W. B. Whitman (2009). Revised road map to the phylum Firmicutes. Bergey's Manual® of Systematic Bacteriology, Springer.
- Olsen, G. J. & Woese, C. R. (1993). Ribosomal RNA: a key to phylogeny.
- O'Sullivan, B. (2005). Whole Genome Phylogeny Reconstruction for the Genus Bacillus. Department of Biology. Maynooth, NUI Maynooth.
- Peris-Tortajada, M. et A. Eliasson (2004). "Measuring starch in food." Starch in Food. CRC Press, Woodhead Publishing Limited, England.
- Peris-Tortajada, M. et A. Eliasson (2004). "Measuring starch in food." Starch in Food. CRC Press, Woodhead Publishing Limited, England.
- Petti, C. A. (2007). Detection and identification of microorganisms by gene amplification and sequencing. Clinical Infectious Diseases.

- Petti, C. A., Polage, C. R. & Schreckenberger, P. (2005). The role of 16S rRNA gene sequencing in identification of microorganisms misidentified by conventional methods. *J.Clin.Microbiol.*

- Renvoisé Aurélie (2012). Applicabilité de la PCR « universelle » 16S comme outil d'identification et de détection bactérienne en laboratoire hospitalier de bactériologie.

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

- Schmeing, T. M. & Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461.

- Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., Waterston, R.H. (2017) DNA sequencing at 40: past, present and future. *Nature.* 550:345-353.

- Singh, S., V. Sharma, M. Soni et S. DAS (2011). "Biotechnological applications of industrially important amylase enzyme." *International Journal of Pharma and Bio Sciences* 2.

- Slepecky, R. A. et H. E. Hemphill (2006). The Genus *Bacillus*—Nonmedical. The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer et E. Stackebrandt. New York, NY, Springer US.

- Smith, N. R., R. E. Gordon et F. E. Clark (1946). Aerobic mesophilic sporeforming bacteria, US Department of Agriculture.

- Sontakke, S., Cadenas, M.B., Maggi, R. G., Diniz, P.P. & Breitschwerdt, E. B. (2009). Use of broad range 16S rDNA PCR in clinical microbiology. *J. Microbiol.Methods.*

- Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., Kampfer, P., Maiden, M. C. J., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H. G., Vauterin, L., Ward, A. C.

& Whitman, W. B. (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol*.

- Sundarram, A. et T. P. K. Murthy (2014). " α -amylase production and applications: A review." *Journal of Applied & Environmental Microbiology* 2(4)

- Taggart, P. et A. Eliasson (2004). "Starch as an ingredient: manufacture and applications." *Starch in food: structure, function and applications*.

- Wang, A. et G. J. Ash (2015). "Whole Genome Phylogeny of *Bacillus* by Feature Frequency Profiles (FFP)." *Sci Rep* 5.

- Wang, Q.-f., W. Li, Y.-l. Liu, H.-h. Cao, Z. Li et al. (2007). "*Bacillus qingdaonensis* sp. nov., a moderately haloalkaliphilic bacterium isolated from a crude sea-salt sample collected near Qingdao in eastern China." *International journal of systematic and evolutionary microbiology* 57(5).

- Wang, T. L., T. Y. Bogracheva et C. L. Hedley (1998). "Starch: as simple as A, B, C?" *Journal of Experimental Botany* 49(320).

- Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol*.

- Woese . C. R. (1987). *Bacterial evolution*. *Microbiol.Rev*.

- Zeigler, D. et J. Perkins (2009). "The Genus *Bacillus*. Practical handbook of Microbiology." Taylor & Francis Group 200.

- Zeigler, D. et J. Perkins (2009). "The Genus *Bacillus*. Practical handbook of Microbiology." Taylor & Francis Group 200.

- Ziebolz B. Droege M (2007). Toward a new era in sequencing. *Biotechnol Annu Rev*. 13:1

