



الجمهورية الجزائرية الديمقراطية الشعبية
RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي و البحث العلمي
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Frères Mentouri Constantine 1
Faculté des Sciences de la Nature et de la Vie

جامعة الأخوة منتوري قسنطينة
كلية علوم الطبيعة و الحياة

Département : MICROBIOLOGIE

قسم : ميكروبيولوجيا

Mémoire présenté en vue de l'obtention du Diplôme de Master

Domaine : Sciences de la Nature et de la Vie

Filière : Ecologie et environnement

Spécialité : Ecologie microbienne

Intitulé :

Une démarche à suivre pour l'identification d'une bactérie inconnue

Présenté par : *HEDDOUD Houcine Islem*

HAMADA Ayoub

Le : 23/09/2012

Jury d'évaluation :

Président du jury : Mr. BENHIZIA Yacine (Professeur- UFM Constantine 1).

Rapporteur : Mme. GUERGOURI Ibtissem (MAA- UFM Constantine 1).

Examineur : Mme. BOULTIFAT Linda (MCB- UFM Constantine 1).

*Année universitaire
2020- 2021*

Remerciements

Tout travail ne peut être réalisé sans l'aide de Dieu le plus puissant. Alors, et en premier lieu, on remercie notre créateur qui nous a donné la force et la santé pour en être la aujourd'hui.

Nous tenons à remercier notamment notre encadreur madame GUERGOURI Ibtissam pour son aide et ses conseils précieux qui ont contribué à l'accomplissement de ce travail.

Nos remerciements les plus sincères s'adressent également aux membres du jury monsieur BENHIZIA Yacine et madame BOULTIFAS Linda qui ont accepté d'examiner notre mémoire et qui nous ont fait l'honneur de siéger dans notre jury.

Nous tenons également à remercier nos familles pour nous avoir montré un énorme soutien tout au long de notre parcours éducatif et surtout pendant notre travail, ainsi que nos amis et collègues pour tout leur amour et leur soutien.

Notre cycle d'études ainsi que la réalisation de ce travail ne se seraient pas aussi bien déroulé sans l'aide des enseignants et le personnel administratif, nos remerciements les plus vifs vont à eux également.

RÉSUMÉ

La combinaison de caractères phénotypiques et phylogénétiques s'avère importante dans le cas de la détermination du genre bactérien d'un isolat inconnu. C'est une approche polyphasique que l'on qualifie de « consensus ». Ce présent travail est une feuille de route montrant les différentes étapes de représentation graphique d'un arbre phylogénétique à partir d'une séquence partielle d'*ARNr16S* et de séquences homologues. Il montre suffisamment de détails pour permettre à une personne débutante d'entreprendre ce genre d'analyses, faisant usage des bases de données telles que NCBI et EZ-BioCloud, et des logiciels de bioinformatique tels que MEGA X. Les résultats d'une étude antérieure de caractérisation phénotypique et d'identification d'un isolat inconnu, conduisant au genre bactérien *Bacillus* sp étaient comparés à ceux obtenus par reconstruction phylogénétique de l'arbre afin de voir la correspondance des deux approches et est-ce que la phylogénie positionne la bactérie au sein du même genre. Les résultats obtenus par les deux approches phénotypique et phylogénétique ont conduit à une affiliation au genre *Bacillus*.

Mot clés : approche polyphasique, arbre phylogénétique, genre *Bacillus*, alignement de séquence, *ARNr 16S*.

ABSTRACT

The combination of phenotypic and phylogenetic traits is important when determining the bacterial genus of an unknown isolate. It is a multi-phased approach that is called "consensus". This present work is a roadmap showing the different stages of graphical representation of a phylogenetic tree from a partial sequence of *16S rRNA* and homologous sequences. It shows enough detail to allow a newbie to undertake this kind of analysis, making use of databases such as NCBI and EZ-BioCloud, and bioinformatics software such as MEGA X. Study results prior phenotypic characterization and identification of an unknown isolate, leading to the bacterial genus *Bacillus* sp. were compared to those obtained by phylogenetic reconstruction of the tree to see the correspondence of the two approaches and whether the phylogeny positions the bacteria within the same genus. The results obtained by both phenotypic and phylogenetic approaches led to the affiliation of genus *Bacillus*.

Keywords: polyphasic approach, phylogenetic tree, genus *Bacillus*, sequence alignment, *16S rRNA*.

الملخص

يعد الجمع بين السمات المظهرية و الوراثةية أمرا مهما عند تحديد الجنس البكتيري لعزلة غير معروف. و هو منهج متعدد المراحل يسمى «توافق الآراء». هذا العمل الحالي عبارة عن خارطة طريق توضح المراحل المختلفة للتمثيل البياني للشجرة الوراثةية من تسلسل لـ ARNr16S و المتواليات المتماثلة. يعرض تفاصيل كافية للسماح لشخص مبتدئ بإجراء هذا النوع من التحليل ، و الاستفادة من قواعد البيانات مثل NCBI وEZ BioCloud ، و برامج المعلوماتية الحيوية مثل MEGAX . نتائج الدراسة السابقة لوصف النمط الظاهري و التعرف على العزلة المجهولة تؤدي الى الجنس البكتيري *Bacillus sp* . تمت مقارنتها مع تلك التي تم الحصول عليها من خلال اعادة بناء الشجرة لمعرفة التطابق بين المنهجين و ما اذا كان تطور السلالة يضع البكتيريا داخل نفس الجنس. ادت النتائج التي تم الحصول عليها من خلال المنهجين الظاهري و الوراثةي الى نتيجة الجنس

Bacillus

الكلمات المفتاحية: نهج متعدد الأطوار ، شجرة النسوء و التطور ، جنس العصوية ، محاذاة التسلسل ، ARNr 16S.

Liste des abréviations

AMC: Acétyle- méthyl- carbinole.

BLAST: *Basic local alignment search tool.*

ClustalW: *Clustal weighting.*

DDBJ: *DNA data bank of Japan.*

ddNTP: Didésoxyribonucléotides triphosphates.

dNTP: Désoxyribonucléotides triphosphates.

EMBL: Laboratoire européen de biologie moléculaire.

MEGA: *Molecular evolutionary genetics analysis.*

ML: maximum de vraisemblance

NCBI: *National center for biotechnology information.*

NJ: *Neighbor joining.*

Pb: Pair de bases.

PFGE : Electrophorèse sur gel a champ pulsé.

RAPD : Amplification aléatoire d'ADN polymorphe.

RFLP : Polymorphisme de longueur des fragments de restriction.

TIGR: *The institute for genome research.*

UPGMA: *Unweight pair group method with arithmetic mean.*

VF: Viande-foie.

VP-RM: Voyes Proskauer- rouge de méthyle.

ΔT_m : Valeur des températures de demi-dénaturation.

Liste des figures

Figure 1 : Les disciplines de la taxonomie.....	3
Figure 2 : L'arbre phylogénétique.....	6
Figure 3 : La hiérarchie taxonomique.....	6
Figure 4 : Événements clés de la taxonomie et de la nomenclature des procaryotes au cours des 100 dernières années.....	11
Figure 5 : Les aspects de la classification bactérienne.....	13
Figure 6 : Méthodologies pour la caractérisation des procaryotes et les niveaux de résolution taxonomiques respectifs approximatifs.....	18
Figure 7 : Courte carte d'identification des espèces de bacilles.....	22
Figure 8 : La diversité de l'apparence colonies au sein du genre <i>Bacillus</i>	24
Figure 9 : Le séquençage d'ADN.....	28
Figure 10 : Les étapes de la technique de Sanger.....	29
Figure 11 : La technique de séquençage Sanger.....	30
Figure 12 : Page web de National Center for Biotechnology Information (NCBI).....	32
Figure 13 : Page web d'EzBioCloud.....	33
Figure 14 : Photographies de l'aspect macroscopique et microscopique des colonies.....	39
Figure 15 : Clé dichotomique d'orientation vers le genre de la souche M3.....	40
Figure 16 : Le chromatogramme de séquence après séquençage.....	41
Figure 17 : Extraction de la séquence sous forme FASTA.....	42
Figure 18 : Le Format FASTA sur Notepad.....	42
Figure 19 : Réglage des paramètres sur la fenêtre de BLASTn du NCBI.....	43
Figure 20 : Résultat de BLAST de la séquence.....	44
Figure 21 : Séquence valide avec le lien vers sa publication.....	44
Figure 22 : La base de données EzBioCloud.....	45
Figure 23 : Fenêtre d'identification d'une nouvelle séquence.....	46
Figure 24 : Résultats de l'identification de la nouvelle séquence.....	46
Figure 25 : Page descriptive de la souche.....	47
Figure 26 : L'ajout des séquences nucléotidiques dans l'explorateur d'alignement de MEGA X.....	48
Figure 27 : Alignement des séquences sur le MEGA X.....	49

Figure 28: Utilisation du Programme CLUSTAL W pour aligner les séquences.....49

Figure 29: Choix de la méthode ML pour la reconstruction de l'arbre phylogénétique.....50

Figure 30 : Paramètres de construction de l'arbre par ML.....50

Figure 31: Test pour choix du modèle de calcul des distances génétiques.....51

Figure 32 : Construction de l'arbre et affichage de l'échelle.....52

Figure 33 : Arbre obtenu par Analyse évolutive par la méthode du Maximum de
Vraisemblance.....57

Figure 34 : Arbre phylogénétique en mode « topologie seulement ».....58

Liste des tableaux

<i>Tableau 1</i> : Ressources de taxonomie et nomenclature en ligne.....	31
<i>Tableau 2</i> : Caractères phénotypiques de la souche M3.....	39
<i>Tableau 3</i> : Les souches bactériennes similaires publiées sur NCBI.....	54
<i>Tableau 4</i> : Les résultats d'EzBioCloud.....	55
<i>Tableau 5</i> : Les rangs taxonomiques de la souche M3.....	59

Remerciements	
Résumé	
Abstract	
المخلص	
Liste des abréviations	
Liste des figures	
Liste des tableaux	
Introduction	1
Chapitre I : La taxonomie des procaryotes, spécificité et intérêts	3
I-1-La taxonomie des procaryotes.....	3
I-2-L'importance de la taxonomie.....	4
I-3-La hiérarchie taxonomique.....	5
I-4-L'espèce chez les procaryotes.....	7
I-4-1-La définition de l'espèce bactérienne selon la commission de Wayne <i>et al.</i> (1987)..	7
I-4-2-La définition de l'espèce bactérienne selon la commission de Stackebrandt (2002).....	7
I-5-Intérêt d'identifier un microorganisme inconnu.....	8
I-6-L'évolution de la taxonomie.....	9
I-7-La différence des approches d'identification bactérienne en fonction du domaine.....	12
I-8-Les techniques appliquées en taxonomie, leurs limites et leur pouvoir de résolution.....	14
I-8-1-L'étude du phénotype.....	14
I-8-2-Les limites des approches phénotypiques.....	14
I-8-3-L'étude du génotype.....	15
I-8-4-Les limites des approches moléculaires.....	16
I-8-5-Le pouvoir discriminatoire des différentes techniques.....	16
I-9-Importance du séquençage d'ARNr 16S.....	18
Chapitre II : Identification d'un bacille à Gram positif, cas particulier des <i>Bacillus</i>	20
II-1- Une approche polyphasique pour l'identification d'un <i>Bacillus</i>	21
II -1-1-La mise en culture et la purification de la bactérie.....	21

II -1-2-Les caractères phénotypiques recherchés pour identifier le genre <i>Bacillus</i>	21
II -2-1-Caractérisation morphologique des <i>Bacillus</i>	23
II -2-2-Caractérisation biochimique et physiologique des <i>Bacillus</i>	25
II -3-La biologie moléculaire et la phylogénie des <i>Bacillus</i>	26
II-4- Démarche pour l'analyse d'une copie de l'ADNr16s chez une bactérie inconnue.....	27
II -4-1-La réaction de polymérisation en chaîne (PCR).....	27
II-4-2- Le séquençage.....	28
II-5- L'analyse bioinformatique des séquences.....	30
II-5-1- Les bases de données.....	30
II-5-2- L'alignement de séquences.....	33
II-5-3- L'arbre phylogénétique.....	35
Chapitre III : Matériel et méthodes	38
III-1-Matériel biologique.....	38
III-2-Caractérisation phénotypique.....	38
III-3-Analyse des résultats de séquençage.....	41
III-3-1-Recherche de séquences homologues et l'identification de la bactérie.....	41
III-3-2-Représentation de l'arbre phylogénétique.....	48
III-4-Résultats, interprétation et discussion.....	54
III-4-1- Résultats de l'alignement par BLASTn.....	54
III-4-2-Résultat de la construction de l'arbre phylogénétique.....	57
III-4-3- Interprétation et discussion.....	58
III-4-4-La comparaison des résultats des deux approches phénotypique et phylogénétique.....	60
Conclusion	61
Références bibliographiques	63

Introduction

Introduction

La taxonomie microbienne traitant de la caractérisation et l'identification des procaryotes a toujours été et restera à jamais un domaine de recherche majeur en microbiologie à travers le monde. Elle fournit aux microbiologistes à l'échelle internationale des points de repère et références pour l'étude de chaque groupe de microorganismes, impliqués dans les différentes disciplines.

La reconnaissance du microorganisme en question est d'une importance fondamentale dans l'industrie, l'agriculture, l'alimentation, clinique, domaine de la recherche scientifique et particulièrement dans la découverte et la préservation des écosystèmes environnementaux.

Selon Philip Hugenholtz et ses collaborateurs (2021), la taxonomie peut être basée sur n'importe quelle combinaison de techniques ; généralement celles qui révèlent le plus de caractères distinctifs permettant de différencier les microorganismes les uns des autres et rapprocher ceux qui se ressemblent. Mais avant tout, il faut commencer par la reconnaissance de leur histoire évolutive (ascendants communs) comme la façon la plus naturelle d'organiser les organismes. Cependant, les micro-organismes ont été actuellement les taxons les plus problématiques à organiser dans un cadre phylogénétique parce que leurs propriétés phénotypiques ne révèlent pas pour la plupart leurs liens de parenté.

La caractérisation phénotypique est l'approche préliminaire largement employée jusqu'à nos jours dans les différents domaines particulièrement dans les pays ne disposant pas de l'arsenal moléculaire. Elle est en effet, d'un grand intérêt dans l'identification bactérienne, cependant, elle reste insuffisantes et des fois même déroutante.

Le développement intensif des technologies se poursuivra pour permettre une résolution toujours plus élevée et une classification hiérarchique de plus en plus raffinée au niveau du genre, de l'espèce et au-dessous de l'espèce.

En mettant l'accent sur l'état actuel de la technologie et le spectre des méthodes utilisées aujourd'hui pour la caractérisation des procaryotes, il s'est avéré que pour l'établissement d'un système réussi de classification hiérarchique stable et pratique, l'application exclusive de paramètres phylogénétiques n'est pas suffisante. En effet, les propriétés phénotypiques de valeur taxonomique acceptable doivent être combinées aux données phylogénétiques pour la délimitation d'au moins le genre et rangs taxonomiques supérieurs.

L'identification précise d'un microorganisme reste une tâche extrêmement compliquée vu qu'il n'existe pas de guide universel expliquant la démarche à suivre pas à pas. D'autant plus que l'analyse phylogénétique et la production d'un arbre phylogénétique est parfois considérée comme un processus sophistiqué nécessitant des années d'expériences.

Selon Barry G. Hall (2013), ce processus d'analyse des données moléculaires est assez simple, il peut être appris rapidement, et appliqué efficacement.

Dans cet esprit, l'objectif qu'on s'est fixé dans ce travail est d'établir et de tracer un chemin à suivre pour arriver à la identification d'une bactérie inconnue isolée d'un sol pollué, ainsi que la détermination de sa position taxonomique au sein de son groupe via son arbre phylogénétique. L'approche est basée sur l'analyse d'une séquence nucléotidique de gène d'*ARNr 16S* issue d'une bactérie dont le profil phénotypique est établi par une étude précédente. Une comparaison des résultats des deux approches est effectuée en dernier lieu.

Chapitre I :
La taxonomie
des
procaryotes...
spécificité et
intérêts

I-1-La taxonomie des procaryotes

La taxinomie, ou la taxonomie est un domaine fondamentale des sciences biologiques, qui étudie les lois d'arrangement des formes de vie variées (taxis = arrangement, ordre ; nomos = lois).

La systématique est la science qui permet de classer les microorganismes, de rechercher leur relation phylogénétique et de retracer leur histoire évolutive. La taxonomie fournit alors un moyen cohérent pour l'étude de la diversité à travers différentes disciplines comme la morphologie, l'écologie, l'épidémiologie, la biochimie, la biologie moléculaire et la physiologie. Au sens large, la dynamique de la taxonomie comprend trois branches distinctes mais étroitement interdépendantes (Figure 1), la classification, la nomenclature et l'identification [7, 8, 20].

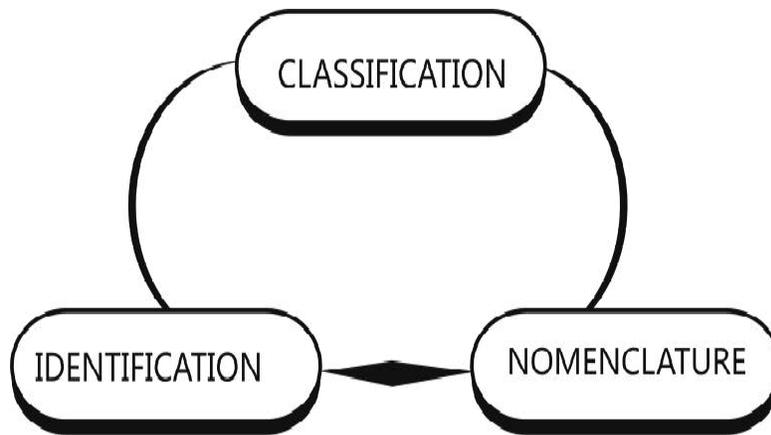


Figure 1 : Les disciplines de la taxonomie [7].

La classification est l'une des préoccupations fondamentale d'un schéma taxonomique, elle signifie l'action d'organiser et répartir les organismes en groupes appelés taxons, en fonction de leur similarité mutuelle, selon des critères phylogénétiques, phénotypiques et génétiques [7,20].

La nomenclature est la discipline taxonomique qui fournit des noms distinctifs aux individus selon des règles publiées, qui sont dérivées du système binomial établie par Linnaeus (Carl Von Linne : grand taxonomiste du dix-huitième siècle) dans lequel un nom latin de genre précède le nom d'espèce accordé à un individu [7,40].

L'identification est le processus qui attribue un organisme inconnu à l'une des unités définies et marquées de la hiérarchie taxonomique par des méthodes de comparaison à d'autres organismes connus pour définir ressemblances et différences entre eux. Ce processus implique des

critères phénotypiques et de nombreuses méthodes d'identification moléculaire afin d'arranger les microorganismes au mieux [7, 20, 38].

La taxonomie conçoit des méthodes appropriées d'organiser et de cataloguer les unités taxonomiques, c'est-à-dire d'établir des relations entre les groupes d'organismes et de faire la distinction entre eux, elle fournit également des points de repère et référence pour une identification des organismes classés [7, 20].

I-2-L'importance de la taxonomie

La taxonomie est une branche fondamentale dans la recherche en microbiologie. Elle met en évidence une très grande variété de micro-organismes sur la planète Terre. Chaque effort de recherche conduit à une variété d'isolats directement ou indirectement, enrichissant ainsi les connaissances sur la biodiversité, chacune est unique et contribue à sa manière à l'environnement. Afin d'identifier un isolat particulier, de le placer dans une catégorie et connaître sa signification au monde, il est nécessaire de l'ajouter au système de taxonomie existant, afin qu'il puisse être étudié et utilisé au profit de la recherche scientifique. Il existe diverses caractéristiques du microbiote, qui aident à leur inclusion et à leur classification taxonomique ainsi qu'à la compréhension de l'origine de leur diversité [34].

Leur classification se fait à l'aide de l'étude de leurs similitudes (ou différences), des relations ainsi que des interactions qui existent entre eux. Plusieurs problèmes comme les défis croissants posés par les micro-organismes multirésistants en microbiologie médicale, les erreurs expérimentales des tests microbiologiques et par conséquent l'identification des agents pathogènes et l'administration de traitement aux maladies, dépendent d'une classification taxonomique appropriée. Des recherches ont été menées dans ce sens, mais il reste encore beaucoup à faire [34].

La taxonomie est une discipline très importante pour le monde scientifique, car elle soutient d'autres plusieurs champs disciplinaires comme le domaine de la médecine et santé, les recherches scientifiques, les champs de production alimentaire ou industriel, etc. elle reflète les attributs de microorganismes placés dans ces groupements. La taxonomie implique une dynamique de description qui aide à catégoriser les organismes afin de faciliter la communication des données biologiques et l'utiliser comme un moyen à comprendre la diversité de la vie sur notre planète. [35].

I-3-La hiérarchie taxonomique

Les rangs taxonomiques sont des niveaux d'organisation biologique, ou des constructions définies subjectivement basées sur des dissemblances des groupes d'organismes. Ces rangs ou taxons ont été introduits dans la systématique des spécimens biologiques pour simplifier la communication entre les scientifiques et pour la hiérarchie des organismes selon des caractères variés présents dans un rang mais absents dans les autres [31].

Les rangs de la classification des procaryotes ainsi que leur nomenclature suivent le schéma général de Carl Von Linne proposé dans son ouvrage *Systema Naturae* (1735) destiné aux botanistes. La classification bactérienne a commencée avec la contribution d'un arrière-plan botanique au dix-neuvième siècle pour établir après un système comparable à ceux des plantes et des animaux, basé principalement sur les similitudes anatomiques et morphologiques [31].

Pour le botaniste Cohn en (2002), cette systématique étaient inappropriés pour démêler un cadre phylogénétique, il a été le premier à soulever la question de savoir si le classement des bactéries peut être comme celui des plantes et des animaux, cependant, il a dicté que la restriction aux propriétés morphologiques et la similitude de forme sont insuffisantes et n'exclut pas que les bactéries diffèrent les unes des autres.[31]. Actuellement, la différence entre les structures hiérarchiques des procaryotes est décrite par l'analyse caractéristique des séquences de l'ADN et l'ARN et des protéines. Cette structure est principalement basée sur l'arbre phylogénétique (Figure 2) du gènes de la petite sous-unité de l'ARN ribosomique (un gène marqueur), une molécule universellement présente entre les taxons procaryotes, ce qui indique une phylogénie de l'ARN ribosomique entraînant une taxonomie et des changements de rangs à cause de leur division et fusion et de l'introduction de nouveaux taxons à tous les niveaux [59].

Plusieurs niveaux ou rangs sont utilisés dans la classification des procaryotes, les microorganismes, dans chaque niveau ou rang, partagent un ensemble de caractères spécifiques. Les rangs sont organisés en une hiérarchie non chevauchante, de sorte que chacun d'eux incluse non seulement les traits qui définissent le rang, mais un nouvel ensemble de traits plus restrictifs (Figure3), le rang supérieur est appelé le domaine. Tous les organismes procaryotes sont placés dans deux domaines : archées et bactéries. Les sous-ensembles successivement plus petits et non chevauchants du domaine sont le phylum, la classe, l'ordre, la famille, le genre, l'espèce et la sous-espèce. L'unité de base de système est l'espèce et le genre, la description binomiale d'une espèce

type n'est pas possible sans décrire un genre et vice versa, un genre ne peut pas être décrit sans une espèce [8, 20].

Bien que les rangs taxonomiques supérieurs soient utiles pour apprécier les relations globales des bactéries, la « classification de travail » repose sur les rangs binomiaux du genre et de l'espèce [46].

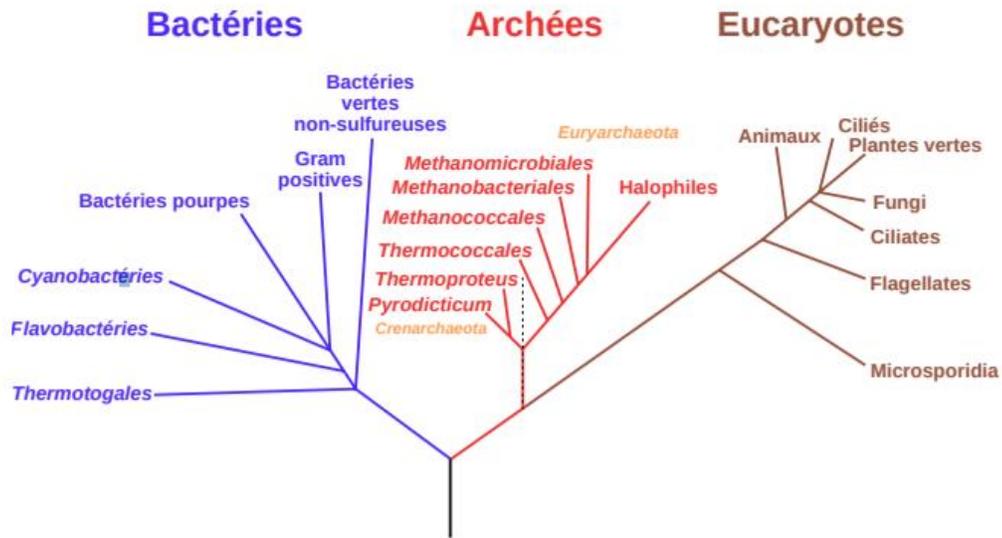


Figure 2 : l'arbre phylogénétique [39].

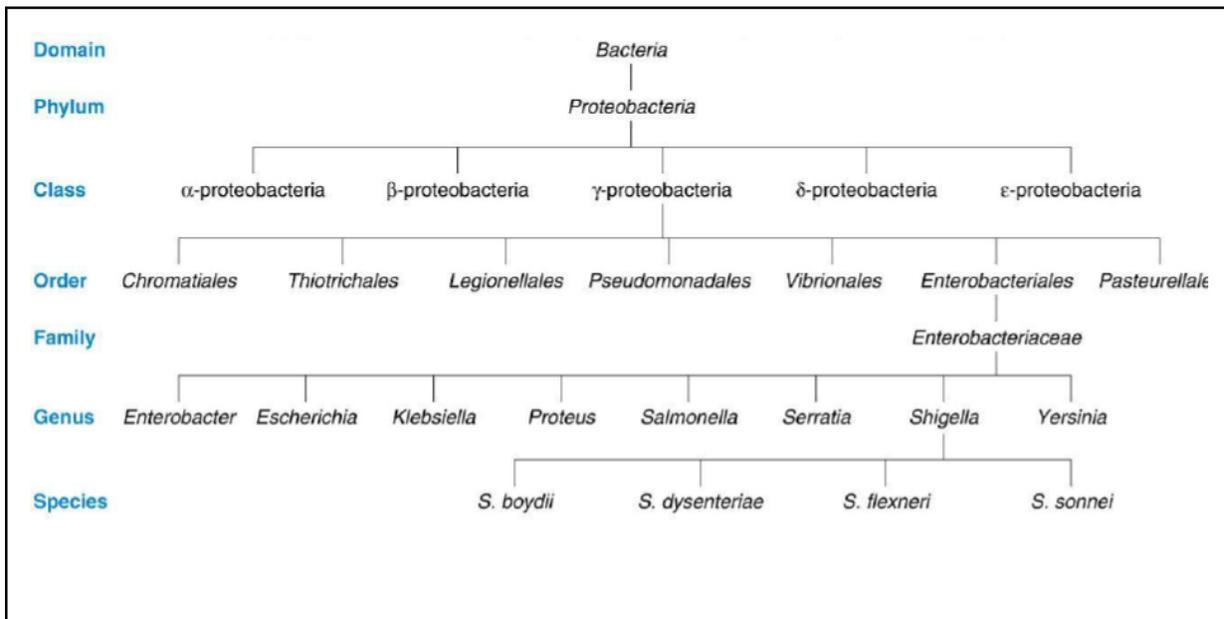


Figure 3 : la hiérarchie taxonomique [20].

I-4-L'espèce chez les procaryotes

Le domaine de la taxonomie des procaryotes a connu un essor au cours de vingtième siècle, le système de la classification ainsi que la nomenclature linnéenne ont été adoptés à tous les taxons à l'exception de l'espèce, c'est le groupe taxonomique de base et le plus important dans la systématique bactérienne. Une espèce procaryote est un concept différent de celui des organismes supérieurs car il n'est pas universel et les espèces sont considérées comme des individus ou des entités pratiques pour lesquelles les circonscriptions peuvent varier selon la notion phénétique et évolutif d'espèce [26, 57].

Actuellement, l'espèce procaryote est le résultat d'amélioration et d'optimisation des méthodes microbiologiques qui révèlent à la fois les propriétés génomiques et phénotypiques des spécimens procaryotiques, ainsi que des relations phylogénétiques décelés par la similitude d'ADN. En parallèle, le terme espèce définit un groupe de souches qui ont certaines caractéristiques considérablement distinctives ; chaque espèce diffère et peut être distinguée de toutes les autres espèces [26, 57].

I-4-1-La définition de l'espèce bactérienne selon la commission de Wayne (1987)

Selon le Comité International pour la Systématique des Procaryotes, la définition phylogénétique de l'espèce bactérienne est la réunion des différentes souches qui forment une même espèce, selon les critères suivants :

- Un seuil d'hybridation d'ADN/ADN d'environ 70% ou plus.
- Une valeur de ΔT_m de 5°C maximum.

Malgré les limites techniques à l'époque, le comité révèle l'importance du potentiel phylogénétique des ARN ribosomiques et la possibilité de séquencer l'intégralité d'un génome bactérien afin d'évaluer l'espèce bactérienne [39].

I-4-2-La définition de l'espèce bactérienne selon la commission de Stackebrandt (2002)

Le comité convient d'encourager des séries de recommandations afin d'intégrer des techniques génomiques innovantes pouvant compléter ou supplanter l'hybridation ADN/ADN avec d'autres empreintes génétiques pour uniformiser les descriptions des espèces identifiables. Ce comité cite notamment le séquençage de gènes de ménage ou tout autre gène conservé et maintient la description des espèces d'une manière plus précise avec le maximum d'approches

possibles, phénotypique, génotypique, ou phylogénétique. Cette description est appelée une approche polyphasique, elle permet d'apporter une identification de l'espèce la plus précise possible et présente une cohérence génétique et phénotypique [39].

I-5-Intérêt d'identifier un microorganisme inconnu

L'identification bactérienne d'un organisme nouvellement isolé peut servir comme un instrument pour l'associer à ses alliés les plus proches ou pour l'admettre comme un nouveau taxon jusqu'alors inconnu. La nécessité d'identifier un microorganisme dans plusieurs domaines distincts de la microbiologie est une application très importante, donc une classification des microorganismes outre son intérêt scientifique, a été liée aux découvertes du rôle de certains d'entre eux, comme par exemple dans les maladies infectieuses, la détérioration des denrées alimentaires, ainsi que leur utilisation dans la production industrielle par fermentation de produits naturels [38, 55]. L'utilisation des suppléments microbiens réglementés comme engrais en agriculture et aussi un volet qui nécessite l'évaluation d'innocuité par une identification précise du microorganisme [11].

L'identification d'un organisme fournit un établissement d'une connaissance approfondie et permet aussi la description de nouveaux taxons basée sur l'interprétation intégrée de données de divers champs disciplinaires :

- En médecine, l'identification bactérienne permet d'identifier la maladie, de prescrire le bon traitement la maladie. La taxonomie différencie les différents agents pathogènes : les bactéries, les champignons, les virus, ainsi que les protozoaires.
- En agriculture, la classification d'un isolat permet d'identifier les variétés à haut rendement, les maladies causant des ravages, les vers de terre utiles pour la production de lombricompost, le type de biofertilisant, etc...
- En microbiologie, l'identification fournit des outils de validation des procédures de routine appliquées ; la stérilisation (par exemple l'usage des spores de *Bacillus stearothermophilus* pour valider l'autoclavage), la désinfection, l'immunologie, etc.
- En écologie, la gestion écologique peut aider à améliorer l'environnement, à valoriser les ressources naturelles et à protéger les êtres vivants. L'identification exacte d'un micro-organisme permet d'utiliser les caractéristiques connues d'un groupe taxonomique pour en évaluer les risques s'il s'agit d'un pathogène d'importance clinique ou environnemental, ou

d'exploiter son potentiel métabolique comme pour remédier les sites pollués par des xénobiotiques [11, 13, 55].

I-6-L'évolution de la taxonomie

A la fin du dix-neuvième siècle, le début de la taxonomie bactérienne a été marqué par la disponibilité des nouvelles techniques de la classification. En 1872, Ferdinand Cohn a été le premier à classer les six genres bactériens entant que membres du monde des plantes en fonction de leurs similitudes morphologiques. Entre 1880 et 1900, la plus part des scientifiques s'intéressaient à la description des nombreuses bactéries pathogènes selon la morphologie, les exigences de croissances et le potentiel pathogène [59].

Au début du vingtième siècle, la caractérisation et l'identification des cultures bactériennes étaient basées sur des données physiologiques et biochimiques en plus de la morphologie, plus tard alors, les enzymes ont été élucidées avec les voies métaboliques.

La première édition du Bergey's Manual of Determinative Bacteriology a classé les bactéries en 1923 sur la base de ces propriétés (Figure 3). Même dans sa 7e édition, publiée en 1957, les bactéries étaient encore classées comme membres des plantes (Protophyte, plantes primitives). Dans la 8e édition du Manuel de Bergey, publiée en 1974, les bactéries étaient reconnues comme membres du royaume des Procaryotes et organisées en groupes basés principalement sur la coloration de Gram, la morphologie et les besoins en oxygène. Les archées ont été classées pour la première fois en tant que royaume distinct sur la base des séquences partielles des gènes de l'ARN ribosomique 16S (ARNr) [9, 59].

Actuellement, il n'existe toujours pas de système officiellement reconnu pour la classification des procaryotes. En effet, pour identifier une bactérie, les approches comprennent des données phénotypiques, chimiotaxonomiques, génétiques, ainsi que les données phylogénétiques, c'est ce que l'on appelle une approche polyphasique [26].

Le séquençage du gène de l'ARNr 16S est appliqué pour déterminer la position phylogénétique des organismes (Figure 4). Chaque taxon doit être décrit et différencié des taxons apparentés par ses caractéristiques phénotypiques, chimiotaxonomiques et génotypiques, ces différences distinctives sont nécessaires pour la description d'une nouvelle espèce. Si de telles différences ne sont pas trouvées, les groupes de bactéries similaires qui semblent être génétiquement distincts doivent être décrits par d'autres termes [26]. Ce système de classification pragmatique pouvant servir d'outil d'identification de routine. Une classification peu utile aux microbiologistes, quel que soit le degré

de sophistication de son schéma, sera bientôt ignorée ou considérablement modifiée. L'approche polyphasique a fait ses preuves dans la taxonomie bactérienne, mais apparemment elle n'est toujours pas satisfaisante. Elle ne peut pas faire face à l'immense diversité microbienne qui reste à révéler [58, 59].

Dans la découverte des liens entre la phylogénie et les caractères phénotypiques, la taxonomie parvient à un système de classification basé sur la théorie de concept phylogénétique/évolutif pragmatique, de plus, les études métagénomiques peuvent être utiles pour améliorer la classification des bactéries non cultivées. Cependant, dans la future classification des bactéries, les processus évolutifs conduisent toujours à des groupes d'organismes clairement résolus qui dérivent d'un ancêtre commun, à cause de l'effet encore peu claire du transfert de gènes, en particulier le transfert horizontal, ce qui rend la description et la distinction des espèces difficile et peut éliminer les frontières entre eux. Il est également possible que des changements dans le flux de gènes causés par des facteurs écologiques peuvent même conduire à un début de fusion de deux espèces bactériennes ou toutes autres unités taxonomiques [59].

En revanche, il existe également d'autres indicateurs qui soutiennent la faisabilité d'une classification stable des bactéries même face à l'immense fluidité génomique substantielle. Les gènes conservés du génome sont des marqueurs phylogénétiques pour la caractérisation géno-phénotypiques d'un taxon. L'analyse comparative des séquences de certains gènes de base, y compris les gènes d'ARNr, peut être utile pour tracer les relations phylogénétiques des taxons supérieurs, aussi divers gènes de caractères, en particulier les gènes qui codent pour des différences phénotypiques clés, peuvent convenir comme marqueurs phylogénétiques pour la classification des taxons inférieurs. Leur expression en protéines fonctionnelles détectées dans le protéome pourrait aussi offrir des informations sur les caractéristiques phénotypiques et ils peuvent être utilisés pour définir un taxon [59].

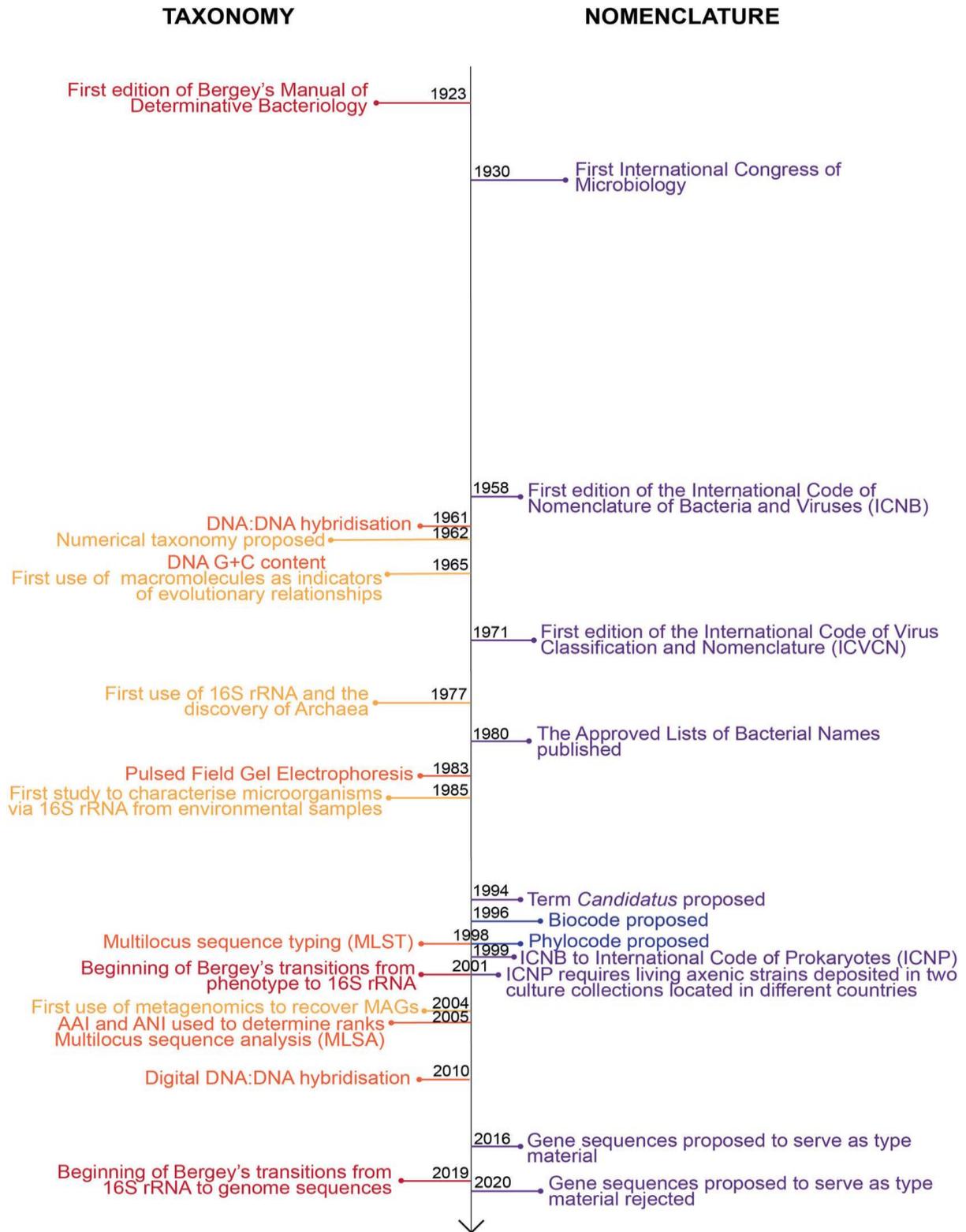


Figure 4 : Événements clés de la taxonomie et de la nomenclature des procaryotes au cours des 100 dernières années [26].

I-7-La différence des approches d'identification bactérienne en fonction du domaine

Dans la pratique actuelle de la classification des bactéries et des archées, une méthode en plusieurs phases est utilisée pour identifier une nouvelle espèce, dans laquelle nous considérons les aspects multidimensionnels de l'organisme, y compris le phénotype, le génotype et les caractéristiques chimiques (Figure 5). Dans ce processus, les caractéristiques génotypiques sont les éléments de base de la description des espèces, car l'information génétique révèle la relation évolutive entre les différentes lignées. L'analyse phylogénétique basée sur la séquence du gène de l'*ARNr 16S* et la détermination de la similarité de séquence sont généralement effectuées comme première étape de l'identification de nouveaux organismes [17]. Cette méthode consensuelle prend en compte les données de différentes origines en utilisant la phylogénie générale dérivée de l'analyse de séquence d'*ARNr 16S* comme cadre. On peut dire qu'il s'agit d'un compromis qui contient le moins de contradictions possible. Les taxonomistes n'ont pas utilisé les valeurs de pourcentage d'hybridation de l'ADN pour normaliser la classification des espèces. L'absence d'un autre « étalon-or » pour la définition des espèces a eu un impact énorme sur la taxonomie bactérienne. Ce problème est rencontré dans la taxonomie polyphasique, qui ne dépend pas d'une théorie, d'une hypothèse ou d'un ensemble de règles, mais plutôt d'une approche pragmatique d'une taxonomie de type consensus, intégrant au maximum toutes les données disponibles [60].

Le choix de la ou les méthodes à suivre pour l'identification microbienne est fonction de certains facteurs ; comme la nature des taxons microbiens (bactérie ou autres) et la différence du domaine microbiologique (clinique, industrielle, alimentaire) ou en domaine de la recherche scientifique...etc. De ce fait, la taxonomie constitue une discipline centrale avec des différentes approches polyphasique d'identification microbienne [11, 12].

En bactériologie clinique, lors d'une infection sévère, la détection et détermination des agents bactériens responsables et leurs sensibilités aux antibiotiques sont des étapes de diagnostic importantes. Cette mise au point passe par des tests moléculaires comme : la PCR à large spectre, l'amplification par PCR des toxines et l'identification des ERV, SARM par PCR à partir des cultures positives. Ces méthodes se rajoutent aux tests conventionnels (sérologiques, antibiogrammes, détection des antigènes, hémoculture automate de détection, identification par immunofluorescence, par LATEX agglutination, les tests ELISA ainsi que la production des toxines et métabolites) [18, 24].

Dans le domaine bactériologique des industries, les méthodes d'identification bactérienne sont encore très prudentes quant à l'usage rapide en vue de l'amélioration significative de la qualité des produits industriels. Afin de contrôler la présence ou l'absence des microorganismes dans le processus industriel, l'identification bactérienne débute par les techniques de dénombrement quantitatif, basée sur la détection et comptage des colonies sur les milieux de culture. De ce fait la caractérisation microbiologique et biochimique qui suit consiste à : l'analyse de la microflore, l'analyse de catalase et coloration de gram, le profil fermentaire et enzymatique. Outre les approches phénotypiques, la méthode de référence pour une caractérisation génotypique reste le séquençage d'ADN. Le protéotypage est l'une des méthodes qui permet l'obtention d'identification à haute qualité en industrie [10, 25, 37].

Dans le cadre des analyses bactériologiques alimentaires, l'identification bactérienne doit détecter et évaluer les flores pathogènes et d'altération présentes dans les aliments. En effet, cette identification est établie selon une approche appropriée qui commence avec des techniques de numération, soit microscopiques, en milieu liquide ou solide, suivi par une identification phénotypique : (caractères cultureux, morphologiques et structuraux, physiologiques et biochimiques, immunologiques, et l'étude du pouvoir pathogène) [61]. La mise au point d'une caractérisation moléculaire consiste à un typage des agents pathogènes d'origine alimentaire par amplification par PCR, séquençage et empreintes électrophorétiques.

L'approche polyphasique est utilisée pour choisir les méthodes appropriées et nécessaires pour identifier un microorganisme. Idéalement, le choix de ces méthodes sert à justifier la discrimination et faciliter l'organisation de nombreuses données recueillis dans le cadre d'identification et caractérisation de microorganisme [61].

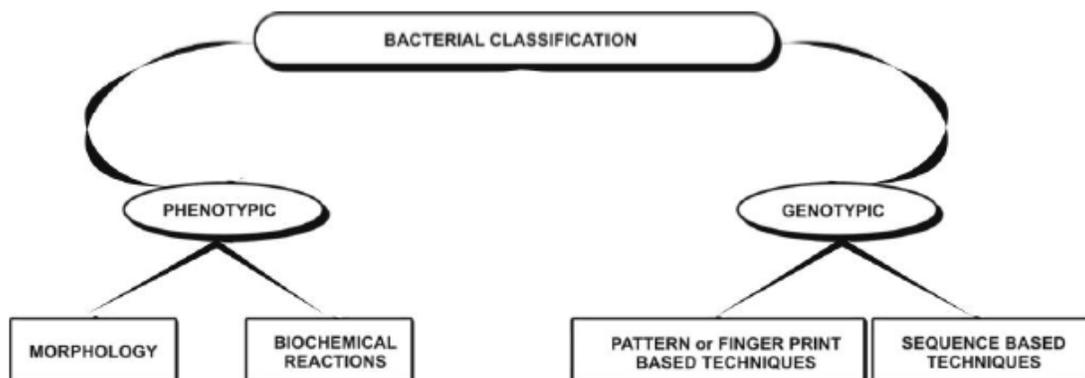


Figure 5 : Les aspects de la classification bactérienne [7].

I-8-Les techniques appliquées en taxonomie, leurs limites et leur pouvoir de résolution

I-8-1-L'étude du phénotype

Jusqu'à il y'a pas si longtemps, les années soixante, la composition des communautés procaryotiques a été uniquement accessible par des observations microscopiques directes et des méthodes culturales. L'identification phénotypique conventionnelle en bactériologie n'utilisait que la mise en évidence de certains caractères considérés comme critères importants pour l'identification des microorganismes, tels que la morphologie, l'habitat, la position écologique, la capacité a sporuler...etc. de même toutes les tentatives de construction taxonomique hiérarchisées étaient établies à base de ce type de caractères, c'est pour cela qu'on qualifie cette ancienne méthode d'identification comme demeure non concluante et insuffisante pour caractériser et différencier les espèces bactériennes [59].

Actuellement, diverses révisions et améliorations ont été apportées à l'identification phénotypique des bactéries, y compris l'étude des propriétés morphologiques qui peuvent inclure les caractéristiques de la colonie (couleur, forme, pigmentation...etc.). En outre, les caractéristiques de la cellule sont décrites en ce qui concerne la forme, la taille, la réaction de Gram, le matériel extracellulaire comme la capsule, la présence d'endospores, la présence et l'emplacement des flagelles, la motilité et les corps d'inclusion. Les propriétés physiologiques et biochimiques, via des techniques standardisées, décrivent la croissance de l'organisme à différentes plages de température, de pH, de salinité et de conditions atmosphériques, la croissance en présence d'agents antimicrobiens, la production de diverses enzymes et la croissance en présence de différentes sources uniques de carbone et d'azote en combinaison avec des propriétés supplémentaires de sérologie, des données de la production des toxines et des métabolites et des propriétés chimiotaxonomiques. Ces tests doivent être effectués en utilisant des procédures standardisées pour obtenir des résultats reproductibles afin que la valeur de l'identification phénotypique soit considérée comme significative et discriminatoire [58].

I-8-2-Les limites des approches phénotypiques

- Dépendent beaucoup des conditions de culture, des milieux de culture, les conditions de croissance et des procédures d'analyse qui doivent être normalisées. Elle ne convient qu'aux organismes cultivables.
- Il peut y avoir des propriétés morphologiques similaires pour des espèces non apparentées.

- Cette approche ne génère que des résultats préliminaires, des tests supplémentaires sont donc nécessaires pour identifier.
- L'identification est limitée aux micro-organismes faisant partie de la base de données. Les espèces ou les groupes taxonomiques qui ne font pas partie de la base de données ne seront pas identifiables.
- Nécessite des connaissances techniques, ainsi que des équipements coûteux.
- Pour certaines espèces bactériennes, des réactions de coloration spéciales sont nécessaires puisqu'elles ne répondent pas bien à la coloration de Gram.
- L'identification au niveau de l'espèce est incertaine voire impossible dû aux variations entre certaines souches. Dans le cas d'une mutation, il y aura généralement une incidence sur une propriété physiologique ou biochimique.
- La disponibilité et coût des anticorps peuvent affecter la sensibilité et à la spécificité des analyses immunologiques.
- L'expression d'une molécule antigénique est nécessaire pour les techniques sérologiques.
- Le choix des toxines ou des métabolites secondaires à rechercher devrait reposer sur le profil de toxines ou de métabolites secondaires documenté du groupe taxonomique.
- Les erreurs d'identification de nouveaux antigènes inconnus sont courantes [12].

I-8-3-l'étude du génotype

Actuellement, il existe des techniques nombreuses essentiellement moléculaires pour l'identification des bactéries en biologie moléculaire et d'autre domaine de recherche bactérienne. Les techniques moléculaires sont largement utilisées comme outils puissants pour l'identification des taxons remplaçant les méthodes traditionnelles qui présentent une certaine incertitude et instabilité [58].

Leur importance réside dans l'applicabilité universelle et développée, ce qui permet une caractérisation plus exhaustive et un pouvoir discriminatoire de la diversité microbienne, en effet, ces approches sont basées sur la spécificité du génome pour distinguer génétiquement les populations bactériennes [58].

L'approche convient à des pratiques qui dépendent de la plateforme et des méthodes utilisées pour le séquençage, y compris la couverture génomique appropriée pour l'identification des espèces et souches bactériennes. Cette analyse comparative de la séquence de génomes par application des différentes techniques comme hybridation ADN-ADN, typage par PCR et

techniques génétiques (RAPD, RFLP, PFGE, Ribotypage) et le séquençage des gènes marqueurs (l'*ARNr 16s*) permet d'identifier des gènes qui codent pour des éléments tels que des facteurs de virulence, enzymes, toxines, métabolites, résistances aux antibiotiques ou antifongiques, résistances aux stress abiotiques et biotiques, facteurs de survie et de persistance, etc. L'identification peut aller jusqu'au niveau de l'espèce et de la souche, elle est considérée comme identification significative et primaire aux microorganismes qui ne peuvent être identifiés au moyen de la culture classique ou par des méthodes de séquençage de gènes individuels [3].

I-8-4-Les limites des approches moléculaires

- Une grande quantité d'ADN pure et de haute qualité est requise, la préparation de la banque génomique et du séquençage, sont des processus fastidieux. Les résultats dépendent aussi de la taille du génome et des fois de la position des séquences nucléotidiques dans celui-ci.
- Les résultats d'analyse d'*ARNr16S* peuvent varier selon la qualité de la séquence d'intérêt. En effet, la multiplicité des copies, les variations inter géniques rendent nécessaire l'analyse d'une séquence consensus.
- Une mauvaise connaissance des séquences cibles et les mutations occasionnelles d'ADN peuvent compromette le processus.
- Cette approche requière une grande puissance des systèmes de réseau avancés sur le plan technique, de traitement de données et une expertise bio-informatique pour interpréter les résultats obtenus pour l'identification microbienne de même que pour mettre en évidence leur pertinence biologique et écologique.
- Les résultats dépendent de la richesse et de la qualité des bases de données de références génomiques utilisées ainsi que des méthodes alignement des séquences [3].

I-8-5-Le pouvoir discriminatoire des différentes techniques

Dans l'objectif de fournir un schéma complet et distinctif de la parenté des organismes, la caractérisation des procaryotes doit être analysée et décrite de manière exhaustive suivant les deux approches, phénotypique et génotypique, qui sont considérées les éléments clé de la taxonomie microbienne [46].

Les caractérisations phénotypiques sont essentielles pour la description et la différenciation microbienne aux niveaux les plus élevés des taxons tels du phylum, la classe, l'ordre, la famille et le genre (Figure 6), ainsi dans certains cas aux niveaux les plus bas comme l'espèce et sous-

espèce. Les approches employées ont inclus des descriptions de caractères morphologiques des colonies et des cellules, les conditions de croissance. Ces traits sont les points de départ pour les premières caractéristiques à être reconnues et enregistrées, cependant, de telles caractéristiques visibles ne permettent pas de description unique pour chacun des procaryotes, de plus, ne sont pas nécessairement corrélées avec les relations évolutives des procaryotes au point qu'il n'est pas possible d'extrapoler que les souches présentant des caractéristiques morphologiques différentes sont également plus disparates génétiquement que les souches présentes des caractéristiques similaires [46].

Les tests physiologiques sont généralement des profils de réactions caractéristiques pour l'utilisation de substrats ou les activités métaboliques de bases des procaryotes. Les applications utilisant des panels de tests physiologiques sont sensibles et sont souvent efficaces pour la discriminer les taxons bactériens les moins différenciables. Concernant la caractérisation chimique ou chimiotaxonomiques, elle exploite les différents composants structurels des cellules procaryotes ou du cytoplasme (les peptidoglycanes cellulaires, les acides gras, les lipides....etc.), l'analyse de ces caractéristiques permet des différenciations systématiques qui sont souvent en corrélation avec les relations évolutives des taxons. Le degré de la résolution est variable pour les membres de différents niveaux taxonomiques en dessous du genre [46].

Le développement et l'application des analyses basées sur les séquences d'ADN sont devenues des approches de choix pour la systématique des procaryotes, offrant un niveau plus haut de résolution phylogénique (Figure 6). L'utilisation des techniques de séquençage du génome, l'hybridation de l'AND, l'amplification PCR et de séquençage des gènes marqueurs en particulier les gènes de l'*ARNr 16S* permet l'estimation rapide des relations phylogénétiques jusqu'au niveau d'espèce, la technologie de séquençage de nouvelle génération offre des possibilités d'exploiter des degrés de résolution plus en plus élevés et fournit une caractérisation des procaryotes au niveau de l'espèce et la souche (Figure 6) [46].

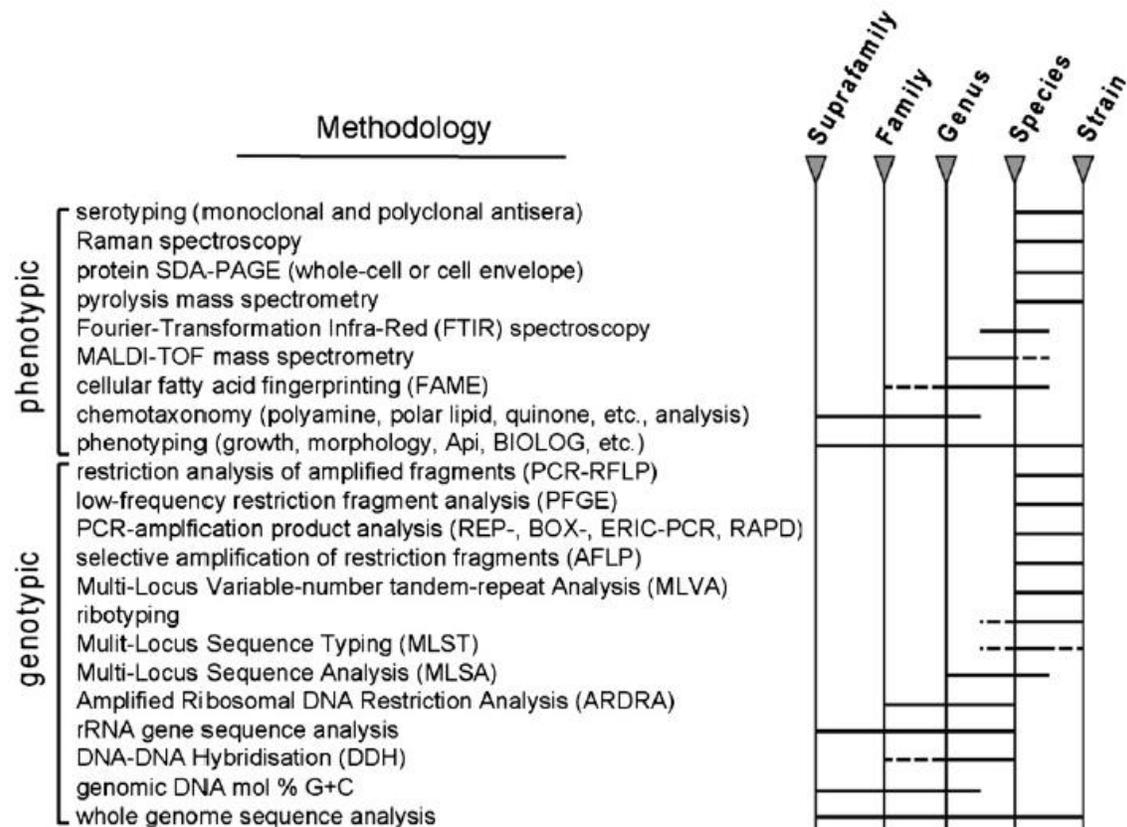


Figure 6: Méthodologies pour la caractérisation des procaryotes et les niveaux de résolution taxonomiques respectifs approximatifs [46].

I-9-Importance du séquençage d'*ARNr 16S*

Le séquençage du gène de l'*ARNr 16S* des bactéries est une technique standard pour l'étude phylogénétique et taxonomique des bactéries. La séquence de ce gène est présente dans toutes les espèces bactériennes, fonctionnellement constante, et est composée de région conservées et variables, constituant ainsi un bon marqueur taxonomique pour dériver des relations taxonomiques. Le gène de l'*ARNr 16S* code pour l'ARN ribosomique 16S, il forme, en association avec des protéines, la petite sous-unité 30S du ribosome bactérien qui est constituée de neuf régions variables entourées de régions conservées. Ce gène présente une taille relativement constante de 1378 nucléotides avec un écart-type de 57 nucléotides, cette valeur est calculée sur l'intégralité des séquences complètes des gènes de l'*ARNr 16S* présentes dans GreenGeens 13.5. McDonald *et al*, 2012 [39].

L'analyse du gène de l'*ARNr 16S* dans le cadre d'une taxonomie bactérienne produit des valeurs numériques de similitudes qui peuvent être utilisées comme limites de circonscription pour les taxons et impose le dépôt de la séquence du gène lors de chaque description d'une nouvelle espèce bactérienne. Les organismes présentant des similarités génomiques supérieures à 70% partagent un seuil de 97% de similarités entre les séquences du gène de l'*ARNr 16S*, ce qui correspond à une différence maximum de 45 nucléotides entre les gènes, cette valeur pourrait obtenir des indications sur la possibilité d'appartenance ou non de deux organismes à une même espèce. Cependant, l'*ARNr 16S* manque de pouvoir de résolution au niveau de l'espèce ; ainsi, il existe des espèces différentes avec des séquences d'*ARNr 16S* identiques ou presque. Ce n'est que si les gènes présentent plus de 97% de similarité de séquence qu'il est alors nécessaire de faire appel à l'hybridation ADN/ADN afin de confirmer un taux d'hybridation supérieur ou égal à 70% entre les deux organismes et donc de valider que ceux-ci appartiennent à la même espèce. L'analyse comparative de l'*ARNr 16S* est une très bonne méthode pour une première affiliation phylogénétique d'organismes potentiellement nouveaux et mal classés. En raison de l'avantage pratique de cette approche à des fins d'identification, il est recommandé d'inclure la séquence ribosomique dans les nouvelles descriptions d'espèces procaryotes [39, 57].

L'analyse de la séquence d'*ARNr 16S* a introduit un nouveau paramètre qui a une énorme influence sur la classification des procaryotes, l'application de cette approche a permis la détermination des relations entre les bactéries apparentées à distance. Le séquençage de l'*ARNr 16S* a dominé la possibilité de dessiner les arbres phylogénétiques qui représentent des lignées de descendance, et produit des valeurs numériques de similarités particulièrement utiles pour les classes au-dessus du niveau de l'espèce [57].

Chapitre II :
Identification
d'un bacille à
Gram positif,
cas particulier
des Bacillus

Le genre *Bacillus* appartient au phylum des Firmicutes, à la classe des Bacilli, à l'ordre des Bacillales et la famille des Bacillaceae. Les *Bacillus* sont des bactéries ubiquitaires vivants dans le sol, l'eau, l'air, les plantes et certains aliments fermentés. Ce sont des bactéries à Gram positif en forme de bâtonnets, souvent aérobie ou anaérobie facultative, formant des spores de résistance et très souvent mobiles. Ce genre est extrêmement hétérogène sur le plan génétique et phénotypique, notamment au niveau respiratoire et métabolique [29, 44].

Les Bacillaceae ont été classifiés en utilisant des approches phénotypiques telles que les caractères morphologiques, physiologiques (composition des cellules), biochimiques (test API) et nutritionnels et d'autres approches moléculaires basées sur l'hybridation ADN-ADN et la composition d'ADN. Cependant, ces approches n'étaient pas concluantes à cause de la très grande hétérogénéité des espèces à l'intérieur du genre, notamment aux niveaux physiologique et écologique, de ce fait, la classification des Bacillaceae est orienté vers l'analyse des séquences du gène 16S afin de construire des arbres phylogénétiques des Bacillaceae à partir de la séquence d'une seule copie du gène 16S par espèce [47].

L'identification des bactéries appartenant au genre *Bacillus* est souvent très importante dans plusieurs domaines en biotechnologies, en industrie, en médecine et en bioremédiation. Certaines souches des Bacillaceae sont non pathogènes pour l'homme, d'autres provoquent des infections et peuvent parfois causer la mort. Ces bactéries sont connues pour produire des molécules qui ont diverses applications industrielles, notamment la production industrielle d'enzymes telles que les amylases utilisées dans l'industrie de panification, les protéases intra et extracellulaires et les cellulases employées dans les industries agroalimentaires et des détergents. Les protéases sont généralement utilisées dans la transformation des produits alimentaires et pharmaceutiques, dans les industries de tannage de cuir et dans la gestion des déchets. Les *Bacillus* sont également impliquées dans la production de métabolites primaires tels que les vitamines et les ribonucléosides et de même que des métabolites secondaires comme les bactériocines et les biosurfactants. En effet, les biosurfactants produits par les bactéries du genre *Bacillus* sont utilisés dans l'industrie pétrolière, pharmaceutique et alimentaire [29].

Certaines bactéries du genre *Bacillus* possèdent des propriétés probiotiques et la conservation des aliments, sont également utilisées par le secteur agronomique dans la lutte biologique, en parallèle l'industrie pharmaceutique exploite la capacité de ces bactéries à produire des antibiotiques, comme la bacitracine, pour fabriquer des médicaments destinés à lutter contre des

maladies mortelles. Le genre *Bacillus* présente un intérêt considérable en recherches dans la compréhension de sa diversité physiologique, sa phylogénie avec les autres espèces qui se trouve principalement dans le sol et les plantes, certaines de ces bactéries possèdent un pouvoir de la dépollution de l'environnement [29].

II-1- Une approche polyphasique pour l'identification d'un *Bacillus*

II -1-1-La mise en culture et la purification de la bactérie

Les milieux de cultures usuels permettent la croissance de la plupart des espèces du genre *Bacillus*. Ces bactéries poussent à des températures de 28-33°C ainsi qu'à 37 °C, et beaucoup d'espèces tolèrent des différences thermiques marquées. Les espèces du genre *Bacillus* ont chacune leurs exigences propres et sont souvent relativement homogènes. Les milieux favorables utiles pour l'identification sont : la gélose nutritive à pH 6, le bouillon hyper salé, le milieu Mossel (sélectif), le bouillon lysogène (utilisé pour la purification et la conservation) [1, 41].

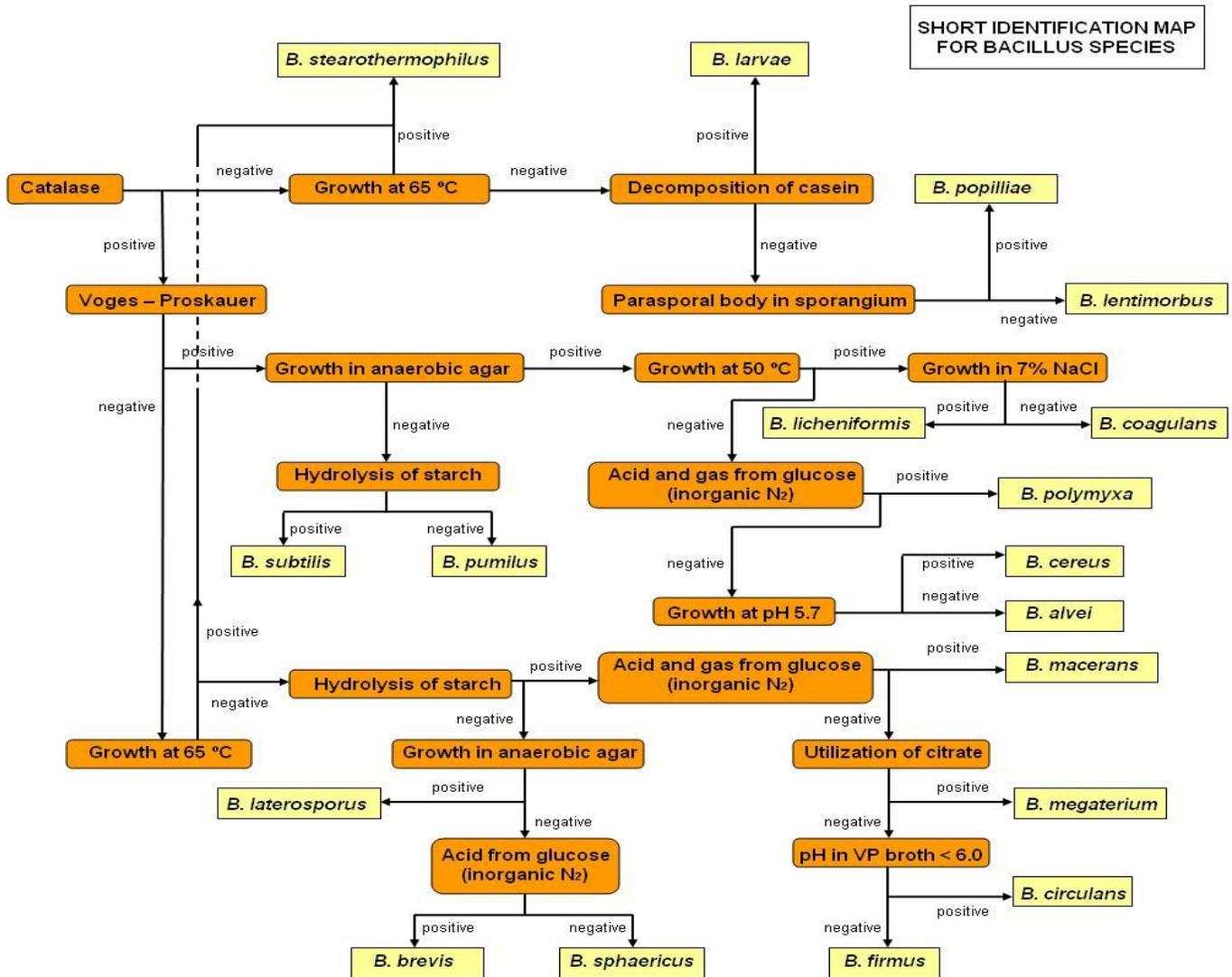
L'isolement se fait par étalement sur milieu Mossel et l'incubation à 37 °C pendant 24 heures en aérobiose. La purification des isolats se fait par repiquage sur le bouillon lysogène puis sur gélose en utilisant la technique des quadrants jusqu'à l'obtention des colonies identiques [29].

II -1-2-Les caractères phénotypiques recherchés pour identifier le genre *Bacillus*

La caractérisation phénotypique consiste à une différenciation basée sur des critères macroscopiques observés directement à l'œil nu de l'aspect morphologique, la taille, la forme, la pigmentation et la consistance des colonies préalablement incubées sur un milieu solide, et sur des critères microscopiques tels que la mobilité, la coloration de Gram et la coloration des spores. La caractérisation biochimique et physiologique consiste à une mise en évidence du type respiratoire, des enzymes respiratoires (la catalase, les nitrates réductases), et la dégradation de certains substrats ainsi que l'étude de la résistance à différentes températures [1, 41].

L'étude physiologique et biochimique ainsi que les tests de résistances sont très variables d'une souche à l'autre. Sur le plan biochimique, les *Bacillus* sont souvent positives au test de catalase et pour certaines à l'oxydase [41]. Les résultats obtenus de ces tests permettent de choisir une galerie classique des caractères biochimiques et activités enzymatiques qui vont pousser l'identification jusqu'au genre. L'étude physiologique et biochimique concernent la production de la catalase, la réduction des nitrates en nitrites, le type respiratoire ainsi que la résistance à différentes températures sont des propriétés étaient très variables de genre, famille et même au sein de l'espèce (Figure 7). Ce qu'exige une identification génétique par les technologies de l'ADN

utilisant le gène conservé et codant pour l'ARNr 16S permettant de constater une particularité diversifiée des bactéries appartenant au genre *Bacillus* [1].



Based on Gordon R.E., Haynes W.C., Pang C.H. (1973) – The genus *Bacillus*. Agriculture Handbook No. 427, U.S.D.A., Washington D.C.
 Since 1973 the following changes occurred:
Bacillus alvei => *Paenibacillus alvei*
Bacillus polymyxa => *Paenibacillus polymyxa*
Bacillus macerans => *Paenibacillus macerans*
Bacillus larvae => *Paenibacillus larvae*
Bacillus lentimorbus => *Paenibacillus lentimorbus*
Bacillus laterosporus => *Brevibacillus laterosporus*
Bacillus brevis => *Brevibacillus brevis*
Bacillus sphaericus => *Lysinibacillus sphaericus*
Bacillus stearothermophilus => *Geobacillus stearothermophilus*

(c) Costin Stoica tgw1916.net

Figure 7 : Clé dichotomique d'identification des espèces de bacilles [56].

II -2-1-Caractérisation morphologique des *Bacillus*

L'identification morphologique des isolats a d'abord consisté en un examen de l'aspect macroscopique par observation directe à l'œil nu des colonies sur l'un des milieux précités ou sur gélose au sang pour l'activité hémolytique des *Bacillus*, puis l'aspect microscopique à l'état frais ou suite à une coloration de Gram. Quant aux spores de résistance, elles sont observées après leur coloration spécifique au vert de malachite [29].

Les espèces de *Bacillus* présentent un très large éventail de morphologies des colonies, à la fois au sein des espèces et entre elles, et bien sûr la composition du milieu et d'autres conditions d'incubation ont une forte influence. En général, les colonies observées sur milieu solide sont larges, lisses, plates et fréquemment circulaires avec des bords irréguliers de couleur crème ou blanc ou encore jaunâtre. Cependant, certaines espèces ont des morphologies caractéristiques mais variables, dans le cas de *Bacillus cereus* (Figure 8a, b, h), les colonies sont caractéristiquement grande (2–7 mm de diamètre) et de forme variable, de circulaire à irrégulière, avec des bords entiers à ondulés, crénelés ou fimbriés, elles ont des textures mates ou granuleuses, mais les colonies lisses et humides ne sont pas rares. Pour les colonies de *Bacillus anthracis*, sont généralement plus petites, non hémolytiques, peuvent présenter plus de pointes ou de traînées le long des lignes d'inoculation et sont très tenaces par rapport aux colonies généralement plus butyreuses (Figure 8a, f, h, i). Les colonies de *Bacillus mycoides* diffèrent de celles des autres membres du groupe *Bacillus cereus*, ils sont typiquement rhizoïdes ou velus et adhérents, et ils couvrent facilement toute la surface de la gélose. Les colonies d'autres espèces varient de humides et brillantes passant de granuleux à ridé (Figure 8d), les formes varient de rondes à irrégulières, parfois étalées, avec des bords entiers ondulés ou crénelés à fimbriés, la couleur varie généralement du chamois ou du gris crème au blanc cassé, mais des souches occasionnelles peuvent produire des pigments noirs, bruns, orange, roses ou jaunes ; une telle pigmentation tend à être caractéristique des espèces ou des sous-espèces [51].

Les cellules en forme de bâtonnet, droites ou légèrement incurvées, apparaissant seules et par paires, certaines en chaînes, et parfois sous forme de longs filaments, Gram-positif, ou Gram-positif uniquement aux premiers stades de la croissance, mobile au moyen de flagelles péritriche ou dégénérés péritriche, ou non mobiles. Les endospores sont formées de manière intracellulaire et qu'elles diffèrent de leurs cellules végétatives parentales, elles sont optiquement rétractiles et sont très résistantes aux stress chimiques et physiques qui sont mortels pour les cellules

végétatives. Ces propriétés sont conférées par la composition chimique et l'ultrastructure spéciales des spores. La formation de spores est la plus importante dans l'identification au niveau du genre. Avant d'essayer d'identifier au niveau de l'espèce, il est important d'établir que l'isolat est vraiment un formateur d'endospores aérobie et que les autres inclusions ne sont pas confondues avec des spores [51].

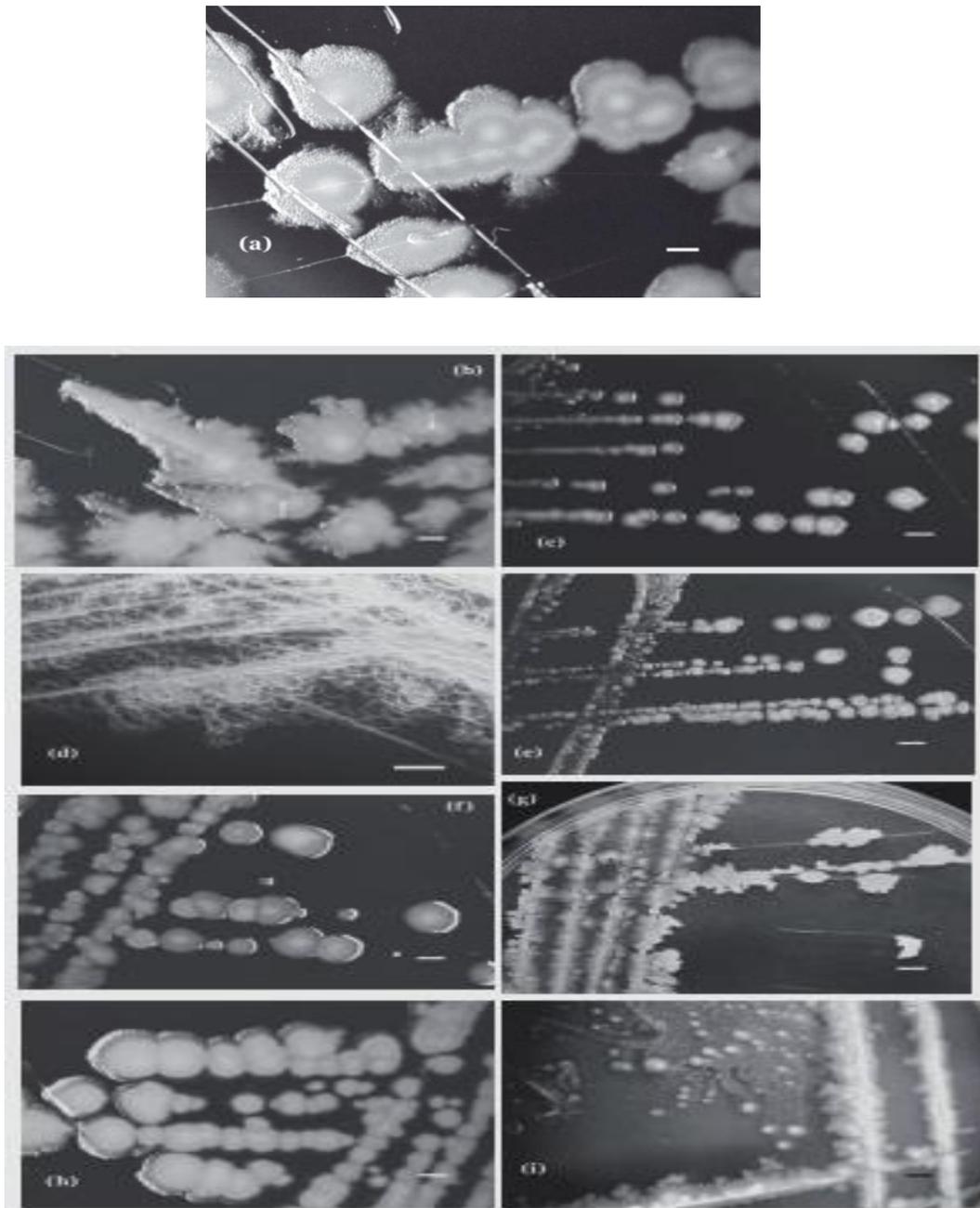


Figure 8: La diversité de l'apparence des colonies au sein du genre *Bacillus* [51].

II -2-2-Caractérisation biochimique et physiologique des *Bacillus*

Il s'agit de plusieurs tests :

Le test de catalase consiste à mettre en suspension la colonie dans une goutte d'eau oxygénée déposée sur une lame de verre propre. Un résultat positif se traduit par un dégagement des bulles gazeuses [41].

Le test d'oxydase se fait par disposition d'une colonie sur un disque pré-imprégné de réactif NN-diméthyl-paraphénylène diamine. Le résultat positif se traduit par l'apparition d'une couleur rose violacé [41].

Le test de type respiratoire est réalisé par un ensemencement dans un milieu viande foie (VF) régénéré durant 30 minutes au bain marie par piquage d'une pipette pasteur prolongée au fond du tube et puis remontée d'une façon spirale, l'incubation à 37 °C pendant 24 à 48 heures. L'identification biochimique est aussi menée par une galerie classique qui regroupe des caractères biochimiques tels que du glucose, lactose, citrate, mannitol-mobilité-nitrate, l'urée indole...etc. [1, 41]. Le genre *Bacillus* possède un profil des enzymes de dégradation souvent utilisé pour la production de métabolites et d'enzymes par fermentation. Ceci est en partie dû au fait que la plupart sont d'excellents sécréteurs de protéines et de métabolites. Les souches de *Bacillus* identifiées couramment sont utilisées à la fois dans la transformation des aliments et la fermentation industrielle. Les amylases sont utilisées pour la production de glucose, ce processus enzymatique est donc devenu important pour la production industrielle de sucre à partir d'amidon, soit comme substrat pour la fermentation ultérieure en éthanol, soit comme agent édulcorant dans les boissons non alcoolisées et autres aliments. Dans le cas des *Bacillus subtilis*, cette bactérie est utilisée dans l'industrie pour produire la subtilisine, protéase produite par ce germe et quelques autres *Bacillus* apparentés. Cette protéase est très stable, et résiste bien à l'action des détergents [29, 44].

La phospholipase C, autrement appelée lecithinase est aussi recherchée chez certaines espèces du genre *Bacillus*. C'est une enzyme qui sépare la lecithine phospholipidique. Une telle activité de lecithinase dégrade la lecithine ; une substance largement répandue dans les tissus animaux, le jaune d'œuf et certaines plantes supérieures, consistant en phospholipides liés à la choline. Sa détection se fait par une culture sur gélose au jaune d'œuf, et les colonies des microorganismes la possédants auront un précipité blanc opaque autour d'elles [45].

La résistance à la variation des températures est aussi variable entre les souches, les maximum et minimum peuvent être étendus exceptionnellement chauds ou froids au-delà des limites habituelles entre les espèces du genre [51].

II -3-la biologie moléculaire et la phylogénie des *Bacillus*

Les premières classifications phylogénétiques du genre *Bacillus* étaient basées sur les propriétés morphologiques et physiologiques. Couramment, l'apparition de l'approche basée sur l'analyse d'une seule copie du gène de l'*ARNr 16S* demeure l'une des approches d'identification et de classification les plus utilisées. Le genre *Bacillus* est à ce jour le genre le plus représenté sur les bases de données génomiques. L'analyse des gènes conservés à travers le niveau taxonomique éclaire les relations évolutives et fonctionnelles entre les espèces proches de *Bacillus*. Ainsi pour mieux comprendre l'histoire évolutive des bactéries du genre *Bacillus*, plusieurs reconstructions phylogénétiques ont été faites. Les marqueurs traditionnels tels que l'opéron ribosomal: L'*ARNr 16s*, *23s* et *5s*, spécialement le gène de l'*ARNr 16S*, et les gènes de ménage [29, 51].

L'opéron ribosomal est un bon marqueur moléculaire dont le nombre est proportionnel à la taille et à la vitesse de croissance de la bactérie, fréquemment utilisé pour le traçage des rapports génétiques et même l'identification des nouvelles souches. Chez la famille des Bacillaceae, son nombre varie entre 8 et 13 copies [47].

Les gènes de l'*ARNr 16S* ont permis de construire des arbres phylogénétiques des bactéries du genre *Bacillus* issues de différentes niches écologiques. Leurs séquences sont très utilisées en phylogénie pour reconstruire l'histoire évolutive des organismes dans la mesure où sa vitesse d'évolution relativement lente permet d'établir d'anciennes divergences génétiques. Le séquençage partiel du gène de l'*ARNr 16S* permet l'identification rapide des bactéries du genre *Bacillus*. Cependant, ce critère est insuffisant pour certaines espèces, au sein de la famille des Bacillaceae, il présente une similitude de l'ordre de 99% car ce gène est hautement conservé. Les gènes *23s* et *5s* sont également utilisés dans la distinction et la différenciation des espèces notamment les plus proches pour lesquelles l'étude du gène *16S* ne permet pas d'en retenir des conclusions quant à leur classification phylogénétique [29, 47].

Les gènes de ménage existent pratiquement chez tous les microorganismes, particulièrement chez les *Bacillus*, ils permettent d'obtenir un résultat rapide et fiable dans l'identification des souches très proches, Ces gènes assurent des fonctions indispensables à la vie de tous les types de cellules et présentent une grande diversité intra-spécifique. Parmi ces gènes

on peut citer : les gènes *rpoB*, *gyrB*, *recA*, cette caractéristique a fait que ces gènes sont considérés comme outil de choix dans le typage moléculaire pour l'identification bactérienne [29].

II-4- Démarche pour l'analyse d'une copie de l'ADN*r16s* chez une bactérie inconnue

II -4-1-La réaction de polymérisation en chaine (PCR)

L'analyse de l'ADN récupéré lors de l'extraction requière une amplification car la quantité extraite d'un échantillon est très faible pour être considérée comme un échantillonnage représentatif. La technique d'amplification par la réaction de polymérisation en chaine (PCR) a été mise au point en 1986 par Mullis, cette technique permet d'amplifier un fragment d'ADN en utilisant le mélange réactionnel contenant des amorces spécifiques et une enzyme de polymérisation en présence de dNTP (dATP, dCTP, dTTP, dGTP) et de tampon. La réaction est faite dans un cycle thermique où la température d'hybridation et le nombre de cycles sont ajustés selon la longueur de l'amplifié le T_m (melting temperature) et la nature des amorces utilisées [3].

La PCR sur colonie

Lorsqu'on travaille sur les colonies isolées de boîtes, il est possible de réaliser directement les PCR à partir de la biomasse relativement fraîche présente à la surface de la gélose. Cette méthode est donc un outil puissant pour cribler rapidement et facilement un nombre potentiellement important de colonies et plus couramment utilisée pour l'identification. En effet, les travaux de Saris et ses collaborateurs en 1990 ont indiqué que malgré l'épaisseur des parois cellulaires chez *B. subtilis* et leur résistance importante à la lyse induite par la chaleur, ils lisent encore assez au cours de l'étape de dénaturation dans la PCR [52].

Dans la plus part des cas une seule colonie remplace l'ADN matrice pour l'amplification sans qu'il soit nécessaire de préparer de l'ADN pur, une toute petite quantité de biomasse est collectée puis directement mise en suspension dans le mélange réactionnel utilisé pour la PCR. Après amplification, une visualisation des produits de la PCR est effectuée par électrophorèse sur gel d'agarose [33, 34]. Dans certains cas, il faut avoir recours à l'extraction d'ADN à partir de la colonie, pour faciliter la libération d'ADN des cellules bactériennes [4].

➤ **Etape 1** : la réaction de polymérisation en chaine (PCR)

Une réaction PCR standard est utilisée pour amplifier le produit d'intérêt de l'ADN. Une colonie est grattée à partir d'une culture sur gélose, avec la pointe de pipette et dispersés directement dans le mélange réactionnel contenant des amorces spécifiques et une enzyme de

polymérisation en présence des désoxyribonucléotides ou dNTP (dATP, dCTP, dTTP, dGTP) et de tampon à température d'hybridation ajustée selon le cycle thermique [4].

➤ **Etape 2 : la visualisation du gel électrophorétique**

Les produits de la PCR générés sont maintenant séparés par électrophorèse selon leurs tailles moléculaires sur un gel d'agarose contenant un colorant intercalant l'ADN tel que le bromure d'éthidium [4].

II-4-2- Le séquençage

La deuxième moitié des années 1970 est marquée par l'invention de séquençage d'ADN. Cette technique a pour but de déterminer la succession linéaire des bases nucléotidiques A/C/G/T constituant la structure de l'ADN ce qui permet la lecture des données spécifiques de la structure d'ADN chez chaque individu [3, 36].

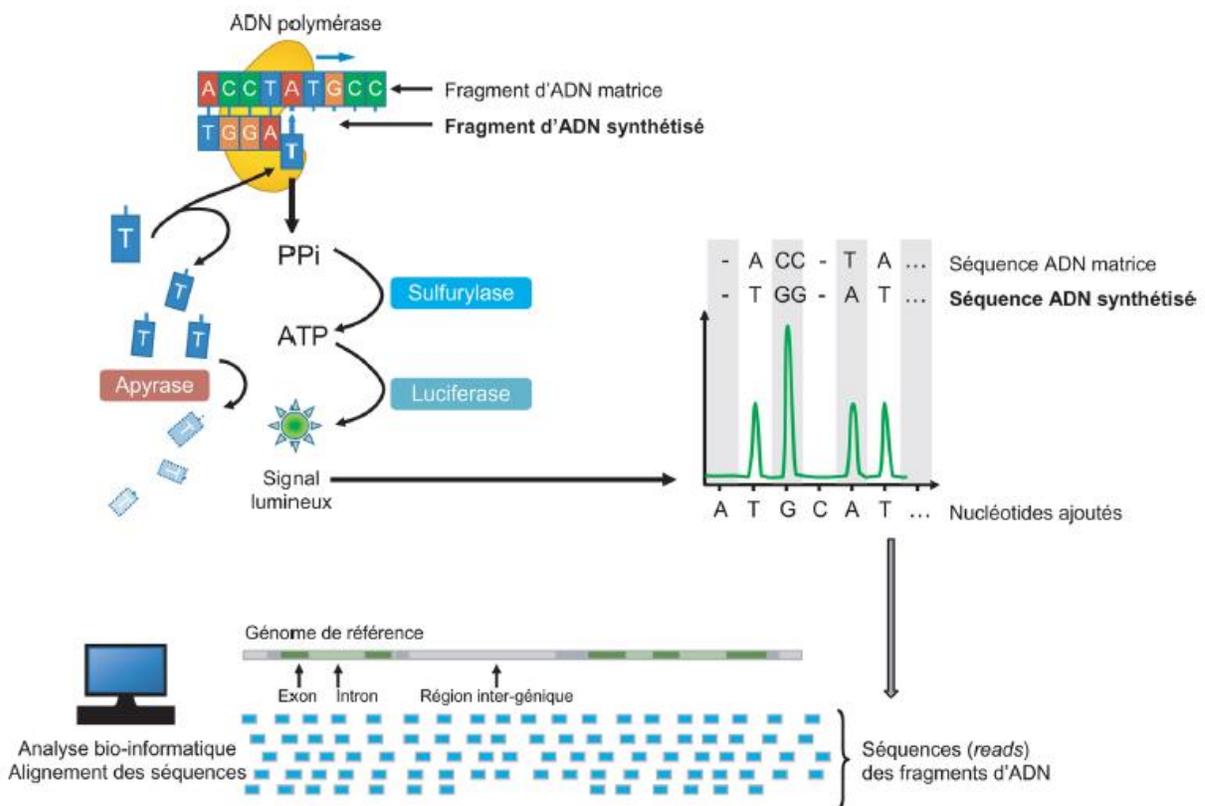


Figure 9 : Le séquençage d'ADN [36].

La technique de séquençage nécessite une fragmentation de l'ADN génomique en séquences de quelques centaines de bases qui sont ensuite amplifiées par la PCR, dans un milieu

réactionnel contenant des désoxyribonucléotides et une ADN polymérase. Le brin d'ADN complémentaire est synthétisé du fragment à séquencer (Figure 9), il s'agit d'un apport successif des nucléotides A, C, G, T, les fragments de séquences sont ensuite réalignés dans un génome de référence à l'aide d'outils bio-informatiques [36].

Le séquençage a été introduit à deux méthodes développées indépendamment, l'une par l'équipe de Walter Gilbert et Maxam en 1977 fondée sur le principe de dégradation chimique sélective, et l'autre par Frederick Sanger en 1977 fondée sur le principe de la synthèse enzymatique sélective [3].

Le séquençage d'ADNr 16s par la technique de Sanger

La méthode de Sanger (également connue sous le nom de méthode de terminaison de chaîne) est basée sur l'interruption de la synthèse enzymatique d'un brin d'ADN complémentaire (Figure 11) [3], voici les étapes à suivre en abrégé (Figure 10):

Étape 1 : PCR d'ADN cible et dénaturation en simples brins.

Milieu réactionnel contenant :

Amorces oligonucléotidiques, tampons, ADN polymérase, désoxynucléotides triphosphates (dNTP), didésoxynucléotides triphosphates (ddNTP) marqués à l'aide de fluorophores.



Étape 2 : synthèse du brin d'ADN complémentaire initiée par l'amorce et arrêt de l'élongation par attachement de didésoxynucléotides. La longueur des fragments varie selon le point d'attachement et chacun porte un fluorophore terminal.



Étape 3 : séparation des fragments par électrophorèse capillaire sur un automate de séquençage, selon leur taille moléculaire.



Étape 4 : détection de signaux lumineux lors de la migration des fragments et analyse par logiciels générant un électrophorégramme de la séquence.

Figure 10 : Les étapes de la technique de Sanger [3, 36].

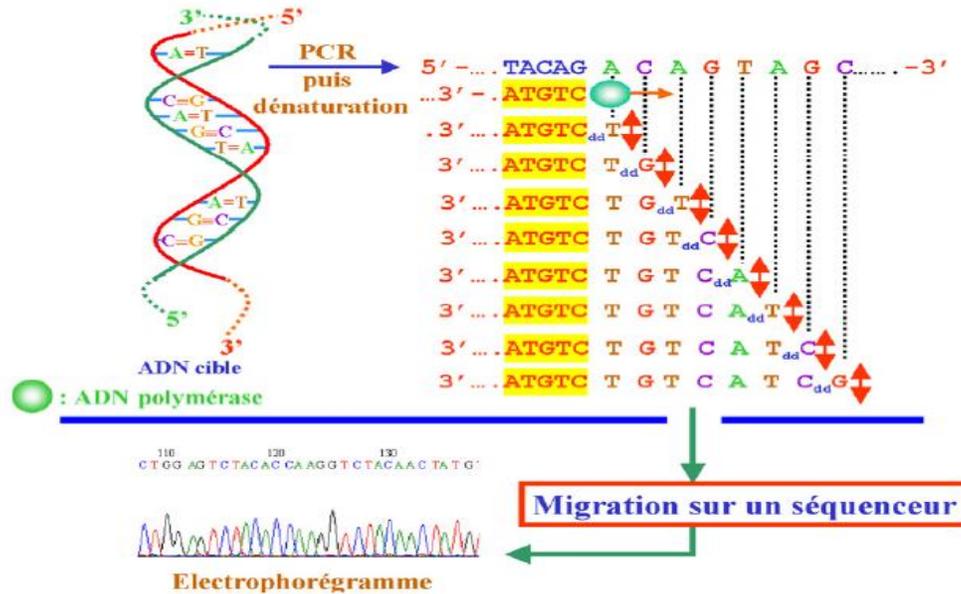


Figure 11 : La technique de séquençage Sanger [36].

II-5- L'analyse bioinformatique des séquences

Les techniques de séquençage génomique produisent une énorme quantité de données nécessitant des outils puissants de traitement et d'analyse. Avec la naissance des banques de séquences génomiques (EMBL, TIGR et GenBank) au début des années 80, la discipline de la bioinformatique est apparue. Elle propose des algorithmes qui permettent d'exploiter (manipuler, stocker, visualiser et analyser) l'information sur les séquences nucléotidiques et protéiques, via des logiciels informatiques [47].

La définition de plusieurs termes employés en bioinformatique est nécessaire pour la compréhension des différentes étapes d'analyse. Voici un aperçu de quelques notions.

II-5-1- Les bases de données

Il existe dans le domaine de la taxonomie plusieurs bases de données utiles (tableau 1). Il s'agit de collections de données structurées et organisées constituant une source des grandes quantités d'information génomique indispensable pour la communauté scientifique. Ces informations sont rendues publiques, afin d'en faciliter leur recherche, leur utilisation, le téléchargement, l'ajout, la mise à jour et éventuellement leur analyse dans les systèmes les plus évalués pour la systématique, [5]. Deux exemples sont décrits ci-dessous.

Tableau 1 : Ressources de taxonomie et nomenclature en ligne [26].

Name of resource	Tax/Nom ^a	Type (16S, G,M) ^b	Taxonomic scope	Number of sequences in current release	Year commenced	Hyperlink to resource	References
RDP	Tax	16S/28S	Bacteria, Archaea, Fungi	RDP Release 11 3,356,809 16S 125,525 28S	1992	https://rdp.cme.msu.edu/	[123, 124]
SILVA	Tax	16S/18S, 23S/28S	Bacteria, Archaea, Eukaryotes	Silva SSU/LSU 132 6,073,181 SSU 907,382 LSU	2008	https://www.arb-silva.de/	[125–127]
EzBioCloud	Tax	16S, G	Bacteria, Archaea	Aug 06, 2019 81,189 taxa 64,416 16S 146,704 genomes	2010	https://www.ezbiocloud.net/	[128]
Greengenes	Tax	16S	Bacteria, Archaea	Out of commission	2006–2013	https://greengenes.secondgenome.com/	[129, 130]
MIDAS	Tax	16S	Bacteria, Archaea	Jun-2020 4,245 species	2015	http://www.midastfieldguide.org/	[131, 132]
NCBI	Tax	16S,G	Bacteria, Archaea, Eukaryotes, Metazoa, Viridiplantae, Viruses	Jun-2020 905,918 species	1993	https://www.ncbi.nlm.nih.gov/taxonomy	[133]
IJSEM	Nom	16S,G,M	Bacteria, Archaea		1951	https://www.microbiologyresearch.org/content/journal/ijsem	[5]
LPSN/DSMZ	Nom	16S,G	Bacteria, Archaea	May-2020 18,678 16S, 77,990 strain deposits	1997	https://lpsn.dsmz.de/	[112, 137, 138]
Namesforlife	Tax, Nom	16S,G	Bacteria, Archaea	Sep-2019 16,335 16S, 10,877 genomes	2004	https://www.namesforlife.com/	[139]

^aTax (Taxonomy), Nom (Nomenclature).

^b16S/18S/23S/28S (16S/18S/23S/28S rRNA gene), G (Genome), M (Morphology).

➤ Le National Center for Biotechnology Information (NCBI)

Le National Center for Biotechnology Information est une base des données créé en 1988 fournissant des ressources d'analyse et de récupération pour les données dans GenBank et d'autres données biologiques faites disponible sur le site Web du NCBI (Figure 12). La base reçoit des données grâce à la collaboration internationale avec la base de données ADN du Japon (DDBJ) et la base de données de séquences de nucléotides du Laboratoire Européen de Biologie Moléculaire (EMBL-Bank) ainsi que de la communauté scientifique. Le NCBI peut être utilisé pour afficher l'arbre de taxonomie ou récupérer des données à partir de l'une des bases de données pour un organisme ou un groupe particulier [49].

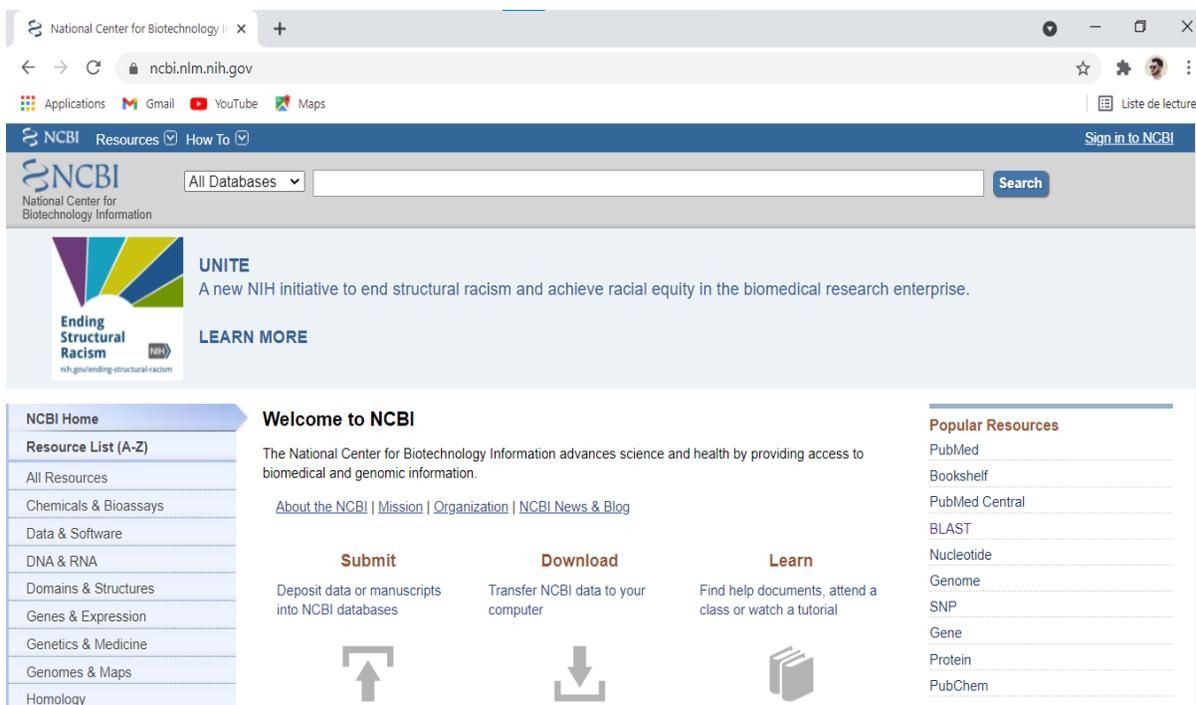


Figure 12: Page web de National Center for Biotechnology Information (NCBI) [48].

➤ EzBioCloud

C'est une construction d'une base de données du génome (Figure 13) contenant la hiérarchie taxonomique des bactéries et des archées représentée par des séquences de gènes d'*ARNr 16S* et de génomes de qualité contrôlée. Les technologies récentes de séquençage de l'ADN facilitent l'utilisation des données de séquençage du génome qui fournissent des moyens pour une classification et une identification plus informatives et précises des membres des bactéries et des archées. En conséquence, la base de données est constituée de 61 700 espèces /phylotypes, dont 13 132 avec des noms validement publiés, et 62 362 assemblages de génomes entiers qui ont été identifiés taxonomiquement au niveau du genre, de l'espèce et de la sous-espèce. Cette base des données permet de renforcer les efforts dans l'exploration de la diversité procaryote et la découverte de nouvelles espèces ainsi que pour les identifications de routine [64].

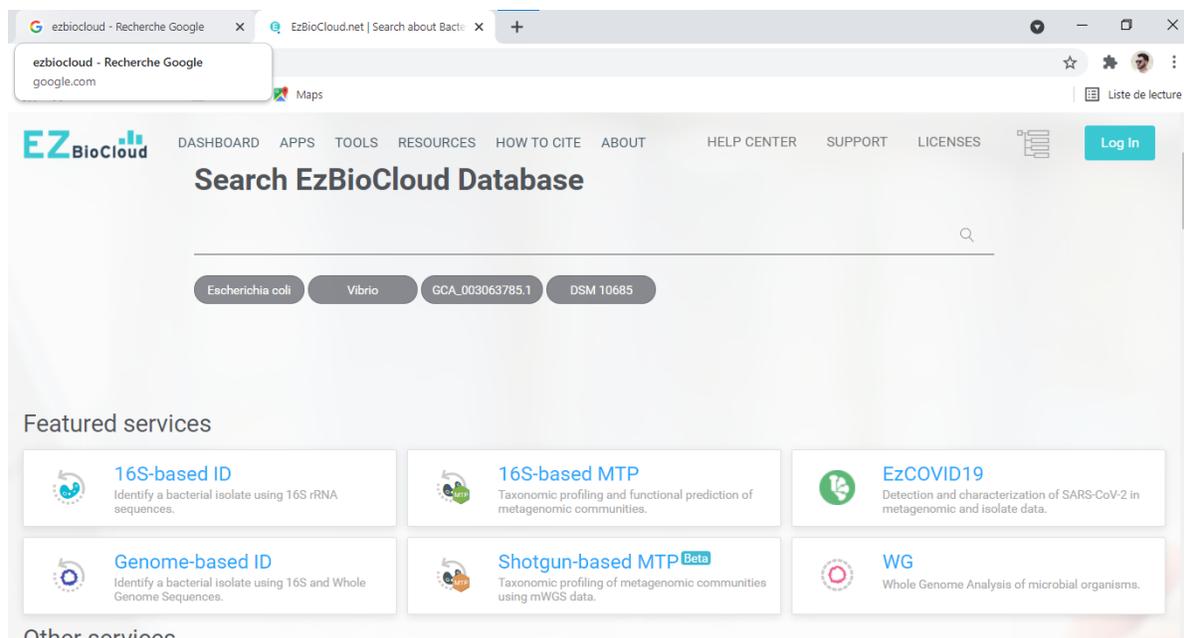


Figure 13 : Page web d'EzBioCloud [21].

II-5-2- L'alignement de séquences

Pour être comparées, les séquences nucléotidiques ou protéiques obtenues pour les différents individus doivent d'abord être alignées. Cet alignement pose l'hypothèse d'homologie primaire permettant de comparer des caractères comparables. Des logiciels tels que BLAST, accessibles gratuitement en ligne peuvent réaliser l'alignement de ces différentes séquences. Ces programmes effectuent des alignements globaux, c'est à dire réalisés sur toute la longueur de la séquence, une fois l'alignement réalisé, l'hypothèse d'homologie primaire est posée et l'hypothèse secondaire ou ascendante, c'est-à-dire trouver les ancêtres communs et construire l'arbre, peut être abordée [16, 50].

- **FASTA**

Le format FASTA est utilisé pour représenter les séquences en bio-informatique et sont requis par plusieurs programmes, un fichier FASTA contient une ou plusieurs séquences nucléotidiques ou peptidiques, la première ligne dans le fichier est une ligne de commentaire, commençant par ">" et décrivant la séquence suivi d'une entête contenant normalement le nom de la séquence et les informations complémentaires qu'on veut y ajouter. Toutes les lignes suivantes constituent la séquence [16].

➤ **Exemple de séquence d'ADN en format FASTA [6].**

```
>gi|23423|ref|NM_23542.0| Homo sapiens protéine  
ATGAATCGATACGATAGCTAGCTATCGATGCA  
GATCAGAGAGGGGCTTTAGCTAGCTAAGCTAG
```

• **BLAST**

BLAST « Basic Local Alignment Search Tool », est un outil de recherche en ligne fourni par le NCBI couramment utilisé pour trouver des régions d'homologie entre des différentes séquences biologiques (nucléotide ou protéine). L'algorithme de recherche est à la fois rapide et sensible, en l'utilisant pour saisir une séquence de gènes d'intérêt qui sera comparée à une banque de séquences nucléotidiques ou protéiques. Le NCBI détient une énorme base de données de séquences biologiques, à laquelle sont comparées les séquences de requêtes afin de trouver les plus similaires [16, 50].

Nucleotide BLAST fait référence à l'utilisation d'un membre de la suite des programmes BLAST. BLASTn est un programme de recherche et d'alignement de nucléotides à usage général qui est sensible et peut être utilisé pour aligner des séquences d'ARNt ou d'ARNr ainsi que des séquences d'ARNm ou d'ADN génomique contenant un mélange de régions codantes et non codantes [62].

La quantité d'informations sur l'interface de BLAST est un peu écrasante, même pour les scientifiques qui l'utilisent fréquemment. [16, 50]. La comparaison peut être effectuée en analysant quelques paramètres figurant sur la feuille des résultats dans les champs suivants :

a. Valeur E ou « E-value »

La valeur attendue est un nombre qui décrit le nombre de fois que vous vous attendez à une correspondance par hasard dans une base de données. Plus la valeur E est faible, plus la correspondance est significative [50].

b. Pourcentage d'identité ou « Percent Identity »

C'est un nombre qui décrit à quel point la séquence de requête est similaire à la séquence cible. Plus le pourcentage d'identité est élevé, plus la correspondance est significative [50].

c. Couverture de la requête ou « Query Cover »

C'est un nombre qui décrit la proportion de la séquence de requête couverte par la séquence cible. Si la séquence cible dans la base de données couvre l'intégralité de la séquence de requêtes, la couverture de la requête est de 100 % [50].

II-5-3- L'arbre phylogénétique

Avec l'avènement du séquençage de génomes entiers, la classification entre dans une nouvelle ère : l'ère de la phylogénie. Les études phylogéniques se proposent de reconstruire la phylogénie d'un ensemble de taxons en utilisant un grand nombre de gènes homologues ou marqueurs (l'ARNr *16S* ou *23S*), c'est à dire appartenant à une même famille génique. Elles permettent de disposer d'un nombre de caractères, et donc d'informations, beaucoup plus important pour reconstruire un arbre que les approches de phylogénie moléculaire traditionnelles. Les méthodes de reconstruction phylogénétique nous permettent de générer des arbres phylogénétiques, représentatifs de l'évolution entre les différents taxons, à partir des séquences nucléotidiques ou protéiques d'un marqueur phylogénétique. Après alignement, les séquences sont ensuite comparées et servent à la construction d'arbres phylogénétiques, chaque position nucléotidique alignée est considérée comme un caractère [20, 53].

Les méthodes de reconstruction phylogénétique actuelles ont bénéficié des énormes progrès de l'amplification génique, du séquençage et de la bio-informatique permettant de mettre de l'ordre dans l'énorme quantité de connaissances que nous avons des organismes, parce que tous les membres d'un groupe donné partagent de nombreuses caractéristiques. Ceci conduit à établir des prédictions et des hypothèses pour une recherche future, basée sur la connaissance d'organismes similaires et répartit les microorganismes en groupes significatifs, utiles, avec des noms précis, de sorte que les microbiologistes peuvent les étudier et communiquer efficacement et essentiellement pour l'identification précise des microorganismes [20, 53].

Quatre principales étapes sont utilisées pour construire l'arbre :

1. Construction de base de données contenant la séquence cible, les séquences homologues et celle du groupe externe (à raffiner)
2. Alignement multiple de séquences (à vérifier et à filtrer)
3. Construction d'un arbre (choix de modèle de distance évolutive et de méthode de reconstruction)
4. Evaluer la robustesse de l'arbre et l'interpréter

- **Le groupe externe ou « Out groupe »**

C'est une ou un groupe de séquences connues *a priori* comme externes au groupe d'intérêt tout en étant la plus proche possible de ce dernier. La racine est alors sur la branche qui relie le groupe externe aux autres séquences [27].

- **Le logiciel MEGA**

Molecular Evolutionary Genetics Analysis, le logiciel MEGA a été développé pour estimer les distances évolutives, reconstruire des arbres phylogénétiques et calculer des quantités statistiques de base à partir de données moléculaires [32].

L'objectif du logiciel MEGA a été de fournir des outils pour explorer, découvrir et analyser des séquences d'ADN et de protéines dans une perspective évolutive. La première version a été développée pour les ressources de calcul limitées qui étaient disponibles sur l'ordinateur personnel moyen au début des années 1990. Au fil du temps, des mises à jours étaient faites a ce logiciel pour améliorer ses performances, allant de la version MEGA 1 puis 2, 4, 5, 6, 7 jusqu'à la version X, actuellement la version 11 est disponible [42].

- **Le programme ClustalW**

C'est la version la plus classique (W pour weighting). Ce programme utilise un algorithme d'alignements multiple progressif qui lui permet d'être très utile car il facilite la comparaison des séquences, l'étude des différences et de similitudes. Les séquences alignées par ce programme sont fournies par l'utilisateur ou ils proviennent à partir des banques de données. Ce programme peut être un programme libre ou hébergé dans un serveur ou logiciel comme MEGA X [47].

- **Méthodes de la reconstruction phylogénétique**

Trois familles principales de méthodes existent :

- Méthodes cladistiques : basée sur l'étude des états de caractères (nucléotide ou acide aminé présent à une position, présence ou absence d'une insertion/délétion...).

Ex : Maximum de parcimonie.

- Méthodes de distances : basées sur des mesures de distances (nombre de substitutions par site).

Ex : UPGMA, NJ, minimum d'évolution, moindres carrés.

- Méthodes statistiques : basée sur l'étude des états de caractères et sur des distances.

Ex : Maximum de vraisemblance, méthodes bayésiennes [14].

- **Maximum de vraisemblance (ML)**

C'est une méthode dite de caractère(s), elle repose sur un ou plusieurs caractères à étudier. Il s'agit d'une méthode probabiliste qui nécessite un modèle d'évolution. Le choix de ce modèle est crucial pour la qualité de l'arbre obtenu. Elle est souvent décrite comme étant la meilleure méthode, c'est-à-dire la plus efficace pour trouver l'arbre le plus proche de la réalité. Son désavantage se situe au niveau des temps de calculs qui sont extrêmement longs [63].

- **Le modèle de Kimura à deux paramètres**

C'est un modèle de calcul d'une distance génétique (évolutive) entre deux séquences proposé par kimura en 1980. Il stipule que la probabilité des transitions est plus élevée que celle des tansversions. Ce modèle est plus sophistiqué que le modèle de Jukes et Cantor (un équivalent) mais biologiquement plus réaliste [47, 27].

- **Le « bootstrap »**

C'est une méthode qui permet la vérification de la robustesse et la fiabilité de l'arbre, plus précisément les branches internes des arbres phylogénétiques. C'est une estimation statistique de la confiance à accorder à une branche [27].

Chapitre III :

Matériel et

méthodes

III-1-Matériel biologique

La séquence utilisée dans ce présent travail est obtenue à partir de la souche bactérienne M3, après séquençage par la méthode de Sanger par une société privé (Beckman coulter, France). Ceci s'inscrit dans le cadre de préparation de la thèse de doctorat de Mme Guergouri I.

Le numéro d'accèsion de la souche (MW229039) est obtenu après soumission de la séquence à la base de données NCBI-GenBank. La taille de la séquence est 1140pb.

III-2-Caractérisation phénotypique

La caractérisation phénotypique est démontré par un parcours des techniques employées pour le phénotype, les caractères morphologiques sont exploités par une observation des colonies sur des milieux spécifiques faisant apparaitre certains caractères propres aux espèces, c'est une description directe faite sur boites d'isolement, permettant au moins une distinction des souches les unes des autres à fin de les purifier.

Une observation microscopique est effectuée sur des bactéries à l'état frais, pour voir leur mobilité. La taille, la forme et le type des cellules, suivies par une coloration de Gram et une coloration au vert de malachite pour voir si les bactéries possèdent des spores et vérifier la pureté des isolats et s'orienter dans le diagnostic.

Le métabolisme respiratoire consiste en un ensemencement dans milieu VF (viande-foie), test de catalase, et test de nitrate réductase sur bouillon nitrate, et le test d'oxydase. Le métabolisme des glucides est exploité par l'utilisation de milieux de culture contenant un ou plusieurs glucides comme source de carbone et d'énergie, l'auxanogramme du carbone détermine la gamme des glucides utilisés par la bactérie.

L'emploi de test VP-RM permet de mettre en évidence la fermentation acide mixte et la fermentation acide-butandioïque qui génère l'AMC (acétyl-méthyl-carbinole) comme composé final.

Les résultats de la caractérisation phénotypique de la souche M3 sont montrés dans le tableau si dessous de même pour l'observation des colonies et observation microscopique (Figure 14). Ils sont issus d'un mémoire de fin d'étude intitulé (étude de la diversité bactérienne d'un sol pollué par les hydrocarbures), des étudiants : Anissa Oussalah et Racim Louadfel, encadrés par Madame Guergouri I. en 2014.

Tableau 2: Caractères phénotypiques de la souche M3.

Colonies	rugueuses ; possédant un contour dentelé et un aspect sec
Cellule	Bâtonnets à Gram positif
Spores	non-déformantes
Mobilité	absente
Type respiratoire	aérobies strictes
Enzymes présentes	Catalase, nitrate réductase, lecithinase et amylase
Enzymes absentes	Oxydase, Uréase, gélatinase
Sucres utilisés	Glucose, galactose et fructose, ribose, xylose et arabinose saccharose, lactose, sucrose et maltose, sorbitol
Fermentation	RM+/VP-

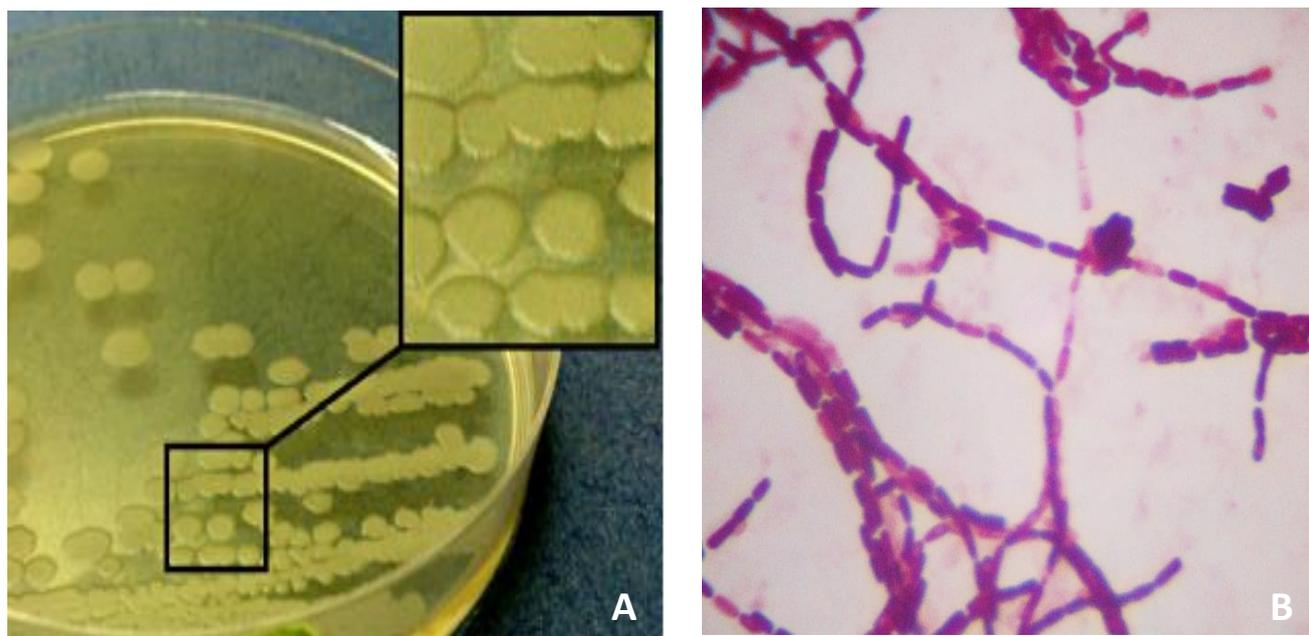


Figure 14 : Photographies de l'aspect macroscopique des colonies (A) et microscopique des cellules (B) [2].

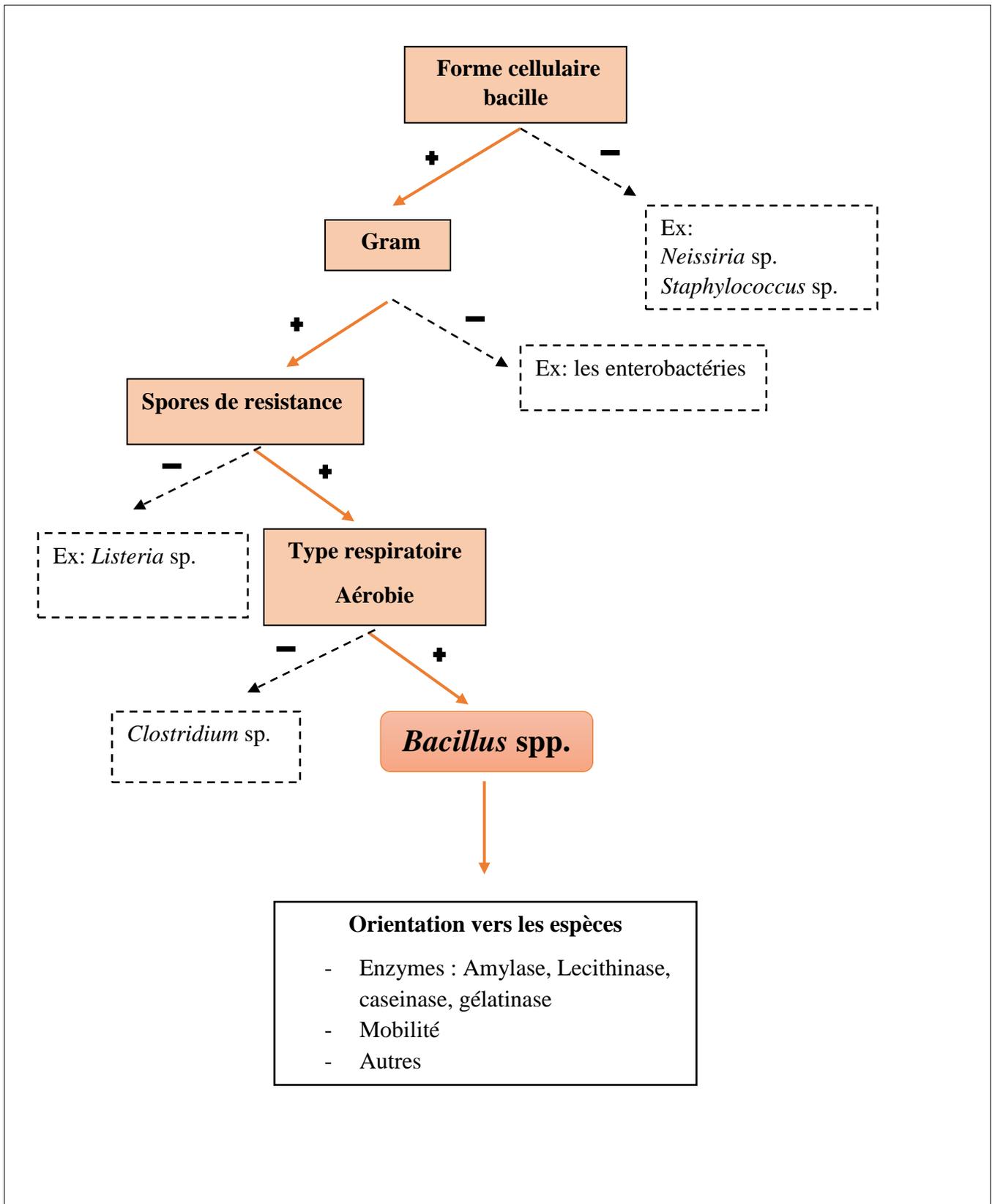


Figure 15 : Clé dichotomique d'orientation vers le genre de la souche M3.

III-3-Analyse des résultats de séquençage

Dans cette démarche on se propose d'établir le suivant :

1. Une recherche de séquences homologues à travers un alignement de la séquence partielle d'*ADNr 16S* sur deux bases de données différentes qui sont NCBI et EzBioCloud
2. Une détermination de la position taxinomique de la souche en représentant un arbre phylogénétique par la méthode de reconstruction maximum de vraisemblance (ML).
3. Une comparaison des résultats phénotypiques et phylogénétiques de la souche M3.

III-3-1-Recherche de séquences homologues et l'identification de la bactérie

La séquence nucléotidique est obtenue sous forme de chromatogramme après l'étude génétique (une PCR sur colonie et le séquençage Sanger) ainsi qu'un fichier d'extension (.abi) généré par le séquenceur lisible par une application qui s'appelle *Chromas* (version 2.6.6) (Figure 16)

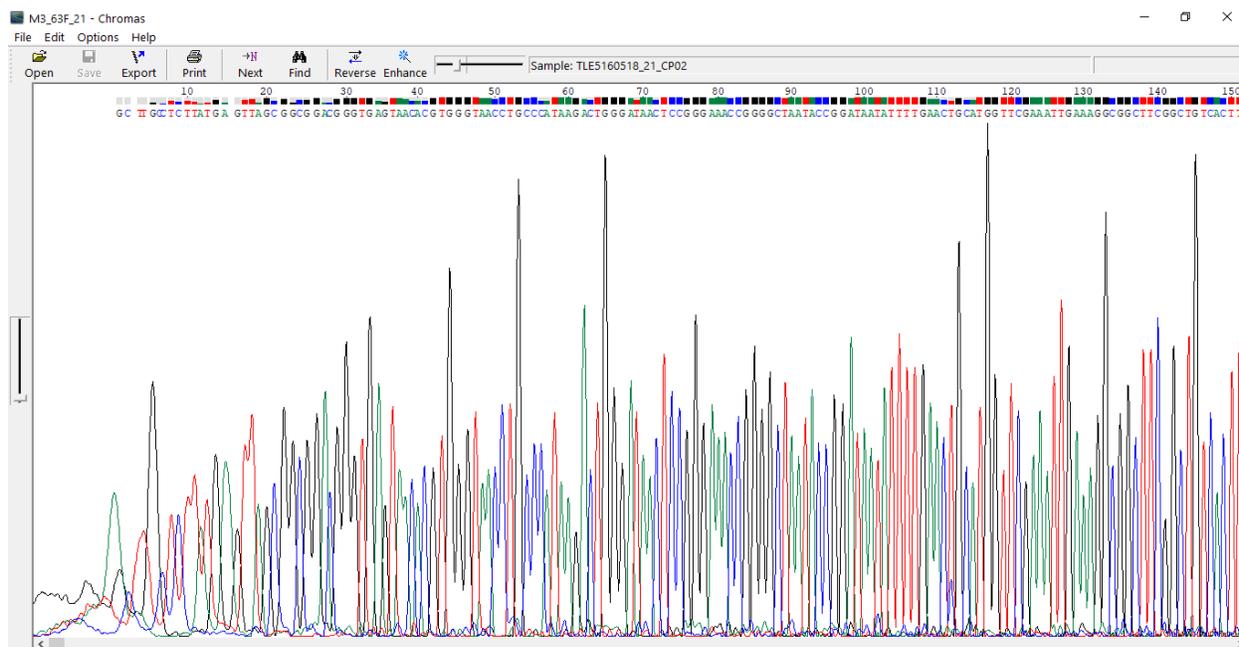


Figure 16 : Le chromatogramme de séquence après séquençage

➤ Correction de la séquence avec *Chromas* et enregistrement en format FASTA.

Un nettoyage des séquences est effectué sur les marges pour couper les parties de la séquence qui sont de mauvaise qualité (avec bruit de fond), puis sont enregistrées en format FASTA (Figure 17) lisible par l'application Notepad de Windows (Figure 18).

l'alignement à savoir : choisir les bases de données qui seront interrogées selon l'organisme en question (16S ribosomal RNA sequences Bacteria and Archaea), exclure les séquences d'échantillons non cultivables/environnementales et choisir l'algorithme de calcul (mégablaste) pour identifier des homologues étroitement liés à la séquence soumise. L'opération est lancée en s'appuyant sur l'icône BLAST (Figure 19).

The screenshot displays the NCBI BLASTn web interface. The browser address bar shows the URL: `blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome`. The page header includes the NIH logo and the text "U.S. National Library of Medicine National Center for Biotechnology Information". The main navigation bar shows "BLAST® » blastn suite" and links for "Home", "Recent Results", "Saved Strategies", and "Help".

The interface is titled "Standard Nucleotide BLAST" and features several sections for configuring the search:

- Enter Query Sequence:** A text area contains the sequence:


```
TAAGTCCCACGAAGCGCAACCCCTGATTGCTTAGTGCCATCCATTA
AGTTGGGCA
CCTCTAAAGGTGACTGCCGTGACAACCGGAGGAAAGGTTGGATGACGT
CAATCAT
```

 Below the text area are options to "Or, upload file" and a "Job Title" field containing "TLE5160518_21_CP02".
- Choose Search Set:** This section allows selecting the database and organism.
 - Database:** "Standard databases (nr etc.):" is selected, with "16S ribosomal RNA sequences (Bacteria and Archaea)" chosen from the dropdown menu.
 - Organism:** A text field for "Enter organism name or id" is present, with an "Add organism" button.
 - Exclude:** "Models (XM/XP)" is unchecked, and "Uncultured/environmental sample sequences" is checked.
 - Limit to:** "Sequences from type material" is unchecked.
 - Entrez Query:** A text field for "Enter an Entrez query to limit search" is present.
- Program Selection:** Under "Optimize for", "Highly similar sequences (megablast)" is selected. Other options include "More dissimilar sequences (discontiguous megablast)" and "Somewhat similar sequences (blastn)".

At the bottom, a "BLAST" button is visible, along with a note: "Note: Parameter values that differ from the default are highlighted in yellow and marked with a sign".

Figure 19: Réglage des paramètres sur la fenêtre de BLASTn du NCBI.

➤ **Analyse des résultats et recherche des séquences homologues valides**

Un tableau de description s'affiche montrant une liste de noms bactériens avec des informations sur le score d'alignement, la E-value, le pourcentage de similitude ainsi que le numéro d'accèsion des différentes séquences (Figure 20).

The screenshot shows the NCBI BLAST results page for the query sequence TLES160518_21_CP02. The table lists 100 sequences with the following columns: Description, Scientific Name, Max Score, Total Score, Query Cover, E value, Per. Ident, Acc. Len, and Accession. The sequences are sorted by E value, with the lowest values (0.0) at the top. The table includes entries for various Bacillus species, such as Bacillus thuringiensis, Bacillus cereus, and Bacillus tropicus, with their respective accession numbers and scores.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Bacillus sp. (in. Bacteria) strain SRSJ4 16S ribosomal RNA gene, partial sequence	Bacillus sp. (in. Ba...	2004	2004	100%	0.0	98.51%	1290	MW371116.1
Bacillus thuringiensis strain HER1410 chromosome, complete genome	Bacillus thuringiensis	2002	27907	100%	0.0	98.51%	5585577	CP050183.1
Bacillus cereus strain FDAARGOS_798 chromosome, complete genome	Bacillus cereus	2002	27973	100%	0.0	98.51%	5381410	CP053954.1
Bacillus tropicus strain FDAARGOS_782 chromosome, complete genome	Bacillus tropicus	2002	27846	100%	0.0	98.51%	5263142	CP053955.1
Bacillus cereus strain FDAARGOS_799 chromosome, complete genome	Bacillus cereus	2002	27973	100%	0.0	98.51%	5381556	CP053951.1
Bacillus cereus strain GD1 16S ribosomal RNA gene, partial sequence	Bacillus cereus	2002	2002	100%	0.0	98.51%	1393	MT434695.1
Bacillus wiedmannii strain J8BS10 16S ribosomal RNA gene, partial sequence	Bacillus wiedmannii	2002	2002	100%	0.0	98.51%	1323	MT415987.1
Bacillus wiedmannii strain J5A12LARS 16S ribosomal RNA gene, partial sequence	Bacillus wiedmannii	2002	2002	100%	0.0	98.51%	1314	MT378557.1
Bacillus sp. RZ2MS9 chromosome, complete genome	Bacillus sp. RZ2MS9	2002	27896	100%	0.0	98.51%	5357194	CP049978.1
Bacillus tropicus strain AOA-CP51 chromosome	Bacillus tropicus	2002	27857	100%	0.0	98.51%	5246960	CP049019.1
Bacillus thuringiensis strain YA-3B 16S ribosomal RNA gene, partial sequence	Bacillus thuringiensis	2002	2002	100%	0.0	98.51%	1363	

Figure 20: Résultat de BLAST de la séquence.

En cliquant sur les noms des bactéries une fenêtre descriptive de chacune apparaît, donnant un aperçu détaillé sur la séquence. La validité de cette séquence est vérifiée en recherchant la publication la comportant (Figure 21).

The screenshot shows the NCBI database entry for the sequence NR_157735.1. The entry includes the following information:

- LOCUS:** NR_157735 1509 bp rRNA linear BCT 12-MAR-2019
- DEFINITION:** Bacillus proteolyticus strain MCCC 1A00365 16S ribosomal RNA, partial sequence.
- ACCESSION:** NR_157735
- VERSION:** NR_157735.1
- DBLINK:** Project: 33175, BioProject: PRJNA33175
- KEYWORDS:** RefSeq.
- SOURCE:** Bacillus proteolyticus
- ORGANISM:** Bacillus proteolyticus; Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
- REFERENCE 1:** (bases 1 to 1509) Liu, Y., Du, J., Lai, Q., Zeng, R., Ye, D., Xu, J. and Shao, Z. Proposal of nine novel species of the Bacillus cereus group. Int. J. Syst. Evol. Microbiol. 67 (8), 2499-2508 (2017). PMID: 28792367
- REFERENCE 2:** (bases 1 to 1509) NCBI RefSeq Targeted Loci Project. Direct Submission. Submitted (01-AUG-2018) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
- REFERENCE 3:** (bases 1 to 1509)

The right sidebar contains a 'Customize view' section with options to 'Analyze this sequence', 'Run BLAST', 'Pick Primers', 'Highlight Sequence Features', and 'Find in this Sequence'. Below this is 'Related information' including 'BioProject', 'PubMed', 'Taxonomy', and 'BioCollections'. At the bottom, there is a 'Recent activity' section showing the current sequence entry.

Figure 21: Séquence valide avec le lien vers sa publication.

Les sequences valides dont le pourcentage de similitude est superieur à 97% sont selectionnées (Figure 20) et un fichier comportant toutes leurs sequences (sequence requette et homologues) en format FASTA est téléchargé.

A ce stade, la lecture des resultats concernant l'identité de la souche inconnue (reqête) peut être fait en regardant le poucentage de similitude des souches proposées.

Afin de preciser le genre bacterien, une similarité de 98% est acceptée. Pour l'espece par contre il faut un pourcentage superieur de 98%.

➤ Recherche de séquences d'ARN 16S des souches types sur EzBioCloud

Pour accéder aux services de la base de donnée EzBioCloud, un compte doit être créé. Plusieurs fenêtres sont proposées à l'interface, celle de l'identification bactérien par ARN 16S est sélectionnée : « 16S-based ID » (Figure 22).

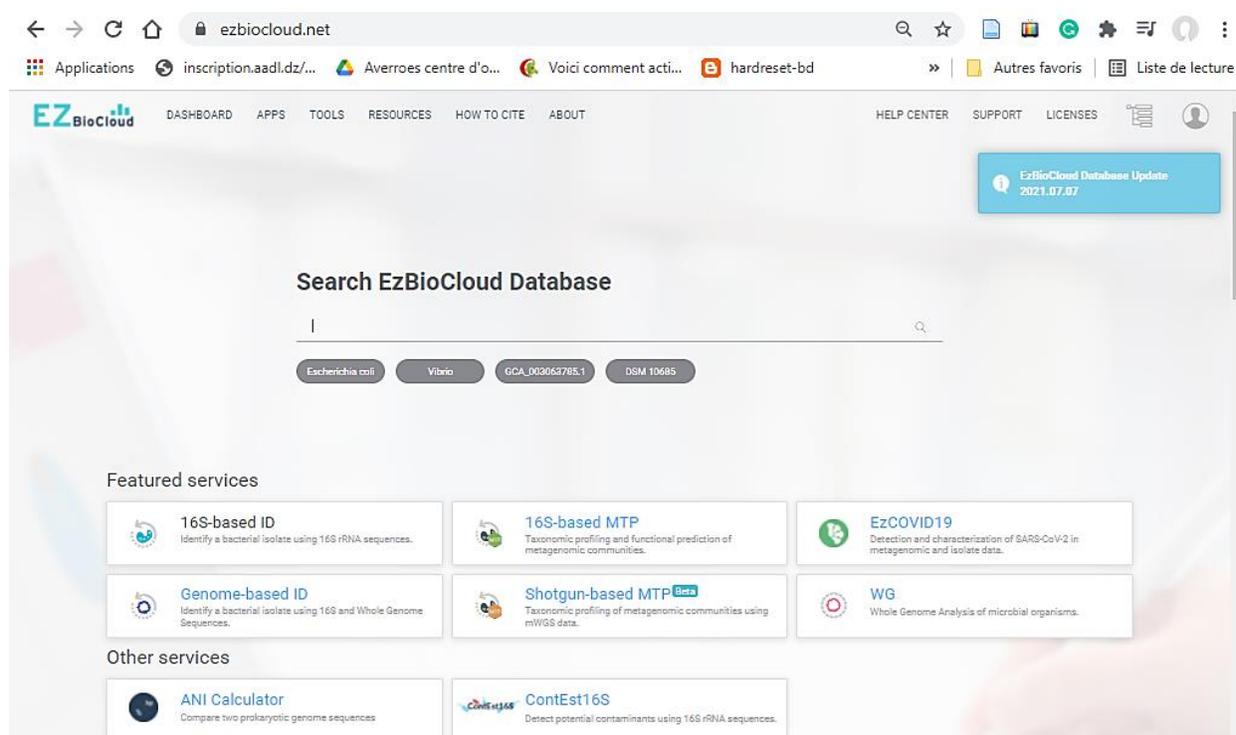


Figure 22 : La base de données EzBioCloud.

Puis un cadrant apparait en cliquant sur « identify new sequence », permettant de coller la séquence requête (Figure 23) et l'aligner.

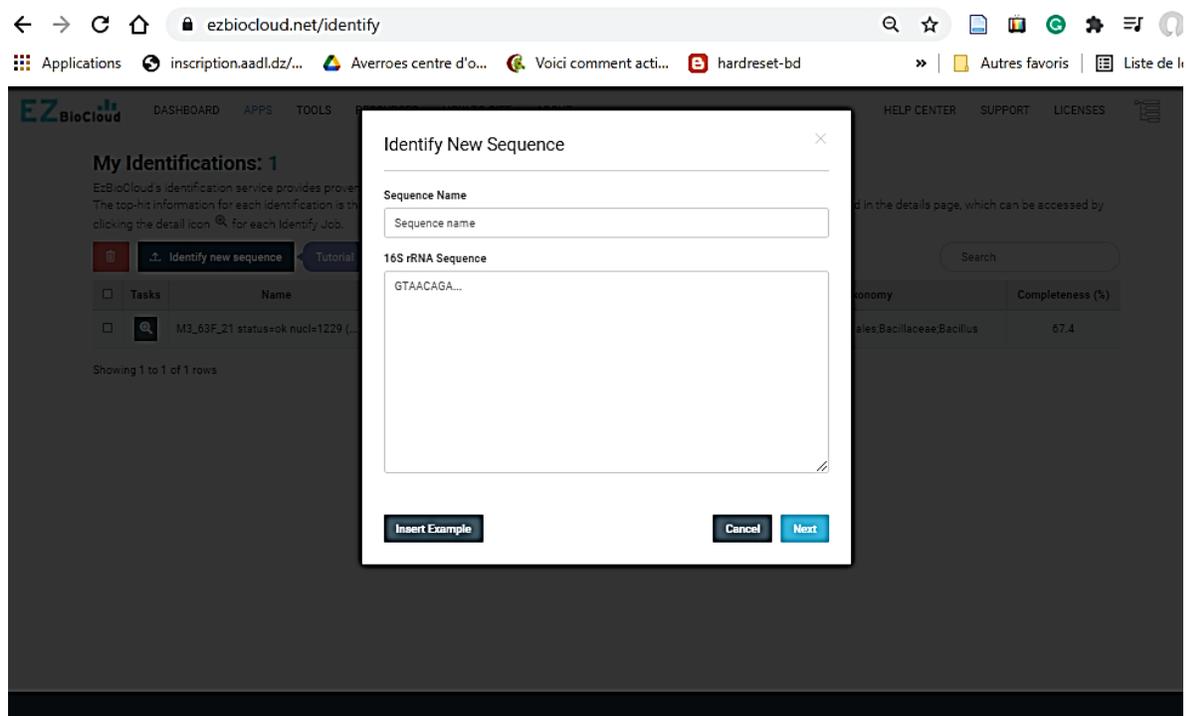


Figure 23: Fenêtre d'identification d'une nouvelle séquence.

Une fenêtre des résultats de l'identification s'ouvre donnant une liste des bactéries rapprochées à notre bactérie, sur la base de similitude des séquences d'ARNr 16S (Figure 24).

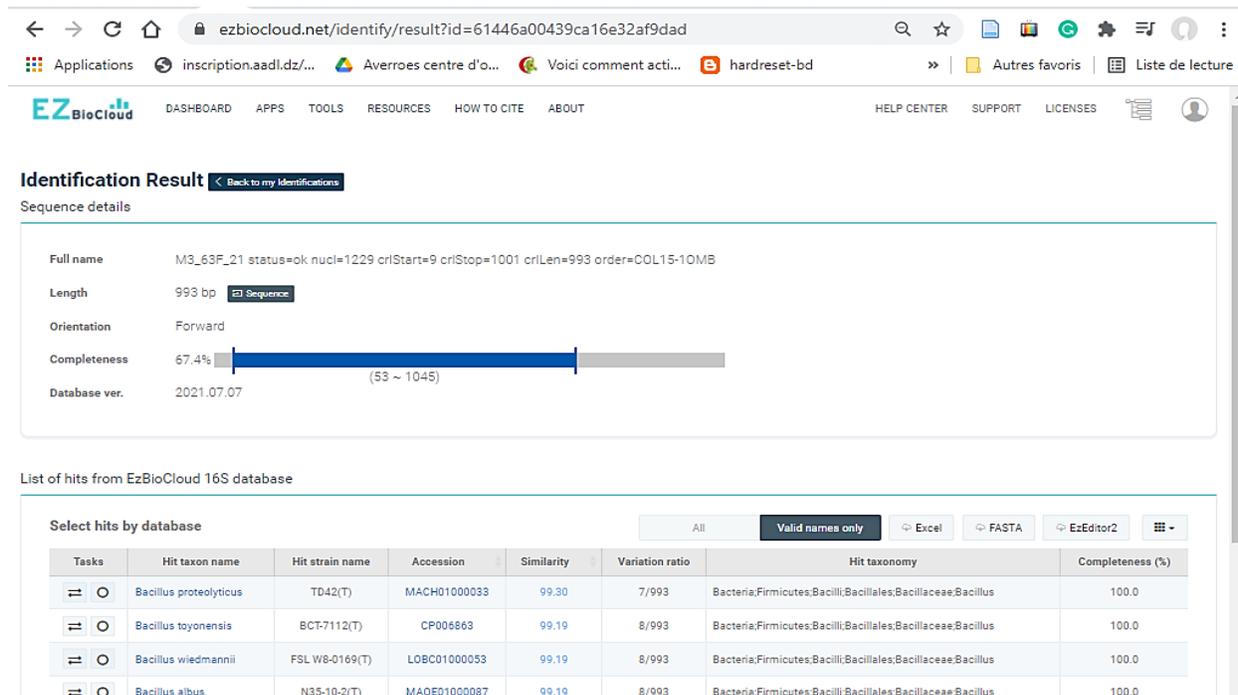


Figure 24 : Résultats de l'identification de la nouvelle séquence.

À ce stade, l'identification de la bactérie par rapport à celles proposées par la base de données est possible en procédant de la même manière que pour le NCBI, sauf qu'ici il est recommandé de choisir la première proposition comme étant la plus proche.

Pour la représentation de l'arbre par contre, les séquences de quelques souches types sont récupérées après limitation de la liste aux noms valides seulement « valide names only ». En cliquant sur le numéro d'accèsion des souches, une page descriptive de cette souche apparaît (Figure 25). L'icône « sequence » permet de copier la séquence recherchée. Ainsi toutes les séquences seront rajoutées au fichier précédemment récupéré de NCBI.

The screenshot shows a web browser window with the URL `ezbiocloud.net/16SrRNA?ac=MACH01000033`. The page title is "16S rRNA sequence". The main content is a table with the following rows:

Accession
MACH01000033 Sequence TE B
Taxon name
Bacillus proteolyticus Liu et al. 2017
Strain
TD42 (Type)
Sequencing method
Genome sequencing
Completeness
100.0%
Taxonomy
Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus

Figure 25 : Page descriptive de la souche.

La séquence du groupe externe est aussi recherchée de la même manière après le choix de la souche à utiliser.

Ainsi, un fichier Notepad qui contient la séquence requête, les séquences homologues de souches types et non-types aussi du groupe externe est prête pour l'utilisation dans la représentation de l'arbre phylogénétique.

III-3-2-Représentation de l'arbre phylogénétique

➤ Importation du fichier contenant les séquences

Le fichier des séquences est importé, dans le MEGA X en cliquant sur l'icône « align » puis “edit/build an new alignment”, elles sont copiées puis collées sur la fenetre d'alignement qui s'ouvre (Figure 26).

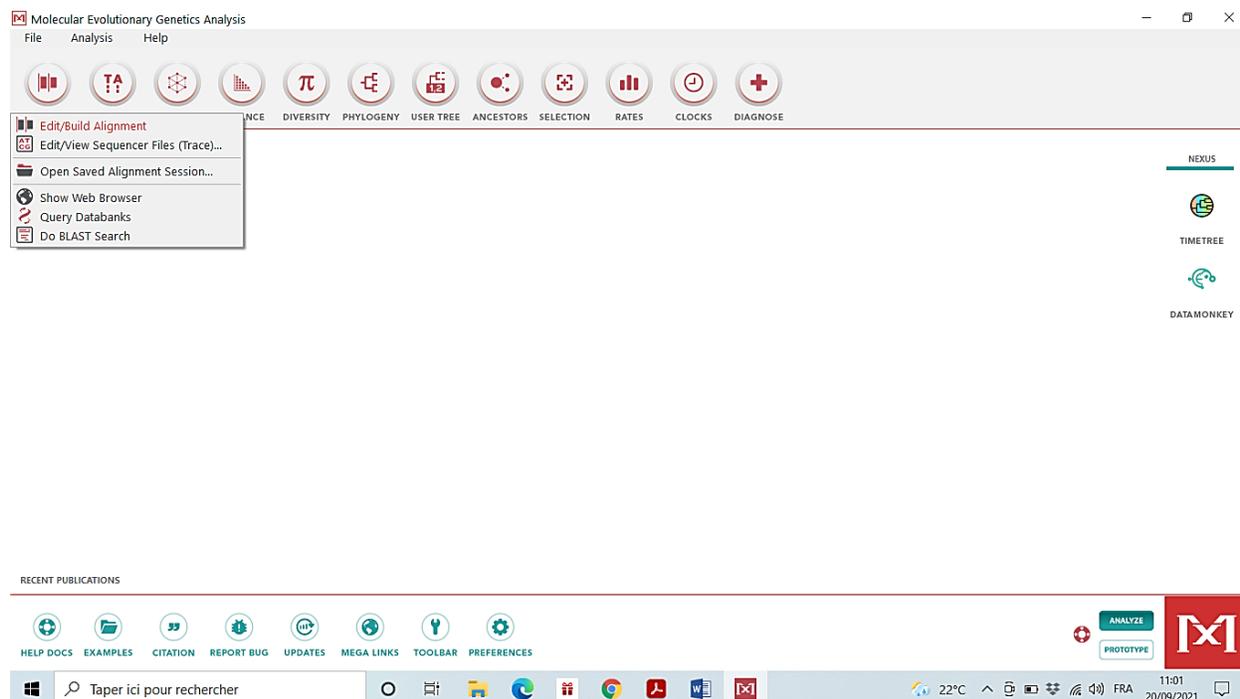


Figure 26 : l'ajout des séquences nucléotidique dans l'explorateur d'alignement de MEGA X.

L'explorateur d'alignement affiche un nom pour chaque séquence à gauche, suivi de la séquence. Les lignes identifiants des différentes séquences sur le fichier FASTA sont nettoyées au préalable de toute information supplémentaire appart le nom de la bactérie et son code d'accension (Figure 27).

➤ Alignement des séquences par le programme CLUSTAL-W hébergé dans MEGA X

Dans la barre d'outils, près du haut de la fenêtre, l'alignement Clustal est symbolisé par l'icône W. En choisissant Aligner l'ADN (Figure 28), une fenêtre de paramètres s'ouvre. Ceux par défaut sont parfaits pour démarrer le processus d'alignement. Selon le nombre de séquences impliquées et la méthode choisie, l'alignement peut prendre de quelques secondes à quelques heures. Une fois l'alignement terminé, la session est enregistrée en exportant les données l'alignement dans le menu DATA, sous un nom déterminé et avec soit le format FASTA ou MEGA.

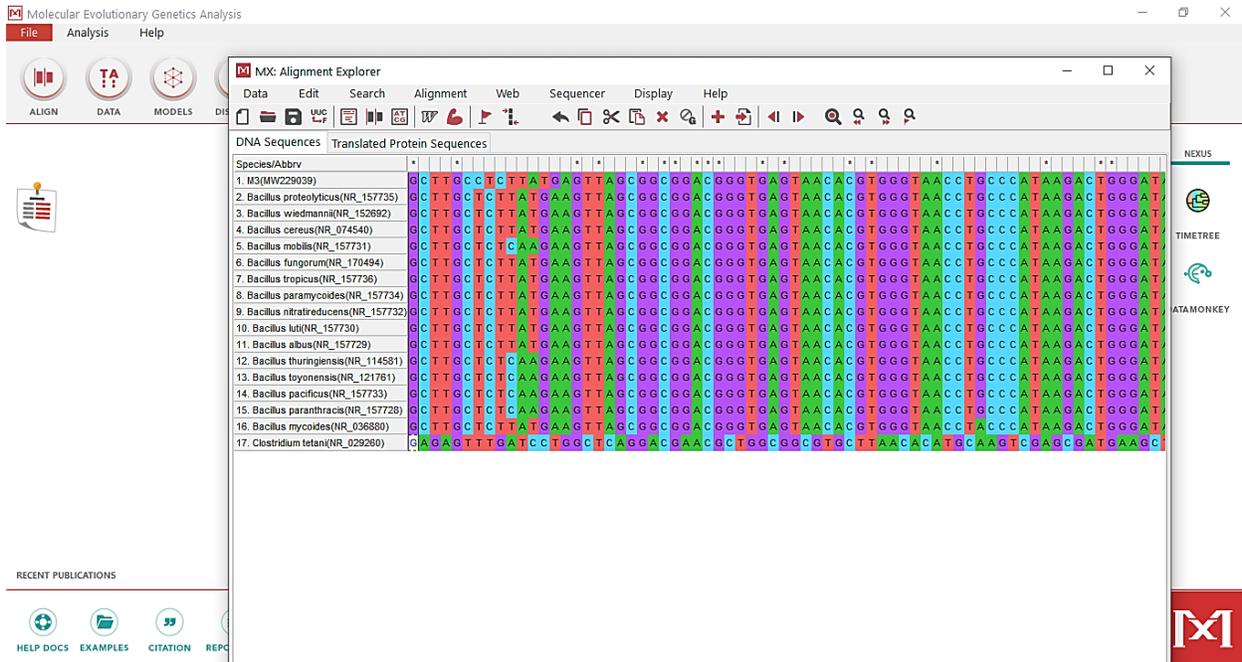


Figure 27 : alignement des séquences sur le MEGA X.

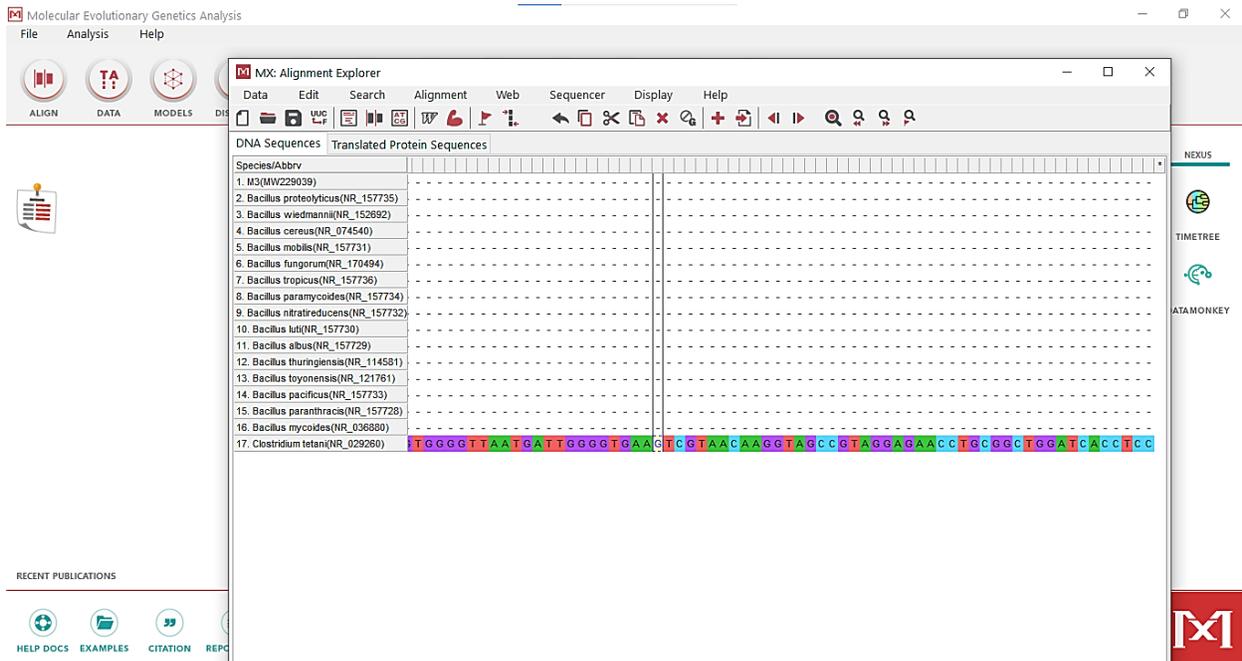


Figure 28: Utilisation du Programme CLUSTAL W pour aligner les séquences.

➤ Choix de la méthode de reconstruction sur MEGA X

Le fichier ou la session enregistrée à l'étape précédente est ouverte à partir du menu File. L'icône « Phylogeny », permet de choisir une parmi cinq méthodes de reconstruction.

La méthode utilisée dans notre cas est la méthode du maximum de vraisemblance (ML) (Figure 29). Une boîte de dialogue de paramètre apparaîtra (Figure 30). Lorsque les préférences sont définies, on clique sur le bouton ok pour la construction de l'arbre. Finalement, une fenêtre d'explorateur d'arbre s'ouvrira qui affichera l'arbre

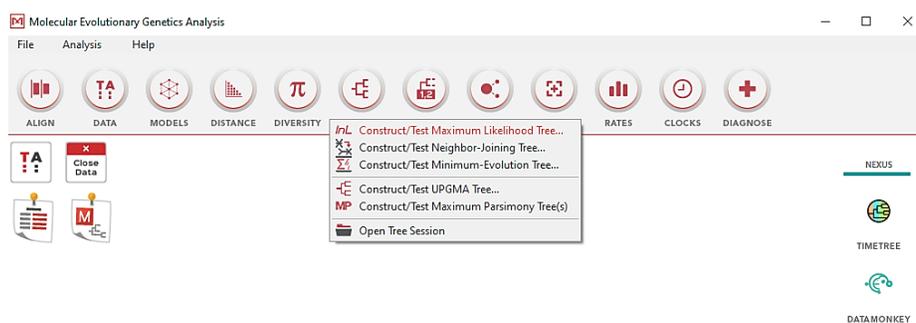


Figure 29: Choix de la méthode ML pour la reconstruction de l'arbre phylogénétique.

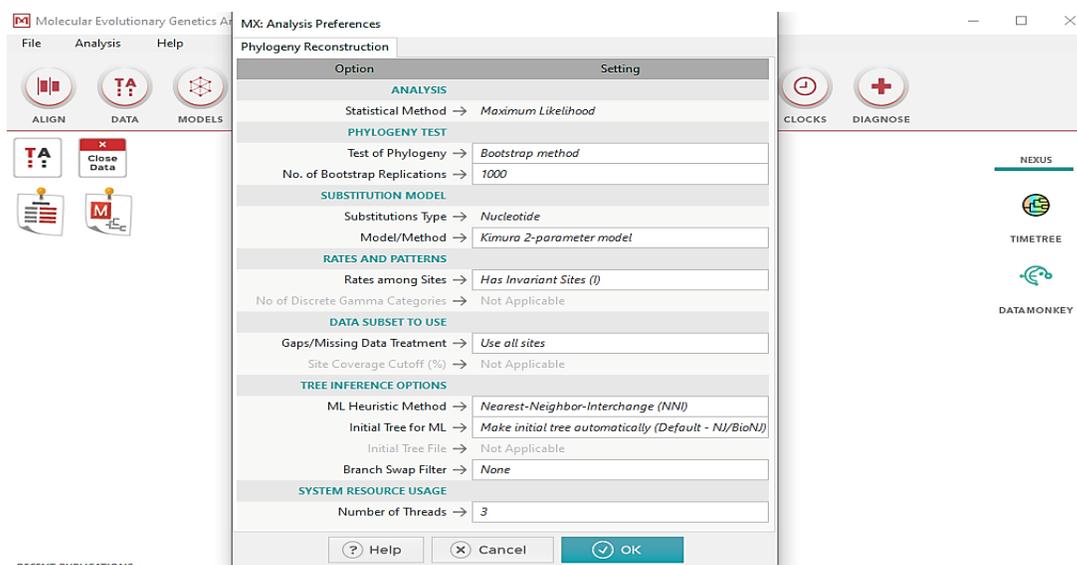


Figure 30 : Paramètres de construction de l'arbre par ML.

➤ **Choix du modèle de calcul des distances génétiques entre les séquences**

Il est dit aussi modèle de substitution. Son choix est basé sur un test réalisé sur le MEGA en cliquant sur l'icône modèle, le logiciel va lancer un ensemble de calculs pour déterminer le meilleur modèle à utiliser figurant comme premier choix dans la liste des propositions.

Dans notre cas c'est le model K2, appelé Kimura à deux paramètres (Figure 31) qui doit être sélectionné sur la figure précédant.

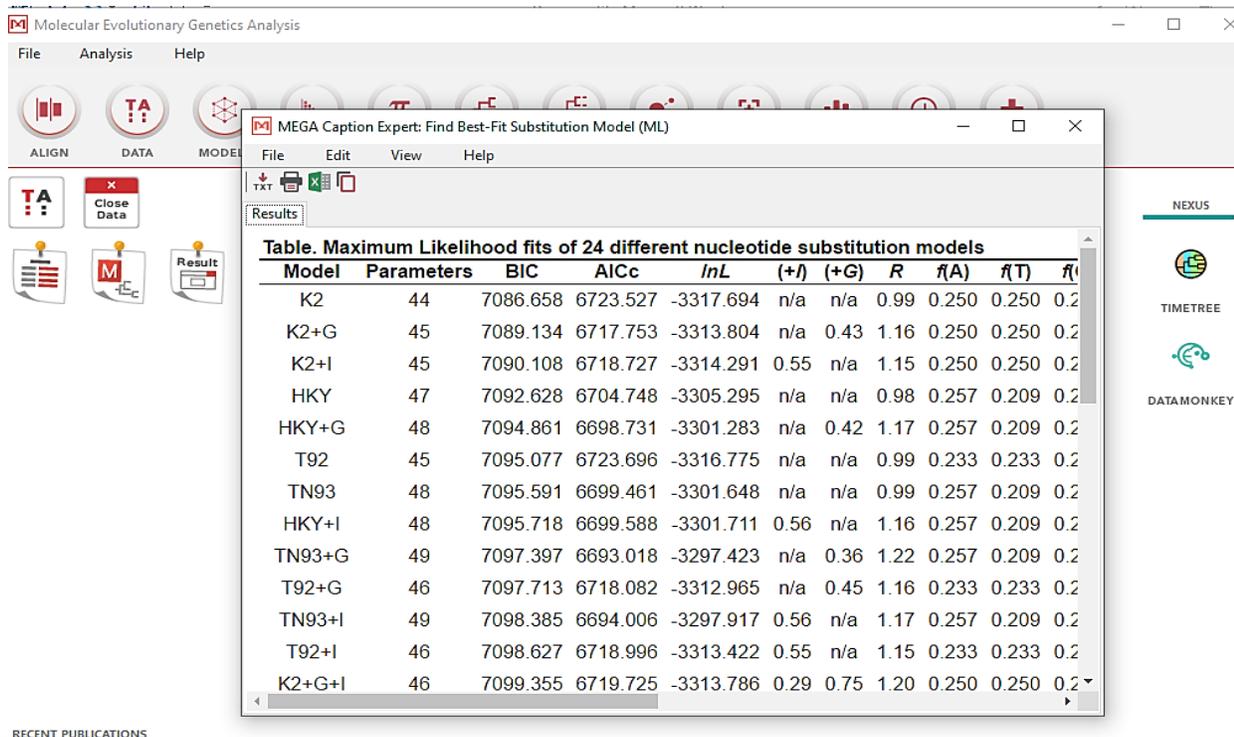


Figure 31: Test pour choix du modèle de calcul des distances génétiques.

L'étape prochaine sera de cliquer sur OK pour lancer la reconstruction de l'arbre qui va prendre un temps relatif au nombre de répétition choisies pour le test de robustesse « bootstrap ». Pour mieux tester la fiabilité et la robustesse des branches sur l'arbre, 1000 répétions sont demandées sur la fenêtre précédente.

Après épuisement du temps d'attente, un arbre est généré sur une fenêtre (Figure 32), comportant en bas la description de toutes les préférences sélectionnées auparavant.

Quelques petites détails au niveau de l'affichage de l'arbre peuvent être changés ; on peut cacher les valeurs de bootstrap non significatives (<50) en cliquant sur « Branches » puis « statistic/Frequency », on peut aussi afficher l'échelle en cliquant sur « Scale »

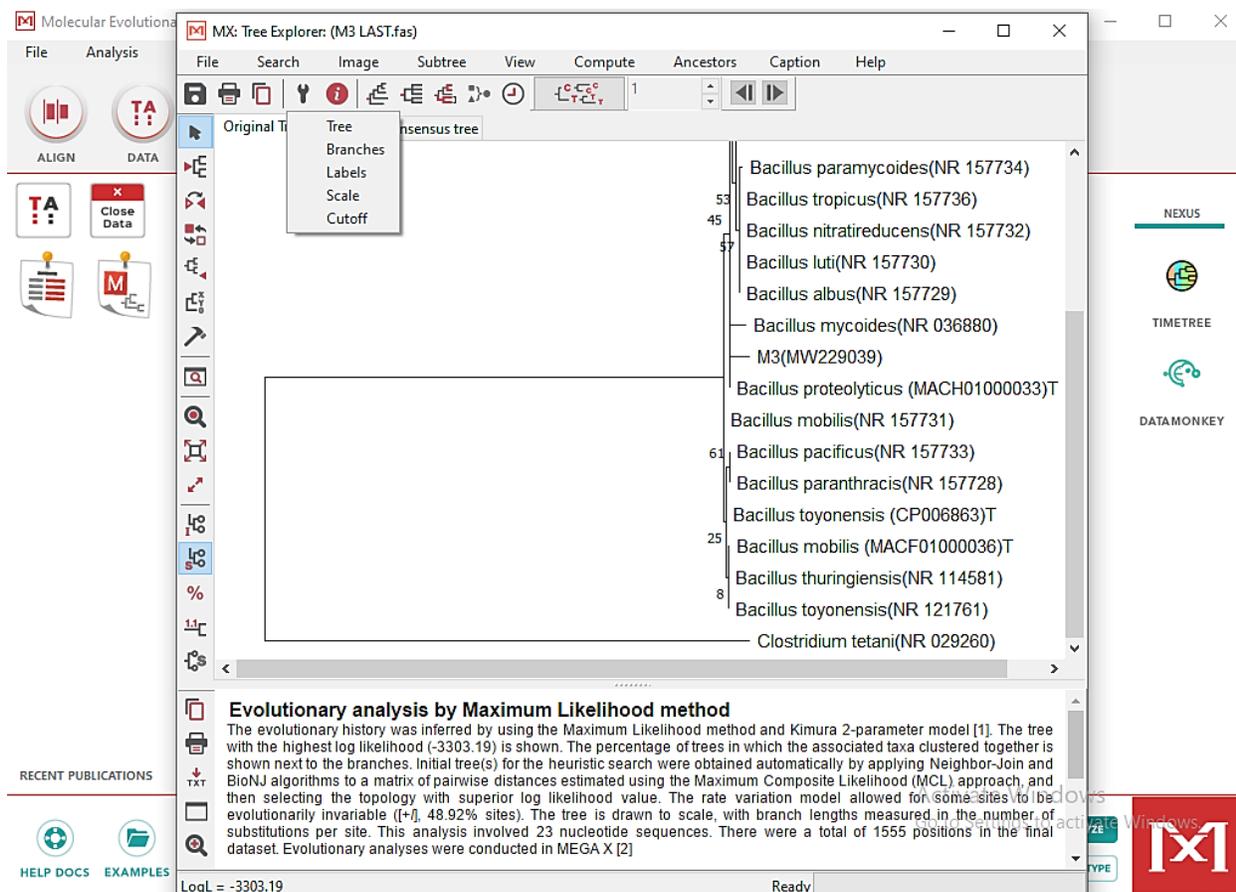
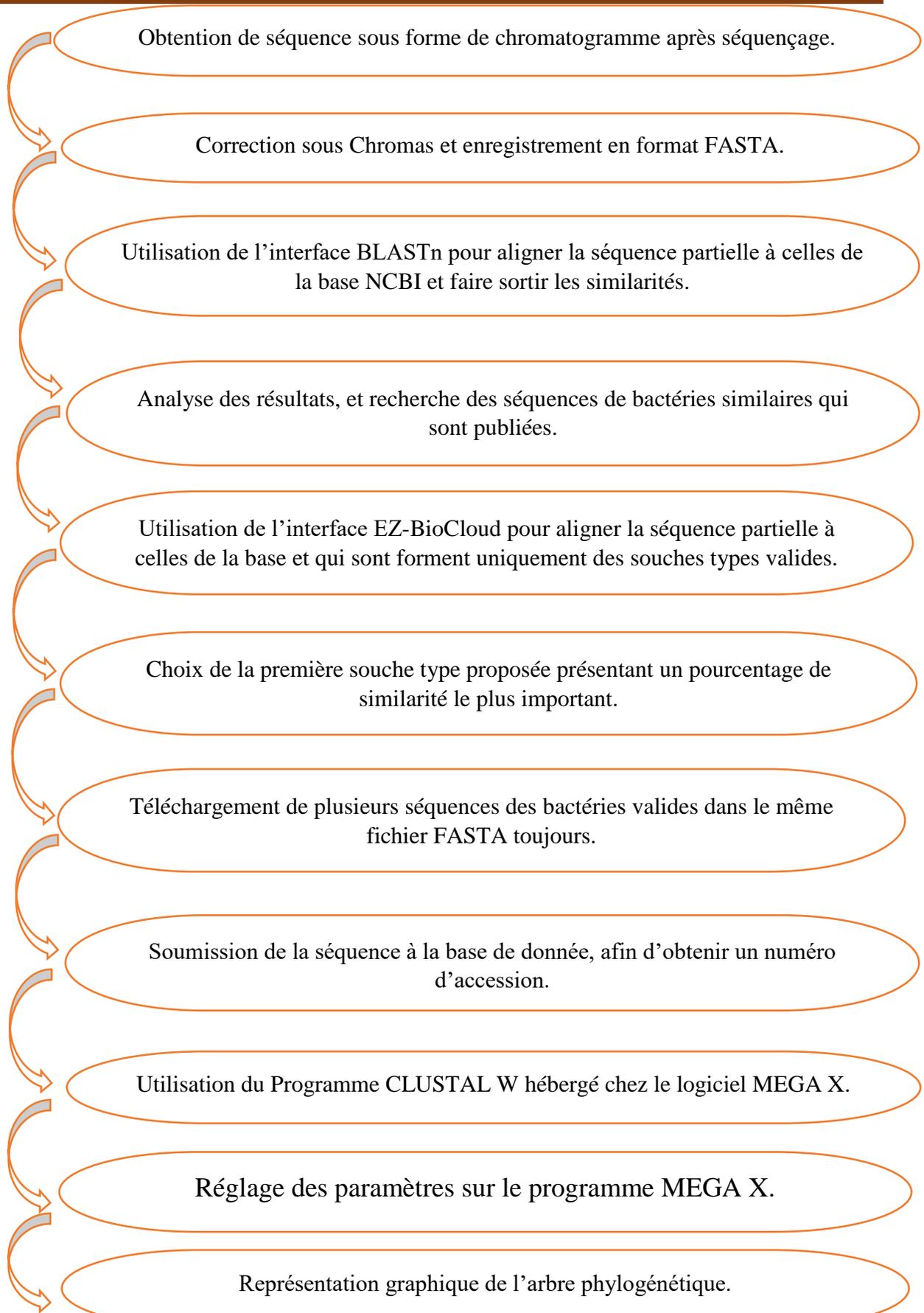


Figure 32 : Construction de l'arbre et affichage de l'échelle.

Il est important de sauvegarder l'arborescence, et la légende, afin qu'il puisse être modifié ultérieurement si nécessaire. L'enregistrement peut se faire en différents formats, au niveau du menu « Image ».



III-4-Résultats, interprétation et discussion

III-4-1- Résultats de l'alignement par BLASTn

Tableau 3 : Les souches bactériennes similaires publiées sur NCBI.

Nom de la bactérie	Numéro d'accèsion	Date	Taille de la séquence	Query cover	E-value	Per.ident
<i>Bacillus proteolyticus</i> strain MCCC 1A00365	NR_157735	12-MAR-2019	1509 bp	100%	0	98.43%
<i>Bacillus wiedmannii</i> strain FSL W8-0169	NR_152692	12-MAR-2019	1540 bp	100%	0	98.73%
<i>Bacillus cereus</i> ATCC 14579	NR_074540	12-MAR-2019	1512 bp	100%	0	98.34%
<i>Bacillus mobilis</i> train MCCC 1A05942	NR_157731	12-MAR-2019	1509 bp	100%	0	98.34%
<i>Bacillus fungorum</i> strain 17-SMS-01	NR_170494	25-SEP-2020	1576 bp	100%	0	98.34%
<i>Bacillus tropicus</i> strain MCCC 1A01406	NR_157736	12-MAR-2019	1509 bp	100%	0	98.25%
<i>Bacillus paramycoïdes</i> strain MCCC 1A04098	NR_157734	12-MAR-2019	1509 bp	100%	0	98.25%
<i>Bacillus nitratireducens</i> strain MCCC 1A00732	NR_157732	12-MAR-2019	1509 bp	100%	0	98.25%
<i>Bacillus luti</i> strain MCCC 1A00359	NR_157730	12-MAR-2019	1509 bp	100%	0	98.25%
<i>Bacillus albus</i> strain MCCC 1A02146	NR_157729	12-MAR-2019	1509 bp	100%	0	98.25%
<i>Bacillus thuringiensis</i> strain ATCC 10792	NR_114581	12-MAR-2019	1482 bp	100%	0	98.25%
<i>Bacillus toyonensis</i> strain BCT-7112	NR_121761	12-MAR-2019	1544 bp	100%	0	98.25%
<i>Bacillus pacificus</i> strain MCCC 1A06182	NR_157733	12-MAR-2019	1509 bp	100%	0	98.16%
<i>Bacillus paranthracis</i> strain MCCC 1A00395	NR_157728	12-MAR-2019	1509 bp	100%	0	98.16%
<i>Bacillus mycoïdes</i> strain 273	NR_036880	12-MAR-2019	1513 bp	100%	0	98.08%
<i>Clostridium tetani</i> strain NCTC 279	NR_029260	12-MAR-2019	1508 bp	OUT GROUP E	OUT GROUPE	OUT GROUPE

Tableau 4 : Résultats de l'EzBioCloud.

Rank	Name	Strain	Accession	Pairwise Similarity(%)	Mismatch/Total nt	Completeness(%)
1	Bacillus proteolyticus	TD42	MACH01000033	98.58	16/1129	100
2	Bacillus toyonensis	BCT-7112	CP006863	98.49	17/1129	100
3	Bacillus wiedmannii	FSL W8-0169	LOBC01000053	98.49	17/1129	100
4	Bacillus albus	N35-10-2	MAOE01000087	98.49	17/1129	100
5	Bacillus luti	TD41	MACI01000041	98.49	17/1129	100
6	Bacillus mobilis	0711P9-1	MACF01000036	98.49	17/1129	100

Le pourcentage de similitude

Selon Catherine Dauga et ses collaborateurs (2005), «...Aujourd'hui, n'importe quelle bactérie est identifiable par la position de sa séquence d'*ADNr 16S* dans un groupe représentatif d'une famille, d'un genre ou d'une espèce bactérienne au sein d'un arbre phylogénétique. Cette position phylogénétique est corrélée à la valeur de similitude des *ARNr 16S*: deux bactéries appartiennent à des espèces différentes si leurs *ARNr 16S* partagent moins de 97 % de similitude. »
Fin de citation.

Lorsqu'une souche bactérienne est isolée, on veut savoir si elle appartient à un genre et une espèce connue ou s'il s'agit d'une nouvelle espèce. La première technique couramment utilisée pour rattacher une souche à un genre bactérien est le séquençage du gène codant pour l'*ARNr 16S*. A partir de la séquence obtenue, on peut rechercher le genre le plus proche dans des banques de données par alignement des séquences. Le pourcentage de similitude obtenu permet de déterminer l'appartenance d'un isolat à un genre bactérien mais la limite reste variable d'un genre à un autre [15].

En effet, il est important d'apprécier cette limite au cas par cas, parce que le taux de mutations observé dépend non seulement du temps écoulé entre les épisodes de spéciation, mais aussi de l'efficacité des mécanismes de réparation de l'ADN propres à chaque espèce [19].

Lorsque ce pourcentage dépasse 97%, une autre méthode (comme l'hybridation ADN-ADN) doit être utilisée, pour confirmer si une souche appartient à une espèce existante ou est une nouvelle espèce. Un pourcentage d'hybridation supérieur à 70% indique une nouvelle espèce. D'autres gènes peuvent être utilisés particulièrement pour les genres pour lesquels le gène de l'*ARNr 16S* est trop conservé pour servir d'indicateur taxonomique [15]

En observant nos données figurant sur les tableaux 3 et 4 on s'aperçoit que notre séquence présente un pourcentage de 98% avec les 10 Premières propositions dans la liste du NCBI,

En le comparant aux séquences de la base EzBioCloud qui sont toutes des souches types, notre séquence présente aussi un pourcentage similaire au premier. On peut conclure que la bactérie M3 appartient au genre *Bacillus*

On reste cependant incertains de l'espèce malgré les pourcentages dépassant le 97%. Ceci est dû au fait que la qualité de la séquence n'est idéale en plus ce n'est qu'une copie parmi de nombreuses copies de ce gène.

III-4-2-Résultat de la construction de l'arbre phylogénétique

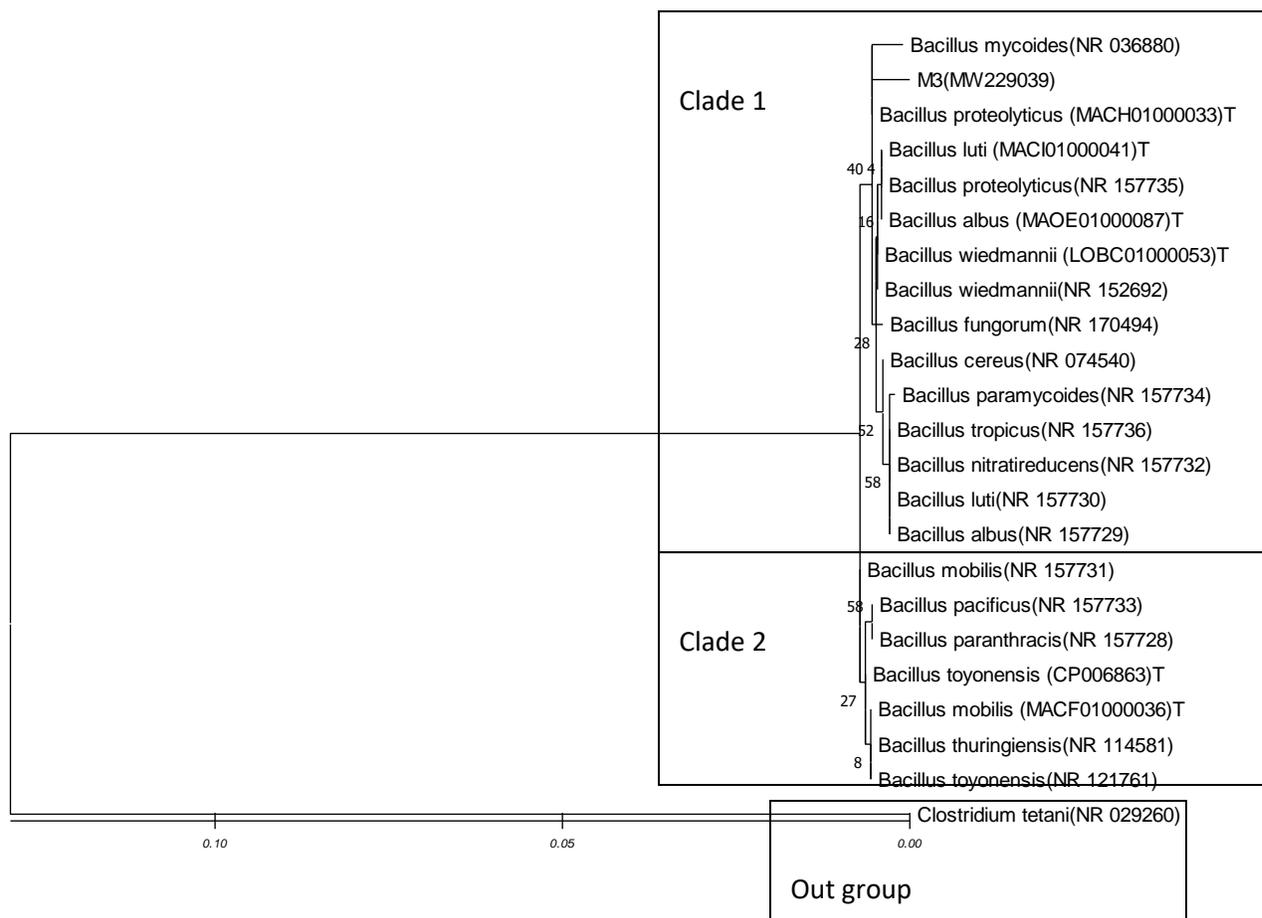


Figure 33 : Arbre obtenu par Analyse évolutive par la méthode du Maximum de Vraisemblance.

Légende : L'histoire de l'évolution a été déduite en utilisant la méthode du maximum de vraisemblance et le modèle à 2 paramètres de Kimura [30]. L'arbre avec la probabilité de log la plus élevée (-3303,19) est affiché. Le pourcentage d'arbres dans lesquels les taxons associés se sont regroupés est indiqué à côté des branches. Les arbres initiaux pour la recherche heuristique ont été obtenus automatiquement en appliquant les algorithmes Neighbor-Join et BioNJ à une matrice de distances par paires estimées à l'aide de l'approche Maximum Composite Likelihood (MCL), puis en sélectionnant la topologie avec une valeur de probabilité de log supérieure. Le modèle de variation de taux a permis à certains sites d'être évolutivement invariables ([+I], 48,92 % de sites). L'arbre est dessiné à l'échelle, avec des longueurs de branches mesurées en nombre de substitutions par site. Cette analyse portait sur 23 séquences nucléotidiques. Il y avait un total de 1555 positions dans l'ensemble de données final. Des analyses évolutives ont été menées dans MEGA X [23, 33].

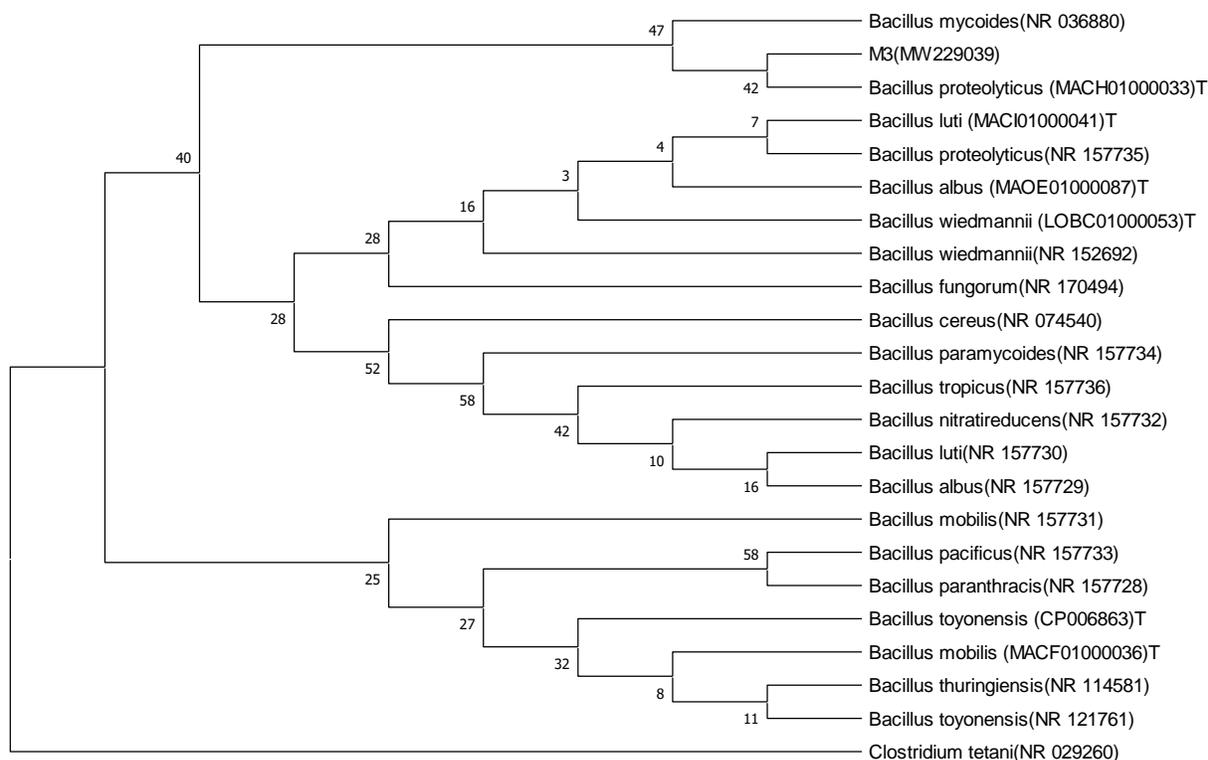


Figure 34: Arbre phylogénétique en mode « topologie seulement ».

III-4-3- Interprétation et discussion

L'arbre phylogénétique est un diagramme des relations évolutives entre les microorganismes à travers des séquences nucléotidiques ou protéiques analysées, constituant des marqueurs moléculaires conservés au fil du temps au sein des groupes de microorganismes proches. Ces séquences sont représentées sur ce que nous appelons les pointes de l'arbre.

Les pointes sont reliées par une ligne horizontale et une ligne verticale à un nœud, le nœud rassemble les clades qui sont connectés de la manière la plus similaire en termes de séquences, le nœud représente l'ancêtre commun et les pointes d'un même clade sont les séquences descendant de cet ancêtre.

Dans notre cas, l'arbre phylogénétique est représenté par le programme MEGA X en utilisant la séquence d'intérêt nommée M3 (MW229039) et des séquences homologues issues des deux bases NCBI (souches non types) et EzBioCloud (souches types) choisies en fonction de leur pourcentage de similitude. Elles appartiennent toutes à différentes espèces du genre *Bacillus* sp. On distingue principalement deux clades (Clade 1 et 2 sur la figure 34) qui sont des groupes de séquences descendant du même ancêtre.

Notre séquence M3 se trouve sur l'un des deux clades, descendant avec les séquences de *Bacillus proteolyticus* et *Bacillus mycoides* d'un même ancêtre. Il est supposé qu'elles sont toutes liées entre elles selon la topologie de l'arbre. La valeur du bootstrap (vérifiée sur l'arbre à topologie seulement) est de 47 désignant la robustesse qu'on accorde à la position de la branche portant notre souche sur l'arbre via les 1000 tests fait du bootstrap. Ça signifie qu'il y a 53 autres possibilités où la séquence de M3 prend d'autres positions alternatives,

Ces valeurs du bootstrap sont faibles et inférieur à 70 signifiants qu'on ne peut pas faire confiance à cette position sur l'arbre. La qualité de la séquence, l'utilisation d'une seule copie du gène d'ARNr16S et l'utilisation de ce gène seul peuvent être les causes.

Le niveau de divergence entre les séquences dans les clades est proportionnel à la longueur des branches horizontales, ça indique le taux de l'évolution entre ces séquences. L'absence de cette branche signifie que les séquences sont identiques.

L'enracinement de l'arbre sur un groupe externe suffisamment mais pas très éloigné de la souche à identifier illustre convenablement cette idée. La branche horizontale reliant le groupe externe *Clostridium tetani* aux deux clades des séquences homologues appartenant au genre *Bacillus* représente la distance évolutive entre ces deux genres bactériens. Cependant, les branches très courtes/absentes reliant les séquences qui se trouvent sur les deux autres clades indiquent qu'il y'a peu/pas de divergence entre les espèces choisies pour représenter l'arbre.

Ça peut signifier que la distance entre les espèces du genre n'a pas était élucidée par l'étude du gène d'ARNr16S.

Donc en conclusion à l'analyse de l'arbre phylogénétique, la taxinomie de notre bactérie est la suivante :

Tableau 5 : Les rangs taxonomiques de la souche M3.

Domain	<i>Bacteria</i>
Phylum	<i>Firmicutes</i>
Classe	<i>Bacilli</i>
Ordre	<i>Bacillales</i>
Famille	<i>Bacillaceae</i>
Genre	<i>Bacillus</i>

Afin de préciser davantage la position taxinomique d'une bactérie, il faut avoir recours :

- au séquençage en utilisant plus d'amorces universelles (ça permet d'avoir plusieurs copies d'*ARNr16S* et construire une séquence consensus)
- séquencer d'autres gènes, comme les gènes de ménages tels que : *pdhC* (codant pour le pyruvate déshydrogénase), *adk* (adénylate kinase), *gyrB* (la sous-unité B de l'ADN gyrase).
- pourquoi pas aussi le génome en entier.

Inhérentes à cette méthode nécessitent d'être soulevées. L'inconvénient lié à l'utilisation du gène de l'*ARNr 16S* est celui dû au fait que ce gène ait souvent un nombre variable de copies. En effet, certaines espèces peuvent posséder plusieurs opérons ribosomiques dans leur génome, ce qui ne permet pas une étude quantitative appropriée, ce critère est insuffisant pour certaines espèces, notamment la classification des espèces bactériennes étroitement proches peut être très difficile. Il faut donc explorer si les autres gènes (*ARN 23S* et *ARN 5S*) de l'opéron ribosomal ou certains gènes de ménage offrent une meilleure alternative pour la classification phylogénétique de genre *Bacillus*.

III-4-4- la comparaison des résultats des deux approches phénotypique et phylogénétique

La caractérisation phénotypique par les techniques de caractérisation morphologique des colonies isolées et l'observation microscopique des bactéries à l'état frais, de même que l'analyse métabolique de profil respiratoire et enzymatique, issues d'un mémoire antécédent (étude de la diversité bactérienne d'un sol pollué par les hydrocarbures), montrent l'appartenance de la souche M3 au genre *Bacillus* sp. (Voir la clé dichotomique en figure 15). Ces résultats sont cohérents avec résultats de l'alignement sur les deux bases de données NCBI et EzBioCloud, ou les souches présentant un pourcentage de 98% avec notre souche appartiennent au genre *Bacillus*. L'analyse de l'arbre représenté par la méthode du maximum de vraisemblance (ML) confirme la position taxinomique de la souche M3 dans le genre *Bacillus* mais sans distinction de l'espèce. En effet, le gène de l'*ARNr 16S* n'a pas permis de distinguer clairement la position de la souche M3 parmi les espèces du genre *Bacillus* donc sa résolution à ce niveau est médiocre.

Conclusion

Dans l'objectif d'identifier un isolat particulier, et de le placer dans une catégorie afin de connaître sa signification au monde, il est nécessaire de l'ajouter à un système de taxonomie existant, pour qu'il puisse être étudié et utilisé au profit de demain qu'il convient. La taxonomie est une branche fondamentale dans la recherche en microbiologie. Elle met en évidence une très grande variété des échelles internationales et des points de repère et références de microorganismes sur la planète Terre. Chaque effort de recherche conduit à une variété d'isolats directement ou indirectement, enrichissant ainsi les connaissances sur la biodiversité, chacune est unique et contribue à sa manière dans l'environnement.

Actuellement, il n'existe toujours pas de système officiellement reconnu pour la classification des procaryotes. En effet, pour identifier une bactérie, les approches comprennent des analyses phénotypiques, largement utilisées d'une façon préliminaire pour conduire l'identification dans la bonne direction, ça comprend en premier lieu l'étude des propriétés morphologiques. En outre, les caractéristiques de la cellule sur un plan physiologique et biochimiques sont également et de même décrites via des techniques standardisées. Cependant, cette approche reste toujours insuffisante pour une distinction exhaustive d'un isolat particulier.

Dans la découverte des liens entre la phylogénie et les caractères phénotypiques, la taxonomie parvient à un système de classification basé sur le concept de la phylogénie évolutive. Ce concept basé sur des technologies modernes purement moléculaires qui retracent les liens de parenté entre les individus d'un groupe donné et leurs ancêtres permettant une résolution toujours plus élevée et une classification plus concluante, en spécifiant le genre l'espèce et la souche.

La disponibilité sans précédent des séquences du génome signifie qu'une taxonomie peut être construite sur un cadre évolutif complet, un objectif à long terme des taxonomistes. Cependant, il existe des difficultés à l'adoption d'un cadre unique pour préserver l'enchaînement taxonomique. Le défi à venir est alors de parvenir à un consensus sur le cadre taxonomique et d'adapter le chemin phylogénétique existant, ou de créer un nouveau chemin, pour intégrer systématiquement les nouveaux taxons au cadre choisi.

Dans ce travail, notre approche été polyphasique en combinant les résultats d'une analyse phénotypique précédente sur une bactérie avec l'analyse phylogénétique d'une séquence de son ARNr 16S par nous-même pour voir à quel niveau peut-on identifier cet isolat inconnu, et est-ce que les résultats aboutissent au même taxon.

En parallèle, cette modeste recherche servira de feuille de route pour tout débutant faisant usage des bases de données et logiciels de bioinformatique, montrant suffisamment de détails sur l'alignement d'une séquence partielle d'*ADNr16S* aux bases de données NCBI et EzBioCloud et la reconstruction d'un arbre phylogénétique par logiciel MEGA X.

Les pourcentages de similitudes obtenus en alignant notre séquence aux bases de données précitées et l'arbre phylogénétique reconstruit nous ont permis de conclure que notre bactérie appartient au genre *Bacillus* sp. Ce résultat est cohérent avec la caractérisation phénotypique établi par l'étude précédente du même isolat ceci dit que il peut y'avoir une correspondance entre l'identification phénotypique et phylogénétique à un certain niveau, mais ceci n'est pas le cas pour tous les genres bactériens.

L'approche basée sur l'analyse du gène *ARNr 16S* a fait son apparition. Elle demeure jusqu'à présent l'une des approches d'identification et de classification les plus utilisées. Cependant, elle reste insuffisante particulièrement pour les espèces bactériennes étroitement proches, tel est le cas pour les *Bacillus*.

Comme perspectives, il sera souhaitable d'explorer d'autres gènes tels que les gènes de ménages pour une meilleure classification.

Références Bibliographiques

1. **Abe, Anne, Assanvo, Justine, Moussa, Sanogo, Koffi, Kouamé**, (2018), Caractérisation phénotypique de 52 souches des Bacillus isolées à partir de racines fraîches de manioc cultivées en Côte d'Ivoire. *International Journal of Biological and Chemical Sciences*, 12, 2284-2293. 10.4314/ijbcs.v12i5.28.
2. **Anissa Oussalah, Racim Louadfel**, (2014), étude de la diversité bactérienne d'un sol pollué par les hydrocarbures, encadrés par Madame Guergouri I, Université des Frères Mentouri Constantine.
3. **Benmechernene, Zineb, Mebrouk, Kihal, Miloud, Hadadji, Eddine, HENNI**, (2015), Livre les methodes moleculaire utilisées dans l'identification des microorganismes.
4. **Bergkessel. M, Guthrie. C**, (2013), Colony PCR. *Methods in enzymology*, 529, 299–309, <https://doi.org/10.1016/B978-0-12-418687-3.00025-2>.
5. **Beroud C**, (2010-2011), Bases de données et outils bio-informatiques utiles en génétique, Collège National des Enseignants et Praticiens de Génétique Médicale, Univ. Médicale Virtuelle Francophone, p 3-6.
6. **BIOSEQ**, Bioinformatics activity, using blast [en ligne], (page consultée le 15/09/2021), <https://ase.tufts.edu/chemistry/walt/sepa/Activities/BLASTpractice.pdf>.
7. **Bisen Prakash S, Dirk Gevers, Peter Dawyndt1, Peter Vandamme1, Anne Willems1, Marc Vancanneyt1, Jean Swings1 and Paul De Vos**, (2015), *Microbial Taxonomy Stepping stones towards a new prokaryotic taxonomy*, IK International, New Delhi p 196-259.
8. **Brenner D.J, Staley J.T, Krieg N.R**, (2005), *Classification of Procaryotic Organisms and the Concept of Bacterial Speciation*, *Bergey's Manual of Systematic Bacteriology*. Springer, Boston, MA, p 27-28, https://doi.org/10.1007/0-387-28021-9_4.
9. **Brian Hedlund, Svetlana Dedysh, Holger Daims, Andreas Brune and Susumu Sakata**, (2019), *Bergey's Manual Trust* [en ligne], (page consultée le 18/08/2021).
10. **C Herreman, D Asselin, G Novel**, (1994), Identification et caractérisation de souches industrielles du genre Bifidobacterium par 2 méthodes de taxonomie moléculaire. *Le Lait*, INRA Editions, 74 (6), p 479-484.

11. **Canada.ca Agence canadienne d'inspection des aliments**, (2020), Identification et classification taxonomique des micro-organismes présentés pour usage comme suppléments en vertu de la Loi sur les engrais, [en ligne], (page consultée le 17/08/2021).
12. **Canada.ca**, (2017), Cadre d'identification microbienne pour l'évaluation des risques, [en ligne], (page consultée le 18/08/2021).
13. **Canada.ca**, (2021), environnement changement climatique services gestion pollution, Substances nouvelles : organismes vivants, [en ligne], (page consultée le 08/09/2021), <https://www.canada.ca/fr/environnement-changement-climatique/services/gestion-pollution/evaluation-substances-nouvelles/biotechnologie-organismes-vivants/cadre-identification->.
14. **Céline Brochier, Jacques van Helden**, Bio-informatique appliquée : Construction des arbres phylogénétiques, (cours), [en ligne], (page consultée 25/09/2021), http://pedagogix-tagc.univ-mrs.fr/courses/bioinfo_intro/BI4U2/ppt_pdf_emese/BI4U2_phylogenie.pdf.
15. **Céline Villeneuve**, (2012), Caractérisation et classification des nouvelles espèces.
16. **Charlebois. P**, (2007), Automatisation des étapes informatiques du séquençage d'un génome d'organite et utilisation de l'ordre de gènes pour analyses phylogénétiques. Univ. Laval, Québec, p 23-25.
17. **Chun J, Rainey FA**, (2014), Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int J Syst Evol Microbiol*, 64(Pt. 2), 316-324, Doi: 10.1099/ijms.0.054171-0. PMID: 24505069.
18. **Cupić. M, Lazarević. I, Knezević. A, Stanojević. M**, (2007), Conventional and molecular methods in diagnosis and monitoring of viral infection, *Srp Arh Celok Lek*; 135(9-10):589-93. Serbian. PMID: 18088048.
19. **Dauga, Catherine, Dore, Joel, Sghir, Abdelghani**, (2005), La diversité insoupçonnée du monde microbien. *médecine/sciences*, 21. 10.1051/medsci/2005213290.
20. **Dr. Torche. E**, (2016), systématique des procaryotes [support de cours], Faculté : science de la nature et de la vie et science de terre et de l'univers, université 8 Mai 1945-Guelma.
21. **EzBioCloud**, [en ligne]. Disponible sur : <https://www.ezbiocloud.net/>, (consulté le 11/09/2021).

22. -**F. Uhell, L. Zafrani**, (2019), Nouvelles techniques de biologie moléculaire, Volume 28, Numéro 6, Méd. Intensive Réa (2019) 28:464-472, <https://doi.org/10.3166/rea-2019-0119>.
23. **Felsenstein J**, (1985), Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.
24. **G. Prod'hom, J. Bille**, (2006), Diagnostic bactériologique rapide : des méthodes conventionnelles aux méthodes moléculaires modernes, Institut de microbiologie, centre hospitalier universitaire Vaudois, Bugnon 48, 1011 Lausanne, Suisse, 180–186.
25. **Guy ROEHRIG**, (2021), Les méthodes rapides en microbiologie. Une opportunité pour tout un chacun au sein de l'entreprise, [en ligne], (page consultée le 18/08/2021).
26. **Hugenholtz, Philip et al**, (2021), "Prokaryotic taxonomy and nomenclature in the age of big sequence data." *The ISME journal* vol. 15.7: 1879-1892. Doi: 10.1038/s41396-021-00941-x.
27. **Introduction à la reconstruction phylogénétique**, (2019), (support de cours), [en ligne], (page consultée 25/09/2021), http://www.m2p-bioinfo.ups-tlse.fr/site/images/8/80/Introduction_Phylogenie_2019_M2Diag.pdf.
28. **Joëlle Brodeur, Martin Toussaint, Projetbleu**, (2006), Mini-préparation d'ADN plasmidique, Extrait de *Biologie moléculaire*, page 85, CCDMD, Catégorie : Pédagogie, Numéro : 55521.
29. **Kaya-Ongoto, Moïse, Kayath, Aimé Christian**, (2018), Identification rapide des bactéries du genre *Bacillus* en utilisant une nouvelle génération des gènes de ménage codant pour les enzymes fibrinolytiques, 10.13140/RG.2.2.18527.71843.
30. **Kimura M**, (1980), A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.
31. **Konstantinidis K.T, Stackebrandt E**, (2013), *Defining Taxonomic Ranks: The Prokaryotes*, Springer, Berlin, Heidelberg, p 249-254, https://doi.org/10.1007/978-3-642-30194-0_4.
32. **Kumar S, Tamura K, Nei. M**, MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput Appl Biosci*, Apr;10(2), p 189-91, doi: 10.1093/bioinformatics/10.2.189. PMID: 8019868.

33. **Kumar S, Stecher G, Li M, Knyaz C, and Tamura K**, (2018), MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, 35:1547-1549.
34. **Kurhekar, Jaya**, (2015), Role of taxonomy with reference to research in microbiology, *Biology discussion* [en ligne], (page consultée le 17/08/2021), <https://www.biologydiscussion.com/bacteria/bacterial-taxonomy/bacterial-taxonomy-meaning-importance-and-levels/54679>.
35. **Kurhekar, Jaya**, (2015), Role of taxonomy with reference to research in microbiology.
36. **Lamoril J, Ameziane N, Deybach, Jean-Charles, Bouizegarène P, Bogard, Michael**, (2008), Les techniques de séquençage de l'ADN : une révolution en marche. Première partie. *Immuno-analyse & Biologie Spécialisée*, 23, p 260-279. 10.1016/j.immbio.2008.07.016.
37. **Laura Tordjman-Valency**, (2016). Défi du dénombrement microbien dans l'industrie pharmaceutique : les nouvelles méthodes alternatives sont-elles appliquées ? *Sciences pharmaceutiques*, ffdumas01373145.
38. **Luciano Paolozzi, Jean-Claude Liébart Matthieu Arlat, Michel Dion, Harivony Rakotoarivonina**, (2019), Introduction à la microbiologie -Microbiologie fondamentale et appliquée, Dunod, 11, rue Paul Bert, 92240 Malakoff, chapitre 2 : Taxinomie des micro-organismes, p 4-11.
39. **M.Malo le Boulch**, (2019), Taxonomie et inférence fonctionnelle des procaryotes: développement de MACADAM, une base de données de voies métaboliques associées à une taxonomie [en ligne], thèse de doctorat, pathologie. Toxicologie. Génétique et nutrition, université de Toulouse, disponible sur : https://oatao.univ-toulouse.fr/25532/1/LeBoulch_Malo.pdf.
40. **Mark J, Pallen, Andrea Telatin, Aharon Oren**, (2021), The Next Million Names for Archaea and Bacteria, *Trends in Microbiology*, April 2021, Vol. 29, No. 4, p 289-298.
41. **Meddas Seyyid Ali, Oulebsir-Mohandkaci Hakima, Hadjouti Rima, Reghmit Nacer, Houas Yasmine, Aamirat Dihia, Naitbouda Lamine, Doumandjt-Mitiche Bahia**, (2010), caractérisation des *Bacillus spp* isolées du sol d'olivier et étude de leurs potentialité de biocontrolle contre la mouche de l'olive *Dacus Oleae* (GMEL) au nord de l'Algérie, Vol-11 N-01, Université Saad Dahlab de Blida, Algérie, 2507-7627.

42. **Megasoftware**, Molecular Evolutionary Genetics Analysis, [en ligne], (page consultée 25/09/2021), <https://www.megasoftware.net/>.
43. *Methylophaga nitratreducenticrescens* et *Methylophaga frappieri*, Université du Québec, p10-11.
44. -**Michael K. Dahl**, (2014), *Bacillus*, Encyclopedia of Food Microbiology, Vol 1, p 113–119, Elsevier Ltd.<http://dx.doi.org/10.1016/B978-0-12-384730-0.00018-5>.
45. **Microbiologie clinique**, (2019), Techniques microbiologiques ; recherche de la lécithinase, [en ligne], (page consultée 25/09/2021), <https://microbiologie-clinique.com/1%C3%A9cithinase.html>.
46. **Moore ER, Mihaylova SA, Vandamme P, Krichevsky MI, Dijkshoorn L**, (2010), Microbial systematics and taxonomy: relevance for a microbial commons. Res Microbiol, 161(6):430-8, Doi: 10.1016/j.resmic.2010.05.007. Epub 2010 Jun 1. PMID: 20670913.
47. **Mouna Cheikh Rouhou**, (2006), Évaluation des classifications phylogénétiques des Bacillaceae basées sur les gènes de l'opéron et de gènes de ménage, Université du Québec à Montréal, p 6-7.
48. **National Center for Biotechnology Information**, U.S. National Library of Medicine, [en ligne], Disponible sur : <https://www.ncbi.nlm.nih.gov/>, (consulté le 11/09/2021).
49. -**NCBI REF : NCBI Resource Coordinators**, (2013), Database resources of the National Center for Biotechnology Information, Nucleic Acids Research, Vol 41, Issue D1, p 8–20, <https://doi.org/10.1093/nar/gks1189>.
50. **Newell P. D, Fricker A. D, Roco C. A, Chandrangsu P, Merkel S. M**, (2013), A Small-Group Activity Introducing the Use and Interpretation of BLAST. Journal of microbiology & biology education, 14(2), 238–243. <https://doi.org/10.1128/jmbe.v14i2.637>.
51. **Niall A, Logan, Paul De Vos**, (2015), *Bacillus*, Firmicutes/"Bacilli"/Bacillales/Bacillaceae, Bergey's Manual of Systematics of Archaea and Bacteria, in association with Bergey's Manual Trust, Doi: 10.1002/9781118960608.gbm00530.

52. **Per E.J. Saris, Lars G. Paulin, Mathias Uhlén**, (1990), Direct amplification of DNA from colonies of *Bacillus subtilis* and *Escherichia coli* by the polymerase chain reaction, *Journal of Microbiological Methods* 11 (1990), p 121 – 126.
53. **Philippe Lopez, Didier Casane, Hervé Philippe**, (2002), Phylogénie et évolution moléculaires, *Phylogénie, Bio-informatique et Génome*, Cnrs UMR 7622, Université Pierre et Marie Curie, Paris, France, Vol 18, N 11, p 1146 – 1154, <https://doi.org/10.1051/medsci/200218111146>.
54. **Radha S, Fathima A.A, Iyappan S, et al**, (2013), Direct colony PCR for rapid identification of varied microalgae from freshwater environment. *J Appl Phycol* 25, 609–613. <https://doi.org/10.1007/s10811-012-9895-0>.
55. **Ranga.nr**, (2013), what is the Importance of Taxonomy| Its role in Biology & Biodiversity, Study Read [en ligne], (page consultée le 17/08/2021), [https://www.studyread.com/importance-of-taxonomy-biology/February 10, 2013](https://www.studyread.com/importance-of-taxonomy-biology/February%2010,%202013) by Ranga.nr.
56. **REGNUM PROKARYOTE, ABIS ENCYCLOPEDIA**, Genus *Bacillus*, Genus *Paenibacillus*, [en ligne], (consultée : 13/09/2021), <https://www.tgw1916.net/Bacillus.html>.
57. **Rossello-Mora R, Amann R**, (2001), The species concept for prokaryotes, *FEMS Microbiology Reviews* 25 (2001) 39-67.
58. **Sarethy I, Sharadwata Pan, M Danquah**, (2014), *Modern Taxonomy for Microbial Diversity*.
59. **Schleifer. K**, (2009), Classification of Bacteria and Archaea: past, present and future, *Systematic and applied microbiology*, *Schleifer2009ClassificationOB*, Vol (32, 8), p 42-533.
60. **Vandamme P, Pot B, Gillis M, Vos P, Kersters K, Swings J**, (1996), Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev*, 60(2):407-38, Doi: 10.1128/mr.60.2.407-438.1996, PMID: 8801440; PMCID: PMC239450.
61. **Véronique Lafarge**, (2006), Identification des microorganismes du lait par des méthodes moléculaires et incidence des traitements thermiques et de la microflore du lait sur la cinétique de croissance de *Listeria monocytogenes*, Thèse de doctorat en Stratégies

d'exploitation des fonctions biologiques, à Compiègne , en partenariat avec AgroParisTech (France).

62. **Wheeler D, Bhagwat M**, (2007), BLAST QuickStart: Example-Driven Web-Based BLAST Tutorial. In: Bergman NH, editor. Comparative Genomics: Volumes 1 and 2. Totowa, (NJ):Chapter9.Availabl from:
<https://www.ncbi.nlm.nih.gov/books/NBK1734/>.
63. **Yoann. M**, (2012), Les arbres phylogénétiques : construction et interprétation, [en ligne], (page consultée 25/09/2021), <https://bioinfo-fr.net/les-arbres-phylogeniques-construction-et-interpretation>.
64. **Yoon S. H, Ha S. M, Kwon S, Lim J, Kim Y, Seo H, Chun J**, (2017), Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. International journal of systematic and evolutionary microbiology, 67(5), 1613–1617. <https://doi.org/10.1099/ijsem.0.001755>.

Mémoire présenté en vue de l'obtention du Diplôme de Master

Filière : Ecologie et environnement
Spécialité : Ecologie microbienne

Titre

Une démarche à suivre pour l'identification d'une bactérie inconnue

Résumé

La combinaison de caractères phénotypiques et phylogénétiques s'avère importante dans le cas de la détermination du genre bactérien d'un isolat inconnu. C'est une approche polyphasique que l'on qualifie de « consensus ». Ce présent travail est une feuille de route montrant les différentes étapes de représentation graphique d'un arbre phylogénétique à partir d'une séquence partielle d'*ARNr16S* et de séquences homologues. Il montre suffisamment de détails pour permettre à une personne débutante d'entreprendre ce genre d'analyses, faisant usage des bases de données telles que NCBI et EZ-BioCloud, et des logiciels de bioinformatique tels que MEGA X. Les résultats d'une étude antérieure de caractérisation phénotypique et d'identification d'un isolat inconnu, conduisant au genre bactérien *Bacillus* sp. étaient comparés à ceux obtenus par reconstruction phylogénétique de l'arbre afin de voir la correspondance des deux approches et est-ce que la phylogénie positionne la bactérie au sein du même genre. Les résultats obtenus par les deux approches phénotypique et phylogénétique ont conduit à une affiliation au genre *Bacillus*.

Mot clés : Mot clés : approche polyphasique, arbre phylogénétique, genre *Bacillus*, Alignement de la séquence, *ARNr 16S*.

Membre du jury :

Mme. GUERGOURI Ibtissem

Mr. BENHIZIA Yacine

Mme. BOULTIFAT Linda

Présenté par :

HEDDOUD Houcine Islem et HAMADA Ayoub

Année universitaire : 2020-2021