



Université Mohamed Khider de Biskra
Faculté des sciences exactes et sciences de la nature et de
la vie
Département des sciences de la nature et de la vie

MÉMOIRE DE MASTER

Domaine : Sciences de la nature et de la vie

Filière : Sciences biologiques

Spécialité : Microbiologie appliquée

Réf. :

Présenté et soutenu par :
Cheriet Roukia et Chelbi Hadjer

Le : lundi 10 juin 2024

Biodiversité bactérienne : Métagénomique et classification phylogénétique des ARNr 16S

Jury :

| | | | |
|------------------------------|-----|----------------------|------------|
| Mme. BOUKHAROUBA Khadidja | Pr | Université de Biskra | Président |
| M. MOUSSI Abdelhamid | Pr | Université de Biskra | Rapporteur |
| M. ATHAMNA Ahmed | MCB | Université de Biskra | Examineur |

Année universitaire : 2023/2024

Remerciements

*Je tiens tout d'abord à remercier **ALLAH***

Qui ma a donné le courage et la force et la patience pour terminer ce mémoire.

*Un grand merci à notre encadreur le professeur **MOUSSI Abdelhamid** pour son encadrement et pour avoir inspiré et suivi avec beaucoup d'intérêt ce travail, son soutien et sa confiance ont permis l'accomplissement de ce mémoire et surtout pour sa disponibilité malgré ses occupations.*

Nous remercions par ailleurs les membres du jury, président et examinateur, de nous avoir fait l'honneur de présider et de juger notre travail.

*Nos professeurs et tous les enseignants du département des Sciences de la Nature et de la Vie
De L'Université Mohamed Khider.*

Ainsi toute personne qui a participé de près ou de loin à l'exécution de ce modeste travail.

Toute l'équipe de bibliothèque De L'université pour ses aides et pour leurs soutiens,

*Un grand merci à mes amis (es) et collègues de département **sciences de la nature et vie**.
Cette page ne serait été complète sans remercier mes parents qui m'ont apporté toute l'aide dont j'avais besoin. Et mes frères.*

Merci à tous et à toutes...

Dédicace

C'est avec l'aide et la grâce du Dieu que j'ai achevé ce modeste travail

*Je dédie ce travail à l'âme de mon père, **CHeriet Mahmoud**,*

Qui a été mon modèle et mon soutien à chaque étape de ma vie, à m'a enseigné le sens de la vie et la valeur de la science et du travail, à celui qui m'a quitté physiquement, mais dont le bon esprit vit toujours dans mon cœur et m'accompagne à chaque instant, Que Dieu tout puissant ait pitié et t'accorde le repos éternel.

*A ma chère mère **Ben Kahla Fatiha***

, la lumière de ma vie, qui a toujours veillé pour moi, qui a toujours été à mes côtés dans les moments de joie et de peine, dont les prières ont été le secret de ma réussite et de mon succès, et sans laquelle je n'aurais pas atteint cet accomplissement, ni pu réaliser mes rêves, que Dieu te soutienne et te récompense de tout ce qui est bon.

*À mon cher frère **Mohammed Ridha***

Merci, Pour ton soutien, et d'avoir toujours été à mes côtés.

*Mes belles sœurs **Farida, Rania** et ma petite sœur **Nadjia**.*

Qui ont été une source de soutien et d'amour et ont partagé avec moi les bons et les mauvais moments.

*Ames chères amis **Wuiame, Chourouk, Amal** et **Zayneb**.*

Ceux qui ont rendu mon parcours universitaire plein de beaux souvenirs et de temps merveilleux.

Aux personnes qui m'ont toujours aidé et encouragé.

Roukia

Dédicace

Je dédie ce modeste travail à :

Mes chers parents qui n'ont ménagé ni leurs efforts ni leurs sacrifices pour m'amener jusqu'à ce jour, je vous présente le fruit de votre travail et de vos efforts

*À mon grand-père, le moudjahid **Ben Salem Saad**, qui a passé sa vie à servir son pays avec sincérité et dévouement*

À ma chère grand-mère dans sa tombe, une femme forte et patiente qui s'est battue pour la libération de ce pays.

Que Dieu leur fasse miséricorde

*À mes chers frères **Hamza, Fath elddine, et Salah***

*Et mes chères sœurs **Soumia et Tita***

qui ont toujours été mon soutien et mon guide, je vous adresse tout mon amour et mes remerciements.

À mes honorables professeurs

À toute ma famille

À tous mes amis, À mes amis et collègues, partenaires dans ce voyage, Je vous félicite et partage ce moment spécial avec vous.

À tous ceux qui ont cru en moi et m'ont encouragé à persévérer.

Enfin, à tous ceux qui ont eu un impact positif sur mon parcours éducatif, je dédie cette note en signe de gratitude.

Et sans oublier :

*Au pays de la liberté, de l'honneur et de la dignité, **Palestine**.*

*Aux enfants de **Gaza**, aux martyrs, aux blessés, aux orphelins et aux réfugiés.*

Notre prière pour eux est que Dieu leur accorde bientôt la victoire.

Hadjer

Clarification

Dans la rédaction de la thèse de master, nous avons utilisé divers outils et ressources pour améliorer la qualité et la clarté du texte. Nous avons notamment fait usage de plateformes de traduction en ligne pour vérifier la précision des termes et expressions utilisés, ainsi que des outils d'intelligence artificielle pour améliorer la formulation des phrases et corriger les erreurs grammaticales et syntaxiques. Ces moyens ont été précieux pour garantir la qualité du contenu rédactionnel et assurer sa compréhension optimale.

Sommaire

| | |
|------------------------------|-----|
| Remerciements | |
| Dédicace | |
| Liste des tableaux | I |
| Liste des figures | II |
| Liste des abréviations | III |
| Introduction générale | 1 |

PARTIE I. BIBLIOGRAPHIE

CHAPITRE 1 : GENERALITES SUR LA METAGENOMIQUE

| | |
|---|---|
| 1.1. Définition..... | 3 |
| 1.2. Historique | 3 |
| 1.3. Différentes approches de la métagénomique :..... | 4 |
| 1.3.1. Métagénomique globale (shotgun, plein-génome)..... | 4 |
| 1.3.2. Métagénomique ciblé (<i>Metabarcoding</i>) | 4 |
| 1.4. Avantages de la métagénomique..... | 5 |
| 1.5. Applications de la métagénomique..... | 6 |
| 1.5.1. Application médicale | 6 |
| 1.5.2. Environnements naturels | 6 |
| 1.5.3. Application agricole | 6 |
| 1.5.4. En biotechnologie | 6 |

CHAPITRE 2 : BIODIVERSITE

| | |
|----------------------|---|
| 2.1. Définition..... | 7 |
|----------------------|---|

| | |
|---|---|
| 2.2. Niveaux de la biodiversité | 7 |
| 2.2.1. Diversité génétique | 7 |
| 2.2.2. Diversité spécifique | 7 |
| 2.2.3. Diversité écosystémique | 8 |
| 2.3. Méthodes d'analyse en écologie microbienne | 8 |
| 2.3.1. Approches microbiologiques classiques | 9 |
| 2.3.2. Approches métagénomique | 9 |

PARTIE II. EXPERIMENTALE

CHAPITRE 3: MATERIEL ET METHODES

| | |
|--|----|
| 3.1. Matériel..... | 10 |
| 3.1.1. Présentation du matériel | 10 |
| 3.1.2. Construction de l'arbre phylogénétique | 10 |
| 3.1.2.1. Collecte des Données | 11 |
| 3.1.2.2. Alignement des séquences | 13 |
| 3.1.2.3. Ajustement de l'Alignement (Alignment Curation)..... | 14 |
| 3.1.2.4. Construction de l'Arbre phylogénétiques | 14 |
| 3.1.2.5. Visualisation de l'Arbre phylogénétiques | 14 |
| 3.1.3. Choix et construction de l'arbre | 15 |
| 3.1.3.1 Cluster Mapping and Clustering with Entropy (BMGE) | 15 |
| 3.1.3.2. Construction d'un arbre phylogénétique..... | 16 |
| 3.1.4 Délimitation des espèces..... | 16 |
| 3.1.4.1. Assemble Species by Automatic Partitioning (ASAP) | 17 |
| 3.2. Taxonomie | 18 |

CHAPITRE 4: RESULTATS ET DISCUSSION

| | |
|---|----|
| 4.1. Téléchargement des séquences | 19 |
| 4.2. Identification et vérification des séquences | 19 |
| 4.3. Alignement des séquences | 19 |
| 4.4. Construction d'arbre phylogénétique | 20 |

4.5. Délimitation des espèces 23

Conclusion..... 37

Références bibliographiques..... 38

Resumés

Liste des Tableaux

| | |
|---|----|
| Tableau 1. Caractéristiques des ordinateurs utilisées | 10 |
| Tableau 2. Résultats du taxonomie des espèces à partir la méthodes de délimitation « ASAP » de la partition 03. | 28 |

Liste des Figures

| | |
|--|----|
| Figure 1. Vue d'ensemble de l'approche globale de la métagénomique. | 4 |
| Figure 2. Vue d'ensemble de l'approche ciblée de la métagénomique. | 4 |
| Figure 3. Régions conservées (gris) et hypervariables (bleu) de l'ADNr 16S bactérien. | 5 |
| Figure 4. Numéros d'accessions des séquences nucléotidiques bactériennes déposées par le chercheur Adjeroud dans une base de données GenBank. | 11 |
| Figure 5. Résultats de l'alignement BLAST pour l'échantillon MH394069.1 avec la souche " <i>Tepidimonas charontis</i> ". | 12 |
| Figure 6. Alignement multiple des séquences par la méthode Muscle par MEGA 11. | 13 |
| Figure 7. Service web NGPhylogeny.fr | 15 |
| Figure 8. Utilisation de BMGE et de l'arbre PhyML sur le service web NGphylogeny.fr | 16 |
| Figure 9. Partie de l'alignement des séquences montre les similitudes et les divergences. | 20 |
| Figure 10. Arbre phylogénétique de la séquence étudiées par la méthode ML, Les couleurs indiquent le regroupement des séquences après l'analyse réalisée par l'approche ASAP (partition 3)..... | 23 |
| Figure 11. Affichage du graphique des sous-ensembles encadrés dans les résultats de l'outil web ASAP. | 26 |

Liste des abréviations

ABGD : Automatic Barcode Gap Discovery

ADNr : Acide Désoxyribonucléique ribosomal

ARNr : Acide ribonucléique ribosomal

ASAP: Assemble Species by Automatic Partitioning

BLAST: Basic Local Alignment Search Tool

BMGE: Block Mapping and Gathering with Entropy

bPTP: bayesian Poisson Tree Processes

GMYC: Generalized Mixed Yule-Coalescent

MEGA: Molecular Evolutionary Genetics Analysis

ML: Maximum Likelihood « maximum de vraisemblance »

mPtP: multirate Model the Poisson Tree Processes

NCBI: National Centre for Biotechnology Information

NGS: Next-Generation Sequencing

NIH: National Institutes of Health

NLM: National Library of Medicine

OTU: Operational Taxonomic Units

PCR: Polymerase Chain Reaction

UICN : Union Internationale pour la Conservation de la Nature

Introduction générale

Introduction générale

Depuis la première utilisation du terme en 1998, la métagénomique est devenue un domaine innovant et intégré de l'écologie microbienne. La métagénomique est un concept qui consiste à analyser le matériel génétique extrait directement d'échantillons de l'environnement. Le progrès de la technologie du séquençage de nouvelle génération (NGS) a ouvert la possibilité de réaliser des analyses du métagénome sur la base de séquences. Les plates-formes de la génomique de nouvelle génération permettent aux scientifiques de collecter des données à haut débit à faible coût, ce qui favorise le développement des analyses métagénomiques fondées sur les séquences et destinées à décrire des communautés microbiennes dans leur totalité (Dudhagara, et al., 2015).

La métagénomique est une association de la génomique, de la bioinformatique, et de la biodiversité des écosystèmes qui peut être employée pour étudier les génomes de plusieurs organismes en même temps (Luiza et Dumitru 2021). Grâce aux outils de la bio-informatique, la phylogénétique a permis de définir de nouvelles espèces bactériennes. Les biologistes, les informaticiens et les bio-informaticiens travaillent ensemble pour analyser, classer et interpréter les données biologiques (Vert, 2013).

Le séquençage du gène de l'ARN ribosomal 16S est crucial pour l'étude de la phylogénie et de la taxonomie bactériennes, surtout lorsque les méthodes biochimiques classiques échouent. Cette méthode permet d'identifier le genre bactérien dans plus de 90% des cas et l'espèce bactérienne dans 65 à 85% des cas. Toutefois, entre 1 et 14% des isolats peuvent rester non identifiés. Globalement, le séquençage de l'ARNr 16S surpasse les méthodes bactériologiques traditionnelles (Lamoril et *al.*, 2008).

L'objectif de notre travail de recherche est d'identifier et de classer des séquences inconnues d'ARNr 16S afin de les attribuer à une famille, un genre ou une espèce spécifique. Pour ce faire, nous nous concentrons sur l'analyse de séquences d'ARNr 16S obtenues à partir de la base de données GenBank.

Plus précisément, notre étude vise à :

1. Rechercher, sélectionner et télécharger des séquences d'ARNr 16S des unités taxonomiques opérationnelles (OTUs) à partir de l'outil BLAST.
2. Utiliser des outils bioinformatiques pour aligner et traiter ces séquences.
3. Construire un arbre phylogénétique représentant les OTUs.

4. Analyser cet arbre phylogénétique à l'aide de la méthode de délimitation ASAP (Assemble Species by Automatic Partitioning).

Cette approche permettra non seulement de localiser les séquences inconnues, mais aussi de les classer précisément, améliorant ainsi notre compréhension des relations phylogénétiques et de la diversité des espèces représentées.

Première partie :
Partie bibliographie

Chapitre 1 : Généralités sur la métagénomique

1.1. Définition

« Meta “ signifie ” après “ et ” génomique » correspond à l'étude du contenant complet de l'ADN d'un organisme (Chiriac et *al.*, 2021).

Ce terme consiste en une analyse génomique d'un groupe de micro-organismes par extraction directement de l'ADN d'une communauté de bactéries (Handelsman, 2004).

La métagénomique permet de ne pas avoir à isoler ou à cultiver des micro-organismes. Les stratégies métagénomiques basées sur un isolement direct des acides nucléiques à travers des échantillons environnementaux se sont montrées des méthodes puissantes pour l'exploration de nouvelles souches et de communautés de bactéries complexes. Ces communautés sont souvent composées de germes non cultivables (99 %) des bactéries présentes dans la majorité des environnements (Simon et Daniel, 2011 ; Diene et *al.*, 2014).

1.2. Historique

L'étude des communautés microbiennes est très ancienne, mais elle s'est longtemps limitée à l'utilisation de techniques d'imagerie qui ne permettaient que l'observation de caractéristiques morphologiques.

Dans ce contexte, seuls les organismes cultivables peuvent être testés. Ainsi, avant les techniques de biologie moléculaire, seule une petite partie des micro-organismes existants pouvait être étudiée à une faible résolution. L'ARN ribosomique a été décrit comme un indicateur pour la classification des espèces. Ces méthodes ont été utilisées pour les données environnementales par Pace et al, qui ont suggéré en 1986 le séquençage direct de l'ARN ribosomique dans le milieu, sans qu'il soit nécessaire de cultiver des bactéries. Le concept de métagénomique fait référence au séquençage direct de l'ADN dans un environnement donnant potentiellement accès aux génomes de tous les membres d'une communauté (Guyomar et Lemaitre, 2023).

Par la suite, le terme métagénomique a progressivement été utilisé pour faire référence à des études telles que la génomique environnementale, la génomique des communautés, la génomique écologique, la génomique environnementale et la génomique microbienne. L'objectif de cette nouvelle approche de l'écologie microbienne est de mieux déchiffrer la diversité microbienne présente dans les niches environnementales. Vous pourrez mieux comprendre les relations évolutives des différents taxons déterminés et caractériser des nouveaux gènes et des nouvelles fonctions (Pauline et *al.*, 2013).

1.3. Différentes approches de la métagénomique :

Il y a deux types de données métagénomiques utilisées pour mener ces analyses :

1.3.1. Métagénomique globale (shotgun, plein-génome)

La métagénomique du génome entier ou métagénomique implique le séquençage de tout l'ADN d'un échantillon. Au lieu d'amplifier des régions spécifiques du génome, l'ADN entier du génome ou du métagénome est découpé au hasard en fragments et séquençé à l'aide de techniques traditionnelles à haut débit. Cela signifie que nous pouvons séquencer les génomes de tous les membres de notre communauté (Guyomar et Lemaitre, 2023).

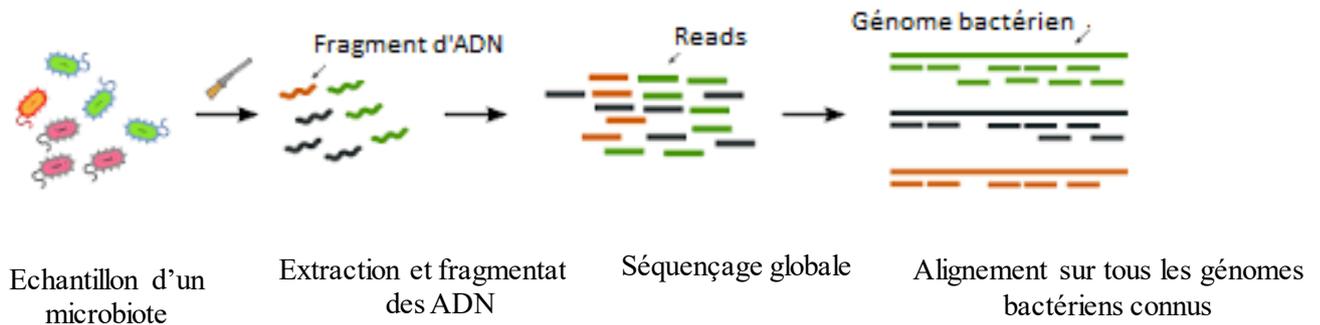


Figure 1. Vue d'ensemble de l'approche globale de la métagénomique (Schutz, 2016).

1.3.2. Métagénomique ciblée (*Metabarcoding*)

La première approche est une stratégie ciblée qui implique l'amplification et le séquençage d'une ou plusieurs parties précises d'un ou plusieurs génome(s) particulier(s). Cette approche est particulièrement utilisée pour la détermination des microbes (tel que, les gènes codant pour l'ARNr 16S afin d'identifier les bactéries) ou la recherche de résistance (par exemple, les gènes liés à des mutations de résistance chez les virus) (Marchand et *al.*, 2024).

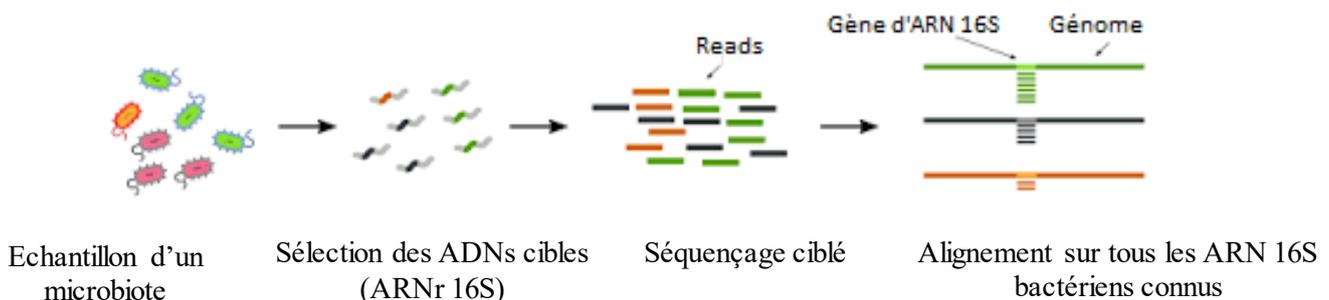


Figure 2. Vue d'ensemble de l'approche ciblée de la métagénomique (Schutz, 2016).

ARNr 16S

Généralement, les marqueurs moléculaires sont considérés comme des études de codes-barres et/ou des études taxonomiques ou évolutives. Bien que plusieurs régions ou gènes aient été proposés, l'acide ribonucléique ribosomal 16S (ARNr 16S) a été initialement proposé par Pace et al en 1986, où il a été proposé comme un bon choix pour la classification bactérienne. L'idée a été rapidement adoptée par la communauté scientifique et les séquences d'ARNr 16S ont été utilisées pour construire des bases de données spécialisées. Ce qui précède permet aux séquences d'ARNr 16S de servir d'outil important pour reconstruire les relations phylogénétiques. L'ARNr 16S est un polyribonucléotide d'environ 1 500 nucléotides codé par le gène de l'ADN ribosomal 16S. Cependant, l'ARNr est suffisamment variable pour distinguer non seulement les organismes les plus éloignés, mais également les plus proches, et peut distinguer les espèces, les souches ou les variantes (Fabiola et al., 2015).

L'ARNr 16S contient neuf régions moins conservées ou hypervariables (V1-V9) qui fournissent les informations les plus utiles pour les études phylogénétiques et taxonomiques (figure 3). Les régions conservées sont d'une grande aide dans la conception d'amorces universelles capables d'amplifier différentes régions hypervariables de l'ARNr 16S de la grande majorité des micro-organismes présents dans la communauté (Fabiola et al., 2015).

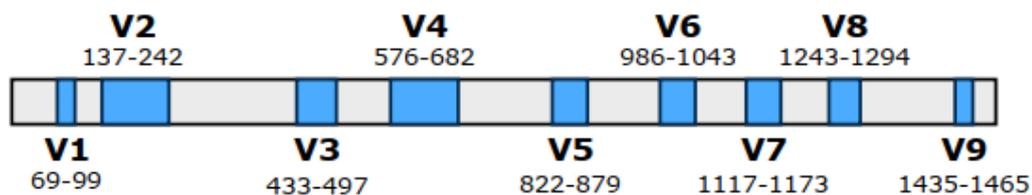


Figure 3. Régions conservées (gris) et hypervariables (bleu) de l'ADNr 16S bactérien (Siegwald, 2017).

1.4. Avantages de la métagénomique

La métagénomique offre plusieurs avantages, notamment :

Étude de la biodiversité : Permet d'analyser la diversité génétique de communautés microbiennes complexes sans besoin de cultiver les micro-organismes en laboratoire (Sauvage et al., 2017).

Découverte de nouvelles espèces : Facilite l'identification de nouvelles espèces de micro-organismes, y compris celles qui sont difficiles ou impossibles à cultiver (Sauvage et al., 2017).

Séquençage en parallèle pour une efficacité optimisée : Le séquençage métagénomique permet d'analyser simultanément plusieurs échantillons au cours d'une seule opération de séquençage. Cette capacité de séquençage réduit considérablement les coûts, rendant cette méthode plus économique pour les études à grande échelle (Lamoril et *al.*, 2008).

1.5. Applications de la métagénomique

Les applications de la métagénomique démontrent l'étendue et la diversité des domaines où cette technologie peut être mise en œuvre. Elle ouvre des perspectives innovantes tant pour la recherche scientifique que pour les développements industriels, en permettant des avancées significatives dans des secteurs variés tels que la santé, l'écologie, l'agriculture, l'industrie alimentaire, et la biotechnologie (Diene et *al.*, 2014).

1.5.1. Application médicale

L'emploi de méthodes de métagénomique permet de comprendre par exemple comment le microbiote intestinal humain fonctionne. De plus, la découverte de traits microbiologiques associés à diverses affections a donné lieu à l'élaboration d'approches préventives, diagnostiques ou thérapeutiques novatrices en médecine, comme la *cirrhose* du foie (Calenge et *al.*, 2017).

1.5.2. Environnements naturels

Analyse de la diversité microbienne dans des écosystèmes variés, tels que les océans, les sols, et les eaux douces, pour comprendre les cycles biogéochimiques et les interactions écologiques (Faure et Joly, 2016).

1.5.3. Application agricole

Étude des communautés microbiennes du sol pour améliorer la fertilité et la santé des cultures (Adjeroud, 2021).

1.5.4. En biotechnologie

La métagénomique présente un potentiel économique notable en raison de la disponibilité de nouvelles enzymes et molécules bioactives tels que les antibiotiques et les anticancéreux. En intégrant les méthodes de criblage à haut débit, elle révèle une grande diversité de nouveaux biocatalyseurs, ou molécules, dans le domaine des applications biotechnologiques et pharmaceutiques. De nombreux environnements extrêmes ont été explorés dans la littérature pour identifier de nouveaux biocatalyseurs qui sont employés dans différentes industries comme la fabrication de biocarburants et la dégradation de plastiques (Adjeroud, 2021).

Chapitre 2 :

Biodiversité

2.1. Définition

Le terme « biodiversité » a été inventé au début des années 1970 par l'Union internationale pour la conservation de la nature (UICN). Il fait généralement référence au nombre total d'espèces (plantes, animaux terrestres et marins, champignons et micro-organismes) qui existent dans l'ensemble de la biosphère ou de l'écosystème, mais il englobe également la diversité à différents niveaux (diversité génétique, populations et sous-populations d'espèces et communautés écologiques mixtes). Le terme « biodiversité » fait référence à la variété et à la variabilité des diverses formes de vie et communautés écologiques qu'elles habitent (Ramade, 2008 ; Rodriguez *et al.*, 2023).

2.2. Niveaux de la biodiversité

La biodiversité est essentiellement une question d'intégration, car elle englobe la variabilité génétique des populations, la diversité des espèces et la diversité fonctionnelle des écosystèmes.

2.2.1. Diversité génétique

La variation génétique est un indicateur de la santé d'une population. Une moindre diversité génétique au sein d'une population entraîne des risques de consanguinité et réduit la durée de vie des individus, des populations et des espèces. À l'inverse, une population démographiquement saine illustre une dynamique positive de renouvellement génétique. Cette diversité génétique se manifeste dans chaque individu, reflétant la diversité morphologique et physiologique (phénotypique) des organismes (Harold, 2007 ; Ramade, 2008).

2.2.2. Diversité spécifique

La diversité des espèces est d'une grande importance et représente une référence fondamentale dans toutes les mesures de protection de la biodiversité. Le nombre d'espèces d'un groupe systématique donné au sein d'un écosystème particulier est en effet d'une grande importance en écologie, qui sert à protéger la nature et ses ressources (Ramade, 2008).

Contrairement à la diversité génétique, qui rend l'évaluation de la biodiversité plus complexe, la richesse et la diversité des espèces au sein d'une communauté dépendent d'un grand nombre de paramètres. Ces paramètres comprennent l'hétérogénéité de l'habitat, la latitude géographique et la résilience des écosystèmes (Harold, 2007).

En biologie de la conservation, on distingue trois catégories générales de diversité des espèces : La diversité au niveau d'une communauté est appelée diversité alpha, et la diversité

entre les communautés est appelée diversité bêta. Généralement définie comme un gradient mesurable et quantitatif à travers le paysage (par exemple, le pH du sol, l'humidité, etc.), la diversité gamma se réfère à la diversité unique que l'on trouve dans les larges paysages (Virginie, 2006).

2.2.3. Diversité écosystémique

An assemblage dynamique consistant d'une communauté d'organismes et de leur milieu abiotique est appelé un écosystème (Virginie, 2006).

Le niveau supérieur de l'écosystème (ou diversité écologique) est inclus dans l'échelle de la biodiversité. En réalité, ce qui le distingue le plus du nombre d'espèces qu'il contient, ce sont les caractéristiques résultant de la disposition des espèces qui composent sa communauté. Des particularités qui différencient l'écosystème d'un autre émergent de ce regroupement d'espèces interconnectées propres à un écosystème donné (Ramade, 2008).

2.3. Méthodes d'analyse en écologie microbienne

La capacité à déterminer le nombre et le type de micro-organismes au sein d'une communauté est essentielle pour comprendre la structure et la fonction d'un écosystème. La forme simple de la plupart des micro-organismes fournit peu d'indices pour leur identification et leurs caractéristiques physiologiques sont souvent floues. En outre, de nombreux organismes sont résistants à la culture, ce qui est nécessaire pour les caractériser. La technologie de l'ADN recombinant a permis de dépasser bon nombre des difficultés liées à la culture et à la caractérisation et a fourni depuis lors une multitude de nouvelles informations sur la complexité des communautés microbiennes naturelles. Les approches moléculaires utilisant l'analyse des séquences d'ARN ribosomique (ARNr) 16S permet d'étudier en direct la structure, la diversité et l'évolution des communautés microbiennes dans presque tous les environnements. L'hybridation acide nucléique peut également être utilisée pour quantifier des espèces individuelles de micro-organismes ou des communautés microbiennes entières. De plus, l'utilisation de sondes d'ARNr spécifiques à une population marquée par fluorescence permet l'examen microscopique de cellules individuelles et d'interactions *in situ* au sein d'assemblages microbiens complexes (Theron et Cloete, 2000).

Il existe plusieurs façons d'étudier différentes communautés microbiennes au sein d'un habitat :

2.3.1. Approches microbiologiques classiques

Après la découverte des boîtes de Pétri, celles-ci ont été largement utilisées pour la culture en laboratoire. Les colonies peuvent être isolées par des coupes répétées sur des milieux solides (ou en les versant dans des boîtes ou des tubes d'agar). Les cellules peuvent également être isolées par dilution dans un milieu liquide. L'enrichissement et les cultures pures sont utilisés pour sélectionner les micro-organismes d'intérêt. Cependant, les milieux et les conditions utilisés dans les cultures de laboratoire ne reproduisent souvent pas les conditions naturelles, ce qui empêche la croissance des microbes. Cette méthode classique semble être plus efficace pour les micro-organismes à forte croissance, qui constituent une petite proportion de toutes les communautés microbiennes naturelles. Il est donc nécessaire de développer de nouvelles méthodes pour cultiver une plus large gamme de micro-organismes (Van et Jaisoo, 2012).

2.3.2. Approches métagénomique

La prise de conscience du fait que la plupart des micro-organismes de l'environnement ne peuvent pas être cultivés à l'aide de méthodes standard a stimulé le développement de la métagénomique, ou analyse génomique des micro-organismes non cultivés. Deux types d'analyse sont utilisés pour récupérer des informations à partir des banques métagénomiques. L'une est l'approche basée sur les fonctions, où les banques métagénomiques sont d'abord sélectionnées sur la base des traits exprimés, et l'autre est l'approche basée sur les séquences, où les banques sont d'abord sélectionnées sur la base de séquences d'ADN spécifiques. De nouveaux antibiotiques et enzymes ont été parmi les premières découvertes en métagénomique. Les améliorations futures des méthodes d'enrichissement en gènes dotés de fonctions spécifiques accéléreront la découverte de molécules utiles (Patrick et Handelsman, 2003).

Deuxième partie :
Partie expérimentale

Chapitre 3 :

Matériel et Méthodes

Pour atteindre les objectifs de notre travail, nous avons étudié les différentes séquences génétiques des bactéries trouvées par le chercheur Adjeroud Moussa dans le cadre de son étude sur un échantillon d'une source thermale prélevé à la station thermale « Hammam Essalihine » dans la wilaya de Khenchela. Cette recherche, publiée dans sa thèse intitulée « Investigation métagénomique de la biodiversité des Bactéries et Archaea productrices de substances bioactives à partir de quelques sources hydrothermales en Algérie », se concentre sur la biodiversité des bactéries et des archées produisant des substances biologiquement actives. Le chercheur a extrait l'ADN génomique d'échantillons d'eau à 68-70°C et ciblé la région V4-V5 de l'ARNr pour amplification, séquençage et analyse.

3.1. Matériel

3.1.1. Présentation du matériel

Cette étude a été effectuée sur des outils en bioinformatiques sur des plateformes en ligne et des applications (installé), à partir deux simples ordinateurs, ses caractéristiques sont détaillées dans le tableau ci-dessous :

Tableau 1. Caractéristiques des ordinateurs utilisées

| Les caractéristiques | Ordinateur 1 | Ordinateur 2 |
|------------------------|---|--|
| Nom de l'appareil | DESKTOP-9CGVMTG | DESKTOP-GPT352O |
| Processeur | Intel(R) Core (TM) i3-6006U CPU @ 2.00GHz 2.00 GHz | Intel(R) Pentium(R) Dual CPU T2370 @ 1.73GHz 1.73 GHz |
| Version | 22H2 | 21H2 |
| Système d'exploitation | Windows 10 Professionnel 64 bits | Windows 10 Professionnel 64 bits |

3.1.2. Construction de l'arbre phylogénétique

La construction d'un arbre phylogénétique est un processus complexe qui permet de représenter les relations évolutives entre différentes espèces, autres unités taxonomiques ou différents gènes (Taguan, 2015).

Ce processus se déroule selon les étapes clés suivantes :

3.1.2.1. Collecte des Données

Une recherche dans GenBank (<https://www.ncbi.nlm.nih.gov/nucleotide>) des séquences nucléotidiques bactériennes déposées par le chercheur sous le nom Adjeroud a révélé 120 séquences bactériennes avec des numéros d'accessions numérotées de MH394069 à MH394186 + (MK309929 et MK309930) (Figure 4).

Toutes ces séquences sont signalées par une définition : Clone A d'une bactérie non cultivée, gène de l'ARN ribosomal 16S, séquence partielle. Elles ne sont assignées à aucun groupe taxonomique (Taxon), mais seulement sous à un organisme bactérien non cultivé.

The screenshot shows the GenBank Nucleotide search interface. The search criteria are 'Nucleotide' and 'Adjeroud'. The results show 120 items, with the first two items displayed: 'Uncultured bacterium clone M01945_72_000000000-BD7H6_1_2104_11672_8269_16S_ribosomal' and 'Uncultured bacterium clone M01945_72_000000000-BD7H6_1_2101_5497_9839_16S_ribosomal'. The search details show the query: 'Adjeroud[All Fields] AND bacteria[filter]'.

Figure 4. Numéros d'accessions des séquences nucléotidiques bactériennes déposées par le chercheur Adjeroud dans une base de données GenBank.

GenBank

Est une vaste base de données accessible au public qui contient des séquences de nucléotides ainsi que des annotations biologiques et bibliographiques. Le National Center for Biotechnology Information (NCBI), une branche de la National Library of Medicine (NLM), est responsable de la construction et de la distribution de la GenBank. La NLM est située sur le campus des National Institutes of Health (NIH) à Bethesda, dans le Maryland (Dennis et al., 2017).

Elle comprend des séquences nucléaires accessibles au public pour plus de 340 000 espèces partiellement décrites. Est disponible sur son site web <https://www.ncbi.nlm.nih.gov/> (Clark et al., 2016).

Nous avons ajouté 75 séquences identifiées soit au niveau de l'espèce, du genre, ou de la famille, en les soumettant toutes les séquences à logiciel BLAST et en les téléchargeant lorsqu'elles présentaient une identité supérieure à 93 % (figure 5).

The screenshot shows the BLAST search results for job MH394069.1. The search parameters are: Job Title: gb|MH394069.1, RID: 5K72FBUP013, Program: BLASTN, Database: nt, Query ID: MH394069.1, Description: Uncultured bacterium clone M01945_72_000000000-BD71..., Molecule type: nucleic acid, Query Length: 375. The filter results section shows the organism filter set to 'only top 20 will appear' and the 'Filter' button is highlighted. The table of sequences producing significant alignments is as follows:

| Description | Scientific Name | Total Score | Query Cover | E value | Per. Ident | Accession |
|---|-----------------------|-------------|-------------|---------|------------|------------|
| Tepidimonas sp. SPSPC-18 16S ribosomal RNA gene, partial sequence | Tepidimonas charontis | 693 | 100% | 0.0 | 100.00% | MH590703.1 |
| Tepidimonas sp. SPSP-6 16S ribosomal RNA gene, partial sequence | Tepidimonas charontis | 693 | 100% | 0.0 | 100.00% | MH590702.1 |

Figure 5. Résultats de l'alignement BLAST pour l'échantillon |MH394069.1 avec la souche "*Tepidimonas charontis*"

BLAST (Basic Local Alignment Search Tool)

Le site Web du National Center for Biotechnology Information (NCBI), BLAST (Basic Local Alignment Search Tool), constitue une ressource précieuse pour l'alignement et la recherche de séquences. Il s'agit d'un outil fréquemment utilisé pour trouver des similitudes de séquences. BLAST est proposé par le National Center for Biotechnology Information (NCBI) à la fois de manière indépendante et par l'intermédiaire de son site web <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (Grzegorz et al., 2013 ; Newell et al., 2013).

Plusieurs variantes de BLAST comparent toutes les combinaisons de requêtes de nucléotides ou de protéines avec des bases de données de nucléotides ou de protéines (Ye et al., 2006).

3.1.2.2. Alignement des séquences

L'alignement multiple des séquences est le processus de comparaison stratégique des séquences d'ADN, d'ARN ou d'acides aminés afin d'identifier les similitudes et les différences. Ce processus met en évidence les zones conservées et les différences entre les séquences, ce qui peut donner des informations essentielles sur leur évolution, leur fonction et leurs relations (Pevzner, 2006).

Les programmes utilisés dans l'alignement multiple pour aligner les séquences sont ClustalW, MUSCLE ou MAFFT. Dans le cadre de notre étude, MEGA X version 11 a été utilisé pour réaliser des alignements de séquences à l'aide de la méthode MUSCLE®, disponible à son site <https://www.megasoftware.net/> (Figure 6).

MEGA

Logiciel MEGA (Molecular Evolutionary Genetics Analysis) offre une collection d'outils pour l'évolution moléculaire, tels que l'assemblage de séquences, la déduction d'arbres évolutifs et l'estimation des distances et des diversités génétiques (Kumar et al., 2018).

Dès le départ, MEGA a été développé dans le but d'être "toujours gratuit" pour le public et de faciliter les analyses statistiques de l'évolution moléculaire à l'aide d'ordinateurs personnels (Caspermeyer, 2018).

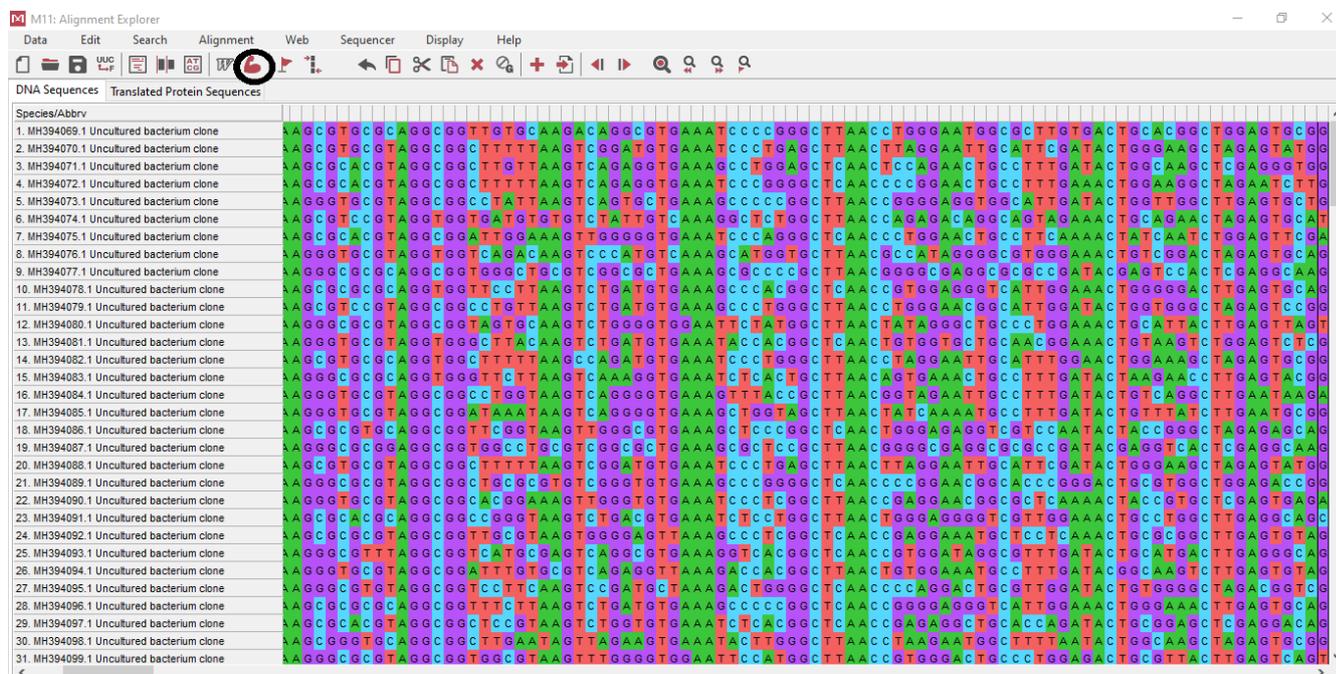


Figure 6. Alignement multiple des séquences par la méthode Muscle par MEGA 11.

3.1.2.3. Ajustement de l'Alignement (Alignment Curation)

L'ajustement de l'alignement est une étape cruciale et délicate dans la construction d'un arbre phylogénétique. Il nécessite à la fois des outils automatisés et une expertise humaine pour assurer que les séquences sont correctement alignées et que les analyses ultérieures sont basées sur des données fiables (pour corriger les erreurs ou les ambiguïtés). L'ajustement automatique de l'alignement peut être réalisé en utilisant des logiciels spécialisés. Ces outils, comme BMGE, Gblocks, Noisy ou trimAl. Ces programmes appliquent des critères statistiques et des algorithmes pour détecter et corriger les erreurs de l'alignement initial, en garantissant une meilleure qualité des données alignées (Dereeper et *al.*, 2008 ; Criscuolo et Gribaldo, 2010).

3.1.2.4. Construction de l'Arbre phylogénétiques

Plusieurs méthodes existent pour construire l'arbre phylogénétique :

- Méthode de Distance : Utilise des matrices de distances entre paires de séquences (ex. : méthode UPGMA, Neighbour-Joining) (Meyer, 1991 ; Mao, 2007).

- Méthode de Parcimonie : Cherche l'arbre avec le plus petit nombre de changements évolutifs (Meyer, 1991 ; Mao, 2007).

- Méthode de Maximum de Vraisemblance : Estime la probabilité de l'arbre donné les données et le modèle d'évolution (Meyer, 1991 ; Mao, 2007).

- Méthode Bayésienne : Utilise des techniques statistiques bayésiennes pour estimer la probabilité des arbres (Meyer, 1991 ; Mao, 2007).

3.1.2.5. Visualisation de l'Arbre phylogénétiques

Pour dessiner et visualiser l'arbre phylogénétique en représentation graphique, des logiciels utilisés comme FigTree, MEGA, Dendroscope ou la plateforme itol.embl.de.

Les arbres peuvent être enracinés ou non enracinés, selon que l'on connaît ou non l'ancêtre commun (Didier et Guindon, 2022).

ITol

Un logiciel en ligne qui est accessible à partir de n'importe quel site web récent (<https://itol.embl.de/>). iTOL fournit une diversité de services communs pour visualiser les arbres phylogénétiques et offre plusieurs formats d'affichage, tel que enracinés ou non, rectangulaire ou radial (Letunić et Bork, 2021).

3.1.3. Choix et construction de l'arbre

Pour construire un arbre phylogénétique de nos séquences collectées, nous avons utilisé le logiciel MEGA pour réaliser l'étape d'alignement. Nous avons ensuite utilisé la plateforme NGPhylogeny pour réaliser les autres étapes de (ajustement de l'alignement et construction de l'arbre) et enfin nous avons utilisé itol.embl.de pour visualiser l'arbre.

NGPhylogeny

NGPhylogeny.fr est un site web spécialisé dans l'étude de la phylogénèse. Il propose une gamme complète d'outils et de méthodes de travail qui conviennent à divers contextes et niveaux d'expertise des usuels. Il est possible de déployer NGPhylogeny.fr sur n'importe quel serveur ou ordinateur personnel, et il est accessible gratuitement sur <https://ngphylogeny.fr> (Lemoine *et al.*, 2019).



Figure 7. Service web NGPhylogeny.fr

3.1.3.1 Cluster Mapping and Clustering with Entropy (BMGE)

Conçue pour déterminer les régions dans l'alignement de séquences multiples qui conviennent à l'inférence phylogénétique, la méthode BMGE est capable d'effectuer des coupures biologiques sur des alignements multiples de séquences d'ADN, de codons ou d'acides aminés (Criscuolo et Gribaldo, 2010).

Le fichier utilisé par BMGE est un alignement de séquences multiples au format FASTA.

Donc, BMGE est un outil efficace qui peut avoir de larges utilisations dans les analyses génétiques. Il est accessible de manière gratuite (Criscuolo et Gribaldo, 2010).

3.1.3.2. Construction d'un arbre phylogénétique

Un arbre phylogénétique est une représentation graphique des relations évolutives entre différents groupes d'organismes, illustrant leurs liens de parenté (Taguan, 2015).

Ensuite, l'arbre est construit en utilisant le format d'inférence de l'arbre (phyML), puis en lançant toutes les séquences pour créer l'arbre phylogénétique (figure 7).

PhyML

Les méthodes du maximum de vraisemblance prennent en compte la longueur des branches de l'arbre phylogénétique et sont basées sur des caractères observés (Taguan, 2015).

Elles permettent de comparer un grand nombre d'arbres génétiques possibles et de sélectionner celui qui maximise la probabilité pour observer les données, en dépendant du modèle de substitution de l'ADN employé. Cette probabilité est appelée "vraisemblance" de l'arbre. (Taguan, 2015)

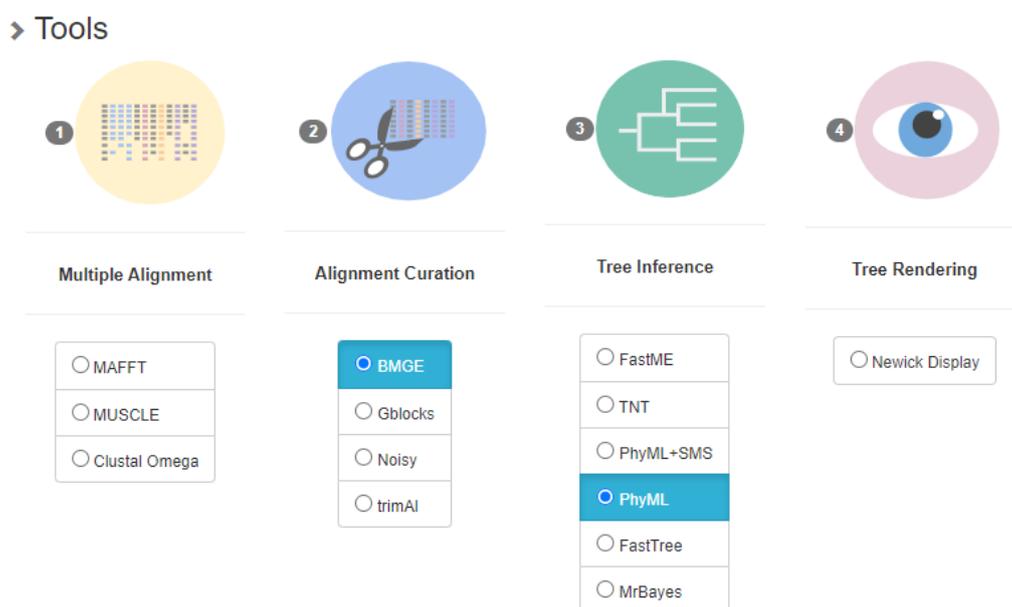


Figure 8. Utilisation de BMGE et de l'arbre PhyML sur le service web NGPhylogeny.fr

3.1.4 Délimitation des espèces

La délimitation des espèces permet de déterminer quels groupes d'organismes individuels composent différentes populations d'une même espèce et lesquels constituent des espèces

distinctes. Les données génomiques jouent un rôle essentiel dans ce processus, en particulier dans le cadre de plusieurs concepts d'espèces récents (Rannala et Yang, 2020).

Il existe plusieurs méthodes de délimitation des espèces : ASAP, ABGD, GMYC, mPtP, bPTP.

- ✓ IL y a deux méthodes basées sur la distance :
 - ASAP (Assemble Species by Automatic Partitioning) (Puillandre et *al.*, 2021).
 - ABGD (Automatic Barcode Gap Discovery) (Puillandre et *al.*, 2012).
- ✓ Trois autres basées sur l'arbre phylogénétique :
 - GMYC (Generalized Mixed Yule Coalescent) (Pons et *al.*, 2006 ; Fujisawa et Barraclough, 2013)
 - mPTP (le modèle multi-rate Poisson Tree Processes) (Kapli et *al.*, 2017)
 - bPTP (le modèle Bayesian Poisson Tree Processes) (Zhang et *al.*, 2013)

Parmi celles-ci, la méthode ASAP a été utilisée pour identifier et différencier les espèces.

3.1.4.1. Assemble Species by Automatic Partitioning (ASAP)

Définition

Une nouvelle technique de création de partitions d'espèces basée sur l'alignement de séquences à des locus individuels, elle permet de partitionner un ensemble de données en une grande séquence de 10^4 espèces putatives en quelques minutes (Puillandre et *al.* 2021).

ASAP est un logiciel indépendant qui peut être utilisé par une interface web graphique (<https://bioinfo.mnhn.fr/abi/public/asap>) ou qui est téléchargeable et disponible en version compilée pour une utilisation locale (Puillandre et *al.* 2021).

Principe

Il s'agit d'une mise en œuvre de l'algorithme de classement hiérarchique qui classe les séquences en "groupes", puis les fusionne en une séquence jusqu'à ce que toutes les séquences forment un groupe. À chaque étape de la fusion, le positionnement de toutes les séquences dans les groupes est appelé "partition". La première partition contient le nombre de groupes correspondant au nombre de séquences (non encore regroupées), alors que la dernière partition est un groupe unique contenant toutes les séquences. Les groupes plus importants sont formés en fusionnant les groupes de la partition précédente (Puillandre et *al.* 2021).

3.2. Taxonomie

GenBank a été utilisé pour classifier les espèces bactériennes incultivables. En utilisant l'outil de taxonomie, il suffit de saisir le nom du genre recherché pour obtenir sa classification taxonomique, incluant le nom de la famille, l'ordre et la classe associés à chaque genre. Le logiciel Excel a été utilisé pour établir un tableau contenant les genres et leur classification.

Unité Taxonomique Opérationnelle (OTU)

Est une méthode utilisée pour classer des groupes d'organismes en fonction de la similarité de leurs séquences génétiques. Dans les études de microbiologie et de métagénomique, les OTU sont souvent utilisées pour identifier et comparer les microorganismes présents dans des échantillons environnementaux. Chaque OTU représente un groupe de séquences de gènes (souvent des séquences d'ADN ribosomique 16S chez les bactéries) qui sont suffisamment similaires pour être considérées comme provenant d'un même groupe taxonomique, bien que le seuil de similarité puisse varier en fonction de l'étude. Les OTU permettent de simplifier l'analyse des données de séquençage en regroupant les séquences en unités comparables, facilitant ainsi l'étude de la diversité microbienne et de la composition des communautés microbiennes (Mysara *et al.*, 2017).

Chapitre 04 :

Résultats et Discussion

Ce chapitre présente les résultats obtenus après l'analyse de séquences d'ADN extraites de bases de données à l'aide d'approches bioinformatiques telles que la construction d'arbres phylogénétiques et la délimitation d'espèces.

4.1. Téléchargement des séquences

Nous avons d'abord téléchargé les séquences bactériennes publiées par Moussa Adjroud dans GenBank avec des numéros d'accès compris entre MH394069 et MH394186 +(MK309929 et MK309930). Nous avons ensuite ajouté (75 séquences) d'autres séquences présentant un degré élevé de similarité en comparant les séquences précitées un par un à l'aide du test BLAST. Les longueurs de toutes ces séquences étaient comprises entre 372 et 381 pb.

Ces séquences ont été identifiées, chacune représentant une unité taxonomique opérationnelle (OTU) (Adjroud, 2021).

4.2. Identification et vérification des séquences

Pour vérifier les différentes séquences téléchargées (120 séquences de MH394069 à MH394189+ MK309929, MK309930 et 75 séquences similaires), nous avons utilisé l'outil BLAST. Parmi ces séquences similaires ont été sélectionnées les séquences les plus proches et excluant les espèces non identifiées incultivées. De plus, une seule espèce a été retenue pour chaque genre.

4.3. Alignement des séquences

Nous avons aligné les séquences que nous avons collectées en utilisant la méthode de Muscle pour pouvoir discriminer d'une part les régions de similarité susceptibles d'être des relations fonctionnelles, structurelles ou évolutives et d'autre part les régions de divergence susceptibles de représenter différentes mutations (substitutions, délétions, insertion, ...).

Dans l'exemple ci-dessous, les mutations peuvent être représentées :

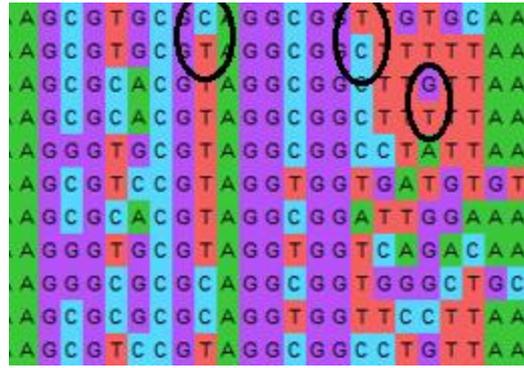
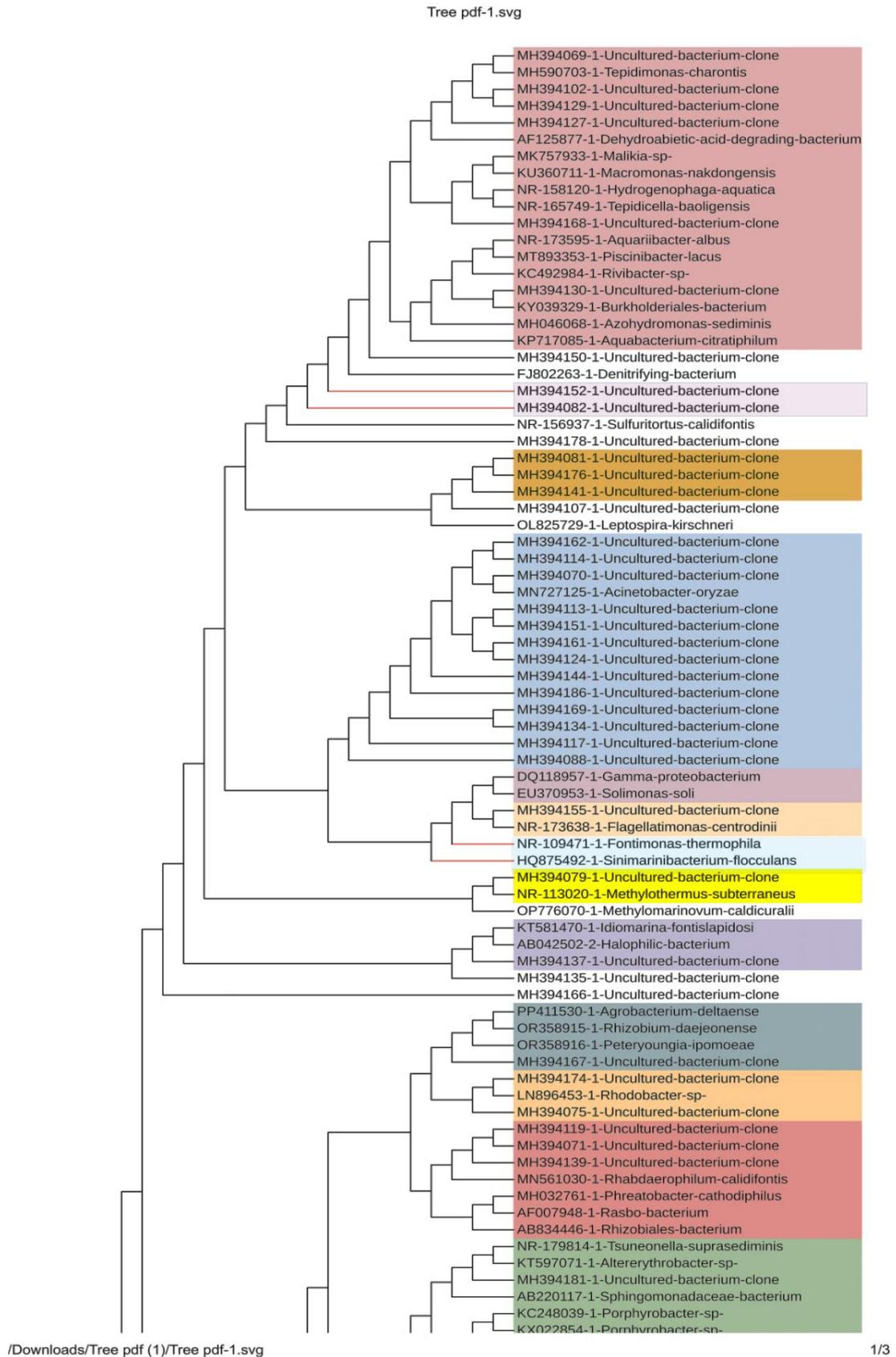


Figure 9. Partie de l'alignement des séquences montre les similitudes et les divergences.

4.4. Construction d'arbre phylogénétique

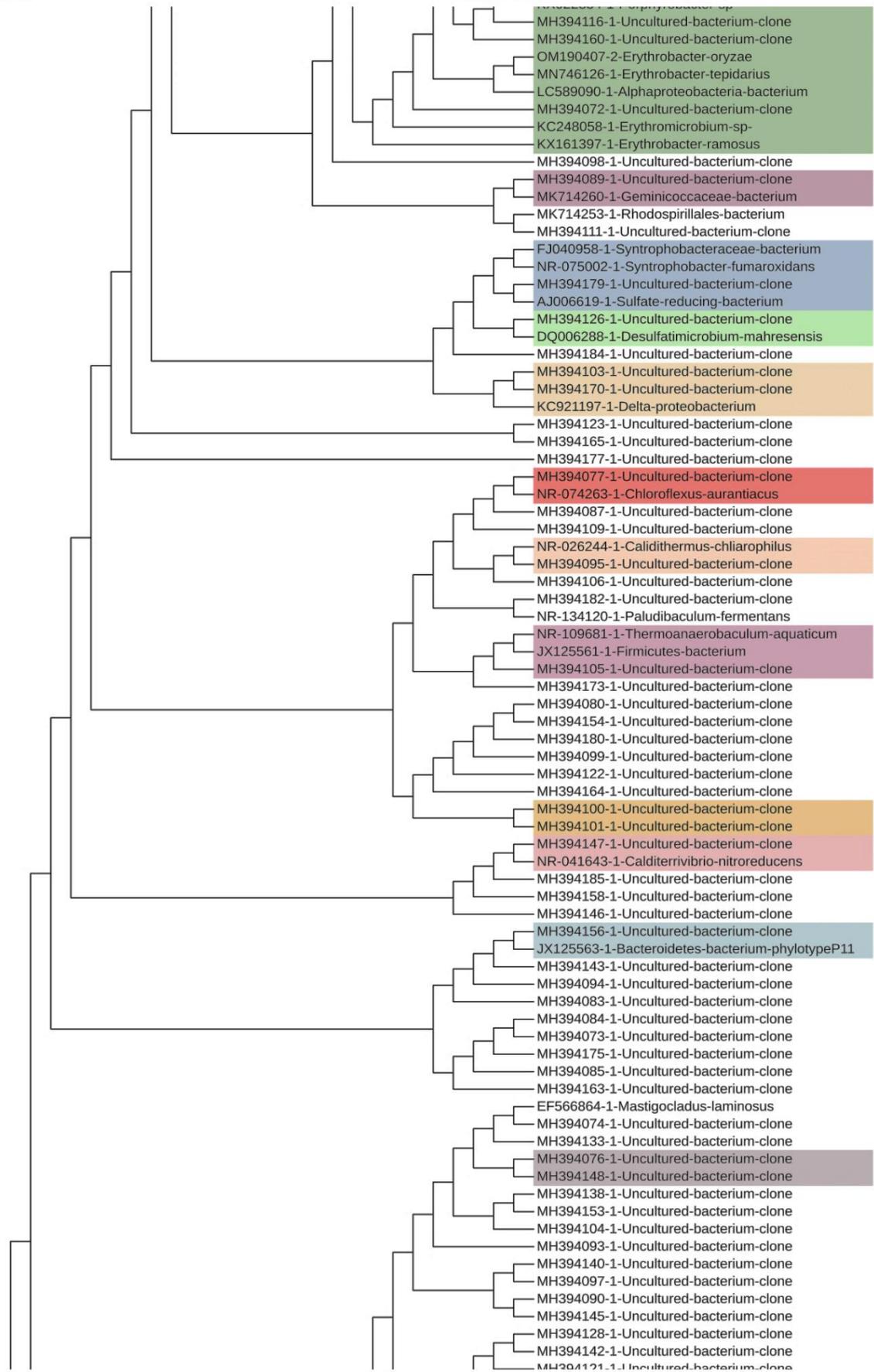
Pour construire l'arbre phylogénétique, nous avons employé la méthode du maximum de vraisemblance (ML), en suivant une série d'étapes spécifiques. L'alignement des séquences a été effectué à l'aide du logiciel MEGA X, version 11, en utilisant l'algorithme MUSCLE. Le curage des séquences alignées a été réalisé avec l'outil BMGE, accessible via la plateforme NGPhylogeny.fr. Ensuite, l'arbre phylogénétique a été construit en utilisant la méthode PhyML, également disponible sur NGPhylogeny.fr. Enfin, la visualisation de l'arbre résultant a été effectuée à l'aide de la plateforme en ligne Itool.de.

La figure 10 montre l'arbre phylogénétique des séquences étudiées à l'aide de la méthode de vraisemblance maximale ML, les clades colorés indiquent les différents groupes (subset) et chaque groupe représente nombre des espèces différentes (18, 14, 7, ...etc.) selon la partition 3 des résultats de délimitation ASAP par contre les clades non colorés représentent des groupes individuels.



24 11:13

Tree pdf-1.svg



Isers/info/Downloads/Tree pdf (1)/Tree pdf-1.svg

2/3

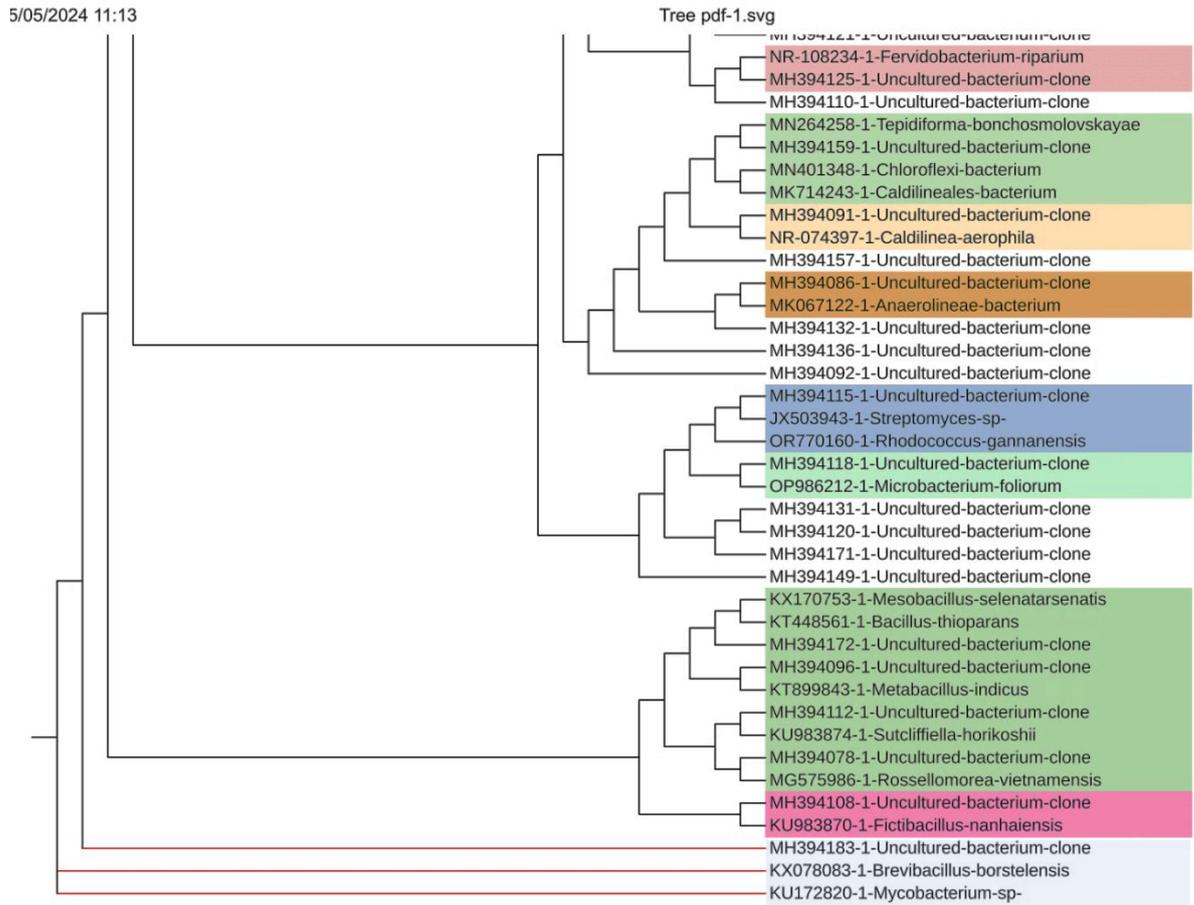


Figure 10. Arbre phylogénétique de la séquence étudiées par la méthode ML, Les couleurs indiquent le regroupement des séquences après l'analyse réalisée par l'approche ASAP (partition 3).

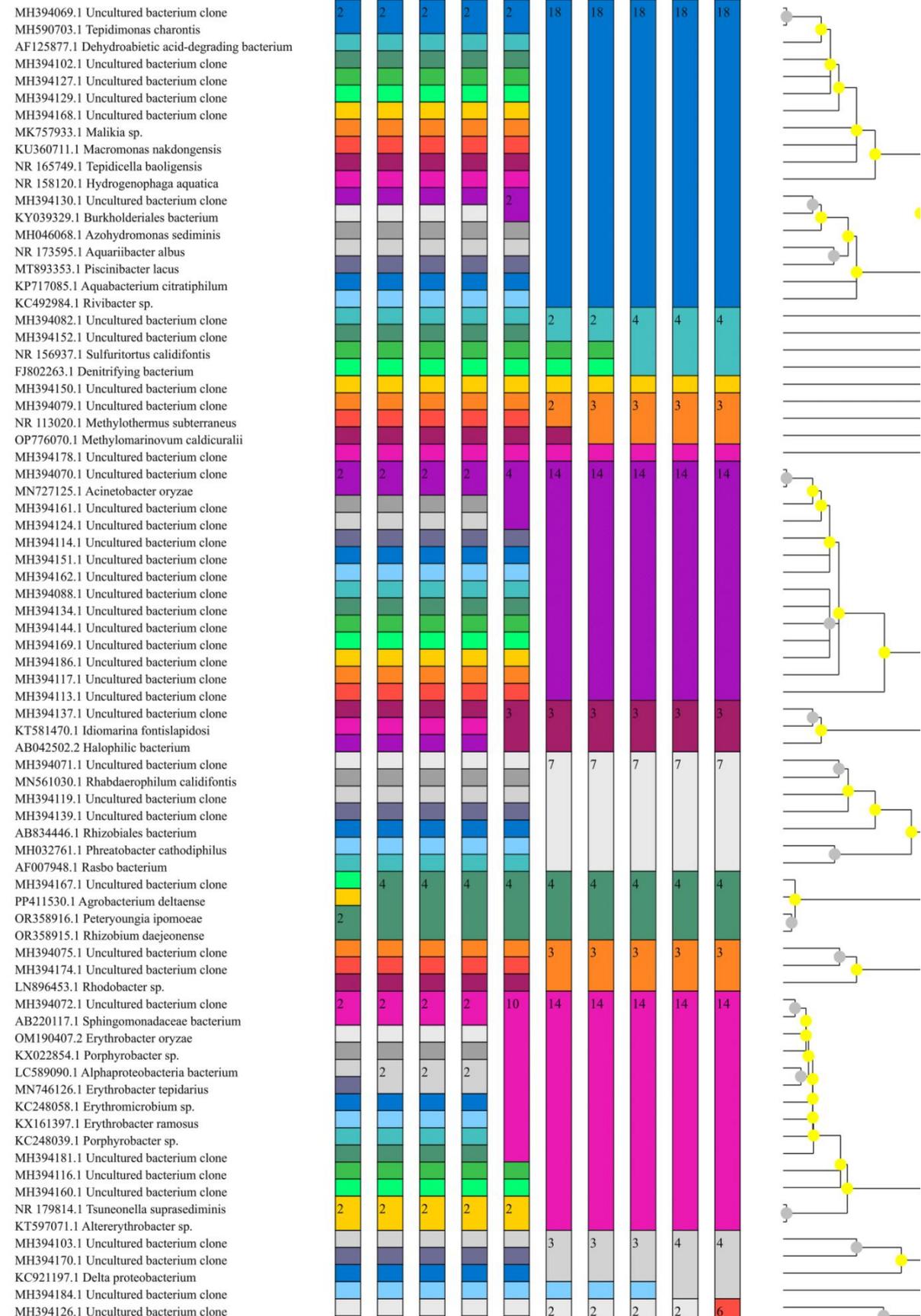
4.5. Délimitation des espèces

La méthode de délimitation des espèces ASAP a donné 10 partitions meilleures, chacune accompagnée d'un score indiquant leur robustesse. Plus le score est bas, meilleure est la partition, nous avons choisi les trois scores les plus faibles, la première partition donne 177 subsets « sous ensemble », le deuxième donne 174 subsets et le troisième donne 95 subsets.

5/23/24, 10:22 PM

Nb subsets/asap score
Rank

bioinfo.mnhn.fr/abi/public/asap/temp/31539.1804289383/groups.svg

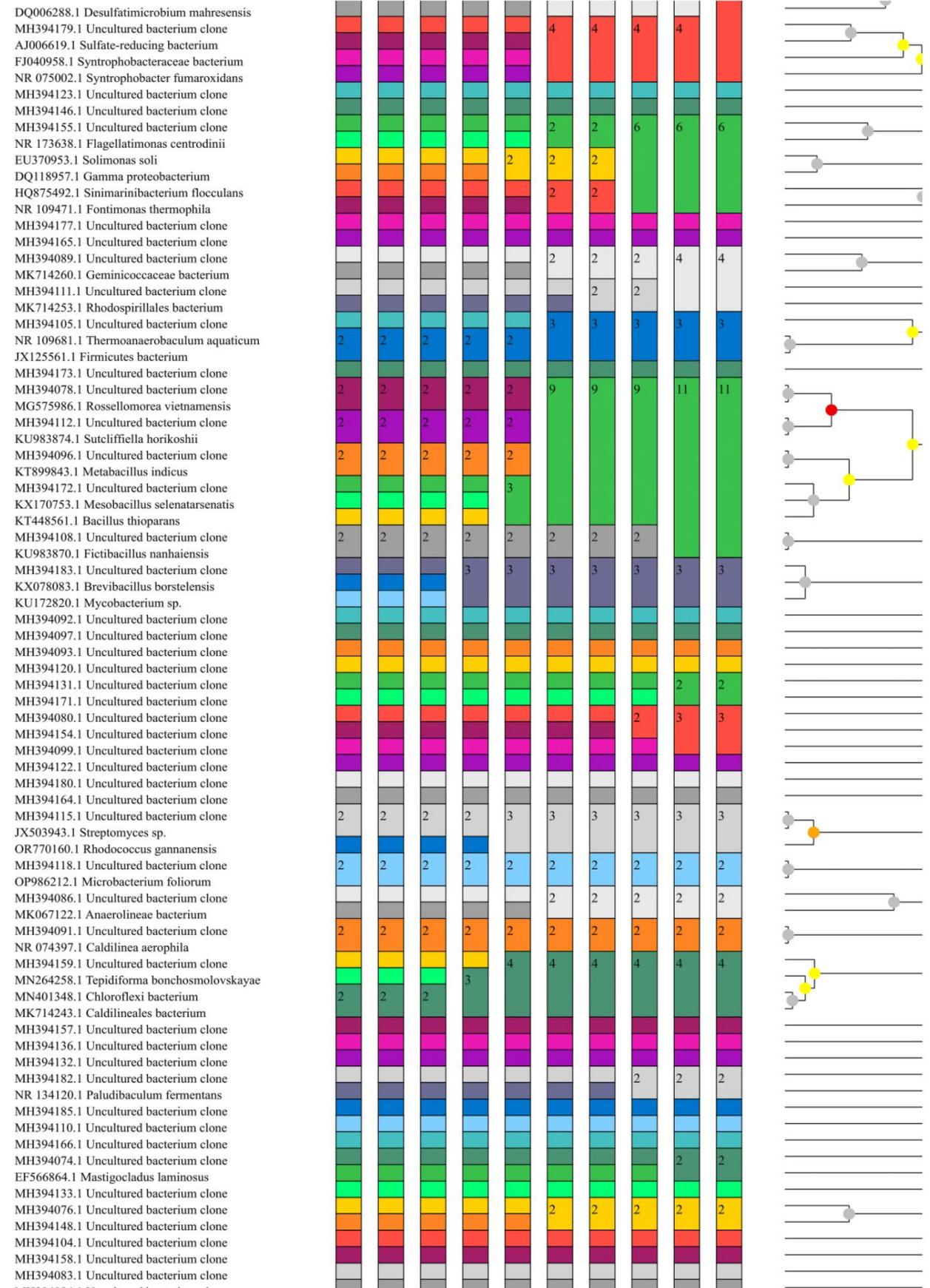


https://bioinfo.mnhn.fr/abi/public/asap/temp/31539.1804289383/groups.svg

1/3

5/23/24, 10:22 PM

bioinfo.mnhn.fr/abi/public/asap/temp/31539.1804289383/groups.svg



https://bioinfo.mnhn.fr/abi/public/asap/temp/31539.1804289383/groups.svg

2/3

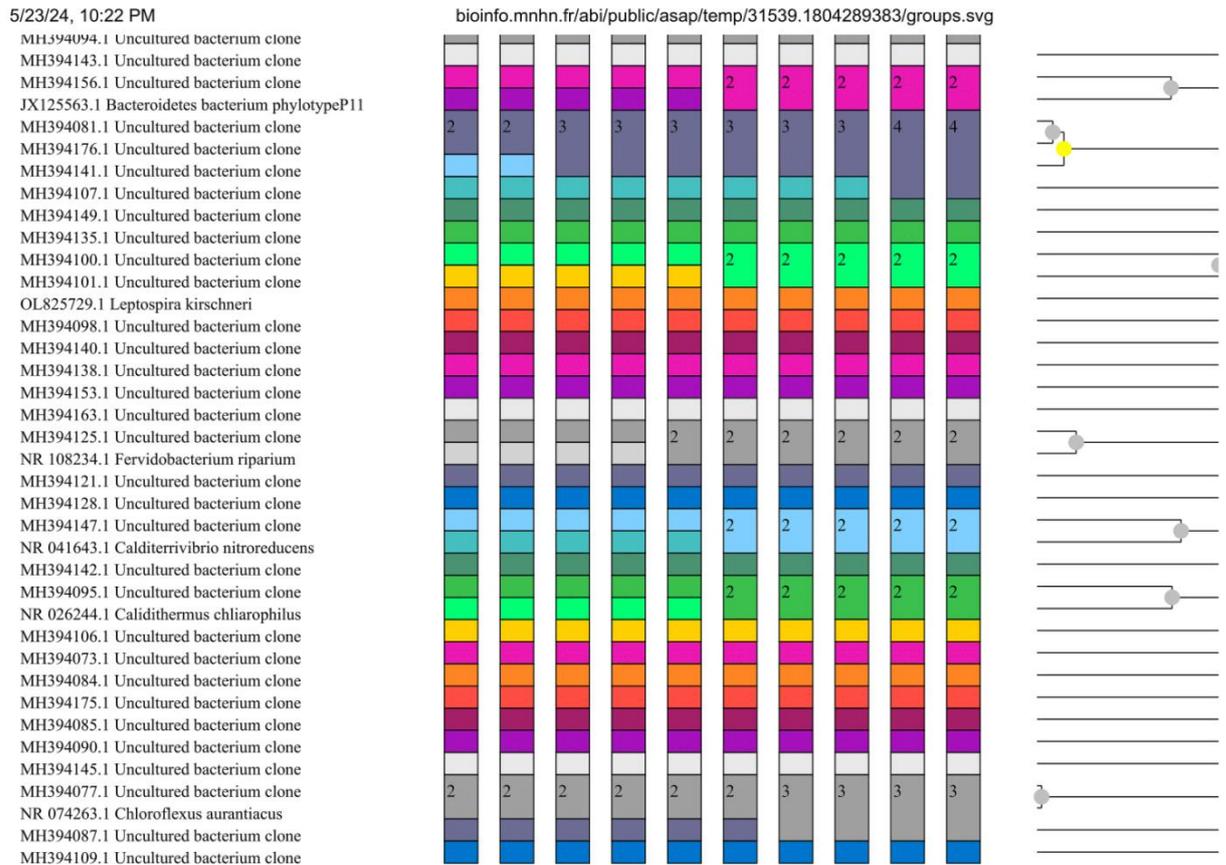


Figure 11. Affichage du graphique des sous-ensembles encadrés dans les résultats de l'outil web ASAP.

• Partition 1

Dans la première partition, qui a obtenu le score le plus bas de 7,5, 177 subsets (sous-ensembles) ont été trouvés. La plupart de ces sous-ensembles représentent des espèces individuelles, chaque espèce constituant un sous-ensemble distinct. Cependant, certains sous-ensembles regroupent deux espèces ensemble, Comme Adjeroud en 2021 a défini les sous-ensembles comme des OTUs (unité taxonomique opérationnelle) au niveau des espèces, nous pouvons également considérer que sous-ensemble représente une espèce.

- Le premier groupe : Une bactérie non cultivable, identifiée par le code d'accession MH394069, est regroupée avec l'espèce *Tepidimonas charontis*. Cette dernière appartient à l'ordre des Burkholderiales et à la classe des Betaproteobacteria. Cela signifie que cette bactérie non cultivable est taxonomiquement proche de *Tepidimonas charontis*.

- Le deuxième groupe a révélé une autre espèce non cultivable, identifiée par le code d'accèsion MH394070, est regroupée avec l'espèce *Acinetobacter oryzae*, qui appartient à la famille Moraxellaceae. Par conséquent, cette espèce non cultivable est également attribuée à cette famille.

- Le troisième groupe montre une espèce non cultivable, identifiée par le code MH394072, est classée au sein de la famille des Sphingomonadaceae.

- Le quatrième groupe comprend une espèce non cultivable, identifiée par le code MH394078, est associée à l'espèce *Rosellomorea vietnamensis*, qui appartient à la famille Bacillaceae.

- Le cinquième groupe comprend une bactérie non cultivable, identifiée par le code MH394112, est regroupée avec l'espèce *Sutcliffiella horikoshii*, également membre de la famille Bacillaceae.

- Le sixième groupe est représenté par l'espèce *Metabacillus indicus* qui appartient à la famille Bacillaceae avec un espèce non cultivable codé par MH394096.

- Le septième groupe contient l'espèce *Fictobacillus nanhaiensis* appartient à la famille Bacillaceae, avec l'espèce non cultivable son code MH394108.

- Le huitième groupe regroupée l'espèce non cultivable son code est MH394091, avec l'espèce *Caldilinea aerophila*, appartenant à la famille des Caldilineaceae, indiquant que cette espèce non cultivable appartient également à cette famille.

- Le neuvième groupe représente une bactérie non cultivable, identifiée par le code d'accèsion regroupée avec le genre *Streptomyces*, ce qui signifie que cette espèce non cultivée est attribuée à cette famille.

- Le dixième groupe comprend une bactérie non cultivable, identifiée par le code MH394118, est regroupée avec une espèce *Microbacterium foliorum*, qui appartient à la famille des Microbacteriaceae. Cette bactérie non cultivable appartient donc à cette famille.

- Le onzième groupe est représenté par deux espèces non cultivables, identifiées par les codes d'accès MH394176 et MH394081, sont classées ensemble dans un même groupe.

- Le douzième groupe représente l'espèce *Chloroflexus aurantiacus* qui appartient à la famille Chloroflexaceae avec espèce non cultivable son code est MH394077.

En résumé, la première partition révèle une structuration précise des espèces, avec toutes les séquences des espèces non cultivables obtenues par Adjeroud réparties dans des OTUs (Unités Taxonomiques Opérationnelles) distinctes comme dans son étude (Adjeroud, 2020) et correctement associées à des familles ou genres connus, ce qui démontre l'efficacité de la méthode ASAP pour la délimitation des espèces.

- **Partition 2** : contient un score égal 10, a fourni 174 subsets, dans cette partition, il y a une seule différence par rapport à la première qui représente par un groupe contient 04 espèces, une espèce non cultivée son code d'accèsion MH394167 regroupée dans un subset numéroté 99 avec 3 espèces connus : *Agrobacterium deltaense*, *Peteryoungia ipomoeae*, *Rhizobium daejeonense*, qui appartient à la même famille : Rhizobiaceae, cela signifie que cette espèce incultivable appartient à cette famille.

- **Partition 03** : le score de cette partition 13 et le nombre de subsets 95, La suite des résultats de cette partition est montrée dans le tableau 02.

Tableau 2. Résultats de la taxonomie des espèces à partir des méthodes de délimitation « ASAP » de la partition 03.

| OTU | Les séquences | La famille | L'ordre |
|-------------------|---|-----------------|-----------------|
| Subset[1] n:18 | MH394069.1 Uncultured bacterium MH590703.1 Tepidimonas charontis AF125877.1 Dehydroabietic acid-degrading bacterium MH394102.1 Uncultured bacterium MH394127.1 Uncultured bacterium MH394129.1 Uncultured bacterium MH394168.1 Uncultured bacterium NR 165749.1 Tepidicella baoligensis MH394130.1 Uncultured bacterium KY039329.1 Burkholderiales bacterium NR 173595.1 Aquariibacter albus KP717085.1 Aquabacterium citratiphilum KC492984.1 Rivibacter sp. | Famille inconnu | Burkholderiales |
| | MK757933.1 Malikia sp. KU360711.1 Macromonas nakdongensis NR 158120.1 Hydrogenophaga aquatica | Comamonadaceae | |
| | MH046068.1 Azohydromonas sediminis MT893353.1 Piscinibacter lacus | Sphaerotilaceae | |

| | | | |
|---------------------------|--|--------------------|---------------------|
| Subset[2] n:14 | MH394070.1 Uncultured bacterium MN727125.1 Acinetobacter oryzae MH394161.1 Uncultured bacterium MH394124.1 Uncultured bacterium MH394114.1 Uncultured bacterium MH394151.1 Uncultured bacterium MH394162.1 Uncultured bacterium MH394088.1 Uncultured bacterium MH394134.1 Uncultured bacterium MH394144.1 Uncultured bacterium MH394169.1 Uncultured bacterium MH394186.1 Uncultured bacterium MH394117.1 Uncultured bacterium MH394113.1 Uncultured bacterium | Moraxellaceae | Moraxellales |
| Subset[3] n: 7 | MH394071.1 Uncultured bacterium MH394119.1 Uncultured bacterium MH394139.1 Uncultured bacterium AB834446.1 Rhizobiales bacterium AF007948.1 Rasbo bacterium | Famille inconnue | Hyphomicrobiales |
| | MN561030.1 Rhabdaerophilum calidifontis | Rhabdaerophilaceae | |
| | MH032761.1 Phreatobacter cathodiphilus | Phreatobacteraceae | |
| Subset[4] n:14 | MH394072.1 Uncultured bacterium LC589090.1 Alphaproteobacteria bacterium KC248058.1 Erythromicrobium sp. KX161397.1 Erythrobacter ramosus MH394181.1 Uncultured bacterium MH394116.1 Uncultured bacterium MH394160.1 Uncultured bacterium | Famille inconnu | Alphaproteobacteria |
| | AB220117.1 Sphingomonadaceae bacterium | Sphingomonadaceae | |
| | OM190407.2 Erythrobacter oryzae KX022854.1 Porphyrobacter sp. MN746126.1 Erythrobacter tepidarius KC248039.1 Porphyrobacter sp. NR 179814.1 Tsuneonella suprasediminis KT597071.1 Altererythrobacter sp. | Erythrobacteraceae | |
| Subset[5] n: 1 | MH394073.1 Uncultured bacterium | Famille inconnue | |
| Subset[6] n: 1 | MH394074.1 Uncultured bacterium | Famille inconnue | |
| Subset[7] n: 3 | MH394075.1 Uncultured bacterium MH394174.1 Uncultured bacterium LN896453.1 Rhodobacter sp. | Paracoccaceae | Rhodobacterales |

| | | | |
|----------------------------|---|-------------------|-----------------|
| Subset[8] n: 2 | MH394076.1 Uncultured bacterium MH394148.1 Uncultured bacterium | Famille inconnue | |
| Subset[9] n: 2 | MH394077.1 Uncultured bacterium NR 074263.1 Chloroflexus aurantiacus | Chloroflexaceae | Chloroflexales |
| Subset[10] n: 1 | MH394087.1 Uncultured bacterium | Famille inconnue | |
| Subset[11] n: 9 | MH394078.1 Uncultured bacterium MG575986.1 Rossellomorea vietnamensis MH394112.1 Uncultured bacterium KU983874.1 Sutcliffiella horikoshii MH394096.1 Uncultured bacterium KT899843.1 Metabacillus indicus MH394172.1 Uncultured bacterium KX170753.1 Mesobacillus selenatarsenatis KT448561.1 Bacillus thioparans | Bacillaceae | Bacillales |
| Subset[12] n: 2 | MH394079.1 Uncultured bacterium NR 113020.1 Methylothermus subterraneus | Methylothermaceae | Methylococcales |
| Subset[13] n: 1 | MH394080.1 Uncultured bacterium | Famille inconnue | |
| Subset[14] n: 3 | MH394081.1 Uncultured bacterium MH394176.1 Uncultured bacterium MH394141.1 Uncultured bacterium | Famille inconnue | |
| Subset[15] n: 2 | MH394082.1 Uncultured bacterium MH394152.1 Uncultured bacterium | Famille inconnue | |
| Subset[16] n: 1 | MH394083.1 Uncultured bacterium | Famille inconnue | |
| Subset[17] n: 1 | MH394084.1 Uncultured bacterium | Famille inconnue | |
| Subset[18] n: 1 | MH394085.1 Uncultured bacterium | Famille inconnue | |
| Subset[19] n: 2 | MH394086.1 Uncultured bacterium MK067122.1 Anaerolineae bacterium | Famille inconnue | Anaerolineae |
| Subset[20] n: 2 | MH394089.1 Uncultured bacterium MK714260.1 Geminicoccaceae bacterium | Geminicoccaceae | |
| Subset[21] n: 1 | MH394090.1 Uncultured bacterium | Famille inconnue | |
| Subset[22] n: 2 | MH394091.1 Uncultured bacterium NR 074397.1 Caldilinea aerophila | Caldilineaceae | |
| Subset[23] n: 1 | MH394092.1 Uncultured bacterium | Famille inconnue | |
| Subset[24] n: 1 | MH394093.1 Uncultured bacterium | Famille inconnue | |
| Subset[25] n: 1 | MH394094.1 Uncultured bacterium | Famille inconnue | |

| | | | |
|----------------------------|--|------------------------|-----------------------|
| Subset[26] n: 2 | MH394095.1 Uncultured bacterium NR 026244.1 Calidithermus chliarophilus | Thermaceae | Thermales |
| Subset[27] n: 1 | MH394097.1 Uncultured bacterium | Famille inconnue | |
| Subset[28] n: 1 | MH394098.1 Uncultured bacterium | Famille inconnue | |
| Subset[29] n: 1 | MH394099.1 Uncultured bacterium | Famille inconnue | |
| Subset[30] n: 2 | MH394100.1 Uncultured bacterium MH394101.1 Uncultured bacterium | Famille inconnue | |
| Subset[31] n: 3 | MH394103.1 Uncultured bacterium MH394170.1 Uncultured bacterium KC921197.1 Delta proteobacterium | Famille inconnue | |
| Subset[32] n: 1 | MH394104.1 Uncultured bacterium | Famille inconnue | |
| Subset[33] n: 3 | MH394105.1 Uncultured bacterium JX125561.1 Firmicutes bacterium | Famille inconnue | |
| | NR 109681.1 Thermoanaerobaculum aquaticum | Thermoanaerobaculaceae | Thermoanaerobaculales |
| Subset[34] n: 1 | MH394106.1 Uncultured bacterium | Famille inconnue | |
| Subset[35] n: 1 | MH394107.1 Uncultured bacterium | Famille inconnue | |
| Subset[36] n: 2 | MH394108.1 Uncultured bacterium KU983870.1 Fictibacillus nanhaiensis | Bacillaceae | Bacillales |
| Subset[37] n: 1 | MH394109.1 Uncultured bacterium | Famille inconnue | |
| Subset[38] n: 1 | MH394110.1 Uncultured bacterium | Famille inconnue | |
| Subset[39] n: 1 | MH394111.1 Uncultured bacterium | Famille inconnue | |
| Subset[40] n: 3 | MH394115.1 Uncultured bacterium | Famille inconnue | |
| | JX503943.1 Streptomyces sp. | Streptomycetaceae | Kitasatosporales |
| | OR770160.1 Rhodococcus gannanensis | Nocardiaceae | Mycobacteriales |
| Subset[41] n: 2 | MH394118.1 Uncultured bacterium OP986212.1 Microbacterium foliorum | Microbacteriaceae | Micrococcales |
| Subset[42] n: 1 | MH394120.1 Uncultured bacterium | Famille inconnue | |
| Subset[43] n: 1 | MH394121.1 Uncultured bacterium | Famille inconnue | |
| Subset[44] n: 1 | MH394122.1 Uncultured bacterium | Famille inconnue | |

| | | | |
|----------------------------|--|------------------------|---------------------|
| Subset[45] n: 1 | MH394123.1 Uncultured bacterium | Famille inconnue | |
| Subset[46] n: 2 | MH394125.1 Uncultured bacterium NR 108234.1 Fervidobacterium riparium | Fervidobacteriaceae | Thermotogales |
| Subset[47] n: 2 | MH394126.1 Uncultured bacterium DQ006288.1 Desulfatimicrobium mahresensis | Syntrophobacteraceae | Syntrophobacterales |
| Subset[48] n: 1 | MH394128.1 Uncultured bacterium | Famille inconnue | |
| Subset[49] n: 1 | MH394131.1 Uncultured bacterium | Famille inconnue | |
| Subset[50] n: 1 | MH394132.1 Uncultured bacterium | Famille inconnue | |
| Subset[51] n: 1 | MH394133.1 Uncultured bacterium | Famille inconnue | |
| Subset[52] n: 1 | MH394135.1 Uncultured bacterium | Famille inconnue | |
| Subset[53] n: 1 | MH394136.1 Uncultured bacterium | Famille inconnue | |
| Subset[54] n: 3 | MH394137.1 Uncultured bacterium AB042502.2 Halophilic bacterium | Famille inconnue | |
| | KT581470.1 Idiomarina fontislapidosi | Idiomarinaceae | Alteromonadales |
| Subset[55] n: 1 | MH394138.1 Uncultured bacterium | Famille inconnue | |
| Subset[56] n: 1 | MH394140.1 Uncultured bacterium | Famille inconnue | |
| Subset[57] n: 1 | MH394142.1 Uncultured bacterium | Famille inconnue | |
| Subset[58] n: 1 | MH394143.1 Uncultured bacterium | Famille inconnue | |
| Subset[59] n: 1 | MH394145.1 Uncultured bacterium | Famille inconnue | |
| Subset[60] n: 1 | MH394146.1 Uncultured bacterium | Famille inconnue | |
| Subset[61] n: 2 | MH394147.1 Uncultured bacterium NR 041643.1 Calditerrivibrio nitroreducens | Calditerrivibrionaceae | Deferribacterales |
| Subset[62] n: 1 | MH394149.1 Uncultured bacterium | Famille inconnue | |
| Subset[63] n: 1 | MH394150.1 Uncultured bacterium | Famille inconnue | |
| Subset[64] n: 1 | MH394153.1 Uncultured bacterium | Famille inconnue | |
| Subset[65] n: 1 | MH394154.1 Uncultured bacterium | Famille inconnue | |
| | MH394155.1 Uncultured bacterium | Nevskiaceae | Nevskiales |

| | | | |
|---------------------------|--|----------------------|---------------------|
| Subset[66] n: 2 | NR 173638.1 Flagellatimonas centrodinii | | |
| Subset[67] n: 2 | MH394156.1 Uncultured bacterium JX125563.1 Bacteroidetes bacterium phyloTYPEP11 | Famille inconnue | |
| Subset[68] n: 1 | MH394157.1 Uncultured bacterium | Famille inconnue | |
| Subset[69] n: 1 | MH394158.1 Uncultured bacterium | Famille inconnue | |
| Subset[70] n: 4 | MH394159.1 Uncultured bacterium MN401348.1 Chloroflexi bacterium MK714243.1 Caldilineales bacterium | Famille inconnue | |
| | MN264258.1 Tepidiforma bonchosmolovskayae | Tepidiformaceae | Tepidiformales |
| Subset[71] n: 1 | MH394163.1 Uncultured bacterium | Famille inconnue | |
| Subset[72] n: 1 | MH394164.1 Uncultured bacterium | Famille inconnue | |
| Subset[73] n: 1 | MH394165.1 Uncultured bacterium | Famille inconnue | |
| Subset[74] n: 1 | MH394166.1 Uncultured bacterium | Famille inconnue | |
| Subset[75] n: 4 | MH394167.1 Uncultured bacterium | | |
| | PP411530.1 Agrobacterium deltaense | | |
| | OR358916.1 Peteryoungia ipomoeae | Rhizobiaceae | Hyphomicrobiales |
| | OR358915.1 Rhizobium daejeonense | | |
| Subset[76] n: 1 | MH394171.1 Uncultured bacterium | Famille inconnue | |
| Subset[77] n: 1 | MH394173.1 Uncultured bacterium | Famille inconnue | |
| Subset[78] n: 1 | MH394175.1 Uncultured bacterium | Famille inconnue | |
| Subset[79] n: 1 | MH394177.1 Uncultured bacterium | Famille inconnue | |
| Subset[80] n: 1 | MH394178.1 Uncultured bacterium | Famille inconnue | |
| Subset[81] n: 4 | MH394179.1 Uncultured bacterium AJ006619.1 Sulfate-reducing bacterium | Famille inconnue | |
| | FJ040958.1 Syntrophobacteraceae bacterium | | |
| | NR 075002.1 Syntrophobacter fumaroxidans | Syntrophobacteraceae | Syntrophobacterales |
| Subset[82] n: 1 | MH394180.1 Uncultured bacterium | Famille inconnue | |

| | | | |
|---------------------------|---|-------------------|------------------------|
| Subset[83] n: 1 | MH394182.1 Uncultured bacterium | Famille inconnue | |
| Subset[84] n: 3 | MH394183.1 Uncultured bacterium | Famille inconnue | |
| | KX078083.1 <i>Brevibacillus borstelensis</i> | Paenibacillaceae, | Bacillales |
| | KU172820.1 <i>Mycobacterium</i> sp. | Mycobacteriaceae | Mycobacteriales |
| Subset[85] n: 1 | MH394184.1 Uncultured bacterium | Famille inconnue | |
| Subset[86] n: 1 | MH394185.1 Uncultured bacterium | Famille inconnue | |
| Subset[87] n: 1 | EF566864.1 <i>Mastigocladus laminosus</i> | Hapalosiphonaceae | Nostocales |
| Subset[88] n: 1 | OP776070.1 <i>Methylomarinovum caldicuralii</i> | Methylothermaceae | <u>Methylococcales</u> |
| Subset[89] n: 1 | OL825729.1 <i>Leptospira kirschneri</i> | Leptospiraceae | Leptospirales |
| Subset[90] n: 1 | NR 156937.1 <i>Sulfuritorus calidifontis</i> | Thiobacillaceae | Nitrosomonadales |
| Subset[91] n: 1 | MK714253.1 <i>Rhodospirillales bacterium</i> | Famille inconnue | Rhodospirillales |
| Subset[92] n: 1 | FJ802263.1 Denitrifying bacterium | Famille inconnue | |
| Subset[93] n: 2 | EU370953.1 <i>Solimonas soli</i> DQ118957.1 <i>Gamma proteobacterium</i> | Famille inconnue | |
| Subset[94] n: 2 | HQ875492.1 <i>Sinimarini bacterium flocculans</i> NR 109471.1 <i>Fontimonas thermophila</i> | Nevskiaceae | Nevskiales |
| Subset[95] n: 1 | NR 134120.1 <i>Paludibaculum fermentans</i> | Bryobacteraceae | Bryobacteriales |

Ce tableau montre les résultats de la taxonomie des espèces de partition 3 :

- Subset 1 comprend 18 séquences appartenant toutes à l'ordre des Burkholderiales. Certaines de ces séquences sont réparties en deux familles distinctes : trois appartiennent à la famille Comamonadaceae et deux à la famille Sphaerotilaceae, tandis que les autres ne sont affiliées à aucune famille.
- Subset 2 regroupe 14 séquences collectées dans la même famille, Moraxellaceae.
- Subset 3 ce groupe contient 7 séquences appartenant au même ordre, les Hyphomicrobiales. Il y a des séquences qui n'appartiennent à aucune famille, et deux

autres séquences sont divisées en deux familles, Rhabdaerophilaceae et Phreatobacteraceae.

- Subset 4 : regroupe les 14 séquences au même ordre Alphaproteobacteria. Il y a des séquences divisées en deux familles différentes : Erythrobacteraceae et Sphingomonadaceae. Alors que les autres séquences n'appartiennent à aucune famille.
- Subset 7 : comporte 3 séquences retient à la même famille, Paracoccaceae.
- Subset 9 : contient deux séquences appartenant à la même famille Chloroflexaceae
- Subset 11 : composé 9 séquences de la même famille Bacillaceae.
- Subset 12 : Comprend deux séquences de la même famille Methylothermaceae
- Subset 19 : contient deux séquences appartenant au même ordre, les Anaerolineae, et à une famille inconnue.
- Subset 20 : composé de deux séquences comporte à la même famille Geminicoccaceae
- Subset 22 : contient deux séquences appartenant à la même famille Caldilineaceae
- Subset 26 : Comprend deux séquences de même famille Thermaceae
- Subset 31 : contient trois séquences regroupées en même classe Deltaproteobacteria, et à la famille inconnue.
- Subset 33 : comporte trois séquences, L'un d'entre eux se trouve dans une famille Thermoanaerobaculaceae, les deux restes appartient à la phylum Bacillota
- Subset 36 : contient deux espèces retient à la même famille Bacillaceae
- Subset 40 : a recueilli 3 séquences, mais l'une d'entre elles appartenait à la famille Streptomycetaceae, une autre à la famille Nocardiaceae et la troisième n'a pas été reconnue.
- Subset 41 : comprend deux séquences au même famille Microbacteriaceae.
- Subset 46 : regroupé deux séquences à la même famille Fervidobacteriaceae.
- Subset 47 : comporte deux séquences au même famille Syntrophobacteraceae.
- Subset 54 : contient trois séquences, dont l'une appartient à la famille Idiomarinaceae et les deux autres à une famille inconnue.
- Subset 61 : comporte deux séquences au même famille Calditerrivibrionaceae.

- Subset 66 rassemble deux séquences appartenant à la même famille, Nevskiaceae.
- Subset 70 : se compose de quatre séquences, dont l'une appartient à la famille Tepidiformaceae, tandis que les trois autres ne sont connues pour appartenir à aucune famille.
- Subset 75 : ce groupe contient quatre séquences collectées dans la même famille.
- Subset 81 : Il contient quatre séquences, dont deux appartiennent à la famille Syntrophobacteraceae et deux dont on ne sait pas à quelle famille elles appartiennent.
- Subset 84 : comporte trois séquences, mais l'une d'entre elles appartenait à la famille Paenibacillaceae, une autre à la famille Mycobacteriaceae, et autre séquence n'a pas été reconnue.
- Subset 87 comporte une séquence appartenant à la famille Hapalosiphonaceae.
- Subset 88 contient une séquence appartenant à la famille Methylothermaceae
- Subset 89 contient une séquence appartenant à la famille Leptospiraceae
- Subset 90 : contient une séquence appartenant à la famille Thiobacillaceae
- Subset 93 comporte deux séquences appartiennent à la classe Gammaproteobacteria.
- Subset 94 compose de deux séquences appartiennent à la famille Nevskiaceae
- Subset 95 contient une séquence appartenant à la famille Bryobacteraceae

Dans tous ces résultats, 31 séquences non cultivées ont été identifiées et classées dans la famille à laquelle elles appartiennent. Cependant, certaines séquences n'ont pas pu être identifiées, ce qui nous amène à penser qu'elles pourraient appartenir à de nouvelles familles.

Conclusion

Conclusion

Notre étude a mis en évidence la capacité des technologies de métagénomique et de séquençage de nouvelle génération (NGS) à analyser et à classer les populations microbiennes. En associant la génomique, la bioinformatique et l'étude de la biodiversité des écosystèmes, nous avons réussi à étudier simultanément les génomes de plusieurs organismes, illustrant ainsi les progrès significatifs effectués dans l'identification et la classification des espèces bactériennes.

En utilisant la méthode de délimitation ASAP (Assemble Species by Automatic Partitioning) pour notre arbre phylogénétique construit à partir des séquences d'ARNr 16S, nous avons pu attribuer des séquences à des familles, des genres ou des espèces précises, mais aussi améliorer notre compréhension des relations phylogénétiques et de la diversité des espèces rencontrées.

En conclut, ce travail apporte une contribution significative à la classification des séquences microbiennes, ce qui ouvre des perspectives pour de futures recherches dans le domaine de la métagénomique.

Références Bibliographiques

Références Bibliographiques

- **Adjeroud M.** 2021. Investigation métagénomique de la biodiversité des bactéries et Archaea productrices de substances bioactives à partir de quelques sources hydrothermales en Algérie. Thèse de doctorat, Université A. MIRA, Algérie, 182.
- **Calenge F., Leloutre L., Quéré P., Velge P., Gabriel I., Dore J.** 2017. Le microbiote intestinal à l'ère de la métagénomique : Perspectives d'application dans la filière avicole. 12. Journées de la Recherche Avicole et Palmipèdes à Foie Gras (No. 12, pp. 1222-p).
- **Caspermeyer J.** 2018. MEGA Software Celebrates Silver Anniversary, *Molecular Biology and Evolution*, 35(6) :1558–1560
- **Chiriac L.S., Murariu D.T.** 2021. Application of metagenomics in ecology: A -Brief Overview. *Current Trends in Natural Sciences* 10(19) :346-353.
- **Clark K., Karsch-Mizrachi I., Lipman D. J., Ostell J., Sayers E. W.** 2016. GenBank. *Nucleic acids research*, 44(D1): D67-D72.
- **Criscuolo, A., Gribaldo, S.** 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10(210): 1471-2148
- **Dennis A. Benson, M. C.M.** 2017. GenBank. *Nucleic Acids Research*, p 41-47
- **Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J. M., Gascuel O.** 2008. *Phylogeny.fr* : analyse phylogénétique robuste pour le non-spécialiste, *Nucleic Acids Research*, 36(2) : W465–W469.
- **Didier G., Guindon S.** 2022. Arbres : combinatoire et modèles. Modèles et méthodes pour l'évolution biologique, pp.7-32 *24(3) : 218-222.*
- **Diene S.M., Bertelli C., Pillonel T., Schrenzel J., Greub G.** 2014. Génomique et métagénomique bactériennes : applications cliniques et importance médicale, *Rev Med Suisse* 10 : 2155-61.
- **Dudhagara P., Bhavsar S., Bhagat C., Ghelani A., Bhatt S., Patel R.** 2015. Web resources for metagenomics studies. *Genomics, Proteomics and Bioinformatics*, 13(5): 296-303.

- Fabiola V.G ., Ramón C. H., Enrique V ., Francisco V.A ., Christine H.** 2015. Le gène de l'ARNr 16S dans l'étude des communautés microbiennes marines 41(4) : 297-313.
- **Fujisawa T, Barraclough T.G.** 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. 62(5): 707-24
- Grzegorz M. B., Camacho C., Cooper P.S., Coulouris G., Fong A., Ning M., Thomas L. M., Wayne T. M., Scott D. M., Yuri M., Yan R., Eric W. S., Tao T., Jian Y., Irena Z.** 2013. BLAST: a more efficient report with usability improvements, *Nucleic Acids Research*, 41, (1): W29–W33.
- Guyomar C., Lemaitre C.** 2023. Métagénomique et métatranscriptomique. Des séquences aux graphes Méthodes et structures discrètes pour la bioinformatique, 252.
- Handelsman J.** 2004. Metagenomics: application of genomics to uncultured microorganisms, *Microbiol. Mol. Biol. Rev* 68 : 669-685.
- Kapli P., Lutteropp S., Zhang J., Kobert K., Pavlidis P., Stamatakis A., Flouri T.**2017. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo, *Bioinformatics*, 33(11) :1630–1638.
- Kumar S., Stecher G.,Li M.,KnyaZ C.,Tamura K.** 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35(6):1547-1549.
- **Lamoril J., Ameziane N., Deybacha J.C., Bouizegarènea P., Bogard M.** 2008. Les techniques de séquençage de l'ADN : une révolution en marche. Première partie DNA sequencing technologies : A revolution in motion. Part one 23 (5): 260-279.
- Lemoine F., Correia D., Lefort V., Doppelt-Azeroual O., Mareuil F., Cohen-Boulakia S., Gascuel, O.** 2019. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research*, 47(1): 260-265.
- **Letunić I., Bork P.** 2021. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Research*. 49(1): 293-296.
- Mao C. X.** 2007. Estimating population sizes by catch-effort methods. *Statistical Methodology*, 4(1) : 111-119.

- **Marchand S., Rodriguez C., Woerther P.L.** 2024. Séquençage à haut débit pour le diagnostic en maladies infectieuses : exemple de la métagénomique shotgun dans les infections du système nerveux central 45 (3) :166-173.
- **Meyer K.** 1991. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genet Sel Evol* 23(1): 67-83
- **Mysara M., Vandamme P., Props R., Kerckhof F.M., Leys N., Boon N., Raes J., Monsieurs P.** 2017. Reconciliation between operational taxonomic units and species boundaries, *FEMS Microbiology Ecology* 93(4): fix029.
- **Newell P. D., Fricker A. D., Roco C. A., Chandransu 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.
- **Patrick D.S., Handelsman J.** 2003. Perspectives biotechnologiques issues de la métagénomique 14(3) : 303-310.
- **Pauline B., Emmanuel A., Marc E., Philippe R.** 2013. Métagénomique virale et pathologie, Une histoire récente, *Med Sci(Paris)* 29 : 501–508.
- **Pevzner P. A.** 2006, Alignement multiple. *Bio-informatique moléculaire*, 125-135.
- **Pons, J. Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamoun S., Sulmin W.D., Volger A.P.** 2006. Délimitation des espèces basée sur la séquence pour la taxonomie de l'ADN des insectes non décrits. *Biologie systématique* 55(4) : 595-609.
- **Puillandre N., Lambert A., Brouillet S., Achaz G.** 2012. ABGD, Automatic Barcode Gap Discovery pour la délimitation primaire des espèces. *Écologie moléculaire* 21 (8) : 1864-1877
- **Puillandre N., Brouillet S., Achaz G.** 2021. ASAP : assemblespecies by automatic partitioning. *Molecular Ecology Resources* 21(2) : 609-620.
- **Sauvage V., Gomez J., Boizeau L., Laperche S.** 2017. Le potentiel de la métagénomique virale dans la sécurité transfusionnelle La métagénomique virale : un nouvel outil de surveillance des agents viraux émergents au service de la sécurité transfusionnelle
- **Siegwald L.** 2017. Solutions d'amélioration des études de métagénomique ciblée. *Bio-Informatique biologie systémique*. Thèse de doctorat, Université de Lille, France, 283 p.

- **Simon C., Daniel R.** 2011. Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol* 77 :1153-1161.
- **Taguan,** 2015, Les arbres phylogénétiques ; Catégories : Programmation et algorithmique et Biologie
- **Theron J., Cloete T. E.** 2000. Molecular techniques for determining microbial diversity and community structure in natural environments. *Critical Reviews in Microbiology* 26(1) :37-57.
- **Van H.T.P., Jaisoo K.** 2012. Cultivation of unculturable soil bacteria *Trends in Biotechnology* 30(9):475-84.
- **Vert J. P.** 2013. Les applications industrielles de la bio-informatique. Réalités industrielles. *Annales des mines*, pp 17-23.
- **Virginie M.** 2006. La protection de la biodiversité : entre science, éthique et politique. Thèse de doctorat, Université de Montréal, Canada, 336 p.
- **Ye J., McGinnis S., Madden T. L.** 2006. BLAST: improvements for better sequence analysis. *Nucleic acids research* 34(2): W6-W9.
- **Zhang L, Huang L, Chen L, Hao D, Chen J.** 2013. Neuroprotection by tetrahydroxystilbene glucoside in the MPTP mouse model of Parkinson's disease. *Toxicol Lett.* 222(2) :155-63.

Livres :

- **Harold L.** 2007. Quels indicateurs pour la gestion de la biodiversité, Institut français de la biodiversité 57, France, p 99.
- **Faure D., Jouly D.** 2016. La génomique environnementale : La révolution du séquençage à haut débit. ISTE Group.
- **Ramade F.** 2008. Dictionnaire encyclopédique des sciences de la nature et de la biodiversité. Paris, Dunod, p737.
- **Rodriguez M., Njike H.B.M., Mbarga D.A.O., Phelps J.** 2023. Réparation juridique des préjudices causés à la biodiversité : Une analyse de la législation camerounaise sur la responsabilité environnementale. Conservation-Ligitation. org, p43.
- **Rannala B., Yang Z.** 2020. Species delimitation. Self-published.

Les liens :

- Site web 1: <https://www.ncbi.nlm.nih.gov/>
- Site web 2: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
- Site web 3: <https://ngphylogeny.fr>
- Site web 4: <https://www.megasoftware.net/>
- Site web 5: <https://itol.embl.de/>
- Site web 6: <https://bioinfo.mnhn.fr/abi/public/asap>
- Site web 7: <https://dridk.me/metagenomique.html>
- Site web 8: dridk.me/metagenomique.html/

تلخيص:

لقد أحدثت الميتاجينوميكات ثورة في علم البيئة الميكروبية من خلال تمكين الدراسة المباشرة للمادة الوراثية في العينات البيئية. وبفضل تقنيات تسلسل الجيل التالي، أصبحت تحليلات الميتاجينوميكات غير مكلفة وسريعة، مما يسهل دراسة المجتمعات الميكروبية. يهدف هذا المشروع إلى تحديد وتصنيف تسلسل الحمض النووي الريبوزومي 16S غير المعروف باستخدام قاعدة بيانات GenBank. تتضمن المنهجية البحث عن التسلسلات واختيارها وتحميلها ومحاذاتها ومعالجتها، يلي ذلك بناء وتحليل شجرة النشوء والتطور باستخدام طريقة ASAP. تُظهر النتائج تصنيفاً دقيقاً للتسلسلات غير المعروفة، مما يحسن فهمنا للعلاقات النشوء والتطور وتنوع الأنواع البكتيرية. وفي الختام، يتفوق هذا النهج بشكل كبير على الطرق التقليدية.

الكلمات المفتاحية: علم الميتاجينوميكات، علم البيئة الميكروبية، تقنيات تسلسل الجيل التالي، الحمض النووي الريبوزومي 16S، شجرة الجينات الوراثية.

Résumé :

Les analyses métagénomiques ont révolutionné l'écologie microbienne en permettant l'étude directe du matériel génétique présent dans les échantillons de l'environnement. Avec les technologies de séquençage de nouvelle génération (NGS), les analyses métagénomiques sont désormais économiques et rapides, ce qui rend l'étude des communautés microbiennes plus facile. Ce projet a pour but d'identifier et de classer des séquences d'ARNr 16S inconnues à l'aide de la base de données GenBank. La méthodologie comprend la recherche, la sélection, le téléchargement, l'alignement et le traitement des séquences, suivis de la construction et de l'analyse d'un arbre phylogénétique à travers la méthode ASAP. Les résultats obtenus révèlent une classification précise des séquences inconnues, ce qui permet d'améliorer notre compréhension des relations phylogénétiques et de la diversité des espèces. Enfin, cette approche est significativement efficace par rapport aux approches traditionnelles.

Mot clés : Métagénomique, l'écologie microbienne, technologies de séquençage de nouvelle génération (NGS), ARNr 16S, Arbre phylogénétique

Abstract:

Metagenomics has transformed microbial ecology by enabling the direct study of genetic material in environmental samples. Thanks to next-generation sequencing (NGS) technologies, metagenomic analyses have become inexpensive and rapid, facilitating the study of microbial communities. The aim of this project is to identify and classify unknown 16S rRNA sequences via the GenBank database. The methodology involves searching, selecting, downloading, aligning and processing sequences, followed by the construction and analysis of a phylogenetic tree using the ASAP method. The results show an accurate classification of unknown sequences, improving understanding of phylogenetic relationships and species diversity. In conclusion, this approach significantly outperforms traditional methods.

Keywords: Metagenomics, microbial ecology, NGS, 16S rRNA, Phylogenetic tree