الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

كلية علوم الطبيعة والحياة

Faculté des Sciences de la Nature et de la Vie

جامعة الإخوة منتوري قسنطينة 1

Université Frères Mentouri Constantine 1

Département de Biologie Appliquée

قسم البيولوجيا التطبيقية

Mémoire en vue de l'obtention du Diplôme de Master en :
Bioinformatique

THÈME

Analyse métagénomique du microbiote intestinal des patients atteints la maladie d'insuffisance rénale chronique : traitement des données de pyroséquençage d'ARNr 16s.

Présenté par :

MERBOUCHE Manel
BOULADJENIB Khouloud

Soutenu le : 21 - 06 - 2022

Devant le jury :

Président : Pr. HAMIDECHI Mohamed Abdelhafid Prof. Univ. Frères Mentouri Constantine 1

Encadrant : Dr. KELLOU Kamel MAA Univ. Frères Mentouri Constantine 1

Examinateur : Dr. CHEHILI Hamza MCA Univ. Frères Mentouri Constantine 1

Année universitaire 2021-2022

Remerciement

Nous tenons tout d'abord à remercier Dieu le tout Puissant et Miséricordieux, qui nous a donné la force et la patience d'accomplir ce travail.

Nous remercions Monsieur le Docteur **KELLO KAMEL** pour leur accueil et encadrement, le temps passé ensemble et le partage de ses connaissances. Grâce aussi à sa confiance nous pouvons accomplir totalement dans nos missions, et surtout ses judicieux conseils, qui ont contribué à alimenter nos réflexions. Il fut d'une aide très précieuse dans les moments les plus délicats.

Sans oublier d'exprimer nos profondes gratitudes à Monsieur le professeur **HAMIDECHI MOHAMED ABDELHAFID** pour la qualité de son enseignement, ses conseils et son intérêt incontestable qu'il porte à tous les étudiants.

Nous tenons à remercier Monsieur le Docteur **CHEHILI HAMZA** pour l'honneur qu'elle nous a fait en acceptant d'examiner notre travail.

Nous tenons à remercier tous mes enseignants actuels et passés du département de BIOLOGIE APPLIQUÉE et de notre spécialité BIO-INFORMATIQUE pour leurs qualités scientifiques et pédagogiques.

Nous désirons aussi remercions le cadre administratif de l'université qui fournit les outils nécessaires à la réussite de nos études universitaires. Merci tous ceux qui ont participé de loin ou de près à la réalisation de ce modeste travail.

Dédicace

بسم الله الرحمن الرحيم

والصلاة والسلام على أشرف وخير الانام نبينا محمد صلى الله عليه وسلم

وانه من الصعب التعبير تحت وطأت هذه اللحظة العظيمة في حياتي تنتحر الكلمات والعبارات امامها لأنها حقا تكبرني فانا الان عاجزه عن وصف الغبطة التي تغمرني لما لا وقد كللت دموعي والليالي البيضاء المتعبة على مدار الخمس سنوات المنصرمة بالنجاح وقد استوقفني مصداق الآية الكريمة "هل يوفى الصابرون اجرهم"، نعم بالفعل فها انا استشهد الان جوهر هذه الآية الكريمة بكل جوارحي..

الحمد لله انه تسنى لي عيش هذه اللحظة فلا يحمد الا سواه واود ان اعبر عن خالص حبي وتقديري وامتناني الخاص الى

والداي:

الى مهجه القلب وحشاشه الروح وقرة عيني ومصدر قوتي ومأمني واماني، جزيل الشكر لكما والداي الحبيبان فضلكما علي عظيم ولا يمكن ان يوفى وسأدعه للرحمن المنان لكي يقضى..

زميلتي:

الى الرفيقة من طوابير الجامعة الى اخت في مدرسه الحياة **بولجنيب خلود** شكرا لكونك خير السند والعون في أيام الشدائد.

والشكر الخاص الى المجهودات العظيمة المبذولة من طرف الزميل **رائد سرار** التي سهلت علينا مشقة البحث العلمي.

أخواي:

قال تعالى: " سنشد عضدك بأخيك" على ضوء الآية أود اشكر سندي، اليد التي تربت علي، الاذرع التي أبطش بها في محني، طرف من روحي أخواي **مربوش محمد عبد الرؤوف، مربوش عبد اللطيف** الحمد لله على وجودكما..

وفي اخر المطاف الي ريثما أخطو اخر خطواتي في هذه الجامعة وجدت نفسي مع كل خطوه اخطوها اتذكر الذكريات الدافئة والجميلة التي امضيتها في ربوعها، من عبق تلك الذكريات اسمع صدى الضحكات كالمتوهم في سراب، اللحظات الجميلة نعمه والسيئة اخذناهم كدروس

اريد ان اقول لجميع الاصدقاء الذين صنعتهم لي جامعه الأخوة منتوري، ليس في معجمي الكلمات الكافية لشكركم والتعبير عن كم انا ممتنه بوجودكم.. شكرا.

منال

Dédicace

En témoignage d'amour et d'affection, je dédie ce modeste travail avec tellement de fierté a tous ce qui me sont chers :

A ma chère Maman

Qui a œuvré pour ma réussite par son amour son soutient tout le long de mon parcours, c'est vrai qu'elle n'est plus là pour récolter les fruits de ces sacrifices mais elle reste toujours la plus présente.

A mon cher père

Tu as toujours été là pour moi, tu m'as soutenu, encouragé que ce travail traduit ma gratitude et mon affection que dieu te protégé et que la réussite soit toujours à ma portée pour que je puisse te combler de bonheur.

A Mes chers frères

Puisse dieu vous donne santé, bonheur courage et surtout beaucoup de réussite

A Mon Chère Binôme

MERBOUCHE Manel pour avoir eu la patience et le courage d'aboutir a ce travail en dépit de tous ce que nous avons subi durant cette merveilleuse année.

A tous mes chères Amies

Un grand Merci pour votre présence surtout **SERRAR Raid** qui mérite tout mon respect pour son aide précieuse et sa présence sans oublier tous les autres.
A toute ma famille que je n'ai pas pu les cités.

KHOULOUD

➢ Résumé

Résumé

Le microbiote intestinal, acteur clé de la santé, est considéré comme un dispositif à part entière de l'organisme humain. Le microbiote intestinal joue un rôle décisif dans notre santé. Il est extrêmement diversifié et varie d'un individu à l'autre. Dans l'objectifs d'étudier sa composition microbienne et de déterminer sa distribution, nous avons utilisé les données de pyroséquençage du gène ARN ribosomique 16s pour étudier les différences de microbiote intestinal entre quatre groupes de sujets atteints ou pas de la maladie d'insuffisance rénale (CKD).

Ce travail met le point sur l'utilisation du pipeline MOTHUR pour le traitement des données d'ARNr 16s. Cet outil nous a permis d'effectuer un prétraitement des séquences pour éliminer les erreurs, une analyse de l'unité taxonomique opérationnelle (OTUs), une description de la diversité des échantillons alpha et bêta, une taxonomie phylogénétique des OTU et la une visualisation de la diversité des échantillons à l'aide de dendrogramme Krona.

Mots clé : Métagénomique, microbiote intestinal, MOTHUR et phylogénie.

Abstract

The intestinal microbiota, a key player in health, is considered to be an integral part of the human body. The intestinal microbiota plays a decisive role in our health. It is extremely diversified and varies from one individual to another. In order to study its microbial composition and determine its distribution, we used pyrosequencing data of the 16s ribosomal RNA gene to study the differences in gut microbiota between four groups of subjects with and without CKD.

This work focuses on the use of the MOTHUR pipeline for processing 16s rRNA data. This tool allowed us to perform sequence preprocessing to eliminate errors, operational taxonomic unit (OTU) analysis, alpha and beta sample diversity description, OTU phylogenetic taxonomy and sample diversity visualization using Krona dendrogram.
Key words: Metagenomics, gut microbiota, MOTHUR and phylogeny.

ملخص

تعتبر الجراثيم المعوية التي تلعب دورا رئيسيا في الصحة جزءا لا يتجزأ من جسم الإنسان تلعب الجراثيم المعوية دورا حاسما في صحتنا. إنه متنوع للغاية و يختلف من فرد الى آخر. من أجل دراسة ARN16s و تركيبته الميبكروبية و تحديد توزيعه لدراسة الإختلافات بميكروبوتا الأمعاء بين أربع مجموعات من الأشخاص الذين يعانون من المرض و بدونه .

يركز هذا العمل على استخدام الحمض الريبوزومي s16 سمحت لنا هده الأداة بإجراء معالجة مسبقة و وصف تنوع عينة (OTU) لتسلسل و التخلص من الأخطر و تحليل وحدة التصنيف التشغيلية و تصور تنوع العينة باستخدام ألفا و بيتا و تصنيف التطور الوراثي (OTU)

الكلمات المفتاحية : علم الجينات ,ميكروبيتا الأمعاء , MOTHUR تطور

TABLE DES MATIERES

Chapitre 3 : Matériels et méthodes

Chapitre 4 : Résultats et Discussion

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ABREVIATIONS

CKD : Chronic Kidney Disease.

gMB : gut Microbiota.

IRC : Insuffisance Rénale Chronique.

IUPAC: International Union of Pure and Applied Chemistry.

NCBI : The National Center for Biotechnology Information.

NGS: Next Generation Sequencing.

NPC : Néphropathie Chronique.

OTU : Unité Taxonomique Opérationnelle.

PCR: Polymerase Chain Reaction.

PH : Potentiel Hydrogène.

RTPCR : Reverse Transcription Polymerase Chain Reaction.

SRA : Sequence Read Archive.

.

➢ Introduction

INTRODUCTION :

Les progrès des plates-formes de séquençage de nouvelle génération (NGS) ont fourni une analyse à haut débit de séquences nucléotidiques. Ces technologies sont capables de séquencer simultanément et indépendamment des milliards de molécules d'ADN, ainsi leur combinaison avec des approches bioinformatiques a permis aux chercheurs d'approfondir l'étude du rôle critique de l'ADN des organismes vivants.

L'avènement de ces nouvelles générations a permis aux chercheurs d'étudier le microbiome des différents environnements avec une résolution et un débit sans précédent. Cela a stimulé le développement d'outils bioinformatiques sophistiqués pour analyser les vastes quantités de données générées par ces technologies. Par conséquent, les chercheurs ont besoin d'une compréhension claire des concepts clés nécessaires à la conception, à l'exécution et à l'interprétation des expériences sur les données de la métagénomique.

Le microbiote intestinal (gMB : *gut Microbiota*) a un rôle clé tous au long de la vie humaine et garantie un intestin en bonne santé. Chaque personne possède une identité propre au niveau de son microbiote intestinal où un tiers des bactéries est commun entre les individus. Certains facteurs peuvent moduler de manière négative le microbiote et son environnement tels que l'alimentation inadaptée, le stress et la prise de médicaments. Le microbiote intestinal est au cœur du métabolisme de l'hôte et de la mésostase immunitaire, il a été impliqué aussi dans des maladies allant du cancer colorectal et du diabète aux troubles du spectre autistique. [1]

Les amplicons du gène de l'ARN ribosomique 16s (ARNr), un sous-composant de la sous-unité ribosomique procaryote 30s, ont été utilisé pour comprendre la diversité microbienne de l'intestin, ils constituent ainsi des marqueurs génétiques informatifs pour déterminer la variation taxonomique des communautés microbiennes intestinales en cas de maladies. [2]

En utilisant les bases de données de séquençage des gènes ARNr 16s d'une diversité microbienne sur les malades d'insuffisance rénale chronique (IRC ou NPC : Néphropathie Chronique) ou CKD (*Chronic Kidney Disease*), un pipeline de bioinformatique est devenu nécessaire pour comprendre cette maladie. Nous avons choisi le pipeline mothur pour mesurer l'abondance de la flore intestinale et étudier les effets de la flore intestinale sur les patients CKD.

Mothur vise à être un progiciel complet qui permet aux usagers d'utiliser un seul logiciel pour analyser les données de séquence communautaire. Il s'appuie sur les outils préliminaires pour fournir un logiciel flexible et un package puissant d'analyse des données de séquençage. Pour notre cas d'étude, nous avons utilisé mothur pour découper, filtrer, aligner des séquences, calculer les distances, attribuer des séquences à des unités taxonomiques opérationnelles et décrire la diversité des quatre échantillons préalablement caractérisés par pyroséquençage de fragments de gènes d'ARNr 16s. Cette analyse avec plus de 340 000 données de rangs apparents a été réalisée avec ce progiciel à l'aide d'un ordinateur portable.[3]

- Ce travail est subdivisé en quatre chapitres. Le premier dévoile la métagénomique et ses approches.
- Dans le deuxième chapitre on a essayé de bien connaître le microbiote intestinal, sa composition, sa distribution, et sa relation en cas de maladies avec un historique à propos des études impliquées sur le microbiote.
- Le troisième chapitre présente notre matériel de bases de données de séquençage et la méthodologie menée pour la réalisation de ce travail.
- Le dernier chapitre contient une description des principaux résultats de notre recherche, tandis que la section de discussion interprète les résultats qui fournissent l'importance des conclusions.

➢ *Chapitre1 :*

Approches

Métagénomiques

Chapitre1 : Approches métagénomiques

1. Introduction

Les domaines de la génomique et de la métagénomique ont apporté un soutien illimité à la progression de nos notions en génétique bactérienne. D'autre part, les bactéries sont de plus en plus liées à la santé humaine, et compte tenu de leur énorme diversité dans les besoins métaboliques, la métagénomique est utilisée pour résoudre les difficultés liées à leur culture. Ainsi, les nouvelles technologies de séquençage ont permis, à grande échelle, la production des séquences d'ADN particulières à des fins de caractérisation et de comparaison pour élucider des questions communément associées à la santé humaine et à son environnement. Les progrès de la génomique et de la métagénomique requièrent de la bioinformatique qui a le potentiel de gérer et de manipuler des données biologiques massives. [4]

2. Métagénomique

C'est une nouvelle approche pour étudier le matériel génétique extrait directement d'échantillons environnementaux, basée sur des techniques et des procédures utilisées pour l'analyse indépendante du contenu du génome microbien total des microorganismes dans un environnements particulier, y compris l'intestin, le sol et de l'eau. Ainsi comment ces communautés changent-elles en réponse aux changements des propriétés physiques et chimiques de leur environnement grâce à l'analyse des gènes d'ARNr, par un criblage génétique fonctionnel ou une analyse des données de séquençage, l'offre un objectif puissant pour observer les communautés microbiennes à plus grande échelle et aborder les questions fondamentales de la diversité, de l'évolution et de l'écologie microbiennes. [5]

Les principaux domaines d'intérêt de la recherche en métagénomique sont la diversité microbienne, la composition des communautés, les relations génétiques et évolutives, les activités fonctionnelles, les interactions et les relations avec l'environnement.[6]

La recherche en métagénomique s'est développée rapidement grâce au séquençage à haut-débit et à la bioinformatique. .[6]

3. Le métagénome

Tout matériel génétique constitué d'un mélange des génomes de nombreux microorganismes individuels présent dans un échantillon environnemental, "connu sous le nom de génome environnemental microbien".

La vie microbienne existe dans presque tous les environnements, de l'environnement le plus familier tel que le sol du jardin, les feuilles des plantes vertes ainsi que leurs racines ou le tuyau sous notre évier.[7]

Mais ces habitats microbiens comprennent également des environnements considérés comme difficiles à survivre en raison de conditions extrêmes.

Par exemple : Au fond de la mer dans la banquise arctique ou dans les déserts de sel les microbes peuplent également notre propre corps et vivent sur notre peau ou à l'intérieur de l'intestin. Presque tous les environnements sur terre sont colonisés par différents types de micro-organismes (figure 1).

Nous étudions le métagénome car la plupart des microorganismes ne peuvent pas se développer sur une culture pure et la culture ne peut jamais capturer le spectre complet de la diversité microbienne, la métagénomique fournit des informations indépendantes de la culture sur les microorganismes environnementaux.

Figure 01 : déférents environnements pour l'échantillonnage du métagénome**.[8]**

4. Approches métagénomiques

L'étude métagénomique est basée sur deux approches qui sont la métagénomique descriptive qui permet "l'estimation des abondances microbiennes relatives en fonction de différentes conditions physiologiques et environnementales pour révéler la structure communautaire et la variabilité du microbiome", d'autre part la métagénomique fonctionnelle vise à "étudier les interactions hôte-microbe et microbe-microbe et construire des modèles d'écosystèmes dynamiques prédictifs pour refléter les liens entre les microbes ou les identités communautaires". C'est avec le développement conjoint de ces approches métagénomiques, en se basant sur l'extraction directe de l'ADN bactérienne, la métagénomique permettra de comprendre le rôle de l'environnement naturel et de découvrir ainsi de nouveaux médicaments pouvant être potentiellement utiles au développement des molécules à intérêt thérapeutique comme des antibiotiques. [9 ,10]

5. Analyse métagénomique

De nombreuses méthodes distinctes ont émergé pour étudier le microbiome, l'une des méthodes les plus couramment utilisées consiste à analyser l'ADN total et c'est donc là que nous prenons notre échantillon. Nous extrayons l'ADN directement de cet échantillon et nous pouvons donc séquencer un marqueur taxonomique comme le gène 16s d'ARNr qui nous indique quelles bactéries se trouvent dans notre échantillon. Grâce à des fonctions stables des gènes de l'ARNr 16s, la présence de ce dernier dans la plupart des microorganismes à une taille suffisante pour l'analyse bioinformatique, en utilisant la PCR pour cibler et amplifier la partie de la région hypervariable du gène de la sous-unité bactérienne de l'ARN ribosomal 16s. Cette stratégie est l'une des méthodes les plus couramment utilisées pour comprendre la taxonomie et la phylogénie microbiennes.

Différentes études se contredisent sur le choix de la région ciblée, indépendamment du type de microbiote, certaines régions sont conseillées selon le microbiote d'intérêt.

Par exemple : la région V3 - V4, pour le microbiote intestinal humain (figure 2).

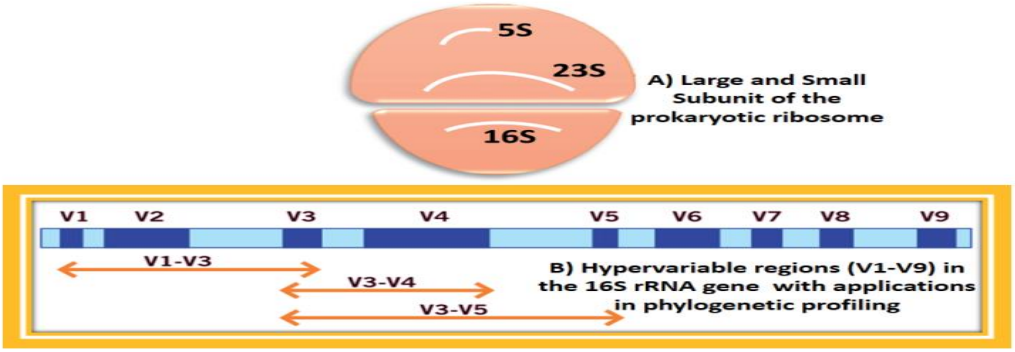

Figure 02 : Représentation générale des ribosomes et régions à séquence variable.[11]

La métagénomique basée sur la méthode *shotgun* permet le séquençage aléatoire de l'ensemble du métagénome d'un échantillon sans amorces spécifiques, réduisant ainsi les biais dans la sélection des amorces. Cette démarche a ajouté des informations à la composition des gènes et la capacité fonctionnelle qui nous indique quels gènes sont codés avec les génomes des bactéries dans nos échantillons. [11,12]

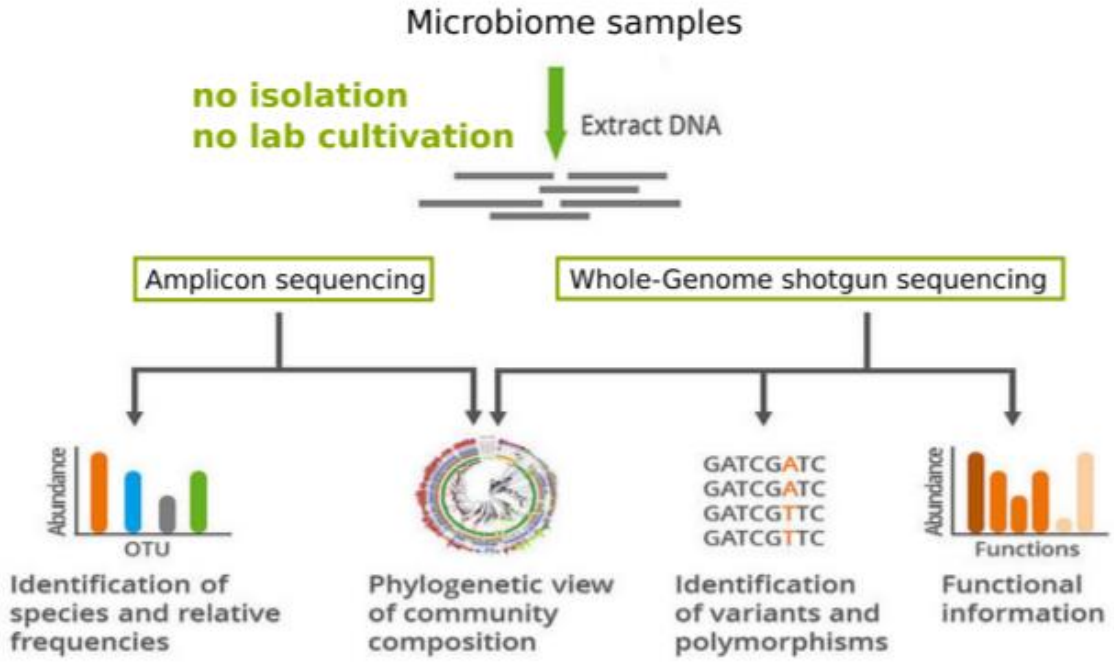

Figure 03 : Techniques couramment utilisées pour l'étude du microbiome.[8]

6. Technologies de séquençage et métagénomique

Il est souvent admis que la diversité des microorganismes est énorme et que la grande majorité (~99%) de ces microorganismes ne sont pas couramment cultivés et ne peuvent donc pas être cultivés à l'aide de méthodes traditionnelles. Par exemple, l'électrophorèse sur gels, le polymorphisme, l'hybridation *in situ* par fluorescence, la RT-PCR sont utilisés depuis des décennies pour analyser des communautés biologiques ou environnementales individuelles, mais les détails limités de toutes les communautés microbiennes et de leurs membres individuels obtenus par ces techniques rendent difficile la caractérisation complète de communautés complexes ou diverses [13],

Les progrès des technologies de séquençage de nouvelles générations (NGS) ont fourni une analyse de séquence à haut débit, permettant le séquençage simultané et indépendant de milliards de molécules d'ADN et décrire précisément la structure et la diversité de métagénome ainsi que son altération en pathologie. La combinaison de ces techniques avec des approches métagénomiques aide les chercheurs à étudier la diversité microbienne et à comprendre les fonctions et les relations entre les différentes communautés microbiennes comprennent le séquençage à base d'amplicon PCR.

Parmi les méthodologies NGS, le séquençage d'amplicon ciblé du gène ARNr 18s, ARNr 16s, dénommé «*16s-seq*», est actuellement la stratégie la plus utilisée pour l'identification et quantification des bactéries résidentes humaines. En effet, le séquençage de nouvelle génération permet d'étudier et d'identifier des organismes directement à partir des habitats sans préparation préalable, cet avènement de plusieurs stratégies a enrichi la métagénomique. Récemment, de nombreuses méthodes NGS ont été développées, notamment le pyroséquençage Roche/454, le séquençage Illumina/Solexa et le séquençage Applied Biosystems/SOLiD. [14]

➢ Chapitre 2 :

Génomique

Du Microbiote

Intestinal humain

Chapitre2 : Génomique du microbiote intestinal humain

1. Bref historique sur l'étude du microbiome humain

Dans les années 1800, Robert Koch et Louis Pasteur ont développé le concept de microbes (ou microorganismes), selon lesquels les maladies infectieuses humaines sont causées par des infections microbiennes. Plus de 100 ans plus tard, la prochaine révolution conceptuelle implique des communautés naturelles de microorganismes, collectivement connus sous le nom de microbiome. [15] La recherche sur le microbiome humain est un domaine relativement nouveau dans la biologie humaine, connu sous le nom d'« organe oublié » du corps humain, et étroitement liée à la microbiologie. Cette recherche commence par des méthodes réductionnistes telles que l'utilisation de milieux de culture et de la microscopie pour identifier et caractériser les souches bactériennes individuelles.

Initialement, seules les espèces bactériennes cultivables ont été étudiées, mais il existe un grand nombre de microorganismes qui n'ont pas été cultivés en laboratoire. Ceci se trouve lorsque le nombre de microbes vus au microscope ne correspond pas au nombre de microbes qui poussent sur la plaque médiane. En 1970, Carl Woese a proposé que les gènes d'ARN ribosomal puissent être utilisés comme marqueurs moléculaires pour la taxonomie bactérienne. Par conséquent, les scientifiques ont développé des techniques indépendantes de la culture pour l'amplification du gène de l'ARNr 16s et son séquençage basé sur des méthodes PCR. [16] Ces stratégies ont été utilisées pour classer phylogénétiquement les microbes intestinaux, puis annoter leurs fonctions dans des écosystèmes microbiens natifs distincts.[17]

2. Distribution des communautés microbiennes dans les tractus gastro-intestinaux humain

Les conditions environnementales du tube digestif humain ne sont pas uniformes, mais il existe des différences significatives entre l'estomac et le colon. Par conséquent, il n'est pas surprenant que les communautés microbiennes dans différentes parties du tube digestif diffèrent, sous l'activité métabolique. Les microbes protecteurs commencent d'abord à se développer sur la peau du bébé et continuent dans l'estomac et sont connus pour assurer leur survie dans l'environnement de l'estomac en produisant de l'uréase, une étude sans culture des séquences du gène microbien de l'ARN 16s dans 23 échantillons de biopsie de la muqueuse gastrique a révélé une communauté diversifiée de 128 phototypes telle que (*Bacteroidetes*, les actinobactéries …).

L'intestin grêle colonisé représente alors la partie la plus longue du tube digestif, augmentant avec l'évolution des conditions et de la densité des cellules partout. D'une manière générale, la composition du microbiote intestinal grêle variait davantage d'un individu à l'autre. Une étude récente comparant le microbiote duodénal rapporte la présence de nombreux genres tels que *Brevibacterium*, *Veillonella* et *Prevetella*. (2017) En Suite, par d'autres genres dans le côlon, ces populations de bactéries sont appelées flore microbienne, et elles effectuent une grande variété d'activités cruciales. Ils nous protègent de l'infection par des microorganismes virulents.[15] (figure 4).

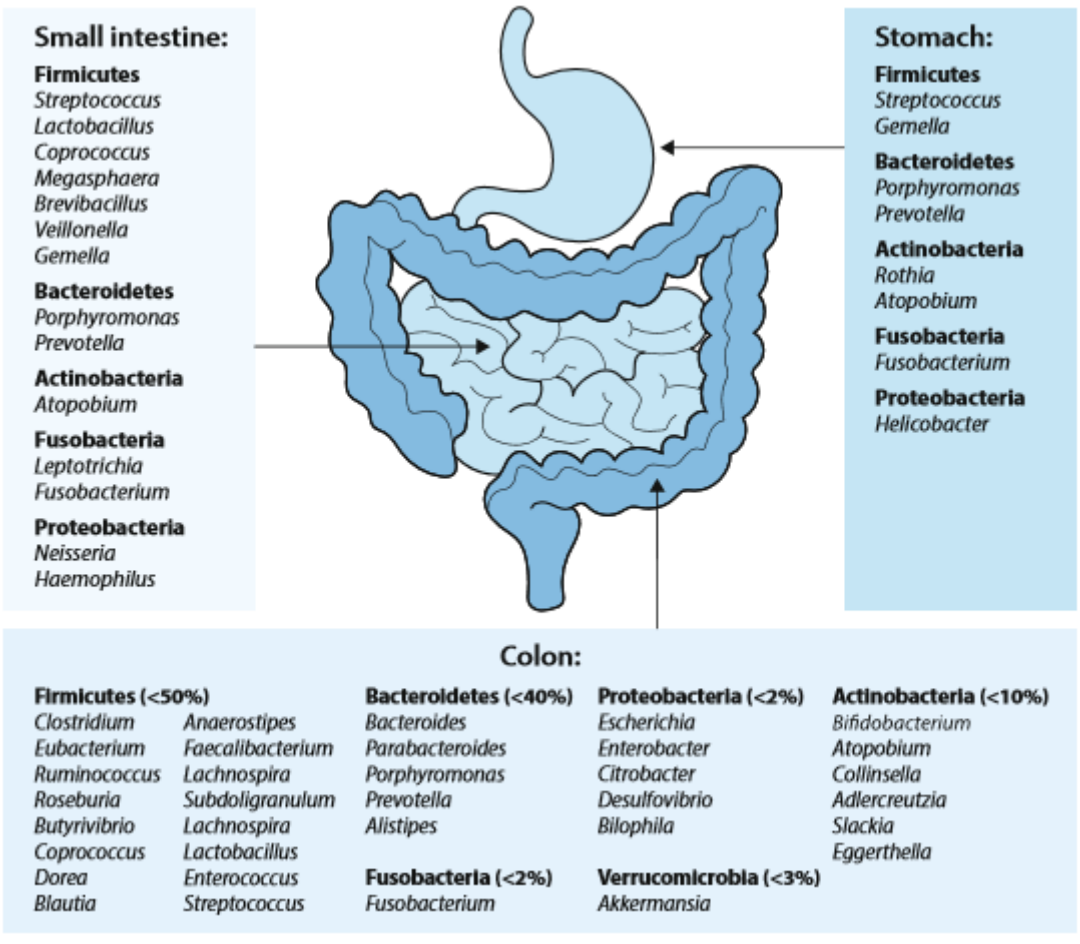

Figure 04 : Principaux genres bactériens rencontrés dans les différentes sections du tractus gastro-intestinal.[15]

3. Composition et Fonction du microbiome intestinal

Le microbiote intestinal ou la flore intestinale est l'ensemble des microorganismes hébergés dans le tube digestif principalement dans le côlon. Un pH et une concentration en oxygène variables affectent l'abondance du microbiome intestinal dans le tractus gastro-intestinal. Notre microbiote comprend près de cent mille milliards de bactérie.

Le microbiome intestinal s'établit au cours des premières années de la vie et contient jusqu'à 100 milliards de microbes, sa composition se complique avec le temps et elle comprend de très nombreuses espèces de bactéries mais aussi des champignons microscopiques, des protozoaires, des archées et des virus. Le classement taxonomique du microbiome intestinal comprend les espèces, les genres, les familles, les ordres, les classes et les embranchements. La plupart des espèces appartiennent aux *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Proteobacteria* et *Verrucomicrobia phyla*. [18,19]

La flore intestinale peut même être considéré comme un véritable organe à part entière impliqué dans de multiples fonctions physiologiques telle que :

- Protection contre les agents pathogènes tuant ou en inhibant les organismes ;
- Régulation du système immunitaire en influençant la production de cytokines et anticorps ;
- Régulation du métabolisme impliqué dans une variété de processus métaboliques.

Ces processus y compris la régulation de l'homéostasie énergétique et pondérale, la production d'acides gras chaîne courte (après fermentation de fibres non digestibles) et vitamines (vitamines B et vitamine K), contrôle glycémique, interaction avec l'augmentation de l'étain et du métabolisme des lipides et l'os. **[**20]

4. Origine de la variation normale du microbiote intestinal

Déterminant qu'un microbiome intestinal sain est très variable, alors nous devons comprendre pourquoi cela change afin que ces informations puissent être utilisées pour adapter la thérapie ou les essais cliniques ? Par exemple, la mesure dans laquelle les membres d'une même famille présentent de microbes similaires déterminera les antécédents familiaux ; L'enrichissement informatif induit par le microbiote, le degré de changement dans le microbiome, l'âge ou la grossesse doivent être pris en compte lors de la conception des groupements distincts. Les facteurs externes, du microbiome au régime alimentaire, éclaireront les stratégies de traitement des maladies liées au microbiome.[21]

4.1. Facteur Âge

Des changements significatifs dans le microbiote intestinal se produisent tôt dans la vie, avec une diversité et une stabilité accrue au cours des trois premières années (Figure 5). La maturation du microbiote humain est un exemple de succession écologique. Après la colonisation initiale, un continuum de composition et de fonction de la communauté se produit. Une communauté climax relativement stable est établie.[22]

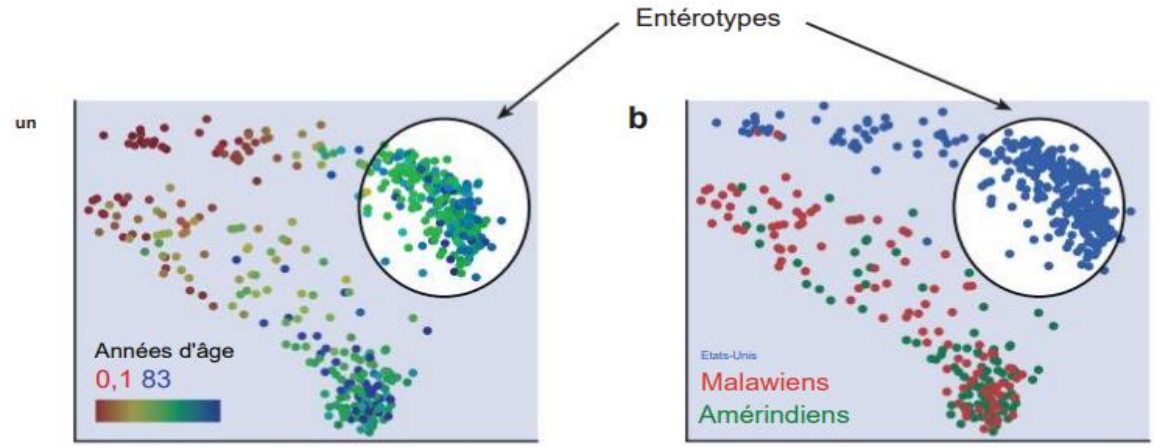

Figure 5 : Diversité microbienne humaine et entérotypes. (Images associées à la diversité du microbiome L'homme est considérablement élargi. Le microbiote de 531 enfants était associé à Adultes en bonne santé (Amérindiens) du Malawi, Amazonas, Venezuela Les États-Unis utilisent la séquence du gène de l'ARNr 16S dans)

Le microbiote du nourrisson est relativement instable et fonctionnel, la variation des compagnies microbiennes et de gènes fonctionnels entre les habitants est plus importante chez le nouveau-né que chez les adultes. Les microbiomes infantiles partagent des propriétés caractéristiques entre les individus et les populations à la fois compositionnelles (de nombreuses bifidobactéries et richesse en espèces plus faible que chez les adultes) et fonctionnelles.

Les microbiomes des nourrissons sont affectés par l'utilisation d'antibiotiques. Le choix de l'alimentation postnatale (lait maternel ou préparation pour nourrissons) influe sur le processus de colonisation du nouveau-né. En vieillissant, l'introduction d'aliments solides dès 2 ans et la production d'hormones sexuelles dès la puberté ainsi que la ménopause féminine apportent une richesse et une complexité supplémentaires au microbiome intestinal. Bien qu'il ne soit pas clair si les différences dans le microbiote au début de la vie affectent sa composition à l'âge adulte, pourtant les apports sont différents. Des effets induits lors de l'âge de petite enfance peuvent influencer la susceptibilité aux troubles immunitaires à l'âge adulte, tels que l'asthme et les maladies atopiques.[21]

4.2. Effet génétique, environnemental et alimentation

Qu'en est-il des facteurs tels que l'environnement et l'alimentation, la forme du microbiote intestinal humain reste floue, en raison que ces facteurs sont souvent contradictoires. Chez les jumeaux et les couples mère-fille, la composition du microbiote est plus similaire que celle des individus non apparentés, ceci suggère qu'il peut avoir des effets génétiques sur la flore intestinale humaine. Cependant, d'autres études ont montré que les jumeaux et les adultes, en vie collective, ont des microbiomes également similaires, ce qui suggère que les similitudes peuvent être environnementales plutôt que génétiques.

Les populations peuvent être distinguées par des différences caractéristiques de la flore intestinale. Par exemple, le microbiome des enfants italiens est similaire à ce des enfants en Afrique rurale, ainsi les enfants et les adultes aux États-Unis différent de la multitude des populations du Malawi et de l'État d'Amazonie au Venezuela.

Tant que génétiquement distinctes, ces peuplements diffèrent également par d'autres agents qui peut affecter le microbiome, par exemple l'hygiène et la propreté adéquates, la qualité de l'alimentation et les normes d'utilisation d'antibiotiques. Les facteurs culturels, notamment, peuvent jouer un rôle clé dans la formation du microbiome intestinal. [21]

5. Microbiome intestinal et les maladies

La relation symbiotique entre l'hôte humain et le microbiote intestinal peut déclencher des réponses biologiques spécifiques locales et systémiques.

Les colonies peuvent également provoquer des maladies spécifiques à l'activité (dysbiose) telles que l'obésité et la malnutrition chez le diabète de type 2, les maladies inflammatoires de l'intestin, les maladies neurologiques et le cancer. La dysbiose peut être la cause et/ou la conséquence de ces maladies ou des activités de protection contre les maladies (probiotiques). [23]

6. Méthodologie d'étude du microbiome intestinal humain

6.1. Analyse phylogénétique de la communauté microbienne

6.1.1. Méthode dépendante de la culture :

Les approches basées sur la culture sont toujours considérées comme le protocole de référence pour identifier de nouvelles espèces et fournir des informations pertinentes dans le monde des microorganismes. Il s'agit d'une méthode peu coûteuse et plus fiable pour l'identification bactérienne. Mais il n'est pas prouvé qu'ils soient pleinement efficaces contre les bactéries anaérobiques et non sensible. On sait que plus de 30 % des espèces bactériennes ne peuvent pas être planté en dehors de leur habitat. De plus, le microbiote intestinal comprend non seulement les bactéries, mais aussi les phages, les archées, les espèces fongiques et les eucaryotes unicellulaires. Par conséquent, nous avons besoin d'une approche d'enquête plus large pour couvrir toutes les préparations microbiennes qui impliquent et contribuent à stabiliser le microbiote intestinal. [16]

6.1.1.1. culturomique

L'importance des méthodes dépendantes de la culture pour identifier les microorganismes dans la communauté microbienne intestinale ne peut être sous-estimée et les chercheurs ont dévoilé les méthodes basées sur la culture en ajoutant de nombreux instruments sophistiqués et des milieux de croissance appropriés. Cela permet de cultiver la plupart des bactéries cultivables auparavant considérées comme impossibles en laboratoire. La culturomique est une méthode de culture exigeante et qui utilise plusieurs conditions de culture, la spectrométrie de masse MALDI-TOF et le séquençage de l'ARNr 16s pour identifier les espèces bactériennes. Au cours de son développement, l'objectif principal était de permettre à la méthode de fournir une variété de conditions de culture qui soutiennent la croissance de bactéries discernantes dans l'intestin humain.

En effet, cette méthode de culture nous aidons à indiquer plus sur les aspects fonctionnels du microbiome intestinal, notamment sa composition.[24]

6.1.1.2. Dosage microfliudiques

La microfluidique ou cellule sur puce c'est la science de la manipulation de fluides à l'échelle micrométrique, alors la microfluidique est une technologie qui a été utilisé pour faire avancer la science qui offrent un microenvironnement spécifique pour les réactions biochimiques. La technique microfluidique, a toujours été liée au vivant. Le corps humain lui-même est un système microfluidique, les vaisseaux sanguins, les bronches et une partie du système digestif sont des systèmes microfluidiques, cette technique est également appliquée dans les études du microbiote intestinale connus aussi sous le nom *gut-on-chip*. Avec les micropuces qui exposent les cellules cultivées, ils fournissent un environnement de croissance de type gastro-intestinal pour la co-croissance des cellules épithéliales humaines et une nutrition spécifiques nécessaires à ces croissances bactériennes. Par exemple (diaphonie microbienne humaine (HuMiX)).

En effet, la microfluidique permet de faire des tests, améliorer de nombreux aspects de la vie humaine, sauvé des vies humaines grâce à des médicaments « encapsulés ». [25]

6.1.2. Méthodes indépendantes de la culture

Concernant les méthodes indépendantes de la culture, il existe plusieurs étapes à suivre permet les étapes majors :

6.1.2.1. Méthode de prélèvement et de standardisation des échantillons

Dans toute analyse microbiologique ou biochimique. La préparation des échantillons est une étape critique et importante qui détermine la précision et l'efficacité de toute technique analytique simple ou complexe. Lors des études du microbiome humain, il existe deux principaux types d'échantillons, les biopsies fécales et muqueuses.[27]

6.1.2.2. Analyse métagénomique de la communauté

D'après l'analyse métagénomique de la communauté, les microbiologistes ont développé plusieurs méthodes avancées pour comprendre la composition du microbiote intestinal et la croissance microbienne, révolutionnant le domaine de la recherche sur la communauté microbienne humaine. La métagénomique est la première technique qui permet d'identifier phylogénétiquement 80% des microorganismes non cultivés. Les techniques métagénomiques classiques reposent sur l'ARN ribosomal 16s dont la fonction principale est de réguler la synthèse

des protéines, méthodes de séquençage de l'ADN nouvelle génération. Les amplicons du gène de l'ARNr 16s sont isolés et séquencés, c'est donc maintenant la méthode la plus réussie et la plus largement indépendante pour la classification taxonomique des cultures récentes. Le séquençage d'ADN de nouvelle génération a abouti à des approches métagénomiques plus rapides et plus complexes que le séquençage métagénomique du génome entier.[28]

6.1.2.3. PCR en temps réel

La PCR en temps réel ou PCR quantitative (qPCR) est une technique utilisée pour l'analyse du microbiome, en particulier pour l'analyse phylogénétique. Elle peut être utilisée à la fois quantitativement et semi-quantitativement, selon l'application, pour quantifier l'ADN dans les échantillons de selles ou de muqueuse intestinale en utilisant des sondes fluorescentes ou des molécules colorées insérées entre des molécules d'ADN double brin ou des amplicons d'ARN 16s. Ces sondes émettent des signaux forts dont l'intensité est proportionnelle à la quantité d'échantillon d'ADN présent. La qPCR a été utilisée pour étudier l'environnement écologique des populations normales et obèses, pour comprendre la diversité microbienne fonctionnelle dans le microbiote intestinal des patients âgés et pour voir l'effet des antibiotiques sur les microbes intestinaux. Les méthodes basées sur la qPCR conviennent à une analyse phylogénétique prédictive précise. [29,30]

➢ Chapitre 3 :

Matériels

et

Méthodes

Chapitre 3 : Matériel et Méthodes

1.Matériel :

1.1. Les données :

Le fichier avec lequel nous avons travaillé est celui de la métagénomique (données du séquençage des ARNr 16s) qui a été séquencés à partir des prélèvements du microbiote de l'intestin humain sur les patients CKD (Chronic Kidney Disease) car il y'avait des travaux voisins sur cette maladie. Avec plus de 340 000 données de rangées à l'aide de la méthode de séquençage Illumina NovaSeq 6000. En raison de la variabilité de la communauté microbienne, nous avons choisi quatre types de variants des échantillons nous nous sommes intéressés à quelques facteurs d'origine de la variation normale (l'Âge et le sexe) chez les malades et les contrôles des deux sexes. Nous avons choisi des adultes avec des âges rapprochés pour éviter l'influence de la déférence d'âge sue la distribution des bactéries.

Ces échantillons ont été téléchargés de la base de données NCBI (SRA) au format FASTQ de l'Université pharmaceutique de Chine, ils ont été publiés le 18-10-2021. Cette étude intrigue toujours les chercheurs.

- **Sequence Read Archive (SRA) :** est l'archive principale pour les données de séquençage à haut débit, qui fait partie de l'International Archives Partnership du NCBI, de l'Institut européen de bioinformatique qui stocke les données de séquence brutes des technologies de séquençage de « nouvelle génération » notamment Illumina 454, pour lire et produire des formats tels que FASTQ.

Le format Fasta est un format textuel pour représenter des séquences nucléotidiques ou des séquences peptidiques. En format Fasta nous commençons l'écriture par une ligne de description débutant par ">" et comprenant l'identificateur de séquence et une description. Les lignes suivantes contiennent les données de séquence qui sont censés être représentés dans les codes d'acides nucléiques et d'acides aminés standards de L'IUPAC.

FASTQ est devenu un format de fichier commun pour le partage des données de lecture de séquençage combinant à la fois la séquence et un score de qualité associé par base (figure 6).

```
@F54C.1 1 length=250
ATGACCTACGGGTGGCTGCAGTGATTAACCTTTAGCAATAAACGAAAGTTTAACTAAGCTATACTAACCCCAGGGTTGGTCAATTTCGTGCCAGCCACCGCGGTCACACGATTA
+F54C.1 1 length=250
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@F54C.2 2 length=250
ATGACCTACGGGAGGCTGCAGTAGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGGGGGATGAATGCCTTCGGGTTGTAAACCCCTTTCAGCAGGGAAG
+F54C.2 2 length=250
FFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFF
@F54C.3 3 length=250
ATGACCTACGGGGGGCTGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGCCTTTTGGTTGTAAAGCACTTTAAGCAGGGAGG
+F54C.3 3 length=250
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF::FFFFFFFFFFFFFFFFFFFFFFFFFFFF
@F54C.4 4 length=250
ATGACCTACGGGTGGCAGCAGTGAGGAATATTGGTCAATGGACGAGAGTCTGAACCAGCCAAGTAGCGTGAAGGATGACTGCCCTATGGGTTGTAAACTTCTTTTATATGGAA
+F54C.4 4 length=250
FFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFF::F,FFFFFFFFFFFFFF
@F54C.5 5 length=250
ATGACCTACGGGTGGCTGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGGGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCGACAGGGACG
+F54C.5 5 length=250
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
```

Figure 06 : Fichier d'exemple de format FASTQ.

Tableau 1 : Informations sur les données utilisées pour l'analyse.

Séquence	SRR16472981	SRR16472980	SRR16472979	SRR16472978
La source	METAGENOMIC	METAGENOMIC	METAGENOMIC	METAGENOMIC
Library Name	CKD1	CKD2	Control1	Control2
Sélection	PCR	PCR	PCR	PCR
Instrument	Illimuna NovaSeq 6000	Illimuna NovaSeq 6000	Illimuna NovaSeq 6000	Illimuna NovaSeq 6000
Stratégie	AMPLICON	AMPLICON	AMPLICON	AMPLICON
Disposition	PAIRE	PAIRE	PAIRE	PAIRE
Nombre de spots	83.1k	86.1k	86.6k	84.7k
Nombre de Bases brutes	41.6Mbp	43.0Mbp	43.3Mbp	42.4Mbp
Taille	15.7M	15.8M	14.5M	14.3M
Contenu GC	54.3%	55.7%	52%	52.1%
ID	22357453	22357454	22357463	22357464
Publié	2021-10-25	2021-10-25	2021-10-25	2021-10-25

1.2. Les environnements :

L'analyse bioinformatique de nos données /échantillons moléculaires a été réalisée à l'aide des plateformes, des softwares.

Linux : Ubuntu :

Linux est un système d'exploitation complet et libre, qui peut être utilisé en lieu et place de systèmes d'exploitation commercialisés, tels que Windows, de Microsoft. Il est accompagné de nombreux logiciels libres complémentaires, offrant un système complet aux utilisateurs.

Anaconda :

Anaconda est un gestionnaire de packages, un gestionnaire d'environnement, une science des données Python/R distribution et une collection de plus de 7 500 packages open source. Anaconda est gratuit et facile à installer, et il offre un support communautaire gratuit.

Logiciel :

Plusieurs outils bioinformatiques ont été développés pour analyser les données métagénomiques au niveau moléculaire (ARNr 16S) : QIIME, MOTHUR, DADA2, UPARSE. Dans ce travail on a choisi le pipeline mothur.

Mothur :

Mothur est un logiciel analysant des séquences brutes et générant des outils de visualisation pour décrire la diversité. Devenu l'un des outils bioinformatiques les plus cités pour l'analyse des séquences de gènes d'ARNr 16S et il s'agit d'une combinaison de plusieurs outils analytiques pour décrire et comparer les communautés microbiennes. Il fournit des exemples de données acquises à partir de différentes plates-formes de séquençage.
Avantages : Capable d'effectuer des analyses basées sur l'OTU.

MEGA

Est un logiciel qui prend en charge l'affichage des arbres au format Newick contenant les longueurs des branches ainsi que le Bootstrap ou d'autres décomptes.

GALAXY :

Galaxy est une plate-forme de flux de travail scientifique, d'intégration de données, de persistance et de publication de données et d'analyses qui vise à rendre la biologie computationnelle accessible aux chercheurs.

Tableau 2 : Caractéristiques de l'ordinateur utilisé pour l'analyse des données.

L'ordinateur	Les caractéristiques
Processeur	Intel(R) Core (TM) i5-7200U CPU @ 2.50GHz 2.70 GHz
RAM	8,00 Go DDR4
Stockage	222 GB HDD
Système	Windows 10

2. Méthodes

Notre objectif est d'étudier les effets de la flore intestinale chez des patients CKD à l'aide de pipeline mothur. Analyse des données d'ARNr 16S sur la diversité microbienne entre les patients sains et les patients atteints CKD.

Dans cette section, nous fournissons une description claire et précise des différents protocoles des basses de la procédure opératoire standard du laboratoire Schloss pour les données Illumina MiSeqces étapes. Workflow suivant résume le processus du pipeline proposé.

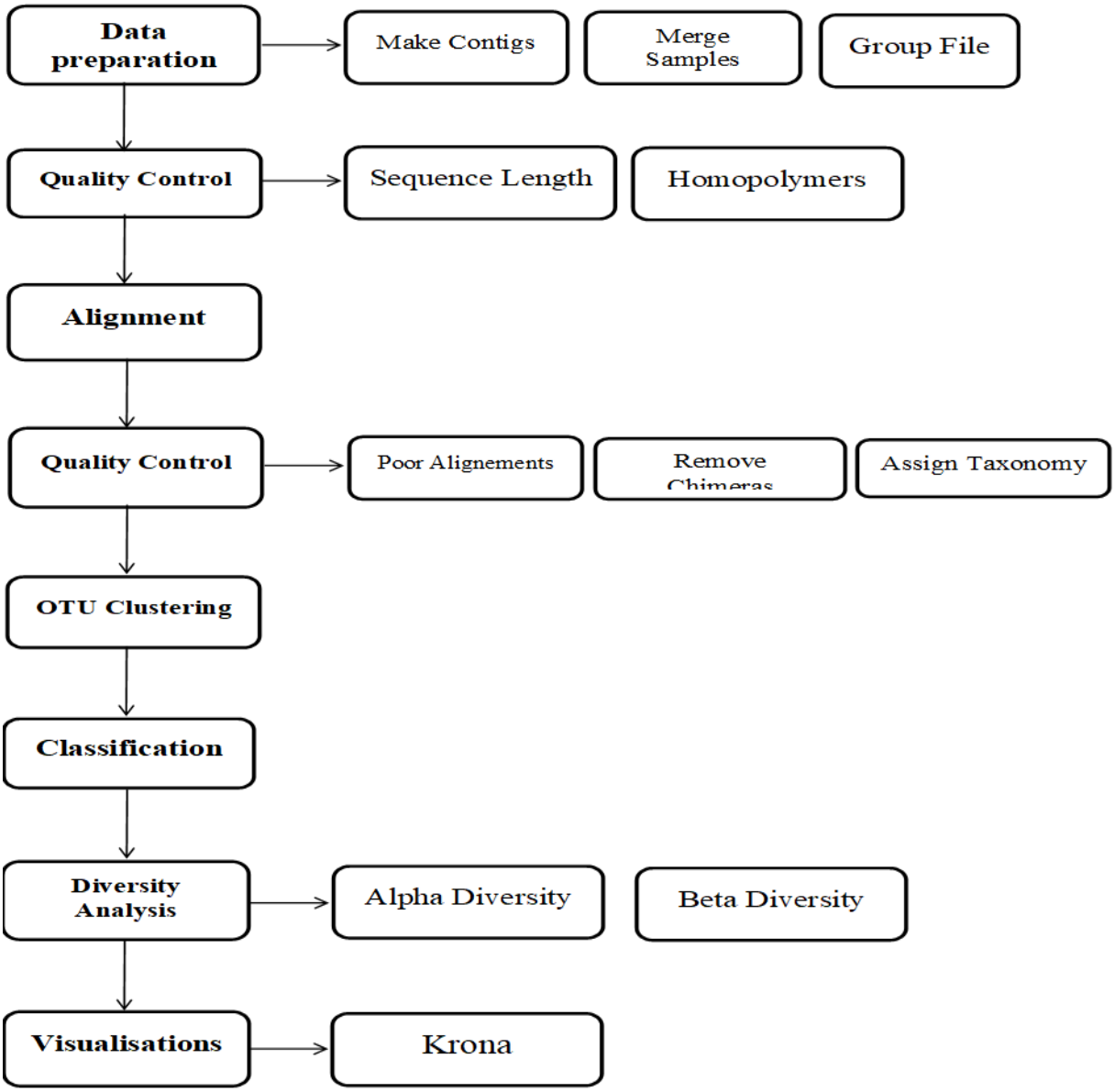

Figure 07 : Résumé des techniques utilisées pour l'analyse des données de séquençage de l'ARNr 16S.

Pour réaliser notre travail on a procédé au Chappidi, S., Villa, E. C., & Cantarel, B. L **[1] [31] [32]** :

https://training.galaxyproject.org

http://mothur.org/

2.1. Collection des données biologiques

En raison des progrès du NGS, un grand nombre d'ensembles de données de séquences métagénomiques ont été générés. Ces ensembles de données sont disponibles dans divers référentiels, y compris NCBI (SRA) https://www.ncbi.nlm.nih.gov/sra. Les données obtenues sont en format FASTQ.

Pour étudier les effets de la flore intestinale sur les patients CKD et déterminer la composition et la structure de la communauté bactérienne dans les ensembles de données d'ARN ribosomique 16s, on a utilisé la base de données https://www.ncbi.nlm.nih.gov/sra.

Nos séquences sont organisées de cette manière : F63CKD, F54C, M57C, M53CKD (où F : femelle, M : mâle, 63 : l'âge, CKD : patient, C : Contrôle).

Après le téléchargement des données on doit reconnaître l'appariement à partir des noms de fichiers, qui ne différeront que par _R1 ou _R2 dans le nom de fichier puisque chaque échantillon se compose de deux fichiers fastq distincts, l'un contenant *forward reads* et l'autre contenant the *reverse reads*. Cette étape est réalisée par la commande suivante (figure 8) :

```
(bioinformatics) manel@manel-VirtualBox:~/Desktop/m$ fastq-dump --split-3 F54C F63CKD M53CKD M57C
Read 86580 spots for F54C
Written 86580 spots for F54C
Read 86077 spots for F63CKD
Written 86077 spots for F63CKD
```

Figure 08 : Commande utilisé pour décompresse les données.

Après avoir assemblé nos données d'entrée, on les intégrera dans la plate-forme Galaxy.

1. Contrôle de qualité

1.1 Créer des contigs à partir paired-end reads

La première chose que nous voulons faire est de combiner nos deux ensembles de lectures pour chaque échantillon, puis de combiner les données de tous les échantillons. Cela se fait à l'aide de la commande ***make.contigs*** de Mothur**.**

- La commande ***make*.contigs** lit un fichier fastq vers l'avant et un fichier fastq vers l'arrière et génère de nouveaux fichiers fasta et de rapport.

➢ Le nombre de fichiers de sortie :

```
Group count:
F54C.fastq.gz    86580
F63CKD.fastq.gz  86077
M53CKD.fastq.gz  83112
M57C.fastq.gz    84746

Total of all groups is 340515

Output File Names:
combo_fastq.trim.contigs.renamed_map
combo_fastq.trim.contigs.fasta
combo_fastq.trim.contigs.qual
combo_fastq.contigs.report
combo_fastq.contigs.groups


mothur > quit
```

Figure 09 : Apparition de la fusion des données et le nombre de fichiers de sortie.

1.2 Nettoyage des données

Le fichier contig fasta nous aide à améliorer la qualité de nos données, ensuite on passe à la commande ***summary.seqs*** qui résume la qualité des séquences dans un fichier de séquence au format fasta aligné ou non.

➢ *Logfile* du fichier summary contient également des statistiques sur la qualité des séquences (figure 10) :

```
                Start   End     NBases  Ambigs  Polymer NumSeqs
Minimum:        1       250     250     0       3       1
2.5%-tile:      1       295     295     0       4       8513
25%-tile:       1       447     447     0       5       85129
Median:         1       453     453     0       5       170258
75%-tile:       1       466     466     0       6       255387
97.5%-tile:     1       472     472     2       32      332003
Maximum:        1       500     500     152     250     340515
Mean:   1       445.339 445.339 0.247798        8.31983
# of Seqs:      340515

Output File Names:
fasta.summary

It took 2 secs to summarize 340515 sequences.

mothur > quit
```

Figure10 : Apparition des statistiques de la qualité des séquences.

- Presque tous les reads sont entre 250 et 472 bases de longueur.
- 2,5% ou plus de nos reads avaient des appels de base ambigus (*Ambigs* colonne).

- La lecture la plus longue dans l'ensemble de données est de 500 bases.

Il y a 340515 séquences.

Notre région d'intérêt du gène 16s, ne fait qu'environ 250 à 472 bases de long. N'importe quel reads significativement plus long que cette valeur attendue ne s'est probablement pas bien assemblé à l'étape ***make.contigs***, de plus on constate que 2,5% de nos reads avait entre 2 et 152 appels de bases ambiguës. Dans les prochaines étapes, nous allons nettoyer nos données en supprimant ces problèmes des reads avec la commande ***screen.seq***, celle-ci est utilisée pour supprimer les lectures de mauvaise qualité :

1. Séquences à bases ambiguës (maxambig).
2. Contigs plus longs qu'un seuil donné (maxlength).

- Pour voir combien de lectures ont été supprimées lors de cette étape de sélection, cela peut être déterminée en regardant le nombre de lignes dans le fichier ***bad.accons*** output of ***screen.seq,*** dans notre cas 6,925 lignes sont éliminés.

1.3 Optimiser les fichiers pour le calcul

Les échantillons de microbiome contiennent souvent un grand nombre d'organismes identiques, nous nous attendons donc à trouver de nombreuses séquences identiques dans nos données. Pour accélérer le calcul, nous identifions d'abord les lectures uniques, puis nous enregistrons le nombre de fois où ces lectures distinctes ont été observées dans l'ensemble de données d'origine. Pour cela, nous utilisons la commande ***Unique.seqs***. À partir du fichier ***good.fasta*** output de ***Screen.seqs***.

À cette étape, la commande ***count.seq*** génère un tableau de regroupement des nombres de séquences identiques. Ce tableau détermine la classification taxonomique et l'abondance de l'OTUs (Unité taxonomique opérationnelle) dans les étapes en aval. ***count.seq*** est appliqué avec les paramètres suivants :

- **Name :** fichier output de ***unique.seqs***
- **Group :** fichier output de ***screen.seqs***

Output de count.seqs résume le nombre de fois où chaque séquence unique a été observée dans chacun des échantillons. Cela ressemblera à ceci (figure 11) :

name	total				
Representative_Sequence	total	F54C.fastq.gz	F63CKD.fastq.gz	M53CKD.fastq.gz	M57C.fastq.gz
1_F54C.fastq.gz	67	67	0	0	0
2_F54C.fastq.gz	2	2	0	0	0
3_F54C.fastq.gz	2	2	0	0	0
4_F54C.fastq.gz	1	1	0	0	0
5_F54C.fastq.gz	1	1	0	0	0
6_F54C.fastq.gz	1	1	0	0	0
7_F54C.fastq.gz	11	11	0	0	0
8_F54C.fastq.gz	1	1	0	0	0
9_F54C.fastq.gz	1	1	0	0	0
10_F54C.fastq.gz	1	1	0	0	0
11_F54C.fastq.gz	1	1	0	0	0
12_F54C.fastq.gz	1	1	0	0	0
13_F54C.fastq.gz	30	30	0	0	0
14_F54C.fastq.gz	1	1	0	0	0
15_F54C.fastq.gz	8	1	0	0	7
16_F54C.fastq.gz	1	1	0	0	0
17_F54C.fastq.gz	114	12	0	0	102
18_F54C.fastq.gz	143	15	0	0	128
19_F54C.fastq.gz	72	72	0	0	0
20_F54C.fastq.gz	1	1	0	0	0

Figure 11 : Exemple de tableau de regroupement des séquences uniques.

- La première colonne contient les noms des séquences représentatives, et les colonnes suivantes contiennent le nombre de répétitions de cette séquence observée dans chaque échantillon.

2. Alignement de séquences

On passe maintenant à l'étape d'alignement qui est une étape importante pour améliorer le regroupement de nos OTUs. Tout d'abord, on doit télécharger une base de données de référence (*SILVA.bacteria*) et assurer qu'elle se trouve dans le même dossier que le nôtre de candidature. En ce moment-là, nous sommes prêts à aligner nos séquences par la commande ***align.seqs.***

Puis la commande ***summary.seqs*** résumera la qualité des séquences dans un fichier de séquence au format fasta aligné ou non.

- Le fichier output de ***summary.seqs*** :

```
                Start   End     NBases  Ambigs  Polymer NumSeqs
Minimum:        0       0       0       0       1       1
2.5%-tile:      1044    1053    2       0       1       8340
25%-tile:       6332    25319   446     0       4       83398
Median:         6332    25319   453     0       5       166796
75%-tile:       6333    25319   466     0       6       250193
97.5%-tile:     43061   43116   472     2       6       325251
Maximum:        43116   43116   472     8       109     333590
Mean:   7500.16 24739.5 415.156 0.208708        4.87787
# of unique seqs:      219717
total # of seqs:       333590

Output File Names:
fasta.summary

It took 298 secs to summarize 333590 sequences.

mothur > quit
```

Figure 12 : Résumé de la qualité des séquences après alignement (table de sortie).

Les colonnes ‘’Start’’ et ‘’End’’ nous indiquent que la plupart des lectures s'alignent entre les positions 1044 et 43116, ce que nous nous attendions à trouver lors de l'utilisation du fichier de référence (silva). Cependant, certaines lectures étaient alignées à des positions très différentes, ce qui peut indiquer des insertions ou des suppressions aux extrémités alignées ou d'autres facteurs de complication, notez également dans la colonne ‘’Polymer’’ dans la table de sortie. Cela représente la longueur moyenne de l'homopolymère.

Notre base de données de référence ne contient aucun fragment homo-polymérique supérieur à huit lectures, toutes les lectures contenant des fragments aussi longs sont probablement le résultat d'erreurs de PCR et nous les supprimons judicieusement. Ensuite nous allons encore nettoyer nos données en supprimant les séquences mal alignées et toutes les séquences avec de longues étendues d'homopolymère.

2.1. Second nettoyage de données

Les séquences qui ne s'alignent pas sur la région 16S souhaitée peuvent être déterminées et supprimées avec les étapes ***screen.seqs*** et ***filter.seqs***

- ***Screen.seqs***

1. Supprimez tout lectures ne chevauchant pas la région V1-V9 (position 1044 et 43116).
2. Supprimez les homopolymères de longueur > 8 à l'aide de l'option maxhomop = 8.

- ***Filter.seqs***

1. Supprimez tout overhang à chaque extrémité de la région pour nous assurer que nos séquences ne chevauchent que cette région V1-V9.

2. Nettoyez notre fichier d'alignement en supprimant toutes les colonnes qui ont un caractère gap (-, où. Pour les espaces terminaux) à cette position dans chaque séquence (également en utilisant ***Filter.seqs***).

2.2 Pré-clustering

Cette étape consiste à réduire la redondance des séquences, sont but et de nettoyer les données pour fusionner des séquences presque identiques en utilisant les deux commandes suivantes : ***Unique.seqs*** et ***Pre.cluster*** avec "diffs"= 2.

1. La commande ***unique.seqs*** réduit le nombre de séquences à analyser.
2. La commande ***pre.cluster*** utilise le fichier fasta output de ce dernier pour regrouper les séquences avec un nombre maximum (seuil) de différences de bases et débarrasser des éventuelles erreurs de PCR. Avec l'option *diffs*, nous pouvons modifier ce seuil.

2.3 Identification et suppression des chimères

Les chimères : sont des séquences d'ADN uniques issues de plusieurs séquences parentes, peuvent exister dans les lectures de séquençage, en raison de manipulations en laboratoire ou d'artefacts de PCR.

- Les chimères potentielles peuvent être :

1. Identifiées à l'aide de la commande. ***chimera.vsearch***
2. Supprimées à l'aide de la commande ***remove.seqs.***

3. Classification taxonomique

La classification taxonomique est fondée sur une base de données des séquences de référence phylogénétique (reference_tax.rdp) qui est déjà téléchargé et importer dans le même dossier. Dans le but de déterminer cette taxonomie des séquences en passant les filtres de contrôle de qualité du protocole de base avec reference_tax.rdp (base de données de référence phylogénétique). La variabilité des séquences dans un groupe taxonomique peut être classée en phylum, classe, ordre, famille, genre ou rang de l'espèce. La classification des espèces est difficile avec seulement une proportion du gène de l'ARNr 16s.

3.1 Élimination des séquences non bactériennes

Malgré tout ce que nous avons fait pour améliorer la qualité des données, il reste peut-être encore beaucoup à faire : il peut y avoir de gène 16S rRNA des archées, des chloroplastes et des mitochondries qui ont persisté à toutes les étapes de nettoyage jusqu'à présent. Nous ne sommes généralement pas intéressés par ces séquences et souhaitons les supprimer de notre jeu de données.

Le programme ***classify.seqs*** est utilisé pour prédire la classification taxonomique de chaque lecture ou contig.

- "fasta": the fasta output de ***Remove.seqs.***
- "taxonomy": reference_tax.rdp de Notre historique.
- "count": count table fishier output de ***Remove.seqs.***

Remove.seqs consiste à supprimer les séquences basées sur la nature du taxon non désiré. La classification des faux positifs peut être supprimée. Étant donné que notre étude cible l'ARNr 16s des eubactéries, nous supprimons toutes les classifications non bactériennes.

- "taxonomy": the taxonomy output de ***Classify.seqs***.
- "taxon-Sélectionner manuellement les taxons pour le filtrage": Chloroplast-Mitochondria-inconnu-Archaea-Eukaryota param-file "fasta": the fasta output from ***Remove.seqs.***
- param-file "count": the count table de ***Remove.seqs***

4. Résolution des abondances d'OTU et leur classification taxonomique

La classification OTU nécessite qu'une matrice de distance soit calculée entre les paires de séquences puis les séquences soient regroupées par distance.

a) La commande ***dist.seqs*** crée une matrice de distance et toutes les distances > 0,03 ne seront pas incluses dans la matrice.

b) La commande ***cluster*** détermine les OTU à des distances différentes. Un fichier partagé fournit l'abondance de chaque OTU par échantillon et est produit à l'aide de la commande ***make.shared***.

c) La commande ***classify.otu*** identifie la classification taxonomique consensuelle pour les OTU.

Pour comprendre la stabilité du microbiome en observant les changements dans la structure communautaire entre les échantillons en amont et en aval.

Étant donné que certains de nos échantillons peuvent contenir plus de séquences que d'autres, il est généralement judicieux de normaliser l'ensemble de données par sous-échantillonnage.

Nous voulons d'abord voire combien de séquences nous avons dans chaque échantillon. Nous allons le faire avec l'outil ***Count.groups***

- Output de fichier ***Count.groups*** est :

```
F54C.fastq.gz contains 38352.
F63CKD.fastq.gz contains 73835.
M53CKD.fastq.gz contains 59609.
M57C.fastq.gz contains 65267.

Total seqs: 237063.

Output File Names:
group.count.summary


mothur > quit
```

Figure 13 : Output du fichier *Count.groups*

- Le plus petit échantillon est F54C et se compose de 38352 séquences, c'est un nombre raisonnable. Nous allons donc procéder au sous-échantillonnage de tous les autres échantillons jusqu'à ce niveau.

***Sub.sample* :** utilisée pour normaliser nos données ou créer un ensemble plus petit à partir de notre ensemble d'origine.

- Output de fichier de ***Sub.sample*** est :

```
Sampling 38352 from each group.
0.03

Output File Names:
input_otu.0.03.subsample.dat


mothur > quit
```

Figure 14: Output du fichier de *Sub.sample*.

L’ensemble des groupes (sous-échantillons) obtenus est de 38352 séquences.

Les fichiers ***make.biom*** créent un fichier biom, ces fichiers Biom peuvent être importés dans des programmes externes pour effectuer des analyses supplémentaires et créer des tracés.

5. Analyse de la diversité

La diversité des espèces est un outil précieux pour décrire la complexité écologique d'un même échantillon (la diversité alpha) ou d'un échantillon à l'autre (la diversité beta). La diversité n'est pas une quantité physique qui peut être mesurée directement, plusieurs mesures différentes ont été proposées pour quantifier cette diversité.

La diversité des espèces comprend trois composantes :

- **La richesse des espèces** : le nombre d'espèces différentes dans une communauté.
- **La diversité taxonomique** : la régularité en nombre de chaque espèce dans une communauté.
- **Diversité phylogénétique** : le degré de parenté des espèces dans une communauté.

5.1. Diversité alpha :

Afin d'estimer la diversité alpha des échantillons, nous calculons les courbes de raréfaction avec le ***Rarefaction.single***. Rappelons que la raréfaction mesure le nombre d'OTU observées en fonction de la taille du sous-échantillonnage.

Rarefaction.single avec le paramètre suivant :

- "shared": the shared file de ***Make.shared***

Il existe autres options nombreuses disponibles sous le paramètre « calc » dans le pipeline mothur. Comme nous sommes intéressés par la richesse bactérienne dans notre échantillon, nous avons utilisé ici le paramètre **Sobs** (la richesse spécifique observée).

numsampled	0.03-F54C.fastq.gz	lci-F54C.fastq.gz	hci-F54C.fastq.gz	0.03-F63CKD.fastq.gz	lci-F63CKD.fastq.gz	hci-F63CKD.fastq.gz	0.03
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
100	59.2300	51.0000	68.0000	20.7860	15.0000	28.0000	
200	95.5490	84.0000	107.0000	33.1560	25.0000	42.0000	
300	125.1580	111.0000	139.0000	43.8960	34.0000	55.0000	
400	150.9200	136.0000	166.0000	53.9140	43.0000	66.0000	
500	174.5460	159.0000	192.0000	63.2870	51.0000	76.0000	
600	196.1530	178.0000	214.0000	72.6370	59.0000	86.0000	
700	216.4630	198.0000	236.0000	81.5100	67.0000	96.0000	
800	235.4630	216.0000	256.0000	90.1260	76.0000	107.0000	
900	253.6890	234.0000	274.0000	98.5990	82.0000	115.0000	
1000	271.0000	251.0000	292.0000	106.9430	90.0000	124.0000	

Figure 15 : Fichier affiche le nombre d'OTU identifiées par quantité de séquences utilisées (*numsampled*).

Hands-on : Plot Rarefaction

a) Dans cette étape on utilise cette commande : ***Plotting tool*** **-** *for multiple series and graph types.*

b) L'outil ***Summary.single*** pour générer un rapport de synthèse.

5.2. Diversité bêta :

La diversité bêta est une mesure de la similitude de l'appartenance et de la structure trouvée entre différents échantillons. Il existe plusieurs indices pour mesurer la similarité entre populations, nous en avons retenu deux : l'indice de Jaccard "jclass" qui ne prend en compte que le nombre d'espèces et les proportions d'espèces des espèces partagées, ainsi que le coefficient de similarité thêta de Yue & Clayton "thetayc" un indice de similarité qui comprend les proportions d'espèces des espèces partagées et non partagées dans chaque population.

Nous calculons cela avec l'outil : ***Dist.shared*** et ***Heatmap.sim.***

On passe maintenant au diagramme de Venn avec la commande Venn, on utilise les paramètres suivants :

- "OTU Shared": output de ***Sub.sample.***
- "groups": F63CKD, F54C, M57C, M53CKD

Cela génère un diagramme de Venn à 4 voies et un tableau répertoriant les OTU partagées. Ensuite, générons un dendrogramme pour décrire la similarité des échantillons entre eux. Nous allons générer un dendrogramme à l'aide des calculatrices ***jclass*** et ***thetayc***. Dans cette étape nous intéressons par ces deux commandes : ***Tree.shared*** et ***Newick display.***

6. Visualisation des résultats taxonomiques

Un outil que nous pouvons utiliser pour visualiser la composition de notre communauté est Krona avec la commande ***Taxonomy-to-Krona*** et ***Krona pie chart.***

➢ Chapitre 4 :

Résultats

et

Discussion

Chapitre 4 : Résultats et Discussion

1.Résultats

Rappelons que notre étude est basée sur l'utilisation des outils bioinformatiques pour établir la procédure opératoire standard (SOP) que le laboratoire Schloss de Chine a utilisé pour traiter leurs séquences de gènes d'ARNr 16s générées par la plate-forme *MiSeq* d'Illumina à l'aide de lectures appariées. Afin de comprendre et d'analyser les effets de la flore intestinale chez des patients CKD et des sujets sains.

1.1. Résultats de l'analyse de la diversité

1.1.1 Diversité alpha

On procède à la mesure du nombre d'OTU observées en fonction de la taille du sous-échantillonnage pour déterminer la courbe de raréfaction (figures 16 et 17).

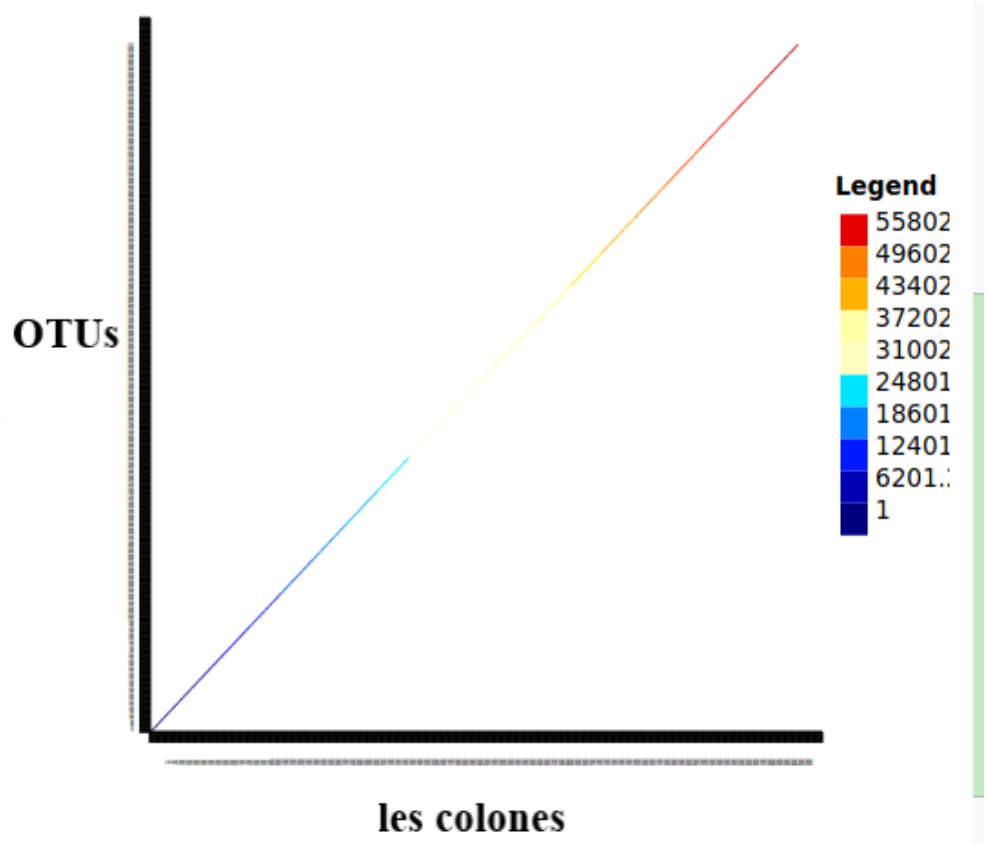

Figure16 : Courbe de raréfaction repère le nombre d'espèces en fonction du nombre d'individus échantillonnés.

numsampled	0.03-F54C.fastq.gz	lci-F54C.fastq.gz	hci-F54C.fastq.gz	0.03-F63CKD.fastq.gz	lci-F63CKD.fastq.gz	hci-F63CKD.fastq.gz	0.03
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
100	59.2300	51.0000	68.0000	20.7860	15.0000	28.0000	
200	95.5490	84.0000	107.0000	33.1560	25.0000	42.0000	
300	125.1580	111.0000	139.0000	43.8960	34.0000	55.0000	
400	150.9200	136.0000	166.0000	53.9140	43.0000	66.0000	
500	174.5460	159.0000	192.0000	63.2870	51.0000	76.0000	
600	196.1530	178.0000	214.0000	72.6370	59.0000	86.0000	
700	216.4630	198.0000	236.0000	81.5100	67.0000	96.0000	
800	235.4630	216.0000	256.0000	90.1260	76.0000	107.0000	
900	253.6890	234.0000	274.0000	98.5990	82.0000	115.0000	
1000	271.0000	251.0000	292.0000	106.9430	90.0000	124.0000	

Figure17 : fichier de sortie de la courbe de raréfaction (rapport de synthèse).

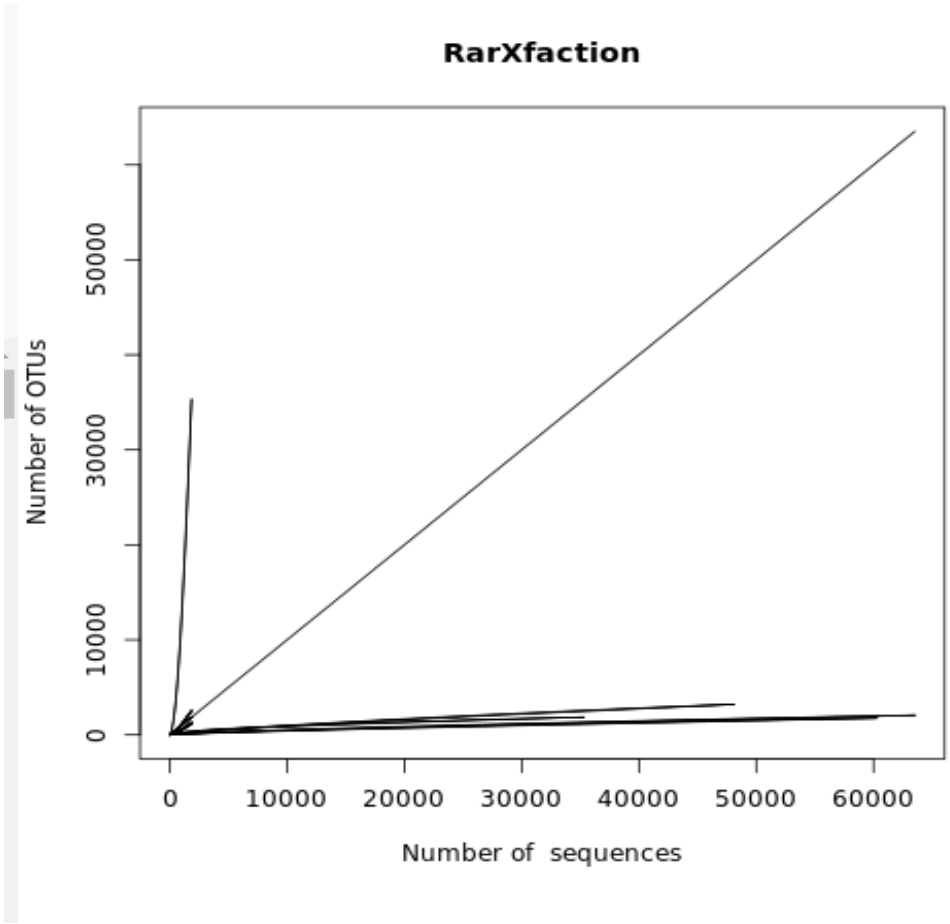

Figure18 : Courbes de raréfaction

Avec le pipeline mothur, nous avons pu analyser la richesse bactérienne spécifique observée dans notre échantillon (figures 18 et 19).

label	group	sobs	coverage	invsimpson	invsimpson_lci	invsimpson_hci	nseqs
0.03	F54C.fastq	1860.000000	0.968253	46.337153	45.302846	47.419793	35310.000000
0.03	F63CKD.fastq	2062.000000	0.974576	3.909395	3.873469	3.945993	63523.000000
0.03	M53CKD.fastq	3213.000000	0.948082	17.086109	16.828941	17.351258	48114.000000
0.03	M57C.fastq	1771.000000	0.976431	11.446414	11.281983	11.615710	60291.000000

Figure19 : Certaines métriques de diversité alpha.

1.1.2 Diversité bêta

Pour mesurer la similitude de l'appartenance et de la structure entre les différents échantillons, on a accordé une approche basée sur l'OTU, laquelle utilise l'indice de Jaccard "*jclass*" et le coefficient de similarité thêta de Yue & Clayton "*thetayc*" (figures 20 et 21).

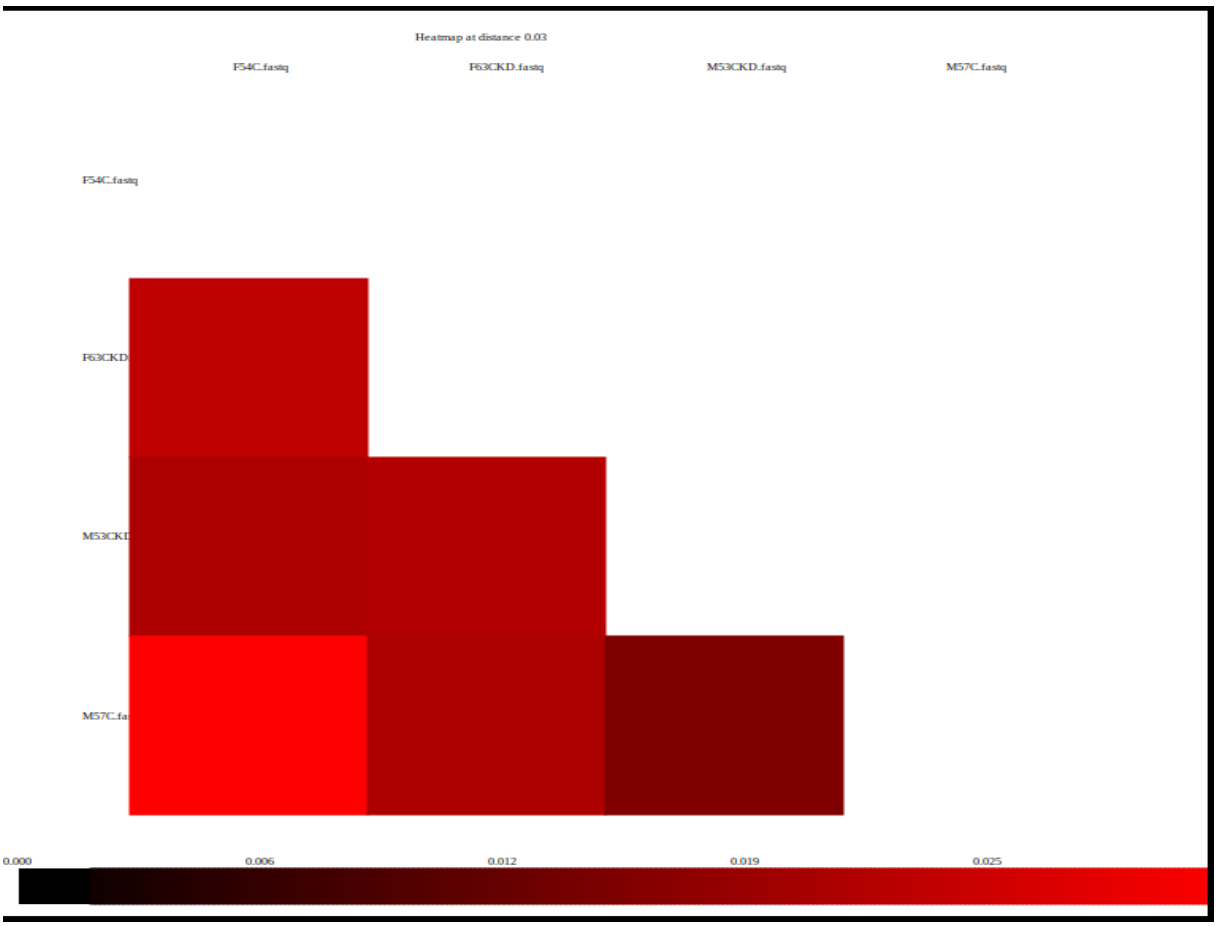

Figure20 : Carte thermique de l'indice de Jaccard ''*jclass*'' calculatrice.

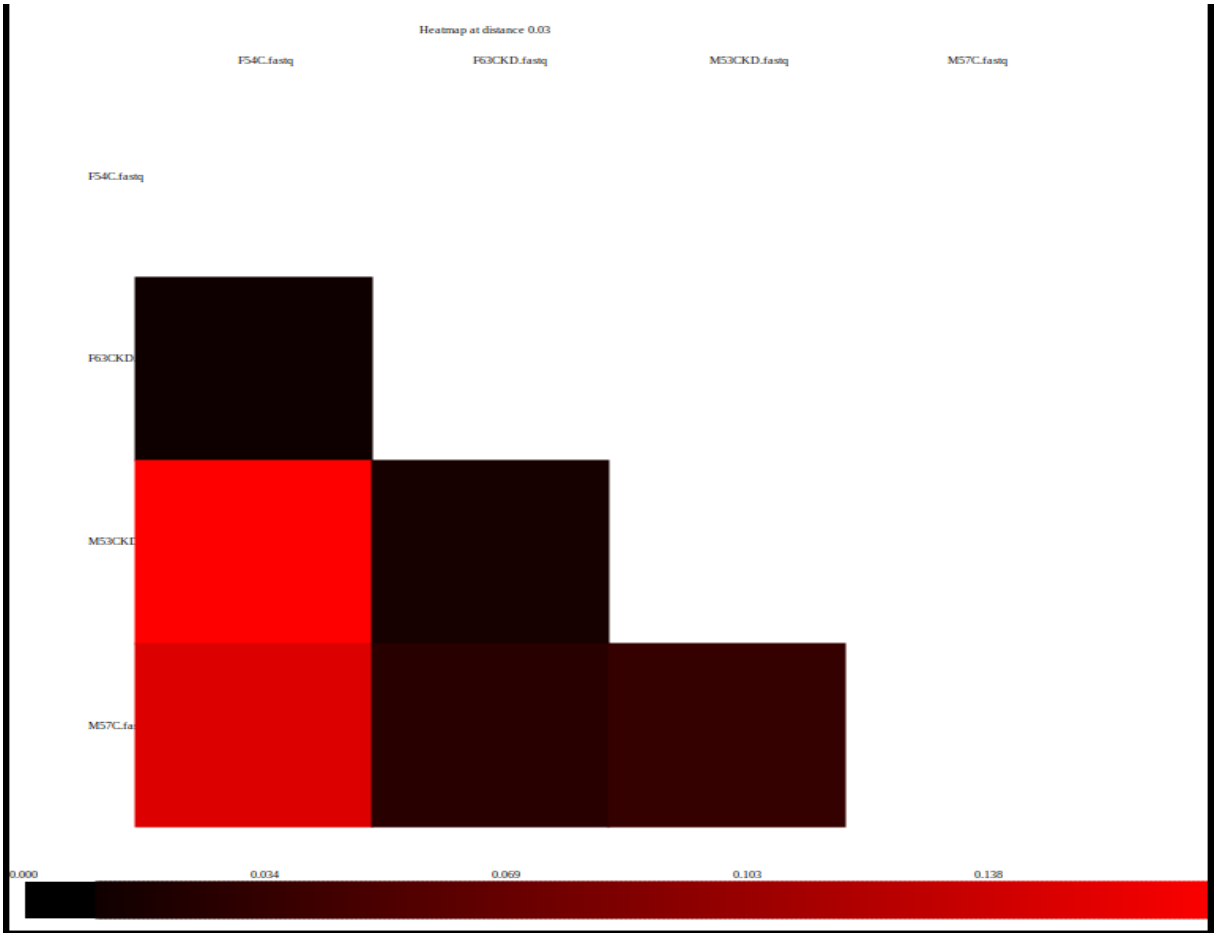

Figure 21: Carte thermique du coefficient de similarité thêta de Yue & Clayton ‘’*thetayc*’’ calculatrice.

Ces deux indices nous ont permis de mesurer la similarité entre nos populations, où l'indice de Jaccard "*jclass*" pris en compte le nombre d'espèces et les proportions des espèces partagées, ainsi que le coefficient de similarité thêta de Yue & Clayton "*thetayc*" comprend les proportions des espèces partagées et non partagées dans chaque population.

Toutefois, avec la commande *Venn*, on a généré un diagramme de Venn à 4 voies et un tableau répertoriant les OTU partagées (figure 22).

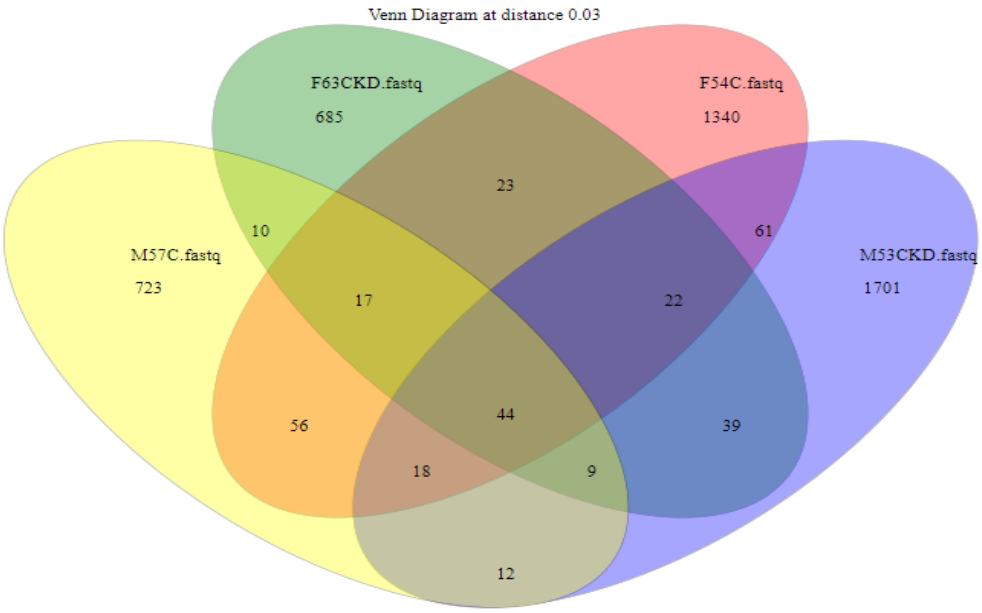

Figure 22: Diagramme de Venn à 4 groupes.

En utilisant l'indice de Jaccard "*jclass*" et le coefficient de similarité thêta de Yue & Clayton "*thetayc*", on a pu générer deux dendrogrammes pour décrire la similarité des échantillons entre eux (figures 23 et 24).

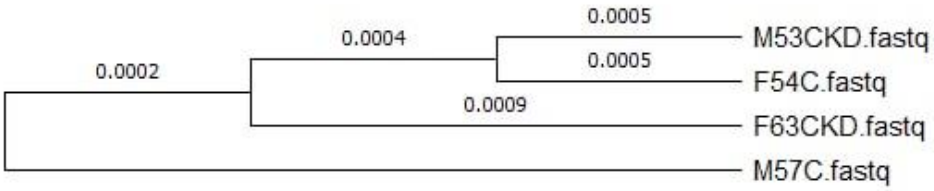

Figure23 : Dendrogramme de similarité des échantillons entre eux obtenu à l'aide de calculatrice jclass.

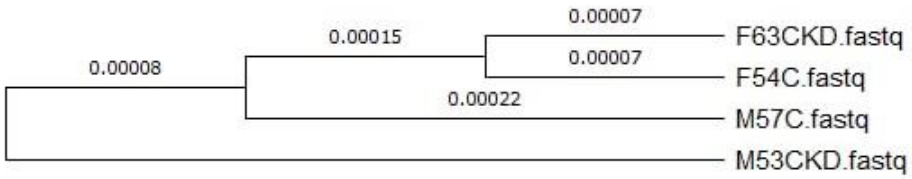

Figure24 : Dendrogramme de similarité des échantillons entre eux obtenu à l'aide de la calculatrice thetayc.

1.2. Classification phylogénique des quatre microbiotes intestinaux étudiés

Pour la visualisation de la composition des quatre communautés microbiennes étudiées et afin d'établir leur répartition phylogénétique on a procédé à l'utilisation des dendrogrammes Krona (figure 25).

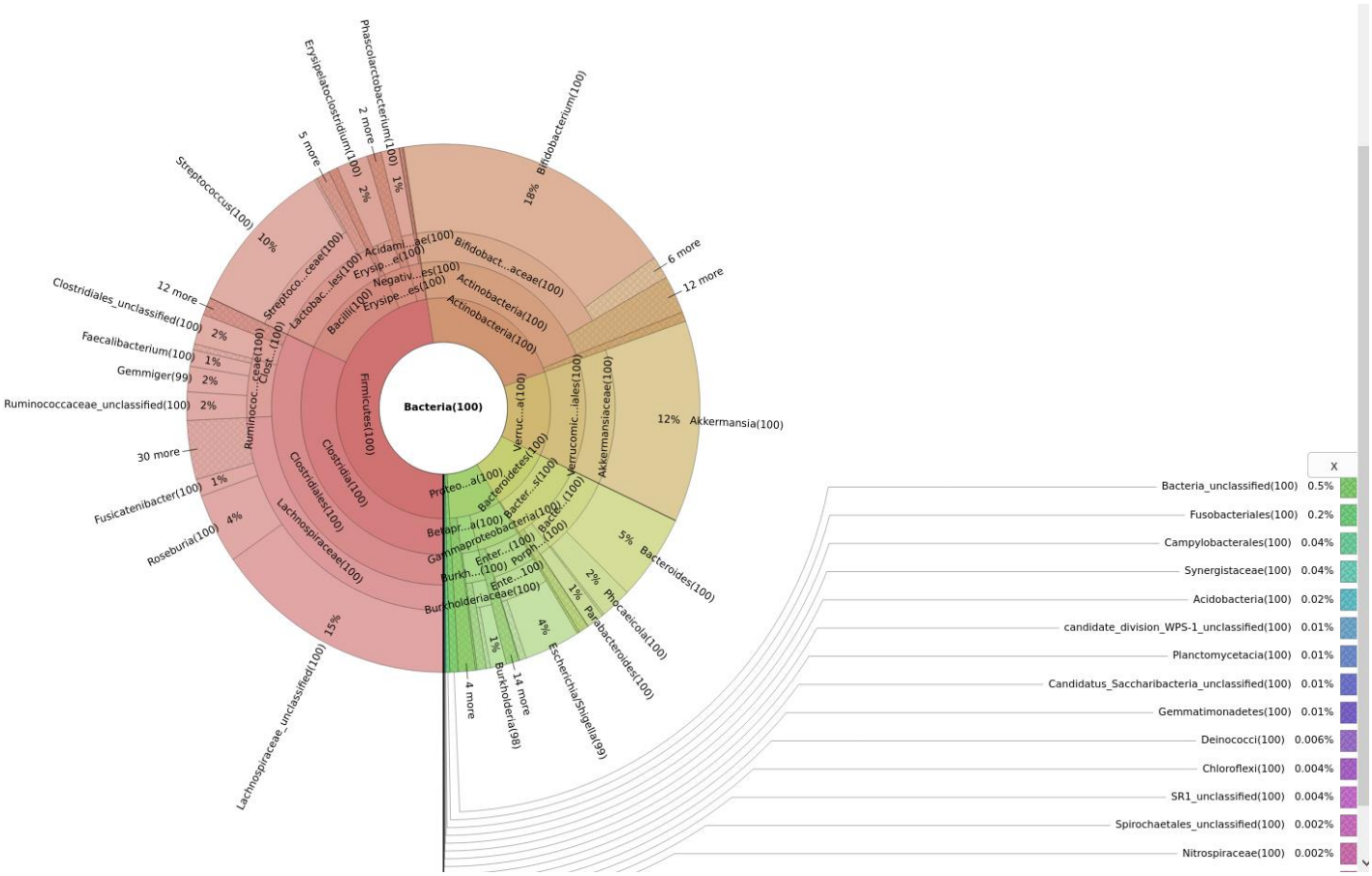

Figure 25 : Dendrogramme krona montre la distribution des bactéries dans la flore intestinale chez les quatre populations.

Cet outil nous a permis de faire un inventaire des espèces bactériennes qui peuplent les intestins de chacune des quatre populations étudiées.

Par exemple, la figure 26 montre la prépondérance des Actinobacteries chez les sujets mâles malades (M53CKD), tandis que la figure 27 nous montre l'abondance des Verrucomicrobies chez les sujets femelles malades F63CKD.

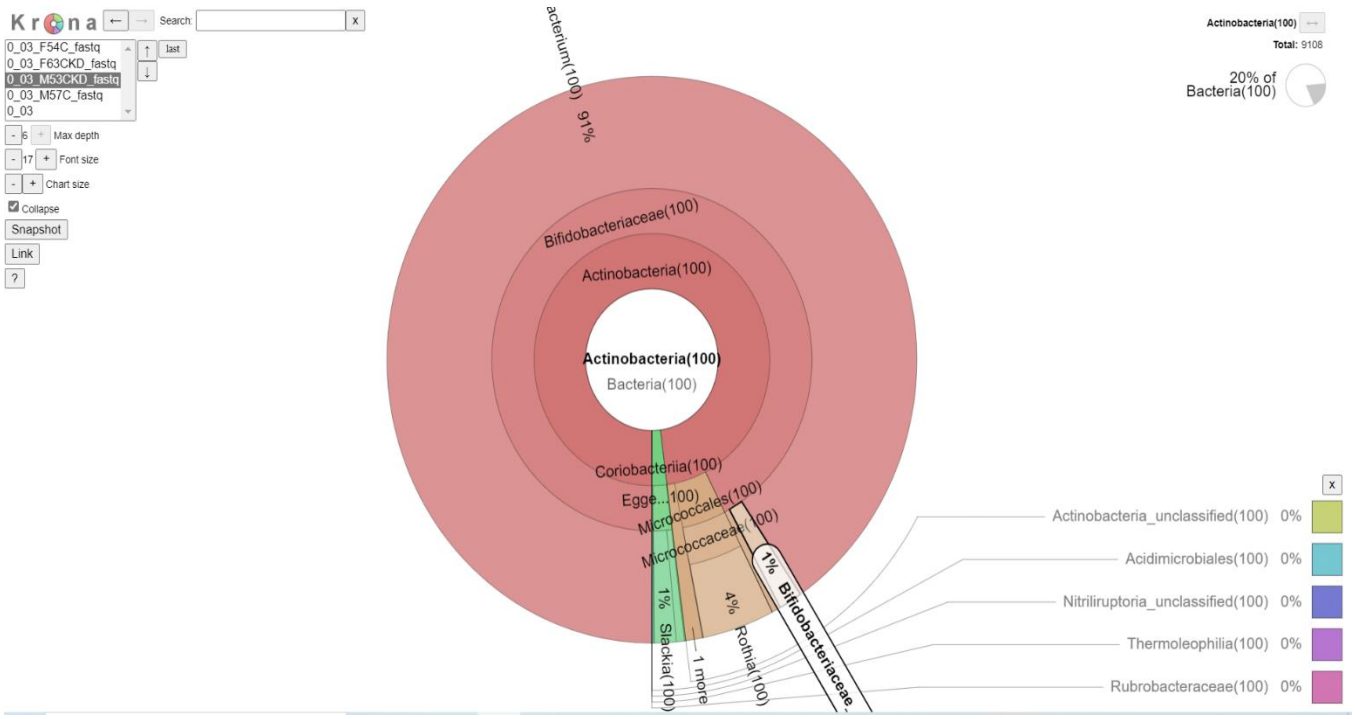

Figure 26 : Exemple des Actinobacteria chez les sujets mâles malades M53CKD.

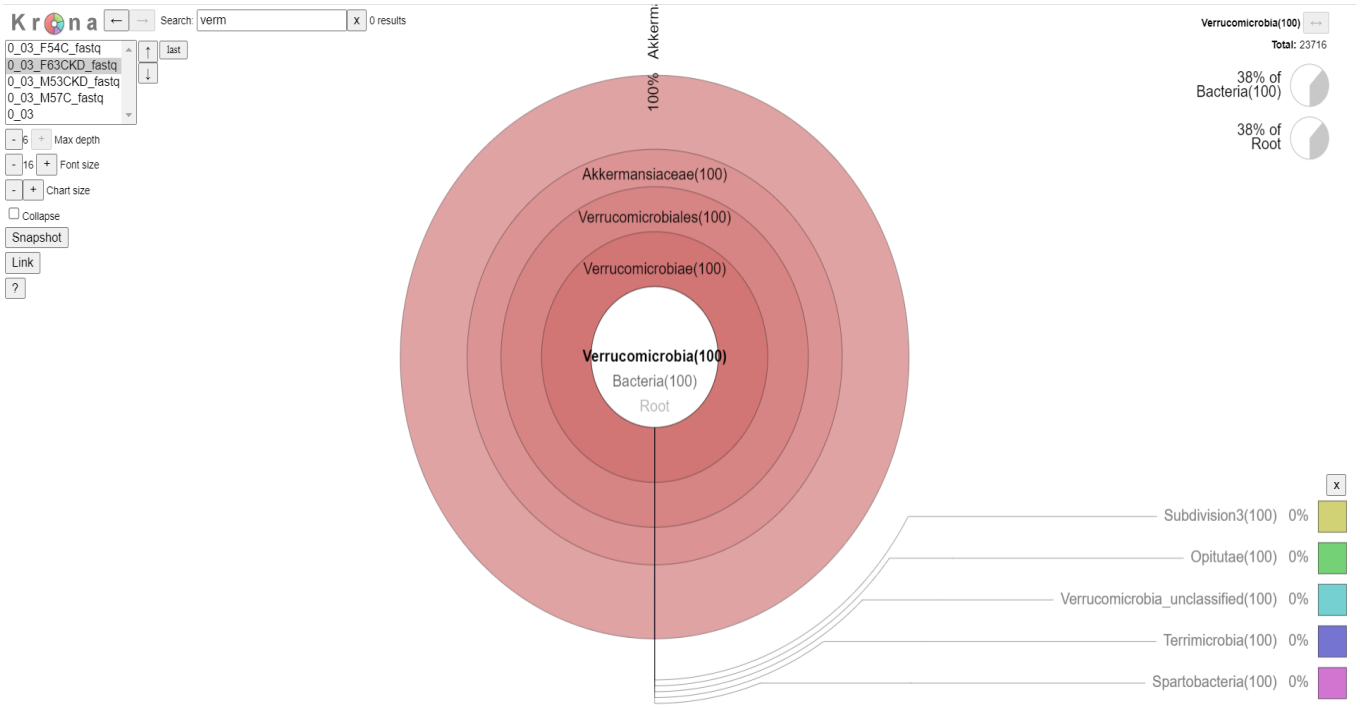

Figure 27 : Exemple des Verrucomicrobia chez les sujets femelles malades F63CKD.

2. Discusion de résultats

Bien que la définition quantitative de la diversité alpha ait été contestée, nous pouvons la décrire qualitativement comme une mesure de la complexité de la composition d'une communauté au sein d'un environnement, qui augmente avec le nombre d'espèces présentes et avec la régularité de leurs abondances relatives.

D'après la figure (16), qui montre une courbe de rarefaction trace le nombre d'espèces en fonction du nombre d'individus échantillonnés, la courbe commence par une pente cela signifie que notre habitat est riche en espèces, la pente et douce car seule une petite fraction a été échantillonnée. Cependant la figure (17), nous donne plus de détails sur un fichier qui affiche le nombre d'OTU identifiées par quantité de séquences utilisées (numsampled).

Le nombre d'OTUs supplémentaires identifiées lors de l'ajout des séquences supplémentaires atteignant un plateau, ainsi que la courbe de raréfaction commence à se stabiliser pour tous les échantillons, donc on est convaincus qu'une grande partie de la diversité des échantillons est couverte figures (18 et 19).

Le nombre d'espèces partagées entre les groupes F54C.fastq et M53CKD.fastq est de 145, et le nombre de séquences est de 50082, ce qui signifie qu'ils partagent le plus grand nombre d'espèces et que l'on peut observer dans l'arbre phylogénétique dans le même clade, car ils sont plus étroitement apparentés entre eux par rapport aux autres groupes '*sister taxa*'. (Figures 22 et 23).

Le nombre d'espèces partagées entre les groupes F63CKD.fastq et M57C.fastq est de 80, et le nombre de séquences est de 46172, ce qui est observé aux figures (22 et 23), car le groupe M53C constitut un sous-groupe.

Le nombre d'espèces dans le groupe F54C.fastq est de 1581 alors que le nombre de séquences est de 35536.

Le nombre d'espèces dans le groupe F63CKD.fastq est de 849 ainsi que le nombre de séquences est de 35536.

Le nombre d'espèces dans le groupe M53CKD.fastq est 1906 et le nombre de séquences est 35536.

Le nombre d'espèces dans le groupe M57C.fastq est de 889 et le nombre de séquences est de 35536.

Le nombre d'espèces partagées entre les groupes F54C.fastq et F63CKD.fastq est de 106 alors que le nombre de séquences est de 35319.

Le nombre d'espèces partagées entre les groupes F54C.fastq et M57C.fastq est de 135 or le nombre de séquences est de 57445.

Le nombre d'espèces partagées entre les groupes F63CKD.fastq et M53CKD.fastq est de 114 ainsi que le nombre de séquences est de 60843.

Le nombre d'espèces partagées entre les groupes M53CKD.fastq et M57C.fastq est de 83 en même temps que le nombre de séquences est de 54222.

Le nombre d'espèces partagées entre les groupes F54C.fastq, F63CKD.fastq et M53CKD.fastq est de 66 tandis que le nombre de séquences est de 58640.

Le nombre d'espèces partagées entre les groupes F54C.fastq, F63CKD.fastq et M57C.fastq est de 61 ainsi que le nombre de séquences est de 60407.

Le nombre d'espèces partagées entre les groupes F54C.fastq, M53CKD.fastq et M57C.fastq est de 62 alors que le nombre de séquences est de 70893.

Le nombre d'espèces partagées entre les groupes F63CKD.fastq, M53CKD.fastq et M57C.fastq est de 53 or le nombre de séquences est de 68050.

La richesse totale de tous les groupes est de 4760 espèces alors que le nombre de séquences dans les OTUs partagées par tous les groupes est de 80197 séquences. Cela montre qu'il y a eu un total de 4760 OTU observées entre les 4 groupes. Où seuls 44 de ces OTU étaient partagés par les quatre groupes au même temps. (Figure 22, diagramme de Venn).

Dans les quartes échantillons, il y’avait des Firmicutes, des Actinobacteries, des Bacteroides et des Proteobacteries avec des pourcentages différents, les résultats des krona ont montré une forte présence des bactéries nocives : Actinobactéries avec 20 % dans M53CKD par contre avec 3% chez M57C, une abondance inferieur des Bactéroides chez les CKD par rapport au C, la même chose pour les Firmicute qui sont moins présents dans les FCKD que les FC. Une présence des verrucomicrobie avec une distribution de 38% parmis les F63CKD alors que’elles sont absentes chez les C. Aucun changement dans les espèces de protéobactéries dans tous les groupes.

➢ Conclusion

Conclusion

Au terme de ce travail, notre étude nous a permis d'attribuer des séquences de gènes d'ARNr 16s du microbiote intestinal humain à des unités taxonomiques opérationnelles et décrire leur diversité au sein des quatre échantillons préalablement caractérisés par pyroséquençage de ces séquences. Diverses modifications de la composition du microbiote intestinal humain (ou *gut Microbiota*, gMB) ont été décrites chez les patients atteints la maladie d'insuffisance rénale chronique (IRC), (néphropathie chronique NPC) ou CKD (*Chronic Kidney Disease*).

La relation entre la CKD et le gMB est probablement bidirectionnelle, puisque les maladies rénales peuvent perturber un gMB équilibré et, à leur tour, les altérations du gMB pourraient affecter la progression de la maladie rénale et le degré des comorbidités associées.

Cependant, nous croyons que la principale conclusion de notre étude est la différence significative de la composition en gMB de Verrucomicrobia entre les patients F-CKD et F-C. Toutefois, nos résultats ne nous permettent pas de déterminer s'il existe une relation de cause à effet directe entre les changements de gMB entre les témoins et les patients atteints la CKD, ces résultats étant simplement associés par d'autres facteurs causaux courants (par exemple le sexe).

Notre étude a une limite conceptuelle concernant la petite taille de l'échantillon du groupe témoin disponible sur la plate-forme, et les facteurs de variation ce qui nous a empêchés d'effectuer une analyse des cas témoins. Cela pourrait biaiser les différences de gMB qui ont été observées entre les CKD et les groupes témoins sains et ouvre des perspectives plus élargies.

➢ Bibliographie

[1] Chappidi, S., Villa, E. C., & Cantarel, B. L. (2019). Using mothur todetermine bacterial community composition and structure in 16Sribosomal RNA datasets. Current Protocols in Bioinformatics, 67,e83. doi: 10.1002/cpbi.83

[2] Garmendia, L., Hernandez, A., Sanchez, M. B ., Martinez, J. L. (2012). Metagenomics and antibiotics. Clin Microbiol Infect, 18 (4):27–31

[3] Schloss,P.D., Westcott, S.L., Ryabin,T., Hall,J.R., Hartmann,M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G. G., Van Horn, D. J., Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 75 (23):7537–7541

[4] Déraspe, M. (2021).Génomique et métagénomique comparatives des bactéries. Open Access Theses and Dissertations [En ligne] (consulté le 18/06/2022)

[5] Hozzein ,W. N. Introductory Chapter: Metagenomics and Metagenomic Approaches . (2020). intechopen, 10 (5572): 87949

[6] L. Zhang et al., « Advances in Metagenomics and Its Application in Environmental Microorganisms », Front. Microbiol., vol. 12, p. 766364, déc. 2021, doi: 10.3389/fmicb.2021.766364.

[7] R. S. Mandal, S. Saha, et S. Das, « Metagenomic Surveys of Gut Microbiota », Genomics, Proteomics & Bioinformatics, vol. 13, no 3, p. 148-158, juin 2015, doi: 10.1016/j.gpb.2015.02.005.

[8] Caboche, S. (2018, 05 et 06 décembre). Cycle de formation NGS Module 5 : Métagénomique Partie 1 : Métagénomique ciblée [DIAPOSITIF]

[9] Xia LC, Cram JA, Chen T, Fuhrman JA, Sun F. Accurate genome relative abundance estimation based on shotgun metagenomic reads. PLoS ONE 2011;6:e27992.

[10] Garmendia L, Hernandez A, Sanchez MB, Martinez JL.Metagenomics and antibiotics. Clin Microbiol Infect 2012; 18:27–31.

[11] nazir Nazir, A. (2016). Review on Metagenomics and its Applications. Imperial Journal of Interdisciplinary Research (IJIR),2 (3): 2454-1362

[12] B. Gao et al., « An Introduction to Next Generation Sequencing Bioinformatic Analysis in Gut Microbiome Studies », Biomolecules, vol. 11, no 4, p. 530, avr. 2021, doi: 10.3390/biom11040530.

[13] S. Jünemann et al., « Bioinformatics for NGS-based metagenomics and the application to biogas research », Journal of Biotechnology, vol. 261, p. 10-23, nov. 2017, doi: 10.1016/j.jbiotec.2017.08.012.

[14] A. Shuikan, S. Ali Alharbi, D. Hussien M. Alkhalifah, et W. N. Hozzein, « High-Throughput Sequencing and Metagenomic Data Analysis », in Metagenomics - Basics, Methods and Applications, W. N. Hozzein, Éd. IntechOpen, 2020. doi: 10.5772/intechopen.89944.

[15] W. R. Streit et R. Daniel, Éd., Metagenomics: Methods and Protocols, vol. 1539. New York, NY: Springer New York, 2017. doi: 10.1007/978-1-4939-6691-2.

[16] A. P. Singh, « Genomic Techniques Used to Investigate the Human Gut Microbiota », in Biochemistry, vol. 20, N. V. Beloborodova et A. V. Grechko, Éd. IntechOpen, 2021. doi: 10.5772/intechopen.91808.

[17] A. Sánchez-Reyes et J. Luis Folch-Mallol, « Metagenomics-Based Phylogeny and Phylogenomic », in Metagenomics - Basics, Methods and Applications, W. N. Hozzein, Éd. IntechOpen, 2020. doi: 10.5772/intechopen.89492.

[18] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207-214. DOI: 10.1038/nature11234

[19] Schmidt TSB, Raes J, Bork P. The human gut microbiome: From association to modulation. Cell. 2018;172:1198-1215. DOI: 10.1016/j.cell.2018.02.044

[20] D. Haller, Éd., The Gut Microbiome in Health and Disease. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-90545-7.

[21] Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. sept 2012;489(7415):220-30

[22] Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. PLoS Biol.5, e177 (2007).

[23] Holmes, E., Li, J.V., Athanasiou, T., Ashrafian, H .,Nicholson, J.K. (2011).Understanding the role of gut microbiome–host metabolic signal disruption in health and disease. cell press, 19 (7): 0966-842

[24] Lagier, J.C., Dubourg,G., Million, M., Cadoret, F., Bilen, M., Fenollar, F., Levasseur, A., Rolain, J.M., Fournier, P.E ., Raoult, D.(2018).Culturing the human microbiota and culturomics. NATuRE REvIEWS | Microbiology , 16

[25] G. Trujillo-de Santiago, M. J. Lobo-Zegers, S. L. Montes-Fonseca, Y. S. Zhang, et M. M. Alvarez, « Gut-microbiota-on-a-chip: an enabling field for physiological research », Microphysiol Syst, vol. 1, p. 1-1, 2018, doi: 10.21037/mps.2018.09.01.

[26] T. Šuligoj et al., « Effects of Human Milk Oligosaccharides on the Adult Gut Microbiota and Barrier Function », Nutrients, vol. 12, no 9, p. 2808, sept. 2020, doi: 10.3390/nu12092808.

[27] Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, de Vos WM. Mucosaassociated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Applied and Environmental Microbiology. 2002;68(7):3401-3407

[28] Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005;308(5728):1635-1638

[29] Tannock , G. W. (2001).Molecular assessment of intestinal microflora. Am J Clin Nutr,73 (1): 410S–4S.

[30] Sedighi, M ., Razavi, S ., Navab-Moghadam, F ., Khamseh, M.E., Alaei-Shahmiri,F ., Mehrtash, A .,Amirmozafari, N . (2017).Comparison of gut microbiota in adult patients with type 2 diabetes and healthy individuals. Microbial Pathogenesis, 111: 362-369

[31] E. Margiotta *et al.*, « Gut microbiota composition and frailty in elderly patients with Chronic Kidney Disease », *PLoS ONE*, vol. 15, n° 4, p. e0228530, avr. 2020, doi: 10.1371/journal.pone.0228530.

[32] H. Lun *et al.*, « Altered gut microbiota and microbial biomarkers associated with chronic kidney disease », *MicrobiologyOpen*, vol. 8, n° 4, p. e00678, avr. 2019, doi: 10.1002/mbo3

Année universitaire : 2021-2022

Présenté par : MERBOUCHE Manel
BOULADJENIB Khouloud

Analyse métagénomique du microbiote intestinal des patients atteints la maladie d'insuffisance rénale chronique : traitement des données de pyroséquençage d'ARNr 16s.

Mémoire pour l'obtention du diplôme de Master en Bioinformatique

Résumé

Le microbiote intestinal, acteur clé de la santé, est considéré comme un dispositif à part entière de l'organisme humain. Le microbiote intestinal joue un rôle décisif dans notre santé. Il est extrêmement diversifié et varie d'un individu à l'autre. Dans l'objectifs d'étudier sa composition microbienne et de déterminer sa distribution, nous avons utilisé les données de pyroséquençage du gène ARN ribosomique 16s pour étudier les différences de microbiote intestinal entre quatre groupes de sujets atteints ou pas de la maladie d'insuffisance rénale (CKD).

Ce travail met le point sur l'utilisation du pipeline MOTHUR pour le traitement des données d'ARNr 16s. Cet outil nous a permis d'effectuer un prétraitement des séquences pour éliminer les erreurs, une analyse de l'unité taxonomique opérationnelle (OTUs), une description de la diversité des échantillons alpha et bêta, une taxonomie phylogénétique des OTU et la une visualisation de la diversité des échantillons à l'aide de dendrogramme Krona.

Mots-clés : Métagénomique, microbiote intestinal, MOTHUR et phylogénie.

Président : Pr. HAMIDECHI Mohamed Abdelhafid Prof. Univ. Frères Mentouri Constantine 1

Encadrant : : Dr. KELLOU Kamel MAA Univ. Frères Mentouri Constantine 1

Examinateur : Mr. CHEHILI H. MCA Univ. Frères Mentouri Constantine 1