

THÈSE DE MÉDECINE

CHU-BREST

UNIVERSITÉ DE BREST
BREST

Description du microbiote pulmonaire chez les patients atteints de mucovidoses

Auteur :

Sacha SCHUTZ

Responsable :

Geneviève HERY-ARNAUD

10 avril 2017



Engagement de non plagiat

Je, soussigné Sacha SCHUTZ, interne en biologie moléculaire au CHU de Brest, déclare être pleinement informé que le plagiat de documents ou de parties de documents publiés sur toute forme de support, y compris l'internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.

En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour la rédaction de ce document.

Date : 17/05/2017

Signature :

ssh pub key fingerprint : a4 :e3 :da :87 :78 :2d :e1 :6f :bb :56 :5c :d1 :72 :f5 :50 :63

License

Copyright (c) 2015 SCHUTZ Sacha. Permission est autorisée de copier, distribuer et/ou modifier ce document sous les termes de la Licence de Documentation Libre GNU, Version 1.2 ou toute version ultérieure publiée par la Free Software Foundation ; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. Une copie de la license est incluse dans la section intitulée "GNU Free Documentation License".



Table des matières

1	Avant-propos	2
2	Définition	4
3	Introduction	1
3.1	La mucoviscidose	1
3.1.1	Une maladie génétique	1
3.1.2	Une maladie infectieuse	2
3.2	Le microbiote pulmonaire	3
3.3	Exploration du microbiote pulmonaire	3
4	Materiel et Méthodes	4
4.1	Recueil des données	4
4.2	Extraction de l'ADN	5
4.3	Séquençage	5
4.4	Analyse bioinformatique	5
4.4.1	Pré-Traitement	6
4.4.2	Verification des qualités	6
4.4.3	Fusions des reads	6
4.4.4	Trimming des reads	6
4.4.5	Filtrage des qualités	6
4.4.6	Assignement taxonomique	7
5	Résultat	7
5.1	Pipeline Mucobiome	7
5.2	Diversité du microbiote respiratoire	8
5.2.1	Courbe de rarefaction	8
5.2.2	Diversité des échantillons	8
5.3	Evolution dans le temps	8
5.4	Comparaison entre catégories Free et Never	8
5.5	Sensibilité et Specificté du Pyo	8
6	Discussion	8
7	Conclusion	8

1 Avant-propos

Depuis Pasteur, les microbes ont toujours été associé aux maladies. Ils sont les agents nuisibles devant être éradiqué dans le fantasme d'un corps stérile comme signe de bonne santé. La mise en évidence des bactéries pathogènes allant de la syphilis jusqu'au grande Peste n'a pas aidé ces êtres microscopique à sortir de ce stéréotype. La médecine les ont donc naturellement choisi comme cible privilégiée en développant l'hygiène, les vaccins et les antibiotiques. Personne ne peut nier que cette triade a permis l'amélioration de notre santé en diminuant la prévalence des maladies infectieuses. Mais la disparitions de nos "vieux amis"[@] ayant co évolué avec nous depuis des milliers d'année, est associé, dans les pays industrialisé, à une recrudescence de nouvelle maladie[@] comme l'asthme, le diabète de type 1 ou encore la maldie de Crohn.[Théorie hyieniste]. Aujourd'hui, les nouvelles méthodes d'exploration de ce monde microscopique, comme le séquençage de l'ARN 16S, permet aux bactéries de redorer leurs blazons. Un role majeur leurs à d'abord été attribué en botanique par l'étude des sols. Ils jouent en effet le rôle principal dans le cycles de l'azote en permettant à la biomasse d'absorber l'azote athmosphérique. Par la chaine alimentaire, les bactéries sont la seul sources d'azotes permettant de consruire nos protéines et notre ADN. Sans le soleil il n'y aurait pas d'homme, sans les bactéries non plus. Les bactéries ont ensuite été retrouvé dans tous les environements, y compris les plus inhospitalier. Notemment les archéobactéries qui peuvent vivre dans des conditions d'acidité et de chaleurs exceptionnelles. C'est d'ailleurs leurs découverte qui nous a éclairé sur l'origine des eucaryotes[baspage] et qui nous a permis de faier un bon de géant en biologie moléculaire[baspage]. Chez l'homme sain, Les régions autrefois considéré stérile d'après la culture, floisonne maintenant de bactérie. En effet, la majorité des bactéries ne pouvant pousser en cultures, elles ont longtemps été indetectable. Elles sont retrouvé dans toutes les régions du corps exposés ou elles forment des communautés. La peau est colonisé par Propionibacterium, Corynebacterium et du Staphylococcus[@]. Le vagin contient les bacille de Döderlein et la bouche princiapелеment du Streptococcus[@]. L'intestin est une flore bactérienne dominé par les anéorobie et pouvant représenté jusqu'à 2 kg du poids corporel.[@] En échange de son hospitalité, notre microbiote participe au bon fonctionnement de son hôte. Il aide à la digestion en dégradant par exemple les sucres du lait maternelle ne pouvant être digéré par le nouveaux née[@ bifidus]. Il participe à la synthèse de vitamine essentiel (K, B12,B8)[ref]. Il éduque notre système immunitaire et fait barrières à tout nouvelle agent pathogène. Tout defaillance de ce nouvel organe ou dysbiose, peut s'accompagner d'un problème de santé. La liste des maladies associés à une dysbiose est longue. On retrouve la maladie de Crohn, la maladie coeliaque, le cancer de l'intestin, le syndrome du colon irritable, l'obésité, le diabète de type 1, l'asthme, l'eczéma, la sclérose en plaque, la polyarthrite rhumatoïde, la maladie d'alzheimer et même l'autisme. La colite à clostridium difficile

qui survient suite à un traitement antibiotique est une application directe des théories microbiotique. En effet, ce traitement détruit aussi bien les pathogènes que la flore intestinale. Un des traitements proposé est la transplantation fécale qui consiste à ré-introduire un microbiote au patient. [Notion de bactérie spécifique à notre génome. qqe famille seulement parmi une centaine.] Le microbiote amène donc à reconsidérer notre individualité. Nous ne sommes pas qu'un eucaryotes multicellulaire composé d'un unique génome. Mais un écosystème où cellules eucaryotes et microbiennes vivent en symbiose mutualiste. Les dernières études estiment que pour chaque individu il y a environ 30 billions de cellules humaines pour 39 billions de cellules microbiennes[1]. De plus, en associant les gènes bactériens, le génome d'un individu passe de 23 000 gènes à 3,3 millions[2] de gènes avec toutes la complexités des interactions que cela engendre.[3] Les scientifiques ont donné le nom d'holobionte à cette entité vivante hétérogène. Pour pousser le vis, l'ensemble du génome humain et microbiens est appelé hologénome. Ce dernier est pour certain la cible de la sélection naturelle. Il faut tout fois rester prudent quant au rôle donné au microbiote et éviter de tomber dans l'excès. Nombreux sont les publications scientifiques qui se contredisent et qui ne montre que des corrélations, sans rapport direct de causalité. Cette excès de publication à même conduit à la création d'un hashtag humoristique sur twitter :GutMicrobiomeAndRandomSomething[4] ou les gens publiait les corrélations les plus absurde. En effet, il n'y pas que l'hygiène qui a changer dans nos sociétés. D'autre facteur, comme notre mode de vie sédentaire ou notre alimentation, peuvent tout aussi bien être impliqué dans la survenue de maladie. Les études sur le microbiotes necessitent d'être réalisé à plus grande echelle avec plus de patient et avec un suivi à long terme important. Les nouvelles technologies de séquençage haut débit vont dans ce sens en permettant d'ammasser des quantités de donnée phénoménale limité seulement par les capacités de calculs. Il n'y a pas de honte à dire qu'aujourd'hui, nous ne connaissons pas grand chose au sujet du microbiote humain. C'est une science naissante et seul le futur nous dira si il s'agit d'un effet de mode ou d'une révolution. Pour ma part, au regard de la théorie de l'Evolution, je parierai dessus. Car ce sont bien des anciennes bactéries, qui permette à l'ensemble de nos cellules de respirer et que nous appelons maintenant mitochondries. [Pas de probiotic miracle] [PTetre qu'en regardant autre chose que le Gut, on trouvera un truc] Les mécanismes d'aide du microbiote :

Effet de barrière stimule le SI

2 Définition

Le microbiote est l'ensemble des micro-organismes (bactéries, levures, champignons, virus) vivant dans un environnement donné.

Le microbiome s'emploie selon deux définitions. En français, le microbiome est l'environnement qui héberge le microbiote. Dans sa définition angle-saxonne, le microbiome fait référence à l'ensemble des génomes microbien contenu dans un environnement. De façon général le microbiome est associé aux génomes bactériens. Les termes de Virome et de Mycobiome sont utilisé pour les génomes viraux et mycotiques.

La biocénose est le terme écologique dans un sens large désignant l'ensemble des organismes vivants dans un environnement appelé **Biotope**. Biocénose et biotope forme ensemble un **écosystème**.

Une symbiose est une association durable entre deux organismes. Leurs relations peuvent être mutualistes, parasitaire ou commensale.

La métagénomique est une méthode d'étude du contenu en ADN présent dans un milieu grâce à la technique de séquençage haut débit. Contrairement à la génomique qui s'intéresse au génome d'un individu, la métagénomique s'intéresse aux génomes d'une population d'individu. Dans son sens stricte, la méta-génomique correspond à l'étude de l'ensemble des séquences d'ADN. L'analyse d'un seul gène, comme celui de l'ARN 16s est associé à tort au terme métagénomique, mais son usage reste courant. On lui préférera le terme de **metagénétique**

Un read est un terme bioinformatique désignant une séquence d'ADN issue d'un séquençage haut débit. Selon les technologies, les reads varient entre 150 et 300 paire de bases.

Un OTU (*Operational taxonomic Unit*), est un terme utilisé en phylogénie, désignant un groupe d'individu proche et fait souvent référence à l'espèce dans la classification de Linée. En microbiologie, un OTU fait référence à un groupe d'individu ayant une similarité dans leurs séquence d'ARN 16s supérieur à 97%

L'Abondance absolu est le nombre de séquences d'ADN d'un OTU retrouvé dans un échantillon. L'abondance relative est le pourcentage en séquences d'ADN d'un OTU retrouvé dans un échantillon. Ce dernier indicateur permet de rendre les échantillons comparable entre eux.

La table des OTU correspond à un tableau à double entrée contenant l'abondance d'un OTU pour chaque échantillon. Dans le tableau suivant, l'échantillon 1 contient 68% de l'OTU 1.

	échantillon 1	échantillon 2	échantillon 3
OTU 1	68%	12%	25%
OTU 2	40%	24%	25%
OTU 3	28%	64%	50%

FIGURE 1 – La table des OTUs

La diversité alpha est une mesure de biodiversité au sein d'un échantillon. Elle correspond donc à l'étude d'une colonne dans la table des OTUs. Plusieurs indicateurs de diversité alpha.

La diversité beta est une analyse descriptive de la biodiversité entre plusieurs échantillons. Elle correspond à l'étude de l'ensemble de la table des OTUs. L'approche la plus courante est de réaliser une analyse multivariée par des méthodes d'ordination. Il s'agit de représenter un graphique à N dimensions, impossible à dessiner, en le projetant dans un espace à une ou deux dimensions.

La richesse est le nombre d'espèce présent dans un échantillon. Les deux échantillons suivant ont la même richesse de 2.

échantillon 1 : 4 Streptococcus , 4 Escherichia

échantillon 2 : 432 Streptococcus, 12 Escherichia

L'uniformité indique si les espèces d'un échantillon sont répartis uniformément.

L'uniformité du premier échantillon est plus grand que le second

échantillon 1 : 50 Streptococcus , 50 Escherichia

échantillon 2 : 432 Streptococcus, 12 Escherichia

L'indice Chao1 est une estimation de la richesse réel (in vivo) par rapport à la richesse observé (in vitro). Cette indice part du principe que si l'échantillon contient beaucoup de singletons (OTU détecté une seul fois), il est fort probable que la richesse réel soit plus grande que la richesse de l'échantillon. La formule est la suivante.

(1)

L'indice de Shannon est un indicateur évaluant à la fois la richesse et l'uniformité dans un échantillon. Il se calcul de la même façon que l'entropie de shannon.

(2)

L'indice de Shannon est un indicateur évaluant la probabilité que deux individus sélectionnées aléatoirement dans un échantillons donnée soient de la même espèces. La formule est la suivante.

La courbe de rarefaction est utilisé pour déterminer si la profondeur de séquen-

çage est suffisante pour caractériser la diversité d'un échantillon. Pour générer cette courbe, des groupes de reads de taille croissante ($1 \dots n$) sont tirés aléatoirement sans remise. Pour chaque groupe en abscisse le nombre d'OTU correspondant est reporté sur l'axe Y. Une courbe s'aplatissant indique qu'une profondeur de séquençage plus grande, n'apportera pas plus d'information.² and¹

3 Introduction

3.1 La mucoviscidose

3.1.1 Une maladie génétique

La mucoviscidose est une maladie génétique autosomique récessive grave touchant en France 1 naissance sur 5400 [1]. La Bretagne est la région la plus touchée avec une prévalence de 1/3000 [2]. La loi de Hardy Weinberg estime qu'en Bretagne 1 patient sur 25 est porteur de la mutation à l'état hétérozygote [3]. Cette haute prévalence s'explique probablement par un effet fondateur et d'un avantage sélectif pour les individus porteurs de l'allèle muté. (Plusieurs hypothèses ont été proposées, notamment lors des grandes épidémies de choléra en diminuant les pertes hydriques [4]. D'autre hypothèse suggère une amélioration du fitness chez les individus atteints de tuberculose. Et d'autres qu'il s'agit d'un exemple de pleiotropie antagoniste. [5]) Le gène CFTR impacté se situe sur le chromosome 7 en position [6]. Il est constitué de 27 exons sur 250,188 [7] paire de bases. Il code pour une canneaux chlore AMP dépendant permettant les échanges des ions chlorures au niveau des membranes cellulaires [8]. Il est également impliqué dans le transport du thiocyanate (SCN^-) et des bicarbonate (HCO_3^-) [9]. On dénombre à ce jour, 2017 mutations impliquées dans la mucoviscidose [10]. La perte d'une phénylalanine en position 508 par déletion du codon ($\Delta F508$ (new nom)) est responsable à elle seule de 80% des mutations sont responsables soit d'une protéine défectueuse ou d'une absence de canaux sur les membranes cellulaires. D'un point de vue clinique, la mutation est responsable chez les patients d'une insuffisance pancréatique exocrine et d'une infertilité par disparition des canaux déferents. Des signes digestifs, hépatiques et articulaires sont également retrouvés. Mais l'atteinte pulmonaire est la plus brillante. En effet au niveau de l'épithélium broncho-pulmonaire, l'absence d'un CFTR fonctionnelle est à l'origine d'une deshydratation du mucus le rendant plus visqueux et empêchant les cils bronchiques de jouer leur rôle. [11] La forte prévalence de la maladie nécessite de réaliser un dépistage précoce chez tous les nouveaux nés (test de Guthrie) afin d'adapter au plus tôt la prise en charge. Seul le test à la sueur permet de poser le diagnostic. Des tests de dépistage prénatal basés sur l'ADN circulant sont actuellement à l'étude. Il n'y a pas de traitement curateur à l'heure actuelle. Le traitement repose avant tout sur

une prise en charge respiratoire (kinésithérapie, dornase, antibiothérapie). Les thérapies génétiques sont encore à l'étude[has] L'Ivacaftor est le seul traitement à ce jour qui agit directement sur le CFTR. Mais concerne uniquement certaine mutation de classe 3 dont la plus courante G551D.[has] La greffe pulmonaire est le dernière recours.

3.1.2 Une maladie infectieuse

[Principe cause de morbidity / mortalité dans la muco [Nixon et al. 2001]] L'atteinte pulmonaire est caractérisé par des infections successive associé à une réaction inflammatoire qui dégrade progressivement la fonction respiratoire. Plusieurs pathogènes sont décrit. Chez les jeunes enfant, Haemophilus influenza et Staphylococcus aureus sont majoritairement retrouvé. Burkholderia cepacia et Stenotrophomonas maltophilia également retrouvé chez le sujet plus âgé. Mais c'est Pseudomonas Aeruginosa qui caractérise l'atteinte pulmonaire dans la mucoviscidose en marquer un tournant décisive dans l'évolution de la maladie. Ce bacille aérobie strictes, est un germe de l'environnement rarement retrouvé chez les patients sains. En revanche il est retrouvé chez 60% Puis survient le passage à la chronicité. Pseudomonas s'adapte à son milieu et s'installe à long terme. Il perd certaine caractère de virulence, mais devient résistant aux antibiotiques. Son phénotype change. Il devient mucoïde en sécrétant un film d'alginate qui le protège du système immunitaire. Les mécanismes sous jacent à cette transformation sont ingénieux. La forte densité en bactérie est responsable d'activation de certain gène amenant au phénotype mucoïde par un processus appelé quorum sensing. un processus dans lequel chaque bactéries communique avec ses voisins via des signaux. (On peut comparé ce processus à un système multiagent. comme un banc de poisson ou le comportement global dépend du comportement d'un individu). Les génomes de pseudomonas aeruginosa deviennent aussi hypermutable. Et par conséquence présente une plus grande diversité génétique au regard de la sélection naturelle. Pour un biologiste de l'évolution, il s'agit d'un cas d'évolvabilité. A ce stade le traitement antibiotique n'est plus curatif et l'évolution tend inexorablement vers un déclin de la fonction respiratoire. On ne sait pas aujourd'hui exactement pourquoi Pseudomonas aeruginosa s'installe préférentiellement chez les patients muco. Plusieurs hypothèse ont été posé.

La dysfonction ciliaire empêche les les Pseudo d'être viré L'hypersalinité du film muqueux désactive les peptides antimicrobiens Le CFTR est un récepteur de Pyo pour les internaliser et les virer L'inflammation de de l'épithélium augmente les métabolites qui permette de se développer. Alanine et lactate sont une source de carbone pour le Pyo. [89 Host microorganisme] D'un point de vue clinique, l'approche est préventif, visant à éliminer le pyo dès qu'il est détecté en culture. Une surveillance rapproché des patients muco avec un prélèvement mensuel ou bimensuel est préconisé. La culture étant peu sensible, d'autre approche peuvent être utilisé pour détecter le pyo . La détection

des anticorps anti-pyocianique par des méthodes élixa a montrer peu de ... || La PCR ciblé s'est montré plus sensible et spécifique que la culture. En pratique, la colonisation chronique est défini comme étant 3 résultat positifs successifs au cours d'un suivi mensuel ou bi mensuel. Une autre classification a été défini par Lee [10]. Groupe Free et Never.

3.2 Le microbiote pulmonaire

Bien qu'il soit en contact avec le milieu extérieur, l'arbre respiratoire (comprenant la trachée, les bronches et les alvéoles) a longtemps été considéré comme stérile avec les méthodes de culture classique. Il a fallu attendre l'avènement du séquençage haut débit pour les mettre en évidence. pulmonaire [ref Host-microorganism 1-3]. Le microbiote pulmonaire, est beaucoup moins abondante que la flore digestive. Il est constitué d'une flore dynamique provenant essentiellement de l'air ambiant mais aussi du tube digestif via des microaspirations. [10] Le microbiote pulmonaire est dominée par le phylum des Firmicutes (Streptococcus) et des Bacteroidetes (Prevotella). [11] Les genres retrouvés majoritairement sont Streptococcus, Prevotella, Fusobacteria, Veillonella, Haemophilus, Neisseria et Porphyromonas. l'arbre respiratoire étant en continuité directe avec les voies aériennes supérieures, certains genres bactériens sont communs, comme Streptococcus, staph, Haemophilus et Moraxella. Tandis que d'autres genres comme corynebacterium et Dolosigranulum ne sont retrouvés qu'au niveau des voies aériennes supérieures.

Le microbiote est variable dans l'espace et le temps. Dans l'espace, suivant les régions prélevées du poumon. [différente zone, foyer inflammatoire]. Le poumon présente des régions variables en Oxygène et en Ph. Et certaines bactéries sont anaérobies strictes, et pas d'autres ?

Plusieurs études suggèrent une différence de microbiote entre patients sains et avec une atteinte respiratoire chronique comme l'asthme, la BPCO et la mucoviscidose [Huang et al 2010].

Plusieurs études suggèrent que certains microbiotes sont associés à des pathologies comme l'asthme, la BPCO ou la mucoviscidose. Le mécanisme sous-jacent, théorie hygiéniste, stipule que les bactéries stimulent le système immunitaire.

propre à chaque individu variable dans le temps variable selon la localisation
résilience des populations résidentes permanentes et les attaques Liens avec les maladies :
– hygiéniste

3.3 Exploration du microbiote pulmonaire

(<http://bacterioweb.univ-fcomte.fr/bibliotheque/remic/08-Bronc.pdf>) ==> A LIRE
COURS BACTERIO Le microbiote pulmonaire est exploré par le séquençage des
ADNs bactériens présents dans un prélèvement broncho-pulmonaire. Toutes les méthodes

de recueils sont possible, mais les prélèvement protégé sont recommandé afin d'éviter une contamination par les voies supérieures. L'ADN est alors extrait ... 2 stratégies de séquençages peuvent être employées : La stratégie shotgun séquence l'ensemble du contenu en ADN présent dans l'échantillon après une fragmentation des ADNs. Les algorithmes bioinformatiques consistent ensuite à d'abord supprimer les ADNs humains. Puis à aligner ensemble les ADNs restant afin de reconstruire les génomes bactériens.

consiste à amplifier une région d'ADN assez discriminante pour identifier une espèce. Cette amplicon est ensuite séquençé. Les algorithmes bioinformatiques assignent alors à chaque reads son taxon approprié.

La première stratégie est plus complexe d'un point de vue calculatoire que la seconde mais est beaucoup plus informatif. En effet en raison des transferts génétiques horizontaux, la fonction du microbiote est plus liée aux gènes présents qu'aux espèces présentes. La stratégie 16s reste une bonne alternative pour décrire les populations bactériennes.

4 Matériel et Méthodes

4.1 Recueil des données

47 patients atteints de mucoviscidose ont été suivis sur 3 ans (2008-2011) dans une étude prospective multicentrique (Nantes, Brest, Roscoff) appelée mucobiome. La figure () montre la date des prélèvements. La CPP VI-Ouest et le comité d'éthique du CHRU de Brest ont approuvé le protocole. Tous les patients (ou les parents pour les mineurs) ont signé un consentement éclairé. Le protocole a fait l'objet d'une déclaration de biocollection à l'ARS et au MESR (n° DC-2008-214). Dans le suivi planifié des patients, les crachats lavés des patients ont été recueillis lors des séances de kinésithérapie respiratoire tous les 3 mois, suivant le calendrier des recommandations officielles. En pratique, sur l'ensemble de la cohorte suivie, l'intervalle médian entre 2 consultations a été de 3,4 mois. Les patients devaient avoir un génotypage CFTR et un test à la sueur positif. Les transplantés ont été exclus de l'étude. Une culture positive à *Pyo* était un critère de non-inclusion. Si pendant l'étude, une culture revenait positive à *Pyo*, le patient était sorti de l'étude pour être réinclus 1 an après en l'absence de colonisation chronique à *Pyo*. 15 patients ont été ainsi réinclus. Chaque patient était classé dans la catégorie Free ou Never (Lee et al 2003). D'autres données ont été également recueillies (tableau). Au total, 188 échantillons ont été recueillis, soit en moyenne 4 échantillons par patients. Pour chaque échantillon, une culture a été réalisée en suivant la procédure standard [ref]. Une qPCR ciblant le *Pyo* a également été réalisée en combinant les marqueurs *gyrB/ecfX* définis au laboratoire [ref].

4.2 Extraction de l'ADN

Les échantillons ont été liquéfié avec du Dithiotrétiol . Les protéines ont été dégradé avec une Protéine kinase. Les parois bactériennes ont été fragmenté par sonication. (DTTpar sonication (Elamsonic S10, Singen, Germany). Après 10 min de centrifugation, L'ADN a été extrait à partir du culot via QUIAamp DNA Minikit (Quagen). Les extraits d'ADN ont été envoyé pour séquençage via un prestataire GATC.

4.3 Séquençage

La séquençage d'une région de l'ARN 16S a été réalisé sur Illumina MiSeq. La librairie a été crée en amplifiant la région V3-V5 de l'ARN 16s à l'aide du couple d'amorces A B et du kit MiSeq Reagent Kits v3. Ce dernier permet de produire des reads païré chevauchant de 300 pb chacun. Environ 25 millions de reads sont produit par run. En multiplexant à l'aide de 94 index, 2 runs ont permis de séquencer les 188 échantillons. Au final 188 x 2 fichiers fastq ont été généré à l'issue du séquençage.

4.4 Analyse bioinformatique

L'analyse des 188x2 fichiers fastq a été réalisé grâce un pipeline bioinformatique appelé "mucobiome" conçu et tester dans le cadre de cette thèse. Par rapport aux autres logiciel comme QIIME ou MOTHUR, le pipeline mucobiome est spécialisé dans l'analyse des données 16s et dépend de très peu de dépendance applicatives. Mais il est surtout beaucoup plus rapide, ceci en raison d'un très haut niveau de parallélisation permis grâce un Snakemake. Snakemake est un programme informatique permettant de distribuer les différentes operations de calcul du pipeline sur plusieurs processeurs en même temps. Pour illustrer ces propos, imaginons qu'un pipeline est une recette de cuisine, composé de plusieurs étapes successif. Sans parallélisation, un cuisinier doit attendre que chaque étape se termine avant de passer à la suivante. Faire fondre le beurre, puis dans un second temps battre les oeufs en neige (Exécution synchrone). En parallélisant, le cuisinier peut faire plusieurs étapes en même temps. Battre les oeufs pendant que le beurre fondent. (Exécution asynchrone). Maintenant, si l'on demande à 40 cuisiniers (processeurs) de faire 188 gateaux, l'organisation des tâches devient complexe si l'on veut distribuer toutes les tâches de façon à maximiser les performances. C'est le problème que résout Snakemake en réalisant ce qu'on appelle un DAG (Direct Acyclic Graph) . Le pipeline mucobiome prend comme entré, l'ensemble des fichiers brut de séquençage pour produire un seul fichier unique contenant la table des OTU et la taxonomie . L'ensemble des étapes est défini ci-dessous .

4.4.1 Pré-Traitement

4.4.2 Verification des qualités

4.4.3 Fusions des reads

Les données bruts provenant d'un séquenceur sont des fichiers FastQ (annexe). Ces fichiers contiennent les séquences nucléotidique et les scores de qualités de l'ensemble des reads lu par le séquenceur. La stratégie de séquençage étant paired-end, pour chaque échantillon séquencé, deux fichiers fastq sont fourni. L'un correspond à la séquence lu dans le sens forward et l'autre lu dans le sens reverse. La première étape du pipeline consiste donc à fusionner ces deux fichiers . c'est à dire fusionner les reads deux à deux afin de produire un plus long reads de 500 pb en moyenne. Cette plus longue séquence correspond à la région V3-V5 de l'ARN 16S. 2 algorithmes ont été utilisé pour la fusion des reads, et sont mis à disposition de l'utilisateur.

Flash

VSearch

4.4.4 Trimming des reads

Afin d'assurer un alignement parfait, les séquences ne contenant pas à leurs extrémités les amorces V3-V4 ont été supprimées ou ajuster. L'algorithme utilisé est celui de cutadapts. Cette algorithmes reconnait avec un tolérance ajustable (0.1 par default) les amorces puis ajuste le reads en retirant les nucléotides en excès. Lorsqu'aucune séquences d'amorces n'est retrouvé, le read est supprimé .

4.4.5 Filtrage des qualités

Les données provenant de séquenceur haut débit peuvent contenir de nombreuses erreurs de séquençage comparé au méthode classique comme la méthode Sanger. Ceci est particulièrement vrai avec la stratégie MiSeq et le kit 300pb. Il est donc important de supprimer les reads de mauvaise qualité pour gagner en spécificité. Tout d'abord une analyse évaluant la qualité des reads à été réalisé sur chaque échantillon avec le logiciel FastQC. Ce programme produit à partir des fichiers fastq une série de graphiques permettant de juger sur la qualité du séquençage. Après cette analyse, le filtrage des reads de mauvaise qualité est exécuté en utilisant le programme sickle. L'algorithme sous jacent repose sur une fenêtre glissante de taille défini (par default : 20 pb) qui glisse tout le long de la séquence. A chaque étape la moyenne des scores de qualité est calculé dans cette fenêtre. Si successivement le score moyen dépasse un certain seuil, le reads est supprimé . Par default le seuil utilisé est de 20 avec une fenêtre glissante de x. Enfin FastQC a de nouveau été lancé afin de comparer la qualité des reads avant et

après filtrage.

4.4.6 Assignement taxonomique

L'assignement taxonomique consiste à labelliser chaque reads à son taxon. Pour cela deux stratégies existent. La stratégie "de novo" consiste à regrouper les reads qui se ressemblent en groupe ou cluster. Chaque cluster définit alors une seule séquence consensus qui est comparée à une base de données d'ARN 16S pour recevoir son assignation taxonomique. La stratégie "close reference" consiste à labéliser chaque reads en les comparant directement un par un à une base de données. Cette dernière stratégie peut sembler plus longue, mais elle est en réalité beaucoup plus rapide que la première stratégie. En effet la complexité de la stratégie "de novo" est de type N^2 (N le nombre de reads). Chaque reads étant comparé à l'ensemble des reads le temps de calcul augmente de façon exponentielle avec le nombre de reads. La complexité de la deuxième stratégie est elle de type N . C'est à dire que le temps de calcul augmente avec le nombre de reads de façon linéaire. En contrepartie, si un read n'est pas retrouvé dans la base de données, celui-ci est ignoré. Alors qu'en stratégie "de novo", tous les reads héritent du taxon de leurs clusters. Les bactéries de la flore humaines étant plus étudiées que d'autres flores plus exotiques, elles sont très souvent retrouvées dans les bases de données. La stratégie close reference est donc suffisante dans le cas de notre étude avec des tests préliminaires montrant une assignation réussie chez plus de 98

5 Résultat

5.1 Pipeline Mucobiome

Après demultiplexage, 188 x 2 fichiers fastq ont été générés soit 2 fichiers par échantillon. La taille des reads pour chaque fichier est de 301 paires de bases. Au total 115'002'297 reads ont été produits sur 2 runs MiSeq. Avec en moyenne 616900 reads par échantillon. Un minimum de 61422 reads pour l'échantillon 2154 et un maximum de 1071188 pour l'échantillon 3165. Les analyses de qualité avec FastQC montrent dans l'ensemble une baisse de qualité en fin de séquence. Les 50 dernières bases ont des scores de qualité médiocres entre 10 et 20 points selon le phred score. Après prétraitement des reads, c'est à dire merging Flash et filtering 20, en moyenne 49.24 L'assignement taxonomique a réussi sur 99.88. Au total, le pipeline mucobiome s'est exécuté en 1h29 sur 40 cœurs et 20 Go de mémoire contre 42h dans les tests précédents sans optimisation.

5.2 Diversité du microbiote respiratoire

5.2.1 Courbe de rarefaction

Les courbes de rarefaction par échantillons [figure] s'aplatissent précocement, témoignant d'un très bon niveau d'échantillonnage.

5.2.2 Diversité des échantillons

Au total, 54 genres bactériens sont retrouvés dans l'ensemble des échantillons. [figure]. Les trois phyla majoritaires, sont Protéobacteria (Haemophilus), Firmicutes (Streptococcus) et Bacteroidetes (Prevotella). [figure pie]. Certains genres bactériens sont très prévalents, c'est à dire présents dans l'ensemble des échantillons. Streptococcus, Neisseria, Prevotella, Granulicatella, Gemella, Veillonella et Fusobacterium sont présents dans plus de 185 échantillons. D'autres sont dominants, c'est à dire qu'ils représentent plus de 90%. Le core microbiota est défini comme l'ensemble des taxons retrouvés dans plus de 50. La figure [violin] montre que Neisseria et Streptococcus ont des abondances très variables dans les échantillons. Staphylococcus et Haemophilus ont dans la plus part des échantillons des abondances relativement faibles, mais la variance est tirée vers le haut par quelques échantillons où ils sont en dominance.

5.3 Evolution dans le temps

Alpha diversité (dot) Abondance relative (boxplot)

5.4 Comparaison entre catégories Free et Never

5.5 Sensibilité et Spécificité du Pyo

6 Discussion

7 Conclusion

Références

- [1] James M Beck, Vincent B Young, and Gary B Huffnagle. The Microbiome of the Lung.
- [2] Robert P Dickson, Fernando J Martinez, and Gary B Huffnagle. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet (London, England)*, 384(9944) :691–702, aug 2014.