

THÈSE DE MÉDECINE

CHU-BREST

UNIVERSITÉ DE BREST
DES DE BIOLOGIE MÉDICALE

Description du microbiote pulmonaire chez les patients atteints de mucoviscidoses

Auteur :

Sacha SCHUTZ

Responsable :

Geneviève HERY-ARNAUD

4 mai 2017



Engagement de non-plagiat

Je, soussigné Sacha SCHUTZ, interne en biologie moléculaire au CHRU 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

Licence

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 licence est incluse dans la section intitulée «“GNU Free Documentation License”.»



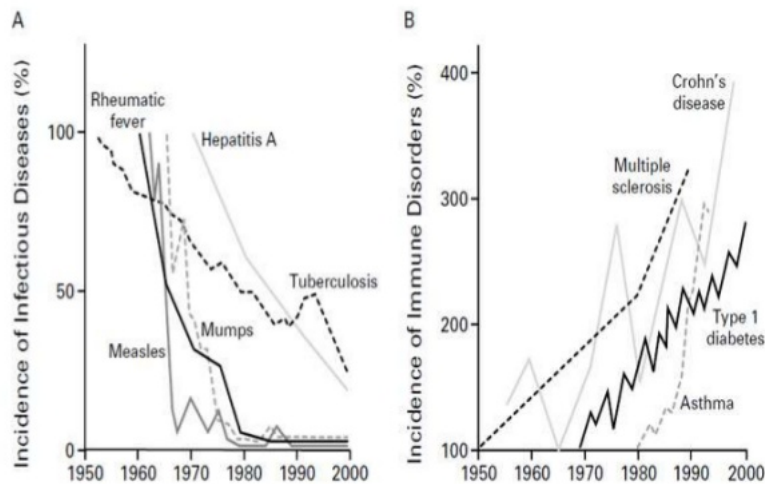
Table des matières

1	Avant-propos	3
2	Définition	6
2.1	Termes utilisés en écologie	6
2.2	Termes utilisés en bioinformatique	8
3	Introduction	1
3.1	La mucoviscidose	1
3.1.1	Une maladie génétique	1
3.1.2	Une maladie infectieuse et inflammatoire	2
3.2	Le microbiote pulmonaire	3
3.3	Exploration du microbiote pulmonaire	4
3.3.1	Méthode de prélèvement	4
3.3.2	Séquençage haut débit	5
3.4	Objectif de l'étude Mucobiome	6
4	Matériel et Méthodes	7
4.1	Recueil des données	7
4.2	Extraction de l'ADN	7
4.3	Séquençage	8
4.4	Analyse bio-informatique	8
4.5	Étape du pipeline	9
4.5.1	Merging : Fusions des reads	9
4.5.2	Cleaning : Filtrage des qualités	9
4.5.3	Reversing : Séquence complémentaire	10
4.5.4	Trimming : Suppression des primers	11
4.5.5	dereplicating : Suppression des doublons	11
4.5.6	Assignement taxonomique	12
4.5.7	Analyse de la table des OTUs	12
5	Résultat	13
5.1	Séquençage et pipeline	13
5.1.1	Données de séquençage	13
5.1.2	Qualité des reads	13
5.1.3	Assignement taxonomique	13
5.1.4	Profondeur d'échantillonnage	14
5.2	Résultats descriptifs	14
5.2.1	Composition du microbiote	14

5.2.2	Évolution de l'alpha diversité	15
5.2.3	Évolution des abondances	17
5.2.4	Beta Diversité	19
5.3	Résultats analytiques	20
5.3.1	Corrélation entre les genres bactériens	20
5.3.2	Comparaison entre échantillons Free et Never	20
6	Discussion EN COURS	23
6.1	Méthode de séquençage	23
6.2	Composition du microbiote pulmonaire dans la mucoviscidose	25
6.3	Limite de l'étude	25
7	Conclusion	26

1 Avant-propos

Depuis Pasteur, les bactéries ont toujours été associées aux maladies. La mise en évidence des agents pathogènes comme la syphilis ou la peste n'a pas aidé à tordre le cou de ce stéréotype. La médecine les a donc longtemps considéré comme l'ennemie à combattre plutôt qu'un allié. Aujourd'hui, personne ne peut nier que les traitements anti-infectieux ont permis l'amélioration générale de notre état de santé. Les avancées majeures dans les domaines de l'hygiène, la vaccination et les antibiotiques ont conduit à faire diminuer la prévalence des maladies infectieuses voire à les faire disparaître. Cependant, la destruction systématique et massive des micro-organismes qui font partie de nous depuis des milliers d'années pourrait bien être la cause de l'émergence de nouvelles maladies (Figure 1).



JEAN-FRANÇOIS BACH
New England Journal of Medicine September 2002

FIGURE 1 – Incidence des maladies infectieuses et auto-immunes en Europe au cours du temps¹.

Les méthodes récentes d'exploration de ce monde microscopique comme le séquençage haut débit ont permis aux bactéries de retrouver leurs lettres de noblesse. Elles sont présentes partout et jouent le premier rôle dans le fonctionnement des écosystèmes. Elles sont, par exemple, impliquées dans le cycle de l'azote et permettent à la biomasse d'absorber le diazote atmosphérique. Les bactéries sont ainsi la source primaire à laquelle les organismes puisent pour construire leurs protéines et leurs acides nucléiques. Elles peuvent survivre dans les milieux les plus inhospitaliers. Les archées (anciennement archéobactéries) peuvent résister dans des environnements d'acidité et de température exceptionnellement élevées. On les retrouve dans les fonds océaniques où, privées de

lumière, elles sont la seule source d'énergie pour la faune grâce à la chimiosynthèse à l'instar de la photosynthèse. Le corps humain est lui aussi un environnement riche en bactéries. La majorité des bactéries a longtemps été indétectable par les méthodes de culture classiques. Mais à présent, nous savons que les régions anatomiques autrefois considérées stériles contiennent beaucoup de bactéries. On les retrouve sur tous les territoires du corps exposés où elles se regroupent en communautés. La peau est colonisée par *Propionibacterium*, *Corynebacterium* et *Staphylococcus*²⁵. Le vagin est colonisé essentiellement par *Lactobacillus* et la bouche principalement par *Streptococcus*²⁵. La muqueuse intestinale est colonisée par une "flore" bactérienne ou microbiote dominée par les anaérobies pouvant contribuer jusqu'à 0.2 kg au poids corporel²³. En échange de son hospitalité, le microbiote contribue au bon fonctionnement de son hôte. Il aide à la digestion en dégradant par exemple les sucres du lait maternel chez le nouveau-né^{3;25}. Il participe à la synthèse des vitamines essentielles (K, B12, B9)^{14;25}. Il éduque notre système immunitaire et fait barrière à l'implantation de tout nouvel agent pathogène. Toute déséquilibre du microbiote intestinale ou *dysbiose*, peut avoir des conséquences pour notre santé. La liste des affections associées est longue²⁵. On y retrouve la maladie de Crohn, la maladie coeliaque, le cancer de l'intestin, le syndrome du côlon 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.

Par exemple, la destruction du microbiote intestinale par des antibiotiques donne l'opportunité à la souche toxigène de *Clostridium difficile* de s'installer et provoquer la colite pseudo-membraneuse. Un des traitements proposés après plusieurs échecs de traitement antibiotique est la transplantation fécale visant à régénérer le microbiote du patient.

Le microbiote nous amène donc à reconsidérer notre individualité. Nous ne sommes plus seulement un organisme multicellulaire composé d'un seul génome ; Mais plutôt un écosystème où cellules eucaryotes et micro-organismes vivent en symbiose. Cette relation n'étant pas figée dans le temps et pouvant varier entre le commensalisme, le parasitisme et le mutualisme. Des études récentes ont permis d'estimer pour un être humain le nombre de cellules microbiennes à 39 milliards comparé aux 30 milliards de cellules humaines²³. En associant les gènes bactériens, le génome d'un être humain passe de 23 000 à 3,3 millions de gènes¹⁸ avec toute la complexité des interactions que cela engendre. Les scientifiques ont nommé *holobionte* cet écosystème vivant. L'ensemble des génomes est appelé *hologénome*.

Il faut toutefois rester prudent quant au rôle donné aux microbiotes et éviter de tomber dans l'excès. Nombreuses sont les publications scientifiques qui se contredisent ou qui confondent corrélation et causalité. Ces publications ont d'ailleurs conduit à la création du hashtag humoristique sur Twitter : *#GutMicrobiomeAndRandomSomething*.

Les études sur le microbiote doivent être étayées par des études fonctionnelles afin de trouver des relations de cause à effet. Les corrélations doivent être réalisées sur des populations plus grandes avec un suivi dans le temps plus conséquent. Les nouvelles technologies de séquençage haut débit vont dans ce sens en collectant toujours plus de données.

Il est encore trop tôt pour dire si cette science va révolutionner la médecine ou s'il s'agit d'un effet de mode. Au regard de l'évolution biologique et de l'importance du microbiote chez d'autres espèces, les paris sont ouverts. Ne l'oublions pas, ce sont d'anciennes bactéries, que nous appelons aujourd'hui des mitochondries (Figure 2), qui permettent à l'ensemble de nos cellules de respirer.



FIGURE 2 – La mitochondrie est l'exemple de symbiose ultime entre eucaryote et procaryote

2 Définition

2.1 Termes utilisés en écologie

Le **microbiote**⁸ est l'ensemble des micro-organismes (bactéries, levures, champignons, virus) vivant dans un environnement donné.

Le **microbiome**⁸ est l'ensemble des génomes du microbiote. Le plus souvent le microbiome fait référence aux génomes bactériens seuls. On parle alors de virome et de mycobioime pour les virus et champignons respectivement.

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

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

L'**OTU** (*Operational Taxonomic Unit*) est un terme utilisé en phylogénie pour désigner un groupe d'individus semblable. Dans la majorité des cas, il s'agit de l'espèce dans la classification de Linné. En microbiologie, un OTU est défini par un groupe de bactéries ayant une similarité dans leurs séquences d'ADNr 16S supérieur à 97%.

L'**abondance** absolue est le nombre d'individus d'un même OTU dans un échantillon. L'abondance est indirectement évaluée en comptant le nombre de molécules d'ADNr 16S. L'abondance relative est l'expression en pourcentage de l'abondance absolue.

La **dominance** caractérise un OTU dont l'abondance est supérieur à 90% dans un échantillon.

La **table des OTU** est à la base de toutes analyses en écologie. Elle correspond à un tableau à double entrée contenant l'abondance par échantillon et par OTU. Dans le tableau 3, l'abondance relative pour *Staphylococcus aureus* est de 68%.

	échantillon 1	échantillon 2	échantillon 3
OTU 1(<i>S.aureus</i>)	68%	12%	25%
OTU 2(<i>E.coli</i>)	40%	24%	25%
OTU 3(<i>P.aeruginosa</i>)	28%	64%	50%

FIGURE 3 – La table des OTUs : L'abondance relative de 3 espèces bactériennes pour 3 échantillons

La **diversité alpha** est une mesure de la biodiversité au sein d'un échantillon. Elle correspond à une colonne de la table des OTUs. La richesse, l'indice de Chao1, de Shannon et de Simpson sont des indices de diversité alpha.

La diversité bêta est une analyse descriptive de la biodiversité entre **plusieurs** échantillons. Elle correspond à 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 deux ou trois dimensions.

La richesse (richness) est le nombre d'OTU présent dans un échantillon. Les deux échantillons suivant ont la même richesse (2) mais pas les mêmes abondances.

échantillon 1 : 4 *Streptococcus* , 4 *Escherichia*

échantillon 2 : 432 *Streptococcus*, 12 *Escherichia*

L'équitabilité (eveness) indique si les OTUs d'un échantillon sont réparties uniformément. L'équitabilité du premier échantillon est plus grande que le second

échantillon 1 : 50 *Streptococcus* , 50 *Escherichia*

échantillon 2 : 432 *Streptococcus*, 12 *Escherichia*

L'indice Chao1 est une estimation de la richesse réelle (*in vivo*) par rapport à la richesse observée dans l'échantillon(*in vitro*). Cet indice part du principe que si l'échantillon contient beaucoup de singletons (OTU détecté une seule fois), il est fort probable que la richesse réelle soit plus grande que la richesse de l'échantillon. La formule est la suivante :

$$Chao1 = S_{observed} + \frac{a^2}{2b} \quad (1)$$

avec **S** la richesse observée, **a** le nombre de singletons et **b** le nombre de doubletons

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

$$Shannon = \sum_{i=1}^S p_i \ln(p_i) \quad (2)$$

avec **p** la fréquence d'un OTU parmi les **S** OTUs présents dans l'échantillon

L'indice de Simpson est un indicateur évaluant la probabilité que deux individus sélectionnés aléatoirement dans un échantillon donné soient de la même espèce. L'indice est de sens contraire aux précédents. La formule est la suivante :

$$Simpson = 1 - \sum_{i=1}^S p_i \quad (3)$$

avec \mathbf{p} la fréquence d'un OTU parmi les \mathbf{S} OTUs présents dans l'échantillon

La courbe de raréfaction est utilisée pour déterminer si la profondeur de séquençage est suffisante pour caractériser la diversité d'un échantillon. Pour comprendre imaginons un sac noir contenant des billes (reads) de différentes couleurs (espèces). Avec deux couleurs, une petite poignée (un petit échantillon) de bille sera suffisante pour évaluer la diversité. Vous ne trouverez pas de nouvelle couleur avec plus de bille dans votre main. En revanche, avec d'avantage de couleurs dans le sac, il faudra une poignées plus importante pour évaluer la diversité. La courbe de raréfaction permet d'évaluer si cette poignée de bille est suffisamment représentative. Pour générer cette courbe, des groupes de reads de taille croissante (1..n) sont tirés aléatoirement depuis l'échantillon. Le groupe est reporté sur l'axe X et le nombre d'OTU retrouvé dans ce groupe est reporté sur l'axe Y. Une courbe s'aplatissant indique une bonne profondeur de séquençage (Figure 4).

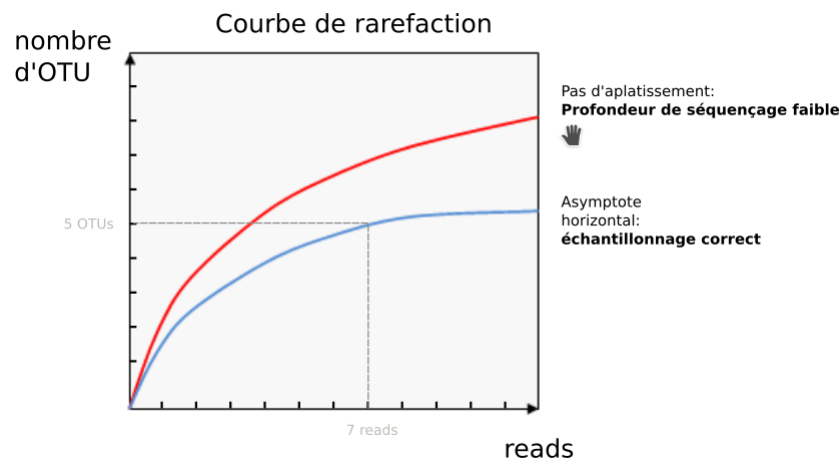


FIGURE 4 – Exemple de courbes de raréfactions. La courbe bleu témoigne d'un bon échantillonnage en s'aplatissant précocement. Plus de reads ne permettrait pas de découvrir de nouveau OTU contrairement à la courbe rouge.

2.2 Termes utilisés en bioinformatique

Pipeline Un pipeline est un ensemble d'étapes de calcul informatique. Chaque étape prend en entrée des fichiers pour en produire des nouveaux dans sa sortie.

On peut comparer cela aux étapes d'une recette de cuisine. Sans parallélisation, un cuisinier (le processeur) doit attendre de faire fondre le beurre avant de battre les œufs en neige (exécution synchrone). En parallélisant, le cuisinier peut réaliser plusieurs étapes en même temps. Battre les œufs pendant que le beurre fond (exécution asynchrone). Maintenant si l'objectif est de produire 188 gâteaux (188 analyses) et que l'on dispose de 64 cuisiniers (64 processeurs), l'organisation des tâches devient complexe si l'on veut maximiser le rendement. Pour cela, on dispose d'outils comme *Snakemake*¹¹, qui permettent de trouver la meilleure façon d'optimiser les tâches en construisant un Graphe orienté acyclique.

NGS (Next Generation Sequencing) désigne l'ensemble des séquenceurs de nouvelle génération permettant de séquencer un nombre très important de fragments d'ADN.

La métagénomique est une méthode d'étude du contenu en ADN présent dans un milieu grâce aux techniques 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'individus. Dans son sens strict, la métagénomique 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** ou **métataxonomique**.

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

Une librairie est l'ensemble des fragments d'ADN préparés pour être séquencé.

Score de qualité Phred ou QScore exprime la confiance que l'on porte au séquençage. Il est lié de façon logarithme à la probabilité d'erreur d'identification d'un nucléotide. Par exemple un score de 10 équivaut à une 1 erreur sur 100. Un score de 40 à 1 erreur sur 10000.

$$Q = -10 \log_{10} P \quad (4)$$

Le score Q est associé de façon logarithmique à la probabilité d'erreur P de s'être trompé en séquençant un nucléotide

Une ordination est une analyse multivariée visant à décrire des objets caractérisés par plusieurs variables. Par exemple un échantillon est caractérisé par différentes abondances de bactéries. Il existe plusieurs méthode d'ordination. La plus connus étant l'analyse en composante principale (PCA) basé sur une matrice de covariances. En écologie l'analyse en Coordonnée principale (PCoA) est préférée.

Elle se base sur une matrice de distances. Elle consiste à projeter des points d'un espace à n dimensions vers un espace visible en 2 dimensions. La figure 5 illustre ce principe avec un exemple à deux dimensions seulement.

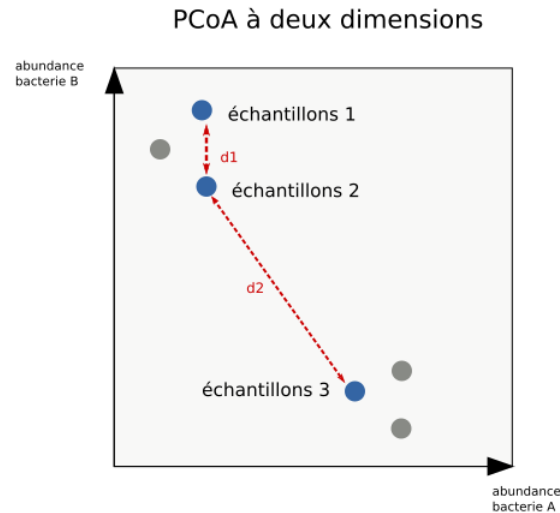


FIGURE 5 – Exemple d'une PCoA à deux dimensions. Chaque échantillon est assimilable à un point de coordonnées (abondance A ; abondance B). Il est alors possible de calculer des distances entre les points. La plus connue étant la distance euclidienne. Mais d'autres existent comme la distance de Bray-Curtis. Sur cette figure, l'échantillon 1 et 2 sont des microbiotes proches car la distance $d1$ qui les sépare est petite.

La distance de Bray-Curtis est un indice de dissimilarité entre deux échantillons qui s'assimile à une distance entre 2 points dans un espace de taille n (Figure 5). Plus la distance est grande et plus les deux échantillons sont dissemblable. La formule est la suivante :

$$BC_{jk} = 1 - \frac{2 \sum_{i=1}^p \min(N_{ij}, N_{ik})}{\sum_{i=1}^p (N_{ij} + N_{ik})} \quad (5)$$

Où N_{ij} est l'abondance d'une espèce i dans l'échantillon j et N_{ik} l'abondance de la même espèce i dans l'échantillon k . Le terme $\min(.,.)$ correspond au minimum obtenu pour deux comptes sur les mêmes échantillons. Les sommes situés au numérateurs et dénominateur sont réalisées sur l'ensemble des espèces présentes dans les échantillons.

Un fichier **fastq** contient les séquences des reads et leurs qualités par nucléotides exprimée en Phred Score.



FIGURE 6 – Exemple d'un read contenu dans un fichier fastq. Les scores de qualités sont encodés avec des symboles ASCII. Par exemple le quatorzième T est associé au symbole « : » correspondant au score phred de 25, soit une probabilité de 0.003% d'erreur.

3 Introduction

3.1 La mucoviscidose

3.1.1 Une maladie génétique

La mucoviscidose est une maladie génétique autosomique récessive grave qui frappe en France 1 naissance sur 5400²⁰. La Bretagne est la région la plus touchée avec une prévalence de 1/3000²⁰. La loi de Hardy Weinberg estime qu'en Bretagne 1 patient sur 25 est porteur de la mutation à l'état hétérozygote. Cette haute prévalence s'explique probablement par un effet fondateur associé à un avantage sélectif pour les individus porteurs de l'allèle muté.ⁱ Le gène *CFTR* impacté se situe sur le chromosome 7 en position q31.2. Il est constitué de 27 exons pour 250 188¹⁶ paires de bases. Il code pour un canal chlore AMP dépendant permettant les échanges des ions chlorures au niveau des membranes cellulaires. Il est également impliqué dans le transport du thiocynate (SCN-) et des bicarbonates (HCO₃-)¹⁹. On dénombre à ce jour 2017 mutations⁹ mises en cause dans la mucoviscidose. La perte d'une phénylalanine en position 508 par délétion du est triplet c.1521-1523delCTT (anciennement $\Delta F508$) cause à elle seule 80% des mucoviscidoses⁹. Ces mutations sont responsables d'une protéine défectueuse ou d'une absence de canaux sur les membranes cellulaires.

Cliniquement, la mutation entraîne une insuffisance pancréatique exocrine et une infertilité par disparition des canaux déférents. Des signes digestifs, hépatiques et articulaires sont également retrouvés. L'atteinte de la fonction respiratoire est la plus bruyante. En effet au niveau de l'épithélium broncho-pulmonaire, l'absence d'un CFTR fonctionnel est responsable d'une déshydratation du mucus le rendant plus visqueux et empêche les cils bronchiques de jouer leurs rôles.

La forte prévalence de la maladie nécessite de réaliser un dépistage précoce chez tous les nouveaux nés (test de Gutri) afin d'adapter au plus tôt la prise en charge. Seul le test à la sueur permet de poser le diagnostic. Le dépistage prénatal basé sur l'ADN circulant est actuellement à l'étude.

Le traitement repose avant tout sur une prise en charge respiratoire (kinésithérapie, doraseⁱⁱ, antibiothérapie). Les recherches en thérapies génétiques sont encourageantes [4]. Les traitements correcteurs/potentiateurs de la protéine CFTR, comme l'*ivacaftor* sont les seuls traitements curateurs. Mais concerne uniquement certaine mutation rare comme la G551D. La greffe pulmonaire est le dernier recours.

i. Plusieurs hypothèses ont été proposées, notamment lors des grandes épidémies de choléra en diminuant les pertes hydriques.]

ii. Est une DNase qui fluidifie les sécrétions trop riche en ADN provenant de l'afflux des neutrophiles in situ et de leur apoptose prématurée

3.1.2 Une maladie infectieuse et inflammatoire

L'atteinte pulmonaire dans la mucoviscidose est caractérisée par des réactions inflammatoires qui dégradent progressivement la fonction respiratoire. Plusieurs pathogènes sont impliqués. Chez les jeunes enfants⁵, *Haemophilus influenza* et *Staphylococcus aureus* sont le plus souvent responsable. *Burkholderia cepacia complex* et *Stenotrophomonas maltophilia* sont retrouvés parmi les sujets plus âgés⁵. Mais c'est *Pseudomonas aeruginosa* qui caractérise l'atteinte pulmonaire dans la mucoviscidose en marquant un tournant décisif dans l'évolution de la maladie. Ce bacille aérobie stricte est une bactérie de l'environnement rarement retrouvé parmi les patients sains[@]. En revanche, dans la mucoviscidose, il est mis en évidence chez 60%¹² des patients jeunes, et plus de 90% des patients adultes¹².

La primocolonisation à *Pseudomonas aeruginosa* est difficilement détectable, mais semble avoir lieu tôt dans l'enfance[@]. Il y a ensuite une phase de latence, variable entre les individus, marquée par des épisodes d'exacerbations et de rémissions. À ce moment, l'éradication par des antibiotiques reste possible. Puis survient le passage à la chronicité. *Pseudomonas aeruginosa* s'adapte à son milieu et s'installe à long terme. Il perd certains caractères de virulence, mais devient résistant aux antibiotiques¹². Son phénotype change. Il se transforme pour devenir mucoïde en sécrétant un film d'alginate qui le protège du système immunitaire. Les mécanismes sous-jacents à cette adaptation sont ingénieux. La forte densité en bactérie est responsable d'activation de certains gènes par un processus appelé *quorum sensing*²¹. Un processus dans lequel chaque bactérie communique avec ses voisines par des signaux chimiques. Le génome de *Pseudomonas aeruginosa* devient aussi hyperpermutable⁵ afin de présenter une plus grande diversité génétique au regard de la sélection naturelleⁱⁱⁱ.

À ce stade, le traitement antibiotique n'est plus curatif et l'évolution tend inexorablement vers un déclin de la fonction respiratoire.

L'approche clinique est donc préventive afin de retarder le passage à la chronicité. Elle vise à éliminer *Pseudomonas aeruginosa* dès qu'il est détecté en culture. Une surveillance rapprochée des patients avec un prélèvement mensuel ou bimensuel est préconisée selon l'HAS[@]. La culture étant peu sensible, d'autres méthodes d'identification peuvent être employées. La détection des anticorps anti-pyocyaniques par ELISA a montré peu de sensibilité¹⁷. La PCR ciblée associant les protéines bactériennes *OPRL1*, *GYRB1* et *ECFX1* s'est montrée plus sensible et plus spécifique que la culture^{24;13}

En pratique, la colonisation chronique est définie lorsque 3 expectorations sont rendues positives en culture, successivement au cours d'un suivi mensuel ou bimensuel¹². Une autre classification, celle de Lee¹⁵ a montré une forte liaison clinico-biologique. Elle est composée de 4 groupes :

iii. En biologie évolutive, il s'agit d'évolvabilité

groupe chronique $> 50\%$ des cultures sont positives sur 12 mois

groupe intermédiaire $\leq 50\%$ des cultures sont positives sur 12 mois

groupe Free Toute culture négative sur 12 mois, avec des antécédents

groupe Nevers Toute culture négative sur 12 mois sans antécédents

On ne sait pas aujourd'hui pourquoi *Pseudomonas aeruginosa* s'installe préférentiellement chez les patients atteints de mucoviscidose. Plusieurs hypothèses ont été proposées :

- La dysfonction ciliaire empêche les pseudo d'être viré par le haut
- L'hypersalinité du film muqueux désactive les peptides antimicrobiens
- Le CFTR est un récepteur de pyo qui les internalise et les viré
- L'inflammation de l'épithélium augmente les métabolites qui permettent de se développer.
- Alanine et l'acta sont une source de carbone pour le pyo.
- Le microbiote influence la colonisation
- Le pyo stimule le Système immunitaire pour virer tous les autres concurrents

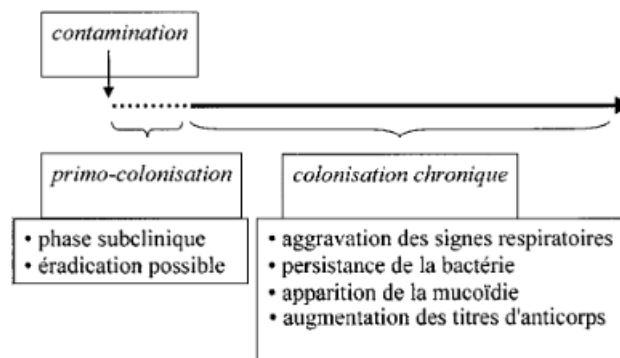


FIGURE 7 – Infection pulmonaire à pyo dans la mucoviscidose. ref []

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 stérile avec les méthodes de culture classique. Il a fallu attendre l'avènement du séquençage haut débit pour mettre en évidence le microbiote pulmonaire^{10;2;7}.

L'arbre respiratoire chez le fœtus^{iv} est stérile. Il se colonise comme l'intestin, à la naissance en traversant la filière génitale puis se différencie avec le temps. Il est beaucoup moins riche que le microbiote digestif mais plus dynamique en raison d'un flux aérien

iv. Un article récent critique l'idée d'un fœtus stérile.

bidirectionnel. Les bactéries du microbiotes proviennent de l'air ambiant, des voies supérieures, mais aussi du tube digestif par des micro-aspirations⁶. Une étude a d'ailleurs montré une concordance entre le microbiote digestif et respiratoire⁶.

Le microbiote pulmonaire est dominé par le phylum des *Firmicutes* (*Streptococcus*) et des *Bacteroidetes* (*Prevotella*). Les genres retrouvés majoritairement sont *Streptococcus*, *Prevotella*, *Fusobacteria*, *Veillonella*, *Haemophilus*, *Neisseria* et *Porphyromonas*. L'arbre respiratoire étant en continuité direct avec les voies aériennes supérieures, certains genres bactériens sont communs, comme *Streptococcus*, *Staphylococcus*, *Haemophilus* et *Moraxella*. Tandis que d'autres genres comme *Corynebacterium* et *Dolosigranulum* ne sont retrouvés qu'au niveau du nez et de l'oropharynx.

Le microbiote pulmonaire varie dans l'espace et le temps.

Dans l'espace, du fait de sa structure, certaines régions de l'arbre bronchique peuvent présenter des microbiotes différents. Un prélèvement au niveau d'un foyer infectieux se distinguera nécessairement d'un foyer sain. Il existe également des différences physico-chimiques selon la région pouvant sélectionner des espèces. Les cavernes tuberculeuses par exemple se trouvent essentiellement dans le lobe supérieur en raison d'une concentration en oxygène plus élevée qui favorise ce bacille aérobie stricte.

Le microbiome varie dans le temps. Avec l'âge la diversité diminue. Declin de la diversité alpha avec l'âge probablement à cause des exacerbations multiples et du traitement. Le microbiote est variable entre les individus ...

Le microbiote est corrélé à la pathologie respiratoire. Plusieurs études suggèrent une différence entre patients sains et patients asthmatiques, BPCO ou atteints de mucoviscidose [10]. Plus riche avec la maladie. Il est par exemple démontré que chez les jeunes enfants, une diminution d'infection respiratoire est associée à une augmentation d'asthme. Une étude montre une augmentation des *proteobacteria* en échange d'une baisse des *bacteroidetes* (*haemo*, *maraxella* et *neisseria*). Chez les muco, augmentation de la richesse et diminution de la diversité. Chez les muco, augmentation des germes opportuniste par rapport aux commensaux (data mining paper)

Le microbiote pulmonaire joue probablement le rôle de barrière vis-à-vis d'autre agent pathogène.¹⁰

3.3 Exploration du microbiote pulmonaire

3.3.1 Méthode de prélèvement

Le microbiote respiratoire est exploré en séquençant l'ADN des micro-organismes présents dans un échantillon. Toutes les méthodes de recueils sont possibles, mais les prélèvements protégés (combicath, Liquide Broncho Alvéolaire (LBA)) sont recommandés afin d'éviter une contamination par les voies aériennes supérieures. En pratique, les

patients atteints de mucoviscidose sont suivis par le recueil des prélèvements moins invasifs comme les prélèvements pharyngées ou les expectoration spontanées ou induites (ECBC). Dans ces dernières, la qualité du prélèvement peut être évaluée en comptant le nombre de cellules épithélium (normalement < 25 par champs à l'objectif x10) et de polynucléaire (normalement > 10 par champ à l'objectif x10) selon le score de Bartlett (REMIC 2015). Des prélèvements sont également réalisables *in situ* sur les explants lors des greffes pulmonaires.

3.3.2 Séquençage haut débit

Le séquençage de nouvelle génération permet de séquencer de grande quantité d'ADN dans un échantillon et ainsi déterminer sa composition en bactéries ou autres micro-organismes. À titre d'exemple, un séquenceur Sanger classique permet de lire des fragments d'ADN d'environ 800 pb parallélisable jusqu'à 96 fois en augmentant le nombre de capillaires. À l'inverse, un séquenceur de nouvelle génération lit des fragments plus courts de l'ordre de 150pb. Mais la parallélisation peut aller jusqu'à 20 milliards de fois en un seul run sur un illumina Novaseq.

2 stratégies de séquençages sont utilisées en écologie microbienne :

La stratégie *shotgun* consiste à séquencer l'ensemble des ADN présents dans l'échantillon sans discernement, que ce soit humain ou bactérien. Les séquences sont filtrées puis les génomes bactériens sont reconstruits par des méthodes bio-informatiques complexes.

La stratégie *Amplicon* est moins coûteuse sur le plan de l'analyse. Elle consiste à séquencer un gène spécifiquement bactérien et suffisamment variable pour discriminer une espèce. Il s'agit du gène de l'ARN 16S.

L'ARN 16S est un ARN non codant participant à la structure de la petite sous unité des ribosomes bactériens. Il est composé de 1500 nucléotides et forme plusieurs boucles dans sa structure secondaire (Figure 9). L'alignement des séquences d'ARN 16S entre plusieurs espèces met en évidence des régions constantes et 9 régions variables (Figure 8). Les régions constantes permettent de concevoir des amorces s'hybridant sur tous les ADNs bactériens. Les régions variables apportent la spécificité taxonomique permettant d'identifier l'espèce bactérienne.

En séquençant l'ensemble des génomes bactériens, la stratégie *shotgun* est plus informative car elle permet de prédire la fonction d'un microbiote. En effet, les transferts génétiques horizontaux amènent à dissocier l'espèce de sa fonction. Deux bactéries d'une même espèce peuvent avoir des fonctions différentes. L'inférence fonctionnelle à partir de la deuxième stratégie est possible mais déconseillée. La stratégie 16S reste toutefois une méthode simple pour décrire les populations bactériennes présentes. C'est cette stratégie qui a été utilisée dans notre étude.

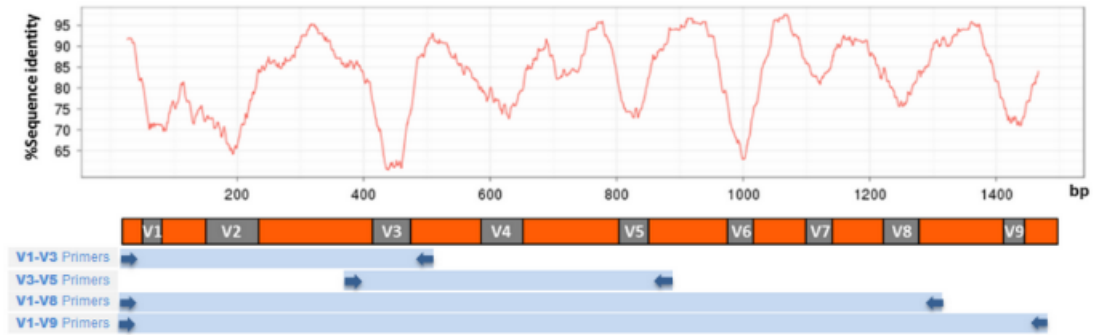


FIGURE 8 – Régions constantes et variables de l'ADNr 16S

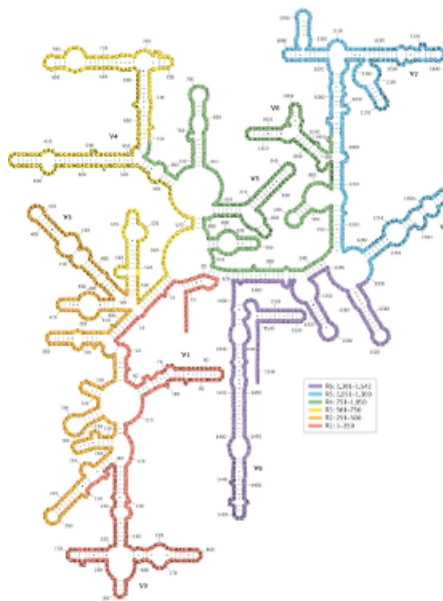


FIGURE 9 – Structure secondaire de l'ARN 16S

3.4 Objectif de l'étude Mucobiome

L'objectif général de l'étude MUCOBIOME est de savoir si dans la mucoviscidose, le microbiote respiratoire influence la colonisation à *Pseudomonas aeruginosa*. Pour cela, nous avons suivi pendant 3 ans, une cohorte de 47 patients atteints de mucoviscidose exempts de *Pseudomonas aeruginosa* depuis au moins 1 an. L'exploration de leur microbiote par la stratégie 16S a été réalisée à partir d'expectorations bronchiques obtenues dans le cadre de leurs suivis longitudinal. À partir des données générées par le séquençage, nous avons effectué l'étude descriptive et analytique de leurs microbiotes en créant un pipeline bio-informatique dédié. L'objectif principal de ce travail a été la mise au point et l'évaluation de ce pipeline à façon pour l'étude MUCOBIOME.

4 Matériel et Méthodes

4.1 Recueil des données

Quatre-vingt dix-sept patients atteints de mucoviscidose ont été suivis sur 3 ans (2008-2011) dans une étude prospective multicentrique (Nantes, Brest, Roscoff) appelée *Mucobiome*. Le comité de protection des personnes (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).

Les ECBC des patients ont été recueillies 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. Pour être inclus, le patient devait être exempt de *P. aeruginosa* depuis au moins un an d'après les résultats de la culture bactérienne des expectorations. Le patient devait être en capacité d'expectorer spontanément (exclusion des prélèvements pharyngés). La première culture bactérienne positive à PA était le « endpoint » de l'étude. A cette étape le patient était sorti de l'étude pour être ré-inclus un an plus tard en cas de maintien de la négativité des cultures bactériennes obtenue par l'antibiothérapie d'éradication anti-*P. aeruginosa*. D'autres données clinico-biologiques ont également été recueillies. Au total, 756 expectorations ont été recueillies à l'issue de l'étude MUCOBIOME. Pour ce travail portant sur l'étude du microbiote broncho-pulmonaire, 188 ont été sélectionnées.

citer qq exemples de données (que tu exploiteras après) et préciser le score cytologique (à mettre dans les résultats : impact oui ou non sur les données du microbiote)

4.2 Extraction de l'ADN

La procédure d'extraction des ADN des échantillons est détaillée dans la publication de Le Gal... Brièvement les échantillons ont été liquéfiés avec du Dithiotréitol(DTT). Les protéines ont été dégradées avec de la protéinase K. Les parois bactériennes ont été fragmentées par sonication. (DTT par sonication (Elamsonic S10, Singen, Germany). Après 10 min de centrifugation, L'ADN a été extrait à partir du culot à l'aide du kit QUIAamp DNA Minikit (Quagen). Les extraits d'ADN ont été envoyés pour séquençage par un prestataire GATC.

	Never	Free
<i>N samples</i>	114	74
<i>Male (%)</i>	77 (67.54)	41 (55.41)
<i>Age mean (year)</i>	22.45	19.65
<i>Weight mean(kg)</i>	39.3	33.66
<i>Height mean (cm)</i>	146.7	140.8
<i>IMC mean (kg/cm²)</i>	17.22	16.32
<i>dF508/dF508 homo (%)</i>	59 (51.75)	35 (47.3)
<i>dF508/other hetero (%)</i>	55 (48.25)	39 (52.7)
<i>Taking Domase (%)</i>	78 (68.42)	59 (79.73)
<i>Taking antibiotics (%)</i>	48 (42.11)	33 (44.59)
<i>Positif PA culture (%)</i>	10 (8.77)	12 (16.22)
<i>Positif PA PCR (OPRL1+) (%)</i>	20 (17.54)	26 (35.14)

FIGURE 10 – Données associées par échantillons (à refaire par patients)

4.3 Séquençage

La librairie a été générée en amplifiant la région V3-V5 à l'aide du couple d'amorces *forward*(*CCTACGGGAGGCAGCAG*) et *reverse*(*CCGTCAATTCMTTTRAGT*) et du kit MiSeq Reagent Kits v3.

Le séquençage a été réalisé sur Illumina MiSeq pour produire un couple de séquences chevauchantes de 300 pb (Figure 11). Après fusion du couple, le séquençage permet de lire 535 pb correspondant à la région V3-V5.

Environ 25 millions de reads sont produits par run MiSeq. En multiplexant à l'aide de 94 index, les 188 échantillons ont été séquencés sur 2 runs pour produire 188 x 2 fichiers fastq.

4.4 Analyse bio-informatique

L'analyse des 188 paires de fichiers fastq a été réalisée grâce à un pipeline bio-informatique, appelé *mucobiome*, conçu et testé dans le cadre de cette étude. Par rapport aux autres logiciels comme *QIIME*⁴ ou *MOTHUR*²², le pipeline *mucobiome* est spécialisé dans l'analyse des données 16S. Il est également plus rapide en raison d'un très haut

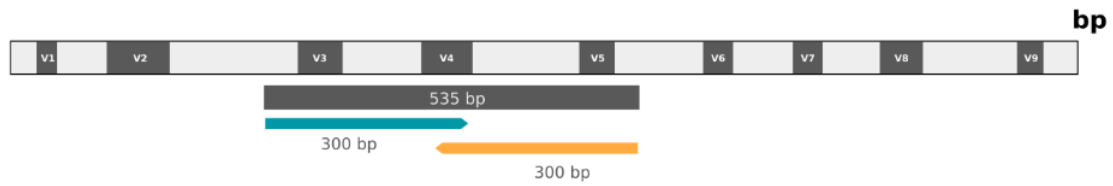


FIGURE 11 – le couple de séquence de 300pb permet de recouvrir l'ensemble de la région V3-V5 de l'ARN 16S

niveau de parallélisation permis grâce à *Snakemake*¹¹. Cet outil modélise l'ensemble du pipeline sous forme d'un graphe direct acyclique (DAG) et le résout afin d'optimiser la parallélisation.

Le pipeline mucobiome prend en entrée, les 188 couples de fichiers fastq provenant du séquençage et produit un fichier BIOM contenant la table des OTUs. Les figures 12 et 13 illustrent toutes les étapes du pipeline qui sont décrites ci-dessous.

4.5 Étape du pipeline

4.5.1 Merging : Fusions des reads

2 fastq en entrée 1 fastq en sortie.

La première étape du pipeline consiste à fusionner le couple de fichier fastq afin de produire un seul fichier contenant des plus longues séquences de 535pb correspondant à la région V3-V5 de l'ARN 16S. Le programme **Flash**[@] a été utilisé avec les paramètres par défaut. À partir de deux fichiers fastq, ce dernier recherche le meilleur alignement entre deux reads et produit 1 fichier fastq contenant les reads fusionnés. Une analyse qualitative des reads a été réalisée avec FastQt² avant et après fusion.

4.5.2 Cleaning : Filtrage des qualités

1 fastq en entrée 1 fastq en sortie.

Les données de séquençage haut débit peuvent contenir beaucoup d'erreurs. Il est important de supprimer les reads de mauvaise qualité pour gagner en spécificité. (expliquer la qualité). Le filtrage des reads de mauvaise qualité est réalisé avec le programme **sickle**. Son algorithme repose sur l'utilisation d'une fenêtre glissante de taille définie (par défaut : 20 Pb). Cette fenêtre glisse le long de la séquence et pour chaque position calcule la moyenne des scores de qualité dans cette fenêtre. Si successivement le score moyen passe sous un certain seuil, le read est supprimé. Les paramètres utilisés sont ceux par défaut. Un score de 20 avec une fenêtre glissante de 20pb. Une analyse qualitative des reads a été réalisée avec FastQt² après le filtrage.

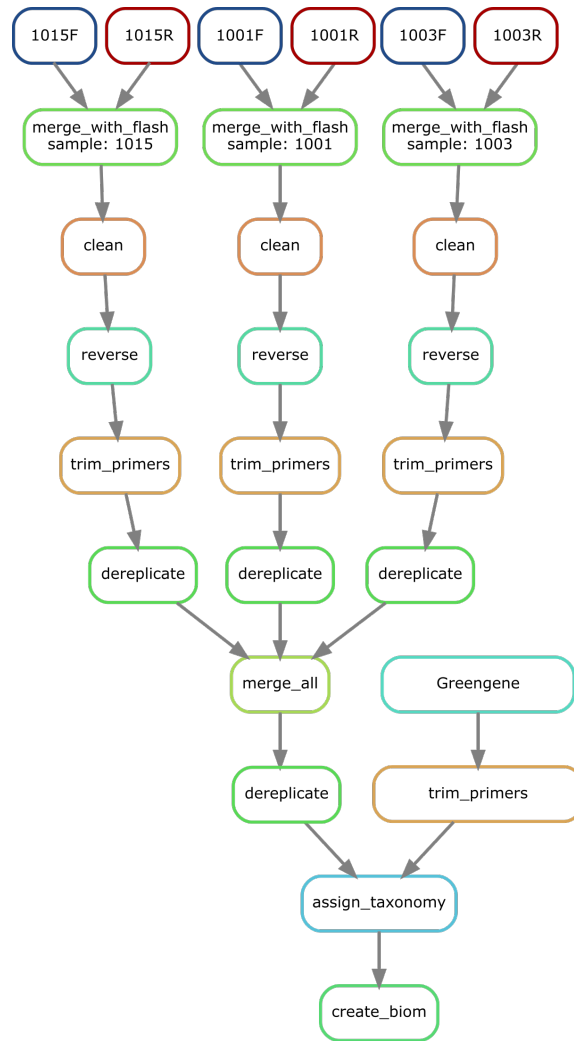


FIGURE 12 – Graphe du pipeline simplifié sur 3 échantillons 1015, 1001 et 1003.

merging : Les reads pairs de 300 pb sont fusionnés pour produire un fichier fastq contenant des reads de 535pb. **cleanning** : les reads de mauvaise qualité sont supprimés. **reversing** : les reads sont transformés en leurs séquences complémentaires pour pouvoir être alignés. **Trimming** : seule la séquence entre les primers V3-V5 est conservée. **dereplicating** : les séquences dupliquées sont retirées. **merging** : L'ensemble des séquences est regroupé dans un seul fichier. **Taxonomy assignment** : Les séquences sont alignées sur la base de données greengene. **create_biom** : la table des OTU est créée

4.5.3 Reversing : Séquence complémentaire

1 fastq en entrée 1 fasta en sortie.

Les reads produits par le séquenceur ne sont pas orientés dans le même sens que la base de données *greengene*. Pour permettre l'alignement, les séquences ont été remplacées par leurs séquences complémentaires grâce au programme **seqtk**. Par la même occasion les scores de qualités devenus inutiles sont supprimés. Les séquences sont sau-

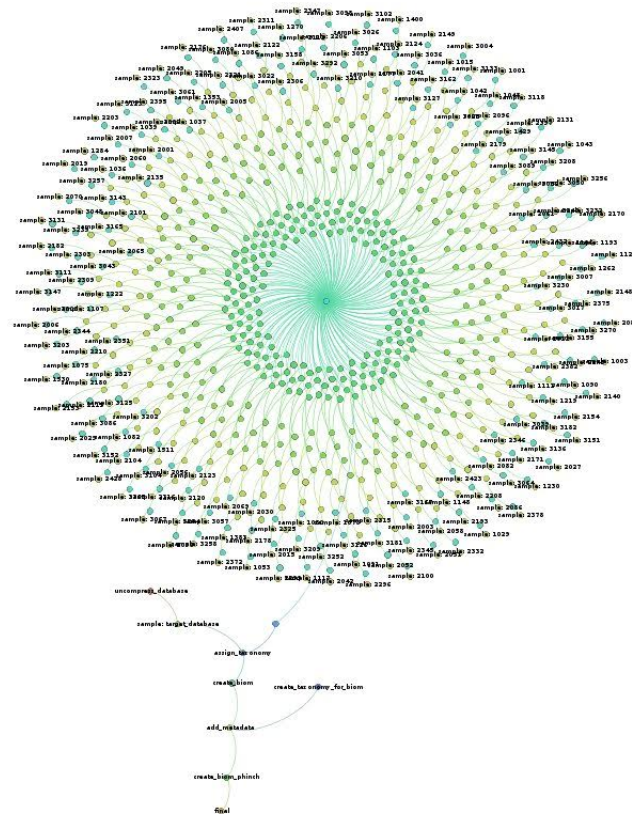


FIGURE 13 – le graphe du pipeline sur l'ensemble des échantillons

vegardées dans un fichier fasta.

4.5.4 Trimming : Suppression des primers

1 fasta en entrée 1 fasta en sortie.

Pour permettre un alignement parfait entre les reads et la base de données, les primers sont retirés et seule la séquence V3-V5 est conservée. Cette étape est réalisée aussi bien pour les données du séquençage que la base de données greengene. Le programme **cutadapts** est utilisé avec une tolérance de 0.1 par défaut.

4.5.5 dereplicating : Suppression des doublons

1 fasta en entrée 1 fasta en sortie.

Cette étape consiste à supprimer toutes les séquences doublons. En procédant ainsi, on s'assure de ne pas répéter l'assignement taxonomique plusieurs fois sur des reads identiques. Le nombre de reads dupliqués est conservé pour être pris en considération

lors du calcul des abondances. Il s'agit d'une étape d'optimisation permettant d'économiser en temps de calcul. La déréplication a été réalisé avec **vsearch** et sa fonction *-derep_fulllength*

<pre>>sample1 ACGTTTTT >sample1 ACGTTTTT >sample1 GTAGAGT >sample1 ACGTTTTT</pre>	<pre>>sample1;size=3 ACGTTTTT >sample1;size=1 GTAGAGT</pre>
<i>avant dereplication</i>	<i>après dereplication</i>

FIGURE 14 – Exemple de déréplication d'un fichier fasta

4.5.6 Assignment taxonomique

2 fasta en entrée, 1 fichier biom en sortie.

L'assignment taxonomique consiste à labelliser chaque read à son taxon. Nous avons utilisé la stratégie *close référence* dont le rôle est de comparer chaque read à une base de données avec un seuil de 98% de similarité. Cet algorithme est de complexité N . C'est-à-dire que le temps de calcul est directement proportionnel au nombre de reads testé. La base de données *Greengene* version mai 2013 a été utilisé. Il s'agit d'un fichier fasta contenant 1 262 986 séquences et 203 452 OTUs.

L'autre stratégie d'assignment *de novo* n'a pas été utilisée. Cette dernière, de complexité N^2 , consiste à comparer les reads entre eux pour former des groupes. Elle s'emploie de préférence pour détecter les bactéries absentes des bases de données.

L'assignment taxonomique a été réalisé avec **vsearch** et sa fonction *-usearch_global*.

4.5.7 Analyse de la table des OTUs

L'analyse de la table des OTUs a été réalisée avec R et le package **phyloseq**. Les OTUs ont été regroupé par genre bactérien. L'étude descriptive a été réalisé en calculant des indicateurs de diversité alpha et en produisant différents graphiques. Le core microbiota a été calculé. Il est défini comme l'ensemble des taxons retrouvés dans plus de 50

5 Résultat

5.1 Séquençage et pipeline

5.1.1 Données de séquençage

Après démultiplexage, 188 x 2 fichiers Fastq ont été générés soit 2 fichiers par échantillons. La longueur des reads dans l'ensemble est de 301 paires de bases. Au total, 115 002 297 reads ont été produits sur 2 runs MiSeq (Figure 15). Avec en moyenne 616 900 reads par échantillon. Un minimum de 61 422 reads pour l'échantillon 2154 et un maximum de 1 071 188 pour l'échantillon 3165.

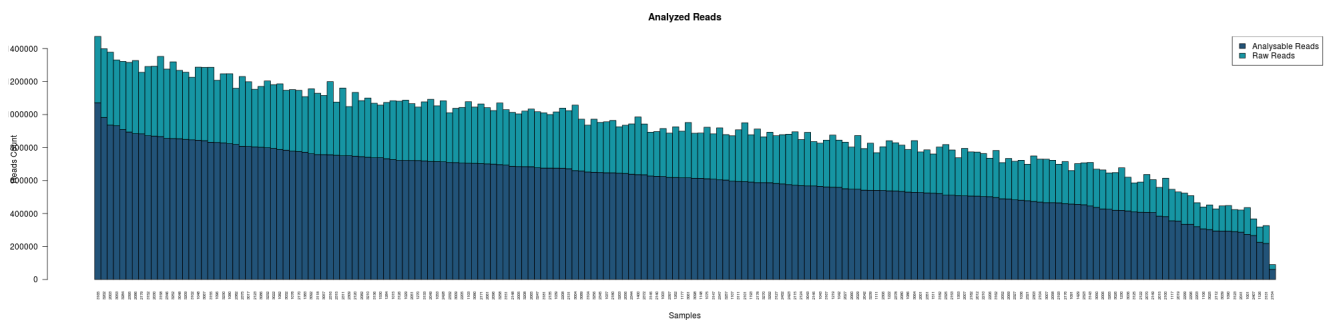


FIGURE 15 – Nombre de reads par échantillons

5.1.2 Qualité des reads

La figure 16 montre la qualité typique d'un fichier fastq produit par le séquenceur. Une baisse de qualité importante est observée à hauteur du 250ème nucléotides. Tous les fichiers fastq présentaient le même profil. A cause de cette baisse de qualité, x% des reads pairs n'ont pas réussi à fusionner. L'étape de filtrage a permis de ramener la qualité médiane au dessus de 20 (Figure 17). Au total seulement 49.24% des reads sont conservé pour l'analyse (Figure 15) avec des bornes allant de 37,30% à 61,13%.

5.1.3 Assignement taxonomique

99.88% (4615960) des reads analysables ont reçu une assignation taxonomique pour 10517 OTUs^v correspondant à 54 genres bactériens (Figure 18). Le temps de calcul a été de 1 h 29 et a nécessité 40 cœurs et 20 gigaoctets de mémoire contre 21h de calcul sans l'optimisation par déréduplication.

v. Pour chaque espèce, il y a plusieurs OTUs définis dans greengene

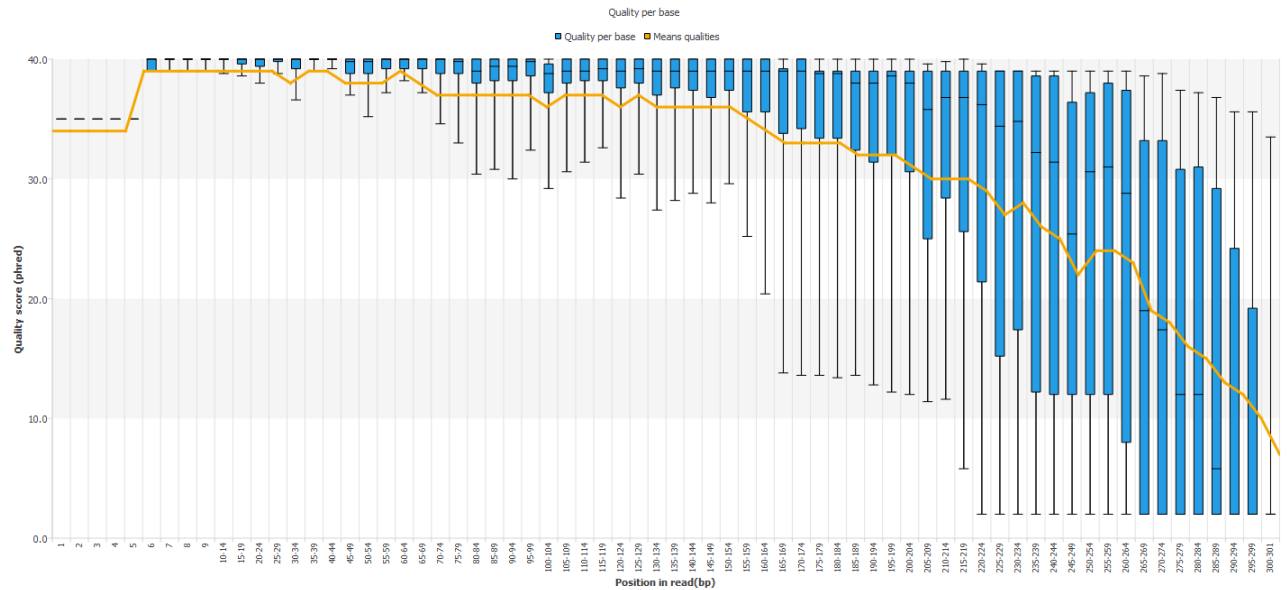


FIGURE 16 – Qualité par nucléotide des reads forward de l'échantillon 1003. **Axe X** : la position sur le read. **Axe Y** : La distribution des qualités

5.1.4 Profondeur d'échantillonnage

La figure 19 montre les courbes de raréfaction pour les 188 échantillons. Dans l'ensemble elles s'aplatissent précocement, témoignant d'un bon niveau d'échantillonnage. Les quelques échantillons n'ayant pas atteint l'asymptote horizontale ont tout de même été conservés .

5.2 Résultats descriptifs

5.2.1 Composition du microbiote

Cinquante-quatre genres (Figure 18) et sept phylums (Figure 20) bactériens sont retrouvés dans l'ensemble des échantillons analysés. Les deux phyla majoritaires sont *Firmicute* (42.59%) et *Proteobacteria* (31.48%). Parmi les *Firmicutes* majoritaires, on retrouve *Streptococcus* et *Staphylococcus*. Chez *Proteobacteria*, *Neisseria* et *Haemophilus* sont les genres les plus abondantes. Le tableau de la figure 21 résume l'ensemble des résultats en y associant la prévalence des genres bactériens parmi les 188 échantillons. Par exemple *Streptococcus*, *Neisseria*, *Prevotella*, *Granulicatella*, *Gemella*, *Veillonella* et *Fusobacterium* sont très prévalents, car présents dans plus de 185 échantillons. D'autres genre sont dominants. Il s'agit de *Streptococcus*, *Neisseria*, *Haemophilus* et *Staphylococcus*. *Sténotrophomonas* et *Achromobacter* sont retrouvées dans 64 et 8 échantillons respectivement. *Pseudomonas* est retrouvé dans 53 échantillons, dont un, au moins avec une abondance de 25.94%. *Burkholderia* est retrouvé seulement dans deux échantillons

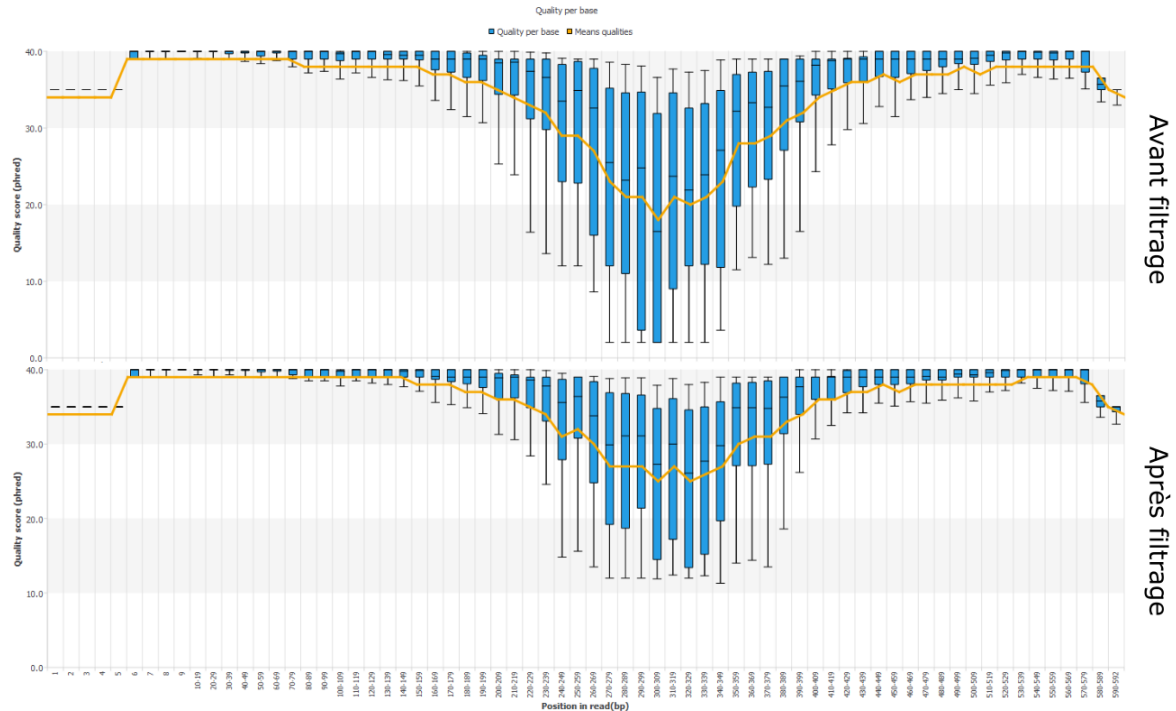


FIGURE 17 – Qualité par nucléotide des reads fusionnés pour l'échantillon 1003

à moins de 1%.

Le core microbiota est défini comme l'ensemble des taxons retrouvés dans plus de 50% des échantillons avec une abondance supérieure à 0,1%. Il est constitué de 15 genres bactériens (Figure 22).

La distribution du core microbiota est illustrée dans la figure 23. *Streptococcus* respecte grossièrement une distribution normale variant de la quasi-absence à la dominance avec une moyenne de 30% par échantillon. *Neisseria* est le deuxième genre le plus abondant avec une moyenne de 18%. Les abondances de *Staphylococcus* et *Haemophilus* sont faibles dans la plupart des échantillons. Mais pour quelques échantillons, ces genres sont dominants. Les autres genres ont une abondance faible qui varie faiblement. Elles ne sont jamais retrouvées comme dominantes.

5.2.2 Évolution de l'alpha diversité

Les figures 34, 35 et 36 montrent l'évolution des diversités alpha par patient au cours du temps en utilisant les indices Chao1, Observed et Shannon.

La richesse (Observed, Chao1) par patient varie de 10 à 40 genres bactériens. La richesse du patient 256 est faible. Celle du patient 26 élevée. Certains patients ont des richesses stables au cours du temps. Les patients 25, 26, 74 conservent leurs richesses sur plus de 5 prélèvements successifs. Les patients 20 et 69 présentent une stabilité globale entrecoupée par des pertes de biodiversité. La richesse du patient 8 diminue progres-

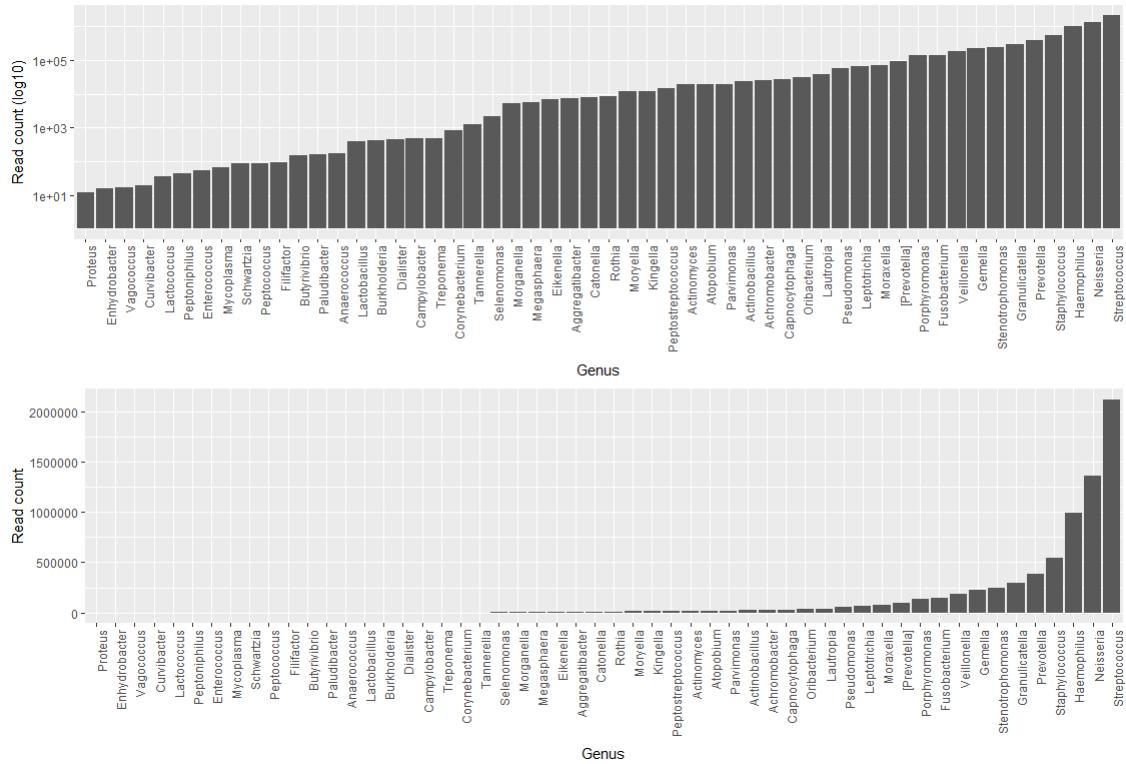


FIGURE 18 – Nombre de reads par genre bactérien

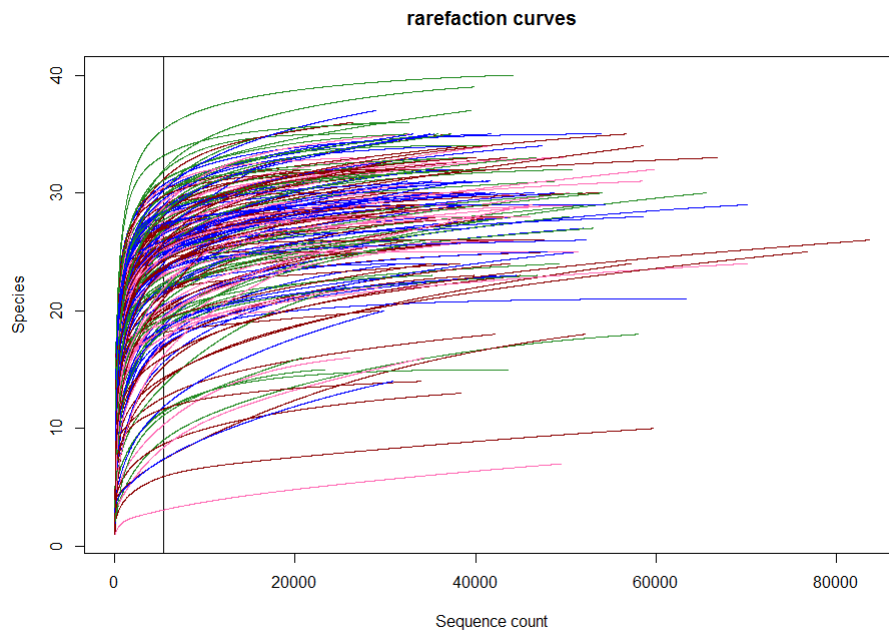


FIGURE 19 – Courbe de rarefaction

sivement. Les patients 23, 232 et 248 ont des fluctuations plus chaotiques. L'indice de Shannon montre que pour un richesse constante dans le temps, l'équitabilité est diffé-

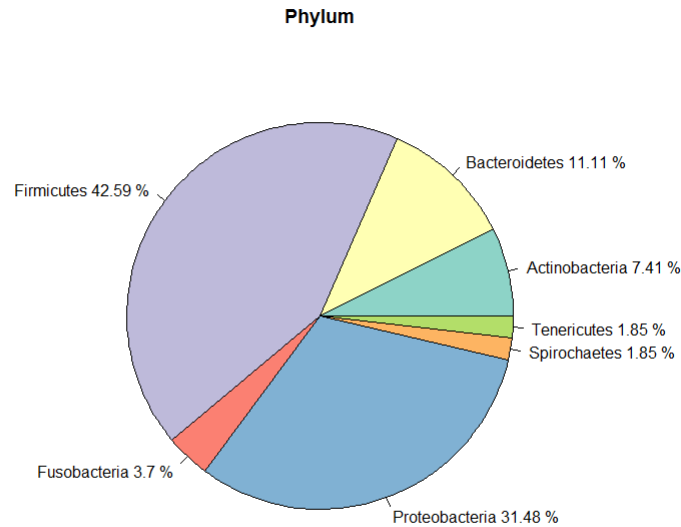


FIGURE 20 – Phylum

rente. Par exemple la richesse du patient 223 est stable sur le graphique 34 mais son indice de shannon varie sur le graphique 36 témoignant d'une distribution différente des ses bactéries. Seul le patient 26 semble conserver à la fois une richesse et une équitabilité stables au court du temps.

5.2.3 Évolution des abondances

Les figures 31, 32 et 33 montrent l'évolution des abondances au cours du temps pour chaque patient. Ces graphiques nous permettent d'interpréter plus finement les graphiques d'alpha diversité précédente. Par exemple la faible richesse du patient 8 est liée à une dominance de *Stenotrophomonas* sur les 4 échantillons. La perte de diversité du 3ème prélèvement du patient 223 est causée par l'apparition de *Neisseria* qui devient dominant. Le patient 3 montre une diminution progressive d'*Haemophilus* parallèlement à une augmentation de *Streptococcus*.

Le patient 256 présente une dominance à *Sténotrophomonas* sur l'ensemble de ses échantillons. *Achromobacter* est dominant dans le premier échantillon du patient 211. D'autres montrent un phénomène de résilience. Par exemple, le patient 223 récupère un microbiote identique après une colonisation complète à *Haemophilus*.

Dans l'ensemble, il existe deux types de population bactérienne : une population de bactéries constantes et minoritaires (*Fusobacterium*, *Granulicatella*, *Gemella*, *Veillonella*, *Parvimonas*, *Leptotrichia*, *Oribacterium*, *Capnocytophaga*, *Catonella*) ; Une autre popula-

	reads	reads.total.p	rel.abundance.mean	rel.abundance.sd	rel.abundance.min	rel.abundance.max	samples
<i>Streptococcus</i>	2119507	29.32	30.22	20.04	0.01	89.87	188
<i>Neisseria</i>	1363473	18.86	18.08	18.49	0	93.96	187
<i>Haemophilus</i>	994209	13.75	12.91	23.97	0	99.78	184
<i>Staphylococcus</i>	549082	7.6	7.28	15.79	0	97.9	178
<i>Prevotella</i>	385850	5.34	5.63	5.98	0	30.85	188
<i>Granulicatella</i>	294632	4.08	4.12	3.83	0	26.73	185
<i>Gemella</i>	226403	3.13	3.29	3.26	0	15.34	186
<i>Stenotrophomonas</i>	241616	3.34	3.12	15.21	0	94.71	65
<i>Veillonella</i>	183601	2.54	2.76	3.18	0	15.62	188
<i>Fusobacterium</i>	144709	2	2.18	2.72	0	15.43	185
<i>Porphyromonas</i>	141016	1.95	1.96	3.46	0	26.08	167
<i>[Prevotella]</i>	92004	1.27	1.4	2.34	0	19.03	170
<i>Leptotrichia</i>	68581	0.95	1.01	1.69	0	14.44	172
<i>Pseudomonas</i>	56914	0.79	0.81	3.83	0	25.94	53
<i>Moraxella</i>	72796	1.01	0.79	6.33	0	75.51	51
<i>Lautropia</i>	39496	0.55	0.53	1.46	0	13.94	142
<i>Achromobacter</i>	25274	0.35	0.52	7.06	0	96.78	8
<i>Oribacterium</i>	31670	0.44	0.45	0.67	0	5.17	178
<i>Capnocytophaga</i>	26941	0.37	0.42	0.82	0	7.25	175
<i>Actinobacillus</i>	24228	0.34	0.37	2.35	0	31.19	95
<i>Atopobium</i>	19381	0.27	0.29	0.54	0	3.61	166
<i>Actinomyces</i>	19094	0.26	0.29	0.51	0	4.18	174
<i>Parvimonas</i>	19479	0.27	0.29	0.76	0	6.07	151
<i>Peptostreptococcus</i>	14961	0.21	0.23	0.49	0	2.93	129
<i>Kingella</i>	11943	0.17	0.17	0.47	0	4.55	149
<i>Moryella</i>	11912	0.16	0.16	0.39	0	3.62	143
<i>Rothia</i>	8346	0.12	0.12	0.25	0	1.83	138
<i>Catonella</i>	8130	0.11	0.12	0.17	0	1.16	159
<i>Aggregatibacter</i>	7597	0.11	0.11	0.32	0	2.51	86
<i>Eikenella</i>	6883	0.1	0.1	0.35	0	2.55	143
<i>Megasphaera</i>	5772	0.08	0.08	0.26	0	2.6	105
<i>Morganella</i>	5192	0.07	0.06	0.83	0	11.44	7
<i>Selenomonas</i>	2164	0.03	0.04	0.1	0	0.81	147
<i>Tannerella</i>	1240	0.02	0.02	0.04	0	0.31	120
<i>Corynebacterium</i>	854	0.01	0.01	0.03	0	0.18	113
<i>Treponema</i>	495	0.01	0.01	0.04	0	0.39	53
<i>Burkholderia</i>	415	0.01	0.01	0.11	0	1.36	2
<i>Campylobacter</i>	495	0.01	0.01	0.02	0	0.11	99
<i>Lactobacillus</i>	391	0.01	0.01	0.03	0	0.39	34
<i>Dialister</i>	461	0.01	0.01	0.02	0	0.16	53
<i>Vagococcus</i>	17	0	0	0	0	0.01	13
<i>Mycoplasma</i>	68	0	0	0	0	0.02	19
<i>Anaerococcus</i>	174	0	0	0.04	0	0.51	2
<i>Enhydrobacter</i>	16	0	0	0	0	0.02	9
<i>Lactococcus</i>	36	0	0	0	0	0.04	13
<i>Paludibacter</i>	158	0	0	0.02	0	0.18	18
<i>Peptoniphilus</i>	45	0	0	0	0	0.02	12
<i>Schwartzia</i>	85	0	0	0.01	0	0.12	11
<i>Peptococcus</i>	88	0	0	0	0	0.04	23
<i>Butyrivibrio</i>	146	0	0	0.02	0	0.19	10
<i>Enterococcus</i>	54	0	0	0.01	0	0.08	27
<i>Proteus</i>	12	0	0	0	0	0.01	9
<i>Filifactor</i>	93	0	0	0.01	0	0.11	5
<i>Curvibacter</i>	19	0	0	0	0	0.01	18

FIGURE 21 – Nombre de reads et prévalence parmi les 188 échantillons.
 nombre de reads - pourcentage de reads - abondance moyenne - ecart-type - abondance
 minimum - abondance maximum - prévalence (phrase ?)

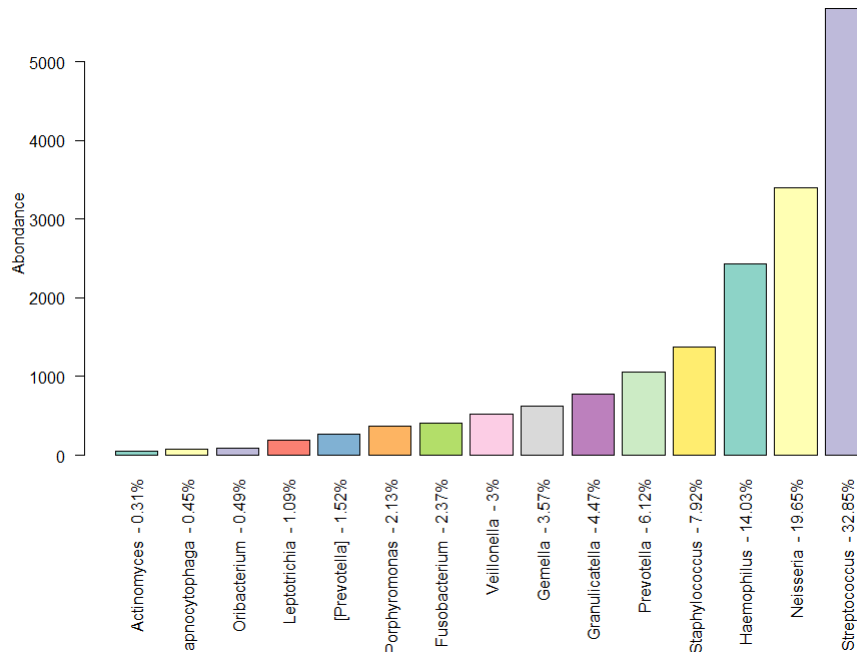


FIGURE 22 – Distribution du core microbiota dans l'ensemble des échantillons

tion de bactéries très fluctuantes dans le temps jouant alternativement le rôle de genre dominants ou au contraire du genre absent. (*Haemophilus*, *Streptococcus*, *Neisseria* et *Prevotella*). La figure 24 illustre ces propos en montrant l'évolution des abondances du patient 43. La figure du patient 25 est mise également à titre exemple, car elle montre l'apparition de *Pseudomonas aeruginosa*.

MATERIEL ET METHODS = EXPLIQUER LA PCA

5.2.4 Beta Diversité

La bêta diversité sur l'ensemble des échantillons a été réalisé par une méthode d'ordination de type PCA en utilisant les distances de Bray-Curtis (Figure 26,27 et 30). 2 axes principaux expliquent respectivement 28.5% et 17.4% de la variabilité. Certains échantillons d'un même patient sont très proches sur le graphique d'ordination. Par exemple le patient 69 et 003 ont des échantillons dont les points se confondent. Aucun des analyses tenant compte des paramètres biocliniques colligés pendant l'étude MUCOBIOIME comme âge, le sexe, le poids, la prise d'antibiotique et le type de mutation du gène *CFTR* n'a mis en évidence des groupes distincts de microbiote. La variabilité est expliqué principalement par la dominance des genres *Neisseria*, *Streptococcus* et *Haemophilus* comme le montre la figure 27.

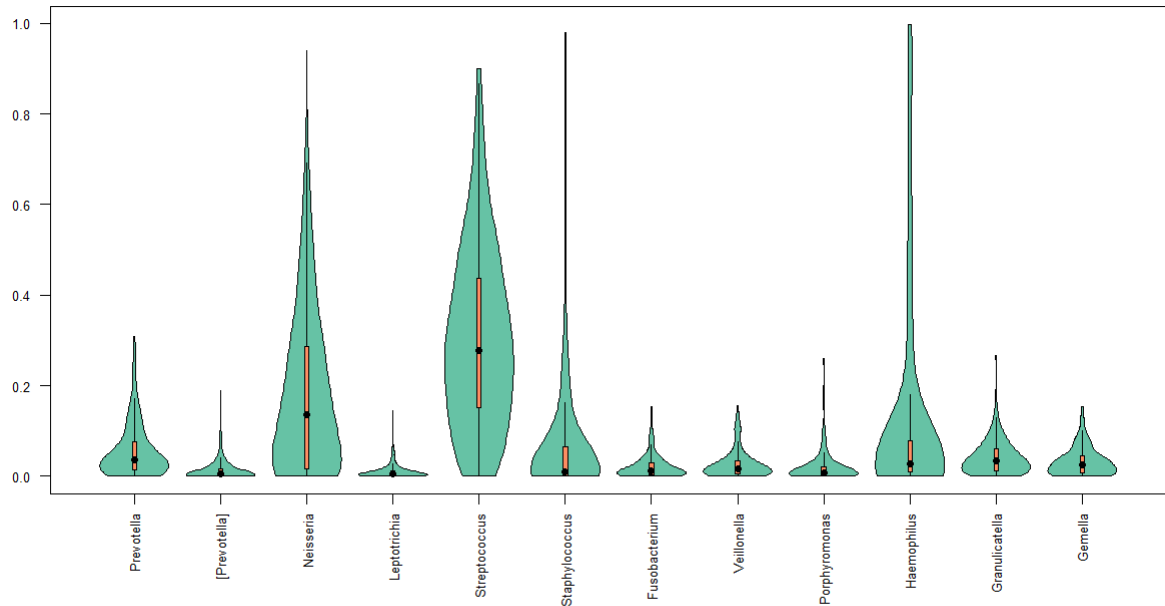


FIGURE 23 – Variation du core microbiota dans l'ensemble des échantillons

5.3 Résultats analytiques

5.3.1 Corrélation entre les genres bactériens

La figure 28 montre les corrélations linéaires réalisées entre les genres du core microbiota. La corrélation la plus forte est entre *Prevotella* et *Veillonella* avec un coefficient de Pearson à 0.60. *Streptococcus* et *Haemophilus* évoluent dans le sens inverse avec un coefficient à -0.41.

5.3.2 Comparaison entre échantillons Free et Never

La figure 29 compare les indices de Shannon entre les échantillons Free et Never. Le T-test / ANOVA n'a pas montré de différence significative. (p-value > 0.05).

La figure 30 est une analyse multivariée discriminant les groupes Free et Never. Les échantillons Never ont plus de différences entre eux que les échantillons Free qui semblent converger. Les deux groupes sont significativement différents p-value < 0.001.

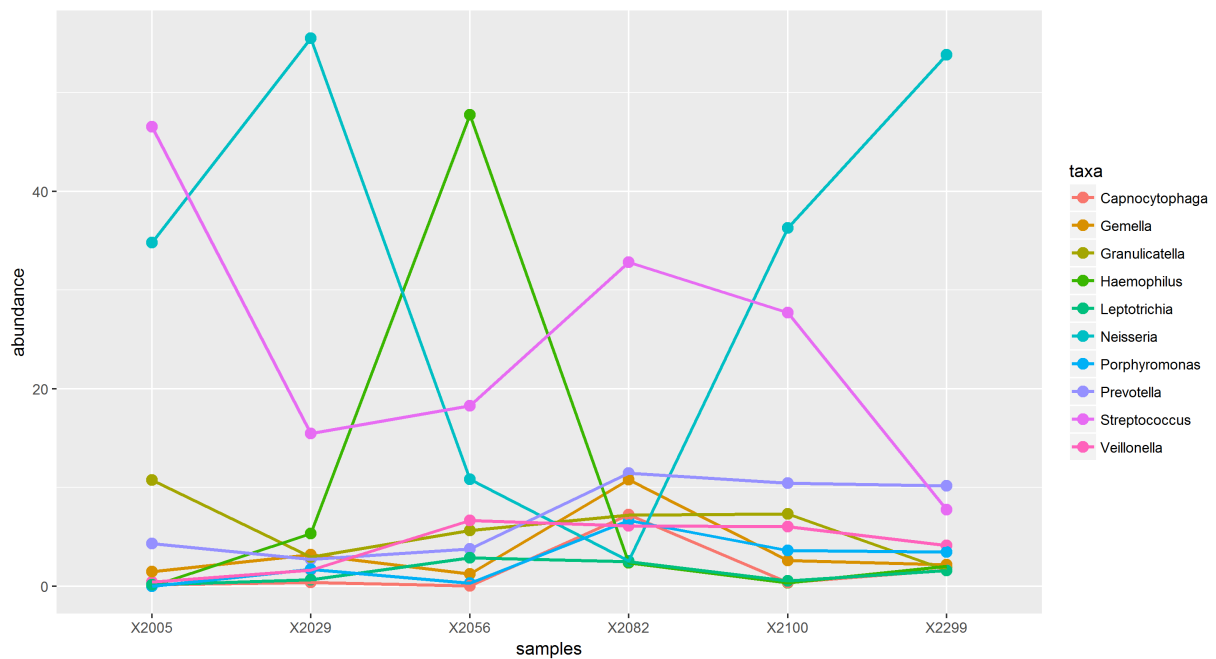


FIGURE 24 – Evolution des abondances pour le patient 43. Notez la population stable et la population fluctuante

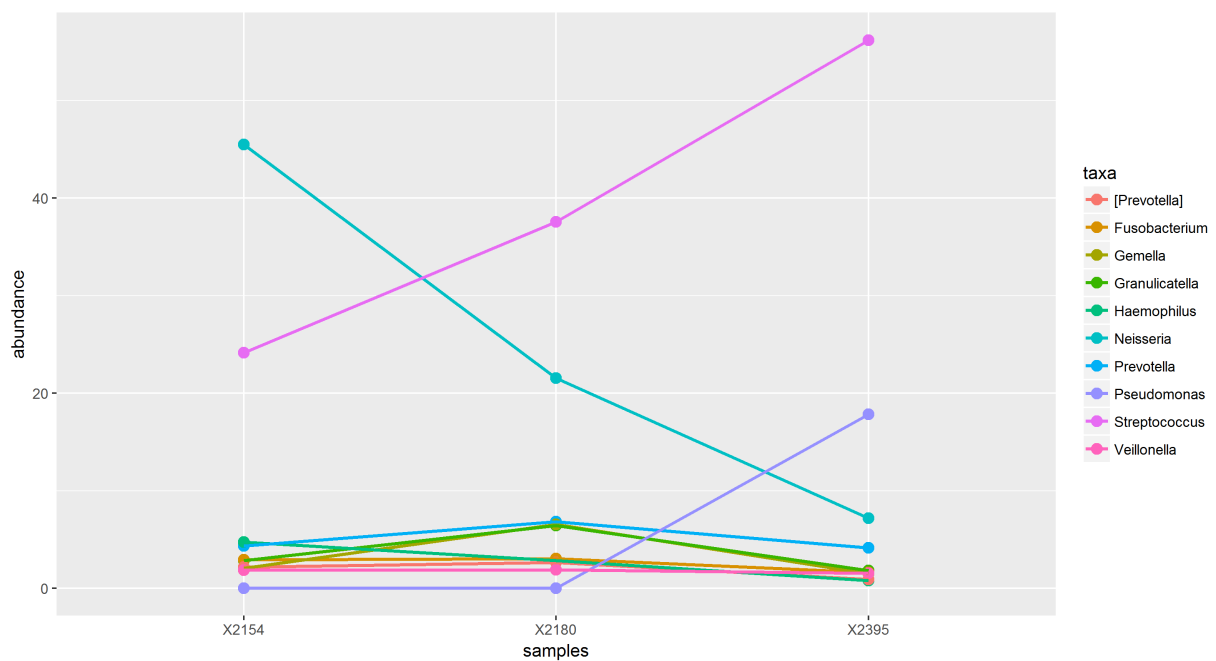


FIGURE 25 – Evolution des abondances pour le patient 54. Notez l'apparition de *Pseudomonas* dans le dernier échantillon

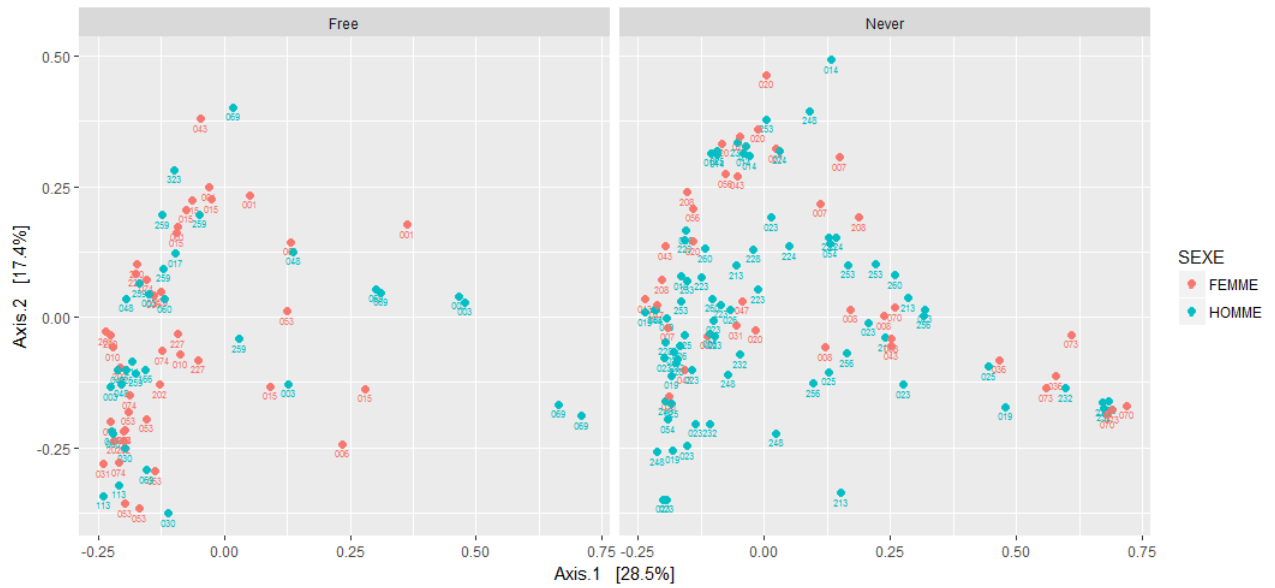


FIGURE 26 – Analyse en coordonnée principale sur les 188 échantillons en utilisant les distance de Bray-curtis. Chaque point est un échantillon labelisé par l'identifiant du patient. A gauche les échantillons Free. A droite les échantillons Never.



FIGURE 27 – Analyse multivarié (PCA + Bray curtis) sur les 188 échantillons. Chaque point est un échantillon labelisé par l'identifiant du patient.

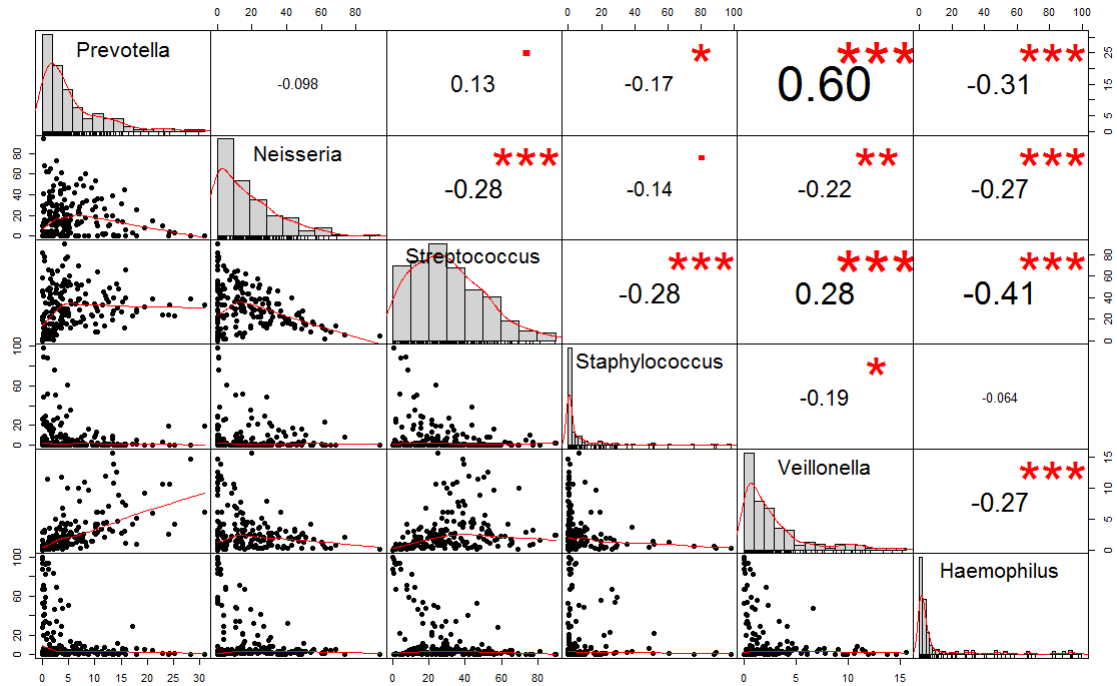


FIGURE 28 – Corrélation des abondances entre les genres du core microbiota

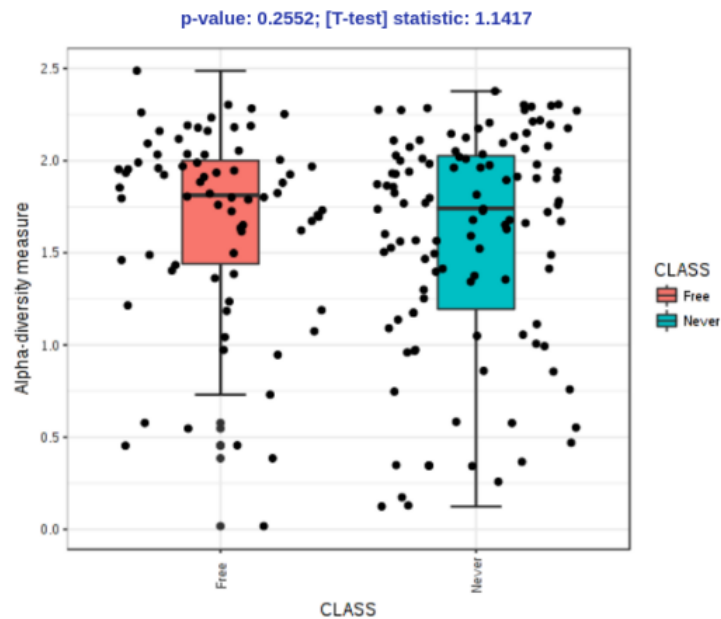


FIGURE 29 – Comparaison des diversités de Shannon entre les échantillons Free et Never

6 Discussion EN COURS

6.1 Méthode de séquençage

La version 3 du kit MiSeq Reagent Kits produits des reads plus long (300pb) que la version 2 (250pb) afin d'augmenter la précision de l'assignement taxonomique. En truc

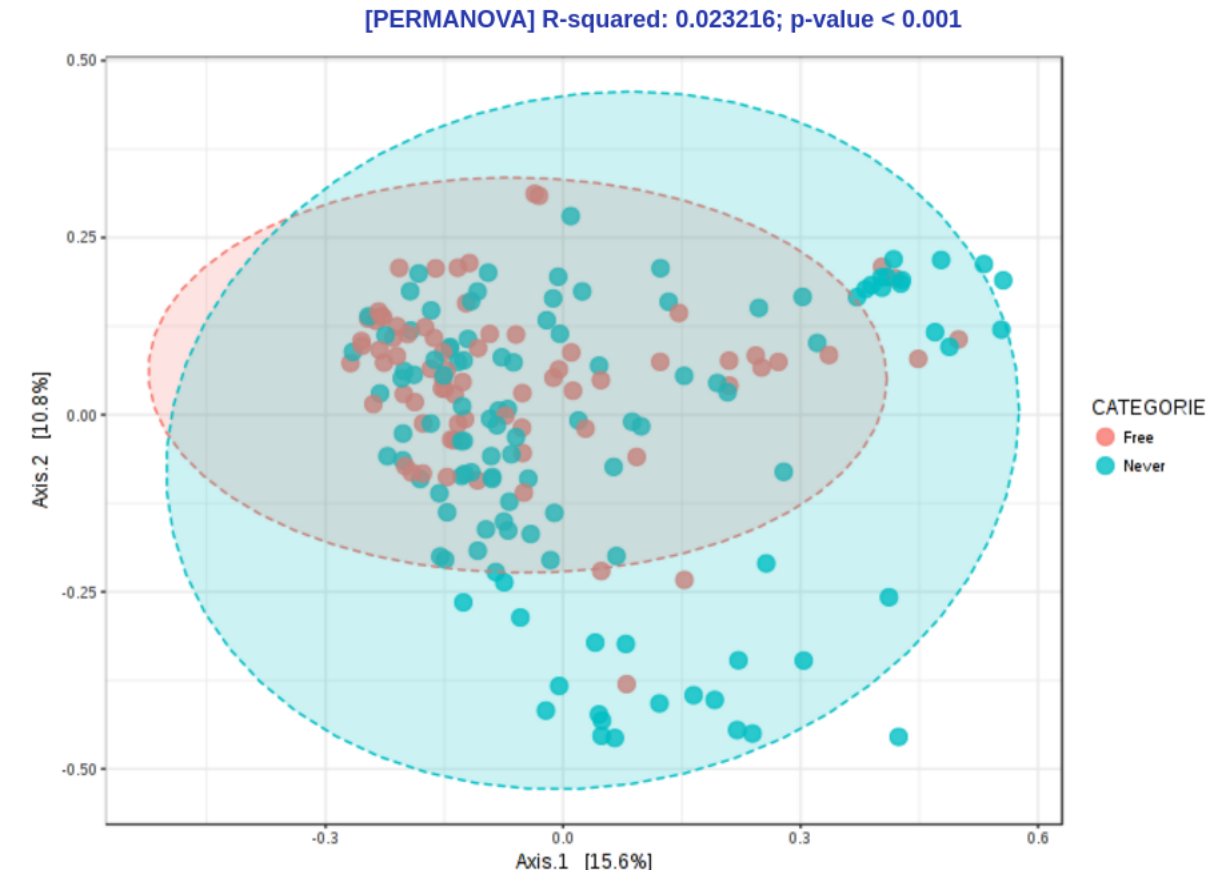


FIGURE 30 – Permanova entre les échantillons Free et Never sur une PCoA en utilisant la distance de Bray-Curtis

contrepartie la qualité du séquençage est médiocre en 3' avec des scores de qualité en dessous de 20. C'est un problème connu liée à la chimie du séquenceur qu'illumina se garde bien de cacher sur sa page web. Cependant, malgré les 43% de reads analysables, la profondeur de séquençage reste suffisante comme l'atteste la courbe de rarefaction. La stratégie close-reference utilisé dans le pipeline n'a pas réussi à assigner les reads jusqu'au rang de l'espèce. Mais 99% des reads ont reçu un assignement allant jusqu'au rang du genre. Cela signifie que la majorité des genres bactéries du microbiotes respiratoires sont connues de *greengene*.

Cette stratégie reste cependant critiquable. En effet l'assignement dépend d'un pourcentage de simmiliarité de 97% fixé arbitrairement. D'autre technique permettant de s'affranchir de ce seuil augmenterait probablement la résolution taxonomique. Par exemple la Maximum entropy clusterisation.

Enfin, la phylogénie des bactéries n'a pas été prise en compte dans l'étude. Au lieu d'utiliser la distance de Bray-Curtis, la distance phylogénétique (methode UPGMA) aurait pu apporter plus d'information.

6.2 Composition du microbiote pulmonaire dans la mucoviscidose

Chez un patient sain, les deux phylums majoritaires du microbiote pulmonaire sont les Bacteroidetes et les Firmicutes en proportion égal au alentour de 40% et les Protéobactéries autour de 10%. Nos résultats chez le patient mucoviscidoses nous montre en revanche une augmentation des Protéobacteria au dépend des Bactéroidetes. Ceci corrobore les résultats de [10] retrouvant ces mêmes proportions dans la Mucoviscidose, l'asthme et la BPCO. Cette excès s'explique principalement par la dominance d'Haemophilus et de Neisseria, et dans dans une moindre mesure Stenotrophomonas. Les pathogènes associées à la mucoviscidoses sont retrouvé et ont tendance à être dominant sans que les patients soient en exacerbations. Staphylococcus, Haemophilus sont les plus fréquent. Stenotrophomonas et Achromobacter sont plus rarement retrouvé. Burkholderia est rare et peu abondant. Malgré avoir sortie de l'étude tous les patients positifs sur la culture, On retrouve pseudomonas dans 53 échantillons sans jamais être dominant. Ceci confirme la faible sensibilité de la culture et montre l'intérêt d'autre méthode de diagnostic comme la PCR [10]. On retrouve également des fortes abondances en anaérobies avec Prevotella et Veillonella. La corrélation forte retrouvé entre ces deux genre a été retrouvé dans l'étude de [10]. Les résultats longitudinaux nous montre la dynamique et la complexité du microbiote respiratoire. Aucun des donnée récolté n'a mis en évidence un groupe particulier de microbiote. L'âge n'a pas été corrélé à la présence d'haemophilus, ni à l'augmentation de diversité évoqué dans plusieurs papier. Probablement car beaucoup de patient avaient le même âge. Les categories Free et Never n'a pas mis en évidence des groupes de microbiote distinct. Cependant les microbiotes des patients Never semble plus diversifié. En revanche pour un même patient, on retrouve des échantillons avec un microbiote très proche. Ceci nous amène à penser, comme l'article de [10], que le microbiote dépend avant tout de l'individu. Et peut être de son génome, son histoire et de son immunité. L'analyse multivarié s'explique principalement par les grandes fluctuations d'abondance de Streptococcus, Neisseria et Haemophilus. En les ignorant, il reste une population plus stable et de plus faible abondance composé de Gemella, Granulicatella, Oribacterium, Veillonella, Fusobacterium, Leptotrichia et Capnocytophaga. Cette architecture évoque le modèle déjà décrit par [10]. Celui ci propose de décrire le microbiote pulmonaire comme la somme des populations entre les territoires sains et inflammatoire.

6.3 Limite de l'étude

Plusieurs point sont à prendre en considération dans l'analyse des résultats. Premièrement, les ADN séquencé proviennent d'expectoration bronchique ayant traversé



FIGURE 31 – Evolution des abondances par genre

les voies aériennes supérieures. Une contamination est toujours possible, malgré les précautions prises. L'hétérogénéité du poumon pourrait également expliquer une différence entre deux échantillons. Il a cependant été montré [10] sur des prélèvements insitu que le microbiote pulmonaire est homogène. Ensuite la stratégie 16S n'est pas parfaite. Certaines espèces peuvent ne pas être détectées par l'amplification comme la montre l'étude de [10]. De plus, cette stratégie ne fait pas la distinction entre les ADNs de bactéries vivantes ou mortes. Enfin, notre étude ne prend pas en considération le virome et le mycobiome. Il est fort probable que ce sont les virus la cause des exacerbations en faisant le lit des surinfections bactériennes.

7 Conclusion

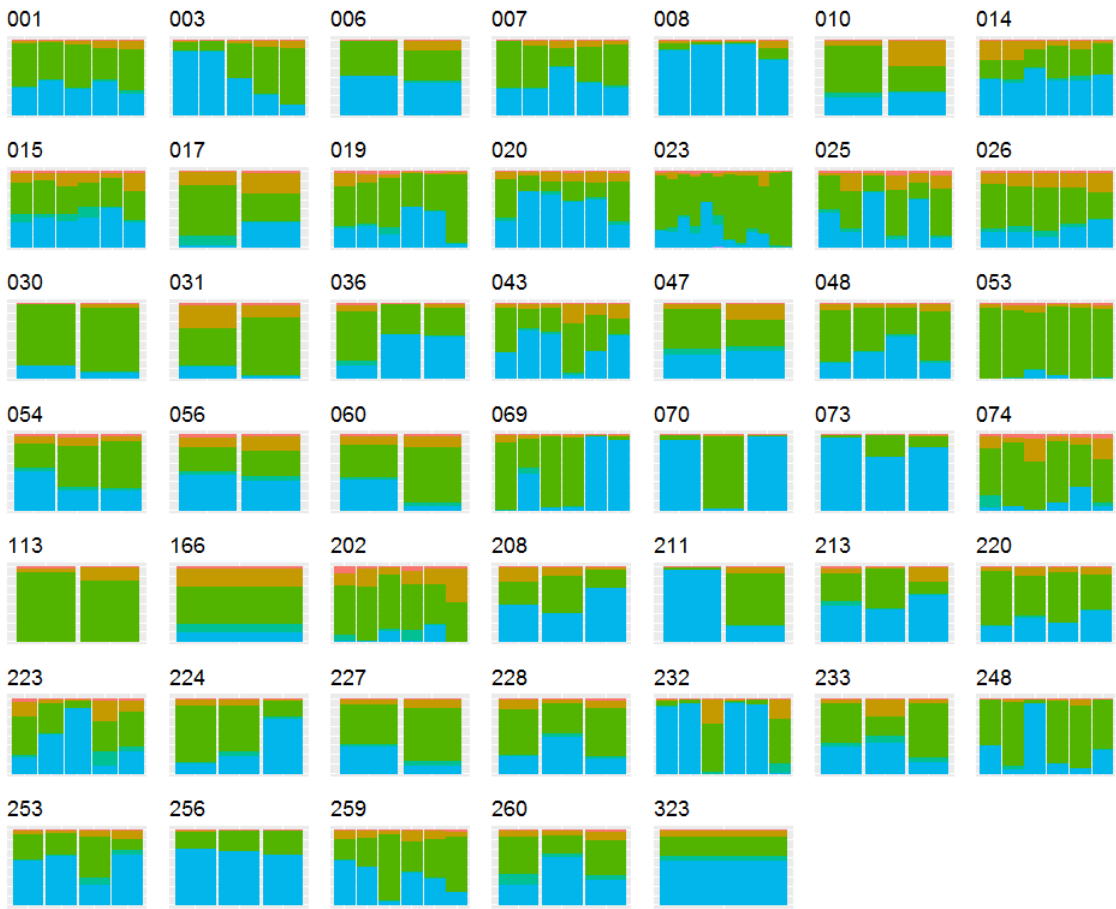


FIGURE 32 – Evolution des abondances par phylum

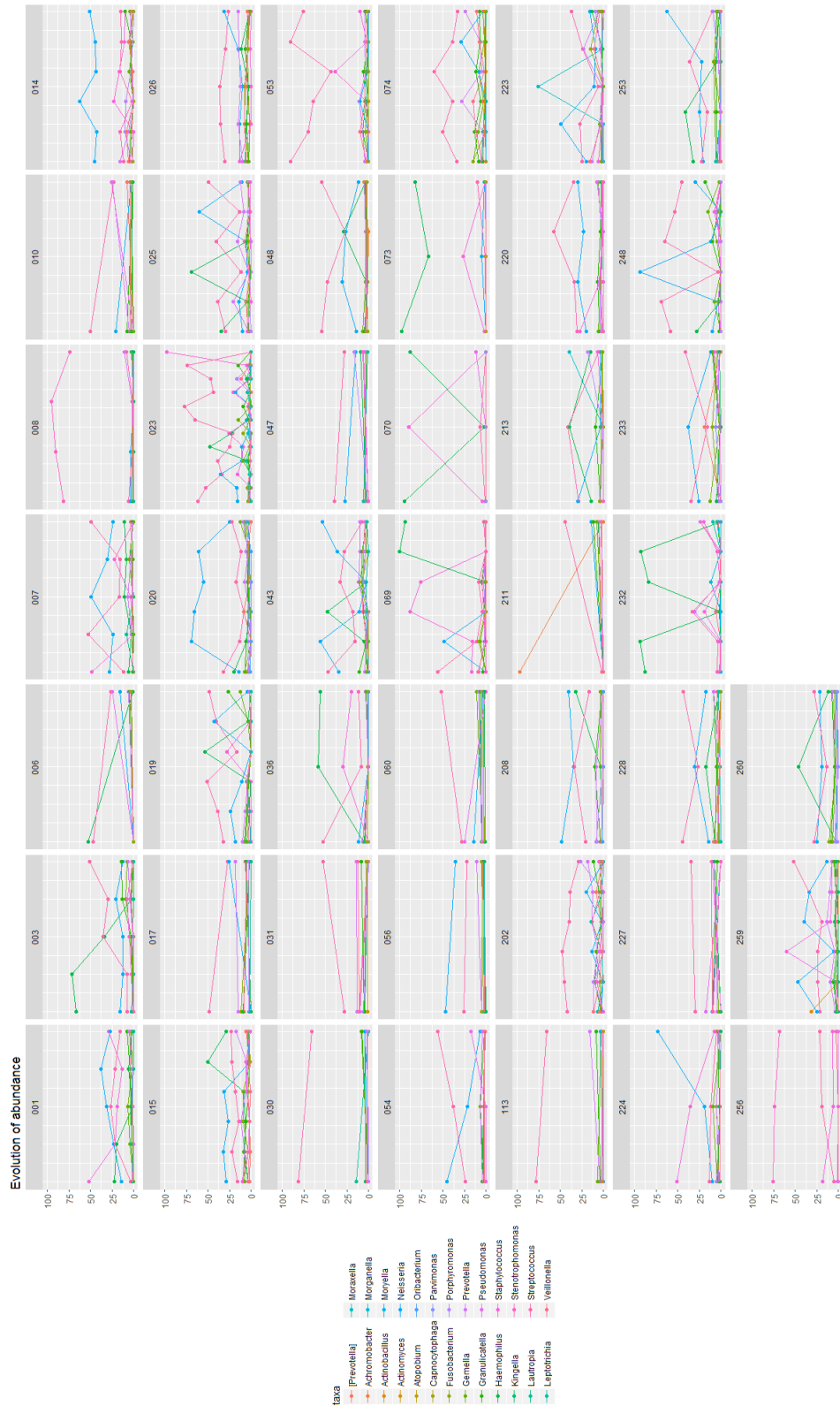


FIGURE 33 – Evolution des abondances par genre

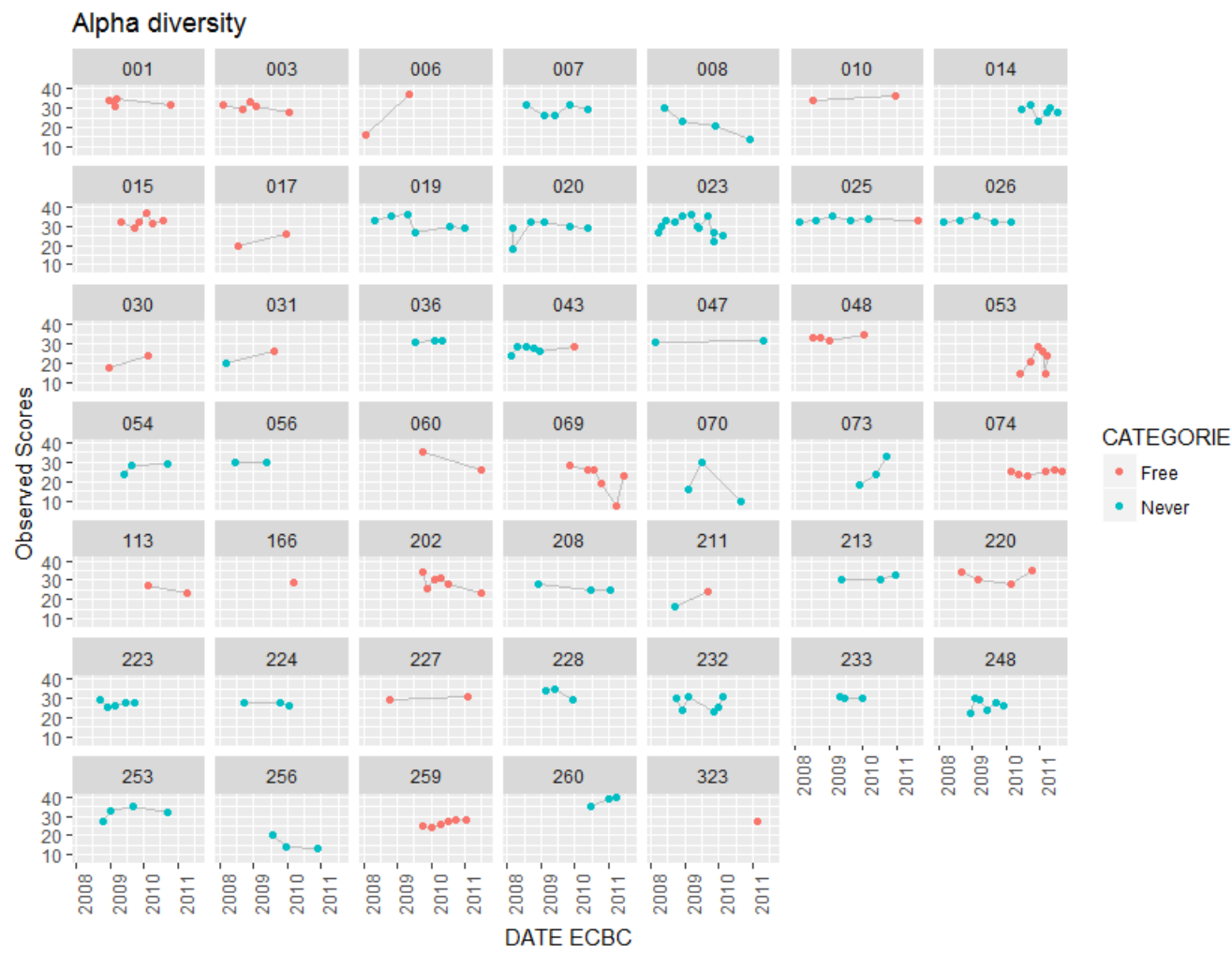


FIGURE 34 – Evolution du nombre d’espèces par patient en fonction du temps

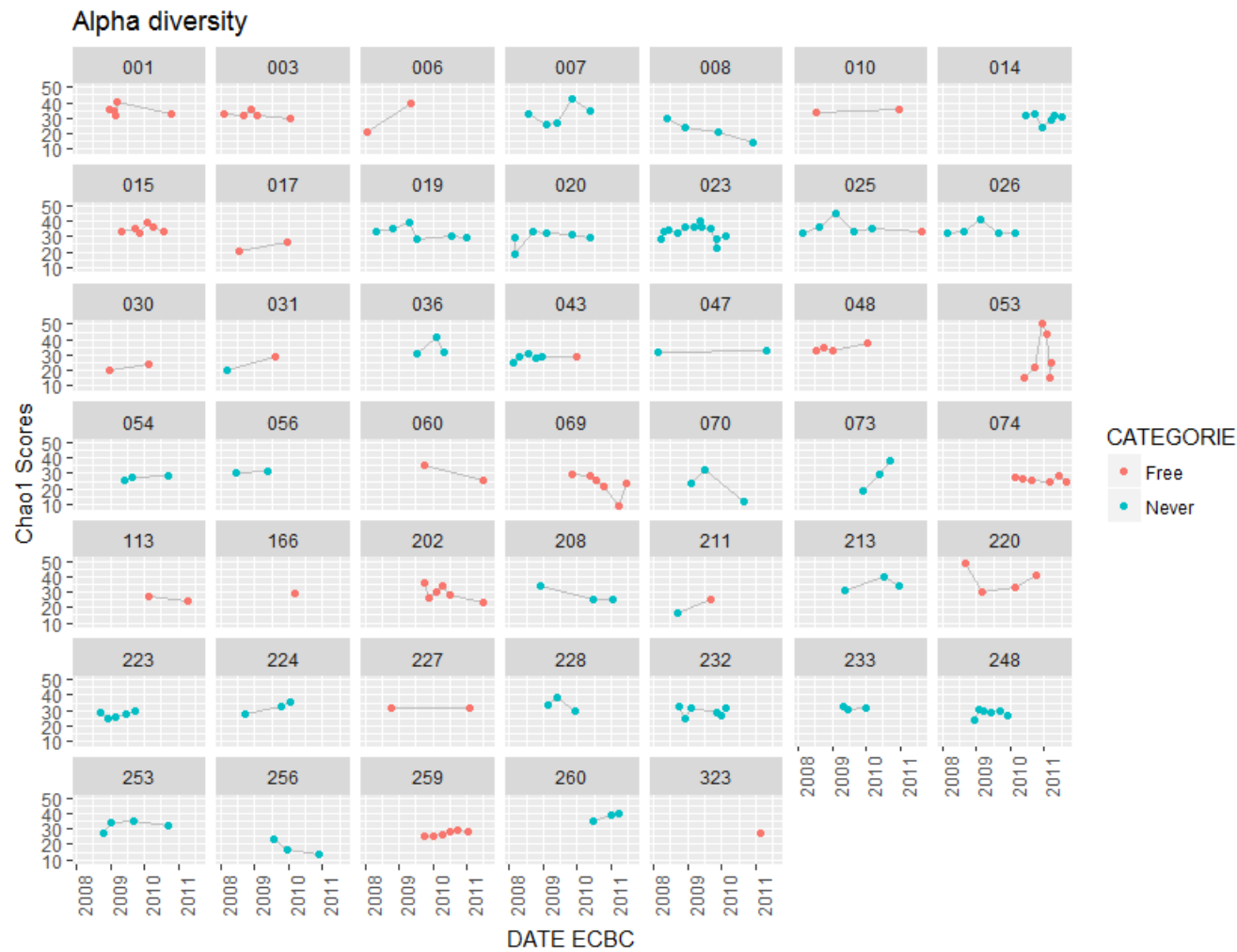


FIGURE 35 – Evolution de l'indice de Chao1 par patient en fonction du temps

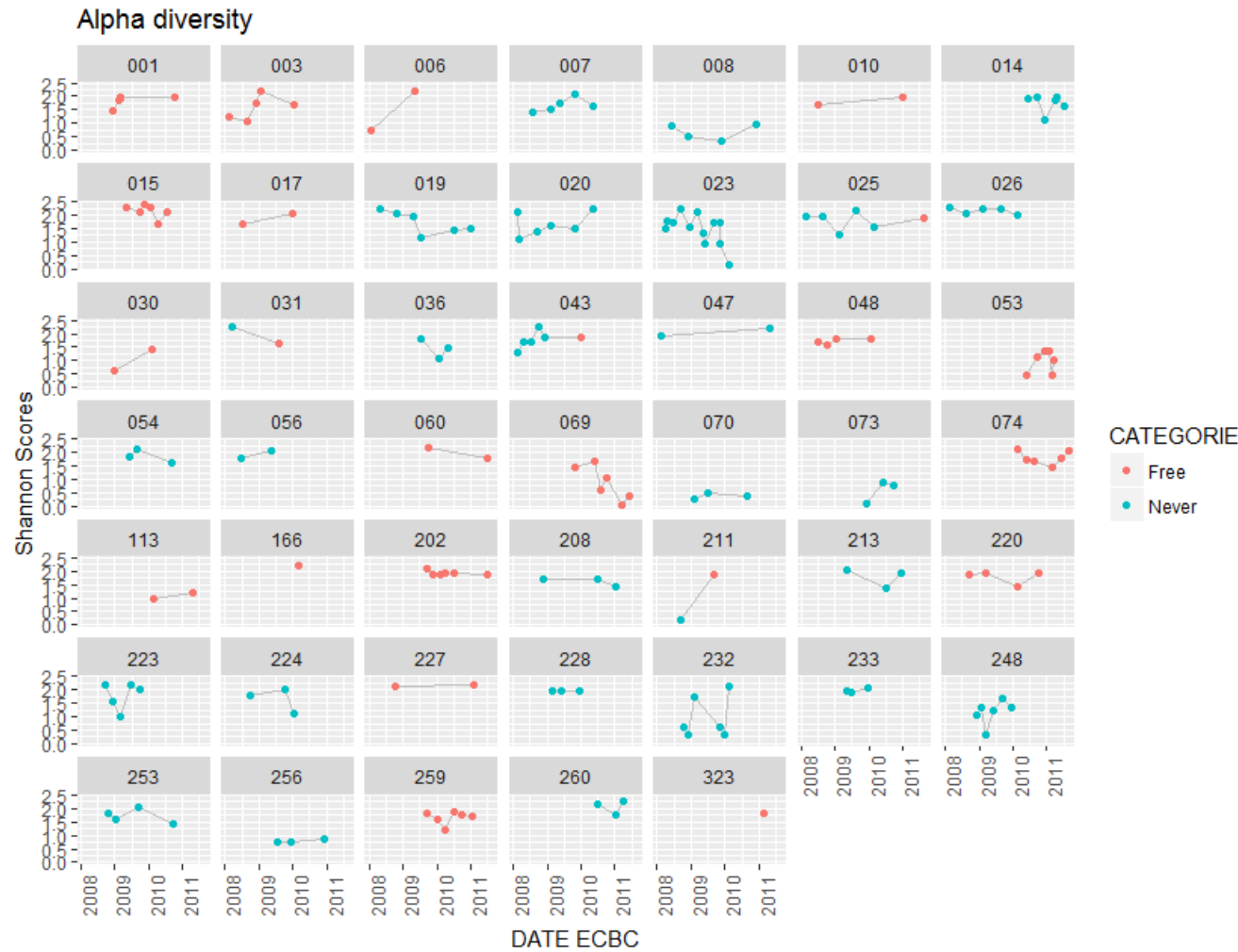


FIGURE 36 – Evolution de l'indice de Shanon par patient en fonction du temps

Références

- [1] Jean-François Bach. The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. *New England Journal of Medicine*, 347(12) :911–920, sep 2002.
- [2] James M Beck, Vincent B Young, and Gary B Huffnagle. The Microbiome of the Lung.
- [3] Lars Bode. Human milk oligosaccharides : every baby needs a sugar mama. *Glycobiology*, 22(9) :1147–62, sep 2012.
- [4] J Gregory Caporaso, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D Bushman, Elizabeth K Costello, Noah Fierer, Antonio Gonzalez Peña, Julia K Goodrich, Jeffrey I Gordon, Gavin A Huttley, Scott T Kelley, Dan Knights, Jeremy E Koenig, Ruth E Ley, Catherine A Lozupone, Daniel McDonald, Brian D Muegge, Meg Pirrung, Jens Reeder, Joel R Sevinsky, Peter J Turnbaugh, William A Walters, Jeremy Widmann, Tanya Yatsunenko, Jesse Zaneveld, and Rob Knight. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5) :335–336, may 2010.
- [5] Jane C Davies. Pseudomonas aeruginosa in cystic ®brosis : pathogenesis and persistence THE MOLECULAR BASIS OF CYSTIC FIBROSIS.
- [6] Robert P Dickson, John R Erb-Downward, and Gary B Huffnagle. The Role of the Bacterial Microbiome in Lung Disease.
- [7] Robert P Dickson and Gary B Huffnagle. The Lung Microbiome : New Principles for Respiratory Bacteriology in Health and Disease.
- [8] Jonathan Eisen. What does the term microbiome mean ? And where did it come from ? A bit of a surprise. <http://www.microbe.net/2015/04/08/what-does-the-term-microbiome-mean-and-where-did-it-come-from-a-bit-of-a-surprise>.
- [9] Genet.sickkids.on.ca. Cystic Fibrosis Mutation Database.
- [10] Wing Ho Man, Wouter AA de Steenhuijsen Piters, and Debby Bogaert. The microbiota of the respiratory tract : gatekeeper to respiratory health. *Nature Publishing Group*, 2017.
- [11] J. Koster and S. Rahmann. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19) :2520–2522, oct 2012.
- [12] Muriel Le Bourgeois and Stéphanie Vrielynck. Infection bronchopulmonaire dans la mucoviscidose.

- [13] Florence Le Gall, Rozenn Le Berre, Sylvain Rosec, Jeanne Hardy, Stéphanie Gouriou, Sylvie Boisramé-Gastrin, Sophie Vallet, Gilles Rault, Christopher Payan, and Geneviève Héry-Arnaud. Proposal of a quantitative PCR-based protocol for an optimal *Pseudomonas aeruginosa* detection in patients with cystic fibrosis.
- [14] Jean Guy LeBlanc, Christian Milani, Graciela Savoy de Giori, Fernando Sesma, Douwe van Sinderen, and Marco Ventura. Bacteria as vitamin suppliers to their host : a gut microbiota perspective. *Current Opinion in Biotechnology*, 24(2) :160–168, apr 2013.
- [15] Tim W.R. Lee, Keith G. Brownlee, Steven P. Conway, Miles Denton, and James M. Littlewood. Evaluation of a new definition for chronic *Pseudomonas aeruginosa* infection in cystic fibrosis patients. *Journal of Cystic Fibrosis*, 2(1) :29–34, mar 2003.
- [16] Nuala A. O’Leary, Mathew W. Wright, J. Rodney Brister, Stacy Ciufu, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M. Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wratko Hlavina, Vinita S. Joardar, Vamsi K. Kodali, Wenjun Li, Donna Maglott, Patrick Masterson, Kelly M. McGarvey, Michael R. Murphy, Kathleen O’Neill, Shashikant Pujar, Sanjida H. Rangwala, Daniel Rausch, Lillian D. Riddick, Conrad Schoch, Andrei Shkeda, Susan S. Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E. Tully, Anjana R. Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J. Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D. Murphy, and Kim D. Pruitt. Reference sequence (RefSeq) database at NCBI : current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1) :D733–D745, jan 2016.
- [17] P Plésiat. Quels critères microbiologiques pour définir une colonisation ou une infection à *Pseudomonas aeruginosa* ?
- [18] Junjie Qin, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Florence Levenez, Takuji Yamada, Daniel R. Mende, Junhua Li, Junming Xu, Shaochuan Li, Dongfang Li, Jianjun Cao, Bo Wang, Huiqing Liang, Huisong Zheng, Yinlong Xie, Julien Tap, Patricia Lepage, Marcelo Bertalan, Jean-Michel Batto, Torben Hansen, Denis Le Paslier, Allan Linneberg, H. Bjørn Nielsen, Eric Pelletier, Pierre Renault, Thomas Sicheritz-Ponten, Keith Turner, Hongmei Zhu, Chang Yu, Shengting Li, Min Jian, Yan Zhou, Yingrui Li, Xiuqing Zhang, Songgang Li, Nan Qin, Huanming

- Yang, Jian Wang, Søren Brunak, Joel Doré, Francisco Guarner, Karsten Kristiansen, Oluf Pedersen, Julian Parkhill, Jean Weissenbach, Maria Antolin, François Artiguenave, Hervé Blottiere, Natalia Borruel, Thomas Bruls, Francesc Casellas, Christian Chervaux, Antonella Cultrone, Christine Delorme, Gérard Denariáz, Rozenn Dervyn, Miguel Forte, Carsten Friss, Maarten van de Guchte, Eric Guedon, Florence Haimet, Alexandre Jamet, Catherine Juste, Ghalia Kaci, Michiel Klee-rebezem, Jan Knol, Michel Kristensen, Severine Layec, Karine Le Roux, Marion Leclerc, Emmanuelle Maguin, Raquel Melo Minardi, Raish Oozeer, Maria Rescigno, Nicolas Sanchez, Sebastian Tims, Toni Torrejon, Encarna Varela, Willem de Vos, Yohanan Winogradsky, Erwin Zoetendal, Peer Bork, S. Dusko Ehrlich, and Jun Wang. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285) :59–65, mar 2010.
- [19] Paul M. Quinton. The neglected ion : HCO_3^- . *Nature Medicine*, 7(3) :292–293, mar 2001.
- [20] Registredelamuco.org. Registre Francais de la mucovidoses 2015.
- [21] R. Ruimy and A. Andremont. Quorum-sensing chez *Pseudomonas aeruginosa* : Mécanisme moléculaire, impact clinique, et inhibition, 2004.
- [22] Patrick D Schloss, Sarah L Westcott, Thomas Ryabin, Justine R Hall, Martin Hartmann, Emily B Hollister, Ryan A Lesniewski, Brian B Oakley, Donovan H Parks, Courtney J Robinson, Jason W Sahl, Blaz Stres, Gerhard G Thallinger, David J Van Horn, and Carolyn F Weber. Introducing mothur : open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75(23) :7537–41, dec 2009.
- [23] Ron Sender, Shai Fuchs, Ron Milo, T Lee, H Ahn, and S Baek. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, 14(8) :e1002533, aug 2016.
- [24] R. Vongthilath, C. Dehillotte, L. Lemonnier, B. Richaud-Thiriez, A. Guillien, B. Degano, P. Plésiat, and J.-C. Dalphin. Étude PYOnever : caractéristiques des adultes atteints de mucoviscidose indemnes de colonisation pulmonaire par *Pseudomonas aeruginosa*. *Revue des Maladies Respiratoires*, 34, 2017.
- [25] Ed Yong. *Moi, microbiote, maître du monde : Les microbes, 30 billions d'amis !* 2017.