

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

30 avril 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”.»



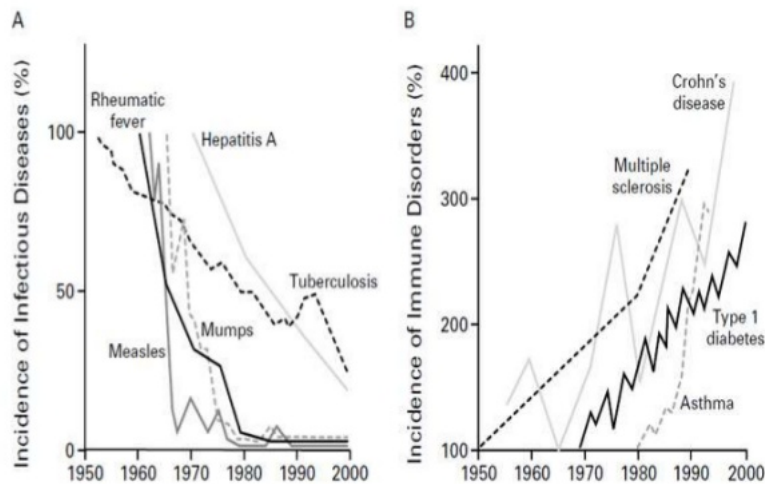
# Table des matières

<b>1</b>	<b>Avant-propos</b>	<b>3</b>
<b>2</b>	<b>Définition</b>	<b>6</b>
2.1	Termes utilisés en écologie . . . . .	6
2.2	Termes utilisés en bioinformatique . . . . .	8
<b>3</b>	<b>Introduction</b>	<b>1</b>
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 . . . . .	4
3.3.1	Méthode de prélèvement . . . . .	4
3.3.2	Séquençage haut débit . . . . .	5
3.4	Objectif de l'étude Mucobiome . . . . .	6
<b>4</b>	<b>Matériel et Méthodes</b>	<b>6</b>
4.1	Recueil des données . . . . .	6
4.2	Extraction de l'ADN . . . . .	8
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 . . . . .	10
4.5.3	Reversing : Séquence complémentaire . . . . .	11
4.5.4	Trimming : Suppression des primers . . . . .	11
4.5.5	dereplicating : Suppression des doublons . . . . .	11
4.5.6	Assignement taxonomique . . . . .	11
4.5.7	Analyse descriptive . . . . .	12
<b>5</b>	<b>Résultat</b>	<b>12</b>
5.1	Séquençage et pipeline . . . . .	12
5.1.1	Données de séquençage . . . . .	12
5.1.2	Qualité des reads . . . . .	12
5.1.3	Assignement taxonomique . . . . .	13
5.1.4	Qualité d'échantillonnage . . . . .	14
5.2	Composition du microbiote pulmonaire . . . . .	14
5.3	Dynamique du microbiote pulmonaire . . . . .	15

5.3.1	Évolution de l'alpha diversité . . . . .	15
5.3.2	Évolution des abondances . . . . .	18
5.4	Corrélation entre genre bactériens . . . . .	19
5.5	Beta Diversité . . . . .	19
<b>6</b>	<b>Discussion</b>	<b>21</b>
<b>7</b>	<b>Conclusion</b>	<b>21</b>

# 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 considérées 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 temps<sup>1</sup>.

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. L'étude des archées a par ailleurs apporté un éclairage nouveau sur l'origine des eucaryotes<sup>i</sup> et a fait faire un bond de géant à la biologie moléculaire<sup>ii</sup>.

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*<sup>8</sup>. Le vagin est colonisé essentiellement par *Lactobacillus* et la bouche principalement par *Streptococcus*<sup>8</sup>. La muqueuse intestinale est colonisée par une "flore" bactérienne ou microbiote dominée par les anaérobies pouvant contribuer jusqu'à 0.2 kg au poids corporel<sup>7</sup>. 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<sup>3;8</sup>. Il participe à la synthèse des vitamines essentielles (K, B12, B8)<sup>5;8</sup>. Il éduque notre système immunitaire et fait barrière à l'implantation de tout nouvel agent pathogène. Toute défaillance de notre microbiote ou *dysbiose*, peut avoir des conséquences pour notre santé. La liste des affections associées est longue<sup>8</sup>. 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 en dépit de lignes 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 humaines<sup>7</sup>. En associant les gènes bactériens, le génome d'un être humain passe de 23 000 à 3,3 millions de gènes<sup>6</sup> avec toute la complexité des interactions que cela engendre. Les scientifiques ont nommé *holobionte* cet écosystème vivant. L'ensemble

i. Le génome des archées est composé d'introns comme chez les eucaryotes

ii. La Taq polymérase est une enzyme d'archée qui résiste à de hautes températures utilisées dans les PCR

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 paries 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**<sup>4</sup> est l'ensemble des micro-organismes (bactéries, levures, champignons, virus) vivant dans un environnement donné.

**Le microbiome**<sup>4</sup> est l'ensemble des génomes des micro-organismes provenant d'un environnement donné. Dans la plus part des cas, le microbiome fait référence aux bactéries seuls. on parle alors de virome et de mycobiome pour les virus et champignons.

**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.

**Un OTU** (*Operational Taxonomic Unit*) est un terme utilisé en phylogénie pour désigner un groupe d'individus semblable. Un OTU est très souvent considéré comme le nom 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'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 l'expression en pourcentage de l'abondance absolu, afin de rendre les échantillons comparable entre eux.

**La dominance** caractérise un OTU majoritaire dans un échantillon.

**La table des OTU** est 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'échantillon 1 contient 68% de *Staphylococcus Aureus*.

	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

**La diversité alpha** est une mesure de biodiversité au sein d'un échantillon. Elle correspond à l'étude d'une colonne dans 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'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 deux ou trois dimensions.

**La richesse** est le nombre d'espèces présent dans un échantillon. Les deux échantillons suivant ont une richesse de 2.

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

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

**L'uniformité / équitabilité** indique si les espèces d'un échantillon sont réparties uniformément. L'uniformité du premier échantillon est plus grande que la seconde

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 (*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 singleton et **b** le nombre de doubleton

**L'indice de Shannon** est un indicateur évaluant à la fois la richesse et l'uniformité 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). Si seulement deux couleurs sont présente, une petite poigné de bille suffit pour évaluer la diversité. En revanche, avec plus de couleur il faudra une poignées plus importante. La courbe de raréfaction permet d'évaluer si cette poigné de bille est suffisamment représentative. Pour générer cette courbe, des groupes de reads de taille croissante (1..n) sont tirés aléatoirement depuis un échantillon. Le groupe est reporté sur l'axe X et la richesse du groupe est reporté sur l'axe Y. Une courbe s'aplatissant indique une bonne profondeur de séquençage ( Figure 4 ).

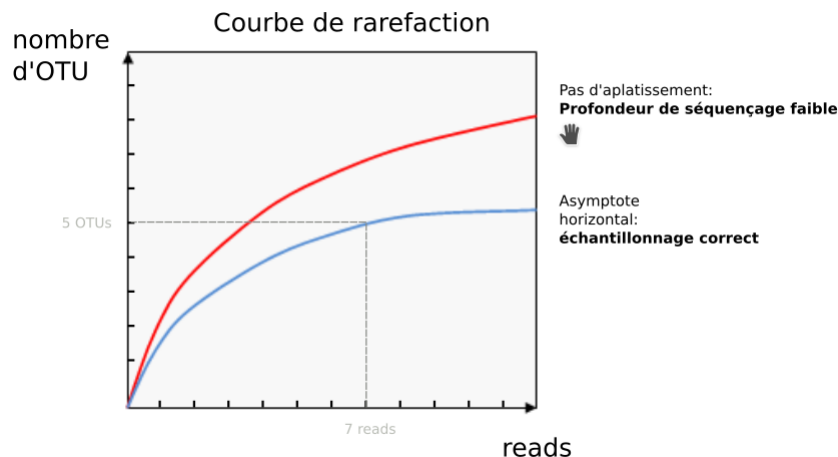


FIGURE 4 – Exemple de courbes de raréfactions. La courbe bleu témoigne d'un bon échantillonnage en s'aplatissant précocement. Plus de reads ne permettrait pas de découvrir de nouveau OTU contrairement à la courbe rouge.

## 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, qui permettent de générer un graphique des étapes (direct Acyclique Graph) et de trouver la meilleure façon d'optimiser les tâches entre les différents cuisiniers (processeur).

**NGS** ( Next Generation sequencing ) désigne l'ensemble des séquenceurs de nouvelle génération permettant de séquencer un nombre très important de fragment 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** Le score de qualité Phred 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.

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

**Un fichier fasta** contient uniquement les séquences des reads.

**Un fichier biom** contient la table des OTUs

## 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 5400 [1]. La Bretagne est la région la plus touchée avec une prévalence de 1/3000 [1]. La loi de Hardy Weinberg estime qu'en Bretagne 1 patient sur 25 est porteur de la mutation à l'état hétérozygote [1]. Cette haute prévalence s'explique probablement par un effet fondateur associé à un avantage sélectif pour les individus porteurs de l'allèle muté. <sup>iii</sup>

Le gène CFTR impacté se situe sur le chromosome 7 en position q31.2. Il est constitué de 27 exons pour 250 188 [1] paires de bases. Il code pour un canal chlore AMP dépendant permettant les échanges des ions chlorures au niveau des membranes cellulaires [1]. Il est également impliqué dans le transport du thiocynate (SCN-) et des bicarbonates (HCO<sub>3</sub>-) [1].

On dénombre à ce jour 2017 mutations mises en cause dans la mucoviscidose [1]. La perte d'une phénylalanine en position 508 par délétion du triplet c.1521-1523delCTT (anciennement  $\Delta F508$ ) cause à elle seule 80% des mucoviscidoses. 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 à l'origine d'une déshydratation du mucus le rendant plus visqueux et empêche les cils bronchiques de jouer leurs rôles. [1]

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. Le dépistage prénatal basé sur l'ADN circulant est actuellement à l'étude [1]. Le traitement repose avant tout sur une prise en charge respiratoire (kinésithérapie, dorasse, antibiothérapie). Les recherches en thérapies génétiques sont encourageantes [2]. L'Ivacaftor est le seul traitement à ce jour qui agit directement sur le CFTR. Mais concerne uniquement certaine mutation rare comme la G551D. [2] La greffe pulmonaire est le dernier recours.

---

iii. Plusieurs hypothèses ont été proposées, notamment lors des grandes épidémies de choléra en diminuant les pertes hydriques. D'autres suggèrent qu'il s'agit d'une pléiotropie antagoniste. [1]

### 3.1.2 Une maladie infectieuse

L'atteinte pulmonaire dans la mucoviscidose est caractérisée par des réactions inflammatoires qui dégrade progressivement la fonction respiratoire. Plusieurs pathogènes sont impliqués. Chez les jeunes enfants, *Haemophilus influenza* et *Staphylococcus Aureus* sont le plus souvent responsable. *Burkholderia Cepacia* et *Stenotrophomonas Maltophilia* sont retrouvés parmi les sujets plus âgé. 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 un germe de l'environnement rarement retrouvé parmi les patients sains[@]. En revanche, dans la mucoviscidose, il est mis en évidence chez 60%[@] des patients jeunes, et plus de 90% des patients adultes[@]. 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 antibiotiques[@]. 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*[@]. Un processus dans lequel chaque bactérie communique avec ses voisines par des signaux. Le génome de *Pseudomonas Aeruginosa* devient aussi hyperpermutable afin de présenter une plus grande diversité génétique au regard de la sélection naturelle<sup>iv</sup>.

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 sensibilité [10]. La PCR ciblée associant les protéines bactériennes *OPRL1*, *GYRB1* et *ECFX1* s'est montrée plus sensible et plus spécifique que la culture[10].

En pratique, la colonisation chronique est définie lorsque 3 expectorations sont rendues positives en culture, successivement au cours d'un suivi mensuel ou bimensuel[10]. Une autre classification, celle de Lee a montré une forte liaison clinico-biologique. Elle est composée de 4 groupes :

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

---

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

**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

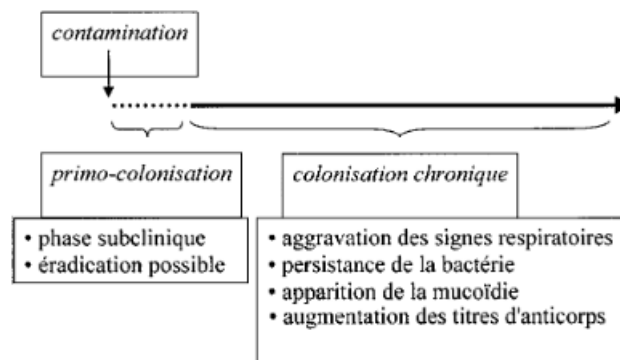


FIGURE 5 – Infection pulmonaire à pyo dans la mucoviscidose. ref [1]

### 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 mettre en évidence le microbiote pulmonaire[ref Host-microorganism 1-3]. L'arbre respiratoire chez le fœtus 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. Il est constitué d'une flore dynamique provenant de l'air ambiant, des voies supérieures, mais aussi du tube digestif par des micro-aspirations.[@] Une étude à montrer une concordance entre le microbiote digestif et respiratoire[@ Dixon page4]. En raison du flux d'air bidirectionnel, le microbiote respiratoire est aussi

plus dynamique que le microbiote digestif. 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 directe 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... 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 ..... [1]. 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* (*Haemophilus*, *Moraxella* 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'autres agents pathogènes. [1]

### 3.3 Exploration du microbiote pulmonaire

#### 3.3.1 Méthode de prélèvement

Le microbiote respiratoire est exploré en séquençant l'ensemble des ADN présents dans un échantillon. Toutes les méthodes de recueils sont possibles, mais les prélèvements protégés (combicath, LBA) sont recommandés afin d'éviter une contamination par les voies supérieures. Dans le cas contraire (ECBC) on peut évaluer la qualité du prélèvement en comptant le nombre de cellules épithéliales (normalement bas) et de polynucléaires (normalement haut) dans le poumon. La meilleure méthode de prélèvement

étant le prélèvement in situ réalisable lors des greffes pulmonaires.

### 3.3.2 Séquençage haut débit

Grâce à son haut débit, Le séquençage de nouvelle génération permet de séquencer l'ensemble des ADN présent dans un échantillon respiratoire et ainsi déterminer sa composition en bactérie. 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 capillaire sur la machine. À l'inverse, un séquenceur de nouvelle génération lit des fragments plus courts de l'ordre de 150pb. Mais cette lecture peut être parallélisé jusqu'à 20 billions de fois en un seul run sur un illumina Novaseq.

2 stratégies de séquençages sont utilisées en écologie microbienne :

**La stratégie shogun** 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. Il s'agit d'amplifier un gène uniquement bactérien et suffisamment variable pour discriminer une espèce. Pour les bactéries, il s'agit 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 7). L'alignement des séquences d'ARN 16S entre plusieurs espèces met en évidence des régions constantes et 9 régions variables (Figure 6). Les régions constante permette de concevoir des amorces s'hybridant sur tous les ADNs bactériens. Les régions variables apportent la spécificité taxonomique permettant d'identifier l'espèce bactérienne.

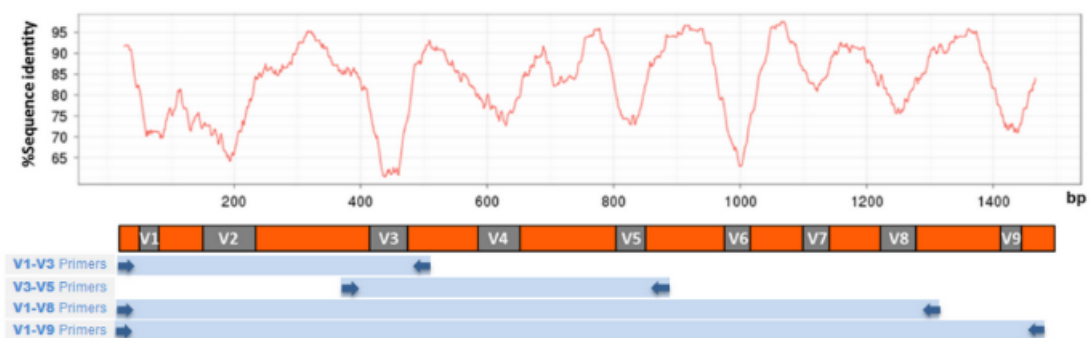


FIGURE 6 – région constante et variable de l'ARN 16S

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



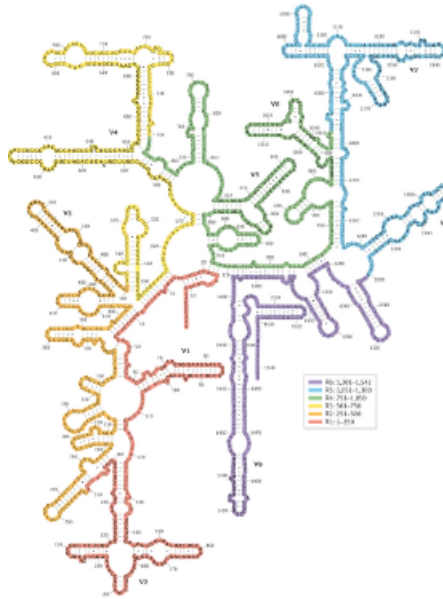


FIGURE 7 – Structure secondaire de l'ARN 16S

génétiques horizontaux amènent à dissocier l'espèce de sa fonction. 2 bactéries d'une même espèce peuvent avoir des fonctions différentes. L'inférence fonctionnelle à partir de la deuxième stratégie est possible mais déconseillé[@]. 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.

### 3.4 Objectif de l'étude Mucobiome

L'objectif de notre étude est de savoir si le microbiote respiratoire influence la primo colonisation à *Pseudomonas Aeruginosa*. Pour cela, nous avons suivi pendant 3 ans, une cohorte de 47 patients atteints de mucoviscidose sans infection chronique. L'exploration de leurs microbiotes par la stratégie 16S a été réalisée à partir d'expectorations bronchiques obtenues dans le cadre de leurs suivis. 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.

## 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 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 génotype CFTR et un test à la sueur positif. Les transplantés ont été exclus de l'étude. Une culture positive à *Pseudomonas Aeruginosa* était un critère de non-inclusion. Si pendant l'étude, une culture revenait positive à *Pseudomonas Aeruginosa*, le patient était sorti de l'étude pour être réinclus 1 an après en l'absence de colonisation chronique. 15 patients ont été ainsi réinclus. Chaque patient a été répertorié dans la catégorie Free ou Never (Lee et al 2003). D'autres données ont également été recueillies ( Figure 8 ). Au total , 188 échantillons ont été récoltés, soit en moyenne 4 échantillons par patients. Pour chaque échantillon, une culture a été réalisée en suivant les procédures standards [10]. Une qPCR ciblant *Pseudomonas Aeruginosa* a parallèlement été réalisée en combinant les marqueurs gyrB/ecfX conçus au laboratoire[10].

	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 8 – Données associées par échantillons ( à refaire par patients)

## 4.2 Extraction de l'ADN

Les échantillons ont été liquéfiés avec du Dithiotrétiol. Les protéines ont été dégradées avec une Protéine kinase. 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 via QUIAamp DNA Minikit ( Quagen). Les extraits d'ADN ont été envoyés pour séquençage par un prestataire GATC.

## 4.3 Séquençage

La librairie<sup>v</sup> 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. Un couple de séquence chevauchant de 300pb est produit (Figure 9). Après fusion du couple, le séquençage permet de lire les 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 séquençés sur 2 runs pour produire 188 x 2 fichiers fastq.

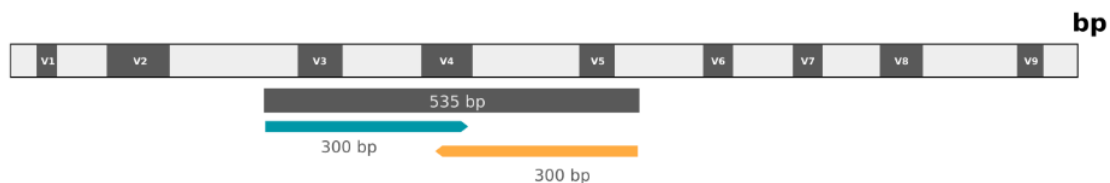


FIGURE 9 – le couple de séquence de 300pb permet de recouvrir l'ensemble de la région V3-V5

## 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** ou **MOTHUR**, le pipeline *mucobiome* est spécialisé dans l'analyse des données 16S. Il est également plus rapide en raison d'un très haut niveau de parallélisation permis grâce à **Snakemake**. 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

v. l'ensemble des fragments d'ADN à séquençer

séquençage et produit un fichier BIOM contenant la table des OTUs. Les figures 10 et 11 illustrent toutes les étapes du pipeline et sont décrites ci-dessous.

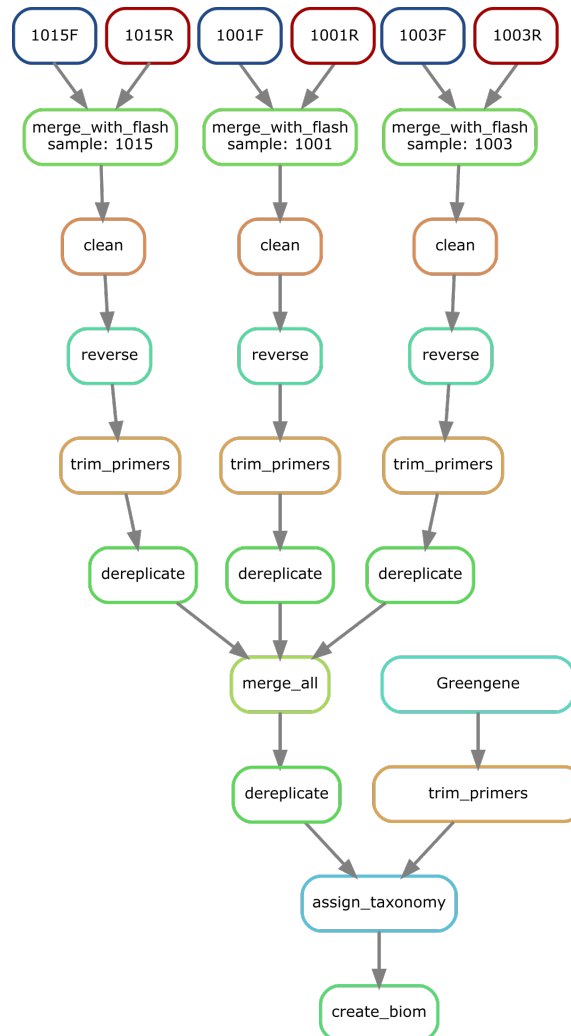


FIGURE 10 – Graphe du pipeline simplifié sur 3 échantillons 1015, 1001 et 1003.

**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 assignent** : Les séquences sont alignées sur la base de données greengene. **create\_biom** : la table des OTU est créée

## 4.5 Étape du pipeline

### 4.5.1 Merging : Fusions des reads

*2 fastq en entrée 1 fastq en sortie.*

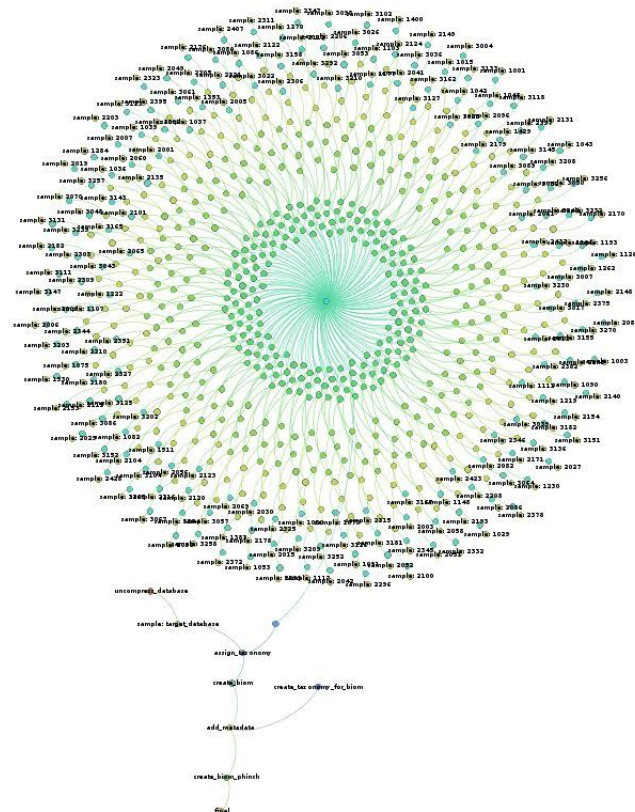


FIGURE 11 – le graphe du pipeline sur l'ensemble des échantillons

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 FastQt<sup>2</sup> avant et après fusions.

### 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 FastQt<sup>2</sup> après le filtrage.

#### 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 sauvegardées 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 dupliqué 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*

#### 4.5.6 Assignement taxonomique

*2 fasta en entrée, 1 fichier biom en sortie.*

L'assignement 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

<code>&gt;sample1</code>	
ACGTTTTT	
<code>&gt;sample1</code>	<code>&gt;sample1;size=3</code>
ACGTTTTT	ACGTTTTT
<code>&gt;sample1</code>	<code>&gt;sample1;size=1</code>
GTAGAGT	GTAGAGT
<code>&gt;sample1</code>	
ACGTTTTT	

*avant dereplication      après dereplication*

FIGURE 12 – Exemple de déréduplication d'un fichier fasta

fasta contenant 1 262 986 séquences et 203 452 OTUs.

L'autre stratégie d'assignement *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'assignement taxonomique a été réalisé avec vsearch et sa fonction `-usearch_global`.

#### 4.5.7 Analyse descriptive

L'analyse de la table des OTUs a été réalisée avec R et le package phyloseq. Les OTUs ont été regroupé agrégé par genre. la diversité globale et individuelle a été évalué à l'aide de plusieurs graphiques. L'ensemble des résultats est disponible à cette adresse : [http ://...](http://...)

## 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 pairs par échantillons. La longueur des reads pour chaque fichier est de 301 paires de bases. Au total, 115 002 297 reads ont été produits sur 2 runs MiSeq (Figure 13). 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.

#### 5.1.2 Qualité des reads

La figure 14 montre le profil de qualité des reads d'un échantillon avant fusion. Une baisse du score de qualité importante est observée à hauteur du 250ème nucléotide. Les 188 échantillons présentaient le même profil. La figure 15 représente les scores de

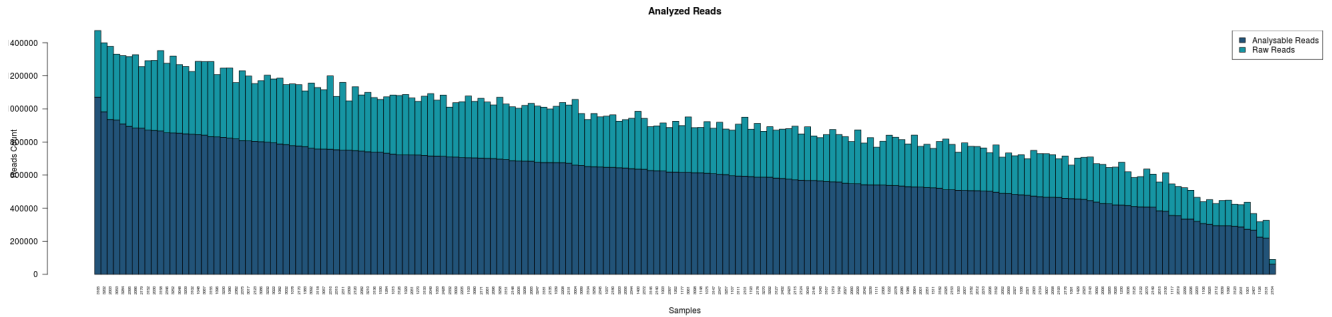


FIGURE 13 – Nombre de reads par échantillons

qualités d'un échantillon après fusions et après filtrage des reads de mauvaise qualité. Le chevauchement des reads pairs n'étant pas suffisant, une baisse de qualité est observée au milieu, à hauteur du chevauchement. Après filtrage la qualité médiane au centre de la séquence remonte au dessus de 20.

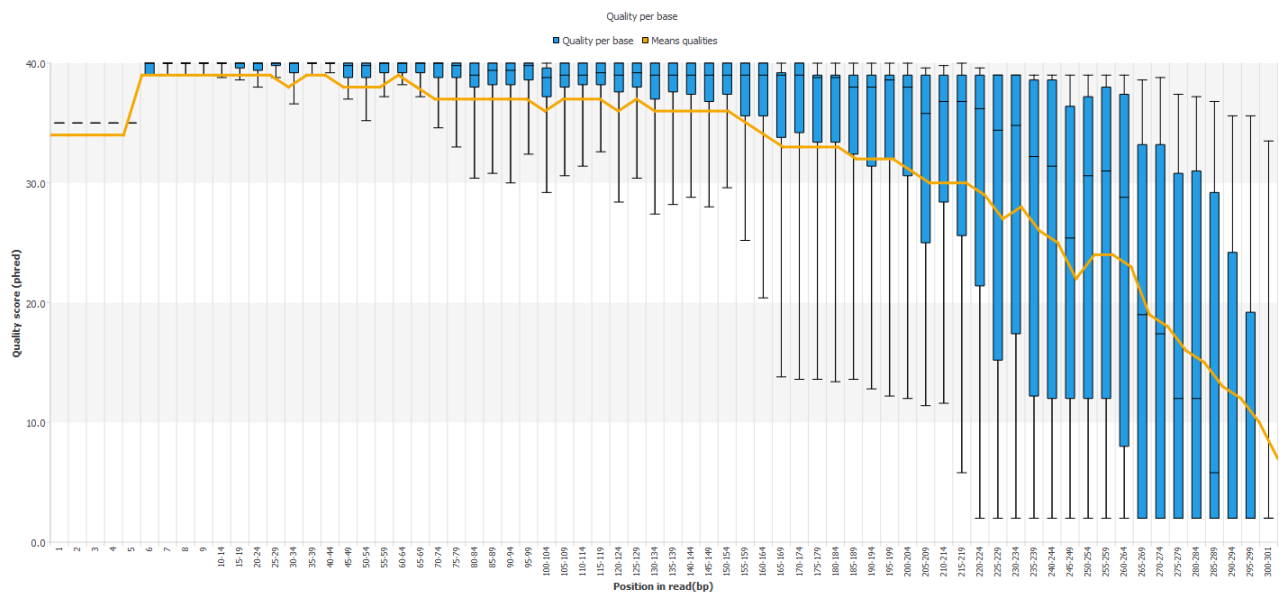


FIGURE 14 – Qualité par nucléotide des reads forward de l'échantillon 1003. **Axe X** : la position sur le read. **Axe Y** : La distribution des qualités

### 5.1.3 Assignment taxonomique

Après traitement des reads, seulement 49,24 % (Figure 16) des reads sont analysable avec des bornes allant de 37,30% à 61,13%. La majorité des reads ayant été supprimé lors de l'étape de filtrage. Au total, 99.88% des reads analysables ont reçu une assignation taxonomique. La figure 16 montre le nombre de reads assignés par genre bactérien.

Au total le pipeline mucobiome s'est exécuté en 1 h 29 sur 40 cœurs et 20 gigaoctets



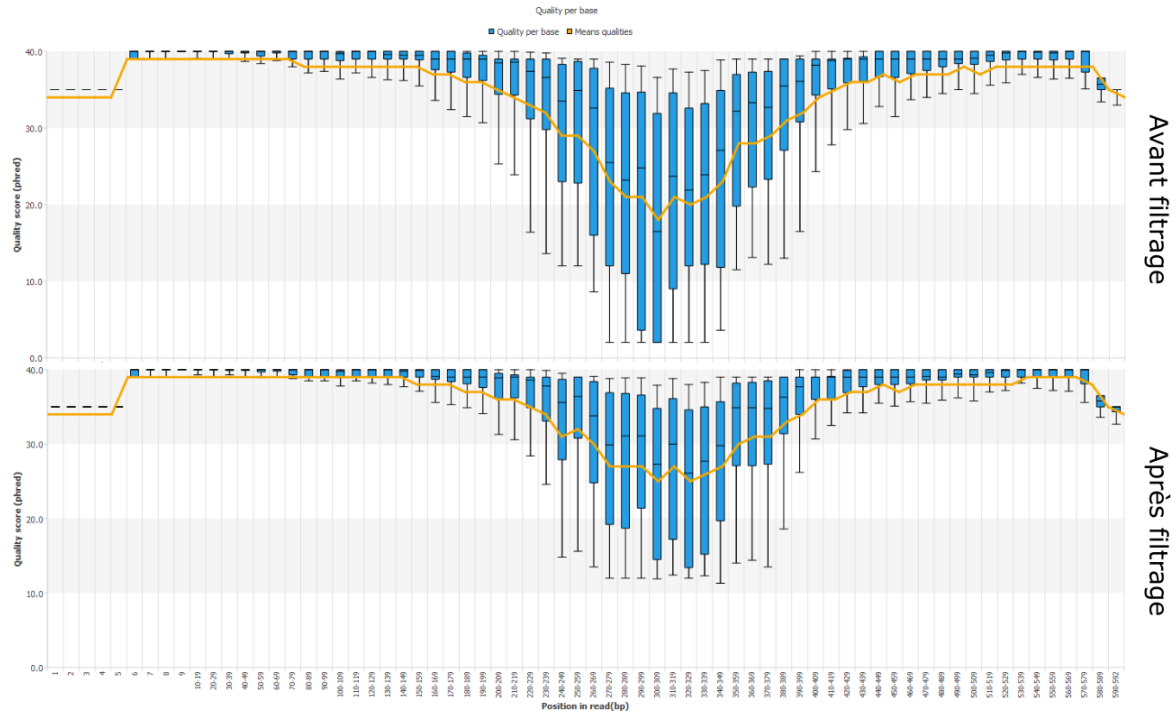


FIGURE 15 – Qualité par nucléotide des reads fusionnés pour l'échantillon 1003

de mémoire contre 21h de calcul sans l'optimisation par déréduplication.

#### 5.1.4 Qualité d'échantillonnage

La figure 17 montre les courbes de raréfaction pour les 188 échantillons. Dans l'ensemble elles s'aplatissent précocement, témoignant d'un très bon niveau échantillonnage.

## 5.2 Composition du microbiote pulmonaire

54 genres ( Figure 16 ) et 7 phylums ( Figure 18 ) bactériens sont retrouvés dans l'ensemble des échantillons analysés. Les trois phylums majoritaires sont *Firmicute* ( 42.59%) et *Proteobacteria* ( 31.48%). Parmi les *Firmicutes* majoritaires, on retrouve *Streptococcus* et *Staphylococcus*. Chez les *Protéobacteria* *Neisseria* et *Haemophilus* sont les plus abondantes. Le tableau de la figure 19 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 prévalent, car présents dans plus de 185 échantillons. D'autres sont dominantes, c'est-à-dire qu'ils représentent plus de 90% du microbiote pulmonaire dans certain échantillon. Il s'agit de *Streptococcus*, *Neisseria*, *Haemophilus* et *Staphylococcus*. *Sténotropomonas* et *Achromobacter* sont retrouvées dans 64 et 8 échantillons respectivement. *Pseudomonas* est retrouvé dans 53 échantillons, dont un, au moins avec une

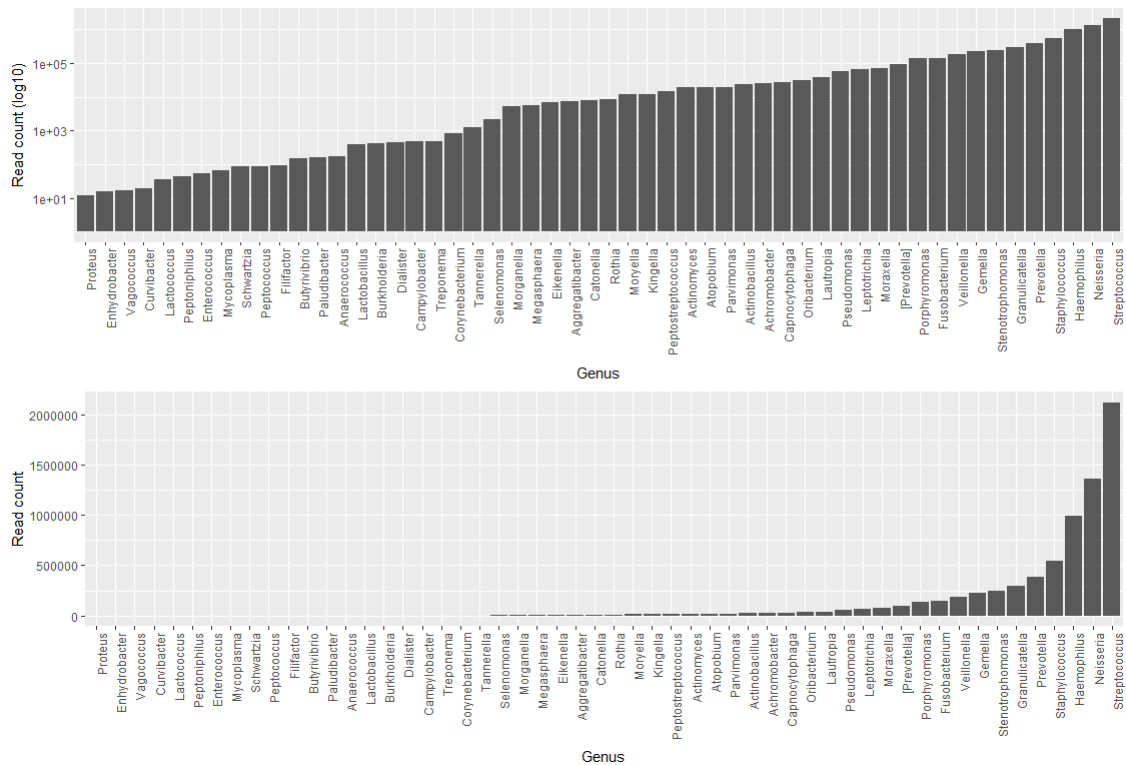


FIGURE 16 – Nombre de reads par genre bactérien

abondance de 25.94%. *Burkholderia* est retrouvé seulement dans 2 échantillons 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 20). Leur distribution est illustrée dans la figure 21. *Streptococcus* respecte grossièrement une distribution normale qui varie 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 uns, sont dominants. Les autres genres ont une abondance faible qui varie faiblement. Elles ne sont jamais retrouvées dominantes.

## 5.3 Dynamique du microbiote pulmonaire

### 5.3.1 Évolution de l'alpha diversité

Les figures 30, 31 et 32 montrent l'évolution des diversités alpha par patient au cours du temps en utilisant 3 indices différents. Les indices Chao1 et Observed reflètent la richesse (nombre de genre) présente dans l'échantillon. Tandis que l'indice de Shannon reflète aussi bien la richesse que l'équitabilité des distributions microbiennes. Les deux

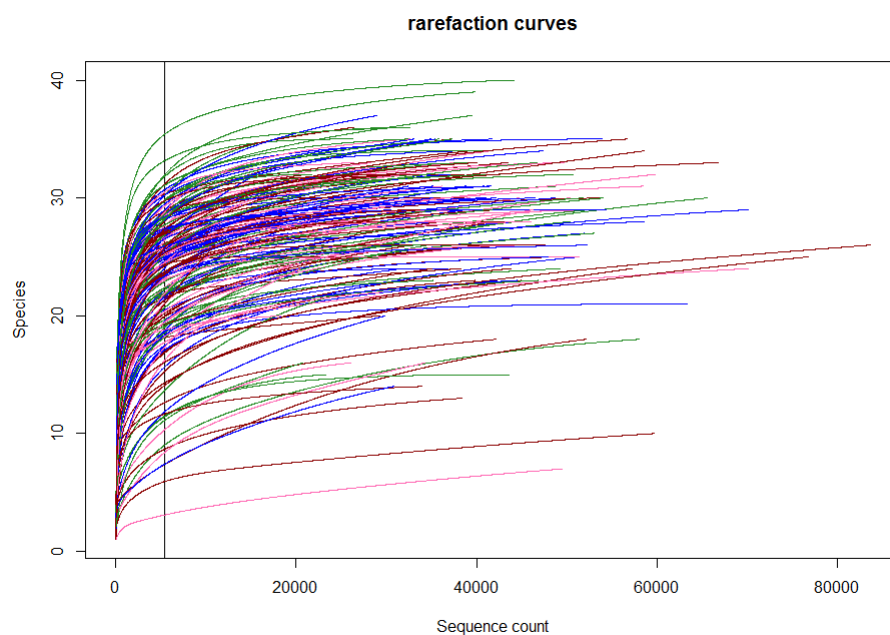


FIGURE 17 – Courbe de rarefaction

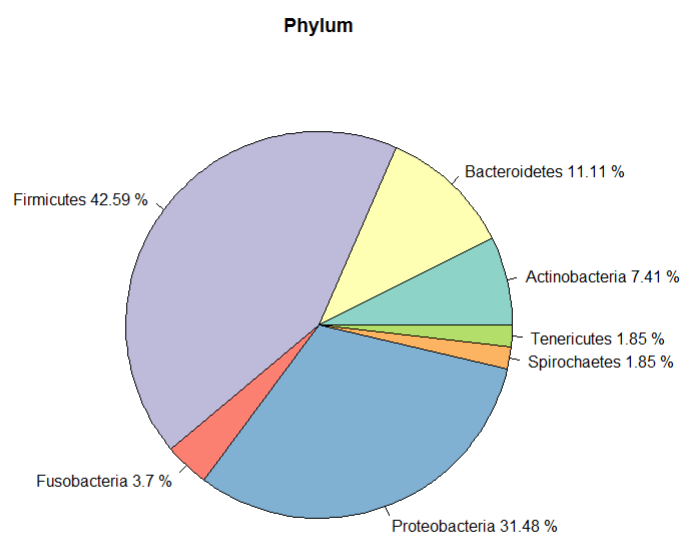


FIGURE 18 – Phylum

premiers graphiques montrent que les échantillons ont un nombre d'espèces pouvant varier 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. Le patient 25,26,74 conserve leurs richesses sur plus de 5 prélèvements successifs. Le patient

	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 19 – Nombre de reads et prévalence des genres bactériens parmi les 188 échantillons

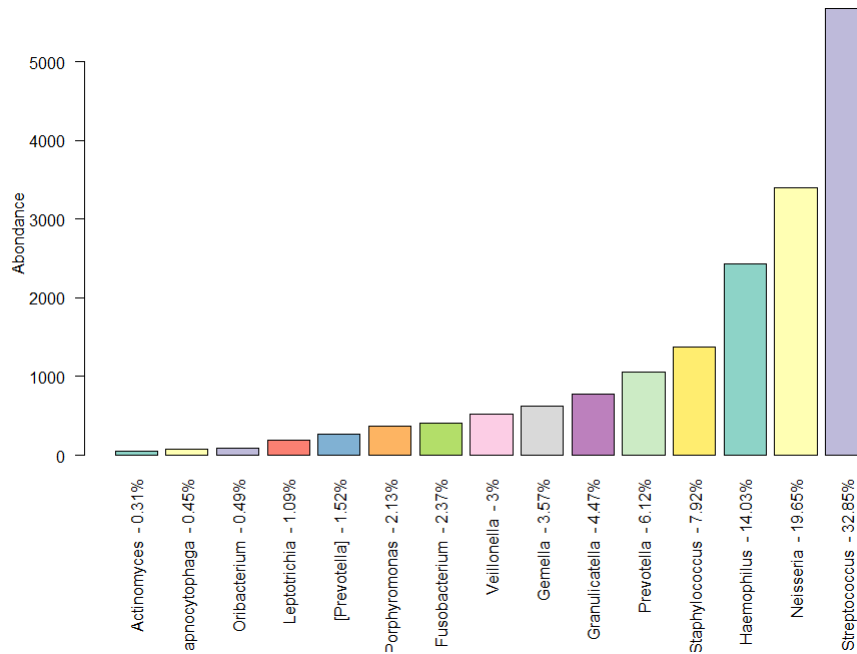


FIGURE 20 – Distribution du Core microbiota dans l'ensemble des échantillons

20 et 69 présente une stabilité entrecoupée par des pics de diminution. La richesse du patient 8 diminue progressivement. Le patient 23, 232 et 248 ont des fluctuations chaotiques. Le graphique de shannon montre que pour un nombre d'espèces constant, leurs distributions varient dans le temps. Par exemple sur le graphique 30, le patient 223 contient un nombre de genres relativement stable. En revanche sur le graphique 32, la variabilité de l'indice de shannon du patient 223 indique une distribution des espèces différente entre les échantillons. Seul le patient 26 conserve à la fois une richesse et une distribution stable au court du temps.

### 5.3.2 Évolution des abondances

Les figures 27, 28 et 29 nous montre 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 un *Neisseria* qui devient dominant à plus 80%. Le patient 3 montre une diminution progressive d'*Haemophilus* parallèlement à une augmentation de *Streptococcus*.

Le patient 256 est caractérisé par une dominance complète à *Sténotrophomonas* sur l'ensemble de ses échantillons. Pour le patient 211, *Achromobacter* est dominant dans le premier échantillon. D'autres échantillons montrent une résilience. Le patient 223 par exemple récupère un microbiote identique après une colonisation complète à *Haemo-*

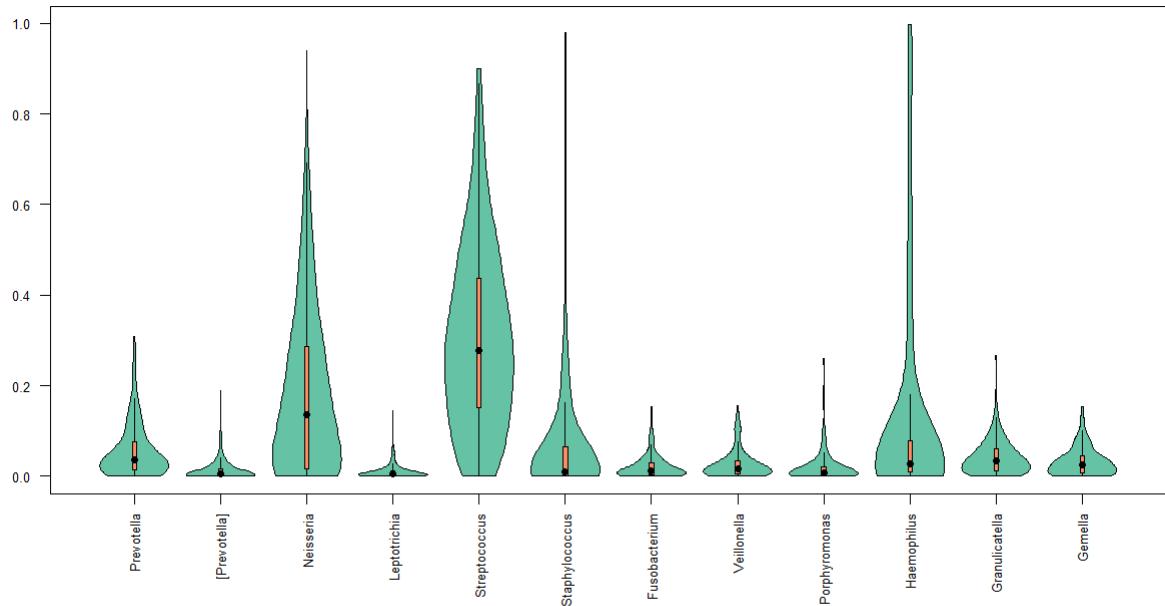


FIGURE 21 – Distribution du core microbiota

*philus*.

Dans l'ensemble, il existe une population de bactéries constante et minoritaire (*Fusobacterium*, *Granulicatella*, *Gemella*, *Veillonella*, *Parvimonas*, *Leptotrichia*, *Oribacterium*, *Capnocytophaga*, *Catonella*). Et une autre population très fluctuante dans le temps jouant alternativement le rôle de dominant. (*Haemophilus*, *Streptococcus*, *Neisseria* et *Prevotella*). La figure 23 illustre ces propos en montrant l'évolution des abondances du patient 43. La figure du patient 24 est mis également à titre exemple, car elle montre l'apparition de *Pseudomonas Aeruginosa*.

L'ensemble des graphiques est disponible sur cette adresse.

## 5.4 Corrélation entre genre bactériens

La figure 22 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 ( $p < 0.5$ ). *Streptococcus* et *Haemophilus* évolue dans le sens inverse avec un coefficient à -0.41 ( $p < 0.5$ ).

## 5.5 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 25 et 26 ). 2

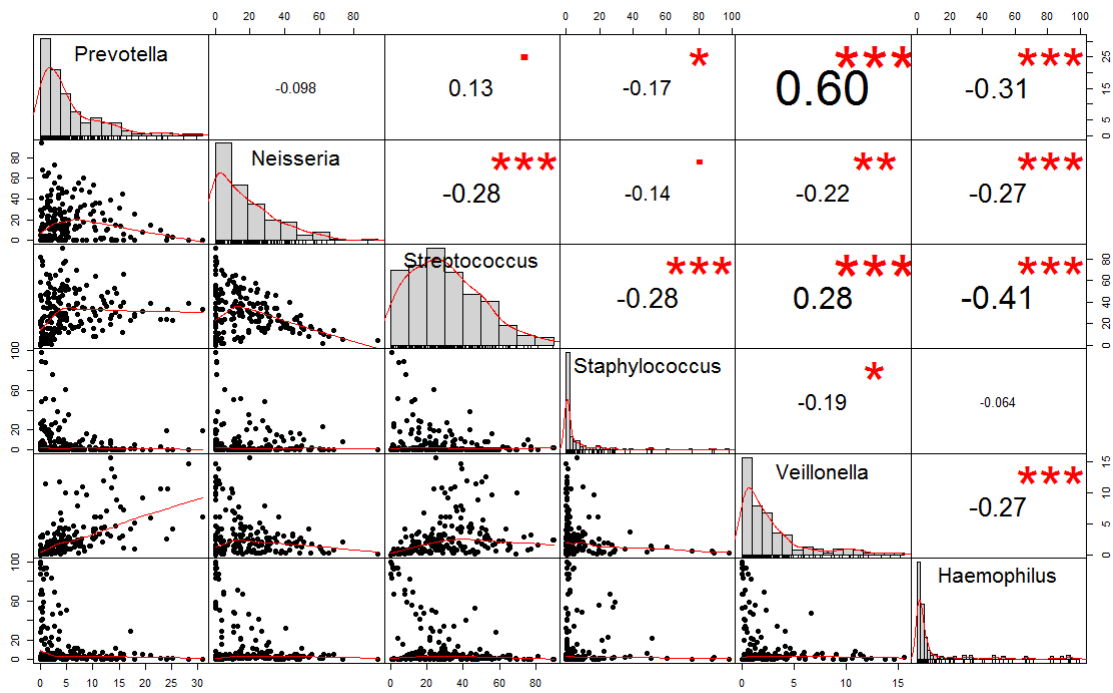


FIGURE 22 – Corrélation des abondances entre les genres du core microbiota

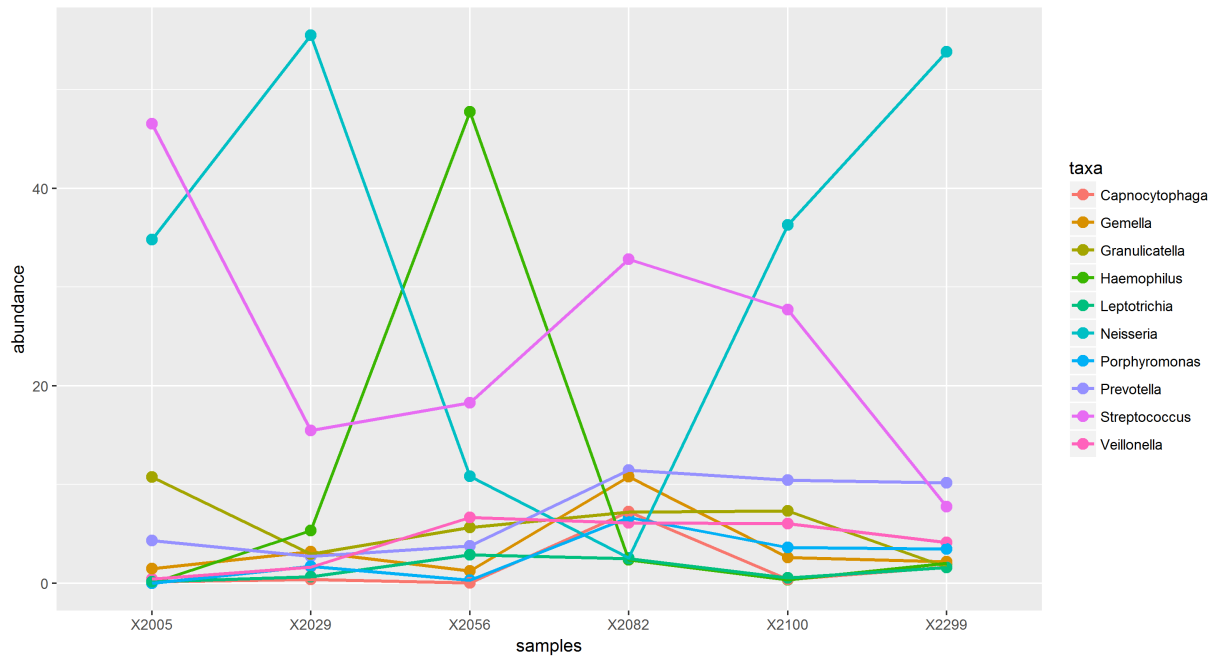


FIGURE 23 – Evolution des abondances pour le patient 43. Notez la population stable et la population fluctuante

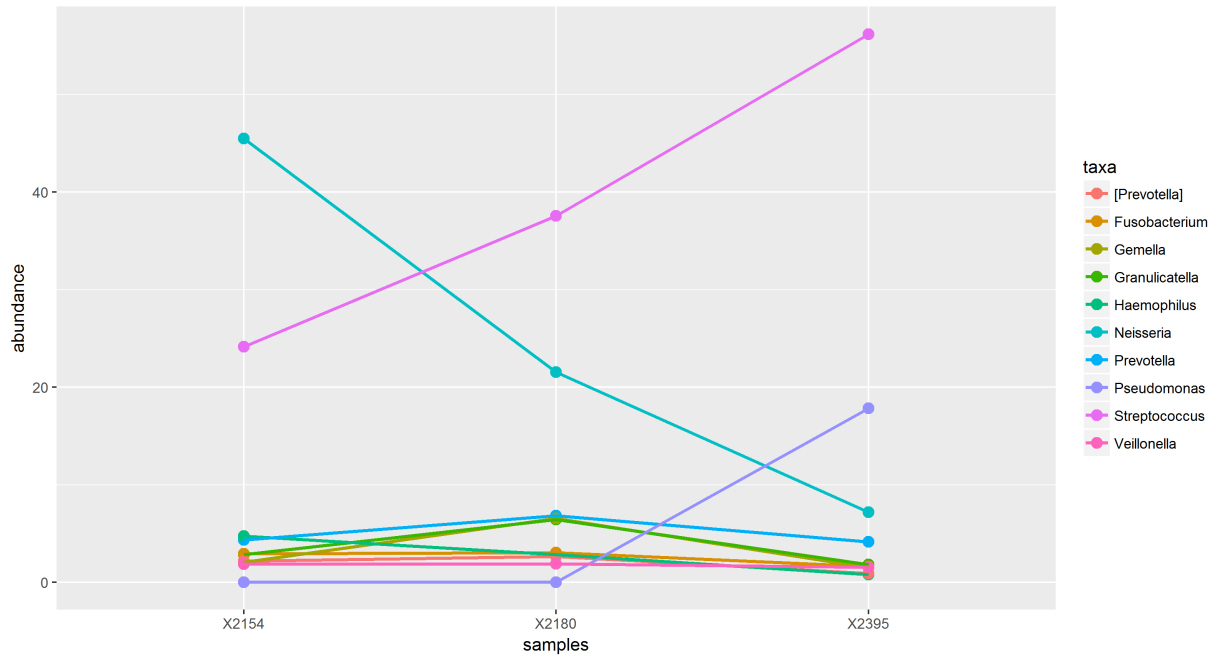


FIGURE 24 – Evolution des abondances pour le patient 54. Notez l'apparition de *Pseudomonas* dans le dernier échantillon

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 paramètres comme âge, le sexe, le poids, la prise d'antibiotique et le type de mutation n'a mis en évidence des groupes distincts de microbiote. La variabilité est expliquée principalement par la dominance des genres *Neisseria*, *Streptococcus* et *Haemophilus*. La diversité des microbiotes des échantillons Free ont une diversité plus faible que les échantillons Never.

Discussion LA RÉOLUTION TEMPORALE n'EST PAS ASSEZ GRANDE Population d'attaque et population locale Notion des 2 calques papiers On retrouve du corynébactérium ... Conclusion

## 6 Discussion

## 7 Conclusion



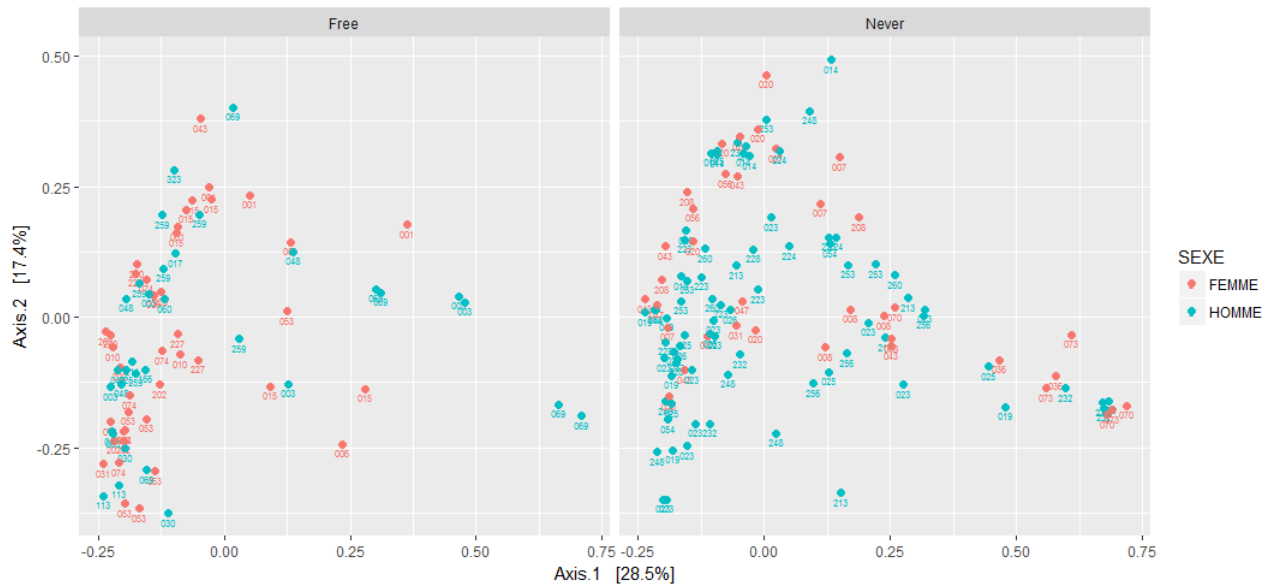


FIGURE 25 – Analyse multivarié ( PCA + Bray curtis) sur les 188 échantillons. Chaque point est un échantillon labelisé par l'identifiant du patient. A gauche les échantillons Free. A droite les échantillons Never.



FIGURE 26 – Analyse multivarié ( PCA + Bray curtis) sur les 188 échantillons. Chaque point est un échantillon labelisé par l'identifiant du patient.



FIGURE 27 – Evolution des abondances par genre

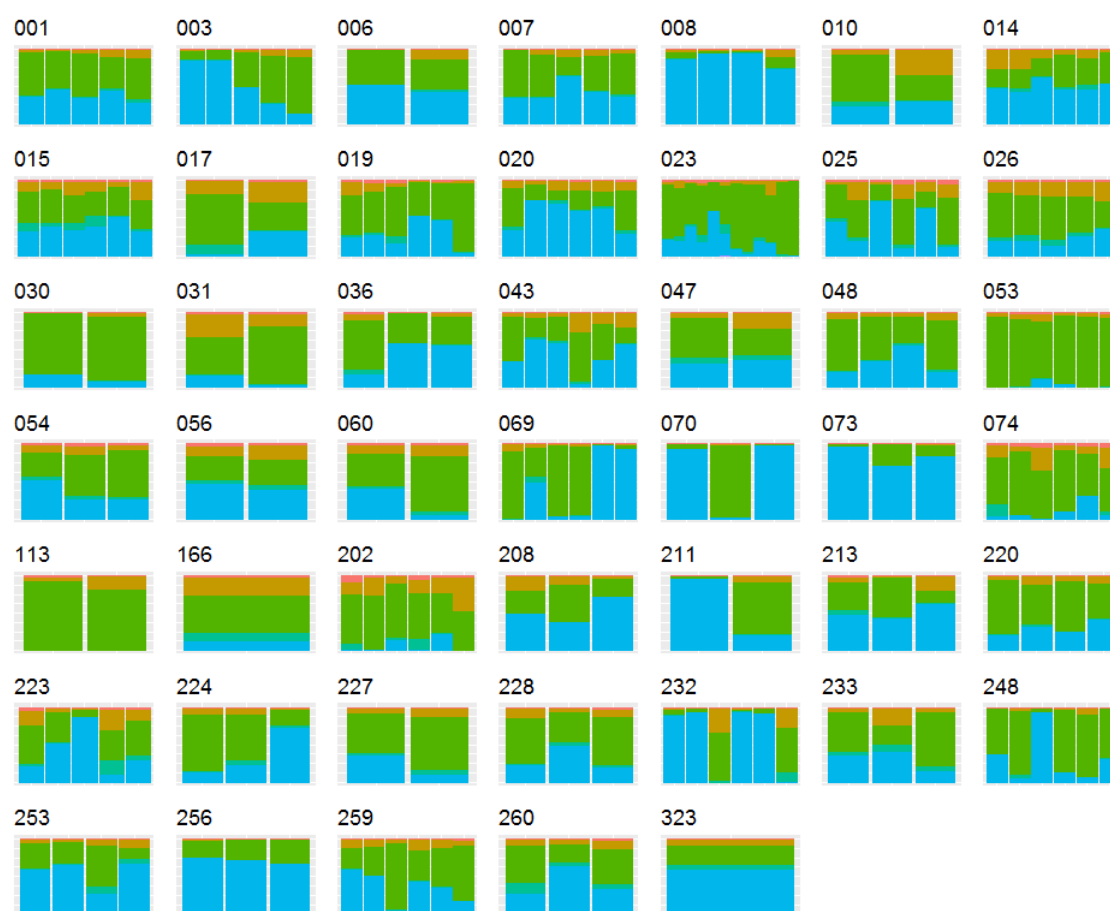


FIGURE 28 – Evolution des abondances par phylum

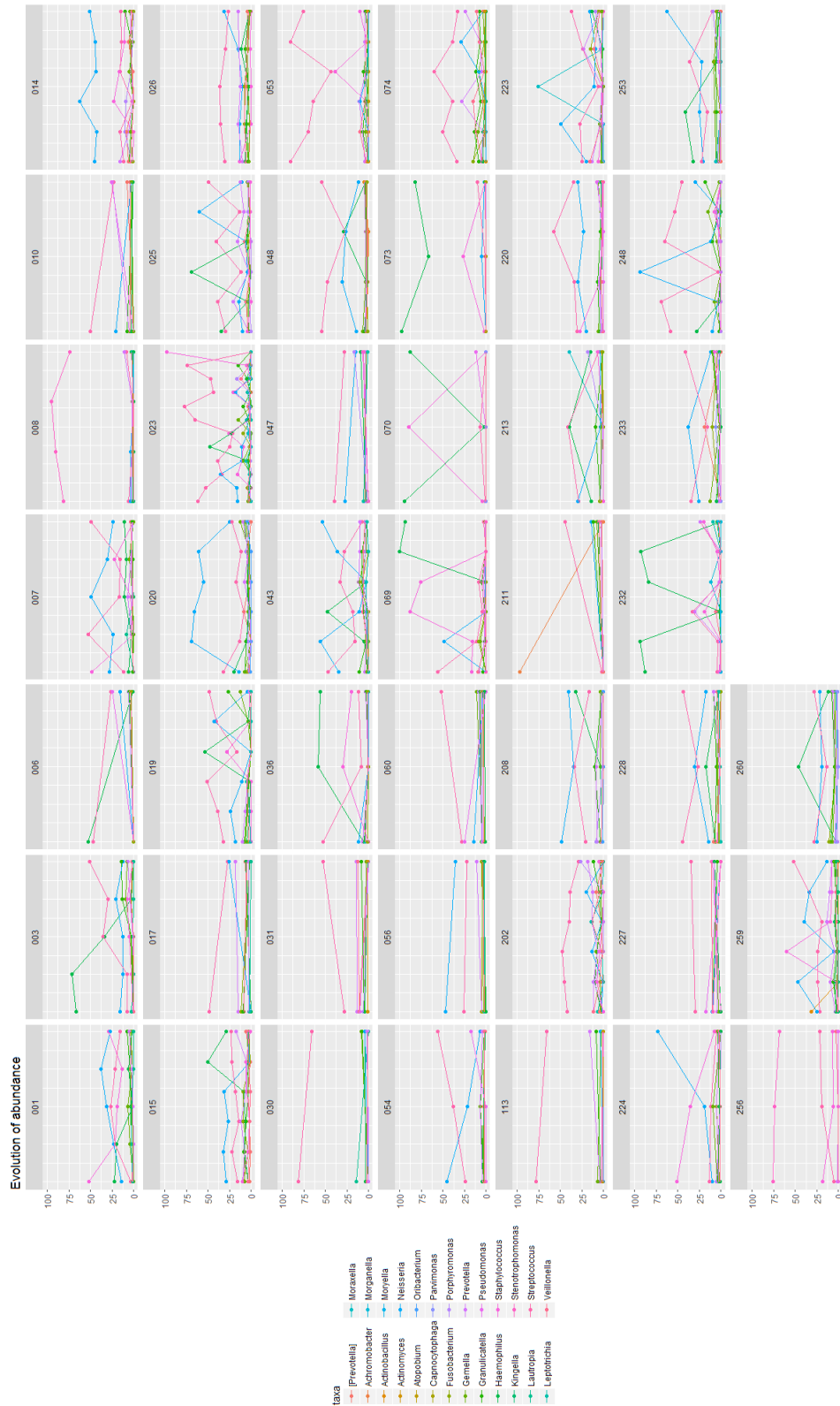


FIGURE 29 – Evolution des abondances par genre

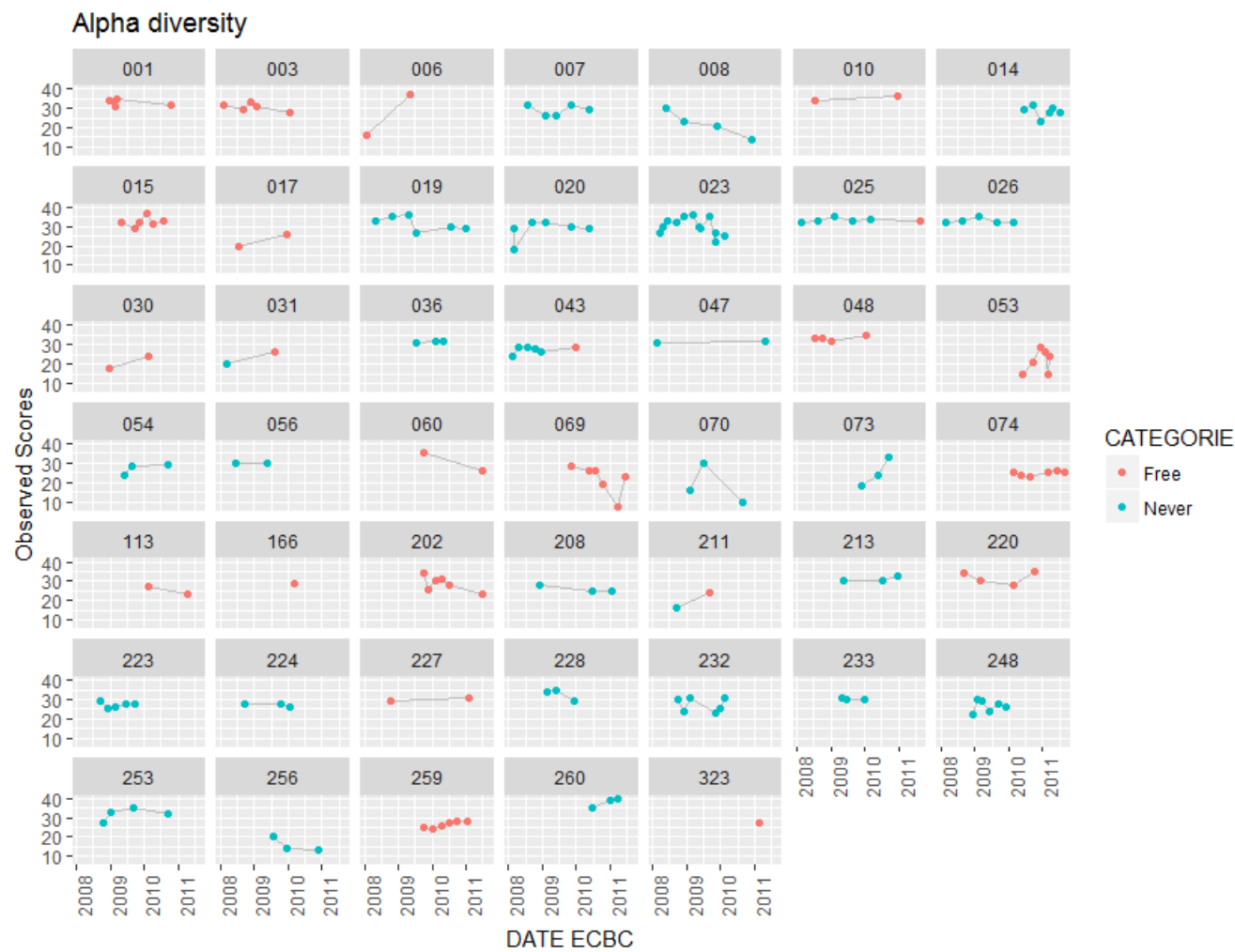


FIGURE 30 – Evolution du nombre d’espèces par patient en fonction du temps

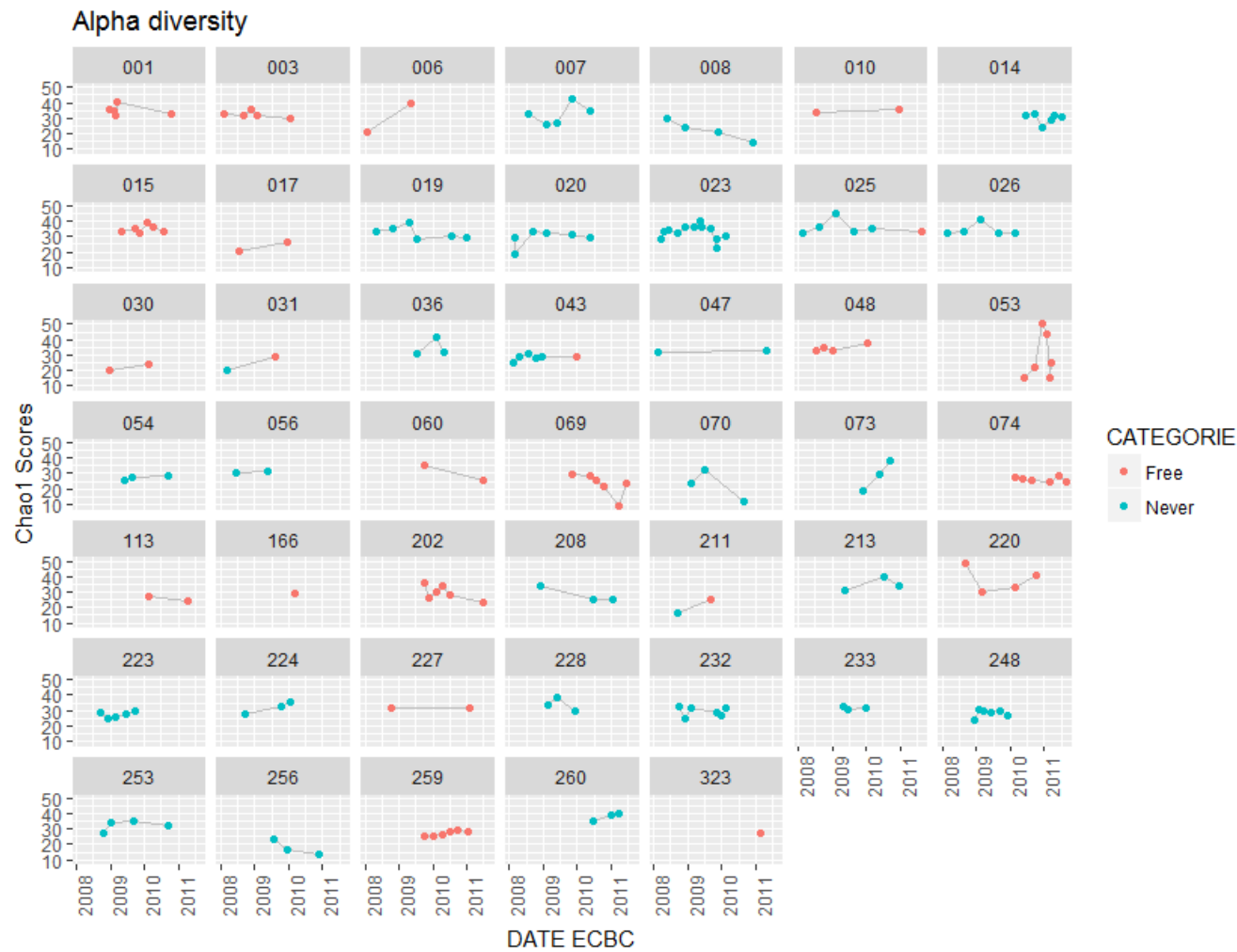


FIGURE 31 – Evolution de l'indice de Chao1 par patient en fonction du temps

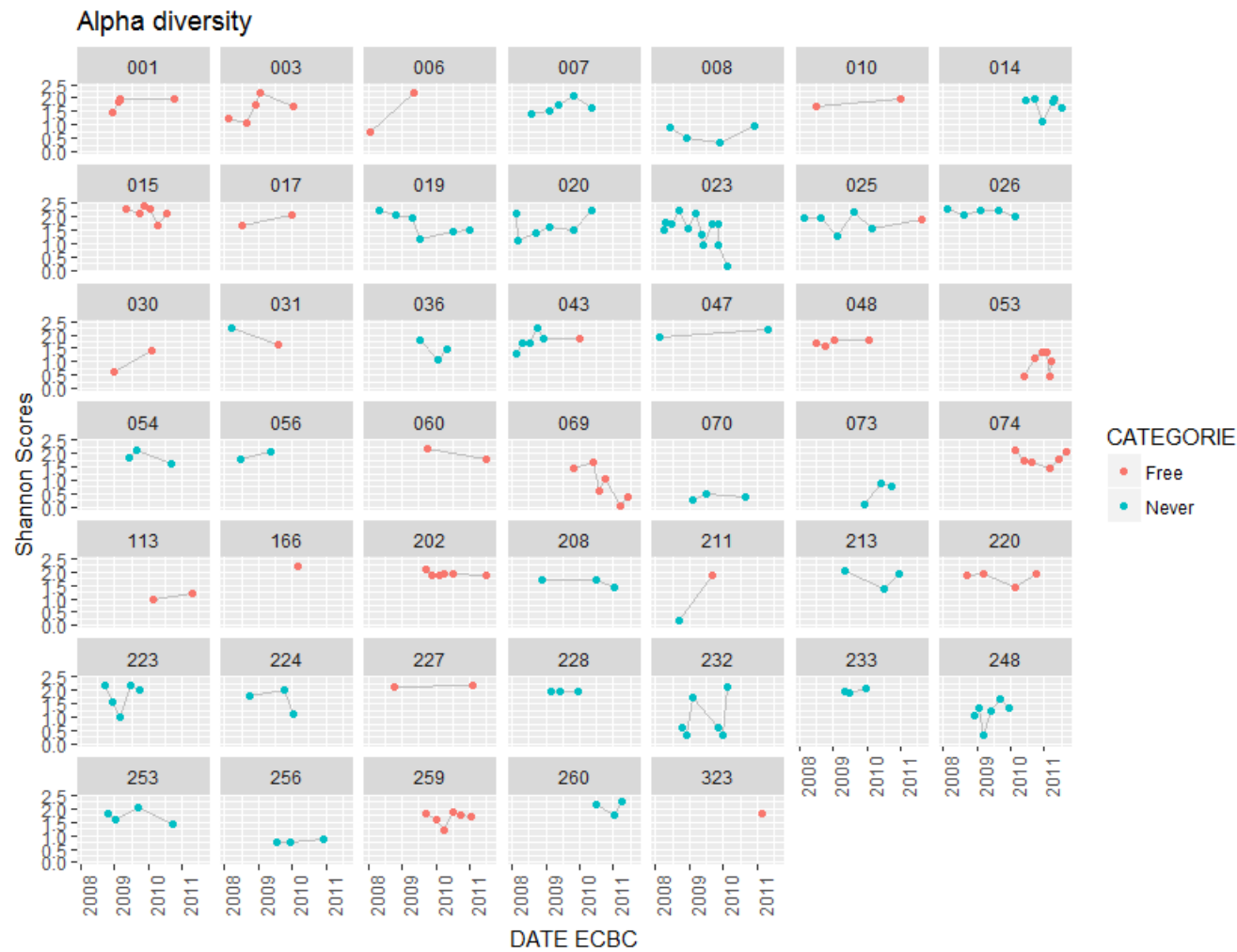


FIGURE 32 – Evolution de l'indice de Shanon par patient en fonction du temps

## 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] 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>.
- [5] 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.
- [6] Junjie Qin, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvs-ten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Florence Le-venez, 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 Artigue-nave, Hervé Blottiere, Natalia Borrue, Thomas Bruls, Francesc Casellas, Christian Chervaux, Antonella Cultrone, Christine Delorme, Gérard Denariáz, Rozenn Der-vyn, Miguel Forte, Carsten Friss, Maarten van de Guchte, Eric Guedon, Florence Haimet, Alexandre Jamet, Catherine Juste, Ghalia Kaci, Michiel Kleerebezem, Jan Knol, Michel Kristensen, Severine Layec, Karine Le Roux, Marion Leclerc, Em-manuelle 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 hu-man gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285) :59–65, mar 2010.



- [7] 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.
- [8] Ed Yong. *Moi, microbiote, maître du monde : Les microbes, 30 billions d'amis!* 2017.