



**HAL**  
open science

# Développement d'une technique de métagénomique infectieuse semi-ciblée basée sur la capture d'acides nucléiques pour le suivi des patients atteints de mucoviscidose

Benjamin Némoz

► **To cite this version:**

Benjamin Némoz. Développement d'une technique de métagénomique infectieuse semi-ciblée basée sur la capture d'acides nucléiques pour le suivi des patients atteints de mucoviscidose. Médecine humaine et pathologie. 2019. dumas-02092449

**HAL Id: dumas-02092449**

**<https://dumas.ccsd.cnrs.fr/dumas-02092449>**

Submitted on 8 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il n'a pas été réévalué depuis la date de soutenance.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact au SID de Grenoble :  
[bump-theses@univ-grenoble-alpes.fr](mailto:bump-theses@univ-grenoble-alpes.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

<http://www.cfcopies.com/juridique/droit-auteur>

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

UNIVERSITÉ GRENOBLE ALPES  
FACULTÉ DE MÉDECINE DE GRENOBLE

Année : 2019

---

Développement d'une technique de métagénomique infectieuse  
semi-ciblée basée sur la capture d'acides nucléiques pour le suivi  
des patients atteints de mucoviscidose

THÈSE  
PRÉSENTÉE EN VUE DE L'OBTENTION DU DOCTORAT EN MÉDECINE  
DIPLOME D'ETAT

Par

**Benjamin NEMOZ**

[Données à caractère personnel]

THÈSE SOUTENUE PUBLIQUEMENT A LA FACULTÉ DE MÉDECINE DE  
GRENOBLE\*

Le 05 Avril 2019

DEVANT LE JURY COMPOSÉ DE

Président du jury      Monsieur le Professeur Patrice MORAND

Membres      Madame le Docteur Sylvie LARRAT (Directrice de thèse)  
Monsieur le Docteur Boubou CAMARA  
Monsieur le Docteur Yvan CASPAR  
Madame le Professeur Muriel CORNET  
Monsieur le Professeur Alexandre MOREAU-GAUDRY

\* La Faculté de Médecine de Grenoble n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions sont considérées comme propres à leurs auteurs.

Doyen de la Faculté : **Pr. Patrice MORAND**

Année 2018-2019

**ENSEIGNANTS DE L'UFR DE MEDECINE**

<b>CORPS</b>	<b>NOM-PRENOM</b>	<b>Discipline universitaire</b>
PU-PH	<b>ALBALADEJO Pierre</b>	<b>Anesthésiologie réanimation</b>
PU-PH	<b>APTTEL Florent</b>	<b>Ophthalmologie</b>
PU-PH	<b>ARVIEUX-BARTHELEMY Catherine</b>	<b>Chirurgie générale</b>
PU-PH	<b>BAILLET Athan</b>	<b>Rhumatologie</b>
PU-PH	<b>BARONE-ROCHETTE Gilles</b>	<b>Cardiologie</b>
PU-PH	<b>BAYAT Sam</b>	<b>Physiologie</b>
PU-PH	<b>BENHAMOU Pierre Yves</b>	<b>Endocrinologie, diabète et maladies métaboliques</b>
PU-PH	<b>BERGER François</b>	<b>Biologie cellulaire</b>
MCU-PH	<b>BIDART-COUTTON Marie</b>	<b>Biologie cellulaire</b>
MCU-PH	<b>BOISSET Sandrine</b>	<b>Agents infectieux</b>
PU-PH	<b>BOLLA Michel</b>	<b>Cancérologie-Radiothérapie</b>
PU-PH	<b>BONAZ Bruno</b>	<b>Gastro-entérologie, hépatologie, addictologie</b>
PU-PH	<b>BONNETERRE Vincent</b>	<b>Médecine et santé au travail</b>
PU-PH	<b>BOREL Anne-Laure</b>	<b>Endocrinologie, diabète et maladies métaboliques</b>
PU-PH	<b>BOSSON Jean-Luc</b>	<b>Biostatistiques, informatique médicale et technologies de communication</b>
MCU-PH	<b>BOTTARI Serge</b>	<b>Biologie cellulaire</b>
PU-PH	<b>BOUGEROL Thierry</b>	<b>Psychiatrie d'adultes</b>
PU-PH	<b>BOUILLET Laurence</b>	<b>Médecine interne</b>
PU-PH	<b>BOUZAT Pierre</b>	<b>Réanimation</b>
PU-PH	<b>BRAMBILLA Christian</b>	<b>Pneumologie</b>
PU-PH	<b>BRAMBILLA Elisabeth</b>	<b>Anatomie et de Pathologique Cytologiques</b>
MCU-PH	<b>BRENIER-PINCHART Marie Pierre</b>	<b>Parasitologie et mycologie</b>
PU-PH	<b>BRICAULT Ivan</b>	<b>Radiologie et imagerie médicale</b>
PU-PH	<b>BRICHON Pierre-Yves</b>	<b>Chirurgie thoracique et cardio- vasculaire</b>
MCU-PH	<b>BRIOT Raphaël</b>	<b>Thérapeutique, médecine d'urgence</b>
MCU-PH	<b>BROUILLET Sophie</b>	<b>Biologie et médecine du développement et de la reproduction</b>
PU-PH	<b>CAHN Jean-Yves</b>	<b>Hématologie</b>
PU-PH	<b>CANALI-SCHWEBEL Carole</b>	<b>Réanimation médicale</b>
PU-PH	<b>CARPENTIER Françoise</b>	<b>Thérapeutique, médecine d'urgence</b>
PU-PH	<b>CARPENTIER Patrick</b>	<b>Chirurgie vasculaire, médecine vasculaire</b>
PU-PH	<b>CESBRON Jean-Yves</b>	<b>Immunologie</b>
PU-PH	<b>CHABARDES Stephan</b>	<b>Neurochirurgie</b>
PU-PH	<b>CHABRE Olivier</b>	<b>Endocrinologie, diabète et maladies métaboliques</b>
PU-PH	<b>CHAFFANJON Philippe</b>	<b>Anatomie</b>
PU-PH	<b>CHARLES Julie</b>	<b>Dermatologie</b>
PU-PH	<b>CHAVANON Olivier</b>	<b>Chirurgie thoracique et cardio- vasculaire</b>
PU-PH	<b>CHIQUET Christophe</b>	<b>Ophthalmologie</b>

<b>CORPS</b>	<b>NOM-PRENOM</b>	<b>Discipline universitaire</b>
PU-PH	<b>CHIRICA Mircea</b>	<b>Chirurgie générale</b>
PU-PH	<b>CINQUIN Philippe</b>	<b>Biostatistiques, informatique médicale et technologies de communication</b>
MCU-PH	<b>CLAVARINO Giovanna</b>	<b>Immunologie</b>
PU-PH	<b>COHEN Olivier</b>	<b>Biostatistiques, informatique médicale et technologies de communication</b>
PU-PH	<b>COURVOISIER Aurélien</b>	<b>Chirurgie infantile</b>
PU-PH	<b>COUTTON Charles</b>	<b>Génétique</b>
PU-PH	<b>COUTURIER Pascal</b>	<b>Gériatrie et biologie du vieillissement</b>
PU-PH	<b>CRACOWSKI Jean-Luc</b>	<b>Pharmacologie fondamentale, pharmacologie clinique</b>
PU-PH	<b>CURE Hervé</b>	<b>Oncologie</b>
PU-PH	<b>DEBATY Guillaume</b>	<b>Médecine d'Urgence</b>
PU-PH	<b>DEBILLON Thierry</b>	<b>Pédiatrie</b>
PU-PH	<b>DECAENS Thomas</b>	<b>Gastro-entérologie, Hépatologie</b>
PU-PH	<b>DEMATTEIS Maurice</b>	<b>Addictologie</b>
PU-PH	<b>DEMONGEOT Jacques</b>	<b>Biostatistiques, informatique médicale et technologies de communication</b>
MCU-PH	<b>DERANSART Colin</b>	<b>Physiologie</b>
PU-PH	<b>DESCOTES Jean-Luc</b>	<b>Urologie</b>
PU-PH	<b>DETANTE Olivier</b>	<b>Neurologie</b>
MCU-PH	<b>DIETERICH Klaus</b>	<b>Génétique et procréation</b>
MCU-PH	<b>DOUTRELEAU Stéphane</b>	<b>Physiologie</b>
MCU-PH	<b>DUMESTRE-PERARD Chantal</b>	<b>Immunologie</b>
PU-PH	<b>EPAULARD Olivier</b>	<b>Maladies Infectieuses et Tropicales</b>
PU-PH	<b>ESTEVE François</b>	<b>Biophysique et médecine nucléaire</b>
MCU-PH	<b>EYSSERIC Hélène</b>	<b>Médecine légale et droit de la santé</b>
PU-PH	<b>FAGRET Daniel</b>	<b>Biophysique et médecine nucléaire</b>
PU-PH	<b>FAUCHERON Jean-Luc</b>	<b>Chirurgie générale</b>
MCU-PH	<b>FAURE Julien</b>	<b>Biochimie et biologie moléculaire</b>
PU-PH	<b>FERRETTI Gilbert</b>	<b>Radiologie et imagerie médicale</b>
PU-PH	<b>FEUERSTEIN Claude</b>	<b>Physiologie</b>
PU-PH	<b>FONTAINE Éric</b>	<b>Nutrition</b>
PU-PH	<b>FRANCOIS Patrice</b>	<b>Epidémiologie, économie de la santé et prévention</b>
MCU-MG	<b>GABOREAU Yoann</b>	<b>Médecine Générale</b>
PU-PH	<b>GARBAN Frédéric</b>	<b>Hématologie, transfusion</b>
PU-PH	<b>GAUDIN Philippe</b>	<b>Rhumatologie</b>
PU-PH	<b>GAVAZZI Gaétan</b>	<b>Gériatrie et biologie du vieillissement</b>
PU-PH	<b>GAY Emmanuel</b>	<b>Neurochirurgie</b>
MCU-PH	<b>GILLOIS Pierre</b>	<b>Biostatistiques, informatique médicale et technologies de communication</b>
PU-PH	<b>GIOT Jean-Philippe</b>	<b>Chirurgie plastique, reconstructrice et esthétique</b>
MCU-PH	<b>GRAND Sylvie</b>	<b>Radiologie et imagerie médicale</b>
PU-PH	<b>GRIFFET Jacques</b>	<b>Chirurgie infantile</b>
MCU-PH	<b>GUZUN Rita</b>	<b>Endocrinologie, diabétologie, nutrition, éducation thérapeutique</b>
PU-PH	<b>HAINAUT Pierre</b>	<b>Biochimie, biologie moléculaire</b>
PU-PH	<b>HALIMI Serge</b>	<b>Nutrition</b>
PU-PH	<b>HENNEBICQ Sylviane</b>	<b>Génétique et procréation</b>
PU-PH	<b>HOFFMANN Pascale</b>	<b>Gynécologie obstétrique</b>
PU-PH	<b>HOMMEL Marc</b>	<b>Neurologie</b>
PU-MG	<b>IMBERT Patrick</b>	<b>Médecine Générale</b>
PU-PH	<b>JOUK Pierre-Simon</b>	<b>Génétique</b>

<b>CORPS</b>	<b>NOM-PRENOM</b>	<b>Discipline universitaire</b>
PU-PH	JUVIN Robert	Rhumatologie
PU-PH	KAHANE Philippe	Physiologie
MCU-PH	KASTLER Adrian	Radiologie et imagerie médicale
PU-PH	KRAINIK Alexandre	Radiologie et imagerie médicale
PU-PH	LABARERE José	Epidémiologie ; Eco. de la Santé
MCU-PH	LABLANCHE Sandrine	Endocrinologie, diabète et maladies métaboliques
MCU-PH	LANDELLE Caroline	Bactériologie - virologie
MCU-PH	LARDY Bernard	Biochimie et biologie moléculaire
MCU - PH	LE PISSART Audrey	Biochimie et biologie moléculaire
PU-PH	LECCIA Marie-Thérèse	Dermato-vénérologie
PU-PH	LEROUX Dominique	Génétique
PU-PH	LEROY Vincent	Gastro-entérologie, hépatologie, addictologie
PU-PH	LETOUBLON Christian	Chirurgie digestive et viscérale
PU-PH	LEVY Patrick	Physiologie
PU-PH	LONG Jean-Alexandre	Urologie
MCU-PH	LUPO Julien	Virologie
PU-PH	MAGNE Jean-Luc	Chirurgie vasculaire
MCU-PH	MAIGNAN Maxime	Médecine d'urgence
PU-PH	MAITRE Anne	Médecine et santé au travail
MCU-PH	MALLARET Marie-Reine	Epidémiologie, économie de la santé et prévention
PU-PH	MALLION Jean-Michel	Cardiologie
MCU-PH	MARLU Raphaël	Hématologie, transfusion
MCU-PH	MAUBON Danièle	Parasitologie et mycologie
PU-PH	MAURIN Max	Bactériologie - virologie
MCU-PH	MC LEER Anne	Cytologie et histologie
PU-PH	MORAND Patrice	Bactériologie - virologie
PU-PH	MOREAU-GAUDRY Alexandre	Biostatistiques, informatique médicale et technologies de communication
PU-PH	MORO Elena	Neurologie
PU-PH	MORO-SIBILOT Denis	Pneumologie
PU-PH	MOUSSEAU Mireille	Cancérologie
PU-PH	MOUTET François	Chirurgie plastique, reconstructrice et esthétique ; brûlologie
MCU-PH	PACLET Marie-Hélène	Biochimie et biologie moléculaire
PU-PH	PALOMBI Olivier	Anatomie
PU-PH	PARK Sophie	Hémato - transfusion
PU-PH	PASSAGGIA Jean-Guy	Anatomie
PU-PH	PAYEN DE LA GARANDERIE Jean-François	Anesthésiologie réanimation
MCU-PH	PAYSANT François	Médecine légale et droit de la santé
MCU-PH	PELLETIER Laurent	Biologie cellulaire
PU-PH	PELLOUX Hervé	Parasitologie et mycologie
PU-PH	PEPIN Jean-Louis	Physiologie
PU-PH	PERENNOU Dominique	Médecine physique et de réadaptation
PU-PH	PERNOD Gilles	Médecine vasculaire
PU-PH	PIOLAT Christian	Chirurgie infantile
PU-PH	PISON Christophe	Pneumologie
PU-PH	PLANTAZ Dominique	Pédiatrie
PU-PH	POIGNARD Pascal	Virologie
PU-PH	POLACK Benoît	Hématologie

<b>CORPS</b>	<b>NOM-PRENOM</b>	<b>Discipline universitaire</b>
PU-PH	<b>POLOSAN Mircea</b>	<b>Psychiatrie d'adultes</b>
PU-PH	<b>PONS Jean-Claude</b>	<b>Gynécologie obstétrique</b>
PU-PH	<b>RAMBEAUD Jean-Jacques</b>	<b>Urologie</b>
PU-PH	<b>RAY Pierre</b>	<b>Biologie et médecine du développement et de la reproduction</b>
MCU-PH	<b>RENDU John</b>	<b>Biochimie et Biologie Moléculaire</b>
MCU-PH	<b>RIALLE Vincent</b>	<b>Biostatistiques, informatique médicale et technologies de communication</b>
PU-PH	<b>RIGHINI Christian</b>	<b>Oto-rhino-laryngologie</b>
PU-PH	<b>ROMANET Jean Paul</b>	<b>Ophthalmologie</b>
PU-PH	<b>ROSTAING Lionel</b>	<b>Néphrologie</b>
MCU-PH	<b>ROUSTIT Matthieu</b>	<b>Pharmacologie fondamentale, pharmaco clinique, addictologie</b>
MCU-PH	<b>ROUX-BUISSON Nathalie</b>	<b>Biochimie, toxicologie et pharmacologie</b>
MCU-PH	<b>RUBIO Amandine</b>	<b>Pédiatrie</b>
PU-PH	<b>SARAGAGLIA Dominique</b>	<b>Chirurgie orthopédique et traumatologie</b>
MCU-PH	<b>SATRE Véronique</b>	<b>Génétique</b>
PU-PH	<b>SAUDOU Frédéric</b>	<b>Biologie Cellulaire</b>
PU-PH	<b>SCHMERBER Sébastien</b>	<b>Oto-rhino-laryngologie</b>
PU-PH	<b>SCOLAN Virginie</b>	<b>Médecine légale et droit de la santé</b>
MCU-PH	<b>SEIGNEURIN Arnaud</b>	<b>Epidémiologie, économie de la santé et prévention</b>
PU-PH	<b>STAHL Jean-Paul</b>	<b>Maladies infectieuses, maladies tropicales</b>
PU-PH	<b>STANKE Françoise</b>	<b>Pharmacologie fondamentale</b>
MCU-PH	<b>STASIA Marie-José</b>	<b>Biochimie et biologie moléculaire</b>
PU-PH	<b>STURM Nathalie</b>	<b>Anatomie et cytologie pathologiques</b>
PU-PH	<b>TAMISIER Renaud</b>	<b>Physiologie</b>
PU-PH	<b>TERZI Nicolas</b>	<b>Réanimation</b>
MCU-PH	<b>TOFFART Anne-Claire</b>	<b>Pneumologie</b>
PU-PH	<b>TONETTI Jérôme</b>	<b>Chirurgie orthopédique et traumatologie</b>
PU-PH	<b>TOUSSAINT Bertrand</b>	<b>Biochimie et biologie moléculaire</b>
PU-PH	<b>VANZETTO Gérald</b>	<b>Cardiologie</b>
PU-PH	<b>VUILLEZ Jean-Philippe</b>	<b>Biophysique et médecine nucléaire</b>
PU-PH	<b>WEIL Georges</b>	<b>Epidémiologie, économie de la santé et prévention</b>
PU-PH	<b>ZAOUI Philippe</b>	<b>Néphrologie</b>
PU-PH	<b>ZARSKI Jean-Pierre</b>	<b>Gastro-entérologie, hépatologie, addictologie</b>

**PU-PH** : Professeur des Universités et Praticiens Hospitaliers

**MCU-PH** : Maître de Conférences des Universités et Praticiens Hospitaliers

**PU-MG** : Professeur des Universités de Médecine Générale

**MCU-MG** : Maître de Conférences des Universités de Médecine Générale

*Aux claques,  
qui font tourner la tête dans la bonne direction ...*

*[Teresa Radice]*



## REMERCIEMENTS

Je souhaiterais dans un premier temps remercier mon jury de thèse :

**Monsieur le Professeur Patrice Morand**, merci pour la confiance et la disponibilité dont vous avez fait preuve à mon égard au cours de ces dernières années. Vous avez conservé, malgré les fonctions chronophages que vous occupez, une porte ouverte à la discussion et cela est particulièrement apprécié. Vous avez sans cesse répondu à mes questions d'ordre virologique, et vous m'avez renouvelé plusieurs fois votre soutien. Je vous en suis très reconnaissant.

**Monsieur le Docteur Boubou Camara**, merci tout d'abord d'avoir accepté de relire et juger de mon travail. Je tenais tout particulièrement à vous remercier pour l'opportunité que vous m'avez offert de travailler sur ce projet. L'ensemble de l'étude Dysbiose et tout ce qu'il y a derrière au laboratoire n'aurait pas été là sans votre initiative.

**Madame le Professeur Muriel Cornet**, vous m'accordez le plaisir de juger mon travail aujourd'hui et c'est un honneur pour moi. J'aurais grand plaisir à travailler avec vous à l'avenir.

**Monsieur le Professeur Moreau-Gaudry**, merci d'avoir accepté d'être présent à cette soutenance. J'espère que vous aurez trouvé dans ce manuscrit un aperçu intéressant de l'activité du laboratoire et que cela ouvrira peut-être les portes d'une collaboration avec votre service.

**Sylvie**, merci pour ton encadrement dans ce travail et pour l'ensemble des conseils dont tu as pu m'abreuver pour mener à bien la rédaction de ce manuscrit. Les échanges que nous avons pu avoir ont toujours été constructifs et ton point de vue intéressant. Merci pour toutes les connaissances que tu as pu me partager, en particulier sur les hépatites virales.

**Yvan**, un grand merci à toi d'avoir accepté de participer à ce jury. J'espère que l'évaluation de mon travail de thèse ne te laissera pas un trop mauvais souvenir et ne nous empêchera pas de poursuivre la vie sociale du DAI ! Un grand merci à toi pour ta bienveillance, ta disponibilité et le dynamisme que tu apportes au 3<sup>ème</sup> étage. J'ai énormément apprécié travailler avec toi et j'espère que nous aurons l'occasion de mettre en place les projets déjà abordés et tous les autres à venir.

Par la suite, j'aimerais remercier les personnes qui, de près ou de loin, ont façonné ma vie et m'ont permis d'aboutir ici aujourd'hui.

A toi, **Samuel**, mon cher grand frère. Je crois que je pourrais écrire mille lignes à ton propos. Tu as rempli avec brio l'ensemble des missions qui sont celles d'un grand frère. Et pourtant, le challenge était quand même de taille ! Conseil (toujours bon), réconfort, discussions sur le sens de la vie, partage, rigolades et bières... Tu as toujours été là, infailible. Et infatigable ! Je te tire mon chapeau parce que je ne t'ai (presque) jamais vu baisser les bras. Ta détermination et ta force de caractère ont toujours été un exemple pour moi. Il est évident que je te souhaite le meilleur pour tous ces projets que tu entreprends, seul ou avec **Anne**, ma future belle-sœur préférée !

Une énorme pensée pour toi **Timothée**, mon frangin adoré, et aussi un extraordinaire merci. L'année 2018 a été pour nous deux une année charnière. Pour tout plein de raisons. Mais tu as été là, répondant présent à l'appel, plus près que jamais ! Tu as été une bouffée d'oxygène incroyable pour moi par ton arrivée à Grenoble. L'occasion de nous retrouver, d'échanger sur nos vies et sur les chemins que l'on choisit, et aussi de bien sûr de remettre sur le planning les courses en montagne ! Escalade, foot, course à pied, ski bien sûr (je rappelle que le

monoskieur prend son pied... dans le même sabot !)... tout est bon ! Alors pour toutes ces bonnes doses d'endorphines, merci à toi.

**Corentin**, pour toi, je réserve une poignée de merci, bien sûr. Mais surtout, je profite de ces quelques lignes qui me sont offertes en tribune pour te témoigner de ma plus grande admiration. Tu es quelqu'un d'extraordinaire (si, si, je t'assure !) qui brille dans tout ce qu'il entreprend et je te voue une certaine admiration pour cela. Malgré la distance de ces dernières années, j'ai eu grand plaisir à te retrouver à chaque réunion familiale pour suivre tes élucubrations et tes projets. Et chaque fois, tu as réussi à me surprendre par tes choix, parfois originaux mais incroyablement pertinents. Non seulement je suis fier de toi, mais presque un peu jaloux ! J'espère que cette soutenance te donnera envie, toi aussi, de prolonger pour quelques années la fabuleuse vie d'étudiant ...

Mon cher petit frère. Un très grand merci **Théophile** pour ta présence et les bonnes tranches de rire que l'on s'est payé. On ne t'a pas laissé la voie la plus facile (passer après nous 4, on a dû t'en faire baver !), mais on peut dire que tu t'en es plutôt bien sorti avec des frangins pareils ! Merci pour les recommandations de lecture, pour m'avoir supporté dans la lutte (le communisme vaincra !) et pour les discussions riches autour de l'actualité ou de sujets chauds. Merci de nous avoir bien fait rire au baby-foot (t'es mauvais, Jack !), il me reste une dernière raclée à te coller ce soir !

A toi ma maman, merci. Il y a tant de choses pour lesquelles j'aimerais te remercier **Blandine**. Tu n'imagines même pas. Je ne t'ai jamais remercié pour tout ce que tu as fait pour moi, pour tout ce que tu as fait pour nous. Tu nous as donné un amour incroyable et infini, qui a bercé ma vie de tendresse et de (ré)confort. Tu représentes à toi seule tout ce que l'on peut espérer

avoir derrière le mot 'famille' et je peux t'assurer que tu as carrément été à la hauteur. Bravo à toi, j'étais pourtant une vraie teigne.

A mon super-héros de papa, j'aurais aussi mille raisons de te remercier. **Bernard**, tu es un papa génial, toujours présent quand il le faut, incroyablement fiable et droit, juste bien sûr, autoritaire un peu, parfois. Tu es un roc sur lequel j'ai pu m'appuyer, qu'il s'agisse de logistique à l'autre bout du monde, de peine de cœur, de dégauchisseuse ou de groupe électrogène, ou pour remplir ma fiche d'impôts ! Véritable touche-à-tout, tu as su faire cohabiter ton travail avec notre vie de famille et cela en plus d'être un modèle pour nous. C'est fort ! J'ai adoré discuter avec toi, de physique, d'astrophysique, de particules, de trous noirs et de matière sombre et de tout le reste ! Mais n'oublie pas : les super-héros *aussi* doivent se reposer !

Merci à vous, les **Cockings'z**. Il faut que je vous remercie un par un, bien sûr, mais vous avez été, en groupe, une aventure incroyable. Vous êtes les représentants d'une extraordinaire amitié que je chérie plus que tout. Vous avez créé une dynamique et une complicité que peu de gens connaissent. Vous compter parmi mes amis les plus proches est un honneur autant qu'une joie. Vous comptez énormément pour moi.

A toi **Masaki**, ou plutôt devrais-je dire Koun, merci beaucoup. Merci d'avoir tracé la voie, par cette extraordinaire alchimie de calme et de douceur mêlés à l'intelligence et à la vivacité d'esprit qui te caractérisent ! Tu as été (et tu es encore !) un modèle pour moi dès nos années collège. Brillant, partout, énervant presque ! Mais tellement accessible et disponible ! L'épreuve Suisse est particulièrement difficile pour **Isona** et toi, mais je sais que vous la traverserez avec succès et je serai là pour vous retrouver dans nos montagnes bien aimées !

A toi **Quentin**, cher ami de longue date. Merci pour ta présence et ton écoute, tes conseils et tes histoires, toujours rocambolesques et hilarantes ! La page berlinoise est tournée, il est temps d'ouvrir en grand le chapitre madrilène avec **Hayley**. Je sais pertinemment qu'il sera riche de rencontres, de rires, de retrouvailles et de bonheur. Ton amitié m'est particulièrement chère Q's, *I'm glad we shared those years together and hope our friendship will last for a couple of years more...*

**Guillaume**, je crois avoir découvert plus récemment, presque tardivement, que, sous la carapace de l'incroyable *cockroach* se laisse petit-à-petit approcher un ami au grand cœur et doué d'une sensibilité incroyable. J'aime à penser que tu es un ami plein de surprises et de ressources, qui ne cessera jamais de m'étonner ! Autant de raisons pour passer plus de temps ensemble, avec **Marie** également, à Bordeaux, à Paris ou ailleurs... A bord d'un paramoteur, d'un kite ou sous une simple voile de monosurface !

Merci à toi **Éric**, pour ta présence et toute l'eau que tu apportes à mon moulin cérébral. Tu me gaves de nouveautés, musicales bien sûr, mais idéologiques aussi beaucoup, avec un point de vue toujours original par rapport au mien. Tu n'imagines pas la chance que j'ai de t'avoir comme ami, tu me pousses chaque jour à avancer. Un grand merci pour tout cela.

Mais que serions-nous sans toi, **Stephen** ? Du 'grand-frère' de la cour de récré, à tes questionnements sur la vie, le boulot et le bonheur, en passant par MFPS, *dacrew*, Première Pro et tout le reste, tu as été incroyable ! Toujours surprenant par tes choix (pertinents !), et plein de ressources, tu es un ami d'une grande valeur. Et après Lyon, Montréal, Singapour, Paris et Londres, où vas-tu atterrir cette fois ?

A toi **Thomas**, merci ! Merci pour toutes ces mimiques, cris *and faces*, pour ton humour éclatant et pour tous les fous-rires qui ont suivi. Merci pour cette résilience dont tu fais preuve

et qui force l'admiration. Je suis particulièrement fier de te compter parmi mes meilleurs amis, vraiment. Go, position du Dragon !

**Medy** enfin, ton absence est douloureuse et j'aurais aimé partager encore tant de choses avec toi. Les classes d'espagnol n'ont pas suffi, on a encore tellement de choses à partager ! Même si Montréal n'est pas *si* loin, tu manques à l'appel et j'aimerais te retrouver pour échanger, avancer, confronter et s'émerveiller de nos vies réciproques.

**Marioncita**, douce amie ! Tu as été longuement présente à mes côtés, lors des épreuves précédentes. Tu n'es pas là aujourd'hui mais je sais que tu m'accompagnes en pensées. Nos discussions prolifiques et nos échanges me manquent, j'espère te revoir très bientôt. Merci pour tout !

Ma chère **Blanquette**, excuse-moi pour mon amitié en pointillé ! Mais je crois que maintenant tu y es habituée ! C'est toujours un bonheur de te retrouver et j'espère que nous créerons de nouvelles occasions pour débattre de la vie avec plus de régularité et d'efficacité !

A vous **mes co-internes**, merci pour votre présence à mes côtés ces 4,5 dernières années ! Vous avez été exceptionnels ! La Love Promo, bien sûr. On a quand même formé un groupe extraordinaire à l'IBP, je pense qu'on y laissera une trace durable. Merci **Créno** pour les histoires du jour (365 x 4 ans = 1460 histoires du jour !), merci **Rezadoudou** pour les surnoms, les repas iraniens, les excès de confiance au billard et les élites, merci **Sinus** pour les foots du mardi, les babyfoots de tous les jours, pour les (très bons) jeux de mots et pour la bonne humeur. Merci **VTH** pour les bouteilles de vinasse descendues, pour l'esprit ronchonchon et

pour la papamobile en chaussette au WES entre autres ! **Clairou**, merci à toi qui nous a préféré à Lyon, qui est mon rayon de soleil matinal depuis que l'on partage ce bureau et qui me remet les pieds sur terre quand je deviens idiot ! *Last but not least*, merci à toi **JD** pour la force tranquille, et pour avoir ouvert la voie de la thèse ! Quelques soient vos aventures par la suite, j'espère bien vous garder proche de mon cœur, même loin de Grenoble.

**Nano**, je ne t'oublie pas, et je me permets de t'inclure dans le groupe IBP, bien que tu aies également une place bien à toi dans mon cœur. Merci cher ami pour ta vivacité et ton humour, pour ta connaissance du système hospitalier et ton sens de la fête. Je pense que nous aurons encore l'occasion de vérifier cela quelques fois... *Et tu pompes, et tu pompes, et tu pompes !*

Le reste de l'équipe IBP ne doit pas se sentir oublié ! Merci à vous **Despipi, Anaboule, Tristanus CromCrom, Axelle, La Truffe, Léa, Twi** surtout (4 ans ! et des barres !), **Fanny** et bien sûr **l'Aplasia** (médullaire) !

Aux **équipes techniques de l'IBP**, qui sont admirables de patience avec les internes que nous sommes. Merci de votre présence et votre résilience en ces temps difficiles. Vous parvenez tous les jours à faire des miracles avec des plannings compliqués, des biologistes et cliniciens exigeants, et des restrictions budgétaires infinies. La passion que vous mettez à effectuer votre travail au service du patient vous honore. Un merci tout particulier pour les équipes du DAI, le pré-ana et la bactério, et surtout les équipes de **sérologie virale** et de **BioMol** (vous avez vu, il fait beau et chaud !). Merci également aux **biologistes** qui m'ont transmis leur savoir et leur expertise pendant ces quatre années d'internat.

A toute l'équipe des **Mastres**, vous êtes phénoménaux ! Vous avez été une belle *team* pour animer les soirées grenobloises (#laChèvredeIHCE) et vous restez une bande de copains hors

du temps. J'avoue que le match Martinique vs Salle Mimosa était un peu perdu d'avance (surtout avec les espèces de Quèchts que vous êtes), mais j'espère être de la partie pour le prochain voyage ! Un grand merci à vous, **LeKing, Braymastre, Duchesse, Pluch', Julian, DJ Galletos, Cam'ctus, Nico & Emilie**. Sans oublier **Thibault & Palo, Gaillard, Laurazépam, Georges, Yarbo, La Duv'**, et les autres !

Merci à toi **Lucie** pour toutes ces années partagées. La vie nous a poussé à prendre des routes différentes, aussi j'espère que celle que tu choisis t'apportera du bonheur, de la félicité, et cette extraordinaire joie de vivre qui te va si bien. Je garde de mon côté des légumes, quelques chansons et listes de lectures, deux ou trois livres et un réveil à 6h30 ! Non bien entendu, ce n'est pas tout. Mais la liste tout entière de ce que tu me laisses ne tiendrait pas ici. Alors je te dis simplement merci pour tout cela et pour tout le reste. Très bonne continuation à toi.

A toi bonne **Boussole**, j'ai aussi envie de te remercier grandement ! Pour le support, l'écoute, les brunchs, les verres de vins dégustés, les restos partagés (on pourrait faire un guide à force, non ?), pour l'éternelle balance « orthodontie *versus* robes », pour ta tête de caboche, pour nos prises de bec. Mais aussi pour ta confiance et ta présence. Pour apprendre à devenir exigeant et critiquer. Pour tout le reste, j'en oublie certainement ! Merci de m'avoir aidé à garder le cap !

**Adrien**, merci pour ton amitié, aussi mauvais sois-je pour garder le contact et pour planifier des petits repas à la maison, tu as toujours été là. Drôle, brillant, inlassablement heureux, je ne t'ai pratiquement jamais vu quitter ton sourire ! Toi aussi, légèrement en avance sur moi, tu as toujours ouvert la route, et m'as ouvert les yeux sur bien des questions. A moi de renouer



avec des rendez-vous plus fréquents ! J'en profite pour vous souhaiter le meilleur avec **Agathe**, mais je sais que tu excelles quand il s'agit de trouver le bonheur !

Mon cher **Guillaume**, cher ami montagnard. Je te dois des remerciements tout particuliers puisque c'est grâce à toi que j'ai atterri ici à Grenoble. Tu as été particulièrement stimulant pendant notre externat, que ce soit en sous-conf' où tu savais déjà tout et me poussais à en apprendre chaque jour d'avantage, ou sur un mur d'escalade où tu me motivais à grimper toujours plus haut. Je garde un souvenir extraordinaire de chacune de nos courses en montagne, je fais simplement le vœu de pouvoir ajouter avec toi de nouveaux itinéraires à notre palmarès !

A vous **Agnès & Jean**, mes chers amis Libériens. Merci de m'avoir accueilli et accepté dans votre projet complètement fou ! L'expérience africaine a été d'une incroyable richesse et l'amitié que j'en retire prends une saveur particulière. Celle du soleil, de l'humidité, de la piste défoncée et du bougainvillier. Je garde un souvenir extraordinaire de cette aventure (Aaahh Foya !), j'espère continuer à trouver des histoires pour vous faire rire encore longtemps !

A tous ceux que je n'ai pas cité, qu'ils ne se sentent pas oubliés. La vie nous éloigne puis nous rapproche et je suis certains que nous nous retrouverons bientôt ! En vrac, merci à vous : **Paulo W, JB, Lolotte, Pich'Pich', Judo** évidemment, **GVG, Béné, Mini-Mouff', AmorAmor, Batcha, Joven, Clairette & JeanMi, Alizée & Benoît ; Guillaume, Pauline et Charlou** bien entendu, **Paola & Nat', Romy, Gautier & Fanny, Maxou** (il faut absolument qu'on se voit plus souvent l'animal !), **Baptou, Katharina, Marie & Benoît, Olivia & Bruno, Tata Ben, Philippe, Tonton Pedro & Béatrice**, et tous les autres !

## RÉSUMÉ

L'exploration du microbiote respiratoire dans le cadre du suivi des patients atteints de mucoviscidose nécessite à ce jour la mise en culture de prélèvements respiratoires ou l'utilisation de PCR sélectives. Ces techniques présentent d'importantes limitations et ne sont pas adaptées à une exploration pan-microbienne qui serait nécessaire pour mieux comprendre les mécanismes physiopathologiques à l'œuvre dans cette pathologie et dans son évolution. Nous présentons ici une approche novatrice pour l'exploration de ce microbiote respiratoire et les premiers essais réalisés au sein du laboratoire pour la caractériser. Cette méthode est basée sur le séquençage de nouvelle génération (NGS) des échantillons après un enrichissement par capture semi-ciblée en milieu liquide. Cette capture est réalisée par hybridation sur sondes des acides nucléiques d'intérêt. La réalisation de cet enrichissement a nécessité la sélection de cibles d'intérêt pour la fabrication d'une puce de capture pan-microbienne. Nous montrons dans une deuxième partie de ce travail les résultats obtenus à l'aide de cette puce DB-Chip pour la détection de bactéries, virus et champignons au sein d'échantillons artificiels. Cette approche retrouve les pathogènes attendus sur les prélèvements dont la composition est connue. L'identification des micro-organismes est possible jusqu'au rang d'espèce pour les bactéries et les virus. Les performances de notre puce de capture ne sont pas inférieures à celles obtenues par d'autres équipes avec des approches axées sur un seul règne (viral ou bactérien). Notre approche est prometteuse mais nécessite des ajustements pour son utilisation en clinique.

## ABSTRACT

Exploration of the microbiota living within the cavities of the human body requires new tools that are to be developed in order to overcome the limitations of the traditional culture-based methods. Although PCR and other recently-developed approaches are commonly used for clinical diagnosis of lung infections in cystic fibrosis patients, the narrow spectrum of such techniques does not depict the full picture of the lung microbiota. Metagenomics however enable the microbiologist to assess the whole diversity held within our airways and lungs. In order to routinely benefit from this tool, and apply this approach in a clinical study of the lung dysbiosis, we set up and tested an analytical pipeline combining next-generation sequencing (NGS) with liquid-based probe-aided capture of the nucleic acids present in samples. The findings of the experiments performed are presented in this work as preliminary results for the full implementation of this approach. We first tested a commercial capture array on both artificial and clinical samples. Experience thus gained helped us develop a novel pan-microbial capture array that was finally used to perform advanced testing on artificial samples. We show that identification of bacteria and viruses is possible up to the species level. Our custom design of a capture array showed no inferiority compared to previously published work with separate bacterial or viral designs. While remaining testing is still to be performed on clinical samples and reproducibility to be assessed, this work brings promising results for the exploration of microbiota and large-scale clinical studies.

# TABLE DES MATIERES

<b>REMERCIEMENTS</b> .....	<b>I</b>
<b>RÉSUMÉ</b> .....	<b>X</b>
<b>ABSTRACT</b> .....	<b>XI</b>
<b>TABLE DES MATIERES</b> .....	<b>1</b>
<b>TABLE DES ILLUSTRATIONS</b> .....	<b>3</b>
<b>LISTE DES ABREVIATIONS</b> .....	<b>5</b>
<b>INTRODUCTION</b> .....	<b>7</b>
<b>PARTIE I – DONNEES BIBLIOGRAPHIQUES</b> .....	<b>8</b>
1. Mucoviscidose.....	8
1.1. Pathologie.....	8
1.1.1. Origine génétique .....	8
1.1.2. Expression extra-pulmonaire.....	9
1.1.3. Symptomatologie respiratoire.....	10
1.2. Evolution naturelle de la maladie.....	12
1.1.1. Evolution temporelle .....	13
1.1.2. Exacerbations.....	14
1.1.3. Traitements.....	17
2. Microbiote respiratoire.....	20
2.1. Définitions.....	20
2.1.1. Microbiote .....	20
2.1.2. Virome.....	21
2.1.3. Dysbiose .....	23
2.2. Microbiote et symptomatologie respiratoire.....	24
2.2.1. Microbiote respiratoire bactérien .....	27
2.2.2. Microbiote respiratoire viral.....	32
2.2.3. Microbiote respiratoire fongique .....	38
3. Exploration du microbiote respiratoire .....	40
3.1. Origine du prélèvement.....	41
3.2. Techniques actuelles de détection .....	42
3.2.1. Culture(s).....	42
3.2.2. Spectrométrie de masse .....	43
3.2.3. PCR spécifiques ciblées .....	44
3.2.4. PCR multiplexes syndromiques.....	45
3.2.5. Identification par amplification et séquençage d'un gène générique (16S, 28S ou ITS) .....	46
3.2.6. Antibiogrammes, anti-fongigrammes et anti-virogrammes.....	48
3.2.7. Méthode(s) de référence.....	49
3.3. Nouvelles approches .....	49

3.3.1.	NGS pour Next Generation Sequencing .....	50
3.3.1.	Métagénomique .....	51
3.3.2.	Shotgun sequencing.....	54
3.3.3.	Enrichissement.....	55
4.	Métagénomique infectieuse semi-ciblée au CHU Grenoble – Alpes.....	56
4.1.	Choix technologiques .....	56
4.1.1.	Enrichissement.....	56
4.1.2.	Séquençage.....	59
4.1.3.	Analyse bioinformatique.....	59
4.2.	L'étude Dysbiose.....	61
4.2.1.	Problématique(s) .....	61
4.2.2.	Population et échantillons .....	61
4.2.3.	Intérêt de l'étude et résultats attendus .....	64
<b>PARTIE II – MATERIELS ET METHODES.....</b>		<b>66</b>
1.	Mise en œuvre expérimentale.....	66
1.1.	Echantillons artificiels .....	67
1.2.	Echantillons patients .....	68
1.2.1.	Condensats d'air exhalé.....	68
1.2.2.	Expectorations .....	69
1.3.	Analytique.....	69
1.3.1.	Extraction et synthèse de cDNA .....	69
1.3.2.	Fragmentation .....	70
1.3.3.	Préparation des bibliothèques.....	70
1.3.4.	Size selection et amplification pré-capture .....	72
1.4.	Capture par hybridation en milieu liquide .....	73
1.5.	Séquençage.....	73
1.6.	Analyse bio-informatique .....	74
2.	Design d'une puce de capture pan-microbienne DB-Chip.....	76
2.1.	Base de données.....	76
2.2.	Clustering.....	77
2.3.	Validation <i>in-silico</i> .....	77
3.	Validation de la puce de capture DB-Chip .....	78
3.1.	Echantillons.....	78
3.1.1.	Echantillons artificiels .....	78
3.1.2.	Echantillons de patients.....	80
3.2.	Protocole analytique.....	81
<b>PARTIE III – RESULTATS .....</b>		<b>83</b>
1.	Evaluation de la méthode analytique .....	83
1.1.	Préparation des échantillons.....	83
1.2.	Résultats du séquençage .....	86
2.	Design de la puce de capture DB-Chip.....	90
3.	Validation de la puce de capture DB-Chip .....	99
<b>PARTIE IV – DISCUSSION .....</b>		<b>108</b>

CONCLUSION .....	119
REFERENCES.....	121
SERMENT D'HIPPOCRATE.....	134

## TABLE DES ILLUSTRATIONS

<b>Figure 1</b> : Séquences ADN et protéiques autour de la mutation $\Delta F508$	9
<b>Figure 2</b> : Types cellulaires constitutifs de l'épithélium respiratoire et expression de la protéine CFTR	11
<b>Figure 3</b> : Synthèse des mécanismes mis en jeu dans la mucoviscidose et à l'origine de la symptomatologie pulmonaire	12
<b>Figure 4</b> : Différentes relations possibles entre organismes	24
<b>Figure 5</b> : Evolution du microbiote respiratoire bactérien chez les patients atteints de mucoviscidose en fonction de l'âge	28
<b>Figure 6</b> : Diversité du microbiote bactérien en fonction de l'âge chez le patient mucoviscidose	29
<b>Figure 7</b> : Principe du shotgun sequencing	54
<b>Figure 8</b> : Capture des séquences nucléotidiques par des sondes biotinylées en milieu liquide	58
<b>Figure 9</b> : Procédures classiques d'analyse de données métagénomiques	60
<b>Figure 10</b> : Logigramme des étapes pré-analytiques, analytiques et post-analytiques	67
<b>Figure 11</b> : Schéma des insertions réalisées dans le cadre de la préparation des librairies	71
<b>Figure 12</b> : Principe du séquençage (Dual-Indexed Sequencing) sur cellule avec brins appariés (Paired-end Flow-Cell)	74
<b>Figure 13</b> : Principe de fonctionnement de l'algorithme Kraken	75
<b>Figure 14</b> : Plan expérimental de la mise au point de la méthode de capture	81
<b>Figure 15</b> : Gel d'électrophorèse des acides nucléiques obtenus après extraction et synthèse du cDNA des 6 échantillons initiaux et avant synthèse des libraires	84

<b>Figure 16</b> : Gel d'électrophorèse des acides nucléiques obtenus après synthèse des bibliothèques et avant capture	85
<b>Figure 17</b> : Enrichissement relatif en séquences d'intérêt contrôlé par qPCR pré- et post-capture (puce VirCapSeq-VERT) d'un contrôle fourni par Roche.	86
<b>Figure 18</b> : Répartition des reads issus du séquenceur pour les 6 échantillons initiaux	87
<b>Figure 19</b> : Distribution des séquences (humaines et non-humaines) pour les six échantillons testés dans le cadre de la validation de méthode analytique	88
<b>Figure 20</b> : Attribution taxonomique et répartition des séquences issues des trois échantillons 'positifs' (témoin positif et expectorations de patients)	89
<b>Figure 21</b> : Répartition des séquences des génomes viraux après sélection des virus à tropisme humain uniquement en fonction de la longueur de séquence	93
<b>Figure 22</b> : Visualisation de la base de données des virus eucaryotes (hôte humain)	94
<b>Figure 23</b> : Répartition des séquences de la base de données de capture pour les phages en fonction de la longueur de celles-ci.	95
<b>Figure 24</b> : Répartition des séquences a) bactériennes du gène 16S, b) bactériennes du gène 23S, c) fongiques pour l'élément ITS, d) fongiques du gène 28S.	97
<b>Figure 25</b> : Résultat du test in-silico pour les virus : mapping de la base de données post-clustering sur des séquences de références absentes de la base de départ	98
<b>Figure 26</b> : Score de qualité obtenu au séquençage	100
<b>Figure 27</b> : Enrichissement relatif en séquences d'intérêt contrôlé par qPCR pré- et post-capture (puce DB-Chip) de différentes cibles bactériennes	101
<b>Figure 28</b> : Classification taxonomique des reads obtenus en fonction du panel considéré et de l'algorithme de classement mis en œuvre.	102
<b>Figure 29</b> : Attribution taxonomique à l'espèce des séquences obtenues pour l'échantillon contenant un mélange de bactéries (99,5 %) et de virus (0,5 %) et enrichi par la puce de capture DB-Chip.	103
<b>Figure 30</b> : Métrique de performance finale	106

## LISTE DES ABREVIATIONS

ADN :	Acide Désoxyribonucléique
ARN :	Acide Ribonucléique
BPCO :	Bronchopneumopathie Chronique Obstructive
CAE :	Condensat d’Air Exhalé
cDNA :	<i>complementary DNA</i>
CF :	<i>Cystic Fibrosis</i>
CFTR :	<i>Cystic fibrosis transmembrane conductance regulator</i>
CMV :	Cyto-Mégalo Virus
CPP :	Comité de Protection des Personnes
CRCM :	Centre de Ressources et de Compétences de la Mucoviscidose
CRP :	<i>C-Reactive Protein</i>
DDB :	Dilatation des bronches
DMIS :	Département de Méthodologie et de l’Information de Santé
DNA :	<i>Deoxyribonucleic acid</i>
EBV :	Epstein-Barr Virus
IFN :	Interféron
IgG :	Immunoglobuline de type G
IgM :	Immunoglobuline de type M
NGS :	<i>Next-Generation Sequencing</i>
OTU :	<i>Operational Taxonomic Unit</i>
PCR :	<i>Polymerase Chain Reaction</i> ou amplification en chaîne par polymérase
qPCR :	<i>Quantitative (Real-Time) Polymerase Chain Reaction</i>
PEG :	Poly-Ethylène Glycol
pb :	Paire de bases
RNA :	<i>Ribonucleic acid</i>
RT-PCR :	<i>Reverse Transcription – Polymerase Chain Reaction</i>
SDRA :	Syndrome de Détresse Respiratoire Aiguë



SIDA :           Syndrome d'Immunodéficience Acquise  
VEMS :           Volume Expiré Maximal Seconde  
VIH :            Virus de l'Immunodéficience acquise Humain  
VRS :            Virus Respiratoire Syncytial

## INTRODUCTION

Les infections virales respiratoires sont la pathologie la plus fréquente dans le monde (1). Avec des sévérités différentes, allant du simple ‘coup de froid’ à des complications extrêmement délétères telles que le syndrome de détresse respiratoire aiguë (SDRA), l’impact sur l’Homme de ces infections est majeur. Ces infections respiratoires, souvent virales, concernent tous les âges de la vie, mais les deux extrêmes – jeunes enfants et personnes âgées – sont particulièrement touchés : les nourrissons et les très jeunes enfants supportant le plus grand nombre de viroses respiratoires dans le monde (2). Certaines pathologies respiratoires chroniques, telles que la mucoviscidose, créent un terrain favorable à ces infections qui deviennent, à leur tour, partie prenante dans l’évolution de la maladie. C’est cette interaction entre les infections respiratoires et les exacerbations de la pathologie que nous souhaitons explorer dans le cadre d’une étude clinique. Les modalités techniques d’étude de cet environnement microbien constituent un important travail préliminaire. C’est de ce travail que nous préciserons, au fil de ce manuscrit, l’approche méthodologique originale développée au CHU Grenoble – Alpes.

# PARTIE I – DONNEES BIBLIOGRAPHIQUES

## 1. Mucoviscidose

La mucoviscidose est la plus fréquente des pathologies génétiques. Elle touche environ une personne sur 5000 en France. Il s'agit d'une pathologie grave, incurable en l'état actuel des connaissances et dont la principale conséquence est le raccourcissement de l'espérance de vie des patients atteints, qui se situe autour de quarante ans (3). Dans les pays moins développés, l'espérance de vie est encore plus réduite avec une majorité de décès survenant avant une vingtaine d'années. On assiste cependant à une augmentation récente de l'espérance de vie : depuis une dizaine d'années, la population de patients atteints comporte plus d'adultes que d'enfants dans les pays développés. Or, l'importance de l'atteinte pulmonaire dans la mucoviscidose est majeure chez l'adulte, encore plus que pour le jeune enfant. Une meilleure compréhension des mécanismes physiopathologiques au niveau pulmonaire est nécessaire à l'amélioration de la qualité de vie de ces patients qui ont gagné en espérance de vie.

### 1.1.Pathologie

La mucoviscidose, ou *cystic fibrosis* (CF) en anglais, est une pathologie complexe dont l'expression chez les patients atteints varie selon le système ou l'organe considéré. Elle engendre de ce fait des manifestations cliniques très diverses, la symptomatologie pulmonaire n'étant qu'une facette de la pathologie parmi d'autres.

#### 1.1.1. Origine génétique

Le *primum movens* de la mucoviscidose est la mutation du gène CFTR (*Cystic Fibrosis Transmembrane conductance Regulator*) situé sur le chromosome 7 (en position q31.2) dont

la traduction protéique code pour un canal transmembranaire perméable aux ions chlorures et situé au sein de la membrane apicale de nombreuses cellules, dont les cellules épithéliales retrouvées dans les épithéliums sécrétants. La mutation la plus fréquemment retrouvée est la délétion  $\Delta F508$  (**Figure 1**) et correspond à la perte nette d'un acide aminé (une phénylalanine) dans la protéine mutée. Cependant, plus de 1800 mutations distinctes ont été décrites (3), toutes étant à l'origine d'une altération de la fonction de cette protéine et donc responsables de la survenue des symptômes. La transmission de la mutation  $\Delta F508$  se fait un sur mode autosomique récessif classique (3).

### Normal sequence

DNA	5'	...	AAT	ATC	ATC	TTT	GGT	GIT	...	3'
Protein			Asn	Ile	Ile	Phe	Gly	Val		
Position			505	506	507	508	509	510		

### Mutated DNA

DNA	5'	...	AAT	ATC	AT	-	-	T	GGT	GIT	...	3'
Protein			Asn	Ile	Ile	-		Gly	Val			
Position			505	506	507	508	509	510				

**Figure 1** : Séquences ADN et protéiques autour de la mutation  $\Delta F508$

#### 1.1.2. Expression extra-pulmonaire

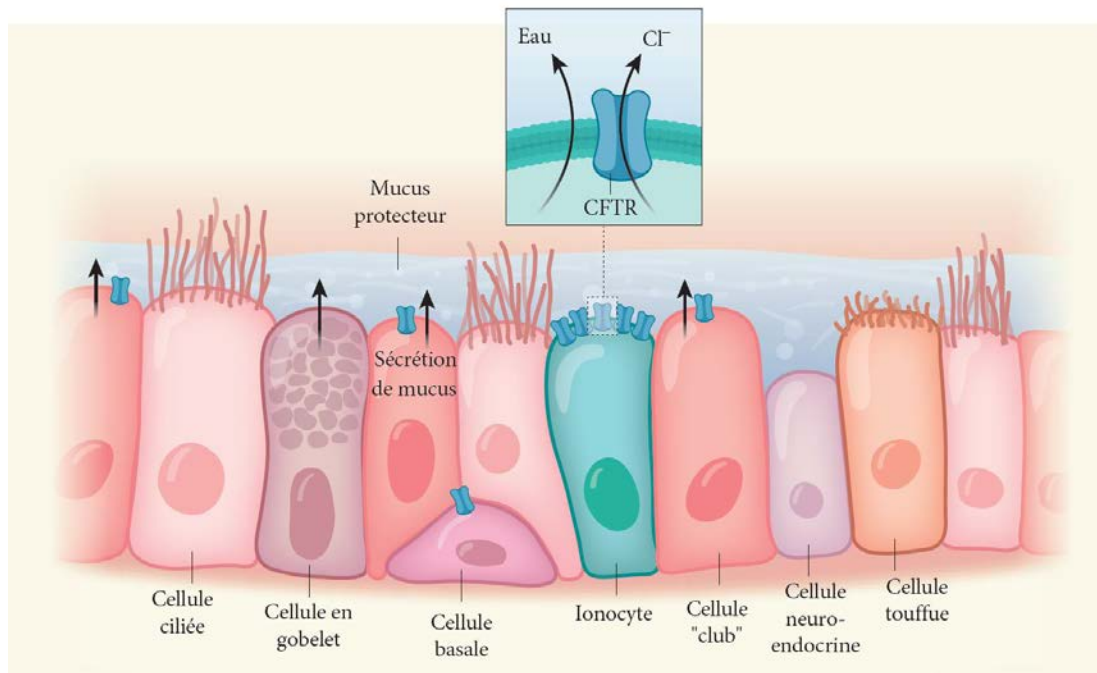
Une des principales manifestations de la pathologie est digestive. C'est d'ailleurs par ce mode d'expression qu'elle a été caractérisée pour la première fois en 1938 (4). La présence d'une mutation sur le gène CFTR est à l'origine de la synthèse d'une protéine défailante à la membrane apicale des cellules épithéliales des canaux sécrétoires du pancréas exocrine. La présence d'une protéine CFTR inactive ou partiellement fonctionnelle modifie la rhéologie du

suc pancréatique excrété, et donc sa viscosité. La sécrétion d'enzymes pancréatiques dont le rôle dans la digestion est crucial est ainsi compromise par l'épaississement de ce suc pancréatique. Cette anomalie sécrétoire est responsable d'une malabsorption digestive caractéristique qui entraîne une malnutrition associée à un déficit de prise de poids chez le patient atteint, et ce, depuis son plus jeune âge. On note également des conséquences à plus long terme avec, en particulier, la dégradation du pancréas exocrine par activation des enzymes protéolytiques au sein du tissu pancréatique. Il en résulte une aggravation de l'insuffisance pancréatique exocrine mais également une insuffisance pancréatique endocrine : c'est le diabète *mellitus*.

### *1.1.3. Symptomatologie respiratoire*

La protéine CFTR présente à la membrane apicale des cellules épithéliales de la muqueuse respiratoire est responsable de la sécrétion d'ions chlorures dont la présence au sein du mucus respiratoire conditionne la rhéologie de ce dernier. Elle est également exprimée en très forte concentration par un type cellulaire récemment décrit (5–7) : le ionocyte pulmonaire. Ces propriétés permettent d'obtenir un mucus hydraté relativement fluide (**Figure 2**) qui est éliminé par la motilité des cils tapissant ce même épithélium (escalator muco-ciliaire). Ce mucus est responsable de l'arrêt et de l'élimination de 90 % des pathogènes respiratoires (1).

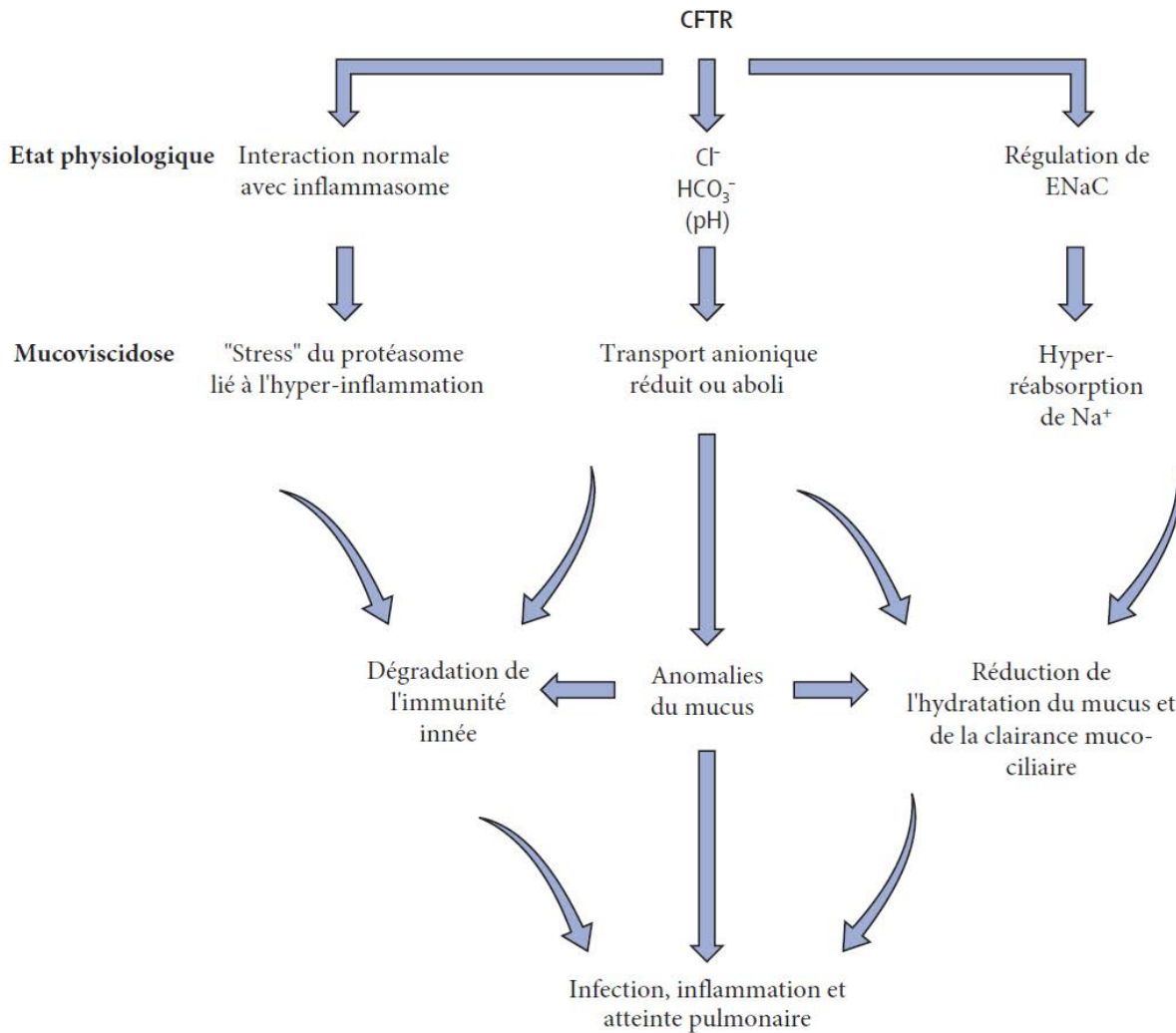
En présence d'une mutation sur le gène CFTR chez les patients atteints de mucoviscidose, la sécrétion d'ions chlorures par le canal CFTR n'a pas lieu, ou l'est en quantité insuffisante. Le mucus de l'épithélium respiratoire s'en trouve très épaissi. Cette augmentation de la viscosité est à l'origine d'une moindre clairance du mucus par la motilité ciliaire et cause la stagnation de ce dernier, alors sujet à sa colonisation par des micro-organismes présents dans l'air inspiré. On parle de défaut de la clairance muco-ciliaire (3).



**Figure 2** : Types cellulaires constitutifs de l'épithélium respiratoire et expression de la protéine CFTR.

Adapté de Travaglini et al. 2018 (7)

La présence de micro-organismes dans le mucus est à l'origine d'une inflammation locale. Cela contribue à la symptomatologie pulmonaire, qui semble être également expliquée par un deuxième phénomène : l'absence de la protéine CFTR ou la présence d'une protéine défectueuse semble générer un microenvironnement local toxique pro-inflammatoire indépendamment de toute colonisation microbienne. En effet, la mutation  $\Delta F508$  altère la capacité de la protéine CFTR à inhiber la cascade de signalisation de la voie NF $\kappa$ B en se liant aux récepteurs pro-inflammatoires (8). Les différents mécanismes à l'œuvre sont résumés dans la **Figure 3** ci-dessous.



**Figure 3 :** Synthèse des mécanismes mis en jeu dans la mucoviscidose et à l'origine de la symptomatologie pulmonaire. Adapté de Elborn, 2016 (3)

## 1.2. Evolution naturelle de la maladie

La pathologie évolue à bas bruit mais en continu pour finalement altérer de façon importante la qualité de vie des patients atteints et ainsi grever le pronostic global à long terme. Cette évolution met en place, dès le jeune âge, un cercle vicieux qui auto-entretient le développement de la pathologie respiratoire : l'obstruction chronique liée au défaut de clairance muco-ciliaire est à l'origine d'une colonisation microbienne chronique. Celle-ci est responsable de l'inflammation locorégionale, elle-même favorisant l'obstruction des voies aériennes (9). Le défaut d'élimination des bactéries (particulièrement dû à la modification du

pH) semble être le premier pas de ce cercle vicieux qui contraint la capacité respiratoire en faisant passer les voies aériennes et les poumons d'un état « stérile » à un état colonisé (10).

Outre la réduction de l'espérance de vie, la qualité de vie de ces patients atteints est très altérée. La fréquence importante des hospitalisations en est le principal responsable. En moyenne, ces patients sont hospitalisés dix fois par an, avec une durée moyenne de séjour de neuf jours pour les exacerbations de la pathologie au plan pulmonaire (11). A ce jour, les facteurs qui contribuent à la dégradation de la fonction respiratoire (comme en témoigne la diminution du VEMS) chez le patient stable par ailleurs restent largement inconnus (12).

### 1.1.1. *Evolution temporelle*

La mucoviscidose s'exprime dès la vie intra-utérine : la pathologie commence avant même l'apparition des symptômes. Une part du remodelage des voies aériennes est congénitale : on observe par exemple l'obstruction de celles-ci ainsi que le phénomène de *trapping* (piégeage) chez les jeunes cochons cftr  $\Delta F508$ , et ce, avant l'apparition de l'inflammation et en l'absence de mucus (10). Ces données semblent transposables à l'homme : dans la même étude, 15 % des enfants atteints de moins de 16 mois présentaient une trachéomalacie visible à la bronchoscopie et associée aux symptômes respiratoires. L'importance d'un diagnostic précoce est évidente et celui-ci fait aujourd'hui partie, en France, et depuis 2002, du dépistage néo-natal, dit test de Guthrie, réalisé entre 72h et 96h de vie chez tous les nouveau-nés (13). Ce dépistage se fait par dosage de la trypsine immuno-réactive dans le sang. Il s'agit d'une enzyme pancréatique normalement absente du sang périphérique et dont la présence fait fortement suspecter la pathologie. En cas de positivité, une étude génétique à la recherche de mutations est réalisée.



Témoins de la précocité des atteintes, on observe des anomalies radiologiques dès 3 mois de vie au scanner thoracique (10). Les broncheectasies ou dilatation des bronches sont à l'origine d'un trouble ventilatoire par augmentation du calibre d'une partie de l'arbre respiratoire (atteinte des bronches et bronchioles). Cette broncheectasie est rapportée chez un enfant atteint sur trois dès l'âge de 3 ans (10). Il semble que plusieurs mécanismes moléculaires soient à l'œuvre pour expliquer ces modifications. Le principal agent causal est l'élastase des polynucléaires neutrophiles, et dans une moindre mesure, la cathepsine-5 et les MMP (*matrix-metalloproteinases*) (3). Cette broncheectasie rend progressivement sensible aux infections par des bactéries à Gram négatives. Or, ces infections ne sont habituellement observées que chez les immunodéprimés et chez les patients pour lesquels la barrière épithéliale a perdu son intégrité. Il semble que ceci soit en partie dû à une sécrétion muqueuse trop acide au niveau de l'épithélium, elle-même causée par l'absence de sécrétion de bicarbonates lorsque la protéine CFTR est déficiente ou déficitaire (10).

### 1.1.2. Exacerbations

La trajectoire clinique des patients atteints de mucoviscidose est ponctuée d'épisodes aigus de réduction de la capacité fonctionnelle respiratoire et dénommés exacerbations. Il n'existe pas de critères stricts de la définition d'exacerbation (14–17), néanmoins, l'ensemble du corps médical s'accorde à lister les symptômes suivants, tous n'étant pas nécessairement présents pour confirmer l'atteinte (**Tableau 1**).

La réduction de la capacité pulmonaire est mesurée par le volume expiré maximal en une seconde (VEMS). La diminution du VEMS est fréquemment retrouvée dans les exacerbations. En revanche, il existe peu de signes radiologiques de l'exacerbation. La radiographie pulmonaire ne montre généralement pas de modifications par rapport à un cliché de

référence, tant au niveau du parenchyme pulmonaire qu'au niveau des voies aériennes. Enfin, la dyspnée peut être décelée lors de l'analyse des gaz du sang avec la mise en évidence d'une hypoxie avec désaturation, sans que l'anomalie biologique soit requise pour poser le diagnostic d'exacerbation (18). Enfin, la définition de l'exacerbation semble différer entre les adultes et en pédiatrie (19).

**Tableau 1** : Symptômes pulmonaires participant à la définition de l'exacerbation chez un patient atteint de mucoviscidose. (adapté de UpToDate : "Cystic fibrosis: Treatment of acute pulmonary exacerbations", consulté le 20/07/2018)

Symptômes
Apparition d'une toux ou aggravation d'une toux préexistante
Apparition de glaires ou augmentation de la production de glaires
Diminution de la tolérance à l'exercice physique ou apparition d'une dyspnée ou aggravation d'une dyspnée existante
Fatigabilité augmentée
Diminution de l'appétit
Augmentation du rythme respiratoire au repos
Modification de la consistance des glaires
Fièvre
Augmentation de la congestion nasale / de l'écoulement nasal

L'exacerbation est une complication de la pathologie qui ne résulte pas complètement en un retour à l'état basal après résolution de l'épisode et sa prévention est une priorité. Dans une étude observationnelle multicentrique récente, 85 % des patients décrivent des symptômes dès 7 jours avant leur admission et un tiers (32 %) jusqu'à 21 jours avant l'hospitalisation (20). Mais une détection précoce ne semble pas pour autant permettre, à ce jour, d'enrayer la diminution de la capacité ventilatoire. Cela est peut-être en lien avec une diminution de la réponse aux antibiotiques lorsque ceux-ci sont trop souvent utilisés (21).

Le plus important facteur de risque de développer une exacerbation au plan pulmonaire et nécessitant l'utilisation d'antibiotiques est d'avoir fait un tel épisode dans l'année précédente (22). Les patients ayant eu trois ou plus épisodes d'exacerbation étaient, dans cette étude, les plus à risque de présenter à nouveau un épisode d'exacerbation. Par ailleurs, il a été montré que plus le nombre d'exacerbations traitées par antibiotiques est élevé, et plus le rythme de déclin de la fonction pulmonaire est important (23). Les facteurs prédictifs d'une mauvaise réponse au traitement et d'une dégradation de la capacité respiratoire sont le sexe féminin, la malnutrition, l'insuffisance pancréatique, la colonisation par *Pseudomonas aeruginosa*, par *Burkholderia cepacia*, ou par un *Staphylococcus aureus* résistant à la méticilline, et la présence d'une aspergillose broncho-pulmonaire (24).

La recherche d'un élément déclencheur de ces exacerbations a conduit à mettre en évidence que 50 % de celles-ci ont pour origine une infection respiratoire virale, expliquant ainsi l'augmentation de l'incidence de ces exacerbations en hiver (25–27). Néanmoins, il persiste encore d'importantes inconnues quant à la survenue de ces exacerbations, de nombreux symptômes retrouvés n'étant pas spécifiques d'un pathogène unique.

La conséquence à court terme de ces exacerbations est la dégradation de la fonction respiratoire. Environ un quart des patients ne retrouvent pas leur capacité fonctionnelle respiratoire après un tel épisode. Il s'en suit une escalade thérapeutique avec une altération de la qualité de vie, en plus d'une évidente majoration du coût des soins. A plus long terme, les exacerbations se traduisent par une diminution de la survie : plus de deux exacerbations par an majorent le risque de décès ou de greffe dans les trois années qui suivent une exacerbation chez l'adulte (28).

Au plan moléculaire, il semble exister un rôle de la protéine CFTR dans l'inflammation liée à l'exacerbation. La localisation normale de celle-ci au sein des radeaux lipidiques membranaires permet normalement une régulation de NFκB pro-inflammatoire (8). Le rôle de CFTR comme celui d'un simple canal ionique est aujourd'hui remis en question. D'abord parce que ce rôle n'est pas retrouvé dans certains modèles animaux (10). D'autre part, parce qu'il joue un rôle central dans la sécrétion des bicarbonates et dans le contrôle du pH muqueux (29,30). Il en résulte probablement une modification ou une diminution de l'activité des peptides antimicrobiens (31).

Au final, une seule cause (la mutation du gène CFTR) est à l'origine de la pathologie mais de multiples conséquences synergiques participent à la survenue d'exacerbations : la diminution de l'hydratation du mucus, l'augmentation de sa viscosité, la diminution de la clairance mucociliaire, l'altération qualitative de la structure du mucus, l'altération de l'adhésion du mucus, le défaut de l'immunité innée, et finalement, une prédisposition à l'inflammation cellulaire locale (3).

### *1.1.3. Traitements*

#### *Traitements de fond*

La prise en charge de l'aspect pulmonaire de la pathologie s'articule autour de sept grands axes. Ceux-ci sont : une prise en charge nutritionnelle adaptée, l'apport de la kinésithérapie respiratoire, la résolution des infections bactériennes, la prévention des pathologies fongiques, la prise en compte des infections respiratoires virales, l'utilisation des anti-inflammatoires dans la gestion de l'inflammation et la réhabilitation respiratoire (28). La prise en charge nutritionnelle permet de garantir un apport énergétique suffisant au regard de la dépense augmentée par l'inflammation et la lutte antimicrobienne permanente. La

kinésithérapie est un des piliers majeurs de la prise en charge de ces patients. Elle vise à mobiliser le mucus, favoriser le drainage bronchique ainsi que la ventilation et le recrutement des surfaces d'échange.

La lutte antibactérienne survient dès le plus jeune âge et consiste en l'administration d'une antibiothérapie par voie inhalée ou systémique. L'antibiothérapie est adaptée à chaque situation clinique (sévérité) et aux antibiogrammes disponibles. Il convient de traiter les patients selon qu'il s'agisse d'une primo-infection (ou primo-colonisation), d'une infection intermittente, ou bien d'une infection chronique. Le traitement antifongique n'est pas encore bien codifié en raison du manque de connaissances disponibles dans ce domaine. La très forte prévalence des colonisations mycologiques ne permet pas de conclure à un rôle pathogène de nombreuses espèces. Outre le traitement de l'aspergillose broncho-pulmonaire allergique à l'aide de corticoïdes, les traitements antifongiques sont administrés en fonction des marqueurs biologiques, cliniques, radiologiques ou selon l'évolution. Les plus utilisés sont les dérivés azolés (28).

Pour la prise en charge spécifique des infections virales, les recommandations sont basées sur une approche triple : prévenir les infections par la vaccination, prévenir les infections par des moyens de lutte génériques et traiter de manière systématique et empirique les épisodes d'allure virale, notamment en période hivernale et d'épidémie grippale.

#### Traitement de l'exacerbation.

Le Centre de Référence de la Mucoviscidose de Lyon a émis des recommandations en Juillet 2017 pour la prise en charge de ces épisodes aigus (28). Il convient bien sûr de contrôler la bonne observance des patients. Mais la priorité est donnée à la mise en place d'une antibiothérapie probabiliste, basée sur les antécédents microbiologiques, et instaurée

précocement. Celle-ci sera adaptée secondairement à l'antibiogramme obtenu et la voie d'administration sera choisie en fonction de la sévérité clinique. Le reste des recommandations rappelle l'aspect pluridisciplinaire de la prise en charge avec l'intervention nécessaire d'un kinésithérapeute, et avec une prise en charge nutritionnelle adaptée.

### Nouveaux traitements mis à disposition

Il s'agit de traitements pharmacologiques de la protéine CFTR. Leur mode d'action varie selon la molécule, certains favorisant l'adressage d'une protéine par ailleurs fonctionnelle à la membrane, d'autres permettant la récupération partielle de la fonction d'une protéine déficiente mais par ailleurs correctement adressée à la membrane. C'est le cas de l'ivacaftor (Kalydeco®), qui est indiquée pour les patients porteurs d'une mutation de classe III. Un deuxième traitement, le lumacaftor (Orkambi®) permet une amélioration de la capacité respiratoire en association avec l'ivacaftor. Ces deux traitements ont reçu des autorisations de mise sur le marché (AMM) entre 2015 et 2016. Une revue de la littérature portant sur 1637 patients recevant la combinaison ivacaftor/lumacaftor ou un placebo montrait une efficacité de ce traitement avec une amélioration de la fonction respiratoire, du score de performance clinique et de l'indice de masse corporelle (IMC) de ces patients (32).

Les essais concernant la thérapie génique sont en cours, sans résultat probant à ce jour. L'administration d'un cDNA de CFTR sauvage par nébulisation au sein d'un vecteur liposomique a été testée au cours d'un essai clinique de phase 2b en Juillet 2015. Un traitement novateur ciblant les patients porteurs d'une mutation de classe I (non-sens) a été testé au sein d'un essai clinique de phase 3. Cette molécule, l'ataluren, n'a pas montré d'efficacité par rapport à un placebo (33).

## 2. Microbiote respiratoire

### 2.1. Définitions

La partie supérieure des voies aériennes respiratoires, depuis l'extrémité antérieure des fosses nasales, jusqu'à l'étage bronchique, est constituée d'un système d'abduction de l'air inspiré et est tapissé d'une barrière muqueuse séparant le milieu extérieur de l'intérieur de notre organisme. La surface de cette barrière muqueuse est estimée à 75 m<sup>2</sup> chez l'adulte (34). L'air inspiré n'étant pas stérile par définition, l'ensemble des voies aériennes supérieures est en perpétuel contact avec un grand nombre de micro-organismes. Cet écosystème constitue, de par sa complexité et par le nombre de micro-organismes concernés, une entité à part entière : le microbiote respiratoire.

#### 2.1.1. *Microbiote*

On désigne, sous le terme microbiote, l'ensemble des micro-organismes, –virus, bactéries et champignons–, qui occupent à un instant donné une loge anatomique d'un organisme plus complexe, et qui entretiennent entre eux et avec l'organisme hôte des relations nombreuses, symbiotiques ou pathologiques. Le microbiote désigne donc tout un écosystème, complexe, et dynamique et dont le terme de 'flore' est parfois synonyme.

Le microbiote est un sujet d'étude relativement récent. Les premiers travaux concernant la définition et l'étude du microbiote datent des années 2010 et se focalisent principalement sur le microbiote digestif bactérien. Aujourd'hui encore, il s'agit d'un domaine de la biologie qui reste peu exploré et dont les implications dans la physiologie humaine et la pathologie sont encore largement sous-documentées et incomprises.

Le microbiote respiratoire commence à être exploré, en particulier chez les patients atteints de mucoviscidose. C'est un écosystème mal compris en ce qui concerne son rôle dans la genèse de la pathologie et dans son évolution. Le microbiote respiratoire viral est peu exploré, tant dans la description de cette flore chez le patient sain, que dans les diverses pathologies respiratoires à l'origine de troubles ventilatoires (mucoviscidose, bronchopneumopathie chronique obstructive, asthme, allergies, ...).

Enfin, le microbiome correspond à l'ensemble des séquences génétiques (ADN et ARN) du microbiote. L'importance du microbiome semble telle que le département d'état du NIH (*National Institute of Health*) américain lui a dédié un projet à part entière (35). On parle également de métagénome. La métagénomique étant la science qui explore la composition en génomes des échantillons complexes (36). On parle enfin d'hologénome pour désigner l'ensemble des séquences génétiques des micro-organismes hébergés par un hôte ajoutées au génome de l'hôte lui-même.

### 2.1.2. Virome

Le mot virome a été créé pour désigner le microbiome viral. Le suffixe -ome est également utilisé dans de nombreux autres domaines de la biologie (protéome, métabolome, transcriptome, exome, ...). Il a été popularisé par l'étude du génome, qui correspond au départ à l'ensemble des gènes, puis finalement à l'ensemble des séquences génétiques d'un organisme. Cette terminologie témoigne de l'explosion du volume de données disponibles sur un sujet ainsi que de l'avènement de l'ère du *big data*. Le terme virome est utilisé ici en ce sens : il s'agit de l'ensemble des séquences d'acides nucléiques (ARN et ADN) correspondantes aux espèces virales présentes au sein d'un microbiote, sans a priori quant à leur rôle pathogène ou non, ni à la cellule cible support de leur répllication. Autrement dit, le virome



est la composante virale du microbiome humain, encore appelé métagénome viral. Dans ce travail, nous utiliserons principalement ce terme de virome.

Nous distinguons trois grands groupes au sein du virome : les virus dits 'eucaryotes' (les virus dont la cellule hôte est une cellule eucaryote), les éléments génétiques dérivés de virus anciens et intégrés au génome humain, et les bactériophages, virus dont la cible est bactérienne (procaryote) (37). Il est établi que le microbiote viral joue un rôle très important dans la disponibilité des nutriments et le cycle énergétique, dans l'immunité et la maturation de celle-ci, et qu'il constitue également un extraordinaire réservoir de diversité génétique. Certaines études ont par ailleurs montré le rôle du virome dans l'inflammation systémique chronique responsable de grands groupes de pathologies telles que les pathologies cardiovasculaires (38).

D'une manière similaire, les termes bactériomes et mycomes sont utilisés pour désigner l'ensemble des séquences génétiques (principalement ADN cette fois-ci) qui appartiennent aux règnes bactérien et fongique respectivement, et qui colonisent une niche écologique donnée. Le microbiote bactérien joue un rôle crucial dans les fonctions de nutrition, d'immunité et d'inflammation. Par exemple, certains taxons bactériens commensaux du tube digestif sont capables d'induire l'expression du gène Foxp3 dans des cellules humaines (39,40). Or, ce gène est un facteur de transcription responsable de l'expression phénotypique caractéristique des lymphocytes T régulateurs qui coordonnent la réponse immunitaire et induisent la tolérance immune. Ainsi, les micro-organismes commensaux semblent favoriser leur reconnaissance par le système immunitaire comme faisant partie du soi.

### 2.1.3. Dysbiose

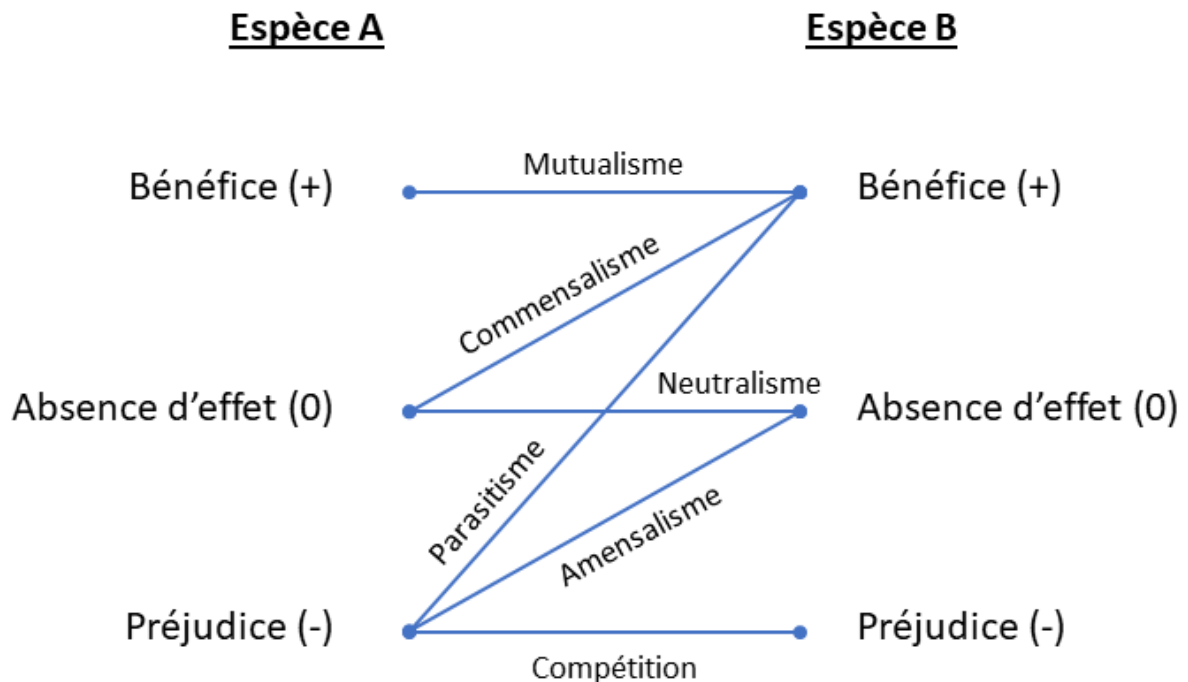
On appelle dysbiose l'état donné d'un microbiote ou d'une flore qui diffère de l'état habituellement retrouvé chez des patients sains, et qui se traduit le plus souvent par l'apparition d'une pathologie chez l'organisme hôte. Cet état, altéré par rapport à un état « basal », se traduit par des modifications qualitatives et quantitatives (absolues ou relatives) des espèces bactériennes, virales, parasitaires ou fongiques qui constituent le microbiote en question (41).

Autrement présenté, le microbiote « sain » est un écosystème fragile qui se trouve dans un état d'équilibre, et dont les interactions entre les différents acteurs symbiotiques se font sur le mode mutualiste ou sur le mode commensal (**Figure 4**). Dans le cas de la dysbiose, l'état d'équilibre qui prédominait jusqu'alors, disparaît au profit d'un nouvel état –déséquilibré– dans lequel prédominent des relations parasitaires.

Cette modification de la flore peut se faire à différentes échelles et selon différents modes. On décrit ainsi des modifications quantitatives relatives, dans lesquelles les proportions des différents micro-organismes présents sont modifiées, sans que l'on note la disparition d'une ou plusieurs espèces données, ni l'apparition de nouvelles espèces jusqu'alors non représentées. Par ailleurs, peuvent être décrites des modifications quantitatives absolues des micro-organismes constituant le microbiote en question. On observe ainsi une raréfaction des composants ou au contraire une prolifération de certains au sein de la loge anatomique concernée.

L'exemple type est celui de la vaginose bactérienne chez la femme : l'important déclin de *Lactobacillus sp*, qui constitue une part conséquente de la flore vaginale physiologique, et la

colonisation par des anaérobies stricts et facultatifs font de la vaginose non pas une infection à proprement parler mais une véritable dysbiose (42).



**Figure 4** : Différentes relations possibles entre organismes. Adapté de Ian Alexander, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=71067142>

Enfin, la dysbiose peut être caractérisée par une altération qualitative du microbiote. Il s'agit alors d'une modification des espèces bactériennes, virales, parasitaires ou fongiques présentes au sein de cette flore. On pourra décrire l'apparition de nouvelles espèces, pathogènes ou non, ou la disparition d'autres, le plus souvent commensales ou mutualistes.

## 2.2. Microbiote et symptomatologie respiratoire

Le microbiote respiratoire des patients atteints de mucoviscidose commence à être exploré. Si peu de données sont encore disponibles concernant l'ensemble du microbiote, en revanche le microbiote bactérien commence à être bien appréhendé. Par ailleurs, des données

commencent à être publiées concernant les virus respiratoires et leur impact sur la fonction pulmonaire de ces patients.

La constitution du microbiote semble suivre le schéma suivant : on assiste d'abord à une explosion rapide et précoce de la diversité avec la survenue de vagues successives d'invasions bactériennes pendant les premiers jours de vie. Puis se constitue un microbiote diversifié (en nombre d'espèces représentées) et stable dans le temps. Enfin, la dernière étape est celle d'un remplacement progressif des espèces par d'autres, qui deviennent majoritaires, associé avec une perte de la diversité initiale (43). Les importantes perturbations de cet écosystème par l'utilisation régulière et répétée d'antimicrobiens altèrent profondément les communautés bactériennes sur le long terme et contribuent à la diminution de la diversité.

Il est probable que le microbiote se constituant au cours des premiers jours de vie d'un enfant atteint de mucoviscidose soit conséquent en nombre de bactéries (quantitativement, en plus de la diversité importante) et qu'il soit à l'origine d'un remodelage précoce et de changements structurels irréversibles survenant très tôt dans la vie du patient, et qui prédisposeraient à un microbiote donné, dont notamment la prédominance quasi systématique à terme de *Pseudomonas aeruginosa* (10).

Il peut sembler contre-intuitif que cette plus grande diversité taxonomique puisse être associée à une meilleure fonction respiratoire. Mais les études s'accumulent pour montrer que la perte de diversité, en particulier bactérienne, est associée à une inflammation chronique, à la croissance de pathogènes et à une atteinte clinique plus prononcée (43).

La mucoviscidose, bien que très clairement définie par une atteinte du gène CFTR, n'est donc pas une pathologie homogène. Selon l'atteinte, c'est-à-dire selon les mutations du gène CFTR présentées par les patients, l'expression clinique n'est pas tout à fait superposable. Et il

semblerait que le microbiote respiratoire de ces patients soit différent en fonction de l'atteinte. Ainsi, les patients homozygotes  $\Delta F508$  montrent spécifiquement une diversité bactérienne des voies aériennes moindre par rapport à des patients atteints de mucoviscidose mais porteur d'autres mutations, ou hétérozygotes. Les patients hétérozygotes ou non- $\Delta F508$  présentent des profils microbiologiques différents (43) avec une perte de diversité étroitement corrélée à l'âge.

Les microbiotes bactériens et viraux sont étroitement liés : l'activité antivirale dégradée dans la mucoviscidose par l'affaiblissement des mécanismes de défense innée contre les virus fait le lit des infections virales respiratoires. Celles-ci sont à leur tour responsables d'une réponse inflammatoire non adaptée qui aggrave la susceptibilité aux infections bactériennes et fongiques, compliquant ainsi les tableaux cliniques (9). Il est clairement établi que les infections virales amènent à la survenue d'exacerbations et à la progression de la maladie (27,44–47) en favorisant également l'adhérence bactérienne aux parois des voies aériennes. L'importance des infections respiratoires virales est probablement sous-estimée car l'infection virale mène à l'infection bactérienne, cette dernière étant le plus souvent prise pour étiologie.

A ce jour, le microbiome respiratoire a surtout été exploré séparément pour ce qui relève du microbiote bactérien, du virome et du microbiote parasitaire et fongique. Pour ce tour d'horizon de la littérature concernant le microbiote respiratoire de l'adulte atteint de mucoviscidose et les liens que l'on retrouve avec la survenue d'exacerbations, nous avons choisi de respecter ce découpage.

Il est cependant une variable à prendre en compte, qui reste largement sous-estimée dans les études jusqu'à présent : c'est la problématique des variations et de l'évolution temporelle du

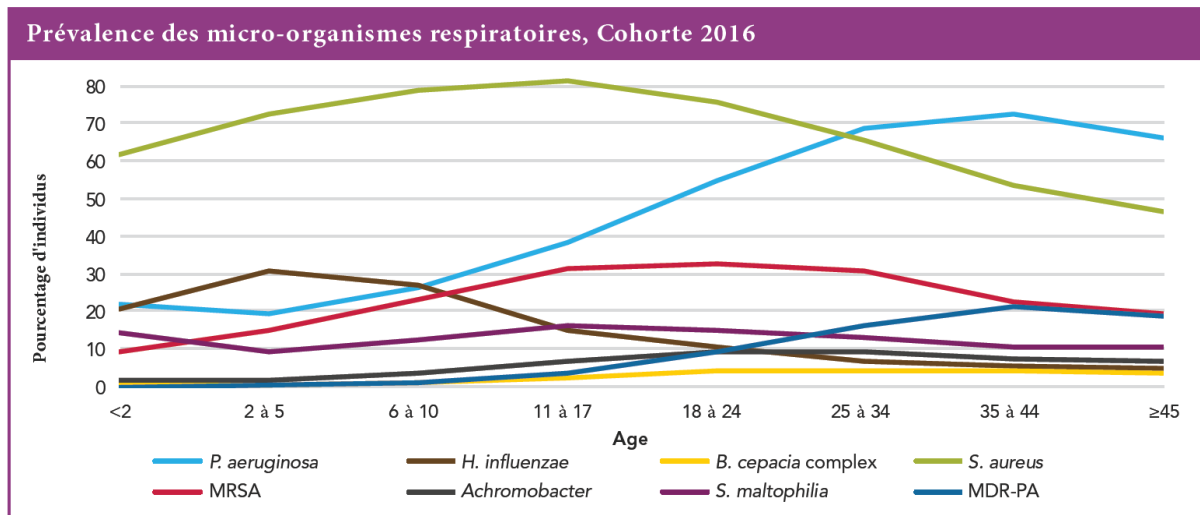
microbiote chez l'homme. Il semble que le microbiote bactérien digestif soit macroscopiquement très stable (48) à long terme, faisant de celui-ci un écosystème où la perte de cette stabilité amène à un état de dysbiose à l'origine de pathologies. Ce n'est pas le cas de tous les microbiotes, Il existe pour certains une évolution rapide. On pourra citer à titre d'exemple le microbiote vaginal (49) pour lequel il existe des variations temporelles sans forcément de corrélation dysbiose – pathologie. Certains microbiotes semblent également présenter des rythmes circadiens. C'est le cas du microbiote intestinal (50), qui en dépit de sa globale stabilité à long terme, présente une fraction qui varie selon le nyctémère de façon périodique. A ce jour cependant, aucune donnée n'est disponible quant à la présence de variations cycliques du microbiote respiratoire, tant chez le patient atteint de mucoviscidose que chez le sujet sain.

### 2.2.1. *Microbiote respiratoire bactérien*

D'importantes études bien conduites commencent à être disponibles concernant le microbiote respiratoire bactérien. La relative simplicité technique des méthodologies employées fait de ce microbiote celui pour lequel le plus de données sont disponibles.

L'histoire naturelle du microbiote respiratoire bactérien (**Figure 5**) semble débiter avec la colonisation par *Haemophilus influenzae* et *Staphylococcus aureus* dans la petite enfance (entre 0 et 4 ans) (52). La colonisation à *Pseudomonas aeruginosa* survient généralement autour de 10 ans (53) et constitue un tournant important dans la pathologie. D'autres bactéries sont fréquemment retrouvées dans les cultures d'expectorations de ces patients. On pourra citer : *Stenotrophomonas maltophilia*, *Achromobacter spp*, *Burkholderia cepacia complex* (3) par exemple. Ces dernières bactéries sont toutes à Gram négatif, et elles ont en commun une importante résistance aux antibiotiques constitutive et acquise. *Burkholderia*

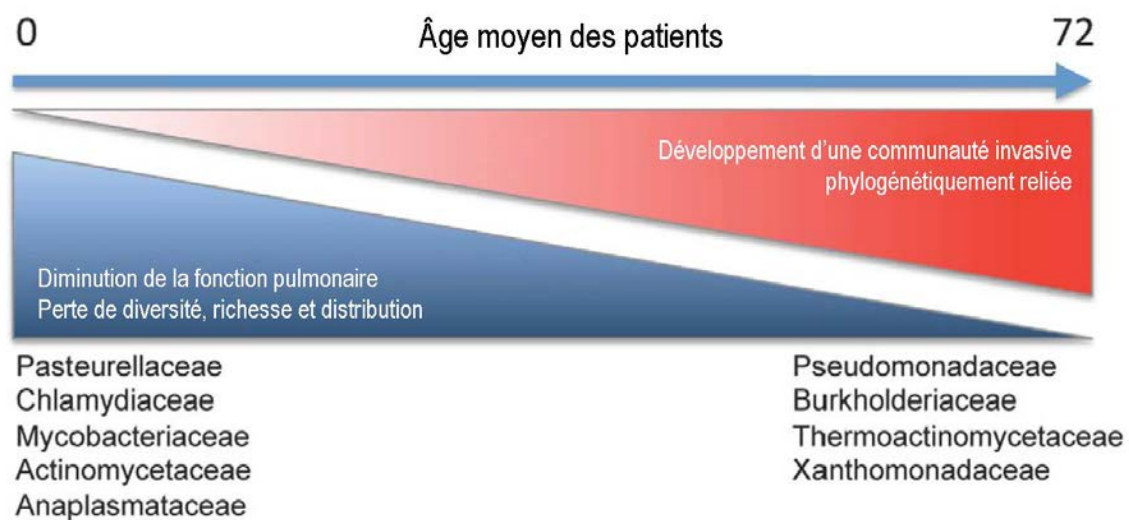
*cepacia complex* et en particulier *Burkholderia cenocepacia* sont particulièrement importantes car la colonisation des voies aériennes par ces dernières est associée à une augmentation de la mortalité (54).



**Figure 5 :** Evolution du microbiote respiratoire bactérien chez les patients atteints de mucoviscidose en fonction de l'âge. Adapté de Cystic Fibrosis Foundation Patient Registry (51). MRSA : *Staphylococcus aureus* résistant à la pénicilline. MDR-PA : Multi-Drug Resistant *Pseudomonas aeruginosa*

Il s'agit là d'un biomarqueur important qui est recherché en pratique courante dans les laboratoires de microbiologie prenant en charge ce type de patient et leurs prélèvements. Par ailleurs, la colonisation par des mycobactéries non-tuberculeuses issues de l'environnement (*Mycobacterium abscessus* et *Mycobacterium avium-intracellulare*) est fréquemment observée. Ceci pose problème car il s'agit de bactéries intrinsèquement résistantes à de nombreux antibiotiques (55). Or les bactéries réunies au sein d'une même flore, ou d'une même loge anatomique échangent régulièrement du matériel génétique. Ce sont des résistances qui sont donc susceptibles de diffuser à d'autres espèces, générant ainsi des bactéries multi-résistantes, qui constituent finalement un véritable challenge thérapeutique au fur et à mesure des hospitalisations et de la vie des patients atteints.

En parallèle de ces données mono-microbiennes, des études ont montré que la présence d'un microbiote bactérien avec une diversité plus large était corrélée à une meilleure fonction respiratoire par rapport à des microbiotes à diversité réduite, voire à la présence d'une seule espèce majoritaire (12,56). Or, la diminution de la diversité bactérienne semble inéluctable avec l'évolution de la pathologie. Cox *et al.* décrivent précisément la diminution de la diversité au cours du temps (43), qu'ils résument dans la figure suivante (**Figure 6**) :



**Figure 6** : Diversité du microbiote bactérien en fonction de l'âge chez le patient mucoviscidose. Adapté de Cox *et al.* (43).

Les mécanismes de cette diminution de la diversité sont mal compris. Toujours d'après Cox *et al.* (43), il existe des modifications des compositions des communautés microbiennes bactériennes différentes entre les états de santé et de pathologie. Les auteurs démontrent l'existence d'assemblages bactériens composés d'un petit nombre de pathogènes spécifiques chez l'enfant et l'adulte atteints de mucoviscidose. Il semblerait enfin que ces assemblages contribuent au déclin de la fonction pulmonaire.

Cependant, d'importantes différences semblent apparaître au sein des patients atteints de mucoviscidose en fonction des mutations portées par les patients. Chez les patients  $\Delta F508$  par



exemple, les patients les plus âgés présentent une colonisation par des phyla de *Moraxellaceae* et de *Sphingobacteriaceae* (43), phyla par ailleurs associés à d'autres pathologies telles que la BPCO (57). Ces patients plus âgés semblent avoir développé un microbiote 'core' stable avec quelques espèces pathogènes, probablement sélectionnées par les multiples vagues d'antibiotiques (43).

Mais ces données sont surtout issues d'études en culture, ou bien réalisées à l'aide de PCR multiplexes ou de PCR 16S. Ces techniques ne sont peut-être simplement pas assez sensibles pour détecter toute la diversité bactérienne de la niche pulmonaire chez les patients atteints de mucoviscidose. D'où l'intérêt et la nécessité de réaliser des études basées sur le NGS pour s'affranchir de la problématique de la surreprésentation des espèces dominantes et détecter la 'biosphère rare' (43).

En termes de quantité de bactéries, la masse bactérienne contenue au sein du microbiote respiratoire est supérieure chez l'animal modèle de la mucoviscidose par rapport à des animaux contrôles exempts de pathologie pulmonaire, et ce, dès le plus jeune âge. Cette observation réalisée sur des furets (10) est, a priori, transposable chez l'humain. En revanche, il ne semble pas y avoir de différence de quantité bactérienne (masse totale) entre les patients atteints de mucoviscidose stables et les patients atteints de mucoviscidose en période d'exacerbation (acides nucléiques de l'ordre de  $10^9$  copies / mL) (58).

Il est intéressant de noter l'absence de différence dans la répartition bactérienne retrouvée, dans une étude de 2008, entre des prélèvements réalisés pendant une période de stabilité et pendant des périodes d'exacerbation en termes d'infection bactérienne (27). Néanmoins toutes les études ne s'accordent pas sur ce résultat : en reliant la diversité initiale avec l'importance des variations de microbiote, Carmody *et al.* ont montré que plus cette diversité

à l'état basal était importante, plus le changement de microbiote bactérien au moment de l'exacerbation était conséquent (58). Cela était par ailleurs d'autant plus vrai qu'il existait une espèce dominante, au premier rang desquelles *Pseudomonas aeruginosa*.

Dans la mucoviscidose, les infections respiratoires sont polymicrobiennes (12,59). Il est supposé d'ailleurs que la composition des communautés bactériennes pourrait être un bien meilleur marqueur que la spirométrie pour la mesure de la progression de la maladie et du déclin de la fonction respiratoire (12).

Enfin, certains taxons semblent être associés à l'exacerbation. Une étude réalisée grâce à la PCR 16S met en évidence *Gemella* – un streptocoque (diplocoque anaérobie facultatif) de la flore orale – comme bactérie corrélée à l'exacerbation. Celle-ci voyait sa prévalence augmentée en termes d'abondance absolue, augmentée également en termes d'abondance relative, et il s'agissait aussi du meilleur facteur prédictif lorsque les données étaient utilisées dans une simulation avec un algorithme d'intelligence artificielle (de type *random forest*) (58). Cette étude est cependant potentiellement biaisée avec la possibilité d'une contamination par la flore orale, et par une relative sur-quantification d'un taxon dans un prélèvement dont la diversité microbiologique diminue. Une autre étude met en avant la présence de *Streptococcus* du groupe *milleri* (*Streptococcus anginosus*) à de fortes concentrations ( $> 10^7$  CFU) chez des adultes au décours d'un épisode d'exacerbation par rapport à des sujets contrôles hors exacerbation (60).

De manière assez inattendue, *Pseudomonas aeruginosa* était significativement moins abondant (en abondance relative et absolue) pendant l'exacerbation qu'à l'état basal dans l'étude de Carmody *et al.* (58) chez les patients pour qui *Pseudomonas* était l'espèce majoritaire. Inversement, dans le groupe des patients dont l'OTU dominant n'était pas

*Pseudomonas*, les abondances relative et absolue de l'OTU en question n'étaient pas significativement différentes entre l'exacerbation et l'état basal. Néanmoins, *Pseudomonas*, dans les communautés où il s'agit de l'OTU dominant, semble jouer un rôle important dans l'évolution de la structure de la communauté avec la survenue des exacerbations. Les patients colonisés par *Pseudomonas aeruginosa* semblent héberger des communautés microbiennes plus résilientes, plus stables, avec un retour à l'état basal après exacerbation proche de celui pré-exacerbation.

A la question peut-il exister différents types d'exacerbations, la réponse est probablement affirmative : certaines induisant des changements majeurs dans la structure des communautés bactériennes ou d'autres n'étant à l'origine que de variations mineures sans disruption de la structure du microbiote (58).

Il existe une hypothèse en vogue depuis le début des années 2010 concernant le rôle de « pathogènes clés » (pathogènes 'fondateurs' : *keystone pathogens*). Décrite dans Nature Reviews | Microbiology (61), l'hypothèse est la suivante : des pathogènes de faible abondance, mais au pouvoir pro-inflammatoire conséquent, seraient à l'origine d'un remodelage en profondeur des communautés microbiennes (62). La mise en évidence de *Gemella* dans l'étude de Carmody (58) pourrait tout à fait être envisagée au regard de cette hypothèse.

### 2.2.2. *Microbiote respiratoire viral*

Le microbiote viral, ou virome, respiratoire est moins connu. Il semble présenter une grande diversité à la fois interindividuelle et spatiale, en plus d'être caractérisé par la présence d'une double population aux rôles et évolutions différents : les virus eucaryotes d'une part et les

phages de l'autre. Un point important sur lequel un grand nombre d'études s'accordent est le rôle primordial de ce microbiote viral dans l'infection bactérienne.

Le nombre d'épisodes infectieux viraux respiratoires, asymptomatiques ou non est élevé : 12 par an chez les enfants et 6 par an chez les adultes selon Byington *et al.* (63). Le rôle de ces infections virales est de plus en plus pointé du doigt dans le déclenchement des exacerbations (64,65). On citera en particulier les rhinovirus qui semblent être étroitement liés à la survenue d'exacerbations (64), sans que les mécanismes à l'origine de cette association ne soient bien compris. Il semble que ces virus, et les infections virales respiratoires en général, jouent un rôle majeur dans la disruption initiale de la barrière muqueuse (1), qui favoriserait l'infection bactérienne. L'impact négatif des infections respiratoires virales dans la mucoviscidose a longtemps été suspecté (les premiers travaux datent de 1991) et confirmés depuis (26). Enfin, la fréquence plus élevée des détections de virus dans les prélèvements respiratoires lors des épisodes d'exacerbation par rapport à des prélèvements systématiques de suivi routinier (27) oriente également en la faveur d'un rôle des virus dans la survenue de ces exacerbations.

Les virus de la grippe (influenza A) et le VRS (virus respiratoire syncytial) interfèrent dans les voies de signalisation de l'interféron (IFN) et des Toll-Like Receptor (TLR), dégradant la réponse immunitaire innée et favorisant la colonisation bactérienne (9). Chez les patients colonisés de manière chronique par *Pseudomonas aeruginosa*, une infection virale respiratoire permet la libération de *Pseudomonas* contenu dans le biofilm et la multiplication de ce dernier ou de d'autres bactéries telles que *Haemophilus influenzae* (28).

Les rares virus eucaryotes mis en évidence sont surtout des *Anellovirus* de la famille de Anelloviridae (66–68), parmi lesquels trois espèces majeures : *Torque Teno Virus*, *Torque Teno*

*Mini Virus* et *Torque Teno Midi Virus* (69–72). Leur véritable rôle reste encore méconnu mais ces virus semblent considérés comme des commensaux non pathogènes (36). Les virus eucaryotes présentent une faible diversité à la naissance et leur nombre augmente progressivement avec l'exposition environnementale jusqu'à atteindre un pic à 6-18 mois (67,68), concomitant de la perte de l'immunité maternelle médiée par les immunoglobulines de type G (IgG) et la maturation du système immunitaire de l'enfant.

Dans une étude réalisée sur des aspirations nasales d'individus sains (73), les auteurs ont trouvé des séquences de *Roseolovirus*, *Herpesvirus*, *alpha-*, *beta-* et *gammapapillomavirus*, de *polyomavirus*, *mastadenovirus*, de *dependoparvovirus*, *alphatorquevirus*, d'*astrovirus* et des virus non classés du genre *Anellovirus*. De manière similaire, les voies aériennes respiratoires des patients sains semblaient fortement habitées par des populations d'*anellovirus* et dans une moindre mesure, d'*herpesvirus* et *papillomavirus* (74). Il semble que la diversité et la richesse virale suivent un gradient descendant avec la localisation dans les voies aériennes : plus l'on descend dans l'arbre trachéo-bronchique, et plus la diversité diminue (36). Il semble également que la diversité virale soit hautement dépendante de la région échantillonnée (75). Dans ces conditions, il est difficile d'obtenir une vue d'ensemble fiable du virome respiratoire (64). Enfin, Young *et al.* dans leur étude (74) publiée en 2015, suggèrent qu'un état de dysbiose puisse apparaître avec la forte divergence des communautés bactériennes, elles-mêmes déstabilisées par l'augmentation du titre des virus.

Concernant les virus eucaryotes, il semble exister un rôle protecteur du virome eucaryote face aux infections bactériennes. Une expérience réalisée chez la souris montre que l'infection du mammifère avec le CMV murin ou avec un *gammaherpesvirus 68* (virus modèle pour les infections à EBV) confère une protection contre *Listeria monocytogenes* et *Yersinia pestis* (76).

Un autre exemple qui soutient cette hypothèse est celui du ralentissement de la progression du VIH vers le stade SIDA chez les patients co-infectés avec le virus de l'hépatite G (GB virus type C) (77). En revanche, certains virus, en maintenant un état pro-inflammatoire chronique, seraient plutôt responsables d'une susceptibilité accrue aux infections par d'autres virus potentiellement plus virulents ou promoteurs de transformation oncotique (78).

Le virome oral apparaît comme hautement individu-dépendant (personnalisé) et stable dans le temps (79–82), bien que l'on retrouve des viromes oraux similaires entre des individus partageant le même régime alimentaire, le même microbiote bactérien oral ou appartenant au même foyer (83). Il existe une importante supériorité numérique des séquences génétiques issues de phages par rapport aux virus eucaryotes. La proportion de séquences sans similarité dans les bases de données publiques est très importante (36), une étude avance le chiffre de 36 à 88 % des séquences retrouvées sans correspondance dans les bases de données actuelles (75). Les phages de la cavité orale sont classifiés en trois familles : *Siphoviridae*, *Myoviridae* et *Podoviridae* (79–82), tous faisant partie de l'ordre des *Caudovirales*. Les phages retrouvés au sein de ce microbiote ont probablement plusieurs relations 'espèce-hôte' (une espèce pouvant interagir avec plusieurs hôtes et un hôte pouvant héberger plusieurs espèces), créant ainsi un réseau d'interactions plutôt qu'une simple relation prédateur-proie (84,85). Celle-ci est d'ailleurs peut-être à l'origine de la stabilité de la population phage. On observe ainsi une stabilité dans le temps avec 95 % et 80 % de conservation de la diversité des phages pendant 1 et 2,5 ans respectivement (86–88). Ces phages du microbiote oral ont été regroupés en trois catégories : un premier groupe très répandu, présent chez plus de la moitié des individus testés dans le monde (qui pourraient correspondre à des prophages intégrés), un deuxième set de 132 bactériophages communs retrouvés chez 20 à 50 % des patients, et enfin, un troisième groupe constitué de phages peu répandus ou uniques par individu (89). A ce jour, il

n'a pas été démontré que cette répartition soit également retrouvée au sein du virome respiratoire plus profond.

Certains parallèles peuvent être faits avec le virome digestif : on retrouve une relative faible abondance de virus eucaryotes chez les individus sains (86,87,90–93). En revanche, une très grande quantité de bactériophages semble coloniser de manière régulière ces niches. La richesse et la diversité des phages est extraordinaire à la naissance et décroît avec le temps, au fur et à mesure que la population bactérienne se stabilise (67,68,90,94). Les phages aident à la stabilisation du microbiote bactérien (95) ainsi qu'au rétablissement de celui-ci après un déséquilibre.

Nous l'avons vu, les phages semblent occuper une position particulière au sein du virome et du microbiote. Ils présentent par ailleurs un rôle prépondérant dans la constitution d'un réservoir de gènes dont des gènes de résistance aux antibiotiques (on parle aussi de résistome) (79,86,92,96–98). Cependant, la plupart des gènes de phages ne codent pas pour une protéine dont la fonction est connue ou homologue d'une protéine connue. Il est donc difficile d'extrapoler la fonction de la majeure partie des protéines codées par les gènes véhiculés par ces phages. Il est également très complexe d'interpréter l'évolution des populations phages dans la santé et la maladie, et dans la réponse au traitement (67,68,80,87,99,100).

Un autre rôle attribué aux phages du microbiote est celui de l'immunité non liée à l'hôte (*non-host innate immunity*). Il s'agit des fonctions de défense antimicrobiennes n'appartenant pas à l'hôte et reliées à une immunité innée (sans premier contact nécessaire pour l'apprentissage). C'est bien cette population phage, présente en grande quantité dans les muqueuses, qui remplit ces fonctions de « système immunitaire » (101). D'ailleurs, la

concentration élevée des phages résidants aux points d'entrée des bactéries dans le corps humain est un argument en faveur de cette hypothèse. Un des aspects de ce système est désigné par l'acronyme BAM : *Bacteriophage Adherence to Mucus*. Des phages à cycle lytique sont regroupés en grande quantité au sein du mucus et donnent ainsi à l'hôte une première barrière de protection contre les bactéries qui s'attachent aux parois cellulaires. Les mécanismes de cette concentration de phages dans le mucus ne sont pas entièrement compris : une part est expliquée par l'adhérence des phages à l'IGLP (Immunoglobulin Like Protein) mais une part reste sans explication (102). Cette interaction avec l'IGLP semble plutôt d'ailleurs promouvoir l'infectivité (potentiel infectieux) plutôt que la concentration géographique. Ainsi des phages à cycle lytique constitueraient notre première barrière de défense contre l'agresseur bactérien. Enfin, les phages semblent présenter de nombreuses fonctions en interaction avec le système immunitaire mais qui dépassent le cadre de cette étude (interaction avec le système anticorps, promotion de la phagocytose / picnocytose par les cellules dendritiques, inhibition de l'activation et de la prolifération des lymphocytes T, stimulation de la réponse immunitaire antivirale...). La population de phage semble ainsi servir d'amorce, de démarreur pour le système immunitaire, engageant ainsi nos défenses contre les bactéries pathogènes (88).

A l'opposé, les phages à cycle lysogénique (donc non lytiques) auraient la possibilité de modifier et contrôler le fonctionnement de ce microbiome en agissant comme réservoir de biodiversité et en favorisant le transfert horizontal de matériel génétique. Cette deuxième population, stable au cours du temps chez un individu donné, pourrait faire varier le *fitness* bactérien pour l'adapter aux conditions d'environnement (92). Leur situation d'entité non vivante et leur localisation tapie au sein du mucus pourraient faciliter leur résistance au cours du temps et face à des événements importants particulièrement délétères pour le reste du



microbiote. En agissant ainsi comme une sorte de tampon du microbiote et de réservoir génétique, la population de phages à cycle lysogénique permettrait la restauration des capacités métaboliques du microbiote après des épisodes majeurs de perte de diversité.

Les connaissances concernant le microbiote viral sont encore limitées mais progressent à grande vitesse. Surtout, c'est notre vision générale du microbiote viral qui évolue, en passant d'un modèle simpliste de parasitisme à une compréhension plus globale d'une interaction durable avec l'hôte avec des rôles généraux et spécifiques par site anatomique.

Au final, dans une étude proche de la nôtre, Willner et collègues ont réalisé l'évaluation du virome respiratoire chez des patients atteints de mucoviscidose avec une approche en NGS (103). Mais celle-ci date de 2009, et la technique utilisée pour enrichir la fraction virale consiste en une filtration à 0,45 µm et une purification au chlorure de césium (CsCl), dont les résultats ne sont pas satisfaisants. En plus de ne purifier et séquencer que les virus à ADN, le séquençage a été réalisé après amplification par déplacements multiples (MDA), une technique connue pour introduire d'importants biais d'amplification. Enfin, cette étude présente les résultats d'explorations menées sur seulement 5 patients avec 5 contrôles. La limite de sensibilité de ce mode opératoire est également pointé du doigt par la revue de Wylie en 2017 (64).

### *2.2.3. Microbiote respiratoire fongique*

Le microbiote respiratoire fongique est à ce jour encore très peu exploré. Une revue de la littérature réalisée en 2017 (104) cite l'existence de moins de dix études du mycome (ou mycobiome) réalisées en NGS. Il semble maintenant acquis que le poumon n'est plus à considérer comme un site stérile, même en dehors de toute pathologie. L'exploration de patients sains met en évidence la prédominance de taxons du sous-règne des Dikarya, des

phyla ascomycota et basidiomycota et de la famille des Davidiellaceae. Les genres les plus représentés semblent être les *Cladosporium*, *Eurotium*, *Penicillium* et *Aspergillus*. Les espèces majoritaires étant les *Candida*, les *Neosartotya*, *Malassezia*, *Hyphodontia*, *Kluyveromyces* et *Pneumocystis*. La diversité des espèces fongiques au sein des prélèvements respiratoires semble être supérieure à la diversité bactérienne observée dans ces mêmes prélèvements (105). La communauté fongique semble par ailleurs plus stable que la communauté bactérienne, en particulier face à l'utilisation d'antimicrobiens (106).

Certaines données expérimentales semblent par ailleurs contradictoires : s'il est acquis que la présence d'espèces fongiques au sein des prélèvements respiratoires est systématique, et que la diversité retrouvée est hypervariable d'un individu à l'autre (104), la variation du microbiote respiratoire fongique au décours de la pathologie est encore sujet à caution. Dans cette même étude, les auteurs concluent à une diminution de la diversité fongique avec la survenue de pathologies respiratoires ou d'exacerbation. En revanche, dans une étude sur 89 échantillons réalisée en 2015, Kim et collègues ne retrouvent pas de différence au niveau fongique lors d'exacerbations de patients atteints de mucoviscidose (107).

Bien qu'*Aspergillus spp* ait été retrouvé abondamment dans les prélèvements respiratoires des patients atteints de mucoviscidose, son rôle exact dans l'infection et la pathologie respiratoire reste encore largement incompris. Les champignons et moisissures sont fréquemment retrouvés chez les patients atteints de mucoviscidose. Les facteurs de risque associés à leur présence sont l'âge du patient, l'exposition prolongée aux antibiotiques et le traitement par corticoïdes (108). Les espèces de filamenteux les plus fréquemment mises en évidence sont *Aspergillus fumigatus* et *Scedosporium apiospermum*. *Aspergillus fumigatus*

semble être présent dans 6 à 60 % des échantillons testés au cours de différentes études. En ce qui concerne les levures, c'est *Candida albicans* qui est le plus représenté.

Si *Aspergillus* semble être le principal champignon impliqué dans les infections du tractus respiratoire, la présence de *Pneumocystis jirovecii* est décrite dans les voies respiratoires de 1,3 % à 21,6 % des patients atteints de mucoviscidose (109). *Pneumocystis* est un champignon opportuniste retrouvé initialement chez l'immunodéprimé et particulièrement les patients infectés par le VIH et au stade SIDA. Il pourrait jouer un rôle dans l'évolution de la mucoviscidose et dans le déclin de la fonction respiratoire, mais à ce jour, aucune étude avec un niveau de preuve suffisant ne vient corroborer cette hypothèse. Green et collègues, dans une étude publiée en 2016 (110), rapportent néanmoins la présence accrue de *Pneumocystis jirovecii* dans les prélèvements respiratoires des patients atteints de mucoviscidose au tout début des symptômes liés à la survenue d'une exacerbation. Enfin, il n'existe à ce jour aucune donnée disponible concernant l'interaction possible de *Pneumocystis* avec les bactéries, qu'elles soient pathogènes ou commensales du tractus respiratoire.

### 3. Exploration du microbiote respiratoire

L'exploration du microbiote respiratoire tel que nous l'avons décrit jusqu'à présent a été réalisée de manière séquentielle, découpée selon les trois classes d'agents infectieux : la bactériologie, la virologie et la parasitologie-mycologie. Aujourd'hui, et au vu des interactions complexes qui font intervenir des espèces relevant de ces différentes disciplines, nous souhaitons essayer de développer une approche plus globale, en s'affranchissant des frontières imposées par ce découpage. En effet, la littérature scientifique fait état de résultats très variables voir même discordants entre les études. En cause, l'hétérogénéité des types de prélèvements réalisés, des âges des patients (des cohortes), la saisonnalité, mais en premier

lieu des techniques de détection utilisées et la sensibilité de celles-ci (9). Aussi, nous détaillerons dans ce chapitre les techniques utilisées jusqu'à présent en routine hospitalière et les innovations apportées récemment pour l'exploration du microbiote respiratoire en recherche fondamentale ou clinique.

### 3.1. Origine du prélèvement

Il existe des recommandations actuelles de prélèvement par expectoration pour l'exploration bactérienne des voies aériennes, mais celles-ci sont inexistantes pour l'exploration virale (9). Pour échantillonner les voies aériennes supérieures, les prélèvements recommandés sont : l'écouvillonnage nasal antérieur, moyen ou postérieur, l'écouvillonnage de gorge ou oropharyngé, et la réalisation d'aspirations nasales. Pour l'exploration de la partie inférieure des voies aériennes sont recommandés les prélèvements suivants : l'expectoration, le lavage broncho-alvéolaire ou la broncho-aspiration, le prélèvement distal protégé ou encore la biopsie pulmonaire.

Quel que soit le type de prélèvement, il existe un risque de contamination : le passage par la cavité orale est une source de contamination très importante. Il a été démontré que les communautés bactériennes sont très différentes dans la cavité orale et dans les expectorations (111), de même que les communautés de virus qui sont très dépendantes du site échantillonné (36). Ce risque, commun à toutes les techniques de prélèvement est difficile à supprimer complètement. La réalisation de prélèvements distaux protégés par exemple ou de lavage broncho-alvéolaire permet de maîtriser ce risque de façon satisfaisante pour les techniques actuelles mais reste une procédure très invasive pour le patient.

L'origine du prélèvement pose également la problématique du *sampling* spatial. C'est-à-dire quel est le site réellement échantillonné et qu'est-ce que cela implique en termes de

communautés microbiennes adjacentes. Échantillonner les voies respiratoires basses, les voies respiratoires hautes, ou des niches anatomiques ou fonctionnelles particulières produit des résultats très différents et il est important de définir à l'avance quel type de prélèvement est le plus adapté. A ce jour, les données manquent concernant le type de prélèvement et le site anatomique effectivement échantillonné : aucun des types de prélèvement disponibles n'est relié à un site anatomique donné avec précision.

### 3.2. Techniques actuelles de détection

Ces techniques sont utilisées en pratique courante dans les centres hospitaliers prenant en charge des patients atteints de mucoviscidose. L'ensemble de ces techniques, utilisées conjointement, permet d'affirmer la présence ou l'absence d'un nombre limité micro-organismes pathogènes majeurs, mais ne présentent pas la finesse suffisante pour rendre compte de l'évolution du microbiote dans toute sa diversité.

#### 3.2.1. Culture(s)

La culture est la méthode historique d'isolement des bactéries, virus et champignons. Simple à mettre en place, elle permet de faire croître, dans un milieu de culture maîtrisé, les micro-organismes afin de faciliter leur identification et leur quantification relative. Il s'agit cependant d'une technique très subjective (observateur dépendant), peu sensible, extrêmement chronophage et non-exhaustive (45). L'existence de bactéries non cultivables met en défaut cette approche. Il s'agit notamment de l'ensemble des bactéries intracellulaires strictes. D'autres bactéries à croissance lente et fastidieuse posent également problème, ainsi que les bactéries fragiles qui pourraient être détruites ou inactivées par des conditions de stockage, de transport et de culture non adaptées (passage à -80°C, ...). Il existe également un problème pratique quant à l'isolement et à l'identification d'un nombre élevé d'espèces différentes à

partir d'un prélèvement unique (110). Malgré tout, la culture reste le pilier central du diagnostic en bactériologie et en mycologie, malgré un délai de rendu de résultat de l'ordre de 24 à 48 heures.

Concernant les virus, la culture est la seule technique qui permet de confirmer le caractère cytopathogène et répliatif d'un virus. Néanmoins, elle nécessite l'entretien constant de lignées cellulaires et d'anticorps spécifiques de chaque type de virus, ce qui représente un coût et une organisation importante (104). Elle a, dans cette discipline, été très largement remplacée par le diagnostic moléculaire.

De plus, un certain nombre de virus et quelques bactéries, sont, lorsqu'ils sont cultivés, considérés comme appartenant à une catégorie de pathogènes de classe 3 nécessitant la mise en œuvre de mesure de sécurité lourdes (travail en Laboratoire National de Sécurité Biologique de Niveau 3, ...)., Pour toutes ces raisons, la mise en culture des micro-organismes pour l'exploration de la flore respiratoire est peu adaptée à une approche globale visant à décrire l'ensemble des micro-organismes présents au sein des voies respiratoires et leur évolution.

### *3.2.2. Spectrométrie de masse*

La spectrométrie de masse a été introduite dans les laboratoires hospitaliers à partir des années 2010. Très largement utilisée en bactériologie où elle a constitué une véritable révolution, et dans une moindre mesure en mycologie, elle a largement supplanté les traditionnelles galeries biochimiques pour l'identification des bactéries et levures pathogènes post-culture. La disponibilité des appareils commerciaux de type MALDI-TOF (Matrix Assisted Laser Desorption and Ionisation – Time Of Flight) associée à des bibliothèques de spectres de plus en plus fournies et robustes ont permis de réduire le temps d'identification d'une bactérie

de 24h minimum à quelques minutes. La plupart du temps cette identification est réalisée après une étape de culture mais des développements actuellement en cours pourraient l'étendre à l'étude directe des prélèvements. La spectrométrie de masse n'occupe qu'une place marginale dans la recherche de virus en pratique clinique. Cette approche manquant de sensibilité pour la détection directe de virus, elle est utilisée de façon anecdotique en post-PCR.

### 3.2.3. PCR spécifiques ciblées

La description du principe de la *polymerase chain reaction (PCR)* a été publiée pour la première fois en 1985 (112). Avec la découverte de la polymérase thermostable de *Thermophilus aquaticus* en 1988 (113), cet outil a très rapidement été adopté pour le diagnostic médical. Grâce à l'utilisation d'amorces spécifiques d'une zone précise du génome à amplifier, elle peut être spécifique d'un pathogène ou d'une famille de pathogène. Son utilisation en microbiologie a été introduite dès les années 1990, en particulier pour la détection des mycobactéries à croissance lente. Son principal atout réside dans la réduction du temps nécessaire au diagnostic, et bien que dans sa forme initiale la technique de PCR s'étalait sur 24h, elle a largement été adoptée dans les laboratoires hospitaliers, tout particulièrement en virologie où elle rivalise avec les performances de la culture. De plus, l'arrivée de la PCR temps réel a encore amélioré la sensibilité et la fiabilité de cette technique. Néanmoins, les infections respiratoires pouvant être dues à de très nombreux micro-organismes parfois associés, un diagnostic des principaux agents pathogènes nécessiterait une démultiplication de techniques de PCR. De fait, la PCR simple est progressivement abandonnée au profit de techniques multiplexées, c'est-à-dire permettant l'amplification et la détection du génome de plusieurs pathogènes simultanément.

#### 3.2.4. PCR multiplexes syndromiques

La PCR multiplexe apparaît dès les années 2005 – 2008, même si le principe et la description théorique datent de quelques années auparavant. Cette technique devient rapidement le ‘gold standard’ en virologie grâce à des délais de rendu courts et bonne sensibilité analytique (27). Cependant, elle n’est pas forcément utilisable dans toutes les situations. D’abord, s’il existe chez le patient un enchaînement rapproché des épisodes infectieux, alors il n’est pas possible de faire la distinction entre le pathogène en cause et les pathogènes de l’épisode précédent ou suivant (114). La persistance de quelques copies de génome viral, bactérien ou parasitaire sera inévitablement amplifiée et détectée par ces techniques de biologie moléculaire ultra-sensibles, sans qu’un lien de pathogénicité ne puisse être formellement établi. Cette problématique rejoint celles posées par la PCR en général, avec la détection d’ADN ou ARN de pathogènes éliminés ou non-viables. Mais ceci pourrait être résolu avec la mise en place de PCR multiplexes quantitatives ou semi-quantitatives. Enfin, les techniques de PCR, ‘simplexes’ ou multiplexes, sont basées sur la spécificité d’amorces nucléotidiques préparées sur mesure pour la détection du pathogène souhaité. Par leur conception, elles ne permettent pas de détecter un micro-organisme non recherché initialement. Cela pose d’importants problèmes pour les pathogènes hautement variables (la présence de mutation peut empêcher la fixation des amorces et donc l’amplification), ainsi que pour les pathogènes émergents dont la séquence génétique n’est pas connue *a priori*. Cela est particulièrement important en virologie, où la diversité et la variabilité des génomes viraux est régulièrement à l’origine de résultats faussement négatifs, impactant ainsi de façon majeure la prise en charge des patients concernés.



L'application des PCR multiplexes au diagnostic en routine clinique fait actuellement face à l'arrivée du séquençage nouvelle génération (ou NGS), chacune des techniques ayant son lot d'avantages et inconvénients. L'application du NGS à la recherche d'acides nucléiques en infectiologie permet d'apporter une solution à la question de la variabilité des génomes, à la découverte de nouvelles espèces et au diagnostic de pathogènes non initialement recherchés. A ce jour, le NGS reste une technique très coûteuse qui ne permet pas son utilisation comme technique de première intention. Par ailleurs, par leur exhaustivité, les approches basées sur le NGS sont sensibles à l'introduction de biais et aux multiples sources de contamination. L'omniprésence d'acides nucléiques dans l'environnement et dans les réactifs utilisés rend cette approche délicate à mettre en place. Enfin, la surreprésentation de l'ADN génomique humain dans les échantillons cliniques est un *challenge* pour ces méthodes basées sur le NGS pour lesquelles cet ADN humain est 'contaminant'.

### 3.2.5. *Identification par amplification et séquençage d'un gène générique (16S, 28S ou ITS)*

La PCR 16S est une application de la PCR et du séquençage à la détection des bactéries. L'utilisation d'amorces complémentaires de zones conservées de l'ARN ribosomique 16S bactérien permet de détecter, sans connaissance *a priori* de l'agent pathogène, la bactérie majoritairement présente dans un prélèvement. Cette méthode est très utilisée en routine clinique couplée avec le séquençage Sanger. Le séquençage dit 'Sanger', du nom de son inventeur, a été mise au point en 1972. Elle permet d'obtenir la séquence en nucléotides d'un brin d'ADN par synthèse d'un brin complémentaire en utilisant des di-désoxynucléotides terminateurs de chaîne marqués à l'aide d'un rapporteur fluorescent. Cette méthode est à la base de l'ensemble des découvertes qu'a permis la génétique au cours de 50 dernières années et continue d'être utilisée aujourd'hui en routine clinique. En revanche, elle ne permet de

mettre en évidence que le pathogène majoritaire et non l'ensemble de la diversité d'une flore polymicrobienne (approche mono-microbienne).

L'approche par PCR 16S a cependant permis les premières études dites 'de microbiote' (bien que cette appellation soit un abus de langage, seules les bactéries étant recherchées, les virus et champignons ne possédant pas de 16S). Une telle étude nécessite le couplage de la PCR 16S avec une technique de détection plus sensible que le séquençage Sanger, par exemple grâce à l'utilisation de *micro*-array. La sensibilité d'une telle détection est importante : une étude a par exemple été réalisée par PCR 16S avec une étape de détection sur une puce PhyloChip. Huit mille cinq cent espèces bactériennes étaient représentées sur ce *micro*-array et la détection des taxa était possible à partir d'un seuil de 0,01 % de représentation d'une population donnée (43). Cette méthode de PCR 16S est en revanche extrêmement sensible aux contaminations, en particulier à la présence d'acides nucléiques de *E. coli* ou *T. aquaticus* dans les réactifs utilisés. Cette source de contamination est néanmoins connue depuis de nombreuses années (115). Enfin, cette technique ne permet pas de rendre compte de la variation du nombre de copies de ce gène. Les bactéries présentant physiologiquement des copies de leur matériel génétique, la quantification de l'ARN 16S ne renseigne pas sur la quantité de bactéries présentes dans l'échantillon considéré et ne peut donc pas être utilisée comme méthode quantitative.

Il existe un équivalent du gène 16S pour le domaine de la mycologie : il s'agit du gène 28S et de la région ITS (Internal Transcribed Spacer). L'ITS est tout particulièrement utilisé et constitue d'ailleurs la référence pour le typage des espèces en mycologie (116). En revanche, il n'existe pas de gène 'générique' dans le domaine de la virologie. Les séquences génétiques étant très éloignées et diversifiées. Par ailleurs, la présence de micro-organismes (tout

particulièrement les virus) dont les génomes sont hypervariables dans le temps et ne présentant aucun gène constant partagé ne permet pas l'utilisation de PCR, simples ou multiplexes, chacune n'étant prévue que pour mettre en évidence une unique cible invariable.

### 3.2.6. *Antibiogrammes, anti-fongigrammes et anti-virogrammes*

L'étude phénotypique de la résistance (ou plutôt de la sensibilité) des micro-organismes aux molécules utilisées en thérapeutique pour inhiber leur croissance et leur développement est réalisée en pratique clinique pour les bactéries sous la forme d'antibiogrammes et pour les champignons filamenteux sous la forme de d'antifongigrammes. Il s'agit de la seule technique disponible pour évaluer les sensibilités phénotypiques avec précision. Cette technique présente en effet une meilleure corrélation avec l'impact réel *in-vivo* de ces antimicrobiens par rapport à des études génotypiques (génétiques) telles que la mise en évidence de gènes de résistance (117). Cependant, la présence de mucus épais et de biofilm sécrété par certaines bactéries rend cette recherche de résistance plus délicate. Une bactérie sensible *in-vitro* peut ainsi être totalement résistante pour la même molécule *in-vivo* à cause d'une faible diffusion des antibiotiques à travers ce biofilm. La publication de *guidelines* concernant la détermination de la sensibilité aux antimicrobiens en présence de biofilm (118) et leur application a montré une réduction de la charge bactérienne après traitement, sans que cela n'apporte de bénéfice pour le patient (119).

Concernant les virus, et à l'exception du VIH, pour lequel la sensibilité aux anti-rétroviraux est régulièrement testée, la susceptibilité des autres virus et particulièrement celles des virus respiratoires aux antiviraux est rarement testée en pratique courante. Des tests génétiques et phénotypiques de résistance du virus de la grippe sont disponibles et réalisés dans certains laboratoires (120). Néanmoins, leur intérêt dans la prise en charge des patients est limité. La

détermination de la sensibilité aux antifongiques est réalisée en pratique clinique par la détermination de CMI en milieu liquide ou sur milieu gélosés (détermination phénotypique). Il n'existe pas de consensus quant à l'indication de réalisation d'un antifongigramme. Cependant, on considère que cette exploration est indispensable dans les cas suivants : isolement d'une souche issue d'un site normalement stérile, infection invasive, infection par une espèce rare ou devant la suspicion d'échec thérapeutique.

### *3.2.7. Méthode(s) de référence*

La technique de référence, c'est-à-dire utilisée actuellement en routine pour le traitement des échantillons respiratoires issus de patients atteints de mucoviscidose, n'est pas unique mais regroupe l'ensemble des techniques utilisées en bactériologie, virologie et parasitologie-mycologie. Il s'agit de l'ensemble des techniques appliquées aux échantillons en routine, c'est-à-dire la culture bactérienne, la culture parasitaire, les PCR virales et les PCR multiplexes. La combinaison de ces méthodes permet une détection, l'identification et la caractérisation de nombreux pathogènes. Néanmoins, il existe un nombre important de limitations techniques et biologiques associées à ces méthodes, qui nous poussent à développer une nouvelle approche.

### **3.3. Nouvelles approches**

La présence de ces nombreux biais et des limites importantes des techniques utilisées actuellement montre la nécessité de nouvelles approches, novatrices au plan technique et surtout plus globales dans leurs spectres de détection, permettant de décrire un microbiote donné en termes bactérien, virologique, parasitaire et fongique. Il est important de noter ce besoin de s'affranchir des frontières artificielles des disciplines historiques de la bactériologie,

virologie et parasito-mycologie, l'ensemble de ces micro-organismes vivant au sein d'un même écosystème et entretenant des relations complexes entre eux et avec l'hôte.

### 3.3.1. NGS pour Next Generation Sequencing

Ces nouvelles approches sont basées sur le séquençage de nouvelle génération (ou NGS). Encore appelée HTS pour *High Throughput Sequencing*, cette technique permet d'obtenir un nombre très important de séquences génétiques avec une qualité élevée des données de séquençage, et ce, beaucoup plus rapidement qu'avec l'ancienne technique de Sanger. Il existe différentes modalités techniques avec différents atouts et inconvénients (longueur de *reads*, profondeur de lecture, ...) dont le choix est à moduler en fonction de la question biologique à laquelle on souhaite répondre.

L'utilisation de cet outil suppose la compréhension de quelques paramètres-clés qui découlent du fonctionnement même de cette technologie. Il s'agit d'une technique de séquençage dite 'par synthèse' (séquençage par ajout d'un terminateur réversible), où en réalité, le signal physique de la mesure correspond à l'ajout d'une base complémentaire de la séquence initiale. Bien sûr, l'informatique qui pilote cet outil prend en compte cet aspect et l'opération s'avère transparente pour l'utilisateur qui récupère la séquence du brin d'ADN introduit et non la séquence complémentaire. La transformation du signal physique reçu par le capteur en une séquence de bases d'acide désoxyribonucléique est une opération qui porte le nom de *base-calling*. Cette opération est normalement effectuée sous la responsabilité du fabricant du séquenceur (Illumina, ...) par des algorithmes propriétaires. Malheureusement, le faible signal et le faible ratio signal / bruit de ces techniques entraîne parfois une 'erreur de lecture' qui se traduit par l'introduction d'une base erronée dans la séquence lue. La présence de ces erreurs à des taux relativement important (au regard du nombre de séquences traitées)

nécessite de séquencer à plusieurs reprises le même brin d'ADN afin de s'en affranchir. Le nombre de lectures d'un même brin est appelé profondeur de lecture et est donc inversement corrélé avec le taux d'erreur finalement obtenu en sortie de séquenceur.

Nous l'avons vu, le NGS peut être couplé à la PCR 16S mais il souffre alors d'une partie des limitations de cette dernière technique (spectre de détection limité aux bactéries, variabilité de l'ARN 16S, problèmes de contamination, ...). De la même manière, les NGS peut être couplé aux PCR 28S ou ciblant l'ITS. Ces techniques permettent d'élargir la détection des pathogènes au règne fongique, mais là encore, des limitations subsistent. La détection et l'identification des virus restent hors de portée en l'absence de gène conservé au sein du règne viral, et l'information obtenue par ces trois gènes reste parcellaire, conduisant à une interprétation largement incomplète du microbiote analysé. A ce jour néanmoins, les approches basées sur la combinaison de la PCR 16S/ITS avec le séquençage nouvelle génération (NGS) sont à la base de la majorité des études de microbiote actuellement publiées.

### *3.3.1. Métagénomique*

Il s'agit de l'exploration des génomes au-delà du seul génome de l'hôte : c'est l'ensemble des ADN et ARN satellites d'un organisme hôte donné. La métagénomique regroupe donc les séquences ADN et ARN des bactéries, virus, phages, champignons, mais également les séquences non humaines intégrées au génome humain ou au sein de cellules humaines (dans le cas de l'homme). Au-delà du champ du vivant que cela vise à découvrir, la métagénomique désigne également par extension la branche de la science qui en réalise l'exploration : cela comprend donc également l'ensemble des outils et techniques qui permettent le traitement des séquences génétiques obtenues, l'assignement taxonomique et l'annotation des génomes. Une autre des branches émergentes de cette discipline est la métagénomique

fonctionnelle, dans laquelle sont croisées des données issues de génétique et des données fonctionnelles concernant le métabolisme.

Le séquençage du gène 16S bactérien est parfois considéré (à tort) comme de la métagénomique lorsque celui-ci est réalisé en NGS (et non en Sanger). Bien que cela ne soit pas techniquement le cas, les études réalisées avec une telle approche sont assimilées à des études de métagénomique en raison de l'important volume de données qu'elles génèrent. Mais elles présentent deux principales limitations : celle de n'explorer que le règne bactérien, et celle de n'explorer qu'une infime fraction des génomes bactériens en présence. Il serait plus rigoureux de parler de métagénétique dans ce cas précis.

La mise en place d'une telle technique d'exploration du microbiote viral correspond à la résolution d'un certain nombre d'obstacles techniques. Quelques *challenges* importants de cette approche ont été soulevés dans la revue dédiée au sujet (121) et publiée dans Microbiome en 2016 :

- Le manque de marqueurs viraux (contrairement au règne bactérien ou fongique par ex) rend l'exploration du virome particulièrement difficile.
- Cette absence de marqueur est en partie en lien avec l'extraordinaire diversité des séquences génétiques virales
- Les virus représentent une très faible partie de la biomasse (petits génomes) par rapport aux bactéries et aux organismes eucaryotes.
- La difficulté à distinguer dans certains cas les séquences virales du génome de l'hôte (rétro-virus intégrés ? ...) des réelles séquences génétiques de virus bien présents dans l'échantillon exploré.

- L'absence d'outils bio-informatiques adaptés à ce jour (bien que ce domaine soit en pleine expansion)
- L'absence de bases de données fiables (*curated*) pour le règne viral ainsi que pour les champignons et levures.

D'autres points d'achoppement sont également pointés du doigt dans une autre étude (73), tel que le très faible nombre de séquences virales en métagénomique par séquençage de type *shotgun* (MSS) sans capture : de l'ordre de 10 séquences virales pour 25 millions de séquences (humaines). Ceci est particulièrement problématique pour des études sur sujet sain, où les virémies attendues peuvent être très faibles (122).

Un autre risque important à prendre en compte est le risque de contamination des échantillons par des acides nucléiques contenus dans les réactifs. Ce risque est déjà décrit dans la littérature à plusieurs reprises (123), et nécessite, pour s'en prémunir de traiter tous les échantillons de la même manière. Pour cette raison, il semble raisonnable de réaliser un seul *batch* unique avec toujours les mêmes lots réactifs ainsi que d'y adjoindre systématiquement des contrôles négatifs pouvant servir de témoins.

Enfin, la problématique de la surreprésentation d'un taxon (bactérien) est une donnée essentielle à appréhender. La présence d'importantes concentrations bactériennes au sein du biofilm et du mucus respiratoire des patients atteints de mucoviscidose risque de saturer l'ensemble des techniques analytiques mises en œuvre dans une telle approche. Une sélection ou un enrichissement sélectif des différentes séquences d'acides nucléiques des micro-organismes recherchés montreront tout leur intérêt.

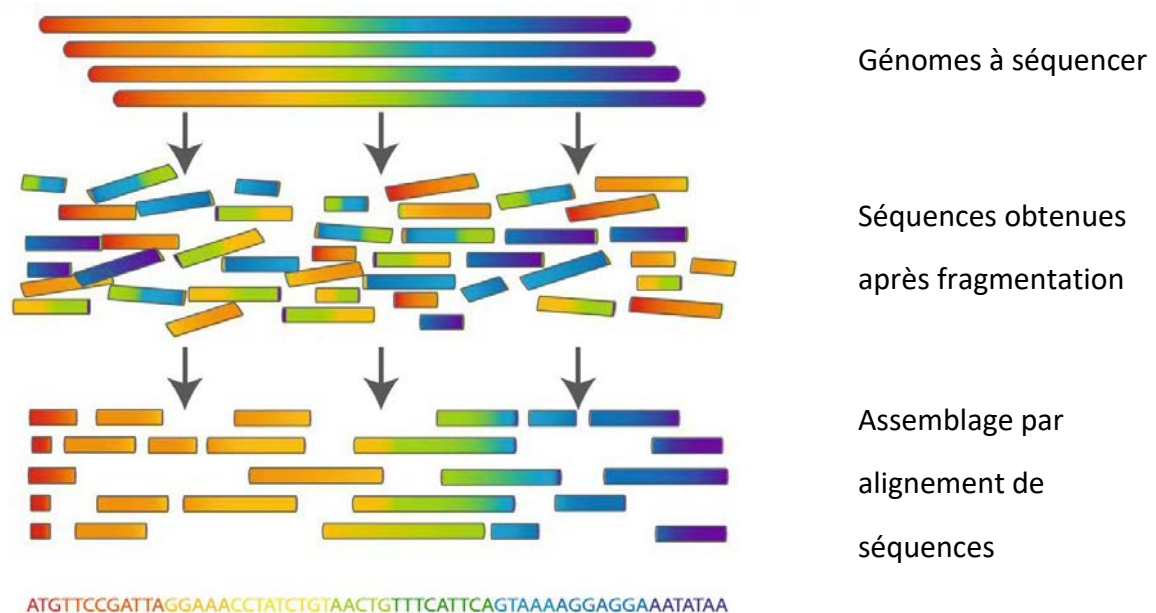
Il existe également une autre limitation à l'utilisation d'une approche basée sur la métagénomique. Cette technique, par les nombreuses étapes pré-analytique qu'elle



nécessite, est assortie d'une plus faible sensibilité qu'une approche moléculaire classique (par PCR spécifique). Cette relative faible sensibilité de la technique n'est cependant pas problématique puisqu'il semble que cette dernière soit équivalente à celle de la culture. Néanmoins, elle pourrait probablement être à l'origine d'une sous-estimation globale des pathogènes au sein du prélèvement muqueux (9).

### 3.3.2. *Shotgun sequencing*

On appelle *shotgun sequencing* une méthode de séquençage des longs brins d'ADN suite à leur fragmentation. La processivité de l'ADN polymérase utilisée dans la réaction de séquençage est telle que la synthèse de longs fragments d'ADN *in vitro* est délicate voire impossible.



**Figure 7 :** Principe du *shotgun sequencing*. Adapté de Commins et al. (124)

L'autre méthode permettant cela est le *primer walking* (ou *chromosome walking*) dans lequel les brins sont séquencés dans l'ordre. Dans la technique du *shotgun sequencing*, il n'a pas d'ordre dans le travail, le séquençage est réalisé sur des fragments courts obtenus

aléatoirement. Cette technique est plus rapide que celle du *primer walking*, mais également plus complexe à mettre en place, en particulier en ce qui concerne le traitement des données obtenues. Appliquée initialement au séquençage Sanger, cette approche a pris un nouveau tournant avec l'avènement du NGS, bien qu'elle soit computationnellement très intensive. La reconstruction des *contigs* (séquences génomiques ordonnées issues de fragments chevauchants) puis l'assemblage de ces derniers se fait par comparaison (alignement) de séquences multiples (**Figure 7**).

### 3.3.3. *Enrichissement*

Dans un échantillon humain clinique, la provenance des acides nucléiques est majoritairement d'origine humaine. La répartition attendue dans ce type de prélèvement est d'environ 99 % humain contre moins de 1 % des séquences pour les micro-organismes. La 'contamination' des échantillons par de l'ADN humain est écrasante et il n'est pas envisageable de travailler sur ces échantillons sans purification préalable. En l'absence de cette étape de purification, deux phénomènes sont observés : d'abord la diminution de la sensibilité de la technique, les séquences des micro-organismes étant noyées dans l'ensemble des séquences humaines. Enfin, l'autre conséquence est l'augmentation du coût économique de la technique. Le nombre de brins d'ADN à séquencer pour obtenir une sensibilité satisfaisante est démultiplié et conduit à un surcoût important. Il est donc très intéressant de dépléter l'échantillon en ADN humain, ou encore d'enrichir relativement la fraction des acides nucléiques correspondant à des micro-organismes d'intérêt.

Il existe plusieurs techniques pour obtenir une telle purification. L'enrichissement par capture est l'une d'entre elle. Il s'agit d'un procédé mis en place par Wylie *et al.* (125) et Briese *et al.* (126) avec des augmentations de 10 à 10.000 fois en termes de nombre de *reads* viraux par

rapport à une analyse en NGS sans capture. Des approches ont été proposées pour enrichir une fraction en virus uniquement : centrifugation sur gradient de sucrose, filtration (127) ou ultrafiltration, utilisation de DNase avant la lyse, ou autres méthodes physiques (128). La nécessité de prendre en compte la présence de bactéries et champignons rend complexe la tâche de trouver un protocole commun. Enfin, une approche par sélection négative est possible (rétention des ADN humains) mais les résultats obtenus par d'autres équipes sont peu concluants.

## 4. Métagénomique infectieuse semi-ciblée au CHU Grenoble – Alpes

Notre souhait était de développer une technique de métagénomique infectieuse au laboratoire pour permettre l'analyse de prélèvements divers, tant supposés initialement stériles (prélèvements sanguins dans le sepsis), que des prélèvements non-stériles d'emblée colonisés (respiratoires, digestifs, ...) à partir desquels il sera possible de caractériser le microbiote. Il existe un nombre important d'incertitudes techniques pour la mise en place d'une telle approche novatrice en infectiologie. C'est ce travail que nous avons initié et dont nous détaillerons les points importants dans ce manuscrit.

### 4.1. Choix technologiques

Un certain nombre de choix techniques ont été réalisés pour la mise au point de cette méthode d'analyse dite de métagénomique infectieuse. Nous en détaillerons ici les principaux arguments.

#### 4.1.1. Enrichissement

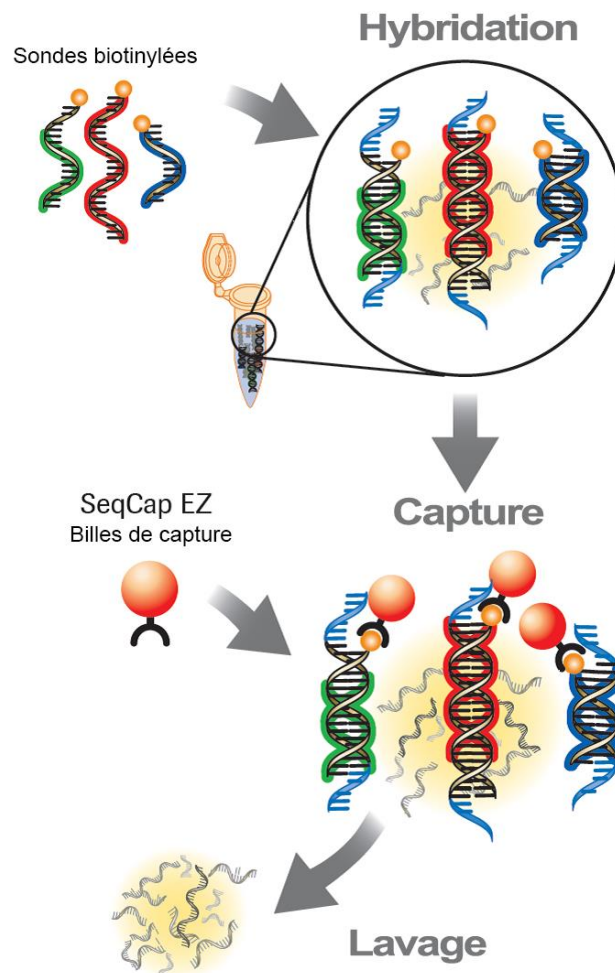
En l'absence de pré-traitement des échantillons, une approche par séquençage des micro-organismes reste de faible sensibilité à cause de la supériorité numérique des acides

nucléiques d'origine humaine dans ce type de prélèvements. Nous avons fait le choix pour notre travail d'un enrichissement (sélection positive) sur puce de capture dans la lignée des travaux de Lipkin (126) et Storch (125). Cette technique utilise la capture de séquences ciblées (*targeted sequence capture*). Elle a été utilisée extensivement pour l'exome humain (129) dès les années 1990. Elle est désormais appliquée à l'étude des virus depuis les années 2010 (130–132). En pratique, il s'agit d'une hybridation des séquences d'intérêt sur des sondes biotinylées en milieu liquide et de leur capture par liaison à la streptavidine. Le processus faisant appel au design de ladite puce de capture par l'industriel NimbleGen ainsi que pour la production (synthèse) des sondes biotinylées.

Cette technique de capture (hybridation) sur sondes en milieu liquide, dont le principe est résumé à la **Figure 8**, a montré d'excellentes performances. Les résultats obtenus par d'autres équipes sont encourageants : des augmentations de la sensibilité pour des acides nucléiques viraux d'un facteur 100 à 1000 ont été observés (126), en association avec une réduction du bruit de fond (*background*) de 33 à 66 %. Dans l'étude coordonnée par Lipkin, cette approche permettait également l'augmentation de la couverture des génomes des virus séquencés. A noter toutefois que ces performances dépendent tout particulièrement de la qualité du *design* des sondes utilisées. Aussi, nous avons appuyé notre réflexion concernant le design de cette puce de capture sur les résultats obtenus par l'équipe de Lipkin à l'aide de leur puce VirCapSeq-VERT (126), spécifiquement élaborée pour capturer les séquences de virus connus ayant pour hôte un représentant du sous-embranchement des vertébrés.

L'objectif dans notre approche était de détecter tous les virus connus pour infecter a priori des cellules humaines, ou capables d'un pouvoir pathogène chez l'Homme, et, détecter des

virus non connus, dont la séquence génique diverge fortement de celles disponibles dans les bases de données.



**Figure 8 :** Capture des séquences nucléotidiques par des sondes biotinylées en milieu liquide. Adapté de Kit SeqCap EZ

Ainsi, la difficulté du design de la puce résidait dans la taille des bases de données de génomes cibles. En incluant les bactéries et les champignons, le nombre et la diversité des séquences génétiques que l'on cherche à capturer est colossal, bien supérieur aux limites techniques fixées par le fabricant de la puce. Nous avons donc choisi de réduire ces données en distinguant une approche par gènes spécifiques en ce qui concerne le règne bactérien et

parasitaire et fongique. Nous avons conservé une approche par génome complet pour ce qui relève de la virologie. Enfin, nous avons procédé à une étape de *clustering* afin de répondre aux spécifications du fournisseur.

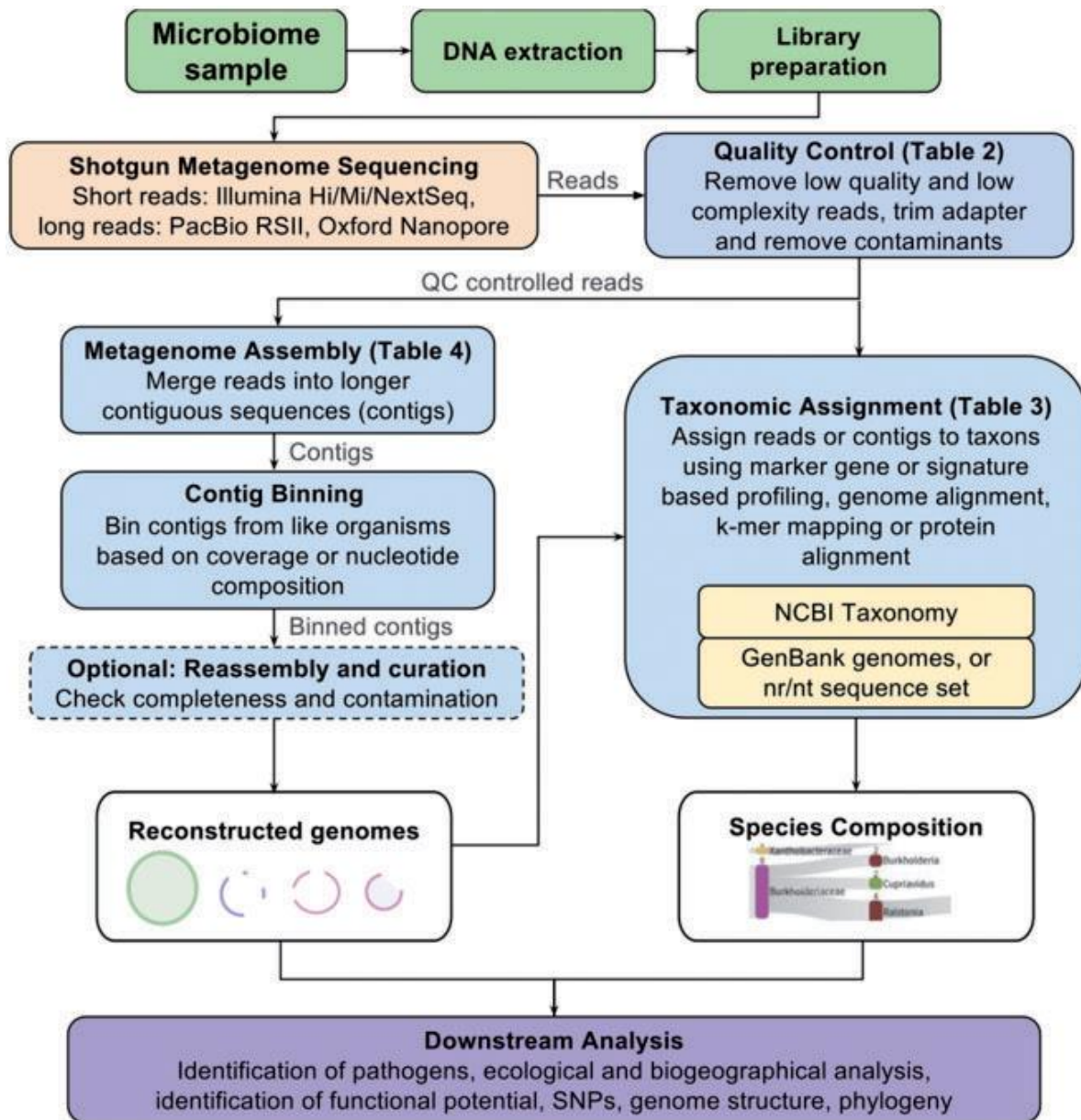
#### 4.1.2. Séquençage

Parmi les différentes technologies de séquençage disponible, nous avons fait le choix du NextSeq 500 d'Illumina car la technique de séquençage utilisée par Illumina est particulièrement robuste et produit des résultats dont la profondeur de lecture est importante (donc avec un taux d'erreur faible). Cette technique est par ailleurs bien adaptée au séquençage de courtes séquences (*reads* d'une longueur de 100 à 200 bases). Enfin, une plateforme de séquençage d'Illumina était déjà disponible sur site au sein de l'Institut de Biologie et Pathologie et permettait un déploiement rapide et à moindre coût de cette nouvelle approche.

#### 4.1.3. Analyse bioinformatique

Le séquençage réalisé dans le cadre d'une approche de métagénomique en *shotgun sequencing* produit un nombre très élevé de séquences génétiques courtes qu'il faut retraiter afin de produire des résultats exploitables. Cette analyse comprend en particulier l'étape d'assignement taxonomique, qui consiste à attribuer, aux séquences obtenues, une espèce bactérienne, virale, parasitaire ou fongique (un taxon) unique. Deux approches ont principalement été développées à cette fin : l'une consiste à réaliser l'assemblage de génomes complets (assemblage en *contigs* puis en génomes complets) avant l'étape d'assignement taxonomique. L'autre méthode propose en revanche de réaliser ce même assignement taxonomique directement à partir des *reads* produits par le séquenceur sans réaliser d'assemblage ([Figure 9](#)). L'avantage apporté par cette deuxième technique est majeur en

termes de ressources informatiques utilisées puisque l'assemblage est une tâche lourde à réaliser.



**Figure 9** : Procédures classiques d'analyse de données métagénomiques. Adapté de Breitwieser *et al*, 2017, Briefings in Bioinformatics

En revanche, l'absence d'assemblage implique de travailler, pour l'assignement taxonomique, sur les séquences génétiques brutes qui sont plus courtes, ce qui a pour conséquence une diminution de la précision de cet assignement. Pour cette étude, différents types

d'algorithmes ont été évalués par le Département de Méthodologie de l'Information de Santé (DMIS) sur la Plateforme Bioinformatique Recherche (Dr Alban Caporossi, Pr Alexandre Moreau-Gaudry) avec lequel nous collaborons.

## 4.2. L'étude Dysbiose

### 4.2.1. Problématique(s)

L'étude « Dysbiose » menée par la clinique de pneumologie du CHU Grenoble – Alpes vise à décrire la flore microbienne respiratoire des patients atteints de mucoviscidose au cours d'un suivi longitudinal et ainsi mettre en évidence les interactions complexes entre les constituants de cet écosystème et l'hôte dans les phénomènes d'exacerbation de la maladie.

Par ailleurs, les prélèvements respiratoires prévus dans cette étude seront concomitants de prélèvements sanguins permettant l'exploration des processus inflammatoires dans le but d'une analyse jointe. Cette dernière est basée sur la mise en évidence de marqueurs protéiques et métaboliques par des analyses de métabolomique ciblée. Parmi les marqueurs recherchés, certains seront dosés dans le sérum (CRP et *serum amyloid A*) et d'autres dans la fraction d'air exhalé ou dans les prélèvements d'expectoration : IL-6, IL-8, CD-40 ligand, YKL-40 (chitinase-3-like protein 1, CHI3L1), MMP-9 (*Matrix metallo-proteinase 9*) et son inhibiteur tissulaire TIMP1. Enfin, le dosage des métabolites du tryptophane (tryptophane, hydroxy-tryptophane, kynurénine, hydroxy-kynurénine, anthranilate, hydroxy-anthranilate, acide kynurénique et sérotonine) sera également réalisé sur la fraction d'air exhalé.

### 4.2.2. Population et échantillons

Le recrutement des patients a pour objectif d'inclure entre 20 et 30 patients avec un suivi longitudinal sur une année. Le recrutement de ceux-ci sera réalisé au sein des centres



pluridisciplinaires de référence et de traitement de la mucoviscidose (CRCM) du CHU de Grenoble adulte et pédiatrique. L'inclusion de patients à partir de 12 ans a été validée par le Comité de Protection des Personnes (CPP).

La collecte d'échantillons aura lieu à chaque visite de routine et également pendant les exacerbations selon le calendrier prévu dans l'étude clinique (**Tableau 2**). Les échantillons pour l'étude seront prélevés en même temps que les échantillons dédiés à la routine, sans contrainte supplémentaire pour les patients.

**Tableau 2** : Suivi des patients : prélèvements et visites prévues dans l'étude « Dysbiose » tels que précisés par le protocole initial

	Sélection	Inclusion M0	Visite M3, M6, M9	Exacerbations	Visite fin de suivi M12
Consentement éclairé		✓			
Examen clinique	✓	✓	✓	✓	✓
Fonction respiratoire		✓	✓	✓	✓
Imagerie Standard		✓		✓	✓
<b>Prélèvements</b>		<b>Microbiote, biomarqueurs, microbiologie classique</b>			
Condensat d'air exhalé		✓	✓	✓	✓
Expectoration		✓	✓	✓	✓
Analyse sérique		✓	✓	✓	✓
Cultures classiques		✓	✓	✓	✓

## Prélèvement(s)

Le timing des prélèvements étudiés semble crucial : les résultats attendus seront différents avant exacerbation, pendant l'exacerbation et après l'exacerbation où l'on espère un retour au niveau basal. Mais l'importance de la réalisation de prélèvements précoces au moment des signes cliniques a montré son intérêt pour étudier la modification de flore, notamment virale (133). En revanche, il semble qu'il n'y ai pas d'impact de la thérapie antimicrobienne de fond sur le microbiote à l'état basal (58). L'état basal, qui correspond à un état stable en dehors de tout épisode aigu a été défini comme l'absence de changement dans la fonction respiratoire (sur la base de tests respiratoires) dans les 2 mois précédents le recueil (43).

La principale problématique est celle de la contamination de l'expectoration par les micro-organismes présents dans les voies aériennes supérieures avec l'introduction d'un biais d'échantillonnage (134). Pour les jeunes enfants, on observe des difficultés à produire des expectorations correctes. Dans ces conditions, l'utilisation de la kinésithérapie respiratoire est une aide précieuse mais parfois insuffisante.

## Condensats d'air exhalé

Ce type de prélèvement présente un intérêt important en raison de son caractère peu invasif. C'est un prélèvement qui semble être peu sensible *a priori* et, du fait de la faible biomasse microbienne collectée, nécessite une durée d'échantillonnage très important. Néanmoins, le prélèvement par recueil de condensats d'air exhalé a été validé par d'autres études même si certains auteurs mettent en doute la validité des résultats obtenus. Glendinning *et al.* (135) par exemple, comparent des condensats d'air exhalé et des prélèvements distaux protégés chez l'animal (le mouton) et ne retrouvent pas de correspondance entre les deux sur le microbiote bactérien (exploré à l'aide de la PCR 16S). D'abord, la quantité d'ADN recouvré est

sans commune mesure, mais, même s'il existe un certain recouvrement des populations certaines disparités ne semblent pas explicables. La question de l'échantillonnage spatial obtenu par une telle technique de prélèvement reste ouverte.

#### *4.2.3. Intérêt de l'étude et résultats attendus*

L'obtention de résultats concernant le microbiote bactérien, viral, parasitaire et fongique en complément des résultats obtenus par les techniques classiques permet de confirmer la présence d'un micro-organisme donné mais également d'élargir le spectre des pathogènes recherchés. Cela pourrait par exemple permettre une meilleure personnalisation du traitement en vue d'obtenir de meilleurs résultats. La mise en évidence de virus au décours d'un épisode d'exacerbation pourrait encourager l'utilisation d'antiviraux classiques ou novateurs. Le palivizumab (Synagis®) pourrait par exemple trouver sa place dans la prévention de l'infection par le VRS, mais cela nécessite de traiter précocement pour obtenir un effet significatif (136).

Il semble qu'il soit possible, d'après les études menées sur le microbiote intestinal, de modifier artificiellement un microbiote donné. En influant sur la présence de certains micro-organismes, des équipes prédisent des gains substantiels en termes de qualité de vie des patients : une diminution du nombre et de la fréquence des hospitalisations, ainsi qu'une amélioration de la fonction respiratoire pourraient être observées. Cela a déjà démontré en partie par l'administration de *Lactobacillus* (137) pour le microbiote intestinal. Les connaissances concernant le microbiote respiratoire sont encore parcellaires (121) et insuffisantes pour envisager des thérapeutiques ciblées à court terme, mais on peut attendre le développement de nouvelles approches dans le traitement des pathologies respiratoires chroniques à plus long terme.

Cette étude sera la première mise à l'épreuve de notre technique de métagénomique infectieuse ciblée par capture en milieu liquide. Celle-ci devra être utilisable pour l'exploration du microbiome et en particulier du virome respiratoire. Elle pourra, par la suite, être utilisée pour l'exploration et la description des autres microbiotes. La description et la compréhension des différents microbiotes (respiratoire, intestinal, vaginal, cutané, ...) apparaissent comme un enjeu important de la médecine personnalisée. Il est donc urgent de se donner les outils adaptés à leur exploration.

Outre la description des microbiotes de patients atteints de mucoviscidose, cette étude devrait permettre de dépasser le cas Individuel pour apporter des réponses globales, concernant l'ensemble de la population étudiée. Plusieurs approches seront mises à contribution : d'abord la comparaison des microbiotes inter-patients, mais également le suivi longitudinal de l'évolution des microbiotes chez un même patient. Enfin, la comparaison des microbiotes en période stable, en pré-exacerbation, pendant l'exacerbation et finalement en post-exacerbation apportera des réponses quant aux évolutions de ce microbiote au cours de la pathologie.

Enfin, le travail effectué ici s'inscrit dans un cadre plus large qui vise à mettre en place un nouvel outil de diagnostic applicable à la pathologie infectieuse au Centre Hospitalier Universitaire de Grenoble – Alpes (CHUGA). C'est l'ensemble du travail préparatoire au déploiement de cet outil qui fait l'objet de cette synthèse. Nous détaillerons donc dans ce documents les différentes étapes qui ont permis la mise en place d'une technique non-invasive d'exploration par la métagénomique du microbiote respiratoire à l'aide du séquençage de nouvelle génération sur des échantillons artificiels d'abord puis sur des prélèvements issus de patients atteints de mucoviscidose.

## PARTIE II – MATERIELS ET METHODES

Nous décrivons dans ce chapitre les méthodes mises en place pour commencer la validation de la technique de métagénomique infectieuse par capture semi-ciblée. L'approche que nous avons choisie de mettre en place au laboratoire doit permettre de répondre à l'ensemble des interrogations soulevées ci-avant. Les choix techniques réalisés sont détaillés ci-après, résumant ainsi la démarche expérimentale menée. Ce travail s'est donc organisé en trois parties. La première a consisté à tester le pipeline biologique et d'analyse sur une puce de capture commerciale issue des travaux de l'équipe de Lipkin : VirCapSeq-VERT (126) qui cible les virus infectant les vertébrés. En parallèle de ce travail, une puce de capture répondant à notre objectif de détection simultanée des virus, bactéries et champignons a été *designée*. Enfin, cette nouvelle puce a été évaluée sur des échantillons contrôles artificiels et, pour la partie virus, comparée aux résultats obtenus avec la puce VirCapSeq-VERT.

### 1. Mise en œuvre expérimentale

La méthode analytique que nous souhaitons mettre en place se découpe donc en plusieurs étapes ou sous-processus résumés à la **Figure 10** ci-dessous. Dans un premier temps, nous avons fait abstraction des étapes liées aux prélèvements de patients pour nous focaliser sur les étapes pré-analytiques et analytiques au laboratoire.



**Figure 10** : Logigramme des étapes pré-analytiques, analytiques et post-analytiques

### 1.1. Echantillons artificiels

Des échantillons artificiels ont été créés pour évaluer cette nouvelle méthode. Un contrôle négatif a été réalisé en utilisant de l'eau stérile ultrapure. Cette eau était la même que celle utilisée dans les différentes étapes du processus technique (rinçages, ...). Elle a subi l'intégralité du processus d'analyse comme un échantillon et permet de mettre en évidence de potentielles contaminations. Un témoin positif a également été réalisé avec les micro-organismes suivants : *Influenza virus A*, *Cytomegalo-virus (CMV)*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Pneumocystis jiroveci* et *Aspergillus fumigatus*. Ces micro-organismes proviennent d'échantillons de patients stockés dans la collection du laboratoire (DC-2008-680, Ministère de l'Enseignement Supérieur et de la Recherche). Ces six micro-organismes étaient introduits dans les échantillons artificiels, créés à partir de prélèvements de lavages broncho-alvéolaires (LBA) de patients hospitalisés. La présence des pathogènes bactériens a été confirmée par des qPCR spécifiques telles que décrites par Gadsby *et al*. La recherche de cytomégalovirus (CMV) a été réalisée à l'aide du kit CMV R-GENE® de BioMérieux. La vérification de la présence du virus *Influenza A* a été faite grâce au kit Altona RealStar® Influenza RT-PCR. Les Ct (*cycle thresholds*) obtenus pour les différents pathogènes sont rapportés dans le **Tableau 3** ci-dessous.

**Tableau 3** : Evaluation semi-quantitative des pathogènes recherchés dans les échantillons artificiels.

Ct (*cycle thresholds*) mesurés pour 200 µL de prélèvements.

Pathogène	Ct ( <i>cycle threshold</i> )
<i>Pneumocystis jirovecii</i>	31,7
<i>Aspergillus fumigatus</i>	31,2
<i>Staphylococcus aureus</i>	23,0
<i>Pseudomonas aeruginosa</i>	30,7
<i>Influenza A</i>	22,1
<i>Cytomegalovirus</i>	35,1

## 1.2. Echantillons patients

### 1.2.1. Condensats d'air exhalé

Afin de valider cette méthode de recueil des prélèvements, des condensats d'air exhalé ont été collectés auprès de deux patients volontaires. Les patients concernés ont été informés et ont donné leur consentement pour l'utilisation de ces prélèvements. Les condensats d'air exhalés ont été collectés au cours du mois de Janvier 2018 au sein de la Clinique de Pneumologie du CHU de Grenoble – Alpes, auprès de patients adultes atteints de la mucoviscidose. Ces patients étaient hospitalisés dans le cadre de la prise en charge de leur pathologie. Le recueil de leur consentement via un formulaire simplifié a été validé par le CPP par un amendement du protocole d'étude. La collection de condensats d'air exhalés fait appel à un dispositif dédié qui permet, grâce à un différentiel de température appliqué aux gaz expirés, de condenser la vapeur d'eau émise, emprisonnant les micro-organismes et les acides nucléiques contenus dans l'air exhalé (Cf. Partie I - 4.2.2 Population et échantillons).

### 1.2.2. *Expectorations*

Des expectorations de ces deux mêmes patients ont été recueillies de façon concomitante. Ces expectorations correspondent aux prélèvements réalisés habituellement pour le diagnostic de routine. Elles ont donc été séparées en deux, une partie servant à la réalisation de notre analyse et la seconde partie restant dévolue au traitement à but diagnostic par des techniques conventionnelles telles que détaillées précédemment.

## 1.3. Analytique

### 1.3.1. *Extraction et synthèse de cDNA*

Les acides nucléiques pour le panel virologique ont été extraits selon un protocole d'extraction manuelle à l'aide du kit QIAamp® Viral RNA Mini Kit (Qiagen, Courtabœuf, France). En ce qui concerne *Pneumocystis* et *Aspergillus*, nous avons utilisé le kit QIAamp DNA easy mini kit, selon les spécifications du fournisseur. Le matériel génétique issu des bactéries était disponible sous forme d'extraits au laboratoire. Les acides nucléiques totaux extraits ont été traités par *reverse transcriptase* pour obtention des cDNA. Cette étape a été effectuée grâce au kit SuperScript® III (Invitrogen® ThermoFisher, Dardilly, France). La synthèse de cDNA a été réalisée grâce à l'utilisation d'amorces aléatoires (*random hexamers*) fournis dans le kit en suivant les étapes indiquées par le fournisseur du kit dans la notice. La synthèse des cDNA double brin a été réalisée en utilisant le kit Next Ultra II Non-directional RNA 2<sup>nd</sup> strand synthesis module (New England Biolabs, Evry, France). L'ensemble des étapes réalisé ici l'ont été en suivant les recommandations du fournisseur dans la documentation technique. Un contrôle de qualité des extraits a été réalisé par dosage du mélange cDNA double brin / ADN par technique Qubit (Invitrogen® ThermoFisher, Dardilly, France).



La purification des ADN ainsi obtenus a été réalisée grâce à la technique AMPure XP (Beckman Coulter, Villepinte, France) et l'utilisation d'une puce *High sensitivity Agilent Bioanalyser* (Agilent Technologies, Les Ulis, France). Celui-ci utilise des billes magnétiques dans un tampon de polyéthylène-glycol (PEG). La concentration de PEG permettant de sélectionner la taille des fragments purifiés, nous avons utilisé un ratio AMPure / échantillon de 1,8. Cela a permis une sélection des fragments  $\geq 100$  pb.

### 1.3.2. Fragmentation

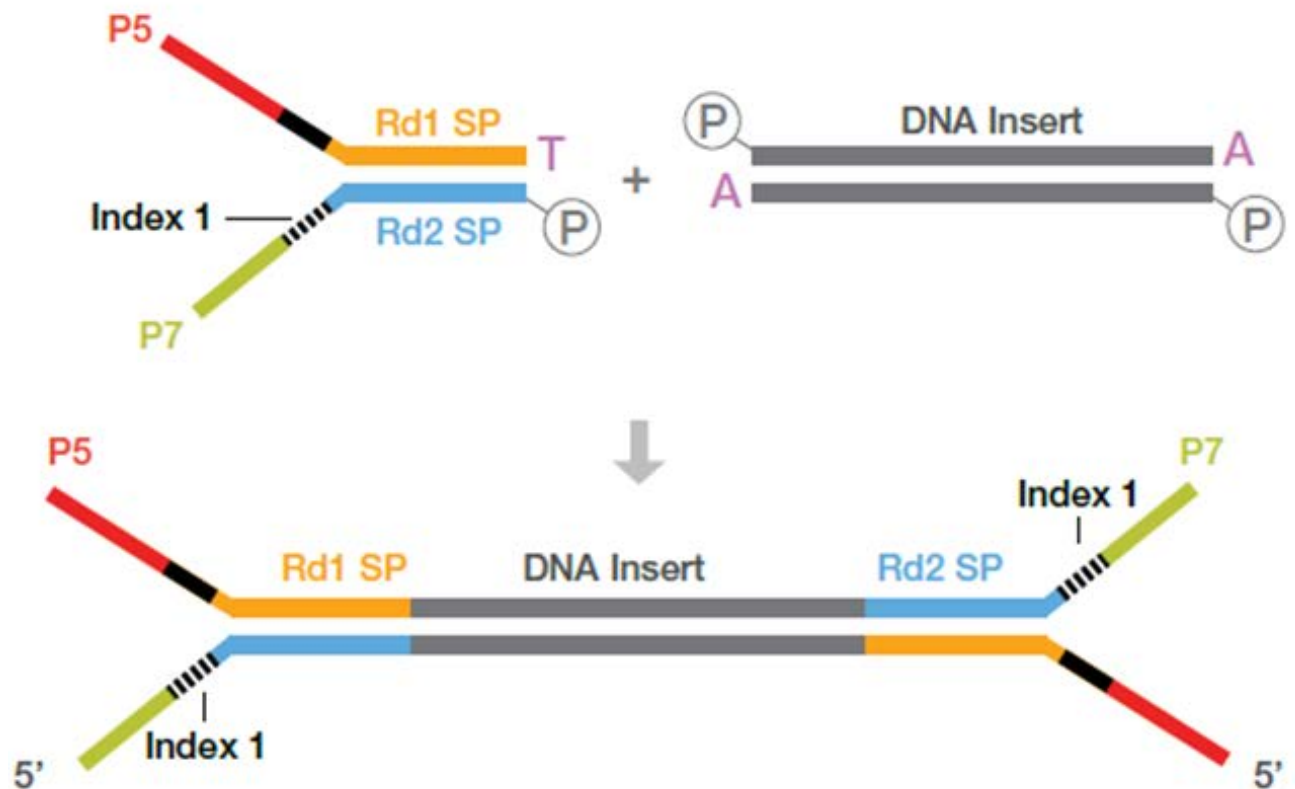
La fragmentation enzymatique des acides nucléiques a été réalisée au moyen d'un kit KAPPA® Hyper+ Library preparation kit (Roche, Mannheim, Allemagne) et ce, de manière à obtenir des fragments de 200 à 250 pb. Après avoir effectué plusieurs tests, la durée de 15 minutes nous a permis d'obtenir la taille souhaitée de façon reproductible. La quantité de cDNA double brin / ADN utilisée pour la fragmentation était de 50 ng quand cela était possible (c'est-à-dire pour les expectorations et les contrôles positifs). Elle était de  $< 10$  ng pour les condensats d'air exhalé (CAE).

### 1.3.3. Préparation des librairies

La préparation des librairies est une étape indispensable préalable au séquençage NGS avec la technique proposée par Illumina. Elle consiste en la ligation d'adaptateurs de part et d'autre du fragment d'ADN à séquencer. Celle-ci permet d'obtenir des échantillons reconnaissables (barre-codés) et prêts à l'emploi pour le séquenceur (**Figure 11**).

L'ajout d'identifiants uniques pour les échantillons permet de *pooler* ces échantillons au sein d'une même étape de séquençage et donc réduit considérablement les coûts. Dans notre cas, un seul index était utilisé. Cette étape a été également réalisée avec le KAPPA® Hyper+ Library preparation kit utilisé précédemment. La quantité d'acides nucléiques introduits était de 50

ng par échantillon et la durée de la réaction de ligation a été de 30 minutes. La concentration des adaptateurs utilisée était adaptée selon la quantité de matrice dans les réactions de ligation. Le ratio adaptateurs/inserts était de 200 : 1.



**Figure 11** : Schéma des insertions réalisées dans le cadre de la préparation des bibliothèques (adapté de Roche®)

Une étape de purification a été réalisée grâce à la technique de *Solid Phase Reversible Immobilisation* (SPRI) Cleanup mise en œuvre par Agencourt AMPure XP. La sélection des acides nucléiques était faite sur la taille des fragments pour permettre une élimination des petits fragments. Cela garantissait l'élimination des dimères et monomères de ligands indésirables pour la suite de l'analyse. Le ratio AMPure / échantillon utilisé était de 0,8.

#### 1.3.4. *Size selection et amplification pré-capture*

La sélection des librairies sur le critère de leur taille a également été réalisée en technique AMPure XP. Cette procédure permettait une double sélection en taille et permettait l'augmentation de l'efficacité du séquençage. Les librairies avaient une taille attendue de 320 à 370 pb. Une étape de PCR quantitative en temps réel (qPCR) était réalisée sur les librairies pour la détermination du nombre de cycles d'amplification à réaliser pour optimiser le processus. Celle-ci a été réalisée au moyen du kit KAPA® library quantification kit (Roche, Mannheim, Allemagne) sur un appareil LightCycler 480 de Roche.

L'amplification avant capture des librairies a été réalisée par PCR. Un nombre limité de cycles a été réalisé pour ne pas introduire de biais. En revanche, la réalisation de 7-8 cycles était recommandée par le fabricant pour obtenir une quantité suffisante de matériel génomique (1 µg) pour les étapes suivantes de capture et de séquençage. En pratique, nous avons réalisé 9 cycles pour les échantillons d'expectorations et les témoins positifs. Les échantillons issus de CAE ont nécessité 19 cycles. Une étape de purification a alors été réalisée, toujours à l'aide de la technique SPRI cleanup sur AMPure XP en utilisant un ratio AMPure / échantillon de 1,8.

Enfin, nous avons réalisé un contrôle de qualité des fragments d'acides nucléiques disponibles à cette étape par la visualisation des répartitions des tailles en profil Agilent Bioanalyser sur une puce *High sensitivity DNA*. Une dernière étape de quantification des librairies a été réalisée grâce à la technique Qubit (ThermoFisher, Dardilly, France) à l'aide d'un kit *High sensitivity DNA*. Un pool a été obtenu par mélange dans des proportions stœchiométriques de quatre échantillons patients barre-codés, d'un témoin positif et d'un contrôle négatif.

## 1.4. Capture par hybridation en milieu liquide

Conformément à l'utilisation du kit utilisé pour le séquençage, nous avons procédé à l'ajout des *blockers*. Ces courts oligonucléotides permettent d'augmenter la spécificité de la technique en réduisant le séquençage *off-target* non souhaité. La capture des bibliothèques d'intérêt a été réalisée par hybridation en milieu liquide avec les sondes biotinylées oligonucléotidiques synthétisées par NimbleGen. Cette méthode a été utilisée sous la forme d'un kit commercialisé par Roche : SeqCap EZ Prime Choice Probes (ref. 8247447001).

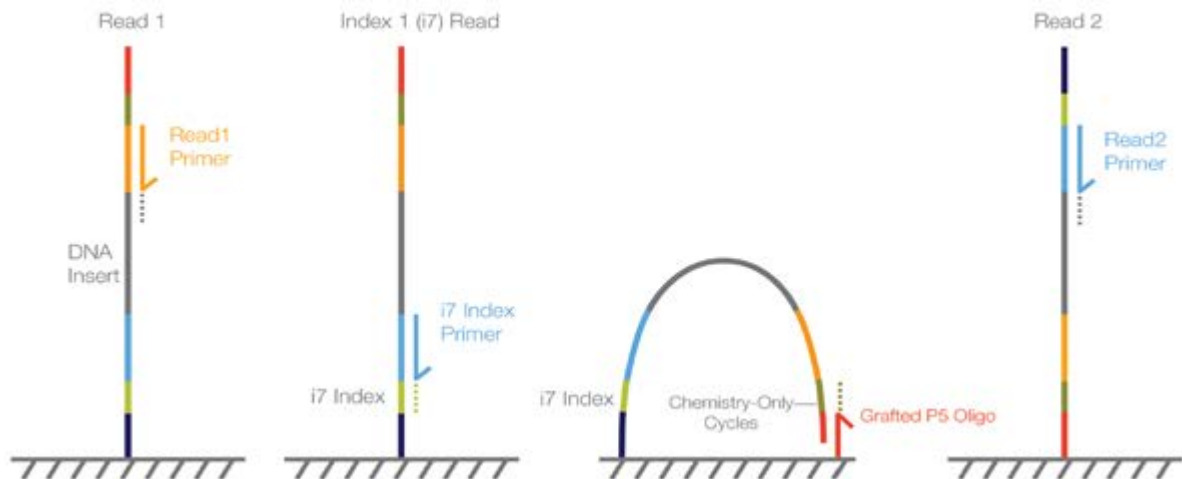
La capture en milieu liquide a été réalisée en une seule fois comme préconisé dans le kit de NimbleGen. Les conditions de réalisation de cette capture étaient les suivantes : la capture a duré 17 heures et a eu lieu à 47°C. Plusieurs étapes de lavage, une fois à 47°C et deux fois à température ambiante avec des tampons de plus en plus stringents permettaient de se débarrasser de toutes les impuretés et des séquences génétiques non souhaitées. La solution obtenue était re-suspendue dans 50 µL d'eau. Une étape de PCR post-capture a été réalisée avant purification des produits obtenus. Cette étape, identique à l'amplification pré-capture, a nécessité 14 cycles.

Une dernière étape de purification par *SPRI cleanup* avec kit AMPure XP a été réalisée selon le même protocole que précédemment utilisé. Un contrôle de la qualité du matériel génétique était obtenu grâce à la réalisation d'un profil Agilent Bioanalyser. Une quantification précise des acides nucléiques était réalisée grâce à la technique Qubit avant séquençage.

## 1.5. Séquençage

Le séquençage en lui-même a été réalisé sur plateforme Illumina NextSeq 500 disponible à l'Institut de Biologie et Pathologie, au CHU de Grenoble. La durée de réaction était de 27h pour accomplir le nombre de cycles de séquençage nécessaires sur une cellule (*flowcell*)

HighOutput avec un séquençage en 2 x 150 pb. L'ensemble des échantillons préparés pour la mise au point de la technique étaient barre-codés et multiplexés au sein d'une seule et même *flowcell*.



**Figure 12** : Principe du séquençage (*Dual-Indexed Sequencing*) sur cellule avec brins appariés (*Paired-end Flow-Cell*). Adapté de la documentation technique fournie par Illumina®

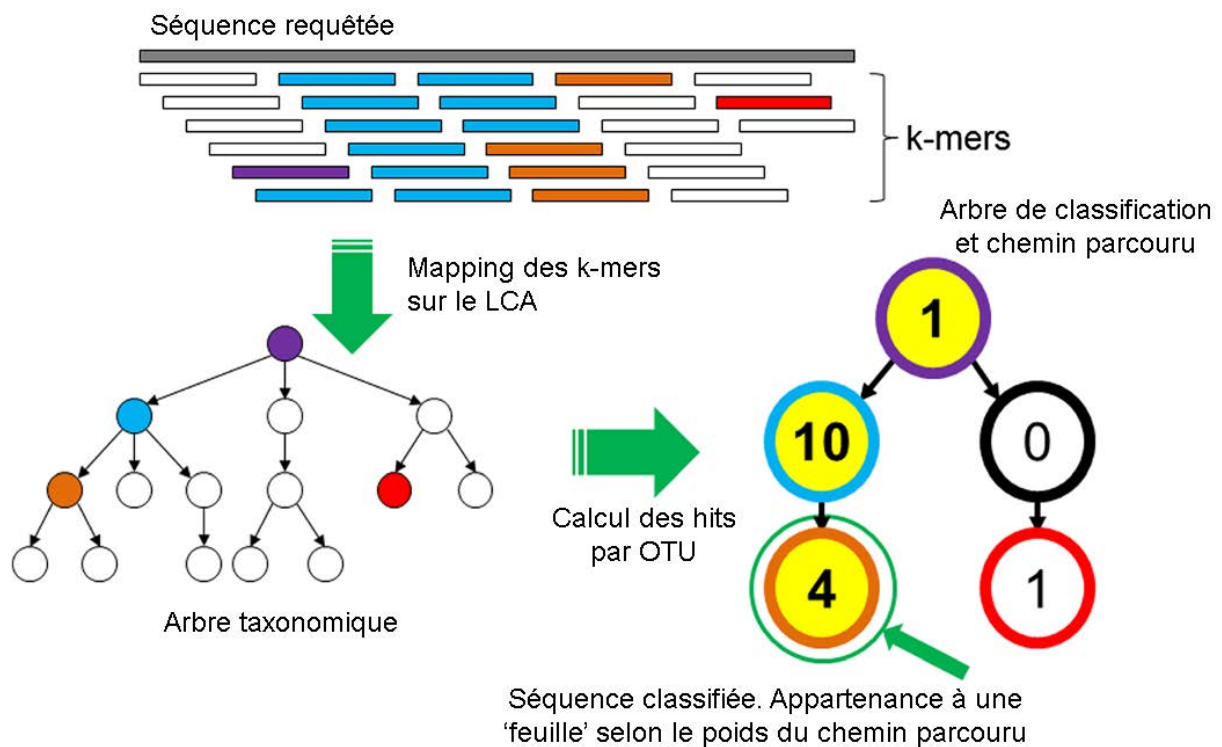
## 1.6. Analyse bio-informatique

Cette analyse a été réalisée par la plateforme Bio-informatique Recherche du Département de Méthodologie de l'Information de Santé (DMIS) du CHU Grenoble – Alpes. Les données brutes de séquences issues de l'étape de séquençage ont été traitées par un pipeline analytique 'maison', résultat d'un assemblage de différents algorithmes de référence publiés dans la littérature et mis à disposition par les équipes qui les ont développés.

En sortie de séquenceur, le démultiplexage des séquences *poolées* a été assuré par le logiciel maison d'Illumina (BaseSpace). Au début du pipeline de traitement, les séquences de chaque patient ont été filtrées sur des critères de qualité puis les séquences associées à du matériel génétique humain ont été éliminées. Cette étape a été réalisée avec le logiciel Kraken (138) dans sa version 2.0 avec une base de donnée constituée à partir du génome de référence

humain GRCh38 (hg38). Les séquences homologues au génome humain (c'est-à-dire dont plus de 95 % des sous-séquences de k nucléotides ou k-mers correspondaient à des séquences présentes dans ce génome humain de référence) étaient éliminées et n'étaient pas analysées dans la suite du pipeline bio-informatique.

L'étape clé de cette étude était l'assignement taxonomique, dans lequel on cherche à faire correspondre, pour chaque séquence obtenue, une espèce taxonomique (ou OTU : *operational taxonomic unit*) donnée.



**Figure 13** : Principe de fonctionnement de l'algorithme Kraken. Adapté de Wood et al., 2014. LCA : Last Common Ancestor, ou plus proche ancêtre commun. OTU : Operational Taxonomic Unit, ou unité taxonomique opérationnelle

Cette étape a également été réalisée avec le logiciel Kraken dans sa version 2.0 avec la base de données standard produite par le développeur de cet outil et avec une base de données 'maison' construite à partir des séquences sélectionnées pour le design de notre puce de

capture (Cf. 2 - Design d'une puce de capture pan-microbienne DB-Chip). Le principe de fonctionnement de Kraken est résumé à la **Figure 13**.

Les résultats obtenus après analyse par Kraken ont été présentés sous forme de représentations visuelles dynamiques grâce au logiciel Krona (139). Cet outil permet de visualiser le taxon final de l'arbre phylogénétique (feuille) et le nombre de représentants de ce même taxon au sein d'une représentation dynamique circularisée.

## 2. Design d'une puce de capture pan-microbienne DB-Chip

Les puces de capture NimbleGen telles que la puce VirCapSeq-VERT (126) sont constituées de sondes chevauchantes d'une taille comprise entre 70 et 150 pb avec un recouvrement compris entre 25 et 50 nucléotides. La société NimbleGen les synthétise avec son propre algorithme à partir des séquences d'intérêt fournies par les laboratoires. Nous avons donc créé notre propre base de données par regroupement des séquences d'intérêt de banques de données existantes en ciblant les séquences de référence d'espèces ou de gènes spécifiques d'espèces. Enfin, avant d'adresser notre sélection à NimbleGen, nous l'avons validée *in-silico*.

### 2.1. Base de données

Une base de données de séquences nucléotidiques a donc été générée pour la synthèse des sondes de capture. Cette base de données devait regrouper des séquences de tous les pathogènes potentiels ainsi que les espèces non-pathogènes susceptibles d'être retrouvées dans les échantillons humains étudiés. Elle devait couvrir les règnes bactérien et fongique ainsi que l'ensemble des virus pouvant infecter l'homme. La limite de données utilisables était fixée par le fabricant de la solution de capture NimbleGen : celle-ci devait contenir au maximum jusqu'à 200 Mb réparties sur un maximum de 2.100.000 sondes.

## 2.2. Clustering

Une étape de clustering a été réalisée par la Plateforme Bioinformatique Recherche du DMIS pour supprimer la redondance et réduire le volume de données de travail en éliminant les doublons et les séquences à forte homologie avec une séquence représentative conservée. L'algorithme CD-HIT (140,141) a été utilisé en autorisant un seuil d'identité de séquences de 90 %. A titre de comparaison, Lipkin *et al.* (126) avaient travaillé avec un seuil à 96 % mais sur des séquences virales uniquement. De leur côté, Wylie et collègues avaient choisi un seuil à 90 % pour leur étude (125) mais en utilisant un autre algorithme (UCLUST). Le clustering a été réalisé sur des fragments de 1kb avec un recouvrement maximum de 100 bases entre fragments successifs.

## 2.3. Validation *in-silico*

La validation de la base de données pour la création de la puce de capture a été réalisée par « mapping » *in-silico* de la base de données post-clustering sur des séquences de référence de virus, bactéries et champignons choisies manuellement et dont les séquences ne sont pas représentées dans la base de départ. Cette validation réalisée par le DMIS avait pour but de contrôler l'exhaustivité de la base de données, garante de l'exhaustivité de la capture réalisée sur les échantillons de patients. Le choix des virus s'est porté principalement sur les virus respiratoires, sur des variants rares ou d'importance clinique significative, auxquels ont été ajoutés quelques virus non-respiratoires pour contrôler la capacité de la puce de capture à efficacement enrichir les prélèvements en virus sans a priori. Une approche similaire a été entreprise pour les bactéries et les champignons.

Enfin, une dernière étape de validation *in-silico* a été réalisée. Il s'agissait d'une validation négative : c'est-à-dire la vérification de l'absence de sélection de séquences non recherchées



(séquences génétiques humaines principalement). Cette analyse a été réalisée par NimbleGen en utilisant le génome humain de référence HG38.

### 3. Validation de la puce de capture DB-Chip

La puce de capture DB-Chip ainsi *designée* a ensuite été testée sur des échantillons artificiels afin d'en valider le fonctionnement. En fonction des résultats obtenus, une seconde version de cette puce pourra être synthétisée. Les échantillons de patients ne seront testés qu'après validation sur des échantillons artificiels du *design* final et après caractérisation de la puce obtenue (sensibilité, spécificité, reproductibilité, quantification relative, ...). Afin de permettre une quantification relative et de limiter les biais introduits, les différents échantillons de suivi d'un même patient seront passés à la fin du recueil lors d'une même procédure, selon le principe analytique décrit ci-dessus (**Figure 10**).

#### 3.1. Echantillons

##### 3.1.1. Echantillons artificiels

La technique est donc, dans un premier temps, évaluée sur des échantillons artificiels qui serviront ensuite de contrôle. Trois panels de pathogènes ont été testés seuls ou en mélange. Pour les virus il s'agit d'un panel commercial de virus « Non WHO Reference Material Viral Multiplex Reference NIBSC code : 11-242-xxx » dont le fabricant précise qu'il contient des ADN humains et bovins et potentiellement du virus de l'hépatite B (AgHBs positif) mais qu'il a été testé négatif pour le VIH et le VHC. Pour les bactéries, le panel utilisé est Metagenomic Control Material for Pathogen Detection (ATCC® MSA4000) (ATCC, Manassas, VA, USA) qui regroupe des ADN génomique totaux à des proportions déterminées en PCR digitale.

**Tableau 4** : Liste des pathogènes utilisés pour la création d'échantillons artificiels lors des tests de la technique au laboratoire

<b>Virologie</b>	<b>Bactériologie</b>	<b>Parasitologie / Mycologie</b>
Adenovirus 2	Acinetobacter baumannii	Pneumocystis jirovecii
Adenovirus 41	strain 5377 (0.1%)	Aspergillus fumigatus
Human herpesvirus 1	Enterococcus faecalis strain	
Human herpesvirus 2	V583 (0.7%)	
Human herpesvirus 3 (VZV)	Escherichia coli strain	
Human herpesvirus 4 (EBV)	CFT073 (1.4%)	
Human herpesvirus 5 (CMV)	Klebsiella pneumoniae	
Rotavirus A	subsp. pneumoniae strain	
Astrovirus	MGH78578 (14.4%)	
Norovirus GI	Neisseria meningitidis strain	
Norovirus GII	FAM18 (28.9%)	
Sapovirus C12	Pseudomonas aeruginosa	
Coronavirus 229E	strain PAO1-LAC (0.3%)	
Coxsackievirus B4	Staphylococcus aureus	
Rhinovirus A39	subsp. aureus (MRSA) strain	
Parechovirus 3	FPR3757 (0.7%)	
Influenza A virus H1N1	Staphylococcus aureus	
Influenza A virus H3N2	subsp. aureus (MSSA) strain	
Influenza B virus	TCH959 (14.4%)	
Metapneumovirus A	Streptococcus agalactiae	
Parainfluenzavirus 1	strain 2603 V/R (2.9%)	
Parainfluenzavirus 2	Streptococcus pneumoniae	
Parainfluenzavirus 3	strain Spain 23F-1 (28.9%)	
Parainfluenzavirus 4	Streptococcus pyogenes	
Respiratory syncytial virus A2	strain SF370 (7.2%)	

Le panel de mycologie utilisé pour la mycologie est identique à la partie mycologie du panel utilisé pour l'évaluation de la méthode analytique (cf. 1.1.1). Les pathogènes présents dans ces différents panels testés sont repris dans le **Tableau 4** ci-dessus.

### 3.1.2. *Echantillons de patients*

En parallèle du test des échantillons artificiels, le recrutement de patients pour l'étude Dysbiose a commencé. Le protocole a été validé en CPP avec déclaration de l'essai clinique sur le site internet <https://clinicaltrials.gov/>. Les échantillons pour l'étude étaient prélevés en même temps que les échantillons dédiés à la routine, sans contrainte supplémentaire pour les patients. Deux types d'échantillons ont été collectés auprès des patients en fonction de leur compliance et de leur état clinique. Le prélèvement de choix était l'obtention de crachats profonds dans le cadre d'une expectoration aidée. Nous avons également obtenu des condensats d'air exhalé afin de tester la pertinence de ce mode de recueil.

Les prélèvements reçus au Département des Agents Infectieux (DAI) de l'Institut de Biologie et Pathologie (IBP) du CHU de Grenoble étaient prétraités dès leur arrivée pour permettre la prise en charge dans le circuit d'analyse de routine. Ce pré-traitement consistait en la fluidification et la dégradation des mucines formant le mucus respiratoire. La fluidification était réalisée avec du Digest-Eur® (Eurobio, Courtabœuf, France) pendant au moins 15 min.

Alors que les échantillons dédiés à l'analyse de routine étaient pris en charge par les laboratoires respectifs, les échantillons dédiés à notre étude étaient stockés à -80°C sur toute la durée de l'étude en attendant l'analyse *batchée* une fois le recrutement et le suivi terminé. Les échantillons prétraités ont été lysés mécaniquement à l'aide de billes métalliques millimétriques. L'extraction des acides nucléiques a ensuite été réalisée à la main (technique

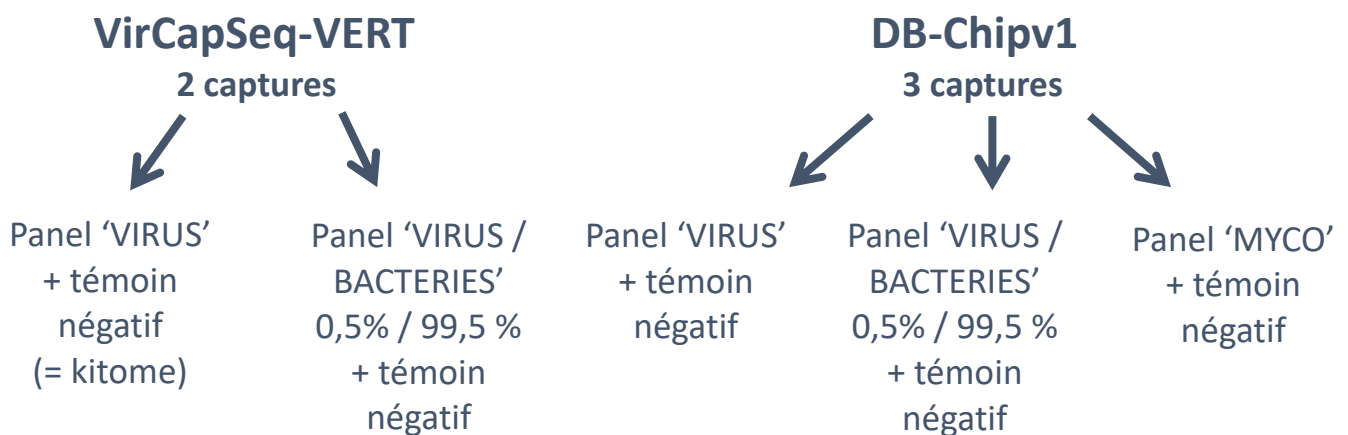
non automatisée) avec le kit Qiagen® All-Prep selon le protocole proposé dans la notice du fabricant.

### 3.2. Protocole analytique

Cette approche avait trois objectifs analytiques :

- Evaluer l'efficacité de notre puce DB-Chip sur des panels de pathogènes connus testés à différentes conditions de mélange
- Comparer sur le panel de virus l'efficacité de notre puce pan-microbienne avec celle de la puce dédiée uniquement aux virus (VirCapSeq-VERT)
- Valider l'ensemble du pipeline biologique et bio-informatique grâce à des données d'entrée maîtrisées.

Pour ce faire, le plan expérimental a été le suivant (**Figure 14**) :



**Figure 14** : Plan expérimental de la mise au point de la méthode de capture

Le protocole d'analyse utilisé pour la partie biologique (*wet-lab*) était identique à celui utilisé pour la première évaluation de la méthode analytique (Cf. chapitre 1 ci-dessus), à l'exception de quelques modifications mineures : au cours de la capture, les lavages nécessaires entre les

différentes étapes sont réalisés à température contrôlée. Ces derniers ont été réalisés à 47°C pour ce nouveau protocole afin d'augmenter la stringence et la spécificité de la capture. De la même manière, des étapes de vortex ont été ajoutées au cours de la capture pour les mêmes raisons. Enfin, nous avons légèrement augmenté le nombre de cycles de PCR réalisés lors de l'amplification en pré- et post-capture (trois cycles ajoutés à chacune des deux étapes).

Des qPCR quantitatives en amont et après l'étape de capture ont été réalisées dans le but de comparer les échantillons et évaluer l'enrichissement relatif des acides nucléiques des pathogènes visés. Ces qPCR ciblaient les gènes suivants : MP1 et MP2 (*Matrix Protein*) pour le virus *Influenza A*, NP (Nucléoprotéine) pour le virus *Influenza B*, ppUL83 pour le CMV, gyrB pour *Pseudomonas aeruginosa*, nuc pour *Staphylococcus aureus*, lytA pour le *Streptococcus pneumoniae*, fucK pour *Haemophilus influenzae* et gltA pour *Klebsiella pneumoniae*. Le CMV a été quantifié en utilisant le kit CMV R-gene (BioMérieux) sans modification du protocole. Les virus de la grippe A et B ont été quantifiés à l'aide du kit R-Dia Flu® (Diagenode, Liège, Belgique). L'enrichissement des génomes bactériens a été mesuré à l'aide d'une qPCR multiplexe telle que détaillée précédemment (142).

Par ailleurs, la classification taxonomique a été étudiée à l'aide du même algorithme Kraken 2, mais en utilisant cette fois ci alternativement deux bases de données. L'une est la base par défaut de Kraken tel que précédemment décrit. L'autre a été réalisée en utilisant la base de données utilisée pour la mise au point de la puce de capture. Ainsi, la même base de données est utilisée pour « filtrer » les échantillons biologiques (c'est la phase de capture), puis pour associer chacune des séquences retrouvées avec un taxon donné. Deux résultats par échantillons sont ainsi obtenus, ce qui permettra de comparer les deux approches.

## PARTIE III – RESULTATS

La mise au point des techniques analytiques s'est donc déroulée de la manière suivante : six échantillons ont servi à la première étape de validation de la méthode à l'aide de la puce de capture VirCapSeq-VERT. Les enseignements tirés de cette première approche ont guidé le développement d'une puce de capture personnalisée DB-Chip. Enfin, des tests ont été réalisés à l'aide de cette nouvelle puce et en comparaison avec la précédente sur des panels d'échantillons calibrés. Les résultats obtenus sont détaillés ci-après.

### 1. Evaluation de la méthode analytique

#### 1.1. Préparation des échantillons

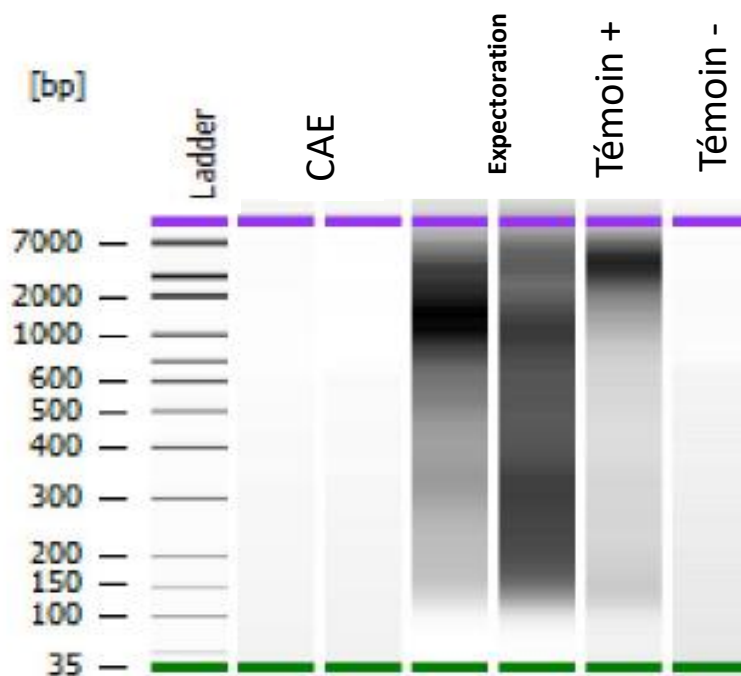
L'évaluation de la méthode ici présentée s'est déroulée au Département des Agents Infectieux du CHU en février 2018. Une première analyse a été réalisée sur 6 échantillons. Quatre échantillons étaient issus de 2 patients atteints de mucoviscidose. Pour chaque patient, un condensat d'air exhalé et une expectoration étaient testés. Un contrôle positif et un contrôle négatif ont également été testés.

Un premier contrôle de qualité évaluant la masse des acides nucléiques collectés après extraction et synthèse du cDNA double brin retrouvait les valeurs suivantes (dosage d'acides nucléiques avec une technique Qubit de quantification par fluorimétrie) avant la préparation des bibliothèques :

- Condensats d'air exhalé : < 0,005 ng/μL
- Témoin négatif : < 0,005 ng/μL
- Témoin positif : 3 ng/μL

- Expectorations : 7 et 11 ng/μL respectivement

La réalisation d'une électrophorèse de ces échantillons était concordante avec les résultats ci-dessus, et rapportée à la **Figure 15** ci-après. La construction des librairies a été réalisée à partir de 100 ng d'acides nucléiques pour les échantillons 'positifs' (témoin positif et expectorations). Moins d'un nanogramme d'acides nucléiques était disponible pour cette même opération pour les échantillons 'pauvres' (témoin négatif et condensats d'air exhalé).

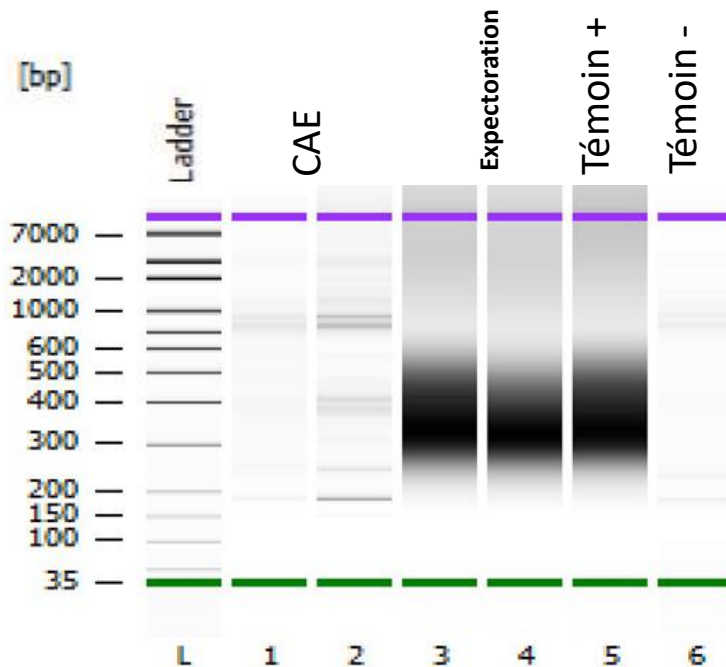


**Figure 15** : Gel d'électrophorèse des acides nucléiques obtenus après extraction et synthèse du cDNA des 6 échantillons initiaux et avant synthèse des librairies. (CAE = Condensat d'Air Exhalé)

La sélection par la taille des fragments en pré-capture a retenu des acides nucléiques de longueur comprise entre 250 et 450 pnb. L'amplification des librairies a nécessité 19 cycles de PCR pour les échantillons 'pauvres' et seulement 9 cycles pour les échantillons 'positifs'. De la

même façon, l'électrophorèse réalisée après synthèse des librairies est présentée à la **Figure**

**16.**

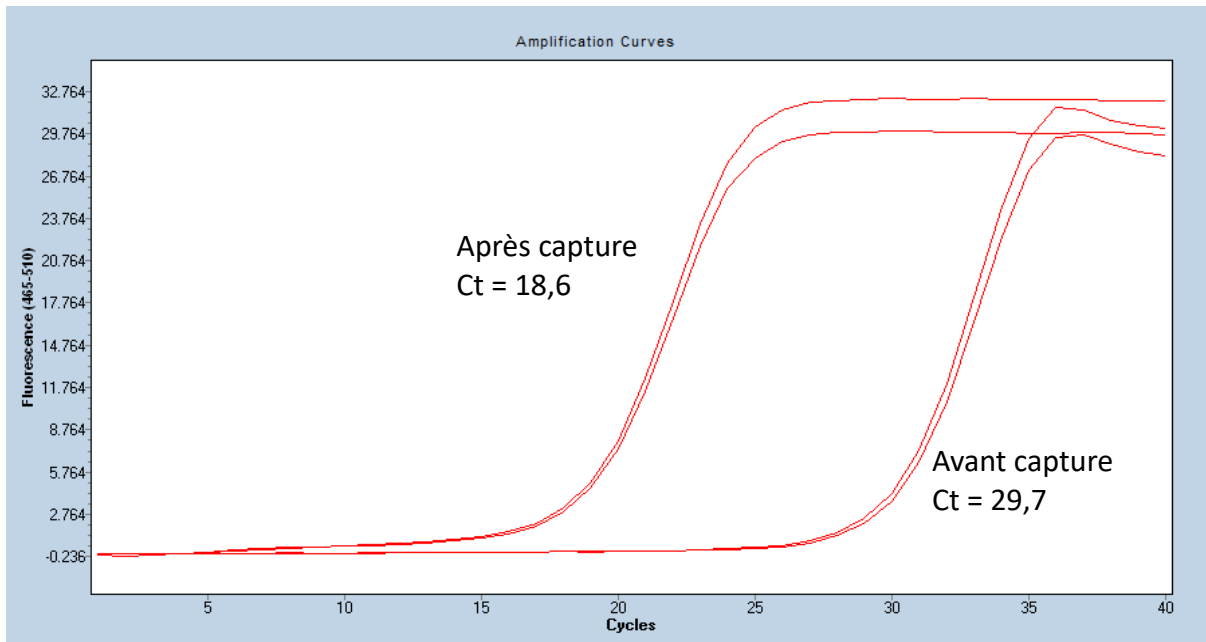


**Figure 16** : Gel d'électrophorèse des acides nucléiques obtenus après synthèse des librairies et avant capture (CAE = Condensat d'Air Exhalé)

L'enrichissement en séquences d'intérêt par capture semi-ciblée en milieu liquide a été réalisé grâce à la puce VirCapSeq-VERT décrite par Lipkin et collègues (126) comme décrit ci-avant.

La qualité de l'enrichissement, mesurée par une qPCR des échantillons en pré- et post-capture d'un contrôle fourni par Roche retrouvait les résultats présentés ici à la **Figure 17.**

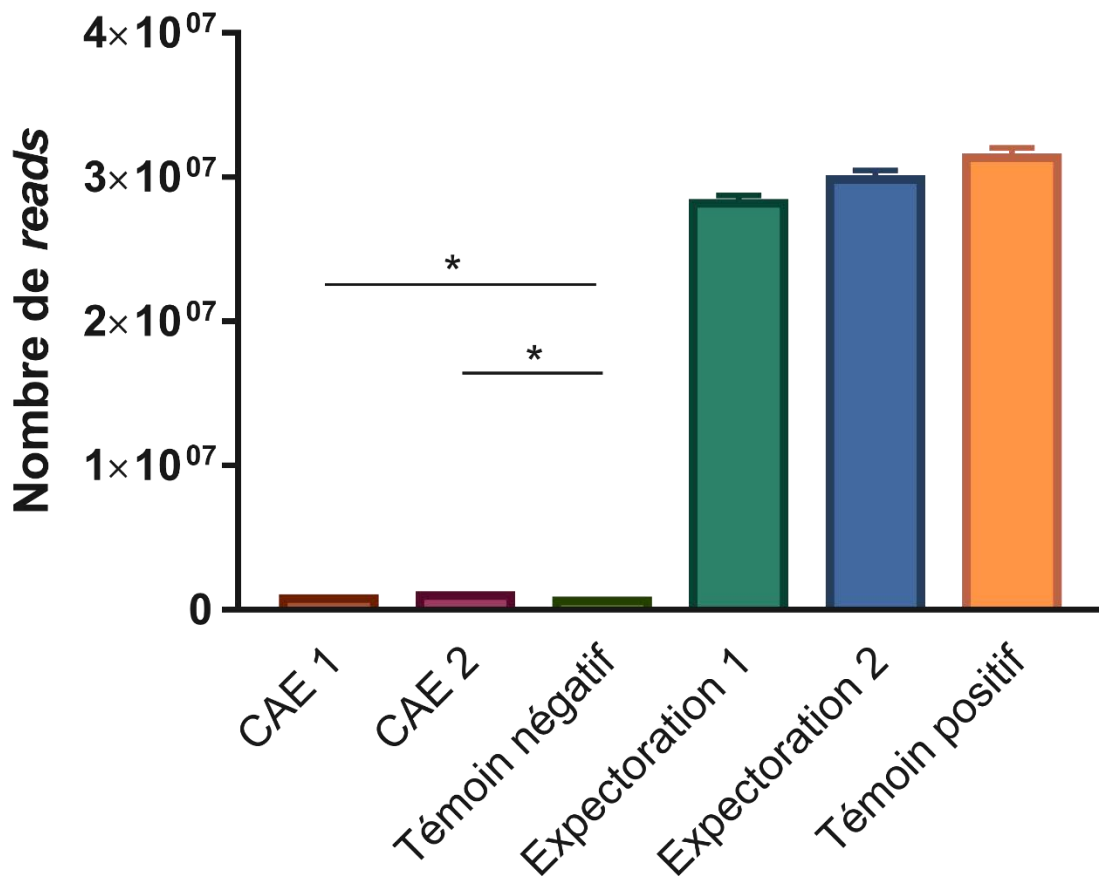




**Figure 17** : Enrichissement relatif en séquences d'intérêt contrôlé par qPCR pré- et post-capture (puce VirCapSeq-VERT) d'un contrôle fourni par Roche.

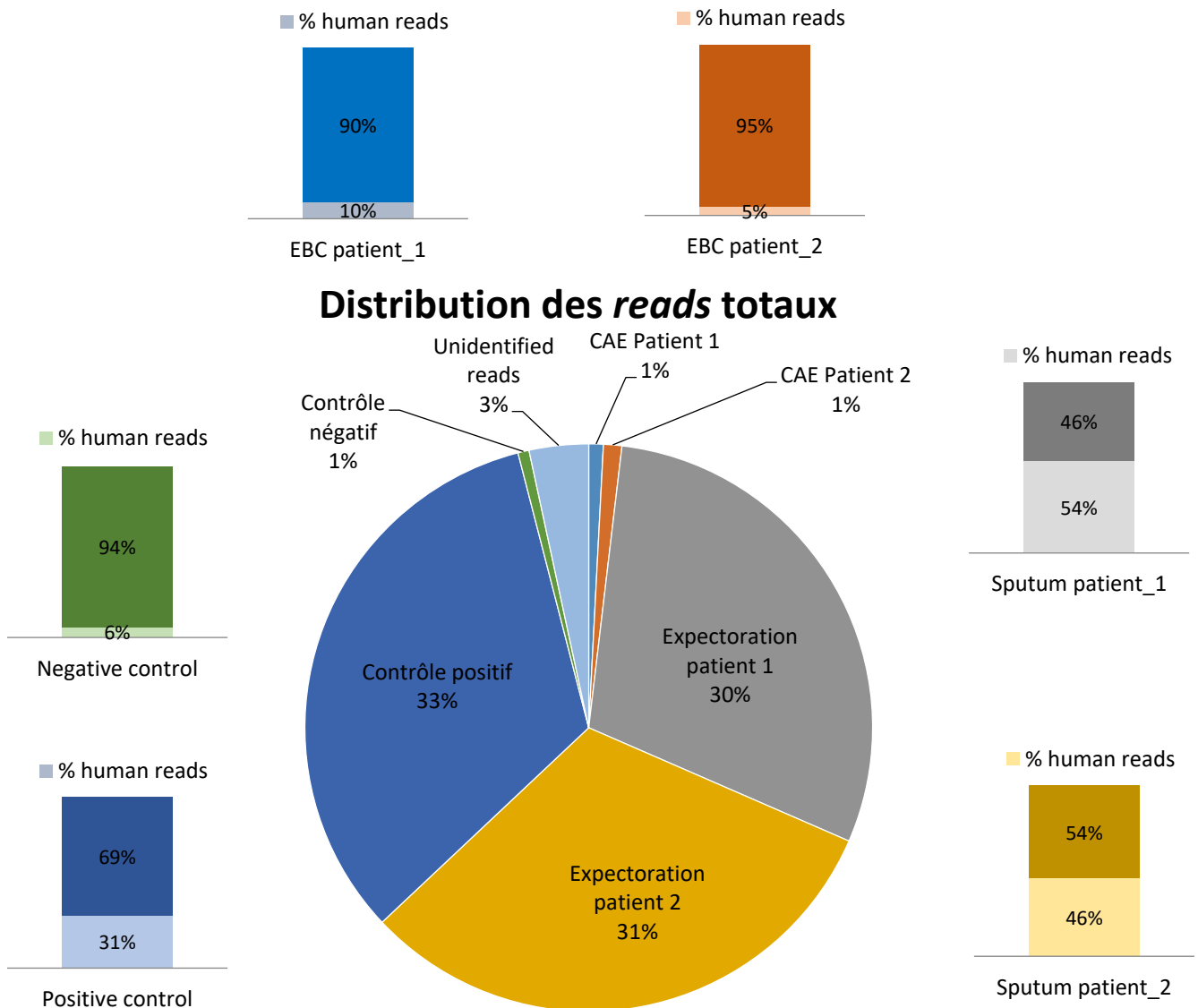
## 1.2. Résultats du séquençage

Le séquençage de ces échantillons a produit un total de 380 millions de séquences (*reads*) dont 86,83 % avaient un score de qualité supérieur ou égal à Q30. La répartition de ces *reads* en fonction des échantillons initiaux est présentée à la **Figure 18**. Le nombre de *reads* obtenus pour les échantillons de condensats d'air exhalé (784462 pour le CAE du patient 1 et 992910 pour le CAE du patient 2) était significativement différent ( $p < 0,001$  dans les deux cas) du nombre de *reads* obtenus pour le témoin négatif (602004).



**Figure 18** : Répartition des *reads* issus du séquenceur pour les 6 échantillons initiaux (CAE = Condensat d'air exhalé).

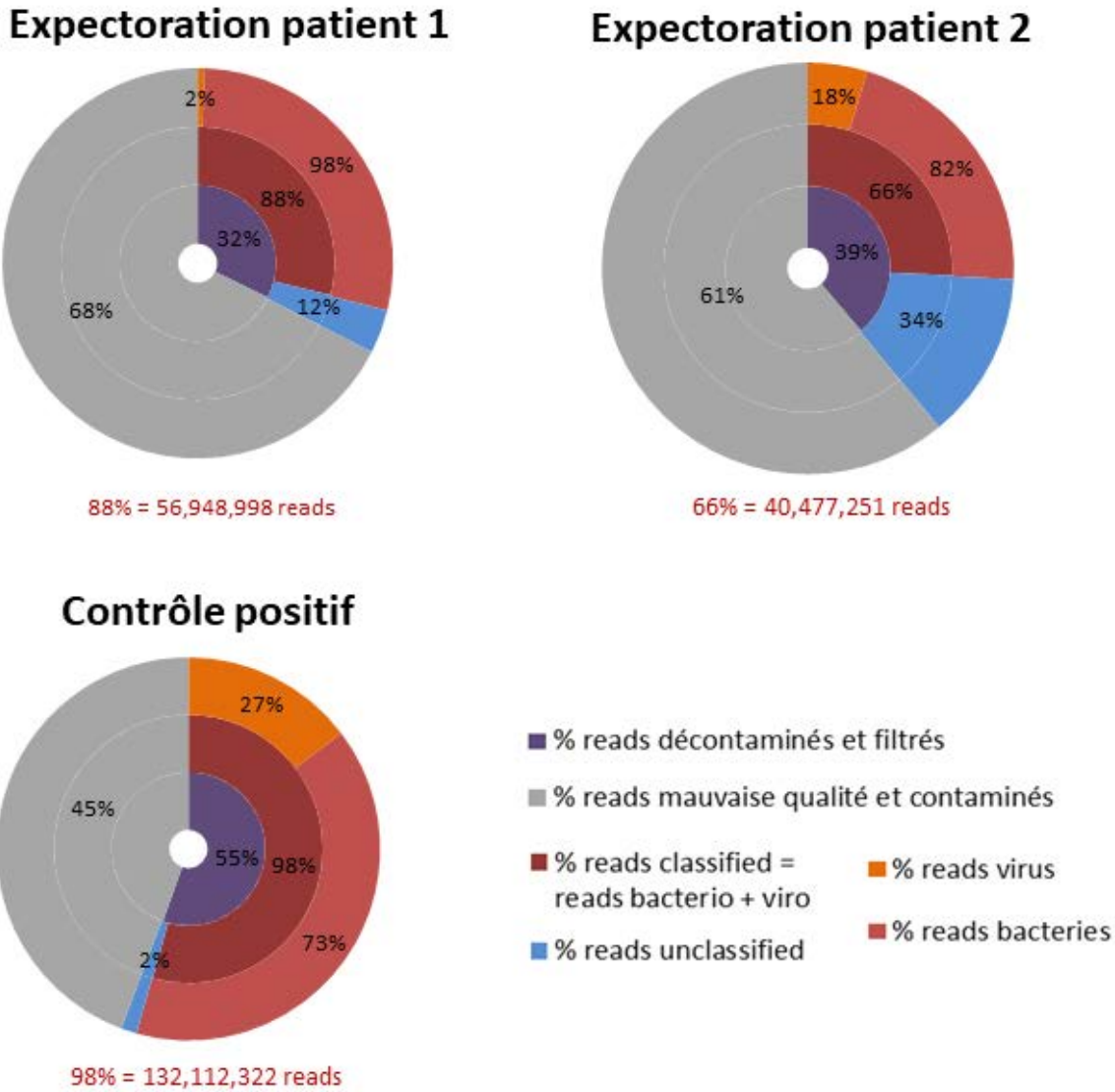
La présence d'acides nucléiques de séquences humaines était importante pour les échantillons riches (témoin positif et expectorations). A l'inverse, les échantillons 'pauvres' en séquences étaient également très peu contaminés par des séquences d'origine humaine (**Figure 19**).



**Figure 19** : Distribution des séquences (humaines et non-humaines) pour les six échantillons testés dans le cadre de la validation de méthode analytique (EBC = *Exhaled Breath Condensate*, condensats d'air exhalé)

L'échantillon témoin négatif ainsi que les échantillons issus de condensats d'air exhalé n'ont pas généré suffisamment de séquences exploitables après soustraction des séquences contaminantes humaines. Les échantillons du témoin positif et des expectorations ont généré

entre 30 et 55 % de *reads* exploitables après élimination des séquences de mauvaise qualité et des séquences correspondantes à du matériel humain. Au total, entre 2 et 27 % de ces séquences étaient des séquences issues de virus et avaient un assignement taxonomique correct. Les répartitions des *reads* de ces trois échantillons sont rapportées à la **Figure 20** :



**Figure 20** : Assignation taxonomique et répartition des séquences issues des trois échantillons 'positifs' (témoin positif et expectorations de patients). En grisé figurent les séquences attribuées à l'hôte humain.

Parallèlement à ces résultats, les explorations traditionnelles menées au laboratoire de bactériologie ont amené aux résultats suivants : l'échantillon d'expectoration du patient 1 mis

en culture était colonisé par *Pseudomonas aeruginosa*, quantifié à hauteur de  $10^6$  UFC / mL. La recherche spécifique de *Burkholderia cepacia* s'est avérée négative. Concernant l'expectoration du patient 2 à l'inverse, la culture a mis en évidence la présence du complexe *Burkholderia cepacia* mais ne mettait pas en évidence la présence de *Pseudomonas aeruginosa*. Les bactéries du complexe *cepacia* étaient retrouvées à une concentration de  $10^6$  à  $10^7$  UCF / mL.

Ces résultats, encourageants mais incomplets, car uniquement basés sur une puce de capture dédiées aux virus ont motivé le *design* d'une puce de capture pan-microbienne, intégrant à la fois les virus mais également les bactéries et les champignons potentiellement présents dans ces échantillons respiratoires. En effet, il était relativement surprenant de retrouver après une étape de capture spécifique aux virus des séquences de bactéries dans des proportions aussi importantes.

## 2. Design de la puce de capture DB-Chip

Les travaux de Lipkin et collègues (126) concernant leur puce de capture VirCapSeq-VERT ont servi de référence pour le *design* d'une nouvelle puce de capture personnalisée. Lipkin et collègues avaient travaillé à partir des séquences virales de la base de données EMBL Coding Domain Sequence Database. Pour permettre une capture des séquences divergentes et variables, les séquences proposées pour la capture étaient filtrées et conservées si elles présentaient plus de 90 % de divergence. La base de données finale utilisée pour le *design* comprenait 1.993.176 séquences de 50 à 100 nucléotides de longueur et dont les  $T_m$  variaient de 58,7°C à 101°C. Un contrôle *in silico* a été réalisé par l'équipe, montrant une couverture de 88 à 98 % selon les régions considérées. Cette même analyse montrait que seuls les virus infectant des vertébrés étaient susceptibles d'être capturés. Les séquences de bactériophages

et de virus de plante ou fongique n'étaient pas représentées dans la base de données et donc, en théorie, non capturées sur cette puce.

Comme pour la puce de Lipkin et collègues, la synthèse de notre puce de capture a été confiée à l'équipe de la société NimbleGen. Pour ce faire, il a été nécessaire de fournir une base de données des séquences complètes à intégrer, le découpage des sondes restant à la discrétion de l'industriel pour répondre à des critères physico-chimiques donnés. La construction de cette base de données regroupe les séquences des virus, bactéries et champignons d'intérêt, et nous avons utilisé une approche se voulant la plus exhaustive possible concernant les micro-organismes pouvant infecter l'homme.

Ce travail a été réalisé en collaboration entre le Département de Méthodologie de l'Information de Santé (DMIS) pour le choix et l'analyse des bases de données et les services de virologie, bactériologie et mycologie pour la validation du choix des bases de données à utiliser puis pour le choix des micro-organismes et/ou des gènes sélectionnés en fonction de leur intérêt en diagnostic médical.

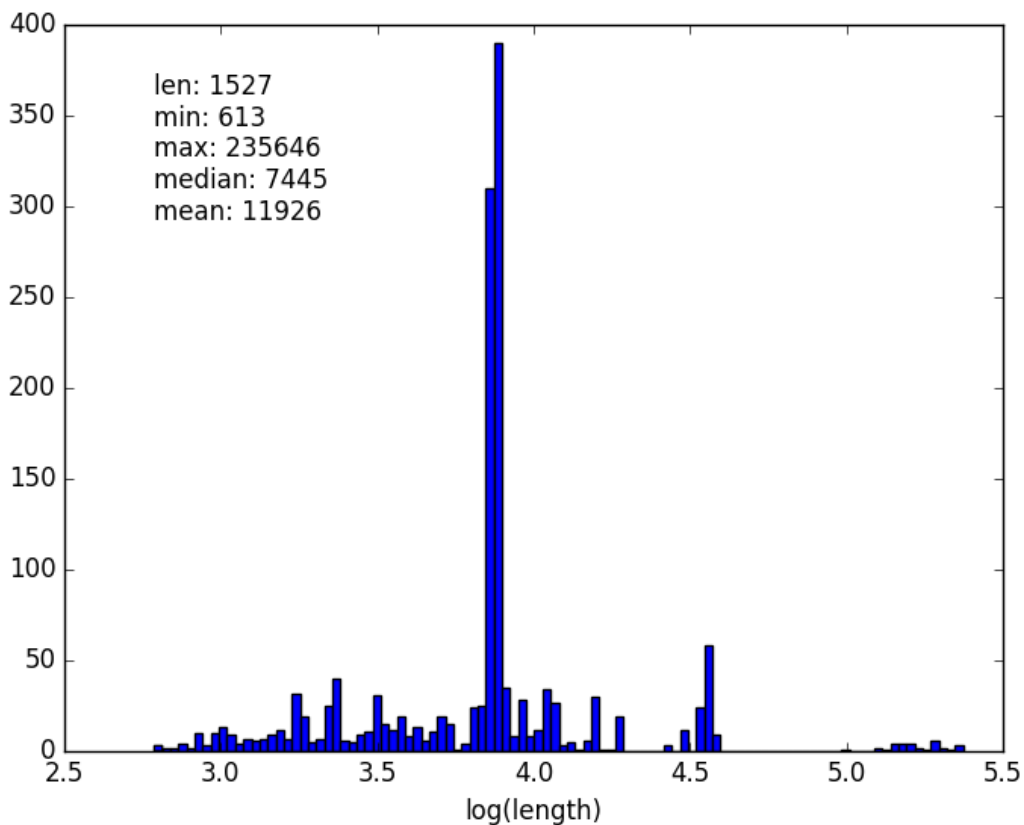
Nous avons pour cela sélectionné les séquences des micro-organismes à partir de bases de données existantes. Pour le règne des virus, nous avons orienté notre choix vers l'utilisation de génomes complets de virus. Ceux-ci ont été récupérés à partir de la base de données NCBI Nucleotide puis filtrés selon le critère de l'hôte potentiel (avec hôte = *homo sapiens*).

Concernant les génomes bactériens, nous avons souhaité dans un premier temps travailler à partir de la base de données existante PATRIC (143). L'utilisation de génomes complets là encore était notre premier choix. Enfin, pour les champignons, nous avons choisi de travailler sur des séquences d'ITS et de 28S.

Ainsi, une première version de la base de données de capture a consisté en la collection des séquences issues de la base américaine de NCBI RefSeq (185.738.687 références au 4 janvier 2019, dont 30.283.153 séquences nucléotidiques correspondant à 86.867 organismes) à laquelle nous avons appliqué un filtre par hôte. En ne sélectionnant que les séquences correspondant à des virus, nous avons obtenu une base de données constituée de 11.617 génomes complets uniques. Cette approche générant un nombre trop important d'entrées, s'est également avérée incomplète car seuls 70 % des séquences possédaient un renseignement quant à l'organisme hôte.

Par la suite, nous avons travaillé avec la base de Virus-Host DB (144) pour relier l'espèce taxonomique à l'hôte (humain, en ce qui nous concerne). Cette base de données recense les génomes complets issus de la base générale NCBI Nucleotide auxquels elle associe les informations concernant l'organisme hôte de la réplication virale. Cette dernière information est extraite des bases de données RefSeq, GenBank, UniProt et ViralZone, ou était manuellement insérée à partir de publications isolées. Cette méthodologie est à l'origine d'une grande qualité des données issues de cette base Virus-Host DB.

Au final, nous avons ainsi obtenu une base de données exploitable pour le design de la puce de capture en ce qui concerne les génomes viraux. Celle-ci était composée de 1527 génomes d'une longueur comprise entre 613 bases à 235.646 bases. Cette base de données des séquences de virus dont l'hôte est l'homme est répartie comme suit (**Figure 21**).

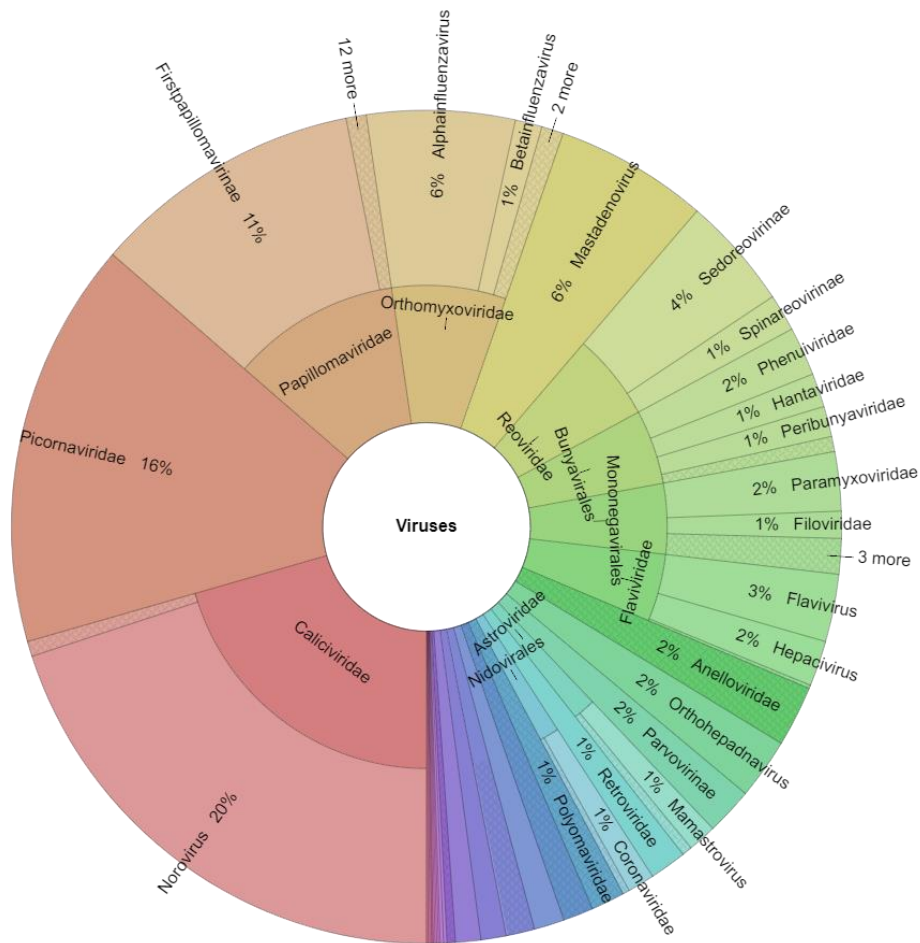


**Figure 21** : Répartition des séquences des génomes viraux après sélection des virus à tropisme humain uniquement en fonction de la longueur de séquence. Séquences de génomes complets issues de la base de données Virus-Host DB.

Les génomes et séquences nucléotidiques choisis pour constituer la base de données pour la production d'une puce de capture peuvent être visualisés selon leur taxon d'appartenance. Cette visualisation circulaire est présentée à la **Figure 22** ci-dessous pour l'ensemble des virus eucaryotes.

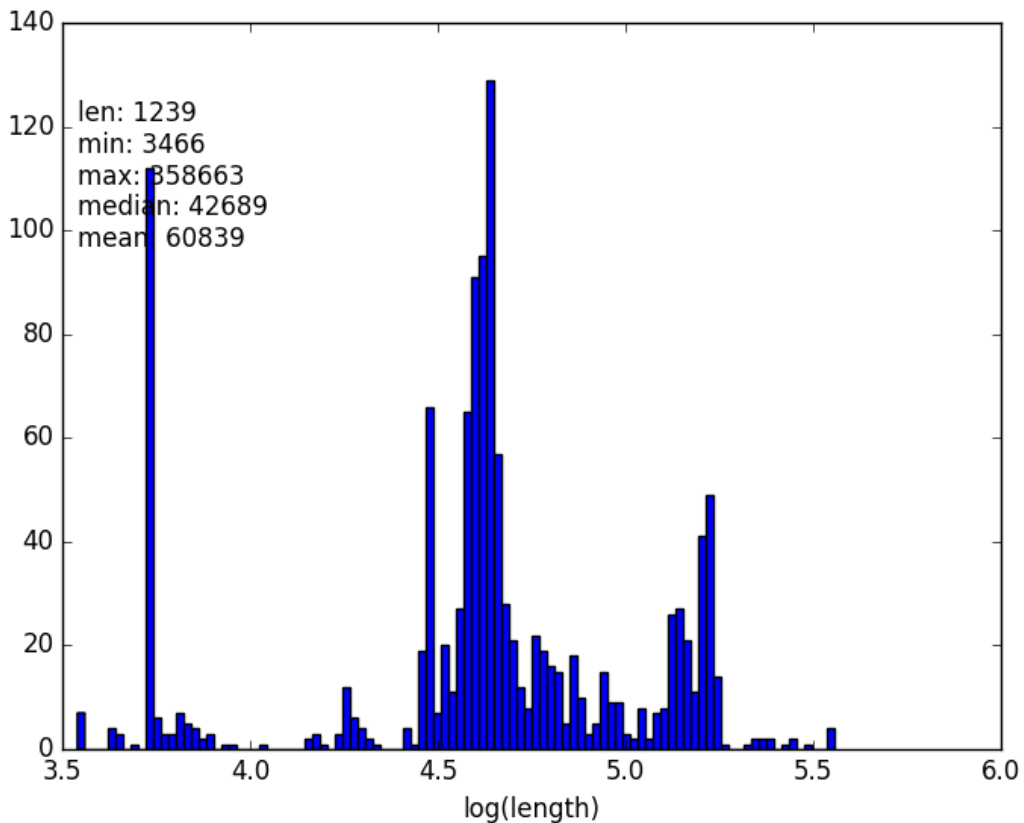
De la même manière, nous avons sélectionné les génomes étiquetés comme provenant de bactériophages. Dans ce cas, l'organisme hôte renseigné n'était pas l'humain, mais il s'agissait d'un hôte bactérien avec un hôte final humain. Cette base de données présentait une répartition des séquences très différente (**Figure 23**).





**Figure 22** : Visualisation de la base de données des virus eucaryotes (hôte humain). Représentation réalisée avec l’outil Krona

En ce qui concerne le règne bactérien, en collaboration avec le service de bactériologie, la base de données initialement choisie a été PATRIC (143). Cette base de données recense les génomes bactériens publiés sous différents supports et ses auteurs procèdent à l’annotation des génomes, ce qui en fait une ressource extensive de très bonne qualité. En 2019, PATRIC disposait de plus de 13600 génomes bactériens complets (et plus de 216.000 génomes partiels). Notre approche utilisant des génomes bactériens complets, elle a été à l’origine d’une masse de données extrêmement conséquente (> 54000 Mb) et dépassant très largement les capacités de la puce de capture que nous souhaitions préparer (200 Mb).



**Figure 23** : Répartition des séquences de la base de données de capture pour les phages en fonction de la longueur de celles-ci.

Aussi, une seconde approche a été nécessaire. Pour cette nouvelle approche, nous avons sélectionné un nombre restreint de gènes bactériens, largement présents chez les différentes espèces et permettant ainsi d'identifier un maximum de bactéries. La liste des gènes sélectionnés est présentée ci-dessous (**Tableau 5**). Cette sélection manuelle des gènes bactériens d'intérêt a été réalisée au sein du service de bactériologie. Elle reposait sur des critères d'exhaustivité, de pertinence et de recoupement entre les différentes bases de données explorées. Ainsi, le gène 16S a été conservé pour son exhaustivité. Des gènes tels que *gyrB* et *HSP65* spécifiques des mycobactéries du complexe *tuberculosis* et des mycobactéries hors complexe *tuberculosis* respectivement ont permis d'accroître la spécificité par rapport au

gène 16S seul. Les autres gènes sélectionnés étaient principalement des gènes de ménage connus et déjà utilisés pour le système de typage MLST.

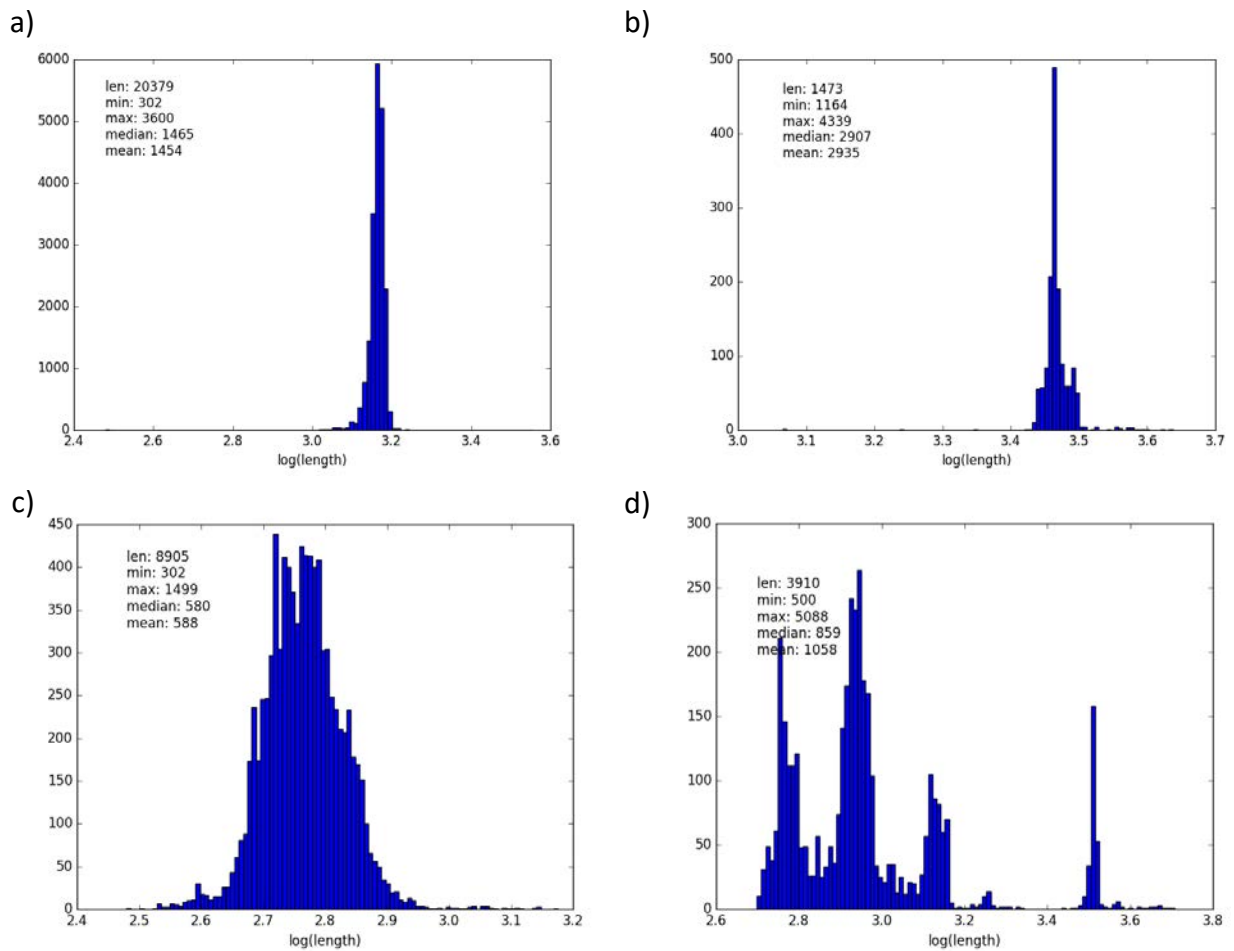
**Tableau 5** : Liste des gènes bactériens sélectionnés pour la création de la puce de capture.

Gènes bactériens		
16S		
23S	glyA	abcZ
adk	glnA	groES
atpA	gltA	parE
gdh	groEL	parC
ftsZ	yqiL	rpoD
dnaK	purA	atpD
leuS	tkt	purA

Enfin, pour la constitution d'une base de données regroupant les champignons, l'utilisation de génomes complets a également été impossible compte tenu de l'importance de ces génomes et de la limitation dans le nombre de sondes disponibles sur la puce de capture. Pour pallier à cette contrainte, ont été sélectionnés, de la même façon que pour le règne bactérien, un nombre limité de gènes partagés par l'ensemble des espèces concernées. Nous avons donc utilisé le gène ribosomique 28S et la région ITS du projet NCBI ITS RefSeq Targeted Loci pour construire la base de données à l'origine du *design* de la puce de capture.

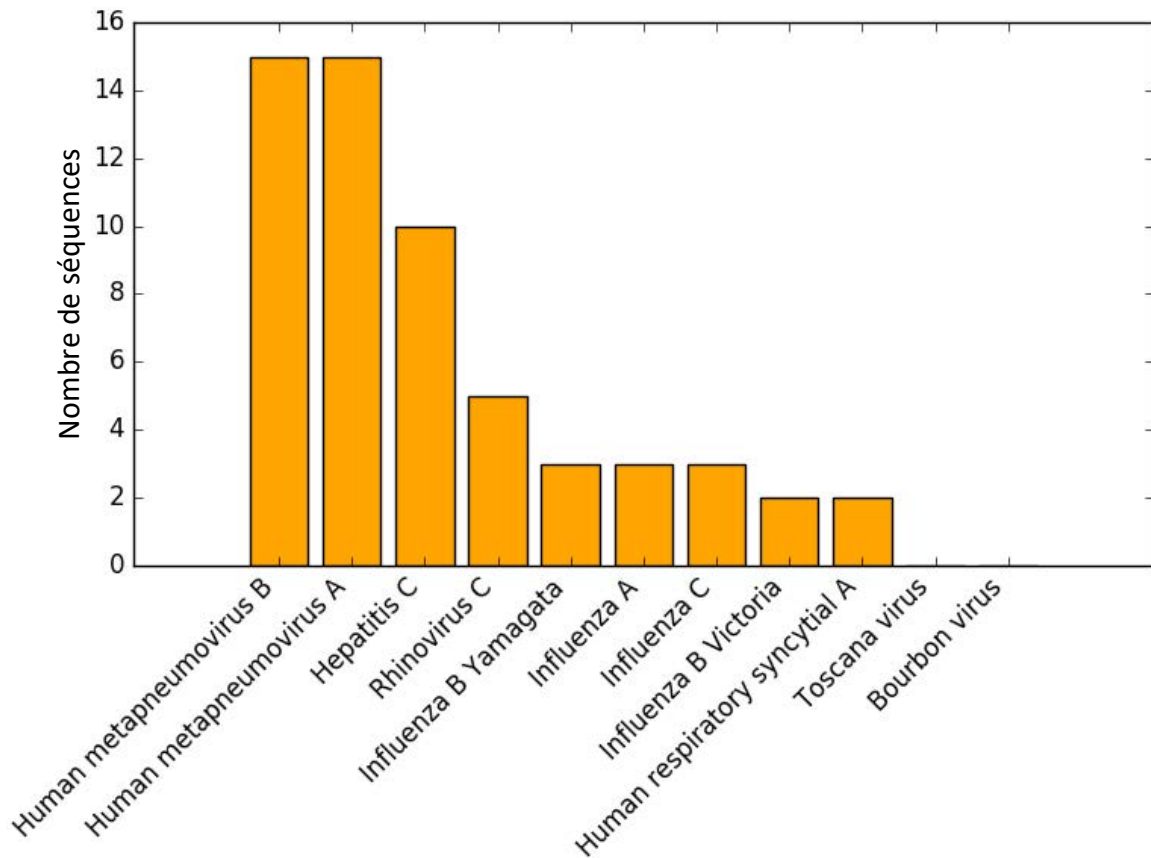
Ainsi collectées, les séquences bactériennes et fongiques étaient distribuées comme indiqué dans la **Figure 24** ci-dessous.

Une fois les séquences d'intérêt sélectionnées, il a été réalisé par le DMIS une fragmentation et un *clustering* de la base de données. Cette étape a permis de réduire le nombre de séquences dans la base de données en éliminant l'information redondante. Pour la partie virale de celle-ci : nous avons 102 millions de bases initialement, puis après *clustering*, un peu moins de 74 millions de bases.



**Figure 24 :** Répartition des séquences a) bactériennes du gène 16S, b) bactériennes du gène 23S, c) fongiques pour l'élément ITS, d) fongiques du gène 28S.

Enfin, une validation *in silico* de la base de données *post-clustering* a été réalisée par le DMIS en alignant cette dernière sur des séquences de référence choisies manuellement et absentes de la base de départ. Les résultats de ces tests *in silico* pour les virus sont présentés à la **Figure 25**.



**Figure 25** : Résultat du test *in-silico* pour les virus : *mapping* de la base de données *post-clustering* sur des séquences de références absentes de la base de départ. En ordonnées, nombre de sondes (fragments de séquences) de la base « captant » (s’alignant) avec la séquence des virus correspondants.

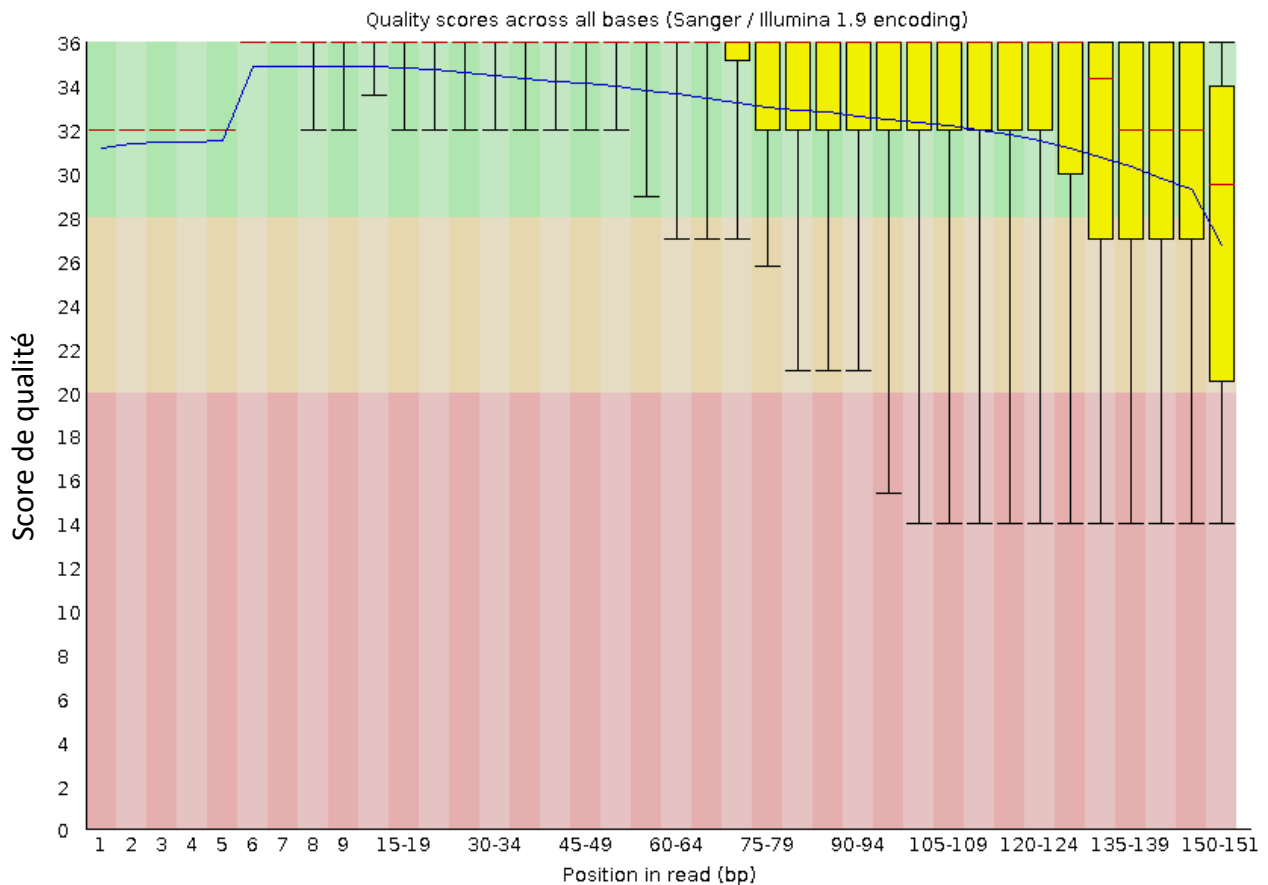
Ainsi, une séquence de métapneumovirus humain de type B par exemple, avait au moins 90 % d’identité sur au moins 70 bases avec 15 fragments de séquences de la base de capture, et ce, bien que cette séquence particulière n’apparût pas dans la base de données initiale. De rares pathogènes (exemple ici du virus Toscana et du virus Bourbon) n’étaient pas retrouvés par les tests *in silico* ce qui était en cohérence avec leur absence de la base de données initiale. La base de données a été par la suite envoyée à NimbleGen pour synthèse de la puce de capture, dont nous avons eu à valider le design avant mise en production. Le design que NimbleGen a fourni a couvert plus de 99 % des bases cibles de notre base de données. Les

espèces faiblement couvertes n'étant pas pertinentes pour ce premier design et son application (par ex. les virus humains betaherpes 6A, herpes 7 ou le rétrovirus endogène humain K113), la proposition de l'industriel a été validée et mise en production.

Nous avons donc obtenu une puce de capture capable de fixer les séquences des virus, des bactéries et des champignons, avec une exhaustivité très importante.

### 3. Validation de la puce de capture DB-Chip

La validation de la méthode analytique utilisant la puce de capture DB-Chip créée par le laboratoire a été réalisée en Octobre 2018. Nous avons réalisé un *run* expérimental avec 5 échantillons artificiels et leurs témoins négatifs respectifs, en utilisant le protocole tel que décrit ci-avant.

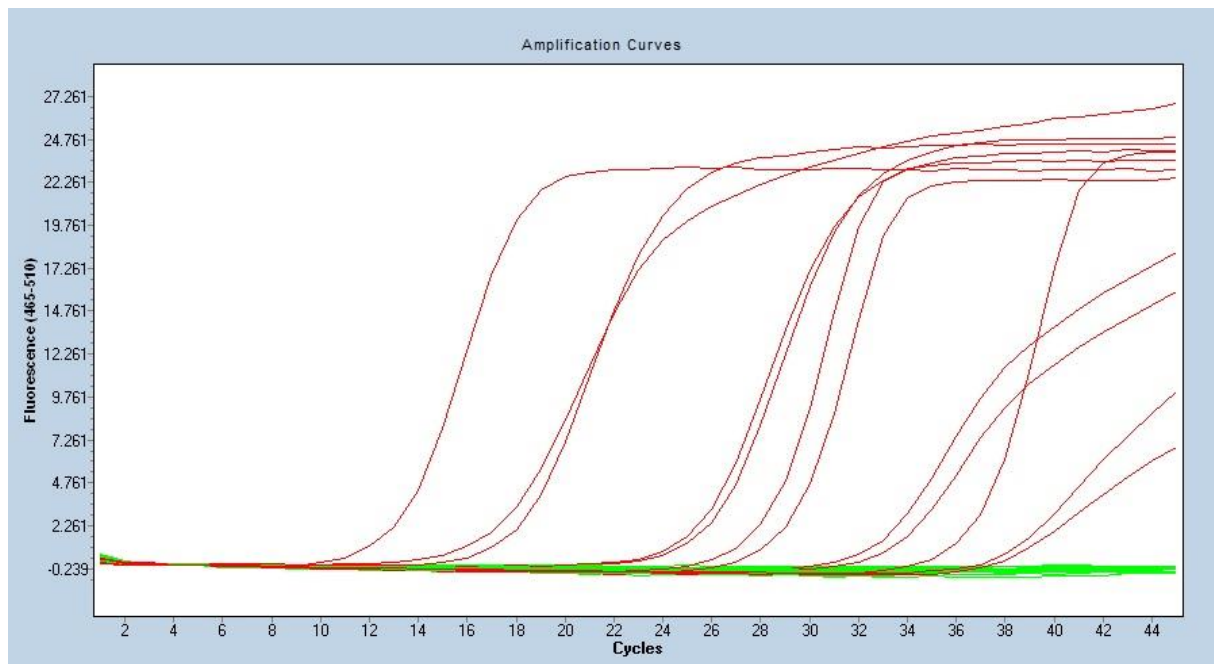


**Figure 26** : Score de qualité obtenu au séquençage. Il s'agit ici du résultat obtenu pour une piste de séquençage, avec le panel de virologie seul, sur notre puce de capture DB-Chip.

Le séquençage de ces échantillons a généré un total de 164,45 millions de séquences, dont 83,39 % avaient un score de qualité supérieur ou égal à Q30. Les séquences obtenues étaient globalement de bonne qualité sur toute la longueur des *reads* (exemple de résultats obtenus pour une piste de séquençage pour le panel virologique seul avec capture sur notre puce DB-Chip : **Figure 26**).

Concernant l'enrichissement par l'étape d'hybridation sur sondes, la **Figure 27** ci-après montre l'enrichissement des différentes séquences en post-capture pour trois cibles bactériennes données (*S. pneumoniae* [lytA], *P. aeruginosa* [gyrB], *K. pneumoniae* [gltA]) avec

la puce de capture DB-Chip. Pour le *S. pneumoniae*, le *cycle threshold* (Ct) initial pré-capture était à 26,55.

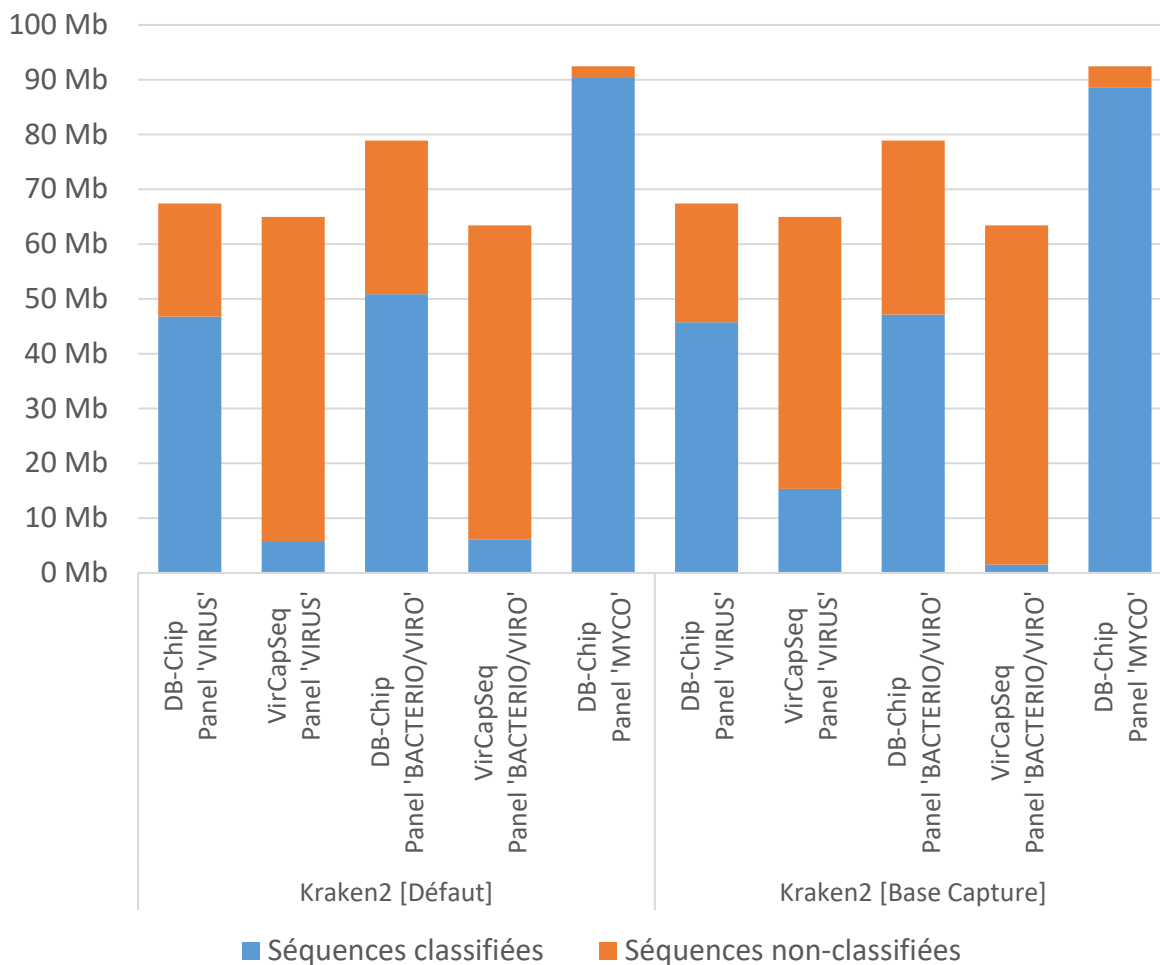


**Figure 27** : Enrichissement relatif en séquences d'intérêt contrôlé par qPCR pré- et post-capture (puce DB-Chip) de différentes cibles bactériennes (*S. pneumoniae* [lytA], *P. aeruginosa* [gyrB], *K. pneumoniae* [gltA])

Après capture, nous retrouvons une valeur de Ct à 18,07. D'une manière similaire, pour *P. aeruginosa*, le Ct mesuré passait de 33,68 à 25,46 et pour de 26,99 à 18,87 pour *K. pneumoniae*, témoins de l'enrichissement de ces gènes suite à la capture. Il est intéressant de noter que l'enrichissement sur la puce de capture et le contrôle de cet enrichissement par qPCR n'a pas été réalisé en employant les mêmes gènes cibles. Ainsi, s'il est difficile d'extrapoler les données obtenues par ces méthodes quantitatives, il est intéressant de noter l'enrichissement en séquences non initialement présentes dans la base de capture.

La proportion des *reads* obtenus et classifiés au plan taxonomique est résumée à la **Figure 28** ci-dessous.

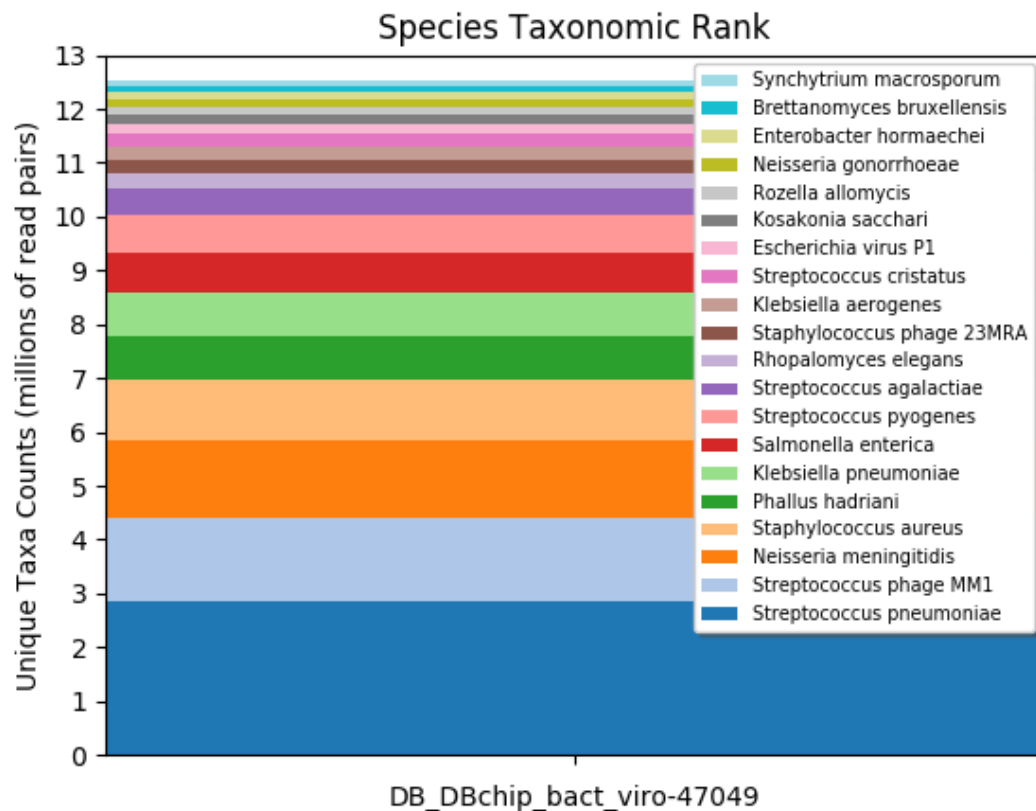




**Figure 28** : Classification taxonomique des *reads* obtenus en fonction du panel considéré et de l'algorithme de classement mis en œuvre.

Ainsi, pour le panel contenant des virus exclusivement, et capturés à l'aide de notre puce de capture DB-Chip, un peu moins de 70 % (69,37 %) des *reads* obtenus ont pu être classifiés et rattachés à une espèce donnée en utilisant la base de données par défaut de Kraken, et une proportion sensiblement identique (67,82 %) l'était en utilisant la base de *design* de la puce de capture comme référence. Les séquences non classifiées, soit environ 30 % des séquences dans ce premier cas, n'étaient pas suffisamment proches des séquences présentes dans la base de données. Ces séquences non classifiées n'ont, dans un premier temps pas été analysées.

Les espèces taxonomiques (OTU) retrouvées pour ces échantillons étaient très variables. Un exemple de résultat est donné dans la **Figure 29** avec les espèces retrouvées au sein du panel contenant un mélange de virus et de bactéries et enrichi avec la puce de capture DB-Chip.



**Figure 29** : Assignment taxonomique à l'espèce des séquences obtenues pour l'échantillon contenant un mélange de bactéries (99,5 %) et de virus (0,5 %) et enrichi par la puce de capture DB-Chip.

L'espèce bactérienne retrouvée avec le plus grand nombre de *reads* était le *Streptococcus pneumoniae*, suivie de *Neisseria meningitidis*. Ces deux pathogènes étaient les plus représentés dans le panel testé (*Streptococcus pneumoniae* : 28,9 % ; *Neisseria meningitidis* : 28,9 %). Viennent ensuite les *Staphylococcus aureus* (15,1 %) et *Klebsiella pneumoniae* (14,4 %) retrouvés en troisième et quatrième position respectivement en nombre de *reads* bactériens. Les 6<sup>ème</sup> et 7<sup>ème</sup> espèces retrouvées par la suite sont le *Streptococcus pyogenes*

(7,2 % dans le panel) et le *Streptococcus agalactiae* (2,9 %). Enfin, les espèces bactériennes présentes à moins de 1 % dans le panel sont retrouvées dans l'analyse réalisée mais avec un nombre de *reads* très faible, similaire à ceux d'espèces non présentes dans le panel. Il pourrait s'agir ici de contaminations ou d'un manque de spécificité.

Au plan virologique, et dans le panel mélangé contenant 99,5 % de bactéries et 0,5 % de virus, les principaux virus retrouvés par cette technique étaient les suivants : *Human betaherpesvirus 5* (CMV), *Human mastadenovirus C*, *Human alphaherpesvirus 3* (VZV), *Human alphaherpesvirus 2* (HSV-2), *Human gammaherpesvirus 4* (EBV), *Human alphaherpesvirus 1* (HSV-1), *Betapapillomavirus 2*, *Primate bocaparvovirus 2*, *Human mastadenovirus F* et *Papiine alphaherpesvirus 2*. Ces résultats, obtenus avec la puce de capture DB-Chip étaient comparables aux résultats obtenus à l'aide de la puce de capture VirCapSeq-VERT. L'ensemble des *herpesvirus* étaient présents dans les échantillons artificiels. Cependant, d'autres virus bien présents initialement n'ont pas été retrouvés par cette analyse.

Les virus retrouvés dans l'échantillon viral pur sont également représentatifs des espèces initialement présentes dans le panel testé. Les dix premières espèces retrouvées en nombre de *reads* étaient : *Parechovirus A*, *Human betaherpesvirus 5*, *Human mastadenovirus C*, *Human alphaherpesvirus 3*, *Human alphaherpesvirus 2*, *Human gammaherpesvirus 4*, *Enterovirus B*, *Rhinovirus A*, *Human alphaherpesvirus 1* et le *Sapporo virus*, toutes présentes dans le panel. Pour les autres 13 autres espèces que contenait le panel de test, celles-ci sont également retrouvées à l'analyse, mais au prix d'une diminution de la spécificité de la méthode avec l'identification d'espèces non présentes (*off-target*).

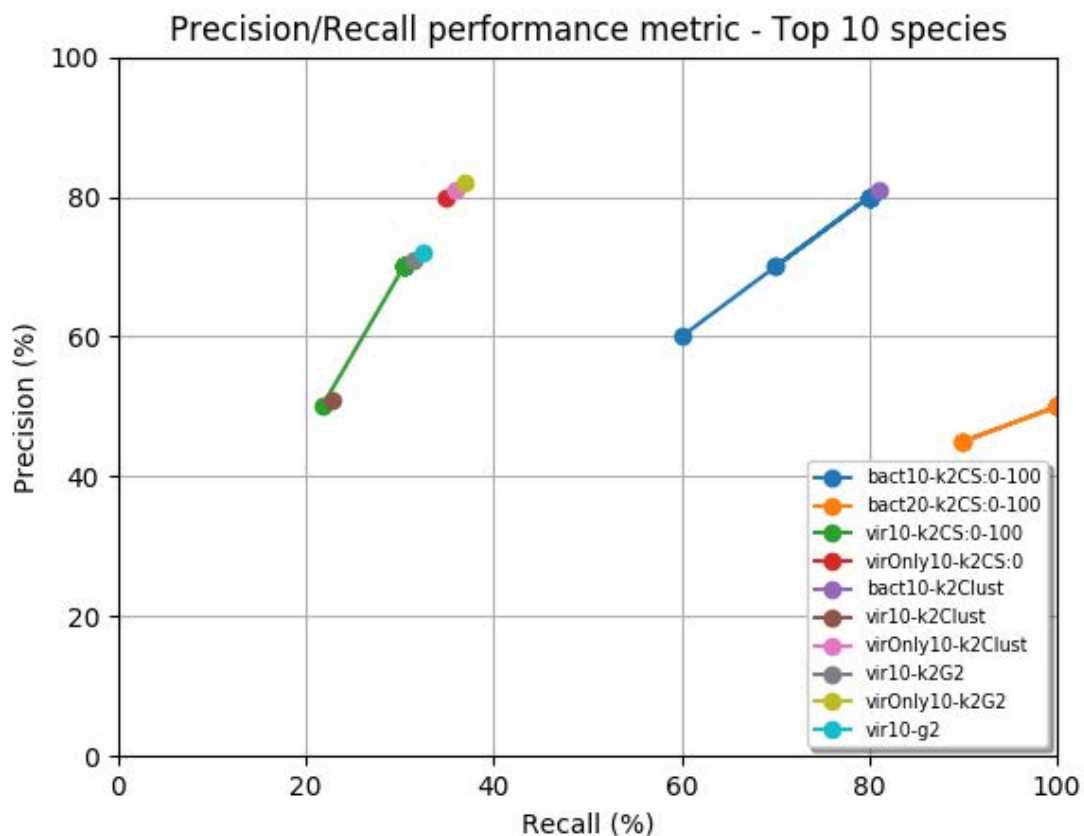
Enfin, concernant la mycologie, *Pneumocystis jirovecii* n'était pas retrouvé par notre analyse cependant que ce pathogène était bien représenté au sein de l'échantillon initial. L'absence

de séquences de *Pneumocystis jiroveci* mais la présence de celles de *Pneumocystis murina* dans la base de données du *design* de la puce expliquerait ce résultat. *Aspergillus fumigatus* était quant à lui correctement détecté avec l'identification jusqu'au rang d'espèce ou de genre en fonction de la base de données utilisée (meilleure performance avec la base de Kraken par défaut). Le faible nombre de séquences (2) d'*Aspergillus fumigatus* dans la base du *design* pourrait expliquer ce manque de performance de notre base de données pour l'identification fine.

La validation de la puce de capture et de la méthode analytique impliquait de retrouver suite au séquençage et à l'assignation taxonomique des séquences, l'ensemble des micro-organismes initialement présents dans les échantillons, et uniquement ceux-ci. Deux paramètres nous ont permis d'estimer la pertinence de l'algorithme utilisé et de monitorer la performance de notre méthode : il s'agit de la précision et du rappel (*recall*). La précision est définie par le nombre d'espèces pertinentes rapportées au nombre d'espèces sélectionnées. Le rappel (*recall*) est un paramètre de performance calculé comme suit : il s'agit d'espèces pertinentes sélectionnées divisé par le nombre d'espèces total. Plusieurs algorithmes bio-informatiques ont ainsi pu être testés et les résultats sont présentés à la **Figure 30** (travail encore en cours).

Ces deux paramètres ont été calculés à partir des dix espèces les plus retrouvées par l'analyse bio-informatique. Dans notre cas, le panel de virus comprenait 23 pathogènes. En ne retenant que 10 de ces 23 pathogènes, nous ne pouvions espérer qu'un maximum de  $10 / 23 = 43,5 \%$  concernant le rappel (*recall*). La précision et le rappel renseignent ainsi sur la capacité de la méthode à retrouver les espèces initialement présentes dans l'échantillon, et ce, tout en

minimisant le nombre de faux positifs (issus de la contamination en particulier, mais également générés lors de l'analyse bio-informatique).



**Figure 30** : Métrique de performance finale : précision et rappel (*recall*) obtenus selon l'algorithme ou la combinaison d'algorithmes mis en œuvre en ne retenant que les 10 premières espèces détectées (travail en cours).

Certaines espèces virales étaient retrouvées lors de l'analyse des séquences obtenues, et ce, malgré l'absence des pathogènes correspondants dans les échantillons artificiels utilisés. Ces espèces sont listées dans le **Tableau 6**. La présence de ces espèces virales et bactériennes ne correspond pas aux pathogènes initialement introduits dans nos échantillons tests.

**Tableau 6** : Liste des espèces virales et bactériennes potentiellement contaminantes retrouvées lors de notre analyse avec la puce de capture DB-Chip

Panel 'Virus'	Panel 'Bactéries / Virus'
<i>Primate bocaparvovirus 2</i>	<i>Cutibacterium acnes</i>
<i>Human immunodeficiency virus 1</i>	<i>Neisseria gonorrhoeae</i>
<i>Papillomaviridae (Betapapillomavirus 2 ou Alphapapillomavirus 4)</i>	
<i>Papiine alphaherpesvirus 2</i>	
<i>Ungulate erythroparvovirus 1</i>	

## PARTIE IV – DISCUSSION

Le travail présenté ici a permis d'initier la mise au point d'une approche de métagénomique infectieuse semi-ciblée basée sur la capture d'acides nucléiques et le séquençage de nouvelle génération. Ce travail préliminaire visait à démontrer la faisabilité d'une telle méthode de détection pan-microbienne puis à caractériser les performances de celle-ci en vue de son utilisation sur des échantillons de patients dans le cadre d'une étude clinique.

Au cours d'une première approche analytique nous avons voulu montrer, sur un petit nombre d'échantillons la faisabilité de cette méthode. Nous sommes partis, pour cela, d'une puce de capture commercialisée et ciblant les virus infectant les cellules des vertébrés. Ainsi, six échantillons ont permis d'établir cette preuve de concept, démontrant la pertinence de cette méthodologie et soulignant les processus techniques clés à maîtriser. Cette première expérience nous a permis de constater que la quantité d'ADN de travail disponible au départ de l'analyse était une donnée importante. La faible quantité d'ADN collectée par le recueil de condensats d'air exhalé (en deçà de la limite de quantification avec la technique QuBit utilisée telle que cité plus haut) montre la difficulté à travailler sur ce type de prélèvements. Bien que le nombre de *reads* obtenus pour ces échantillons après séquençage soit significativement plus élevé que celui obtenu pour le témoin négatif, cela s'est révélé insuffisant pour poursuivre l'analyse de ce type de prélèvements. En revanche, et comme attendu, la contamination par des séquences d'origine humaine était plus faible que celle observée dans les expectorations (8 % vs 50 %). Au total, le nombre de *reads* finalement exploitables pour rechercher des micro-organismes dans les CAE était 16 fois inférieur à celui des expectorations. L'utilisation des condensats d'air exhalé avait été pourtant proposée et utilisée pour du diagnostic microbiologique par d'autres équipes (135,145). L'équipe de

Zakharkina et collègues retrouvait de l'ADN bactérien dans 90 % des CAE testés et de l'ARN dans 72 % des CAE testés. La quantité d'ARN était similaire entre les CAE et les expectorations appariées, mais la quantité d'ADN était supérieure dans les expectorations par rapport aux CAE correspondants (145). De leur côté, l'équipe de Glendinning et collaborateurs avait quantifié les ARN ribosomiques 16S en qPCR et séquencé ces extraits en utilisant la technologie de Illumina. Ils décrivent une quantité d'acides nucléiques plus faibles dans les CAE par rapport à des prélèvements distaux protégés (PDP) mais avec un seul ordre de grandeur de différence. Bien sûr, les quantités retrouvées étaient très faibles ( $7.10^{-8}$  ng/ $\mu$ L).

La spécificité de la puce de capture VirCapSeq-VERT pose également question. Nous avons recouvert une quantité importante d'ADN bactérien lors de nos essais alors que cette puce est dédiée à la capture des acides nucléiques issus de virus uniquement. Ceci n'est pas rapporté par l'équipe de Lipkin et collègues (126) qui ont conçu cette puce de capture, ni par d'autres équipes qui ont publié des résultats en utilisant cette technologie (146,147). Pour pallier à ce défaut nous avons augmenté la stringence de l'étape de capture. Pour cela nous avons modifié les étapes de lavages intermédiaires, en augmentant la température de réaction à 47°C. L'ajout d'étapes de vortex dans le protocole a également permis d'augmenter la spécificité et de réduire la capture de séquences indésirables (*off-target*). Cependant, les bactéries identifiées correspondaient aux bactéries présentes dans les échantillons et retrouvées par les techniques classiques de bactériologie (culture) pour ces deux patients. Ces résultats nous ont donc confortés dans notre approche pan-microbienne et dans notre souhait de développer une puce de capture regroupant les trois règnes de la microbiologie.

Enfin, nous avons montré que l'application d'une technique de capture par hybridation sur sondes en milieu liquide était pertinente. Dans un mélange avec une prédominance très



importante de bactéries par rapport aux virus, nous démontrons la détection et l'identification des virus. Cela est rassurant quant aux capacités futures d'une telle technique qui pourra être appliquée à un grand nombre d'échantillons, quelle que soit la composition relative de ces derniers en bactéries, virus ou champignons.

Dans un second temps, nous avons donc réalisé le *design* d'une puce de capture pour hybridation en milieu liquide de sondes oligo-nucléotidiques afin de permettre un enrichissement sélectif des génomes bactériens, viraux et parasitaires. Cette étape de conception a nécessité la sélection des gènes et génomes d'intérêt et la constitution d'une base de données restreinte mais exhaustive. Ce travail basé sur les bases de données disponibles en montre les limites dans les applications cliniques qui nous intéressent : certaines bases sont insuffisamment complétées (présence et diversité des annotations), ou simplement truffées d'erreurs (bases de données constituées de manière automatique sans validation des données agrégées). La quantité de sondes disponibles sur une puce de capture étant limitée, nous avons choisi de travailler sur des génomes viraux complets mais nous avons dû sélectionner une quantité de gènes restreints pour les règnes bactérien, parasite et fongique. Concernant les bactéries, les gènes choisis correspondaient à des gènes constitutifs retrouvés chez presque toutes les espèces (ARN 16S et 23S ribosomiques par exemple) ainsi qu'à d'autres gènes de ménage retrouvés chez la majorité des bactéries également mais permettant un typage plus fin lorsque mis en évidence (*gltA*, *groES*, ...). Enfin, pour les champignons, seuls deux gènes génériques ont été choisis : ce sont les gènes 28S et ITS. Ces deux gènes suffisent habituellement à génotyper une large majorité des pathogènes mais sont probablement insuffisants pour bénéficier d'un typage fin jusqu'au rang d'espèce. Enfin, la diminution de la taille de la base de données initiale a également été rendue possible par l'utilisation d'une étape de *clustering*.

A partir de notre base de données, la société NimbleGen a proposé une première version de ce qui pourrait être la puce de capture DB-Chip. Nous avons réalisé les tests *in-silico* pour vérifier la pertinence de nos choix et la capture effective de pathogènes non inclus dans la base de données de départ. L'absence de capture de certains pathogènes (le *Toscana virus* par exemple) était certes mise en évidence, mais pour une application à visée d'exploration respiratoire dans un premier temps, ce manque n'était pas délétère. Cependant, les limites de cette première version de la puce DB-Chip seront à prendre en compte lors de la réalisation d'une deuxième version, pouvant par ailleurs servir à l'exploration d'autres pathologies infectieuses comme les encéphalites.

Enfin, dans un troisième acte de cette approche, nous avons commencé à évaluer, sur des échantillons artificiels, la pertinence de cette nouvelle puce de capture ainsi que ses limites. En utilisant des panels de virus, de bactéries et de champignons connus, nous avons pu montrer que notre puce permettait de retrouver les bactéries les plus représentées dans le mélange avec un lien entre le nombre de *reads* obtenus et la proportion de bactéries présentes dans le mélange. Le *design* de la puce de capture pan-microbienne DB-Chip et son application lors de tests a permis de montrer que l'identification des germes à l'espèce était possible, et non limitée à une identification au genre comme certaines études publiées antérieurement pouvaient le laisser supposer (148). La spécificité des sondes oligonucléotidiques utilisées est telle que les séquences spécifiques des différentes espèces peuvent être capturées, enrichies et séquencées par cette méthode. L'utilisation d'une approche pangénomique quand celle-ci était possible (pour le règne viral en particulier) a permis d'augmenter le nombre de ces séquences spécifiques d'espèces capturées par la puce d'hybridation. Il est probable que la démultiplication des gènes ciblés pour les différentes bactéries nous permette d'atteindre ce niveau de classification taxonomique. En modifiant

certaines paramètres lors de l'analyse, tels que la longueur des k-mers utilisés ou les seuils à atteindre pour raccorder une séquence à un taxon, nous sommes à même de faire varier la sensibilité et la spécificité de l'analyse. Cette approche et cette adaptabilité des algorithmes utilisés permettent de montrer dans notre étude l'absence d'infériorité de notre puce de capture DB-Chip par rapport à la puce de capture décrite par Lipkin et collègues et commercialisée par NimbleGen : VirCapSeq-VERT.

En effet, pour l'identification des virus, les tests réalisés en parallèle entre la puce commerciale VirCapSeq-VERT et notre puce pan-microbienne DB-Chip ne mettent pas en évidence de différence en ce qui concerne le nombre d'espèces détectées ainsi qu'en termes de contamination par des séquences non initialement présentes. Bien entendu, des tests supplémentaires sur la reproductibilité de ces résultats seront requis avant de pouvoir conclure définitivement sur cette non-infériorité de l'approche pan-microbienne par rapport à une approche ciblée par classe d'agent infectieux.

La détection et l'identification des champignons lors de nos tests ont été problématiques du fait d'un défaut dans la préparation des séquences à inclure dans la base de données. Ce défaut de conception explique les résultats décevants obtenus pour la détection et l'identification de *Pneumocystis jirovecii* et d'*Aspergillus fumigatus*, pourtant présents dans les échantillons testés. La synthèse d'une nouvelle puce de capture avec la réintégration des gènes utilisés pour cibler ces espèces permettra de concrétiser pleinement l'approche pan-microbienne que nous défendons.

Une question soulevée par les derniers résultats obtenus est celle de la contamination possible par des espèces microbiennes non présentes dans les échantillons. Nous observons en effet le rattachement taxonomique de certains *reads* à des espèces non présentes dans les

panels utilisés. Ces espèces ne sont pas toujours comparables d'un échantillon à l'autre ni avec les témoins négatifs. Ces résultats non-spécifiques peuvent avoir deux origines : il peut s'agir d'une véritable contamination au plan biologique, par la présence non souhaitée d'un micro-organisme à l'une des étapes de l'analyse biologique par exemple. Mais cette contamination peut également survenir lors de l'analyse bio-informatique avec l'attribution erronée de certains *reads* à des taxons non attendus. Ces non spécificités sont en cours d'investigation, sur le plan biologique, avec la réalisation de PCR d'aliqots prélevés tout au long des expériences menées (avant capture, après capture, avant séquençage). Elles sont également en cours d'investigation en ce qui concerne l'analyse bio-informatique où la similarité des séquences entre différentes espèces peut aboutir à des erreurs de classification. Ce travail permettra de retravailler l'ensemble de pipeline analytique pour identifier et éviter au maximum les sources de contaminations. Cela permettra également de proposer de nouveaux gènes à inclure dans une future version de la puce de capture pour affiner le pouvoir discriminant de l'analyse taxonomique.

Nous travaillons actuellement sur une technique de séquençage (Illumina) dont la principale contrainte est de produire des séquences courtes (200 à 300 pb au maximum). Cette approche pourrait s'avérer limitante et l'utilisation d'une autre technologie de 3<sup>ème</sup> génération (Pacific Bioscience, Oxford Nanopore, Roche Genia, ...) permettrait de travailler sur des séquences plus longues. L'extension des régions génomiques séquencées serait en effet reliée à une augmentation de l'information disponible et donc une meilleure classification taxonomique. En outre, certains gènes d'intérêt comme les gènes liés aux résistances bactériennes pourraient être inclus dans l'analyse. A ce jour cependant, nous ne sommes pas en mesure d'anticiper la longueur maximale et optimale des fragments capturés par une hybridation sur

sondes en milieu liquide utilisée ici. De nombreux tests resteraient à réaliser pour répondre à ces interrogations.

Nous avons inclus dans le premier *design* de notre puce de capture DB-Chip les séquences des phages, ces virus ayant pour cibles des cellules procaryotes. Or, l'extraordinaire diversité de ceux-ci a nécessité l'utilisation d'une partie des ressources techniques disponibles sur cette puce de capture, et ce, pour des résultats dont la pertinence clinique est discutable. A ce jour, la présence ou l'absence de certains bactériophages n'est pas reliée à une symptomatologie particulière, ni à une trajectoire clinique. Nous ne disposons dans la littérature d'aucune donnée concernant les évolutions des populations de phages au sein de microbiotes et l'implication clinique de ces derniers (149). Par ailleurs, l'absence de matériel calibré pour les phages et disponible au laboratoire ne nous permet pas de vérifier les données obtenues. Ainsi, au regard des données disponibles et des résultats obtenus, l'apport de l'identification des phages dans les prélèvements respiratoires est discutable. Compte tenu des contraintes techniques imposées lors du *design* d'une puce de capture pan-microbienne, en particulier sur la réduction du nombre de séquences retenues pour les règnes bactérien et fongique, l'inclusion des séquences dédiées à la capture des génomes des phages lors de la synthèse d'une nouvelle puce de capture sera discutée. Avec à la clé de meilleures performances attendues pour la détection des bactéries et des champignons.

Les résultats présentés ici sont très préliminaires et le travail restant à fournir est important. Le coût élevé des puces de capture et des étapes de séquençage ne nous ont pas permis de réaliser une analyse plus extensive de notre puce de capture DB-Chip. Un seul *run* a été réalisé sur des échantillons artificiels. Ces panels calibrés nous ont permis de mettre en évidence les points forts et les lacunes de notre approche ce qui nous pousse à envisager une deuxième

version de cette puce de capture. Il sera alors nécessaire de répéter les mesures sur panels calibrés et sur échantillons cliniques pour s'assurer de la sensibilité, spécificité, et reproductibilité de la méthode. Nous espérons donc pouvoir prolonger l'étude initiée jusqu'à présent en répondant aux questions soulevées. L'intégration des sondes de captures dédiées à la parasitologie et la mycologie sera une des principales préoccupations de la conception d'une nouvelle puce de capture. L'augmentation de la taille de la base de données de capture avec l'ajout de ces gènes et le dépassement des capacités fixées pourra être solutionnée par l'élimination des sondes dédiées à la capture des génomes des phages, dont l'apport est pour l'instant très limité.

Enfin, et c'est là le principal *challenge* auquel nous faisons face en termes de développement, nous souhaitons travailler à la quantification relative des espèces identifiées. L'apport de la quantification des taxons identifiés est particulièrement important car il permet de distinguer la simple présence, fortuite ou quantitativement peu importante, des acides nucléiques de micro-organismes non impliqués dans la pathologie, des micro-organismes présents dans des proportions importantes et fortement impliqués dans l'apparition des signes cliniques, la dégradation de la fonction respiratoire ou encore l'entretien des processus inflammatoires délétères. Enfin, la possibilité d'introduire une approche quantitative au sein d'une technique de biologie moléculaire répond à une longue quête menée depuis l'introduction de ces dernières techniques pour s'affranchir des limitations dont elles souffrent tel que nous l'avons abordé plus haut (Cf. Partie I – 3.2.4. PCR multiplexes syndromiques).

L'approche choisie pour ce travail à l'aide d'une étape de capture présente certaines limites inhérentes à cette phase de sélection. En ciblant l'analyse sur des séquences fixées au préalable, elle ne permet d'explorer que des micro-organismes connus des bases de données

publiées. Néanmoins, le travail de Lipkin et collègues sur la puce VirCapSeq-VERT montrait la présence de séquences capturées et non identifiées. Certaines de ces séquences pouvaient présenter seulement 75 % d'identité avec des séquences connues, laissant supposer que la détection de pathogènes inconnus et de séquences non attendues mais proches génétiquement pourrait être réalisée. Enfin, l'élimination du génome humain pour l'analyse taxonomique contraint *de facto* à l'élimination des séquences issues de rétrovirus endogènes intégrés. Ces séquences comptent pour environ 8 % de l'ensemble du génome humain (37) et pourraient posséder un rôle dans la régulation des microbiotes.

Malgré ses limites et les obstacles restants à franchir pour une utilisation en clinique, cette approche nous est apparue comme un bon compromis au vu des limitations et biais des autres techniques présentées en introduction. Depuis la première étude sur le virome humain (91) en 2001, le chemin parcouru est impressionnant. Cependant, de nombreuses zones d'ombre subsistent. Concernant l'étude « Dysbiose », nous pourrions proposer quelques pistes d'amélioration. Dans le cadre du recueil des expectorations par exemple, il semble nécessaire de contrôler la cellularité du prélèvement pour éviter (limiter) l'introduction de biais. Il pourrait par ailleurs être envisageable de combiner expectorations, écouvillonnage nasal et oro-pharyngé pour maximiser la sensibilité, comme proposé dans une étude (65) publiée en 2014. Enfin, dans la perspective de valider complètement cette approche pour des échantillons cliniques, il sera nécessaire de définir un référentiel adéquat. Le *gold standard* actuel étant principalement basé sur des techniques de culture et de PCR simplexes ciblant un éventail plus étroit de micro-organismes, cette question comparative est loin d'être triviale. Il sera intéressant de proposer des critères de validation contre une référence complexe mêlant PCR multiplexes et PCR unitaires à la culture et peut être à d'autres méthodes de détection moins utilisées.

L'hypothèse principale que nous chercherons à démontrer dans l'étude « Dysbiose » est qu'un changement dans le microbiote respiratoire intervient en amont de l'exacerbation, tout en étant probablement la cause même de celle-ci. Ainsi, une modification du microbiote, conduisant à l'état de dysbiose serait l'origine de l'exacerbation. Cette hypothèse, si elle venait à être confirmée par nos travaux, pourrait permettre de prévoir la survenue d'une exacerbation en monitorant une ou plusieurs espèces du microbiote respiratoire, avec comme objectif de pouvoir prévenir l'apparition de la dysbiose responsable de la dégradation de la fonction respiratoire. Toutes les données de la littérature ne sont pas en faveur de cette hypothèse. Carmody *et al.* n'avaient pas retrouvé cela (58) avec l'utilisation de l'indice de Shannon dans une étude sur le microbiote respiratoire réalisée avec le gène 16S bactérien à l'aide de séquençage en 454. Mais leur approche était limitée par une identification au genre. D'autres données sont, elles, en faveur de cette hypothèse (150).

Au-delà de cette hypothèse principale, nous espérons obtenir d'importantes quantités de données nous permettant de répondre à d'autres questions. De larges pans de données sont actuellement manquants pour la compréhension des interactions au sein du microbiote respiratoire humain et avec l'hôte qui l'héberge. Les mécanismes moléculaires et physiopathologiques qui sous-tendent les relations virus-virus, virus-hôte et virus-bactéries sont largement inconnus. Leurs rôles dans les réseaux cellulaires, dans les interactions cellule-cellule, dans les *pathways* de développement, des intégrations chromosomiques de matériel viral, dans la modulation du système immunitaire, ... sont supputés mais manquent de données observationnelles fiables pour être validés au-delà du rang de simple assertion. Le lien avec le processus inflammatoire fait également partie du projet et pourra permettre d'évoluer vers une intégration des données de métagénomique infectieuse et de



métabolomique afin de mieux comprendre les mécanismes moléculaires de ces exacerbations.

**THÈSE SOUTENUE PAR :** NEMOZ Benjamin

**TITRE :**

DEVELOPPEMENT D'UNE TECHNIQUE DE METAGENOMIQUE INFECTIEUSE SEMI-CIBLEE BASEE  
SUR LA CAPTURE D'ACIDES NUCLEIQUES POUR L'APPLICATION AU SUIVI DES PATIENTS  
ATTEINTS DE MUCOVISCIDOSE.

**CONCLUSION :**

L'exploration des microbiotes incluant la totalité des micro-organismes bactériens, viraux et fongiques, qu'ils soient digestifs, respiratoires, oraux ou autres, est balbutiante et constitue très probablement une part importante de l'avenir de la biologie et de la médecine. L'avènement des nouvelles techniques de séquençage massif des acides nucléiques rend cette exploration possible. Ces bactéries, virus et champignons que nous hébergeons en très grandes quantités, présentent des interactions encore mal connues mais leurs rôles semblent multiples et décisifs pour la santé humaine. De la disponibilité des nutriments à la stimulation du système immunitaire en passant par un réservoir de diversité génétique, les fonctions assumées par ces micro-organismes sont extrêmement diversifiées et importantes pour nos propres organismes.

Dans le cadre de l'étude Dysbiose, portant sur l'exploration du microbiote respiratoire chez les patients atteints de mucoviscidose, nous avons initié la mise en place, au laboratoire, d'une technique de métagénomique par capture semi-ciblée pour appréhender cette flore respiratoire dans son intégralité. Le développement au plan technique de cette méthodologie novatrice nécessite un travail conjoint avec les équipes de bactériologie, de parasitologie-mycologie, de virologie et de biostatistiques.

Les premières validations de cette approche ont permis d'apporter des réponses quant à l'utilisation de cette technique sur des échantillons cliniques et artificiels. Des essais supplémentaires seront nécessaires pour caractériser la méthode et pouvoir répondre aux

besoins de l'étude clinique Dysbiose après la phase de recrutement de l'ensemble des patients. Néanmoins, les résultats préliminaires obtenus sont gage de confiance pour l'utilisation de cette approche novatrice. Son application à d'autres types de prélèvements (sanguins, de liquide cérébro-spinal, ...) et dans d'autres cadres nosologiques permettra de construire de nouvelles approches diagnostiques et participera à l'élaboration d'une médecine personnalisée au service du malade.

**VU ET PERMIS D'IMPRIMER**

Grenoble, le : 13/03/19

**LE DOYEN**



**Pr. Patrice MORAND**

**LE PRÉSIDENT DE LA THÈSE**



**Pr. Patrice MORAND**

## REFERENCES

1. Vareille M, Kieninger E, Edwards MR, Regamey N. The airway epithelium: Soldier in the fight against respiratory viruses. *Clin Microbiol Rev.* 2011 Jan 1;24(1):210–29.
2. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012 Dec;380(9859):2095–128.
3. Elborn JS. Cystic fibrosis [Internet]. Vol. 388, *The Lancet.* 2016. p. 2519–31.
4. Andersen DH. Cystic fibrosis of the pancreas and its relation to celiac disease. *Am J Dis Child.* 1938 Aug 1;56(2):344.
5. Plasschaert LW, Žilionis R, Choo-Wing R, Savova V, Knehr J, Roma G, et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature.* 2018 Aug 1;
6. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket S, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature.* 2018;
7. Travaglini KJ, Krasnow MA. Profile of an unknown airway cell. *Nature.* 2018 Aug 1;
8. Bodas M, Vij N. The NF-kappaB signaling in cystic fibrosis lung disease: pathophysiology and therapeutic potential. *Discov Med.* 2010;9(47):346–56.
9. Billard L, Le Berre R, Pilorgé L, Payan C, Héry-Arnaud G, Vallet S. Viruses in cystic fibrosis patients' airways. *Crit Rev Microbiol.* 2017;43(6):690–708.
10. Stoltz DA, Meyerholz DK, Welsh MJ. Origins of Cystic Fibrosis Lung Disease. *N Engl J Med.* 2015;372(4):351–62.
11. Bradley JM, Blume SW, Balp MM, Honeybourne D, Elborn JS. Quality of life and healthcare utilisation in cystic fibrosis: A multicentre study. *Eur Respir J.* 2013;41(3):571–7.
12. Paganin P, Fiscarelli EV, Tuccio V, Chiancianesi M, Bacci G, Morelli P, et al. Changes in cystic fibrosis airway microbial community associated with a severe decline in lung function. *PLoS One.* 2015;10(4):1–19.
13. HAS. Le Dépistage Néonatal Systématique De La Mucoviscidose En France : Etat Des Lieux Et Perspectives Après 5 Ans De Fonctionnement. 2009;1–174.
14. Rosenfeld M, Emerson J, Williams-Warren J, Pepe M, Smith A, Montgomery AB, et al. Defining

- a pulmonary exacerbation in cystic fibrosis. *J Pediatr*. 2001 Sep;139(3):359–65.
15. Goss CH, Burns JL. Exacerbations in cystic fibrosis. 1: Epidemiology and pathogenesis. *Thorax*. 2007 Jan 24;62(4):360–7.
  16. Bilton D, Canny G, Conway S, Dumcius S, Hjelte L, Proesmans M, et al. Pulmonary exacerbation: Towards a definition for use in clinical trials. Report from the EuroCareCF Working Group on outcome parameters in clinical trials. *J Cyst Fibros*. 2011 Jun;10:S79–81.
  17. Skolnik K, Quon BS. Recent advances in the understanding and management of cystic fibrosis pulmonary exacerbations. *F1000Research*. 2018;7(May):575.
  18. Anstead M, Saiman L, Mayer-Hamblett N, Lands LC, Kloster M, Goss CH, et al. Pulmonary exacerbations in CF patients with early lung disease. *J Cyst Fibros*. 2014 Jan;13(1):74–9.
  19. Smyth A. Treatment of pulmonary exacerbations in cystic fibrosis – could do better? *Paediatr Respir Rev*. 2016;20:6–7.
  20. Sanders DB, Solomon GM, Beckett V V., West NE, Daines CL, Heltshe SL, et al. Standardized Treatment of Pulmonary Exacerbations (STOP) study: Observations at the initiation of intravenous antibiotics for cystic fibrosis pulmonary exacerbations. *J Cyst Fibros*. 2017 Sep;16(5):592–9.
  21. Lechtzin N, Mayer-Hamblett N, West NE, Allgood S, Wilhelm E, Khan U, et al. Home Monitoring of Patients with Cystic Fibrosis to Identify and Treat Acute Pulmonary Exacerbations. eICE Study Results. *Am J Respir Crit Care Med*. 2017 Nov;196(9):1144–51.
  22. VanDevanter DR, Morris NJ, Konstan MW. IV-treated pulmonary exacerbations in the prior year: An important independent risk factor for future pulmonary exacerbation in cystic fibrosis. *J Cyst Fibros*. 2016 May;15(3):372–9.
  23. Stanojevic S, McDonald A, Waters V, MacDonald S, Horton E, Tullis E, et al. Effect of pulmonary exacerbations treated with oral antibiotics on clinical outcomes in cystic fibrosis. *Thorax*. 2017 Apr;72(4):327–32.
  24. Sanders DB, Bittner RCL, Rosenfeld M, Hoffman LR, Redding GJ, Goss CH. Failure to Recover to Baseline Pulmonary Function after Cystic Fibrosis Pulmonary Exacerbation. *Am J Respir Crit Care Med*. 2010 Sep;182(5):627–32.
  25. Chin M, De Zoysa M, Slinger R, Gaudet E, Vandemheen KL, Chan F, et al. Acute effects of viral respiratory tract infections on sputum bacterial density during CF pulmonary exacerbations. *J*

- Cyst Fibros. 2015 Jul;14(4):482–9.
26. Asner S, Waters V, Solomon M, Yau Y, Richardson SE, Grasemann H, et al. Role of respiratory viruses in pulmonary exacerbations in children with cystic fibrosis. *J Cyst Fibros*. 2012 Sep;11(5):433–9.
  27. Wat D, Gelder C, Hibbitts S, Cafferty F, Bowler I, Pierrepoint M, et al. The role of respiratory viruses in cystic fibrosis. *J Cyst Fibros*. 2008 Jul;7(4):320–8.
  28. CRCM Lyon. Protocole National de Diagnostic et de Soins ( PNDS ) Mucoviscidose Centre de référence Mucoviscidose Sommaire. 2017;
  29. Quinton PM. Cystic fibrosis: impaired bicarbonate secretion and mucoviscidosis. *Lancet*. 2008 Aug;372(9636):415–7.
  30. Quinton PM. Role of epithelial HCO<sub>3</sub><sup>-</sup> transport in mucin secretion: lessons from cystic fibrosis. *AJP Cell Physiol*. 2010;299(6):C1222–33.
  31. Keiser NW, Birket SE, Evans IA, Tyler SR, Croke AK, Sun X, et al. Defective innate immunity and hyperinflammation in newborn cystic fibrosis transmembrane conductance regulator-knockout ferret lungs. *Am J Respir Cell Mol Biol*. 2015;52(6):683–94.
  32. Wu H xia, Zhu M, Xiong X feng, Wei J, Zhuo K quan, Cheng D yun. Efficacy and Safety of CFTR Corrector and Potentiator Combination Therapy in Patients with Cystic Fibrosis for the F508del-CFTR Homozygous Mutation: A Systematic Review and Meta-analysis. *Adv Ther*. 2019;36(2):451–61.
  33. Aslam A, Sinha IP, Southern KW. Ataluren and similar compounds (specific therapies for premature termination codon class I mutations) for cystic fibrosis. In: Aslam A, editor. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd; 2016.
  34. Marsland BJ, Gollwitzer ES. Host–microorganism interactions in lung diseases. *Nat Rev Immunol*. 2014;14(12):827–35.
  35. Human Microbiome Project C. A framework for human microbiome research. *Nature*. 2012;486(7402):215–21.
  36. Rascovan N, Duraisamy R, Desnues C. Metagenomics and the Human Virome in Asymptomatic Individuals. *Annu Rev Microbiol*. 2016;70(1):125–41.
  37. Wylie KM, Weinstock GM, Storch GA. Emerging view of the human virome. *Transl Res*. 2012 Oct;160(4):283–90.

38. Virgin HW, Wherry EJ, Ahmed R. Redefining Chronic Viral Infection. *Cell*. 2009;138(1):30–50.
39. Round JL, Mazmanian SK. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci*. 2010 Jul 6;107(27):12204–9.
40. Atarashi K, Tanoue T, Shima T, Imaoka A, Kuwahara T, Momose Y, et al. Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species. *Science* (80- ). 2011 Jan 21;331(6015):337–41.
41. Petersen C, Round JL. Defining dysbiosis and its influence on host immunity and disease. *Cell Microbiol*. 2014;16(7):1024–33.
42. Muzny CA, Schwebke JR. Pathogenesis of Bacterial Vaginosis: Discussion of Current Hypotheses. *J Infect Dis*. 2016 Aug 15;214(suppl 1):S1–5.
43. Cox MJ, Allgaier M, Taylor B, Baek MS, Huang YJ, Daly RA, et al. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS One*. 2010;5(6).
44. Wang EEL, Prober CG, Manson B, Corey M, Levison H. Association of Respiratory Viral Infections with Pulmonary Deterioration in Patients with Cystic Fibrosis. *N Engl J Med*. 1984 Dec 27;311(26):1653–8.
45. Ramsey BW, Gore EJ, Smith AL, Cooney MK, Redding GJ, Foy H. The effect of respiratory viral infections on patients with cystic fibrosis. *Am J Dis Child*. 1989;143(6):662–8.
46. Collinson J, Nicholson KG, Cancio E, Ashman J, Ireland DC, Hammersley V, et al. Effects of upper respiratory tract infections in patients with cystic fibrosis. *Thorax*. 1996 Nov;51(11):1115–22.
47. Armstrong D, Grimwood K, Carlin JB, Carzino R, Hull J, Olinsky A, et al. Severe viral respiratory infections in infants with cystic fibrosis. *Pediatr Pulmonol*. 1998 Dec;26(6):371–9.
48. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, et al. Optimizing methods and dodging pitfalls in microbiome research. *Microbiome*. 2017;5(1):1–14.
49. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, et al. Interplay Between the Temporal Dynamics of the Vaginal Microbiota and Human Papillomavirus Detection. *J Infect Dis*. 2014 Dec 1;210(11):1723–33.
50. Thaïss CA, Zeevi D, Levy M, Segal E, Elinav E. A day in the life of the meta-organism: diurnal rhythms of the intestinal microbiome and its host. *Gut Microbes*. 2015 Mar 4;6(2):137–42.

51. Barley M, McNally J, Marshall B, Faro A, Elbert A, Fink A, et al. Annual Data Report 2016 Cystic Fibrosis Foundation Patient Registry. *Cyst Fibros Found Patient Regist.* 2016;1–94.
52. Salsgiver EL, Fink AK, Knapp EA, LiPuma JJ, Olivier KN, Marshall BC, et al. Changing epidemiology of the respiratory bacteriology of patients with cystic fibrosis. *Chest.* 2016;149(2):390–400.
53. Ong T, Ramsey BW. Update in Cystic Fibrosis 2014. *Am J Respir Crit Care Med.* 2015 Sep 15;192(6):669–75.
54. Parkins MD, Floto RA. Emerging bacterial pathogens and changing concepts of bacterial pathogenesis in cystic fibrosis. *J Cyst Fibros.* 2015;14(3):293–304.
55. Sherrard LJ, Tunney MM, Elborn JS. Antimicrobial resistance in the respiratory microbiota of people with cystic fibrosis. *Lancet.* 2014;384(9944):703–13.
56. O’Neill K, Bradley JM, Johnston E, McGrath S, McIlreavey L, Rowan S, et al. Reduced bacterial colony count of anaerobic bacteria is associated with a worsening in lung clearance index and inflammation in cystic fibrosis. *PLoS One.* 2015;10(5).
57. Murphy TF, Parameswaran GI. *Moraxella catarrhalis*, a Human Respiratory Tract Pathogen. *Clin Infect Dis.* 2009 Jul;49(1):124–31.
58. Carmody LA, Zhao J, Schloss PD, Petrosino JF, Murray S, Young VB, et al. Changes in Cystic Fibrosis Airway Microbiota at Pulmonary Exacerbation. *Ann Am Thorac Soc.* 2013 Jun;10(3):179–87.
59. Stressmann FA, Rogers GB, Klem ER, Lilley AK, Donaldson SH, Daniels TW, et al. Analysis of the Bacterial Communities Present in Lungs of Patients with Cystic Fibrosis from American and British Centers. *J Clin Microbiol.* 2011 Jan 1;49(1):281–91.
60. Sibley CD, Grinwis ME, Field TR, Parkins MD, Norgaard JC, Gregson DB, et al. McKay agar enables routine quantification of the “*Streptococcus milleri*” group in cystic fibrosis patients. *J Med Microbiol.* 2010 May 1;59(5):534–40.
61. Hajishengallis G, Darveau RP, Curtis MA. The keystone-pathogen hypothesis. *Nat Rev Microbiol.* 2012;10(10):717–25.
62. Hajishengallis G, Liang S, Payne MA, Hashim A, Jotwani R, Eskin MA, et al. Low-Abundance Biofilm Species Orchestrates Inflammatory Periodontal Disease through the Commensal Microbiota and Complement. *Cell Host Microbe.* 2011 Nov;10(5):497–506.



63. Byington CL, Ampofo K, Stockmann C, Adler FR, Herbener A, Miller T, et al. Community Surveillance of Respiratory Viruses Among Families in the Utah Better Identification of Germs-Longitudinal Viral Epidemiology (BIG-LoVE) Study. *Clin Infect Dis*. 2015 Oct 15;61(8):1217–24.
64. Wylie KM. The Virome of the Human Respiratory Tract. *Clin Chest Med*. 2017;38(1):11–9.
65. Flight WG, Bright-Thomas RJ, Tilston P, Mutton KJ, Guiver M, Morris J, et al. Incidence and clinical impact of respiratory viruses in adults with cystic fibrosis. *Thorax*. 2014 Mar;69(3):247–53.
66. Kapusinszky B, Minor P, Delwart E. Nearly Constant Shedding of Diverse Enteric Viruses by Two Healthy Infants. *J Clin Microbiol*. 2012 Nov 1;50(11):3427–34.
67. Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat Med*. 2015 Oct 14;21(10):1228–34.
68. Reyes A, Blanton L V., Cao S, Zhao G, Manary M, Trehan I, et al. Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proc Natl Acad Sci*. 2015 Sep 22;112(38):11941–6.
69. Deng X, Terunuma H, Handema R, Sakamoto M, Kitamura T, Ito M, et al. Higher prevalence and viral load of TT virus in saliva than in the corresponding serum: another possible transmission route and replication site of TT virus. *J Med Virol*. 2000 Dec;62(4):531–7.
70. Touinssi M, Gallian P, Biagini P, Attoui H, Vialettes B, Berland Y, et al. TT virus infection: prevalence of elevated viraemia and arguments for the immune control of viral load. *J Clin Virol*. 2001 May;21(2):135–41.
71. Goto K, Sugiyama K, Ando T, Mizutani F, Terabe K, Tanaka K, et al. Detection rates of TT virus DNA in serum of umbilical cord blood, breast milk and saliva. *Tohoku J Exp Med*. 2000 Aug;191(4):203–7.
72. Inami T, Konomi N, Arakawa Y, Abe K. High prevalence of TT virus DNA in human saliva and semen. *J Clin Microbiol*. 2000 Jun;38(6):2407–8.
73. Wylie KM, Mihindukulasuriya KA, Sodergren E, Weinstock GM, Storch GA. Sequence analysis of the human virome in Febrile and Afebrile children. *PLoS One*. 2012;7(6).
74. Young JC, Chehoud C, Bittinger K, Bailey A, Diamond JM, Cantu E, et al. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am J Transplant*. 2015 Jan;15(1):200–9.

75. Willner D, Haynes MR, Furlan M, Hanson N, Kirby B, Lim YW, et al. Case Studies of the Spatial Heterogeneity of DNA Viruses in the Cystic Fibrosis Lung. *Am J Respir Cell Mol Biol*. 2012 Feb;46(2):127–31.
76. Barton ES, White DW, Cathelyn JS, Brett-McClellan KA, Engle M, Diamond MS, et al. Herpesvirus latency confers symbiotic protection from bacterial infection. *Nature*. 2007 May 17;447(7142):326–9.
77. Bhattarai N, Stapleton JT. GB virus C: the good boy virus? *Trends Microbiol*. 2012 Mar;20(3):124–30.
78. McLaughlin-Drubin ME, Munger K. Viruses associated with human cancer. *Biochim Biophys Acta - Mol Basis Dis*. 2008 Mar;1782(3):127–50.
79. Abeles SR, Ly M, Santiago-Rodriguez TM, Pride DT. Effects of Long Term Antibiotic Therapy on Human Oral and Fecal Viromes. Heimesaat MM, editor. *PLoS One*. 2015 Aug 26;10(8):e0134941.
80. Abeles SR, Robles-Sikisaka R, Ly M, Lum AG, Salzman J, Boehm TK, et al. Human oral viruses are personal, persistent and gender-consistent. *ISME J*. 2014 Sep 20;8(9):1753–67.
81. Ly M, Abeles SR, Boehm TK, Robles-Sikisaka R, Naidu M, Santiago-Rodriguez T, et al. Altered Oral Viral Ecology in Association with Periodontal Disease. *MBio*. 2014 May 20;5(3).
82. Pride DT, Salzman J, Haynes M, Rohwer F, Davis-Long C, White RA, et al. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *ISME J*. 2012 May 8;6(5):915–26.
83. Robles-Sikisaka R, Ly M, Boehm T, Naidu M, Salzman J, Pride DT. Association between living environment and human oral viral ecology. *ISME J*. 2013 Sep 18;7(9):1710–24.
84. Lum AG, Ly M, Santiago-Rodriguez TM, Naidu M, Boehm TK, Pride DT. Global transcription of CRISPR loci in the human oral cavity. *BMC Genomics*. 2015 Dec 21;16(1):401.
85. Wang J, Gao Y, Zhao F. Phage-bacteria interaction network in human oral microbiome. *Environ Microbiol*. 2016 Jul;18(7):2143–58.
86. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*. 2010 Jul 15;466(7304):334–8.
87. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. *Proc Natl Acad Sci*. 2013 Jul 23;110(30):12450–5.

88. Mitchell AB, Oliver BGG, Glanville AR. Translational Aspects of the Human Respiratory Virome. *Am J Respir Crit Care Med*. 2016 Dec 15;194(12):1458–64.
89. Zárata S, Taboada B, Yocupicio-Monroy M, Arias CF. Human Virome. *Arch Med Res*. 2017 Nov;48(8):701–16.
90. Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, et al. Viral diversity and dynamics in an infant gut. *Res Microbiol*. 2008 Jun;159(5):367–73.
91. Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, et al. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol*. 2003 Oct;185(20):6220–3.
92. Modi SR, Lee HH, Spina CS, Collins JJ. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature*. 2013 Jul 9;499(7457):219–22.
93. Zhang T, Breitbart M, Lee WH, Run J-Q, Wei CL, Soh SWL, et al. RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. Dangl J, editor. *PLoS Biol*. 2005 Dec 20;4(1):e3.
94. Sharon I, Morowitz MJ, Thomas BC, Costello EK, Relman DA, Banfield JF. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Res*. 2013 Jan 1;23(1):111–20.
95. David LA, Weil A, Ryan ET, Calderwood SB, Harris JB, Chowdhury F, et al. Gut Microbial Succession Follows Acute Secretory Diarrhea in Humans. *MBio*. 2015 Jul 1;6(3).
96. Fancello L, Desnues C, Raoult D, Rolain JM. Bacteriophages and diffusion of genes encoding antimicrobial resistance in cystic fibrosis sputum microbiota. *J Antimicrob Chemother*. 2011 Nov 1;66(11):2448–54.
97. Hannigan GD, Meisel JS, Tyldsley AS, Zheng Q, Hodgkinson BP, SanMiguel AJ, et al. The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. *MBio*. 2015 Oct 30;6(5).
98. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res*. 2011 Oct 1;21(10):1616–25.
99. Abeles SR, Pride DT. Molecular Bases and Role of Viruses in the Human Microbiome. *J Mol Biol*. 2014 Nov;426(23):3892–906.
100. Norman JM, Handley SA, Baldrige MT, Droit L, Liu CY, Keller BC, et al. Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell*. 2015 Jan;160(3):447–

- 60.
101. Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci*. 2013 Jun 25;110(26):10771–6.
  102. Barr JJ, Auro R, Sam-Soon N, Kassegne S, Peters G, Bonilla N, et al. Subdiffusive motion of bacteriophage in mucosal surfaces increases the frequency of bacterial encounters. *Proc Natl Acad Sci*. 2015 Nov 3;112(44):13675–80.
  103. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One*. 2009;4(10).
  104. Tipton L, Ghedin E, Morris A. The lung mycobiome in the next-generation sequencing era. *Virulence*. 2017;8(3):334–41.
  105. Cui L, Lucht L, Tipton L, Rogers MB, Fitch A, Kessinger C, et al. Topographic Diversity of the Respiratory Tract Mycobiome and Alteration in HIV and Lung Disease. *Am J Respir Crit Care Med*. 2015 Apr 15;191(8):932–42.
  106. Willger SD, Grim SL, Dolben EL, Shipunova A, Hampton TH, Morrison HG, et al. Characterization and quantification of the fungal microbiome in serial samples from individuals with cystic fibrosis. *Microbiome*. 2014;2(1):1–15.
  107. Kim SH, Clark ST, Surendra A, Copeland JK, Wang PW, Ammar R, et al. Global Analysis of the Fungal Microbiome in Cystic Fibrosis Patients Reveals Loss of Function of the Transcriptional Repressor Nrg1 as a Mechanism of Pathogen Adaptation. Hogan DA, editor. *PLOS Pathog*. 2015 Nov 20;11(11):e1005308.
  108. Felton IC, Simmonds NJ. Aspergillus and cystic fibrosis: Old disease - New classifications. *Curr Opin Pulm Med*. 2014;20(6):632–8.
  109. Gutiérrez S, Respaldiza N, Campano E, Martínez-Risquez MT, Calderón EJ, De La Horra C. *Pneumocystis jirovecii* colonization in chronic pulmonary disease. *Parasite*. 2011 May 15;18(2):121–6.
  110. Green HD, Bright-Thomas RJ, Mutton KJ, Guiver M, Jones AM. Increased prevalence of *Pneumocystis jirovecii* colonisation in acute pulmonary exacerbations of cystic fibrosis. *J Infect*. 2016;73(1):1–7.

111. Rogers GB, Carroll MP, Serisier DJ, Hockey PM, Jones G, Kehagia V, et al. Use of 16S rRNA gene profiling by terminal restriction fragment length polymorphism analysis to compare bacterial communities in sputum and mouthwash samples from patients with cystic fibrosis. *J Clin Microbiol.* 2006;44(7):2601–4.
112. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science.* 1985 Dec 20;230(4732):1350–4.
113. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science.* 1988 Jan 29;239(4839):487–91.
114. Krause JC, Panning M, Hengel H, Henneke P. The Role of Multiplex PCR in Respiratory Tract Infections in Children. *Dtsch Arzteblatt Online.* 2014 Sep 19;111(38):639–45.
115. Corless CE, Guiver M, Borrow R, Edwards-Jones V, Kaczmarski EB, Fox AJ. Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. *J Clin Microbiol.* 2000 May;38(5):1747–52.
116. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci.* 2012 Apr 17;109(16):6241–6.
117. Dantas G, Sommer MO. Context matters — the complex interplay between resistome genotypes and resistance phenotypes. *Curr Opin Microbiol.* 2012 Oct;15(5):577–82.
118. Høiby N, Bjarnsholt T, Moser C, Bassi GL, Coenye T, Donelli G, et al. ESCMID\* guideline for the diagnosis and treatment of biofilm infections 2014. *Clin Microbiol Infect.* 2015;21(S1):S1–25.
119. Waters V, Ratjen F. Standard versus biofilm antimicrobial susceptibility testing to guide antibiotic therapy in cystic fibrosis. *Cochrane Database Syst Rev.* 2012;(10).
120. Hwang SG, Ha K, Guk K, Lee DK, Eom G, Song S, et al. Rapid and simple detection of Tamiflu-resistant influenza virus: Development of oseltamivir derivative-based lateral flow biosensor for point-of-care (POC) diagnostics. *Sci Rep.* 2018 Dec 29;8(1):12999.
121. Zou S, Caler L, Colombini-Hatch S, Glynn S, Srinivas P. Research on the human virome: Where are we and what is next. *Microbiome.* 2016;4:1–4.
122. Wylie KM, Mihindukulasuriya KA, Zhou Y, Sodergren E, Storch GA, Weinstock GM.

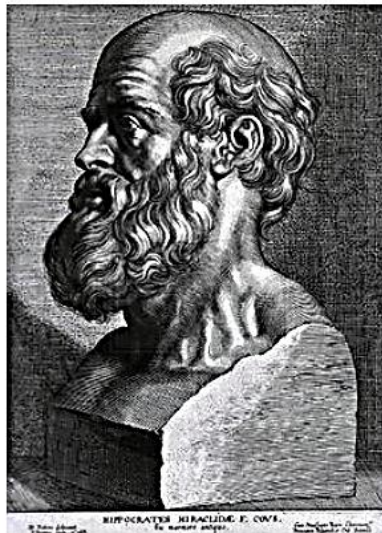
- Metagenomic analysis of double-stranded DNA viruses in healthy adults. *BMC Biol.* 2014 Dec 10;12(1):71.
123. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014 Dec 12;12(1):87.
124. Commins J, Toft C, Fares MA. Computational Biology Methods and Their Application to the Comparative Genomics of Endocellular Symbiotic Bacteria of Insects. *Biol Proced Online.* 2009 Dec 11;11(1):52–78.
125. Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced Virome Sequencing Using Targeted Sequence Capture Enhanced virome sequencing using sequence capture. *Genome Res.* 2015;4:1910–20.
126. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, et al. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. *MBio.* 2015 Oct 30;6(5):e01491-15.
127. Lewandowska DW, Schreiber PW, Schuurmans MM, Ruehe B, Zagordi O, Bayard C, et al. Metagenomic sequencing complements routine diagnostics in identifying viral pathogens in lung transplant recipients with unknown etiology of respiratory infection. *PLoS One.* 2017;12(5):1–16.
128. Conceição-Neto N, Zeller M, Lefrère H, De Bruyn P, Beller L, Deboutte W, et al. Modular approach to customise sample preparation procedures for viral metagenomics: A reproducible protocol for virome analysis. *Sci Rep.* 2015;5(November):1–14.
129. Lovett M, Kere J, Hinton LM. Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc Natl Acad Sci U S A.* 1991 Nov 1;88(21):9628–32.
130. Duncavage EJ, Magrini V, Becker N, Armstrong JR, Demeter RT, Wylie T, et al. Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. *J Mol Diagnostics.* 2011;13(3):325–33.
131. Depledge DP, Palser AL, Watson SJ, Lai IYC, Gray ER, Grant P, et al. Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. *PLoS One.* 2011;6(11).
132. Koehler JW, Hall AT, Rolfe PA, Honko AN, Palacios GF, Fair JN, et al. Development and Evaluation of a Panel of Filovirus Sequence Capture Probes for Pathogen Detection by Next-Generation Sequencing. Kuhn JH, editor. *PLoS One.* 2014 Sep 10;9(9):e107007.

133. Emerson J, Cochrane E, McNamara S, Kuypers J, Gibson RL, Campbell AP. Home Self-Collection of Nasal Swabs for Diagnosis of Acute Respiratory Virus Infections in Children With Cystic Fibrosis. *J Pediatric Infect Dis Soc.* 2013 Dec;2(4):345–51.
134. Burns JL, Emerson J, Kuypers J, Campbell AP, Gibson RL, McNamara S, et al. Respiratory viruses in children with cystic fibrosis: viral detection and clinical findings. *Influenza Other Respi Viruses.* 2012 May;6(3):218–23.
135. Glendinning L, Wright S, Tennant P, Gill AC, Collie D, McLachlan G. Microbiota in exhaled breath condensate and the lung. *Appl Environ Microbiol.* 2017;83(12).
136. Bjornson C, Chan P, Li A, Paes B, Lanctôt KL, Mitchell I. Palivizumab prophylaxis for respiratory syncytial virus in infants with cystic fibrosis: is there a need? *Eur J Clin Microbiol Infect Dis.* 2018 Jun 19;37(6):1113–8.
137. Bruzzese E, Raia V, Spagnuolo MI, Volpicelli M, De Marco G, Maiuri L, et al. Effect of *Lactobacillus GG* supplementation on pulmonary exacerbations in patients with cystic fibrosis: A pilot study. *Clin Nutr.* 2007;26(3):322–8.
138. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15(3).
139. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics.* 2011;12(1):385.
140. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012 Dec;28(23):3150–2.
141. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006 Jul 1;22(13):1658–9.
142. Gadsby NJ, McHugh MP, Russell CD, Mark H, Conway Morris A, Laurenson IF, et al. Development of two real-time multiplex PCR assays for the detection and quantification of eight key bacterial pathogens in lower respiratory tract infections. *Clin Microbiol Infect.* 2015 Aug;21(8):788.e1-788.e13.
143. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D535–42.
144. Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, et al. Linking Virus

- Genomes with Host Taxonomy. *Viruses*. 2016 Mar 1;8(3):66.
145. Zakharkina T, Koczulla AR, Mardanova O, Hattesoehl A, Bals R. Detection of microorganisms in exhaled breath condensate during acute exacerbations of COPD. *Respirology*. 2011;16(6):932–8.
  146. Cummings MJ, Tokarz R, Bakamutumaho B, Kayiwa J, Byaruhanga T, Owor N, et al. Precision surveillance for viral respiratory pathogens: virome capture sequencing for the detection and genomic characterization of severe acute respiratory infection in Uganda. *Clin Infect Dis*. 2018 Aug 7;(August 2018).
  147. Williams SH, Cordey S, Bhuvu N, Laubscher F, Hartley M-A, Boillat-Blanco N, et al. Investigation of the Plasma Virome from Cases of Unexplained Febrile Illness in Tanzania from 2013 to 2014: a Comparative Analysis between Unbiased and VirCapSeq-VERT High-Throughput Sequencing Approaches. Duprex WP, editor. *mSphere*. 2018 Aug 22;3(4).
  148. Prakash T, Taylor TD. Functional assignment of metagenomic data: Challenges and applications. *Brief Bioinform*. 2012;13(6):711–27.
  149. Shkoporov AN, Hill C. Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host Microbe*. 2019;25(2):195–209.
  150. Quinn RA, Whiteson K, Lim YW, Zhao J, Conrad D, Lipuma JJ, et al. Ecological networking of cystic fibrosis lung infections. *npj Biofilms Microbiomes*. 2016;2(1):0–1.



## SERMENT D'HIPPOCRATE



### **SERMENT D'HIPPOCRATE**

*En présence des Maîtres de cette Faculté, de mes chers condisciples et devant l'effigie d'HIPPOCRATE,*

*Je promets et je jure d'être fidèle aux lois de l'honneur et de la probité dans l'exercice de la Médecine.*

*Je donnerai mes soins gratuitement à l'indigent et n'exigerai jamais un salaire au dessus de mon travail. Je ne participerai à aucun partage clandestin d'honoraires.*

*Admis dans l'intimité des maisons, mes yeux n'y verront pas ce qui s'y passe ; ma langue taira les secrets qui me seront confiés et mon état ne servira pas à corrompre les mœurs, ni à favoriser le crime.*

*Je ne permettrai pas que des considérations de religion, de nation, de race, de parti ou de classe sociale viennent s'interposer entre mon devoir et mon patient.*

*Je garderai le respect absolu de la vie humaine.*

*Même sous la menace, je n'admettrai pas de faire usage de mes connaissances médicales contre les lois de l'humanité.*

*Respectueux et reconnaissant envers mes Maîtres, je rendrai à leurs enfants l'instruction que j'ai reçue de leurs pères.*

*Que les hommes m'accordent leur estime si je suis fidèle à mes promesses.*

*Que je sois couvert d'opprobre et méprisé de mes confrères si j'y manque.*



---

**Développement d'une technique de métagénomique infectieuse semi-ciblée  
basée sur la capture d'acides nucléiques pour l'application au suivi des  
patients atteints de mucoviscidose.**

---

**RESUME :**

L'exploration du microbiote respiratoire dans le cadre du suivi des patients atteints de mucoviscidose nécessite à ce jour la mise en culture de prélèvements respiratoires ou l'utilisation de PCR sélectives. Ces techniques présentent d'importantes limitations et ne sont pas adaptées à une exploration pan-microbienne qui serait nécessaire pour mieux comprendre les mécanismes physiopathologiques à l'œuvre dans cette pathologie et dans son évolution. Nous présentons ici une approche novatrice pour l'exploration de ce microbiote respiratoire et les premiers essais réalisés au sein du laboratoire pour la caractériser. Cette méthode est basée sur le séquençage de nouvelle génération (NGS) des échantillons après un enrichissement par capture semi-ciblée en milieu liquide. Cette capture est réalisée par hybridation sur sondes des acides nucléiques d'intérêt. La réalisation de cet enrichissement a nécessité la sélection de cibles d'intérêt pour la fabrication d'une puce de capture pan-microbienne. Nous montrons dans une deuxième partie de ce travail les résultats obtenus à l'aide de cette puce DB-Chip pour la détection de bactéries, virus et champignons au sein d'échantillons artificiels. Cette approche retrouve les pathogènes attendus sur les prélèvements dont la composition est connue. L'identification des micro-organismes est possible jusqu'au rang d'espèce pour les bactéries et les virus. Les performances de notre puce de capture ne sont pas inférieures à celles obtenues par d'autres équipes avec des approches axées sur un seul règne (viral ou bactérien). Notre approche est prometteuse mais nécessite des ajustements pour son utilisation en clinique.

---

**Mot-clés :** Métagénomique ; Mucoviscidose ; Dysbiose ; NGS (Next-Generation Sequencing ; Capture semi-ciblée ; Microbiote respiratoire ; Virome

---

**Jury :** Président du jury : Monsieur le Professeur Patrice MORAND  
Membres du jury : Madame le Docteur Sylvie LARRAT  
Monsieur le Docteur Boubou CAMARA  
Madame le Professeur Muriel CORNET  
Monsieur le Docteur Yvan CASPAR  
Monsieur le Professeur Alexandre MOREAU-GAUDRY

---

Thèse soutenue le 05 Avril 2019 à la Faculté de Médecine de Grenoble