



HAL
open science

Apport des technologies de séquençage haut débit en microbiologie clinique : revue de la littérature et projet d'implantation au CHU de Rouen

Nicolas Delettre

► **To cite this version:**

Nicolas Delettre. Apport des technologies de séquençage haut débit en microbiologie clinique : revue de la littérature et projet d'implantation au CHU de Rouen. Sciences pharmaceutiques. 2022. dumas-03978975

HAL Id: dumas-03978975

<https://dumas.ccsd.cnrs.fr/dumas-03978975>

Submitted on 8 Feb 2023

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.

UNIVERSITE DE ROUEN NORMANDIE
UFR SANTE – Département PHARMACIE

Année 2022

THESE

pour le DIPLOME D'ETAT DE DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement le 19/10/2022

par

Nicolas Delette

Né le 12/01/1994 à Paris

***Apport des technologies de séquençage haut débit
en microbiologie clinique : revue de la littérature et
projet d'implantation au CHU de Rouen***

Président du jury : *Pr. Pestel-Caron Martine, PUPH*

Membres du jury : *Dr. Lemée Ludovic, PH*

Dr. Tazi Asmaa, MCU-PH

Pr. Plantier Jean-Christophe, PUPH

ANNEE UNIVERSITAIRE 2021 - 2022

U.F.R. SANTÉ DE ROUEN

DOYEN : **Professeur Benoît VEBER**

ASSESEURS : **Professeur Loïc FAVENNEC**
Professeur Agnès LIARD
Professeur Guillaume SAVOYE

I - MEDECINE

PROFESSEURS DES UNIVERSITES – PRATICIENS HOSPITALIERS

Mr Frédéric ANSELME	HCN	Cardiologie
Mme Gisèle APTER	Havre	Pédopsychiatrie
Mme Isabelle AUQUIT AUCKBUR	HCN	Chirurgie plastique
Mr Jean-Marc BASTE	HCN	Chirurgie Thoracique
Mr Fabrice BAUER	HCN	Cardiologie
Mme Soumeya BEKRI	HCN	Biochimie et biologie moléculaire
Mr Ygal BENHAMOU	HCN	Médecine interne
Mr Jacques BENICHOU	HCN	Bio statistiques et informatique médicale
Mr Olivier BOYER	UFR	Immunologie
Mme Valérie BRIDOUX HUYBRECHTS	HCN	Chirurgie Vasculaire
Mme Sophie CANDON	HCN	Immunologie
Mr François CARON	HCN	Maladies infectieuses et tropicales
Mr Philippe CHASSAGNE	HCN	Médecine interne (gériatrie)
Mr Florian CLATOT	CB	Cancérologie – Radiothérapie
Mr Moïse COEFFIER	HCN	Nutrition
Mr Vincent COMPERE	HCN	Anesthésiologie et réanimation chirurgicale
Mr Jean-Nicolas CORNU	HCN	Urologie
Mr Antoine CUVELIER	HB	Pneumologie
Mr Jean-Nicolas DACHER	HCN	Radiologie et imagerie médicale
Mr Stéfan DARMONI	HCN	Informatique médicale et techniques de communication

Mr Pierre DECHELOTTE	HCN	Nutrition
Mr Stéphane DERREY	HCN	Neurochirurgie
Mr Frédéric DI FIORE	CHB	Cancérologie
Mr Fabien DOGUET (<i>disponibilité</i>)	HCN	Chirurgie Cardio Vasculaire
Mr Jean DOUCET	SJ	Thérapeutique - Médecine interne et gériatrie
Mr Bernard DUBRAY	CHB	Radiothérapie
Mr Frank DUJARDIN	HCN	Chirurgie orthopédique - Traumatologique
Mr Fabrice DUPARC	HCN	Anatomie - Chirurgie orthopédique et traumatologique
Mr Eric DURAND	HCN	Cardiologie
Mr Bertrand DUREUIL	HCN	Anesthésiologie et réanimation chirurgicale
Mme Hélène ELTCHANINOFF	HCN	Cardiologie
Mr Manuel ETIENNE	HCN	Maladies infectieuses et tropicales
Mr Jean François GEHANNO	HCN	Médecine et santé au travail
Mr Emmanuel GERARDIN	HCN	Imagerie médicale
Mme Priscille GERARDIN	HCN	Pédopsychiatrie
M. Guillaume GOURCEROL	HCN	Physiologie
Mr Dominique GUERROT	HCN	Néphrologie
Mme Julie GUEUDRY	HCN	Ophthalmologie
Mr Olivier GUILLIN	HCN	Psychiatrie Adultes
Mr Claude HOUDAYER	HCN	Génétique
Mr Fabrice JARDIN	CHB	Hématologie
Mr Luc-Marie JOLY	HCN	Médecine d'urgence
Mr Pascal JOLY	HCN	Dermato – Vénérologie
Mme Bouchra LAMIA	Havre	Pneumologie
Mr Vincent LAUDENBACH	HCN	Anesthésie et réanimation chirurgicale
Mr Hervé LEFEBVRE	HB	Endocrinologie et maladies métaboliques
Mr Thierry LEQUERRE	HCN	Rhumatologie
Mme Anne-Marie LEROI	HCN	Physiologie
Mr Hervé LEVESQUE	HCN	Médecine interne
Mme Agnès LIARD-ZMUDA	HCN	Chirurgie Infantile
Mr Pierre Yves LITZLER	HCN	Chirurgie cardiaque
M. David MALTETE	HCN	Neurologie
Mr Christophe MARGUET	HCN	Pédiatrie
Mme Isabelle MARIE	HCN	Médecine interne
Mr Jean-Paul MARIE	HCN	Oto-rhino-laryngologie
Mr Loïc MARPEAU	HCN	Gynécologie - Obstétrique
Mr Stéphane MARRET	HCN	Pédiatrie
Mme Véronique MERLE	HCN	Epidémiologie

Mr Pierre MICHEL	HCN	Hépto-gastro-entérologie
M. Benoit MISSET (<i>détachement</i>)	HCN	Réanimation Médicale
Mr Marc MURAINÉ	HCN	Ophthalmologie
Mr Gaël NICOLAS	UFR	Génétique
Mr Christian PFISTER	HCN	Urologie
Mr Jean-Christophe PLANTIER	HCN	Bactériologie - Virologie
Mr Didier PLISSONNIER	HCN	Chirurgie vasculaire
Mr Gaëtan PREVOST	HCN	Endocrinologie
Mr Jean-Christophe RICHARD (<i>détachement</i>)	HCN	Réanimation médicale - Médecine d'urgence
Mr Vincent RICHARD	UFR	Pharmacologie
Mme Nathalie RIVES	HCN	Biologie du développement et de la reproduction
Mr Horace ROMAN (<i>détachement</i>)	HCN	Gynécologie - Obstétrique
Mr Jean-Christophe SABOURIN	HCN	Anatomie – Pathologie
Mr Mathieu SALAUN	HCN	Pneumologie
Mr Guillaume SAVOYE	HCN	Hépto-gastro-entérologie
Mme Céline SAVOYE-COLLET	HCN	Imagerie médicale
Mme Pascale SCHNEIDER	HCN	Pédiatrie
Mr Lilian SCHWARZ	HCN	Chirurgie Viscérale et Digestive
Mr Michel SCOTTE	HCN	Chirurgie digestive
Mme Fabienne TAMION	HCN	Réanimation médicale
Mr Luc THIBERVILLE	HCN	Pneumologie
M. Gilles TOURNEL	HCN	Médecine Légale
Mr Olivier TROST	HCN	Anatomie -Chirurgie Maxillo-Faciale
Mr Jean-Jacques TUECH	HCN	Chirurgie digestive
Mr Benoît VEBER	HCN	Anesthésiologie - Réanimation chirurgicale
Mr Pierre VERA	CHB	Biophysique et traitement de l'image
Mr Eric VERIN	Les Herbiers	Médecine Physique et de Réadaptation
Mr Eric VERSPYCK	HCN	Gynécologie obstétrique
Mr Olivier VITTECOQ	HC	Rhumatologie
Mr David WALLON	HCN	Neurologie
Mme Marie-Laure WELTER	HCN	Physiologie

MAITRES DE CONFERENCES DES UNIVERSITES – PRATICIENS HOSPITALIERS

Mme Najate ACHAMRAH	HCN	Nutrition
Mme Elodie ALESSANDRI-GRADT	HCN	Virologie
Mme Noëlle BARBIER-FREBOURG	HCN	Bactériologie – Virologie
Mr Emmanuel BESNIER	HCN	Anesthésiologie - Réanimation
Mme Carole BRASSE LAGNEL	HCN	Biochimie
Mr Gérard BUCHONNET	HCN	Hématologie
Mme Mireille CASTANET	HCN	Pédiatrie
Mme Nathalie CHASTAN	HCN	Neurophysiologie
Mr Damien COSTA	HCN	Parasitologie
Me Pierre DECAZES	CB	Médecine Nucléaire
M. Vianney GILARD	HCN	Neurochirurgie
Mr Serge JACQUOT	UFR	Immunologie
Mr Joël LADNER	HCN	Epidémiologie, économie de la santé
Mr Jean-Baptiste LATOUCHE	UFR	Biologie cellulaire
M. Florent MARGUET	HCN	Histologie
Mme Chloé MELCHIOR	HCN	Hépto-gastro-entérologie
M. Sébastien MIRANDA	HCN	Médecine Vasculaire
Mr Thomas MOUREZ (<i>détachement</i>)	HCN	Virologie
Mme Muriel QUILLARD	HCN	Biochimie et biologie moléculaire
Mme Laëtitia ROLLIN	HCN	Médecine du Travail
Mme Pascale SAUGIER-VEBER	HCN	Génétique
M. Abdellah TEBANI	HCN	Biochimie et Biologie Moléculaire
Mme Anne-Claire TOBENAS-DUJARDIN	HCN	Anatomie
Mr Julien WILS	HCN	Pharmacologie

PROFESSEUR AGREGE OU CERTIFIE

Mr Thierry WABLE	UFR	Communication
Mme Mélanie AUVRAY-HAMEL	UFR	Anglais

II - PHARMACIE

PROFESSEURS DES UNIVERSITES

Mr Jérémy BELLIEN (PU-PH)	Pharmacologie
Mr Thierry BESSON	Chimie Thérapeutique
Mr Jean COSTENTIN (Professeur émérite)	Pharmacologie
Mme Isabelle DUBUS	Biochimie
Mr Abdelhakim EL OMRI	Pharmacognosie
Mr François ESTOUR	Chimie Organique
Mr Loïc FAVENNEC (PU-PH)	Parasitologie
Mr Jean Pierre GOULLE (Professeur émérite)	Toxicologie
Mme Christelle MONTEIL	Toxicologie
Mme Martine PESTEL-CARON (PU-PH)	Microbiologie
Mr Rémi VARIN (PU-PH)	Pharmacie clinique
Mr Jean-Marie VAUGEOIS	Pharmacologie
Mr Philippe VERITE	Chimie analytique

MAITRES DE CONFERENCES DES UNIVERSITES

Mme Margueritta AL ZALLOUHA	Toxicologie
Mme Cécile BARBOT	Chimie Générale et Minérale
Mr Frédéric BOUNOURE	Pharmacie Galénique
Mr Thomas CASTANHEIRO MATIAS	Chimie Organique
Mr Abdeslam CHAGRAOUI	Physiologie
Mme Camille CHARBONNIER (LE CLEZIO)	Statistiques
Mme Elizabeth CHOSSON	Botanique
Mme Marie Catherine CONCE-CHEMTOB	Législation pharmaceutique et économie de la santé
Mme Cécile CORBIERE	Biochimie
Mme Nathalie DOURMAP	Pharmacologie
Mme Isabelle DUBUC	Pharmacologie
Mme Dominique DUTERTE- BOUCHER	Pharmacologie
Mr Gilles GARGALA (MCU-PH)	Parasitologie
Mme Nejla EL GHARBI-HAMZA	Chimie analytique
Mr Chervin HASSEL	Virologie
Mme Maryline LECOINTRE	Physiologie
Mme Hong LU	Biologie
Mme Marine MALLETER	Biologie Cellulaire

M. Jérémie MARTINET (MCU-PH)	Immunologie
M. Romy RAZAKANDRAINIBÉ	Parasitologie
Mme Tiphaine ROGEZ-FLORENT	Chimie analytique
Mr Mohamed SKIBA	Pharmacie galénique
Mme Malika SKIBA	Pharmacie galénique
Mme Christine THARASSE	Chimie thérapeutique

PROFESSEURS ASSOCIES

Mme Cécile GUERARD-DETUNCQ	Pharmacie officinale
Mme Caroline BERTOUX	Pharmacie officinale
M. Damien SALAUZE	Pharmacie industrielle

PAU-PH

M. Mikaël DAOUPHARS	Pharmacie
M. Pierre BOHN	Radiopharmacie

PROFESSEUR CERTIFIE

Mme Mathilde GUERIN	Anglais
----------------------------	---------

ASSISTANTS HOSPITALO-UNIVERSITAIRES

M. Eric BARAT	Pharmacie
M. Guillaume FEUGRAY	Biochimie Générale
M. Henri GONDÉ	Pharmacie
M. Paul BILLOIR	Hématologie
M. Romain LEGUILLON	Pharmacie
M. Thomas DUFLOT	Pharmacologie
Mme Alice MOISAN	Virologie

ATTACHES TEMPORAIRES D'ENSEIGNEMENT ET DE RECHERCHE

Mme Chaïma EZZINE	Pharmacologie
M. Abdelmounaim MOUHAJIR	Informatique Bio-Informatique
M. Olivier PERRUCHON	Pharmacognosie
M. Maxime GRAND	Bactériologie

LISTE DES RESPONSABLES DES DISCIPLINES PHARMACEUTIQUES

Mme Cécile BARBOT	Chimie Générale et minérale
Mr Thierry BESSON	Chimie thérapeutique
Mr Abdeslam CHAGRAOUI	Physiologie
Mme Elisabeth CHOSSON	Botanique
Mme Marie-Catherine CONCE-CHEMTOB	Législation et économie de la santé
Mme Isabelle DUBUS	Biochimie
Mr Abdelhakim EL OMRI	Pharmacognosie
Mr François ESTOUR	Chimie organique
Mr Loïc FAVENNEC	Parasitologie
Mme Christelle MONTEIL	Toxicologie
Mme Martine PESTEL-CARON	Microbiologie
Mr Mohamed SKIBA	Pharmacie galénique
Mr Rémi VARIN	Pharmacie clinique
M. Jean-Marie VAUGEOIS	Pharmacologie
Mr Philippe VERITE	Chimie analytique

III – MEDECINE GENERALE

PROFESSEUR MEDECINE GENERALE

Mr Jean-Loup **HERMIL** (PU-MG) UFR Médecine générale

MAITRE DE CONFERENCE MEDECINE GENERALE

Mr Matthieu **SCHUERS** (MCU-MG) UFR Médecine générale

PROFESSEURS ASSOCIES A MI-TEMPS – MEDECINS GENERALISTE

Mr Pascal **BOULET** UFR Médecine générale

Mr Emmanuel **LEFEBVRE** UFR Médecine Générale

Mme Elisabeth **MAUVIARD** UFR Médecine générale

Mme Yveline **SEVRIN** UFR Médecine générale

MAITRE DE CONFERENCES ASSOCIE A MI-TEMPS – MEDECINS GENERALISTES

Mr Julien **BOUDIER** UFR Médecine Générale

Mme Laëtitia **BOURDON** UFR Médecine Générale

Mme Elsa **FAGOT-GRIFFIN** UFR Médecine Générale

Mr Emmanuel **HAZARD** UFR Médecine Générale

ENSEIGNANTS MONO-APPARTENANTS

PROFESSEURS

Mr Paul MULDER (phar)	Sciences du Médicament
Mme Su RUAN (med)	Génie Informatique

MAITRES DE CONFERENCES

Mr Sahil ADRIOUCH (med)	Biochimie et biologie moléculaire (Unité Inserm 905)
Mr Jonathan BRETON (med)	Nutrition
Mme Gaëlle BOUGEARD-DENOYELLE (med)	Biochimie et biologie moléculaire (UMR 1079)
Mme Carine CLEREN (med)	Neurosciences (Néovasc)
M. Sylvain FRAINEAU (med)	Physiologie (Inserm U 1096)
Mme Pascaline GAILDRAT (med)	Génétique moléculaire humaine (UMR 1079)
Mr Nicolas GUEROUT (med)	Chirurgie Expérimentale
Mme Rachel LETELLIER (med)	Physiologie
Mr Antoine OUVRARD-PASCAUD (med)	Physiologie (Unité Inserm 1076)
Mr Frédéric PASQUET	Sciences du langage, orthophonie
Mme Anne-Sophie PEZZINO	Orthophonie
Mme Christine RONDANINO (med)	Physiologie de la reproduction
Mr Youssan Var TAN	Immunologie
Mme Isabelle TOURNIER (med)	Biochimie (UMR 1079)

DIRECTEUR ADMINISTRATIF : M. Jean-Sébastien VALET

HCN - Hôpital Charles Nicolle

HB - Hôpital de BOIS GUILLAUME

CB - Centre Henri Becquerel

CHS - Centre Hospitalier Spécialisé du Rouvray

CRMPR - Centre Régional de Médecine Physique et de Réadaptation

SJ - Saint Julien Rouen

Mise en garde

L'université de Rouen et l'UFR et Médecine et de Pharmacie de Rouen n'entendent donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions sont propres à leurs auteurs.

Table des matières

1.	Introduction	18
2.	Histoire du séquençage.....	19
2.1.	Séquençage de première génération.....	19
2.2.	Séquençage de deuxième et de troisième génération.....	22
3.	Principes techniques du séquençage haut débit.....	23
3.1.	Préparation des échantillons	23
3.2.	Séquençage par synthèse	30
3.3.	Nanopores/minion.....	33
3.4.	PacBio	33
4.	Éléments de bioinformatique	35
4.1.	Tag	35
4.2.	Contrôle qualité des séquences	35
4.3.	Assemblage et <i>mapping</i>	35
5.	Applications	37
5.1.	Taxonomie	37
5.2.	Métagénomique	38
5.3.	Analyse phylogénique.....	40
5.4.	Gènes de résistance	44
5.5.	Gènes de virulence	45
6.	Besoins informatique.....	46
6.1.	Bioinformatique sur site avec une machine dédiée.....	46
6.2.	Bioinformatique en <i>cloud-computing</i> sur plateforme publique.....	46
7.	Performances dans le diagnostic microbiologique.....	47
7.1.	Infections du compartiment vasculaire	47
7.2.	Infections du système nerveux central.....	50

7.3.	Infections ostéoarticulaires	52
7.4.	Infections de la peau et des tissus mous.....	54
7.5.	Infections de pied diabétique	54
7.6.	Collection liquidiennes et abcès profonds	55
7.7.	Infections respiratoires.....	55
7.8.	Infections ophtalmologiques.....	57
7.9.	Virus hépatotropes	58
7.10.	Infections par le VIH	59
7.11.	Infections virales tropicales	60
8.	Discussion	61
	Références Bibliographiques.....	65

Table des tableaux

Tableau I : Résumé des caractéristiques des différentes méthodes de séquençages disponibles (Mestan et al.).....	34
--	----

Table des figures

Figure 1 : Autoradiographe sur gel de polyacrylamide d'une réaction de séquençage selon la méthode de Maxam et Gilbert : une bande foncée sur la première colonne et faible sur la deuxième correspond à une adénine, l'inverse correspond à une guanine, une bande sur la troisième et la quatrième colonne correspond à une cytosine et une bande isolée sur la quatrième colonne correspond à une thymine (Maxam and Gilbert).

..... 20

Figure 2 : Autoradiographe sur gel de polyacrylamide du produit d'une réaction de séquençage selon la méthode de Sanger. Chaque bande correspond à une séquence se terminant par un acide aminé radioactif (Sanger et al.)..... 21

Figure 3 : Représentation schématique du séquençage Sanger par nucléotides terminateurs de chaîne marqués par des fluorochromes suivi d'une électrophorèse capillaire (Wangler and Bellen)..... 22

Figure 4 : Exemple de protocole d'extraction et d'enrichissement d'acides nucléiques microbiens totaux permettant le séquençage d'une série de prélèvements cliniques d'après (Miller et al.). 27

Figure 5 : Etapes successives de polissage et de ligation des adaptateurs à un fragment d'ADN. A : les fragments d'ADN coupés à l'emporte-pièce ne présentent pas d'ADN double brin à leurs extrémités. B : La T4 polymérase réplique les parties simple brin. C : l'extrémité 5' des brins d'ADN est phosphorylée. D : On ajoute une adénine en 3'. E : un adaptateur avec une thymine en 3' vient se fixer sur l'adénine libre. F : une ligase termine la procédure ("The Enzymes Making the Cut in NGS Library Preparation")..... 29

Figure 6 A : Fixation des segments d'ADN sur le transposome et ligation de séquences reconnues par les amorces. B : Amplification PCR avec des amorces contenant un *barcode*, des index sens et anti-sens et adaptateurs (*Illumina DNA Prep | Flexibility for Many Whole-Genome Sequencing Applications*). 30

Figure 7 : Fonctionnement schématique d'une réaction de séquence par nucléotides terminateurs de chaîne réversibles. a : Les fragments d'ADN double brin (rouges) sont additionnés des adaptateurs (bleus) et dénaturés. b : L'adaptateurs de chaque simple brin

se fixe sur la puce de séquençage, la synthèse du brin complémentaire est réalisé et l'adaptateur de celui se fixe à son tour sur la puce. c : Le premier brin est coupé et la réaction se répète (Bentley et al.).	32
Figure 8 : Représentation schématique d'un fragment d'ADN passant au travers d'un pore avec les signaux électriques correspondant aux bases ("How It Works").	33
Figure 9 : Mécanisme schématique du séquençage de troisième génération par la technologie PacBio (Rhoads and Au).	34
Figure 10 : Lecture de la séquence par la technologie PacBio (Rhoads and Au).	34
Figure 11 : Contrôle qualité de deux séquençages par Illumina via FastQC. A gauche avant élimination des séquences de mauvaise qualité, à droite après.	35
Figure 12 : Représentation schématique d'un assemblage utilisant un graphe de de Bruijn (Ponty).	36
Figure 13 : Exemple de graphique obtenu par Krona à partir d'une exploration métagénomique (Ondov et al.).	39
Figure 14 : Dendrogramme réalisé à partir de 4 souches de <i>Streptococcus agalactiae</i> CC-17 suspectes d'une épidémie hospitalière (en rouge) et de 25 souches représentatives des infections à <i>Streptococcus agalactiae</i> CC-17 en France (en noir).	43
Figure 15: Bactéries retrouvées lors de l'utilisation du séquençage haut débit sur des valves cardiaques prélevées d'endocardites infectieuses d'après (Cai et al.).	49
Figure 16 Exemple de flore plurimicrobienne retrouvée dans un abcès cérébral d'après (Lin et al.).	52
Figure 17 : Diagramme récapitulatif des essais permettant la validation de la phase préanalytique du diagnostic métagénomique.	63

1. Introduction

La microbiologie traditionnelle est une discipline utilisant la culture de micro-organismes dans des milieux adaptés. Dans les laboratoires d'analyses médicales, jusqu'à l'avènement de la biologie moléculaire, le seul moyen d'identifier et de caractériser un micro-organisme était de le cultiver puis d'effectuer des tests phénotypiques. Il s'est avéré que 99% des micro-organismes ne sont pas cultivables par des moyens connus (Schloss and Handelsman).

De ce constat, des techniques de biologie moléculaire se sont développées avec pour chef file la réaction de polymérisation en chaîne (PCR). Celle-ci a permis la réalisation de diagnostics ciblés sur des pathogènes non cultivables ou de culture fastidieuse. Le principal désavantage des techniques PCR est la nécessité de cibler les recherches et de faire une réaction par cible. Si des PCR multiplex ont vu le jour, diminuant le temps technique pour la réalisation de multiples PCR, elles n'en restent pas moins spécifiques (Zhu et al.). Aussi, l'utilisation du séquençage en routine, par méthode Sanger, est fastidieuse dans un laboratoire d'analyse médicale et ne peut être réalisée que sur un faible nombre de séquences courtes (Sanger et al.).

Depuis le milieu des années 2000, des méthodes de séquençage dites de deuxième puis de troisième génération, permettent l'analyse d'échantillons sans définir la cible à l'avance et sont très utilisées dans les études sur le microbiote notamment. Ces nouvelles méthodes de séquençage permettent aussi le séquençage de génomes entiers de façon rapide et à moindre coût. Ceci a permis une accélération du séquençage d'organismes vivants, le premier séquencé est le bactériophage ϕ X174 en 1973 (Sanger et al.), puis la première bactérie est *Haemophilus influenzae* en 1995 (Fleischmann et al.) pour un coût estimé de 800 000\$ de 1995. Enfin, le séquençage du premier génome humain n'est arrivé qu'en 2001. Aujourd'hui, le séquençage complet du génome d'une bactérie peut être réalisé en 24h pour un coût inférieur à 50€ en fonction de la technologie utilisée. Ces nouvelles informations ainsi obtenues donnent la possibilité d'effectuer des diagnostics plus précis.

Ce travail bibliographique a pour objectif d'inventorier ces nouvelles techniques de séquençages, de guider la mise en place de celles-ci dans un laboratoire de microbiologie clinique et d'aider à leur usage tant dans leurs indications que dans leur interprétation. Celui-ci servira de base au développement de l'usage du séquençage haut débit au laboratoire de microbiologie de l'Hôpital Charles Nicolle - CHU de Rouen.

2. Histoire du séquençage

2.1. Séquençage de première génération

2.1.1. Méthode de Maxam et Gilbert

En 1977, Allan Maxam et Walter Gilbert décrivent une nouvelle méthode de détermination de séquence d'un fragment d'ADN. Après marquage radioactif des brins d'ADN, ceux-ci sont soumis à une dégradation chimique spécifique d'une ou deux bases azotées.

Pour les purines, l'action du diméthyl-sulfate engendre une méthylation d'une base guanine ou adénine par brin, un traitement par la chaleur à pH neutre puis à pH alcalin entraîne le clivage du brin après la base méthylée. La guanine étant méthylée cinq fois plus rapidement que l'adénine, lors de la résolution électrophorétique, un brin terminé par une guanine marquera une bande plus foncée qu'un brin terminé par une adénine. Inversement, la forme méthylée de l'adénine étant moins stable que celle de la guanine, un traitement plus doux après méthylation à pH légèrement acide suivi du même traitement à pH alcalin clivera préférentiellement les brins d'ADN au niveau des adénines. Cette fois ci, lors de la résolution électrophorétique, les bandes foncées correspondront aux adénines.

Pour les pyrimidines, l'hydrazine permet de cliver indifféremment l'ADN au niveau des thymines ou des cytosines. En présence de NaCl à 2 mol/L, l'hydrazine réagit préférentiellement avec les cytosines. Lors de la résolution électrophorétique, le produit de la première réaction formera des bandes au niveau des thymines et des cytosines alors que celui de la deuxième en formera qu'au niveau des cytosines. La présence de deux bandes correspondra donc à une cytosine alors qu'une seule bande orientera vers une thymine (Maxam and Gilbert).

Cette méthode étant fastidieuse, on lui préférera celle proposée par Sanger la même année.

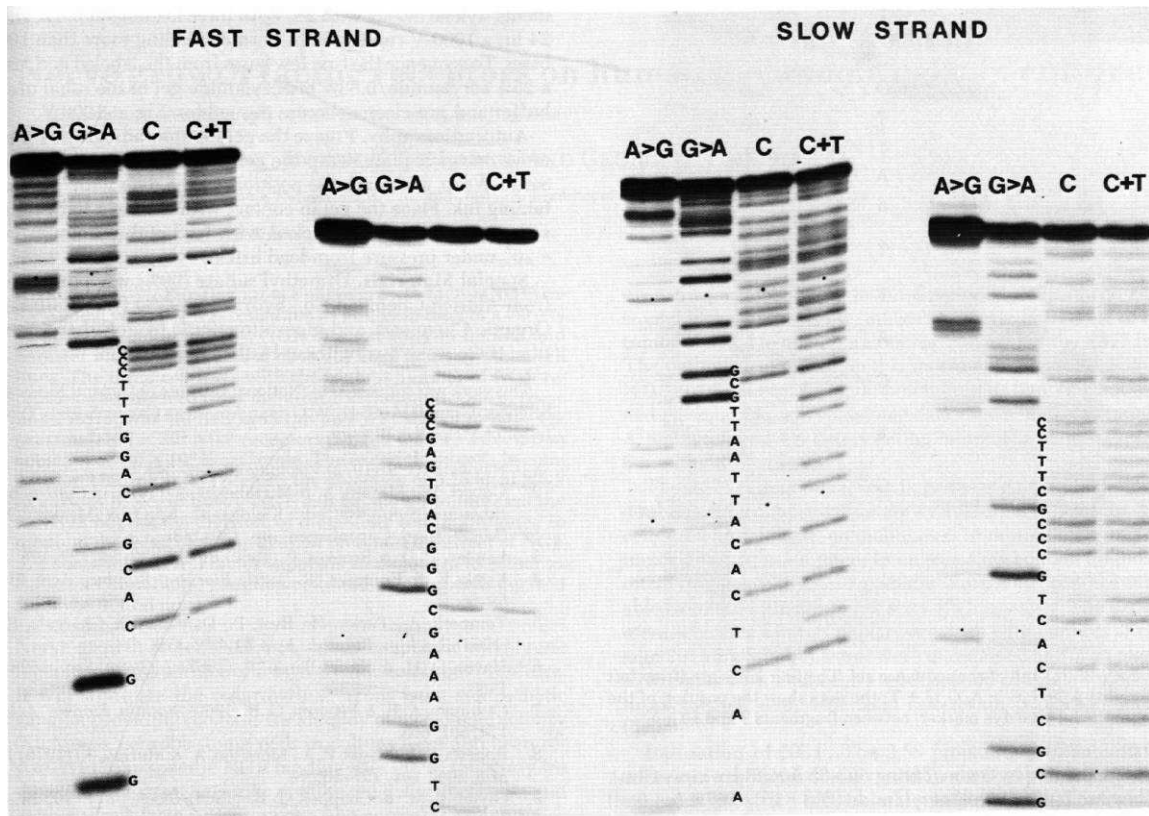


Figure 1 : Autoradiographie sur gel de polyacrylamide d'une réaction de séquençage selon la méthode de Maxam et Gilbert : une bande foncée sur la première colonne et faible sur la deuxième correspond à une adénine, l'inverse correspond à une guanine, une bande sur la troisième et la quatrième colonne correspond à une cytosine et une bande isolée sur la quatrième colonne correspond à une thymine (Maxam and Gilbert).

2.1.2. Méthode de Sanger

En 1977 est publié l'article princeps décrivant le séquençage Sanger (Sanger et al.). Cette publication décrit alors une réaction de séquençage selon plusieurs principes : une didéoxythymidine triphosphate (ddTTP) incorporé par une ADN polymérase I bactérienne à une chaîne nucléotidique entraîne l'impossibilité d'incorporer un autre nucléotide du fait de l'absence de groupe 3'OH, la chaîne se termine donc à une position où une déoxythymidine aurait dû être incorporée. Si on met en présence un brin matrice avec une amorce appropriée, des déoxynucléosides triphosphate (dATP, dTTP, dCTP et dGTP), une ADN polymérase I et des ddTTP marqués par du ^{32}P radioactif, on obtient un mélange de fragments d'ADN ayant la même extrémité 5' avec des résidus didéoxythymidine à l'extrémité 3'. En séparant ces fragments d'ADN par une électrophorèse sur gel de polyacrilamide, on obtient la distribution des résidus thymidine dans le brin d'ADN synthétisé. Selon le même principe, on peut obtenir en parallèle la distribution des résidus adénine, cytosine et guanine du fragment d'ADN et lire la séquence (Figure 2).

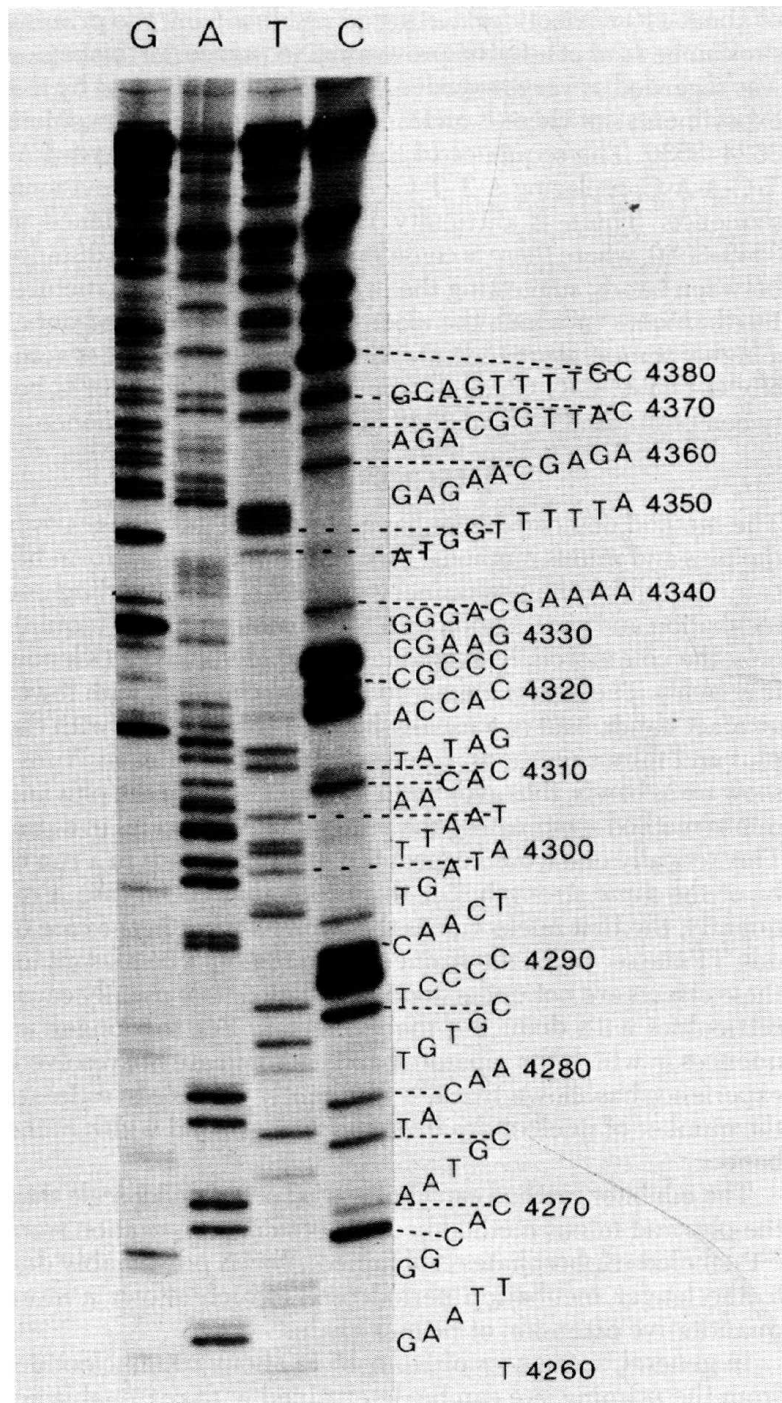


Figure 2 : Autoradiographe sur gel de polyacrylamide du produit d'une réaction de séquençage selon la méthode de Sanger. Chaque bande correspond à une séquence se terminant par un acide aminé radioactif (Sanger et al.).

Aujourd'hui la méthode de Sanger a été largement améliorée et se sert de didéoxynucléotides triphosphate marqués par des fluorochromes spécifiques à chaque base azotée et d'une électrophorèse capillaire. Ces deux innovations principales permettent d'accélérer le séquençage et d'agrandir la taille des fragments, typiquement jusqu'à 500 à 800 paires de bases (Willey et al.).

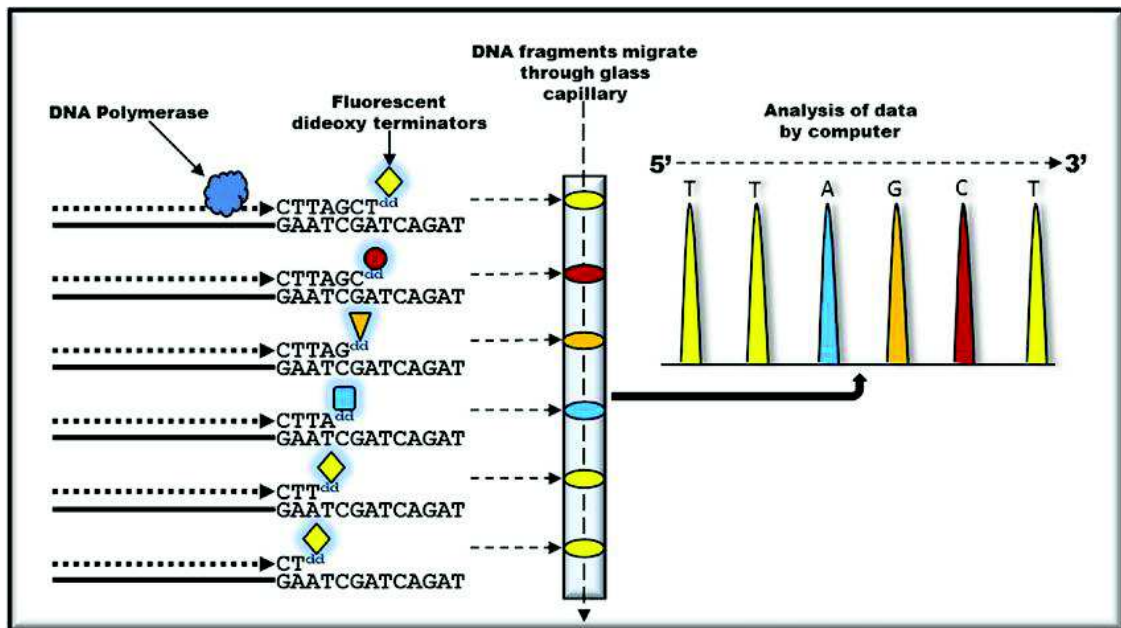


Figure 3 : Représentation schématique du séquençage Sanger par nucléotides terminateurs de chaîne marqués par des fluorochromes suivi d'une électrophorèse capillaire (Wangler and Bellen).

2.2. Séquençage de deuxième et de troisième génération

En 2005, la société 454 (aujourd'hui rachetée par Roche) commercialise une solution de pyroséquençage et en 2006, Solexa (aujourd'hui rachetée par Illumina) sort son premier automate de séquençage par synthèse. C'est le début du séquençage haut débit avec le séquençage de deuxième génération.

En 2010, Ion Torrent Systems Inc. commercialise son automate de séquençage haut débit basé sur sa propre technologie. Cette technologie se situe entre la deuxième et la troisième génération.

En 2011, PacBio commercialise sa solution de séquençage et en 2014, Oxford nanopore commercialise le Minion, un automate portable de séquençage par nanopore. Ces deux technologies représentent la troisième génération.

Ces nouvelles technologies de séquençages ont permis de réduire drastiquement le temps et le coût du séquençage d'ADN, leurs différents principes de fonctionnement sont expliqués en détail dans le chapitre dédié.

Dans la suite de ce document, tout ce qui concerne le séquençage haut débit sera décrit sous l'angle de la microbiologie.

3. Principes techniques du séquençage haut débit

3.1. Préparation des échantillons

3.1.1. Phase préanalytique

Dans la plupart des laboratoires de microbiologie, les prélèvements cliniques susceptibles de donner lieu à un diagnostic moléculaire sont régulièrement congelés à la réception puis décongelés lors de l'analyse. Ce cycle de congélation-décongélation pourrait avoir un impact sur la sensibilité des méthodes de diagnostic basées sur le séquençage haut débit. S'il faut tout de même l'évaluer au cas par cas en fonction de la technologie utilisée, nous supposons que l'impact de la phase pré-analytique est transposable en majorité des méthodes PCR aux méthodes par séquençage haut débit. Certaines technologies utilisant des séquences longues nécessitent tout de même de couper le moins possible les brins d'acides nucléiques et donc requièrent une attention particulière lors de la phase préanalytique.

La plupart des bactéries et virus restent détectables après un cycle de congélation-décongélation sans altération du seuil de détection (Murugesan et al.). Cependant, pour les virus à ARN, plus fragiles, les rendements sont plus dépendants des conditions de prélèvements et les cycles de congélation-décongélation ont un impact dans certains cas (Holohan et al.). La congélation et décongélation des prélèvements peut avoir un effet différent chez les parasites, en effet, ces cycles peuvent altérer la sensibilité des techniques de biologie moléculaire chez *P. falciparum* (Pett et al.) mais l'améliorer chez *Echinococcus multilocularis* ou *Entamoeba dispar* et *E. histolytica* par exemple (Klein et al.; Cnops and Esbroeck).

3.1.2. Contrôle interne

Lors de la préparation des échantillons, avant l'extraction des acides nucléiques, il est possible d'ajouter aux prélèvements un contrôle interne constitué d'une espèce microbienne strictement non pathogène pour l'homme comme un virus bactériophage. La présence de ce bactériophage lors de l'analyse permettra de valider l'extraction et la préparation de la librairie. On peut ainsi ajouter un bactériophage à ADN et un bactériophage à ARN pour valider les deux parties de la préparation (Miller et al.).

3.1.3. Extraction d'ADN bactérien et fongique

Les différentes méthodes de lyse (ou de pré-extraction) utilisées couramment sont la lyse mécanique, par sonication ou par broyage à billes, la lyse chimique, par détergents et la lyse enzymatique avec du lysozyme ou de la protéinase K par exemple (Di Bella et al.).

La qualité des résultats de séquençage dépend de la méthode d'extraction, pour apprécier correctement la diversité des espèces microbiennes présentes dans un échantillon, il est nécessaire d'extraire le matériel génétique de toutes ces espèces. Différentes méthodes d'extractions existent et diffèrent notamment par le procédé de lyse et de séparation de l'ADN.

Bien que ces différences soient notables, des études comparant plusieurs kits d'extraction d'ADN pour étudier le microbiote fécal retrouvent des différences de microbiomes faibles et inférieures aux différences entre les individus (Wagner Mackenzie et al.; Kennedy et al.). Les différences les plus importantes semblent se produire entre les protocoles utilisant ou non une étape de lyse mécanique par billes, cette dernière améliorant notamment la lyse des bactéries à gram positif (de Boer et al.) et des champignons (Scharf et al.).

La lyse mécanique par billes doit cependant être maîtrisée, une lyse trop violente peut dégrader le matériel génétique et réduire le rendement de l'extraction (Rosenbaum et al.).

Selon les prélèvements cliniques, les protocoles d'extraction diffèrent et peuvent influencer le résultat. Certains prélèvements comme des morceaux de tissu sont plus difficile à lyser et nécessitent une lyse plus longue (Helmersen and Aamot). D'autres comme les hémocultures ou les abcès contiennent respectivement volontiers des inhibiteurs de PCR ou de l'ADN humain en grande quantité et nécessitent, soit une dilution des inhibiteurs, soit une purification de l'ADN microbien, souvent par déplétion en ADN humain (Anson et al.). La présence d'inhibiteurs de PCR pourrait n'influer que sur les étapes de PCR dans les protocoles qui y font appel et ne pas influencer sur le séquençage en lui-même.

Il n'existe actuellement pas de consensus permettant de définir un protocole d'extraction d'ADN optimal (Lim et al.). Chaque laboratoire doit s'assurer de la qualité de son extraction.

Suite à l'extraction de l'ADN, il est utile de mesurer la quantité extraite pour s'assurer du bon déroulement de l'extraction. Cela permet, en cas de pertes, de ne pas lancer une réaction de séquence coûteuse sur un échantillon ne contenant pas assez de matériel génétique. Cette mesure peut être réalisée par spectrophotométrie. En effet, les automates de séquençage nécessitent une certaine quantité d'acides nucléiques pour fonctionner et, lors du multiplexage de plusieurs échantillons dans un même *run*, la quantité d'ADN doit être équivalente dans tous les échantillons pour qu'ils soient séquencés en quantité équivalente. Selon les kits de préparation de librairie et les automates, la quantité d'ADN requise varie de 1ng à 500ng, celle-ci est donnée dans les documentations des kits.

3.1.4. Extraction d'acides nucléiques viraux

Les virus pouvant encoder leur génome sur de l'ADN aussi bien que sur de l'ARN, une méthode souhaitant détecter la majorité des virus doit se doter d'une extraction permettant d'isoler à la fois l'ARN et l'ADN.

Pour l'extraction d'ARN, se pose le problème des RNases ubiquitaires présentes dans le contenu intracellulaire libéré lors de la lyse, celles-ci dégradent rapidement l'ARN que l'on cherche à extraire (Di Bella et al.).

Une méthode d'extraction utilisant le thiocyanate de guanidinium permet de lyser les cellules et d'inactiver les RNases en même temps. Historiquement, il est possible de réaliser en même temps une extraction liquide de l'ARN avec du phénol et du chloroforme. L'ADN et les protéines se retrouvent dans la phase phénolique et l'ARN se retrouve dans la phase aqueuse (Chomczynski and Sacchi).

Aujourd'hui, des kits commerciaux d'extractions permettent une isolation de l'ARN sans extraction au phénol-chloroforme par liaison spécifique de l'ARN sur la membrane de billes en silice. Des étapes d'élimination de l'ADN restant par DNase et de lavage peuvent ensuite être réalisées pour purifier l'ARN extrait (Liana E. Kafetzopoulou et al.).

Certains kits d'extraction virale permettent d'extraire les acides nucléiques viraux totaux qu'ils soient ADN ou ARN. Plusieurs sont comparés entre eux et leurs performances comparées diffèrent en fonction des publications et des virus servant d'étalons. Suite à l'extraction d'ARN, il faut effectuer une transcription inverse pour obtenir l'ADN complémentaire requis par les automates de séquençage.

Le protocole TUViD-VM développé par l'institut Robert Koch de Berlin pour augmenter le seuil de détection de pathogènes viraux montre de bons résultats pour l'isolation de quatre virus (virus Sendai, virus de la vaccine, virus de la grippe A et Réovirus) couvrant des virus à ADN et à ARN enveloppés ou non. Il combine plusieurs étapes de centrifugation et d'ultracentrifugation, une digestion de l'ADN de l'hôte puis une extraction au TRIzol et enfin une amplification spécifique par RT-PCR avec des amorces aléatoires. Par rapport à des méthodes d'extraction classiques, après séquençage, le nombre de *reads* pouvait passer de 13/1 000 000 à 500 000/1 000 000. Ces bons résultats sont à mettre en perspective avec le temps technique nécessaire à l'extraction dans ce protocole, autour de 12h, ce qui paraît difficile à mettre en place en routine (Kohl et al.).

3.1.5. Extraction d'acides nucléiques microbiens totaux

Certains kits commerciaux sont capables d'extraire à la fois l'ARN et l'ADN des micro-organismes. Une étude des rendements d'extraction de ces kits est préalablement nécessaire à leur utilisation.

3.1.6. Enrichissement des acides nucléiques microbiens

On peut enrichir le prélèvement en ADN microbien par rapport à l'ADN de l'hôte par lyse spécifique de ce dernier par différentes méthodes. La déplétion spécifique via des anticorps de l'ADN contenant des îlots CpG méthylées, provenant de l'hôte, en est un bon moyen (Miller et al.). Des kits commerciaux existent pour cette application (26). Ces différents kits ont des performances très dépendantes des prélèvements et notamment de la quantité d'ADN de l'hôte présent, leur efficacité d'enrichissement d'ADN bactérien peut varier de $\times 10$ à $\times 10\,000$ (Heravi et al.; Thoendel et al.). Une autre méthode consiste à profiter de la sensibilité des cellules humaines à la saponine. En effet, après traitement par saponine puis choc osmotique, les cellules humaines lysent, une centrifugation permet de récupérer un culot bactérien en éliminant le surnageant contenant le lysat de cellules humaines. Cette méthode réduit la quantité d'ADN humain d'environ 1000 fois (Charalampous et al.).

En ce qui concerne l'ARN, celui-ci peut être enrichi en éliminant l'ADN via des DNases. Si l'on veut enrichir l'ARN et l'ADN d'une extraction d'acides nucléiques totaux, on peut séparer l'extrait en deux et enrichir les deux parties séparément (Miller et al.).

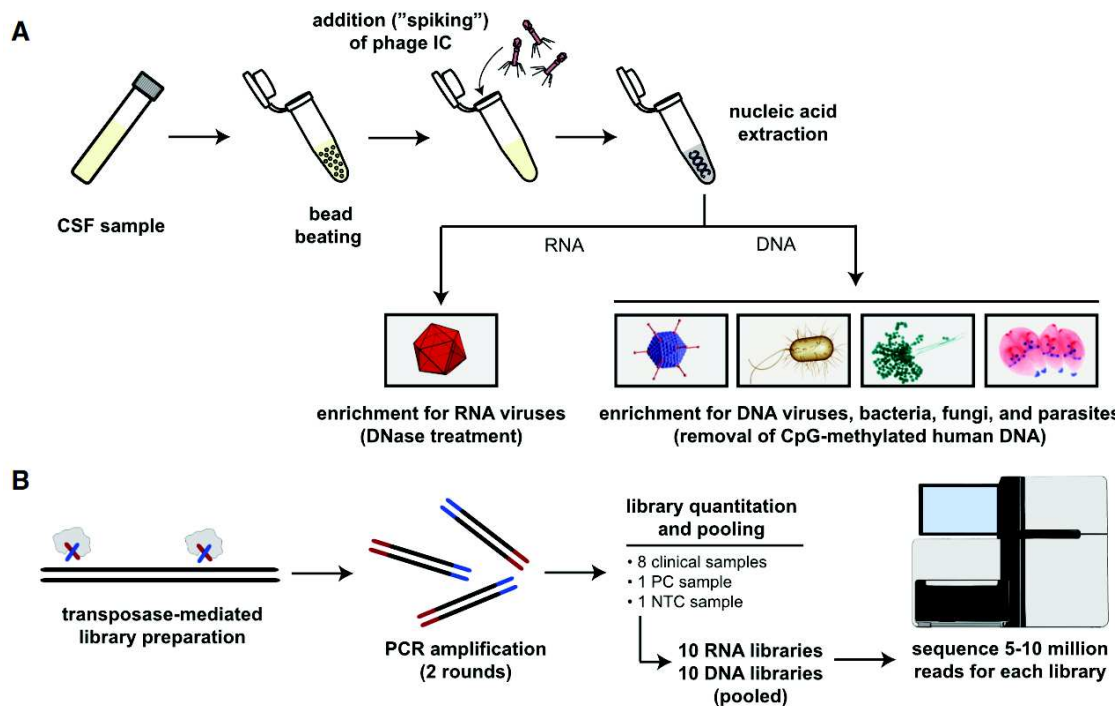


Figure 4 : Exemple de protocole d'extraction et d'enrichissement d'acides nucléiques microbiens totaux permettant le séquençage d'une série de prélèvements cliniques d'après (Miller et al.).

3.1.7. Fragmentation de l'ADN

La plupart des automates de séquençage haut débit séquençant des brins d'ADN courts, de 100 à 5000 bases selon les méthodes de séquençage, il faut fragmenter les brins longs provenant de l'étape d'extraction. Plusieurs méthodes existent, notamment mécaniques et enzymatiques, les meilleures donnent une fragmentation la plus aléatoire possible avec un chevauchement des séquences (Head et al.).

Parmi les techniques de fragmentation mécanique, la nébulisation permet de fragmenter l'ADN en le forçant à passer rapidement à travers un orifice à l'aide d'un gaz sous pression. La fragmentation par cisaillement hydrodynamique utilise un principe similaire à la nébulisation en remplaçant le gaz par un liquide. Ces deux méthodes permettent une fragmentation aléatoire (Thorstenson et al.). La sonication utilise des ultrasons pour fragmenter l'ADN par deux mécanismes principaux, un phénomène de cavitation et un effet thermique (Elsner and Lindblad). Longtemps pensée comme aléatoire, cette fragmentation par ultrasons découperait plus spécifiquement l'ADN au niveau de certaines séquences (Grokhovsky et al.).

Suite à une fragmentation mécanique, les cassures ne se réalisent pas strictement entre deux nucléotides. Les extrémités des brins d'ADN sont donc dégradées et

nécessitent une étape de réparation et de polissage comprise dans les kits de préparation de librairie des fournisseurs (Head et al.).

Il existe aussi des méthodes de fragmentation enzymatiques, plus récentes, utilisant différentes enzymes. Certaines clivent l'ADN de façon aléatoire tandis que d'autres sont spécifiques de très courtes séquences. Comme les méthodes de fragmentation mécaniques, certaines méthodes enzymatiques entraîneraient une dégradation des extrémités des fragments (N et al.).

Une méthode de fragmentation par micro-ondes a aussi été développée mais engendre encore une diminution de la qualité du séquençage (Yang and Hang).

3.1.8. *Ligation*

Suite à la fragmentation, les brins d'ADN ainsi formés sont liés à des adaptateurs, ce sont des courtes séquences servant de fixation au support sur l'automate, d'identification du brin et d'amorces pour les réactions de séquençage. Ces adaptateurs sont spécifiques d'un automate de séquençage haut débit et sont fournis par les industriels.

En plus des adaptateurs, on peut lier de courtes séquences d'identification appelées *tags* ou *barcode* qui permettent de marquer les brins d'ADN comme appartenant à un échantillon. Cette méthode permet de séquencer plusieurs échantillons en une fois et de séparer les résultats obtenus en fonction des *tags* des *reads*.

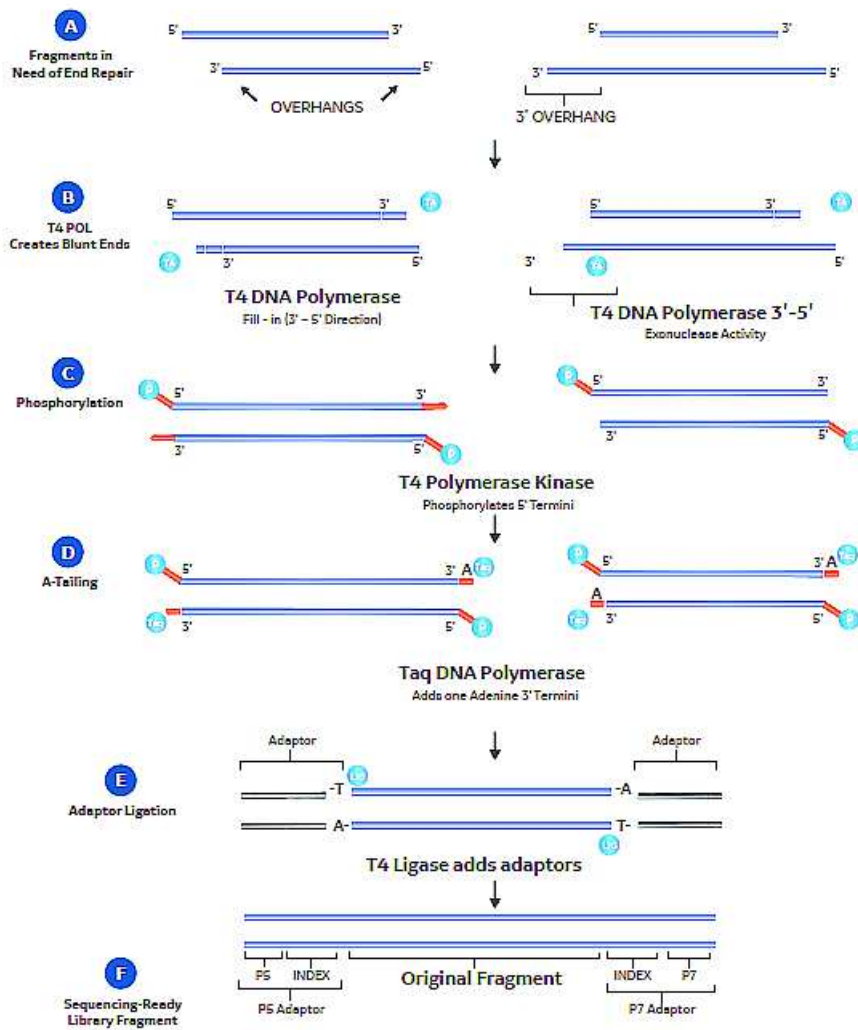


Figure 5 : Etapes successives de polissage et de ligation des adaptateurs à un fragment d'ADN. A : les fragments d'ADN coupés à l'emporte-pièce ne présentent pas d'ADN double brin à leurs extrémités. B : La T4 polymérase réplique les parties simple brin. C : l'extrémité 5' des brins d'ADN est phosphorylée. D : On ajoute une adénine en 3'. E : un adaptateur avec une thymine en 3' vient se fixer sur l'adénine libre. F : une ligase termine la procédure ("The Enzymes Making the Cut in NGS Library Preparation").

3.1.9. Tagmentation

La tagmentation (Figure 6) permet de fragmenter l'ADN, de lier les adaptateurs et d'ajouter les *tags* en une seule étape. On utilise des enzymes appelées transposases, couplées en homodimères pour former des transposomes. La plus utilisée est une enzyme Tn5 génétiquement modifiée pour la rendre non spécifique d'une séquence et augmenter son activité.

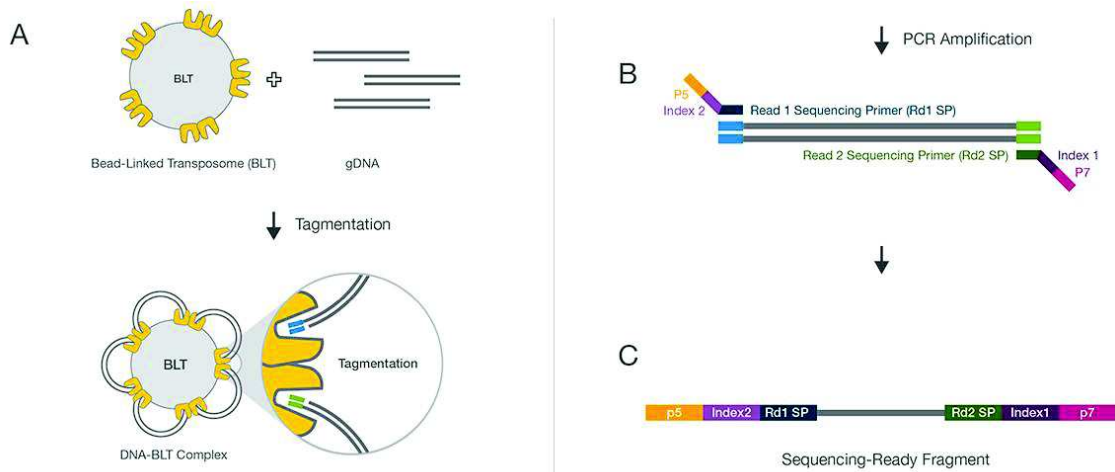


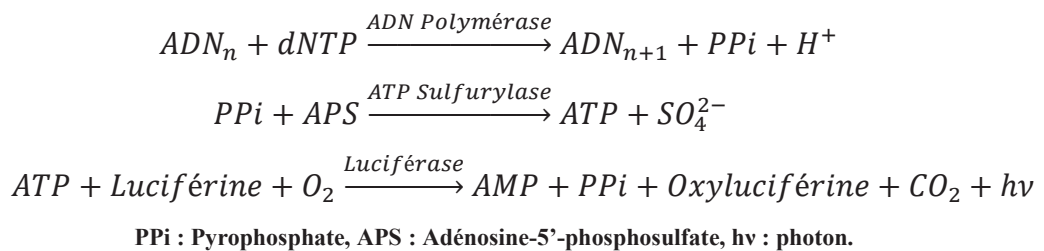
Figure 6 A : Fixation des segments d'ADN sur le transposome et ligation de séquences reconnues par les amorces. B : Amplification PCR avec des amorces contenant un *barcode*, des index sens et anti-sens et adaptateurs (*Illumina DNA Prep | Flexibility for Many Whole-Genome Sequencing Applications*).

3.2. Séquençage par synthèse

3.2.1. Pyroséquençage

Le pyroséquençage notamment développé et commercialisé par 454 puis racheté plus tard par Roche est une des premières méthodes appliquées au séquençage de nouvelle génération. Initialement adapté à une séquence ADN unique comme le séquençage Sanger, il a évolué vers une méthode de séquençage haut débit permettant le séquençage massivement parallèle de nombreux brins.

Après PCR ciblant la région à séquencer, les brins sont liés à des billes magnétiques via une interaction biotine-streptavidine. La méthode se base ensuite sur une suite de réactions chimiques :



Lorsqu'un nucléotide est incorporé, le jeu de réactions chimiques produit un photon qui est détecté. Les dNTP sont ajoutés séquentiellement dans le milieu réactionnel puis lavés, de telle façon que lorsqu'un nucléotide est incorporé, on sait quel dNTP était présent et donc on sait quel nucléotide était sur le brin matrice à séquencer. Si plusieurs nucléotides identiques sont incorporés, l'intensité lumineuse est proportionnelle à leur nombre.

Pour une utilisation du pyroséquençage en séquençage nouvelle génération, on utilise cette technique massivement en parallèle avec chaque brin couplé à une bille et un détecteur individuel (Harrington et al.).

3.2.2. Séquençage par nucléotides terminateurs de chaîne réversibles

Le séquençage par nucléotides terminateurs de chaîne réversibles est la technologie utilisée par les automates Illumina.

Une fois la librairie réalisée, les fragments simple brin d'ADN de quelques centaines de bases à séquencer sont hybridés à une séquence ADN attachée à la puce de séquençage puis sont répliqués en prenant les séquences attachées pour amorce. S'en suit une dénaturation qui permet au deuxième adaptateur de s'hybrider à la deuxième séquence attachée. Le brin est répliqué dans l'autre sens puis les opérations sont répétées en boucle. De ce fait, chaque fragment à séquencer est répliqué dans les deux sens.

Lors de la réplication, les nucléotides utilisés sont modifiés pour être des nucléotides terminateurs de chaîne réversibles couplés à un fluorochrome, c'est-à-dire qu'une fois incorporés, ils bloquent la réplication de la même façon que lors d'un séquençage Sanger mais qu'une fois incorporés, le fluorochrome peut être clivé, ce qui libère la fluorescence et permet de poursuivre la réplication.

Pour séquencer, les nucléotides sont donc incorporés un par un ; lorsqu'un l'est, l'élongation s'arrête, une modification du milieu entraîne le clivage du fluorochrome correspondant à la base azotée incorporée, la fluorescence est lue, puis l'élongation reprend. La séquence correspond à la suite des fluorescences lues par le détecteur.

Le séquençage simultané de très nombreux fragments courts permet de couvrir une grande partie du génome (Willey et al.; Bentley et al.; *Illumina Sequencing by Synthesis*).

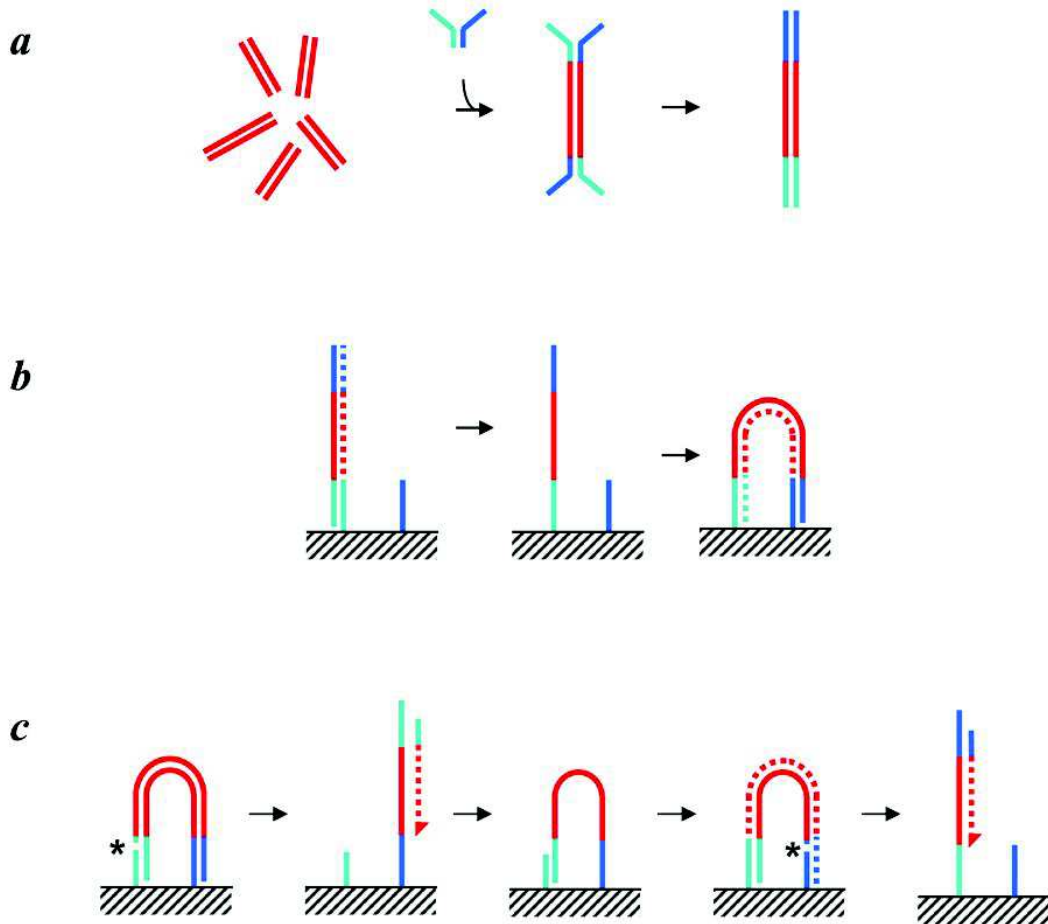


Figure 7 : Fonctionnement schématique d'une réaction de séquence par nucléotides terminateurs de chaîne réversibles. a : Les fragments d'ADN double brin (rouges) sont additionnés des adaptateurs (bleus) et dénaturés. b : L'adaptateur de chaque simple brin se fixe sur la puce de séquençage, la synthèse du brin complémentaire est réalisée et l'adaptateur de celui se fixe à son tour sur la puce. c : Le premier brin est coupé, et la réaction se répète (Bentley et al.).

3.2.3. Séquençage Ion Torrent

La technologie Ion Torrent se base aussi sur du séquençage par synthèse. De façon similaire, les fragments d'ADN à séquencer sont fixés à une matrice par hybridation des adaptateurs et s'en suit une réplication. Dans le milieu, sont ajoutés puis retirés séquentiellement les nucléotides individuellement. Lorsqu'un nucléotide est incorporé, la réaction libère un ion H^+ qui est détecté par un pH-mètre en semi-conducteur. Le signal est attribué à la base azotée présente dans le milieu au moment de la détection. Si plusieurs nucléotides identiques se suivent, autant sont incorporés et le signal est théoriquement proportionnel à leur nombre, ce qui permet de le déduire (Rothberg et al.).

3.3. Nanopores/minion

Utilisée par les automates MinIon, GridIon et PromethIon, la technologie de séquençage par nanopores permet un séquençage en temps réel d'un fragment d'ADN par passage au travers d'un pore protéique.

En passant par le pore, le fragment d'ADN engendre une modification du courant électrique au travers de celui-ci dont la forme et l'intensité sont prédictives de chaque base et définies par un algorithme d'intelligence artificielle. La séquence des signaux électriques est ensuite convertie en la séquence du fragment passant ("How It Works"; Lu et al.).

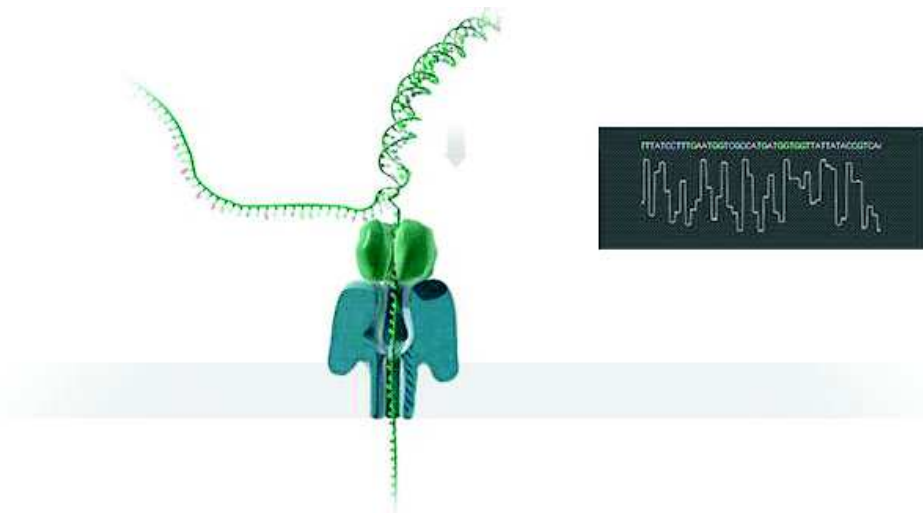


Figure 8 : Représentation schématique d'un fragment d'ADN passant au travers d'un pore avec les signaux électriques correspondant aux bases ("How It Works").

3.4. PacBio

La technologie de Pacific Bioscience se base le séquençage dit de troisième génération qui se distingue de la génération précédente par la lecture en temps réel de la séquence ; les technologies de seconde génération lisent la séquence base par base avec une pause entre chaque cycle. Au niveau des performances comparées à la seconde génération, cette méthode permet des *reads* plus longs (de l'ordre de 10k bases) et des *run* plus courts au prix d'un débit plus faible et d'un coût plus élevé.

Initialement, l'ADN double brin à séquencer (en jaune et violet sur la Figure 9) est transformé en ADN simple brin circulaire par l'adjonction d'un adaptateur en épingle à cheveux (en vert sur la Figure 9) à chaque extrémité. L'ADN polymérase (en gris sur la Figure 9) fixée au fond du puits de réaction initie la réplication sur une des boucles de l'adaptateur et réplique en continu l'ADN circulaire. Les bases incorporées sont

marquées par un fluorochrome qui libère une pulsation lumineuse spécifique lors de son incorporation. La lecture du « film » donne la séquence nucléotidique.

Les *reads* très longs engendrés par cette technologie permettent notamment de résoudre des régions génomiques comprenant de nombreuses répétitions (Rhoads and Au).

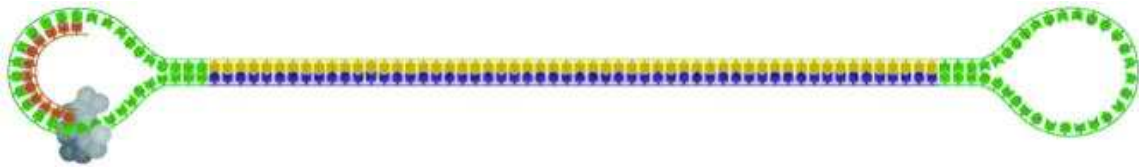


Figure 9 : Mécanisme schématisé du séquençage de troisième génération par la technologie PacBio (Rhoads and Au)

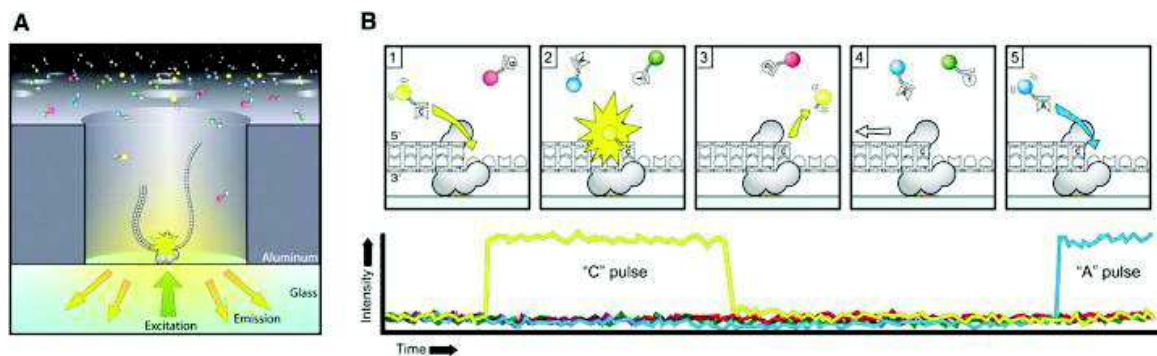


Figure 10 : Lecture de la séquence par la technologie PacBio (Rhoads and Au)

Tableau I : Résumé des caractéristiques des différentes méthodes de séquençages disponibles (Mestan et al.).

	<i>Sanger</i>	<i>Illumina/IonTorrent</i>	<i>Minion/PacBio</i>
<i>Technologie</i>	Séquençage par terminaison de chaînes	Séquençage par synthèse	Séquençage en temps réel
<i>Résolution</i>	Moyenne sur de nombreuses copies d'ADN	Moyenne sur de nombreuses copies d'ADN	Molécule d'ADN unique
<i>Précision</i>	Elevée	Elevée	Faible
<i>Longueur des reads</i>	800-1000 pb	100-150 pb	>1000pb
<i>Débit</i>	Faible	Elevé	Elevé
<i>Coût</i>	Coût par base élevé Coût par <i>run</i> faible	Coût par base faible Coût par <i>run</i> élevé	Coût par base faible Coût par <i>run</i> élevé
<i>Délai de résultat</i>	Heures	Jours	< 1 jour
<i>Préparation</i>	Modérément complexe	Complexe	Variable
<i>Analyse de données</i>	Routinière	Complexe	Complexe

4. Eléments de bioinformatique

4.1. Tag

Lors de la préparation de la librairie, quand plusieurs échantillons doivent être séquencés lors du même *run*, il est nécessaire d'identifier les séquences à leur prélèvement d'origine. Pour cela, pendant l'ajout des adaptateurs, on ajoute aussi des séquences servant d'étiquettes (*tag*) uniques à chaque prélèvement.

Ces étiquettes sont ensuite séquencées et se retrouvent à l'extrémité de chaque *read*. Il faut donc retirer les adaptateurs et les *tags* avant de faire les analyses bio-informatiques suivantes. Des programmes simples comme Trimmomatic permettent de le faire (Bolger et al.).

4.2. Contrôle qualité des séquences

Les différentes méthodes de séquençage haut débit produisent des erreurs qui peuvent interférer avec la bonne analyse des résultats. C'est pourquoi à chaque base incorporée est assigné un score de qualité. Certains programmes dont le plus utilisé est FastQC (*Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data*) permettent de visualiser plusieurs indicateurs de qualité d'un *run*. Ensuite, on peut éliminer les bases et les séquences de mauvaise qualité avec des logiciels de troncature comme Trimmomatic.

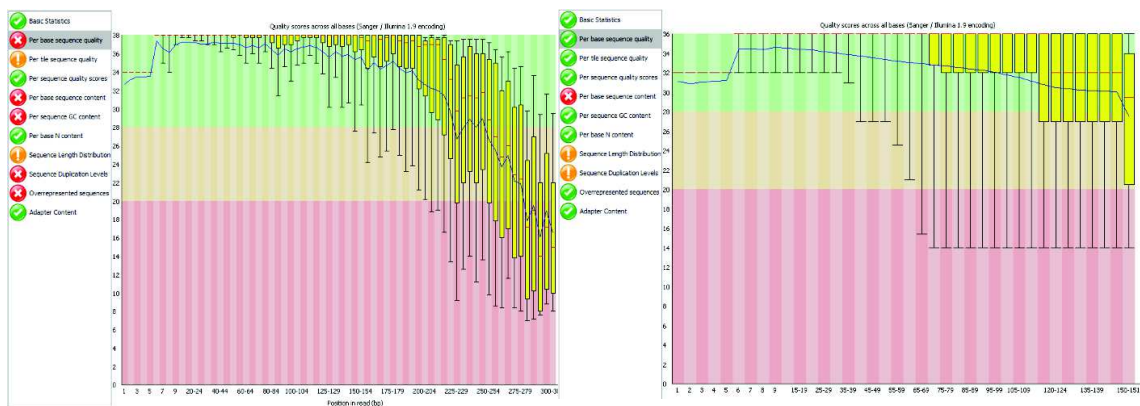


Figure 11 : Contrôle qualité de deux séquençages par Illumina via FastQC. À gauche avant élimination des séquences de mauvaise qualité, à droite après.

4.3. Assemblage et *mapping*

A partir des *reads* obtenus lors du séquençage, on peut assembler des *contigs* par chevauchement des extrémités des *reads*. Ces *contigs* sont des séquences plus longues, voire, dans des conditions parfaites, la totalité du génome d'une souche. Les deux

programmes les plus utilisés sont SPADES (Bankevich et al.) (et metaSPADES dans sa version adaptée à des échantillons plurimicrobiens (Nurk et al.)) et Velvet (Zerbino and Birney) (et metaVelvet dans sa version adaptée à des échantillons plurimicrobiens (Namiki et al.)). Cette opération est appelée assemblage *de novo*.

La plupart des algorithmes d'assemblage, dont les deux précédemment cités, fonctionnent à partir de graphes de de Bruijn. Les *reads* sont segmentés en « k-mers » d'un nombre fixe de bases k puis chaque k-mer est associé avec les k-mer partageant k-1 bases identiques. En aboutant tous les k-mers de cette façon, on reconstitue le génome.

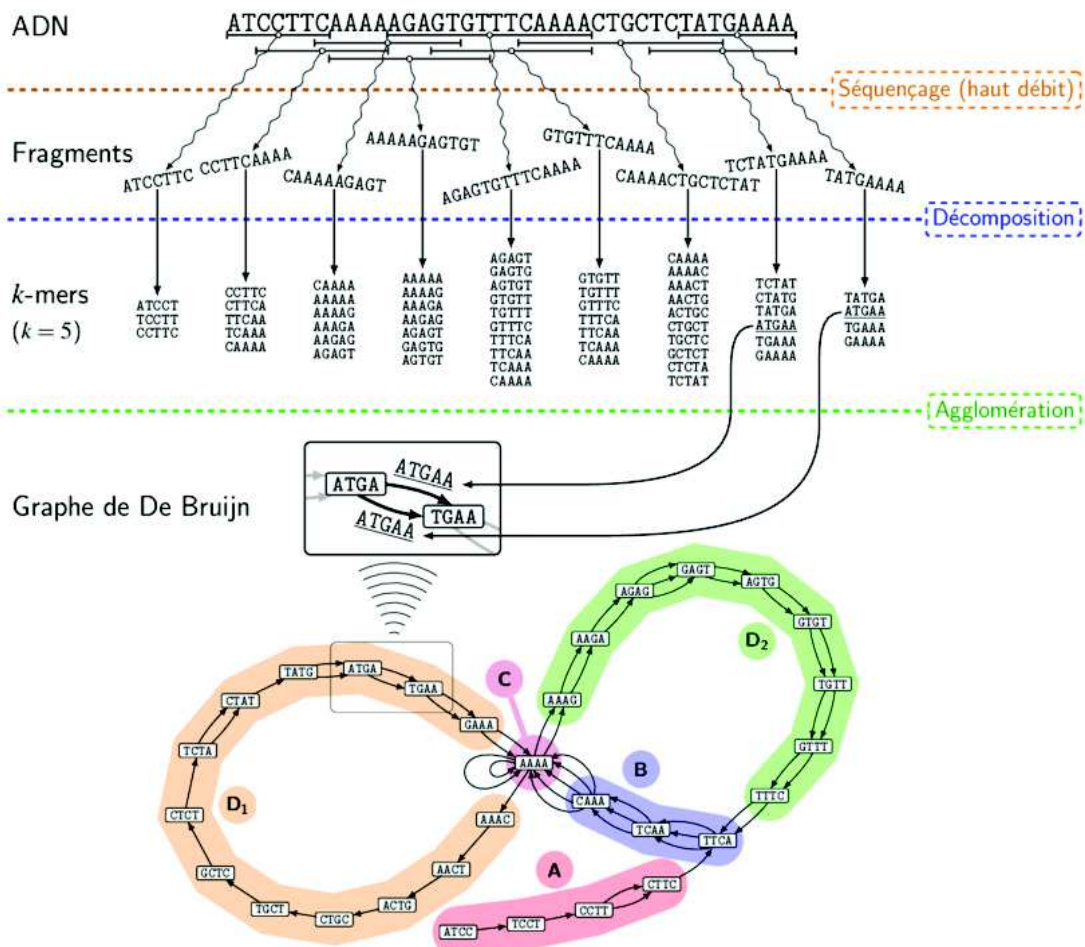


Figure 12 : Représentation schématique d'un assemblage utilisant un graphe de de Bruijn (Ponty).

Lorsque l'espèce microbienne est connue, l'assemblage peut se réaliser en prenant modèle sur un génome de référence. Cette opération est appelée *mapping*, cependant, elle est souvent appelée alignement et est à distinguer de l'alignement de séquences connues. Les programmes les plus utilisés pour le *mapping* sont BWA (Li and Durbin) et bowtie2 (Langmead and Salzberg). Le *mapping*, s'il permet la réalisation d'une séquence continue, exclut de l'assemblage les séquences absentes de la souche de

référence. Il est donc primordial de bien choisir sa souche de référence et de garder à l'esprit l'existence des *reads* non mappés.

Après l'assemblage, on peut définir la profondeur de lecture et la couverture du séquençage. La profondeur de lecture est le rapport entre l'ensemble des longueurs des *reads* associés à un organisme et la taille du génome de l'organisme. La couverture d'une base est le nombre de fois où celle-ci apparaît dans un *read*, ce qui équivaut au nombre de fois où elle a été séquencée.

5. Applications

5.1. Taxonomie

Le concept d'espèce et de classification est une nécessité pragmatique pour étudier le vivant mais rappelons que la nature est continue et non discrète (Rosselló-Mora and Amann). Les premières classifications se basaient sur des caractères morphologiques et phénotypiques (COWAN) puis des méthodes plus précises sont apparues. L'hybridation ADN-ADN a été longtemps été le *gold standard* pour définir si deux souches bactériennes appartenaient à la même espèce mais aujourd'hui, la méthode de référence est l'alignement des génomes complets avec le calcul de l'identité nucléotidique moyenne ; on considère comme appartenant à la même espèce deux souches ayant une identité nucléotidique moyenne supérieure à 95% (Richter and Rosselló-Móra).

Le développement de cette approche génomique a récemment permis de reclasser des espèces bactériennes dans des taxons plus appropriés comme pour *Enterobacter aerogenes* reclassé en 2017 *Klebsiella aerogenes* (Tindall et al.). Cette méthode a aussi engendré la séparation d'une espèce en plusieurs avec l'apparition d'*Enterobacter xiangfangensis* (Wu et al.).

En ce qui concerne les virus, leur grande diversité est à l'origine de la difficulté à les classer et à définir ce qu'est une espèce. Le code international de classification et de nomenclature des virus définit le concept d'espèce comme un groupe monophylétique de virus dont les propriétés peuvent être distinguées des autres espèces par de multiples critères (Adams et al.). La définition des espèces est majoritairement décidée en fonction de l'emprise géographique, de la symptomatologie de la maladie éventuellement engendrée et de la similarité génétique. Pour celle-ci, il n'y a pas de seuil qui fasse consensus et qui s'adapte à toutes les familles de virus (Peterson).

5.2. Métagénomique

“The true taxonomist is a man with a mission; he often leads a cloistered life, protected from the vexations and frustrations of the everyday world, and he may well wear blinkers as opaque as any worn by a horse. He is more likely to have academic interests than to be an applied worker, a follower of Adanson rather than of Kauffmann, and an Englishman rather than an American. Living a life of seclusion, safe in his small laboratory, and surrounded by his books, his microscope (and perhaps his computer tape), he affects an unconcern for the mundane application of his work. But science has a way of making itself useful, and the useful application of classification is identification.” (COWAN)

L’assignation taxonomique des *reads* permet l’identification du ou des microorganismes présents dans un prélèvement. La démarche est similaire que ce soit un prélèvement d’intérêt clinique ou à visée de détermination du microbiote. Deux approches existent : la première consiste à séquencer après amplification préalable d’une région d’intérêt comme une boucle de l’ADN 16S bactérien (Namdari et al.), l’ITS fongique (Schoch et al.) ou d’un panel de gènes d’intérêt (Chao et al.) (*amplicon-based metagenomics*), la deuxième s’affranchit de cette amplification et séquence l’ensemble du matériel génétique présent (*shotgun metagenomics*). Dans le cas d’une amplification préalable, l’interférence éventuelle du génome de l’hôte diminue du fait de l’enrichissement d’ADN d’intérêt par rapport à celui de l’hôte mais seuls les organismes dans le spectre de la PCR pourront être retrouvés. En l’absence d’amplification, n’importe quel pathogène dont le génome apparaît dans les bases de données est susceptible d’être identifié, cependant la présence d’ADN de l’hôte en forte quantité peut brouiller le signal et réduire la sensibilité.

Lors de l’utilisation de régions d’intérêt taxonomique, il est possible, notamment dans des études de microbiote, de retrouver des microorganismes non référencés dans les bases de données, par exemple des bactéries non cultivables. C’est pourquoi est né le concept d’OTU (*Operational Taxonomy Unit*) qui se base sur la similarité génétique entre les organismes séquencés dans un échantillon. En général, le seuil semi-arbitraire de similarité pour appartenir à la même OTU est de 97%. Ces OTU sont aussi utilisées lors de l’assignation taxonomique d’un *read*. Une fois les séquences classées en OTU, on peut assigner une taxonomie à ces OTU. Fréquemment, une OTU est classée à

niveau d'un phylum ou d'un genre et ne descend que rarement jusqu'à l'espèce (Schloss and Westcott).

Plusieurs programmes existent pour assigner des *reads* à une OTU ou une espèce, que ce soit à partir de *contigs* assemblés ou de *reads*. Les deux plus utilisés sont kraken2 (Wood et al.) et MetaPhlan (Beghini et al.), d'autres comme mothur (Schloss et al.) et diamond (Buchfink et al.) sont également utilisables.

Une fois les données traitées, pour les visualiser, il existe des logiciels de représentation graphique comme Krona (Ondov et al.), ou MEGAN (Huson et al.). Ceux-ci permettent d'apprécier visuellement la diversité et l'identification des espèces retrouvées.

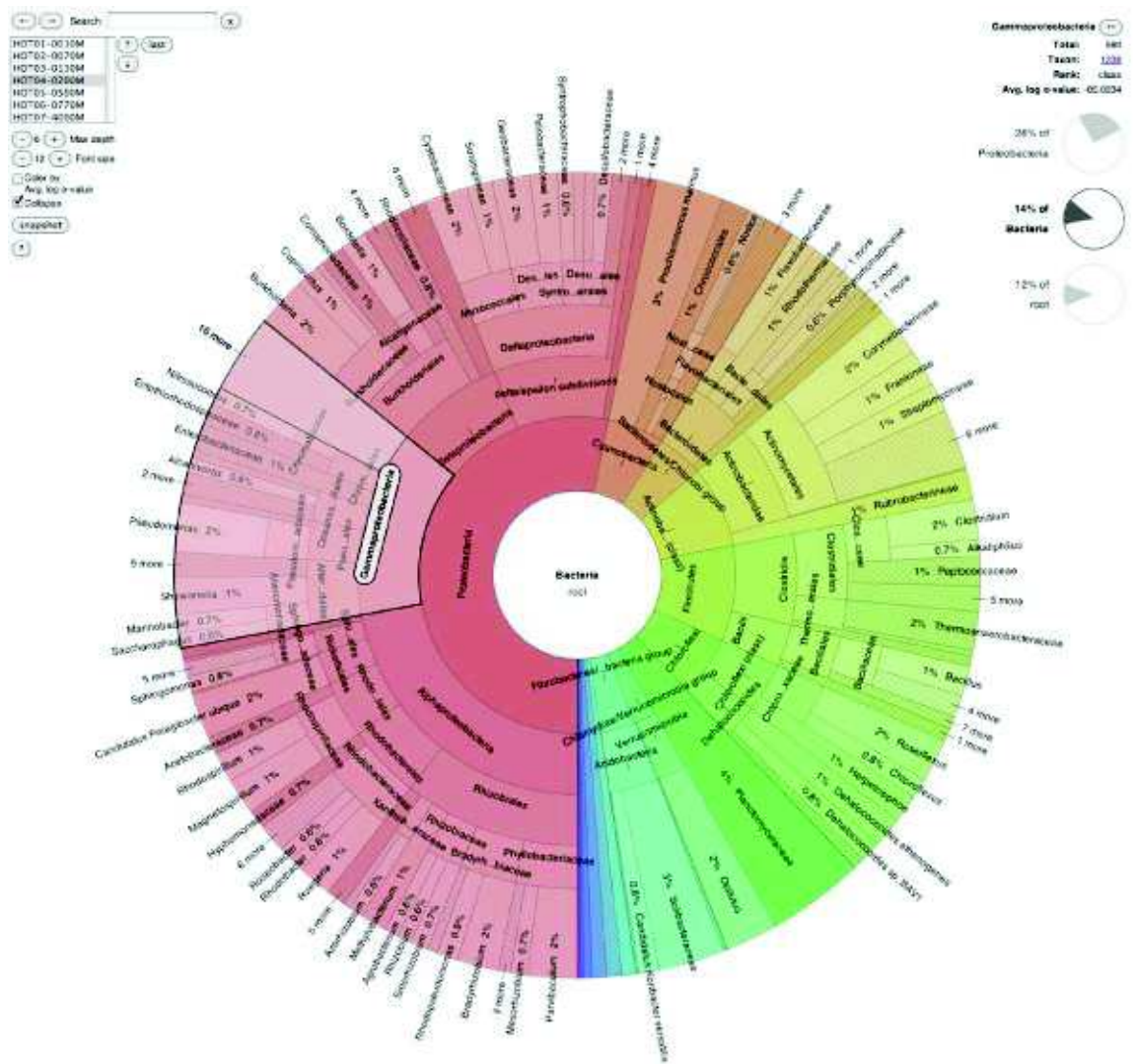


Figure 13 : Exemple de graphique obtenu par Krona à partir d'une exploration métagénomique (Ondov et al.).

Les prélèvements pouvant être contaminés par le microbiote humain et par des acides nucléiques contaminants des réactifs de laboratoire, l'interprétation des résultats est primordiale pour déterminer si un microorganisme retrouvé est impliqué dans la pathogénie ou s'il est un simple contaminant (Maggi and Bendinelli; Moustafa et al.; Foulongne et al.; Zheng et al.; Salter et al.; Strong et al.; Wilson, Fedewa, et al.).

Une façon de s'affranchir d'une partie des contaminants de laboratoire est de rechercher dans le témoin négatif des *reads* appariés aux espèces suspectées pathogènes retrouvées dans les prélèvements. Un rapport de 10 entre le nombre de *reads* pour un microorganisme entre le prélèvement et le témoin négatif permet d'argumenter qu'il ne provient pas d'une contamination de laboratoire (Miller et al.).

En utilisant un contrôle interne, on doit retrouver l'espèce microbienne utilisée comme contrôle interne. L'adjonction du contrôle interne ne devant pas interférer avec la sensibilité, une faible quantité devrait être utilisée pour donner environ 1000 *reads* par million. Un seuil de 100 *reads* de contrôle interne par million de *reads* totaux pourrait être utilisé pour valider une analyse. Si le résultat de séquençage d'un échantillon rendait un nombre de *reads* par million inférieur au seuil, cela traduirait une prédominance écrasante d'ADN de l'hôte et donc devrait faire suspecter une perte de sensibilité (Miller et al.).

Après l'assignement taxonomique des *reads*, il est possible, si la couverture est suffisante, d'assembler *de novo* le génome des micro-organismes supposés présents par l'assignement taxonomique ou de mapper les *reads* sur des génomes de référence des espèces suspectes. Cette opération permet de s'assurer que le micro-organisme détecté est bien présent dans le prélèvement avec son génome complet.

5.3. Analyse phylogénique

Les espèces microbiennes sont des populations regroupant des souches partageant un fond génétique commun mais dont l'hétérogénéité permet de définir des lignées à l'intérieur de l'espèce. Certaines analyses phylogénétiques, dont le génotypage, sont des méthodes utilisées pour replacer une souche spécifique dans sa lignée.

La méthode de référence pour faire du génotypage bactérien d'une souche est la méthode MLST (*Multi Locus Sequence Typing*) décrite en 1998 qui examine les variations de séquences de généralement 7 gènes de ménage déterminés pour chaque espèce (M. C. Maiden et al.; Pérez-Losada et al.). Les nouvelles méthodes basées sur le séquençage haut débit permettent une meilleure compréhension de la diversité

bactérienne avec une analyse plus en profondeur via le séquençage de génome complets et la métagénomique.

5.3.1. MLST

Le typage MLST, historiquement et encore souvent réalisé par séquençage Sanger, permet de définir des grandes lignées à l'intérieur d'une espèce bactérienne. Ces lignées peuvent suffire à déterminer certaines souches comme appartenant par exemple à une lignée hypervirulente comme le clone CC-17 de *Streptococcus agalactiae* (Da Cunha et al.) mais ne sont pas capable de distinguer finement des souches proches. De plus, elle n'est pas applicable à tous les micro-organismes, certains comme *Mycobacterium tuberculosis* ont un génome plus stable donc des souches éloignées peuvent n'exprimer aucune différence allélique sur les gènes de ménage sélectionnés (Gagneux). Le typage MLST est réalisable à partir du séquençage complet d'une souche par séquençage haut débit ; celui-ci concorde bien avec le typage effectué par séquençage Sanger des gènes de ménage (Pérez-Losada et al.; Kwong et al.). Les données de séquençage complet permettent d'utiliser l'approche MLST avec un nombre plus important de gènes, ce qui autorise une analyse plus profonde de la diversité dans une population. Ainsi, le MLST ribosomal (rMLST) prenant en compte une cinquantaine de gènes (Jolley et al.), le MLST *core genome* (cgMLST) l'ensemble des gènes du *core genome* d'une espèce (de Been et al.) et le MLST *whole genome* l'ensemble des gènes découverts chez une espèce (M. C. J. Maiden et al.) ont vu le jour.

5.3.2. Polymorphisme nucléotidique simple (SNP)

L'analyse des variations ponctuelles de nucléotides (*Single Nucleotide Polymorphism*, SNP) de différentes souches par rapport à une souche de référence permet de produire un arbre phylogénétique comprenant toutes les souches étudiées. Cette application permet d'étudier les relations entre des souches beaucoup plus finement que par un génotypage de type MLST, des souches de même type MLST peuvent être classées en lignées distinctes en différenciant des souches à la mutation ponctuelle près. Là où le typage MLST pouvait être réalisé par séquençage Sanger, seul le séquençage complet d'une souche, par séquençage haut débit, permet ce type d'analyses. En effet, à partir de la séquence d'une souche et de celle d'une référence, un logiciel comme snippy (Seemann, *Snippy*) est capable de lister tous les SNP distinguant les deux séquences. On peut utiliser ce même logiciel sur une population entière. Snippy

va d'abord aligner les génomes de toutes les souches pour en extraire le core génome de la population étudiée, le génome commun à toutes les souches. Ensuite, il va extraire tous les SNP par rapport à la référence qui sont inclus dans ce core génome. A partir de ces SNP, un logiciel comme FastTree (Price et al.) ou MEGA (Tamura et al.) peut construire un arbre phylogénétique.

Cependant, cette méthode est basée sur le postulat que, pour toutes les souches, les taux de mutations aléatoires sont similaires or, lors d'événements de recombinaisons génétiques, où une séquence est remplacée d'un coup par une nouvelle, la nouvelle séquence peut être très différente de la souche utilisée comme référence pour les SNP. Dans ce cas, localement, la densité de SNP sera plus élevée ou plus basse au niveau de cet événement de recombinaison et pourra fausser l'arbre phylogénétique en faisant paraître la souche plus éloignée ou plus proche de la référence qu'en réalité. Pour se défaire de ce biais, des logiciels comme gubbins (Croucher et al.) permettent de détecter ces recombinaisons en recherchant des zones où la fréquence de SNP est différente du reste du génome. Après exclusion de ces zones, on peut construire un arbre phylogénétique sans biais.

5.3.3. *Analyse souches épidémiques*

Un des apports principaux du séquençage de génome complet est son utilisation pour caractériser une épidémie. En effet, les méthodes actuelles basées sur la PCR pour caractériser les relations entre les souches d'une éventuelle épidémie, dont le chef de file est l'électrophorèse en champs pulsé, sont peu discriminantes, difficile d'interprétation et fastidieuses.

Aujourd'hui, les approches génomiques de typage basées sur le séquençage de génome complet permettent une très grande résolution avec notamment le *core génome* MLST, le *whole genome* MLST et surtout la phylogénie basée sur les SNP. Ces approches sont en train de remplacer les méthodes de biologie moléculaire classiques pour la surveillance des souches circulantes en raison de la faible résolution de l'électrophorèse en champs pulsé et des autres méthodes de typages basées sur la PCR (Neoh et al.).

L'utilisation de la phylogénétique sur le *core genome* a permis la caractérisation très rapide d'une épidémie de 4 cas d'infections néonatales dues à un clone de *Streptococcus agalactiae* en réanimation par la réalisation d'un dendrogramme. Celui-ci montre les 4

souches partageant une même branche exclusive et très proches les unes des autres (Figure 14).

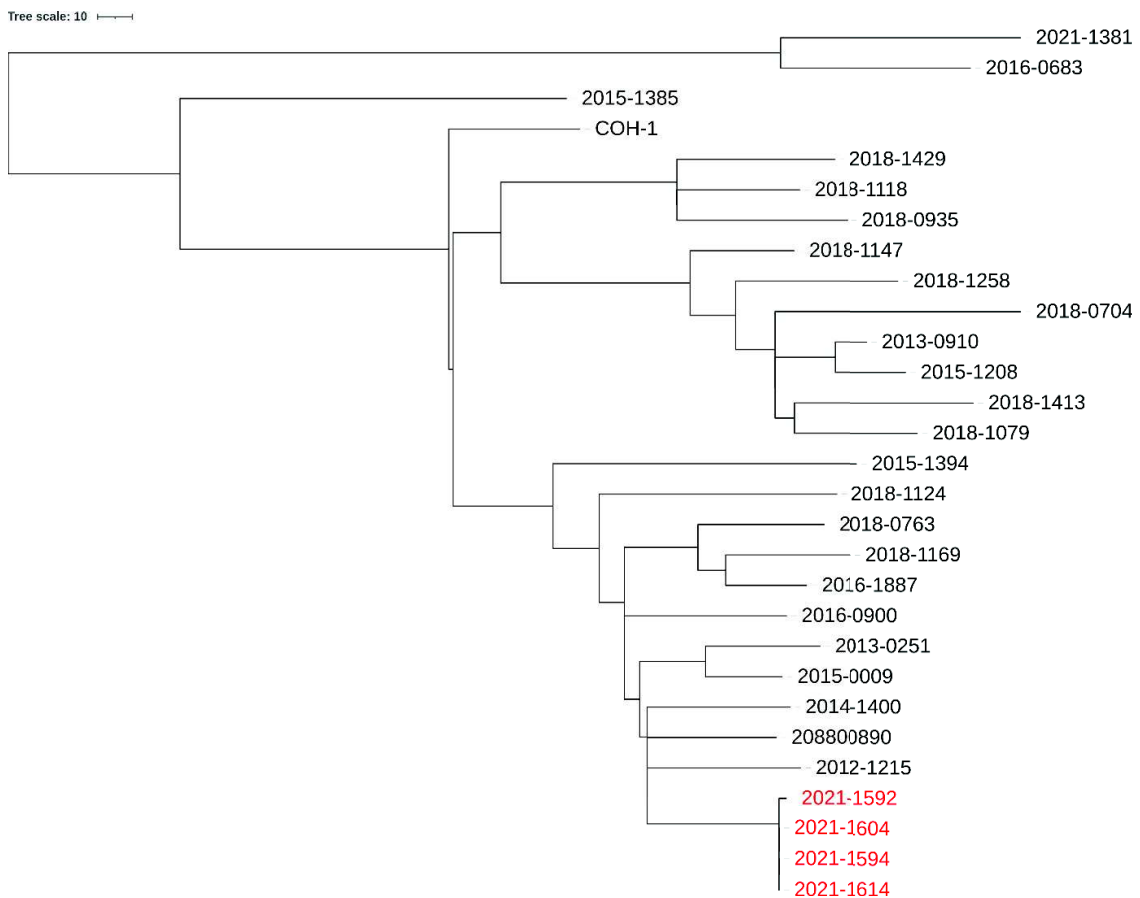


Figure 14 : Dendrogramme réalisé à partir de 4 souches de *Streptococcus agalactiae* CC-17 suspectes d'une épidémie hospitalière (en rouge) et de 25 souches représentatives des infections à *Streptococcus agalactiae* CC-17 en France (en noir).

5.3.4. Analyse de pathogènes émergents

La remarquable précocité avec laquelle SARS-CoV-2 a été caractérisé est en grande partie due aux méthodes de séquençage haut débit. En effet, l'alignement du génome complet du SARS-CoV-2 avec ceux d'autres coronavirus a permis de classer le virus parmi ses congénères ; aussi, les alignements séparés de différentes séquences codantes sur des séquences codantes similaires des autres *Coronavirus* ont été à l'origine d'hypothèses probables sur les réarrangements génétiques précédant l'émergence du SARS-CoV-2 et donc sur son hôte d'origine (Tabibzadeh et al.). Par la suite, le séquençage de nombreuses souches des SARS-CoV-2 de par le monde a pu nous donner de très nombreuses informations sur l'évolution de la pandémie, le suivi des variants et leur dynamique géographique notamment (Bai et al.).

5.4. Gènes de résistance

Il existe deux approches pour retrouver des gènes de résistance aux antibiotiques via un séquençage haut débit. La première consiste à assembler le génome en *contigs* à partir des *reads* puis de chercher (annoter) les gènes de résistance dans ce génome assemblé. Cette méthode nécessite une couverture importante et un temps de calcul conséquent. Cependant, elle donne beaucoup d'informations, notamment sur la position des gènes et sur les séquences qui les régulent (Boolchandani et al.).

Par opposition, on peut mapper directement les *reads* sur les bases de données de gènes de résistance. Cette approche permet de s'affranchir des étapes d'assemblage et d'alignement à un génome de référence et économise donc du temps de calcul, ses principaux avantages sont la précocité des résultats et la capacité de retrouver des résistances sur un micro-organisme de faible abondance dans un microbiote complexe. Avec cette méthode, cependant, on ne peut pas déterminer à quelle bactérie appartient quel gène de résistance dans un échantillon plurimicrobien (Boolchandani et al.).

Dans les deux cas, le choix de la base de données est important, il faut que celle-ci indexe les variations récemment découvertes pour avoir un résultat fiable. Aussi important que les gènes de résistance directement, qui peuvent ou non s'exprimer, il est important de prendre en considération les séquences associées à une résistance phénotypique à un antibiotique (Boolchandani et al.).

Parmi les bases de données, certaines sont généralistes, dont trois des plus utilisées sont Resfinder (Bortolaia et al.; Zankari et al.), ARG-ANNOT (*Antibiotic Resistance Gene Annotation*) (Gupta et al.) et CARD (*Comprehensive Antibiotic Resistance Database*) (Jia et al.), et d'autres sont spécialisées dans une famille de gènes ou dans une espèce. Le choix d'une base de donnée se fait donc au cas par cas en fonction des besoins (Boolchandani et al.).

Les bases de données sont généralement biaisées par la surreprésentation de marqueurs de résistance de bactéries pathogènes pour l'homme et de bactéries cultivable rendant difficile l'identification de marqueurs nouveaux ou de résistances chez des pathogènes inhabituels, particulièrement lorsque les marqueurs de résistance sont complexes.

Une des voies de progression prometteuse est l'utilisation de *machine learning* dans la recherche de marqueurs associés à une résistance (Davis, Boisvert, et al.). L'étude transcriptomique permettrait de détecter des modifications de l'expression de gènes

associées à des résistances que l'on pourrait ne pas voir sur l'étude génomique (Khaledi et al.).

La prédiction de résistance à partir de données de métagénomique est encore difficile et manque de fiabilité. Si les bases de données sont aujourd'hui mises à jour régulièrement avec l'ajout de nouveaux gènes découverts, la réévaluation des gènes présents dans les bases de données depuis longtemps n'est pas toujours réalisée ; par exemple dans la base de données Resfinder, le gène *mre(a)* est faussement associé à une résistance à l'érythromycine chez *Streptococcus agalactiae*. De plus, la recherche de gènes de résistance est fortement orientée vers certaines familles d'antibiotiques, ce qui entraîne des lacunes pour d'autres antibiotiques. Enfin, les bases de données peuvent aussi contenir le même gène ou la même séquence sous plusieurs noms différents, ce qui complique l'analyse (Xavier et al.).

5.5. Gènes de virulence

La recherche de gènes de virulence pourrait probablement permettre de détecter sur culture ou directement sur prélèvement différents facteurs d'intérêt clinique. La mise en évidence du gène codant la leucocidine de Panton-Valentine ou la toxine TSST-1 (Schmidt et al.) chez une souche clinique de *S. aureus* par exemple devrait entraîner une prise en charge médicale spécifique lors d'une infection. Pour certaines bactéries comme *C. diphtheriae*, la présence de toxine définit l'infection.

Dans ces deux cas, la recherche des facteurs de virulence d'intérêt clinique nécessite l'utilisation d'une PCR spécifique. Les laboratoires de biologie médicale ne pouvant pas disposer de couples d'amorces et de protocoles adaptés à chaque situation, il est fréquent de devoir envoyer des souches dans un laboratoire spécialisé pour faire le diagnostic. La maîtrise du séquençage complet de génome permet de s'affranchir de la nécessité d'envoyer à un laboratoire spécialisé pour la détection d'un facteur de virulence particulier.

Il existe des bases de données généralistes de gènes de virulence, la plus connue étant VFDB (Chen et al.). Cependant, lors de l'analyse de certains micro-organismes, des bases de données spécifiques d'espèce peuvent donner des informations supplémentaires.

6. Besoins informatique

6.1. Bioinformatique sur site avec une machine dédiée

Pour les diverses applications nécessitant une étape bioinformatique, le temps de calcul peut être long. Pour ceci, nous avons chronométré certains calculs permettant de définir l'ordre de grandeur des puissances et des temps de calcul nécessaires. Les calculs ont été réalisés sur 32 cœurs d'un processeur Intel(R) Xeon(R) Gold 6142 CPU @ 2.60GHz sur un serveur distant exploité par CentOS 7.9.2009.

Pour évaluer les besoins en calculs, nous avons étudié une série de 21 produits de séquençage sur illumina avec *reads* appariées avec en moyenne 3 168 502 *reads* par échantillon. Les produits séquencés provenaient d'extractions d'ADN de différentes souches de *Streptococcus agalactiae*.

Les étapes de contrôle qualité et de trimming sont très rapides et ne nécessitent que quelques secondes. L'étape d'assignation taxonomique avec MetaPhlan (Beghini et al.) a été réalisée en 1h00 (en moyenne 2min 51s par échantillon) pour l'ensemble de la série en utilisant les paramètres par défaut. L'assemblage de toutes les souches réalisé par spades (Bankevich et al.) dans les conditions par défaut a pris 3h16 soit 9min 20s par souche. Le typage *mlst* (Seemann, *Mlst*) n'a pris que quelques secondes. L'annotation des 21 souches par Prokka (Seemann, "Prokka") a été réalisée en 2h12 (6 min 16s par souche en moyenne) avec les bases de données par défaut et la base de donnée *Streptococcus agalactiae* d'UniprotKB (The UniProt Consortium). L'alignement du *core-genome* des 21 souches et de la souche de référence SGB-COH1 avec l'extraction des SNP par rapport à la référence a nécessité 13 minutes avec Snippy (Seemann, *Snippy*). La construction de l'arbre phylogénétique de ces 21 souches basée sur l'analyse des SNP a mis 1 min avec gubbins (Croucher et al.) et RAxML (Stamatakis).

6.2. Bioinformatique en *cloud-computing* sur plateforme publique

Des plateformes de calcul bioinformatique sont disponibles sur internet et permettent de réaliser des calculs sans posséder de machine capable de les exécuter sur place. Les deux plus connues sont les plateformes Galaxy, dont celle du Projet Galaxy (Afgan et al.) et celle de l'Institut Pasteur (Mareuil et al.), et le site PATRIC (Davis, Wattam, et al.). Ces sites sont gratuits mais la puissance de calcul disponible est partagée entre tous les utilisateurs et les calculs sont mis sur des files d'attente le temps d'allouer la

puissance nécessaire. Ces plateformes ont cependant l'avantage de proposer une interface graphique et de nombreux outils d'analyse facile d'accès. Il est à noter que la durée de ces files d'attente ainsi que des calculs est difficile à anticiper et peut dépasser plusieurs jours dans certains cas. De plus, l'utilisateur ne peut pas maîtriser la totalité des paramètres de tous les programmes utilisés ni ajouter des programmes qui ne seraient pas disponibles sur les plateformes.

Pour ce qui est de la plateforme Galaxy, il existe une version installable sur son propre serveur pour créer sa propre plateforme Galaxy privée. Celle-ci a tous les avantages de la facilité d'accès et de l'interface graphique, permet d'installer et de paramétrer les programmes que l'on souhaite utiliser et s'affranchit du côté aléatoire de la puissance informatique disponible (Afgan et al.).

7. Performances dans le diagnostic microbiologique

7.1. Infections du compartiment vasculaire

7.1.1. Bactériémies et fongémies

Le diagnostic de septicémie par métagénomique n'est pas évident et devrait être proposé surtout pour la recherche de bactéries de culture fastidieuse ou impossible. On retrouve donc par exemple les bacilles Gram négatif du groupe HACCEK ou les mycobactéries. La faible charge bactérienne dans le sang lors de la grande majorité des bactériémies et la présence de grande quantité d'ADN humain entraîne une sensibilité faible, probablement plus faible que l'hémoculture. Les bactéries transitoires asymptomatiques rendent aussi l'interprétation difficile en cas de détection d'un germe réputé non pathogène (Greninger and Naccache).

Malgré tout, il existe un certain nombre de kits commerciaux de diagnostics de bactériémies basés sur l'utilisation du séquençage haut débit avec ou sans amplification préalable. Parmi ceux-ci, les kits avec amplification préalable 16S et 18S SepsiTest™ (Molzylm, Bremen, Germany), Micro-Dx™ (Molzylm, Bremen, Germany) permettent un résultat entre 8 et 12h mais ne sont pas capable de détecter les résistances aux antibiotiques et le premier semble présenter une sensibilité et une spécificité de respectivement 48% et 86% (Stevenson et al.; Peri et al.). Un dernier kit avec amplification préalable, le HybcellPathogen DNA assay (CubeDx, St Valentin, Austria), fonctionnant sur le même principe mais couplé à une PCR multiplex capable de détecter

certaines gènes de résistance (vanA/B, mecA/C, CTX-M, KPC, OXA-48, NDM, IMP) semble présenter une sensibilité 63% et une spécificité de 83% dans le diagnostic de bactériémie (Knabl et al.; Peri et al.). Des kits sans amplification préalable sont aussi disponibles, iDTECT® Dx Blood test (PathoQuest, Paris, France) fait preuve d'une très forte valeur prédictive négative de 98,4% et d'une sensibilité lui permettant de détecter des pathogènes non détectés par les méthodes classiques notamment chez des patients immunodéprimés (Parize et al.; Peri et al.). Un autre kit, le Karius NGS Plasma Test™ (Karius, Redwood City, CA, USA), montre une concordance avec les méthodes classiques aux alentours de 93% (Blauwkamp et al.; Peri et al.). Ces kits sans amplifications préalables sont capables de chercher d'éventuelles résistances aux antibiotiques mais leurs performances dans ce domaine restent à évaluer (Peri et al.).

Une étude a évalué la recherche de pathogènes à partir de l'ADN libre circulant plasmatique sans amplification préalable et retrouve une meilleure sensibilité pour les bactéries et les champignons de la métagénomique associée à l'hémoculture (23% de prélèvements positifs) par rapport à l'hémoculture seule (13% de prélèvements positifs). De plus, la métagénomique a permis de détecter des virus (CMV et HSV-1) dans 15 prélèvements sur 78. Il ne semble pas y avoir de différence de nature entre les pathogènes retrouvés via métagénomique et ceux retrouvés par l'hémoculture (Long et al.).

Finalement, l'apport de ces méthodes dans le diagnostic de septicémies est difficile à évaluer. En effet, si la rapidité de certaines méthodes et l'apport du séquençage pourraient permettre une meilleure prise en charge de certains patients, le coût et la nécessité d'avoir un personnel capable de mettre en œuvre ces méthodes relativise les avantages. De plus, les méthodes plus classiques de culture ou de PCR sont déjà très satisfaisantes dans la très grande majorité des cas de septicémies (Peri et al.).

7.1.2. Endocardites infectieuses

Les endocardites infectieuses représentent un défi diagnostique important, notamment du fait de la fréquence des hémocultures et culture de valves négatives. Deux études sur 49 et 7 patients recouvrant trois méthodes de séquençage haut débit sans amplification préalable sont disponibles (Cai et al.; Cheng et al.).

Pour la première, impliquant 49 patients dont les valves cardiaques ont été prélevées, 28 valves étaient positives en culture alors que 21 étaient négative malgré le diagnostic d'endocardite posé. Une analyse métagénomique sans amplification

préalable par méthode IonTorrent retrouve les mêmes résultats que la culture sur toutes les valves positives mais retrouve aussi la présence d'une bactérie connue comme pourvoyeuse d'endocardite dans toutes les valves négatives en culture (Figure 15). 8 valves de patients atteints de valvulopathies non infectieuses ont été analysées de la même façon en tant que témoins négatifs et aucune d'elle n'était positive en métagénomique. Il est à noter la similarité entre les espèces bactériennes retrouvées en culture et celles retrouvées uniquement par séquençage, ce qui laisse envisager une meilleure sensibilité du séquençage par rapport à la culture. Cette hypothèse est aussi étayée par la plus faible couverture et le plus faible nombre de *reads* obtenus pour les prélèvements négatifs en culture (Cai et al.).

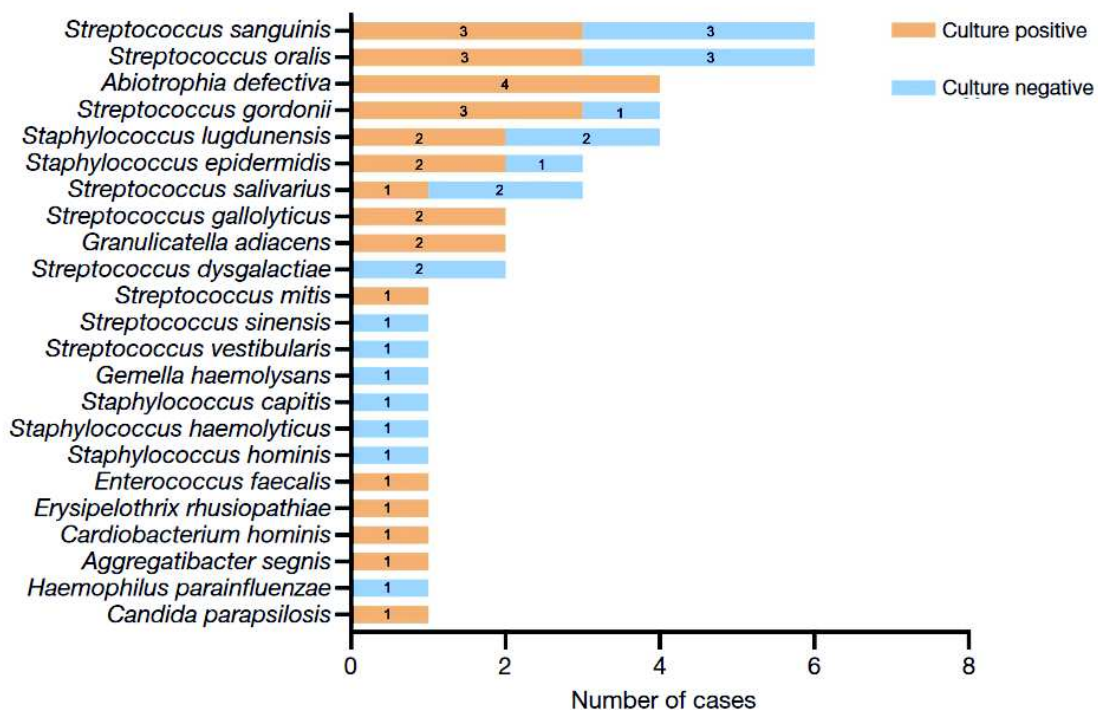


Figure 15: Bactéries retrouvées lors de l'utilisation du séquençage haut débit sur des valves cardiaques prélevées d'endocardites infectieuses d'après (Cai et al.).

La deuxième étude, sur 7 patients, a effectué une analyse métagénomique via séquençage nanopore et BGISEQ (Automate de séquençage aujourd'hui minoritaire basé sur l'utilisation de sondes fluorescentes et produisant des *reads* courts (Mak et al.)) en comparaison à la culture et à la PCR-Séquençage Sanger ADNr 16S. Deux prélèvements seulement étaient positifs par la culture et 6 par PCR-séquençage 16S alors que les 7 ont été positifs lors des méthodes de séquençage haut débit. Le prélèvement positif uniquement par séquençage haut débit retrouve une *Coxiella burnetii*. Les couvertures retrouvées et le nombre de *reads* analysés sont faibles ; il serait intéressant

de connaître le nombre total de *reads* obtenus avec les *reads* humains pour évaluer la sensibilité (Cheng et al.).

L'utilisation de la métagénomique clinique dans le cadre des endocardites infectieuses dont l'étiologie est régulièrement inconnue permet apparemment de récupérer une majorité de diagnostic. Cependant, une couverture et une profondeur de lecture suffisantes pour rechercher des marqueurs de résistance aux antibiotiques semblent difficiles à atteindre.

7.2. Infections du système nerveux central

7.2.1. Méningites et encéphalites

Les méningites et encéphalites restent des tableaux cliniques dont l'étiologie est difficile à déterminer. Dans 40 à 60% des cas, l'étiologie n'est pas retrouvée par les méthodes conventionnelles comprenant la culture bactérienne et les PCR spécifiques, simplex ou multiplex (C. A. Glaser et al.; Carol A. Glaser et al.; Granerod et al.). L'absence d'étiologie retrouvée est plus fréquente chez les patients âgés de plus de 65 ans et est associée à une mortalité moins importante et une durée d'hospitalisation plus courte (Khetsuriani et al.).

Une étude prospective multicentrique menée par l'Université de Californie décrivant sur un an 204 patients atteints de méningite ou de méningo-encéphalite retrouve une cause infectieuse sur 58 patients. Treize d'entre eux (22%) ont eu un diagnostic étiologique par séquençage haut débit dans le LCR alors que les méthodes traditionnelles ont échoué. Le séquençage haut débit dans le LCR est revenu négatif chez 26 patients pour qui un diagnostic a été trouvé par une méthode traditionnelle, cependant, pour 18 d'entre eux, le diagnostic a été réalisé par sérologie ou par recherche de pathogènes dans un autre prélèvement que le LCR ; un manque de sensibilité est incriminé dans seulement 8 prélèvements. Sur les 13 patients dont le diagnostic a été permis par le séquençage haut débit, 7 ont bénéficié d'une prise en charge spécifique déterminée par le résultat. Les pathogènes retrouvés par métagénomique clinique uniquement sont variés et concernent des germes cultivables facilement comme *C. tropicalis*, *Enterococcus faecalis* ou *Streptococcus agalactiae* mais aussi des pathogènes difficilement cultivables ou des virus rarement recherchés comme *Nocardia farcinica* ou le virus de l'hépatite E (Wilson, Sample, et al.).

Un protocole publié aussi par une équipe de l'Université de Californie annonce une limite de détection de 14 copies/mL pour le CMV, 313 copies/mL pour le VIH, 10 UFC/mL pour *Streptococcus agalactiae* et 8 UFC/mL pour *Klebsiella pneumoniae* (Miller et al.). Avec ce protocole d'analyse, l'équipe a testé 20 LCR et retrouve les pathogènes identifiés par les méthodes classiques dans 11 cas sur les 12 prélèvements positifs par ailleurs (8 entérovirus, 1 HSV-1, 1 VIH et un Virus West Nile). L'échec s'est produit pour un Virus West Nile dont le diagnostic avait été étayé par la présence d'IgM spécifiques dans le LCR. Trois prélèvements retrouvant la présence d'*Enterobacter sp.*, *Corynebacterium sp.* et EBV ont été considérés comme des faux positifs devant le tableau clinique et un profil sérologie d'infection ancienne à EBV. Pour les pathogènes n'étant classiquement pas retrouvés par PCR dans le LCR (*Borrelia burgdorferi*, *Mycoplasma sp.*), le séquençage haut débit ne semble pas faire mieux (Miller et al.).

En ce qui concerne les infections parasitaires du système nerveux central, un cas de toxoplasmose cérébrale chez un patient VIH a été rapporté avec un diagnostic par métagénomique sans amplification préalable et montre la capacité de la technique à faire ce type de diagnostic (Z. Hu et al.).

7.2.2. *Abcès cérébraux*

Une première étude avait identifié un intérêt à l'étude métagénomique des abcès cérébraux après amplification 16S et 18S avec notamment la mise en évidence d'infections polymicrobienne, là où le séquençage Sanger est mis en défaut, et avait montré une association entre l'origine de l'abcès cérébral et les espèces retrouvées (Al Masalma et al.). Deux études utilisant la métagénomique après amplification préalable dans des abcès cérébraux montrent l'utilité de cette technique pour améliorer la sensibilité en complément de la culture notamment dans les infections plurimicrobiennes dont les prélèvements ont été réalisés sous antibiothérapie (Stebner et al.; Andersen et al.). En cas d'infection réputée monomicrobienne en culture et en séquençage, il est proposé de rétrocéder l'antibiothérapie à la bactérie retrouvée (Andersen et al.).

Aucune grande série n'a été publiée sur la métagénomique des abcès cérébraux sans amplification préalable. Un cas publié a rapporté la présence d'une flore polymicrobienne en plus d'un *Streptococcus constellatus* et d'une *Prevotella sp.*. Cette flore comprenait des mycobactéries, *Debaromyces sp.*, *Aspergillus sp.*, *Malassezia sp.* et des virus des genres *Herpesvirus*, *Moumouvirus* et *Granulovirus*. Dans ce cas, la

prévotelle disposait des gènes de résistance *ErmF*, *tetQ*, *CfxA2* et *CfxA2* mais lors du séquençage directement sur le prélèvement, seul *tetQa* pu être identifié par manque de profondeur de lecture (Lin et al.). A propos de ce cas, il semble difficile d'élaborer une stratégie thérapeutique à partir d'un tel résultat. Trois cas rapportent la détection et le traitement d'abcès cérébraux impliquant des *Nocardia*. Ces bactéries étant de diagnostic et de culture difficiles, c'est un apport utile dans la prise en charge de celles-ci (Yang et al.; Huang et al.; Zhou et al.). Enfin, une série de 4 cas a été publiée chez l'enfant ; sur les deux cas à culture positive, le séquençage direct retrouve la bactérie incriminée par la culture et n'en décèle pas d'autre de façon claire (*Bacteroides fragilis* et *Streptococcus intermedius*). Sur les deux autres cas, avec des cultures négatives, le séquençage retrouve un *Streptococcus intermedius* dans un cas et une *Prevotella oralis* et un *Fusobacterium nucleatum* dans l'autre (H.-L. Hu et al.).

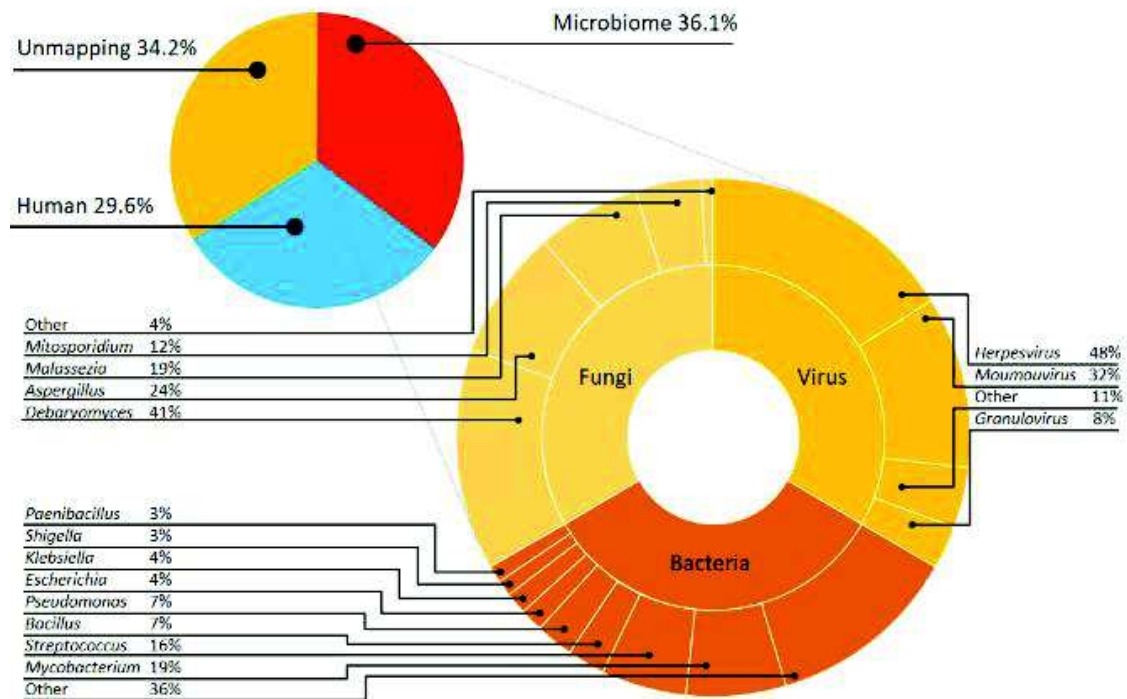


Figure 16 Exemple de flore plurimicrobienne retrouvée dans un abcès cérébral d'après (Lin et al.).

7.3. Infections ostéoarticulaires

L'utilisation de NGS dans les infections de prothèse de genou et de hanche n'est pas étayée par beaucoup d'études. Après amplification préalable, il en ressort une sensibilité probablement plus grande par rapport à la culture mais une spécificité plus faible. Il est difficile de confirmer cette sensibilité supposément meilleure du fait que la présence de fragments génome de bactéries dans le prélèvement ne signifie pas que ces bactéries

sont impliquées sur le site de l'infection. Chez les patients présentant une infection de prothèse de hanche ou de genou selon les critères MSIS, le nombre de germes prédominant dans un prélèvement est significativement plus faible que chez les patients non infectés selon les mêmes critères (Tarabichi et al.). Cette dernière donnée pourrait s'expliquer par le fait que les prélèvements revenants positifs alors qu'il n'y a pas d'infection sont des prélèvements contaminés, les contaminations étant volontiers plurimicrobiennes.

Le problème s'était déjà posé lors de la mise en place de tests basés sur des PCR multiplex avec un taux très important de faux positifs (Jacovides et al.; Melendez et al.). L'interprétation de la détection de bactéries absentes des cultures par séquençage haut débit dans les infections de prothèses est donc difficile.

Les infections de prothèses d'épaules sont un cas particulier dans les infections de prothèses étant donné la fréquence de l'implication de *Cutibacterium acnes*. En effet, *C. acnes* étant un germe à croissance lente et peu virulent, les infections entraînent une symptomatologie moins bruyante et des signes peropératoires moins francs. Le diagnostic de certitude est donc plus difficile, ce qui rend plus délicat la recherche de la sensibilité et de la spécificité d'une nouvelle technique. Une étude prospective sur 44 patients comparant le séquençage haut débit par pyroséquençage après amplification 16S et la culture semble retrouver une faible concordance entre les deux méthodes excepté pour les prélèvements de liquide synovial qui montre une bonne concordance malgré une faible sensibilité. Il est difficile de déterminer si une méthode est plus performante que l'autre ou si les deux pourraient être complémentaires (Namdari et al.).

Plus récemment, une étude de métagénomique clinique sans amplification préalable a été réalisée rétrospectivement sur 24 infections ostéoarticulaires. La sensibilité comparée à la culture était de 58,2% au niveau de l'espèce et 74,5% au niveau du genre. Sur ces 24 prélèvements, après élimination des possibles contaminants, 183 bactéries ont été détectées par métagénomique sans l'avoir été par la culture avec notamment des bactéries anaérobies dans 6 prélèvements et des bactéries du microbiote oropharyngée dans une infection de mandibule. Dans 8 prélèvements, l'analyse des variants a pu montrer la présence de populations polyclonales de la même espèce. Enfin, la recherche de marqueurs de résistance aux antibiotiques chez les bactéries aussi retrouvées en culture a montré une juste prédiction de l'antibiogramme dans 76,5% des cas pour l'ensemble des antibiotiques et une juste prédiction de 100% lors de la recherche

spécifique de résistance aux quinolones. L'étude rétrospective des patients traités par une antibiothérapie non adaptée aux bactéries retrouvées uniquement en métagénomique a retrouvé la rechute d'un patient ; il est difficile de conclure sur l'imputabilité de cette rechute aux bactéries retrouvées en séquençage (Ruppé et al.).

Les études présentées précédemment, laissent entendre un intérêt probable du séquençage haut débit sur les infections ostéoarticulaires et notamment sur les infections de prothèses articulaires. Si cette technique retrouve des bactéries lors de prélèvements sur des matériels à priori non infectés dans un nombre de cas significatif, ces faux positifs devraient pouvoir être relativisés par une analyse approfondie mêlant des arguments cliniques et biologiques. Dans les cas, fréquents (Shanmugasundaram et al.), où une infection est suspectée alors que la culture ne retrouve pas de germe responsable, une analyse NGS pourrait permettre d'identifier des germes dans la majorité des situations. Cette approche nécessite tout de même de discuter l'implication de ces germes dans l'infection (134).

7.4. Infections de la peau et des tissus mous

L'utilisation de la métagénomique clinique dans les infections de la peau et des tissus mous a été étudiée dans une étude portant sur 69 patients. Une analyse métagénomique sans amplification préalable comparée à la culture a mis en évidence une sensibilité bien meilleure pour le séquençage (68% de prélèvements positifs) que pour la culture (32% de prélèvements positifs) avec notamment une meilleure détection des bactéries anaérobies et des mycobactéries tuberculeuses ou non. La détection de virus est aussi en partie responsable de cette bonne sensibilité avec dans l'étude la détection de HSV-1, EBV et VZV. La métagénomique apparaît adaptée dans la recherche étiologie de lésion cutanées infectieuses (Wang et al.).

7.5. Infections de pied diabétique

Les infections de pied diabétique concernent jusqu'à un patient diabétique sur trois. Leur prise en charge est difficile et manque de standardisation, de plus l'interprétation des éventuels prélèvements microbiologiques ne fait pas consensus. Les méthodes de séquençage haut débit permettent tout de même une meilleure identification des bactéries anaérobies et des champignons et ont une bonne concordance avec la culture en ce qui concerne l'identification d'un pathogène dominant. La recherche de gènes de résistance aux antibiotiques est utile au traitement, même s'il est toujours difficile de

faire le lien entre génotype et phénotype, et a une valeur pronostique sur l'évolution du patient ; la présence de bactéries multirésistantes étant prédictive d'une mauvaise évolution. La présence de gènes de virulences dont le gène codant la toxine TSST-1 chez *Staphylococcus aureus* est elle aussi prédictive d'une mauvaise évolution du patient. L'évaluation de la diversité microbienne présente est aussi un facteur pronostic, une plus grande diversité est caractéristique d'une absence de pathogène prédominant et est associée à une meilleure évolution. Lors d'ostéomyélites sur pied diabétique, la métagénomique montre une bonne performance pour détecter un pathogène en cas de culture stérile. Ces avantages sont tout de même à mettre en relation avec une grande disparité des résultats en fonction de la méthode de prélèvement et aussi du moment du prélèvement. Comme pour les méthodes classiques en ce qui concerne les infections de pied diabétique, la métagénomique souffre d'un manque de standardisation et de consensus depuis le prélèvement jusqu'à l'interprétation des résultats (Schmidt et al.).

7.6. Collection liquidiennes et abcès profonds

Nous n'avons retrouvé aucune étude informative sur la pertinence et la performance de l'utilisation de séquençage haut débit lors de la prise en charge de prélèvements de collections liquidiennes ou d'abcès. Une étude rétrospective est en cours par l'équipe de l'Hôpital Henri Mondor (Créteil) visant à comparer les performances de la métagénomique clinique à celles de la culture dans les abcès hépatiques.

7.7. Infections respiratoires

7.7.1. Infections aiguës

La recherche de virus à ARN par séquençage haut débit dans les prélèvements respiratoires hauts après dégradation de l'ADN et rétrotranscription a été réalisée en comparaison avec des PCR multiplex, les deux techniques sont bien corrélées malgré une plus faible sensibilité de la technique NGS. Pour le moment, cette sensibilité moindre de l'ordre d'un ou deux logs ne permet pas de se substituer aux PCR multiplex dans le cadre d'une recherche d'un virus respiratoire, cependant, le séquençage haut débit permet d'accéder à d'autres informations (Thorburn et al.). On peut affiner la taxonomie du virus retrouvé, typer les souches pour suivre l'épidémiologie des infections virales ou détecter d'éventuelles épidémies et, dans certaines pathologies, retrouver des mutations entraînant une résistance à un antiviral. C'est le cas pour la

grippe et la résistance à l'oseltamivir (Thorburn et al.; Petty et al.). La recherche d'un polymorphisme associé à une forme de sévérité inhabituelle serait possible aussi (Houng et al.; Selleri et al.). On retrouve une corrélation entre les CT en RT-PCR et le nombre de *reads* en NGS, ce qui donne un aspect quantitatif aux résultats de NGS (Prachayangprecha et al.) permettant d'interpréter l'imputabilité de la symptomatologie au virus détecté (DeVincenzo et al.). L'utilisation du séquençage haut débit dans le suivi épidémiologique des infections à SARS-CoV-2 a engendré une très grande quantité de données, ce qui a permis un suivi très fin de la pandémie (91).

En ce qui concerne la recherche bactériologique, la comparaison avec la culture permet d'évaluer les performances du séquençage haut débit. Après amplification 16S, la diversité bactérienne retrouvée par NGS est régulièrement plus élevée que par bactériologie standard. L'amplification préalable 16S permet d'enrichir le prélèvement en ADN bactérien mais au détriment de l'exhaustivité de la diversité microbienne retrouvée. En effet, on n'obtient pas les virus ni les champignons ni certaines bactéries qui ne répondraient pas suffisamment bien aux amorces utilisées pour la PCR. Une profondeur de lecture importante permettrait de s'affranchir de cette étape, les *reads* humains peuvent être éliminés via un outil bio-informatique et le nombre de *reads* étant important, il resterait suffisamment de *reads* bactériens pour identifier la diversité bactérienne (Toma et al.).

Sans amplification préalable, certains protocoles permettent de donner des résultats satisfaisants. Après déplétion en ADN humain avec de la saponine, de très bons résultats sont obtenus par séquençage nanopore MinION sur des expectorations pour diagnostiquer des infections respiratoires basses avec une sensibilité et une spécificité proches de 100%. La couverture du génome des pathogènes est suffisante pour envisager la recherche de gènes de résistance, cependant, la présence d'une flore respiratoire contaminante, vivante ou non, mime la présence de gènes de résistance n'appartenant pas aux pathogènes identifiés (Charalampous et al.).

7.7.2. *Mucoviscidose*

Aujourd'hui, le microbiote respiratoire des patients atteints de mucoviscidose est suivi en routine par culture, principalement sur des expectorations, induites ou non. Ce suivi est utile pour voir l'évolution du pronostic des patients ainsi que pour avoir une documentation préalable à un épisode d'exacerbation infectieuse.

L'utilisation de séquençage haut débit pour étudier le microbiote respiratoire chez ces patients est actuellement à l'état de recherche mais des applications cliniques pourraient survenir. Il est possible de détecter la présence de bactéries anaérobies, non recherchés traditionnellement par culture sur prélèvements respiratoires ou des bactéries portant des facteurs de virulence particuliers comme des gènes codant pour des toxines associées à une évolution plus ou moins rapide de la maladie. En plus du microbiote bactérien, le séquençage haut débit permet de mettre en évidence les microbiotes viraux, fongiques et archaebactériens (Françoise and Héry-Arnaud).

Il y a un essai clinique randomisé et contrôlé en cours comparant la prise en charge antibiotique des exacerbations de mucoviscidose avec et sans métagénomique clinique (CFMATTERS - *Cystic Fibrosis Research Project*).

Dans un *case report* relatant la prise en charge et le décès d'un patients atteint de mucoviscidose lors d'une exacerbation, une analyse métagénomique et métatranscriptomique a été réalisée *a posteriori* sur les prélèvements respiratoires. La culture retrouvait *P. aeruginosa*, *S. maltophilia*, *A. terreus* et des levures. La recherche métagénomique retrouve une flore avec les genres *Bacillus*, *Escherichia* et *Streptococcus* principalement. L'analyse transcriptomique retrouve cependant que 97,1% des ARNm séquencés appartiennent au genre *Pseudomonas* et 1,9% au genre *Escherichia*, montrant bien que l'abondance d'une bactérie n'est pas forcément le facteur le plus important pour déterminer son implication dans un processus infectieux. La présence d'*Escherichia coli* étant inhabituelle, une analyse taxonomique poussée a été réalisée ce qui montrera la présence d'un *Escherichia coli* producteur de Shiga-toxine 1 et 2. Chez *Pseudomonas*, l'analyse retrouve la présence de transcrits codant notamment pour MexAB-OprM et d'autres mécanismes de résistance. Ces analyses montrent bien la possibilité d'obtenir des informations supplémentaires par rapport à la culture permettant une prise en charge particulière (Cobián Güemes et al.).

7.8. Infections ophtalmologiques

Les méthodes de séquençage haut débit ont permis d'étudier le microbiote oculaire, notamment celui de la surface cornéenne. De façon inattendue, des études ont retrouvé un microbiote riche comportant des bactéries commensales, environnementales et des pathogènes opportunistes. Ces études sont à mettre en perspective avec la possibilité de contamination de réactifs de laboratoires par des fragments d'ADN (Gallon et al.).

Lors de diagnostic direct, une étude a montré la possibilité de retrouver localement du matériel génétique appartenant à des virus (EBV et HHV-8) qui auraient pu jouer un rôle oncogène chez deux patients atteints de lymphomes oculaires (Gonzales et al.).

Chez une patiente présentant un érythème diffus, des yeux rouges gonflés et douloureux avec des larmoiements, après épuisement des hypothèses diagnostiques, une recherche de pathogènes sans amplification préalable par séquençage haut débit a pu mettre en évidence la présence de *Brugia malayi* signant une filariose. La réponse favorable au traitement confirma le diagnostic (Gao et al.).

Partant du constat que les prélèvements ophtalmologiques étaient très fréquemment non contributifs lors de la recherche d'infection intraoculaire et que les patients étaient traités par des antibiothérapies probabilistes à large spectre, plusieurs études ont été réalisées pour déterminer l'apport du séquençage haut débit dans cette indication. La détermination de l'étiologie permettrait de rétrocéder à une antibiothérapie adaptée et de réduire l'impact écologique du traitement. Par comparaison à un panel de PCR spécifiques, sur des prélèvements d'humeur vitrée une méthode de séquençage haut débit sur ADN par méthode Illumina sans amplification préalable permet une bonne sensibilité et retrouve probablement des pathogènes supplémentaires, que ce soit par une plus grande sensibilité ou par une recherche non ciblée. Ces résultats concernent des uvéites bactériennes, virales et parasitaires (Doan, Acharya, et al.; Doan, Wilson, et al.). De plus, par séquençage complet, il est possible de mettre en évidence la résistance à certains antimicrobiens, par exemple au ganciclovir chez CMV (Doan, Acharya, et al.) ou aux antibiotiques dans les infections bactériennes (Kirstahler et al.).

Une étude a démontré l'efficacité et la précision permise par la métagénomique appliquée à des milieux de conservation de dons de cornée. Par rapport à la culture, la métagénomique offre une précision taxonomique accrue mais la présence de matériel génétique bactérien dans le milieu de conservation ne signe pas forcément la contamination ni la dégradation du milieu ; le résultat n'est pas non plus informatif sur la possibilité d'utiliser le greffon et enfin, le coût de l'analyse est élevé. Par conséquent, l'utilisation de la métagénomique sur les milieux de conservation de dons de cornées n'est aujourd'hui pas pertinent en routine (Parekh et al.).

7.9. Virus hépatotropes

Dans les pays développés, l'étiologie des hépatites aiguës n'est pas retrouvée dans 10 à 15% des cas. L'hypothèse d'une infection par un virus hépatotrope non retrouvé est

souvent évoquée (VHA, VHB, VHC, VHE, Adénovirus, HSV, EBV, VZV). Le diagnostic étiologie par séquençage haut débit sans amplification préalable sur sérum a été évalué dans une étude américaine en 2017. Tous les témoins négatifs sont bien revenus négatifs et tous les témoins positifs sont bien revenus positifs. Sur les 204 sérums, 27 ont été positifs soit en métagénomique, soit par les méthodes de référence. La métagénomique a réussi à diagnostiquer 7 des 8 témoins positifs à VHA et VHB, le témoin positif non diagnostiqué est un prélèvement positif à VHB par PCR pour lequel le séquençage haut débit n'avait attribué que deux *reads* à VHB avec un seul de positivité défini à 5. De plus, sur un des témoins positifs à VHB, ils ont retrouvé une coïnfection à HSV-1. Sur les 11 prélèvements dont l'étiologie était suspectée mais incertaine lors de la prise en charge, l'analyse métagénomique a réussi à confirmer le diagnostic dans 8 cas. Pour les trois cas non confirmés, la recherche étiologique par amplification génique n'avait pas non plus donné de résultat. Finalement, le séquençage haut débit retrouve un virus pour 8 cas dont l'étiologie était inconnue : 3 cas d'HSV-1, un cas de VHB, un cas de parvovirus B19, un cas de HHV-7, un cas de CMV et un cas de HPV-159. Pour ce dernier, HPV-159, la pertinence clinique de cette identification est à remettre en cause étant donné que c'est un virus appartenant au microbiote cutané.

Le *mapping* des séquences obtenues sur des génomes de référence a permis de reconstituer les génomes complets ou au moins les gènes d'intérêt pour la résistance pour 10 échantillons de VHB, VHC, HSV-1. Parmi ceux-ci, on retrouve une mutation associée à une résistance à l'aciclovir chez une souche HSV-1 et une associée à une résistance au telaprevir chez une souche de VHC.

L'interprétation des données de métagénomique doit tout de même prendre en compte la totalité du dossier clinique pour éviter de conclure à tort à une étiologie. Par exemple, lors d'une réactivation d'HSV-1, il n'est pas rare de retrouver une partie de son génome dans le sérum. Dans cette même étude, sur 4 patients immunodéprimés pour lesquels on a retrouvé de l'ADN d'HSV-1 dans le sérum, 2 ont survécu sans que l'étiologie ne soit retrouvée pendant la prise en charge ; la pertinence d'une découverte d'ADN d'HSV-1 est donc à remettre en perspective, au moins dans ces deux cas (Somasekar et al.).

7.10. Infections par le VIH

Le séquençage du génome du VIH est réalisé en routine pour la prédiction d'éventuelles résistances aux antirétroviraux. Celui-ci est habituellement réalisé par

séquençage Sanger mais de plus en plus de laboratoires ont adapté des techniques de séquençages haut débit. Outre la grande quantité de séquences à faible coût, l'utilisation de certaines techniques permettent de séquencer, non seulement la population majoritaire comme avec la méthode Sanger, mais aussi les populations minoritaires ; ceci permettrait une meilleure prédiction des échecs de traitements (Ávila-Ríos et al.). Le séquençage du VIH est un sujet spécifique nécessitant une expertise et dont le volume de la bibliographie disponible est très important ; ce sujet ne sera donc pas traité dans ce travail.

7.11. Infections virales tropicales

7.11.1. Fièvres hémorragiques virales

La fièvre hémorragique de Lassa est la fièvre hémorragique la plus étudiée par métagénomique clinique dans la littérature. Sans enrichissement, le séquençage sur 120 prélèvements cliniques a permis l'identification du pathogène mais aussi le séquençage complet de chaque souche (L. E. Kafetzopoulou et al.). Une méthode d'enrichissement virale à large spectre mais non exhaustive, basée sur l'amplification de nombreux arbovirus et virus de fièvres hémorragiques différents par des amorces complémentaires à des séquences conservées, donne aussi de bon résultats pour la fièvre jaune et Ebola (Deng et al.).

7.11.2. Arboviroses

La recherche de dengue ou de chikungunya peut être réalisée par séquençage haut débit directement sur plasma ou sur sérum après extraction sélective d'ARN avec une sensibilité équivalente à la PCR. Aussi, le séquençage complet de la souche (>95% du génome) est estimé possible dans 90% des cas de chikungunya et le séquençage de 75% du génome est estimé possible dans 80% des cas de dengue. Ces deux estimations sur Illumina MySeq et MinION ont été réalisées à partir des CT de PCR permettant une couverture suffisamment large (CT jusqu'à 30 cycles). Dans une étude britannique, l'Illumina MySeq a donné un nombre médian de 1,4 millions de *reads* totaux et de 373 mille *reads* appariés au génome du virus du chikungunya pour une couverture médiane à 20x de 97,6% ; pour le virus de la dengue, il retrouve 782 mille *reads* totaux médians et avec une médiane de 429 mille *reads* pour une couverture médiane à 20x de 99,0% (Liana E. Kafetzopoulou et al.).

La recherche de dengue et de chikungunya mais aussi du virus Zika, du virus de Powassan et du virus Usutu a aussi été démontrée par métagénomique après enrichissement viral (Deng et al.).

8. Discussion

Les récentes techniques de séquençage haut débit ont révolutionné la recherche fondamentale avec la possibilité d'augmenter fortement la quantité de génomes séquencés pour un coût par génome très inférieur au séquençage Sanger. Cependant, dans la pratique courante en microbiologie clinique, ces méthodes de séquençage ont encore une place limitée du fait de leur coût et de la nécessité de formation sur ces technologies.

Lors de l'implantation du séquençage haut débit dans un laboratoire de microbiologie, plusieurs questions se posent. Tout d'abord, le choix de l'automate de séquençage et donc de la technologie utilisée doit dépendre de l'usage souhaité. En effet, selon les applications, on tirera plus ou moins profit de *reads* courts avec peu d'erreurs desquels l'on ne pourra pas assembler de génome complet circularisé ou de *reads* longs contenant plus d'erreurs desquels les mutations ponctuelles seront plus difficile à distinguer.

Le support bio-informatique est primordial dans l'usage de méthodes de séquençage haut débit et nécessite la présence d'un personnel formé et de matériel adéquat avec une puissance de calcul suffisante. Rappelons que l'usage de programmes en lignes de commandes UNIX étant courant, un personnel acculturé à leur utilisation facilite la prise en main.

Les usages majoritaires du séquençage haut débit dans les laboratoires de microbiologie que sont la métagénomique clinique, l'étude phylogénétique ainsi que la recherche de résistance apportent des options nouvelles dans la prise en charge des patients. Pour ce qui est de la métagénomique, celle-ci apporte une nouvelle ligne d'investigation diagnostique lorsque les méthodes habituelles, biologie moléculaire comprise, sont prises en défaut ; son positionnement reste à définir pour qu'elle soit utilisée de façon raisonnable mais suffisamment précocement pour améliorer la prise en charge du patient. Le séquençage complet des souches, quant à lui, permet notamment une analyse phylogénétique particulièrement fine capable de classer les souches entre elles à la mutation ponctuelle près et de réaliser, non seulement des dépistages d'épidémies, mais aussi de décrire celles-ci de façon précise en de retracer la

chronologie de l'apparition de chaque souche. Ce séquençage complet permet aussi la recherche de résistances génotypiques lorsque la recherche de résistance phénotypique n'est pas disponible ou non concluante.

Ces promesses sont tout de même à évaluer au contact de la réalité de la prise en charge des patients. Effectivement, selon le type d'infection ou le type de micro-organisme, les méthodes décrites dans ce travail sont plus ou moins efficaces et leur interprétation est difficile et nécessite un personnel expérimenté et à même d'identifier les limites de son outil.

Le travail bibliographique dont résulte ce document devra permettre le développement et l'utilisation de méthodes de séquençage haut débit pour le diagnostic microbiologique, le géotypage, la recherche de résistances aux anti-infectieux et l'investigation d'éventuelles épidémies au CHU de Rouen. Cette synthèse théorique est complétée par la réalisation d'un Master 2 portant sur la génomique et l'épidémiogénétique des infections à *Streptococcus agalactiae* à l'Institut Cochin (Paris XIVe) et d'un stage d'internat hors subdivision à l'Hôpital Henri Mondor (Créteil) axé sur l'utilisation pratique de la métagénomique clinique en routine. Les compétences ainsi acquises serviront à faciliter le projet.

Cette thèse bibliographique sera complétée dans un second temps par un travail expérimental visant à valider les différentes phases de l'analyse métagénomique en tenant compte de l'expérience acquise lors du stage hors subdivision à l'Hôpital Henri Mondor (Créteil).

La première partie de ce travail se focalisera sur la phase préanalytique avec la mesure de l'influence de chaque étape de traitement avant l'analyse. Un prélèvement clinique simulé sera réalisé par le mélange d'un liquide de lavage broncho-alvéolaire d'un patient positif à *Aspergillus fumigatus* en PCR avec un CT autour de 25 cycles, d'un prélèvement respiratoire de patient positif à adénovirus en PCR, et d'une suspension bactérienne de 300µL à 0,5 MF de chacune des bactéries *Streptococcus pneumoniae*, *Neisseria meningitidis* et *Staphylococcus aureus*. Ce prélèvement clinique simulé sera scindé et chaque partie subira, ou non, différents traitements parmi une étape de centrifugation, un cycle de congélation et de décongélation, une pré-extraction par broyage à billes et un protocole de déplétion de l'ADN de l'hôte. Le rendement de ces étapes sera évalué par PCR en temps réel ciblant les microorganismes pathogènes constituant le prélèvement. Ces analyses permettront de savoir quels pathogènes sont

affectées par ces étapes et notamment l'impact des centrifugations et de la déplétion en ADN de l'hôte. En effet, un protocole de déplétion d'ADN humain utilisant de la saponine pour lyser les cellules humaines fait intervenir des étapes de centrifugation qui entraînent la concentration des bactéries dans le culot mais dont l'impact sur les virus est incertain ; ceux-ci pourrait se concentrer dans le surnageant. Ces analyses sont résumées dans la Figure 17.

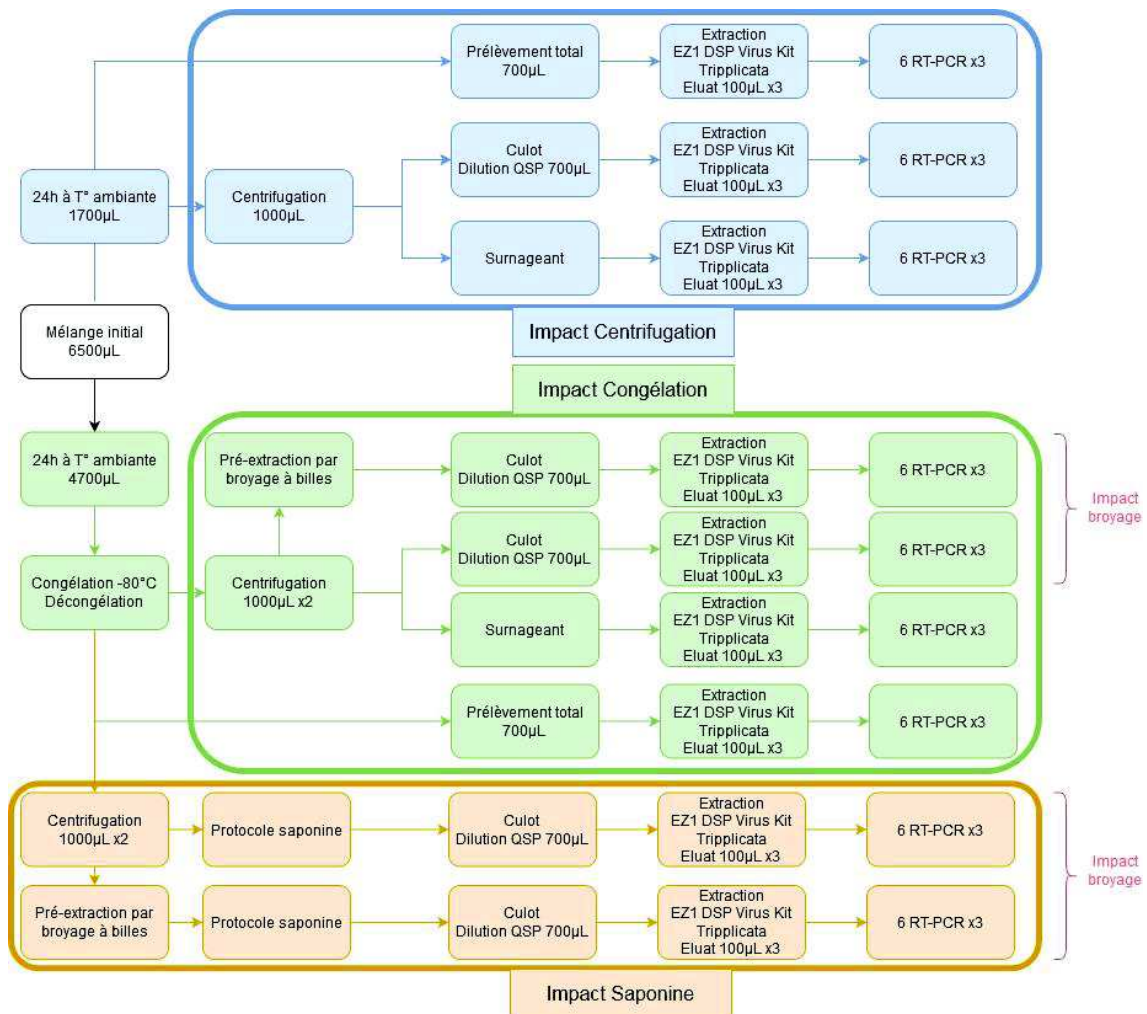


Figure 17 : Diagramme récapitulatif des essais permettant la validation de la phase préanalytique du diagnostic métagénomique.

Dans un deuxième temps, l'influence d'une pré-amplification par PCR 16S (18S ou ITS chez les champignons) sur la détection de pathogènes sera évaluée. Une gamme de différentes concentrations de *Staphylococcus aureus* sera incorporée à une matrice polymicrobienne connue en culture (un liquide péritonéal ou un liquide de lavage broncho-alvéolaire polymicrobien) puis subira une analyse métagénomique avant et après amplification 16S.

Enfin, deux technologies sont disponibles au CHU de Rouen, le séquençage par synthèse Illumina et le séquençage Nanopore. Ces deux méthodes seront comparées pour si l'une, l'autre ou la combinaison des deux est la plus adaptée à un diagnostic métagénomique.

Références Bibliographiques

- Adams, M. J., et al. “Recently Agreed Changes to the International Code of Virus Classification and Nomenclature.” *Archives of Virology*, vol. 158, no. 12, Dec. 2013, pp. 2633–39. *PubMed*, <https://doi.org/10.1007/s00705-013-1749-9>.
- Afgan, Enis, et al. “The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update.” *Nucleic Acids Research*, vol. 46, no. W1, July 2018, pp. W537–44. *Silverchair*, <https://doi.org/10.1093/nar/gky379>.
- Al Masalma, Mouhamad, et al. “Metagenomic Analysis of Brain Abscesses Identifies Specific Bacterial Associations.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, vol. 54, no. 2, Jan. 2012, pp. 202–10. *PubMed*, <https://doi.org/10.1093/cid/cir797>.
- Andersen, Camilla, et al. “Culture on Selective Media and Amplicon-Based Sequencing of 16S rRNA from Spontaneous Brain Abscess—the View from the Diagnostic Laboratory.” *Microbiology Spectrum*, vol. 10, no. 2, pp. e02407-21. *PubMed Central*, <https://doi.org/10.1128/spectrum.02407-21>.
- Anson, Luke W., et al. “DNA Extraction from Primary Liquid Blood Cultures for Bloodstream Infection Diagnosis Using Whole Genome Sequencing.” *Journal of Medical Microbiology*, vol. 67, no. 3, Mar. 2018, pp. 347–57. *PubMed*, <https://doi.org/10.1099/jmm.0.000664>.
- Ávila-Ríos, Santiago, et al. “Next-Generation Sequencing for HIV Drug Resistance Testing: Laboratory, Clinical, and Implementation Considerations.” *Viruses*, vol. 12, no. 6, June 2020, p. 617. *PubMed Central*, <https://doi.org/10.3390/v12060617>.
- Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 23 June 2021.
- Bai, Yunmeng, et al. “Comprehensive Evolution and Molecular Characteristics of a Large Number of SARS-CoV-2 Genomes Reveal Its Epidemic Trends.” *International Journal of Infectious Diseases*, vol. 100, Nov. 2020, pp. 164–73. *PubMed Central*, <https://doi.org/10.1016/j.ijid.2020.08.066>.
- Bankevich, Anton, et al. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” *Journal of Computational Biology: A*

- Journal of Computational Molecular Cell Biology*, vol. 19, no. 5, May 2012, pp. 455–77. *PubMed*, <https://doi.org/10.1089/cmb.2012.0021>.
- Beghini, Francesco, et al. “Integrating Taxonomic, Functional, and Strain-Level Profiling of Diverse Microbial Communities with BioBakery 3.” *ELife*, vol. 10, May 2021, p. e65088. *PubMed*, <https://doi.org/10.7554/eLife.65088>.
- Bentley, David R., et al. “Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry.” *Nature*, vol. 456, no. 7218, Nov. 2008, pp. 53–59. *PubMed Central*, <https://doi.org/10.1038/nature07517>.
- Blauwkamp, Timothy A., et al. “Analytical and Clinical Validation of a Microbial Cell-Free DNA Sequencing Test for Infectious Disease.” *Nature Microbiology*, vol. 4, no. 4, Apr. 2019, pp. 663–74. *PubMed*, <https://doi.org/10.1038/s41564-018-0349-6>.
- Bolger, Anthony M., et al. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics*, vol. 30, no. 15, Aug. 2014, pp. 2114–20. *PubMed Central*, <https://doi.org/10.1093/bioinformatics/btu170>.
- Boolchandani, Manish, et al. “Sequencing-Based Methods and Resources to Study Antimicrobial Resistance.” *Nature Reviews. Genetics*, vol. 20, no. 6, June 2019, pp. 356–70. *PubMed Central*, <https://doi.org/10.1038/s41576-019-0108-4>.
- Bortolaia, Valeria, et al. “ResFinder 4.0 for Predictions of Phenotypes from Genotypes.” *The Journal of Antimicrobial Chemotherapy*, vol. 75, no. 12, Dec. 2020, pp. 3491–500. *PubMed*, <https://doi.org/10.1093/jac/dkaa345>.
- Buchfink, Benjamin, et al. “Fast and Sensitive Protein Alignment Using DIAMOND.” *Nature Methods*, vol. 12, no. 1, Jan. 2015, pp. 59–60. *PubMed*, <https://doi.org/10.1038/nmeth.3176>.
- Cai, Sishi, et al. “The Clinical Value of Valve Metagenomic Next-Generation Sequencing When Applied to the Etiological Diagnosis of Infective Endocarditis.” *Annals of Translational Medicine*, vol. 9, no. 19, Oct. 2021, p. 1490. *PubMed*, <https://doi.org/10.21037/atm-21-2488>.
- CFMATTERS - Cystic Fibrosis Research Project. <https://www.cfmatters.eu/>. Accessed 23 June 2021.
- Chao, Lingshan, et al. “Application of next Generation Sequencing-Based Rapid Detection Platform for Microbiological Diagnosis and Drug Resistance Prediction in Acute Lower Respiratory Infection.” *Annals of Translational*

- Medicine*, vol. 8, no. 24, Dec. 2020, p. 1644. *PubMed*, <https://doi.org/10.21037/atm-20-7081>.
- Charalampous, Themoula, et al. “Nanopore Metagenomics Enables Rapid Clinical Diagnosis of Bacterial Lower Respiratory Infection.” *Nature Biotechnology*, vol. 37, no. 7, 7, July 2019, pp. 783–92. *www.nature.com*, <https://doi.org/10.1038/s41587-019-0156-5>.
- Chen, Lihong, et al. “VFDB: A Reference Database for Bacterial Virulence Factors.” *Nucleic Acids Research*, vol. 33, no. Database issue, Jan. 2005, pp. D325-328. *PubMed*, <https://doi.org/10.1093/nar/gki008>.
- Cheng, Jun, et al. “Identification of Pathogens in Culture-Negative Infective Endocarditis Cases by Metagenomic Analysis.” *Annals of Clinical Microbiology and Antimicrobials*, vol. 17, no. 1, Dec. 2018, p. 43. *PubMed*, <https://doi.org/10.1186/s12941-018-0294-5>.
- Chomczynski, P., and N. Sacchi. “Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction.” *Analytical Biochemistry*, vol. 162, no. 1, Apr. 1987, pp. 156–59. *PubMed*, <https://doi.org/10.1006/abio.1987.9999>.
- Cnops, Lieselotte, and Marjan Van Esbroeck. “Freezing of Stool Samples Improves Real-Time PCR Detection of *Entamoeba Dispar* and *Entamoeba Histolytica*.” *Journal of Microbiological Methods*, vol. 80, no. 3, Mar. 2010, pp. 310–12. *PubMed*, <https://doi.org/10.1016/j.mimet.2010.01.001>.
- Cobián Güemes, Ana Georgina, et al. “Cystic Fibrosis Rapid Response: Translating Multi-Omics Data into Clinically Relevant Information.” *MBio*, vol. 10, no. 2, Apr. 2019, pp. e00431-19. *PubMed Central*, <https://doi.org/10.1128/mBio.00431-19>.
- COWAN, S. T. YR 1965. “Principles and Practice of Bacterial Taxonomy—a Forward Look.” *Microbiology*, vol. 39, no. 1, pp. 143–53. *Microbiology Society Journals*, <https://doi.org/10.1099/00221287-39-1-143>.
- Croucher, Nicholas J., et al. “Rapid Phylogenetic Analysis of Large Samples of Recombinant Bacterial Whole Genome Sequences Using Gubbins.” *Nucleic Acids Research*, vol. 43, no. 3, Feb. 2015, p. e15. *PubMed*, <https://doi.org/10.1093/nar/gku1196>.

- Da Cunha, Violette, et al. “Streptococcus Agalactiae Clones Infecting Humans Were Selected and Fixed through the Extensive Use of Tetracycline.” *Nature Communications*, vol. 5, Aug. 2014, p. 4544. *PubMed*, <https://doi.org/10.1038/ncomms5544>.
- Davis, James J., Sébastien Boisvert, et al. “Antimicrobial Resistance Prediction in PATRIC and RAST.” *Scientific Reports*, vol. 6, June 2016. *PubMed Central*, <https://doi.org/10.1038/srep27930>.
- Davis, James J., Alice R. Wattam, et al. “The PATRIC Bioinformatics Resource Center: Expanding Data and Analysis Capabilities.” *Nucleic Acids Research*, vol. 48, no. D1, Jan. 2020, pp. D606–12. *PubMed*, <https://doi.org/10.1093/nar/gkz943>.
- de Been, Mark, et al. “Core Genome Multilocus Sequence Typing Scheme for High-Resolution Typing of *Enterococcus Faecium*.” *Journal of Clinical Microbiology*, vol. 53, no. 12, Dec. 2015, pp. 3788–97. *PubMed*, <https://doi.org/10.1128/JCM.01946-15>.
- de Boer, Richard, et al. “Improved Detection of Microbial DNA after Bead-Beating before DNA Isolation.” *Journal of Microbiological Methods*, vol. 80, no. 2, Feb. 2010, pp. 209–11. *PubMed*, <https://doi.org/10.1016/j.mimet.2009.11.009>.
- Deng, Xianding, et al. “Metagenomic Sequencing with Spiked Primer Enrichment for Viral Diagnostics and Genomic Surveillance.” *Nature Microbiology*, vol. 5, no. 3, Mar. 2020, pp. 443–54. *PubMed*, <https://doi.org/10.1038/s41564-019-0637-9>.
- DeVincenzo, John P., et al. “Viral Load Drives Disease in Humans Experimentally Infected with Respiratory Syncytial Virus.” *American Journal of Respiratory and Critical Care Medicine*, vol. 182, no. 10, Nov. 2010, pp. 1305–14. *PubMed*, <https://doi.org/10.1164/rccm.201002-0221OC>.
- Di Bella, Julia M., et al. “High Throughput Sequencing Methods and Analysis for Microbiome Research.” *Journal of Microbiological Methods*, vol. 95, no. 3, Dec. 2013, pp. 401–14. *PubMed*, <https://doi.org/10.1016/j.mimet.2013.08.011>.
- Doan, Thuy, Michael R. Wilson, et al. “Illuminating Uveitis: Metagenomic Deep Sequencing Identifies Common and Rare Pathogens.” *Genome Medicine*, vol. 8, no. 1, Aug. 2016. *PubMed Central*, <https://doi.org/10.1186/s13073-016-0344-6>.
- Doan, Thuy, Nisha R. Acharya, et al. “Metagenomic DNA Sequencing for the Diagnosis of Intraocular Infections.” *Ophthalmology*, vol. 124, no. 8, Aug. 2017, pp. 1247–48. *PubMed Central*, <https://doi.org/10.1016/j.ophtha.2017.03.045>.

- Elsner, H. I., and E. B. Lindblad. "Ultrasonic Degradation of DNA." *DNA (Mary Ann Liebert, Inc.)*, vol. 8, no. 10, Dec. 1989, pp. 697–701. *PubMed*, <https://doi.org/10.1089/dna.1989.8.697>.
- Fleischmann, R. D., et al. "Whole-Genome Random Sequencing and Assembly of Haemophilus Influenzae Rd." *Science (New York, N.Y.)*, vol. 269, no. 5223, July 1995, pp. 496–512. *PubMed*, <https://doi.org/10.1126/science.7542800>.
- Foulongne, Vincent, et al. "Human Skin Microbiota: High Diversity of DNA Viruses Identified on the Human Skin by High Throughput Sequencing." *PLoS One*, vol. 7, no. 6, 2012, p. e38499. *PubMed*, <https://doi.org/10.1371/journal.pone.0038499>.
- Françoise, Alice, and Geneviève Héry-Arnaud. "The Microbiome in Cystic Fibrosis Pulmonary Disease." *Genes*, vol. 11, no. 5, May 2020, p. E536. *PubMed*, <https://doi.org/10.3390/genes11050536>.
- Gagneux, Sebastien. "Ecology and Evolution of Mycobacterium Tuberculosis." *Nature Reviews. Microbiology*, vol. 16, no. 4, Apr. 2018, pp. 202–13. *PubMed*, <https://doi.org/10.1038/nrmicro.2018.8>.
- Gallon, Paola, et al. "Metagenomics in Ophthalmology: Hypothesis or Real Prospective?" *Biotechnology Reports*, vol. 23, June 2019. *PubMed Central*, <https://doi.org/10.1016/j.btre.2019.e00355>.
- Gao, Dian, et al. "Diagnosis of a Malayan Filariasis Case Using a Shotgun Diagnostic Metagenomics Assay." *Parasites & Vectors*, vol. 9, Feb. 2016. *PubMed Central*, <https://doi.org/10.1186/s13071-016-1363-2>.
- Glaser, C. A., et al. "Beyond Viruses: Clinical Profiles and Etiologies Associated with Encephalitis." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, vol. 43, no. 12, Dec. 2006, pp. 1565–77. *PubMed*, <https://doi.org/10.1086/509330>.
- Glaser, Carol A., et al. "In Search of Encephalitis Etiologies: Diagnostic Challenges in the California Encephalitis Project, 1998-2000." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, vol. 36, no. 6, Mar. 2003, pp. 731–42. *PubMed*, <https://doi.org/10.1086/367841>.
- Gonzales, John, et al. "Metagenomic Deep Sequencing of Aqueous Fluid Detects Intraocular Lymphomas." *The British Journal of Ophthalmology*, vol. 102, no.

- 1, Jan. 2018, pp. 6–8. *PubMed Central*, <https://doi.org/10.1136/bjophthalmol-2017-311151>.
- Granerod, Julia, et al. “Causes of Encephalitis and Differences in Their Clinical Presentations in England: A Multicentre, Population-Based Prospective Study.” *The Lancet. Infectious Diseases*, vol. 10, no. 12, Dec. 2010, pp. 835–44. *PubMed*, [https://doi.org/10.1016/S1473-3099\(10\)70222-X](https://doi.org/10.1016/S1473-3099(10)70222-X).
- Greninger, Alexander L., and Samia N. Naccache. “Metagenomics to Assist in the Diagnosis of Bloodstream Infection.” *The Journal of Applied Laboratory Medicine*, vol. 3, no. 4, Jan. 2019, pp. 643–53. *Silverchair*, <https://doi.org/10.1373/jalm.2018.026120>.
- Grokhovsky, Sergei L., et al. “Sequence-Specific Ultrasonic Cleavage of DNA.” *Biophysical Journal*, vol. 100, no. 1, Jan. 2011, pp. 117–25. *PubMed Central*, <https://doi.org/10.1016/j.bpj.2010.10.052>.
- Gupta, Sushim Kumar, et al. “ARG-ANNOT, a New Bioinformatic Tool to Discover Antibiotic Resistance Genes in Bacterial Genomes.” *Antimicrobial Agents and Chemotherapy*, vol. 58, no. 1, 2014, pp. 212–20. *PubMed*, <https://doi.org/10.1128/AAC.01310-13>.
- Harrington, Colleen T., et al. “Fundamentals of Pyrosequencing.” *Archives of Pathology & Laboratory Medicine*, vol. 137, no. 9, Sept. 2013, pp. 1296–303. *PubMed*, <https://doi.org/10.5858/arpa.2012-0463-RA>.
- Head, Steven R., et al. “Library Construction for Next-Generation Sequencing: Overviews and Challenges.” *BioTechniques*, vol. 56, no. 2, 2014, p. 61. www.ncbi.nlm.nih.gov, <https://doi.org/10.2144/000114133>.
- Helmersen, Karin, and Hege Vangstein Aamot. “DNA Extraction of Microbial DNA Directly from Infected Tissue: An Optimized Protocol for Use in Nanopore Sequencing.” *Scientific Reports*, vol. 10, no. 1, 19 2020, p. 2985. *PubMed*, <https://doi.org/10.1038/s41598-020-59957-6>.
- Heravi, Fatemah Sadeghpour, et al. “Host DNA Depletion Efficiency of Microbiome DNA Enrichment Methods in Infected Tissue Samples.” *Journal of Microbiological Methods*, vol. 170, Mar. 2020, p. 105856. *PubMed*, <https://doi.org/10.1016/j.mimet.2020.105856>.
- Holohan, Cian, et al. “Influence of Viral Transport Media and Freeze-Thaw Cycling on the Sensitivity of QRT-PCR Detection of SARS-CoV-2 Nucleic Acids.”

- Nanoscale*, vol. 13, no. 37, Oct. 2021, pp. 15659–67. *PubMed*, <https://doi.org/10.1039/d1nr03933d>.
- Houng, Huo-Shu H., et al. “Emergent 2009 Influenza A(H1N1) Viruses Containing HA D222N Mutation Associated with Severe Clinical Outcomes in the Americas.” *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, vol. 53, no. 1, Jan. 2012, pp. 12–15. *PubMed*, <https://doi.org/10.1016/j.jcv.2011.09.004>.
- “How It Works.” *Oxford Nanopore Technologies*, <http://nanoporetech.com/how-it-works>. Accessed 14 Jan. 2021.
- Hu, Hui-Li, et al. “Evaluation of Next-Generation Sequencing for the Pathogenic Diagnosis of Children Brain Abscesses.” *The Journal of Infection*, vol. 78, no. 4, Apr. 2019, pp. 323–37. *PubMed*, <https://doi.org/10.1016/j.jinf.2019.01.003>.
- Hu, Zhiliang, et al. “Metagenomic Next-Generation Sequencing as a Diagnostic Tool for Toxoplasmic Encephalitis.” *Annals of Clinical Microbiology and Antimicrobials*, vol. 17, no. 1, Dec. 2018, p. 45. *PubMed*, <https://doi.org/10.1186/s12941-018-0298-1>.
- Huang, T., et al. “Rapid and Accurate Diagnosis of Brain Abscess Caused by *Nocardia Asiatica* with a Combination of Ziehl-Neelsen Staining and Metagenomics next-Generation Sequencing.” *European Journal of Neurology*, vol. 28, no. 1, Jan. 2021, pp. 355–57. *PubMed*, <https://doi.org/10.1111/ene.14533>.
- Huson, Daniel H., et al. “MEGAN Analysis of Metagenomic Data.” *Genome Research*, vol. 17, no. 3, Mar. 2007, pp. 377–86. *PubMed*, <https://doi.org/10.1101/gr.5969107>.
- Illumina DNA Prep | Flexibility for Many Whole-Genome Sequencing Applications*. <https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/nextera-dna-flex.html>. Accessed 11 Jan. 2021.
- Illumina Sequencing by Synthesis*. Directed by Illumina, 2016. *YouTube*, https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=emb_title.
- Jacovides, Christina L., et al. “Successful Identification of Pathogens by Polymerase Chain Reaction (PCR)-Based Electron Spray Ionization Time-of-Flight Mass Spectrometry (ESI-TOF-MS) in Culture-Negative Periprosthetic Joint Infection.” *The Journal of Bone and Joint Surgery. American Volume*, vol. 94, no. 24, Dec. 2012, pp. 2247–54. *PubMed*, <https://doi.org/10.2106/JBJS.L.00210>.

- Jia, Baofeng, et al. "CARD 2017: Expansion and Model-Centric Curation of the Comprehensive Antibiotic Resistance Database." *Nucleic Acids Research*, vol. 45, no. D1, Jan. 2017, pp. D566–73. *PubMed*, <https://doi.org/10.1093/nar/gkw1004>.
- Jolley, Keith A., et al. "Ribosomal Multilocus Sequence Typing: Universal Characterization of Bacteria from Domain to Strain." *Microbiology (Reading, England)*, vol. 158, no. Pt 4, Apr. 2012, pp. 1005–15. *PubMed*, <https://doi.org/10.1099/mic.0.055459-0>.
- Kafetzopoulou, L. E., et al. "Metagenomic Sequencing at the Epicenter of the Nigeria 2018 Lassa Fever Outbreak." *Science (New York, N.Y.)*, vol. 363, no. 6422, Jan. 2019, pp. 74–77. *PubMed*, <https://doi.org/10.1126/science.aau9343>.
- Kafetzopoulou, Liana E., et al. "Assessment of Metagenomic Nanopore and Illumina Sequencing for Recovering Whole Genome Sequences of Chikungunya and Dengue Viruses Directly from Clinical Samples." *Eurosurveillance*, vol. 23, no. 50, Dec. 2018. *PubMed Central*, <https://doi.org/10.2807/1560-7917.ES.2018.23.50.1800228>.
- Kennedy, Nicholas A., et al. "The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing." *PloS One*, vol. 9, no. 2, 2014, p. e88982. *PubMed*, <https://doi.org/10.1371/journal.pone.0088982>.
- Khaleidi, Ariane, et al. "Transcriptome Profiling of Antimicrobial Resistance in *Pseudomonas Aeruginosa*." *Antimicrobial Agents and Chemotherapy*, vol. 60, no. 8, Aug. 2016, pp. 4722–33. *PubMed*, <https://doi.org/10.1128/AAC.00075-16>.
- Khetsuriani, Nino, et al. "Burden of Encephalitis-Associated Hospitalizations in the United States, 1988–1997." *Clinical Infectious Diseases*, vol. 35, no. 2, July 2002, pp. 175–82. *Silverchair*, <https://doi.org/10.1086/341301>.
- Kirstahler, Philipp, et al. "Genomics-Based Identification of Microorganisms in Human Ocular Body Fluid." *Scientific Reports*, vol. 8, Mar. 2018. *PubMed Central*, <https://doi.org/10.1038/s41598-018-22416-4>.
- Klein, C., et al. "Egg Intensity and Freeze-Thawing of Fecal Samples Affect Sensitivity of *Echinococcus Multilocularis* Detection by PCR." *Parasitology Research*, vol.

- 113, no. 10, Oct. 2014, pp. 3867–73. *PubMed*, <https://doi.org/10.1007/s00436-014-4055-x>.
- Knabl, Ludwig, et al. “Evaluation of a Multiplex OnSpot Primer-Extension PCR Assay in the Diagnosis of Sepsis.” *Journal of Microbiological Methods*, vol. 120, Jan. 2016, pp. 91–93. *PubMed*, <https://doi.org/10.1016/j.mimet.2015.12.001>.
- Kohl, Claudia, et al. “Protocol for Metagenomic Virus Detection in Clinical Specimens¹.” *Emerging Infectious Diseases*, vol. 21, no. 1, Jan. 2015, pp. 48–57. *PubMed Central*, <https://doi.org/10.3201/eid2101.140766>.
- Kwong, Jason C., et al. “Prospective Whole-Genome Sequencing Enhances National Surveillance of *Listeria Monocytogenes*.” *Journal of Clinical Microbiology*, vol. 54, no. 2, Feb. 2016, pp. 333–42. *PubMed*, <https://doi.org/10.1128/JCM.02344-15>.
- Langmead, Ben, and Steven L. Salzberg. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods*, vol. 9, no. 4, 4, Apr. 2012, pp. 357–59. www.nature.com, <https://doi.org/10.1038/nmeth.1923>.
- Li, Heng, and Richard Durbin. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics (Oxford, England)*, vol. 25, no. 14, July 2009, pp. 1754–60. *PubMed*, <https://doi.org/10.1093/bioinformatics/btp324>.
- Lim, Mi Young, et al. “Comparison of DNA Extraction Methods for Human Gut Microbial Community Profiling.” *Systematic and Applied Microbiology*, vol. 41, no. 2, Mar. 2018, pp. 151–57. *PubMed*, <https://doi.org/10.1016/j.syapm.2017.11.008>.
- Lin, Jyun-Hong, et al. “Complex Microbiome in Brain Abscess Revealed by Whole-Genome Culture-Independent and Culture-Based Sequencing.” *Journal of Clinical Medicine*, vol. 8, no. 3, Mar. 2019, p. E351. *PubMed*, <https://doi.org/10.3390/jcm8030351>.
- Long, Yun, et al. “Diagnosis of Sepsis with Cell-Free DNA by Next-Generation Sequencing Technology in ICU Patients.” *Archives of Medical Research*, vol. 47, no. 5, 2016, pp. 365–71. *PubMed*, <https://doi.org/10.1016/j.arcmed.2016.08.004>.

- Lu, Hengyun, et al. "Oxford Nanopore MinION Sequencing and Genome Assembly." *Genomics, Proteomics & Bioinformatics*, vol. 14, no. 5, Oct. 2016, pp. 265–79. *PubMed Central*, <https://doi.org/10.1016/j.gpb.2016.05.004>.
- Maggi, Fabrizio, and Mauro Bendinelli. "Human Anelloviruses and the Central Nervous System." *Reviews in Medical Virology*, vol. 20, no. 6, Nov. 2010, pp. 392–407. *PubMed*, <https://doi.org/10.1002/rmv.668>.
- Maiden, M. C., et al. "Multilocus Sequence Typing: A Portable Approach to the Identification of Clones within Populations of Pathogenic Microorganisms." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 6, Mar. 1998, pp. 3140–45. *PubMed*, <https://doi.org/10.1073/pnas.95.6.3140>.
- Maiden, Martin C. J., et al. "MLST Revisited: The Gene-by-Gene Approach to Bacterial Genomics." *Nature Reviews. Microbiology*, vol. 11, no. 10, Oct. 2013, pp. 728–36. *PubMed Central*, <https://doi.org/10.1038/nrmicro3093>.
- Mak, Sarah Siu Tze, et al. "Comparative Performance of the BGISEQ-500 vs Illumina HiSeq2500 Sequencing Platforms for Palaeogenomic Sequencing." *GigaScience*, vol. 6, no. 8, Aug. 2017, pp. 1–13. *PubMed*, <https://doi.org/10.1093/gigascience/gix049>.
- Mareuil, Fabien, et al. "A Public Galaxy Platform at Pasteur Used as an Execution Engine for Web Services." *F1000Research*, vol. 6, June 2017. *f1000research.com*, <https://doi.org/10.7490/f1000research.1114334.1>.
- Maxam, A. M., and W. Gilbert. "A New Method for Sequencing DNA." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 2, Feb. 1977, pp. 560–64. *PubMed*, <https://doi.org/10.1073/pnas.74.2.560>.
- Melendez, Dante P., et al. "Detection of Prosthetic Joint Infection by Use of PCR-Electrospray Ionization Mass Spectrometry Applied to Synovial Fluid." *Journal of Clinical Microbiology*, vol. 52, no. 6, June 2014, pp. 2202–05. *PubMed*, <https://doi.org/10.1128/JCM.00570-14>.
- Mestan, Karen K., et al. "Genomic Sequencing in Clinical Trials." *Journal of Translational Medicine*, vol. 9, Dec. 2011, p. 222. *PubMed*, <https://doi.org/10.1186/1479-5876-9-222>.
- Miller, Steve, et al. "Laboratory Validation of a Clinical Metagenomic Sequencing Assay for Pathogen Detection in Cerebrospinal Fluid." *Genome Research*, vol.

- 29, no. 5, May 2019, pp. 831–42. *PubMed Central*, <https://doi.org/10.1101/gr.238170.118>.
- Moustafa, Ahmed, et al. “The Blood DNA Virome in 8,000 Humans.” *PLoS Pathogens*, vol. 13, no. 3, 2017, p. e1006292. *PubMed*, <https://doi.org/10.1371/journal.ppat.1006292>.
- Murugesan, Kanagavel, et al. “Investigation of Preanalytical Variables Impacting Pathogen Cell-Free DNA in Blood and Urine.” *Journal of Clinical Microbiology*, vol. 57, no. 11, Oct. 2019, pp. e00782-19. *PubMed Central*, <https://doi.org/10.1128/JCM.00782-19>.
- N, Sapojnikova, et al. “A Comparison of DNA Fragmentation Methods - Applications for the Biochip Technology.” *Journal of Biotechnology*, vol. 256, J Biotechnol, 20 Aug. 2017, <https://doi.org/10.1016/j.jbiotec.2017.06.1202>.
- Namdari, Surena, et al. “Comparative Study of Cultures and Next-Generation Sequencing in the Diagnosis of Shoulder Prosthetic Joint Infections.” *Journal of Shoulder and Elbow Surgery*, vol. 28, no. 1, Jan. 2019, pp. 1–8. *PubMed*, <https://doi.org/10.1016/j.jse.2018.08.048>.
- Namiki, Toshiaki, et al. “MetaVelvet: An Extension of Velvet Assembler to de Novo Metagenome Assembly from Short Sequence Reads.” *Nucleic Acids Research*, vol. 40, no. 20, Nov. 2012, p. e155. *PubMed*, <https://doi.org/10.1093/nar/gks678>.
- NEBNext® Microbiome DNA Enrichment Kit | NEB. <https://international.neb.com/products/e2612-nebnext-microbiome-dna-enrichment-kit#Product%20Information>. Accessed 17 Dec. 2020.
- Neoh, Hui-Min, et al. “Pulsed-Field Gel Electrophoresis (PFGE): A Review of the ‘Gold Standard’ for Bacteria Typing and Current Alternatives.” *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, vol. 74, Oct. 2019, p. 103935. *PubMed*, <https://doi.org/10.1016/j.meegid.2019.103935>.
- Nurk, Sergey, et al. “MetaSPAdes: A New Versatile Metagenomic Assembler.” *Genome Research*, vol. 27, no. 5, May 2017, pp. 824–34. *PubMed Central*, <https://doi.org/10.1101/gr.213959.116>.

- Ondov, Brian D., et al. "Interactive Metagenomic Visualization in a Web Browser." *BMC Bioinformatics*, vol. 12, Sept. 2011, p. 385. *PubMed*, <https://doi.org/10.1186/1471-2105-12-385>.
- Parekh, Mohit, et al. "Next-Generation Sequencing for the Detection of Microorganisms Present in Human Donor Corneal Preservation Medium." *BMJ Open Ophthalmology*, vol. 4, no. 1, Apr. 2019. *PubMed Central*, <https://doi.org/10.1136/bmjophth-2018-000246>.
- Parize, P., et al. "Untargeted Next-Generation Sequencing-Based First-Line Diagnosis of Infection in Immunocompromised Adults: A Multicentre, Blinded, Prospective Study." *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, vol. 23, no. 8, Aug. 2017, p. 574.e1-574.e6. *PubMed*, <https://doi.org/10.1016/j.cmi.2017.02.006>.
- Pérez-Losada, Marcos, et al. "Microbial Sequence Typing in the Genomic Era." *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, vol. 63, Sept. 2018, pp. 346–59. *PubMed Central*, <https://doi.org/10.1016/j.meegid.2017.09.022>.
- Peri, Anna Maria, et al. "Culture-Independent Detection Systems for Bloodstream Infection." *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, vol. 28, no. 2, Feb. 2022, pp. 195–201. *PubMed*, <https://doi.org/10.1016/j.cmi.2021.09.039>.
- Peterson, A. Townsend. "Defining Viral Species: Making Taxonomy Useful." *Virology Journal*, vol. 11, July 2014, p. 131. *PubMed Central*, <https://doi.org/10.1186/1743-422X-11-131>.
- Pett, Helmi, et al. "Comparison of Molecular Quantification of Plasmodium Falciparum Gametocytes by Pfs25 QRT-PCR and QT-NASBA in Relation to Mosquito Infectivity." *Malaria Journal*, vol. 15, no. 1, Nov. 2016, p. 539. *PubMed*, <https://doi.org/10.1186/s12936-016-1584-z>.
- Petty, Tom J., et al. "Comprehensive Human Virus Screening Using High-Throughput Sequencing with a User-Friendly Representation of Bioinformatics Analysis: A Pilot Study." *Journal of Clinical Microbiology*, vol. 52, no. 9, Sept. 2014, pp. 3351–61. *PubMed Central*, <https://doi.org/10.1128/JCM.01389-14>.

- Ponty, Yann. *Bio-Algorithmique Des ARN: Petite Promenade Aux Interfaces*. Oct. 2014.
- Prachayangprecha, Slinporn, et al. “Exploring the Potential of Next-Generation Sequencing in Detection of Respiratory Viruses.” *Journal of Clinical Microbiology*, vol. 52, no. 10, Oct. 2014, pp. 3722–30. *PubMed*, <https://doi.org/10.1128/JCM.01641-14>.
- Price, Morgan N., et al. “FastTree: Computing Large Minimum Evolution Trees with Profiles Instead of a Distance Matrix.” *Molecular Biology and Evolution*, vol. 26, no. 7, July 2009, pp. 1641–50. *PubMed*, <https://doi.org/10.1093/molbev/msp077>.
- Rhoads, Anthony, and Kin Fai Au. “PacBio Sequencing and Its Applications.” *Genomics, Proteomics & Bioinformatics*, vol. 13, no. 5, Oct. 2015, pp. 278–89. *PubMed Central*, <https://doi.org/10.1016/j.gpb.2015.08.002>.
- Richter, Michael, and Ramon Rosselló-Móra. “Shifting the Genomic Gold Standard for the Prokaryotic Species Definition.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 45, Nov. 2009, pp. 19126–31. *PubMed Central*, <https://doi.org/10.1073/pnas.0906412106>.
- Rosenbaum, Jennifer, et al. “Evaluation of Oral Cavity DNA Extraction Methods on Bacterial and Fungal Microbiota.” *Scientific Reports*, vol. 9, Feb. 2019. *PubMed Central*, <https://doi.org/10.1038/s41598-018-38049-6>.
- Rosselló-Mora, R., and R. Amann. “The Species Concept for Prokaryotes.” *FEMS Microbiology Reviews*, vol. 25, no. 1, Jan. 2001, pp. 39–67. *PubMed*, <https://doi.org/10.1111/j.1574-6976.2001.tb00571.x>.
- Rothberg, Jonathan M., et al. “An Integrated Semiconductor Device Enabling Non-Optical Genome Sequencing.” *Nature*, vol. 475, no. 7356, July 2011, pp. 348–52. *PubMed*, <https://doi.org/10.1038/nature10242>.
- Ruppé, Etienne, et al. “Clinical Metagenomics of Bone and Joint Infections: A Proof of Concept Study.” *Scientific Reports*, vol. 7, no. 1, Aug. 2017, p. 7718. *PubMed*, <https://doi.org/10.1038/s41598-017-07546-5>.
- Salter, Susannah J., et al. “Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses.” *BMC Biology*, vol. 12, Nov. 2014, p. 87. *PubMed*, <https://doi.org/10.1186/s12915-014-0087-z>.

- Sanger, F., et al. “DNA Sequencing with Chain-Terminating Inhibitors.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 12, Dec. 1977, pp. 5463–67.
- Scharf, Sebastian, et al. “Introduction of a Bead Beating Step Improves Fungal DNA Extraction from Selected Patient Specimens.” *International Journal of Medical Microbiology: IJMM*, vol. 310, no. 6, Sept. 2020, p. 151443. *PubMed*, <https://doi.org/10.1016/j.ijmm.2020.151443>.
- Schloss, Patrick D., et al. “Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities.” *Applied and Environmental Microbiology*, vol. 75, no. 23, Dec. 2009, pp. 7537–41. *journals.asm.org* (Atypon), <https://doi.org/10.1128/AEM.01541-09>.
- Schloss, Patrick D., and Jo Handelsman. “Biotechnological Prospects from Metagenomics.” *Current Opinion in Biotechnology*, vol. 14, no. 3, June 2003, pp. 303–10. *PubMed*, [https://doi.org/10.1016/s0958-1669\(03\)00067-3](https://doi.org/10.1016/s0958-1669(03)00067-3).
- Schloss, Patrick D., and Sarah L. Westcott. “Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis.” *Applied and Environmental Microbiology*, vol. 77, no. 10, May 2011, pp. 3219–26. *aem.asm.org*, <https://doi.org/10.1128/AEM.02810-10>.
- Schmidt, Brian M., et al. “Metagenomics to Identify Pathogens in Diabetic Foot Ulcers and the Potential Impact for Clinical Care.” *Current Diabetes Reports*, vol. 21, no. 8, June 2021, p. 26. *PubMed*, <https://doi.org/10.1007/s11892-021-01391-7>.
- Schoch, Conrad L., et al. “Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi.” *Proceedings of the National Academy of Sciences*, vol. 109, no. 16, Apr. 2012, pp. 6241–46. *www.pnas.org*, <https://doi.org/10.1073/pnas.1117018109>.
- Seemann, Torsten. *Mlst*. 2014. 1 May 2022. *GitHub*, <https://github.com/tseemann/mlst>.
- . “Prokka: Rapid Prokaryotic Genome Annotation.” *Bioinformatics (Oxford, England)*, vol. 30, no. 14, July 2014, pp. 2068–69. *PubMed*, <https://doi.org/10.1093/bioinformatics/btu153>.
- . *Snippy*. 2014. 9 Mar. 2022. *GitHub*, <https://github.com/tseemann/snippy>.
- Selleri, M., et al. “Detection of Haemagglutinin D222 Polymorphisms in Influenza A(H1N1)Pdm09-Infected Patients by Ultra-Deep Pyrosequencing.”

- Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, vol. 19, no. 7, July 2013, pp. 668–73. *PubMed*, <https://doi.org/10.1111/j.1469-0691.2012.03984.x>.
- Shanmugasundaram, Saseendar, et al. “Evaluation and Management of Periprosthetic Joint Infection-an International, Multicenter Study.” *HSS Journal: The Musculoskeletal Journal of Hospital for Special Surgery*, vol. 10, no. 1, Feb. 2014, pp. 36–44. *PubMed*, <https://doi.org/10.1007/s11420-013-9366-4>.
- Somasekar, Sneha, et al. “Viral Surveillance in Serum Samples From Patients With Acute Liver Failure By Metagenomic Next-Generation Sequencing.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, vol. 65, no. 9, Nov. 2017, pp. 1477–85. *PubMed Central*, <https://doi.org/10.1093/cid/cix596>.
- Stamatakis, Alexandros. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” *Bioinformatics*, vol. 30, no. 9, May 2014, pp. 1312–13. *Silverchair*, <https://doi.org/10.1093/bioinformatics/btu033>.
- Stebner, A., et al. “Molecular Diagnosis of Polymicrobial Brain Abscesses with 16S-RDNA-Based next-Generation Sequencing.” *Clinical Microbiology and Infection*, vol. 27, no. 1, Jan. 2021, pp. 76–82. *ScienceDirect*, <https://doi.org/10.1016/j.cmi.2020.03.028>.
- Stevenson, Matt, et al. “Sepsis: The LightCycler SeptiFast Test MGRADE®, SepsiTTM and IRIDICA BAC BSI Assay for Rapidly Identifying Bloodstream Bacteria and Fungi - a Systematic Review and Economic Evaluation.” *Health Technology Assessment (Winchester, England)*, vol. 20, no. 46, June 2016, pp. 1–246. *PubMed*, <https://doi.org/10.3310/hta20460>.
- Strong, Michael J., et al. “Microbial Contamination in next Generation Sequencing: Implications for Sequence-Based Analysis of Clinical Samples.” *PLoS Pathogens*, vol. 10, no. 11, Nov. 2014, p. e1004437. *PubMed*, <https://doi.org/10.1371/journal.ppat.1004437>.
- Tabibzadeh, Alireza, et al. “Evolutionary Study of COVID-19, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) as an Emerging Coronavirus: Phylogenetic Analysis and Literature Review.” *Veterinary Medicine and Science*, vol. 7, no. 2, Nov. 2020, pp. 559–71. *PubMed Central*, <https://doi.org/10.1002/vms3.394>.

- Tamura, Koichiro, et al. "MEGA11: Molecular Evolutionary Genetics Analysis Version 11." *Molecular Biology and Evolution*, vol. 38, no. 7, July 2021, pp. 3022–27. *Silverchair*, <https://doi.org/10.1093/molbev/msab120>.
- Tarabichi, Majd, et al. "Diagnosis of Periprosthetic Joint Infection: The Potential of Next-Generation Sequencing." *The Journal of Bone and Joint Surgery. American Volume*, vol. 100, no. 2, Jan. 2018, pp. 147–54. *PubMed*, <https://doi.org/10.2106/JBJS.17.00434>.
- "The Enzymes Making the Cut in NGS Library Preparation." *Cytiva*, <https://www.cytivalifesciences.com/en/us/news-center/enzymes-in-ngs-library-prep-10001>. Accessed 5 Oct. 2022.
- The UniProt Consortium. "UniProt: The Universal Protein Knowledgebase in 2021." *Nucleic Acids Research*, vol. 49, no. D1, Jan. 2021, pp. D480–89. *Silverchair*, <https://doi.org/10.1093/nar/gkaa1100>.
- Thoendel, Matthew, et al. "Comparison of Microbial DNA Enrichment Tools for Metagenomic Whole Genome Sequencing." *Journal of Microbiological Methods*, vol. 127, Aug. 2016, pp. 141–45. *PubMed Central*, <https://doi.org/10.1016/j.mimet.2016.05.022>.
- Thorburn, Fiona, et al. "The Use of next Generation Sequencing in the Diagnosis and Typing of Respiratory Infections." *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, vol. 69, Aug. 2015, pp. 96–100. *PubMed*, <https://doi.org/10.1016/j.jcv.2015.06.082>.
- Thorstenson, Yvonne R., et al. "An Automated Hydrodynamic Process for Controlled, Unbiased DNA Shearing." *Genome Research*, vol. 8, no. 8, Aug. 1998, pp. 848–55.
- Tindall, B. J., et al. "Enterobacter Aerogenes Hormaeche and Edwards 1960 (Approved Lists 1980) and Klebsiella Mobilis Bascomb et al. 1971 (Approved Lists 1980) Share the Same Nomenclatural Type (ATCC 13048) on the Approved Lists and Are Homotypic Synonyms, with Consequences for the Name Klebsiella Mobilis Bascomb et al. 1971 (Approved Lists 1980)." *International Journal of Systematic and Evolutionary Microbiology*, vol. 67, no. 2, Feb. 2017, pp. 502–04. *PubMed*, <https://doi.org/10.1099/ijsem.0.001572>.
- Toma, Ian, et al. "Single-Molecule Long-Read 16S Sequencing to Characterize the Lung Microbiome from Mechanically Ventilated Patients with Suspected

- Pneumonia.” *Journal of Clinical Microbiology*, vol. 52, no. 11, Nov. 2014, pp. 3913–21. *PubMed*, <https://doi.org/10.1128/JCM.01678-14>.
- Wagner Mackenzie, Brett, et al. “Evaluating Variation in Human Gut Microbiota Profiles Due to DNA Extraction Method and Inter-Subject Differences.” *Frontiers in Microbiology*, vol. 6, 2015, p. 130. *PubMed*, <https://doi.org/10.3389/fmicb.2015.00130>.
- Wang, Qingqing, et al. “The Clinical Value of Metagenomic Next-Generation Sequencing in the Microbiological Diagnosis of Skin and Soft Tissue Infections.” *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, vol. 100, Nov. 2020, pp. 414–20. *PubMed*, <https://doi.org/10.1016/j.ijid.2020.09.007>.
- Wangler, Michael F., and Hugo J. Bellen. “Chapter 12 - In Vivo Animal Modeling: *Drosophila*.” *Basic Science Methods for Clinical Researchers*, edited by Morteza Jalali et al., Academic Press, 2017, pp. 211–34. *ScienceDirect*, <https://doi.org/10.1016/B978-0-12-803077-6.00012-6>.
- Willey, Joanne M., et al. *Microbiologie de Prescott*. 2018.
- Wilson, Michael R., Hannah A. Sample, et al. “Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis.” *New England Journal of Medicine*, vol. 380, no. 24, June 2019, pp. 2327–40. *Taylor and Francis+NEJM*, <https://doi.org/10.1056/NEJMoa1803396>.
- Wilson, Michael R., Greg Fedewa, et al. “Multiplexed Metagenomic Deep Sequencing To Analyze the Composition of High-Priority Pathogen Reagents.” *MSystems*, vol. 1, no. 4, Aug. 2016. *PubMed*, <https://doi.org/10.1128/mSystems.00058-16>.
- Wood, Derrick E., et al. “Improved Metagenomic Analysis with Kraken 2.” *Genome Biology*, vol. 20, no. 1, Nov. 2019, p. 257. *BioMed Central*, <https://doi.org/10.1186/s13059-019-1891-0>.
- Wu, Wenjing, et al. “Precise Species Identification for Enterobacter: A Genome Sequence-Based Study with Reporting of Two Novel Species, *Enterobacter Quasirogerkampii* Sp. Nov. and *Enterobacter Quasimori* Sp. Nov.” *MSystems*, vol. 5, no. 4, Aug. 2020, pp. e00527–20. *PubMed*, <https://doi.org/10.1128/mSystems.00527-20>.

- Xavier, Basil Britto, et al. "Consolidating and Exploring Antibiotic Resistance Gene Data Resources." *Journal of Clinical Microbiology*, vol. 54, no. 4, Apr. 2016, pp. 851–59. *PubMed*, <https://doi.org/10.1128/JCM.02717-15>.
- Yang, Jianhua, et al. "Brain Abscess Caused by *Nocardia Farcinica* and Diagnosed by Metagenomic Next-Generation Sequencing: A Case Report." *Frontiers in Medicine*, vol. 9, 2022, p. 803554. *PubMed*, <https://doi.org/10.3389/fmed.2022.803554>.
- Yang, Yu, and Jun Hang. "Fragmentation of Genomic DNA Using Microwave Irradiation." *Journal of Biomolecular Techniques : JBT*, vol. 24, no. 2, July 2013, p. 98. www.ncbi.nlm.nih.gov, <https://doi.org/10.7171/jbt.13-2402-005>.
- Zankari, Ea, et al. "PointFinder: A Novel Web Tool for WGS-Based Detection of Antimicrobial Resistance Associated with Chromosomal Point Mutations in Bacterial Pathogens." *The Journal of Antimicrobial Chemotherapy*, vol. 72, no. 10, Oct. 2017, pp. 2764–68. *PubMed*, <https://doi.org/10.1093/jac/dkx217>.
- Zerbino, Daniel R., and Ewan Birney. "Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs." *Genome Research*, vol. 18, no. 5, May 2008, pp. 821–29. *PubMed*, <https://doi.org/10.1101/gr.074492.107>.
- Zheng, HaoQiang, et al. "Detection of Murine Leukemia Virus or Mouse DNA in Commercial RT-PCR Reagents and Human DNAs." *PloS One*, vol. 6, no. 12, 2011, p. e29050. *PubMed*, <https://doi.org/10.1371/journal.pone.0029050>.
- Zhou, Chaojun, et al. "Idiopathic Thrombocytopenic Purpura with Brain Abscess Caused by *Nocardia Farcinica* Diagnosed Using Metagenomics Next-Generation Sequencing of the Cerebrospinal Fluid: A Case Report." *BMC Infectious Diseases*, vol. 21, no. 1, Apr. 2021, p. 380. *PubMed*, <https://doi.org/10.1186/s12879-021-06071-1>.
- Zhu, Hanliang, et al. "PCR Past, Present and Future." *BioTechniques*, vol. 69, no. 4, Oct. 2020, pp. 317–25. *PubMed*, <https://doi.org/10.2144/btn-2020-0057>.

DELETTRE Nicolas

Apport des technologies de séquençage haut débit en microbiologie clinique : revue de la littérature et projet d'implantation au CHU de Rouen

Th. D. Pharm., Rouen, 2022.

RESUME

Le diagnostic microbiologie est le prérequis indispensable à un traitement anti-infectieux adapté. Les techniques de diagnostic microbiologiques sont encore grandement dépendantes de méthodes, certes robustes, mais anciennes et qui atteignent leurs limites. Les nouvelles méthodes de séquençages, dites « haut débit » ou NGS, permettent d'obtenir de nouvelles données sur les prélèvements cliniques. Cependant, leur utilisation est coûteuse et complexe et nécessite un personnel qualifié pour les mettre en œuvre.

Ce travail bibliographique a pour objectif d'inventorier ces nouvelles techniques de séquençages, de guider la mise en place de celles-ci dans un laboratoire de microbiologie clinique et d'aider à leur usage tant dans leurs indications que dans leur interprétation. Celui-ci servira de base au développement de l'usage du séquençage haut débit au laboratoire de microbiologie de l'Hôpital Charles Nicolle - CHU de Rouen.

MOTS CLES : Diagnostic microbiologique, Séquençage haut débit, NGS

JURY

Président : Pr. Pestel-Caron Martine, PUPH

Membres : Dr. Lemée Ludovic, PH

Dr. Tazi Asmaa, MCU-PH

Pr. Plantier Jean-Christophe

DATE DE SOUTENANCE : 19 octobre 2022