



HAL
open science

Microbiote intestinal et obésité : Identification d'une signature de la dysbiose chez le sujet obèse morbide par la technologie Oxford Nanopore.

Rohia Alili

► To cite this version:

Rohia Alili. Microbiote intestinal et obésité : Identification d'une signature de la dysbiose chez le sujet obèse morbide par la technologie Oxford Nanopore.. Alimentation et Nutrition. 2018. hal-02050935

HAL Id: hal-02050935

<https://ephe.hal.science/hal-02050935>

Submitted on 27 Feb 2019

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

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

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE

ECOLE PRATIQUE DES HAUTES ETUDES

Sciences de la Vie et de la Terre

MEMOIRE

Présenté par

Rohia ALILI

Pour l'obtention du diplôme de l'Ecole Pratique des Hautes Etudes

Microbiote intestinal et obésité : Identification d'une signature de la dysbiose chez le sujet obèse morbide par la technologie Oxford Nanopore

Soutenu le : 12 décembre 2018 devant le jury suivant

Bruno CANQUE – Président

Frédéric DEVAUX – Rapporteur

David VALLENET - Examineur

Pr. Karine CLEMENT – Tuteur scientifique

Pr. Thierry WIRTH – Tuteur pédagogique

Table des matières

Introduction	1
1) Obésité et maladies métaboliques : des maladies complexes et fréquentes	2
2) Le microbiote intestinal : un organe à part entière ?	14
3) L'exploration du microbiote : méthodes de Séquençage	23
<i>A - Quelques éléments historiques</i>	23
<i>B - Méthode de Maxam et Gilbert</i>	25
<i>C - Méthode de Sanger</i>	26
<i>D - Séquençage de génome entier</i>	29
<i>E - Autres méthodes</i>	32
<i>E - 1. Séquençage par hybridation</i>	32
<i>E - 2. Séquençage haut débit (HTS)</i>	32
<i>E - 3. Séquençage par nanopore</i>	35
<i>F - Technologie Oxford Nanopore</i>	42
Hypothèse de travail	48
Objectif de notre étude	48
Démarche expérimentale	50
Matériels et méthodes	51
1 - Kit de prélèvement et extraction de l'ADN bactérien	53
<i>A) Prélèvements et conditionnement des selles</i>	53
<i>B) Extraction de l'ADN bactérien</i>	54
2 - Préparation de la librairie de séquençage	55
3 - Analyse bioinformatique	57
Résultats	61
1) Mise au point de l'extraction de l'ADN et de la préparation de la librairie de séquençage	62
<i>A) Comparaison des kits d'extraction d'ADN choisis et test de l'étape de fragmentation de l'ADN du protocole de préparation de la librairie</i>	62
<i>B) Test de l'étape de « End repair » du protocole de préparation de la librairie</i>	69
2) Améliorations de l'extraction de l'ADN avec le Kit Invitrogen et du protocole de préparation de la librairie	71
3) Mode de recueil des selles	78
<i>A) Comparaison de différents modes de stabilisation de l'ADN microbien et de deux modes d'acheminement des prélèvements de selles jusqu'au laboratoire pour l'analyse</i>	78
<i>B) Comparaison mode de stockage des échantillons prélevés sur tube avec stabilisateur</i>	83
Validation du séquençage Oxford Nanopore	85
Discussion et perspectives	90
Discussion	91
Perspectives	98

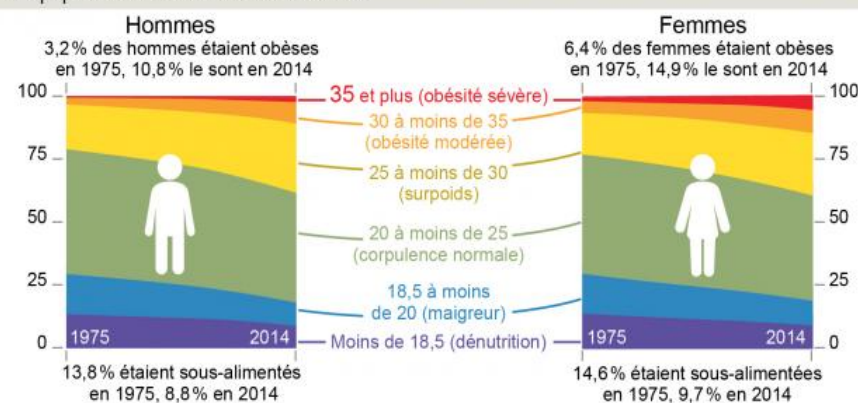
Introduction

1) Obésité et maladies métaboliques : des maladies complexes et fréquentes

L'obésité se définit par une accumulation anormale ou excessive de graisse corporelle qui peut nuire à la santé¹. Elle est, aujourd'hui, considérée comme une pandémie. Depuis 1997, l'obésité est reconnue par l'Organisation mondiale de la santé (OMS) comme étant une maladie en raison de ses répercussions sanitaires et économiques et de son incidence mondiale. Elle concerne aujourd'hui la quasi-totalité de la planète, y compris de nombreux pays émergents : selon l'OMS, 35 % des adultes dans le monde sont atteints d'obésité ou de surpoids. Les complications associées, en particulier le diabète et les maladies cardiovasculaires, entraînent le décès d'au moins 2,8 millions de personnes chaque année. En France, l'obésité concerne 17 % des adultes², contre seulement 6,1 % en 1980 (OBEPI-Roche). Lors du Congrès européen sur l'obésité à Prague en 2015, l'OMS a rendu publics les chiffres prédictifs de l'obésité de l'adulte, en Europe, en 2030. Il est estimé que l'obésité concernera 28 % des femmes et 24 % des hommes en France. Ces chiffres sont la conséquence de l'épidémie d'obésité infantile observée entre 1980 et 2000. L'obésité est devenue la première maladie non infectieuse de l'Histoire. C'est une véritable épidémie qui frappe aussi bien les pays industrialisés que les pays en voie de développement. L'Organisation Mondiale de la Santé place actuellement sa prévention et sa prise en charge comme une priorité dans le domaine de la pathologie nutritionnelle.

ALIMENTATION | PLUS D'UN ADULTE SUR DIX EST OBÈSE DANS LE MONDE

La population mondiale selon son IMC*



Les 10 pays où il y a le plus d'obèses En millions en 2014



*Indice de masse corporelle (poids en kilos divisé par la taille en mètres au carré) Source : The Lancet (avril 2016)

VISACTH

Figure 1 : L'obésité touche aujourd'hui près de 650 millions d'adultes, soit 13% de la population mondiale adulte, un pourcentage qui pourrait atteindre 20% d'ici 2025. (Visactu)

En pratique clinique l'obésité est définie par un Indice de Masse Corporelle (IMC) ou Body Mass Index (BMI). L'IMC est calculé en divisant le poids de la personne par le carré de sa taille (Kg / m²). L'IMC permet de définir plusieurs catégories de poids :

Tableau 1 : Valeurs de référence de l'IMC (OMS 2004)

	Dénutrition	Normalité	Surpoids - Obésité
Enfants	IMC < 3 ^{ème} percentile sur la courbe.	Varie en fonction de l'âge et du sexe. Associé aux courbes de corpulence (carnet de santé).	IMC > 97 ^{ème} percentile sur la courbe. Distinction entre obésité de grade 1 ou 2 selon l'intensité.
Adultes de 18 à 70 ans	IMC < 18,5 : dénutrition. IMC < 16,5 : dénutrition sévère.	18,5 ≥ IMC ≥ 24,9	25 ≤ IMC ≤ 29,9 : surpoids. 30 ≤ IMC ≤ 34,9 : obésité modérée. 35 ≤ IMC ≤ 39,9 : obésité sévère. IMC ≥ 40 : obésité morbide.
Adultes de plus de 70 ans	IMC < 21 : dénutrition.	21 ≥ IMC ≥ 30	IMC ≥ 30 : obésité.

Néanmoins, cet indice est imparfait car il ne rend pas toujours compte de la quantité de masse grasse (exemples du sportif, de la femme enceinte, du sujet âgé), ni de sa répartition, ni de la qualité du tissu adipeux (altération inflammatoire ou métabolique par exemple).

Il existe également d'autres indicateurs de surpoids : le rapport tour de taille/tour de hanches. Il doit être inférieur à 1 chez l'homme et à 0,85 chez la femme. Le diamètre du cou (ou Tour de cou) peut donner un indice fiable d'IMC anormalement élevé chez l'enfant de plus de 8 ans. Il est également associé à l'apnée obstructive du sommeil, au diabète et à l'hypertension chez les adultes.

Les deux grands facteurs – « the Big Two »³ – traditionnellement impliqués dans le développement de l'obésité sont la consommation alimentaire et la dépense énergétique :

- L'apport énergétique quotidien est la somme des calories apportées par les principaux macronutriments présents dans l'alimentation : les glucides, les lipides, les protéines. La prise alimentaire est un comportement régulé finement par des mécanismes biologiques complexes et redondants permettant un apport adapté en nutriments nécessaire à la vie. Des modifications de la prise alimentaire, en termes de structure ou de composition alimentaire, peuvent aboutir à une prise de poids. Parmi les modifications de la prise alimentaire, on retrouve des modifications quantitatives, telles que l'augmentation de la densité calorique de l'alimentation et des modifications qualitatives, comme la diminution de la consommation de glucides complexes (féculents, fibres), l'augmentation de l'apport lipidique et la déstructuration des rythmes alimentaires.

- La dépense énergétique totale est constituée de trois composantes, variables d'un individu à l'autre, contribuant de façon déterminante au bilan énergétique. Ces trois composants sont :

* La dépense énergétique de repos, variable selon l'âge, le sexe et la composition corporelle notamment la masse musculaire : notre organisme consomme 60 % de la dépense énergétique quotidienne pour assurer la circulation du sang via l'activité cardiaque, la respiration, les fonctions gastro-intestinales et rénales ainsi que la thermorégulation de notre corps.

* La thermogénèse post-prandiale : 10 % de la dépense énergétique quotidienne représente le coût de la digestion, du transport et du stockage des nutriments.

* La dépense énergétique liée à l'activité physique (DEAP), qui est très variable d'un individu à l'autre, utilise 30 % de la dépense énergétique quotidienne chez un individu « moyen ».

Dans ce cadre d'analyse, le maintien d'un poids-santé est une question de comptabilité énergétique. Lorsque l'organisme reçoit plus qu'il ne dépense, il stocke une partie de l'apport, sous forme de graisses dans le tissu adipeux. Cependant le métabolisme, très différent selon les individus, joue un rôle important, et certaines personnes vont donc plus facilement devenir obèses que d'autres. La prise en compte des deux grands facteurs, l'apport et la dépense énergétiques, n'explique pas complètement l'explosion de l'obésité contemporaine. Il y a eu bien d'autres périodes d'abondance alimentaire dans l'histoire qui n'ont pas été accompagnées d'obésité. Il faut donc prendre en compte des dérèglements de la régulation qui peuvent être de différentes origines combinées entre elles : hormonales (hyperinsulinisme), psychologiques (boulimie par exemple), métaboliques (perturbateurs endocriniens), alimentaires (aliments nouveaux à l'échelle historique, comme le sucre, les sodas, les jus de fruits, le chocolat, les crèmes glacées, les produits transformés), déstructuration des repas, grignotage, environnementales (sédentarité, chauffage, automobile, marketing, télévision, publicité, stress, manque de sommeil, phénomènes inhérents à la vie moderne).

Il existe un autre déterminant de l'obésité qui est d'ordre génétique ou héréditaire. La contribution de l'hérédité à l'obésité peut être schématisée à ce jour de deux façons⁴ :

- soit par des mutations uniques dans certains gènes expliquant à elles seules le développement de l'obésité (obésité monogénique ou syndromiques). Ces obésités sont rares, très sévères et débutent généralement dans l'enfance ;
- soit par l'interaction de plusieurs variants géniques avec un environnement à risque (obésité polygénique). Dans ce cas, chaque gène de susceptibilité pris individuellement aurait de faibles effets sur le poids, et la contribution cumulative de ces variants ne deviendrait significative qu'en interaction avec des facteurs environnementaux prédisposant à leur expression phénotypique (suralimentation, baisse de l'activité physique). C'est le cas des obésités dites communes.

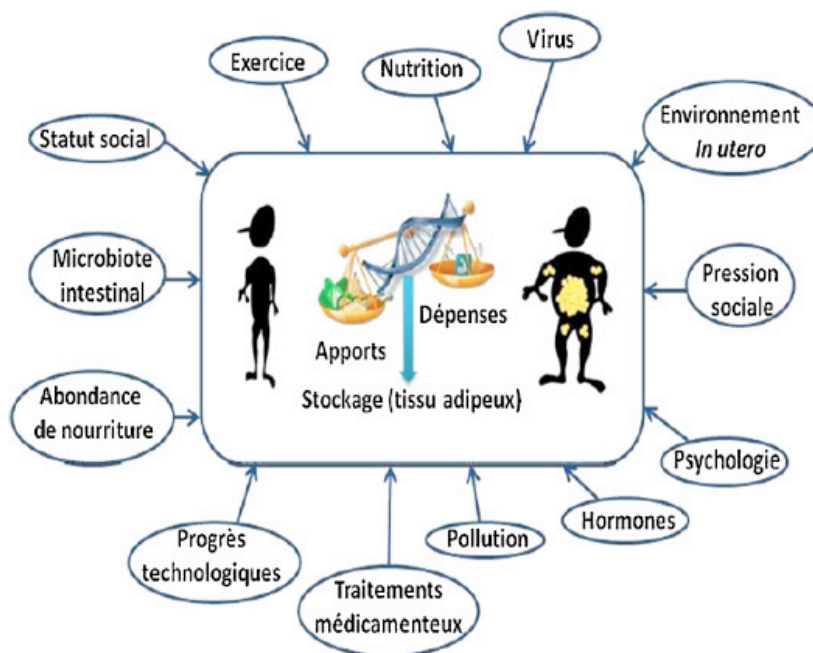


Figure 2 : Facteurs impliqués dans le développement de l'obésité⁵

Les risques liés à l'obésité dépendent de l'importance du tissu adipeux et de sa répartition particulièrement abdominale et péri viscérale. Un excès de masse grasse à ces niveaux définit l'obésité « androïde » avec un tour de taille supérieur ou égale à 80 centimètres chez la femme ou à 90 centimètres chez l'homme, et qui a pour conséquences une augmentation de la prévalence des maladies métaboliques et cardiovasculaires alors que la répartition du tissu adipeux au niveau des hanches et des cuisses définit l'obésité « gynoïde » qui expose moins à ces risques.

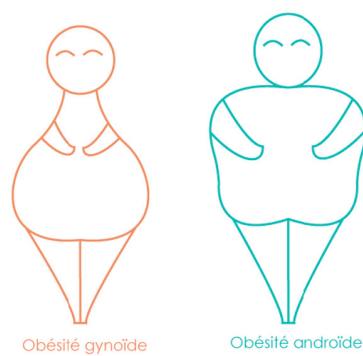


Figure 3 : Différentes localisation de la masse adipeuse⁶.

Les connaissances sur la physiologie et la physiopathologie du tissu adipeux se sont considérablement accrues ces dernières années. De par ses fonctions métaboliques et son rôle endocrine avéré, le tissu adipeux est apparu comme un organe central de l'homéostasie énergétique et métabolique. Les adipocytes s'hypertrophient (augmentent de volume) au fur et à mesure qu'ils accumulent des lipides. Lorsqu'ils ont atteint leur volume maximal, ils ont la capacité de recruter de nouvelles cellules, les pré-adipocytes, qui se différencient en adipocytes matures capables de se charger de triglycérides lors de l'adipogénèse (Fig. 2). Ainsi, la masse du tissu adipeux peut s'accroître non seulement par l'augmentation du volume des adipocytes (hypertrophie), mais aussi par l'augmentation du nombre d'adipocytes qui le compose (hyperplasie). Une altération de l'adipogénèse, sous l'influence par exemple de certains nutriments, d'agents infectieux ou de polluants, de facteurs nerveux ou hormonaux, peut contribuer à l'expansion de la masse grasse.

Lorsque les capacités de stockage du tissu adipeux sous-cutané sont dépassées, les acides gras sont stockés dans des dépôts ectopiques du tissu adipeux au niveau viscéral (graisse omentale), mais également au niveau d'organes multiples tels que le muscle, le cœur (épicaarde), le pancréas, les vaisseaux, et le foie (stéatose hépatique). Ces dépôts ectopiques de tissu adipeux sont responsables de comorbidités qui font de l'obésité une maladie chronique. Bien qu'il existe des variations entre individus, les problèmes de santé commenceraient à se manifester après environ dix ans d'excès de poids, en particulier chez les individus atteints d'obésité sévère ou morbide. Ainsi, de nombreuses études ont mis en évidence une étroite corrélation entre l'IMC et la mortalité associée à de multiples pathologies⁷.

L'obésité sévère ou morbide est un facteur important de risque de maladies métaboliques et chroniques telles que les maladies hépatiques, la résistance à l'insuline, le diabète de type 2 (DT2), l'hypertension artérielle, l'athérosclérose, les maladies cardiovasculaires, et même plusieurs types de cancers⁸. Ces pathologies sont couramment appelées « comorbidités » associées à l'obésité.

Les complications respiratoires sont dominées par le syndrome d'apnées du sommeil, retrouvé dans 25 à 40 % des cas d'obésité massive et dans 10 % des cas d'obésité commune. Il peut être responsable de problèmes respiratoires et cardiaques, voire de mort subite⁹.

L'obésité est également à l'origine de différentes perturbations endocriniennes : hyperœstrogénie, hypofertilité et troubles de l'ovulation chez la femme, hypogonadisme chez l'homme.

On distingue également le « syndrome métabolique » ou (MetS). C'est un regroupement de facteurs qui reflètent la surnutrition, le mode de vie sédentaire et l'excès d'adiposité. Il en existe plusieurs, qui ont varié dans le temps et selon les organismes de santé ou les pays mais il on s'accorde à parler de « syndrome métabolique » dans les cas où au moins trois des problèmes suivants sont associés chez la même personne¹⁰ :

- hypercholestérolémie avec un faible taux de cholestérol HDL (le « bon » cholestérol) inférieur à 1,04 mmol/l pour les hommes, 1,29 mmol/l (46 mg/dl) pour les femmes.
- hypertension (>140 mmHg de pression systolique (maxima) et/ou >90 mmHg de pression diastolique (minima)).
- excès de poids surtout s'il s'agit d'une obésité abdominale (tour de taille supérieur à 90 cm pour les hommes, et à 80 cm pour les femmes).
- hyperglycémie, c'est-à-dire un taux excessif de sucres dans le sang (triglycérides sanguins supérieurs à 1,6 ou 1,7 mmol/l (ou 150 mg/dl) et/ou glycémie à jeun supérieure ou égale à 6,1 mmol/l (ou 110 mg/dl).

Le MetS est associé à d'autres comorbidités, y compris l'état prothrombotique, l'état proinflammatoire, la stéatose hépatique non alcoolique et les troubles de la reproduction¹¹. La prévalence des MetS augmente à des proportions épidémiques non seulement aux États-Unis et dans le reste du monde urbanisé, mais aussi dans les pays en développement. Bien qu'il ne soit pas clair s'il existe un mécanisme physiopathologique unificateur entraînant le MetS, l'adiposité abdominale et la résistance à l'insuline semblent être au centre de ce syndrome et de ses composantes.

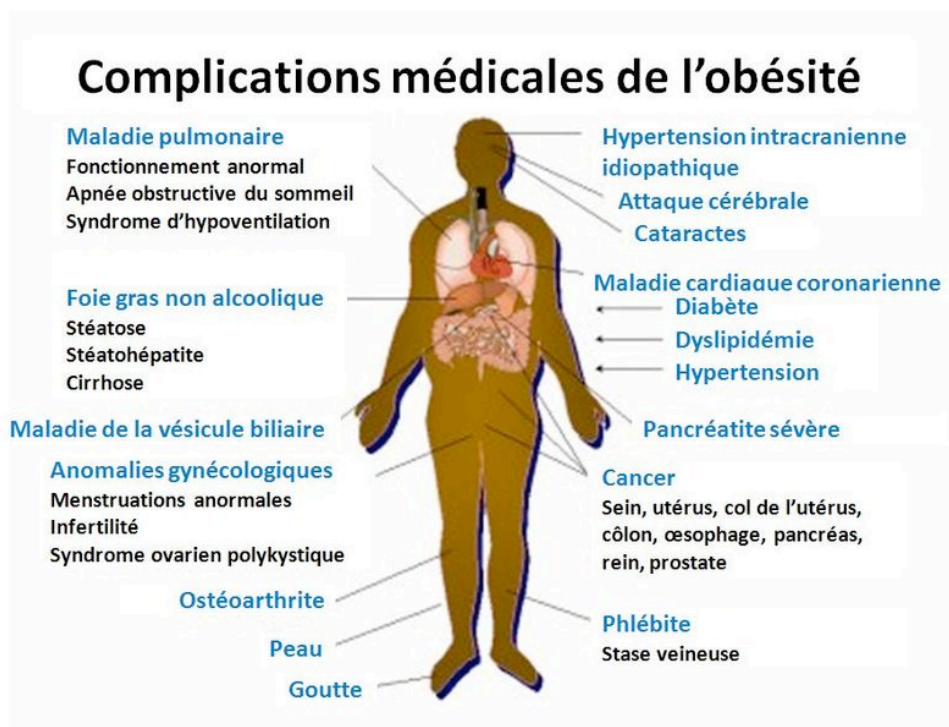


Figure 4 : Les complications de l'obésité sont fréquentes, les plus graves étant cardio-vasculaires et respiratoires.¹²

La prise en charge des patients obèses est nécessairement multidisciplinaire, faisant intervenir des compétences médicales et paramédicales, permettant le diagnostic, le traitement et la prévention de l'obésité et des pathologies associées. La chirurgie de l'obésité ne doit être envisagée que lorsqu'il y a un échec patent d'un traitement médical associant une restriction calorique, une augmentation de l'activité physique, une thérapie comportementale et un traitement pharmacologique. La prise en charge de ces patients est fonction de l'IMC. Quand l'obésité est de classe 1 (surpoids récent, $30 \leq \text{IMC} \leq 34,9$ en l'absence de comorbidité) avec absence de prise en charge médicale bien conduite d'au moins une année, le patient est alors pris en charge médicalement avec recherche de facteurs de comorbidité, prise en charge endocrinologique, diététique, comportementale, pharmacologique et suivi psychologique individuel. Lorsque l'obésité est sévère, alors l'indication chirurgicale pourrait se justifier.¹³

Il existe de nombreuses techniques de chirurgie bariatrique et d'aussi nombreuses variantes. Elles sont basées sur deux principes : techniques de malabsorption qui réduisent la longueur fonctionnelle ou la surface d'absorption de l'intestin grêle et interventions visant à réduire la capacité gastrique. On distingue :

- La gastroplastie verticale calibrée. C'est la technique de référence. Développée initialement par Mason¹⁴, elle consiste à créer une petite poche gastrique de 20 à 30 millilitres par une incision de la paroi de l'estomac. L'estomac est ensuite agrafé de façon à délimiter une petite poche verticale. (Fig.

5). Un anneau est ensuite placé au niveau du tube ainsi généré. Cet anneau va empêcher que le montage réalisé ne se relâche. Il va également contrôler le diamètre de l'orifice d'entrée vers l'estomac. Les aliments vont donc mettre plus de temps à passer dans l'estomac et donner au patient une sensation de satiété. Après ce type d'intervention, les patients doivent modifier leurs habitudes alimentaires. Les risques possibles sont : fissure de l'anneau et rupture de la ligne d'agrafes. Cette technique a été abandonnée

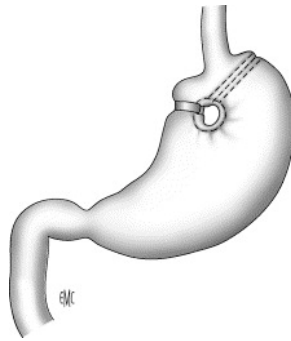


Figure 5 : Gastroplastie verticale calibrée selon Masson¹⁵.

- Les anneaux de gastroplastie ajustables. La technique consiste à placer un anneau de diamètre modifiable autour de la partie supérieure de l'estomac, délimitant ainsi une petite poche. Les aliments vont s'écouler très lentement selon le principe du sablier (Fig. 6). L'anneau est relié par un petit tube à un boîtier de contrôle placé sous la peau. Il s'agit d'une technique assez ancienne mais dont la diffusion est plus récente. Cette technique chirurgicale restrictive diminue le volume de l'estomac en le réduisant à une petite poche de 15 à 20 millilitres et ralentit le passage des aliments sans perturber la digestion des aliments. L'anneau est à la fois ajustable et réversible. Ces anneaux sont le plus souvent implantés par voie coelioscopique sans aucune ouverture du tube digestif, améliorant ainsi le confort péri et postopératoire.

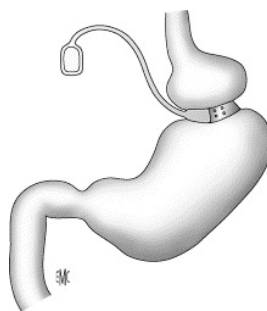


Figure 6 : Anneau de gastroplastie modulable¹³.

- Le court-circuit gastrique ou by-pass gastrique. C'est une intervention associant une réduction gastrique et un défaut d'absorption intestinale. L'intervention consiste à réduire, par coelioscopie, le volume de l'estomac et à modifier le circuit alimentaire. Cette technique, dite dérivation Roux-en-Y du nom de son inventeur César Roux, consiste à faire une transection verticale de l'estomac créant une poche proximale de capacité restreinte dans laquelle arrivent les aliments, reliée directement au jéjunum (par une anse en Y) (Fig. 7). Les aliments ne passent plus par l'estomac et la partie supérieure du tube digestif, ils vont directement dans la partie moyenne de l'intestin grêle. L'amaigrissement est probablement en grande partie en rapport avec une réduction de la prise alimentaire mais également à un degré de malabsorption. C'est l'intervention la plus efficace, donnant une perte de l'excès de poids respectivement de 55 à 70 % en un et cinq ans,¹⁶ mais elle reste associée à un risque de carences en fer, folates et vitamine B12, de complications anastomotiques imposant des ré-interventions et des complications digestives (ulcérations, nausées ou vomissements).

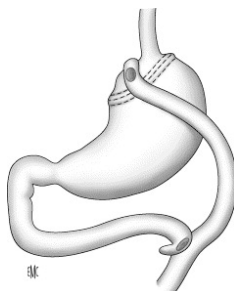


Figure 7 : Dérivation gastrique¹³.

- La sleeve gastrectomie. C'est une gastroplastie verticale calibrée avec résection gastrique. Autrement dit, il s'agit de retirer définitivement une partie de la poche que forme l'estomac pour former un tube (d'où le terme anglophone « sleeve » qui en français signifie « manche ») (Fig. 8). Les aliments ingérés s'engagent dans ce tube. Ils sont ralentis puis évacués très rapidement vers l'intestin grêle. La poche de l'estomac est réduite de deux tiers. La sleeve gastrectomie agit sur l'obésité grâce à deux mécanismes qui, en se combinant, donnent de probants résultats en termes de perte de poids. Comme la gastroplastie, le premier processus est une restriction et un niveau de satiété facilement atteint. Le second entraîne une diminution du taux de ghréline qui est une hormone de la faim.

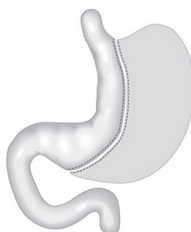


Figure 8 : La sleeve gastrectomie¹⁷

Toutefois, pour comprendre entièrement l'obésité, c'est certainement l'environnement d'un individu dans son ensemble qu'il faudrait prendre en compte.

Des travaux récents ont permis de montrer que la qualité de la flore intestinale ou microbiote intestinal peut être associée à une susceptibilité plus ou moins importante à développer des maladies métaboliques liées à l'obésité. Les études menées chez les souris axéniques (dépourvues de microbiote intestinal) ont clairement démontré le rôle important du microbiote intestinal dans l'homéostasie énergétique de l'hôte¹⁸. De plus, les souris axéniques (comparativement aux souris pourvues en microbiote intestinal) sont protégées de l'obésité induite par un régime à haute teneur en graisses, mais également des désordres métaboliques associés comme l'intolérance au glucose et le diabète de type 2¹⁹. Le transfert du microbiote de souris obèses induit une obésité chez les souris minces (alors que la prise alimentaire de ces dernières n'est pas modifiée)²⁰.

Chez l'homme, des données récentes suggèrent également un lien entre microbiote, homéostasie énergétique (métabolisme glucido-lipidique) et homéostasie inflammatoire. Il a été démontré que le diabète de type 2 est associé à la présence d'espèces bactériennes différentes (en comparaison avec des sujets non diabétiques)²¹.

On savait déjà que les proportions des différentes grandes lignées bactériennes n'étaient pas les mêmes chez les sujets minces et chez les sujets obèses. On a constaté par la suite et dans plusieurs études, que pour une partie des sujets obèses, la diversité bactérienne était réduite, une diminution qui a été associée à des dysfonctionnements métaboliques plus marqués comparativement à des sujets obèses dont la diversité bactérienne est préservée. Parmi ces dysfonctionnements : une insulino-résistance et un taux de triglycérides sanguins plus élevés. Notre équipe a retrouvé cette perte de diversité du microbiote chez 40% des patients en surpoids ou obèse et chez 75% des sujets massivement obèses²².

Dans une étude menée en 2013 par notre équipe, 49 patients obèses et en surpoids ont été soumis à un régime réduit en calories pendant six semaines (pauvre en graisses mais enrichi en fibres, protéines et glucides à faible index glycémique), suivi de six semaines de stabilisation²³. Sur l'ensemble des sujets, une perte de poids moyenne de 5% du poids initial a été constatée, ainsi qu'une amélioration des paramètres métaboliques et inflammatoires comme attendu lors de ce type d'intervention. Les sujets qui présentaient initialement une faible diversité génétique du microbiote ont vu cette diversité augmenter. Néanmoins, l'intervention a été moins efficace sur le plan clinique chez ces sujets avec

une perte de diversité (moindre diminution du taux de triglycérides, moindre amélioration de l'inflammation bas-grade). L'ensemble des signes cliniques liés à l'obésité pourrait être corrigé, ou encore mieux prévenu, par la détection précoce de l'altération du microbiote et grâce à des recommandations nutritionnelles adaptées.

L'idée de moduler la composition et/ou l'activité du microbiote intestinal en vue d'influencer positivement la santé de l'hôte, apparaît aujourd'hui comme une nouvelle voie de développement thérapeutique. Une expérience de « transfert allogénique » de microbiote d'individus sains à des sujets obèses a permis de mettre en évidence une amélioration de la sensibilité hépatique et périphérique à l'insuline qui persiste au moins six semaines après le transfert²⁴. Cette procédure apparaît, à ce jour, davantage comme une preuve de concept que le microbiote lui-même joue un rôle régulateur du métabolisme de l'hôte, que comme une approche thérapeutique potentielle, vu les nombreuses questions qu'elle soulève (innocuité de la procédure ? Assurance de transfert d'un microbiote « sain » ?). En revanche, des constituants de l'alimentation pourraient être intéressants dans la gestion de l'écologie microbienne en relation avec l'adiposité et les pathologies associées. A cet égard, des nutriments de type « fibres alimentaires » qui modulent le microbiote pourraient s'avérer utiles, et permettraient de revoir sous un nouveau jour les conseils nutritionnels à prodiguer dans la prise en charge du syndrome métabolique.

2) Le microbiote intestinal : un organe à part entière ?

Le microbiote est l'ensemble des micro-organismes - bactéries, virus, parasites, champignons non pathogènes, dits commensaux - qui vivent dans un environnement spécifique. Les écosystèmes microbiens diffèrent chez l'hôte en fonction de leur localisation, ainsi le microbiote buccal ou celui de la peau est très différent du microbiote intestinal.

Le microbiote s'installe progressivement dans l'intestin dès la naissance. Il estensemencé principalement à la naissance et est initialement indifférencié dans les divers habitats du corps. La colonisation par le microbiote commensal a lieu de façon progressive, dans un ordre précis : les premières bactéries intestinales ont besoin d'oxygène pour se multiplier (bactéries aérobies : comme les entérocoques, et les staphylocoques). En consommant l'oxygène présent dans l'intestin, elles favorisent ensuite l'implantation de bactéries qui ne prolifèrent qu'en absence de ce gaz (bactéries anaérobies : Bacteroides, Clostridiales, Bifidobacterium). Divers facteurs, y compris la méthode d'accouchement (voie vaginale versus césarienne), l'allaitement et le sevrage, influencent le microbiote du nourrisson (Fig. 9). Par exemple, le microbiote des bébés délivrés par voie vaginale est dominé par *Lactobacillus*, *Prevotella* et *Atopobium*, alors que les bébés délivrés par césarienne ont un microbiote qui ressemble plus au microbiote cutané maternel, les staphylocoques étant un membre précoce dominant²⁵. Des preuves commencent à apparaître que l'environnement in utero peut ne pas être aussi stérile qu'on le pensait à l'origine, avec des bactéries telles que *Enterococcus fecalis*, *Staphylococcus epidermidis* et *Escherichia coli* isolées du méconium (les premières fèces) de nouveau-nés en bonne santé²⁶. Sous l'influence de la diversification alimentaire, de la génétique, du niveau d'hygiène, des traitements médicaux reçus (comme l'antibiothérapie) et de l'environnement, la composition du microbiote intestinal va évoluer qualitativement et quantitativement pendant les premières années de vie. Ensuite, la composition qualitative et quantitative du microbiote reste assez stable au cours de la vie.

Cependant le microbiote intestinal semble devenir plus stable tout au long de l'âge adulte. Un dernier ensemble de changements liés à l'âge dans la composition et la fonction du microbiote intestinal survient pendant la vieillesse. Le vieillissement est associé à des fonctions physiologiques altérées, y compris la fonction du système immunitaire, qui affectent la composition du microbiote intestinal. Les différences liées à l'âge rapportées dans la composition du microbiote intestinal incluent une augmentation du nombre total d'anaérobies facultatifs, des changements dans le ratio des Bacteroidetes/Firmicutes, et une diminution marquée des bifidobactéries chez les personnes de plus

de 60 ans, au moment où le système immunitaire commence à décliner (Fig. 9)²⁷. Des changements métaboliques coïncidant avec l'évolution et la maturation du microbiote intestinal peuvent être trouvés dans les profils d'excrétion des produits bactériens du métabolisme des acides aminés et dans les métabolites liés à l'énergie. Par exemple, les produits bactériens du métabolisme de la choline peuvent être inversement associés à l'âge chez les enfants de moins de 12 ans²⁸. Des changements dans l'excrétion urinaire de l'acide 4-hydroxyphénylacétique, de l'acide indolacétique et des métabolites du cycle de l'acide tricarboxylique ont été observés au début de la puberté²⁹.

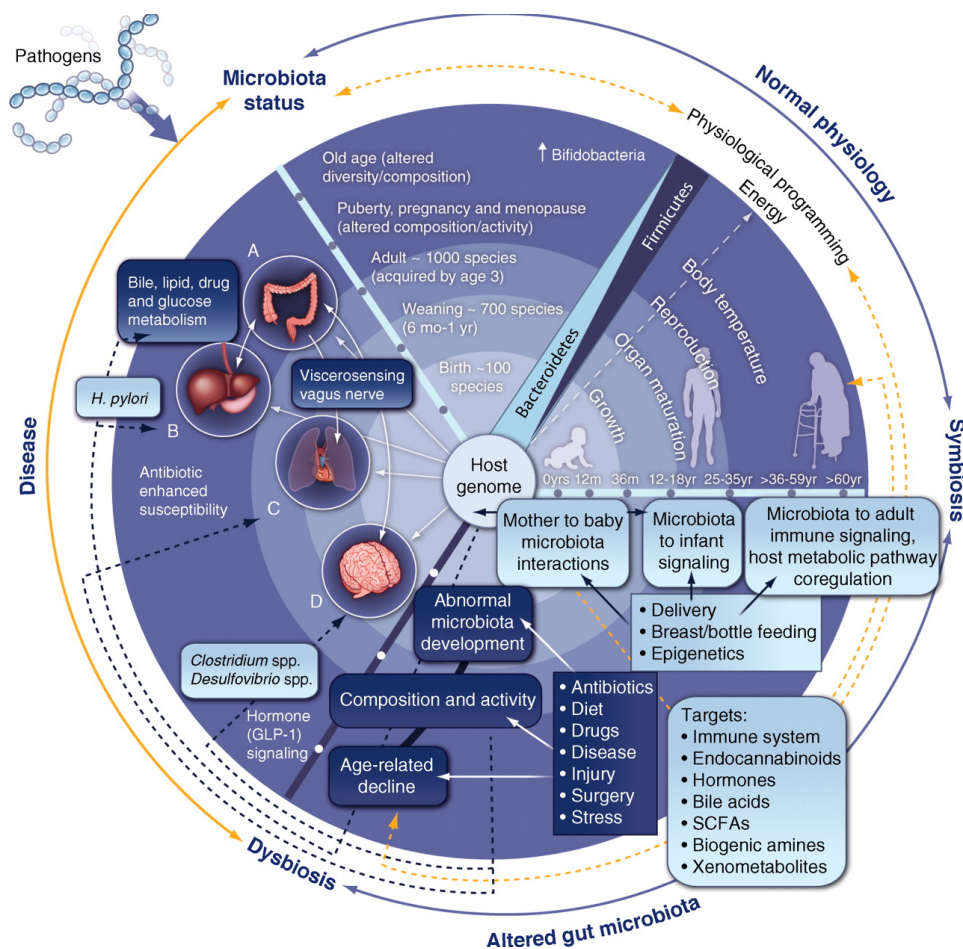


Figure 9 : Le microbiote intestinal dans le développement et la maladie. L'influence du microbiote intestinal sur la santé humaine est continue de la naissance à la vieillesse. À la naissance, environ 100 espèces microbiennes peuplent le côlon. Pendant l'allaitement, la communauté microbienne se développe rapidement. Les changements dans la diversité microbienne se produisent tout au long de la vie des enfants et des adultes. Le microbiote intestinal est important pour maintenir la physiologie normale et la production d'énergie tout au long de la vie. La perturbation du microbiote intestinal (dysbiose) peut conduire à une variété de maladies différentes, y compris (A) la maladie inflammatoire de l'intestin, le cancer du côlon et le syndrome du côlon irritable ; (B) les ulcères gastriques, la stéatose hépatique non alcoolique, et l'obésité et les syndromes métaboliques ; (C) l'asthme et l'hypertension ; et (D) l'humeur et le comportement par signalisation hormonale (par exemple, GLP-1). Le microbiote intestinal est également important pour le métabolisme des médicaments et empêche la colonisation par des microbes pathogènes.³⁰

Notre microbiote intestinal peut être représenté comme un organe microbien placé dans un organe hôte : il est composé de différentes communautés bactériennes capables de communiquer aussi bien entre elles qu'avec les cellules de l'hôte. Il consomme, stocke et redistribue l'énergie. Il intervient dans les transformations chimiques physiologiquement importantes et il peut se maintenir et se réparer par autoréplication. Le Microbiome intestinal, qui peut contenir plus de 100 fois le nombre de gènes de notre génome, nous confère des caractéristiques fonctionnelles que nous n'avons pas eu à développer nous-mêmes.

La plupart des microbes sont des bactéries, l'intestin peut également héberger des levures, des eucaryotes unicellulaires, des virus et de petits vers parasites. Le nombre, le type et la fonction des microbes varient le long du tractus gastro-intestinal mais la majorité se trouve dans le gros intestin où ils contribuent à la fermentation des composants alimentaires non digérés, en particulier les glucides et les fibres, et à la constitution de la masse fécale.

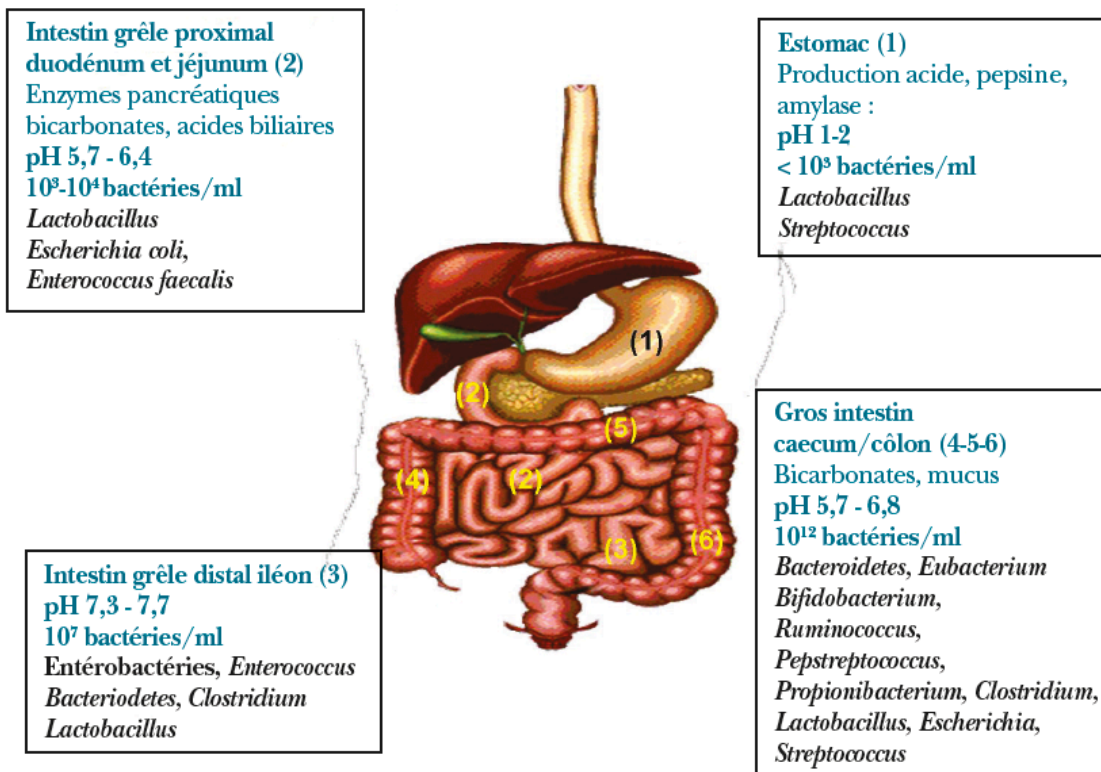


Figure 10 : La composition du microbiote évolue le long du tube digestif en fonction de l'écologie locale.³¹

Le microbiote intestinal est dominé par deux phyla bactériens, les Firmicutes et les Bacteroidetes, qui constituent près de 90 % de la population totale, les Firmicutes étant plus abondants que les Bacteroidetes³². Les autres phyla bactériens présents dans l'intestin humain sont les Actinobacteria, les Protéobacteria, les Verrucomicrobia, les Fusobacteria, les Cyanobacteria, les Spirochaetes. Le phylum des Firmicutes est majoritairement composé de bactéries appartenant aux Clostridia tel que par exemple *Ruminococcus* et *Faecalibacterium prausnitzii*. Le phylum des Bacteroidetes est, quant à lui, représenté par plusieurs espèces de Bacteroides, y compris *B. thetaiotaomicron*, *B. fragilis*, *B. ovatus*, et *B. caccae*. Le phylum des Actinobacteria comprend les Bifidobactéries. Chez l'adulte, le microbiote intestinal est dominé par des espèces bactériennes anaérobies strictes voir d'espèces extrêmement sensibles à l'oxygène. Cependant des bactéries anaérobies facultatives comme des Entérobactéries sont trouvées mais en nombre relativement faible dans le microbiote adulte. L'écosystème intestinal d'un adulte sain est généralement composé de 1000 espèces moléculaires dont on estime que 60-70% restent incultivées à ce jour²⁷.

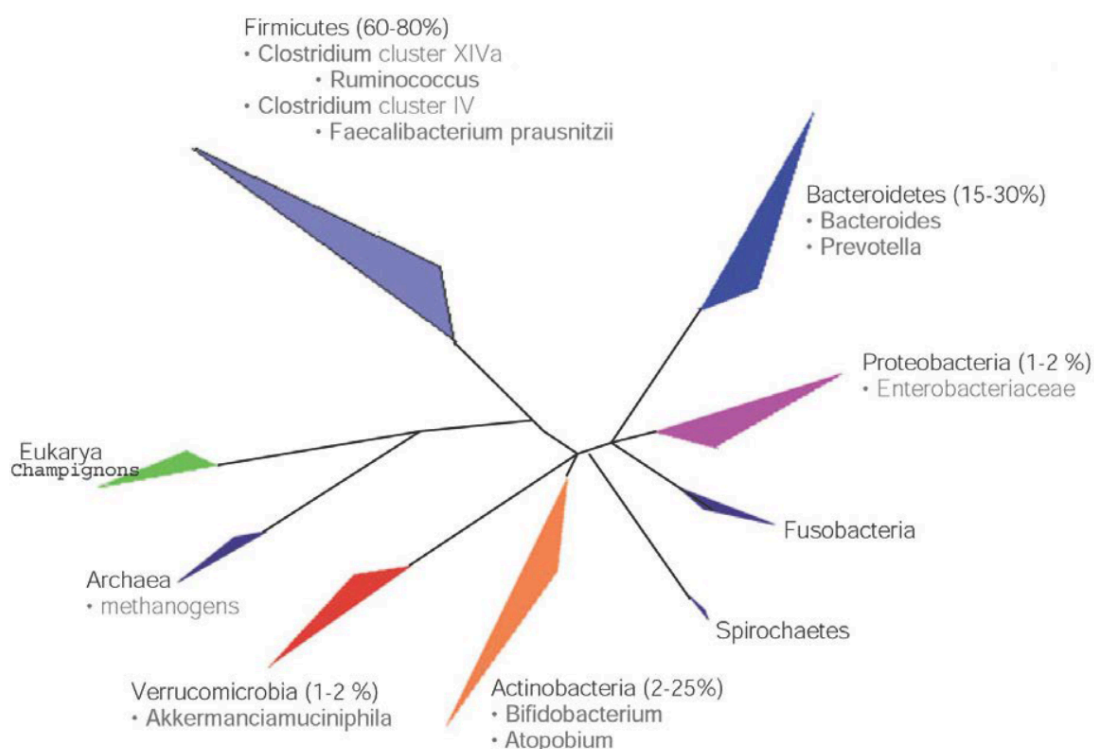


Figure 11 : Représentation schématisée de l'arbre phylogénétique des bactéries résidant dans le côlon montrant l'abondance relative des phyla majeurs du microbiote intestinal humain. Entre parenthèse, la proportion des différents phyla dans le microbiote intestinal total est précisée. Les groupes, genres et espèces bactériennes qui peuvent être importants pour la santé humaine sont indiqués³³.

Certains des genres les plus communs ou reconnus de bactéries intestinales chez les adultes sont *Bifidobacterium*, *Lactobacillus*, *Bacteroides*, *Clostridium*, *Escherichia*, *Streptococcus* et *Ruminococcus*. Bien que les individus puissent avoir jusqu'à plusieurs centaines d'espèces de microbes dans leur intestin, des découvertes récentes du Human Microbiome Project et d'autres projets^{34,35} montrent que des milliers de microbes différents peuvent coloniser collectivement les populations humaines et montrer un degré élevé de variation dans la composition de ces populations entre individus. Malgré cette variation dans les taxons, l'abondance de nombreux gènes microbiens impliqués dans les activités métaboliques de base ou d'entretien est assez similaire entre les individus⁴. Il y a de plus en plus de preuves que les déséquilibres dans la répartition des populations microbiennes intestinales peuvent être associés à la maladie, y compris la maladie inflammatoire de l'intestin³⁶.

Le microbiote intestinal est principalement localisé dans l'intestin grêle et le côlon – l'acidité gastrique rendant la paroi de l'estomac quasi stérile, même si certaines bactéries (comme *Helicobacter pylori*) peuvent s'y développer. Il est réparti entre la lumière intestinale et le biofilm protecteur que forme le mucus intestinal sur sa paroi intérieure (l'épithélium intestinal).

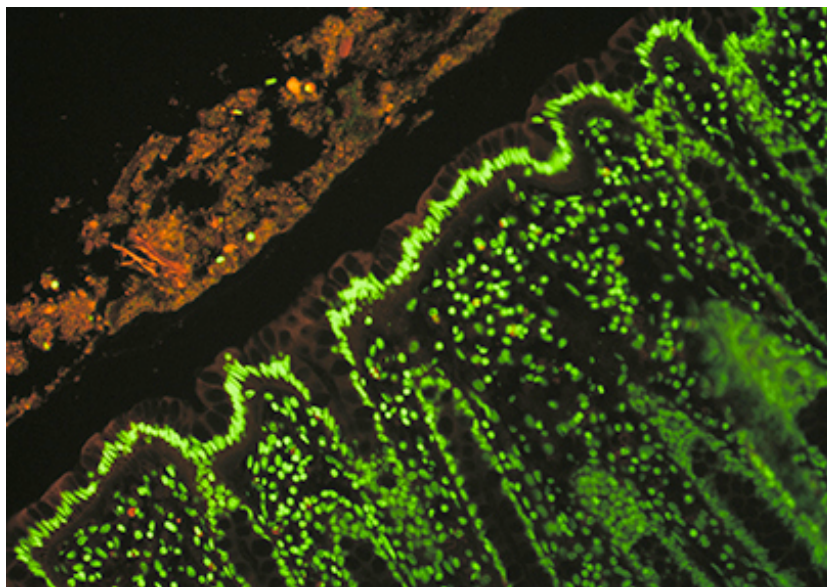


Figure 12 : Microbiote (marquage rouge par la méthode FISH) et intestin (marquage vert/conversion DAPI). Mise en évidence du positionnement de la flore intestinale commensale pour l'essentiel à distance de la surface épithéliale du fait de la présence de mucus et des molécules antimicrobiennes à qui il sert de matrice. Illustration de l'environnement auquel *Lactobacillus* est confronté lors de son processus de colonisation et d'établissement © Inserm, T. Pédro.

L'intestin d'un humain adulte abrite environ 1 kg de bactéries actives sur une surface développée d'environ 400 m² (villosités et microvillosités). Un individu en bonne santé abrite en moyenne 100 milliards de bactéries dans un gramme de selles humaines³⁷.

Le microbiote intestinal est assez spécifique à chaque individu : il est unique sur le plan qualitatif et quantitatif³⁸. Sur les 300 à 1000 espèces qui constituent le microbiote intestinal d'un individu sain, environ 50% des espèces sont communes²⁷ et retrouvées d'un individu à l'autre. Bien que cela soit discuté, il semble que l'on puisse distinguer des groupes homogènes de populations, selon la nature des espèces qui prédominent dans leur microbiote : on distingue trois grands groupes – ou entérotypes – principaux : *Bacteroides*, *Prevotella* et *Ruminococcus*. Indépendants de l'origine géographique, ils seraient en revanche associés aux habitudes alimentaires au long cours. Ainsi, une alimentation riche en protéines et graisses animales est souvent associée à l'entérotype dominé par le genre bactérien *Bacteroides*, quand une alimentation riche en fibres, fruits et légumes l'est davantage à l'entérotype *Prevotella*.

D'un point de vue immunologique, les microorganismes sont considérés comme des agents pathogènes par le système immunitaire de l'hôte qui les reconnaît et les élimine en situation de défense physiologique. Cependant, la majorité des bactéries intestinales sont non pathogènes et cohabitent avec les entérocytes dans une relation symbiotique et le système immunitaire joue aussi un rôle clé dans cette homéostasie. Ce microbiote intestinal assure son propre métabolisme en puisant dans les aliments de l'hôte (notamment parmi les fibres alimentaires). En contrepartie, ses micro-organismes jouent un rôle direct dans la digestion de l'organisme qui les héberge³⁹⁻⁴⁰.

Les bactéries du microbiote intestinal sont capables de produire une vaste gamme de produits dont la production peut dépendre de nombreux facteurs, dont la disponibilité des nutriments et l'environnement luminal, en particulier le pH⁴¹. Les produits microbiens peuvent être absorbés par les tissus gastro-intestinaux, atteindre potentiellement la circulation et d'autres tissus, et être excrétés dans l'urine ou la respiration. La fermentation des fibres et des protéines par les bactéries du gros intestin produit certains des produits physiologiquement importants à savoir les acides gras à chaîne courte (AGCC) qui agissent comme sources d'énergie pour les tissus colorectaux et les bactéries^{42,43}. Les AGCC peuvent atteindre la circulation et avoir un impact sur la fonction immunitaire et l'inflammation dans des tissus tels que les poumons⁴⁴. Cependant, certains produits de fermentation protéique tels que l'ammoniac, les phénols et le sulfure d'hydrogène peuvent être toxiques. Les bactéries telles que *Bifidobacterium* peuvent générer des vitamines (exemple, K, B12, biotine, folate,

thiamine)⁴⁵. La synthèse des acides biliaires secondaires, composants importants du transport des lipides chez l'homme, est assurée par des bactéries, comme celles des genres *Lactobacillus*, *Bifidobacterium* et *Bacteroides*²⁵. De nombreux lipides ayant une activité biologique sont également produits par des bactéries, notamment le lipopolysaccharide (LPS), un composant de la paroi cellulaire des bactéries Gram négatif qui peut provoquer une inflammation des tissus⁴⁶.

De nombreuses enzymes produites par les bactéries de l'intestin influencent la digestion et la santé. En effet, une grande partie de la diversité microbienne dans l'intestin humain peut être attribuée au spectre de la capacité enzymatique microbienne nécessaire pour dégrader les nutriments, en particulier les nombreuses formes de polysaccharides complexes consommées par l'Homme⁴⁷. Certaines bactéries telles que *Bacteroides thetaiotamicron* ont la capacité de produire une gamme d'enzymes nécessaires à la dégradation des hydrates de carbone⁴⁸. Les phytases bactériennes du gros intestin dégradent l'acide phytique présent dans les céréales, libérant des minéraux tels que le calcium, le magnésium et le phosphate⁴⁹ pour les rendre disponibles pour les tissus de l'hôte. Les enzymes qui dégradent les mucines aident les bactéries à satisfaire leurs besoins énergétiques et aident au renouvellement normal de la barrière muqueuse qui tapisse l'intestin.

Cependant, les modifications majeures de la structure du microbiote intestinal, telles que des modifications au niveau de l'ordre ou du phylum bactérien affectent la fonctionnalité du microbiote. De cette dysbiose peut résulter une réponse immunitaire inappropriée de l'hôte. Ainsi il est maintenant suggéré que des pathologies liées à des troubles de l'immunité, ou des désordres métaboliques peuvent être déclenchées, ou tout du moins aggravées, par les bactéries que nous hébergeons. Ainsi, au cours des dernières années, plus de 25 pathologies ou troubles fonctionnels ont été associés de façon plus ou moins convaincante avec des modifications du microbiote intestinal. Le Tableau 1 résume les dix pathologies pour lesquelles il existe des preuves solides d'interactions microbiote–pathologies (Tableau issu de De Vos, 2012).

Tableau 2 : Pathologies ou troubles fonctionnels associés à des dysbioses du microbiote intestinal.⁵⁰

Pathologies	Observations les plus pertinentes et corrélations potentielles
Maladie de Crohn	Diminution de la diversité du microbiote Réduction de <i>F. prausnitzii</i>
Rectocolite hémorragique	Diminution de la diversité du microbiote Réduction de <i>A. muciniphila</i>
Syndrome de l'intestin irritable	Augmentation de <i>Dorea</i> et de <i>Ruminococcus</i>
Infection à <i>Clostridium difficile</i>	Forte diminution de la diversité du microbiote Présence de <i>C. difficile</i>
Cancer colorectal	Variation des <i>Bacteroides</i> Augmentation des Fusobacteria
Allergie / Atopie	Diversité altérée Signatures microbiennes spécifiques
Maladie cœliaque	Composition altérée particulièrement dans l'intestin grêle
Diabète de type 1	Signature microbienne particulière
Diabète de type 2	Signature microbienne particulière
Obésité	Rapport <i>Bacteroidetes/Firmicutes</i> spécifique

La présence de micro-organismes dans l'intestin est connue depuis plus d'un siècle et on a vite présumé qu'il existait une véritable symbiose entre notre organisme et ce microbiote intestinal. Aujourd'hui, le rôle du microbiote intestinal est de mieux en mieux connu. On sait désormais qu'il joue un rôle dans les fonctions digestive, métabolique, immunitaire et neurologique. En conséquence, l'altération qualitative et quantitative de la flore intestinale ou dysbiose est une piste sérieuse pour comprendre l'origine de certaines maladies, notamment celles sous-tendues par des mécanismes auto-immuns ou inflammatoires.

En conditions saines, les bactéries constituant le microbiote intestinal humain adulte sont, pour une très grande majorité d'entre elles, anaérobies strictes (environ 70%) et ne survivent pas plus de quelques minutes en présence d'oxygène. Leur mise en culture est donc rarement possible et dès les années 1980, un défaut de dénombrement de ces bactéries a été décrit. Pour un même échantillon microbien, le nombre de bactéries observées au microscope était ainsi largement supérieur au nombre de colonies dénombrées en culture sur boîte de Pétri.

Le développement du séquençage avec des méthodes indépendantes de la culture (Figure 13) basées, entre autres, sur l'analyse du gène codant l'ARN de la sous-unité 16S du ribosome (*16S rRNA gene*) a rapidement permis d'élargir nos connaissances sur la diversité du microbiote digestif. Seulement dix ans après leur première application à l'intestin humain⁵¹ (18), le nombre de phylotypes (ou espèces moléculaires, appelées OTUs pour *Operational Taxonomic Units*) décrits au niveau du tractus digestif en utilisant des techniques moléculaires, a très largement dépassé celui des espèces cultivées. Parmi plus de 1200 microbes décrits, seulement 12 % ont été détectés par l'application à la fois d'une approche moléculaire et de la culture, alors que la grande majorité (~ 75 %) a été uniquement observée par séquençage du gène codant l'ARNr 16S (19).

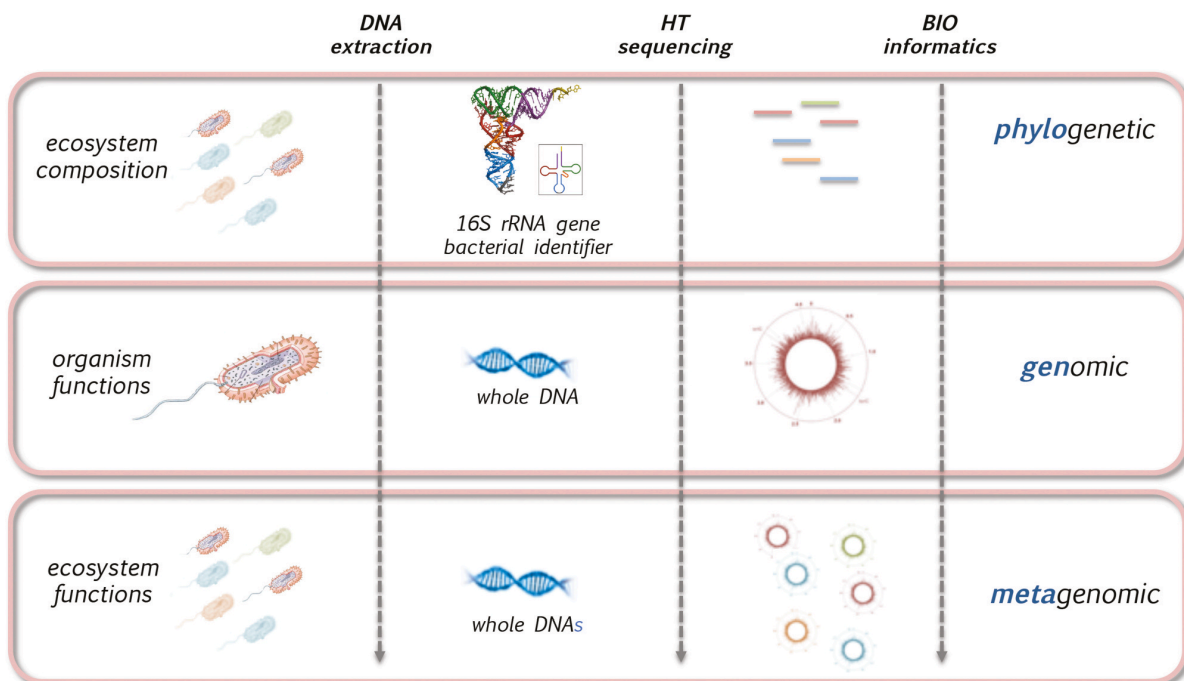


Figure 13 : Description des écosystèmes microbiens complexes par des approches moléculaires indépendantes de la culture. Les approches utilisées comportent trois étapes : extraction des acides nucléiques, séquençage à haut débit (high-throughput sequencing, HT) et analyse bio-informatique. Elles permettent la caractérisation des écosystèmes microbiens au niveau de leur composition en microorganismes (phylogénétique) via l'étude du gène codant l'ARNr 16S, et de leurs fonctions potentielles (métagénomique) par le séquençage sans a priori de tous les génomes bactériens présents dans l'environnement.⁵²

3) L'exploration du microbiote : méthodes de Séquençages

A - Quelques éléments historiques

La séquence d'ADN contient l'information nécessaire aux êtres vivants pour survivre et se reproduire. Déterminer cette séquence est donc utile aussi bien pour les recherches visant à savoir comment vivent les organismes que pour déterminer leur implication dans des pathologies, par exemple. En médecine, elle peut être utilisée pour identifier, diagnostiquer et potentiellement trouver des traitements à des maladies génétiques et à la virologie. En biologie, l'étude des séquences d'ADN est devenue un outil important pour la classification des espèces.

En biochimie, le séquençage consiste à déterminer l'ordre linéaire des composants d'une macromolécule (les acides aminés d'une protéine, les nucléotides d'un acide nucléique comme l'ADN, les monosaccharides d'un polysaccharide, etc.).

En génétique, le séquençage concerne la détermination de la séquence des gènes voire des chromosomes, voire du génome complet, ce qui techniquement revient à effectuer le séquençage de l'ADN constituant ces gènes ou ces chromosomes. Le séquençage d'un génome consiste donc en la détermination de la séquence nucléotidique de l'ADN présent dans chaque cellule d'un organisme donné.

Cette détermination est en général d'autant plus difficile que le génome étudié est grand et riche en séquences répétées. Les virus, qui possèdent de petits génomes dénués de séquences répétées (entre 3 000 et 150 000 paires de bases - souvent moins de 10 000), ont ainsi été les premiers « organismes » séquencés, et représentent aujourd'hui encore la majorité des organismes séquencés. La première bactérie a été séquencée en 1995⁵³, et de nombreux autres procaryotes ont depuis été séquencés intégralement⁵⁴. La taille de leur génome est de l'ordre de quelques millions de paires de bases (mégabases, Mb). La difficulté est tout autre pour les organismes eucaryotes : la grande taille de leur génome (2 à 3 milliards de paires de bases pour les mammifères, par exemple) nécessite un travail cartographique préalable et, souvent, un effort concerté de plusieurs centres de séquençage. Toutefois, les « petits » génomes eucaryotes (tel celui de la paramécie, qui ne mesure « que » 100 Mb) peuvent aujourd'hui être séquencés sans cartographie préalable par un seul grand centre. Si les eucaryotes intégralement séquencés demeurent moins nombreux que les procaryotes et virus, leur nombre est en constante augmentation. Certaines séquences sont à l'état d'ébauches très fragmentaires et incomplètes ; d'autres donnent une vision beaucoup plus complète du génome. Cela reflète l'effort

de séquençage consenti ainsi que la stratégie adoptée par les auteurs du séquençage. En 2018, on compte plus de 219 000 projets de séquençage achevés ou en cours⁵⁵.

Pour ce qui est de l'Homme, la première séquence du génome humain annoncée à grands cris médiatiques fin 2000, n'était en rien une séquence complète. Celle-ci fut disponible en avril 2003.

La liste qui suit donne quelques exemples des premiers génomes entièrement séquencés :

- Virus : 3 778 virus séquencés au 4 mars 2014, dont le VIH (virus d'immunodéficience humaine).

- Procaryotes :

- Archaeobactéries : 319 génomes entièrement séquencés et 447 partiellement au 4 mars 2014.
- Eubactéries : 12 286 génomes entièrement séquencés et 20 403 partiellement au 4 mars 2014.

Exemples :

- *Escherichia coli* :

- *Agrobacterium tumefaciens*
- *Haemophilus influenzae* Rd. (Premier génome cellulaire séquencé, en 1995)

- Eucaryotes : 314 génomes entièrement séquencés et 6660 partiellement au 4 mars 2014. Exemples (les 5 premiers cités sont les 5 premiers publiés, avec la réserve citée pour celui de l'Homme) :

- *Saccharomyces cerevisiae* (levure, premier eucaryote à être séquencé en 1997. Plusieurs souches séquencées de nos jours)
- *Caenorhabditis elegans* (ver nématode)
- *Drosophila melanogaster* (mouche du vinaigre)
- *Arabidopsis thaliana* (arabette - petite plante de la famille du chou)
- *Homo sapiens* (nous...)
- *Neurospora crassa* (champignon ascomycète)
- *Anopheles gambiae* (moustique)
- *Takifugu rubripes* (fugu : poisson-ballon consommé au Japon)
- *Mus musculus* (souris)
- *Plasmodium falciparum* (parasite intracellulaire responsable du paludisme)
- *Oriza sativa* (riz - deux sous-espèces séquencées, japonica et indica)

Le séquençage de l'ADN a été inventé dans la deuxième moitié des années 1970. Deux méthodes ont été développées indépendamment, l'une par l'équipe de Walter Gilbert en 1975 aux Etats-Unis et l'autre par celle de Frederick Sanger en 1977, au Royaume-Uni. Ces deux méthodes sont fondées sur des principes diamétralement opposés : l'approche de Sanger est une méthode par synthèse enzymatique sélective, tandis que celle de Maxam et Gilbert est une méthode par dégradation

chimique sélective. Pour cette découverte, Gilbert et Sanger ont été récompensés par le prix Nobel de chimie en 1980.

B - Méthode de Maxam et Gilbert

En 1976, Allan Maxam et Walter Gilbert, publièrent une méthode de séquençage basée sur la dégradation chimique de l'ADN et l'usage des réactivités différentes des quatre bases A, T, G et C, pour réaliser des coupures sélectives⁵⁶. L'ADN à séquencer est marqué en 5' avec (gamma-32P) dATP puis clivé après le « A », « G », « C » ou le « T » par divers réactifs chimiques. Par exemple, le diméthylsulfate est un réactif qui méthyle la guanine sur l'azote en 7 et l'adénine sur l'azote en 3. La liaison glycosidique d'une purine méthylée est instable et est facilement rompue par chauffage, laissant ainsi libre la fonction semi-alhéydrique de l'ose. Mais dans le squelette sucre-phosphate, l'endroit où se trouve un ose non relié à une base représente un point faible. Ainsi le clivage ose-phosphate se fera facilement de part et d'autre de ce point. En jouant sur les conditions opératoires, on peut favoriser le clivage en G ou en A. Pour la cytosine et la thymine, le réactif utilisé est l'hydrazine. Pour chaque fragment, les produits des différentes réactions sont séparés par électrophorèse en conditions dénaturantes et analysés pour reconstituer la séquence de l'ADN. Cette analyse est analogue à celle que l'on effectue pour la méthode de Sanger, détaillée plus loin. La méthode de Maxam et Gilbert nécessite des réactifs chimiques toxiques et reste limitée quant à la taille des fragments d'ADN qu'elle permet d'analyser (< 250 nucléotides). Moins facile à robotiser, son usage est devenu aujourd'hui confidentiel.

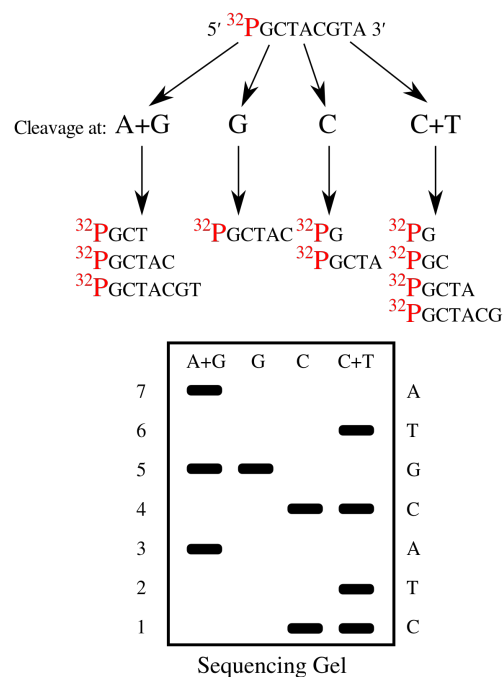


Figure 14 : Principe de la technique de séquençage de Maxam et Gilbert
(From Wikipedia, the free encyclopedia)

C - Méthode de Sanger

Dans les années 50, Fred Sanger a été le premier à déterminer la séquence d'acides aminés dans une protéine. Il a montré que les 51 acides aminés de la protéine d'insuline sont disposés dans un ordre spécifique, ce qui lui a valu le prix Nobel de biochimie en 1958. Puisque le code génétique détermine l'ordre des acides aminés, la séquence d'ADN est colinéaire avec la séquence d'acides aminés. Cependant, connaître la séquence d'acides aminés d'une protéine n'indique pas la séquence nucléotidique exacte de son gène. Le code génétique étant redondant - plusieurs codons peuvent coder pour un acide aminé. Au début des années 70, Fred Sanger a développé une méthode pour déterminer la séquence exacte des nucléotides dans un gène. La méthode était basée sur les travaux antérieurs d'Arthur Kornberg sur la réplication de l'ADN⁵⁷.

Initialement, la méthode de Sanger nécessitait de disposer d'un ADN simple brin qui servait de matrice pour la synthèse enzymatique du brin complémentaire. Pour cette raison, le premier organisme biologique dont le génome a été séquencé en 1977 est le virus bactériophage ϕ X174⁵⁸. Ce virus a la propriété d'avoir un génome constitué d'ADN simple brin qui est encapsulé dans la particule virale.

Le principe de cette méthode consiste à initier la polymérisation de l'ADN à l'aide d'un petit oligonucléotide (amorce) complémentaire à une partie du fragment d'ADN à séquencer. L'élongation de l'amorce est réalisée par le fragment de Klenow (une ADN polymérase I dépourvue d'activité exonucléase 5'→3') et maintenue par des ADN polymérases thermostables, celles qui sont utilisées pour la PCR. Les quatre désoxyribonucléotides (dATP, dCTP, dGTP, dTTP) sont ajoutés, ainsi qu'une faible concentration de l'un des quatre didésoxyribonucléotides (ddATP, ddCTP, ddGTP ou ddTTP)⁵⁹.

Un didésoxyribonucléotide triphosphate est un analogue structural du désoxynucléoside triphosphate mais il ne possède pas de groupement OH en 3'. Il s'incorpore normalement dans une chaîne en cours de synthèse, mais l'élongation de la chaîne est stoppée aussitôt après son incorporation. Cette terminaison se fait spécifiquement au niveau des nucléotides correspondant au didésoxyribonucléotide incorporé dans la réaction.

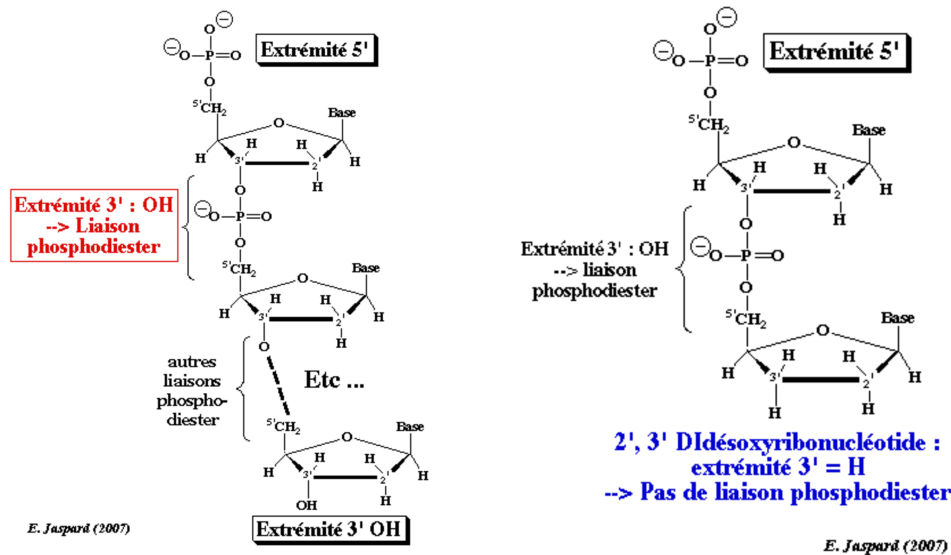


Figure 13 : Structures d'un dNTP et ddNTP

La réaction enzymatique s'effectue en traitant de manière parallèle quatre tubes contenant chacun : un brin d'ADN, quatre désoxynucléotides non radioactifs, l'amorce universelle et une ADN polymérase. Chacun des tubes contient en outre un didésoxy nucléotide radioactif, par exemple ddATP dans le premier, ddGTP dans le deuxième, ddCTP dans le troisième et ddTTP dans le quatrième. Les rapports des quantités didésoxy/désoxy nucléotides à utiliser sont bien précis. Par exemple, dans la réaction où on a ajouté du ddATP, quand l'ADN polymérase arrive en face d'un « T » situé sur le brin de l'ADN matrice, elle incorpore un « A » sur le brin complémentaire en voie de synthèse. Le mélange réactionnel contenant, à la fois du dATP et un peu de ddATP, l'ADN polymérase va utiliser arbitrairement l'un ou l'autre. Si c'est le ddATP qui est incorporé, la réaction s'arrête. Si c'est le dATP qui est incorporé, la réaction de synthèse se poursuit jusqu'au prochain « A » à incorporer et de nouveau le choix s'offre entre dATP et ddATP. Il en résulte un mélange de fragments d'ADN de tailles croissantes, qui se terminent tous au niveau d'un des « A » dans la séquence. Ces fragments sont ensuite séparés par électrophorèse sur un gel de polyacrylamide⁶⁰, ce qui permet ainsi de repérer la position des « A » dans la séquence. Il en va de même pour les autres bases.

La détection des fragments ainsi synthétisés se fait en incorporant un traceur dans l'ADN synthétisé. Initialement ce traceur était radioactif ; aujourd'hui, on utilise des traceurs fluorescents, attachés soit à l'oligonucléotide, soit au didésoxyribonucléotide.

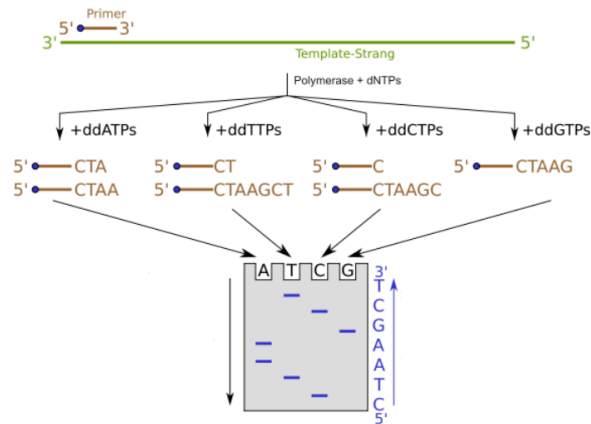


Figure 14 : Principe de la technique de séquençage de Sanger⁶¹

Au cours des 25 dernières années, la méthode de Sanger a été largement développée grâce à plusieurs avancées technologiques importantes :

- la mise au point de vecteurs de séquençage adaptés, comme le phage M13 développé par Joachim Messing au début des années 1980⁶². C'est un bactériophage spécifique d'*Escherichia. coli*. Il est constitué d'un simple brin d'ADN circulaire comprenant 10 gènes et deux petites régions non codantes « intergénique ». Un avantage remarquable est que la séquence immédiatement adjacente au polylinker est la même pour tous les vecteurs M13 de la série de Messing.

- l'introduction de traceurs fluorescents à la place des marqueurs radioactifs utilisés initialement. Ce progrès a permis de sortir le séquençage des pièces confinées nécessaires à l'usage de radio-isotopes⁶³.

- Le développement de la synthèse chimique automatisée des oligonucléotides qui sont utilisés comme amorces dans la synthèse. Cette technique utilise des didésoxyribonucléotides dont chacun est marqué par un fluorochrome spécifique. Les fragments d'ADN synthétisés portent ce fluorochrome terminal. On les appelle des terminateurs d'élongation ou "Big Dye Terminators" ou "Dye-labeled terminator".

- L'adaptation de la technique PCR pour le séquençage.

- L'utilisation de l'électrophorèse capillaire pour la séparation et l'analyse⁶⁴.

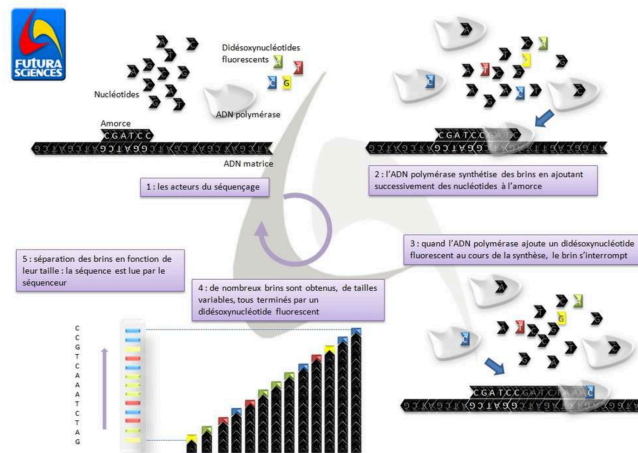


Figure 13 : Principe du séquençage par la méthode de Sanger amélioré avec l'usage de didésoxyribonucléotides liés à un traceur fluorescent et qui permet de détecter les fragments d'ADN synthétisé (Claire Peltier/Futura-Sciences).

D - Séquençage de génome entier

La connaissance de la structure d'un génome dans sa totalité peut passer par son séquençage. Cependant, la taille des génomes étant de plusieurs millions de bases (ou mégabases), il est nécessaire de coupler les approches de biologie moléculaire avec celle de l'informatique pour pouvoir traiter un nombre aussi important de données.

Deux grands principes de séquençage de génome entier sont utilisés, l'ordonnance hiérarchique et la méthode globale. Dans les deux cas, l'ADN génomique est préalablement fragmenté par des méthodes enzymatiques (enzymes de restriction) ou physiques (ultrasons). La principale différence entre ces deux principes est que l'ordonnancement hiérarchique essaie d'aligner un jeu de clones de grande taille (~ 100 kb) alors que dans la méthode globale le génome entier est réduit en fragments de petite taille qui sont séquencés puis alignés.

La méthode de séquençage par ordonnancement hiérarchique consiste à classer les fragments génomiques obtenus avant de les séquencer. Après extraction, l'ADN génomique est découpé par sonication en fragments de 50 à 200 kilo base ou kb (50000 à 200000 paires de base) puis cloné dans des vecteurs spéciaux. Les vecteurs les plus utilisées en biologie moléculaire, plasmides et cosmides, ne permettent pas de cloner des fragments de plus de 45 kb. Ils sont donc inadaptés au clonage d'inserts suffisamment grands pour réaliser une carte physique. Dans un premier temps, les chercheurs ont utilisé les chromosomes artificiels de levure YAC : Yeast Artificial Chromosome. Ces vecteurs permettaient l'insertion de fragments allant jusqu'à 1000 kilobases. Mais ils ont été abandonnés car ils génèrent des clones "chimériques" ne correspondant pas aux inserts clonés. Les vecteurs qui ont été utilisés pour établir la carte physique du génome humain sont principalement les chromosomes

artificiels bactériens (BAC : Bacterial Artificial Chromosome). Bien que de plus faible capacité que les YAC (seulement 300 kb au maximum, 150 kb en moyenne), ils n'en présentent pas les défauts. Des vecteurs dérivés du phage P1 (les PAC), de même capacité, ont aussi été utilisés.

Le nombre de clones doit permettre une couverture de 5 à 10 fois la longueur totale du génome étudié. Le chevauchement et l'ordonnement des clones est réalisé soit par hybridation de sondes spécifiques, soit par analyse des profils de restriction, soit plus fréquemment par un ordonnancement après séquençage et hybridation des extrémités des vecteurs spéciaux. Après ordonnancement des clones, ils sont fragmentés et séquencés individuellement, puis assemblés par alignement bio-informatique.

Les avantages de cette méthode sont une plus grande facilité d'assemblage des fragments grâce aux chevauchements des vecteurs spéciaux, la possibilité de comparer les fragments aux banques de données disponibles et la possibilité de partager le travail de séquençage entre plusieurs laboratoires, chacun ayant en charge une région chromosomique.

L'inconvénient majeur est la difficulté de cloner des fragments contenant des séquences répétées très fréquentes dans certains génomes, comme ceux des mammifères, ce qui rend difficile l'analyse bio-informatique finale.

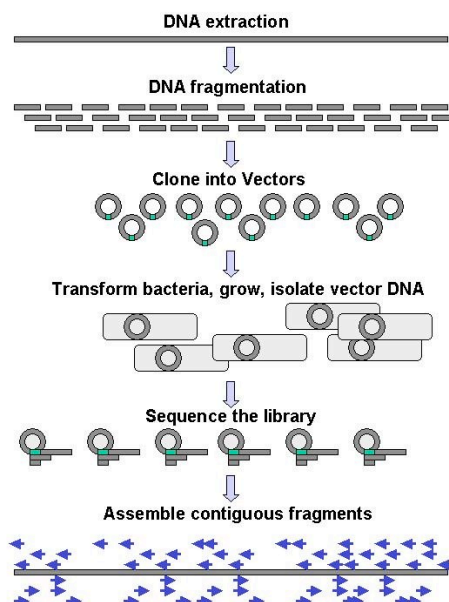


Figure 15 : Les différentes étapes du séquençage par ordonnance hiérarchique.

(From Wikipedia, the free encyclopedia)

La méthode globale (ou whole-genome shotgun) ne fait pas de classement des fragments génomiques obtenus mais les séquence dans un ordre aléatoire. Une analyse bio-informatique faisant appel à un assembleur permet ensuite de réordonner les fragments génomiques par chevauchement des séquences communes. Il s'agit d'une méthode de séquençage d'ADN génomique initialement imaginée dans le laboratoire de Frederick Sanger à Cambridge à la fin des années 1970 pour séquencer les premiers génomes de virus⁶⁵. Cette méthode a été popularisée par Craig Venter pour le séquençage des grands génomes, en particulier au sein de la société Celera Genomics⁶⁶. La première application fut le séquençage de génomes bactériens, puis du génome de la drosophile et enfin du génome humain et murin. Pour réaliser un séquençage de génome complet à l'aide de cette technique, deux à trois banques composées de fragments aléatoires d'ADN génomique sont réalisées. Entre les banques, les fragments divergent aussi bien en taille qu'en localisation sur le génome. À partir de ces banques, de nombreux clones sont séquencés puis assemblés. La séquence totale est obtenue en traitant l'ensemble des banques à l'aide d'outils bio-informatiques et en alignant les fragments à l'aide des séquences chevauchantes.

Les avantages de cette méthode par rapport au séquençage par ordonnancement hiérarchique sont la rapidité de la technique et son coût plus faible. L'inconvénient est que le traitement informatique ne permet pas d'aligner des fragments comportant des séquences répétées de grande taille qui sont fréquemment présentes dans les génomes des mammifères.

Cette méthode est couramment désignée sous le nom de shotgun (fusil à canon scié), ou encore Whole Genome Shotgun (WGS). Cette métaphore illustre le caractère aléatoire de la fragmentation initiale de l'ADN génomique.

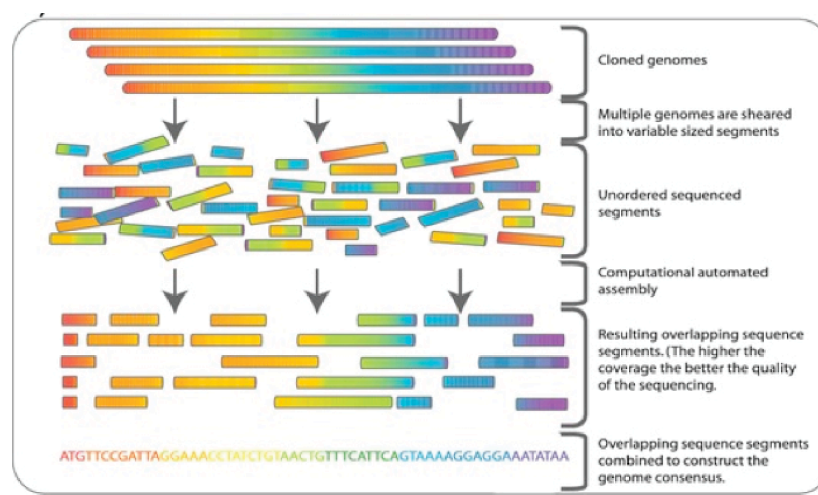


Figure 16 : Différentes étapes du séquençage Shotgun⁶⁷

E - Autres méthodes

E - 1. Séquençage par hybridation

Le séquençage par hybridation repose sur l'utilisation de puces à ADN contenant plusieurs centaines (pour les puces de première génération) à plusieurs milliers d'oligonucléotides. L'ADN à analyser est coupé en de multiples fragments qui sont ensuite incubés sur la puce où ils vont s'hybrider avec les oligonucléotides dont ils sont complémentaires. La lecture de la puce (la détection des oligonucléotides hybridés), permet d'obtenir le spectre de la séquence d'ADN, c'est-à-dire sa composition en sous-séquences de n nucléotides, où n est la taille des sondes sur la puce utilisée. Le traitement informatique du spectre permet ensuite de reconstituer la séquence entière⁶⁸.

E - 2. Séquençage haut débit (HTS)

On désigne par séquençage haut débit (HTS pour high-throughput sequencing) aussi appelé NGS pour next-generation sequencing un ensemble de méthodes apparues à partir de 2005 produisant des millions de séquences en un « run » et à faibles coûts. Elles se caractérisent par l'utilisation d'approches massivement parallèles, permettant de séquencer des centaines de milliers de fragments simultanément. Elles s'affranchissent des étapes de clonage et de constitution de banques génomiques. Elles permettent de séquencer à partir de molécules uniques d'ADN. On distingue :

- Le pyroséquençage : est une méthode rapide et précise qui est basée sur le principe du "séquençage par synthèse". Dans cette technique, les nucléotides ne sont pas ajoutés tous ensemble comme dans une réaction de séquençage normale mais l'un après l'autre. Si le nucléotide ajouté dans le milieu réactionnel correspond à celui attendu par la polymérase, il est incorporé dans le brin en cours de synthèse (d'élongation) et libère un pyrophosphate. Une ATP sulfurylase vient alors transformer ce pyrophosphate (PPi) en ATP qui est alors utilisé, couplé à une luciférine, par une luciférase. On a alors production d'oxyluciférine et d'un signal lumineux. Une apyrase dégrade les nucléotides en surplus). C'est le séquenceur ou plus précisément un capteur CCD (Charge-Coupled Device) qui va capter ce signal lumineux et le reproduire sous forme d'un pic sur le pyrogramme. La hauteur de ce pic est fonction de l'intensité du signal lumineux, elle-même proportionnelle au nombre de nucléotides incorporés en même temps. On peut donc déduire la séquence à partir de la taille des pics obtenus. Par ailleurs, en cas de mélange de nucléotides à une même position (polymorphisme de séquence), la taille des pics permet d'avoir une quantification de la proportion de brins porteurs de l'un ou l'autre des nucléotides.

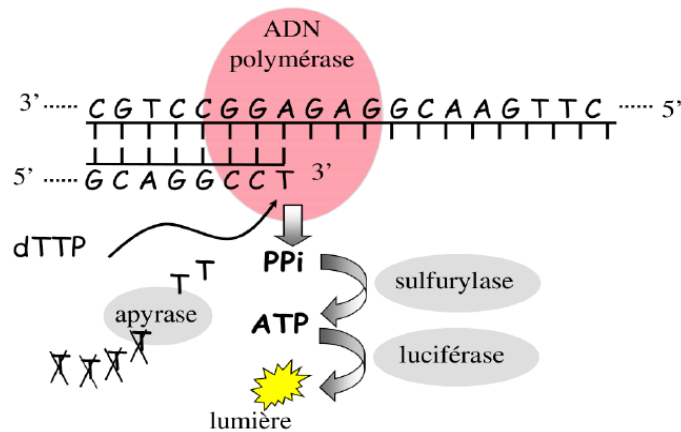


Figure 17 : Principe du pyrosequençage⁶⁹

- Le séquençage de colonie (Colony sequencing) développé dans le laboratoire de Church à Harvard. C'est l'un des premiers systèmes de séquençage à haut débit et il a été utilisé pour séquencer le génome d'*Escherichia coli* en 2005⁷⁰. En biologie, une colonie est un regroupement ponctuel d'ADN, par analogie avec une colonie bactérienne. Le mot colonie est une contraction de « polymérase » et « colonie ». Les colonies sont le résultat de l'amplification d'une seule molécule d'ADN par une enzyme polymérase, au sein d'une matrice de gel. Le concept de l'analyse de régions contenant des populations clonales d'acide nucléique a été décrit en premier par Brown et al, attribué à Genomic Nanosystems, utilisant alors une phase liquide. Le mot « colonie » désigne généralement une colonie d'ADN amplifiée sur billes, en phase solide. Cette technologie, qui s'appelait initialement « génération de colonies d'ADN », a été inventée et développée en 1996 à l'institut de recherche biomédicale de Genève (GBRI) de Glaxo-Wellcome, par le Dr Pascal Mayer et le Dr Laurent Farinelli et a été présentée pour la première fois en 1998. C'est finalement l'entreprise Illumina qui l'a mise sur le marché⁷¹. Pour créer des colonies, une solution contenant un grand nombre de fragments d'ADN différents est diluée et étalée sur une lame de microscope de sorte que les molécules individuelles soient séparées. De l'ADN polymérase est ajoutée pour créer de nombreuses copies des fragments, formant ce que l'on appelle les colonies. L'une des applications des colonies est le séquençage de l'ADN. Cette technique de séparation des molécules d'ADN a inspiré d'autres développements tels que la PCR émulsion et la capture sur billes. Il a ainsi été possible de séquencer 30,1 millions de paires de bases d'une souche d'*E.coli* en 2,5 jours avec une précision de 99,7 %⁷².

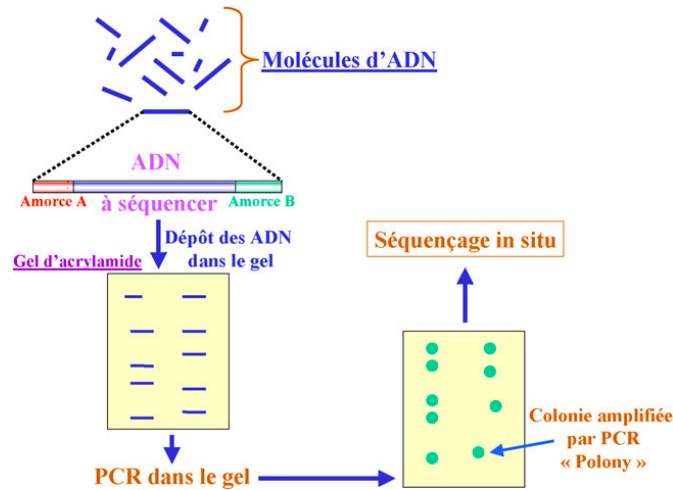


Figure 18 : Principe du séquençage sur polony. Un ensemble de molécules d'ADN issues d'une bibliothèque d'un ADN génomique attachées à des amorces universelles A et B en 3' et 5' par réaction de ligation sont incorporées dans un gel de polyacrylamide contenant tous les réactifs nécessaires à une PCR. Chaque colonie correspond à une seule molécule d'ADN amplifiée. Après PCR, les polonies sont dénaturées in situ et séquencées après addition d'une amorce de séquence et d'une série séquentielle d'un seul nucléotide fluorescent (chaque nucléotide ayant une fluorescence spécifique). Après addition d'un nucléotide, un scanner lira la fluorescence si le nucléotide a été incorporé. Après élimination par lavage des nucléotides non incorporés, le nucléotide suivant est ajouté. La réaction continue⁷³.

- Séquençage SOLiD (sequencing by oligonucleotide ligation and detection), développé par la société Applied Biosystems, permet le séquençage de fragments d'ADN amplifiés en utilisant une technique de ligation. La méthode repose sur la ligation séquentielle d'oligonucléotides couplés à un fluorochrome. Elle peut générer jusqu'à 20 gigabases de données par séquençage. La ligation avec des di-nucléotides marqués permet d'obtenir un très faible taux d'erreurs, car elle ne peut se produire que si les deux nucléotides sont hybridés (Fig. 19). Ce système peut être utilisé pour des applications nécessitant un grand nombre de séquences. Il permet notamment d'obtenir une plus grande sensibilité pour la détection de gènes faiblement exprimés. Cependant, en tenant compte de la longueur relativement courte (35-50 bp) des séquences obtenues, il n'est pas conseillé pour du séquençage de novo. Dans les faits, ce système est peu à peu abandonné au profit du système Illumina.

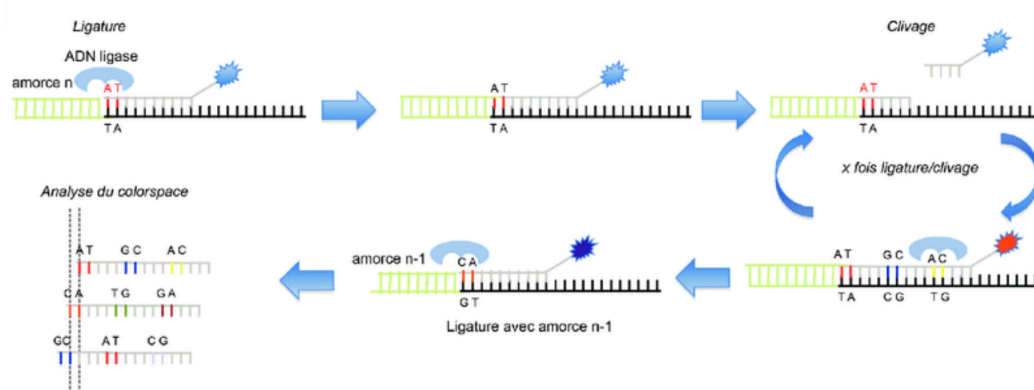


Figure 19 : Principe du séquençage de type Applied Biosystems SOLiD. La technologie repose sur une PCR en émulsion sur billes. Le séquençage s'effectue par ligation. Il utilise 4 fluorophores pour encoder les 16 combinaisons de dinucléotides. Plusieurs cycles de ligation sont nécessaires⁷².

- Le séquençage Illumina : développé initialement par la société Solexa, est fondé sur le séquençage par synthèse et est actuellement le système le plus utilisé au monde, avec environ 90 % des séquençages NGS réalisés. Il nécessite la fabrication préalable d'une banque d'ADN. L'ADN est fragmenté et des adaptateurs de séquençage sont fixés aux extrémités des fragments obtenus. Ces fragments « étiquetés » sont ensuite greffés sur un support solide (la cellule de flux). À cette étape, une amplification est réalisée pour multiplier ces fragments d'ADN immobilisés, puis des didésoxynucléotide couplés à des fluorochromes sont chargés sur la cellule de flux. Ces nucléotides possèdent une propriété de terminaison réversible, permettant à la réaction de séquençage de se produire en présence des quatre nucléotides simultanément.

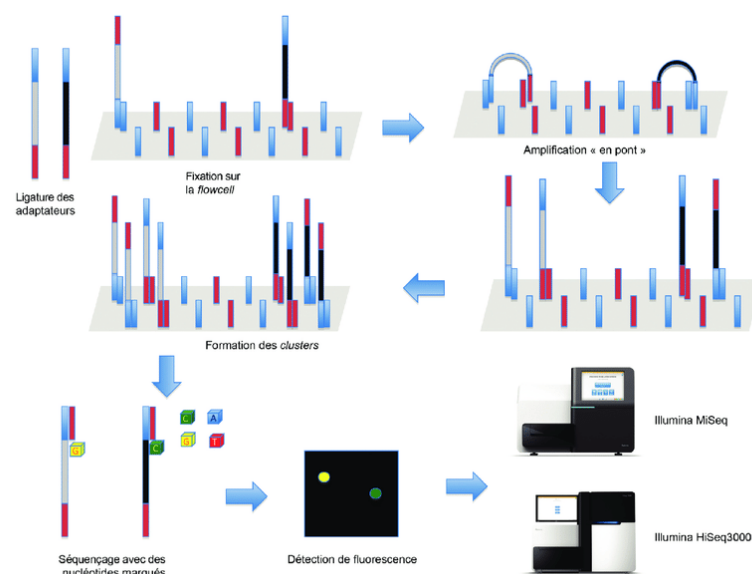


Figure 20 : Représentation schématique du séquençage de type Illumina. Après création de la banque (ligature des adaptateurs en 3' et 5'), celle-ci est chargée sur la cellule de flux et une amplification « en pont » est effectuée. Cette amplification repose sur les adaptateurs 3' et 5' qui sont complémentaires aux sondes fixées sur la cellule de flux. Après n cycles d'amplification, un cluster d'une molécule identique à une position donnée est obtenu. Une étape de séquençage avec des nucléotides marqués succède alors et l'incorporation de ces derniers est détectée par l'émission d'une fluorescence propre au nucléotide⁷⁴.

- Le séquençage PacBio : L'une des avancées les plus importantes est probablement le système de séquençage SMRT (Single Molecule Real Time) mis au point par Pacific Biosciences, réalisant le séquençage en temps réel d'une molécule sans amplification préalable. Une ADN polymérase est fixée au fond d'un puits avec une seule molécule d'ADN comme matrice. Chacun des quatre nucléotides est couplé à un fluorochrome distinct. Lorsque le nucléotide est incorporé par l'ADN polymérase, le fluorochrome est clivé. Un détecteur capte le signal fluorescent correspondant et le nucléotide est identifié selon le fluorochrome détecté. La longueur des séquences obtenues atteint plusieurs kb (jusqu'à 60 selon le fabricant). En 2013, la société a commercialisé PacBio RS II, une

nouvelle version du séquenceur qui utilise les 150 000 puits de la cellule de séquençage simultanément, doublant le débit (150 000 séquences simultanées). Les applications de ce système sont multiples. Cependant, la longueur des séquences produites (10 kb en moyenne) est particulièrement intéressante pour l'assemblage de génome de novo car elle permet de couvrir des régions répétées des séquences d'insertion. Le nombre de séquences totales étant plus faible que celui obtenu avec le système Illumina, la couverture de séquençage maximale ou « coverage » (c'est-à-dire le nombre de fois où chaque nucléotide est séquencé) reste relativement basse (au maximum 50 fois pour un génome de procaryote). De plus, le pourcentage d'erreur est relativement élevé (environ 10-15 %) comparativement à celui observé avec le séquençage Illumina (< 0,01 %).

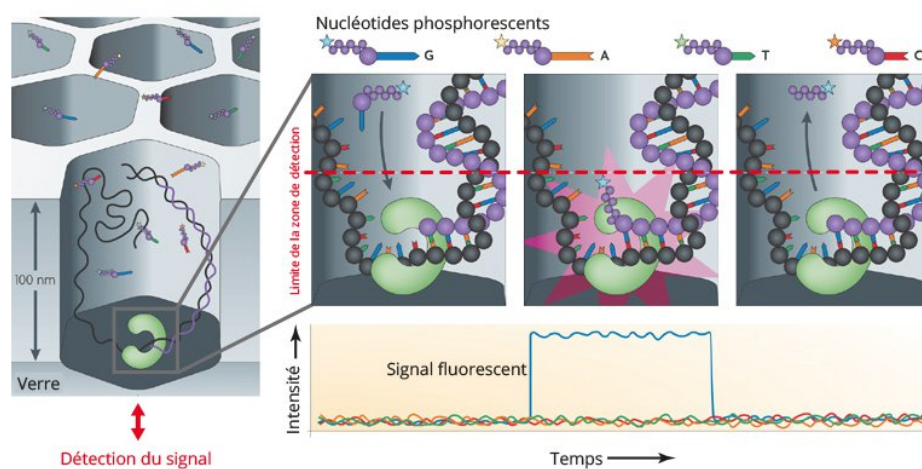


Figure 21 : Séquençage en temps réel de Pacific Biosciences - d'après Nature Reviews : Une enzyme ou catalyseur biologique, responsable de l'élongation de l'ADN est fixée au fond d'un puits. En présence d'une molécule d'ADN simple brin, celle-ci va être capable d'incorporer des nucléotides libres complémentaires au brin à analyser, un par un. Les nucléotides ont la particularité d'être marqués par une molécule fluorescente différente (bleu pour G, jaune pour A, vert pour T et rouge pour G). Lorsque l'enzyme incorpore un nouveau nucléotide, elle va libérer le marqueur fluorescent et générer un signal lumineux. Ce signal sera détecté dans une zone restreinte au fond du puits par le séquenceur (<http://www.napoleome.ch/fr/projet/le-sequençage/methodes/>)

E – 3. Séquençage par nanopore

Le séquençage Nanopore est une méthode en cours de développement depuis 1995 pour le séquençage d'ADN. Un nanopore se définit, en général, comme une interface entre deux solutions séparées par une membrane. Les nanopores sont de minuscules ouvertures de quelques nanomètres de diamètre. En biologie, Ils existent naturellement en tant que protéines membranaires appelées porines. De l'extérieur vers l'intérieur, les porines vont permettre l'entrée des nutriments dans la cellule. De l'intérieur vers l'extérieur, les porines assurent la détoxification du cytoplasme.

Ce sont des protéines pouvant avoir un diamètre interne de 1 nm, environ 100 000 fois plus petit que celui d'un cheveu humain. Ce diamètre est à la même échelle que de nombreuses molécules, y compris l'ADN.

L' α -hémolysine, le nanopore le plus utilisé, est une porine biologique. C'est une exotoxine de la bactérie *Staphylococcus aureus*. Elle se présente sous la forme d'un heptamère à géométrie cylindrique s'insérant dans une bicouche lipidique (cf. figure 18⁷⁵). A l'état naturel, cette protéine est produite par la bactérie dans le but de provoquer l'hémolyse des globules rouges de l'hôte pour les vider des nutriments dont le microorganisme a besoin. En laboratoire, l' α -hémolysine est utilisée dans la technique du patch-clamp et dans les neurosciences pour l'étude des neurotransmetteurs. Sa taille, plus grande que la plupart des canaux ioniques existants, permet le passage de molécules plus volumineuses.

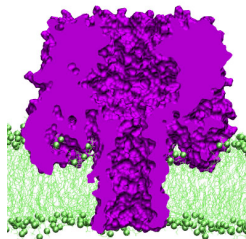


Figure 22 : Vue en coupe d'un pore d' α -hémolysine (en violet) inséré dans une bicouche lipidique (en vert).

Au cours des dernières dizaines d'années, les progrès scientifiques ont permis l'étude du signal électrique fourni par ces nanopores. De nombreuses études se servent de ce signal essentiellement avec deux objectifs. Tout d'abord, il est intéressant de comprendre les mécanismes de transport des ions porteurs de ce signal. Ensuite, ce signal peut être utilisé pour réaliser une détection d'objets plus grands que les ions.

Il existe plusieurs types de nanopores. Ils sont divisés en deux grandes catégories : les nanopores biologiques et les nanopores synthétiques.

Les nanopores biologiques ou la translocation de macromolécules

Le terme de translocation, désigne le passage séquentiel d'une molécule linéaire au travers d'un nanopore, situé au sein d'une membrane séparant deux milieux. En biologie, les macromolécules linéaires présentes dans les cellules sont l'ADN, l'ARN et les protéines. Dans la cellule eucaryote, la translocation de l'ADN n'est pas un processus naturel, car la molécule d'ADN reste à l'intérieur du noyau. En revanche, certains virus lorsqu'ils infectent des cellules cibles, y injectent un brin de leur ADN. Ce processus se fait par translocation. L'ARN messager, résultat de la transcription de l'ADN dans le noyau permet de communiquer la séquence génétique au réticulum endoplasmique, situé dans le cytoplasme, pour y permettre la synthèse des protéines. Il subit donc une translocation à travers la membrane nucléaire. Les protéines, quant à elles, peuvent subir de nombreuses translocations avant d'acquérir leur structure tertiaire. En effet, leur site de destination peut être très éloigné du réticulum endoplasmique, là où la plupart sont synthétisées.

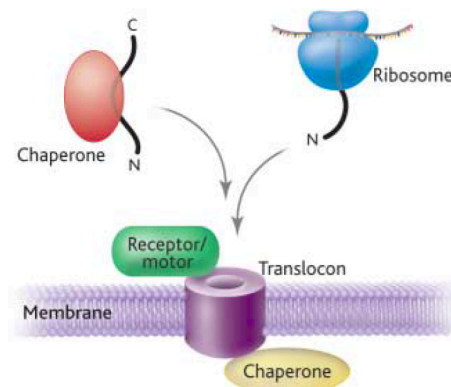


Figure 23 - Modèle général de la translocation d'une protéine dans une cellule⁷⁶.

Le nanopore naturel tel que l' α -hémolysine se révèle particulièrement adapté à la translocation, puisque le diamètre de la partie transmembranaire est d'environ 16 Å⁷⁷. Les premiers résultats de translocation à travers l' α -hémolysine ont été obtenus avec des polymères en solution.

En 1992, l'équipe de Krasilnikov a utilisé du polyéthylène glycol (PEG) pour déterminer la taille interne du pore⁷⁸. Pour la première fois de l'histoire, en 1996, l'équipe de Kasianowicz a réalisé la translocation de molécules d'ADN, dans un pore d' α -hémolysine⁷⁹. Un nanopore unique est d'abord préparé sur une membrane lipidique. Un courant est appliqué et un flux ionique est enregistré. Ensuite, l'ajout de molécules d'ADN ou d'ARN provoque des perturbations des signaux enregistrés dû à l'obstruction plus ou moins partielle du nanopore par le passage des différentes bases (Figure 20). Cette étude a eu deux résultats principaux :

- La durée de translocation d'un brin d'ADN est proportionnelle au nombre de ses nucléotides.
- Seul l'ADN simple brin, peut passer à travers l' α -hémolysine.

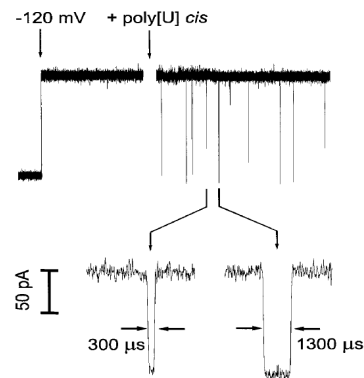


Figure 24 : Observation de blocages de courant lors du passage de brins homopolymères d'ARN⁷⁶.

Grâce à de multiples améliorations, la méthode a pu donner une détection directe et rapide de la séquence de bases de molécules uniques d'ADN ou d'ARN en observant les variations de l'intensité électrique dans le temps et en déterminant une valeur différente pour chaque base de l'acide nucléique (Figure 25).

La découverte importante de Kasianowicz a été la première étape de nombreux efforts pour la course au séquençage par translocation. Les résultats obtenus se sont révélés très prometteurs. L'équipe de Kasianowicz, en 1999, a utilisé différents brins d'ARN et est parvenue à faire la distinction entre des brins qui ne diffèrent que par leur séquence⁸⁰. Ainsi, l'acide polycytidylique, l'acide polyadénylique, et l'acide polyuridylique peuvent être identifiés, en tant qu'homopolymères ou même en tant que segments d'une unique molécule d'ARN. Cette distinction nécessite un échantillonnage du signal de l'ordre de la microseconde.

En 2000, J. Golovchenko et son équipe ont tenté aussi de différencier les bases du génome. Ils sont parvenus à faire la distinction d'homopolymères d'ADN d'une centaine de monomères, ainsi que des hétéropolymères séquencés d'une cinquantaine de bases⁸¹.

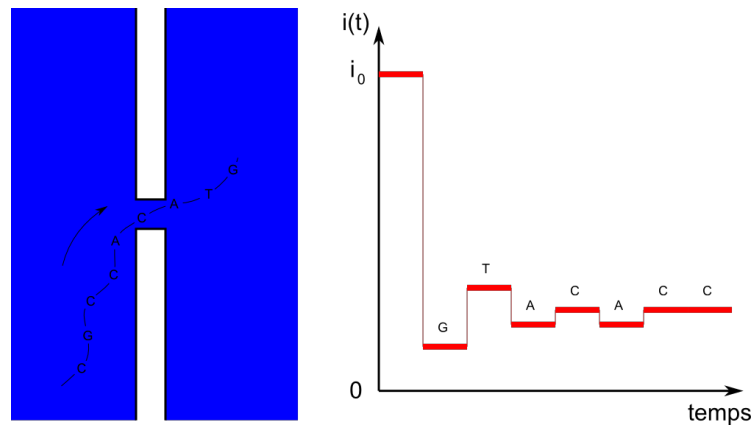


Figure 25 : Théorie du séquençage par translocation.

(C. TASSERIT, 2011)

Avantages et limitations des nanopores naturels

La principale qualité de l' α -hémolysine est la reproductibilité des expériences. Chaque monomère qui la constitue est identique, et par conséquent chaque heptamère constitué est assemblé de la même façon. Pour une concentration en KCl de 1 M, la conductance du pore est de 1nS^{82} , et cette valeur reste toujours la même. De plus, comme ce pore est naturel, il est possible d'intervenir artificiellement sur le code génétique, afin de lui ajouter de nouvelles fonctionnalités. Ceci a été fait notamment par l'équipe de Braha en 1997⁸³.

Les modifications peuvent inclure :

- Des modifications génétiques pour produire ou améliorer des sites de liaison spécifiques dans la cavité du pore.
- La fixation chimique ou génétique de sondes d'affinité ou de ligands, tels que : Oligos ADN, aptamères, biotine, his-tags, sucres, etc.
- La fixation de molécules plus grandes à l'extérieur du pore, comme des protéines motrices ou des sites récepteurs.
- L'entrée des pores peut également être modifiée pour améliorer la liaison des moteurs moléculaires, tels que les hélicases utilisées dans la chimie de séquençage de l'ADN.

Le principal inconvénient des nanopores naturels est leur faible longévité. En effet, ils sont insérés dans une membrane bilipidique synthétique. Cette dernière est très fragile et peut éclater à la moindre perturbation. Elle est sensible à la température, au pH et à la force ionique. De plus, sa durée de vie est de l'ordre de l'heure.

Les nanopores artificiels

Pour combler les insuffisances des pores biologiques, et pour ouvrir de nouvelles possibilités dans le monde de la translocation, d'autres types de nanopores, dits artificiels, ont été mis au point. En voici une brève description, inspirée par les revues de Bayley et Martin de 2000⁸⁴, et celle de Howorka et Siwy de 2009⁸⁵ :

- *Les nanopipettes de verre* : C'est avec des nanopipettes que Neher et Sakmann ont réussi à isoler des canaux biologiques. Elles sont obtenues par échauffement laser et étirement d'une pipette millimétrique. L'équipe de Karhanek a réussi à observer des pics de résistivité correspondant aux passages de brins d'ADN liés à des nanoparticules d'or⁸⁶. On peut obtenir des tailles de pores aussi faibles qu'environ 25 nm de rayon.

- *Les nanopores sur support de silicium* : C'est en 2001 qu'est créé le premier nanopore sur support solide, par l'équipe de Golovchenko⁸⁷. Ils nomment la technique « ion-beam sculpting », c'est-à-dire sculpture par faisceau d'ions. Le matériau utilisé est le nitrure de silicium. Des pores très petits peuvent être obtenus, avec des rayons inférieurs à 2 nm⁸⁸.

- *Les nanopores à trace attaquée* : Il est réalisé à partir d'un film polymère. Après l'irradiation par un ou plusieurs ions lourds, on réalise une attaque chimique qui perce le pore. Cette méthode fournit des pores également très petits, pouvant atteindre un rayon inférieur au nanomètre⁸⁹.

- *Les nanotubes de carbone et les pores de graphène* : On a pu observer des signaux correspondant au passage de brins d'ADN dans des nanotubes de carbone à simple épaisseur⁹⁰, ainsi que dans un nanopore percé dans un plan de graphène⁹¹.

F - Technologie Oxford Nanopore

Oxford Nanopore (ONT) utilise des nanopores biologiques. Pour détecter la molécule, le nanopore est placé dans une membrane électriquement résistante de sorte qu'un courant ionique puisse passer à travers le nanopore lorsqu'une tension est appliquée de part et d'autre de la membrane. La perturbation du courant se produit lorsque la molécule et le nanopore entrent en contact les uns avec les autres, et cette perturbation est mesurée.

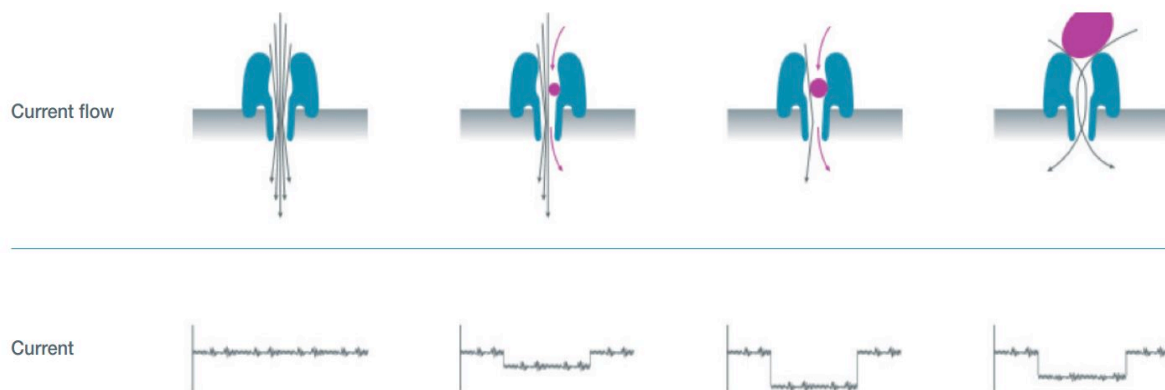


Figure 26 : Schémas de détection de macromolécules par le système d'Oxford Nanopore

Le pore actuel (version R9) est un mutant de la lipoprotéine CsgG d'*Escherichia coli*, qui, en condition naturelle, aide à transporter des polypeptides à travers la membrane bactérienne. La lipoprotéine CsgG est constituée de neuf sous-unités. Le pore a été conçu pour permettre à l'ADN d'être « transloqué » à travers la structure.

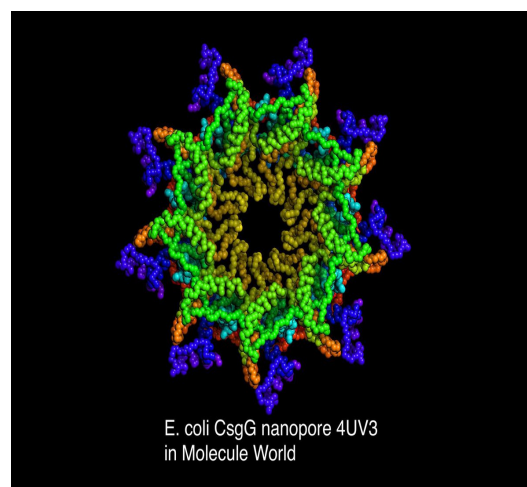


Figure 27 : Structure de la lipoprotéine CsgG

Pour obtenir ces pores, les cellules bactériennes sont broyées et l'extrait résultant purifié pour générer une solution de nanopores à haute concentration. Un traitement supplémentaire est ensuite effectué pour assurer la consistance du matériau. La société a également développé un système électronique qui permet de réaliser des expériences de détection de signaux générés par de multiples nanopores, de collecter et d'analyser les données en temps réel. Cette structure est appelée « Flow Cell » ou cellule de flux.

La cellule de flux est un support jetable qui fournit l'interface fluidique entre le capteur et les électrodes, permettant l'analyse de l'échantillon. Elle comporte un réseau de capteurs dont chacun a été conçu pour enregistrer les données émanant d'un seul nanopore. C'est une cellule de 92,4 mm x 28,0 mm x 5,8 mm qui s'insère dans les appareils MinION ou GridION.

Les cellules de flux sont livrées avec un tampon jaune de stockage pour maintenir un équilibre osmotique et qui couvre le réseau de capteurs ou « Sensor array ». Le tampon de stockage contient du sel et le couple redox standard qui permet au courant de traverser le nanopore. Il contient également une molécule d'ADN QC qui émet un signal spécifique permettant d'identifier les pores fonctionnels pendant le test de QC de la plateforme. La cellule de flux a deux ports :

- Le port d'amorçage ou « Priming port » : il permet de préparer la cellule de flux pour le séquençage grâce à un mélange spécifique d'amorçage qui va rincer la cellule de flux et déplace le tampon de stockage vers le réservoir de déchets (Waste port).
- Le port de dépôt ou « SpotON sample » qui permet de charger la librairie dans la cellule de flux.

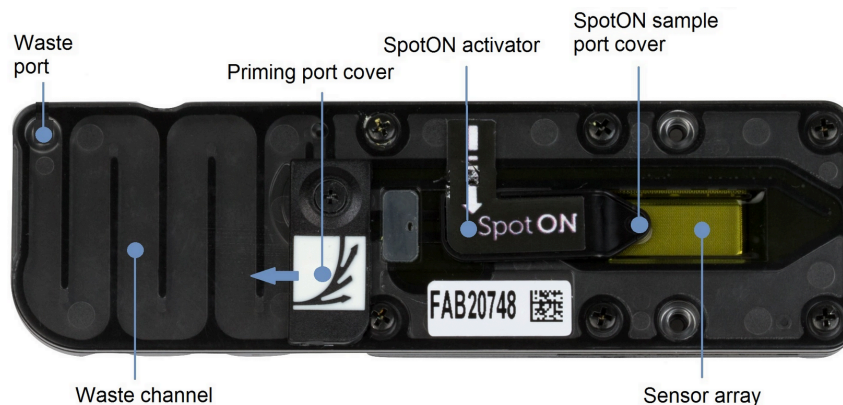


Figure 28 : cellule de flux ou « flow cell » d'Oxford Nanopore »⁹²

La cellule de flux comporte 512 canaux contenant chacun quatre nanopores. Les 2048 pores sont reliés chacun à une électrode. La modification du courant ionique (le signal nanopore) est détectée par l'électrode. Chaque électrode est connectée à un canal individuel sur une puce ASIC (Application Specific Intergrated Circuit), qui contrôle et mesure le signal nanopore. L'ASIC est capable d'enregistrer, simultanément, les signaux émis par tous les canaux reliés aux 2048 électrodes des puits actifs de la cellule de flux. Le choix d'un seul puits pour chacun des canaux d'enregistrement est appelé multiplexage. Le multiplexage du réseau de nanopores est utilisé pour améliorer le rendement des canaux contenant un seul pore.

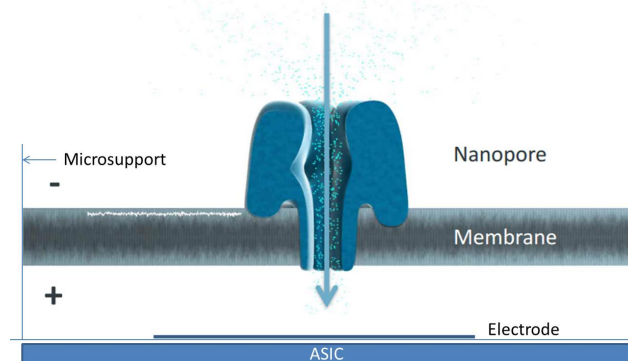


Figure 29 : Schémas d'un capteur nanopore⁸⁶.

Cette unité complète constituée de la membrane, du nanopore, de l'électrode et du canal ASIC est connue sous le nom de capteur nanopore. Le passage du brin d'ADN à travers le pore produit une série de mesures de courant dont chacune est unique et peut être décodée pour produire la séquence. Ces mesures sont converties en lectures par un logiciel « MinKNOW » qui contrôle le capteur nanopore, collecte et traite le signal. La conversion des lectures en séquence peut se faire en temps réel grâce à un logiciel « METRICHOR » disponible sur le « Cloud » ou en téléchargement local.

La technologie Nanopore offre la possibilité d'analyser la molécule d'ADN native sans aucune amplification, évitant ainsi l'introduction de biais et la perte d'informations précieuses, telles que les modifications de base et les sites de méthylation. Les nanopores séquentent des fragments très long d'ADN et permettent d'obtenir des lectures dont la longueur peut dépasser les 100 kb, ouvrant la possibilité de séquencer, par exemple, des sections répétitives d'ADN ou d'ARN. Les données de chaque lecture sont enregistrées instantanément dans un fichier dès qu'un brin a terminé son passage à travers le nanopore, et sont disponibles pour l'analyse en temps réel. Comme chacun des nanopores agit indépendamment dans le réseau de capteurs, dès qu'un brin a fini de traverser un pore, le pore devient disponible pour un autre brin à analyser. La vitesse de translocation est contrôlée par

l'inclusion d'une protéine motrice qui agit comme un cliquet qui déplace le brin d'ADN à travers le nanopore. Le séquençage continue jusqu'à ce qu'il n'y ait plus de brins, ou jusqu'à ce que suffisamment d'informations aient été collectées et que l'expérience de séquençage puisse être arrêtée.

La protéine Motrice est fournie sur un adaptateur leader qui est attaché à l'extrémité de la matrice d'ADN double brin. Cette protéine permet l'ouverture du double brin à l'entrée du pore pour faciliter le passage de l'ADN à travers le nanopore. Le premier brin séquencé est connu sous le nom de modèle. Habituellement, le brin complémentaire se dissocie, et peut être capturé par le nanopore à un moment ultérieur. La protéine motrice déplace le brin une base à la fois, contrôlant ainsi la vitesse à laquelle le brin passe le lecteur qui est abrité dans la partie étroite du nanopore et qui détecte les perturbations caractéristiques du courant électrique créées par le passage des nucléotides qui le traversent.

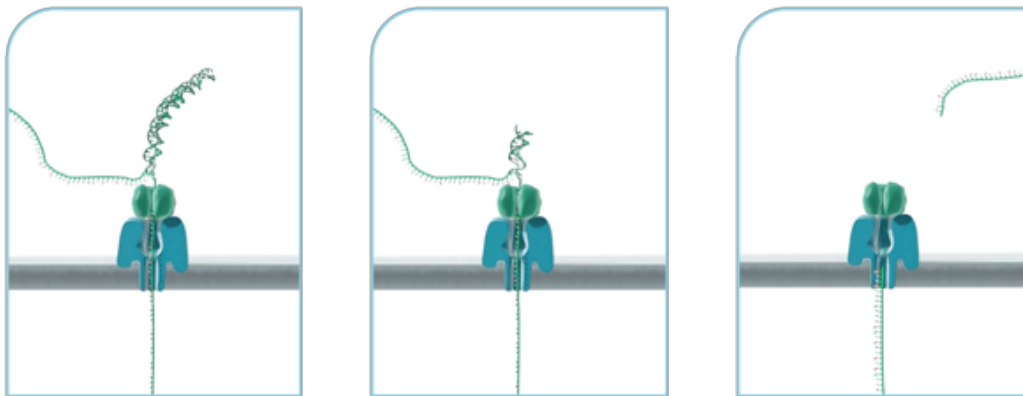


Figure 30 : Schémas du passage de la molécule d'ADN ou d'ARN, munis d'un adaptateur (en vert), à travers le nanopore d'Oxford technologies⁸⁶.

Afin d'optimiser la quantité d'ADN lue par un pore tout en maintenant des besoins faibles en matrice, Oxford Nanopore Technologies a développé une méthode de maintien des molécules d'ADN à la surface de la membrane. Ce mécanisme concentre le matériau d'intérêt à proximité du pore, augmentant ainsi la sensibilité du dispositif et le débit de séquençage. Il s'agit de billes de chargement qu'on ajoute au mix de pré-séquençage juste avant le dépôt sur la cellule de flux.

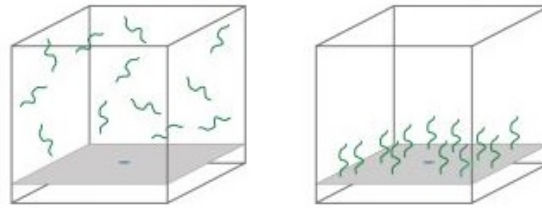


Figure 31 : Schémas de l'attache de l'ADN à la surface de la membrane par le système de billes de chargement⁸⁶.

Les mesures de détection d'Oxford Nanopore sont basées sur une approche électronique ce qui offre à cette technologie la possibilité d'être mise à l'échelle pour répondre aux exigences expérimentales. Aujourd'hui, la technologie a été miniaturisée dans un appareil portable ainsi que dans des instruments de paillasse, qui permettent d'ajuster le débit et le nombre d'échantillons en fonction des besoins. La portabilité de la technologie ouvre également une gamme d'applications dans de nouveaux environnements en dehors du laboratoire traditionnel.

Les principaux produits d'ONT sont :

- MinION : c'est un appareil de séquençage par nanopore portable. Il a été commercialisé en mai 2015, après avoir été lancé initialement dans le cadre d'un programme d'accès anticipé, le Programme d'accès au MinION (MAP) commencé en 2014. Les publications de ce programme décrivent son utilisation dans l'identification rapide des agents pathogènes viraux, la surveillance de l'environnement, de la salubrité des aliments et de la résistance aux antibiotiques ainsi que d'autres applications. Les publications indiquent un taux de lecture de 90 nucléotides par seconde et par nanopore avec un taux d'erreur de 30% au cours de la première phase de sa diffusion autour de 2014. Avec la dernière version R9 de la cellule de flux en 2016, les taux d'erreurs brutes ont été réduits entre 2 et 13% en fonction du type de séquençage de l'ADN (1D, permettant le séquençage simple brin ou 1D², permettant un séquençage double brin de l'ADN). En octobre 2016, la version R9.4 a été diffusée et elle permet de lire 450 bases par seconde et par nanopore produisant 10 Gb de données par cellule de flux.

- GridION X5 : c'est un système de séquençage compact qui a été commercialisé en mars 2017. Il permet de réaliser jusqu'à cinq expériences de séquençage simultanément. L'appareil comporte cinq plateformes de séquençage où des cellules de flux MinION peuvent être connectées et permet de générer jusqu'à 100 Go de données par cycle⁹³.

- PromethION : c'est un instrument autonome pour les analyses à haut débit avec un grand nombre d'échantillons. Il comporte 48 cellules de flux. Chaque cellule permet de séquencer simultanément jusqu'à 3000 nanopores, et produire jusqu'à 11 Tb de données en 48 heures. Ce périphérique est disponible via un programme d'accès (PEAP : PromethION Early Access Programme) qui a été ouvert pour l'enregistrement en juillet 2015.

- VolTRAX : cet appareil permet de gérer la préparation de la librairie pour le séquençage de façon automatisée, permettant d'affranchir les utilisateurs de toute compétence particulière de laboratoire pour la réalisation du séquençage. Cet appareil est également disponible via un programme d'accès anticipé (VIP : VolTRAX Introduction Programme) qui a été ouverte en octobre 2016.

- SmidgION : Il s'agit d'un séquenceur capable d'être raccordé à un téléphone portable. Il a été annoncé en mai 2016 et est toujours en cours de développement.

Tous ces produits sont destinés à être utilisés pour l'analyse de l'ADN, l'ARN, des protéines et des petites molécules avec une gamme d'applications pour la médecine personnalisée, la science des cultures et la recherche scientifique.

Hypothèse de travail

Notre hypothèse de travail est que la nouvelle technologie de séquençage par Nanopore technologie, permet de détecter de façon adéquate et moins coûteuse que la technologie de référence Illumina, des anomalies dysbiotiques du microbiote chez la personne obèse, éventuellement soumise à des interventions diététiques ou chirurgicales.

Objectif de notre étude

Notre étude vise donc à caractériser par une nouvelle méthode (la technologie Nanopore), les anomalies du microbiote intestinal chez des patients à différents stade d'obésité et lors d'interventions visant à perdre du poids. Ce développement technologique permettrait dans le futur d'établir des moyens diagnostics de la dysbiose chez ces patients.

En effet, notre équipe ainsi que d'autres équipes ont montré que les proportions des différentes grandes lignées bactériennes n'étaient pas les mêmes chez les sujets minces et chez les sujets obèses. Parmi les autres modifications du microbiote en lien avec cette altération de la balance des groupes bactériens, nous avons également montré chez des sujets obèses, que la richesse bactérienne, évaluée par le comptage du nombre de gènes, était réduite. Cette diminution a été associée à des phénotypes métaboliques plus marqués comparativement à des sujets obèses dont la diversité bactérienne est préservée. Parmi ces phénotypes : une insulino-résistance et un taux de triglycérides sanguins plus élevés. Dans une étude menée en 2013 par notre équipe, 49 patients obèses et en surpoids ont été soumis à un régime réduit en calories pendant 6 semaines (pauvre en graisses mais enrichi en fibres, protéines et glucides à faible index glycémique), suivi de 6 semaines de stabilisation⁹⁴. Sur l'ensemble des sujets, une perte de poids moyenne de 5% du poids initial a été constatée, ainsi qu'une amélioration des paramètres métaboliques et inflammatoires comme attendu lors de ce type d'intervention. Les sujets qui présentaient initialement une faible diversité du microbiote ont vu leur diversité augmenter. Néanmoins, l'intervention a été moins efficace sur le plan clinique chez ces sujets avec une perte de diversité (moindre diminution du taux de triglycérides, moindre amélioration de l'inflammation bas-grade). Dans une étude plus récente, nous avons retrouvé une perte de richesse du microbiote encore plus marquée (supérieures en moyenne à 70%) chez les patients obèses de la cohorte « Microbaria » constituée de patients sévèrement obèses⁹⁵ et candidats à la chirurgie bariatrique. La richesse bactérienne du microbiote intestinal de ces patients a été évaluée un an après la chirurgie, période durant laquelle les comorbidités liées à l'obésité sont améliorées. On a pu

constater, que chez la plupart des patients avec une diversité bactérienne faible, même si celle-ci augmente après la chirurgie, un maintien d'une faible diversité bactérienne ainsi qu'une amélioration partielle de leurs comorbidités. En d'autres termes, nous pensons que la persistance d'une dysbiose chez ces patients pourraient être en cause dans ces améliorations partielles post chirurgie ou lors d'interventions nutritionnelles.

L'ensemble des signes cliniques liés à l'obésité pourrait être corrigé, ou encore mieux prévenu, par la détection précoce de l'altération du microbiote et grâce à des recommandations nutritionnelles adaptées. Ainsi l'établissement de tests de dépistage de la dysbiose pourrait être intéressant chez ces sujets afin d'analyser la situation avant une intervention chirurgicale.

Notre objectif est de mettre en place un test diagnostique basé sur le séquençage de la flore intestinale bactérienne de nos patients afin de déterminer l'état de leur microbiote intestinal. Ainsi, nous devons mettre en place cette activité au sein de notre laboratoire, partant du recueil des selles au séquençage, en passant par l'extraction de l'ADN bactérien. Nous avons souhaité avoir une technologie qui nous permettrait de générer des résultats rapidement, qui soit facile d'utilisation, dont le coût ne serait pas trop élevé et qui soit fiable, afin de diagnostiquer rapidement les sujets avec un appauvrissement du microbiote et des anomalies des grands équilibres bactériens. Nous nous sommes donc intéressés à la technologie d'Oxford Nanopore qui répond aux trois premiers critères, la fiabilité des résultats faisant l'objet d'une validation par comparaison avec les technologies « Solid » ou « Illumina ».

Démarche expérimentale

- Mettre en place la procédure de prélèvement de selles dans les différentes cohortes prévues pour cette étude et pour de futures approches d'investigation clinique en vie réelle.
- Mettre au point une technique d'extraction de l'ADN bactérien à partir des selles humaines qui permet d'obtenir une quantité et qualité convenable pour le séquençage.
- Mettre en place l'extraction de l'ADN bactérien à partir de selles collectées auprès de patients obèses et de sujets en surpoids.
- Mettre en place le séquençage bactérien, en utilisant l'outil « MinION » d'Oxford Nanopore Technologies. Cette expérimentation permettra de mettre en évidence la diversité bactérienne ainsi que la présence ou l'absence de certains groupes bactériens d'intérêt pour chaque sujet.
- Effectuer une « validation » du séquençage « MinION » sur deux cohortes de patients obèses déjà séquencées dans le cadre de deux précédentes études dans notre équipe. L'une séquencée par la technologie « Illumina » et l'autre par la technologie « Solid ».

Matériels et méthodes

Matériels et méthodes

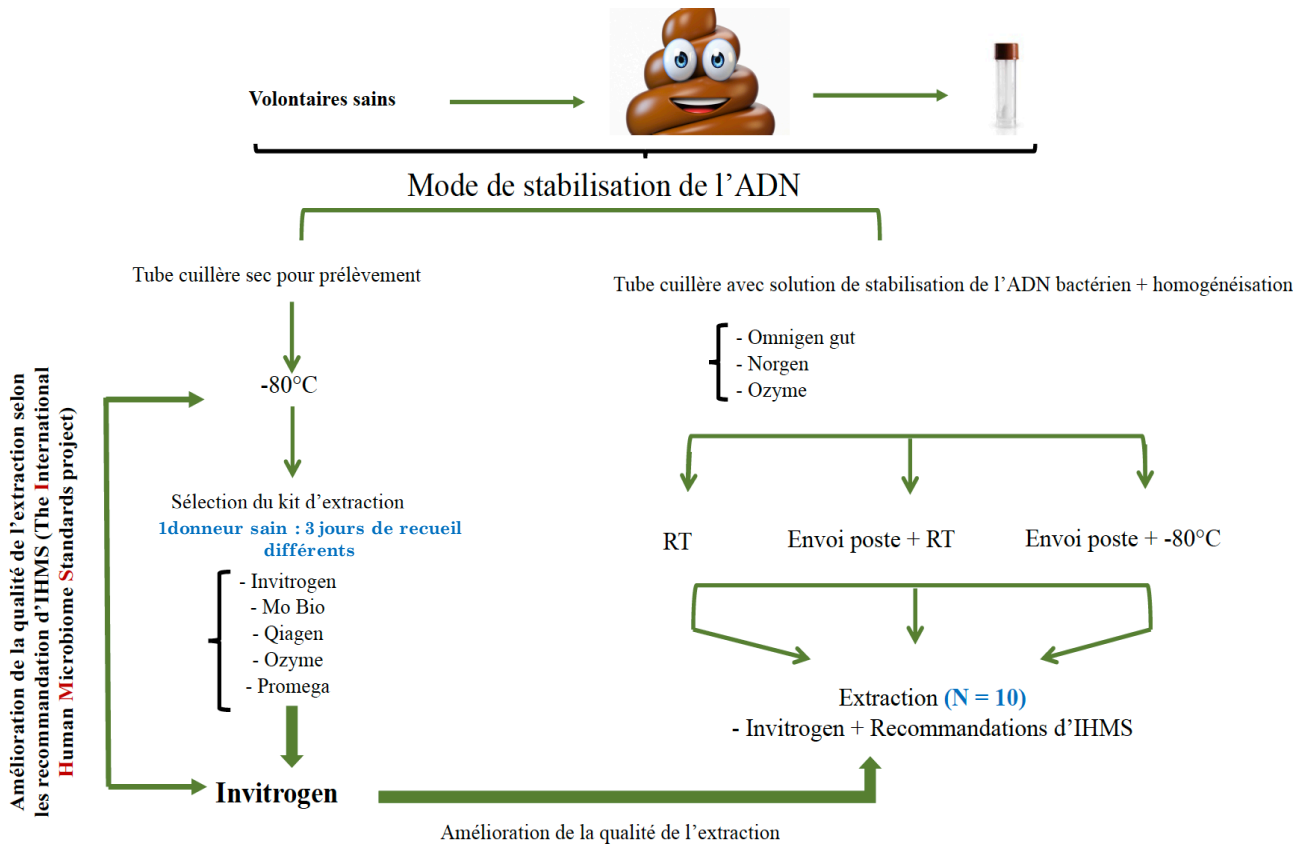


Figure 32 : Schéma général résumant les différentes mises au point, de la procédure de recueil des selles au protocole d'extraction de l'ADN bactérien.

1 - Kit de prélèvement et extraction de l'ADN bactérien.

A) Prélèvements et conditionnement des selles

Les échantillons de selles utilisés pour cette mise au point ont été recueillis auprès de dix volontaires sains non obèses, quatre hommes et six femmes avec un IMC compris entre 18 et 25 kg/m².

Les échantillons prélevés ont été procédés et conservés selon les expérimentations envisagées :

- Prélèvement à domicile de selles fraîches dans un tube cuillère sec (SARSTED) avec conservation des selles à -20°C. Le prélèvement a été ramené au laboratoire dans un sac isotherme et congelé immédiatement à -80°C.

- Prélèvement au laboratoire dans deux tubes contenant une solution pour la stabilisation de l'ADN bactérien. Un tube a été remis au laboratoire juste après le recueil et conservé à température ambiante. L'autre tube a été envoyé par la poste afin de tester l'impact des conditions de ce mode de transport sur la qualité et l'intégrité de l'ADN bactérien. A réception de ce prélèvement, un aliquot a été prélevé et conservé à -80°C afin de pouvoir tester l'impact de cette condition supplémentaire sur la qualité de l'extraction.

Trois fournisseurs différents ont été testés :

- **Ozyme** : « *DNA/RNA Shield - Fecal Collection Tube* » (produit ZYMO commercialisé par Ozyme) : permet de collecter 1g de selle, de le transporter et le conserver à température ambiante jusqu'à deux ans et indéfiniment à -80°C.
- **NORGEN BIOTEK** : « *Stool Nucleic Acid Collection and Preservation Tubes* » : permet de collecter 2 g de selle de les transporter et les conserver à température ambiante jusqu'à deux ans. La conservation à -80°C pour un stockage de longue durée est également possible.
- **DNA GENOTEK** : « *OMNIGEN Gut for Microbiome* » : permet de prélever 500 mg de selles et de les conserver à température ambiante pendant 60 jours. Le stockage à long terme peut se faire à -80°C.

B) Extraction de l'ADN bactérien

L'extraction de l'ADN a été réalisée avec différents kits commerciaux à tester :

- INVITROGEN : PureLink™ Microbiome DNA Purification Kit.
- QIAGEN : Qiamp PowerFecal DNA Kit.
- OZYME : ZimoBiomics DNA Mini Kit.
- MO BIO : Power Microbiome RNA/DNA isolation kit.
- PROMEGA : Chez ce fournisseur, j'ai testé le « MAXWELL Instrument » une station robotisée développée par Promega et destinée à automatiser l'extraction de l'ADN. Pour notre type de prélèvement, à savoir les selles humaines, le passage par Le Maxwell est précédé par une phase de prétraitement des échantillons qui consiste à réaliser une lyse chimique suivit d'une lyse mécanique de la paroi des bactéries. S'ensuit une centrifugation et récupération du surnageant qui contient l'ADN bactérien. Les échantillons sont ensuite chargés dans la station pour la suite de l'extraction. J'ai testé deux kits différents : « Maxwell RSC Buffy Coat DNA Kit » et « Maxwell RSC PureFood GMO and Authentication Kit ».

Le rendement des extractions a été estimé par un fluoromètre, le Qubit (Life technologies) et la qualité a été évaluée par le Nanodrop (Thermo Scientific) et par Tape station (Agilent).

2 - Préparation de la librairie de séquençage

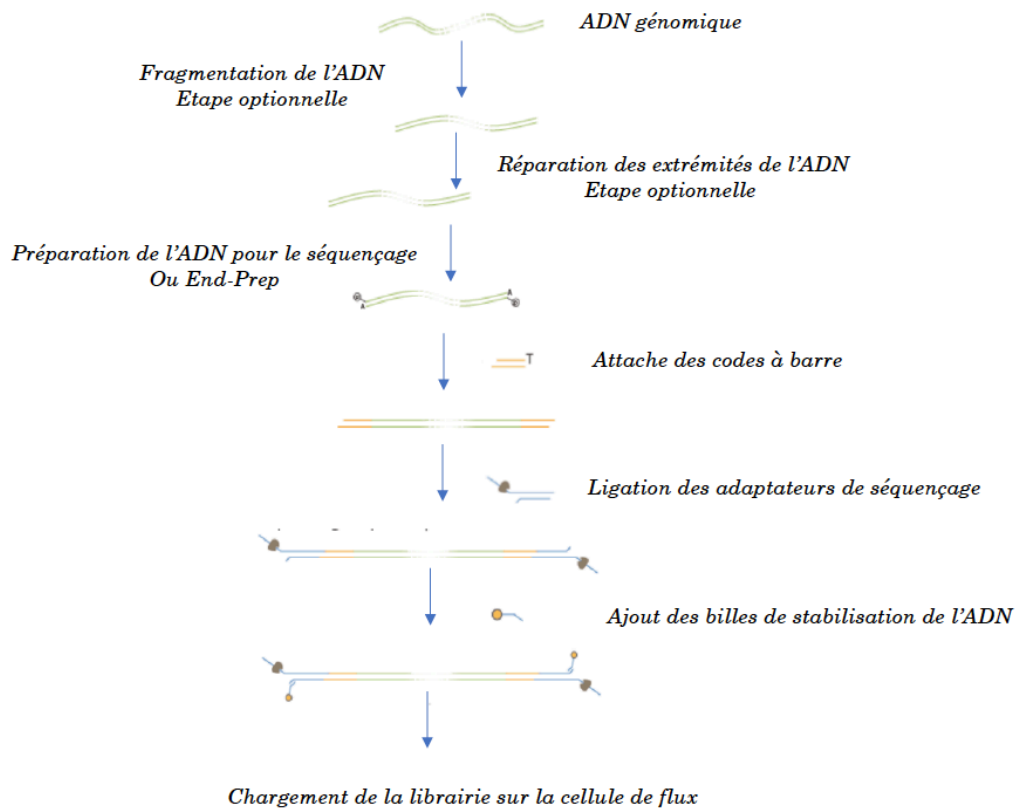


Figure 33 : Schéma général du protocole de préparation de la librairie d'ONT.

La préparation de la librairie de séquençage qui représente une collection de fragments d'ADN représentative de l'ADN à séquencer, a été réalisée selon le protocole proposé par Oxford Nanopore technologies. Ce protocole propose en option une première phase de fragmentation de l'ADN. Pour cela on utilise des g-tube de chez Covaris. C'est un dispositif à usage unique qui coupe l'ADN génomique en fragments sélectionnés de 6kb à 20kb. g-TUBE utilise la force centrifuge (le "g" dans g-TUBE) pour pousser l'échantillon d'ADN à travers un orifice fabriqué avec précision. Cela produit des forces de cisaillement dans l'échantillon qui fragmentent l'ADN. La taille des fragments est sélectionnée en ajustant la vitesse de centrifugation qui modifie le débit de l'ADN à travers l'orifice et donc les forces de cisaillement. Une centrifugation plus élevée produira des fragments plus courts. Le protocole proposé par ONT permet de générer des fragments de **8kb**. Pour cela 1,5µg d'ADN contenu dans 46 µl d'eau DNase free sont placés dans le g-tube Covaris. Le tube est ensuite centrifugé à la vitesse de 6000rpm pendant une minute. Cette première étape permet un premier passage de l'ADN à travers l'orifice. Le tube est ensuite inversé et replacé, à nouveau, dans la centrifugeuse pour une deuxième centrifugation d'une minute à 6000rpm. Ce deuxième passage permet de récupérer l'ADN fragmenté.

La fragmentation de l'ADN est suivie par une étape de réparation des extrémités de l'ADN endommagées par la fragmentation pour en faire des bouts « blunt ». Pour cela nous utilisons le produit « NEBNext FFPE Repair Mix » de chez New England Biolabs (NEB). C'est un mix qui contient des enzymes et des tampons optimisés pour la réparation de l'ADN en vue du séquençage.

L'étape de « End prep » permet de préparer l'ADN au séquençage. On utilise le produit « NEBNext Ultra II End Repair/dA-Tailing Module » de chez NEB qui permet également une réparation de l'ADN et sa préparation avec une phosphorylation en 5' et l'ajout d'une adénine en 3'.

Dans le cadre de cette étude, j'ai séquencé 10 à 12 ADN différents en même temps et sur la même cellule de flux. Pour cela et après l'étape de « End prep », chaque ADN a été marqué par un code à barre unique en utilisant le kit « 1D Native barcoding genomic DNA » d'ONT et kit « NEB Blunt/TA Ligase Master Mix » de chez NEB. A la fin de cette étape les ADN sont quantifiés et regroupés dans un volume de 50µl de sorte que la quantité totale en ADN **soit de 700ng**.

L'étape suivante est l'attache des adaptateurs de séquençage aux ADN préparés, marqués et regroupés. Les adaptateurs contiennent la protéine motrice qui va guider les brins d'ADN vers leur passage à travers le nanopore. Pour cela j'utilise « NEBNext Quick Ligation Module » de chez NEB. Cette préparation permet d'obtenir le mix de pré-séquençage.

La dernière étape est la préparation de la librairie. Au mix de pré-séquençage, on ajoute les billes qui vont permettre d'alourdir les brins d'ADN et d'assurer leur maintien à proximité des pores de séquençage au niveau de la cellule de flux. Pour cela j'utilise le kit « 1D Genomic DNA by ligation » de chez ONT.

Chacune des étapes de la « End repair » à la préparation du mix de pré-séquençage est suivit d'une étape de purification de l'ADN traité sur un système de billes, « Agentcourt AMPure XP », de chez Beckman Coulter. Ces billes répondent à la technologie SPRI (Solid Phase Reversible Immobilisation beads) développées à l'Institut Whitehead par De Angelis et al en 1995 pour la purification de colonies amplifiées par PCR et dont l'objet était le séquençage de leurs ADN. Les billes de SPRI sont paramagnétiques (magnétiques uniquement dans un champ magnétique), ce qui les empêche de s'agglomérer et de sédimenter dans la solution. Chaque bille est faite de polystyrène entouré d'une couche de magnétite, qui est revêtue de molécules carboxyle (figure 3). Ce sont eux qui lient de manière réversible l'ADN en présence du polyéthylène glycol (PEG) et du sel (20% de PEG, 2,5 M de NaCl). Le PEG provoque la liaison de l'ADN chargé négativement avec les groupes carboxyle de

la surface des billes. Comme l'immobilisation dépend de la concentration de PEG et de sel dans la réaction, le rapport volumétrique des billes et de l'ADN est critique.

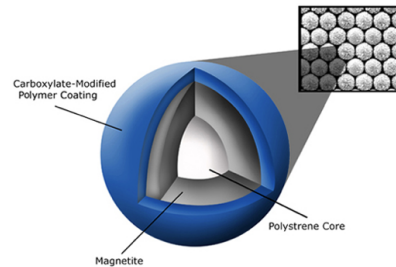


Figure 34 : structure d'une bille avec la technologie SPRI.
([Http://core-genomics.blogspot.com/2012/04/how-do-spri-beads-work.htm](http://core-genomics.blogspot.com/2012/04/how-do-spri-beads-work.htm))

3 - Analyse bioinformatique

Plusieurs indicateurs peuvent être mesurés par l'analyse informatique. Ainsi, la diversité peut être évaluée à plusieurs niveaux emboîtés nommé alpha, bêta et gamma :

- **La diversité α** est la diversité locale, mesurée à l'intérieur d'un système délimité. Plus précisément, il s'agit de la diversité dans un habitat uniforme de taille fixe. Cette diversité peut être quantifiée de différentes manières. Les deux principaux facteurs pris en compte lors de la mesure de la diversité sont **la richesse et la régularité**. La richesse est une mesure du nombre des différents types d'organismes présents dans un écosystème. Par exemple, la richesse en espèces est le nombre d'espèces différentes présentes dans l'écosystème. Cependant, dans l'estimation de la diversité biologique on doit prendre en compte aussi bien la richesse que la régularité. La régularité prend en compte dans le calcul de la diversité, non seulement **la présence des espèces mais aussi leurs abondances**. Plusieurs index ont été développés dans le champ de l'écologie pour estimer la richesse et la régularité.

Pour la richesse on trouve :

* **L'index de Chao (1984)** estime le nombre d'espèces non observées à partir de celles observées une ou deux fois. Il s'agit d'un estimateur minimum, valide à condition que les singletons et doubletons représentent une part importante de l'information.

* **L'index ACE** : Chao et Lee (1992) développent l'estimateur ACE (Abundance-based coverage estimator) à travers l'estimation de la proportion des espèces découvertes.

Comme estimateurs de la régularité, on distingue :

- **L'index de Shannon** sensible aux variations d'importance des espèces les plus rares.
- **L'indice de Simpson** sensible aux variations d'importance des espèces les plus abondantes.

La diversité β mesure les différences entre la composition des écosystèmes. De la même façon que les index de diversité alpha, il y a plusieurs index de diversité beta qui ont été développés dans le champ de l'écologie et qui diffèrent dans la façon de prendre en compte ou pas les abondances des espèces présentes dans l'écosystème et aussi la distance évolutive de ces mêmes espèces. Parmi les index de diversité beta les plus connus :

- **Index de Jaccard** : C'est un estimateur catégorique (ne prend pas en compte la distance évolutive entre les éléments biologiques de l'écosystème quantifié) qui ne prend pas en compte l'abondance des espèces dans les écosystèmes. La distance Jaccard entre deux écosystèmes est calculée à partir de la fraction d'entités biologiques (espèces, genera, OTUs, etc) partagée ou présente dans les deux écosystèmes ($1 - \text{fraction d'espèces partagée}$, donc le plus d'espèces partagées => Les plus proches sont les écosystèmes => L'index de Jaccard plus proche à zéro).
- **Index de Bray-Curtis** : C'est un estimateur catégorique similaire à l'index de Jaccard mais qui prend en compte l'abondance des entités biologiques quantifiées pour calculer la distance entre deux écosystèmes.
- **Indexes Unifrac** : Estimateurs de diversité beta très populaires spécifiquement dans l'écologie microbienne qui prend en compte la distance évolutive entre les entités biologiques quantifiées dans le calcul des distances entre écosystèmes. L'index UniFrac Unweighted ne prend pas en compte les abondances dans le calcul des distances (comme l'index Jaccard) et l'index UniFrac Weighted prend en compte les abondances (comme l'index Bray-Curtis).

Finalement, **la diversité δ** combine la diversité alpha et la diversité beta pour quantifier la diversité totale de l'ensemble de l'écosystème étudié.

Dans le cadre de ce projet, notre équipe travaille en collaboration avec l'équipe « Integromics » de l'institut ICAN qui a mis en place un pipeline d'analyse des données métagénomiques du séquençage avec la technologie Oxford Nanopore (Figure 31). Ainsi les enregistrements électriques bruts du séquençage collectés sous format Fast5, au niveau du logiciel « MinKNOW » sont téléchargés sur les serveurs de notre équipe et transcrits, localement, en séquences ADN grâce au logiciel « Albacore » d'ONT. « Albacore » est un pipeline de traitement de données basé sur les algorithmes d'ONT qui permet de décoder les signaux en ADN et d'en extraire différentes informations permettant une évaluation ultérieure de la qualité des séquences obtenues et l'efficacité de l'expérience de séquençage. Une sélection de fichiers de configuration permet de faire appel à des bibliothèques d'ADN basées sur la gamme actuelle de kits de séquençage et de cellules de flux d'ONT.

Le pipeline Albacore permet de réaliser :

- le **basecalling, ou appel de bases**, qui transforme les signaux électriques en bases.
- la détection des brins d'étalonnage : les lectures sont alignées sur des séquences d'étalonnage de référence via l'alignement minimap2 intégré. Les séquences d'étalonnage servent de contrôle qualité pour les pores (QC). Si la lecture en cours est identifiée comme un brin d'étalonnage, aucune étape de codage à barres ou d'alignement n'est effectuée.
- le codage à barres / démultiplexage : le début et la fin de chaque séquence sont alignés sur les codes à barres fournis par ONT et utilisés pour identifier les différents ADN séquencés. Les lectures sont ainsi identifiées en fonction du code à barres qu'elles portent.
- l'alignement : l'utilisateur peut fournir un fichier de référence au format d'index FASTA, lastdb ou minimap2. Si c'est le cas, les lectures sont alignées sur cette référence via l'alignement minimap2 intégré. Dans le cas de cette étude l'alignement se fait via l'**outil « Centrifuge »**.

Le logiciel Centrifuge permet le classement taxonomique des lectures individuelles de Nanopore à l'aide d'une base de données de référence complète de plus de 8 000 génomes de référence provenant de procaryotes et de virus, incluant également le génome humain de référence. Les résultats de Centrifuge sont, ensuite traités pour éliminer les lectures provenant de l'hôte humain et pour générer un tableau de dénombrement des espèces à partir du nombre de lectures brutes des espèces de référence présentent dans les échantillons étudiés. Ce tableau est filtré pour éliminer les génomes de référence faiblement couverts afin d'éviter les fausses affectations positives. Le tableau de dénombrement des espèces filtrées est ensuite raréfié pour homogénéiser la profondeur de séquençage dans tous les échantillons dans le but d'éviter les biais associés aux différentes profondeurs de séquençage et sera utilisé pour les analyses d'écologie microbienne standard avec les emballages Phyloseq et Vegan R.

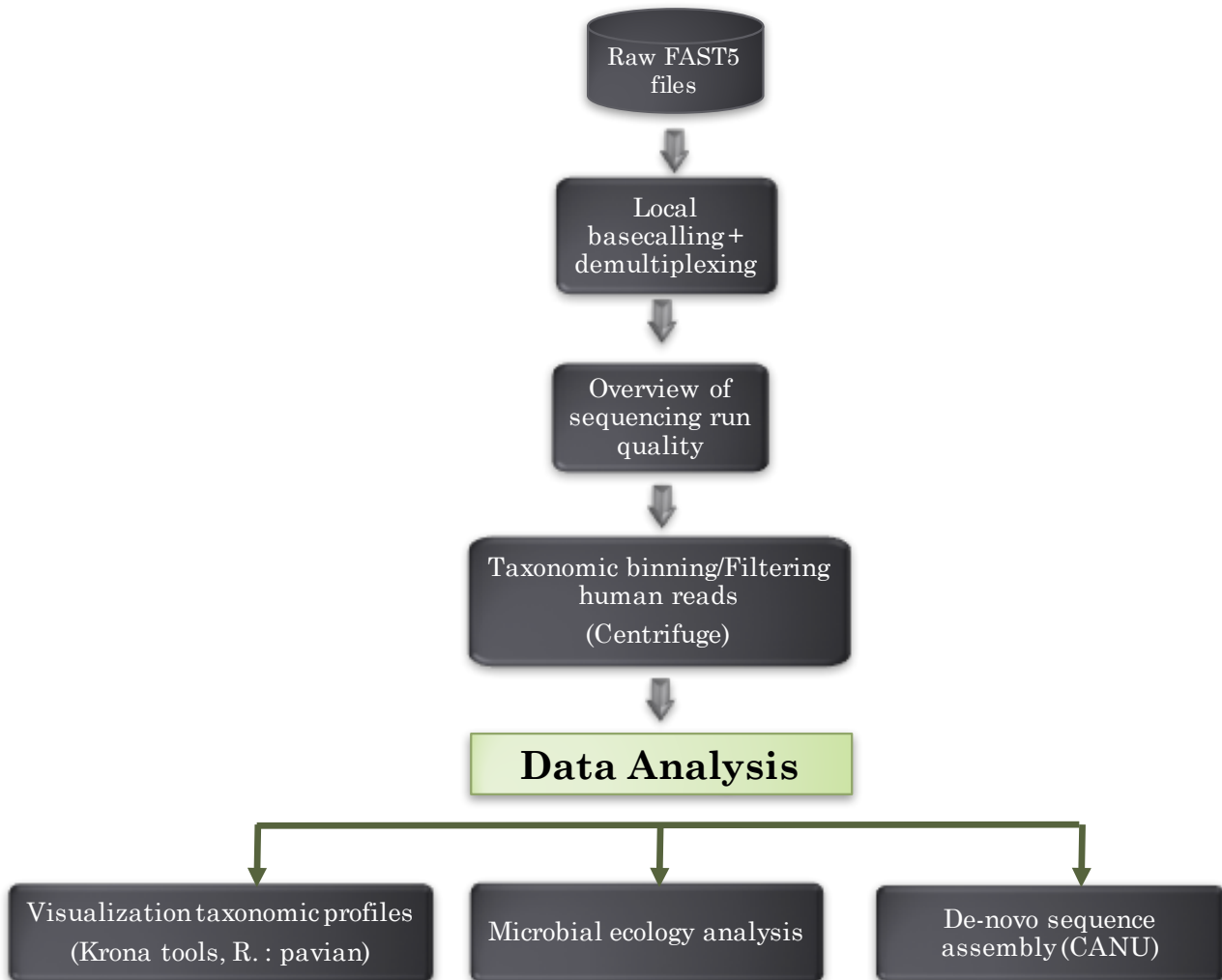


Figure 35 : Schéma général du pipeline d'analyse bio-informatique des données du séquençage MinION d'ONT qui permet de traiter les données depuis l'acquisition par le logiciel « MinKNOW » jusqu'à l'évaluation de l'écologie microbienne, la détermination des profils taxonomiques et la réalisation d'assemblage « de novo ».

Les données ainsi préparées vont ensuite servir à différentes analyses :

- Visualisation des profils taxonomiques avec l'outil « Krona » qui permet de représenter les proportions des différents ordres bactériens ainsi que celles des familles, genres et espèces.
- Analyse de l'écologie microbienne pour caractériser la biodiversité microbienne du biosystème.
- Réalisation d'assemblage « de novo » pour reconstituer le génome d'espèces bactériennes d'intérêt. Pour cela nous utilisons le logiciel « CANU ».

Résultats

1) Mise au point de l'extraction de l'ADN et de la préparation de la librairie de séquençage

A) Comparaison des kits d'extraction d'ADN choisis et test de l'étape de fragmentation de l'ADN du protocole de préparation de la librairie

Dans les études métagénomiques, la mesure des profils de composition du microbiote intestinal est précédée d'un processus complexe de génération de données, constitué de plusieurs étapes, dont chacune est soumise à la variabilité technique⁹⁶.

Dans ce processus, aucune des étapes décrites n'est standardisée, ce qui peut introduire une variabilité entre les différentes études, rendant leurs résultats difficiles à comparer. Par exemple, les différences entre la congélation et la fixation au « RNAlater » ont été signalées pour biaiser la composition de l'échantillon mesuré ainsi que l'endroit de prélèvement du spécimen de selle dans l'échantillon⁹⁷.

Dans cette expérience, j'ai donc comparé les kits d'extraction d'ADN suivants :

- INVITROGEN : PureLink™ Microbiome DNA Purification Kit.
- QIAGEN : Qiamp PowerFecal DNA Kit.
- MO BIO : Power Microbiome RNA/DNA isolation kit.
- PROMEGA : «MAXWELL Instrument» : « Maxwell RSC Buffy Coat DNA Kit» et «Maxwell RSC PureFood GMO and Authentication Kit ».

J'ai testé pour les trois premiers, l'effet de la fragmentation et de l'étape « end repair » sur le résultat du séquençage.

Pour comparer les différents kits, nous avons choisi les critères suivants :

- La taille des lectures obtenues par séquençage ONT qui pourrait rendre confortable l'étape d'alignement des lectures sur les séquences du catalogue de gènes de référence.
- L'écologie microbienne avec la détection des grands groupes bactériens communément retrouvés dans la flore intestinale.

Devant le nombre important d'échantillons que nous sommes amenés à extraire, nous sommes intéressés par l'automatisation de l'extraction. Nous avons donc testé le « MAXWELL Instrument » de chez Promega qui permet, après une phase de prétraitement de 30 minutes, de traiter 16 échantillons en un cycle de 40 minutes.

Prélèvement et extraction : Les prélèvements de selles testés ont été recueillis auprès d'un volontaire sain avec un IMC < 20kg/m² à quatre dates d'émissions différentes dans un tube sec. Ils ont été, ensuite, aliquotés et stockés immédiatement à -80°C. Chaque échantillon a été extrait selon les protocoles fournis par les fournisseurs des kits utilisés. Après extraction, les échantillons ont été évalués :

- Par le fluoromètre « Qubit » qui permet d'estimer la quantité d'ADN obtenue en ng/μl.
- Par le Nanodrop qui permet d'évaluer la qualité de l'échantillon en se référant à trois mesures d'absorbance : 260 nm qui permet de détecter les acides nucléiques, 280nm qui détecte les protéines et 230nm qui est sensible aux solvants. Deux rapports sont alors générés pour déterminer la pureté de l'échantillon : 260/280 qui doit être supérieure ou égal à 1,8 et 260/230 qui doit être compris entre 2 et 2.2.
- Par la Tape Station qui permet de faire migrer les échantillons d'ADN sur un gel pour avoir un visuel de la qualité et fournit un DIN (DNA integrity number) qui témoigne de cette qualité.

Préparation de la librairie pour le séquençage : Dans le protocole de préparation de la librairie pour le séquençage MinION d'ONT, la première étape consiste à fragmenter l'ADN, dans des g-tubes (Covaris), pour générer des fragments d'une taille de 8kb. L'objectif de cette expérience est donc de tester si la fragmentation de l'ADN est indispensable dans le cadre de notre expérimentation.

Résultats

Cette première expérience a permis, dans un premier temps, d'avoir une première idée de la longueur des séquences obtenues par la technique de séquençage d'ONT. Nous constatons donc que le catalogue de 9.9 millions de gènes comporte des séquences de références allant jusqu'à environ 6Kb. Or les séquences obtenues par le séquençage Oxford Nanopore montrent des lectures allant jusqu'à 24Kb environ (Figure 36). Ceci aurait pu poser problème pour la quantification des gènes avec les outils actuellement disponibles, plus adaptés aux séquences de longueurs courtes générées (300 bp en moyenne) par le séquençage Illumina, par exemple. Un travail d'adaptation a été effectué par l'équipe Integromics, d'un pipeline initialement dédié à l'analyse des données illumina pour pouvoir l'utiliser pour les séquences longues du MinION.

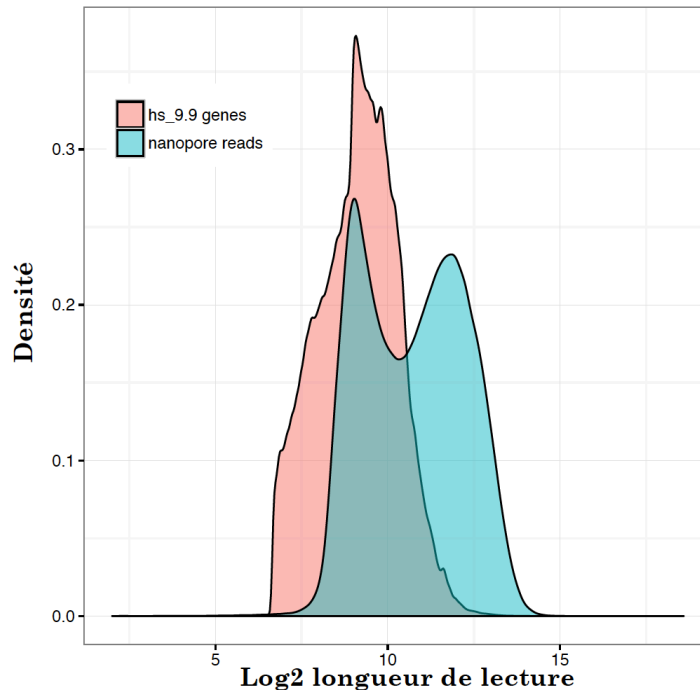


Figure 36 : Longueurs des séquences générées par la technologie MinION d'ONT vs. gènes de références du catalogue de 9.9 millions de gènes établit à partir des résultats de séquençage de 249 échantillons nouvellement séqués, 1 018 échantillons publiés des études Meta HIT⁹⁸ et HMP⁹⁹ et celles d'une grande étude sur le diabète en Chine¹⁰⁰, ainsi que 511 génomes séquencés de bactéries et archées.

Le deuxième résultat de cette expérimentation concerne l'effet de l'étape de fragmentation sur le rendu du séquençage. Le premier constat est que nous obtenons la même distribution bimodale des lectures générées selon que l'on fragmente l'ADN ou non. On peut diviser cette courbe en deux parties : à gauche, courbe regroupant les lectures les plus courtes d'environ 100 à 1100 paires de bases, à droite celle regroupant les lectures les plus longues supérieures à 1100 paires de bases et allant jusqu'à 24Kb environ. Comme on peut le voir sur la figure 37, la densité des lectures les plus courtes est plus élevée avec les ADN non fragmentés. Cependant, lorsque nous regardons les profils de chaque échantillon selon son mode de préparation et en fonction du kit d'extraction utilisé (figure 39), nous remarquons que cette densité est surtout supportée par les lectures obtenues avec l'ADN extrait avec le kit Qiagen. Pour les autres échantillons, les deux courbes de distribution des lectures se superposent complètement, pointant ainsi l'absence d'effet de l'étape de la fragmentation de l'ADN sur la longueur des séquences obtenues.

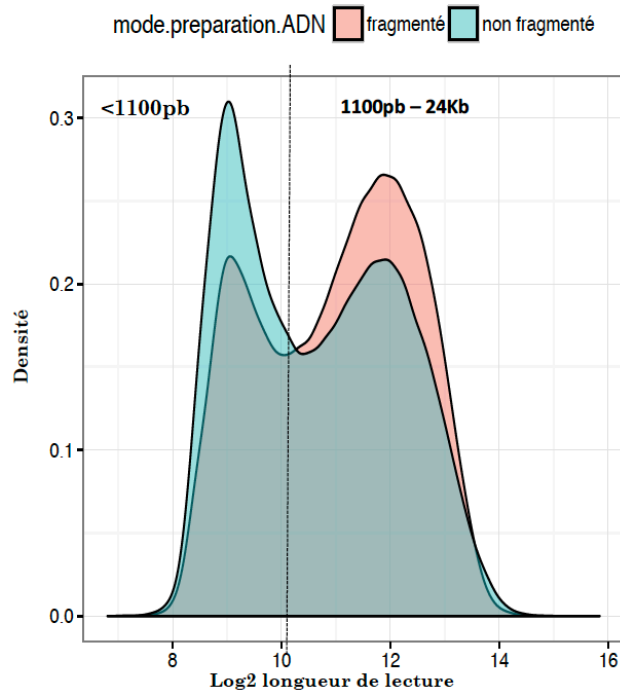


Figure 37 : Distribution des longueurs des séquences obtenues avec ou sans fragmentation de l'ADN. La courbe bimodale peut être divisée en deux parties regroupant à gauche des lectures courtes (<à 1100pb) et à droite des lectures longues (> à 1100pb).

L'analyse de la diversité alpha (fig. 38) montre pour chaque échantillon et quel que soit son mode de traitement, un niveau similaire de diversité microbienne. Ceci est en faveur d'un effet négligeable de l'étape de fragmentation sur le résultat du séquençage.

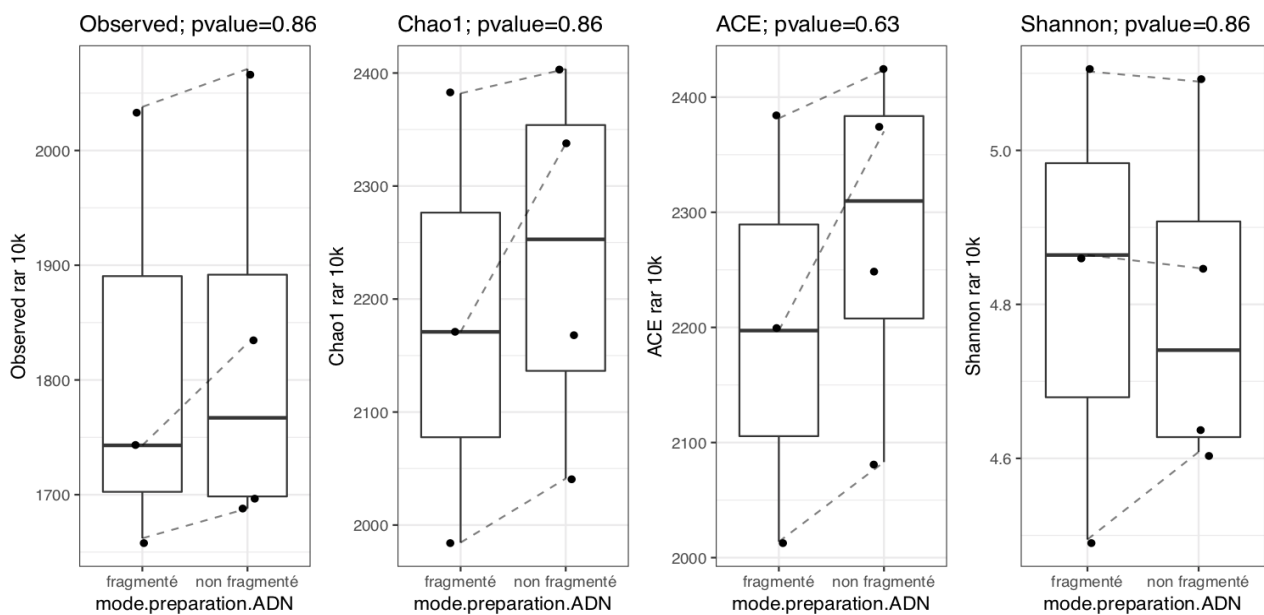


Figure 38 : Diversité alpha observée et estimée selon différents index (voir matériel et méthode). Les variations observées sont non significatives pointant ainsi l'absence d'apport supplémentaire de la fragmentation de l'ADN à l'estimation de la diversité microbienne (Test statistique de Wilcoxon).

Nous avons ensuite comparé les résultats obtenus avec les différents kits utilisés pour extraire l'ADN (figure 39). On constate que selon le kit utilisé, la distribution des longueurs des lectures n'est pas la même.

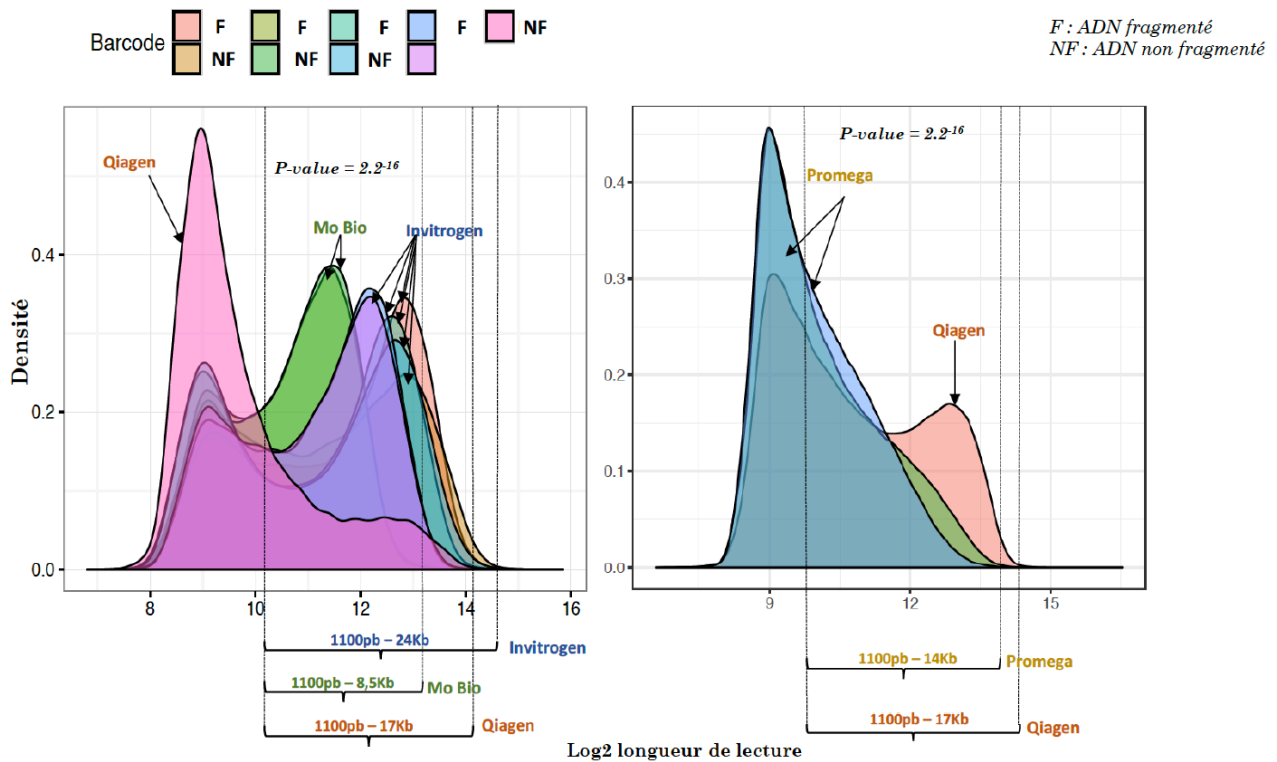


Figure 39 : Distribution des longueurs des séquences obtenues en fonction des kits d'extraction utilisés, avec ou sans fragmentation de l'ADN obtenus à partir des prélèvements d'un volontaire sain à quatre dates différentes. La densité reflète le nombre de lecture retrouvé dans tout l'échantillon et sa probabilité de correspondre à une longueur donnée sur l'axe des abscisses (orange : Qiagen, vert : Mo Bio et bleu : Invitrogen). Variations significatives des distributions des lectures en fonction du kit d'extraction d'ADN utilisé (Test de Kruskal-Wallis).

Les kits Invitrogen et Mo Bio montrent une densité plus grande de séquences supérieures à 1,1Kb. Cependant, le kit Mo Bio donne des longueurs maximales de 8,5Kb alors que le kit Invitrogen permet d'obtenir des lectures allant jusqu'à 24Kb. Le kit Qiagen donne une densité plus grande de séquences courtes et des séquences longues allant jusqu'à 17Kb. Les échantillons extraits avec le MAXWELL Instrument de chez Promega et quel que soit le kit utilisé, montre une courbe unimodale avec une densité élevée pour les lectures inférieures à 1,1Kb.

Enfin, au niveau du phylum, le profil taxonomique des librairies obtenues à partir d'un même échantillon montre une grande similarité que l'ADN soit fragmenté ou pas. Nous observons, également, que les échantillons collectés le même jour, et extrait avec les kits Invitrogen et Mo Bio, présentent des profils taxonomiques très similaires. L'échantillon extrait avec le kit Qiagen montre une dominance de Bacteroidetes comparé aux autres échantillons.

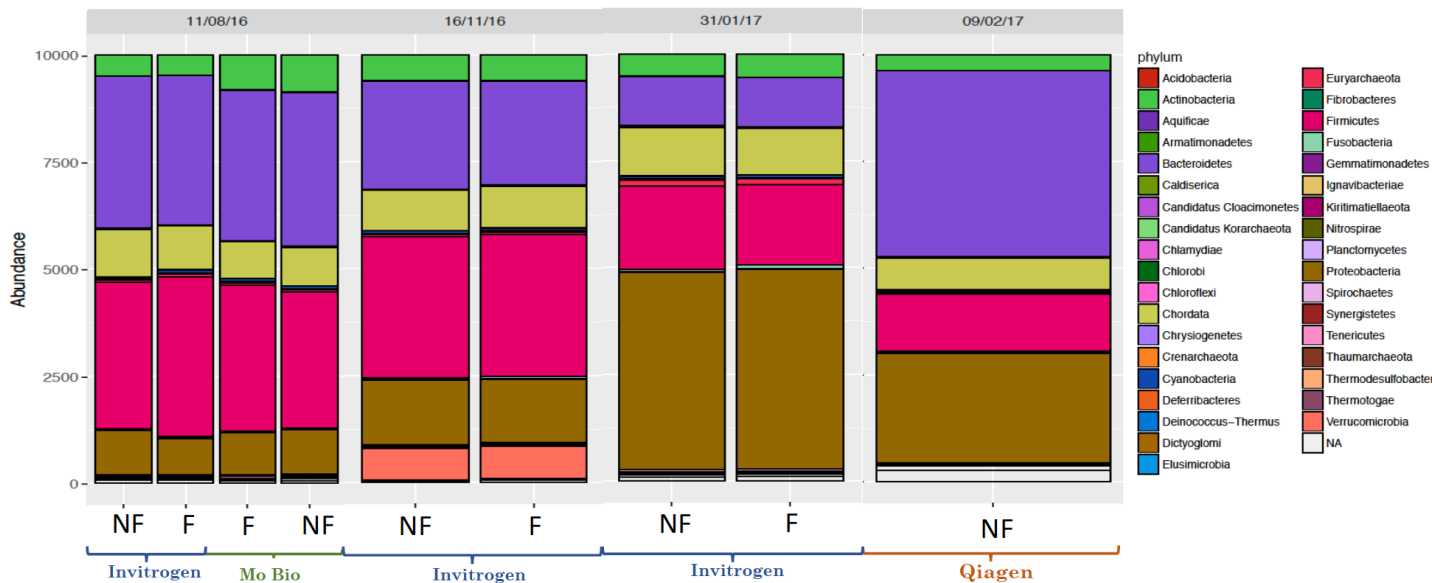


Figure 40 : Profil taxonomique des échantillons en fonction du kit d'extraction. NF = ADN non fragmenté, F = ADN fragmenté. L'abondance représente le nombre d'espèces observées pour chacun des ordres détectés. Profil obtenu à partir des prélèvements d'un volontaire sain à quatre dates différentes de prélèvements.

La stratification obtenue à partir des différentes librairies montre des différences dans l'abondance en espèces bactériennes en fonction de la date de prélèvement de l'échantillon (figure 41).

Par exemple :

- Akkermansia (de l'ordre des Verrucomicrobiales) est principalement séquencée dans les librairies du 16/11/2016.
- l'abondance d'*Escherichia. coli* est plus importante dans les librairies du 31/01/2017.
- *E. rectale* est présente dans toutes les librairies mais son abondance varie en fonction de la date de recueil des selles, indépendamment du kit d'extraction utilisé.

Nous observons que les échantillons collectés à la même date présentent des profils taxonomiques très similaires, que l'ADN soit fragmenté ou pas, préalablement au séquençage.

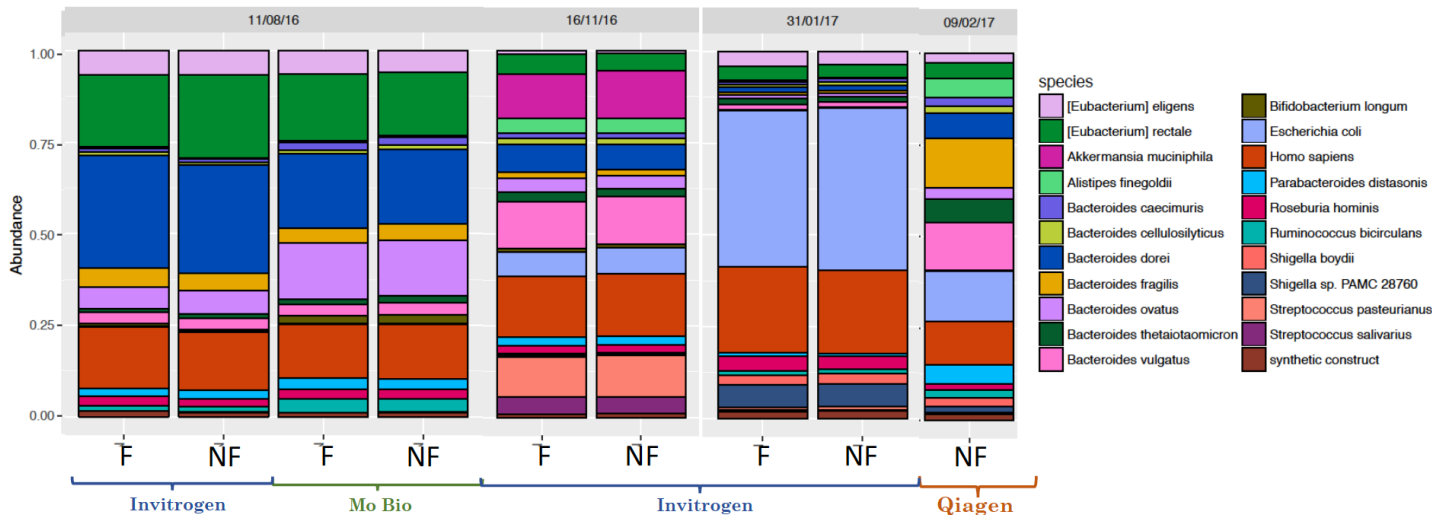


Figure 41 : Profil taxonomique des échantillons en fonction de la date de prélèvement. NF = ADN non fragmenté, F = ADN fragmenté. L'abondance représente le nombre de bactéries observées pour chacune des espèces détectées. Profils obtenus à partir des prélèvements d'un volontaire sain à quatre dates différentes de prélèvements.

Conclusion :

- Nous n'observons pas de différences ni au niveau de la longueur des lectures ni au niveau de la distribution taxonomique entre les bibliothèques obtenues à partir de l'ADN préalablement fragmenté et l'ADN non fragmenté. Je m'affranchi donc de l'étape de fragmentation de l'ADN dans les expérimentations à venir.
- Nous obtenons plus de lectures longues avec le kit « PureLink™ Microbiome DNA Purification » d'Invitrogen qu'avec les kits Mo Bio, Qiagen et Promega. Nous choisissons donc, le kit Invitrogen pour réaliser l'extraction de l'ADN microbien dans notre laboratoire.
- Nous observons une variabilité dans l'abondance de certaines espèces en fonction de la date de prélèvement des selles chez le même sujet.

B) Test de l'étape de « End repair » du protocole de préparation de la librairie

La deuxième étape dans le protocole de préparation de la librairie d'ONT est la réparation des extrémités de l'ADN qui auraient pu s'abîmer pendant la fragmentation. Cette étape est également optionnelle et recommandée par ONT aussi pour l'ADN non fragmenté. Les échantillons de selles de trois volontaires sains, ont été recueillis à différents temps. L'extraction a été réalisée avec le kit « PureLink™ Microbiome DNA Purification ». L'ADN n'a pas subi de fragmentation lors de la préparation de la librairie. Deux librairies ont été préparées pour chaque ADN : une incluant l'étape de « End repair » et l'autre sans cette étape optionnelle.

Résultats :

Lorsque les longueurs de lectures sont stratifiées par échantillon (fig. 42), nous observons que le mode de préparation de l'ADN (« end repair » vs. « non end repair ») n'a aucun effet sur les longueurs de lectures car nous obtenons des distributions qui se superposent complètement.

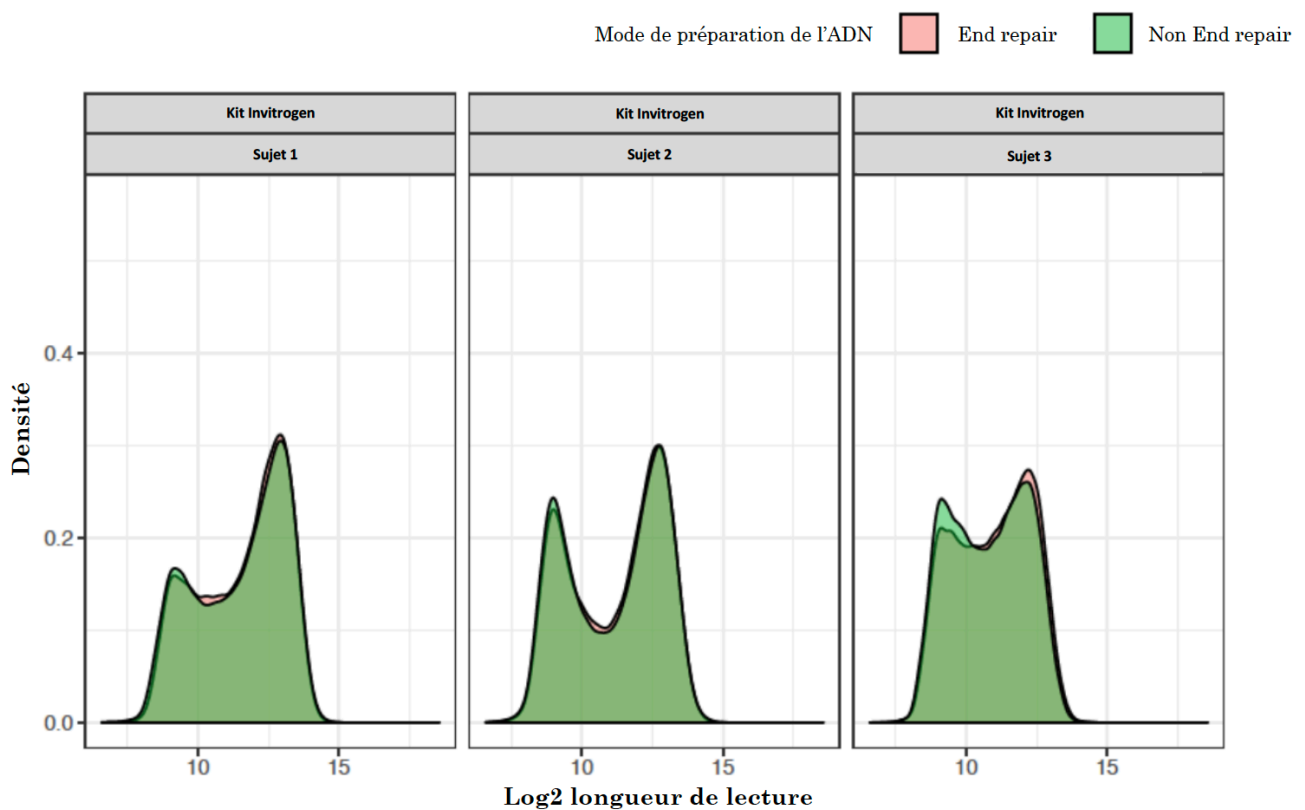


Figure 42 : Distribution des longueurs des séquences obtenues en fonction du protocole de préparation de la librairie, « End repair » vs « non end repair » à partir des prélèvements de 3 sujets sains.

L'analyse de la diversité alpha (fig. 43) comme pour le test de fragmentation de l'ADN, ne montre pas de variation significative dans l'estimation de la diversité microbienne que l'échantillon d'ADN concerné ait subi l'étape de « End repair » ou pas. Ceci montre qu'il n'y a pas d'effet supplémentaire du traitement des extrémités de l'ADN non fragmenté sur la diversité microbienne observée.

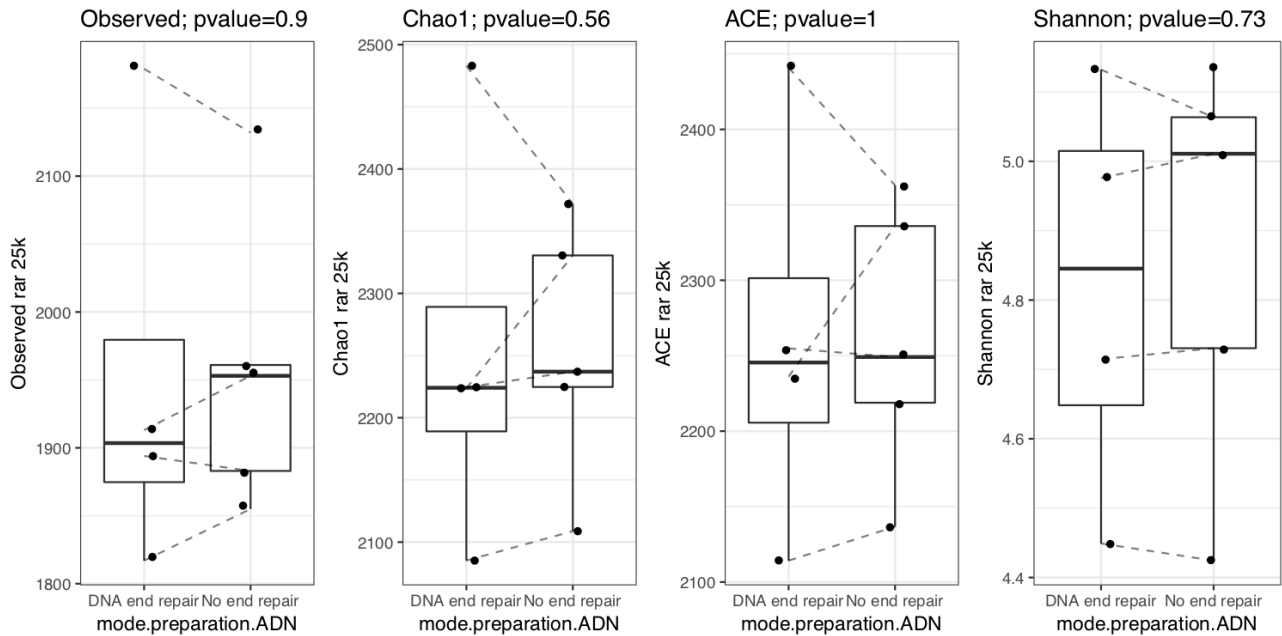


Figure 43 : Diversité alpha observée selon le mode de préparation de l'ADN « End repair » vs « non End repair » et estimée selon différents index, à partir des prélèvements de 3 sujets sains (Test de Wilcoxon).

Conclusion

L'étape de « End repair » n'est pas nécessaire dans le cadre de nos expérimentations. Je m'affranchis donc de cette étape pour les expériences à venir.

2) Améliorations de l'extraction de l'ADN avec le Kit Invitrogen et du protocole de préparation de la librairie.

Nous nous sommes, par la suite, intéressés à l'étude publiée par le consortium IHMS (International Human Microbiome Standards projet)¹⁰¹. La détermination des profils taxonomiques et de la diversité bactérienne est précédée d'un processus complexe de plusieurs étapes de génération de données, dont chacune est soumise à la variabilité technique³. « L'International Human Microbiome Standards (IHMS) project » a évalué dans quelle mesure l'extraction de l'ADN pouvait influencer la quantification de la composition microbienne, et comparé cette variable avec d'autres sources de variation technique et biologique. Cette étude a débouché sur l'élaboration de « SOPs » (pour Standard Operating Procedure) comportant des recommandations qui permettraient d'améliorer l'extraction de l'ADN en termes de quantité et de qualité (<http://www.microbiome-standards.org/>).

J'ai donc, dans cette deuxième étape de mise au point, tenté d'appliquer ces recommandations au protocole d'extraction de l'ADN microbien choisit, à savoir le « PureLink Microbiome DNA Purification Kit » de chez Invitrogen.

Les étapes supplémentaires proposées par le consortium sont représentées dans la figure 44, ci-dessous :

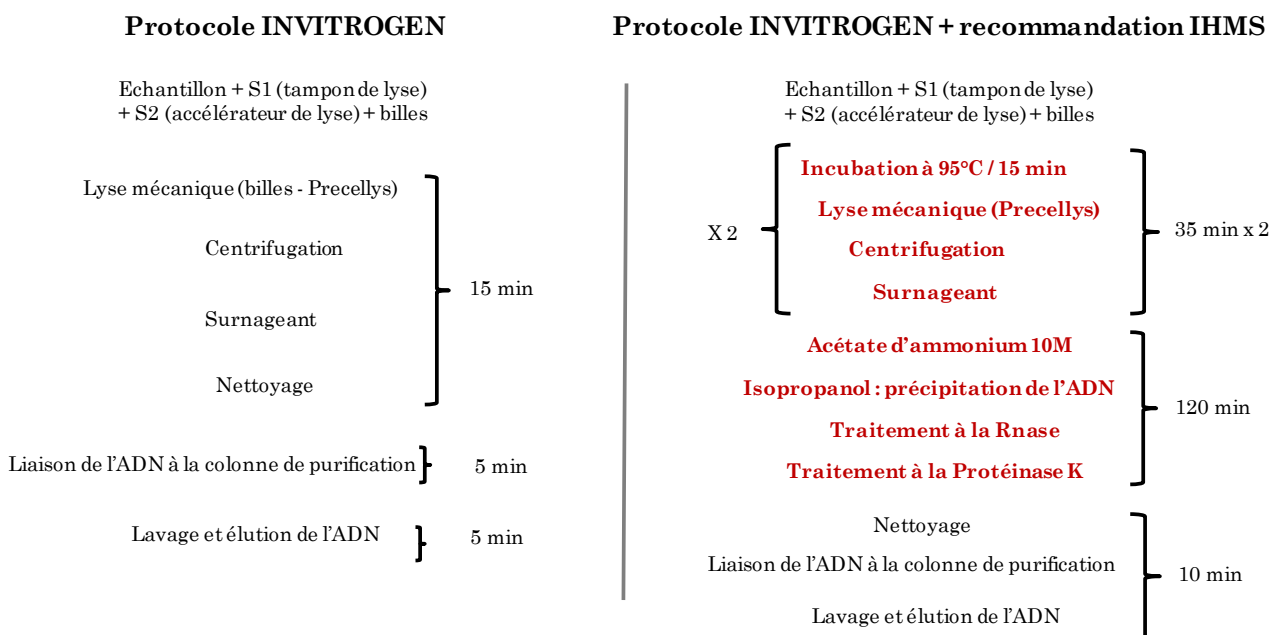


Figure 44 : Comparaison entre le protocole d'extraction du « PureLink™ Microbiome DNA Purification kit » de chez Invitrogen comme préconisé par le fournisseur et celui adapté selon les recommandations de IHMS.

12 échantillons provenant de deux volontaires sains ont été utilisés pour cette expérimentation. Ils ont été extraits avec le protocole d'extraction « classique » tel que proposé par le fournisseur et avec le protocole « modifié », avec les recommandations IHMS.

Résultat

L'application des recommandations IHMS à l'extraction de l'ADN avec le kit Invitrogen montre (tableau 3) :

- Une nette augmentation du rendement de l'extraction d'ADN avec le protocole supplémenté des recommandations IHMS par rapport au protocole classique.
- Une amélioration de la qualité de l'ADN attestée par le rapport 260/230 entre 1,61 et 2,11. Ceci peut s'expliquer par le fait que dans le protocole modifié, l'étape de nettoyage qui permet d'éliminer les inhibiteurs de PCR, tel que les solvants, intervient après une série de traitement de l'ADN. Cette étape de « nettoyage » s'opère donc sur du matériel plus propre comparé à celui obtenu avec le protocole classique où le « nettoyage » à lieu directement sur le surnageant issu de la lyse mécanique des bactéries.
- Une diminution de la contamination par les protéines montrée par un rapport 260/280 compris entre 1,81 et 1,93. Cette amélioration est le résultat d'un double traitement des échantillons pour éliminer les protéines (Acétate d'ammonium qui précipite les protéines et la protéinase K qui les dégrade).

Tableau 3 : Amélioration de l'extraction à partir des selles humaines de l'ADN bactérien avec le protocole du kit « PureLink™ Microbiome DNA Purification » supplémenté des recommandation IHMS. ADN obtenu à partir de 12 échantillons recueillis auprès de 2 volontaires sains. P-value estimée par test de Wilcoxon.

Paramètres	Protocole Invitrogen Classique	Protocole Invitrogen Modifié	P-value
ng/μl	74,5 ± 0,48	300,15 ± 0,75	<0,0001
260/280	1,79 ± 0,03	1,89 ± 0,02	0,0046
260/230	1,31 ± 0,05	1,87 ± 0,04	0,0007

Je me suis ensuite intéressée à l'amélioration de la préparation de la librairie pour le séquençage en apportant des modifications au protocole proposé par ONT. Deux étapes importantes dans la préparation de la librairie ont été concernées :

- L'étape de « End-prep » qui consiste à phosphoryler l'extrémité 5' de l'ADN et à ajouter une Adénine à l'extrémité 3' pour préparer la liaison de l'adaptateur. Cette étape est suivie de deux périodes d'incubation. Dans le protocole d'ONT, il est conseillé d'incuber la réaction (ADN + NEBNext Ultra II End Repair/dA-Tailing Module) pendant 5 minutes à 20°C suivit de 5 minutes à 65°C. Le « NEBNext Ultra II End Repair/dA-Tailing Module » utilisé est de chez New England Biolabs (NEB). Les recommandations de NEB pour l'utilisation de ce kit sont : une incubation à 20°C pendant 30 minutes suivie d'une incubation à 65°C pendant 30 minutes. Ce sont ces recommandations que j'ai testées dans cette expérience.

- L'ADN traité par le kit End prep est ensuite purifié grâce à l'utilisation de billes paramagnétiques « Agentcourt AMPure XP » de chez Beckman. Le principe de la sélection des tailles de fragments avec la technologie SPRI des billes AMPur XP est la suivante : La taille des fragments élués à partir des billes (ou qui se lient en premier lieu) est déterminée par la concentration de PEG qui dépend, à son tour, du mélange d'ADN et de billes. Par exemple, si on ajoute 50µl de billes à un échantillon d'ADN de 50µl on obtient un rapport SPRI/ADN de 1. Lorsque ce rapport est modifié, la longueur des fragments se liant et / ou restant dans la solution change également, plus le rapport SPRI/ADN est faible, plus les fragments finaux sont grands. Des fragments plus petits sont conservés dans le tampon et on pourrait ainsi obtenir des gammes de tailles différentes à partir d'un seul échantillon avec des purifications multiples. Une partie de la raison de cet effet est que la taille du fragment d'ADN affecte la charge totale portée par la molécule. Ainsi, des fragments longs d'ADN auront des charges négatives plus importantes, ce qui favorise leur interaction électrostatique avec les billes et permet une liaison prioritaire aux molécules carboxyle. Le protocole de séquençage proposé par ONT est élaboré sur la base de la fragmentation de l'ADN et de la sélection de fragments d'ADN de 8Kb. Pour cela un rapport SPRI/ADN doit être égale à 1.

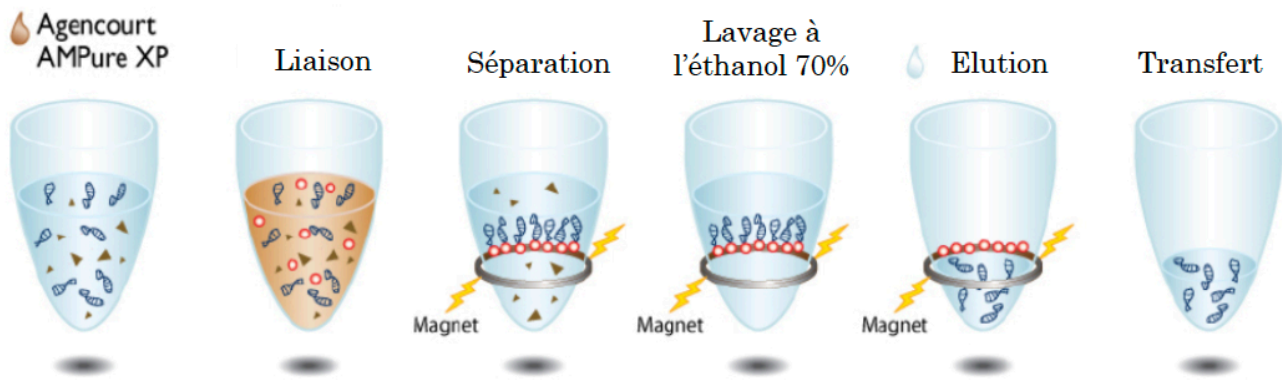


Figure 45 : Principe de la sélection des fragments d'ADN avec la technologie SPRI. L'ADN chargé négativement se lie au groupement carboxyle des billes. Les billes sont retenues sur un aimant. On procède, ensuite au lavage et à l'éluion de l'ADN. ([Http://core-genomics.blogspot.com/2012/04/how-do-spri-beads-work.htm](http://core-genomics.blogspot.com/2012/04/how-do-spri-beads-work.htm)).

Afin de favoriser la sélection des grands fragments d'ADN par les billes paramagnétiques, j'ai réduit le rapport SPRI/ADN à 0,4. Ce calcul se base sur une documentation sur la technologie SPRI ainsi que des recommandations d'utilisateurs de la technologie ONT, recueillies sur le forum « Community ». J'ai donc, combiné deux modifications de deux étapes du protocole de séquençage d'ONT pour tenter l'améliorer la longueur des lectures.

Les échantillons de selles ont été recueillis auprès d'un volontaire sain (témoin1) à six dates différentes. L'extraction de l'ADN a été réalisée avec le protocole Invitrogen modifié mais également avec le protocole classique afin de valider l'amélioration de la longueur des lectures observée dans la mise au point précédente.

Résultat

L'analyse de la distribution des lectures en fonction de leurs longueurs nous permet de constater que (fig. 46) :

- Avec le protocole « classique » d'extraction de l'ADN, nous obtenons toujours une courbe bimodale, avec une densité variable de lectures courtes en fonction du jour de recueil.
- Le protocole d'extraction de l'ADN amélioré selon les recommandations d'IHMS, plus la modification du protocole de préparation de la librairie permet d'obtenir une courbe unimodale avec une distribution des lectures en faveur des séquences les plus longues.

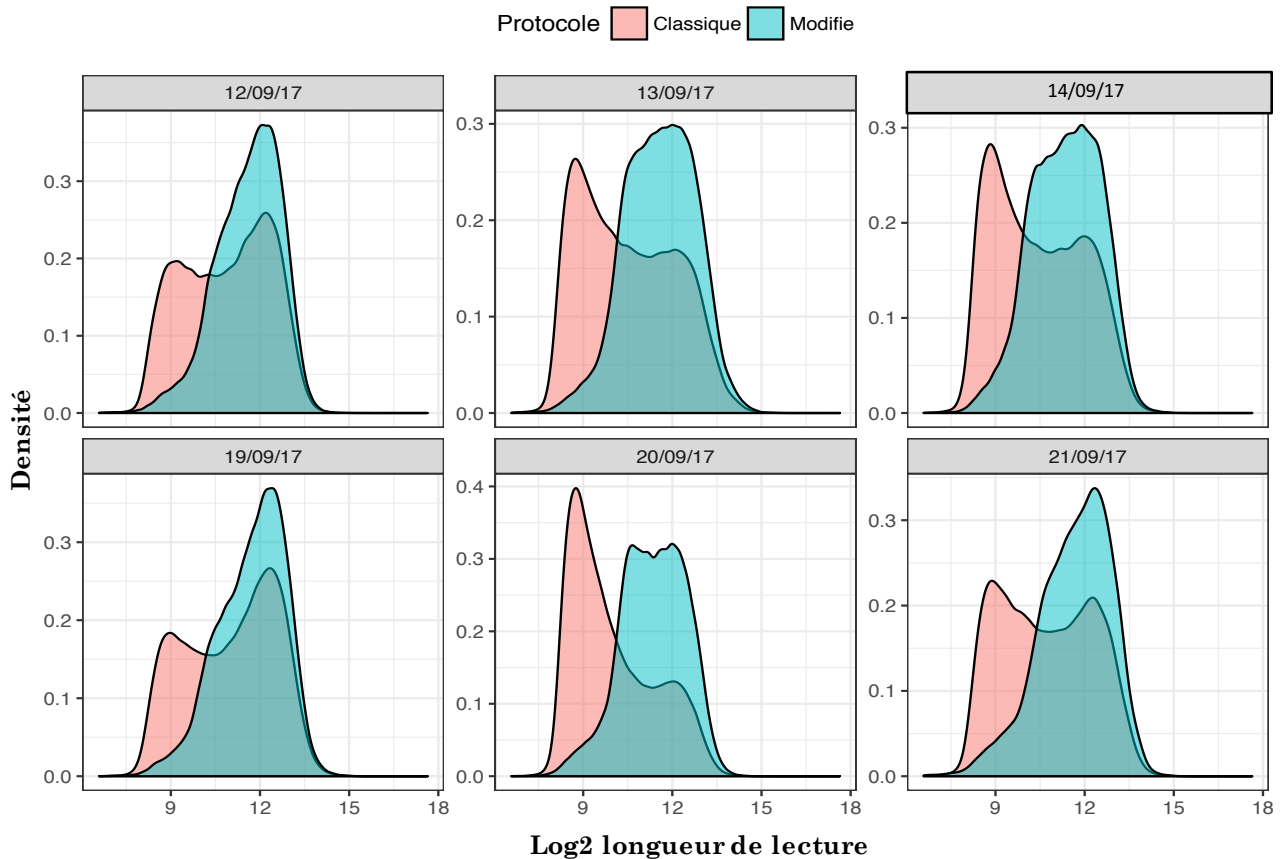


Figure 46 : Distribution des longueurs des séquences obtenues en fonction du protocole d'extraction classique ou modifié utilisé pour l'extraction de l'ADN à partir de 6 prélèvements de selles différents recueillis auprès du témoin 1. Courbe bimodale pour le protocole classique (en rose), courbe unimodale pour les protocoles modifiés (en vert).

L'analyse taxonomique montre également une amélioration des résultats. En effet, nous observons une diminution de la fraction des lectures non classifiées avec les protocoles modifiés d'extraction de l'ADN et de préparation de la librairie comparés au profil obtenu avec les protocoles classiques (figure 47).

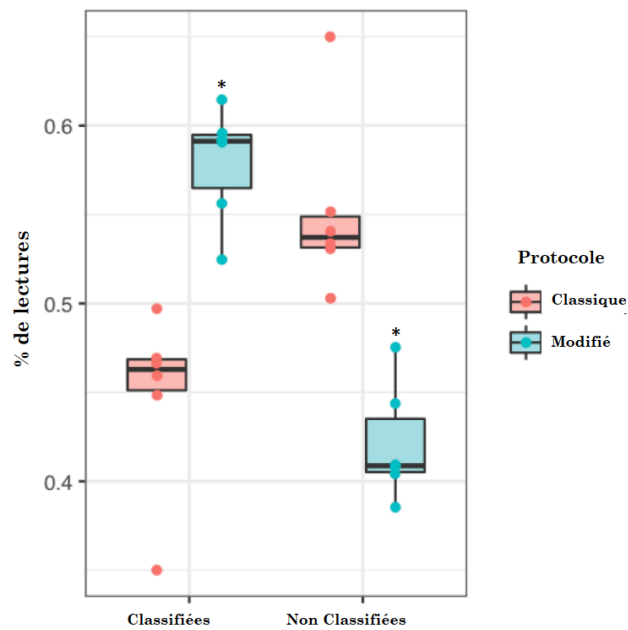


Figure 47 : Profils taxonomiques des lectures obtenues en fonction de leurs classifications à partir des prélèvements à 6 dates différentes chez le témoin 1. Le pourcentage de lectures classifiées à partir des protocoles modifiés est supérieur au pourcentage obtenu avec le protocole classique. P-value estimée par test de Wilcoxon (Classifiée : p-value = 0,0022, Non Classifiée : p-value = 0,002).

La mesure de l’alpha diversité (fig. 48), estimée avec les indices Chao1 et ACE montre une plus grande diversité dans les échantillons de selles dont l’ADN a été extrait avec le protocole Invitrogen modifié. Ceci peut s’expliquer par le fait que les protocoles modifiés permettent d’avoir accès à plus de séquences classifiées (fig. 47), comme montré par l’analyse taxonomique, et donc d’identifier davantage d’espèces.

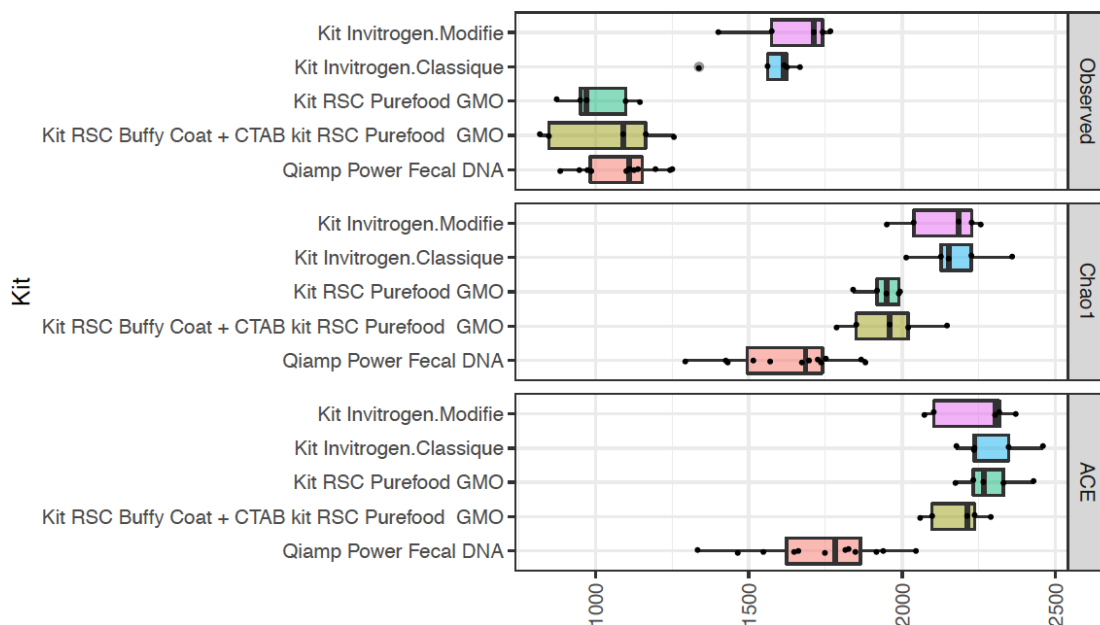


Figure 48 : Alpha diversité observée et estimée selon les indices Chao1 et ACE (axe des abscisses) de différentes techniques d’extraction de l’ADN à partir de 6 prélèvements différents obtenus auprès du témoin 1. Test de Kruskal-wallis (observed: p-value = 8.8^e-05, Chao1: p-value = 1.1^e-05, ACE: p-value = 6.2^e-05).

Conclusion : Dans cette deuxième étape de mise au point du séquençage du microbiote intestinal par la technologie Oxford Nanopore, j'ai pu :

- Améliorer le rendement et la qualité de l'ADN en appliquant les recommandations du consortium IHMS au protocole d'extraction du kit « PureLink Microbiome DNA Purification ».
- Améliorer la qualité du séquençage en modifiant le protocole de préparation de la librairie. Ceci a pour résultat, des lectures plus longues, plus d'espèces identifiées et donc une meilleure estimation de la diversité, et comme nous l'espérons, cela permettra d'avoir accès à des espèces rares.

3) Mode de recueil des selles

A) Comparaison de différents modes de stabilisation de l'ADN microbien et de deux modes d'acheminement des prélèvements de selles jusqu'au laboratoire pour l'analyse.

Jusqu'à présent, les prélèvements de selles chez nos patients ont été recueillis dans un tube cuillère sec avec la nécessité d'avoir accès à un -20°C ou -80°C pour assurer la stabilité de l'ADN bactérien de l'échantillon. Aujourd'hui, on trouve sur le marché plusieurs produits qui pourraient nous affranchir de l'impératif de congélation. Effectivement, on trouve de plus en plus de systèmes contenant une solution qui permet de stabiliser l'ADN bactérien pendant des périodes allant de 60 jours (DNA Genotek) à 2 ans (NORGEN et OZYME).

Pour évaluer la pertinence de l'utilisation du mode de stabilisation de l'ADN bactérien par l'utilisation du liquide stabilisateur, j'ai mis en place un protocole expérimental visant à comparer le résultat de séquençage obtenu à partir de bibliothèques provenant de :

- Selles totales stabilisées à -20°C puis à -80°C.
- Selles stabilisées par le système OMNIGEN :
 - * Prélévées et remises au laboratoire (RT)
 - * Prélévées à domicile et envoyées par la poste (Poste)
- Selles stabilisées par le système NORGEN :
 - * Prélévées et remises au laboratoire (RT)
 - * Prélévées à domicile et envoyées par la poste (Poste)

L'objectif de l'envoi de l'échantillon par la poste est de vérifier si ce mode d'acheminement, avec tout ce qu'il comporte comme variation de température et de traitement des colis, modifie la stabilité de l'ADN bactérien. Deux sujets sains, témoin 1 et témoin 2, se sont portés volontaires pour ces tests.

Résultats

Lorsque nous représentons les données stratifiées par sujet et par protocole de prélèvement fécal (fig. 48), nous observons certaines variations dans les distributions des longueurs des lectures qui diffèrent d'un sujet à l'autre :

- Pour le témoin 1, les kits de prélèvement NORGEN et OMNIGEN ne semblent pas montrer de variation dans les longueurs des lectures selon le mode de transport car dans les deux cas, on observe des lectures longues entre 1,1 et 10Kb pour NORGEN et d'environ 1,1 à 24Kb pour OMNIGEN. La selle totale avec une stabilisation à -80°C de l'ADN bactérien donne des lectures allant jusqu'à 21Kb environ.

- Pour le témoin 3, nous observons un schéma différent soulignant un impact du mode d'acheminement du prélèvement vers le laboratoire, sur la longueur des lectures. La selle totale donne les lectures les plus courtes, comprise entre 1,1 et 10Kb environ. Les kits NORGEN et OMNIGEN, lorsqu'ils sont immédiatement stockés à température ambiante après le recueil, donnent des lectures de longueurs relativement équivalentes d'environ 21 et 24Kb, respectivement.

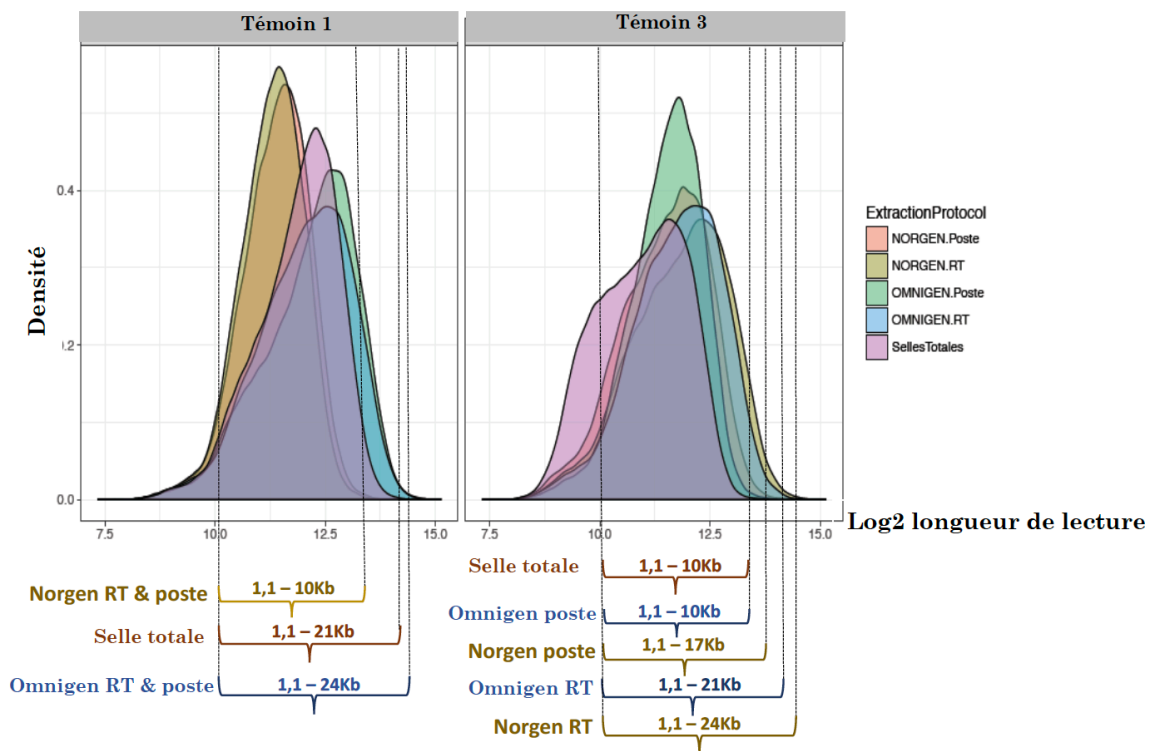


Figure 49 : Distribution des longueurs des séquences obtenues en fonction du kit de prélèvement utilisé et son mode de transport : remise et stockage à RT au laboratoire ou envoi par la poste et stockage à RT à réception. Échantillons obtenus à partir de 2 prélèvements de 2 sujets sains.

Les fractions de lectures classifiées et non classifiées obtenues avec les différentes librairies ne montrent pas de variations dans les résultats des deux individus (p -value $> 0,05$). On obtient donc, les mêmes pourcentages de lectures classifiées et non classifiées (fig. 50).

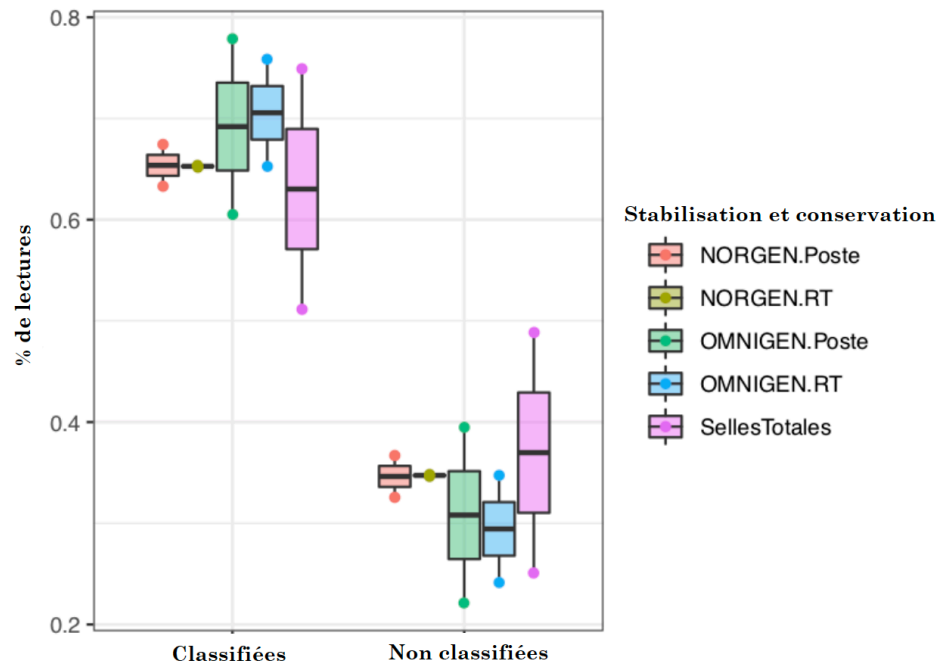


Figure 50 : Profils taxonomiques des lectures obtenues en fonction de leurs classifications ne montrant pas de variations significatives dans les proportions des lectures classifiées et non classifiées, entre les deux systèmes de stabilisation de l'ADN testés et leurs modes d'acheminement au laboratoire, comme l'atteste le test de Wilcoxon : classifiées : p -value = 0,93, non classifiées : p -value = 0,92. Échantillons obtenus à partir de 2 prélèvements de 2 sujets sains.

L'analyse de la diversité alpha (fig. 51) montre une différence significative de la richesse bactérienne entre les deux sujets. En effet, le témoin 3 présente une plus grande diversité bactérienne que le témoin 1. Nous constatons, également, que le témoin 1 présente une meilleure estimation de la diversité avec les kits NORGEN et OMNIGEN lorsque les échantillons sont conservés directement à température ambiante dans la solution de préservation après le recueil. Dans le cas du témoin 3 cette tendance est inversée.

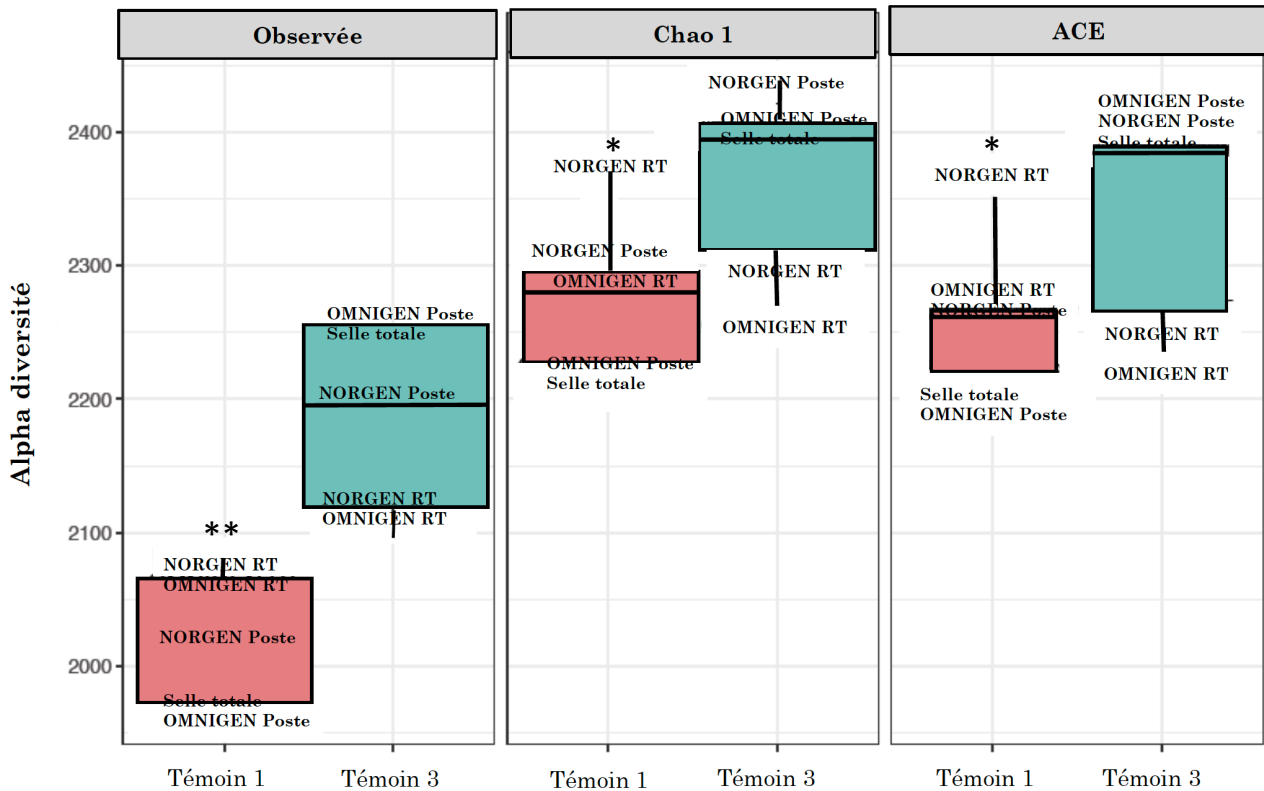


Figure 51 : Alpha diversité observée et estimée selon Chao1 et ACE pour chaque sujet en fonction des systèmes testés de stabilisation de l'ADN et leurs modes d'acheminement vers le laboratoire, montrant une plus grande diversité bactérienne pour le témoin 3 comparé au témoin 1. Variations significatives entre les deux sujets (Test de Wilcoxon : Observée (p-value = 0,0079), Chao 1 (p-value = 0,095), ACE (p-value = 0,015)).

La représentation inverse (stratification par système de recueil des selles) (fig. 51) montre une variabilité moindre entre les deux sujets avec les systèmes OMNIGEN GUT et NORGEN lorsque les échantillons sont remis au laboratoire et stockés à température ambiante. Par contre cette variabilité est plus grande lorsque les échantillons sont envoyés par la poste. Une variabilité similaire est observée avec le prélèvement effectué sur tube sec et stabilisé immédiatement à -80°C ; il est important de noter que ce dernier subit une décongélation préalable à l'extraction, même si elle reste partielle. Cependant, d'un point de vue statistique, les variations observées ne sont pas significatives.

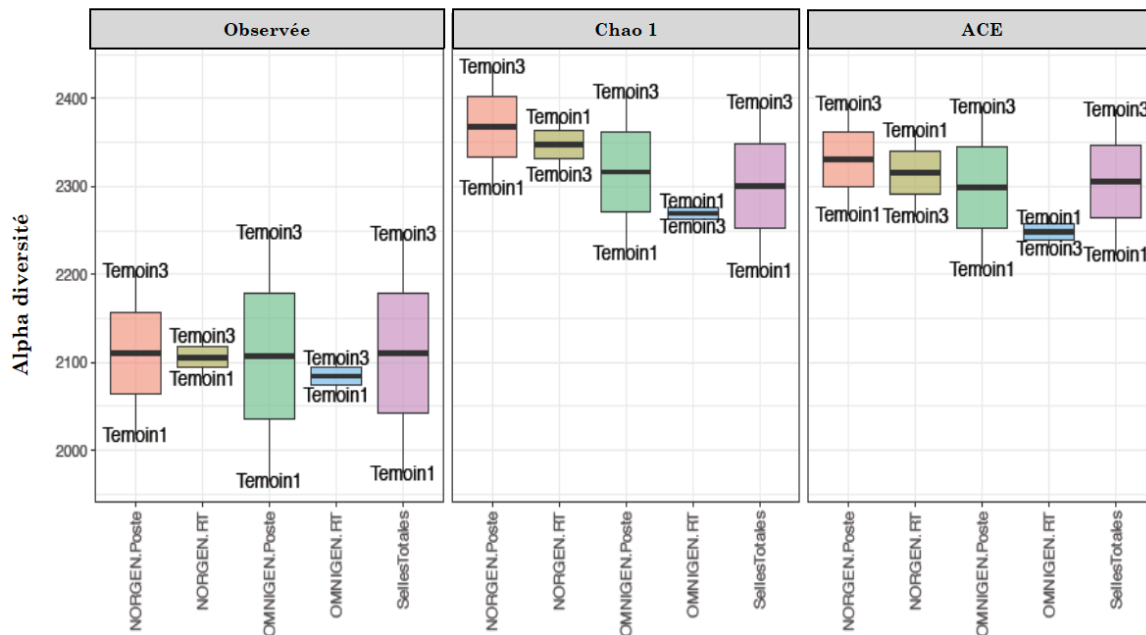


Figure 52 : Alpha diversité estimée selon Chao1 et ACE selon les kits de prélèvement avec estimation de la variabilité entre les deux sujets. Le test de Kruskal-Wallis ne montre pas de variation significative dans l'estimation de la diversité selon les kits utilisés et leurs modes d'acheminement vers le laboratoire (Observée (p -value = 1) Chao 1 (p -value = 0,7) Ace (p -value = 0,66)).

Lorsque nous regardons les 50 espèces les plus abondantes, en termes de nombre de lectures par sujet, et dans les groupes les plus représentés, nous observons la même distribution avec les différents systèmes de prélèvement (p.ex., *B. vulgatus*, *E. rectale*, *B. dorei*). Cependant, on note des différences importantes entre les deux sujets, comme la présence d'*Akkermansia muciniphila* et *Methanobrevibacter smithii* chez le témoin 3 et la présence de différentes lignées de *Bifidobacterium* chez le témoin 1.

Conclusion

Nous n'observons pas de différences significatives dans l'estimation de la diversité entre les modes de recueil et de stockage testés. A ce stade de notre expérimentation, nous éliminons le kit de prélèvement OMNIGEN GUT pour deux raisons :

- Son coût élevé : 22€ le tube.
- Ne permet de recueil que 500 mg de selle, ce qui permet, à peine de faire une extraction supplémentaire si besoin.
- La durée de conservation à température ambiante est de deux mois seulement alors qu'elle est de deux ans pour le système NORGEN.
- La difficulté de réaliser le recueil : utilisation d'une spatule pour le prélèvement qui rend difficile l'estimation de la quantité nécessaire. Difficulté de manipulation du tube par le préleveur.

B) Comparaison de différents modes de stockage des échantillons prélevés sur tube avec stabilisateur.

Sur un autre groupe de sept sujets volontaires, j'ai ensuite comparé le système NORGEN sélectionné dans l'expérimentation précédente, avec le système « DNA/RNA Shield - Fecal Collection Tube » commercialisé par OZYME. J'ai, également, testé en parallèle l'effet de la congélation à -80°C, des selles stabilisées à température ambiante. Chaque sujet a effectué 1 recueil dans un tube NORGEN et trois sujets parmi les sept, ont eu un recueil supplémentaire dans un tube « DNA/RNA Shield - Fecal Collection ».

Les prélèvements ont été remis au laboratoire. Un aliquot de 300µl a été préparé et stocké à -80°C. Le reste des échantillons a été conservé à température ambiante jusqu'au moment de l'extraction.

Résultats

L'estimation de la diversité alpha (fig. 53) observée et estimée selon Chao1 et ACE ne montre pas de différences significatives selon que les échantillons soient prélevés sur tube NORGEN ou Ozyyme ou qu'ils soient conservés à température ambiante ou stockés à -80°C avant l'extraction.

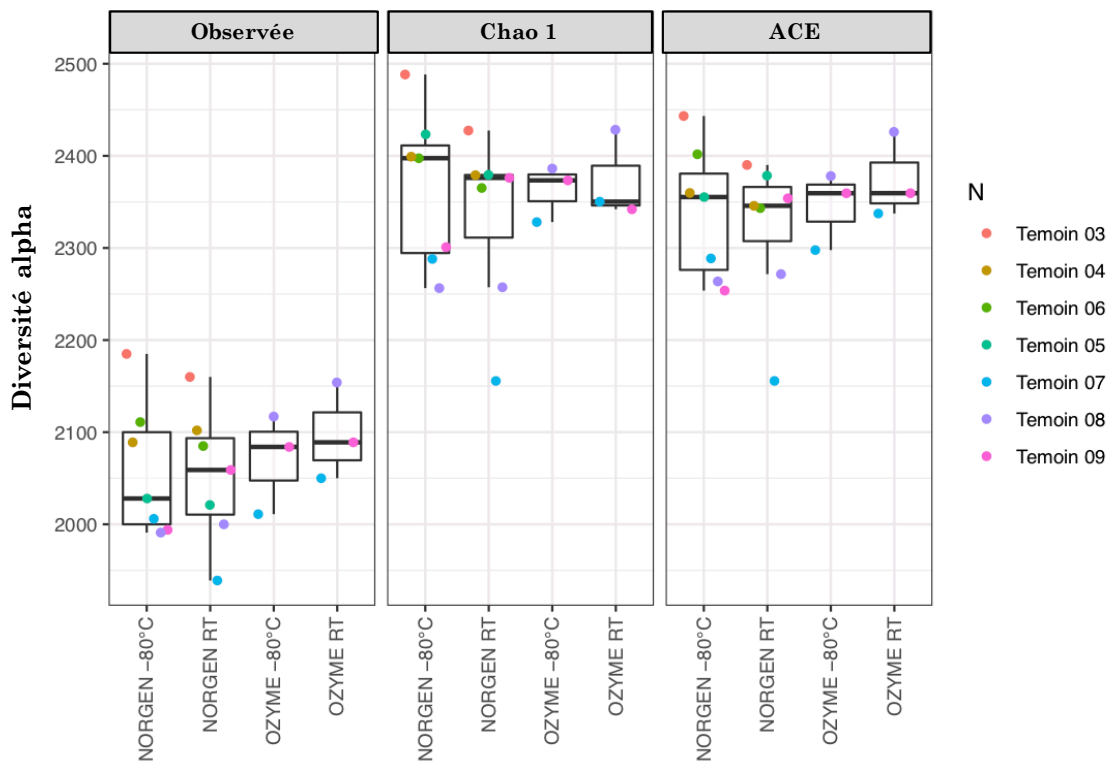


Figure 53 : Alpha diversité estimée selon Chao1 et ACE selon les kits de prélèvement et conditions de stockage testés. Selon le test Kruskal-Wallis, nous n'observons pas de différence significative dans l'estimation de la richesse entre les différentes conditions (Observée (p-value = 0,78), Chao 1 (p-value = 0,94), ACE (p-value = 0,84)).

Conclusion :

Nous choisissons donc l'utilisation du kit DNA/RNA Shield - Fecal Collection de chez OZYME car il permet une estimation de la diversité bactérienne qui permet d'établir l'état de la richesse microbienne du microbiote intestinal. En plus, le coût du tube « DNA/RNA Shield - Fecal Collection » est moins coûteux (9€ le tube) comparé au tube NORGEN (15€ le tube). Une expérimentation est en cours sur les échantillons de 7 autres volontaires sains afin de valider le résultat obtenu avec ce mode de prélèvement.

Validation du séquençage Oxford Nanopore

Comparaison séquençage « MinION » d'Oxford Nanopore et Séquençage « Solid »

La dysbiose intestinale a été identifiée dans le surpoids et l'obésité modérée¹⁰² comme en témoignent les modifications substantielles de la composition microbienne intestinale (avec enrichissement ou diminution de groupes bactériens spécifiques) et la faible richesse microbienne associée à des altérations métaboliques (insulinorésistance, inflammation de bas grade et hypertrophie adipocytaire). Cependant, les caractéristiques du microbiote intestinal avaient été peu explorées dans les formes extrêmes d'obésité (sévère : indice de masse corporelle (IMC) > 35 kg / m²) et morbide (IMC > 40 kg / m²) malgré la progression de l'obésité dans le monde. Alors que certains patients gravement obèses restent métaboliquement en bonne santé¹⁰³, en général, atteindre un IMC > 35 kg/m² induit une augmentation significative de risque de troubles chroniques¹⁰⁴.

Dans l'équipe, nous avons exploré la richesse génétique microbienne chez des patients obèses et son évolution après une intervention bariatrique avec anneau ou by pass gastrique, ainsi que sa relation avec d'autres caractéristiques métaboliques cliniques. Nous avons pour cela, utilisé le séquençage SOLiD (whole shotgun metagenomic), en collaboration avec les équipes de l'INRA (équipe MICALIS). L'objectif de cette étude était de :

- Vérifier si la richesse génétique microbienne se détériore dans l'obésité sévère et comment elle est éventuellement reliée à l'aggravation des comorbidités.
- Vérifier si différents types de chirurgie bariatrique pouvaient corriger différemment les caractéristiques microbiennes intestinales sévères liées à l'obésité en termes de composition et de fonction.

Nous avons inclus de manière prospective 61 femmes sévèrement obèses. Les patientes ont été séparées en deux groupes, anneau (AG) ou by pass (BP) gastriques selon les directives internationales pour la chirurgie bariatrique (BP : IMC ≥ 40 kg/m² ou AG : IMC ≥ 35 kg/m²), avec au moins une complication de l'obésité sévère. Des évaluations cliniques, anthropométriques et biologiques ont été obtenues avant intervention (T0) et à 1 mois (T1), 3 mois (T3) et 12 mois (T12) pendant le suivi post-opératoire. Les fèces ont été collectées à chaque visite. Les échantillons ont été soit stockés immédiatement à -80 °C lorsque le recueil a été réalisé pendant l'hospitalisation, soit conservés brièvement dans des congélateurs domestiques, dans des conditions anaérobies, avant d'être

transportés au laboratoire où ils ont été immédiatement congelés à -80°C . Aucun patient n'avait reçu de traitement antibiotique durant les 3 mois précédents la chirurgie bariatrique, ni n'avait eu d'antécédents de maladies gastro-intestinales aiguës ou chroniques.

L'ADN fécal total des échantillons a été extrait, séquencé et analysé à l'INRA. Une moyenne de 68,72 millions de lectures uniques de 35 bases de long (SD 26M) a été obtenue pour les échantillons. L'analyse primaire, de la qualité de lecture à l'alignement des lectures sur un catalogue de référence de 3,9 millions de gènes a été réalisée à l'aide de Meteor Studio. Les analyses secondaires, de la normalisation de l'abondance des gènes à l'établissement de la richesse microbienne et des analyses statistiques ont été réalisées à l'aide du logiciel « momr R ».

Dans ces travaux, une distribution bimodale de la richesse microbienne a été obtenue. En utilisant la même méthodologie et le même seuil génétique (480 000 gènes) que celui utilisé dans la publication de « Le Chatelier & al., Nature 2013 » on distingue deux groupes de sujets : un groupe avec une faible richesse génétique (LGC) et un groupe avec une richesse génétique plus élevée (HGC). On constate, également, que la grande majorité des patientes (75%) appartenaient au groupe LGC (figure 54).

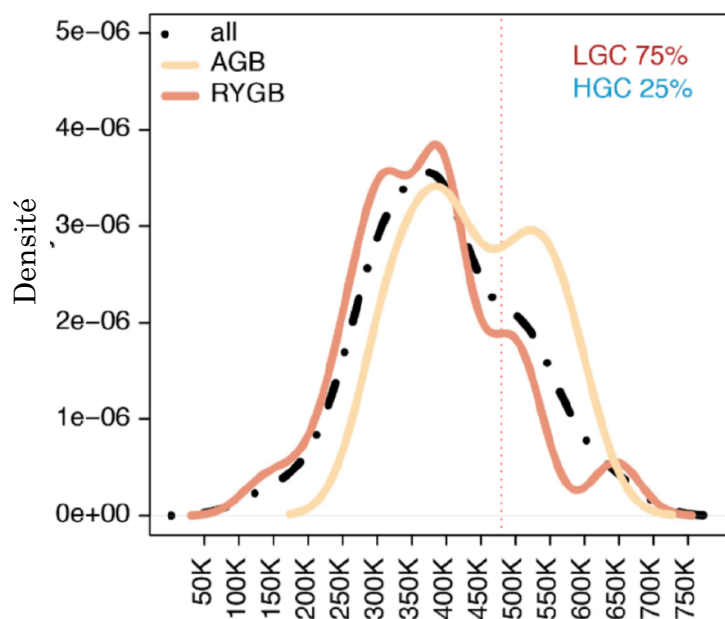


Figure 54 : Distribution basale de la richesse microbienne dans la cohorte « Microbaria » (N=61 sujets), Aron & al., Gut, 2018. Sur l'ensemble de la cohorte étudiée, on relève 75% de sujets LGC et 25% de sujets HGC. Dans le groupe AGB (gastric banding ou anneau gastrique), on dénombre 60% de patientes LGC alors que dans le groupe RYGB (bypass) ce taux est de 83%.

Nous avons donc utilisé cette étude pour laquelle nous avons déjà des données de séquençage obtenues par SOLiD pour valider le séquençage mis au point au laboratoire avec la technologie d'Oxford Nanopore. Pour cela, parmi cette série de sujets, nous avons sélectionné dans la cohorte « Microbaria » 15 patientes HGC et 15 patientes LGC en se basant sur la distribution de la richesse génique. L'extraction de l'ADN bactérien a été effectuée sur les mêmes prélèvements de selles qui ont servis pour l'étude « Microbaria », avec le kit « PureLink Microbiome DNA Purification » supplémenté des recommandations IHMS. La préparation de la librairie de séquençage a été réalisée avec le protocole amélioré d'Oxford Nanopore technologies présenté supra. Le séquençage a été réalisé grâce à l'outil « MinION » d'ONT.

Résultat

L'analyse de la diversité alpha (espèces observées, Chao1, ACE) (fig. 55) nous permet de distinguer deux groupes distincts de patientes qui se regroupent en fonction de leur diversité en espèces bactériennes. La différence entre les deux groupes est plus prononcée avec la diversité alpha observée qui tient compte du nombre d'espèces présentes dans l'échantillon, ce qui la rend équivalente à la richesse génétique microbienne, qui rend compte de la variété à l'intérieur de l'espèce. C'est ce qui a été utilisée dans le cadre de l'étude « Microbaria » pour estimer la richesse génétique bactérienne.

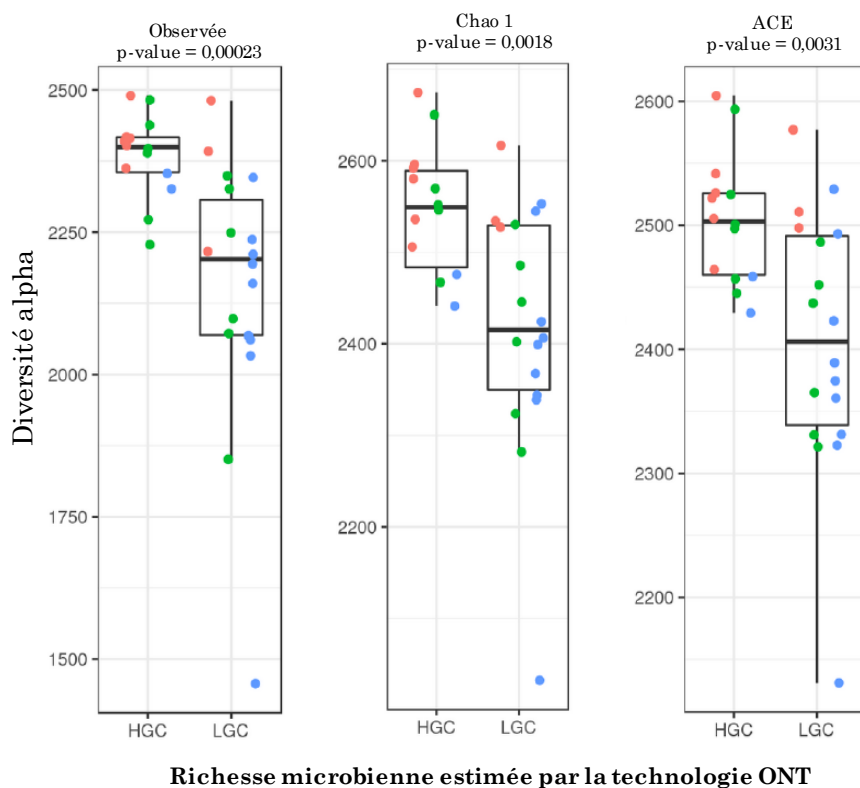


Figure 55 : Diversité alpha des deux groupes de chirurgie bariatrique. Chaque point représente un individu en fonction de la date du séquençage (rouge : expérience 1, vert : expérience 2, bleu : expérience 3) HGC= High Gene Count, LGC = Low Gene count.

Afin de valider davantage ces résultats, nous avons comparé les indices de diversité alpha calculés à partir du séquençage Nanopore (espèces observées, Chao1, ACE) (fig. 55) avec les valeurs de richesse génétique du papier de l'étude « Microbaria ». Pour ce faire, nous avons fusionné les données de la diversité alpha obtenues par le séquençage MinION d'ONT et celles obtenues par le séquençage SOLiD pour les patientes de la sélection.

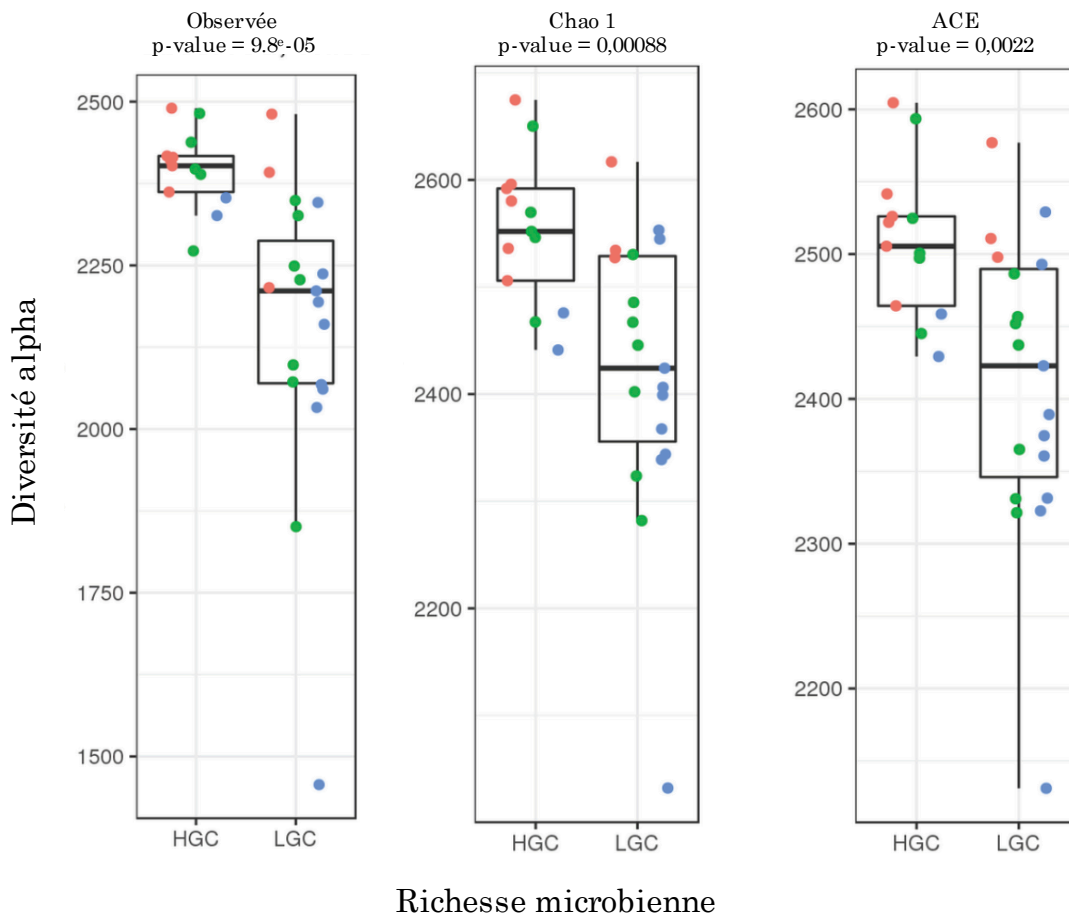


Figure 56 : Diversité alpha des deux groupes de chirurgie bariatrique. Chaque point représente un individu en fonction de la date d'expérimentation (rouge : expérience 1, vert : expérience 2, bleu : expérience 3). Différence significative de diversité alpha entre les groupes HGC (diversité génétique élevée) et LGC (diversité génétique faible), attestée par le test de Wilcoxon.

Nous constatons donc, que nous retrouvons toujours la séparation des patientes en deux groupes, HGC et LGC et nous reproduisons les mêmes résultats que précédemment en augmentant les niveaux de significativité.

Nous avons, par la suite, corrélé les valeurs de la diversité alpha obtenue par séquençage Nanopore avec les valeurs de la richesse génétique obtenue par séquençage SOLID (fig. 57) et nous observons une corrélation positive et significative, plus forte encore avec la diversité alpha observée (équivalente de la richesse génétique).

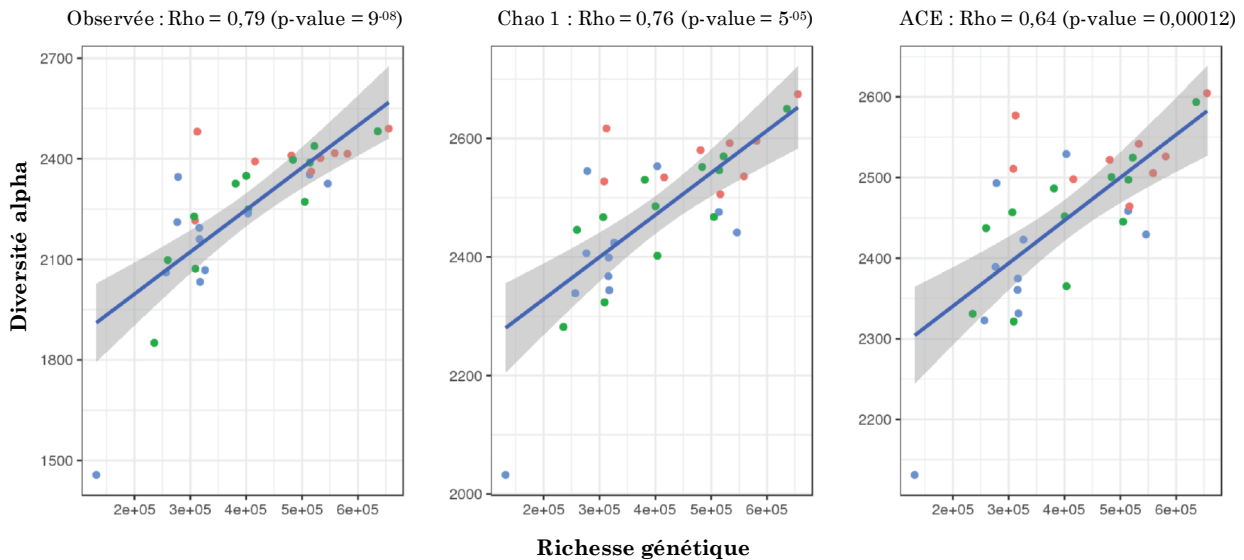


Figure 57 : *Corrélation positive entre la diversité alpha et la richesse génétique (test statistique de Spearman). Chaque point représente un individu en fonction de la date d'expérimentation (rouge : expérience 1, vert : expérience 2, bleu : expérience 3).*

Conclusion :

Nous constatons dans cette série de comparaison que le séquençage d'ONT nous permet d'obtenir la même séparation des patientes en deux groupes, HGC et LGC et donc, que le séquençage ONT nous permet de déterminer le statut des patients en termes de richesse microbienne intestinale au même titre que le séquençage SOLID.

Discussion et perspectives

Discussion

Dans ce travail, nous avons mis en place un protocole de séquençage « maison » basé sur l'utilisation de la technologie d'Oxford Nanopore grâce à l'utilisation de l'outil MinION. Nous avons travaillé sur plusieurs étapes critiques.

Les études du microbiote intestinal chez nos sujets obèses ont, jusqu'alors étaient réalisées en collaboration où les extractions et le séquençage ont été gérés ailleurs que dans notre laboratoire. Le développement du séquençage dans nos locaux, a nécessité le choix et la mise au point d'un protocole d'extraction de l'ADN bactérien à partir des selles humaines qui permet d'obtenir un matériel en bonne quantité et de bonne qualité. Dans la littérature, plusieurs études au cours des cinq dernières années ont utilisé l'ADN ou l'ARN bactérien pour explorer les communautés microbiennes dans des écosystèmes divers et variés allant jusqu'au corps humain. Le microbiote intestinal est particulièrement intéressant en raison de son abondance, sa grande diversité et sa pertinence pour la santé humaine et les maladies. Les études dans ces domaines ont presque exclusivement utilisé leur propre protocole d'extraction d'ADN et une cohorte démographiquement bien déterminée basée sur un questionnaire bien précis. Des techniques différentes de séquençage ont été utilisées (Illumina, Solid). De nombreuses études ont pointé les effets de « batch »¹⁰⁵ et les différences dans l'analyse des données générées en fonction des protocoles utilisés¹⁰⁶⁻¹⁰⁷, ce qui introduit un biais dans les comparaisons analytiques de ces analyses entre elles.

Nous nous sommes donc intéressés à l'étude d'IHMS dans laquelle les équipes scientifiques ont comparé 21 méthodes d'extraction d'ADN en utilisant le séquençage métagénomique « Shot Gun » par la technologie Illumina HiSeq2000. Ils ont évalué le profil taxonomique et la variabilité fonctionnelle tout en standardisant les étapes de collecte des selles, de stabilisation de l'ADN bactérien, de préparation de la librairie et enfin, le séquençage. Il en a résulté des « Standard Operating Procedure » « SOPs » comportant des recommandations qui permettraient d'améliorer l'extraction de l'ADN en termes de rendement (>150ng/μl) et de qualité (Rapports 260/280 >1,8) et 260/230 > 1,8). Les principaux conseils d'amélioration concernent :

- La lyse de la paroi des bactéries. En effet le protocole conseille de réaliser une première fois, une lyse chimique suivie d'une lyse mécanique, de centrifuger les tubes et de récupérer le surnageant contenant les bactéries dont la paroi a été cassée. Sur le culot restant, on procède à un deuxième tour de lyses chimique et mécanique, on centrifuge les tubes et on récupère le

surnageant contenant l'ADN des bactéries qui avaient résisté au premier tour de lyse. C'est cela qui permet d'augmenter le rendement de l'extraction.

- L'élimination des protéines et des ARN. Le consortium propose un premier traitement des deux surnageant récupérés après les deux tours de lyses chimique et mécanique, par de l'acétate d'ammonium qui permet de précipiter les protéines. S'ensuit une précipitation de l'ADN à l'isopropanol. L'ADN ainsi précipité est traité d'abord à la RNase et ensuite à la protéinase K. Cette étape contribue à une meilleure élimination des protéines et des ARN comme peut en témoigner les rapports 260/280 et 260/230.

Ce sont ces recommandations que nous avons appliqué au kit « PureLink™ Microbiome DNA Purification » de chez Invitrogen et nous avons pu améliorer et optimiser notre protocole d'extraction. En effet, nous obtenons un rendement moyen de $300,15 \pm 0,75$ ng/μl contre $74,5 \pm 0,48$ ng/μl (p-value<0,0001) avec le protocole classique proposé par Invitrogen. Le rapport 260/280 qui témoigne de la contamination en protéines passe de $1,79 \pm 0,03$ à $1,89 \pm 0,02$ (p-value = 0,005). Le rapport 260/230 quant à lui passe de $1,31 \pm 0,05$ à $1,87 \pm 0,04$, ce qui est en faveur d'une diminution de la contamination par des solvants.

Pour une autre étape critique, nous avons voulu explorer d'autres méthodes de stabilisation de l'ADN pour deux raisons :

- 1) faciliter et harmoniser les conditions de prélèvement. Dans de nombreux cas, y compris en milieu clinique, les prélèvements fécaux peuvent être prélevés à partir de la selle totale, puis immédiatement congelés à -20°C afin de prévenir les changements dans la communauté microbienne jusqu'à ce que l'extraction de l'ADN puisse être réalisée. Cependant, les échantillons prélevés à domicile nécessitent un transport avec un accès continu à une température stable à -20°C . Cela peut s'avérer compliqué pour certains sujets des études. Même dans les cas où un congélateur domestique est disponible, les échantillons peuvent être soumis à des variations de température pendant le stockage (les congélateurs domestiques sont soumis à des cycles de dégivrage automatique) ou pendant le transport ce qui pourrait provoquer la décongélation des échantillons. Ce mode de prélèvement pose l'inconvénient de maintenir la selle en contact avec l'oxygène plus ou moins longtemps ce qui est délétère pour les bactéries anaérobies.
- 2) Faciliter la gestion des prélèvements au niveau du laboratoire. En effet, la traçabilité des échantillons prélevés à domicile est très difficile car nous ne pouvons pas contrôler les étapes qui sépare le recueil de la réception du prélèvement et son stockage à -80°C au laboratoire.

- Nous avons par la suite amélioré le protocole de séquençage, en améliorant la préparation de la librairie. En effet, la préparation proposée par ONT, nous a permis d'obtenir une courbe bimodale de distribution des lectures de séquençage que nous avons divisée en deux parties séparant une zone de lectures courtes ($\leq 1,1\text{Kb}$) et une zone de lectures longues ($>1,1\text{Kb}$). Comme nous avons choisi le critère de longueur des lectures pour sélectionner le kit d'extraction d'ADN, nous avons essayé d'améliorer le protocole de préparation de la librairie afin d'augmenter la proportion des lectures longues obtenue par le séquençage ONT. Ce critère a été sélectionné pour un aspect pratique au moment du « mapping » de nos lectures sur les séquences de références du catalogue. Oxford Nanopore propose un forum aux utilisateurs de cette technologie. C'est un lieu d'échange de compétence et d'expérience. En explorant ce forum, j'ai pu profiter de l'expérience et recommandations de certains utilisateurs. J'ai pu, grâce à cela, à améliorer la longueur des lectures obtenues par la technologie de séquençage d'ONT.
- Nous avons ensuite voulu comparer nos données de séquençage MinION à d'autres méthodes de séquençage SOLID et Illumina en utilisant les mêmes échantillons. L'objectif étant de vérifier si le séquençage d'ONT permet de classer correctement les patients en fonction de leurs richesses microbiennes.
- Une première comparaison des résultats obtenus avec le séquençage ONT a été réalisée avec ceux obtenus avec le séquençage SOLID sur 30 échantillons de selles humaines de la cohorte de l'étude « Microbaria ». Le séquençage SOLID avait généré une moyenne de 68,72 millions de lectures uniques de 35 bases. Avec le MinION, nous obtenons en moyenne 2 millions de lectures de 200pb à 24 Kb. Le résultat obtenu ne montrait pas de différence d'estimation de la diversité bactérienne entre les deux méthodes de séquençage dans la sous cohorte étudiée. En effet, nous obtenons, avec le séquençage MinION la même répartition des sujets dans les groupes « faible richesse bactérienne (LGC) » et « richesse bactérienne élevée (HGC) ».

Cette première comparaison montre que les résultats obtenus avec le séquençage MinION, nous permettent d'estimer la diversité bactérienne du microbiote intestinal de nos patients et d'établir un diagnostic quant à leurs appartenances au groupe LGC ou au groupe HGC. Ce résultat nous conforte donc par rapport à la question que nous nous posons en premier lieu pour établir un test diagnostique pour les sujets obèses suivis dans le service de Nutrition de la Pitié Salpêtrière et qui est de connaître l'état de richesse de leur microbiote.

Une deuxième expérimentation de validation du séquençage d'ONT est prévue pour comparer, cette fois, les données obtenues par de séquençage Illumina d'une seconde cohorte de patients obèses avec les données de séquençage ONT qui sera réalisé sur un sous-groupe de cette cohorte recrutée dans le cadre d'une étude Européenne.

La technologie MiniION présente des avantages mais également quelques limites.

A) - Avantages

Tout au long des différentes expérimentations effectuées à ce jour, nous avons pu constater que le séquençage MinION nous permettait de visualiser un profil taxonomique regroupant les groupes bactériens les plus communs du microbiote intestinal. Ainsi on retrouve les principaux ordres bactériens du microbiote intestinal : les Bactéroidetes et les Firmicutes qui représentent la majorité du microbiome humain ainsi que les Actinobactéries et les Proteobactéries. On identifie, également, les principales espèces habituellement retrouvées tels que les Bacteroides, *Escherichia. coli* et *Akkermansia muciniphila* (espèce plus rare). Des proportions similaires ont été rapportées à partir de données générées par séquençage 16S Illumina et Pyroséquençage 154¹⁰⁸.



Figure 58 : Profil taxonomique obtenu par le séquençage MinION montrant les principaux groupes bactériens communément présents dans la flore bactérienne intestinale. Profil obtenu à partir des séquences générées des échantillons de deux sujets sains et sur 6 prélèvements différents pour chacun. « more » indique la présence d'autres espèces présent à un faible pourcentage qu'on a choisi de ne pas représenter sur le profil.

Le séquençage MinION permettant d'obtenir des lectures de longueurs supérieures à 1,1kb, nous avons pu après quelques expérimentations réaliser des assemblages « de novo ». Effectivement, la longueur des lectures obtenues par séquençage d'Oxford Nanopore Technologies ont permis, par exemple, d'assembler le génome d'*Akkermansia muciniphila* en sept contigs seulement. Dans une étude réalisée en 2015, l'assemblage du génome d'*Akkermansia muciniphilla* a nécessité l'utilisation du séquençage « Solid » et du séquençage « Shotgun » pour couvrir le génome de la bactérie avec un scaffold et 56 gaps¹⁰⁹.

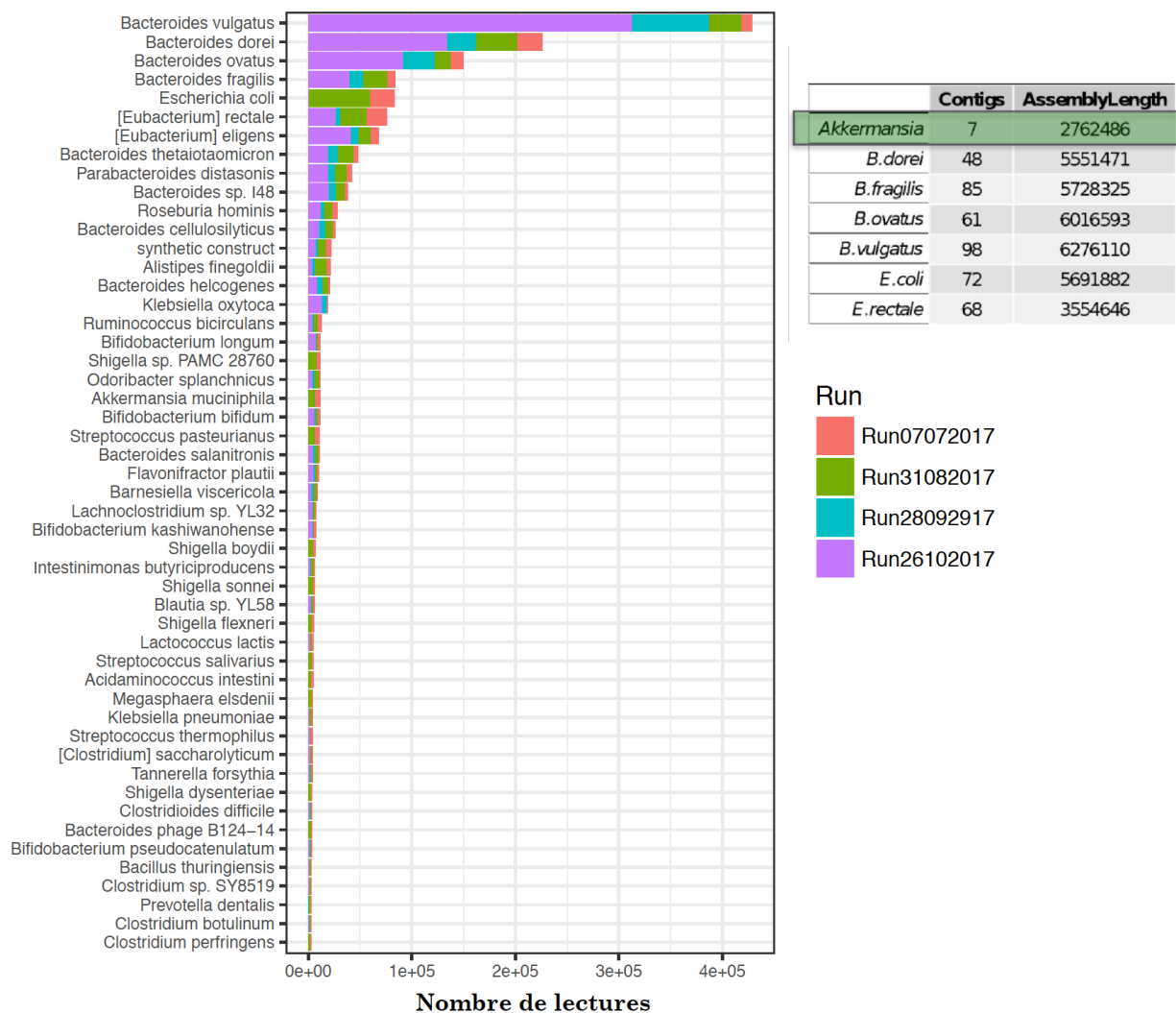


Figure 59 : Quelques exemples d'assemblages « de novo » réalisés à partie des données collectées de 4 expériences différentes permettant la reconstitution certains génomes tel que le génome d'*Akkemansia municiphila* en seulement 7 contigs.

L'analyse par composante principale (PCoA) par rapport à la matrice de distance de diversité Bray-Curtis intégrant l'ensemble des données des mêmes échantillons provenant de différentes analyses, montre une séparation nette des échantillons par sujet le long du premier axe du tracé PCA. Ceci est en faveur de l'absence d'un effet « batch » ou regroupement des données par date de séquençage, dans les résultats obtenus avec la technologie ONT.

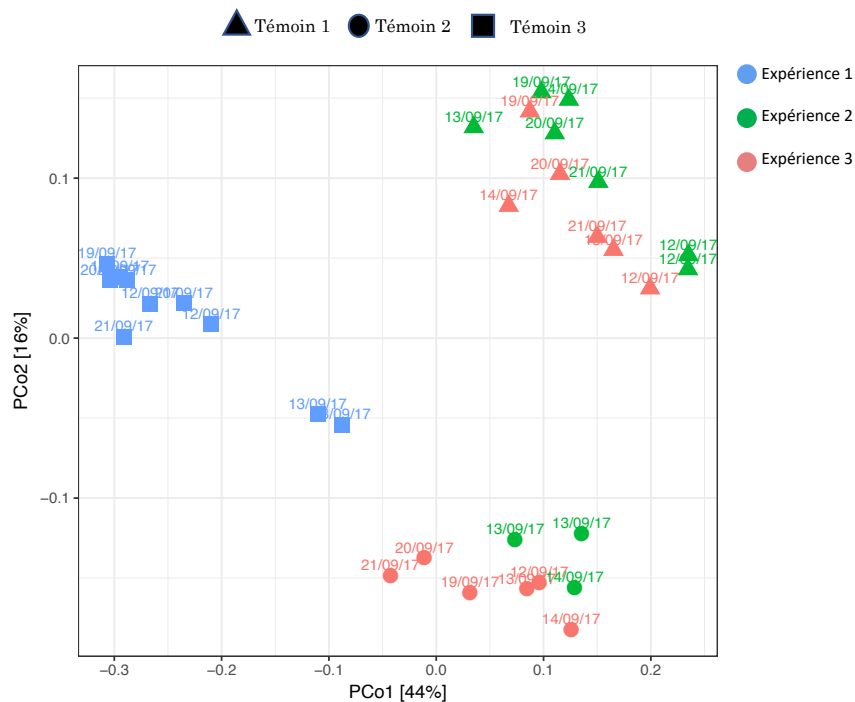


Figure 60 : Répartition de différents échantillons séquencés selon l'analyse par composante principale ne présentant pas d'effet « batch » mais séparant les échantillons par donneur et non par la date d'expérience.

B) -Limites

- nous obtenu un pourcentage assez élevé, environ 25% de lectures qui reste non classifiées. Ceci correspond essentiellement à :

- la perte du code barre par le fragment d'ADN qui est séquencé mais ne peut être identifié lors du démultiplexage et par conséquent ne peut être reconnu et attribué à un échantillon. Nous avons constaté que cette fraction comportait beaucoup de lectures courtes. Probablement, et en raison de leurs longueurs courtes, ces fragments d'ADN ne sont pas correctement marqués ce qui rend leur identification difficile.

- la détection de séquences d'ADN que nous ne pouvons pas aligner sur les séquences de références du catalogue.

Ceci a pour conséquence une perte d'information qui peut impacter l'estimation de la diversité bactérienne intestinale.

- La profondeur de séquençage d'Oxford Nanopore technologies reste assez réduite. Ce qui pourrait être limitant si on souhaite réaliser des assemblages « de novo » de nouvelles espèces ou dans le cas de détection de mutation. Dans ce cas, un recours complémentaire à une autre technologie telle que le séquençage Illumina pourrait s'avérer nécessaire. Cependant et dans le cas de notre étude, cet inconvénient n'est pas vraiment un problème du fait que nous souhaitons avoir une image générale de la diversité bactérienne du microbiote intestinal, et connaître la présence ou l'absence de séquences bactériennes fréquentes. De plus les séquences longues fournies par le séquençage MinION permettent une couverture assez grande du génome de référence ce qui nous permet d'apporter une réponse assez juste à notre question. Par ailleurs, ONT a de nouveaux développements qui permettraient d'augmenter la profondeur du séquençage

Un autre point faible du MinIon est son taux d'erreurs, qui est actuellement de l'ordre de 5 % pour une seule lecture (une erreur toutes les 20 bases environ). Ceci est directement lié au mode de lecture, qui ne procède pas base par base comme on pourrait l'imaginer en raison du diamètre du nanopore. En effet les pores biologiques actuellement utilisés ont une dimension telle que la longueur d'ADN présente à l'intérieur du pore correspond à cinq ou six bases : le signal observé est donc caractéristique d'une suite de bases et non d'une seule d'entre elles. La lecture se fait par comparaison informatique entre le courant observé et des courants de référence établis pour tous les hexamères possibles. Ces problèmes sont en partie résolus grâce à la redondance des lectures mais aussi par la possibilité de séquencer successivement les deux brins d'un ADN et d'améliorer ainsi la qualité du résultat en diminuant le taux d'erreur.

Conclusion

Oxford Nanopore Technologies propose aujourd'hui une solution évidente dans les applications où la facilité de mise en œuvre (portabilité) et l'absence d'investissement sont déterminantes. Elle peut aussi, comme Pacific Biosciences, être complémentaire d'Illumina dans des projets de séquençage de novo où le gros des données est fourni par le séquençage de petits fragments, mais où l'organisation générale du génome étudié repose sur l'obtention de longues séquences. On peut enfin imaginer que des progrès au niveau de la nature et de la structure des pores, tel que le passage aux nanopores artificiels dont on peut contrôler le diamètre, permettrait à l'avenir une véritable lecture base par base, ce qui autoriserait sans doute une nette baisse du taux d'erreurs et mettrait alors ONT au même niveau de qualité qu'Illumina.

Perspectives

Le travail réalisé dans le cadre de mon diplôme EPHE constitue la première étape mais aussi la base d'un projet de thèse de plus grande envergure qui est de mettre au point un test diagnostique pour détecter les patients obèses morbides avec une atteinte du microbiote intestinal : dysbiose, faible richesse bactérienne et altération de la composition microbienne.

Deux observations importantes de notre équipe sont à la base de ce projet et des applications cliniques que nous pourrions développer dans le laboratoire :

- d'une part et au regard des différentes études déjà publiées par notre équipe, nous savons que les patients avec une faible richesse microbienne ont un IMC plus élevé et des complications métaboliques plus sévères que ceux avec une richesse préservée.

- d'autre part, aujourd'hui la chirurgie bariatrique et plus particulièrement le by pass gastrique constitue la solution la plus efficace non seulement pour la perte de poids mais aussi pour la résolution des complications métaboliques dont souffrent les patients obèses. Or, cette chirurgie n'est pas dénuée de risques et de complications tels que : i) le « Dumping syndrome » un malaise qui survient après ingestion de matière dense en sucre et qui peut évoluer en épisode d'évanouissement si la quantité de sucre ingérée est trop élevée, ii) des carences vitaminiques, en raison du montage chirurgical qui court-circuite les 75 premiers centimètres de l'intestin grêle, siège de l'absorption des vitamines au niveau de l'intestin. Ceci nécessite une supplémentation à vie en vitamines avec une surveillance médicale régulière, et bien sûr iii) des risques inhérents à la chirurgie chez des patients fragiles. Enfin tous les patients n'ont pas d'amélioration métaboliques post-chirurgicales.

Ainsi, en établissant un test diagnostique nous serons en mesure d'identifier les patients dont la richesse microbienne est faible. Ces patients se verront proposer un accompagnement diététique et thérapeutique spécifiques à leur pathologie pour les aider à rétablir l'équilibre de leur flore intestinale éventuellement avant l'intervention, ou bien au décours. Pour cela, je vais réaliser le séquençage du microbiote intestinal de 100 patients obèses morbides du service de nutrition de la Pitié Salpêtrière. Ce sont des patients qui ont déjà été opérés d'un « by pass gastrique » et pour lesquels nous avons le suivi clinique complet de leur évolution en termes de poids et de complications métaboliques. Le résultat du séquençage permettra de classer les patients soit dans le groupe LGC soit dans le groupe HGC. L'analyse des paramètres cliniques post chirurgicaux, nous permettrait de vérifier notre hypothèse qui est que les patient LGC répondent moins bien à la chirurgie bariatrique en ne résolvant

que partiellement leurs complications métaboliques. Une fois cette hypothèse validée, nous nous intéresserons à un deuxième objectif de cette étude et qui est d'identifier une signature bactérienne de la dysbiose. En effet, les précédentes études de l'équipe ont permis de mettre en évidence certaines espèces de bactéries retrouvées de façon régulière chez des sujets à faible richesse microbienne. Ces espèces pourraient être des signatures d'une flore intestinale peu diversifiée. Pour cela un travail de validation de ces candidats sera réalisé afin de sélectionner les espèces les plus pertinentes en tant que marqueurs et qui pourront, par la suite, servir pour un diagnostic encore plus rapide de la dysbiose chez le sujet obèse morbide. Ceci pourrait être un premier pas vers la médecine de précision qui est au cœur de beaucoup de programme de recherche.

Aujourd'hui ce protocole mis au point, de l'extraction de l'ADN microbien à son séquençage par le MinION d'Oxford Nanopore Technologies, est utilisé dans d'autres projets de recherche de notre équipe et des projets de collaborations nationale et internationale impliquant du séquençage du métagénome humain ou murin.

Bibliographie

- ¹ Basdevant A, Guy-Grand B. Chapitre 1 : Définition et classification des obésités. Médecine de l'obésité. Paris : Médecine-Sciences Flammarion ; 2004. p. 3-7.
- ² Etude de santé sur l'environnement, la biosurveillance, l'activité physique et la nutrition (Esteban 2014-2016) - Volet nutrition - Chapitre corpulence.
- ³ Tremblay A, Chaput JP, About unsuspected potential determinants of obesity, *Appl Physiol Nutr Metab*, vol. 33, n° 4, août 2008.
- ⁴ C.Lubrano-Berthelier, K. Clément, Génétique de l'obésité humaine. *Rev Med Interne*. 2005 Oct ;26.
- ⁵ Pauline Faucher et Christine Poitou, Physiopathologie de l'obésité. *Revue du Rhumatisme Monographies* Volume 83, Issue 1, February 2016, Pages 6-12.
- ⁶ D'après AFERO, ALFEDIAM, SNDLF. Recommandations pour le diagnostic, la prévention et le traitement de l'obésité. *Diabète & Métabolisme*, Suppl. 2 : 1-48, 1998.
- ⁷ Berrington de Gonzalez & al, Body-mass index and mortality among 1.46 million white adults. *N Engl J Med*. 2010.
- ⁸ Guh DP1, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH, The incidence of comorbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health*. 2009.
- ⁹ Rössner S1, Lagerstrand L, Persson HE, Sachs C, The sleep apnoea syndrome in obesity: risk of sudden death. *J Intern Med*. 1991.
- ¹⁰ Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), National Heart, Lung, and Blood Institute, National Institutes of Health. National Cholesterol Education Program, mai 2001.
- ¹¹ Cornier MA&al., The metabolic syndrome. *Endocr Rev*. 2008.
- ¹² Basdevant A &al., *Traité médecine et chirurgie de l'obésité*. Médecine sciences 2011.
- ¹³ M.Merrouche, B.Coffin, Obésité : prises-en charge, indications et méthodes du traitement endoscopique et chirurgical. *EMC - Hépatogastroentérologie* Volume 2, Issue 3, July 2005, Pages 189-200.
- ¹⁴ E. Mason, Vertical banded gastroplasty for obesity. *Arch Surg*, 117 (1982), pp. 701-706
- ¹⁵ M.Merrouche & B.Coffin, Obesity: management, indications and methods for endoscopic and surgical treatment. *EMC - Hépatogastroentérologie*, Volume 2, July 2005.
- ¹⁶ W.J. Pories, M.S. Swanson, K.G. Mac Donald, S.B. Long, P.G. Morris, B.M. Brown, Who would have thought it? An operation proves the most effective therapy for adult-onset diabetes mellitus. *Ann Surg*, 222 (1995).
- ¹⁷ <https://regimeminceur.info/sleeve-gastrectomie/>.

-
- ¹⁸ F. Backhed, R.E. Ley, J.L. Sonnenburg, et al., Host-bacterial mutualism in the human intestine *Science*, 307 (2005).
- ¹⁹ M. Diamant, E.E. Blaak, W.M. de Vos, Do nutrient-gut-microbiota interactions play a role in human obesity, insulin resistance and type 2 diabetes ? *Obes Rev.* 2011.
- ²⁰ V.K. Ridaura, J.J. Faith, F.E. Rey, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice *Science*, 341 (2013).
- ²¹ R. Burcelin, M. Serino, C. Chabo, et al. Gut microbiota and diabetes: from pathogenesis to therapeutic perspective *Acta Diabetol*, 48 (2011).
- ²² Aron-Wisnewsky J, & al., Major microbiota dysbiosis in severe obesity: fate after bariatric surgery. *Gut*. 2018.
- ²³ Cotillard A. Dietary intervention impact on gut microbial gene richness. *Nature* 500, 585–588 (29 August 2013).
- ²⁴ Vrieze A, Van NE, Holleman F, et al Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* 143 (2012).
- ²⁵ Ravel et al., Vaginal microbiome of reproductive-age women. *Proc. Natl. Acad. Sci. U.S.A.* 108 (suppl. 1), 4680 (2011).
- ²⁶ E. Jiménez et al., Is meconium from healthy newborns actually sterile? *Res. Microbiol.* 159, 187 (2008).
- ²⁷ D. Mariat et al., The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol.* 9, 123 (2009).
- ²⁸ H. Gu et al., 1H NMR metabolomics study of age profiling in children. *NMR Biomed.* 22, 826 (2009).
- ²⁹ Y. Qi et al., Urinary metabolite markers of precocious puberty. *Mol. Cell. Proteomics* 11, M111 .011072 (2012).
- ³⁰ Pettersson S., Host-Gut Microbiota Metabolic Interactions. *Science Rev* 336. 2012 June 8.
- ³¹ Aron-Wisnewsky J. Microbiote intestinale : rôle dans l'obésité et les maladies métaboliques. *Feuillets de Biologie*. Vol LV N°321 – Nov. 2014.
- ³² Qin J & al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010.
- ³³ Cheng J., Palva A.M., de Vos W.M., Satokari R., 2013. Contribution of the intestinal microbiota to human health: from birth to 100 years of age. *Curr Top Microbiol Immunol* 358, 323-346.
- ³⁴ Huttenhower, C.; Gevers, D.; Knight, R.; Abubucker, S.; Badger, J.H.; Chinwalla, A.T.; Creasy, H.H.; Earl, A.M.; Fitzgerald, M.G.; Fulton, R.S.; et al. Structure, function and diversity of the healthy human microbiome. *Nature* 2012.

- ³⁵ Huse, S.M.; Ye, Y.; Zhou, Y.; Fodor, A.A. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One* 2012.
- ³⁶ Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R, Jarrin, C., Chardon, P., Marteau, P., et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006.
- ³⁷ Gérard Corthier, *Bonnes bactéries et bonne santé*, Éditions Quae, 2010, p. 7.
- ³⁸ Eric A. Franzosa, Identifying personal microbiomes using metagenomic codes, E2930–E2938, doi: 10.1073/pnas.
- ³⁹ Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* 307, 1915–1920 (2005).
- ⁴⁰ Hooper, L. V., Midtvedt, T. & Gordon, J. I. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* 22, 283–307 (2002).
- ⁴¹ Duncan, S.H.; Louis, P.; Thomson, J.M.; Flint, H.J. The role of pH in determining the species composition of the human colonic microbiota. *Environ. Microbiol.* 2009.
- ⁴² Cummings, J.H.; Macfarlane, G.T. The control and consequences of bacterial fermentation in the human colon. *J. Appl. Bacteriol.* 1991.
- ⁴³ Donohoe, D.R.; Garge, N.; Zhang, X.; Sun, W.; O'Connell, T.M.; Bunker, M.K.; Bultman, S.J. The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell. Metab.* 2011.
- ⁴⁴ Trompette, A.; Gollwitzer, E.S.; Yadava, K.; Sichelstiel, A.K.; Sprenger, N.; Ngom-Bru, C.; Blanchard, C.; Junt, T.; Nicod, L.P.; Harries, N.L.; et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat. Med.* 2014.
- ⁴⁵ Nicholson, J.K.; Holmes, E.; Kinross, J.; Burcelin, R.; Gibson, G.; Jia, W.; Pettersson, S. Host-gut microbiota metabolic interact. *Science.* 2012.
- ⁴⁶ Trent, M.S.; Stead, C.M.; Tran, A.X.; Hankins, J.V. Diversity of endotoxin and its impact on pathogenesis. *J. Endotoxin Res.* 2006.
- ⁴⁷ Cantarel, B.L.; Lombard, V.; Henrissat, B. Complex carbohydrate utilization by the healthy human microbiome. *PLoS One* 2012.
- ⁴⁸ Xu, J.; Bjursell, M.K.; Himrod, J.; Deng, S.; Carmichael, L.K.; Chiang, H.C.; Hooper, L.V.; Gordon, J.I. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 2003.
- ⁴⁹ Sandberg, A.S.; Andlid, T. Phytogetic and microbial phytases in human nutrition. *Int. J. Food Sci. Technol.* 2002.
- ⁵⁰ de Vos WM1, de Vos EA., Role of the intestinal microbiome in health and disease: from correlation to causation. *Nutr Rev.* 2012.

- ⁵¹ Suau A, Bonnet R, Sutren M, Godon JJ, Gibson GR, Collins MD, et al. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol* 1999.
- ⁵² P. Lepage, Le microbiote intestinal humain : interactions avec l'hôte et dysfonctions. *Feuillets de Biologie* N° 325 – Mars 2015.
- ⁵³ Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb J-F, Dougherty BA, Merrick JM. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. 1995.
- ⁵⁴ Binnewies TT, Motro Y, Hallin PF, Lund O, Dunn D, La T, Hampson DJ, Bellgard M, Wassenaar TM, Ussery DW, Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct Integr Genomics*. 2006 Jul.
- ⁵⁵ Genome Online Database (GOLD) (<http://www.genomesonline.org/>).
- ⁵⁶ Allan M. Maxam et Walter Gilbert, A new method for sequencing DNA , *Proc.Natl. Acad. Sci. USA*, vol. 74, 1977, p. 560-564.
- ⁵⁷ Kornberg A. Biological synthesis of DNA: an isolated enzyme catalyzes synthesis of this nucleic acid in response to directions from pre-existing DNA. *Science* 1960.
- ⁵⁸ F. Sanger, G.M. Air, B.G. Barrell, N.L. Brown, A.R. Coulson, C.A. Fiddes, C.A. Hutchison, P.M. Slocombe et M. Smith, Nucleotide sequence of bacteriophage phi X174 DNA , *Nature*, vol. 265, 1977, p. 687-695.
- ⁵⁹ F. Sanger, S. Nicklen & A.R. Coulson, DNA sequencing with chain-terminating inhibitors , *Proc. Natl. Acad. Sci. USA*, vol. 74, 1977, p. 5463-5467.
- ⁶⁰ F. Sanger & A.R. Coulson, The use of thin acrylamide gels for DNA sequencing , *FEBS Lett.*, vol. 87, 1978, p. 107-110.
- ⁶¹ "Séquençage Sanger - Sujets ScienceDirect. Web. 28 févr. 2017.
- ⁶² Joachim Messing, New M13 vectors for cloning , *Methods Enzymol.*, vol. 101, 1983, p. 20-78.
- ⁶³ L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B. Kent & L.E. Hood, Fluorescence detection in automated DNA sequence analysis, *Nature*, vol. 321, 1986, p. 674-679.
- ⁶⁴ H. Swerdlow, J.Z. Zhang, D.Y. Chen, H.R. Harke, R. Grey, S.L. Wu, N.J. Dovichi et C. Fuller, Three DNA sequencing methods using capillary gel electrophoresis and laser-induced fluorescence , *Anal. Chem.*, vol. 63, 1991, p. 2835-2841.
- ⁶⁵ Roger Staden, A strategy of DNA sequencing employing computer programs, *Nucleic Acids Res*, vol. 6, 1979, p. 2601-2610.
- ⁶⁶ Venter JC and & al., The sequence of the human genome, *Science*. 2001.

-
- ⁶⁷ Commins, J., Toft, C., Fares, M. A. - "Computational Biology Methods and Their Application to the Comparative Genomics of Endocellular Symbiotic Bacteria of Insects." *Biol. Procedures Online* (2009).
- ⁶⁸ Ji-Hong Zhang, Ling-Yun Wu et Xiang-Sun Zhang, « Reconstruction of DNA sequencing by hybridization, *Bioinformatics*, vol. 19, n° 1, 2003, p. 14–21.
- ⁶⁹ M. falque, Évolution des méthodes de génotypage· Researchgate July 2011, <https://www.researchgate.net/publication/230690734>.
- ⁷⁰ Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM, Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* (9 Sep 2005).
- ⁷¹ Fan JB, Chee MS, Gunderson KL, « Highly parallel genomic assays », *Nat. Rev. Genet.*, vol. 7, n°8, août 2006, p. 632–44.
- ⁷² Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005.
- ⁷³ Lamorila J., Ameziane N., Deybacha J.-C, Bouizegarènea P., Bogard M., « Les techniques de séquençage de l'ADN : une révolution en marche. Première partie DNA ». *Immuno-analyse et biologie spécialisée* (2008) 23, 260—279.
- ⁷⁴ R.A. Bonnin, T. Naas & L. Dortet, « Impact du séquençage d'ADN à haut débit sur la surveillance des épidémies de bactéries multi-résistantes aux antibiotiques ». *Feuillets de biologie* 354. January 2017.
- ⁷⁵ Theoretical and computational biophysics group, university of illinois at urbanachampaign, 2006. <http://www.ks.uiuc.edu/Research/hemolysin/>.
- ⁷⁶ W. Wickner and R. Schekman. Protein translocation across biological membranes. *Science*, 310(5753) :1452-1456, December 2005.
- ⁷⁷ E. Gouaux. [alpha]-hemolysin from staphylococcus aureus : An archetype of [beta]-barrel, channel-forming toxins. *Journal of Structural Biology* , 121(2) :110_122, 1998.
- ⁷⁸ O. V. Krasilnikov, R. Z. Sabirov, V. I. Ternovsky, P. G. Merzliak, and J. N. Muratkhodjaev. A simple method for the determination of the pore radius of ion channels in planar lipid bilayer membranes. *FEMS Microbiol Immunol* , 5(1-3) :93_100, September 1992.
- ⁷⁹ J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* , 93(24) :13770_13773, November 1996.
- ⁸⁰ M. Akeson, D. Branton, J. J. Kasianowicz, E. Brandin, and D. W. Deamer. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single rna molecules. *Biophysical Journal* , 77(6) :3227_3233, December

1999.

⁸¹ A. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton., Rapid nanopore discrimination between single polynucleotide molecules. *Proceedings of the National Academy of Sciences of the United States of America* , 97(3) :1079_1084, February 2000.

⁸² J. J. Kasianowicz, E. Brandin, D. Branton, and D. W. Deamer, Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24) :13770_13773, November 1996.

⁸³ O. Braha, B. Walker, S. Cheley, J. J. Kasianowicz, L. Z. Song, J. E. Gouaux, and H. Bayley, Designed protein pores as components for biosensors. *Chemistry & Biology* , 4(7) :497_505, July 1997.

⁸⁴ H. Bayley and C. R. Martin. Resistive-pulse sensing - from microbes to molecules. *Chemical Reviews* , 100(7) :2575_2594, July 2000.

⁸⁵ S. Howorka and Z. Siwy, Nanopore analytics : sensing of single molecules. *Chemical Society Reviews* , 38(8) :2360_2384, 2009.

⁸⁶ M. Karhanek, J. T. Kemp, N. Pourmand, R. W. Davis, and C. D. Webb, Single dna molecule detection using nanopipettes and nanoparticles. *Nano Letters* , 5(2) :403_407, February 2005

⁸⁷ J. Li, D. Stein, C. McMullan, D. Branton, M. J. Aziz, and J. A. Golovchenko. Ionbeam sculpting at nanometre length scales. *Nature* , 412(6843) :166_169, July 2001.

⁸⁸ D. M. Stein, C. J. McMullan, J. L. Li, and J. A. Golovchenko. Feedback-controlled ion beam sculpting apparatus. *Review of Scientific Instruments* , 75(4) :900_905, April 2004.

⁸⁹ J. E. Wharton, P. Jin, L. T. Sexton, L. P. Horne, S. A. Sherrill, W. K. Mino, and C. R. Martin. A method for reproducibly preparing synthetic nanopores for resistive-pulse biosensors. *Small* , 3(8) :1424_1430, August 2007.

⁹⁰ Z. S. Siwy and M. Davenport. Biosensors making nanopores from nanotubes. *Nature Nanotechnology* , 5(3) :174_175, March 2010.

⁹¹ Z. S. Siwy and M. Davenport. Nanopores graphene opens up to dna. *Nature Nano- technology*, 5(10) :697_698, October 2010.

⁹² <https://community.nanoporetech.com/protocols>.

⁹³ "GridION X5". nanoporetech.com. Retrieved 2017-07-06.

⁹⁴ Cotillard A & al., Dietary intervention impact on gut microbial gene richness. *Nature*. 2013 Aug 29.

⁹⁵ Aron-Wisnewsky & al., Major microbiota dysbiosis in severe obesity: fate after bariatric surgery. *Gut*. 2018 Jun 13.

⁹⁶ (Raes, J. & Bork, P., Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* 2008.

- ⁹⁷ Voigt, A.Y. et al., Temporal and technical variability of human gut metagenomes. *Genome Biol.* 2015
- ⁹⁸ Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65 (2010).
- ⁹⁹ The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 486, 215–221 (2012).
- ¹⁰⁰ Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60 (2012).
- ¹⁰¹ Costea PI & al., Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol.* 2017 Nov
- ¹⁰² le chatelier, et al. richness of human gut microbiome correlates with metabolic markers. *Nature* 2013.
- ¹⁰³ Blüher M., are metabolically healthy obese individuals really healthy? *Endocr Soc* 2014.
- ¹⁰⁴ Kopelman Pg., Obesity as a medical problem. *Nature* 2000.
- ¹⁰⁵ Wesolowska-Andersen, A. et al., Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* 2, 19 (2014).
- ¹⁰⁶ McOrist, A.L., Jackson, M. & Bird, A.R., A comparison of five methods for extraction of bacterial DNA from human faecal samples. *J. Microbiol. Methods* 50, 131–139 (2002).
- ¹⁰⁷ Ariefdjohan, M.W., Savaiano, D.A. & Nakatsu, C.H. Comparison of DNA extraction kits for PCR-DGGE analysis of human intestinal microbial communities from fecal specimens. *Nutr. J.* 9, 23 (2010).
- ¹⁰⁸ Ehrlich D., & Bork P., Enterotypes of the human gut microbiome. *MetaHIT Consortium, Nature* 2011.
- ¹⁰⁹ Caputo A., Whole-genome assembly of *Akkermansia muciniphila* sequenced directly from human stool. *Biology Direct* 2015.