

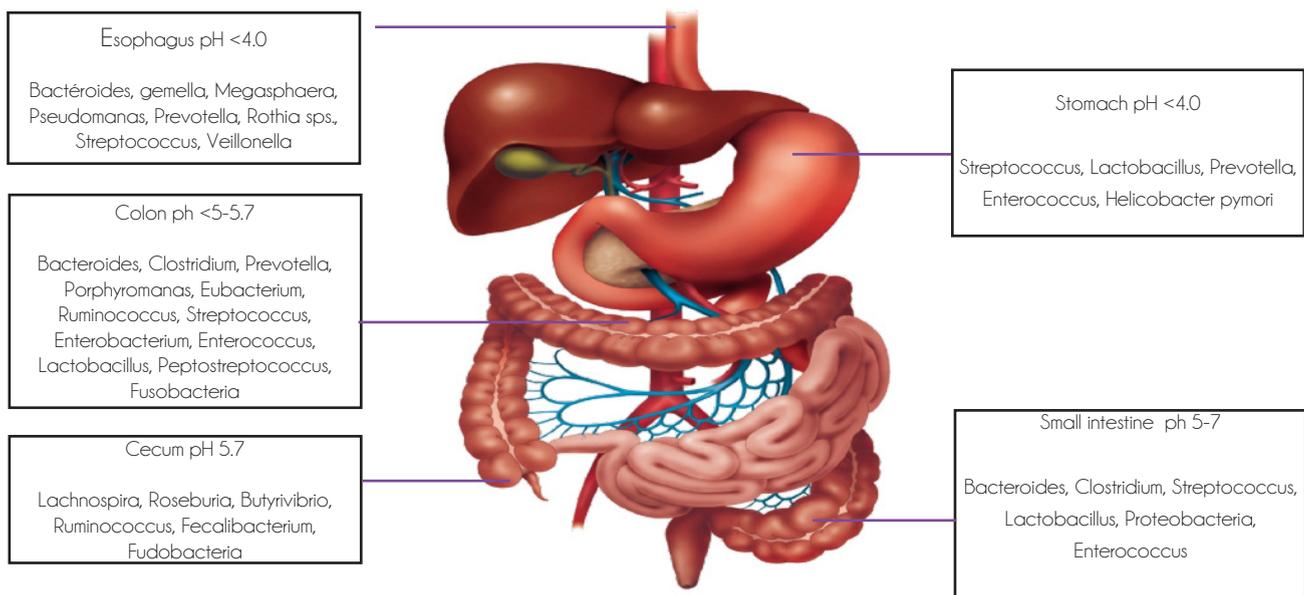
Explorer le microbiote par séquençage métagénomique

Les scientifiques pressentent l'importance des microbiotes sur la santé humaine et animale. Explorer ces flores est extrêmement utile pour la médecine et la nutrition. Ici aussi, on parle d'écologie, de symbiose, de biodiversité et de respect de l'équilibre. Découvrez les méthodes utilisées pour dévoiler les habitants de notre planète « CORPS ».

Attention ! ça vole haut !

Le Microbiote

La flore est d'une extrême complexité : composée essentiellement de bactéries mais également d'autres microorganismes comme des archées, des protozoaires, des champignons... Tout dépend de sa situation : intestin, peau, certains orifices... Ne serait-ce qu'au sein du système digestif, la composition varie beaucoup.



Source : Jandhyala SM, Talukdar R, Subramanyam C, Vuyuru H, Sasikala M, and Reddy DN. Role of the normal gut microbiota. World Journal of Gastroenterology. 2015, 21(29): 8787-8803

Remarque : les archées produisent du méthane dans le microbiote, ce qui a pour conséquence de ralentir le transit intestinal.

Notre microbiote serait responsable d'une partie des variations inter-individus : l'étudier et mieux le comprendre pourrait permettre de grandes avancées scientifiques. Or, le microbiote humain est caractérisé par une fraction cultivable limitée (30 % dans le tractus gastro-intestinal d'après Hayashi, 2020). En d'autres termes, il est impossible d'en cultiver une partie en laboratoire afin de l'étudier. Pour en apprendre davantage sur sa composition, il faut nécessairement aborder la fraction non-cultivable. Pour cela, on réalise des inventaires par approches moléculaires.

Il existe plusieurs méthodes pour caractériser le microbiote.

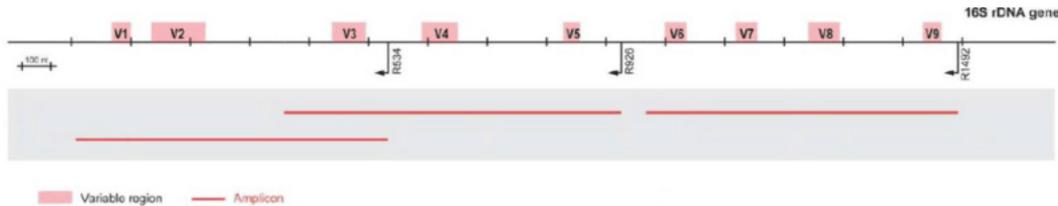
Méthode 1 : ciblage par ARNr 16S

Cette méthode est généralement appliquée lorsque aucune information n'est disponible sur la colonie que l'on veut étudier. Le principe est de cibler un certain pourcentage de microorganismes pour lesquels on amplifie le signal pour mieux pouvoir les détecter. Pour cela, on utilise une région spécifique du génome qui code pour l'ARNr 16S (Acide Ribonucléique ribosomique) car elle est conservée dans tous les génomes et qu'on la retrouve chez les bactéries et les archées. On pourrait choisir d'autres ARN comme le 23S ou 5S car eux aussi sont bien conservés, mais la banque de données de séquences du 16S est aujourd'hui la plus développée. De plus, l'identification est fiable et représentative : les résultats obtenus par séquençage du gène 16S sont similaires à ceux obtenus avec le génome entier.

CONSULTATION NUTRITION

Remarque : selon les êtres vivants, on utilise différents marqueurs. Chez les eucaryotes par exemple, on utilise l'ARNr 18S (protozoaires) ou ITS (champignons). Pour les virus, qui sont de petite taille, on filtre la solution cellulaire avec un tamis très fin à plusieurs reprises pour ne garder que les virus. L'ARN 16S est composé de 1500 nucléotides, formant une séquence contenant des régions variables et des régions

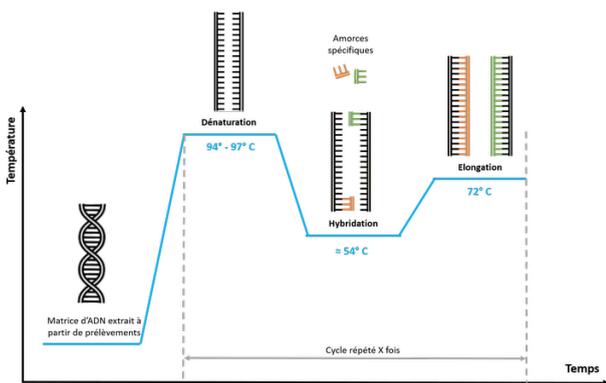
conservées. Les régions variables n'ont pas de rôle fonctionnel important et peuvent évoluer sous l'effet de mutations. Les régions conservées, quant à elles, permettent de créer des amorces universelles pour amplifier les régions à analyser car ce sont des régions identiques chez toutes les bactéries. On peut ainsi faire des amorces d'amplification puis pratiquer une PCR (Réaction en chaîne par polymérase).



Source : Consortium HMP

Une fois que l'amorce a été déterminée, les étapes de la PCR sont les suivantes : dénaturation, hybridation et élongation. C'est en répétant ce cycle x fois que l'on va pouvoir multiplier l'amorce et donc l'amplifier. A chaque cycle, un brin d'ADN en donne 2 : l'amplification est donc exponentielle, n cycles donnent donc 2n-1 molécules d'ADN.

évolutivement, mais toutefois, différente. On pense que c'était à l'origine une bactérie du microbiote oral qui se serait retrouvée dans le lait puis adaptée. Ces deux espèces ont des séquences quasi identiques à 100 % pour le 16S, d'où la difficulté à les différencier.



Source : Biomnigene.fr

Quelle est la nécessité d'amplifier ? Cela permet de révéler seulement les micro-organismes que l'on veut analyser, sans contamination par d'autres sources. Normalement on trouve 99 % de bactéries parmi tous les micro-organismes présents mais le contenu peut, par exemple, être enrichi en cellules de l'hôte (épithéliales ou immunitaires) qui masquent le signal. C'est le cas, par exemple, pour le microbiote nasal. Sans amplifier, on analyserait donc notre ADN et pas celui du microbiote, ce qui n'est pas l'effet recherché. Même si, pour le microbiote intestinal, environ 99 % sont des bactéries du microbiote et non des cellules de l'hôte, on préfère quand même amplifier car la composition peut évoluer : à titre d'exemple, si on a des diarrhées, les cellules épithéliales peuvent occuper une plus grande part des microorganismes présents et masquer le signal. C'est donc une question pratique. L'inconvénient de cette méthode est qu'elle est assez imprécise quant aux bactéries analysées. Par exemple, *S. thermophilus* est une bactérie présente dans le yaourt (elle fermente le lait) qui a évolué avec le temps pour s'adapter à l'écosystème laitier. *S. salivarius* est une bactérie du même genre donc proche

Streptococcus thermophilus

Streptococcus salivarius

Score	Espèce	Identité	Cover	E value	Per Ident
784	<i>Streptococcus thermophilus</i> strain ATCC 19258	0.0	426/427 (99%)	0/427 (0%)	Plus/Plus
784	<i>Streptococcus salivarius</i> strain NCTC8618	0.0	426/427 (99%)	0/427 (0%)	Plus/Plus

Streptococcus salivarius strain NCTC8618 genome assembly, chromosome: 1

Query	Score	Evalue	Identité	Cover	E value	Per Ident
Query 1	TAGGGAATCTTCGGCAATGGGGGCAACCTGACCGAGCAGCAGCCGGCGGATGAGGAGG	60				
Sbjct	TAGGGAATCTTCGGCAATGGGGGCAACCTGACCGAGCAGCAGCCGGCGGATGAGGAGG	15213				
Query 61	TTTTGGGATGTAAGCTCTGTGTAAAGTCAAGACCGGCTGGAGAGTGGAAAGTTCACA	120				
Sbjct	TTTTGGGATGTAAGCTCTGTGTAAAGTCAAGACCGGCTGGAGAGTGGAAAGTTCACA	15273				
Query 121	CTGTACGGTACTTACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	180				
Sbjct	CTGTACGGTACTTACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	15333				
Query 181	GATGGCTCGAGGCTGTCCGATTTATTTGGGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGG	240				
Sbjct	GATGGCTCGAGGCTGTCCGATTTATTTGGGGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGG	15393				
Query 241	GCTGAAATTAAGGCTGTGGCTCAACCATAGTTCGCTTGGAAACTGTCAAACTTGAAT	300				
Sbjct	GCTGAAATTAAGGCTGTGGCTCAACCATAGTTCGCTTGGAAACTGTCAAACTTGAAT	15453				
Query 301	GCAGAGGGGAGAGTGGATTCATGTGTAGGGGTAAGTGGATGATATATGGAGGAGC	360				
Sbjct	GCAGAGGGGAGAGTGGATTCATGTGTAGGGGTAAGTGGATGATATATGGAGGAGC	15513				
Query 361	ACGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	420				
Sbjct	ACGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	15573				
Query 421	GGGACCA 437					
Sbjct	GGGACCA 15630					

• Quasi 100% identiques!

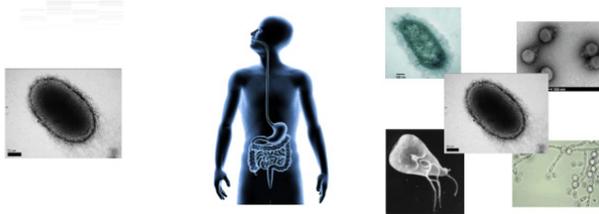
Cela devient problématique car on peut révéler des microorganismes mais - s'ils sont trop proches - on ne pourra pas les séparer. Une autre limite est que cette méthode ne permet pas de quantifier les microorganismes : on peut seulement savoir lesquels sont présents. Si on voulait le savoir, il faudrait compléter cette technique avec un suivi par fluorescence pour mesurer le nombre d'amplicons (fragments d'ADN amplifiés). Enfin, cette méthode permet seulement d'amplifier l'ADN d'une bactérie ou autre micro-organisme, pas de savoir si ce dernier est vivant ou mort car même mort, son ADN peut être intact.

Méthode 2 : Métagénomique globale

Cette méthode est utilisée lorsque l'on veut analyser l'intégralité d'un génome afin de voir toutes les différences qu'il pourrait avoir avec un autre génome. Pour ce faire, on séquence aléatoirement différents microorganismes présents : in fine, c'est donc le contenu global du génome que l'on étudie. Si l'on considère qu'une bille = 1 gène, il faudrait en moyenne 3 000 tirages pour voir toute la diversité du microbiote. Le problème en métagénomique, c'est qu'on ne sait pas combien il existe d'espèces et combien il y a de gènes par espèce : on avance à tâtons.

CONSULTATION NUTRITION

Un des paramètres clés de cette méthode est le choix de la profondeur de séquençage. Plus on va réaliser de séquençages de parties de microbiote, plus l'effort pour retrouver les microorganismes présents sera faible. De la même façon que pour les pixels d'une image : plus il y en a, plus la résolution est bonne et plus on arrive à déterminer ce que représente la photographie.

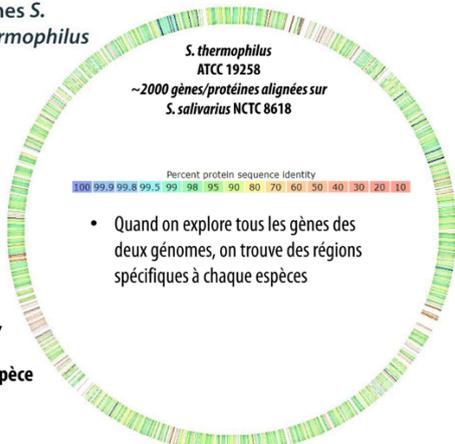


1 génome bactérie
~3 000 000 bases
~3 000 gènes

1 génome humain
~2,9 Gigabases (Gb)
~25 000 gènes

1 métagénome
? Gb
? gènes

Exemple: Génomes *S. salivarius* vs *S. thermophilus*



Légende :
En bleu = séquences très identiques entre deux bactéries (comme 16S)
En rouge = séquences qui diffèrent - En blanc = pas de gène correspondant

Le choix de la méthode dépend des microbiotes auxquels on s'intéresse. Par exemple, le microbiote pulmonaire contient beaucoup d'eucaryotes donc il est difficile de voir l'écosystème : on privilégie la méthode 16S. Au contraire pour le microbiote intestinal, on peut privilégier l'approche métagénomique globale car il est intéressant de connaître entièrement la composition du génome pour voir des différences au niveau de la souche. Contrairement à la première méthode, celle-ci permet de quantifier les microorganismes et donc de déterminer la composition en termes de proportion. Une autre application de l'approche métagénomique est le cas de la consommation de probiotiques via l'alimentation : le séquençage global est un moyen de distinguer les bactéries commensales (hôtes habituels d'un organisme) de celles que l'on vient d'ingérer.

En termes de coût, la méthode globale est plus chère que la méthode de ciblage car elle demande d'analyser tout le génome et pas seulement une partie. Cependant, comme elle est de plus en plus utilisée, notamment pour mieux comprendre le

microbiote intestinal, son coût diminue progressivement pour se rapprocher de celui de la méthode par ciblage.

Chaque écosystème possède sa complexité : on obtient des signaux des micro-organismes commensaux mais aussi des microorganismes de l'environnement, cachés sous le premier. Si on a un effet seuil, on peut se dire que lorsque l'on voit l'image de ce qui est le plus abondant, alors ce microorganisme aura un rôle majeur. Néanmoins, il existe des cas où un pathogène peut être présent en faible abondance mais avoir un effet néfaste conséquent.

Le microbiote vaginal est moins complexe et la profondeur de séquençage n'est pas la même. On note entre une dizaine et une cinquantaine d'espèces différentes dans le microbiote vaginal alors qu'il y en a plusieurs centaines dans le microbiote intestinal. De plus, il est plus simple, et plus facile à observer. Cependant, simple ne veut pas dire que les espèces ne sont pas proches. On peut avoir différentes espèces de lactobacilles, rendant leur identification difficile. Dans ce cas, la méthode 16S ne permet pas de les distinguer.

Outils de séquençage

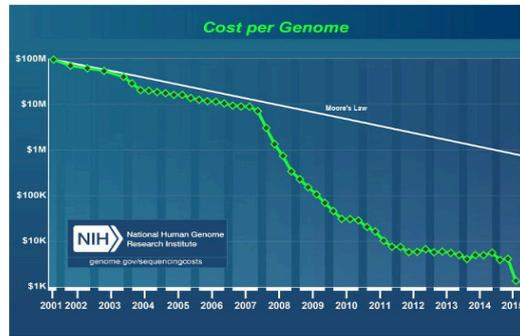
Les outils de séquençage ont évolué avec le temps et on a vu se développer plusieurs générations de séquenceurs. Les séquenceurs de deuxième génération permettent un débit très important et une grande possibilité de résolutions (nombre de séquences). Les séquenceurs de troisième génération équivalent à du temps réel, sont beaucoup plus rapides et la longueur des séquences fragmentées est plus importante que ce qui est possible avec les séquenceurs de deuxième génération.

Pour évaluer les avantages et les inconvénients de chaque technologie, on utilise un mix de micro-organismes bien connus et on les compare. Ce travail a été effectué par le Dr. Mircea Podar (Oak Ridge National Laboratory) avec un mix synthétique contenant 71 espèces microbiennes (archées et bactéries). 4 séquenceurs ont été comparés (voir résultats ci-dessous).

Echantillon	Séquenceur	Temps de séquençage (heure)	Nombre de séquence (Million)	Taille cumulée (Gb)	Taille moyenne des séquences (nt)	Pourcentage moyen d'erreur
MIX 71 espèces	Illumina HiSeq 3000	96	20,917	3,032	144,97	0,06
	Ion S5	24	28,517	4,139	145,16	0,24
	Nanopore Minion R9	48	0,389	1,701	4363,86	9,52
	PacBio Sequel II	96	0,525	5,4	10289,61	0,22

Avec les séquenceurs de deuxième génération (Illumina HiSeq 3000 et Ion S5), on obtient beaucoup plus de séquences mais elles ne sont pas très longues. De plus, les séquenceurs de 2nde génération sont les plus utilisés donc leur coût est de moins en moins élevé. Les pourcentages moyen d'erreur des séquenceurs de 2nde génération et du PacBio sont très faibles. Par contre, le Nanopore (3ème génération) fait beaucoup plus d'erreurs. Le meilleur séquenceur est celui qui propose le meilleur compromis entre la taille de séquences maximum et le moins d'erreurs : le séquenceur PacBio Sequel II est une révolution car son débit est important et il permet d'obtenir des séquences très longues. L'évolution dans les séquenceurs a permis une baisse considérable des coûts de séquençage. Il y a 20 ans, il fallait déboursier près de 100 millions de dollars pour séquencer un génome. Cela ne coûte plus que 1000 dollars aujourd'hui.

CONSULTATION NUTRITION



Projet MetaHIT (2010)

Ce projet financé par la Commission européenne entre 2008 et 2012 avait pour but l'obtention d'un catalogue de génomes permettant de reconstituer la cartographie de tous les génomes dans le monde.

Catalogue MetaHIT: une référence pour explorer le microbiote humain

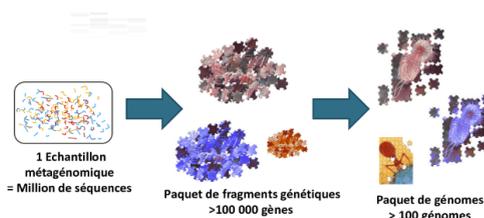
- En 2010: 3,3M de gènes reconstruits avec 124 échantillons fécaux [Qin et al, Nature, 2010]
- La plupart provenant d'organismes inconnus avec des fonctions inconnues

<http://ae2bm.org/wp-content/uploads/2019/10/Blottiere-AEBBM-12-sept-2019.pdf>

Pour cela, la méthode suivie était proche de celle d'un puzzle : tout commence par le regroupement des génomes par séquences similaires (gènes, opérons...) puis la constitution de paquets permettant la quantification des génomes. On effectue cela juste pour un métagénome (un seul être humain) dans un premier temps, puis on l'étend à sa famille, son entourage, la France, etc. Ainsi, on peut obtenir une grande quantité d'informations permettant de compléter le catalogue. L'objectif principal de

ce projet était d'établir des corrélations entre les gènes du microbiote intestinal humain et l'état de santé qui héberge ce microbiote.

La plupart des gènes du catalogue avaient une fonction inconnue en 2010. Certains étaient associés à la formation des acides aminés, des lipides... Mais on ne connaissait pas la majorité. Depuis, beaucoup d'efforts ont été effectués et la population étudiée est de plus en plus importante et diverse. Au départ, beaucoup de génomes étudiés étaient ceux de Finlandais, mais aujourd'hui, on en compte plus de 400, provenant de 26 pays d'Europe.

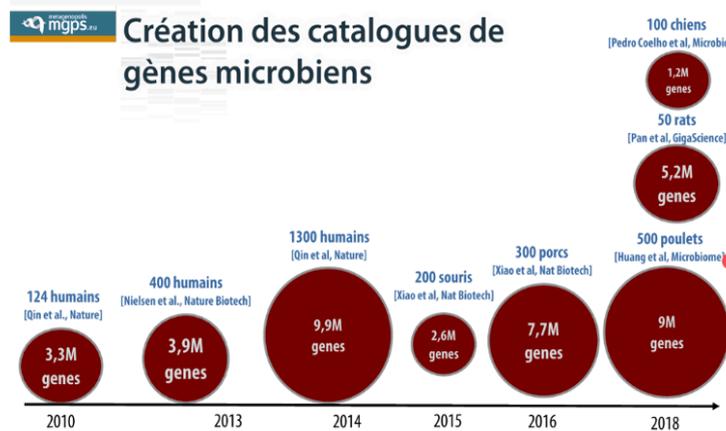


On a pu observer de grandes disparités entre les microbiotes de la population européenne et de ceux d'autres populations. Par exemple, une population de chasseurs cueilleurs possède un microbiote très divers et donc plus riche que celui des européens. Il en est de même lorsqu'on séquence des génomes de microbiotes de populations Sud-Américaines, Africaines,

CONSULTATION NUTRITION

d'Océanie... Au-delà des différences culturelles, la diversité de maladies et des régimes alimentaires peut également apporter une richesse et une composition différente.

L'avancée de la création des différents catalogues jusqu'à ce jour est décrite ci-dessous :



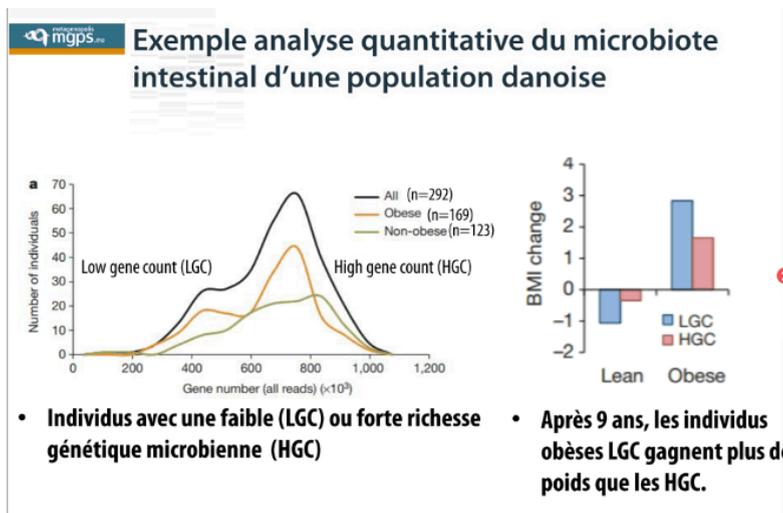
L'approche métagénomique n'est pas réalisée que chez les Hommes, mais aussi chez les animaux d'élevage comme la souris, le porc et le poulet (dont le microbiote très complexe en nombre de microorganismes détectés). Chez les animaux de compagnie apparaissent les mêmes problèmes que chez les Hommes : digestion, inflammation, etc. D'ailleurs, on voit ces problèmes progresser quasiment de la même manière. L'objectif est d'avoir un catalogue très diversifié.

D'autres projets de recherche sont en cours, comme celui cherchant à obtenir la composition du microbiote de 100 000 individus français afin d'améliorer les connaissances.

Actuellement il y a un projet de coffre-fort de la biodiversité microbienne intestinale pour sauver les espèces microbiennes qui sont menacées de disparition. Cela est important car, souvent on ne connaît pas les fonctions des microorganismes et elles pourraient être très utiles.

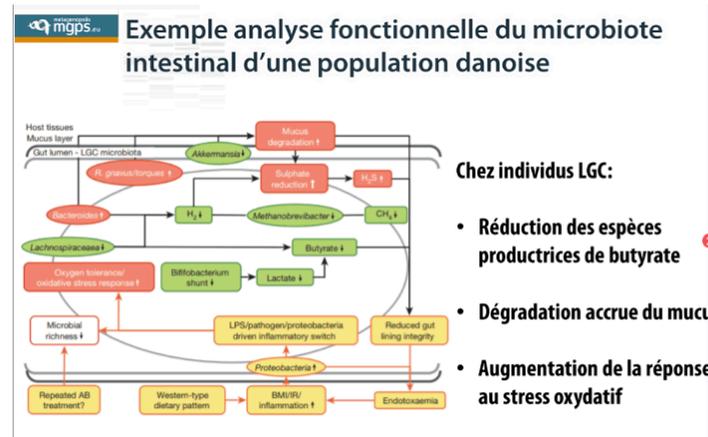
Exemple d'utilisation chez une population danoise :

Une étude s'est intéressée au lien entre microbiote et santé au sein d'une population danoise, mais aussi aux fonctions de ce microbiote. Les principales observations sont, qu'au cours du temps, les individus qui avaient peu de gènes dans leur microbiote ont pris plus de poids. On peut se demander si avoir un microbiome pauvre est lié à l'obésité. Une richesse microbienne élevée permet en général d'être en meilleure santé. Par « richesse », on entend « nombre de gènes différents que l'on observe, lié à la biodiversité ». C'est un bon indicateur pour savoir si le microbiote est favorable ou plus à risque, et pour savoir si un individu doit faire l'objet d'une vigilance plus accrue.



CONSULTATION NUTRITION

On a constaté qu'une faible diversité de gènes était liée à une faible production de butyrate, métabolite très important. De plus, on constate dans ces conditions une dégradation du mucus (qui est alors plus fin). Comme une de ses fonctions est de servir de barrière vis-à-vis des pathogènes, le fait qu'il soit altéré permet d'ouvrir la voie à des micro-organismes pathogènes potentiellement inflammatoires et qui augmentent le stress oxydatif.



Ouverture :

Quel est votre groupe intestinal ? Cette question pourrait être posée dans le futur au même titre que notre groupe sanguin. Des chercheurs du Laboratoire européen de biologie moléculaire (EMBL) à Heidelberg, en Allemagne ont en effet découvert que les bactéries qui colonisent nos intestins constituent, en fonction des individus, trois écosystèmes différents nommés d'après le nom du microbe le plus représenté (Bacteroides, Prevotella ou Ruminococcus). Comme pour les groupes sanguins ces trois entérotypes sont indépendants de traits tels que l'âge, le sexe, l'origine ethnique ou l'indice de masse corporelle. Cette découverte ouvre de nombreuses perspectives thérapeutiques : elle va aider la communauté scientifique à comprendre les différences d'efficacité des traitements selon les individus ou même de prédire l'apparition de certaines maladies, notamment les pathologies chroniques inflammatoires du tube digestif, à partir de l'analyse de la flore intestinale.

Merci à Mathieu Almeida - Chargé de recherche bio-informatique à Metagenopolis. Il a effectué une thèse sur l'analyse du microbiote alimentaire nommée « Caractérisation de flores microbiennes intestinale humaine et fromagère par méthode de métagénomique quantitative », puis il s'est intéressé

au microbiote humain. Metagenopolis est un institut créé en 2013 sous tutelle de l'Université Paris-Saclay et de l'INRAE avec pour objectif de caractériser le microbiote humain et en définir son rôle, mais également de proposer des standards pour les différentes méthodes d'analyse du microbiote.

Nos travaux à MGP

Cancer

2018: Routy et al. *Science*, Microbiome and epithelial cancer immunotherapy
2018: Gopalakrishnan et al. *Science*, Microbiome and melanoma cancer immunotherapy

Metabolic disorders

2012: Qin et al. *Nature*, Type II Diabetes
2013: Le Chatelier et al. *Nature*, Richness of gut microbes and metabolic markers
2014: Qin et al. *Nature*, Human gut microbiome alterations in liver cirrhosis
2015: Qin et al. *Nature*, Accurate liver cirrhosis diagnostic
2015: Forslund et al. *Nature*, Drug cofounders in microbiome analysis
2016: Pedersen et al. *Nature*, Microbiome & insulin resistance

Diet

2011: Arumugam et al. *Nature*, Enterotypes
2013: Cotillard et al. *Nature*, Impact of diet on gut microbiome
2013: Le Chatelier et al. *Nature*, Method for identifying microbial and metabolic markers
2019: Cox et al. *Gastroenterology*, Low FODMAP Diet in Inflammatory Bowel Disease patients

Antibiotic resistance

2019: Ruppé et al. *Nature Microbiology*, Prediction of the intestinal resistome

Technologies

2010: Qin et al. *Nature*, The human gut reference catalogue
2013: Sunagawa et al. *Nature Methods*, Universal phylogenetic markers
2014: Nielsen et al. *Nature Biotech*, Method for identifying metagenomic species
2014: Li et al. *Nature Biotech*, 10 millions genes reference catalog
2015: Xiao et al. *Nature Biotech*, A mouse gut gene catalogue
2016: Xiao et al. *Nature Microbiology*, A pig gut gene catalogue
2017: Costea et al. *Nature Biotech*, Standards for microbiome studies
2018: Plaza Onate et al., *Bioinformatics*, Reconstruction of metagenomic pangenome species

Sources :

AEBBM - Journées Pédagogiques et Scientifiques 2019 - Nantes - 12 septembre 2019 - <http://ae2bm.org/wp-content/uploads/2019/10/Blottiere-AEBBM-12-sept-2019.pdf>

Mathieu Almeida - Thèse - Caractérisation de flores microbiennes intestinale humaine et fromagère par méthode de métagénomique quantitative - Juin 2013 www.sfb.fr/node/11795

www.biomnigene.fr/fr/nos-solutions/pcr.html

Voula Alexandraki et Al. - The complete genome sequence of the yogurt isolate *Streptococcus thermophilus* ACA-DC 2 - Standards in Genomic Sciences volume 12, Article number: 18 (2017)

DEPARTEMENT NUTRITION NUTRIMARKETING

Rédaction : Constance Goujard Conception graphique : Douchane Momcilovic Mise en page : Alix de Reynal
contact@nutrimarketing.eu www.nutrimarketing.eu T : 01 47 63 06 37

Crédit photographique : AgroParisTech - Mathieu Almeida - Biomnigene - Club PAI Food Ingredients - MCP - NutriMarketing - DR Média d'information pour les professionnels de santé - N°107 - Septembre 2020 - Tous droits réservés
NutriMarketing - RCS Paris 412 053 62