

Introduction à la métagénomique

Claire Hoede, PF Bioinfo Genotoul

Quelques définitions

- **Les micro-organismes**

Quelques définitions

- **Les micro-organismes :**
ensemble des organismes microscopiques (bactéries, virus, champignons et algues microscopiques).
- Ils jouent un rôle vital dans le fonctionnement général de la biosphère en participant par exemple au cycle du carbone ou de l'azote.
- Leur diversité spécifique et fonctionnelle, les mécanismes régissant leur dispersion ainsi que leur histoire évolutive demeurent encore mal connus.

Quelques définitions

- **Le microbiote :**

Quelques définitions

- **Le microbiote :**

ensemble des micro-organismes vivant au sein d'un environnement spécifique (incluant les non-cultivables)

Quelques définitions

- **Le microbiote :**

ensemble des micro-organismes vivant au sein d'un environnement spécifique (incluant les non-cultivables)

- **Le microbiome :**

Quelques définitions

- **Le microbiote :**

ensemble des micro-organismes vivant au sein d'un environnement spécifique (incluant les non-cultivables)

- **Le microbiome :**

environnement dans lequel ils évoluent et- interagissent.

Par exemple le microbiome intestinal.

Métagénomique / métagénétique

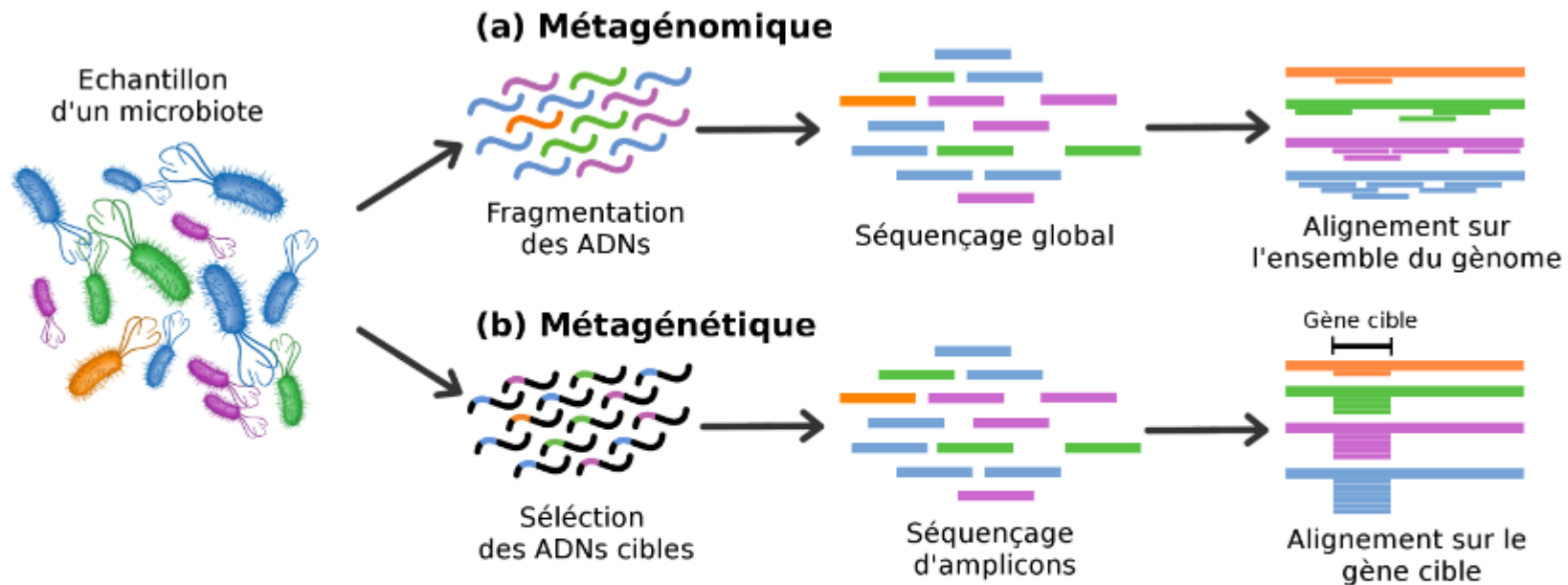


Figure 1.1 Stratégies de séquençage de métagénomies. La métagénomique (a) fragmente l'ensemble des ADNs présents dans un échantillon en courts fragments afin de les séquencer. Les séquences ainsi obtenues peuvent provenir de n'importe quel locus du génome. La métagénétique (b) sélectionne des gènes cibles avant de les séquencer. Les séquences représentent seulement les gènes ciblés.

Métagénomique

- Beaucoup de recul sur la métagénéétique, beaucoup moins avec le WGS en météénomique.
- Les biais ne sont pas aussi bien connus

Métagénomique

- Quelles sont les questions qu'on peut se poser avec ce type de données ?

Métagénomique

- Quelles sont les questions qu'on peut se poser avec ce type de données ?
 - Qui est là ? ==> Composition en espèces
 - Quelles sont les fonctions présentes ?
 - Répartition géographique, phylogéographique
 - Dynamique des populations au cours du temps en réponse à l'environnement
 -

Métagénomique : les étapes du côté des biologistes

- Échantillonnage et extraction de l'ADN :
 - Il dépend du milieu analysé (eau, sol, excréments...)
 - Un filtre sur la taille est souvent fait pour discriminer les virus, les bactéries et les eucaryotes unicellulaires.
 - Le stockage, l'exposition à l'oxygène entraînent des biais ==> il est donc indispensable que les échantillons à comparer est tous subis le même protocole d'extraction.

Métagénomique : les étapes du côté des biologistes

- Préparation des librairies :
 - Selon la quantité d'ADN extraite, il est possible d'une PCR soit nécessaire
 - L'amplification est à éviter à cause des biais importants qu'elle engendre

Métagénomique : les étapes du côté des biologistes

- Séquençage :
 - L'assemblage étant une étape complexe, des reads longs (pacbio) sont à considérer
 - Mais le coût est très important
 - En pratique : pas encore au point
 - Possible seulement quand il y a très peu de diversité
 - En pratique un séquençage Illumina HiSeq est pratiqué.
 - Ce qui permet une grande profondeur de séquençage mais les reads sont courtes (2x150 pb)
 - Intéressant lorsqu'on veut avoir accès aux espèces rares dans le milieu étudié.

Les challenges de la métagénomique

- Les données brutes issues des analyses de métagénomiques sont complexes et bruitées
 - Imaginez de l'ordre de 10000 puzzles de milliards de pièces avec des erreurs à reconstruire. De plus pour la plupart on ne connait pas le modèle !
 - Extraire de l'information biologique de jeux de données de cette taille est un challenge et nécessite de grosses ressources de calcul

Les principales étapes bioinformatiques

- **Nettoyage** sur la base de la qualité de séquençage, la présence de contamination etc
- **Assemblage :**
 - Approches basées sur les graphes de De Bruijn ==> chimères fréquentes
 - MetaVelvet, IDBA-UD, MetaSPAdes...
- **Prédiction de genes**
 - Par homology (ex blast or diamond comme dans MEGAN6)
 - Ab initio (ex GeneMark...)
 - Annotation (Interproscan, EggNOG (HMM protein domains and families), KEGG, metaCyc (metabolic pathways...))

Les principales étapes bioinformatiques

- **Assignment taxonomique et calcul de l'abondance relative (binning)**
 - Recherche d'homologie dans des bases de données (ex MEGAN)
 - Utilisation de marqueurs spécifiques (ex MetaPhlAn)
 - Calcul de l'abondance relative des OTUs avec plus ou moins de correction (ex SLIMM : correction par le profil de couverture)
 - Certains outils utilisent aussi la composition en oligonucléotides (ex Anvi'o, Kraken ...)

Les principales étapes bioinformatiques

- **Intégration de données**
 - Les données de séquences métagénomiques sont associées à des métadonnées

Les principales étapes bioinformatiques

- **Intégration de données**

- Les données de séquences métagénomiques sont associées à des métadonnées : température, profondeur, coordonnées GPS, etc...Des bases de données dédiées existent :
 - <https://img.jgi.doe.gov/cgi-bin/m/main.cgi> (IMG data warehouse)
 - <https://gold.jgi.doe.gov/> (GOLD : genomeOnLine Database) : avec un lien vers les séquences stockées au JGI.
- De plus on peut vouloir interpréter les résultats à la lumière d'autres informations sur les échantillons (metatranscriptome, distance évolutive entre certains organismes ...)

SUJET DE RECHERCHE EN EMERGENCE

Les principales étapes bioinformatiques

- **Métagénomique comparative**

- Comparaison de la composition en espèce
- Comparaison des fonctions présentes
- ...

SUJET DE RECHERCHE EN EMERGENCE

Quelques applications

- **Caractérisation du microbiote intestinal**
 - Lien avec l'obésité, le diabète, l'autisme...
- **Fermenteurs**
 - Biocarburant
- **Alimentation**
 - fromages
- **Ecologie**

Quelques papiers phares

- **Tara Ocean :**

- Lima-Mendez, G. et al. Determinants of community structure in the global plankton interactome. *Science* 348, doi:10.1126/Science.1262073 (2015).
- de Vargas, C. et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* 348, doi:10.1126/Science.1261605 (2015).
- Brum, J. R. et al. Patterns and ecological drivers of ocean viral communities. *Science* 348, doi:10.1126/Science.1261498 (2015).
- Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* 348, doi:10.1126/Science.1261359 (2015).
- Villar, E. et al. Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* 348, doi:10.1126/Science.1261447 (2015).

Quelques papiers phares

- **Meta Hit project :**

- Qin, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* (4 March 2010) 464, 59-70.

- **Human microbiome project :**

- Ward, Doyle V. et al. “Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic E. Coli in Necrotizing Enterocolitis and Mortality in Preterm Infants.” *Cell reports* 14.12 (2016): 2912–2924.