

# GÉNOMIQUE DE DIVERSITÉ : MÉTAGÉNOMIQUE

# Métagénomique

approche classique : isolation d'un seul espèce + clonage + séquençage

métagénomique : extraction d'ADN d'un échantillon environmental (+ amplification ) + séquençage

- ★ pour bcp d'organismes (99%), on n'arrive pas avoir une culture au laboratoire
- ★ bcp d'organismes ne fonctionnent que dans une communauté avec d'autres

analyses :

- ★ composition taxonomique
- ★ composition fonctionnelle
- ★ métagénomique comparative

# Analyse taxonomique

approche classique : filtrage d'une région (primers spécifiques) qui identifie l'espèce  
+ comparaison à une grande base de données sur cette région

ARN ribosomique petite sous-unité (16S chez procaryotes) — régions hypervariables (V1–V9)

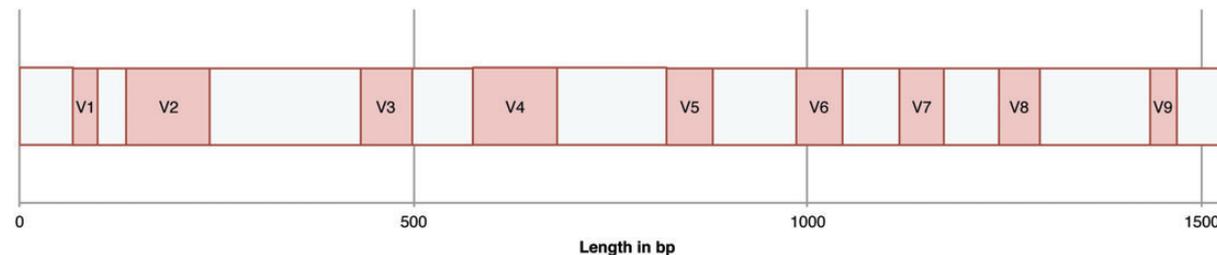


Figure 2. Approximately 1.5 kb 16S rRNA gene of *E.coli* showing the nine variable regions that make it an ideal target as a phylogenetic marker gene.

bases de données d'ARNs alignés :

★ Ribosomal Database Project (RDP)

★ greengenes

★ ARB-SILVA

+ outils spécialisés pour la recherche / alignement rapide

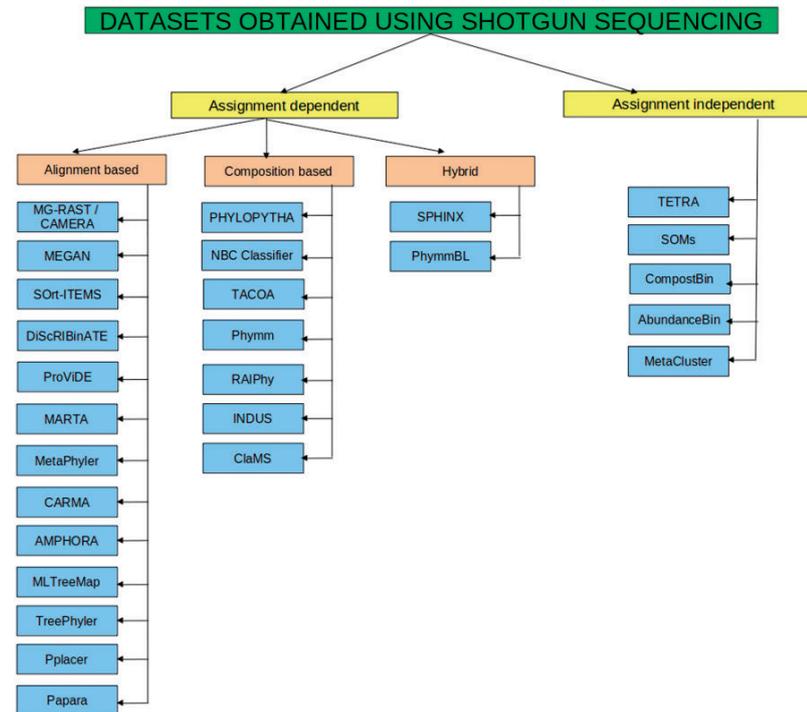
Cox & al. *Human Molecular Genetics* 22 :R88 (2013)

# Analyse taxonomique / binning

séquençage + comparaison à un arbre phylogénétique

on veut lier les séquence à des unités taxonomiques (OTU = operational taxonomic unit)

méthodes de classifications : alignement/similarité, composition ( $k$ -mers, codons)



Mande & al *Briefings in Bioinformatics* 13 :669 (2012)

# Analyse fonctionnelle

**P  
O  
P  
U  
L  
A  
T  
E  
D  
\*  
S  
U  
B  
S  
Y  
S  
T  
E  
M**

Subsystem: Histidine Degradation		
1	<b>HutH</b>	Histidine ammonia-lyase (EC 4.3.1.3)
2	<b>HutU</b>	Urocanate hydratase (EC 4.2.1.49)
3	<b>HutI</b>	Imidazolonepropionase (EC 3.5.2.7)
4	<b>GluF</b>	Glutamate formiminotransferase (EC 2.1.2.5)
5	<b>HutG</b>	Formiminoglutamase (EC 3.5.3.8)
6	<b>NfoD</b>	N-formylglutamate deformylase (EC 3.5.1.68)
7	<b>ForI</b>	Formiminoglutamic iminohydrolase (EC 3.5.3.13)

**A**

Subsystem Spreadsheet								
Organism	Variant	HutH	HutU	HutI	GluF	HutG	NfoD	ForI
<i>Caulobacter crescentus</i>	1	P58082	Q9A9MI	P58079			Q9A9M0	Q9A9L9
<i>Pseudomonas putida</i>	1	Q88CZ7	Q88CZ6	Q88CZ9			Q88D00	Q88CZ3
<i>Xanthomonas campestris</i>	1	Q8PAA7	P58988	Q8PAA6			Q8PAA8	Q8PAA5
<i>Halobacterium sp.</i>	2	Q9HQD5	Q9HQD8	Q9HQD6		Q9HQD7		
<i>Deinococcus radiodurans</i>	2	Q9RZ06	Q9RZ02	Q9RZ05		Q9RZ04		
<i>Bacillus subtilis</i>	2	P10944	P25503	P42084		P42068		
<i>Bacteroides thetaiotaomicron</i>	3	Q8A4B3	Q8A4A9	Q8A4B1	Q8A4B0			
<i>Desulfotalea psychrophila</i>	3	gi51246205	gi51246204	gi51246203	gi51246202			

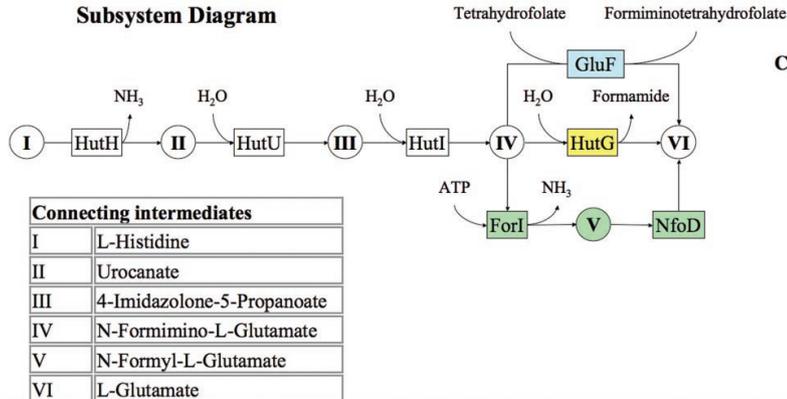
**B**

base de données sur processus biologiques

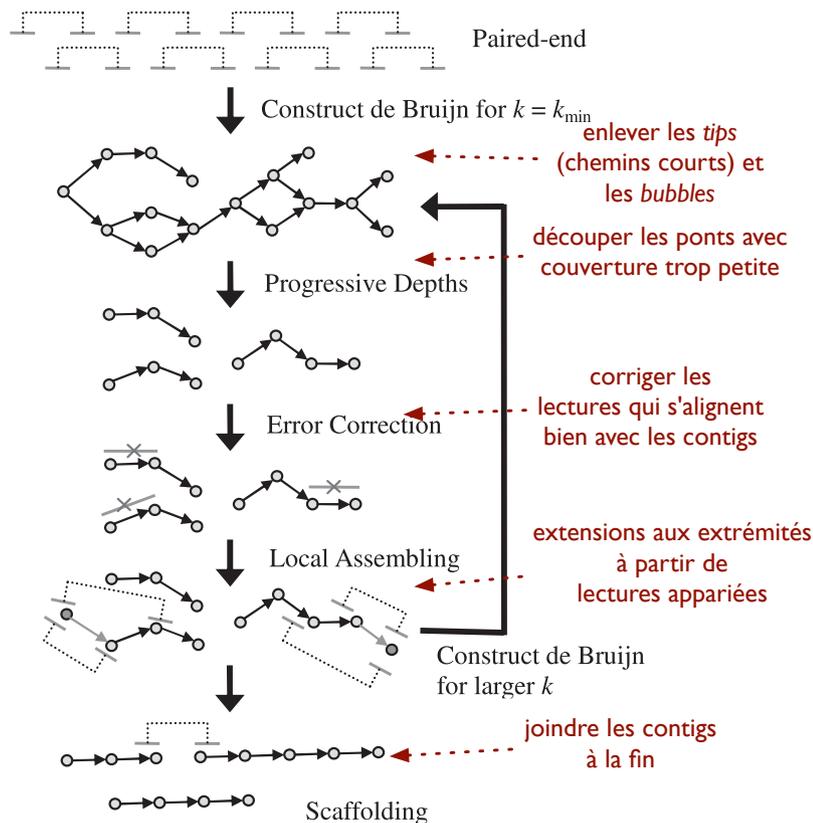
vocabulaire contrôlée

← sous-systèmes (SEED)

(autre BD utile : KEGG)



# Assemblage metagénomique (IDBA-UD)

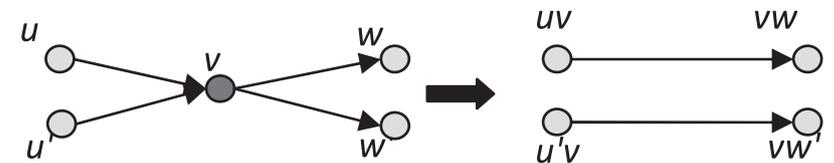


problèmes : bcp de séquences hétérogènes ; couverture variable à travers des génomes

de Bruijn :

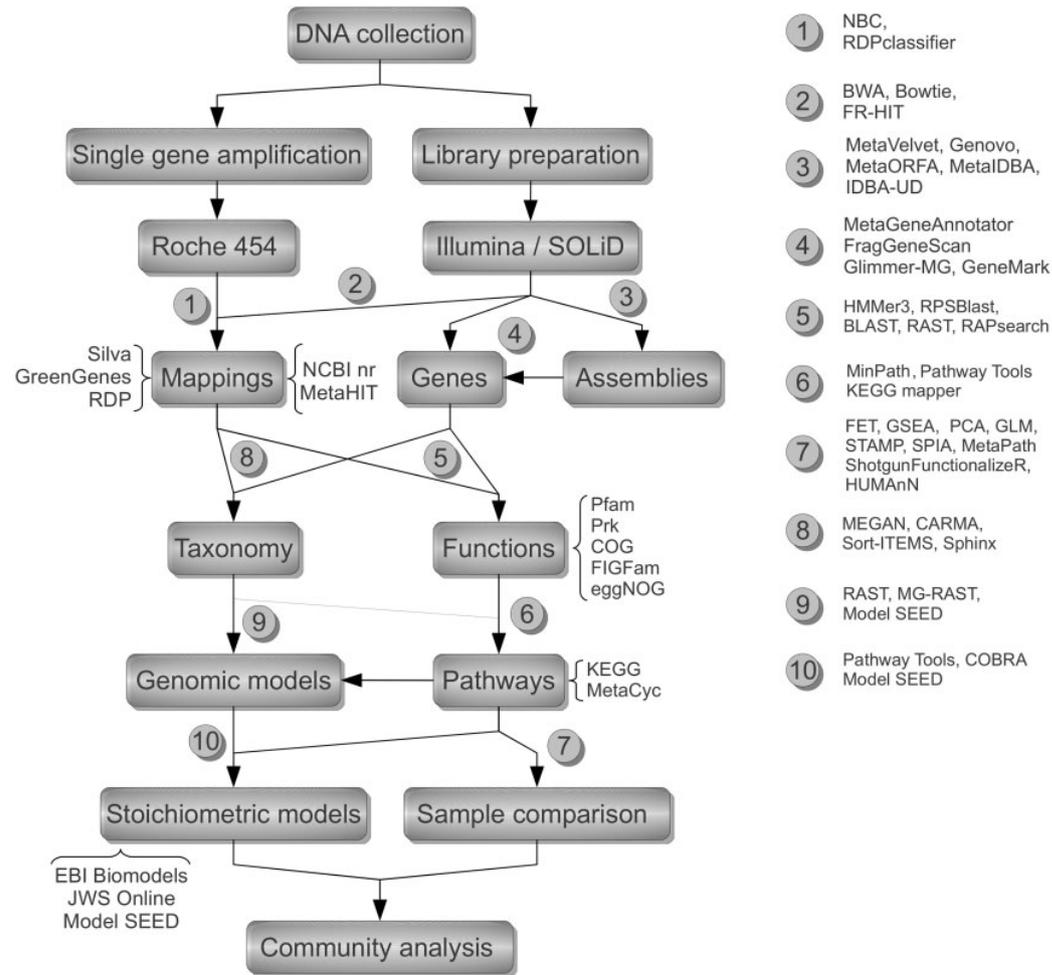
petit  $k \rightarrow$  trop de branchements (ambiguïté) ; grand  $k \rightarrow$  trop d'erreurs, moins de contigs

IDBA-UD : travailler avec  $k, k + 1, k + 2 \dots$  ( $\Rightarrow$  fusion / désambiguation de chemins)



**Fig. 3.** Example of resolving repeats by iteration from  $k$  to  $k + 1$ . The repeat region is a single  $k$ -mer,  $uvw$  and  $u'vw'$  appear in the genome. After the iteration, repeat  $v$  is resolved

# Outils



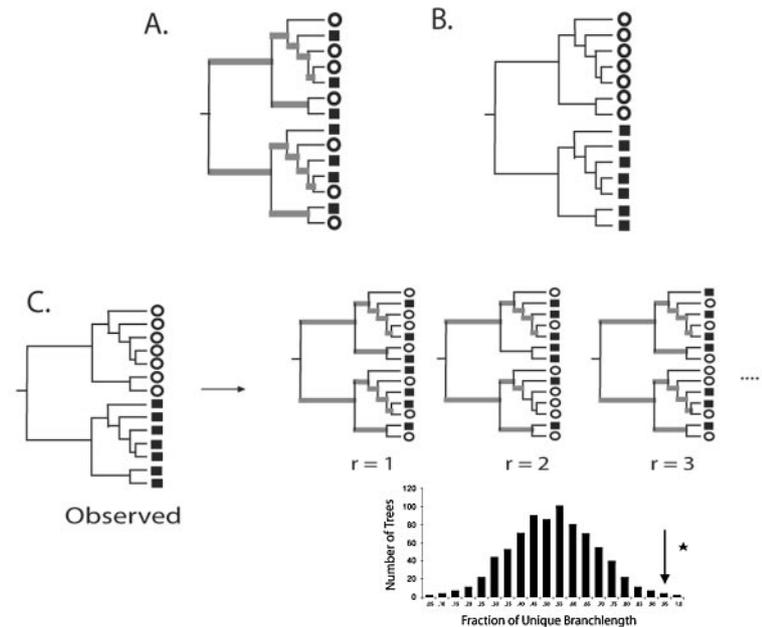
De Filippo *Briefings in Bioinformatics* 13 :696 (2012)

# Comparaisons

UniFrac : % de l'arbre qui est unique (mesure la distance entre la composition taxonomique de deux échantillons)

→ visualisation par positionnement multidimensionnel (*multidimensional scaling*)

— PCoA = principal coordinate analysis



Lozupone & Knight *Applied and Environmental Microbiology* 71 :8228 (2005)

# Ecosystèmes

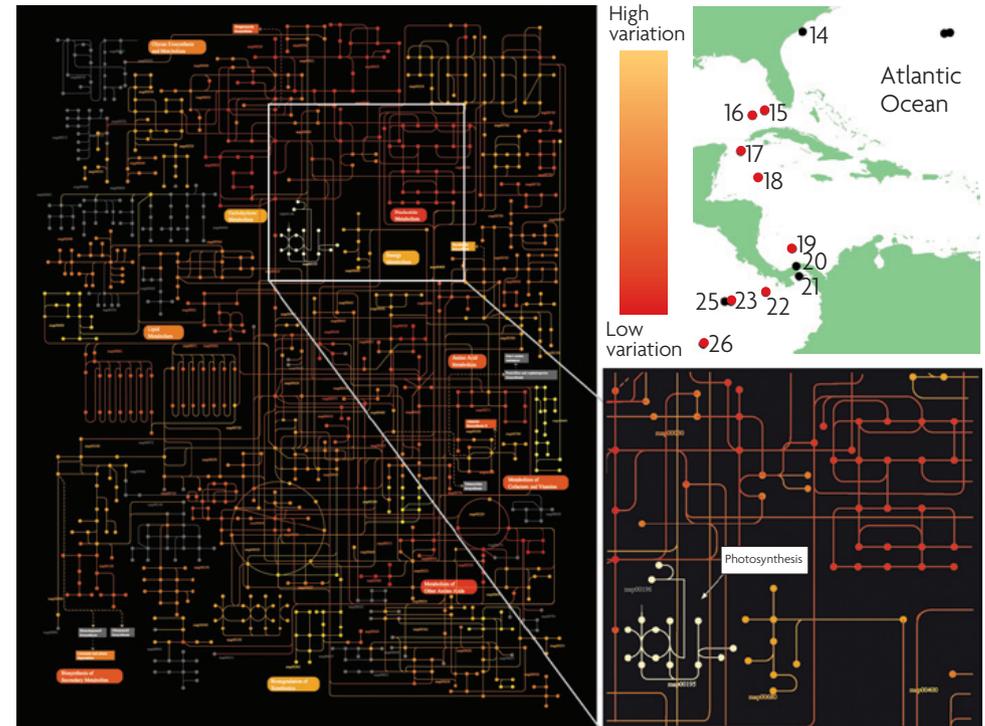
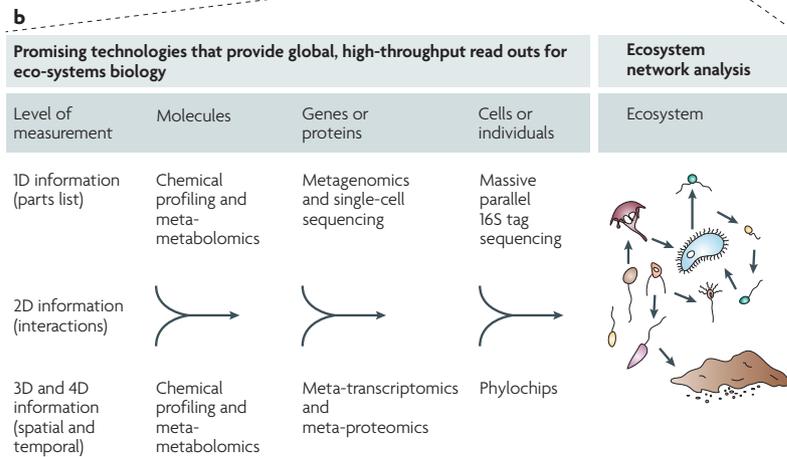
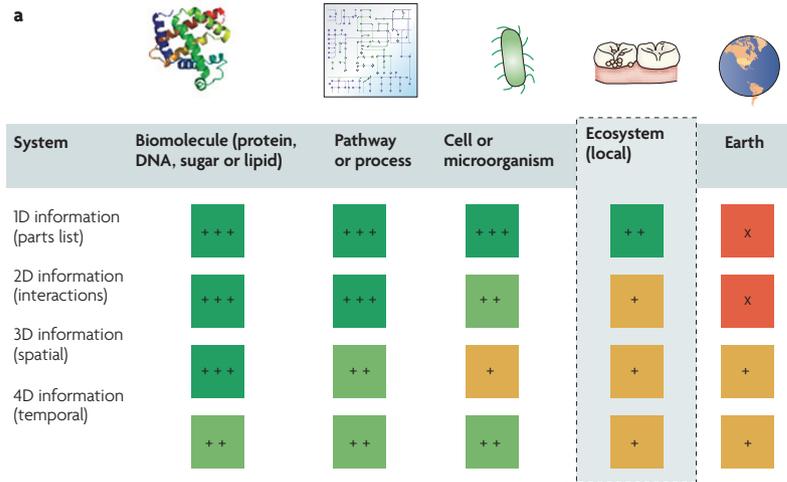
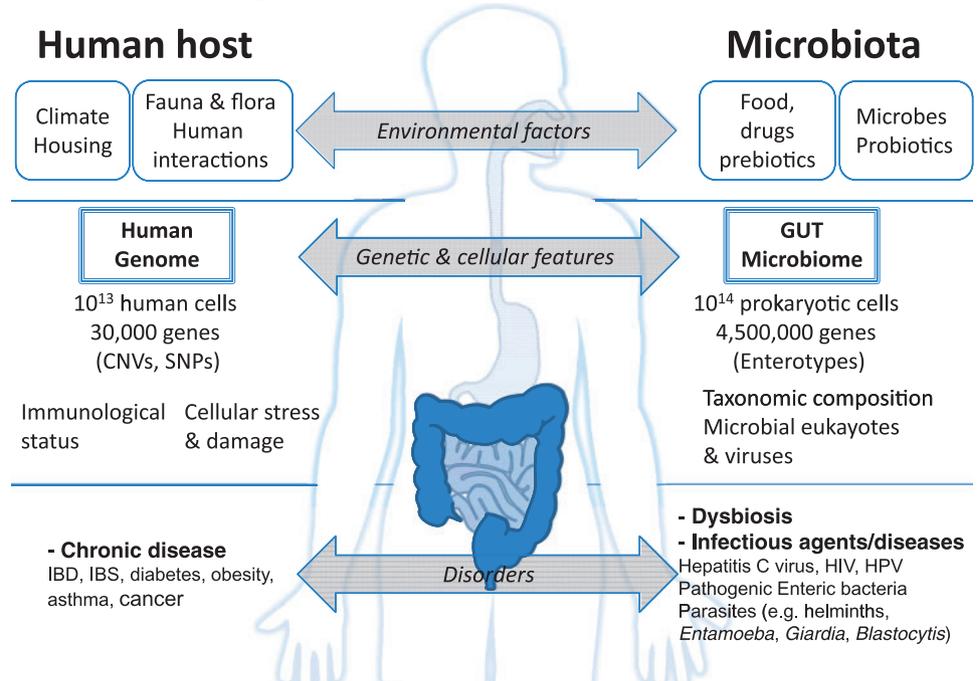
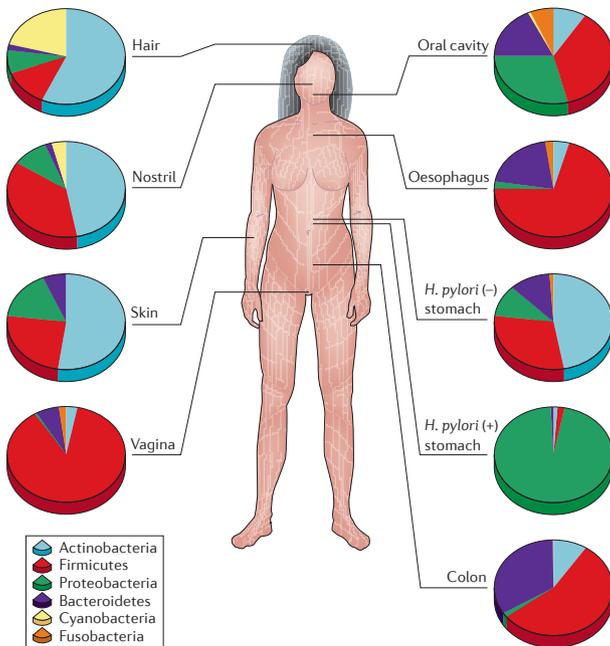


Figure 3 | **Visualizing complex environmental patterns.** Novel visualization techniques will be needed to describe complex data and patterns. The example shown here is a summary of the metabolic variation along a longitudinal transect of ocean surface water samples (data from REF. 42; the samples used (red) were selected for similarity in habitat type). Colour intensity shows the contribution to the overall variance for different KEGG<sup>84</sup> maps that are involved in central metabolism (for example, red indicates maps with low contribution and therefore low variability over sites, whereas yellow indicates maps with high contribution) and grey indicates no significant KEGG mapping for these samples. The inset shows the large contribution of photosynthesis to the overall functional variation among samples.

Raes & Bork *Nat Rev Microbiol* 6 :1 (2008)

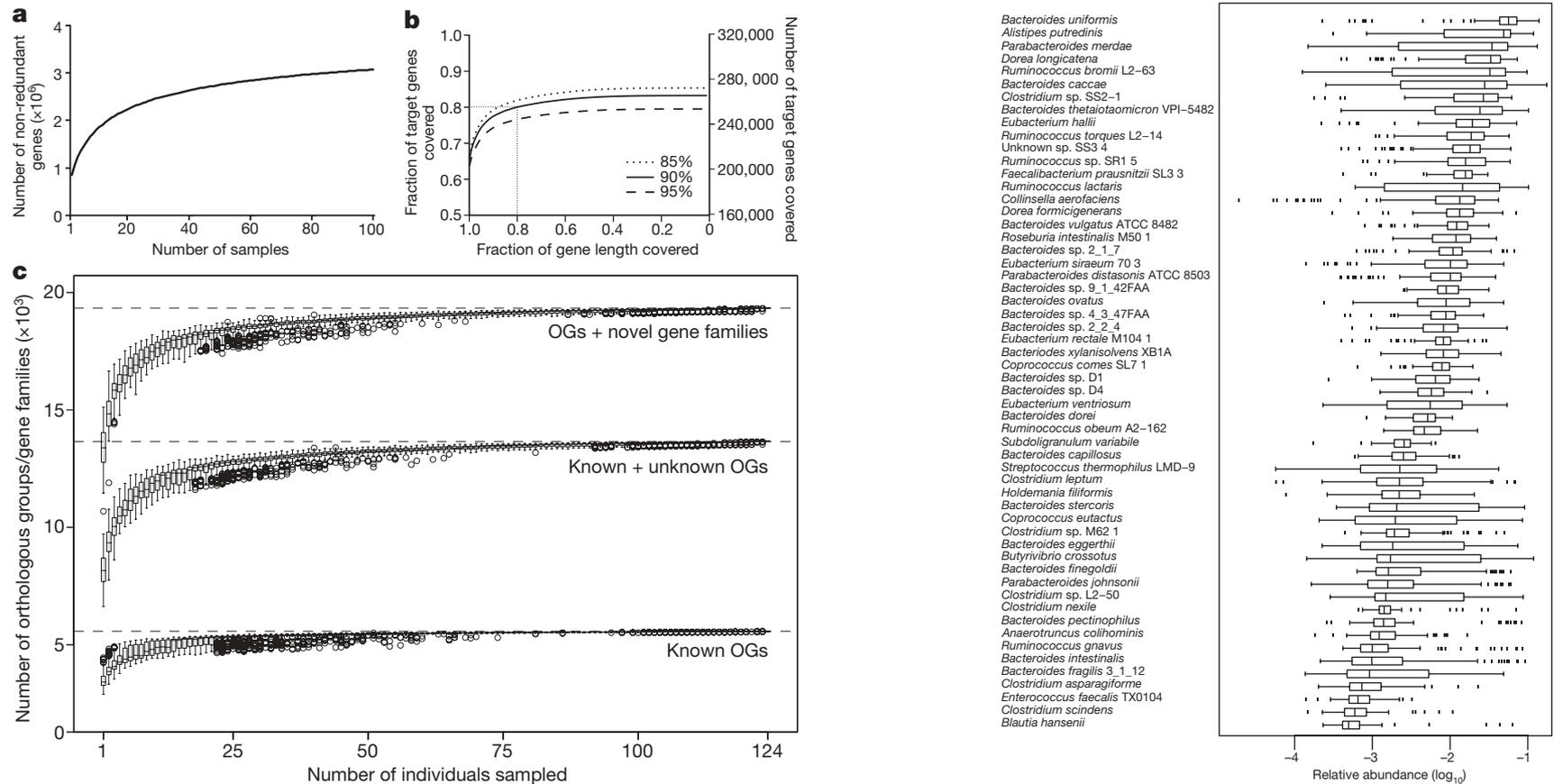
# Microbiome humain

10 fois plus de cellules, 100 fois plus de gènes que l'humain



Cho & Blaser *Nat Rev Genet* 13 :260 (2012) ; Collison & al *Briefings in Bioinformatics* 13 :751 (2012)

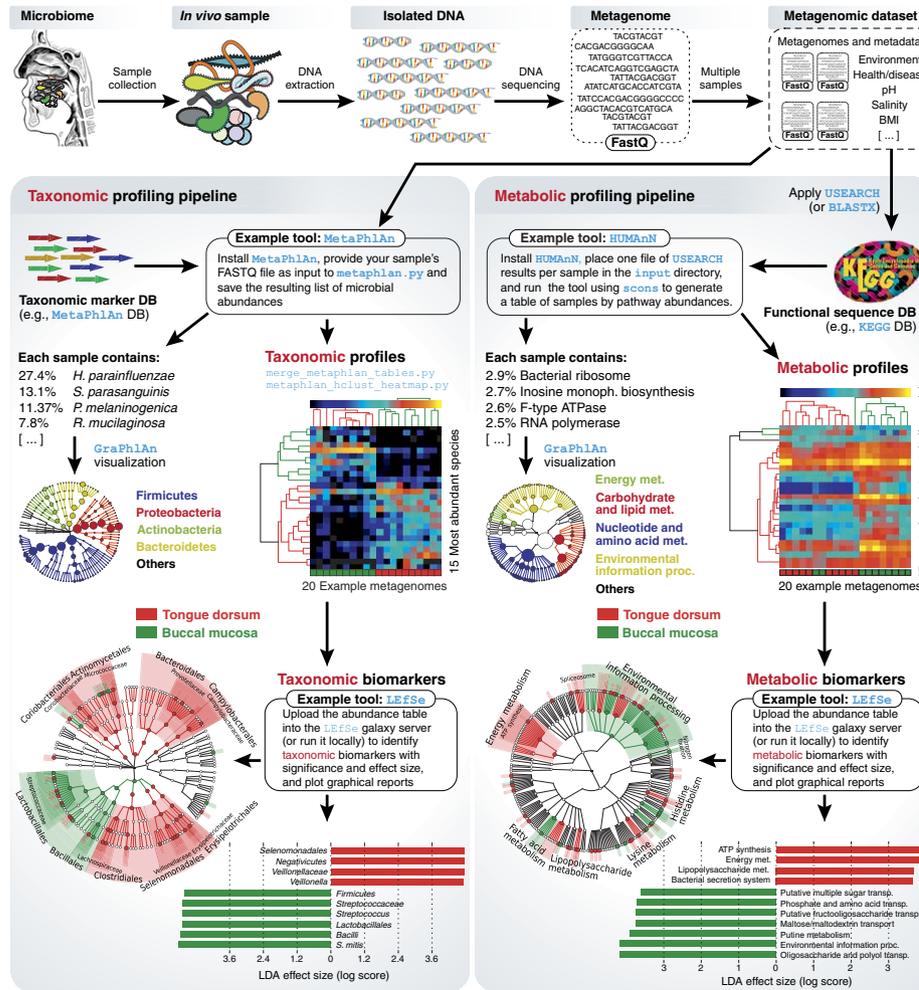
# Catalogue de gènes et organismes



prédiction de gènes : modèles pour séquences partielles, groupage par similarité

Qin, Li & al. *Nature* 464 :59 (2010)

# Analyse comparatif

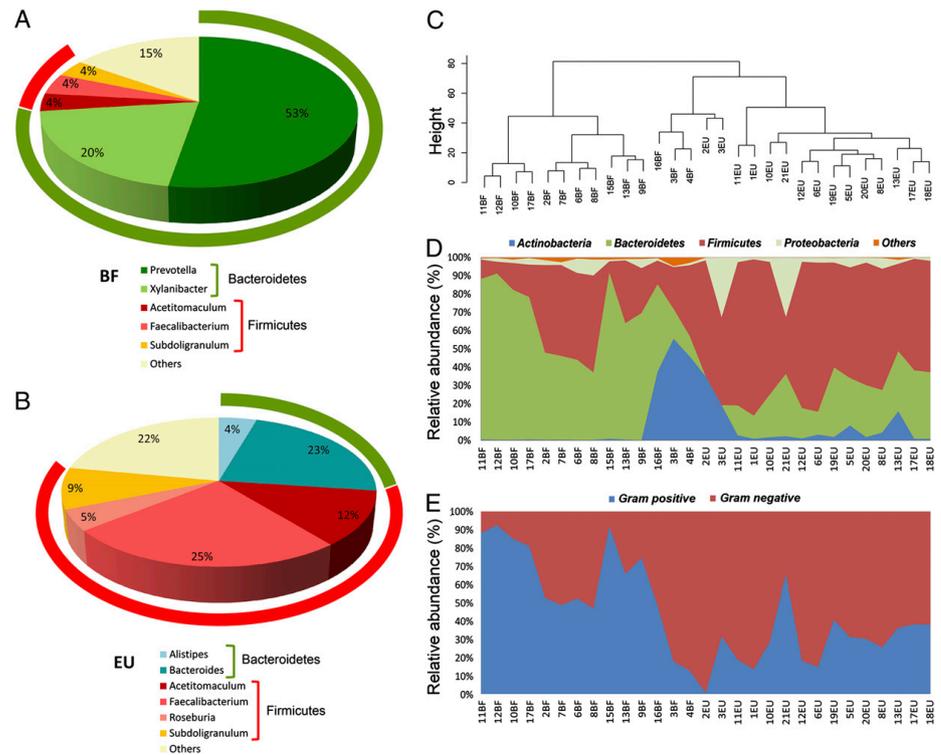


Segata & al *Molecular Systems Biology* 9 :666 (2013)

# Diète

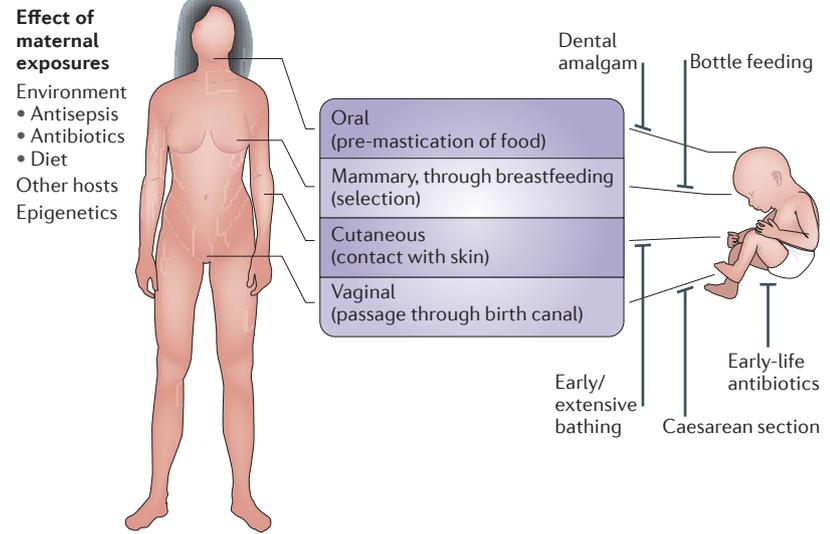
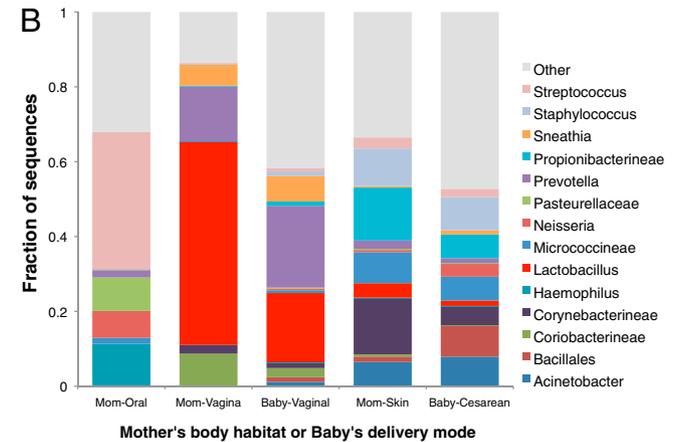
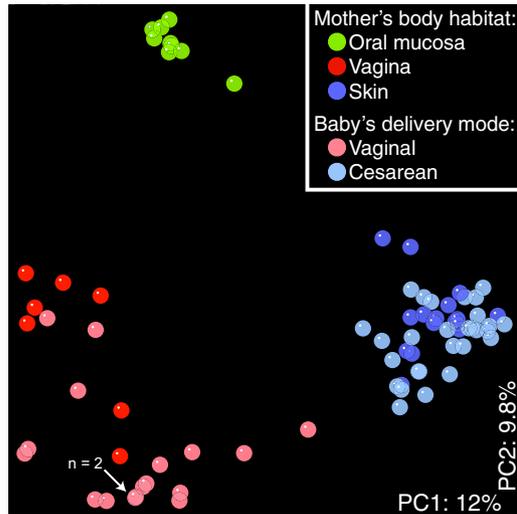
enfants de la ville de Florence, Italie et du village du Boulpon, Burkina Faso

→ diète riche en fibre → présence de bactéries qui consomment les polysaccharides



De Filippo & al. *PNAS* 107 :14691 (2010)

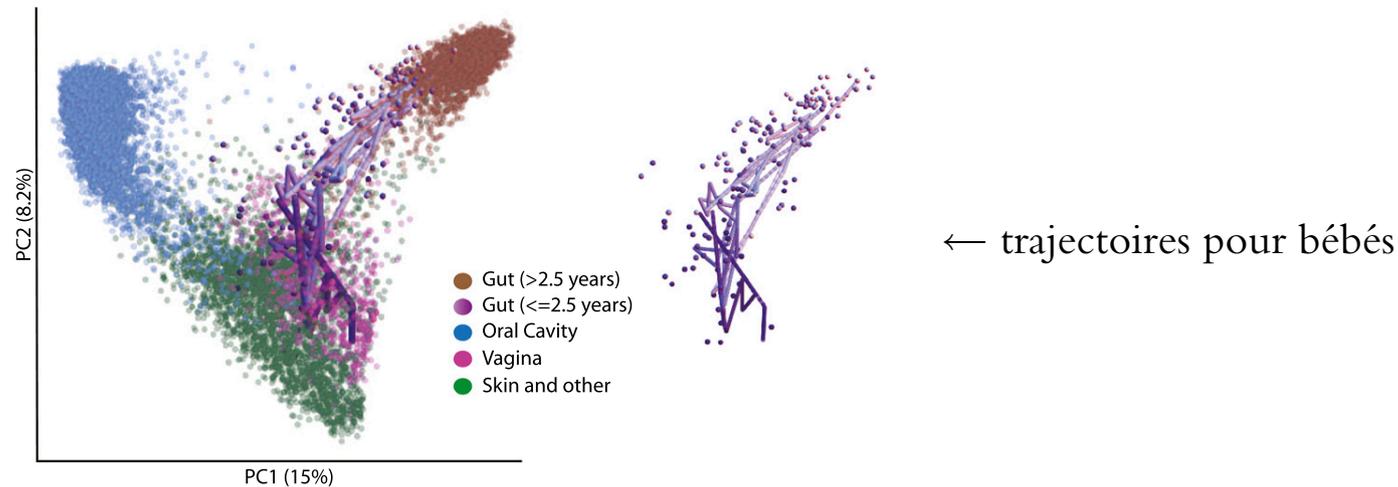
# Développement de notre microbiome



Dominguez-Bello & al. *PNAS* 107 :11971 (2010) ; Cho & Blaser *Nat Rev Genet* 13 :260 (2012)

# Quelques leçons

1. il existe des signatures universelles à travers des individus



2. corrélations avec maladies

Disease	Relevant finding
Psoriasis	Increased ratio of Firmicutes to Actinobacteria
Reflux oesophagitis	Oesophageal microbiota dominated by gram-negative anaerobes; gastric microbiota with low or absent <i>Helicobacter pylori</i>
Obesity	Reduced ratio of Bacteroidetes to Firmicutes
Childhood-onset asthma	Absent gastric <i>H. pylori</i> (especially the cytotoxin-associated gene A ( <i>cagA</i> ) genotype)
Inflammatory bowel disease (colitis)	Larger populations of Enterobacteriaceae
Functional bowel diseases	Larger populations of <i>Veillonella</i> and <i>Lactobacillus</i>
Colorectal carcinoma	Larger populations of <i>Fusobacterium</i> spp.
Cardiovascular disease	Gut-microbiota-dependent metabolism of phosphatidylcholine

Lozupone & al. *Genome Res* 23 :1704 (2013) ; Cho & Blaser *Nat Rev Genet* 13 :260 (2012)