

Metagenómica

Metabarcoding

Dr. Luis Castañeda

Programa de Genética Humana – Instituto de Ciencias Biomédicas
Facultad de Medicina – Universidad de Chile

Santiago, 10 de Mayo de 2021

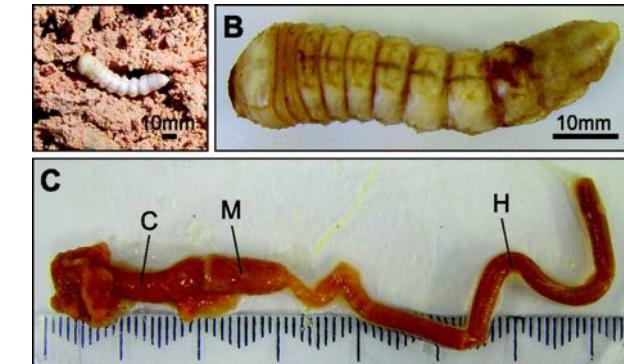
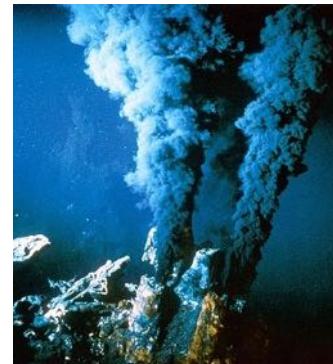


1%!!

99%

A green curly brace is positioned vertically on the right side of the image, grouping the two percentages together. The top percentage, "1%!!", is written in red. The bottom percentage, "99%", is also written in red.

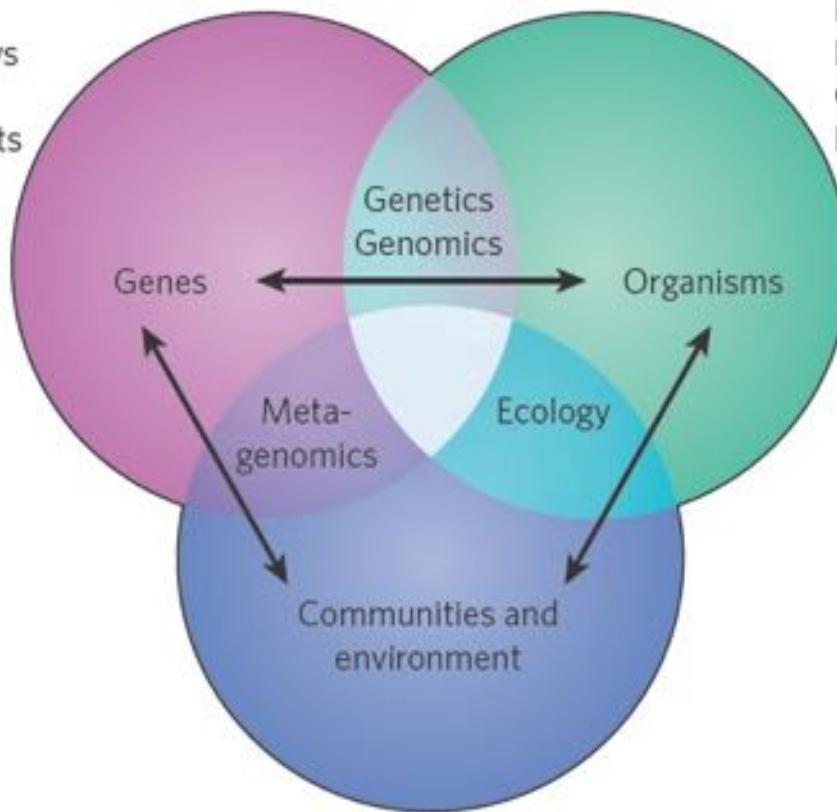
Microorganismos viven en ambientes extremos



Metagenómica

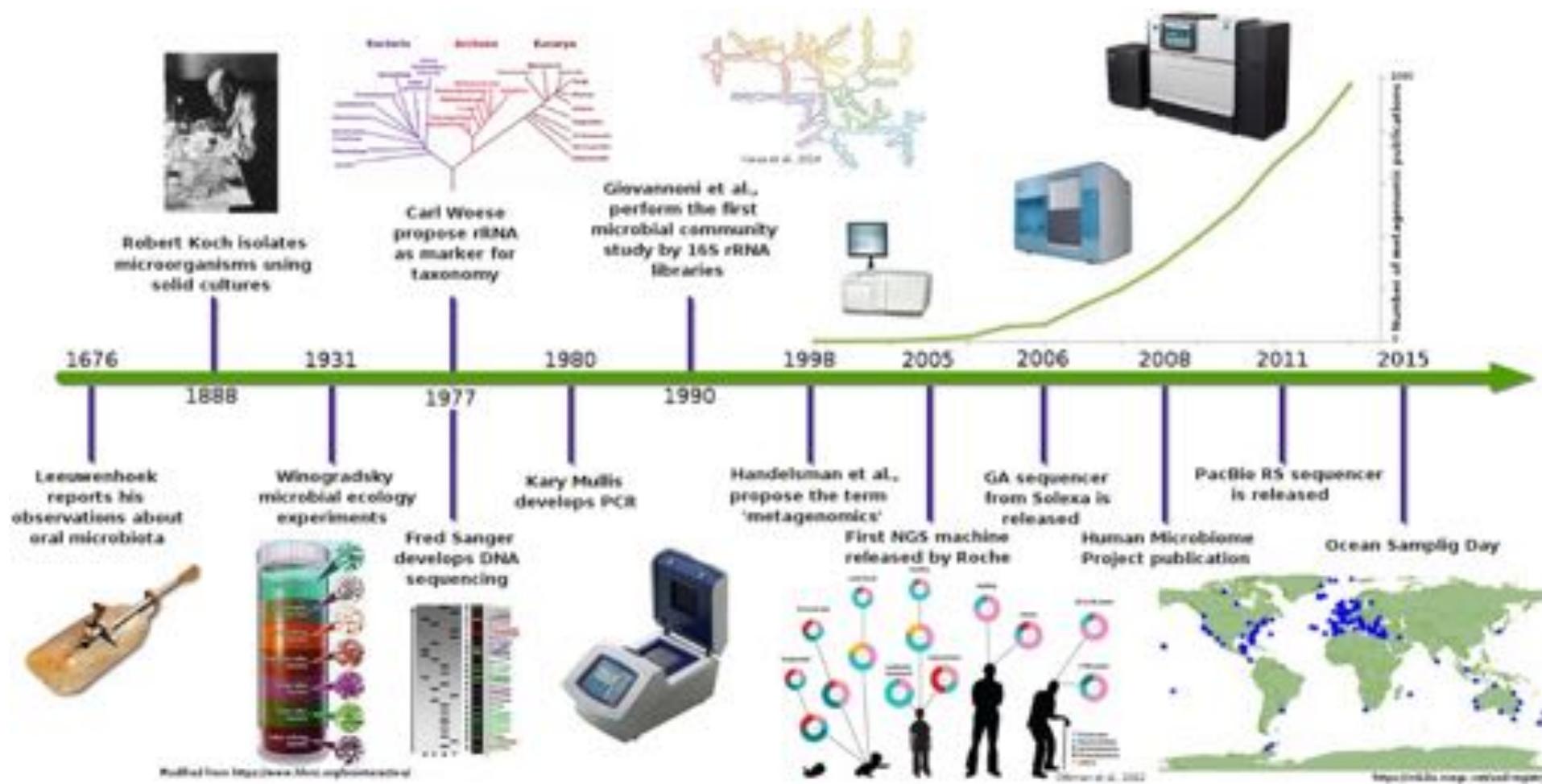
Genotype
Allelic diversity
Metabolic pathways
Functional guilds
Regulatory elements

Phenotype
Physiology
Genetics
Regulation



Environmental variability
Community composition
Population genetics
Functional redundancy
Biogeochemistry
Community dynamics
Ecosystem response

Secuenciación



Secuenciación masiva



Roche GS-FLX+ 454

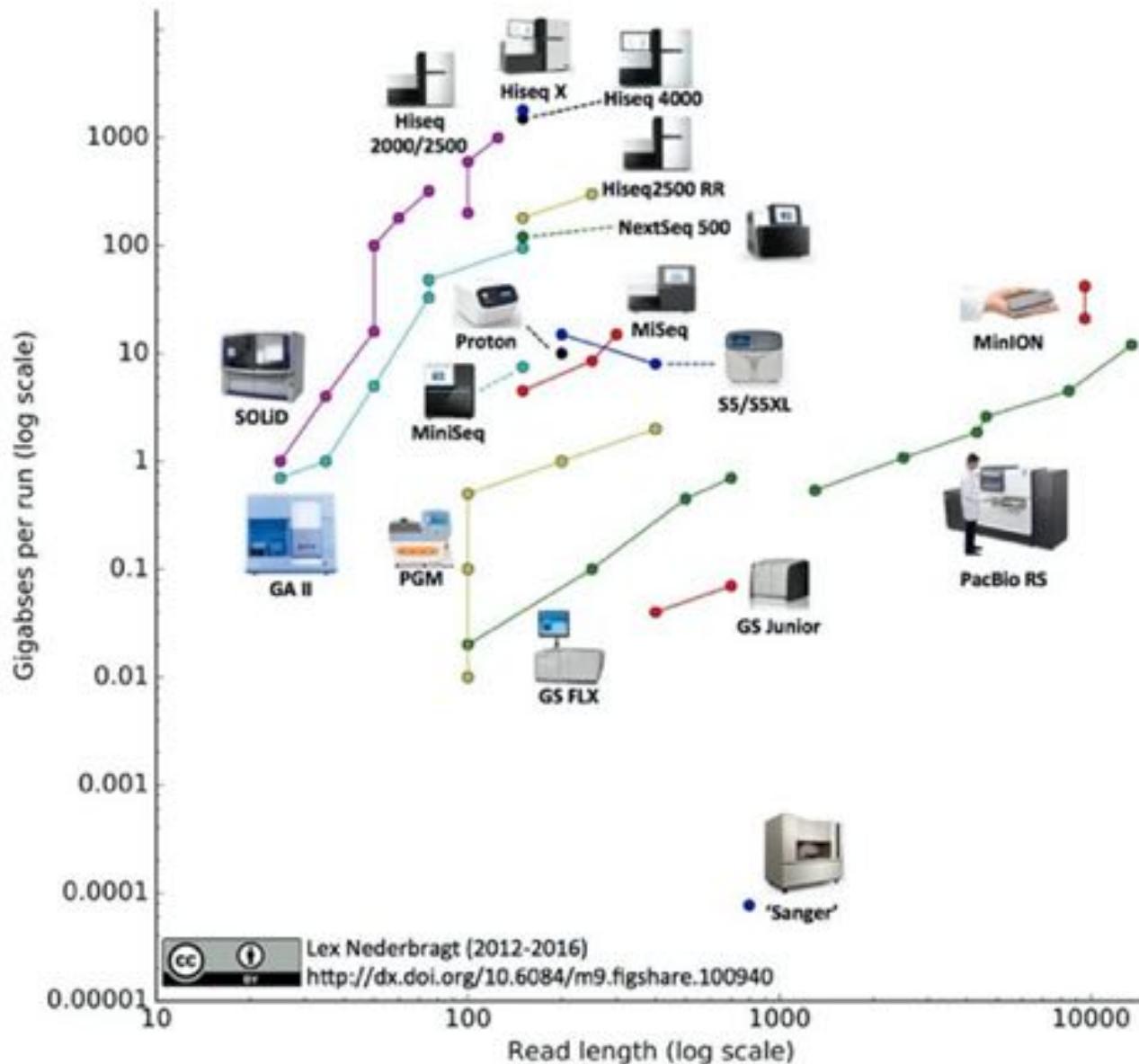
- Up to 1 M seqs
- 1000 bp max length
- 700 Mb per run



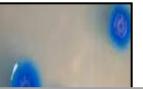
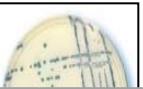
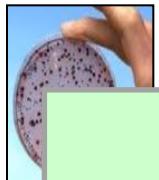
Illumina MiSeq

- Up to 25 M seqs
- 2x 300 bp max length
- 15 Gb per run

Secuenciación masiva



Método clásico



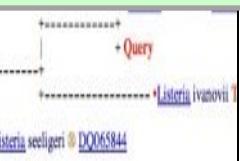
1 colonia

1 día

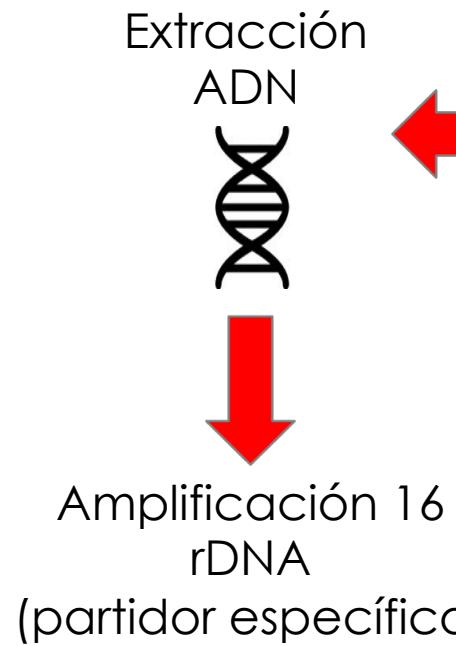
Temporalmente largas
Mucho trabajo de laboratorio
Pocas muestras

cteriana

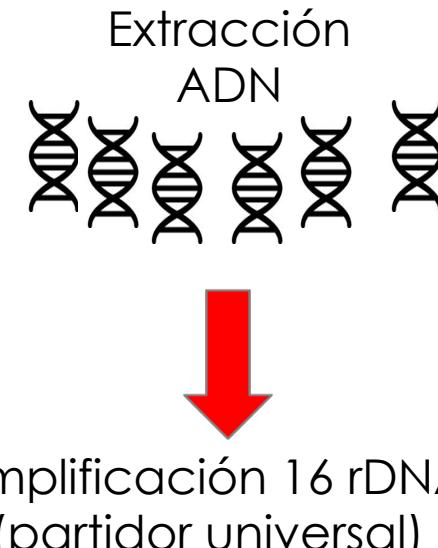
cterianas



DNA barcoding

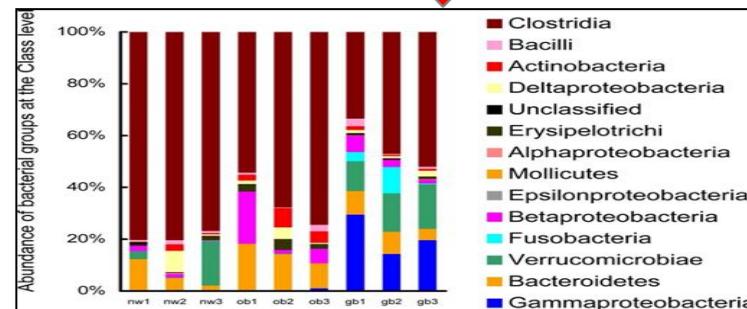


DNA metabarcoding

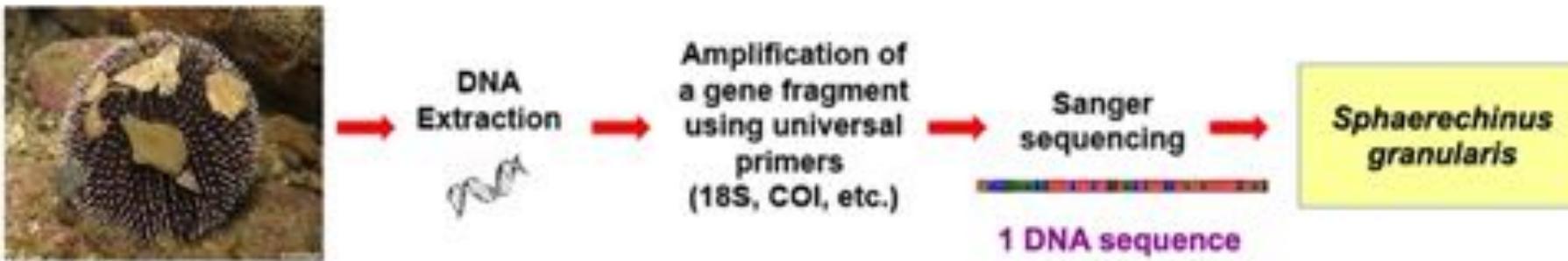


Amplificación 16 rDNA (partidor universal)

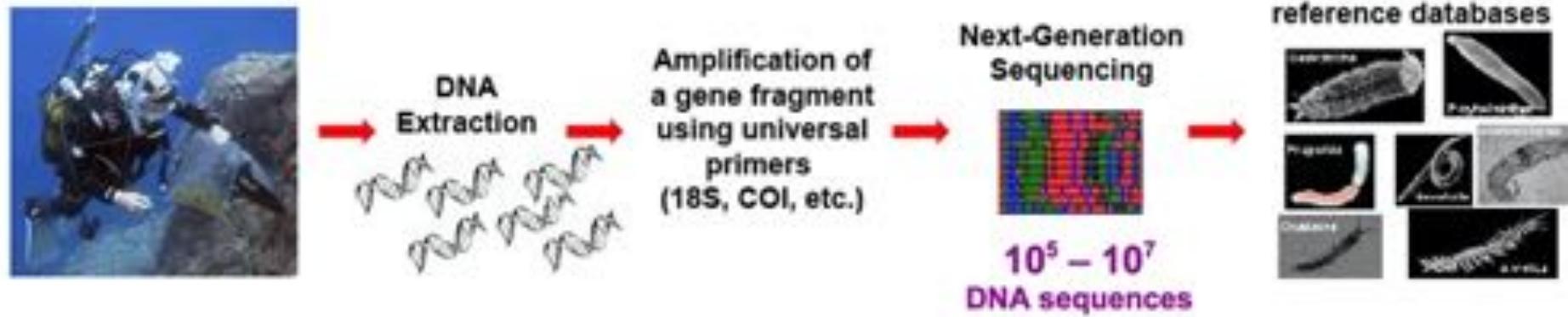
Secuenciación masiva (millones de secuencias)



DNA BARCODING



DNA METABARCODING



Ecología del microbioma

El microbiota representa la diversidad ecológica total de microorganismos (patogénicos, comensalista, simbióticos, etc.) que ocupan un cierto hábitat, comunidad o ecosistema.

Los microbiomas son extremadamente diversos en términos taxonómicos, genómicos, fisiológicos y morfológicos.

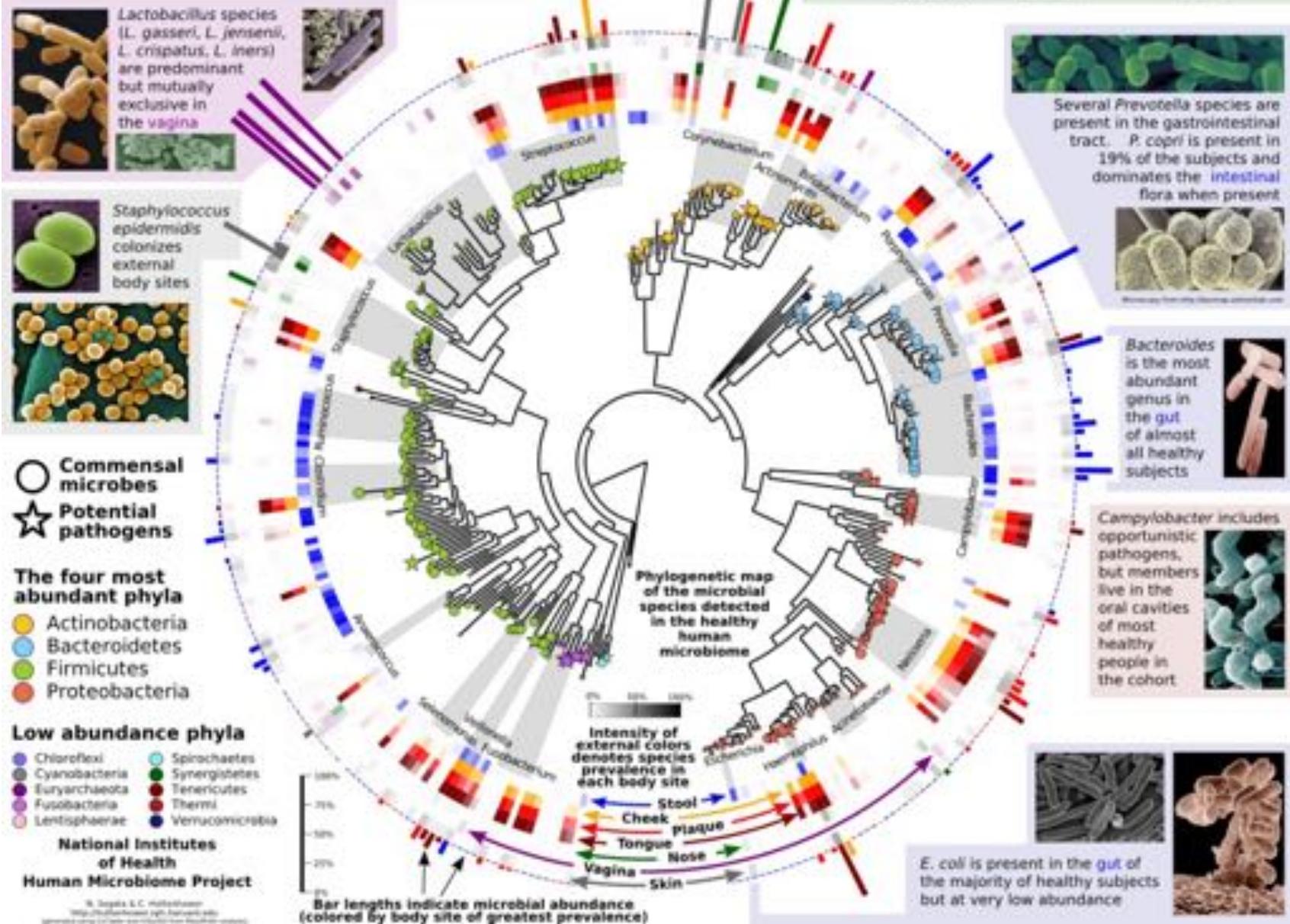
Se estima que la diversidad actual es alrededor de 10^7 a 10^9 especies.

La composición y abundancia de cada microbioma cambia desde microescalas hasta escalas globales.

Table 1.1 Abundance of microbial cells in different habitats (with permission from Fierer and Lennon (2011))

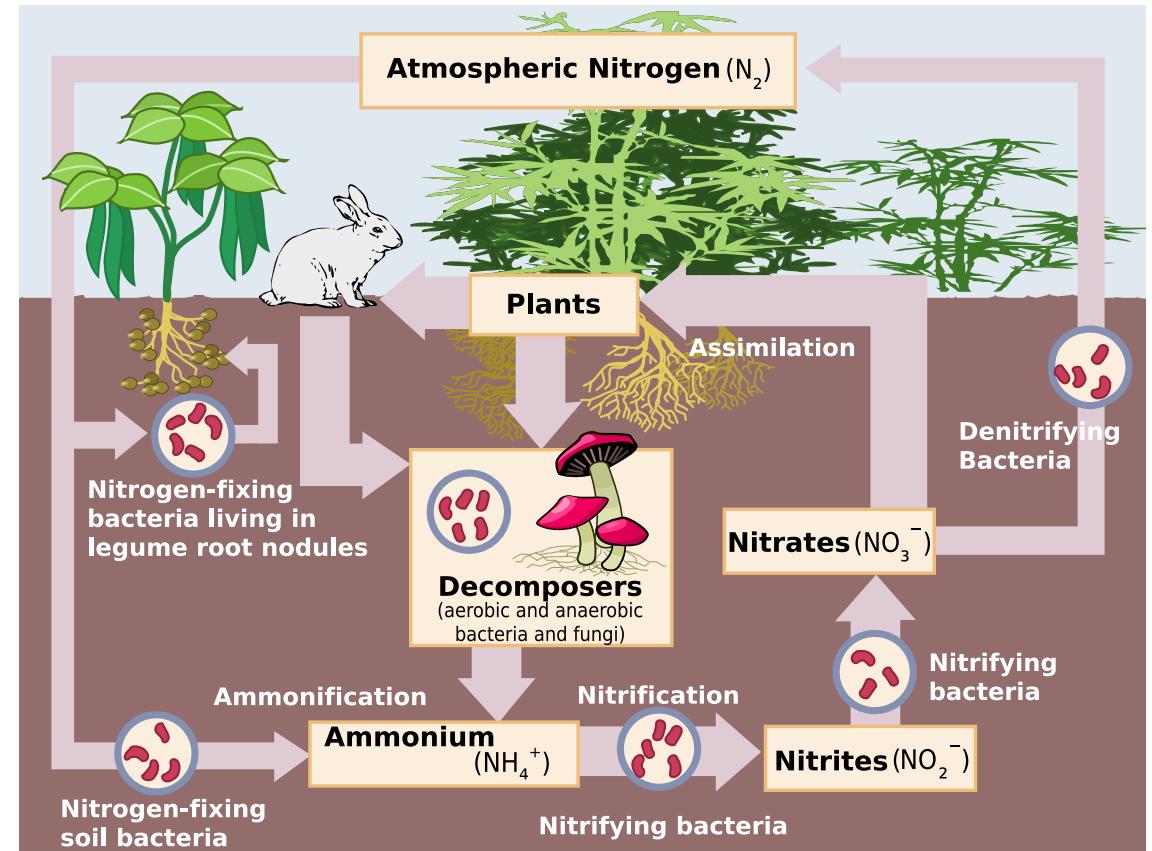
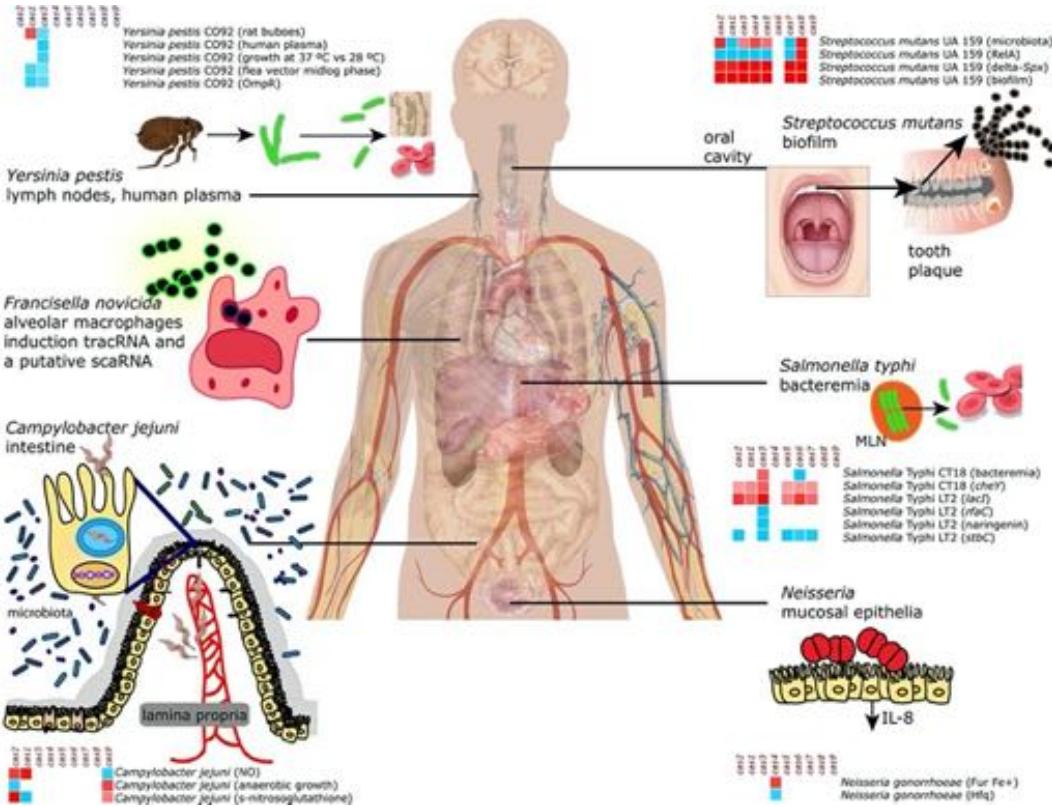
Habitat	Cell conc. (source)	Mean no. per sample (source)	Total volume of sample
Soil	10^{10} cells/cm ³ (Torsvik et al. 2002)	10,000 (Fierer et al. 2007) (Roesch et al. 2007)	1 cm ³
Air	10^6 cells/m ³ (Bowers et al. 2009)	100 (Bowers et al. 2009)	2.5–5.4 m ³
Human skin	10^7 cells/cm ² (Fredricks 2001)	200 (Fierer et al. 2008)	70 cm ²
Human gut	10^{12} cells/g feces (Vaahtovuo et al. 2005)	1000 (Dethlefsen et al. 2007; Turnbaugh et al. 2009)	0.2 g
Human mouth (saliva)	10^8 cells/mL saliva (Lauber et al. unpublished manuscript)	5700 (Keijser et al. 2008)	10 mL
Marine water	10^5 – 10^6 cells/mL sea water (Sogin et al. 2006)	4000 (Sogin et al. 2006)	1000 mL
Marine hydrother- mal vent fluids	10^5 cells/mL (Huber et al. 2007)	260 (Huber et al. 2007)	3 L
Glacial ice	10^6 cell/mL (Sharp et al. 1999)	100 (Simon et al. 2009)	100 mL
Sewage	10^9 cells/mL (McLellan et al. 2010)	2500 (McLellan et al. 2010)	100 mL
Anoxic subglacial lake	10^6 cells/mL (Gaidos et al. 2008)	70 (Gaidos et al. 2008)	250 mL
Baltic sea (brackish surface water)	10^6 cells/mL (Andersson et al. 2009)	1500 (Andersson et al. 2009)	5000 mL
Leaf surfaces	10^5 cells/cm ² (Redford et al. 2010)	250 (Redford et al. 2010)	5000 cm ²

A map of diversity in the human microbiome



Microbioma

El microbioma desempeña muchos procesos biológicos, ecológicos y biogeoquímicos desde sistemas de virulencia en hospederos hasta ciclaje de nutrientes a escala planetaria.

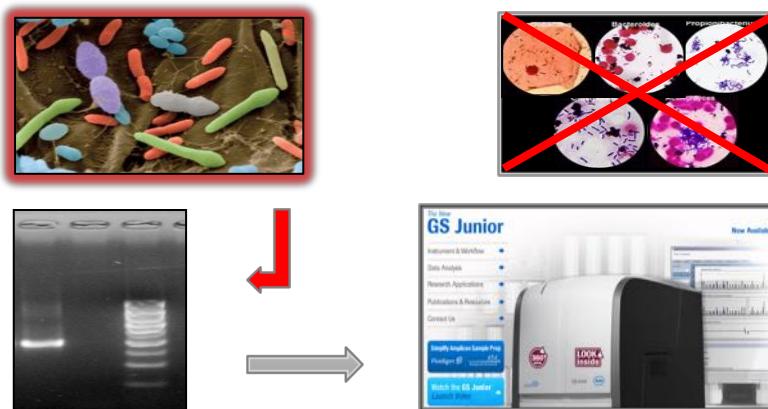


Microbioma

La diversidad del microbioma muestra distintos patrones a escala local, regional y global.



Secuenciación masiva



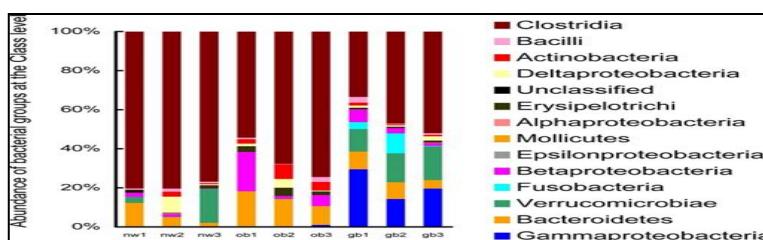
1 análisis implica...

identificación de toda la diversidad presente en la muestra
(dependiendo si hay secuencias en las bases de datos).

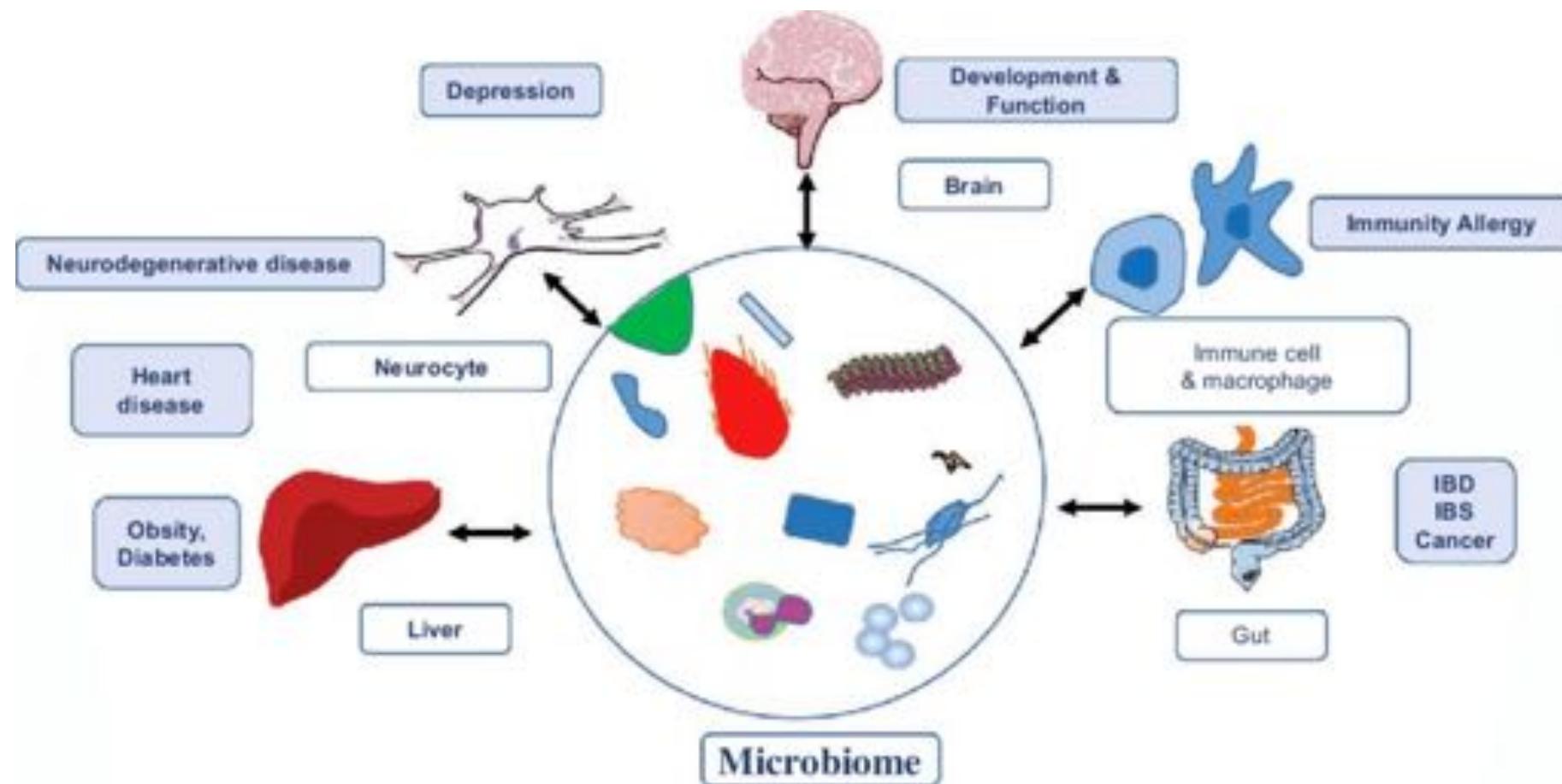
Permite evaluar la presencia de organismos prácticamente imposibles de aislar en el laboratorio:

Bacterias
Hongos

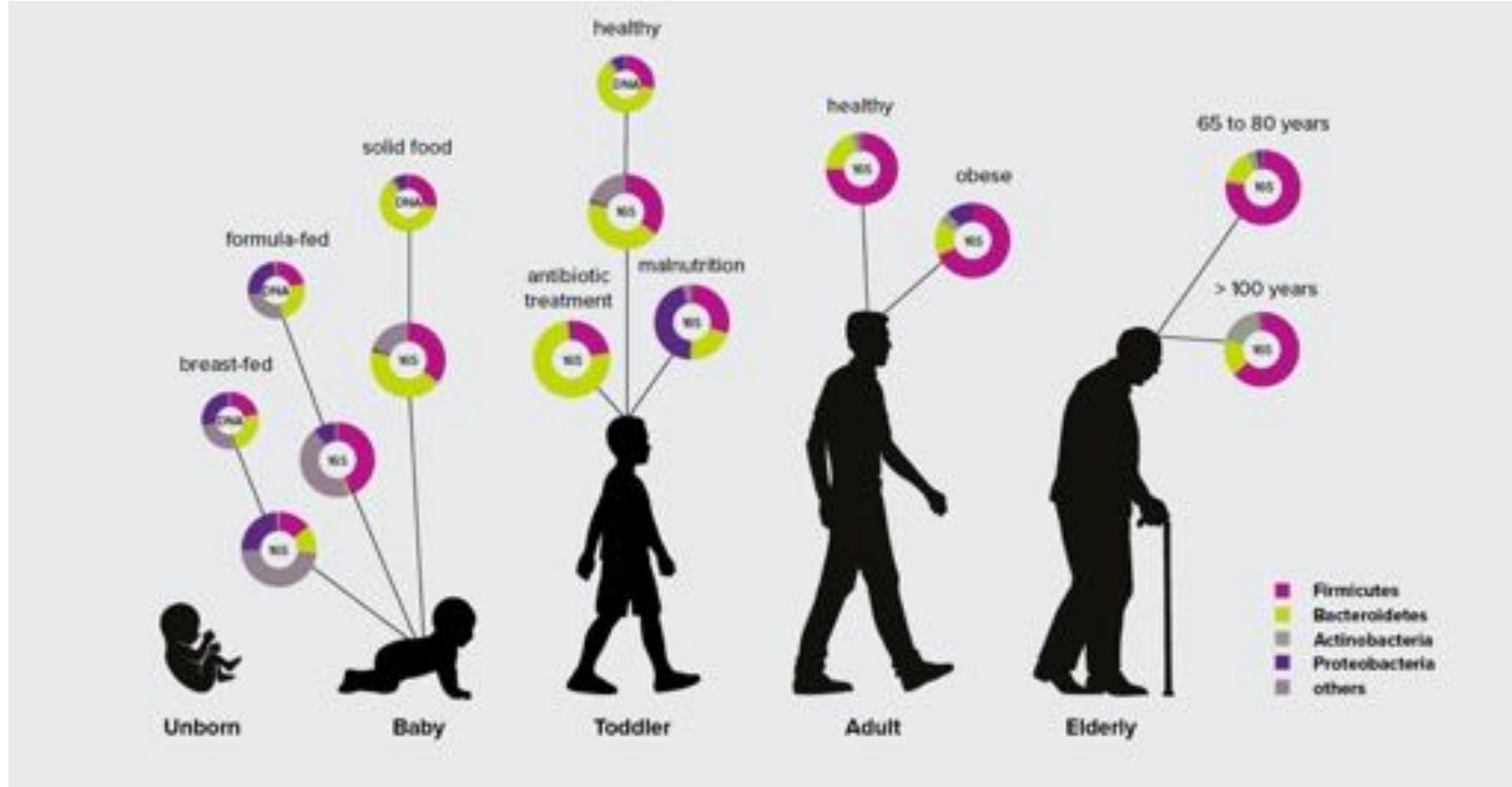
Viruses
Archeas



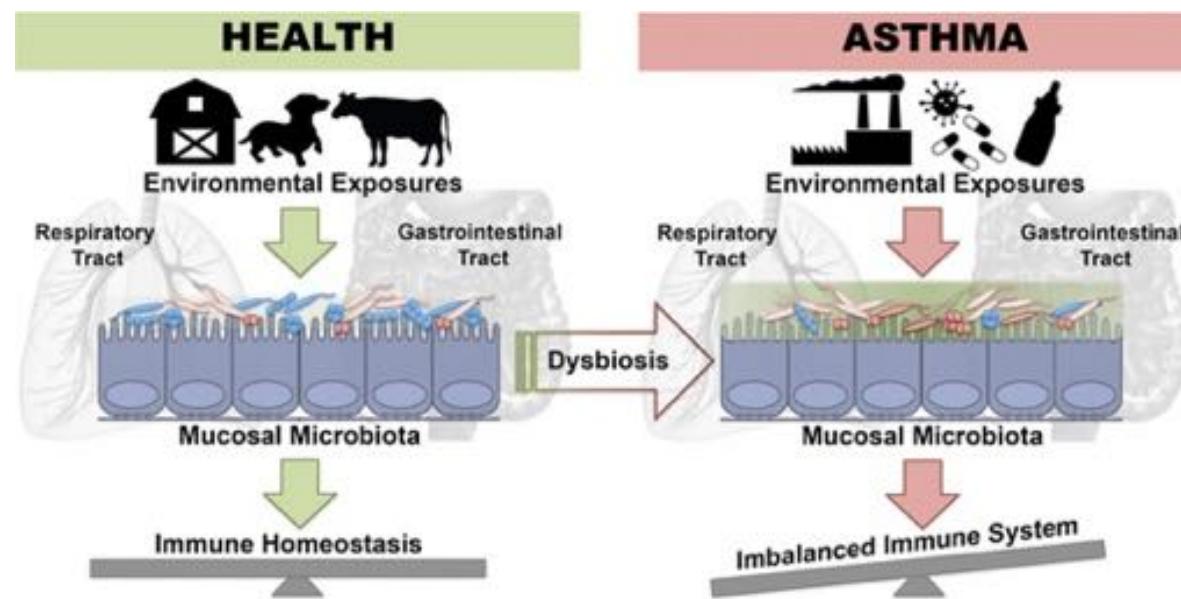
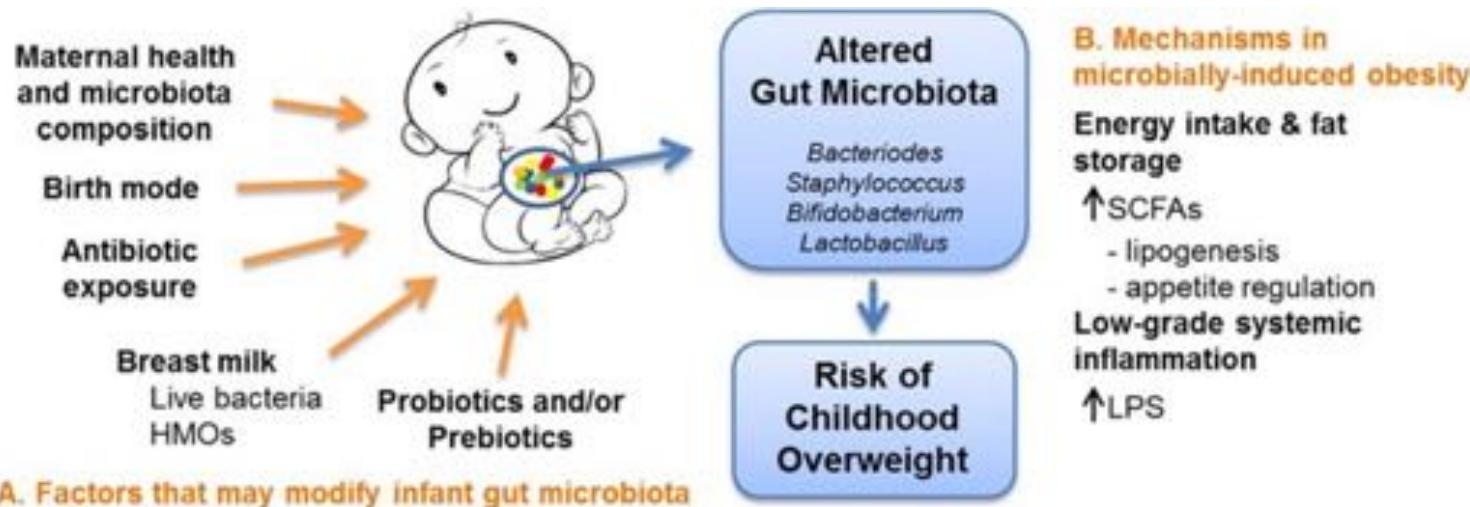
Microbiota intestinal en humanos



Microbiota intestinal en humanos



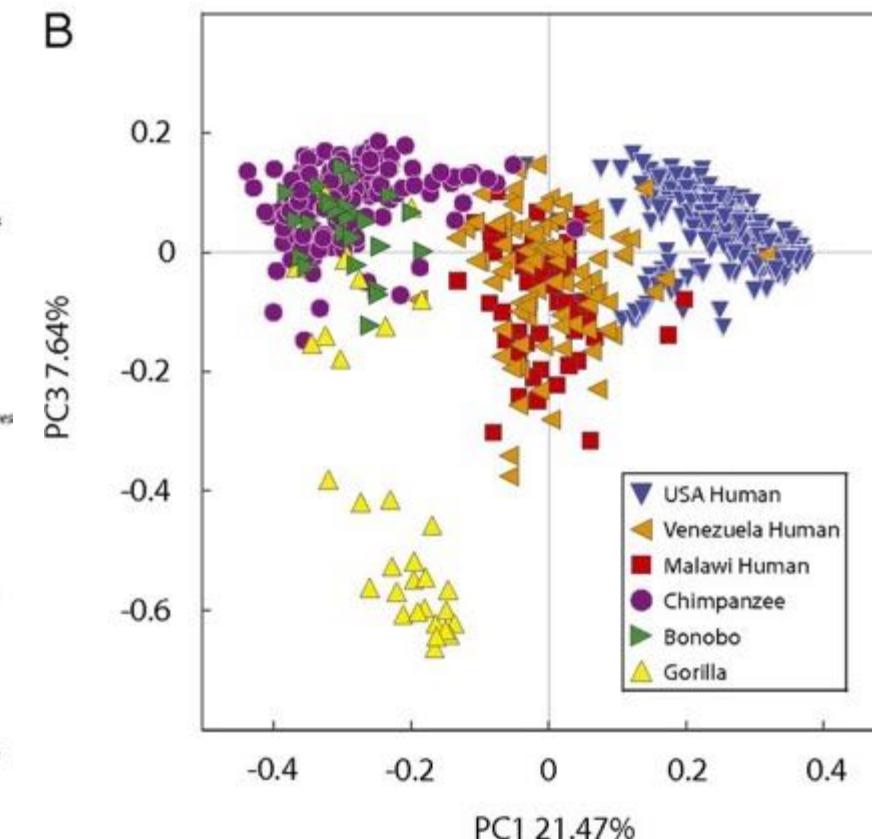
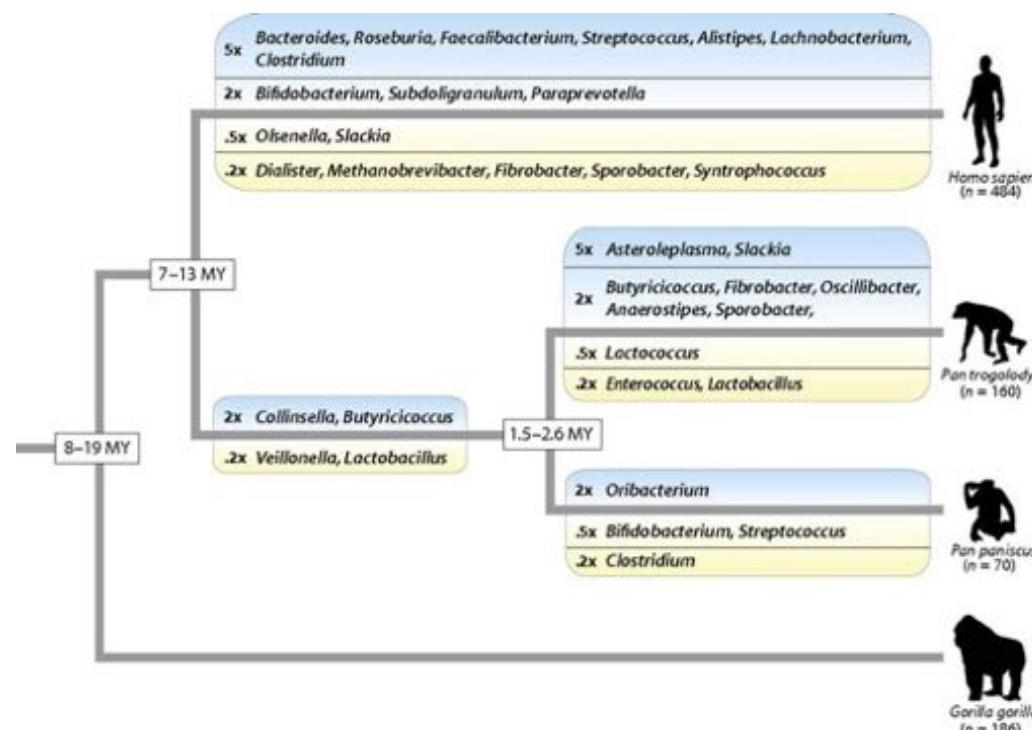
Microbiota en humanos

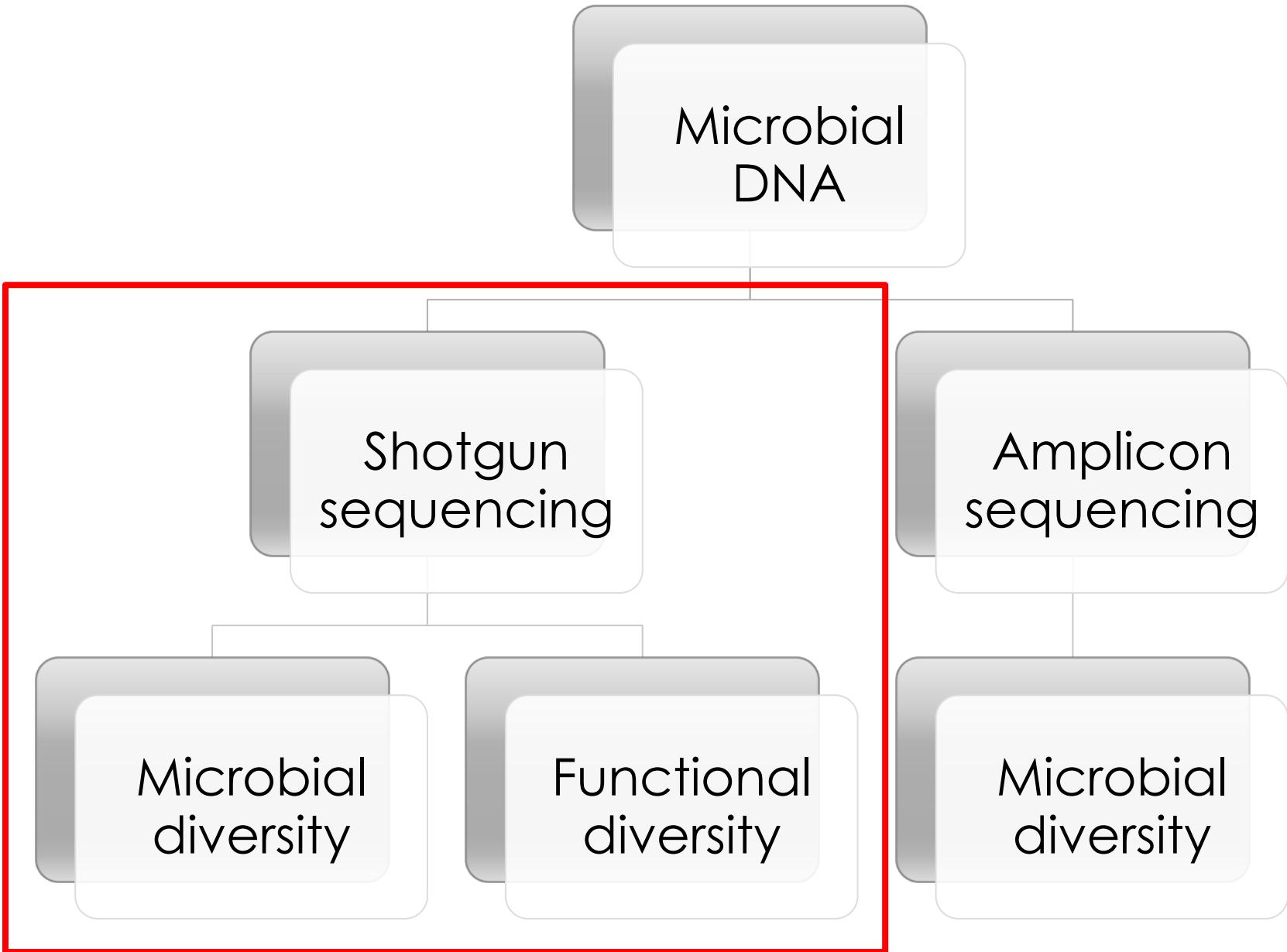


Coevolución microbioma-humano

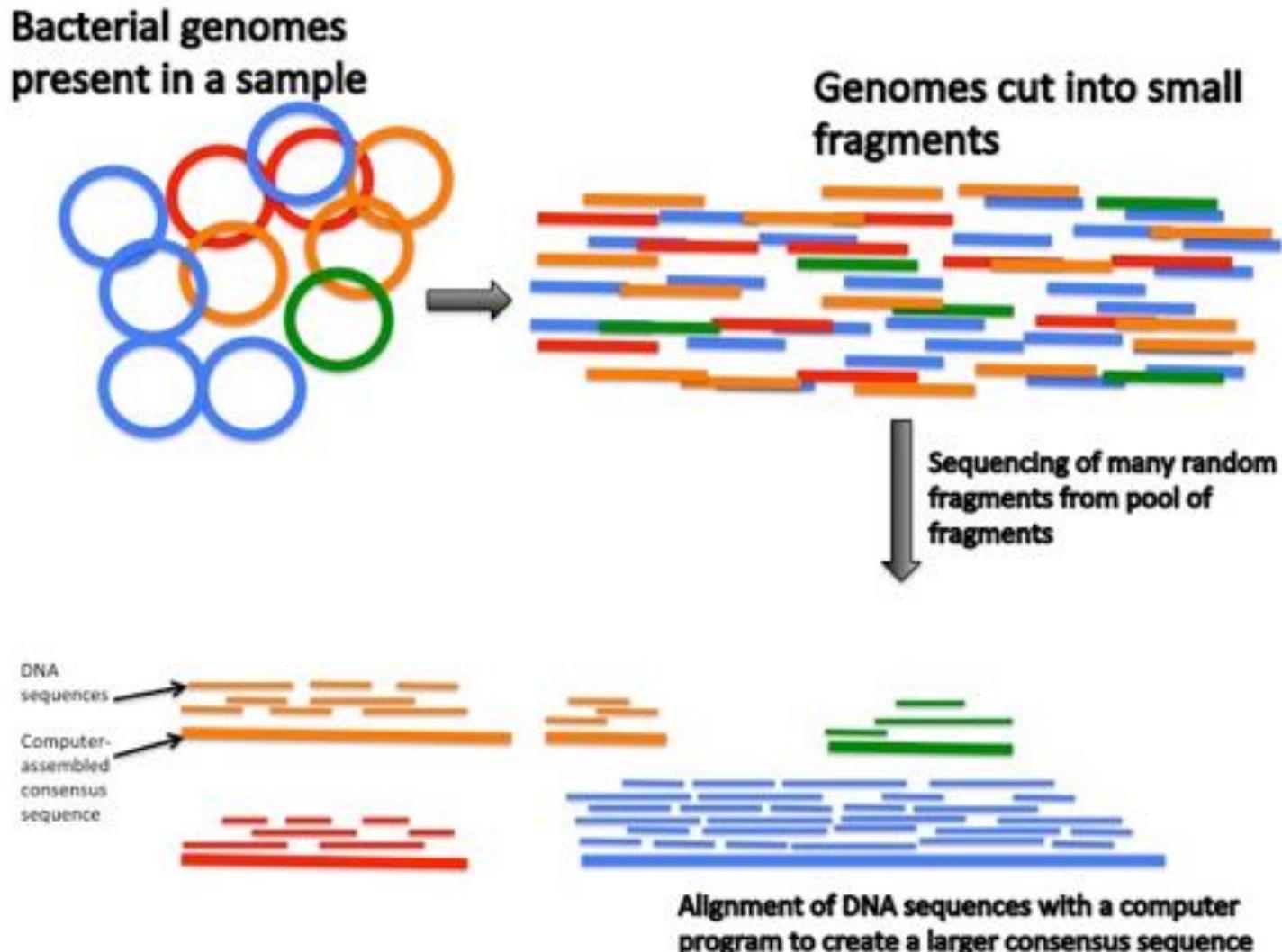
Rapid changes in the gut microbiome during human evolution

Andrew H. Moeller^{a,b}, Yingying Li^{c,d}, Eitel Mpoudi Ngole^e, Steve Ahuka-Mundeke^{f,g}, Elizabeth V. Lonsdorf^{h,i}, Anne E. Pusey^j, Martine Peeters^g, Beatrice H. Hahn^{c,1}, and Howard Ochman^{b,1}

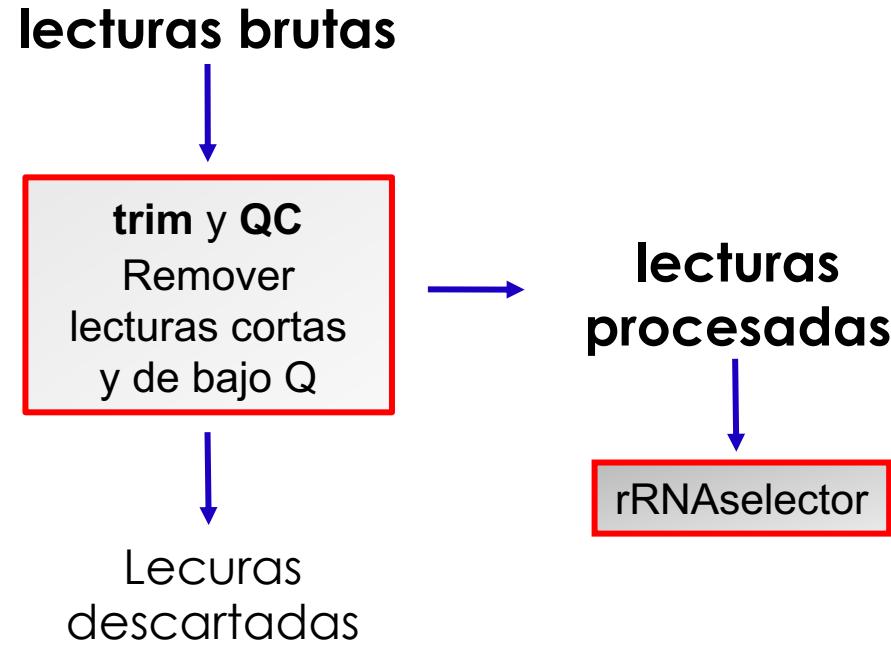




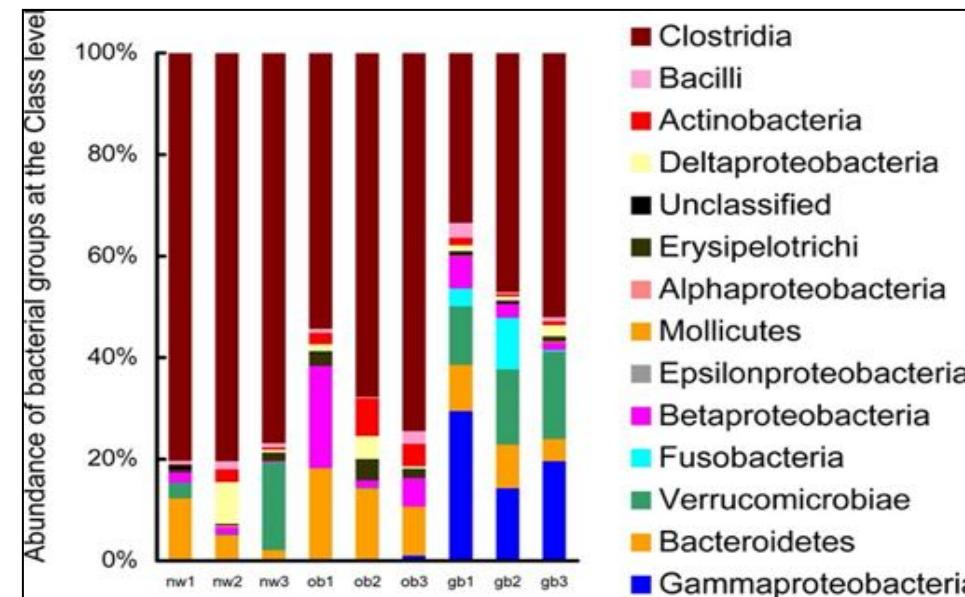
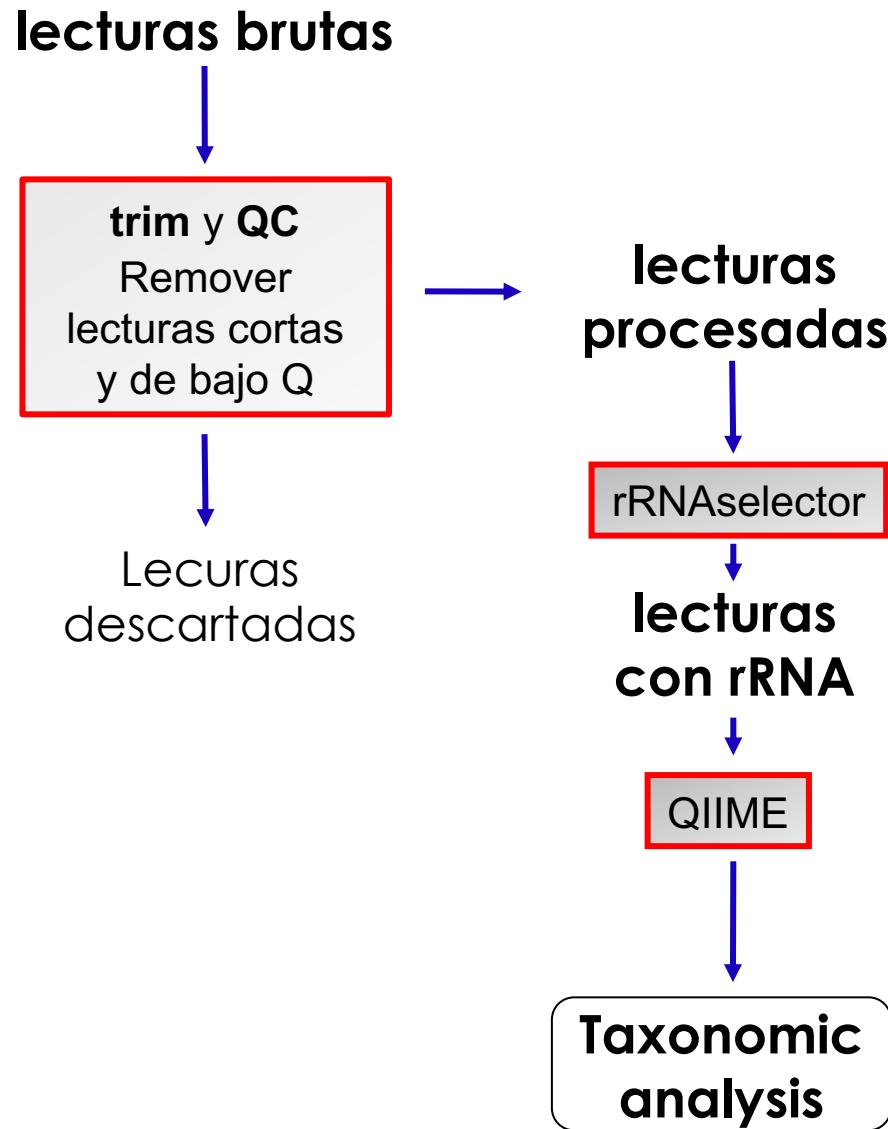
Shotgun sequencing



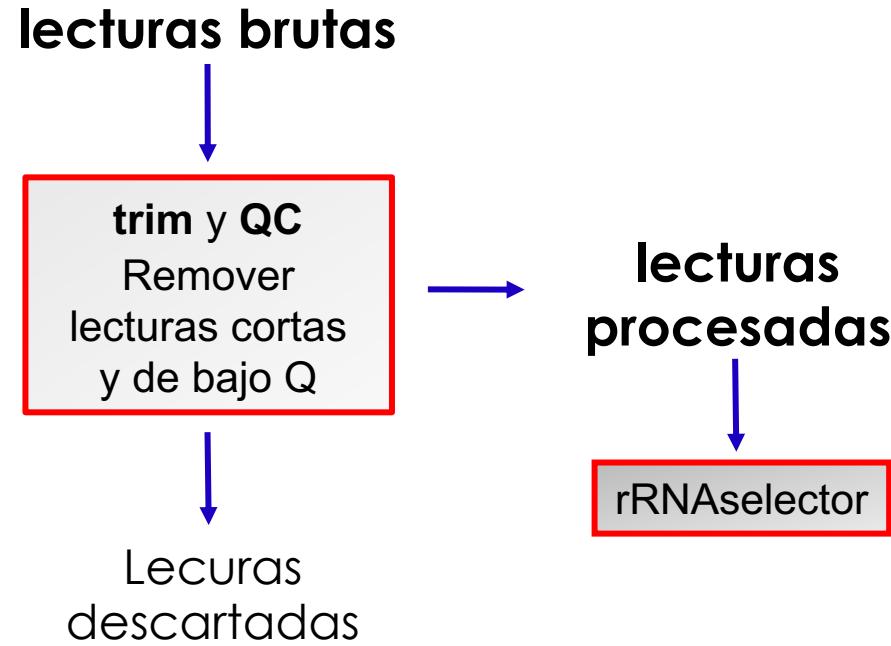
Secuenciación masiva (shotgun)



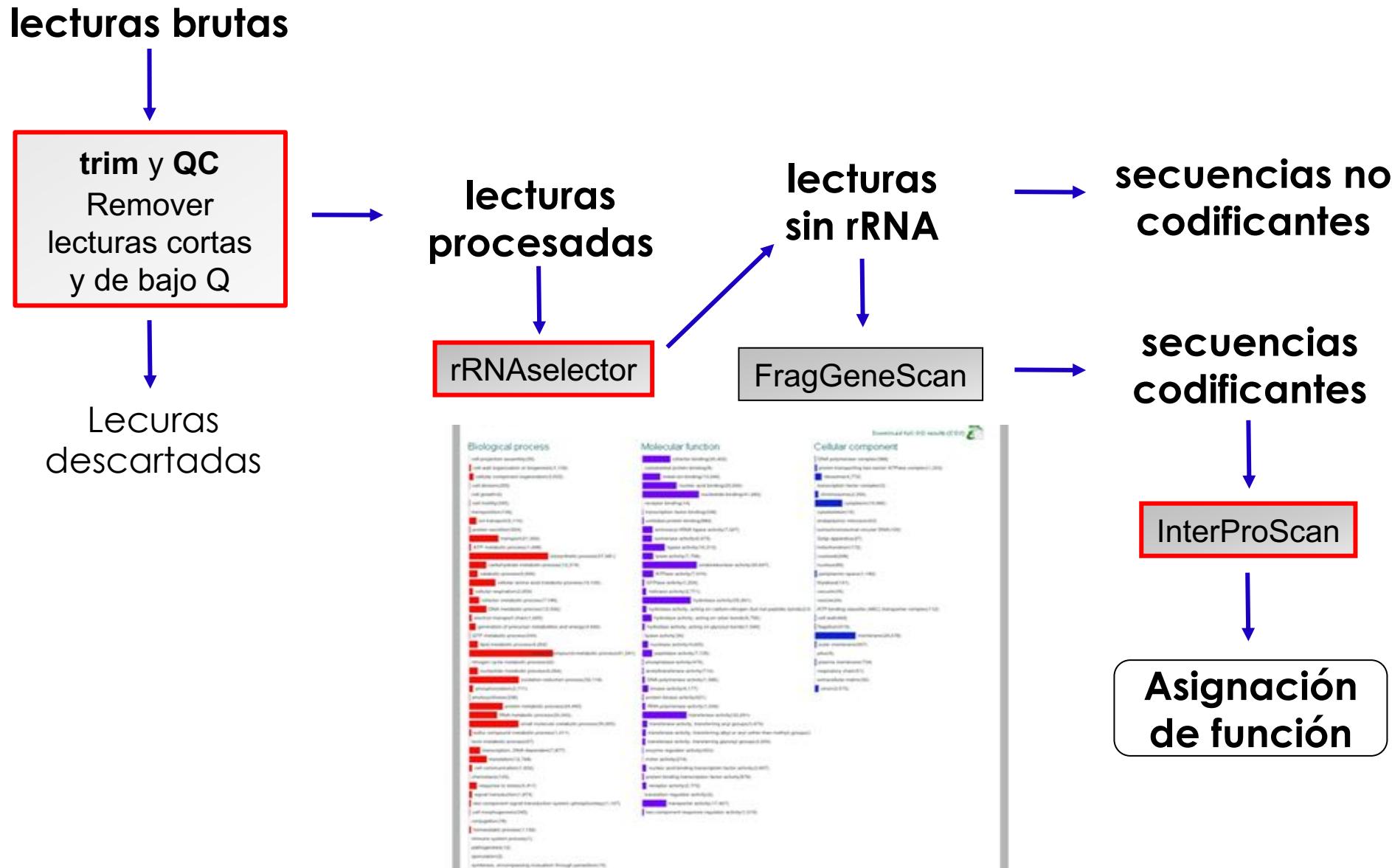
Secuenciación masiva (shotgun)



Secuenciación masiva (shotgun)



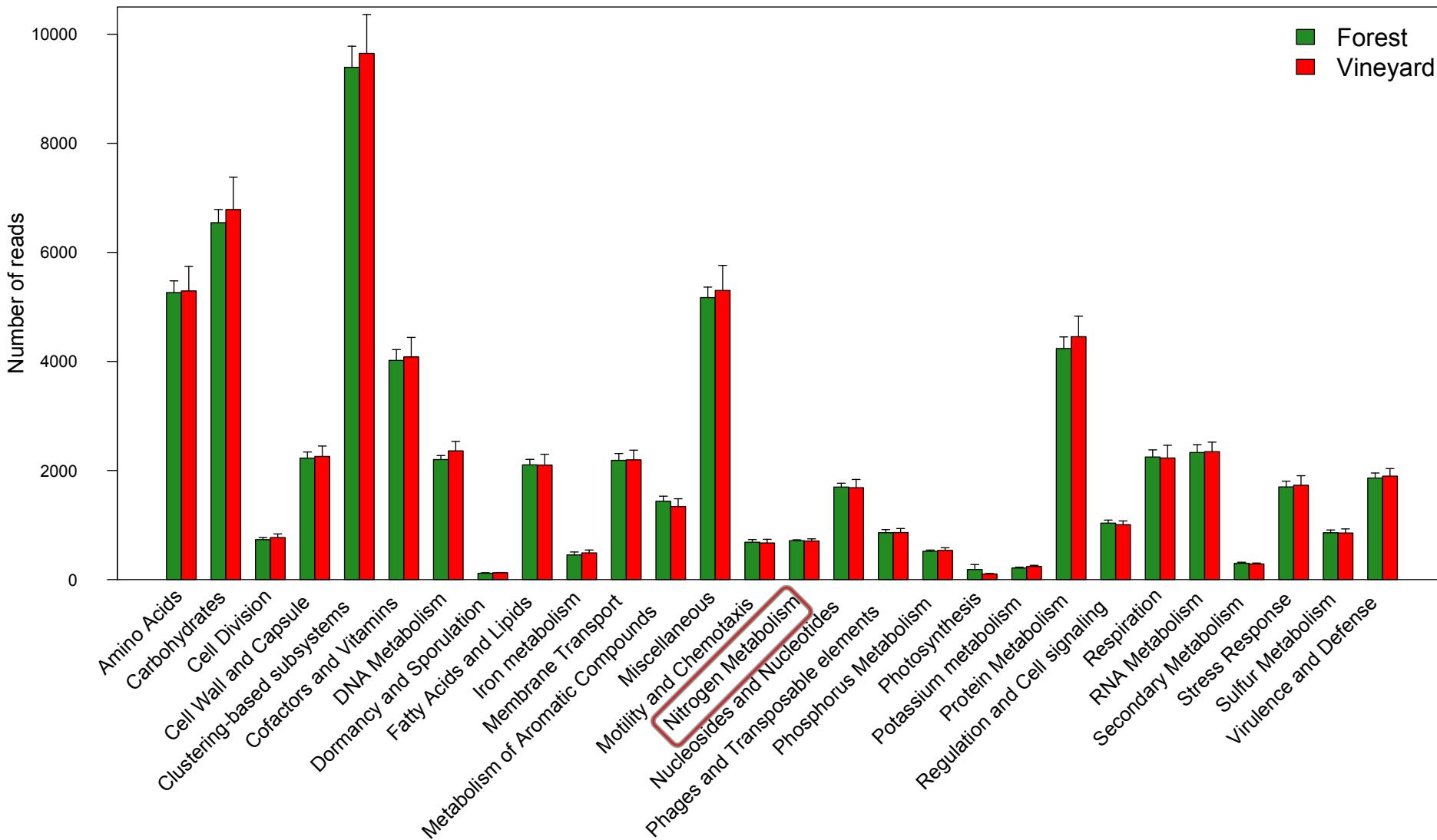
Secuenciación masiva (shotgun)



¿Cuáles son las funciones ecológicas desempeñadas por las comunidades microbiológicas que habitan estos ambientes tan particulares de la zona central de Chile?

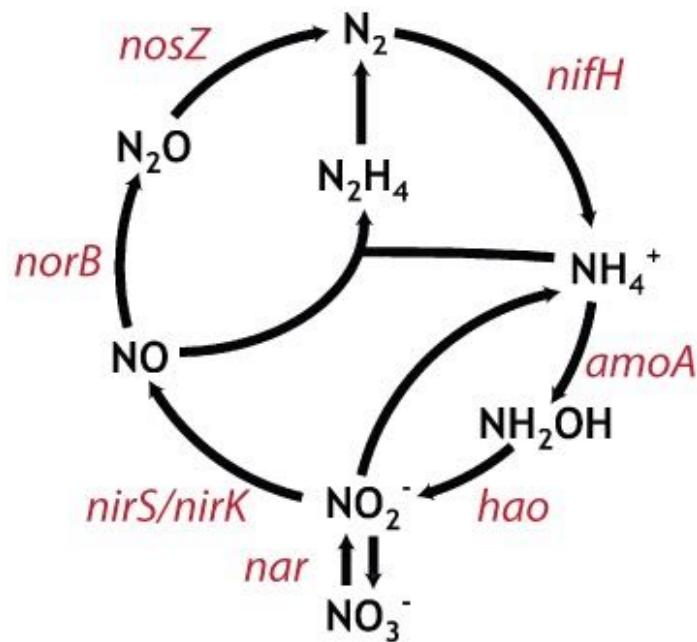
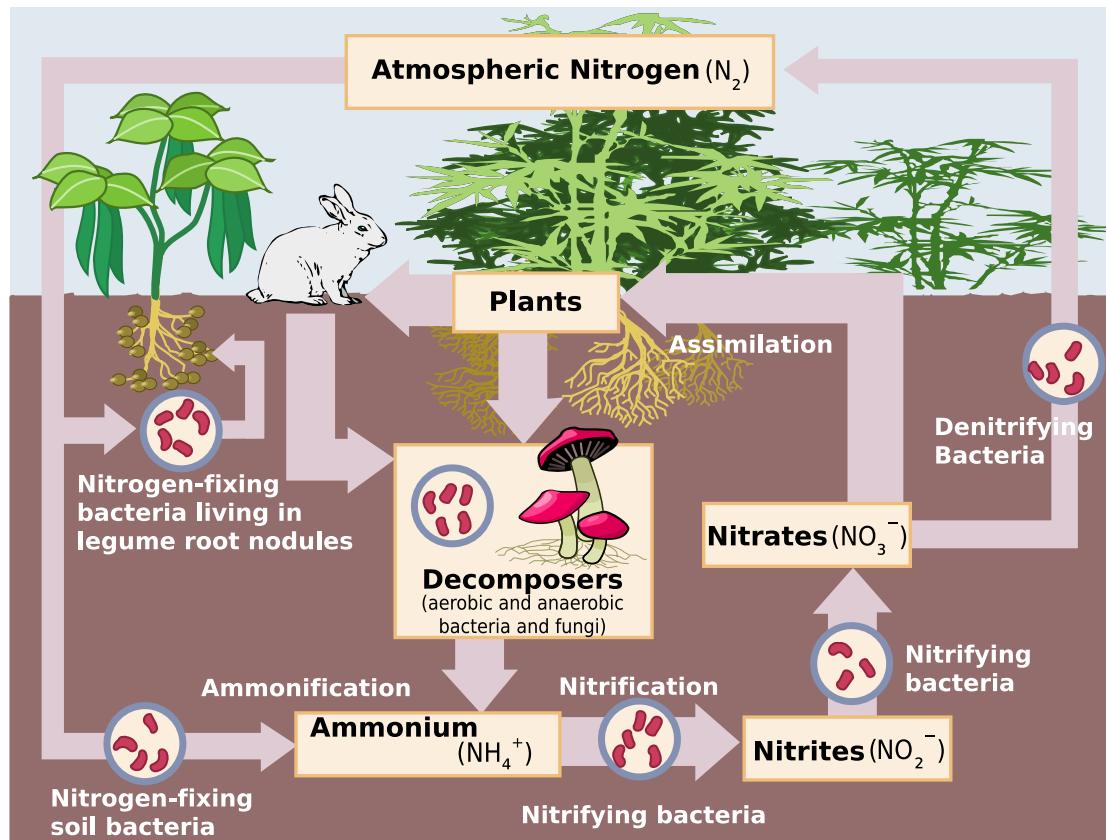


Categorías funcionales

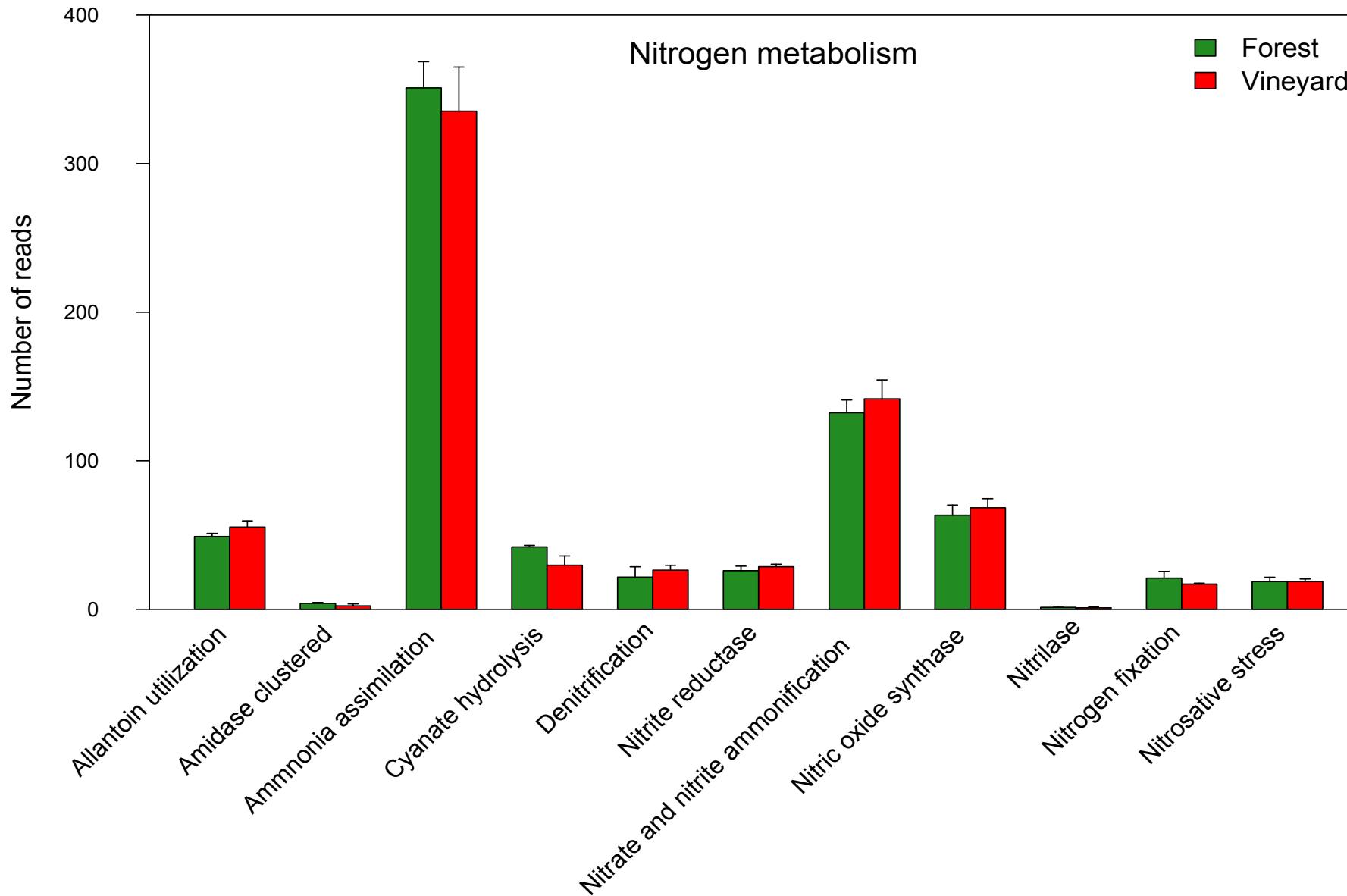


Ciclo de nitrógeno

La disponibilidad del nitrógeno limita el crecimiento de las plantas y la producción primaria, afectando la estructura y funciones ecosistemas.



Metabolismo de nitrógeno



Metagenomic analysis exploring taxonomic and functional diversity of soil microbial communities in Chilean vineyards and surrounding native forests

Luis E. Castañeda¹ and Olga Barbosa^{1,2}

¹ Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

² Instituto de Ecología y Biodiversidad (IEB-Chile), Santiago, Chile

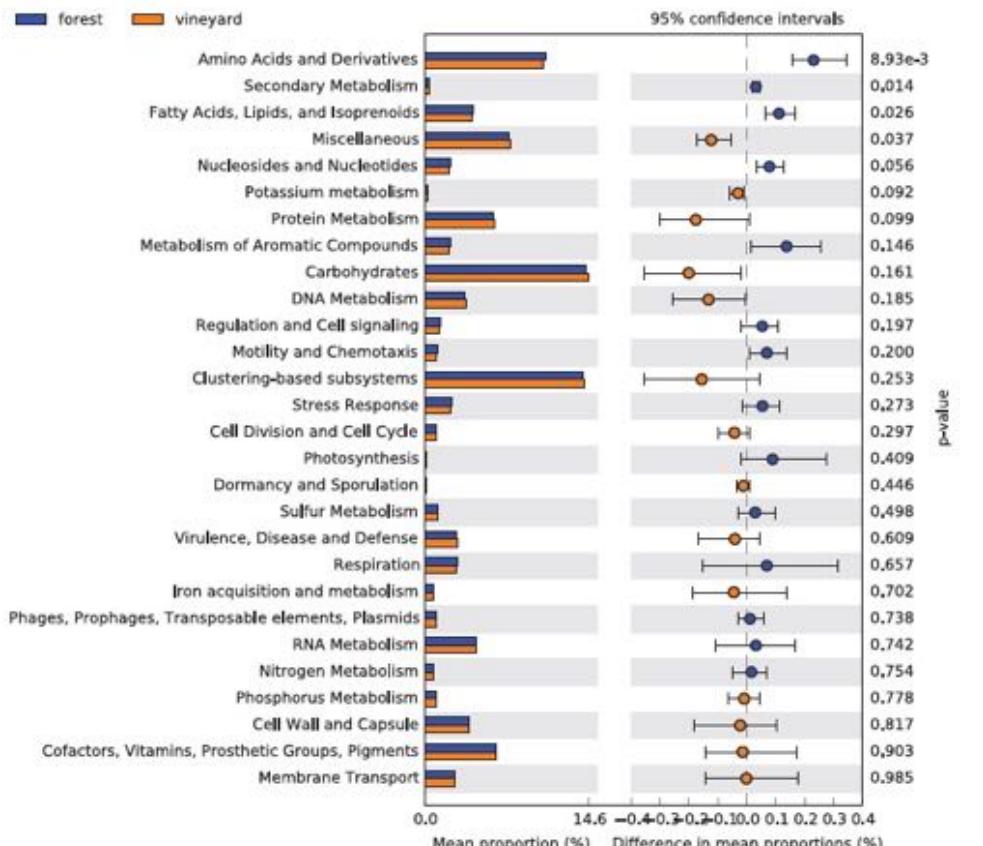
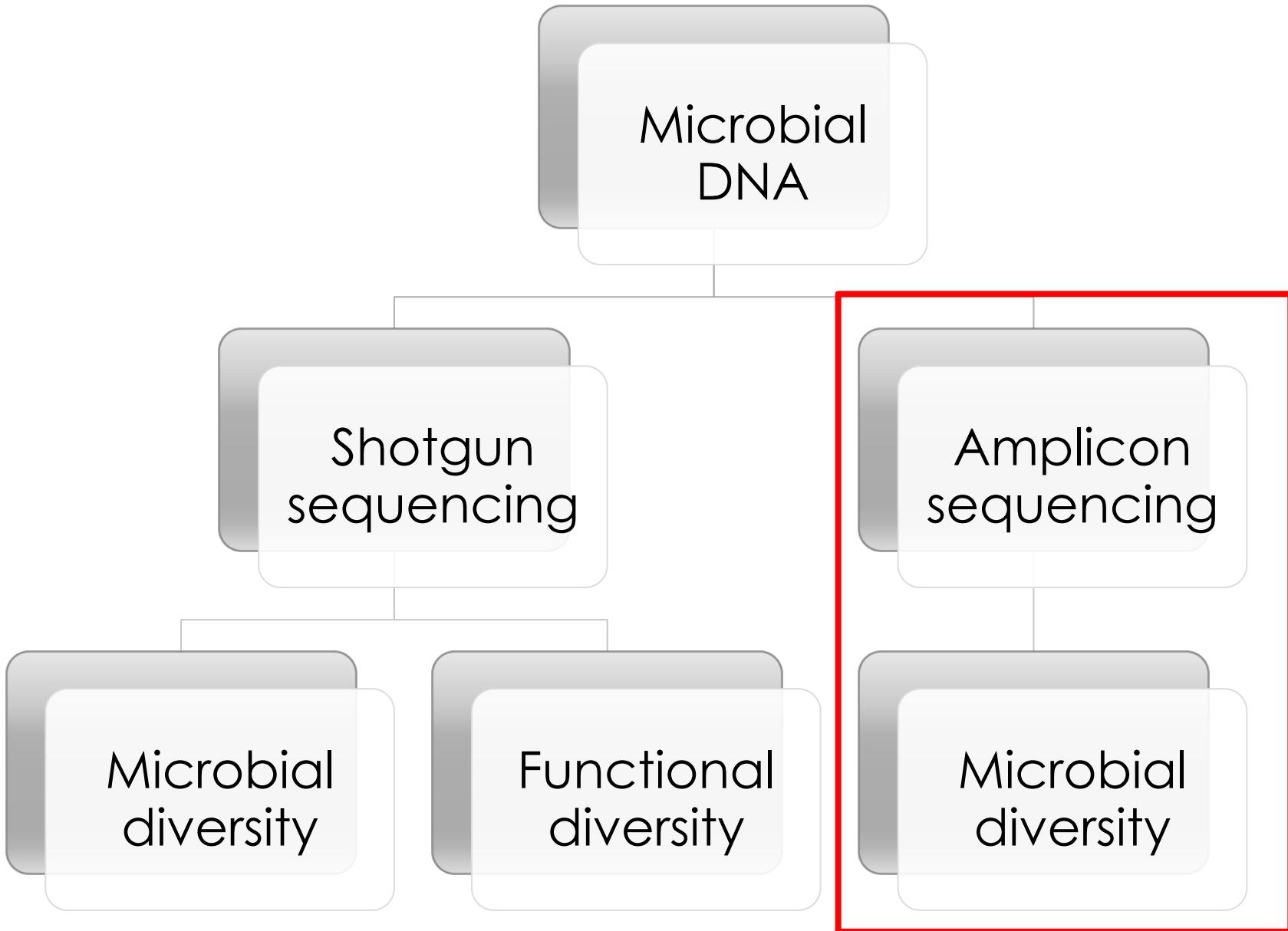


Table 2 Abundances of taxonomic groups in forest and vineyard soils.

Taxa	Forest	Vineyard	p-value
Archaea	0.454 ± 0.050	0.486 ± 0.077	0.83
Crenarchaeota	0.050 ± 0.007	0.048 ± 0.003	0.51
Euryarchaeota	0.376 ± 0.040	0.395 ± 0.054	0.83
Korarchaeota	0.004 ± 0.002	0.004 ± 0.002	0.83
Thaumarchaeota	0.021 ± 0.015	0.036 ± 0.019	0.28
Unclassified	0.003 ± 0.001	0.004 ± 0.001	0.66
Bacteria	90.34 ± 0.561	90.43 ± 0.073	0.72
Acidobacteria	5.118 ± 0.748	5.080 ± 0.841	0.83
Actinobacteria	20.49 ± 2.532	20.17 ± 1.732	0.83
Aquificae	0.098 ± 0.005	0.105 ± 0.009	0.38
Bacteroidetes	3.049 ± 0.502	3.369 ± 0.576	0.51
Chlamydiae	0.051 ± 0.014	0.050 ± 0.013	0.83
Chlorobi	0.315 ± 0.030	0.321 ± 0.034	0.83
Chloroflexi	1.918 ± 0.304	2.017 ± 0.250	0.83
Chrysogenetes	0.015 ± 0.005	0.016 ± 0.005	0.83
Cyanobacteria	1.774 ± 0.236	1.819 ± 0.190	0.51
Deferribacteres	0.048 ± 0.006	0.048 ± 0.005	0.83
Deinococcus-Thermus	0.574 ± 0.043	0.599 ± 0.049	0.51
Dictyoglomi	0.035 ± 0.010	0.046 ± 0.002	0.27
Elusimicrobia	0.012 ± 0.002	0.016 ± 0.004	0.27
Fibrobacteres	0.006 ± 0.001	0.010 ± 0.003	0.08
Firmicutes	2.945 ± 0.296	3.313 ± 0.295	0.13
Fusobacteria	0.028 ± 0.003	0.030 ± 0.011	0.51
Gemmmatimonadetes	1.465 ± 0.208	1.542 ± 0.180	0.28
Lentisphaerae	0.030 ± 0.010	0.037 ± 0.003	0.28
Nitrospirae	0.200 ± 0.018	0.193 ± 0.037	0.51
Planctomycetes	3.001 ± 0.062	3.425 ± 0.617	0.13
Proteobacteria	0.019 ± 0.003	0.021 ± 0.007	0.83
Spirochaetes	0.235 ± 0.019	0.228 ± 0.015	0.83
Synergistetes	0.054 ± 0.003	0.062 ± 0.010	0.13
Tenericutes	0.001 ± 0.001	0.002 ± 0.001	0.27
Thermotogae	0.104 ± 0.010	0.134 ± 0.025	0.13
Verrucomicrobia	2.224 ± 0.656	2.210 ± 0.249	0.83
Unclassified	0.419 ± 0.040	0.439 ± 0.042	0.28
Eukaryota	0.582 ± 0.162	0.434 ± 0.139	0.51
Ascomycota	0.462 ± 0.102	0.339 ± 0.131	0.28
Basidiomycota	0.032 ± 0.013	0.026 ± 0.003	0.83
Unclassified	0.088 ± 0.056	0.069 ± 0.014	0.51
Viruses	0.002 ± 0.008	0.001 ± 0.009	0.49
Unassigned/unclassified	8.624 ± 0.370	8.623 ± 0.115	0.51



Secuenciación de amplicones

Es una aproximación para el estudio la diversidad taxonómica en muestras ambientales. Se basa en la secuenciación de genes taxonómicamente informativos, los cuales se obtienen a partir de una PCR utilizando partidores específicos.

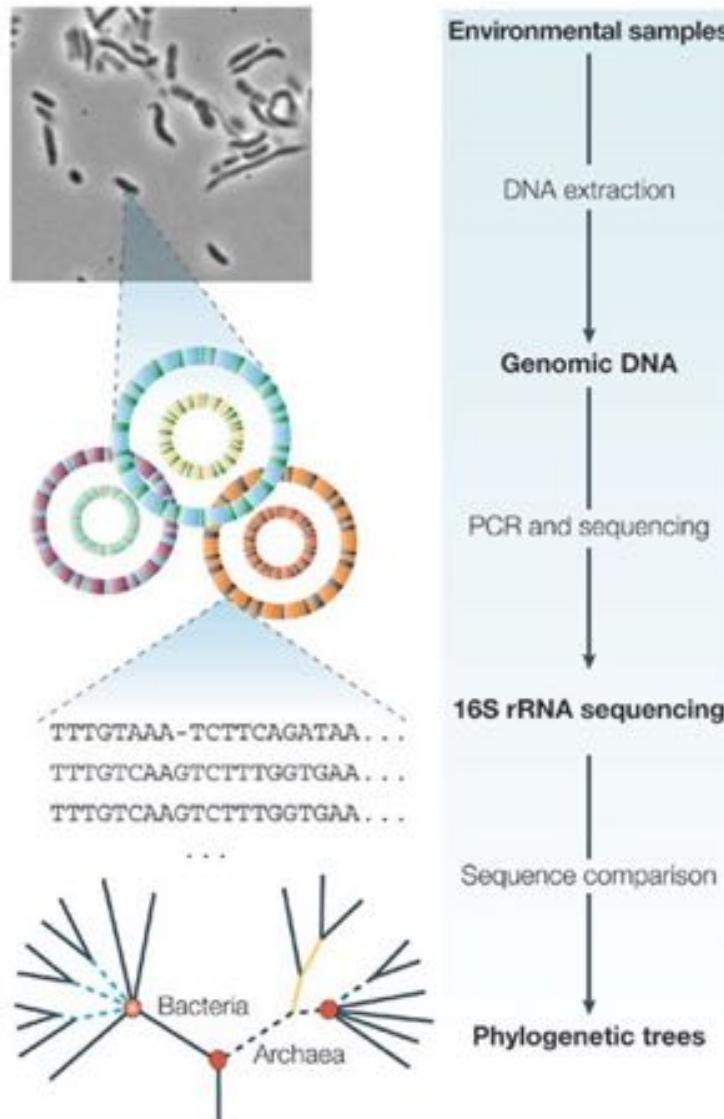
Bacteria: 16S rDNA.

Fungi: Internal transcribed spacer (ITS1-ITS2) y 28S.

Eukaryota: 18S rDNA.

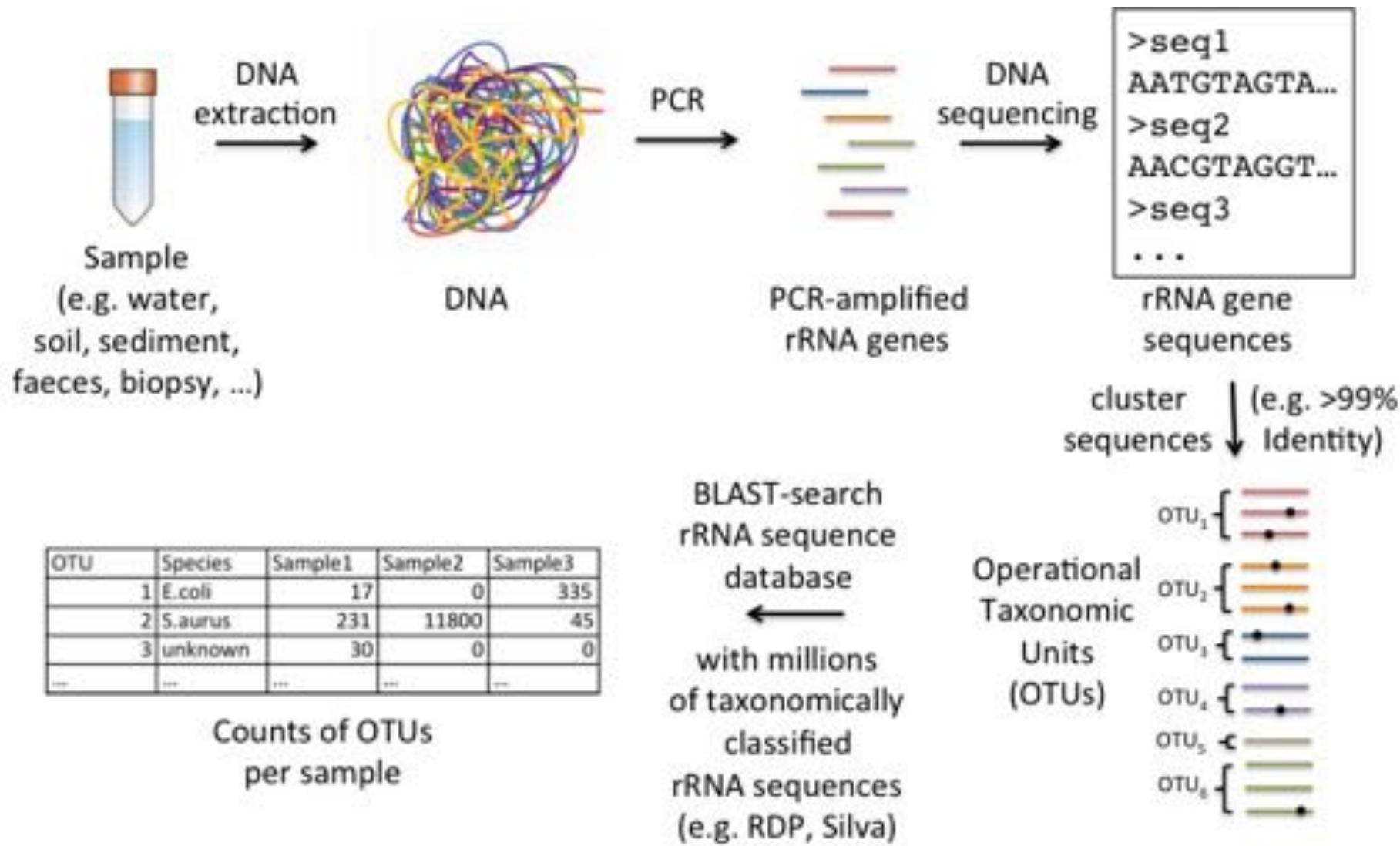
Estas secuencias poseen tasas evolutivas lentas por lo que son útiles para reconstruir filogenias en estos grupos.

Secuenciación de amplicones

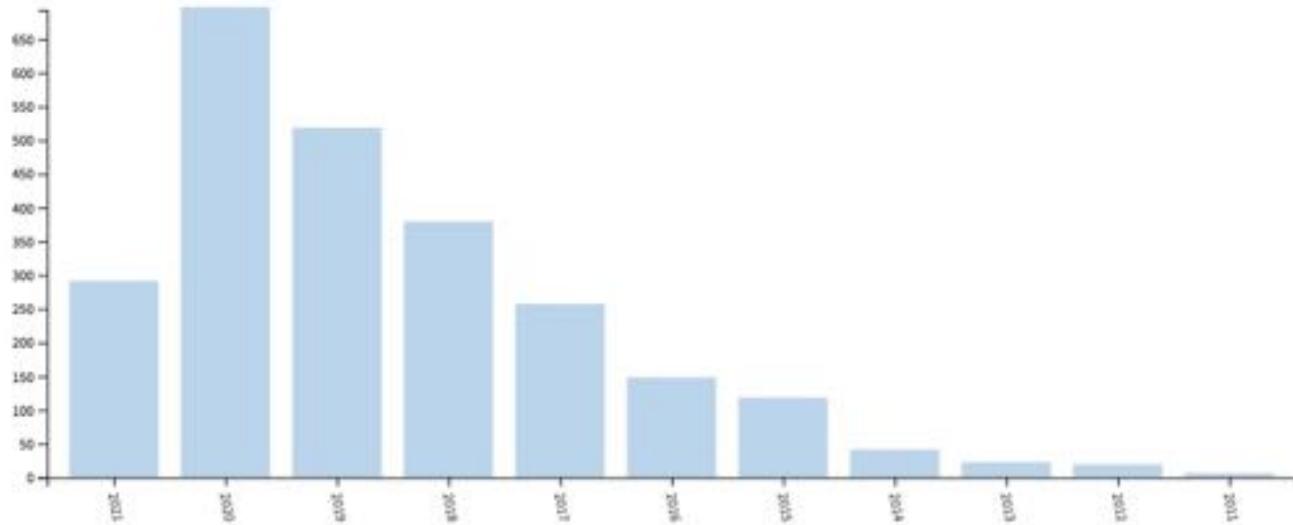


- Permite determinar la composición taxonómica de una muestra, aportando información de presencia y abundancia de una determinada especie.
- Asume que existe una sola copia del gen, lo que puede sobre estimar la abundancia de ciertos grupos que poseen múltiples copias.
- El uso de adaptadores y códigos de barras durante la PCR permite identificar muestras de distintos ambientes que hayan sido secuenciados en una misma placa.

Secuenciación de amplicones



Metabarcoding



Unidad taxonómica operativa (OTU):

Es una aproximación práctica al concepto de especie basada en las secuencias genómicas. En bacterias, las secuencias son agrupadas con un 97% de similitud.

Bases de datos de secuencias

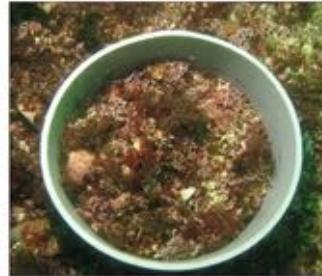
Son bases de datos que contienen secuencias exhaustivamente revisadas, las cuales también contienen asignaciones taxonómicas.

Tabla OTU:

Tabla que contiene la identidad de cada OTU y sus asignaciones taxonómicas (filas) y sus abundancias en cada muestra (columnas).

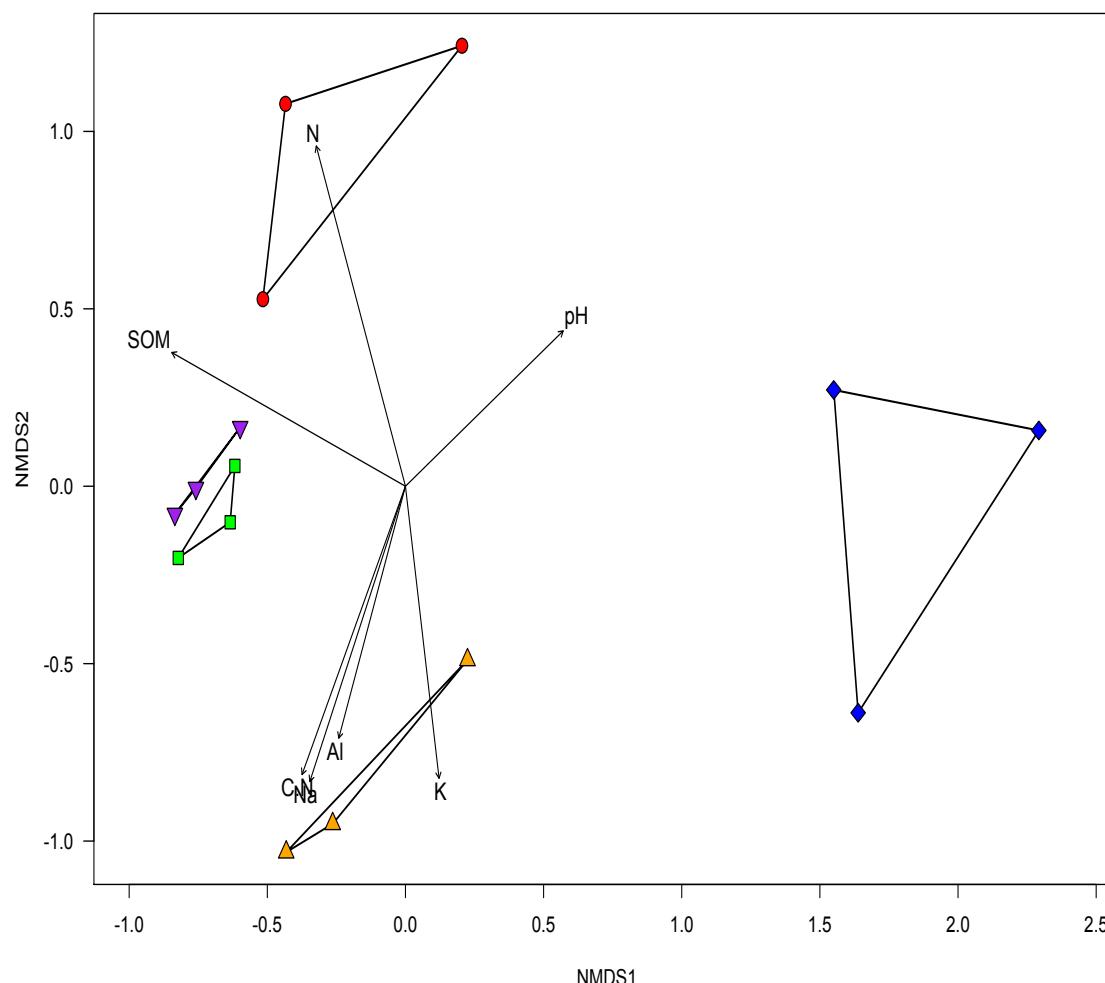
Muestreo y extracción de ADN

- Preservar: congelar, etanol.



Muestreo y extracción de ADN

- Recolectar la mayor cantidad de información ambiental: pH, temperatura, CO₂, contenido nitrógeno, etc., etc.

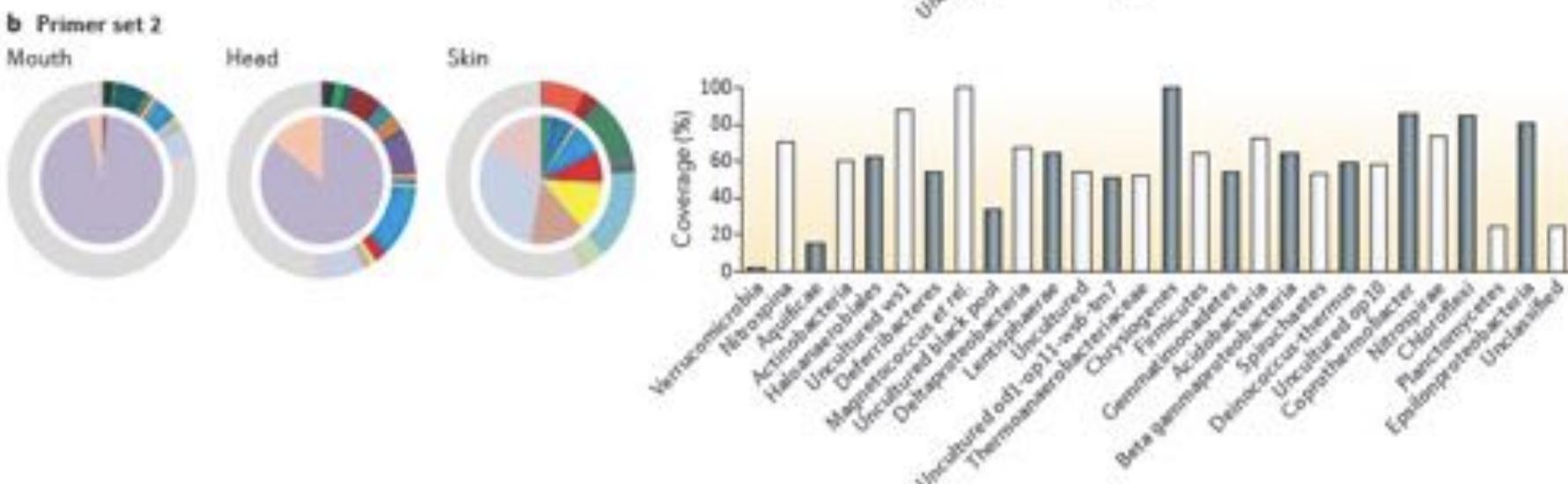
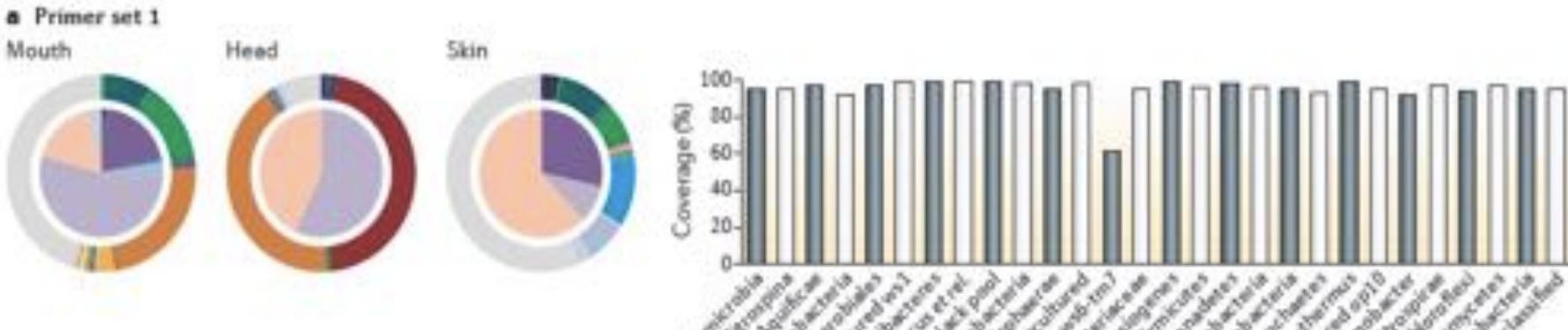


Amplificación de ADN

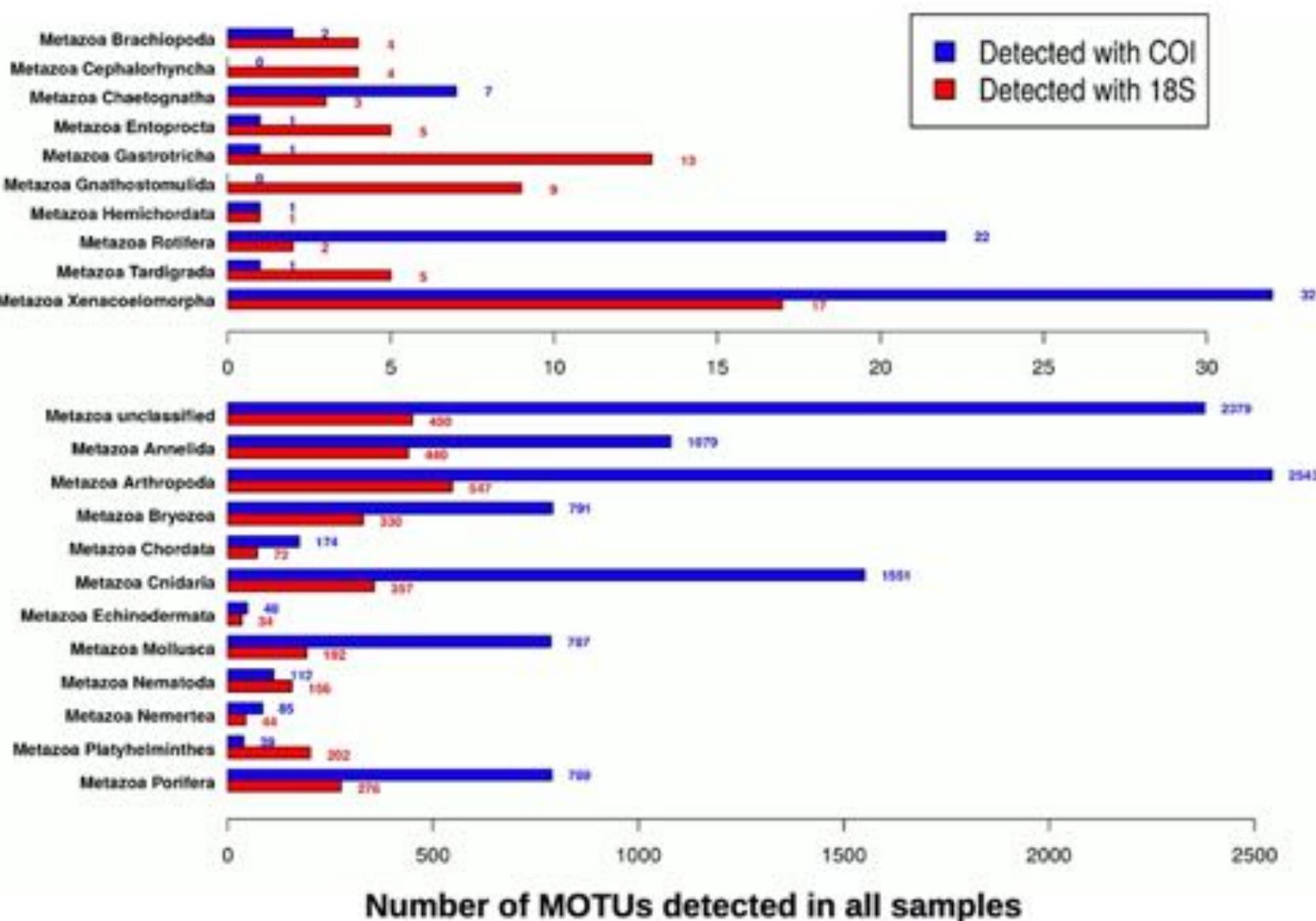
Target	Gene/ region	Reference	Data bases
Bacteria	16S	Sogin <i>et al.</i> (2006)	RDP, Greengenes, SILVA
Archaea	16S	Sogin <i>et al.</i> (2006)	RDP, Greengenes, SILVA
Fungi	ITS	Epp <i>et al.</i> (2012), Schoch <i>et al.</i> (2012)	UNITE, GenBank, BOLD (few records)
	18S	Not recommended (Schoch <i>et al.</i> 2012)	SILVA
Protists	18S	Pawlowski <i>et al.</i> (2012)	SILVA
	ITS	Pawlowski <i>et al.</i> (2012)	GenBank
	CO1	Pawlowski <i>et al.</i> (2012)	BOLD
Meiofauna	CO1	Hebert <i>et al.</i> (2003)	BOLD
	18S	Deagle <i>et al.</i> (2014)	GenBank
Macrofauna	CO1	Hebert <i>et al.</i> (2003)	BOLD
	16S	Epp <i>et al.</i> (2012), Deagle <i>et al.</i> (2014)	GenBank
	12S	Epp <i>et al.</i> (2012), Deagle <i>et al.</i> (2014)	GenBank
	18S	Deagle <i>et al.</i> (2014)	GenBank
Plants	matK	Hollingsworth <i>et al.</i> (2009)	GenBank, BOLD (few records)
	+ rbcL		
	ITS	China Plant Barcode of Life Group (2011)	GenBank

Amplificación de ADN

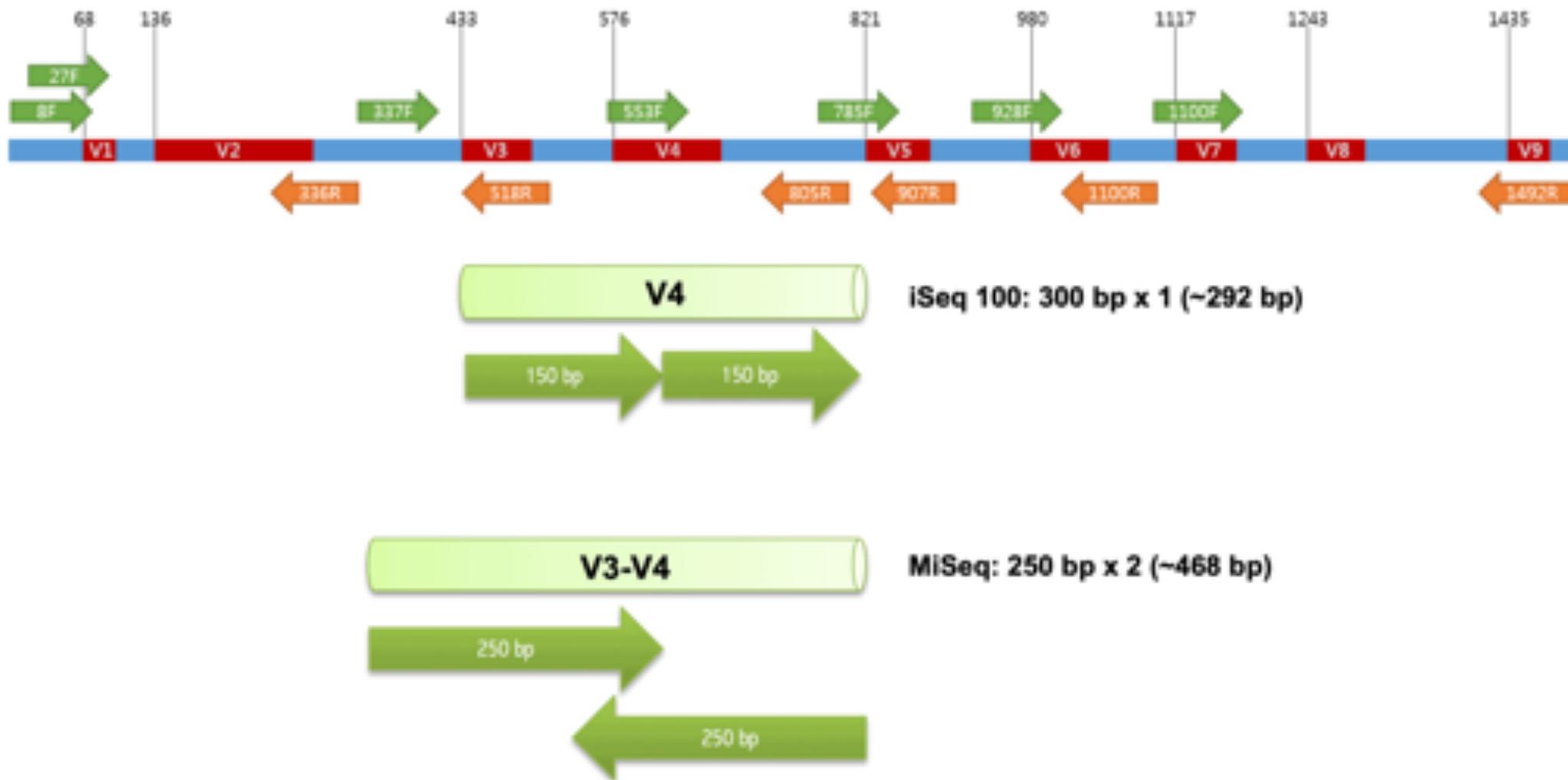
- Para 16S hay muchos partidores... así que, cuál debemos elegir?



Amplificación de ADN

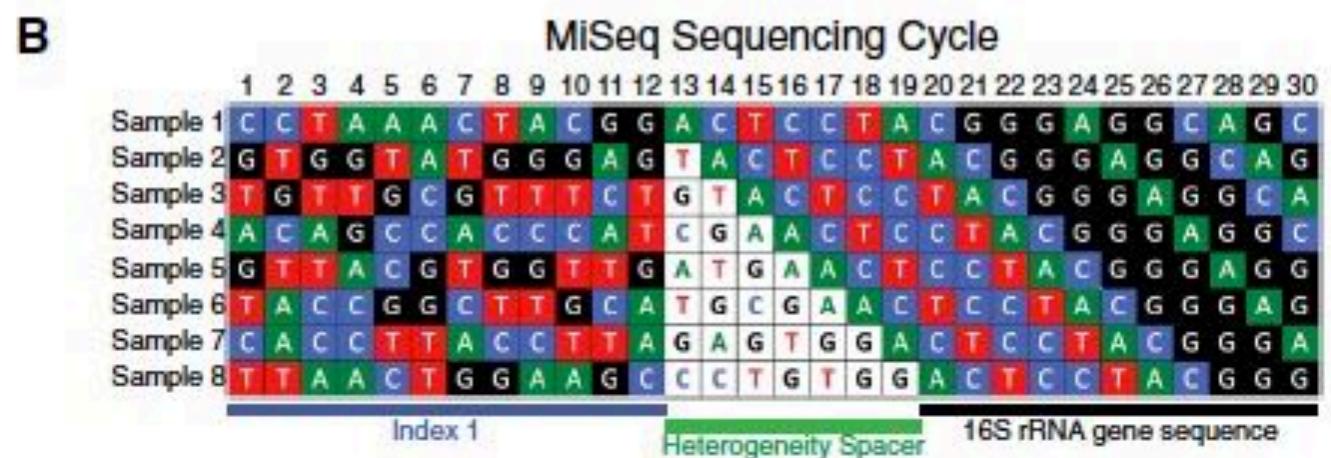
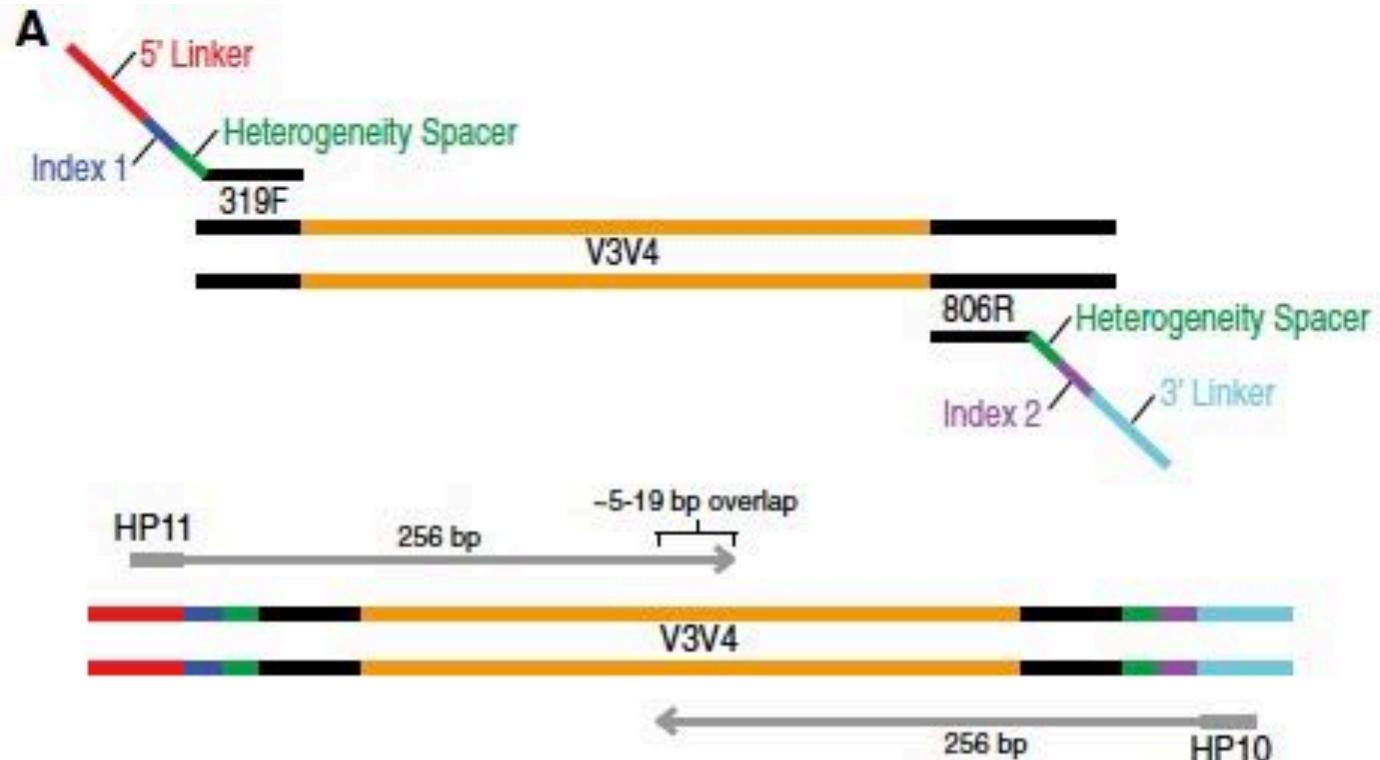


Amplificación de ADN



Secuenciación

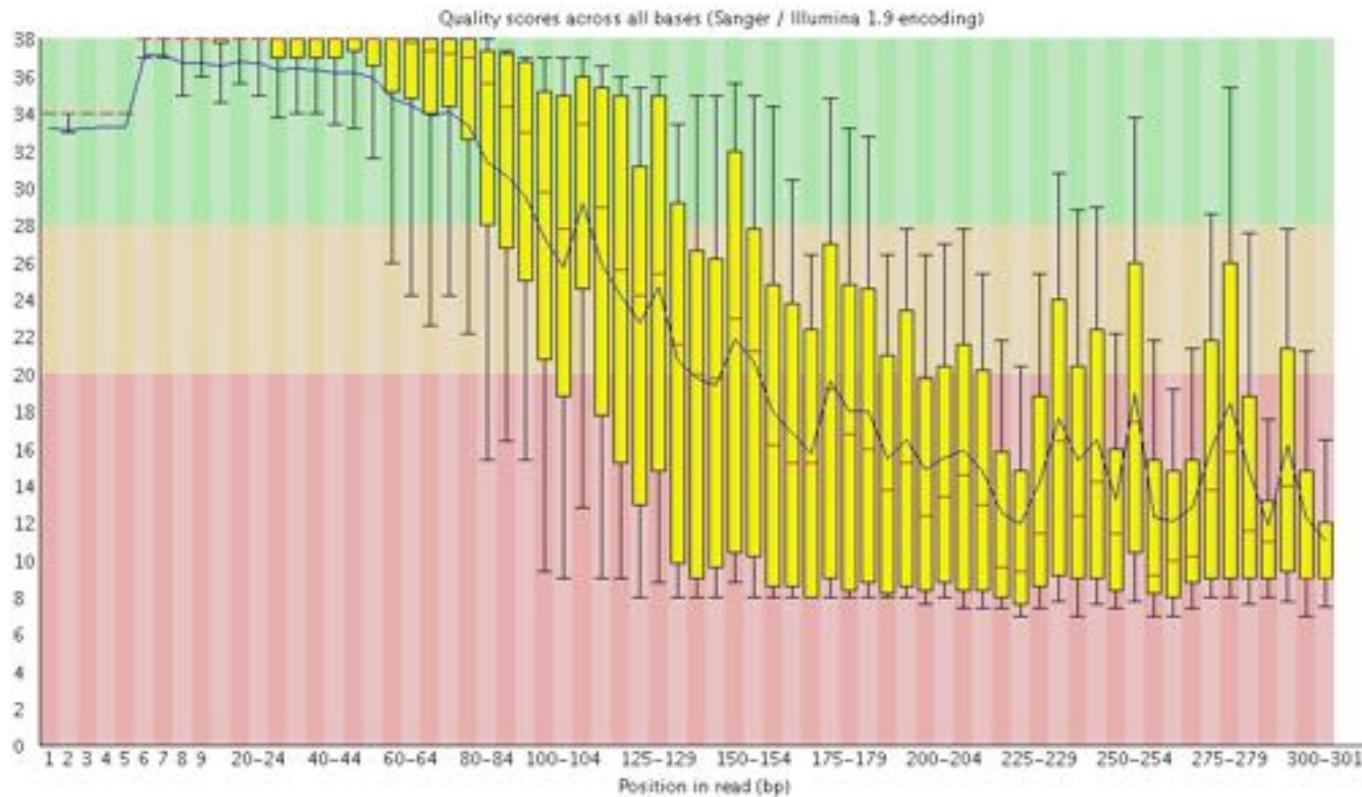
- Illumina: MiSeq
HiSeq.
- Lecturas de secuencias paredas o simples.



Procesamiento de secuencias

- Chequeo de calidad.

FastQC

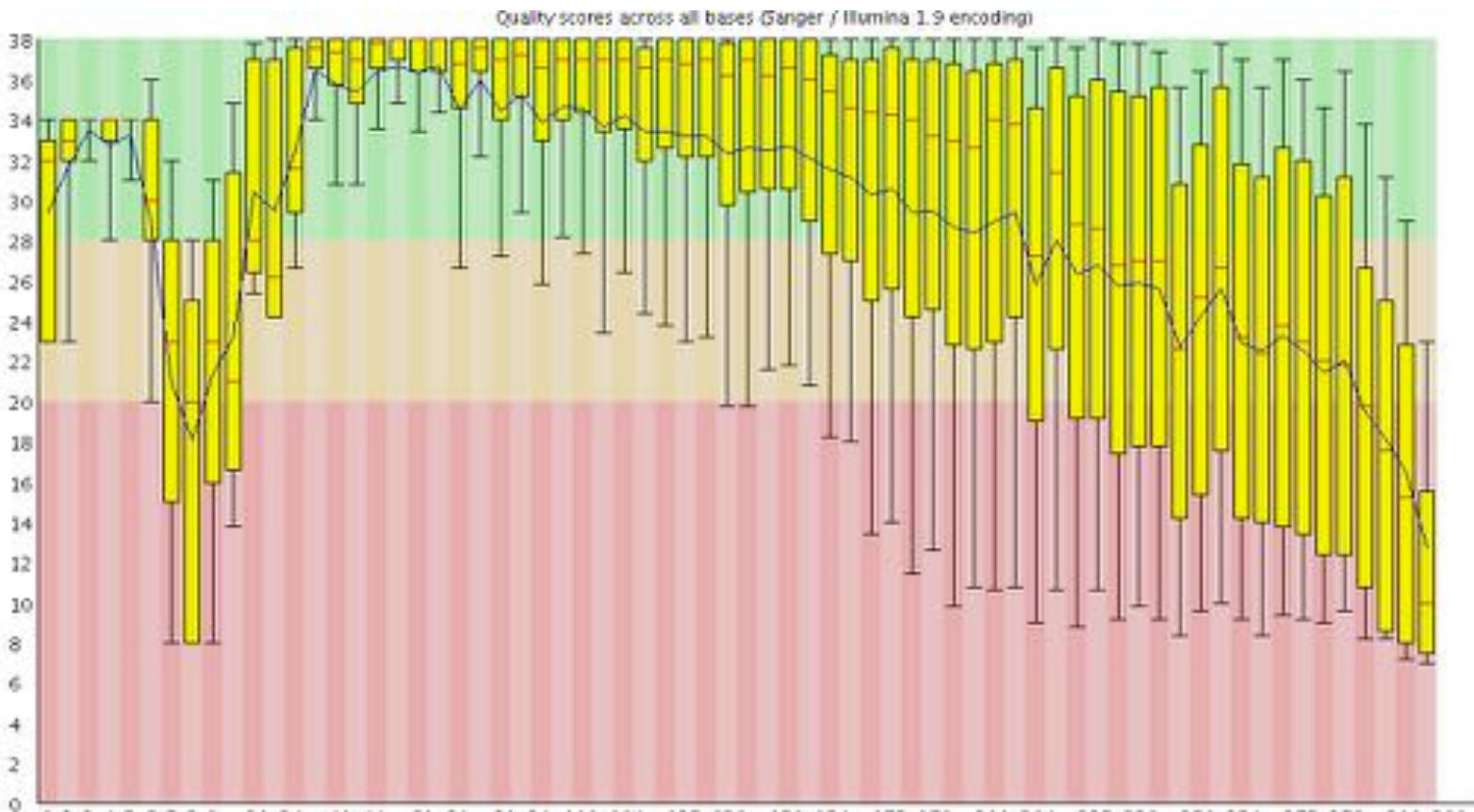


Quality Score	Error Probability
Q40	0.0001 (1 in 10,000)
Q30	0.001 (1 in 1,000)
Q20	0.01 (1 in 100)
Q10	0.1 (1 in 10)

Procesamiento de secuencias

- Chequeo de calidad.

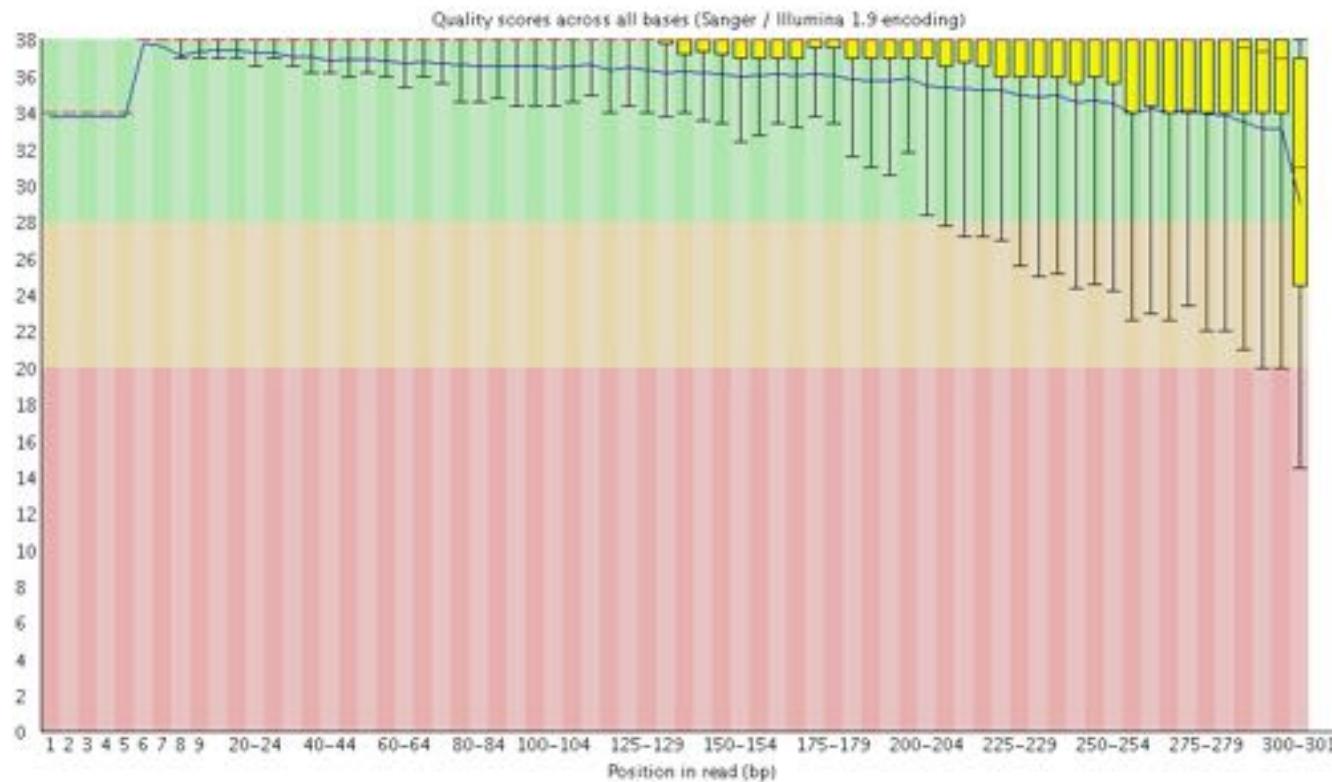
FastQC



Procesamiento de secuencias

- Chequeo de calidad.

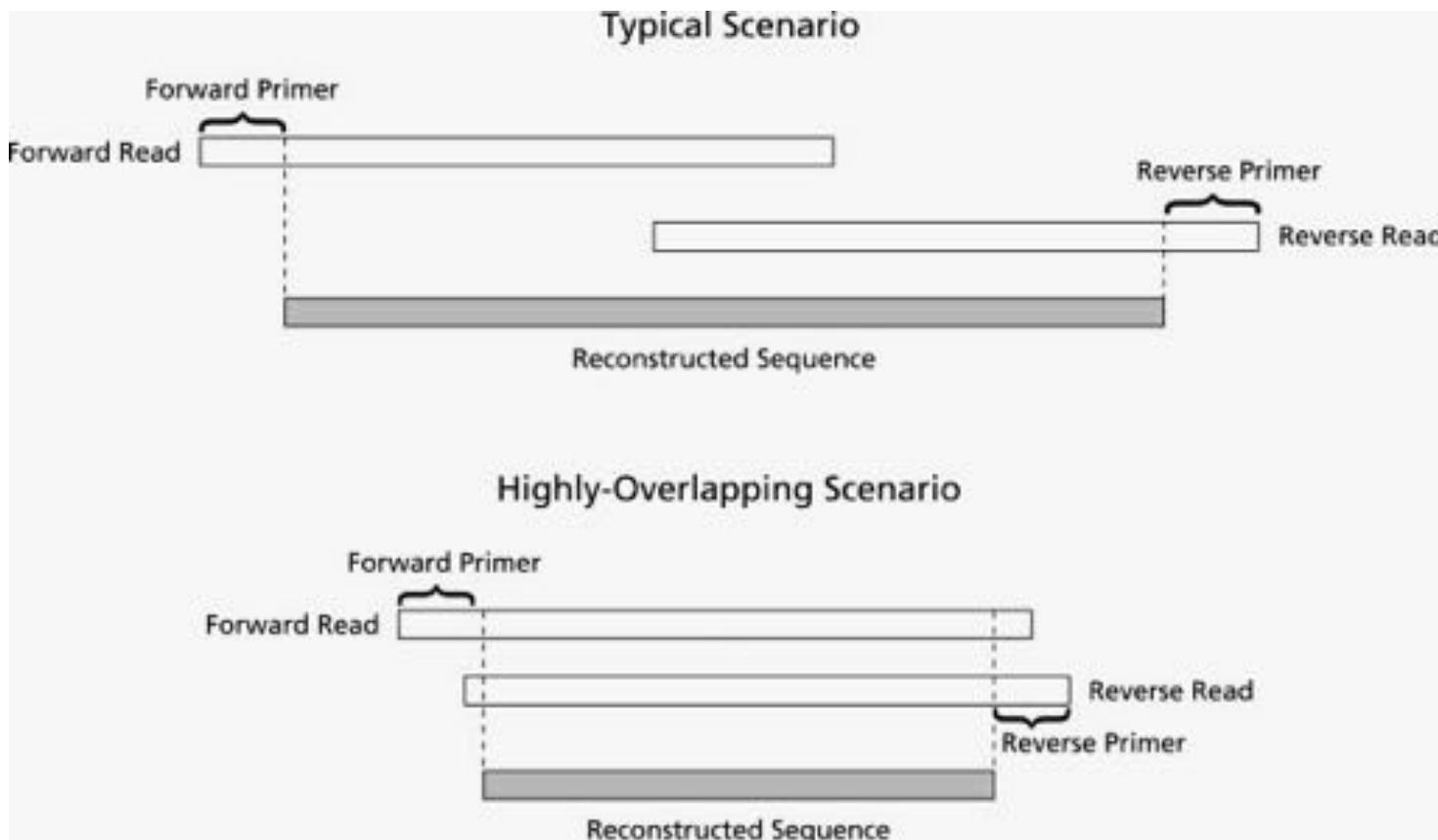
FastQC



Quality Score	Error Probability
Q40	0.0001 (1 in 10,000)
Q30	0.001 (1 in 1,000)
Q20	0.01 (1 in 100)
Q10	0.1 (1 in 10)

Procesamiento de secuencias

- Ensamblar secuencias pareadas.



Análisis de comunidades

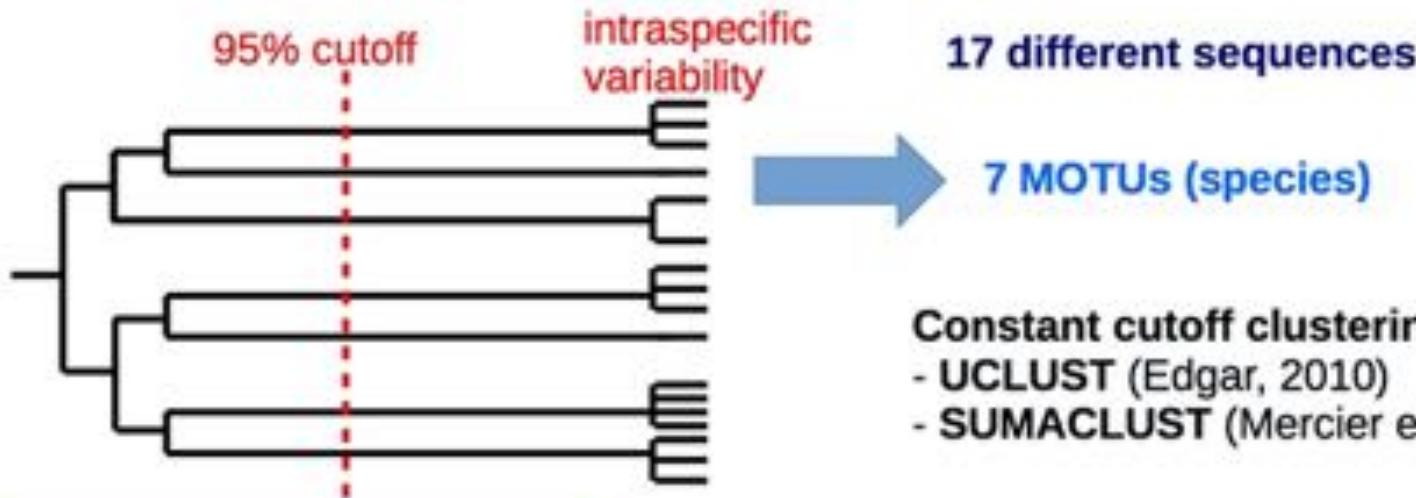
- mothur.
- VAMPS.
- QIIME.

Quantitative Insights In Molecular Ecology.

- Funciona en Linux, o a través de Virtual Machine.
- Puede usar diferentes base de datos como Greengenes, Ribosomal Data Project (RDP), SILVA, entre otras.
- Puede hacer inferencias filogenéticas.
- Da un acceso rápido a diversos análisis estadísticos.

Agrupamiento de secuencias

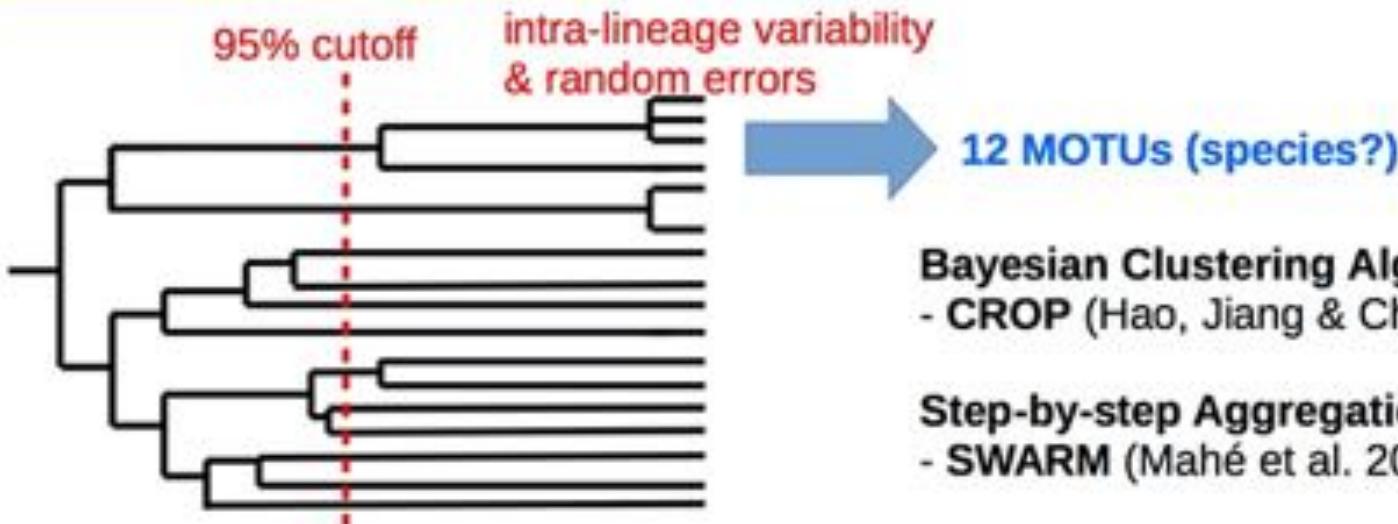
Theory: Constant ID % cutoff



Constant cutoff clustering software:

- UCLUST (Edgar, 2010)
- SUMACLUST (Mercier et al., 2013)

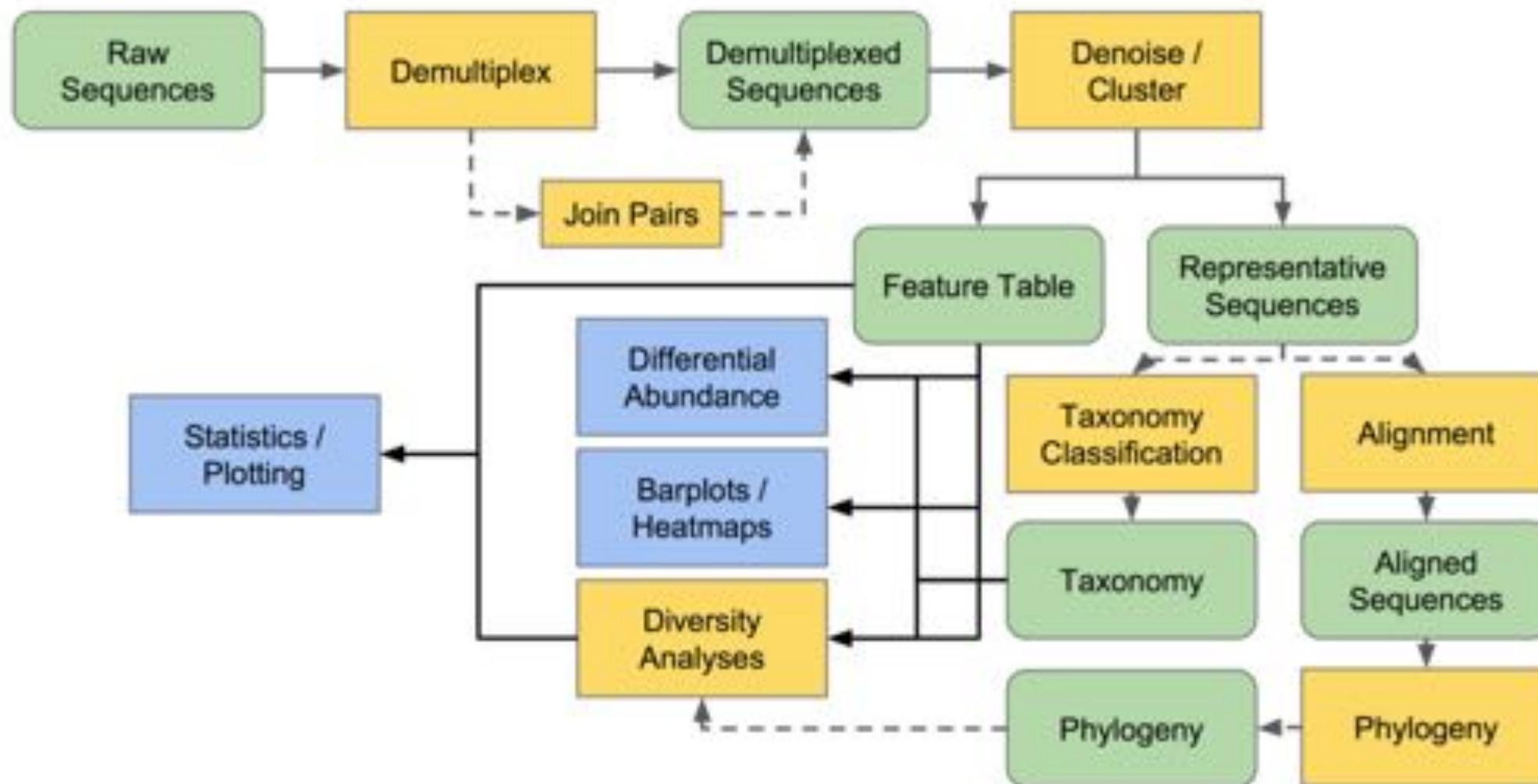
Reality: Variable cutoff



Bayesian Clustering Algorithms:
- CROP (Hao, Jiang & Chen, 2011)

Step-by-step Aggregation Algorithms: - SWARM (Mahé et al. 2014)

QIIME workflow



Análisis comunitario

QIIME.

Realiza análisis estadísticos y genera figuras de alta calidad, pero es poco flexible a demandas específicas solicitadas por usuario.

Por ejemplo, no analiza modelos lineares mixtos.

R.

100% flexible a las demandas específicas de cada usuario.

Esta basado en un comunidad científica cooperativa, la cual constantemente genera nuevos paquetes para análisis específicos como vegan, ape, phyloseq.

El único cons... necesitan aprender el lenguaje de R!!!

Ecología de comunidades

Comunidad: es un grupo de organismos que habitan en un lugar y tiempo dado, a veces interactúando entre ellas.

Riqueza: es el número de entidades (especies, OTUs) en una comunidad. Muchas veces la riqueza varía en función del área muestreada, siendo sensible al proceso de muestreo.

Diversidad: es el número de entidades (especies, OTUs) que pueden ser identificadas y las proporciones de objetos (individuos) en cada entidad.

Diversidad α : diversidad de entidades en un hábitat o comunidad.

Diversidad β : tasa de recambio de entidades entre comunidades.

Diversidad γ : tasa de recambio de entidades en un rango de comunidades a escala geográfica.

Ecología de comunidades

Indice de Shannon-Wiener: representa la incertidumbre de predecir la entidad de un individuo tomado al azar dentro de la comunidad.

Indice de Simpson: representa la probabilidad de que dos individuos tomados a la azar dentro de la comunidad pertenezcan a diferentes entidad.

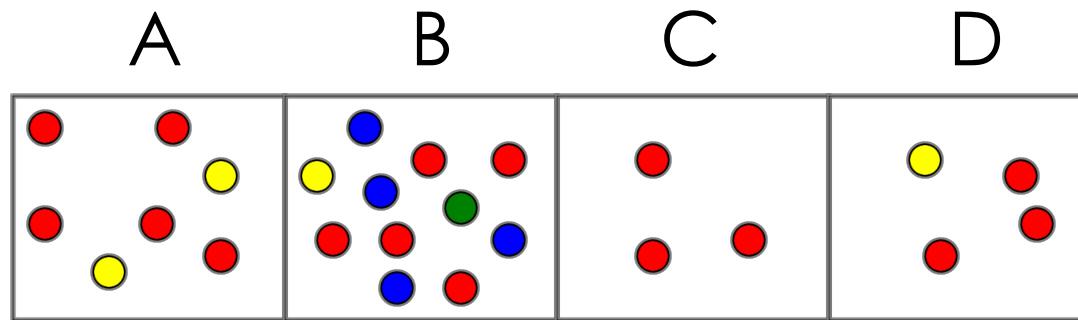
$$H = - \sum_{i=1}^S p_i \log_b p_i \quad \text{Shannon-Weaver} \quad (1)$$

$$D_1 = 1 - \sum_{i=1}^S p_i^2 \quad \text{Simpson} \quad (2)$$

Equitatividad: representa cuan equitativamente se reparten los individuos entre las entidades.

$$J = H / \log(S) = H / H_{\max}$$

Análisis de diversidad



Índices de diversidad:

$$H = - \sum_{i=1}^S p_i \log_b p_i \quad \text{Shannon-Weaver} \quad (1)$$

$$D_1 = 1 - \sum_{i=1}^S p_i^2 \quad \text{Simpson} \quad (2)$$

Riqueza (S):

$$2 - 4 - 1 - 2$$

Uniformidad

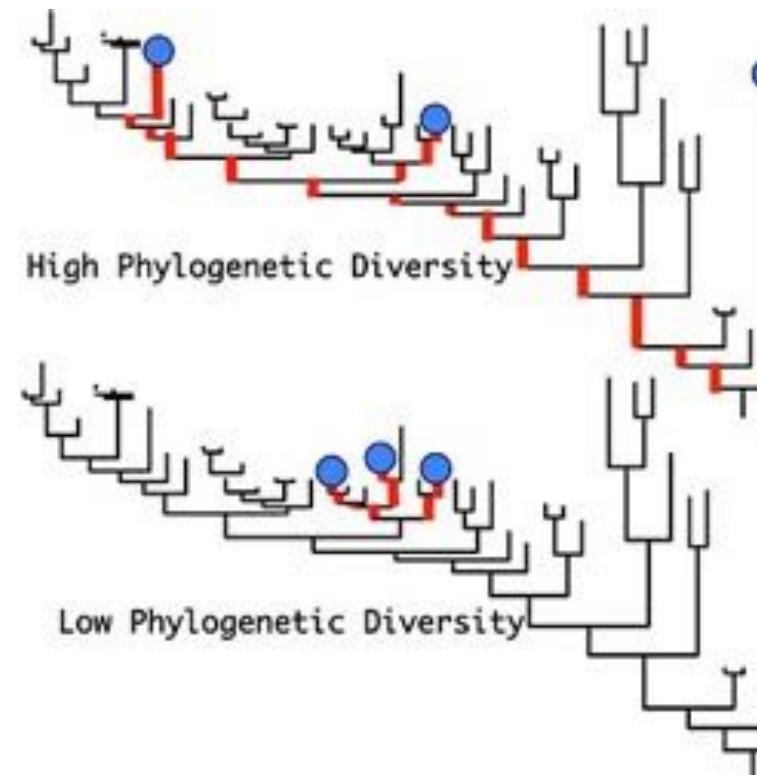
:

$$J = H / \log(S)$$

Análisis de diversidad

Diversidad filogenética.

Este índice calcula la diversidad presente en una comunidad tomando en cuenta las distancias filogenéticas existentes entre las distintas especies que se encuentran en una comunidad.



Ecología de comunidades

Diversidad filogenética: estima la diversidad de una comunidad basado en la relación filogenética entre las especies que forman parte de ella.

$$Faith = \sum_i^n I_i$$

Beta diversidad filogenética: estas mediciones son relativamente nuevas en ecología.

Unifrac: esta medida cuantifica la fracción de la filogenia contenida en cada una de las comunidades.

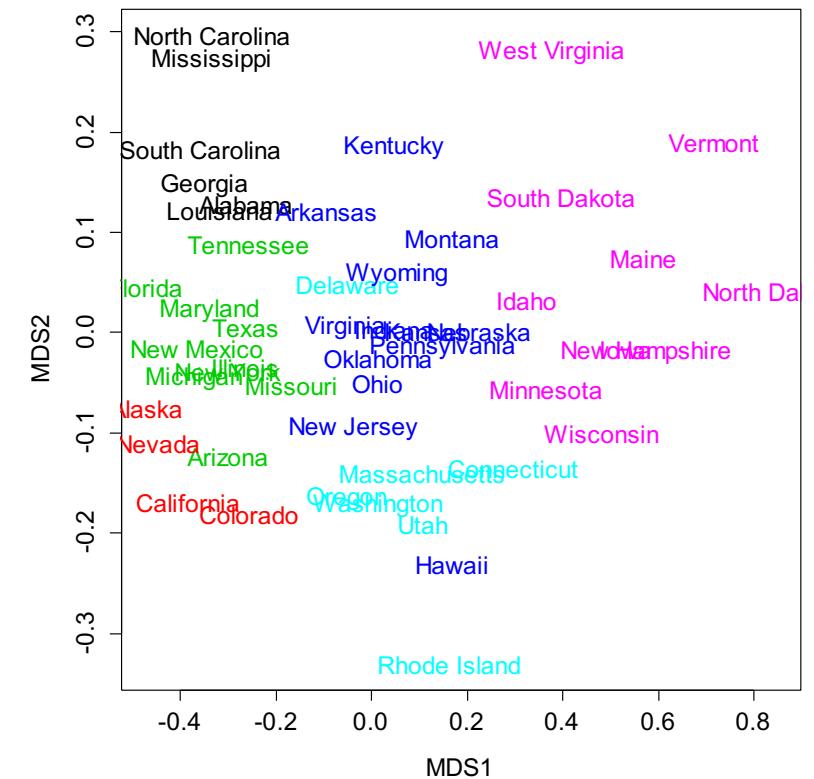
$$UniFrac_{A,B} = \frac{PD_{A \cup B} - PD_{A \cap B}}{PD_{A \cup B}}$$

Ecología de comunidades

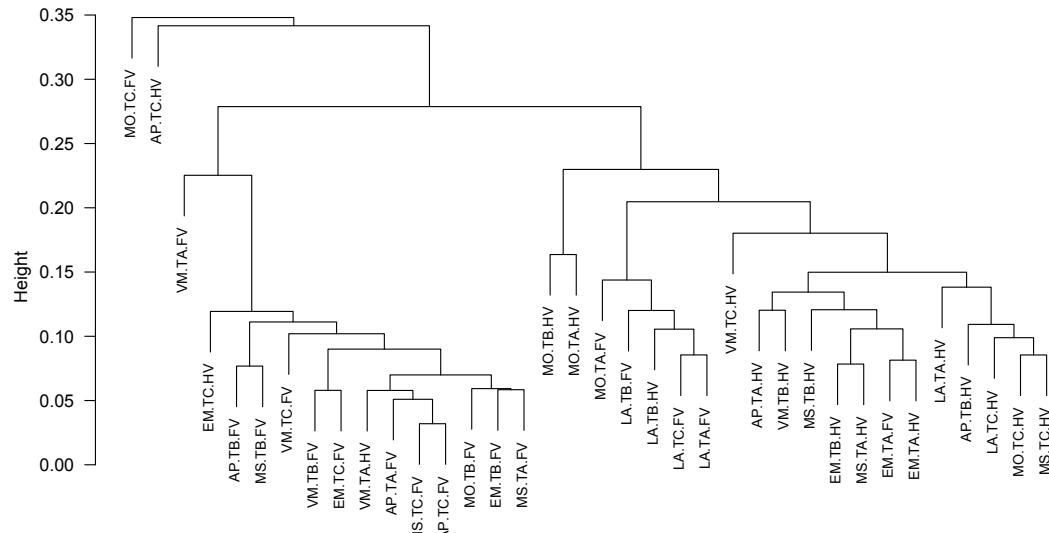
Indice de Bray-Curtis: se basa en el cálculo de porcentaje de cambio en la abundancia de las entidades de una muestra respecto a otra. Es un buen índice cuando la matriz d abundancia contiene demasiados valores cero (ausencia).

$$S_{jk} = \frac{\sum_{i=1}^p 2\min(y_{ij}, y_{ik})}{\sum_{i=1}^p (y_{ij} + y_{ik})}$$

Escalamiento multidimensional no-métrico (nMDS): es una ordenación que se basa en la matriz de semejanzas rankeadas.

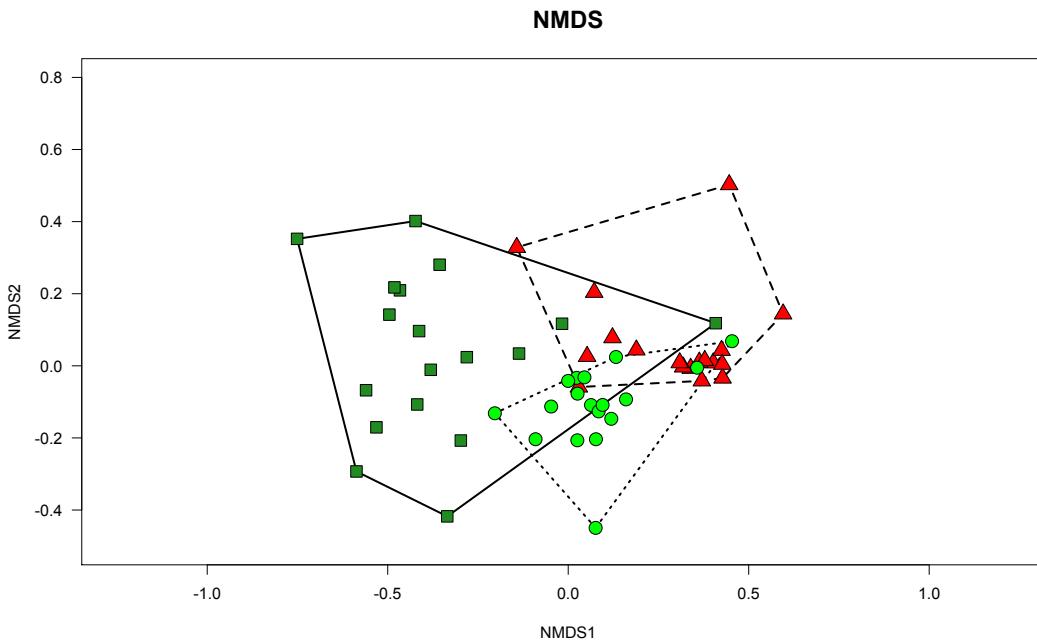


Community ecology



Agrupamiento jerárquico

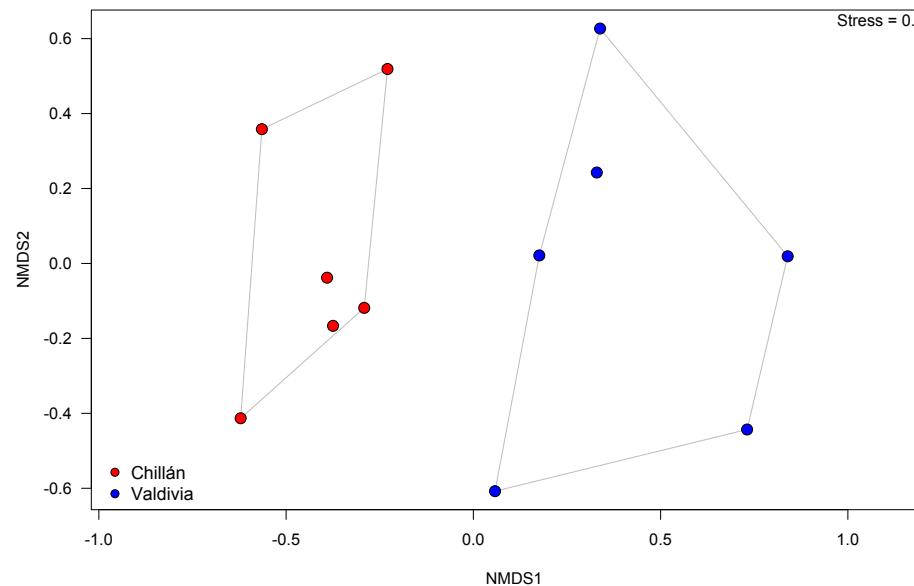
Métodos de ordenación:
PCoA.
NMDS.



Ecología de comunidades

PERMANOVA.

Es un análisis de varianza que usa matrices de distancias, el cual partitiona la variancia entre distintas fuentes de variación usando modelos lineares.



Source	df	MS	F	R ²	P-value
Location	1	0.61	3.21	0.24	0.004
Error	10	0.19		0.76	
Total	1			1	