



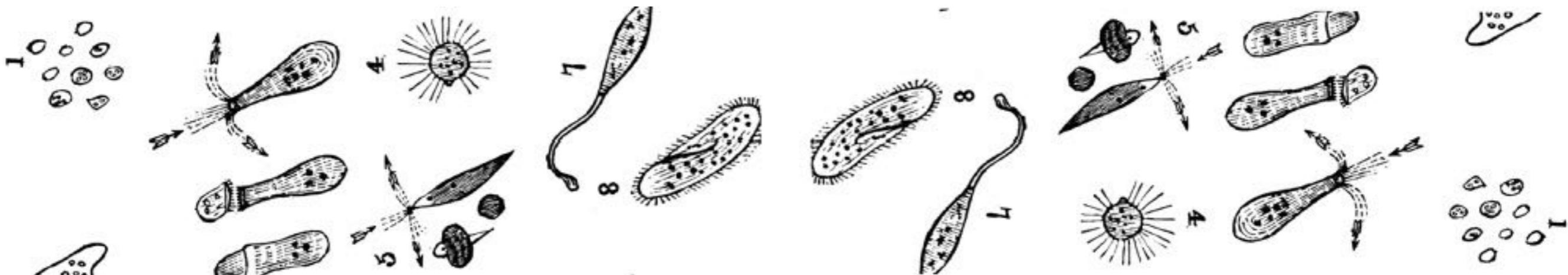
CBIB

CENTER FOR BIOINFORMATICS
& INTEGRATIVE BIOLOGY

Metagenómica

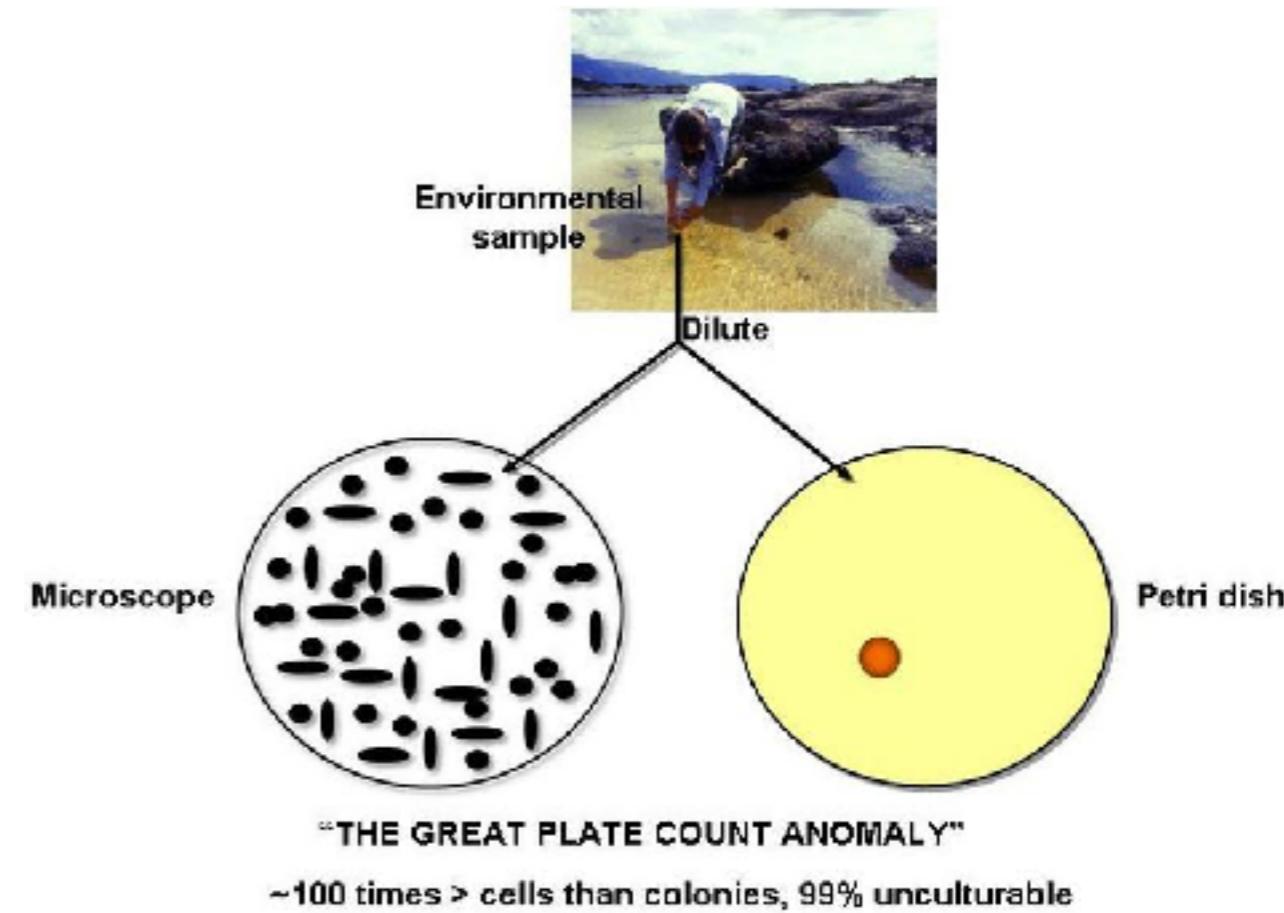
www.castrolab.org
www.cbib.cl

Eduardo Castro, PhD
Universidad Andrés Bello
14 de septiembre de 2018



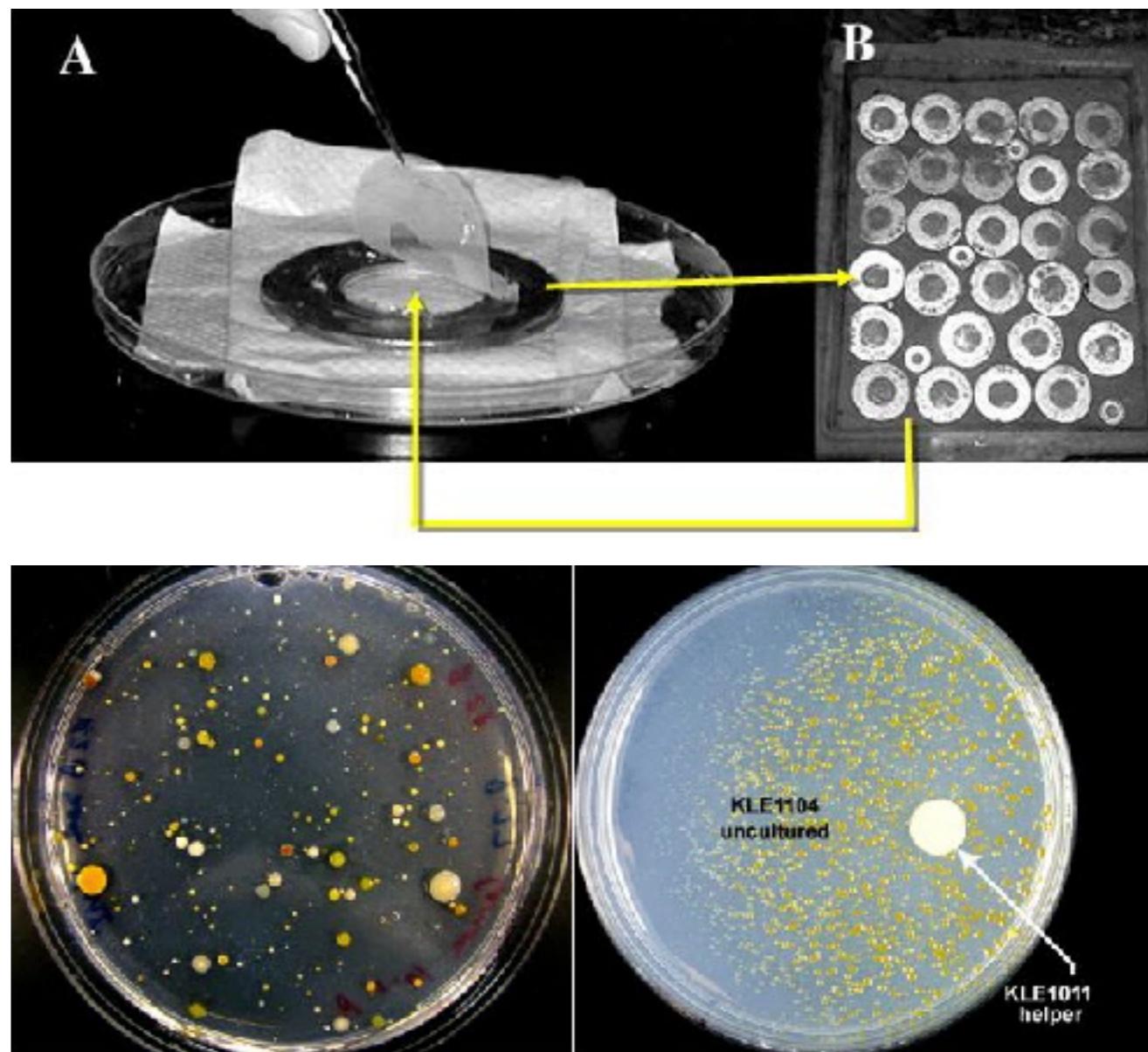
Los microorganismos son incultivables (la mayoría)

- “The Great Plate Count Anomaly”
- Toma una muestra de suelo, mezcla con agua, vórtex, dejar decantar, diluye el sobrenadante y toma dos gotas
- Una gota a placa de Petri con medio LB, y la otra gota al microscopio óptico
- 100:1



Los microorganismos son incultivables (la mayoría)

- Paradoja: genomas similares, uno crece y otro no. Deberían crecer en el mismo medio.
Bacillus; *Pseudomonas*
- No solo no conocemos sus requerimientos nutricionales
- Crecer microorganismos en su ambiente natural
- “Factores de crecimiento”



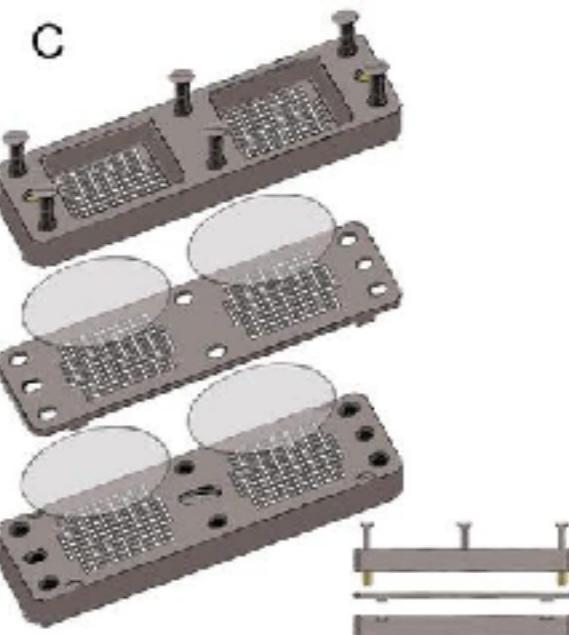
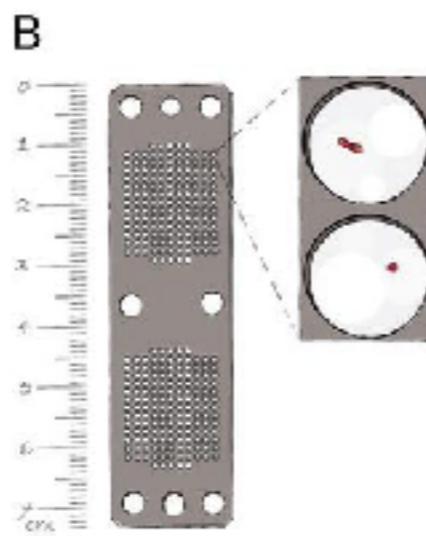
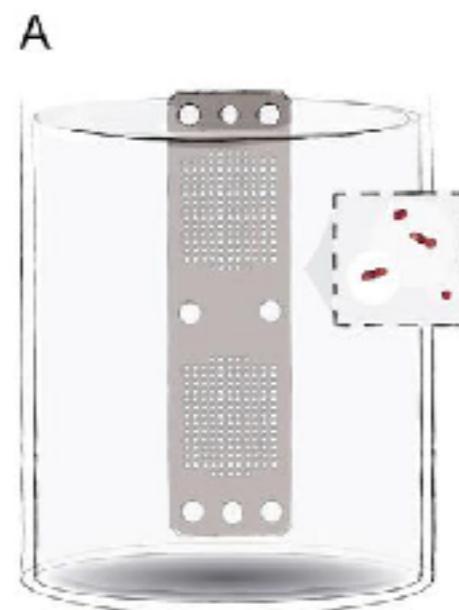
Manibacter polysiphoniae sp.
99.8% identity by 16S rRNA gene

M. luteus sp.

Los microorganismos son incultivables (la mayoría)

Aternativas

- “Persuadirlos” para que crezcan en el lab
- Crecerlos *in situ*
- iChip



Diversidad microbiana

Es posible encontrar microorganismos en la mayoría de los ambientes en la Tierra



¿Cuántos microorganismos hay?

Table 1 Estimating the magnitude of microbial diversity

Number of bacteriophages on Earth	10^{31}
Number of microbes on Earth	5×10^{30}
Number of stars in the universe	7×10^{21}
Number of microbes in all humans	6×10^{23}
Number of humans	6×10^9
Number of microbial cells in one human gut	10^{14}
Number of human cells in one human	10^{13}
Number of microbial genes in one human gut	3×10^6
Number of genes in the human genome	2.5×10^4
Combined length of all bacteriophages on Earth	10^8 Ly
Diameter of the Milky Way	10^5 Ly

Importancia

- Microorganismos manejan y sostienen toda la vida en el planeta
- Ciclo del nitrógeno, carbono, azufre
- Conversión de elementos y compuestos inorgánicos en compuestos orgánicos accesibles a otros organismos
- 99% no cultivable. Grupos taxonómicos completos no contienen un solo representante cultivable
- ¿Cuál es la significancia de la microbiología si estamos limitados a microorganismos cultivables?

Importancia

- Fotosíntesis, fijación de nitrógeno, biodegradación y producción primaria



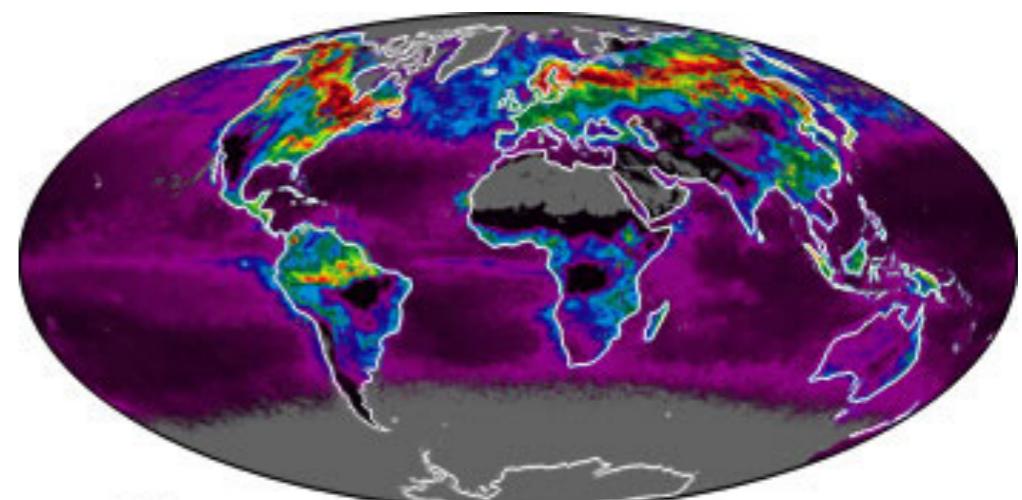
N_2 a compuestos orgánicos
nitrogenados en suelo



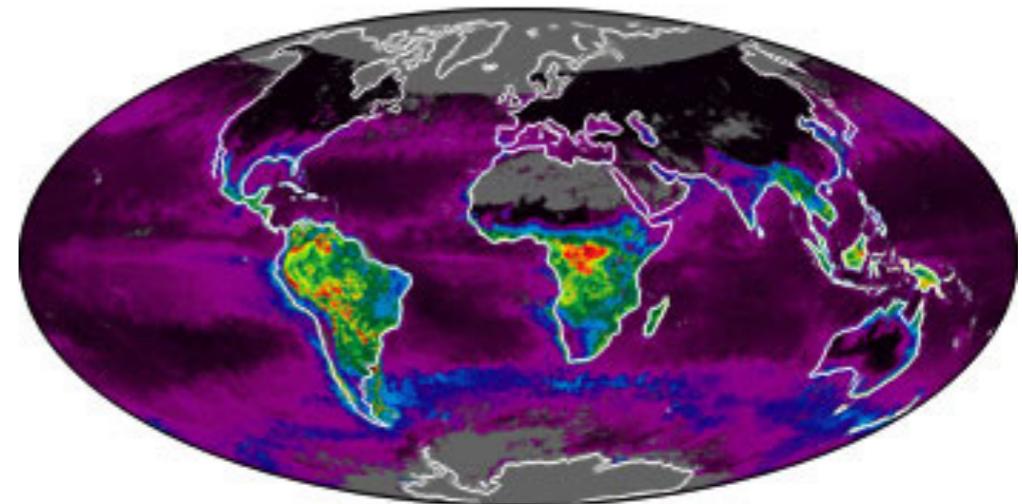
Synechococcus
50% del O₂
mundial



Kg de C fijado



June 2002



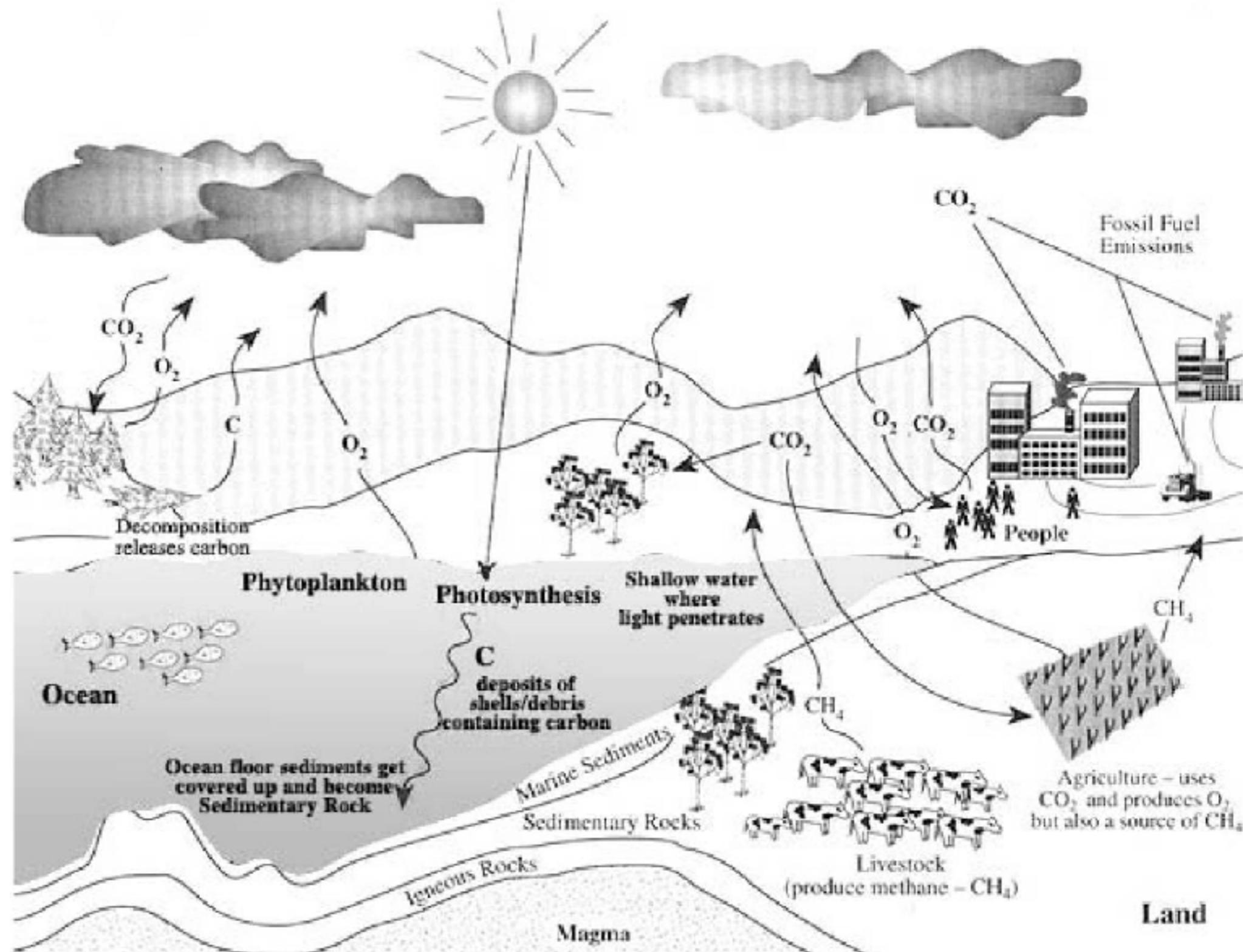
December 2002



Enzimas microbianas degradan
compuestos orgánicos

Importancia

Presentes en el
ciclo del
Carbono → en
todas partes



Antes de la secuenciación masiva

- Denaturing Gradient Gel Electrophoresis (DGGE)

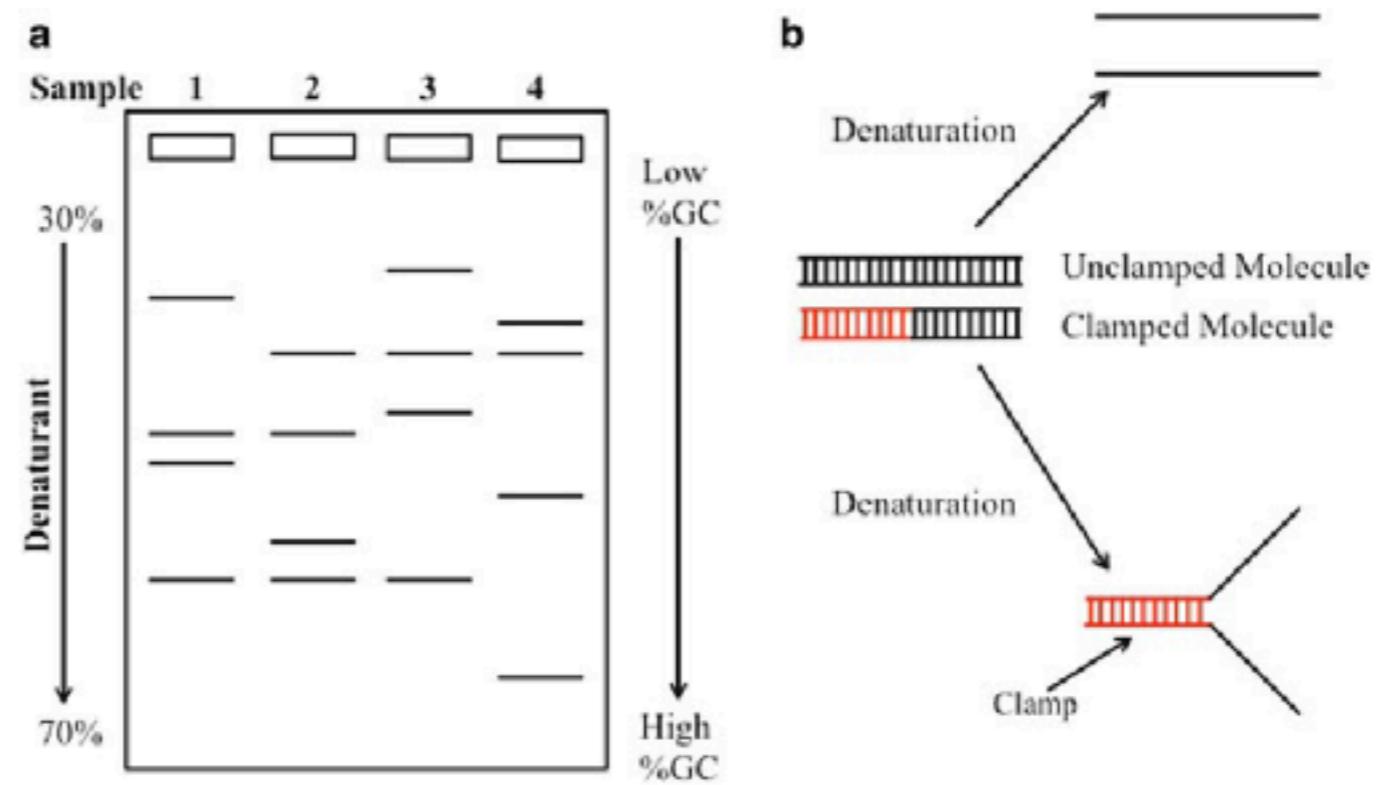
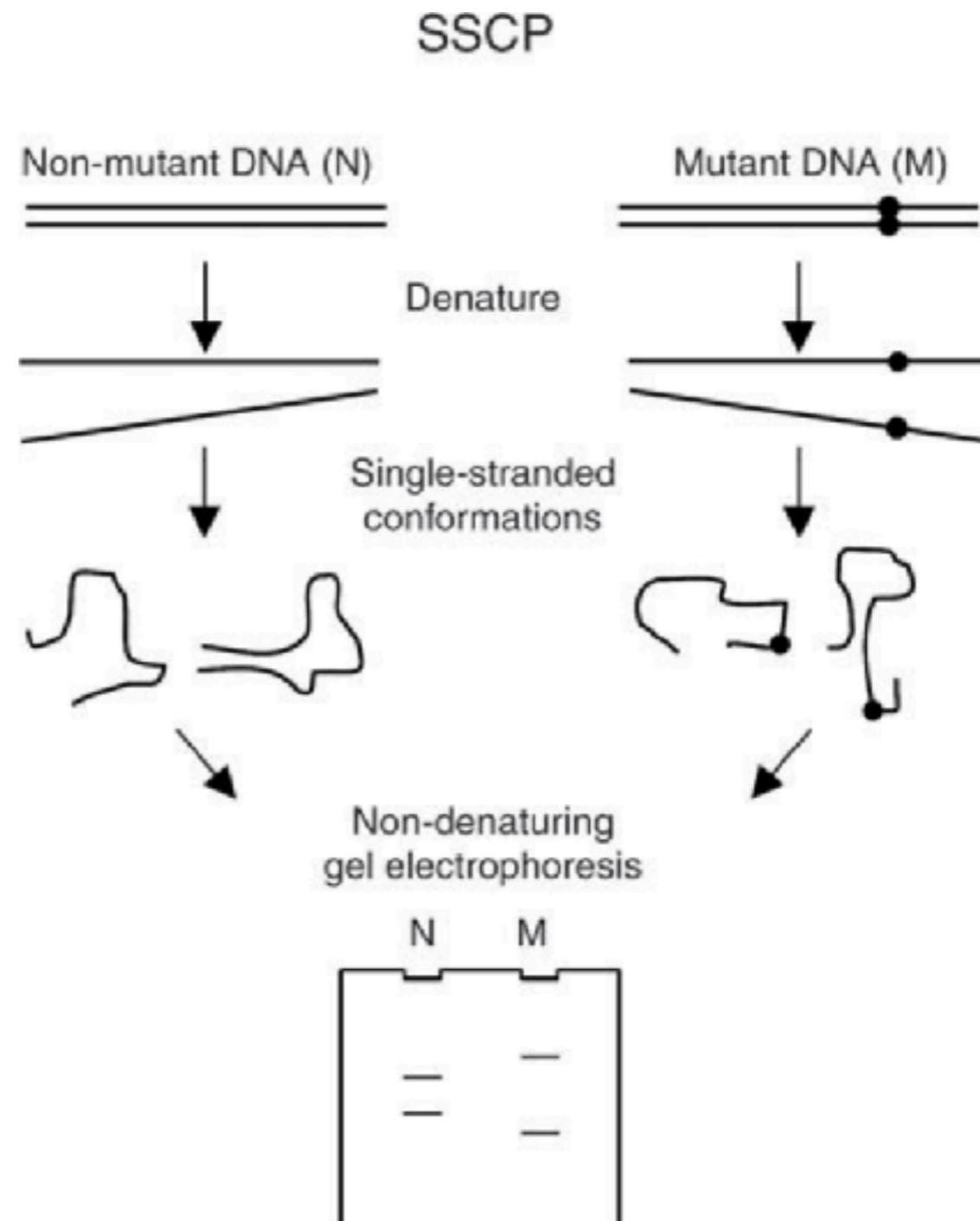
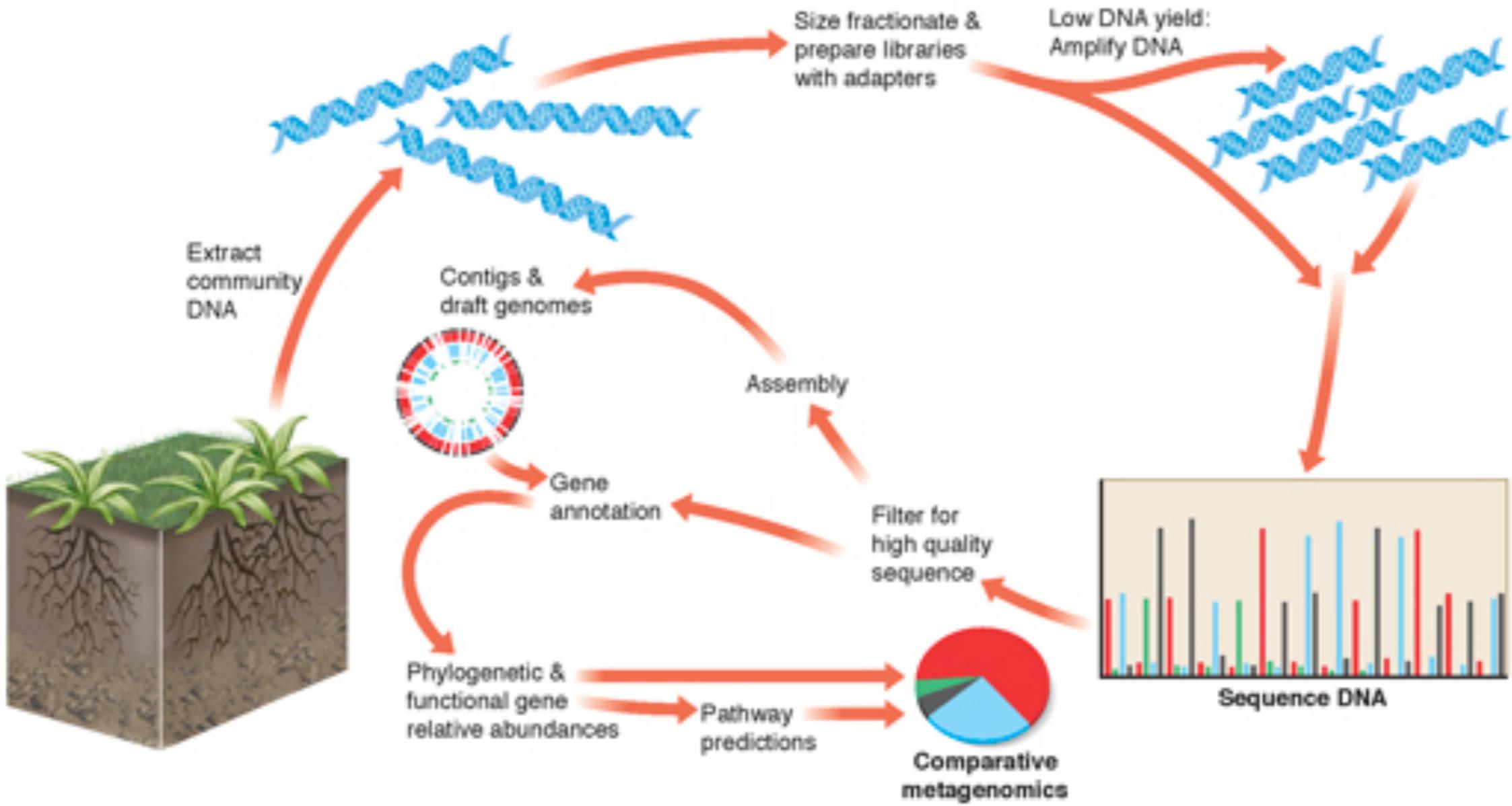


Fig. 1 Principle of denaturing gradient gel electrophoresis. (a) Typical parallel DGGE gel, in which the gradient of denaturant (typically 30–70 %; see Subheading 3) runs parallel to the direction of electrophoresis. Bands are observed at the positions in the gradient at which individual molecules partially denature (b) and hence cease to migrate. Bands of the same mobility correspond to identical sequences (in a fingerprinting assay) or polymorphisms, homo- or heteroduplex forms (in an SNP assay). Where sequences of varying %GC content are separated, the bands lower down the gel correspond to higher %GC sequences, although other sequence-specific factors determine exact melting behavior. (b) Partial denaturation of a double-stranded DNA molecule at a particular point in the gradient; the GC-rich "clamp" region (red), which is added to the PCR amplicons, remains base paired to prevent full denaturation to single DNA strands. A molecule with no clamp denatures fully to single strands and does not form a clear band in the gel

Antes de la secuenciación masiva

- Single Strand Conformation Polymorphism (SSCP)





MICROBE

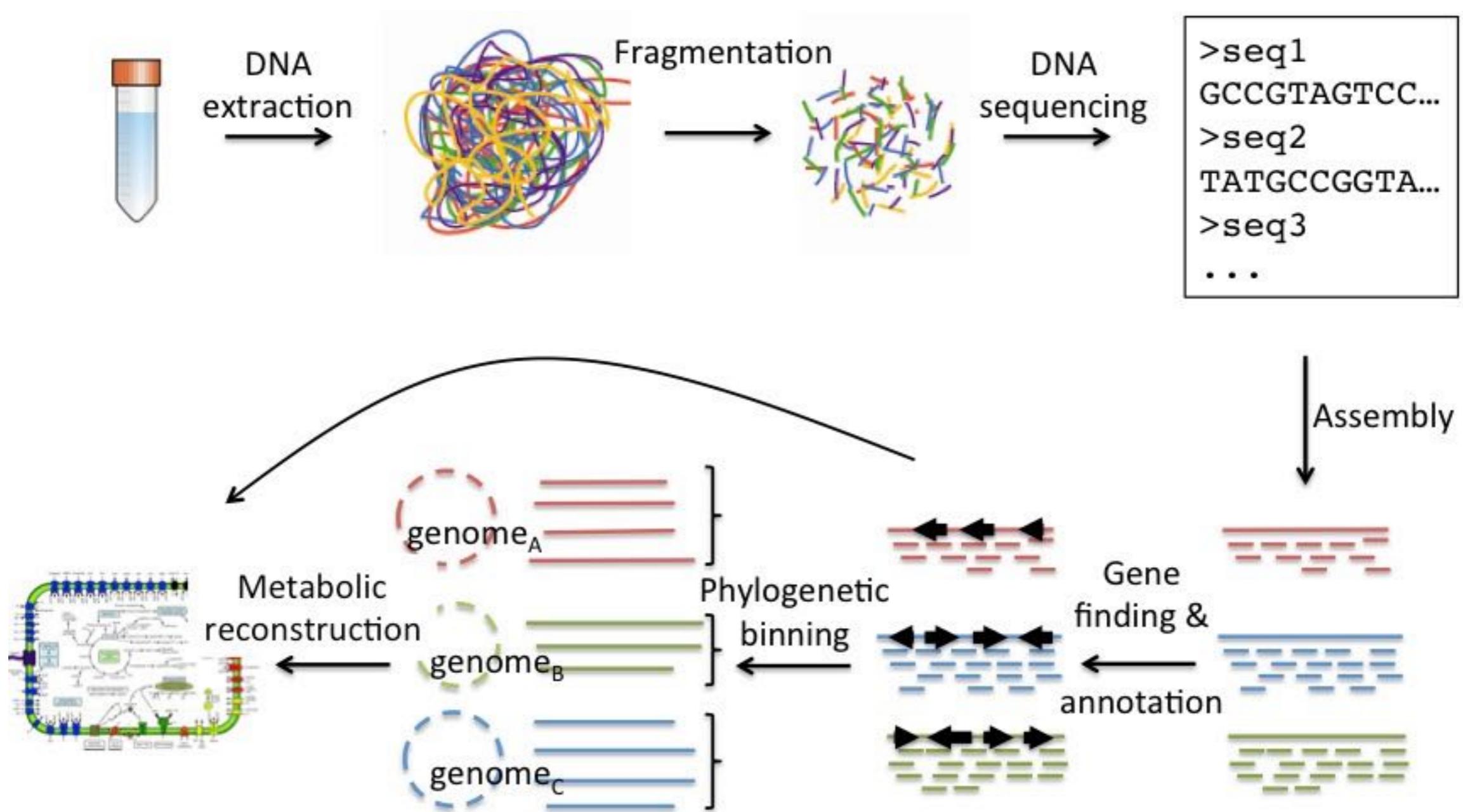
Issue: July 2011, Dr. Jansson

Penumbra Design, Inc. 06/09/11

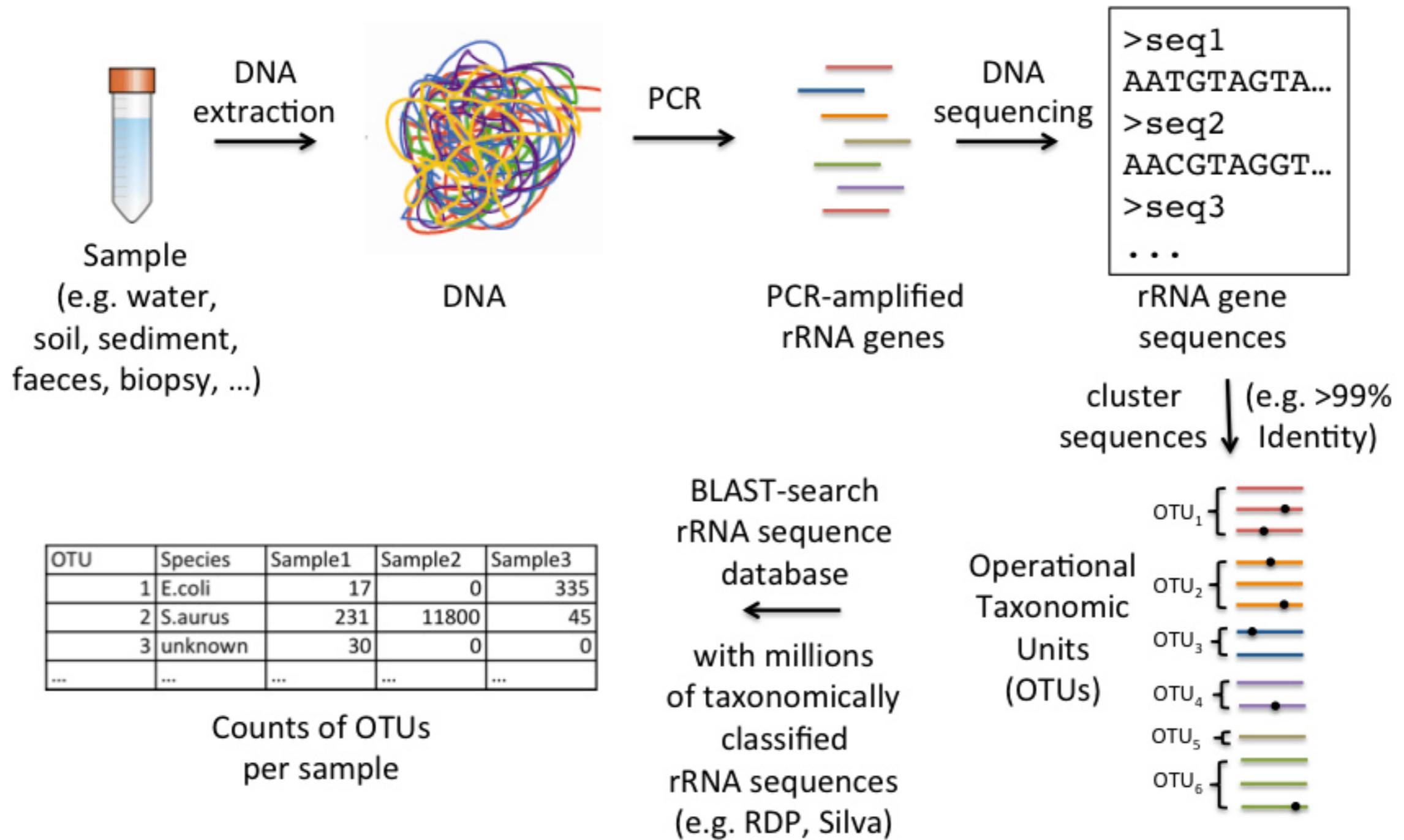
Fig.#01

16S rRNA (amplicones) ≠ Metagenómica

Shotgun Sequencing



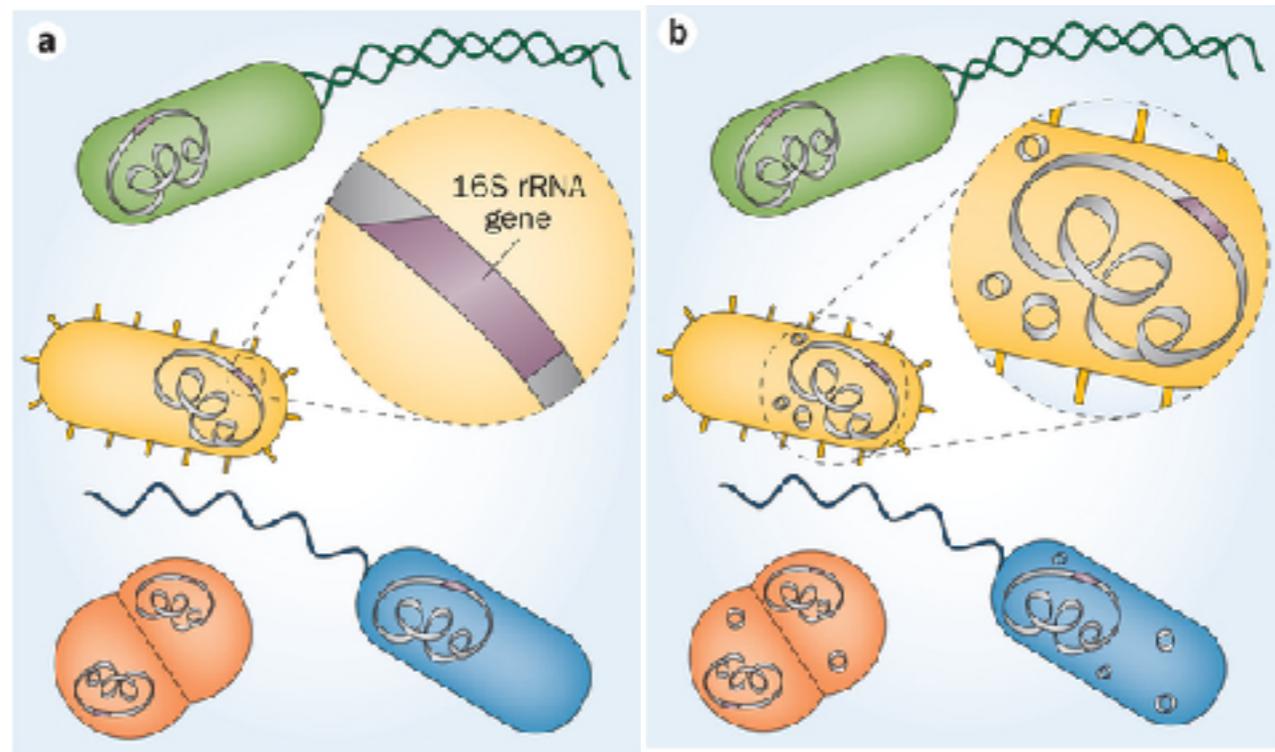
Amplicon Sequencing



AMPLICON SEQUENCING OF RRNA GENES

Definiciones

- Exploración por 16S rRNA o metataxonómica
- Metagenómica
- Microbioma



Marchesi and Ravel. *Microbiome* (2015) 3:31
DOI 10.1186/s40168-015-0094-5



EDITORIAL

Open Access

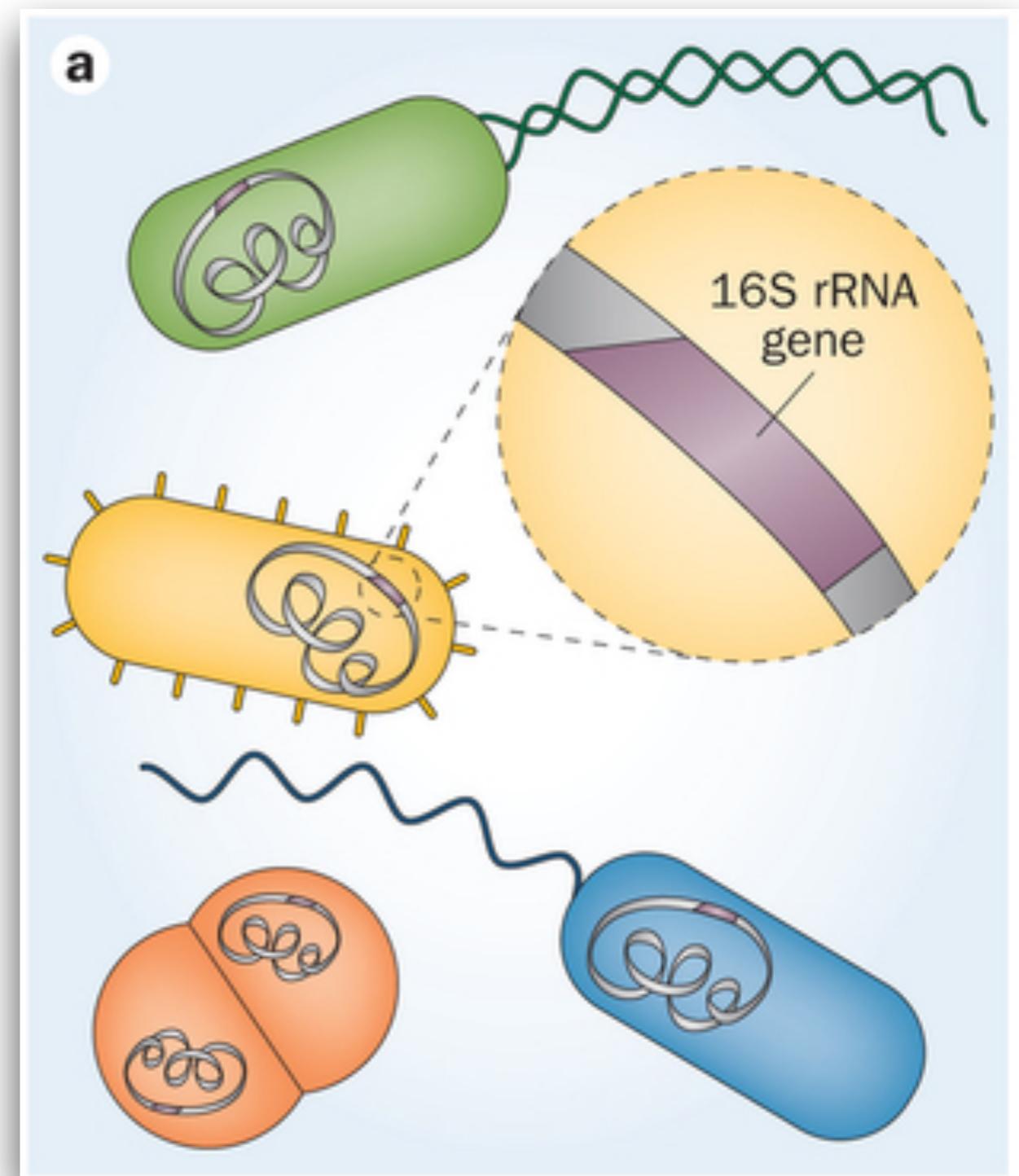
The vocabulary of microbiome research: a proposal



Julian R. Marchesi^{1,2} and Jacques Ravel^{3,4*}

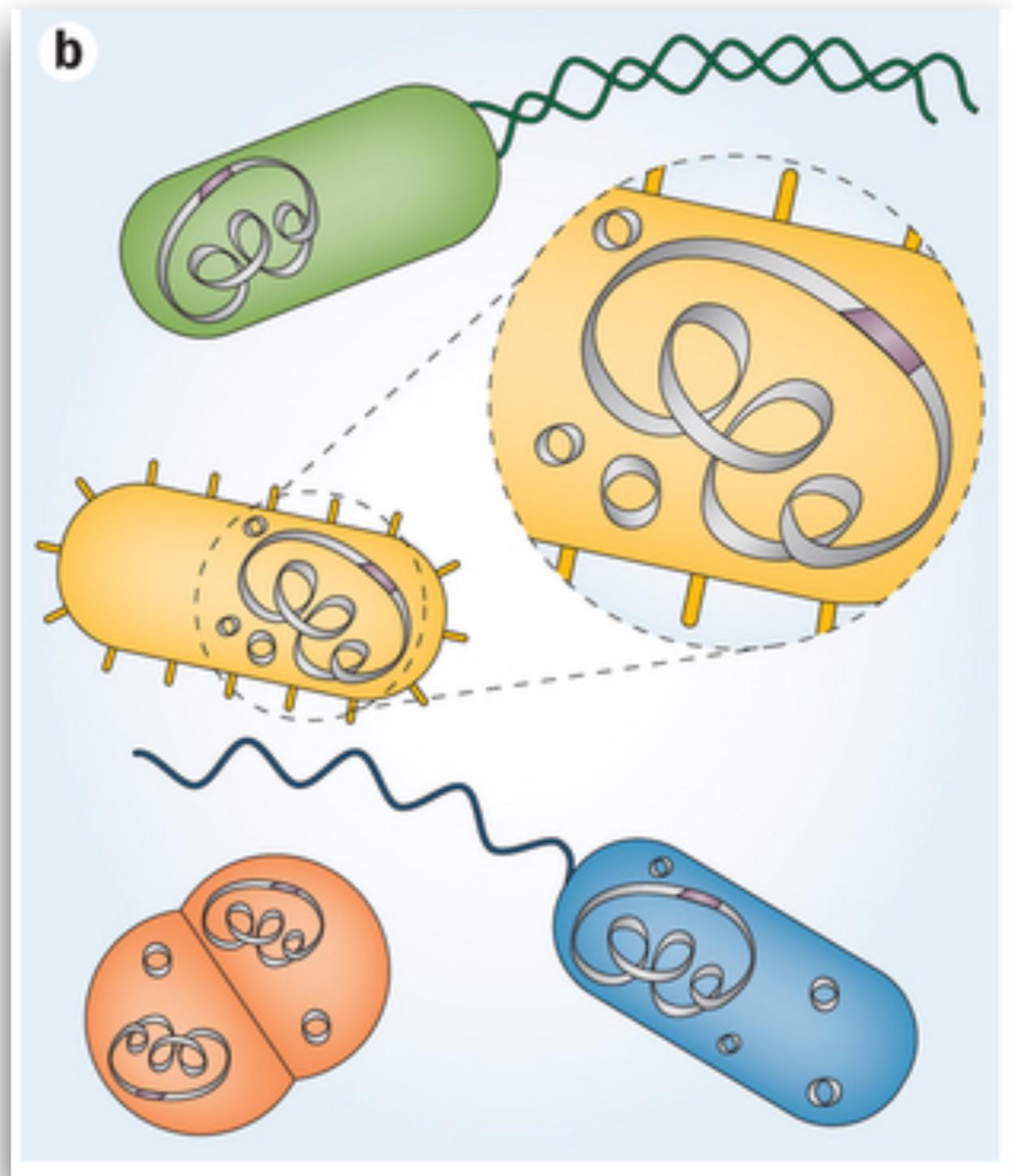
Metataxonómica

- Usar un gen marcador que se pueda asociar con taxonomía - te permite saber qué hay en una muestra - ITS / 16S /18S



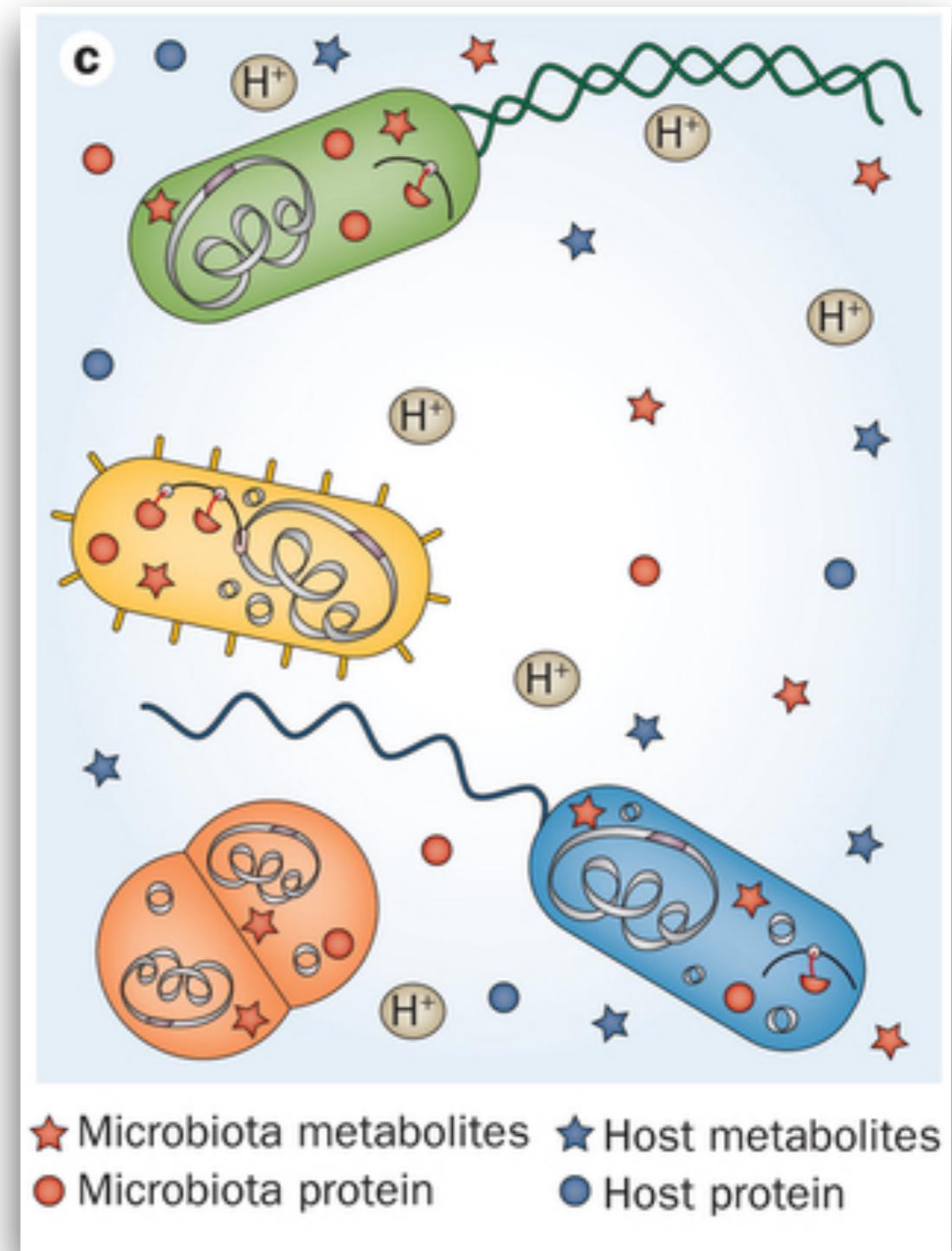
Metagenoma

- Colección de genes y genomas de un ecosistema - más que la taxonomía, también la función - qué hay y qué (potencialmente) hacen



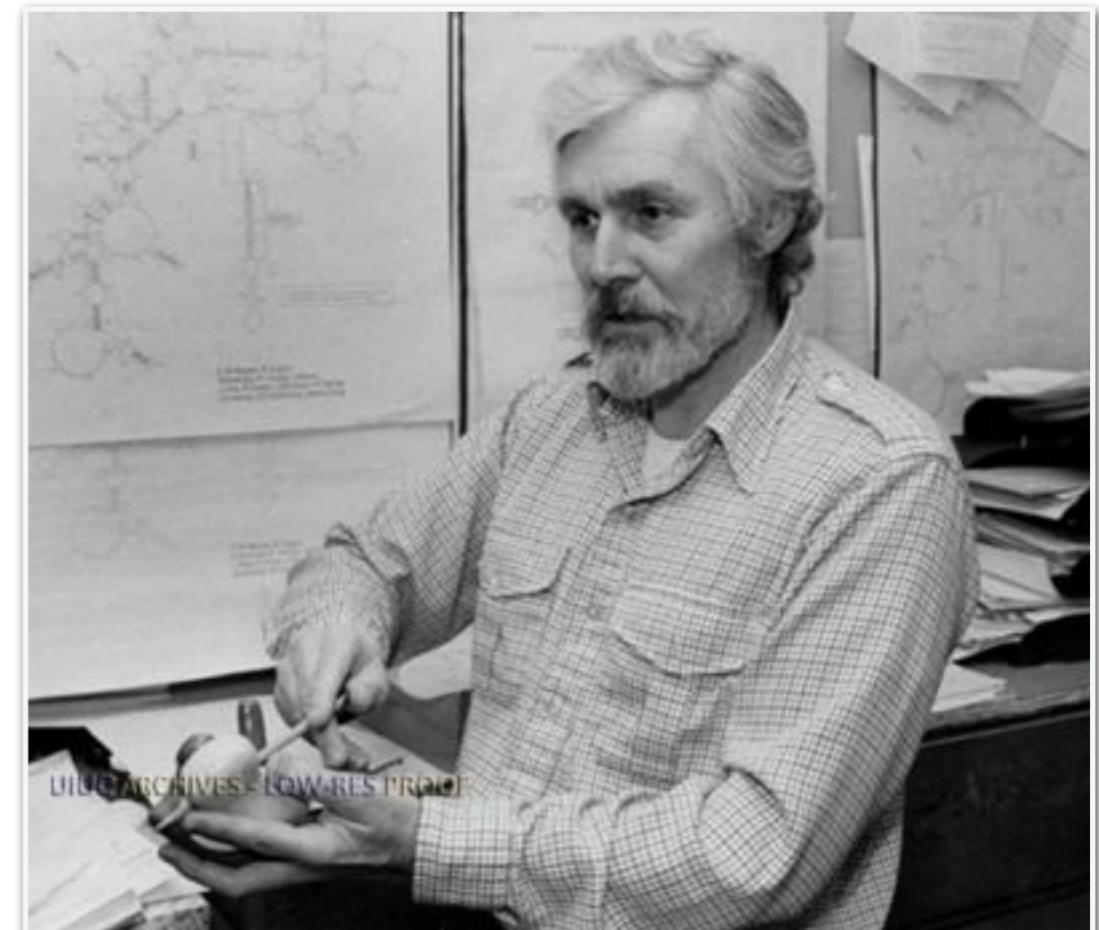
Microbioma

- Similar a metagenóma - aplicado casi exclusivamente a estudios en humanos.
- También como extensión de bioma (factores bióticos y abióticos)



16S rRNA

- Un gen de copia única o múltiple
- Históricamente usado para identificar bacterias y arqueas (género)
- Universal (no se transfiere horizontalmente). Por qué universal?



Carl Woese
1928-2012

Teoría de los tres dominios

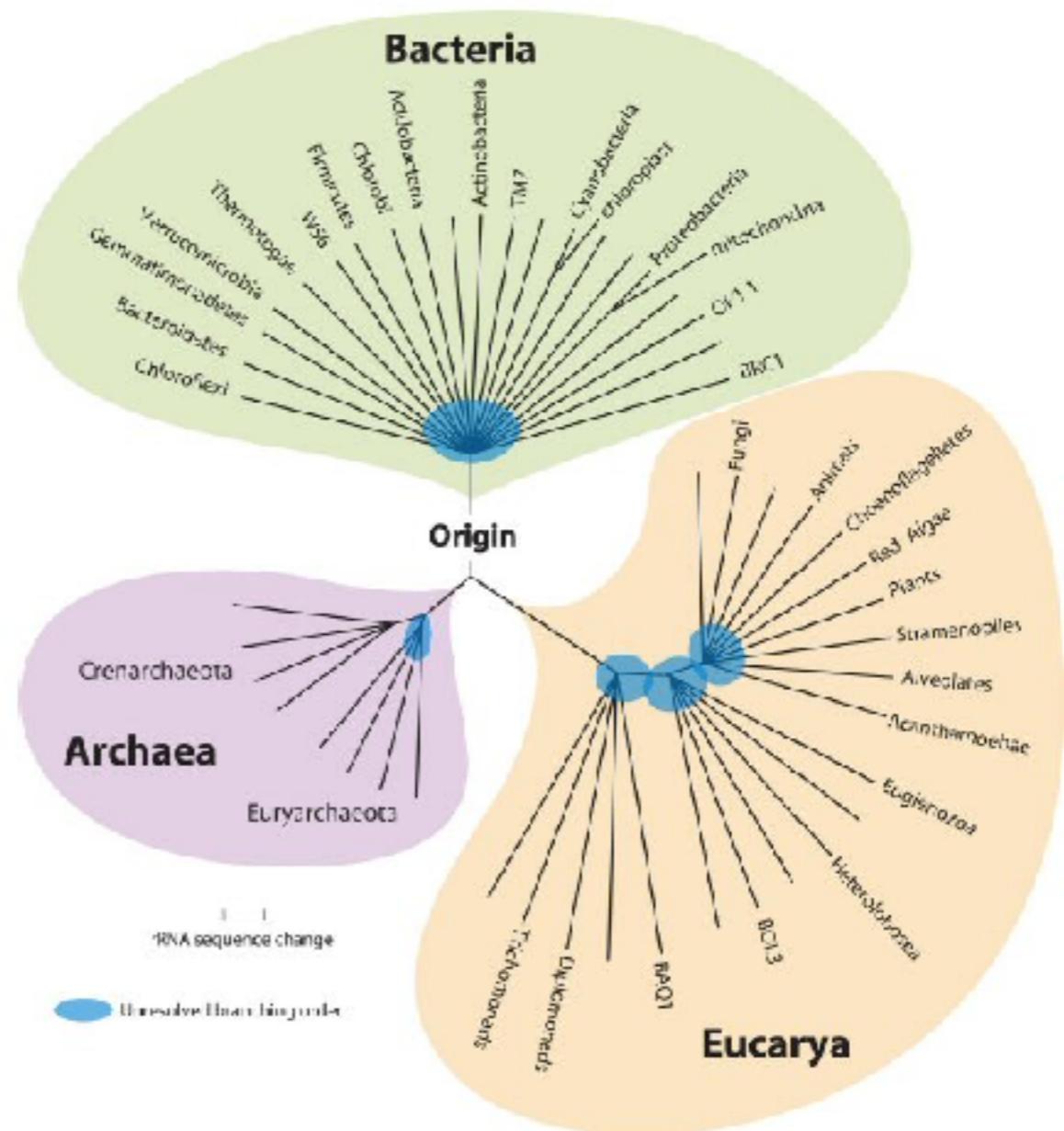
- En 1977 a partir de filogenias de 16/18S
- Arqueas eran consideradas parte de Bacterias

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

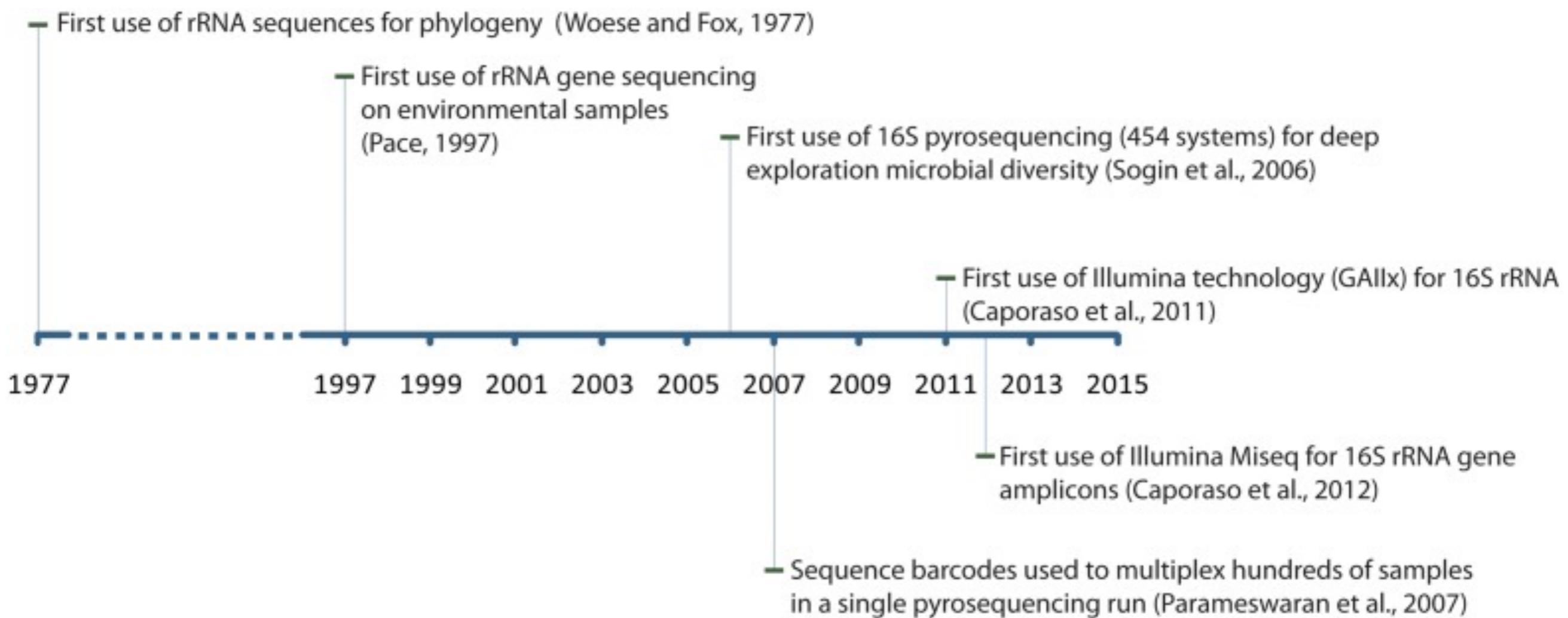
Carl R. Woese and George E. Fox*

Abstract

A phylogenetic analysis based upon ribosomal RNA sequence characterization reveals that living systems represent one of three aboriginal lines of descent: (i) the eubacteria, comprising all typical bacteria; (ii) the archaeabacteria, containing methanogenic bacteria; and (iii) the eukaryotes, now represented in the cytoplasmic component of eukaryotic cells.

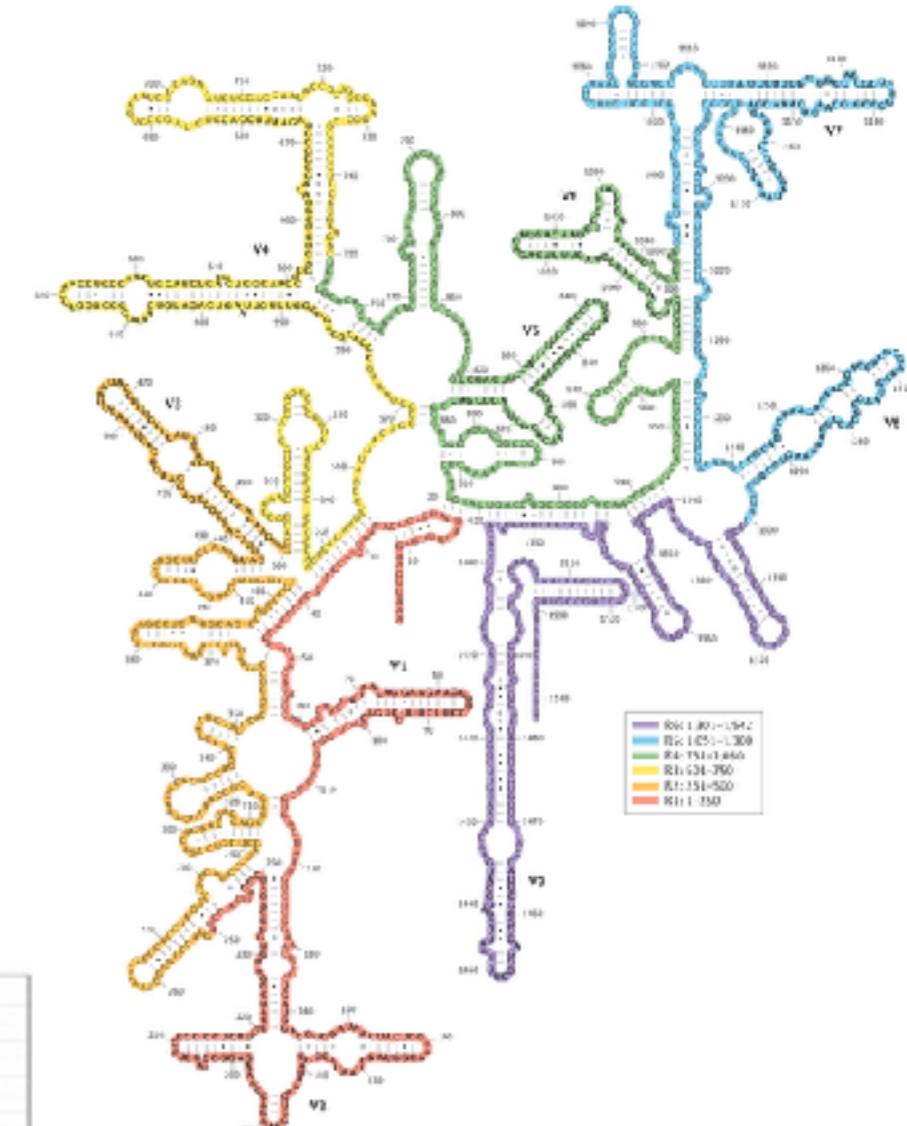
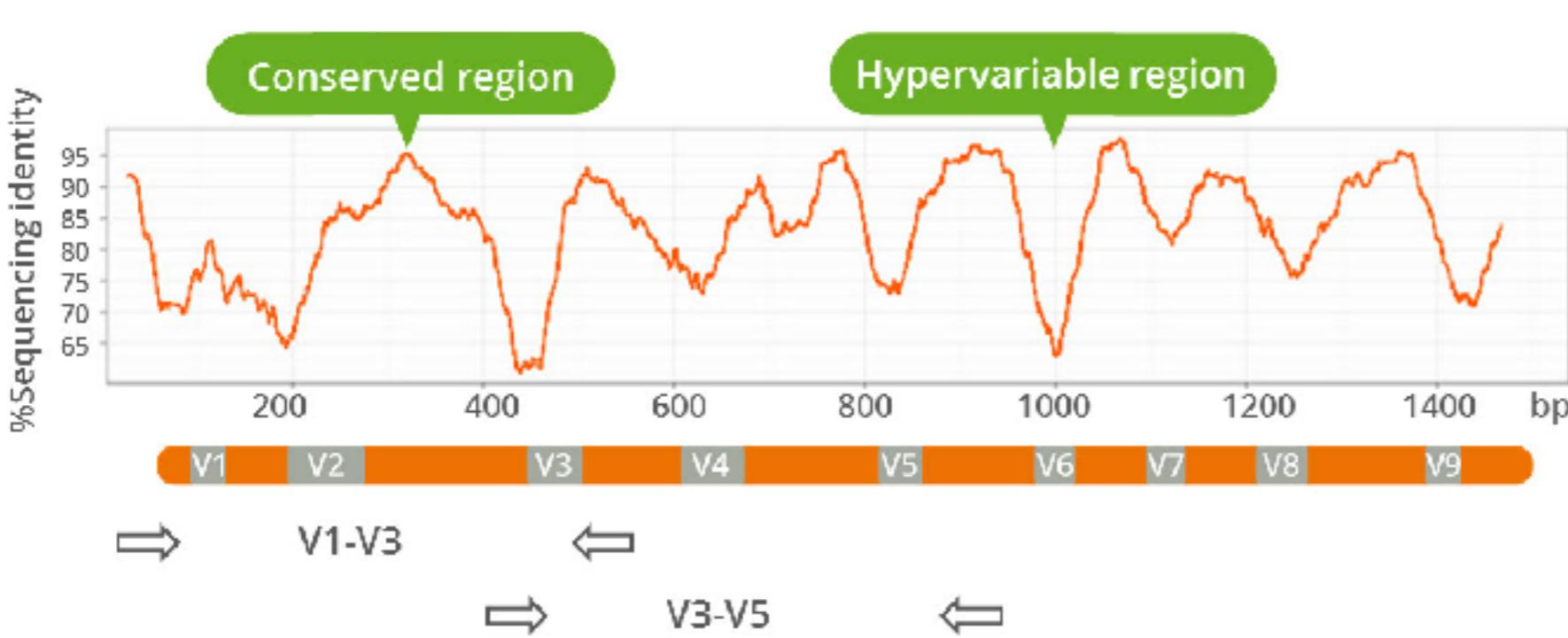


16S rRNA



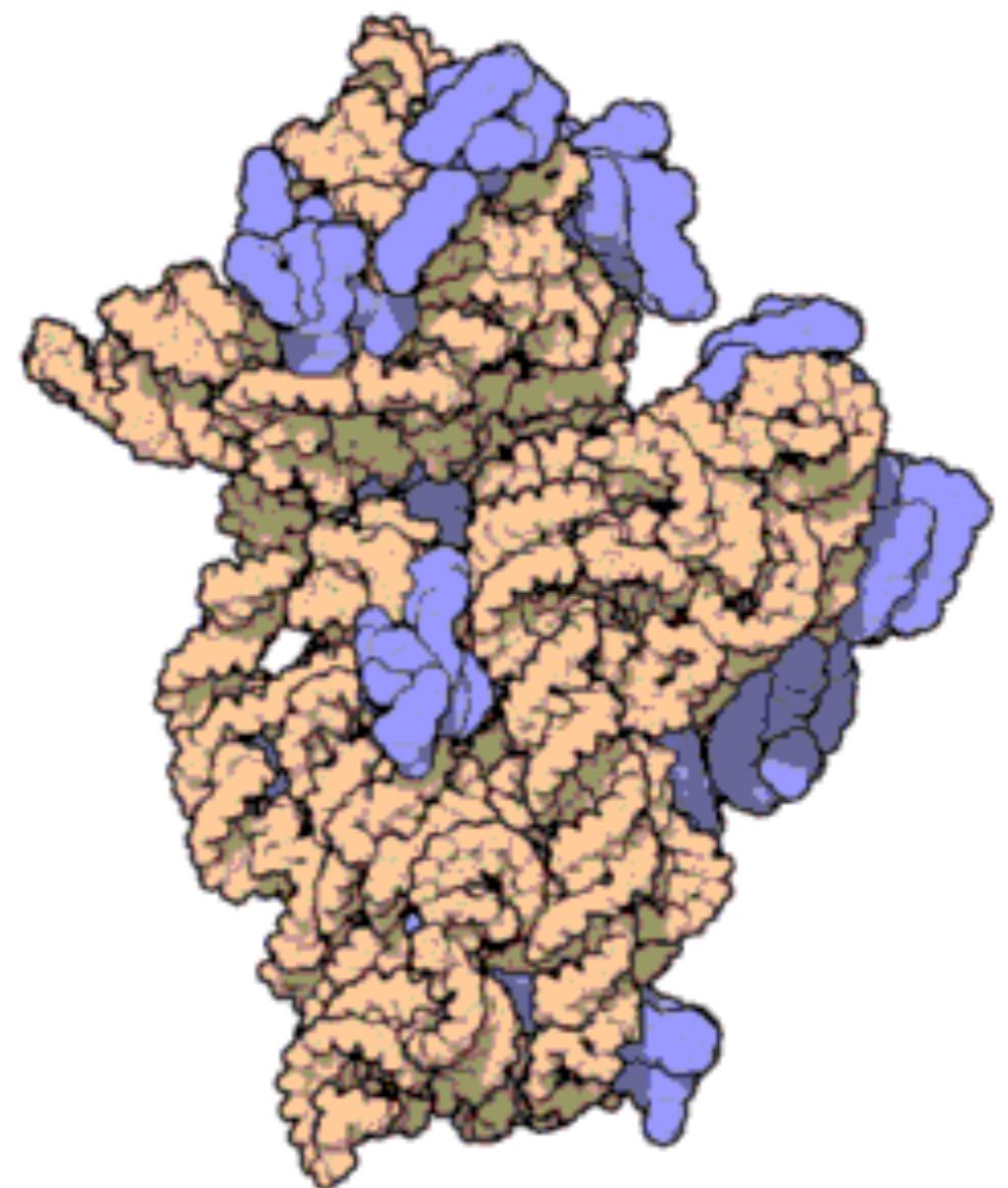
Estructura y ventajas del 16S

- Regiones variables y constantes
- Bases de datos comprehensivas
- Codifica para proteínas?



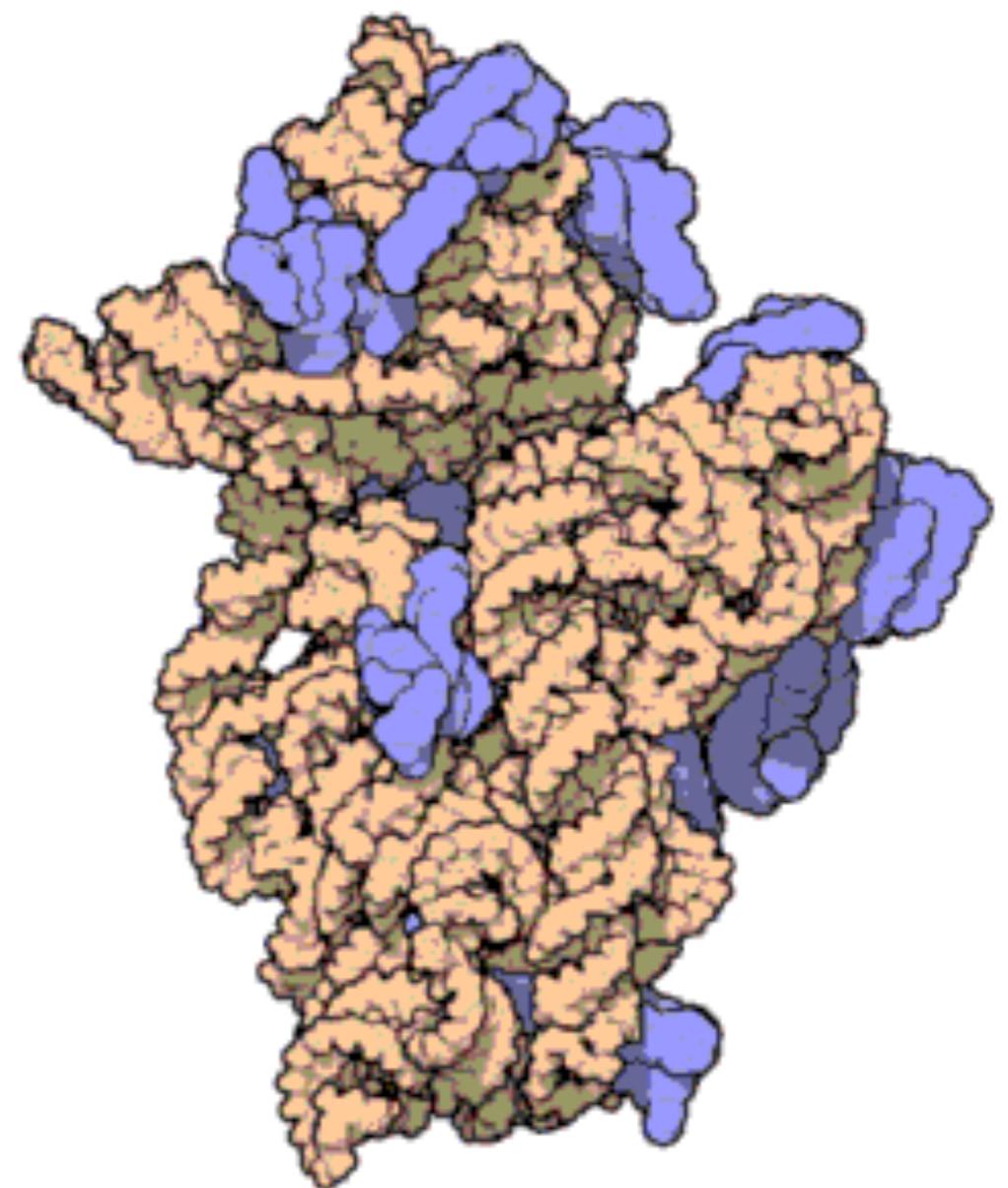
Estructura del 16S

- Parte de la subunidad 30S que se une a la secuencia Shine-Dalgarno (RBS en proks)



Estructura del 16S

- Parte de la subunidad 30S que se une a la secuencia Shine-Dalgarno (RBS en proks)



Poder estadístico

- ¿Cuántas muestras necesito para detectar diferencias de X%?

Alpha = 1%		Reads						
Subjects	500	1000	2500	5000	10000	20000	50000	1000000
10	8.57%	9.56%	10.06%	10.98%	10.51%	10.50%	10.62%	10.17%
15	15.88%	17.42%	18.91%	19.55%	19.85%	19.29%	19.32%	20.10%
25	36.36%	38.81%	41.65%	41.65%	42.91%	42.93%	42.66%	43.54%
50	81.81%	85.60%	87.38%	88.16%	87.50%	87.98%	88.30%	88.59%

Alpha = 5%		Reads						
Subjects	500	1000	2500	5000	10000	20000	50000	1000000
10	23.60%	24.60%	26.30%	22.80%	24.50%	28.20%	25.50%	25.70%
15	32.90%	38.70%	38.60%	40.10%	40.00%	39.10%	37.90%	43.00%
25	61.40%	63.50%	63.90%	65.60%	66.40%	64.90%	66.90%	67.10%
50	93.20%	94.80%	96.50%	95.30%	96.50%	95.40%	96.60%	97.40%

doi:10.1371/journal.pone.0052078.t004

Poder estadístico

- ¿Cuántas muestras necesito para detectar diferencias de X%?

[Bioinformatics](#). 2015 Aug 1;31(15):2461-8. doi: 10.1093/bioinformatics/btv183. Epub 2015 Mar 29.

Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA.

[Kelly BJ¹](#), [Gross R¹](#), [Bittinger K²](#), [Sherrill-Mix S²](#), [Lewis JD³](#), [Collman RG¹](#), [Bushman FD²](#), [Li H³](#).

A web application for sample size and power calculation in case-control microbiome studies 

[Federico Mattiello](#) , [Bie Verbist](#), [Karoline Faust](#), [Jeroen Raes](#), [William D. Shannon](#), [Luc Bijnens](#), [Olivier Thas](#)

Bioinformatics (2016) 32 (13): 2038-2040. DOI: <https://doi.org/10.1093/bioinformatics/btw099>

Published: 19 February 2016 Article history ▾

Método muy sensible - extracción de DNA

- Consenso en que el método de extracción afecta la composición microbiana resultante

Original Research

DNA extraction protocols cause differences in 16S rRNA amplicon sequencing efficiency but not in community profile composition or structure

[Benjamin E. R. Rubin](#) , [Jon G. Sanders](#), [Jarrad Hampton-Marcell](#), [Sarah M. Owens](#), [Jack A. Gilbert](#), [Corrie S. Moreau](#)

The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies

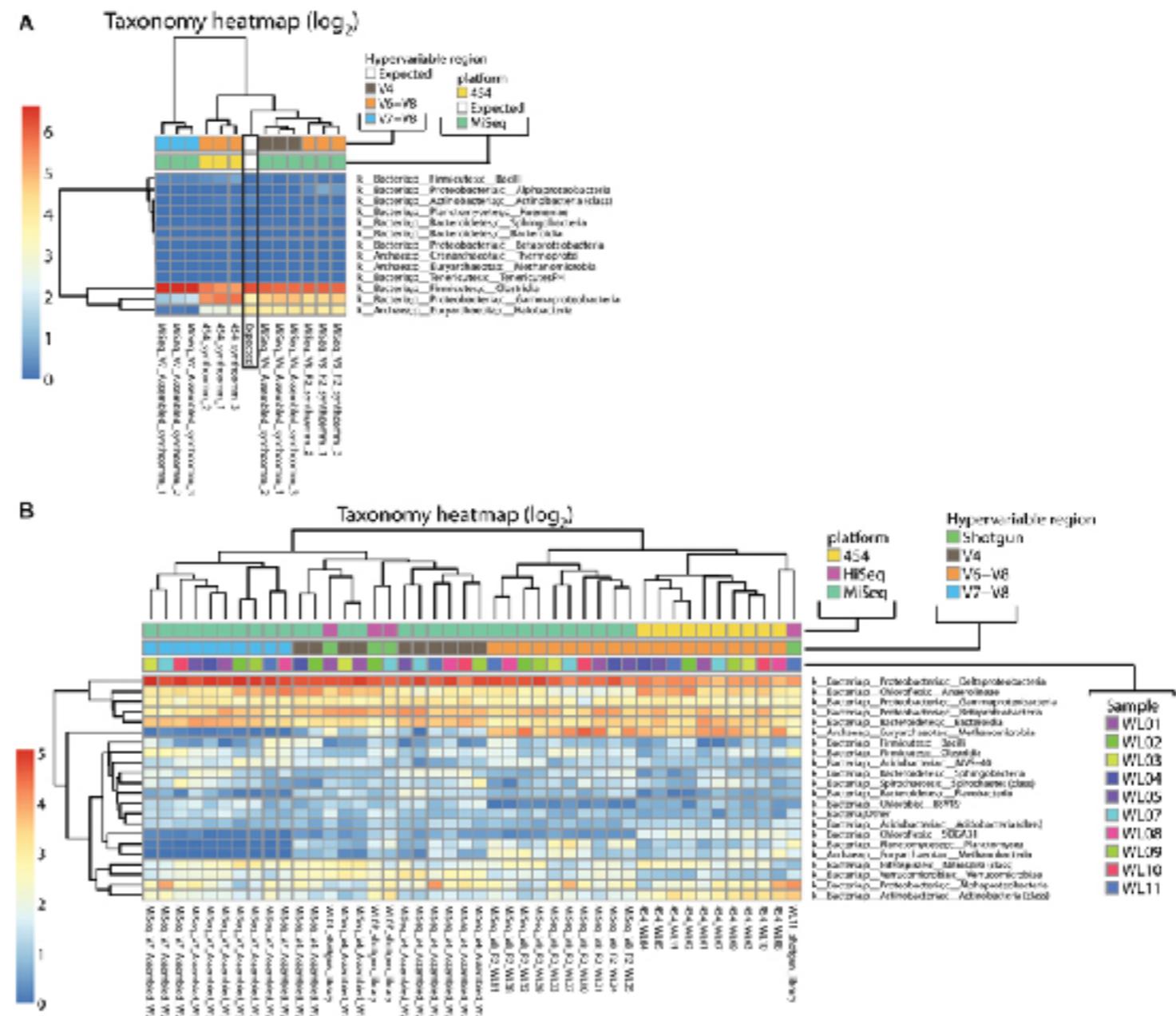
J Paul Brooks , David J Edwards, Michael D Harwich Jr, Maria C Rivera, Jennifer M Fettweis, Myrna G Serrano, Robert A Reris, Nihar U Sheth, Bernice Huang, Philippe Girerd, Vaginal Microbiome Consortium (additional members), Jerome F Strauss III, Kimberly K Jefferson and Gregory A Buck

BMC Microbiology 2015, **15**:66 | DOI: [10.1186/s12866-015-0351-6](https://doi.org/10.1186/s12866-015-0351-6) | © Brooks et al.; licensee BioMed Central, 2015

Received: 17 September 2014 | Accepted: 16 January 2015 | Published: 21 March 2015

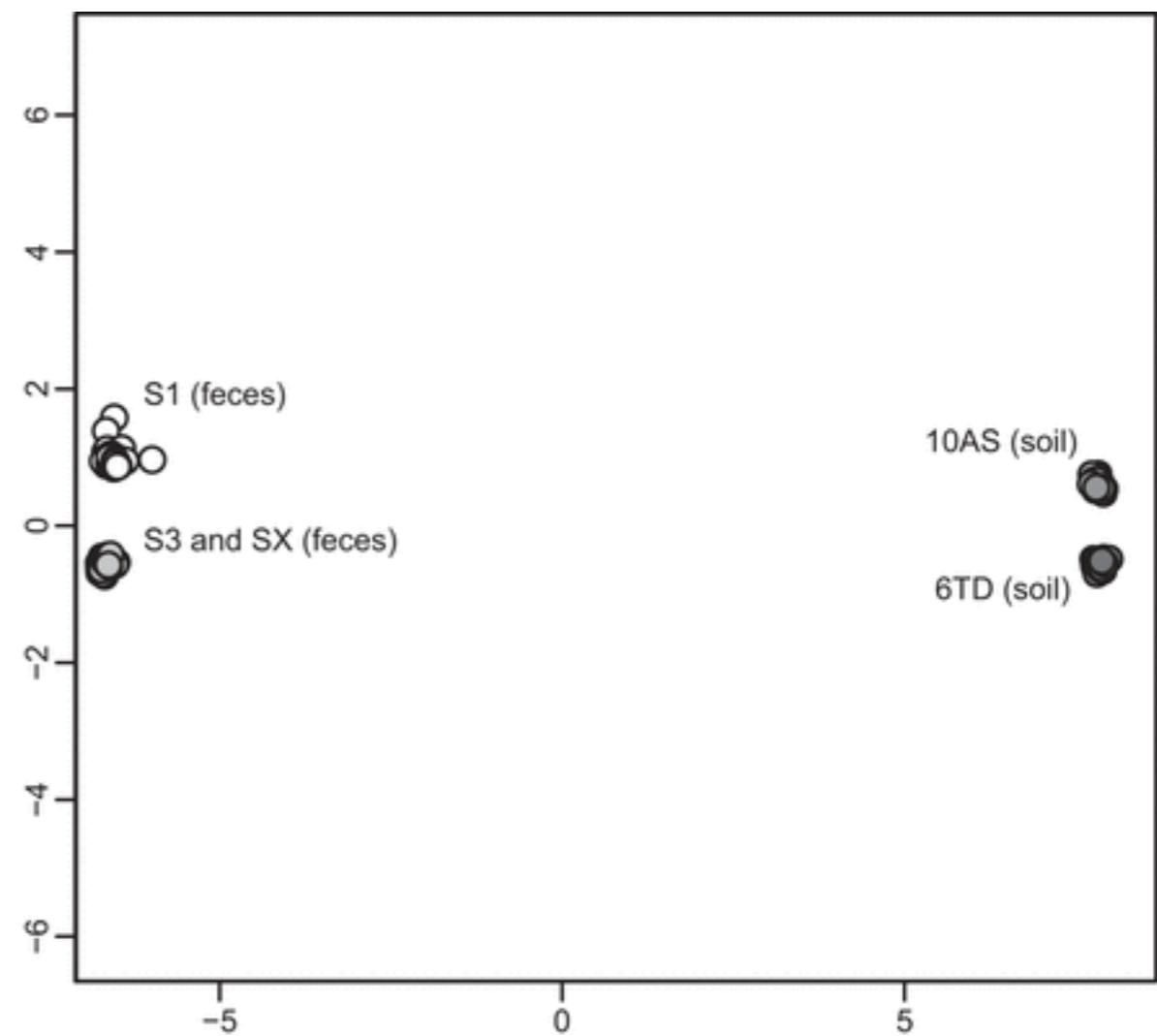
Partidores de PCR

- Distintos partidores tienen preferencia por distintos grupos taxonómicos
- Plataforma de secuenciación tiene mínimo efecto cuando se usan los mismos partidores
- Taxonomía cambia si primers cambian



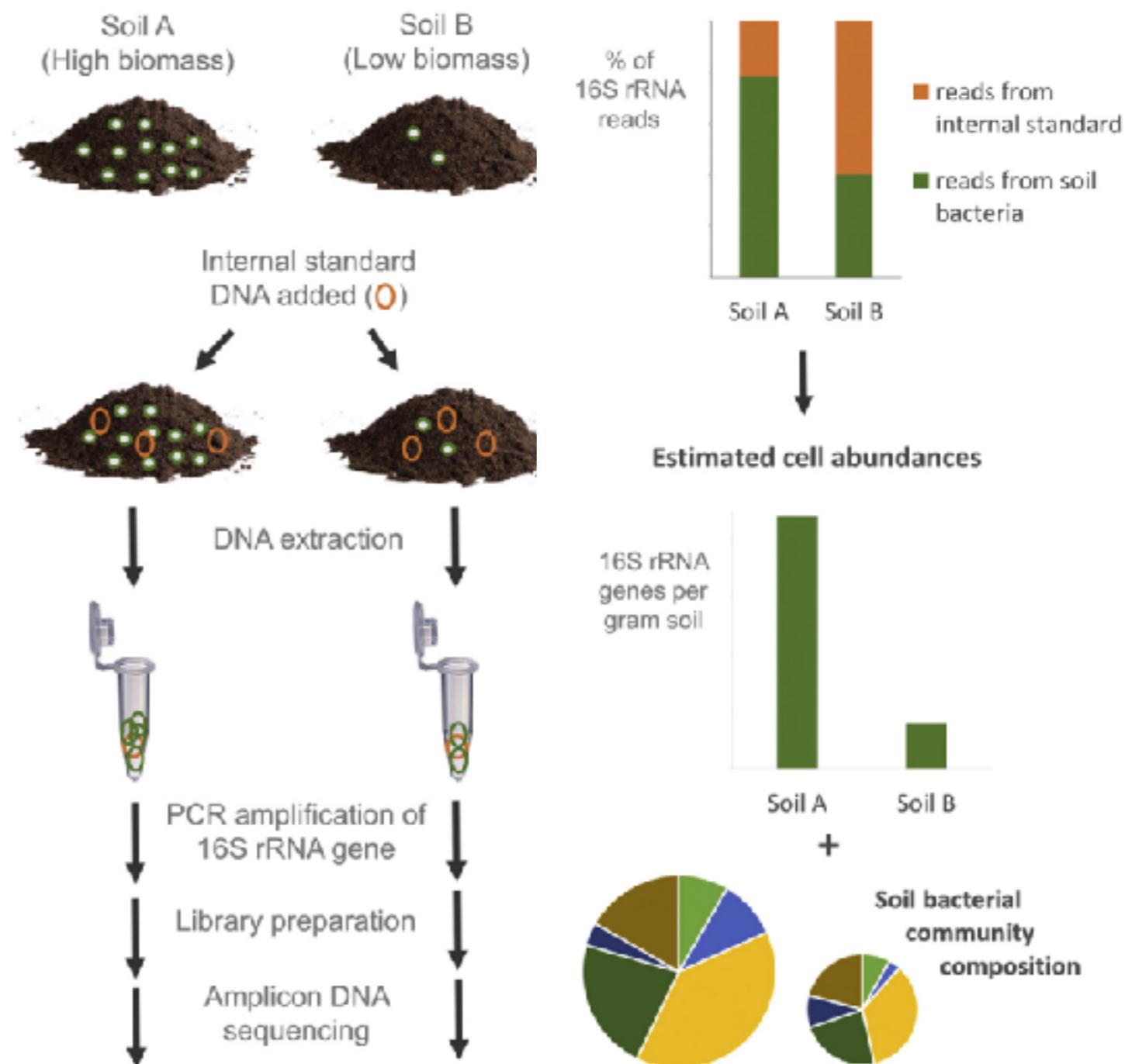
Cantidad de templado

- Distintas concentraciones de templado impactan la composición de microbiota pero introduce baja variabilidad



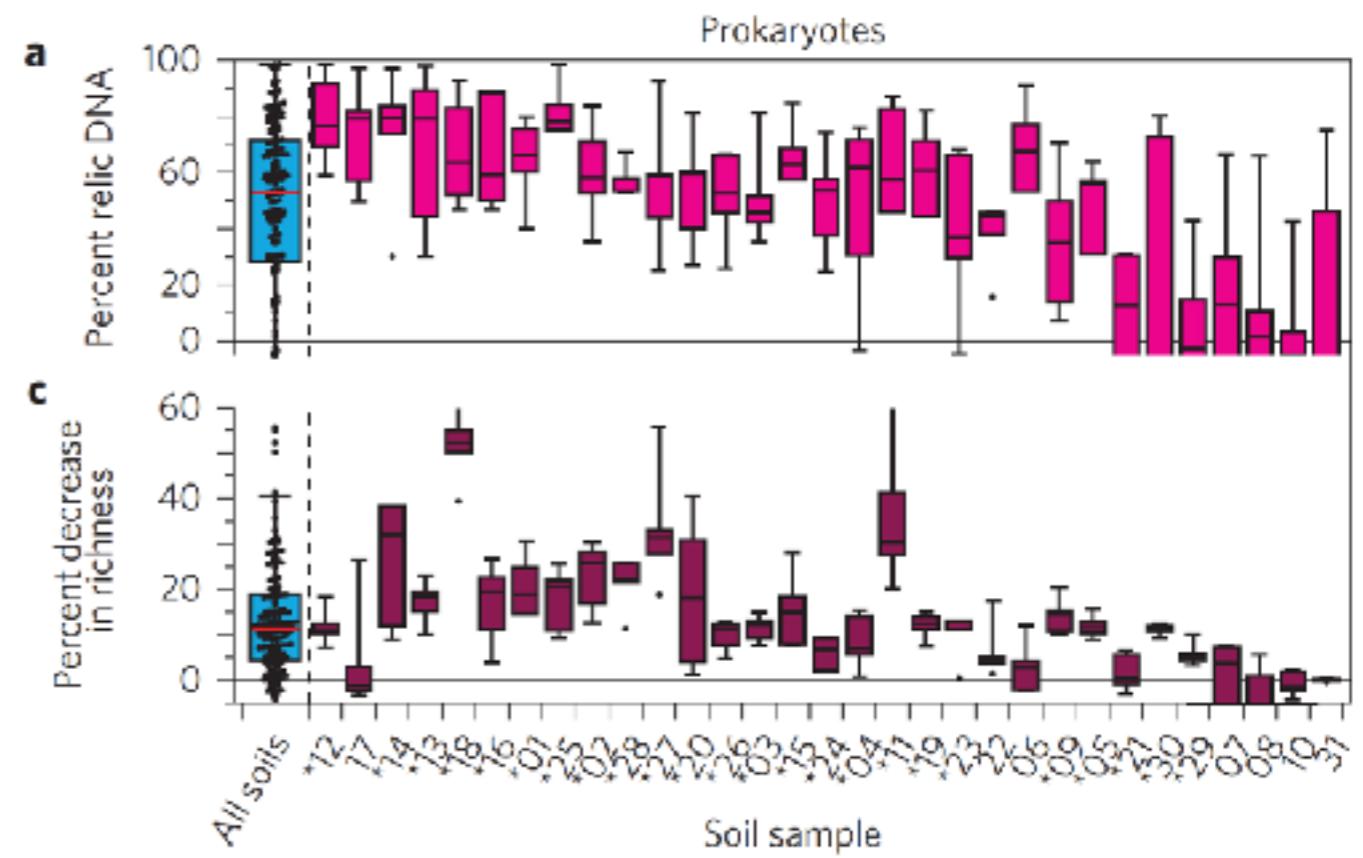
Cuantificación absoluta

- Cuantificación es siempre relativa a menos que se introduzcan controles internos



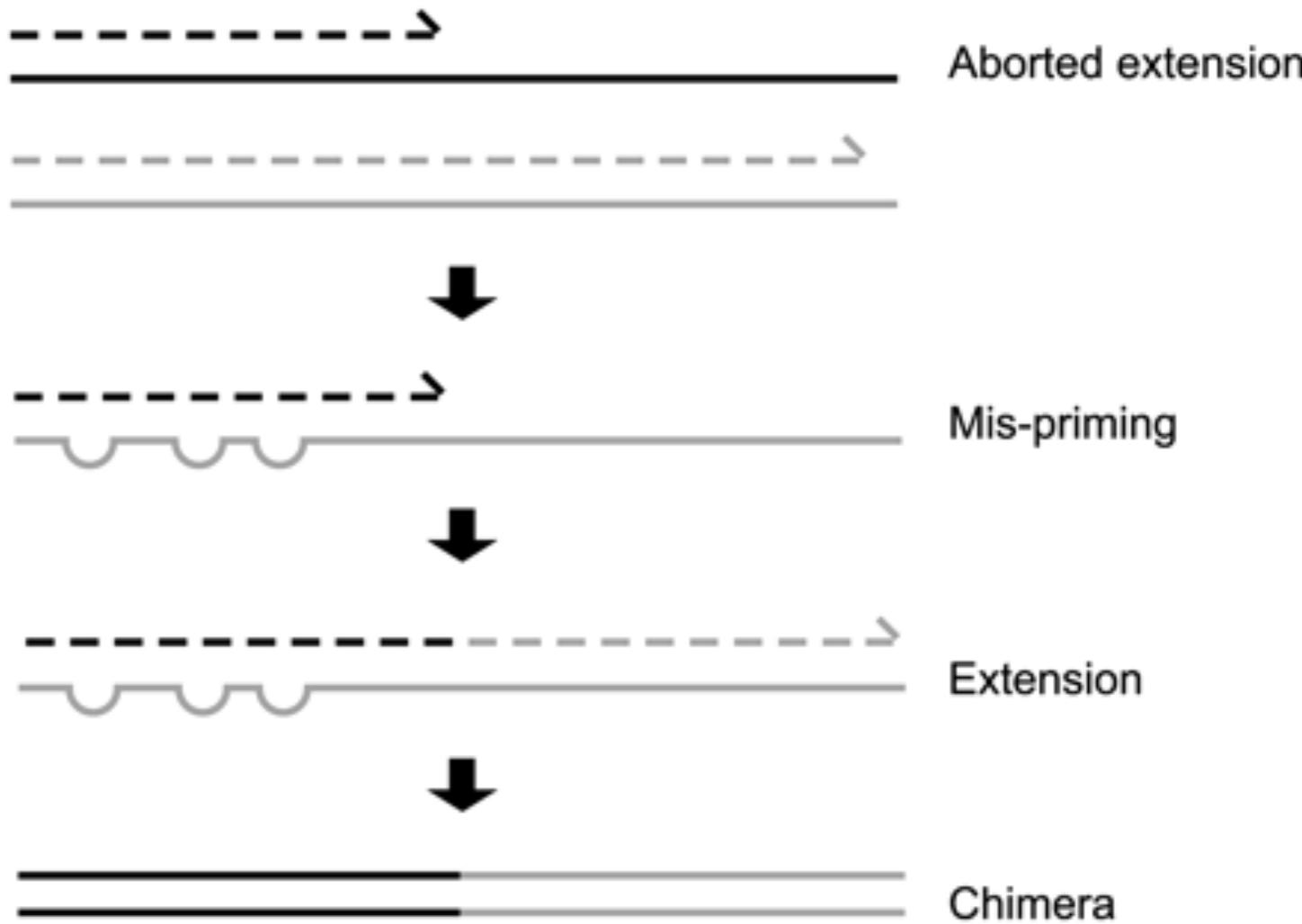
Limitaciones - DNA de células muertas y extracelular

- Hasta un 40% del DNA puede venir de células no intactas o extracelular
- Infla estimados de diversidad alfa
- Estima mal abundancia taxonómica



Limitaciones - Químeras de PCR

- Infla estimados de diversidad
- Afecta filogenias
- Organismos nuevos falsos
- Más prevalente en organismos menos abundantes (hasta 70%)



mothur

- Quality control
- Align sequences
- Clean alignment
- Pre-cluster sequences
- Chimera detection
- Classify sequences
- Remove non bacterial sequences
- Generate distance matrix
- OTU clustering
- Classify OTU
- Taxonomy table

mothur

- Clustering por similitud de secuencia
97% identidad
- Formar OTUs
- Tomar representante
- Clasificar reads

Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities

[PD Schloss, SL Westcott, T Ryabin... - Applied and ..., 2009 - Am Soc Microbiol](#)

ABSTRACT mothur aims to be a comprehensive software package that allows users to use a single piece of software to analyze community sequence data. It builds upon previous tools to provide a flexible and powerful software package for analyzing sequencing data. As a case study, we used mothur to trim, screen, and align sequences; calculate distances; assign sequences to operational taxonomic units; and describe the α and β diversity of ...

[Cited by 6741 Related articles All 20 versions Import into EndNote Save More](#)



www.mothur.org

¿Cómo clasificamos las read?

- Alineamiento estructural en contra de una base de datos curada
- SILVA más popular - más de 20 años
- > 600 mil secuencias

SILVA SSU / LSU 128 - full release

	SSU Parc	SSU Ref	SSU Ref NR 99	LSU Parc	LSU Ref
Minimal length	300	1200/900	1200/900	300	1900
Quality filtering	basic	strong	strong	basic	strong
Guide Tree	no	no	yes	no	yes
Release date	28.09.16	28.09.16	28.09.16	28.09.16	28.09.16
Aligned rRNA sequences	5,616,941	1,922,213	645,151	735,238	154,297

[HTML] The SILVA ribosomal RNA gene database project: improved data processing and web-based tools

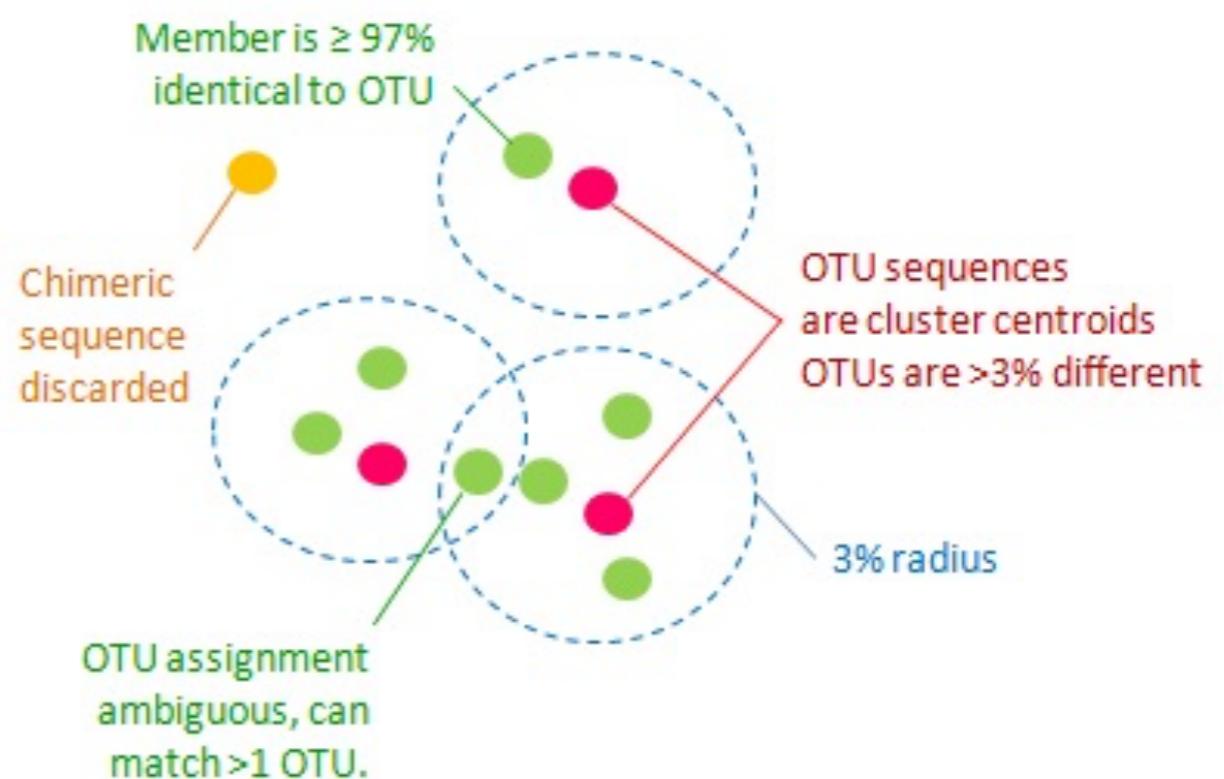
C Quast, E Pruesse, P Yilmaz, J Gerken... - Nucleic acids ..., 2013 - academic.oup.com

Abstract **SILVA** (from Latin **silva**, forest, <http://www.arb-silva.de>) is a comprehensive web resource for up to date, quality-controlled databases of aligned **ribosomal RNA (rRNA) gene** sequences from the Bacteria, Archaea and Eukaryota domains and supplementary online

Cited by 2107 Related articles All 17 versions Import into EndNote Save More

Operational Taxonomic Units

- Unidades taxonómicas operacionales
- ¿Por qué no especies?
- Mismo 16S, diferente genoma
- Misma especie, diferente 16S
- OTUs son clusters de secuencias que divergen como máximo a X% de identidad (3% normalmente)



Clin Microbiol Rev. 2004 Oct; 17(4): 840-862.
doi: [10.1128/CMR.17.4.840-862.2004](https://doi.org/10.1128/CMR.17.4.840-862.2004)

Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases

Jill E. Clamidge,¹*

Author information ► Copyright and Licensee Information ►

PMCID: PMC523561

Appl Environ Microbiol. 2004 Aug; 70(8): 4831–4839.
doi: [10.1128/AEM.70.8.4831-4839.2004](https://doi.org/10.1128/AEM.70.8.4831-4839.2004)

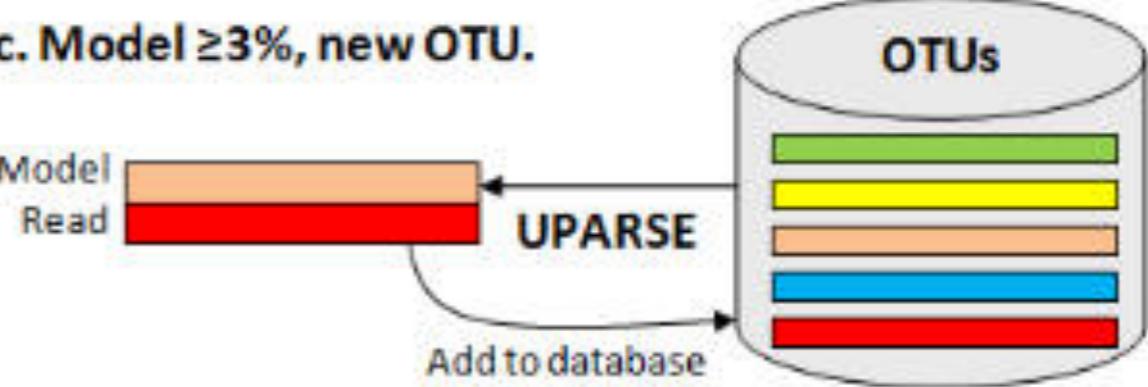
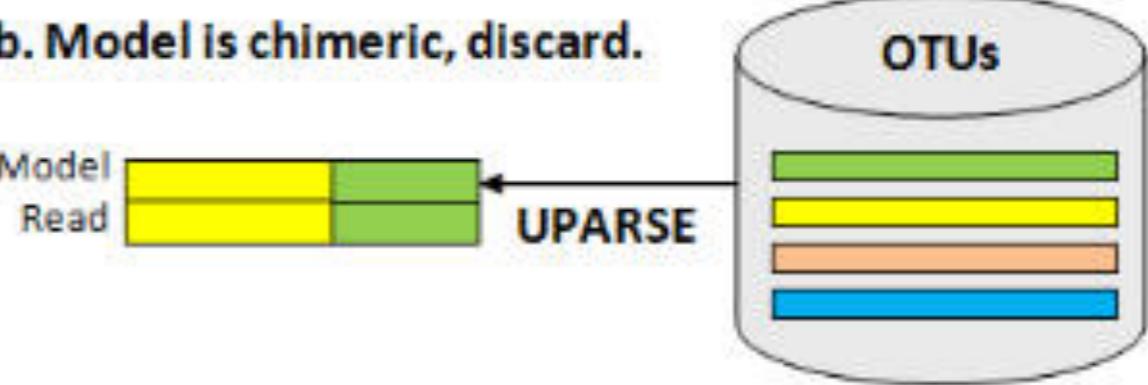
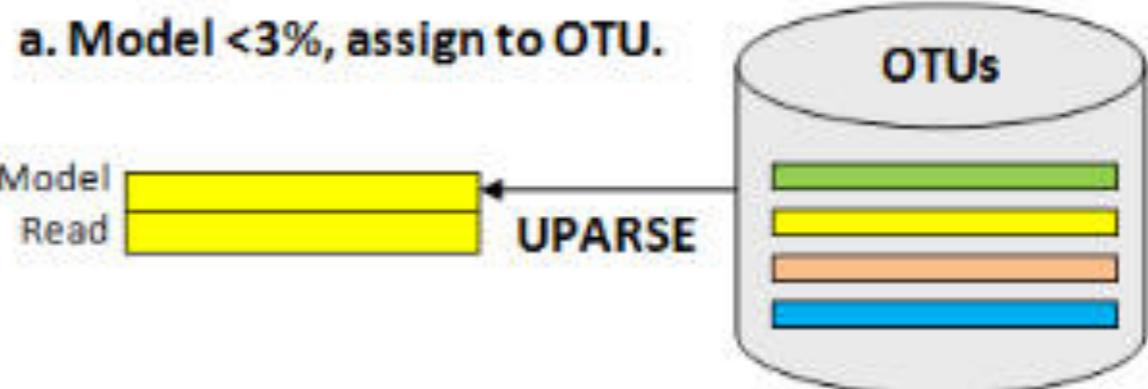
Ecological Significance of Microdiversity: Identical 16S rRNA Gene Sequences Can Be Found in Bacteria with Highly Divergent Genomes and Ecophysiolgies

Elke Jaspers,[†] and Jörg Overmann,[†]

PMCID: PMC492463

Operational Taxonomic Units

- Unidades taxonómicas operacionales
- ¿Por qué no especies?
- Mismo 16S, diferente genoma
- Misma especie, diferente 16S
- OTUs son clusters de secuencias que divergen como máximo a X% de identidad (3% normalmente)



Clin Microbiol Rev. 2004 Oct; 17(4): 840-862.
doi: 10.1128/CMR.17.4.840-862.2004

PMCID: PMC523561

MC492463

Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases

Jill E. Clamidge, Jill*

Author information ► Copyright and Licensee Information ►

Elke Jaspers¹ and Jörg Overmann

Gene

Operational Taxonomic Units

RStudio Source Editor

tax_table(physeq) ▾

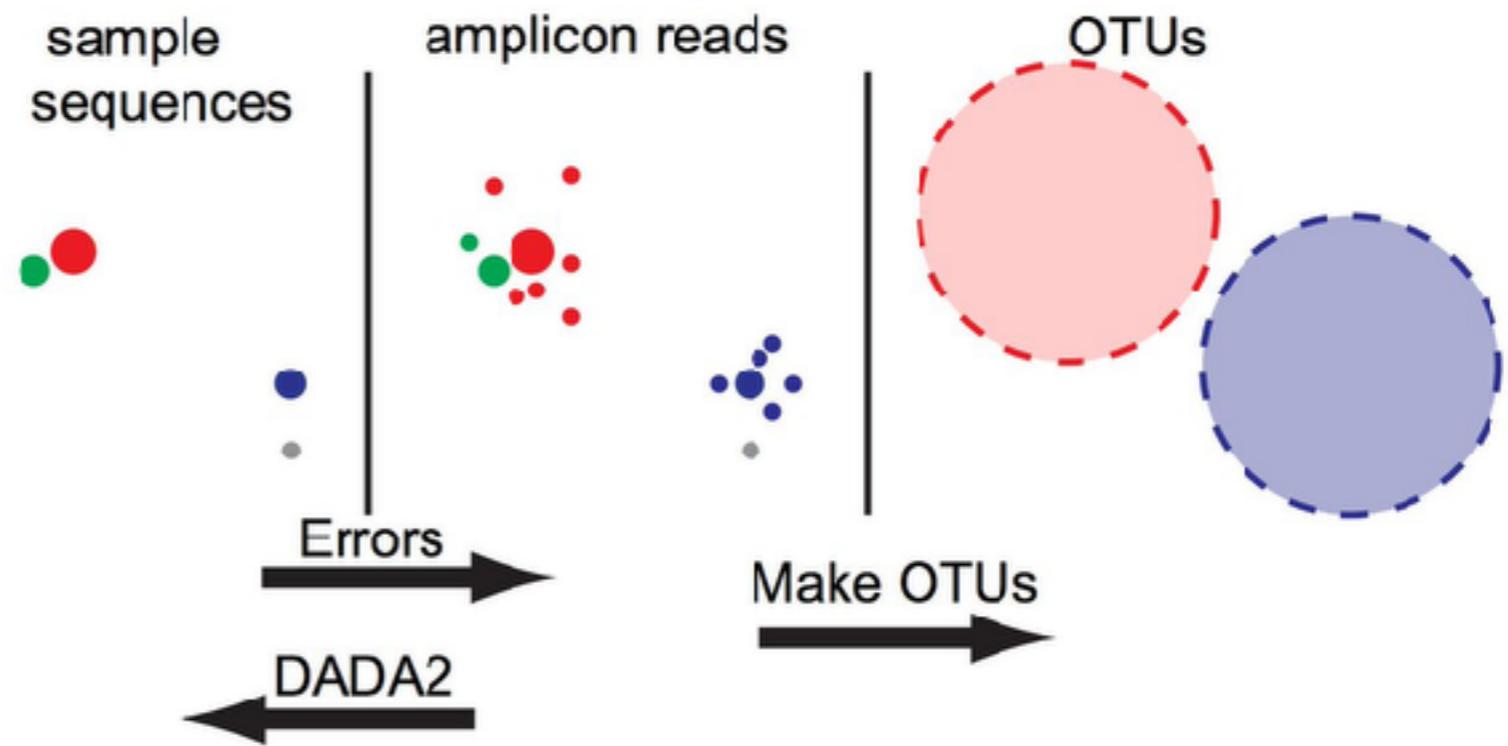
Filter

	Kingdom	Phylum	Class	Order	Family	Genus
Otu08098	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	NA
Otu17259	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiales_unclassified	NA
Otu07427	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	NA
Otu08947	Bacteria	Firmicutes	Firmicutes_unclassified	Firmicutes_unclassified	Firmicutes_unclassified	NA
Otu22896	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	NA
Otu28542	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiales_unclassified	NA
Otu06520	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	NA
Otu21167	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Hallaella
Otu21163	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	NA
Otu16680	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	Helicobacter
Otu05526	Bacteria	Chlamydiae	Chlamydiae	Chlamydiales	Chlamydaceae	Chlamydia
Otu24307	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiales_unclassified	NA
Otu19152	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Pseudoflavorifractor
Otu20183	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Marinibryantia
Otu20140	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidales_unclassified	NA
Otu30064	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	NA
Otu32335	Bacteria	Proteobacteria	Proteobacteria_unclassified	Proteobacteria_unclassified	Proteobacteria_unclassified	NA
Otu18835	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiales_unclassified	NA
Otu30638	Bacteria	Bacteria_unclassified	Bacteria_unclassified	Bacteria_unclassified	Bacteria_unclassified	NA
Otu18095	Bacteria	Bacteroidetes	Bacteroidetes_unclassified	Bacteroidetes_unclassified	Bacteroidetes_unclassified	NA
Otu20049	Bacteria	Bacteroidetes	Bacteroidetes_unclassified	Bacteroidetes_unclassified	Bacteroidetes_unclassified	NA
Otu30558	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	NA
Otu31046	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	NA
Otu32321	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Clostridium_XIVb
Otu30917	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	NA

Showing 1 to 25 of 32,361 entries

Alternativa - Amplicon Sequence Variants

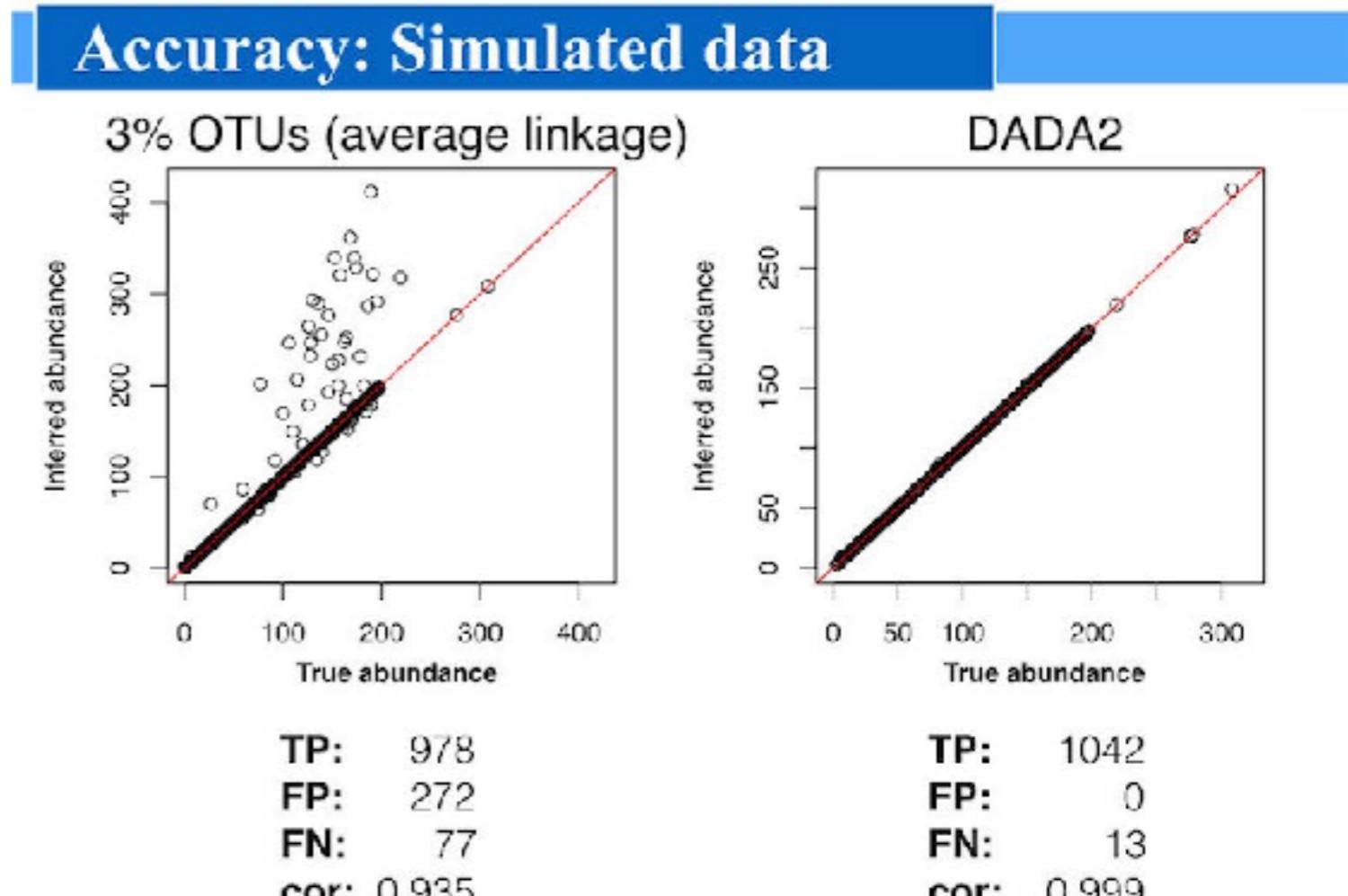
- Mayor resolución taxonómica
- ASVs son consistentes



Callahan, et al. Nature Methods, 2016.

Alternativa - Amplicon Sequence Variants

- Predicción de abundancia relativa en dada2 es más exacta que con OTUs



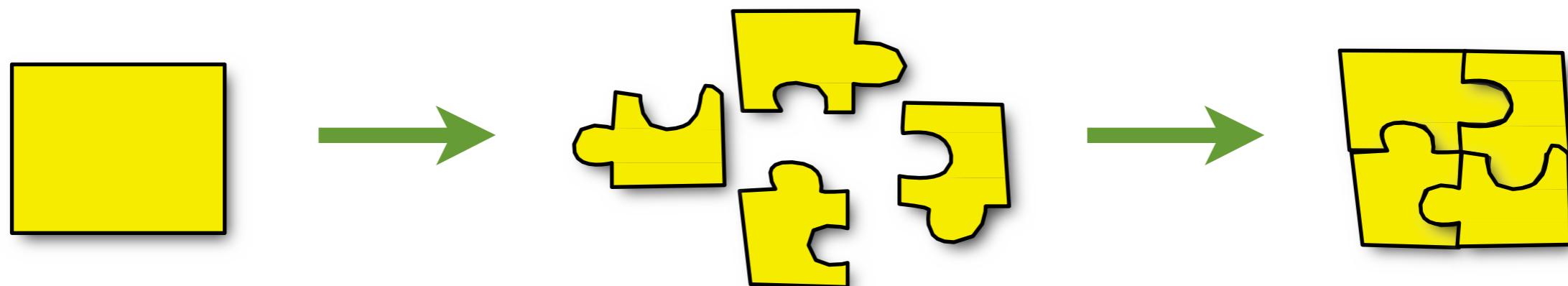
Data: Kopylova, et al. mSystems, 2016.

Recreo de 5 minutos

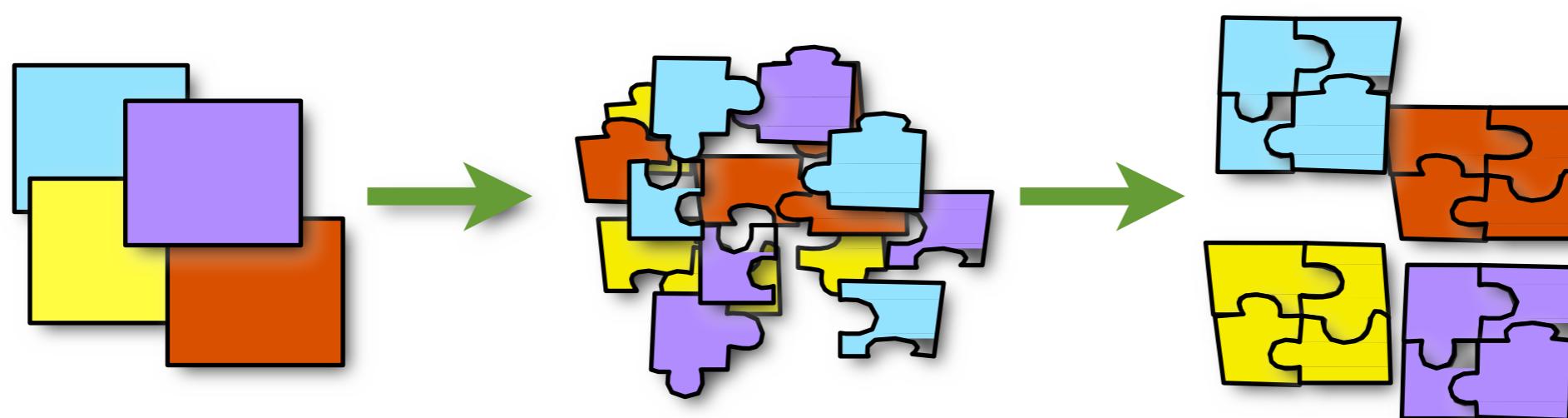
16S rRNA (amplicones) ≠ Metagenómica

Metagenómica

Genómica: Estudio de el genoma de un organismo



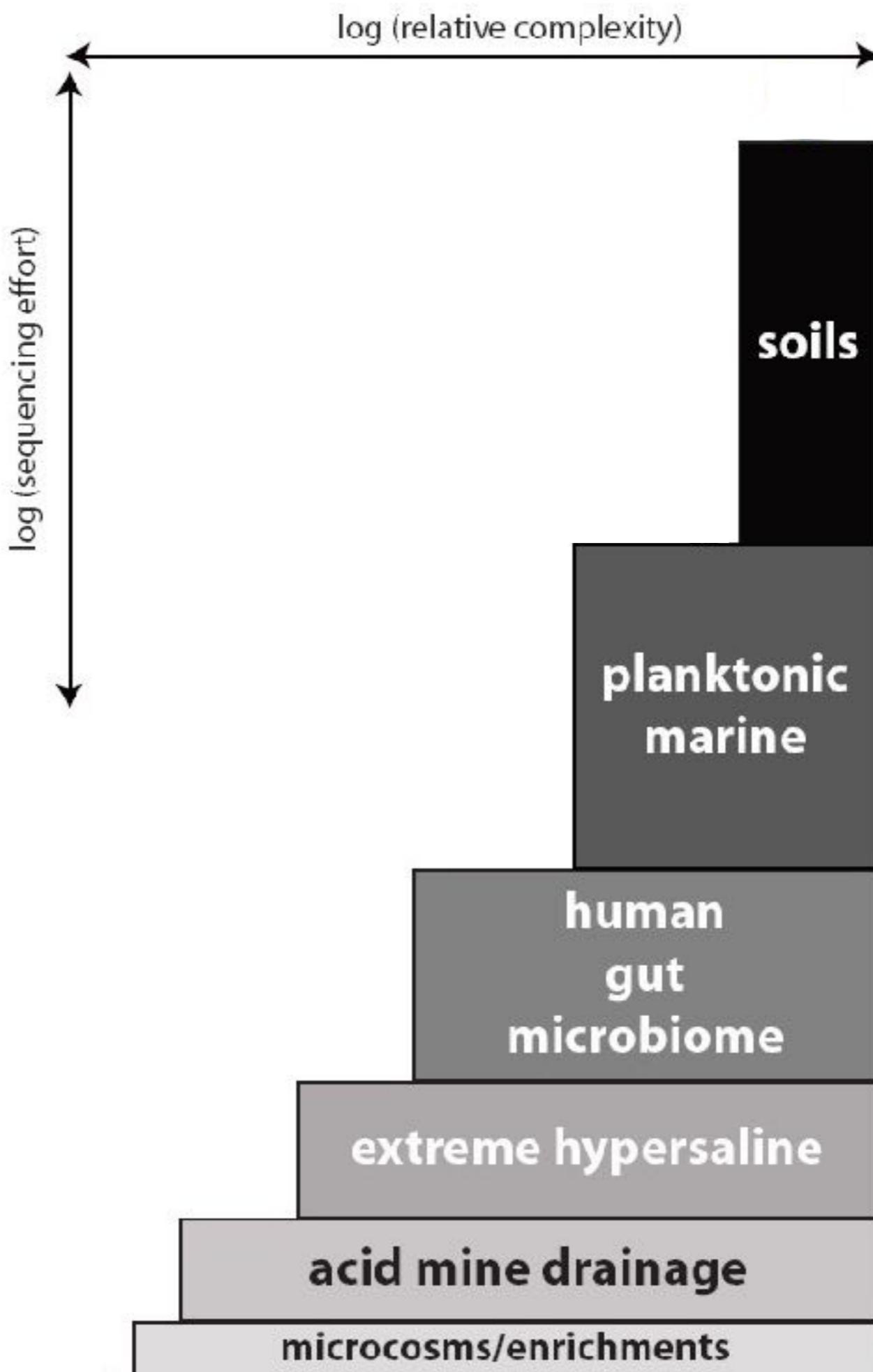
Metagenómica: Estudio de comunidades microbianas, analizando los genomas combinados de los organismos presentes en la comunidad



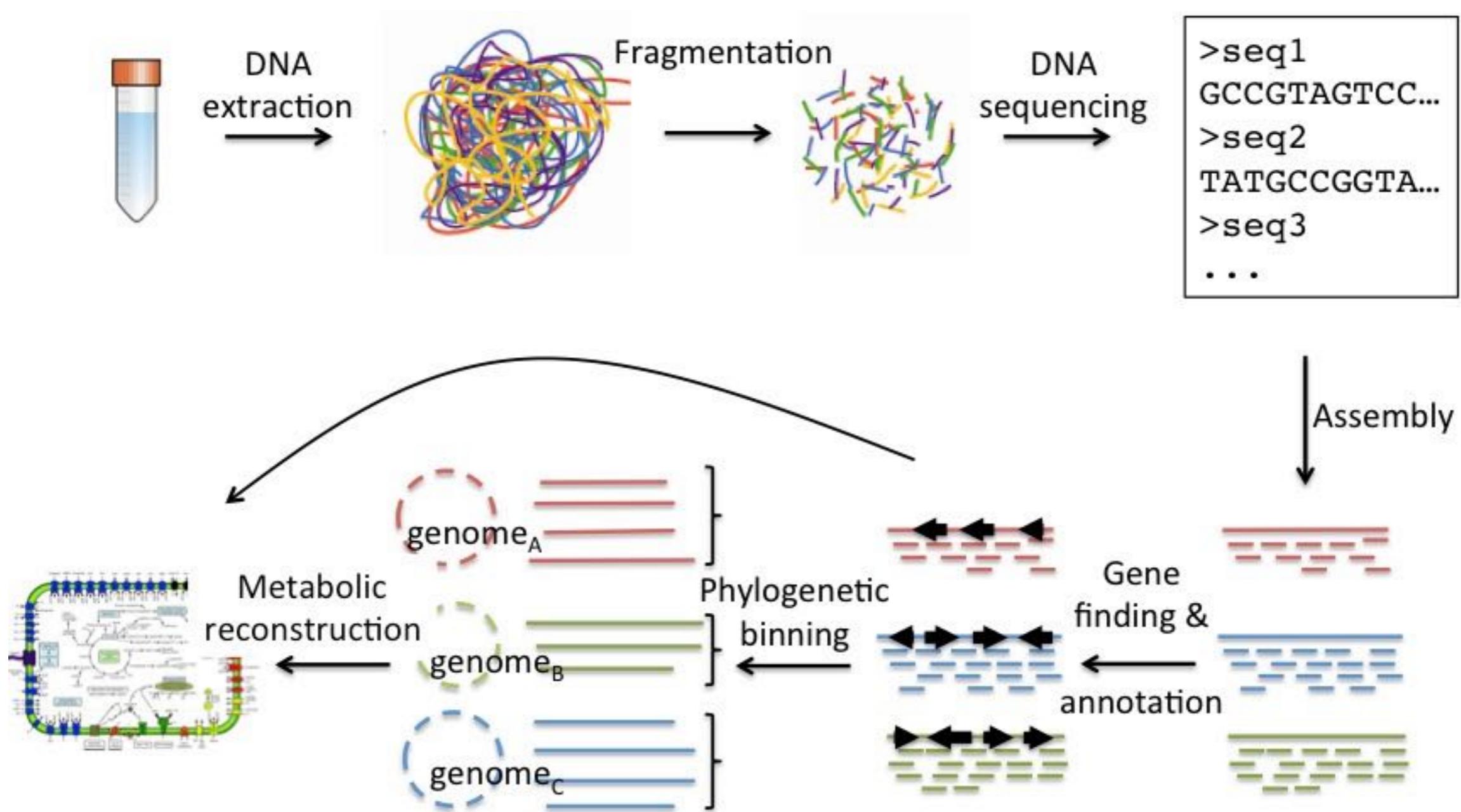
T
H
E

M
E
T
A
G
E
N
O
M
I
C

C
L
I
M
B



Shotgun Sequencing



Dos estrategias de análisis

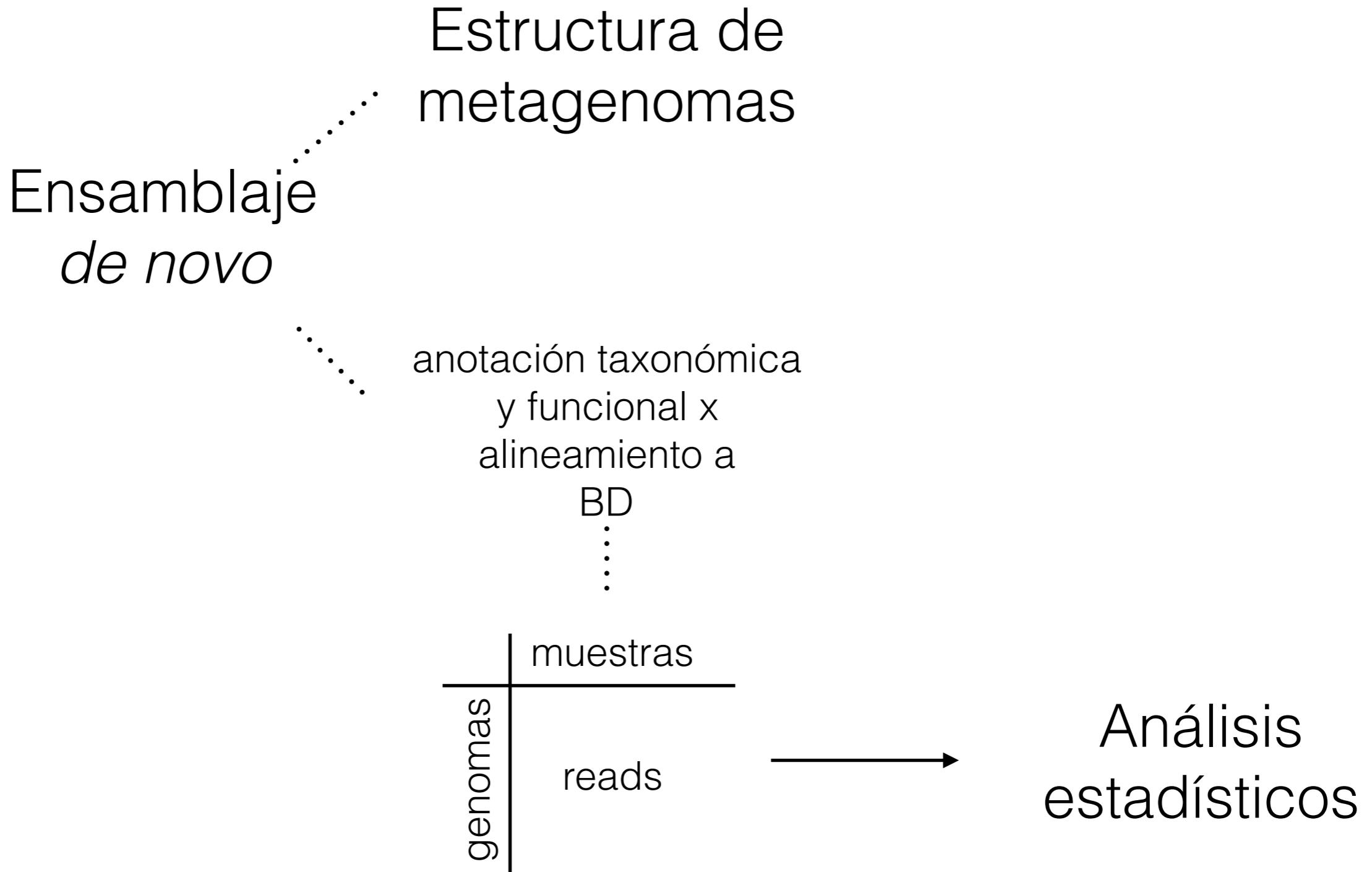
- Basado en ensamblaje *de novo*
- Basado en mapeo en contra de referencias

Estrategias analíticas para datos de metagenomas

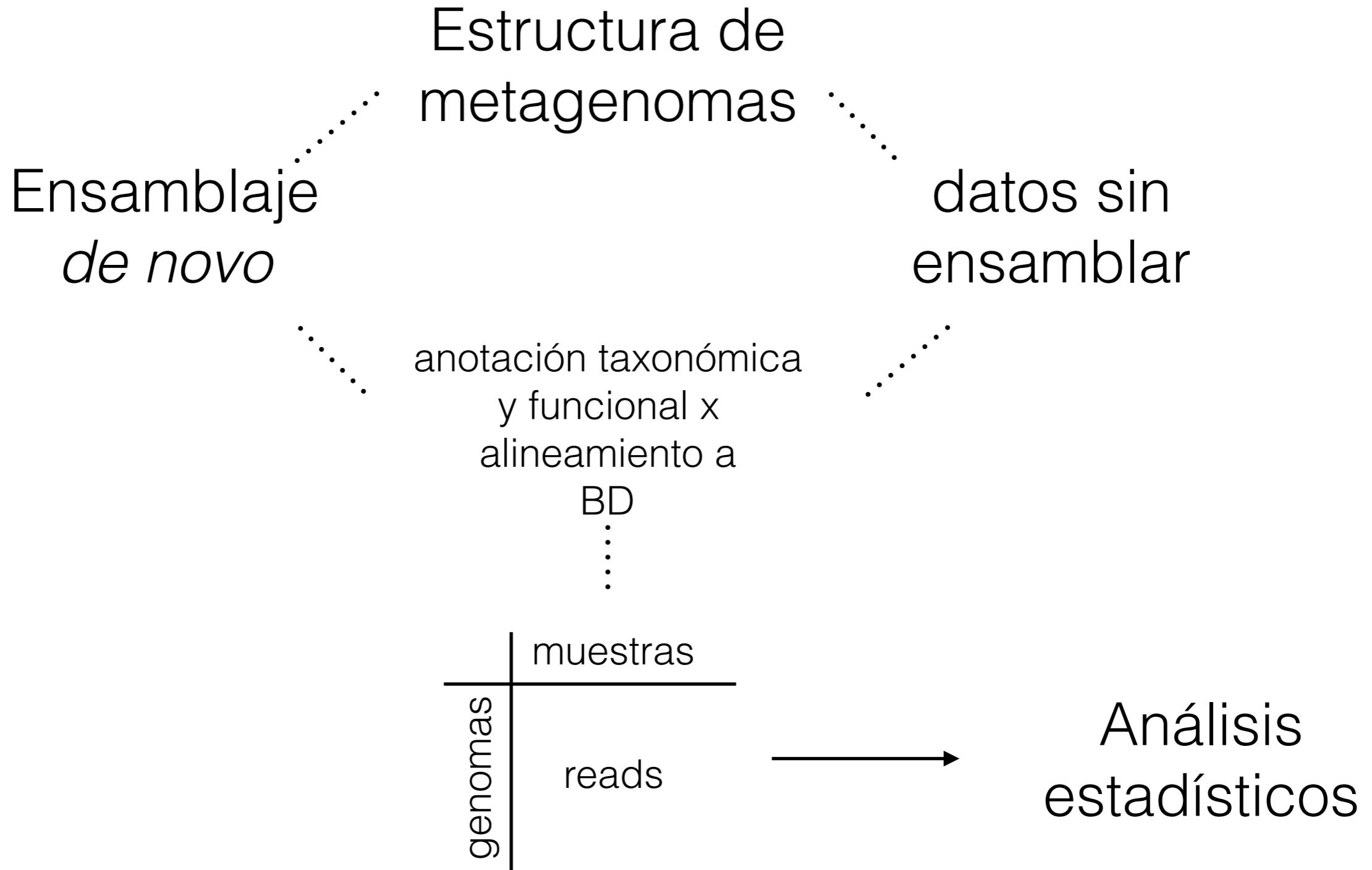
Estrategias analíticas para datos de metagenomas

Estructura de
metagenomas

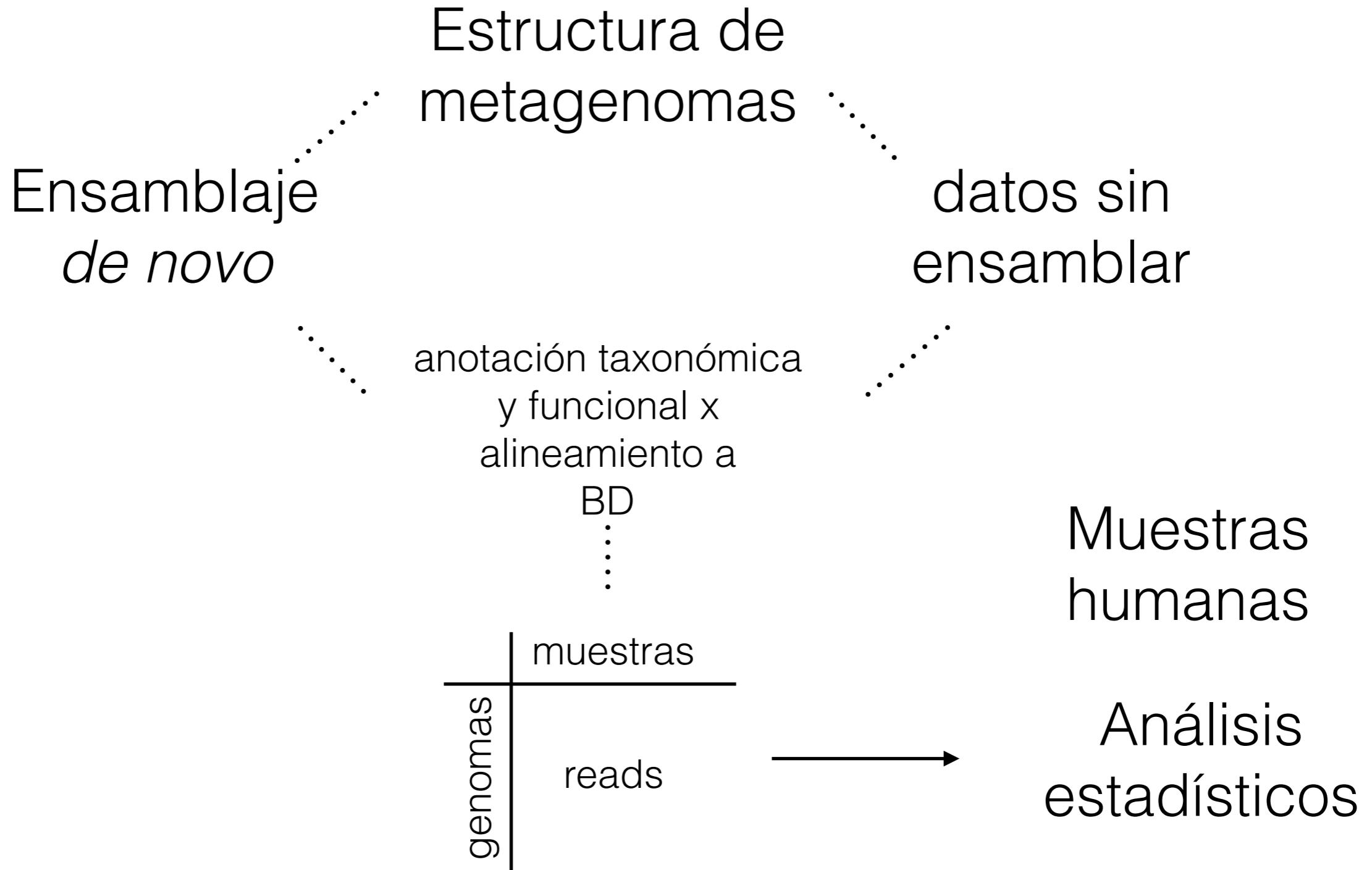
Estrategias analíticas para datos de metagenomas



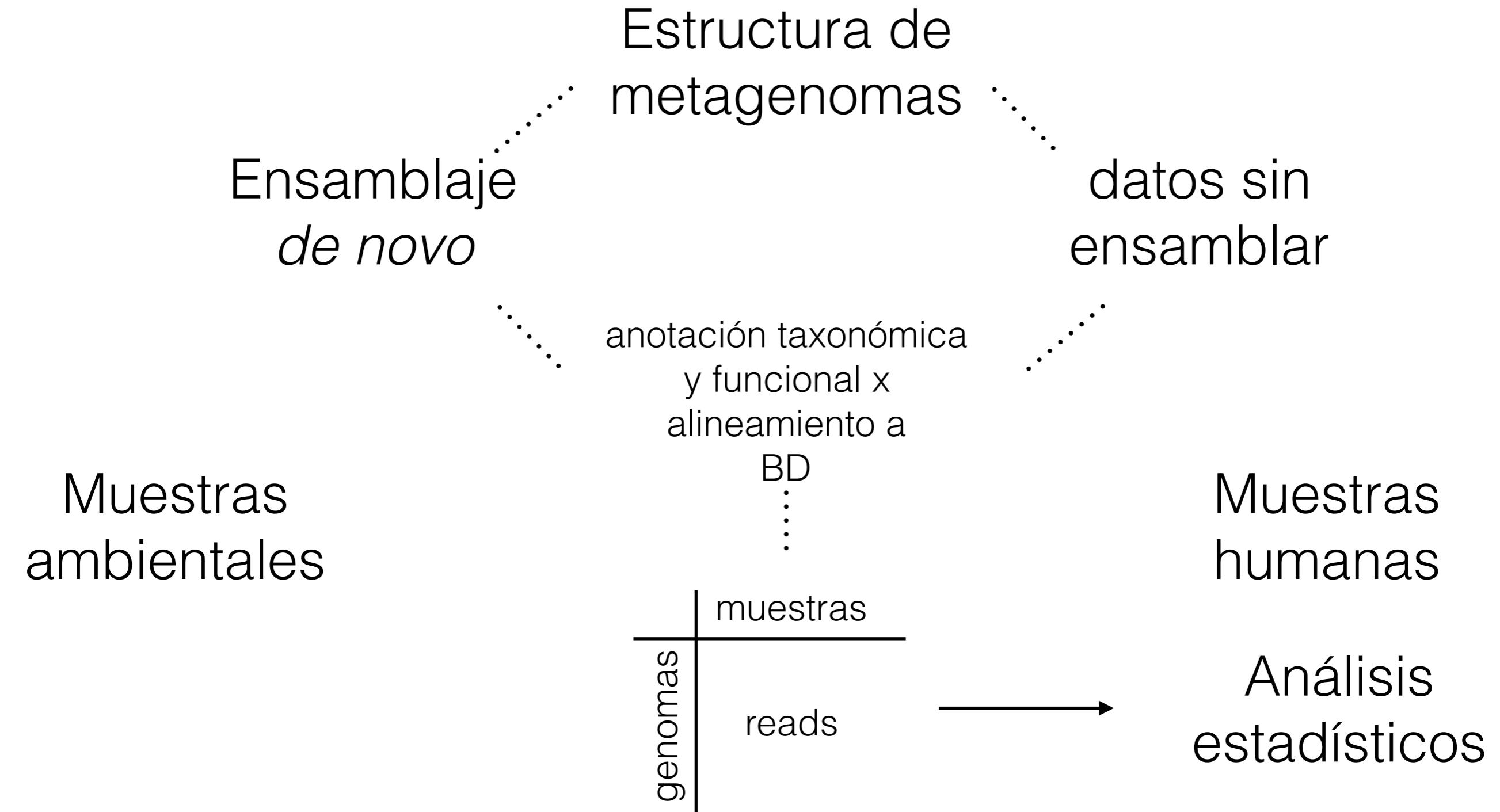
Estrategias analíticas para datos de metagenomas



Estrategias analíticas para datos de metagenomas



Estrategias analíticas para datos de metagenomas



Metagenómica

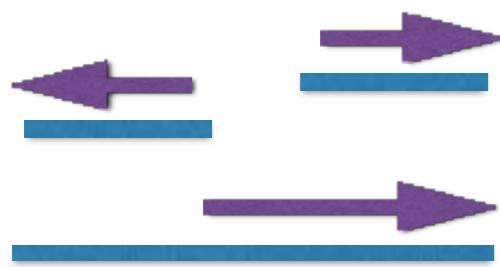
- Secuenciar todo el DNA (cromosomal, plasmidial, etc.)
- Generar un perfil de miembros del metagenóma
- Qué hay y en qué proporción

Metatranscriptómica

- Secuenciar todo el RNA
- Generar un perfil de expresión de genes en la comunidad microbiana
- Qué genes se expresan y en qué medida

Two approaches in metagenomic studies

Read-based

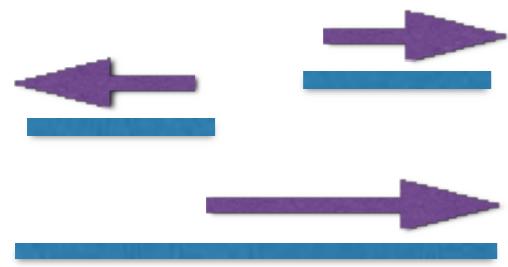


- Taxonomic and functional community profile
- Allows quick comparative analysis
- Provides a broad picture of the community

Two approaches in metagenomic studies

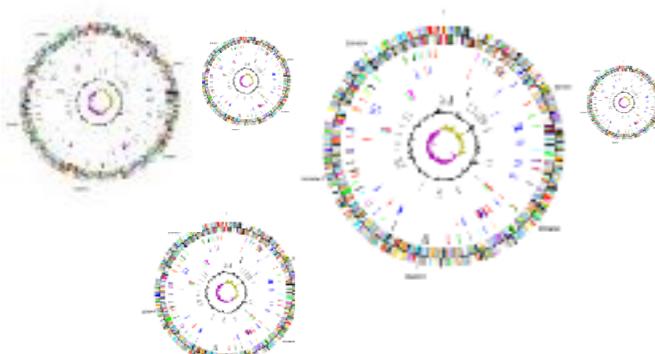
Read-based

- Taxonomic and functional community profile
- Allows quick comparative analysis
- Provides a broad picture of the community

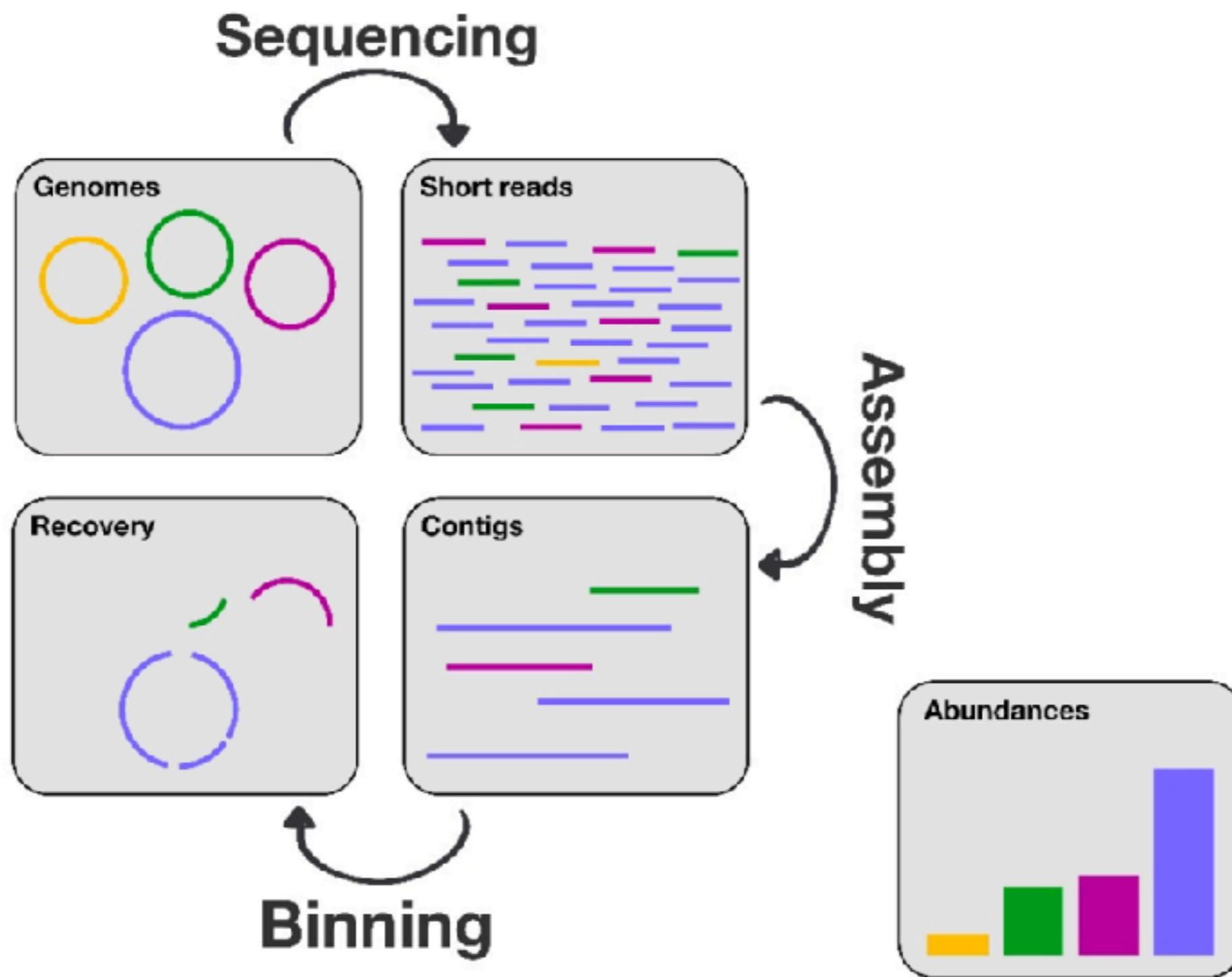


Assembly-based

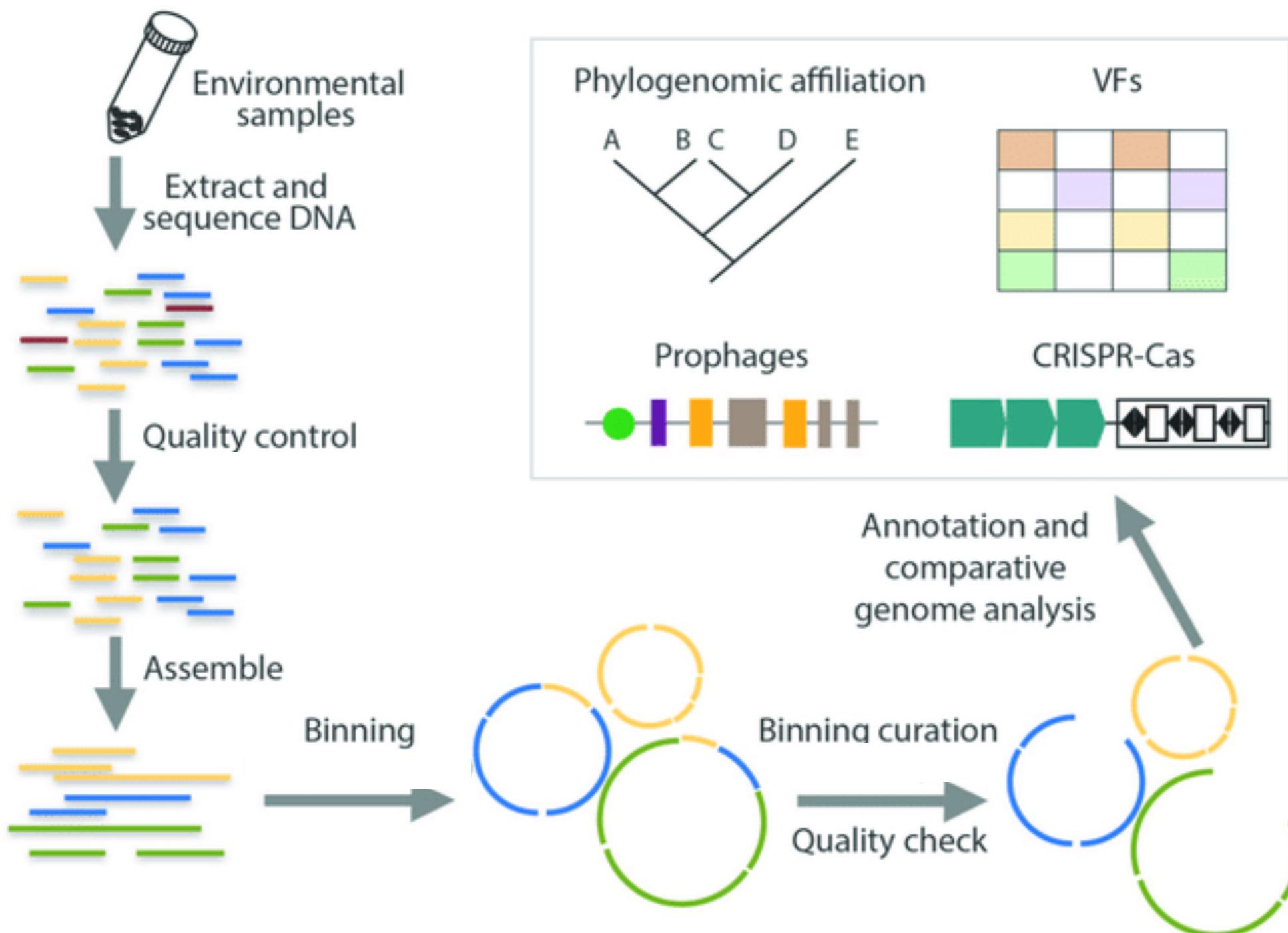
- Improved taxonomic classification
- Association between taxonomy and function
- Novel genes and novel taxa



Genome-resolved metagenomics through binning



Genome-resolved metagenomics through binning



Binning

- Agrupar secuencias de acuerdo a características comunes
- % GC, frecuencia de di-, tri-, o tetra-nucleótidos
- Cobertura diferencial

