



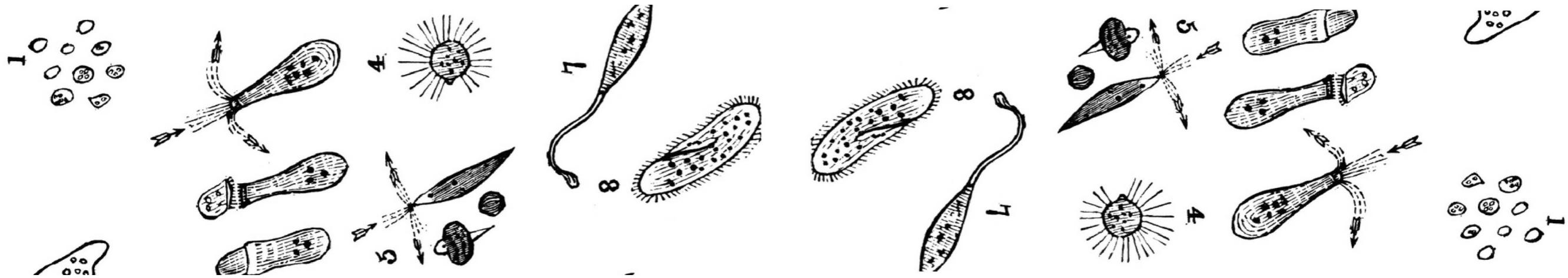
16S rRNA ventajas, limitaciones y análisis

www.castrolab.org

www.cbib.cl

www.ucdavischile.org

Eduardo Castro, PhD
Universidad Andrés Bello
16 de marzo de 2018



Diversidad microbiana

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



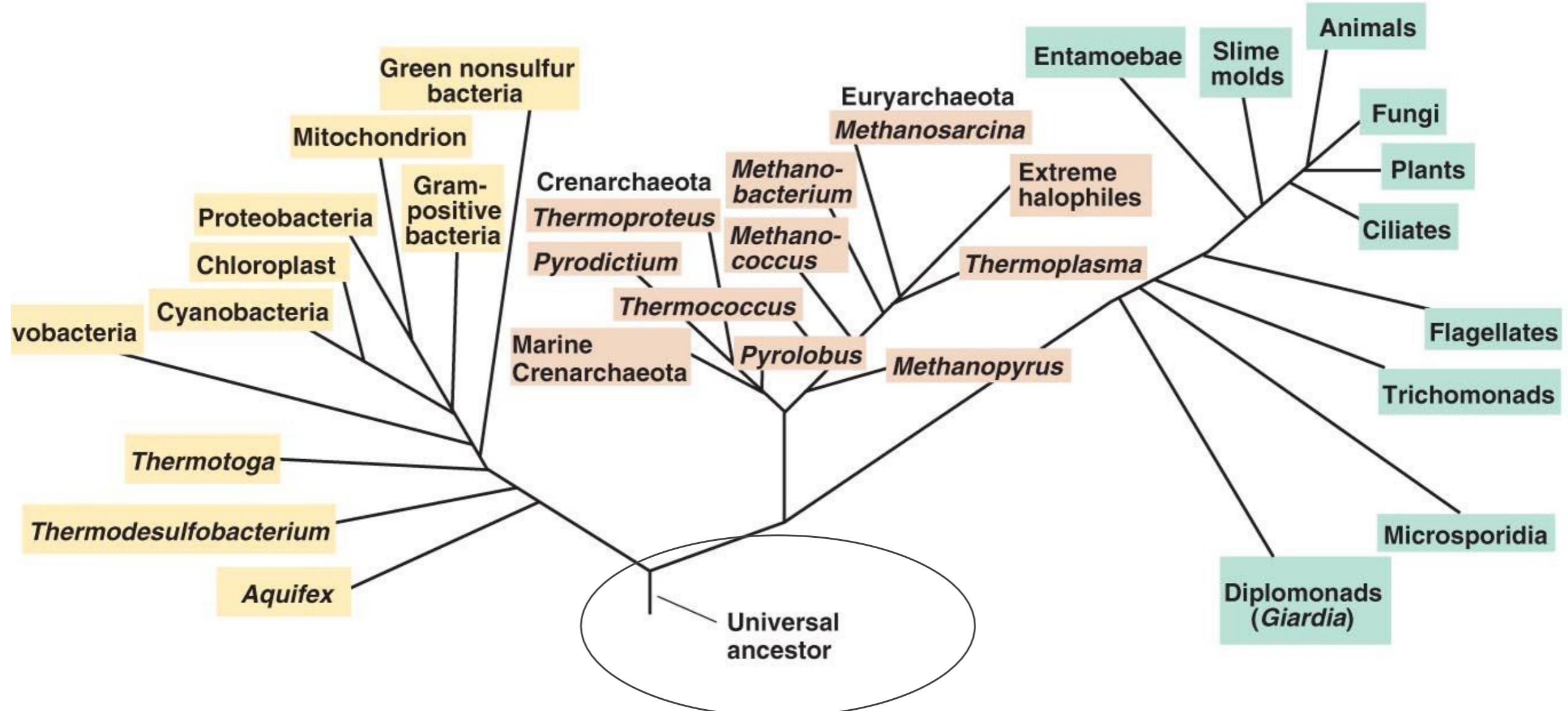
PROKARYOTES

Bacteria

Archaea

EUKARYOTES

Eukarya



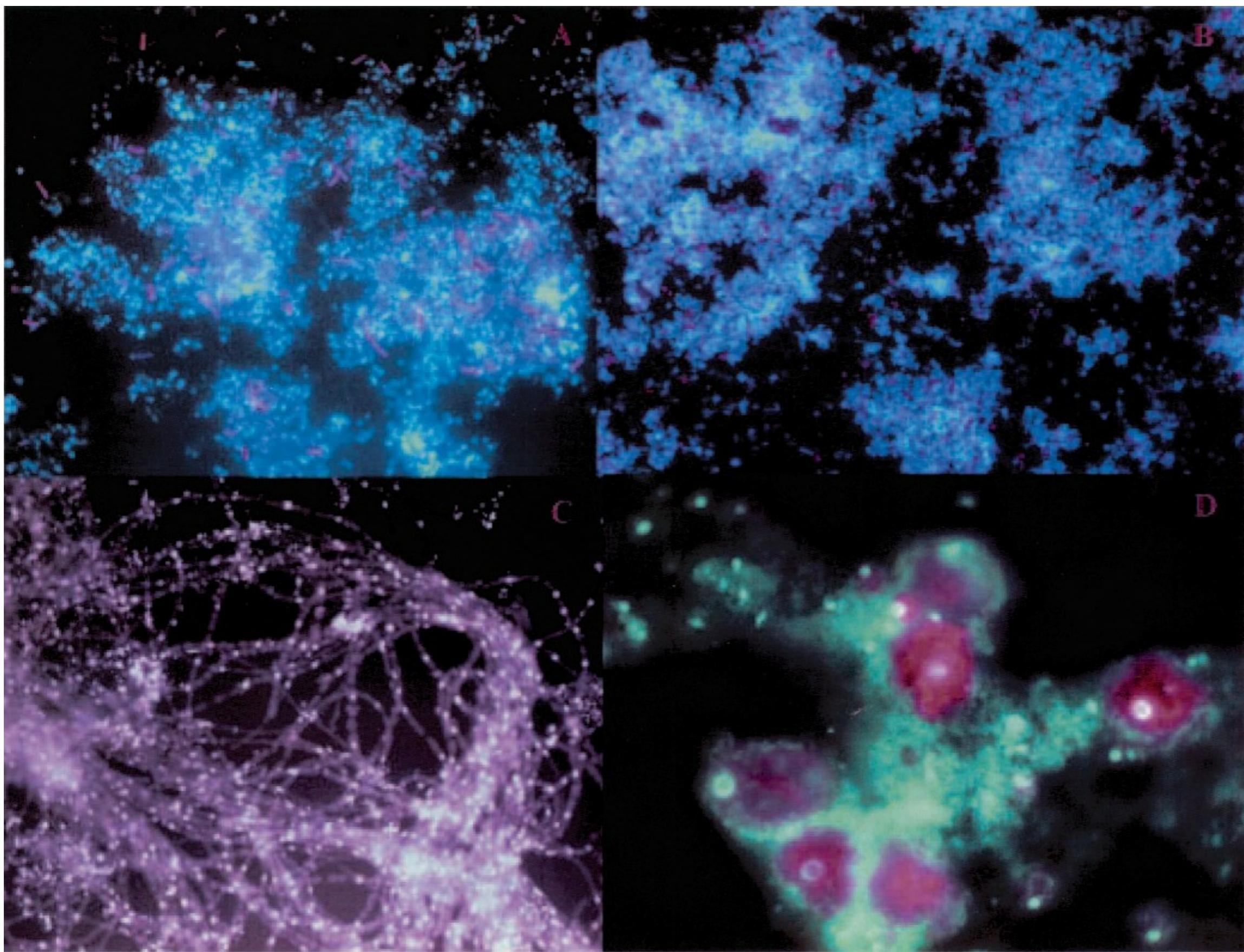
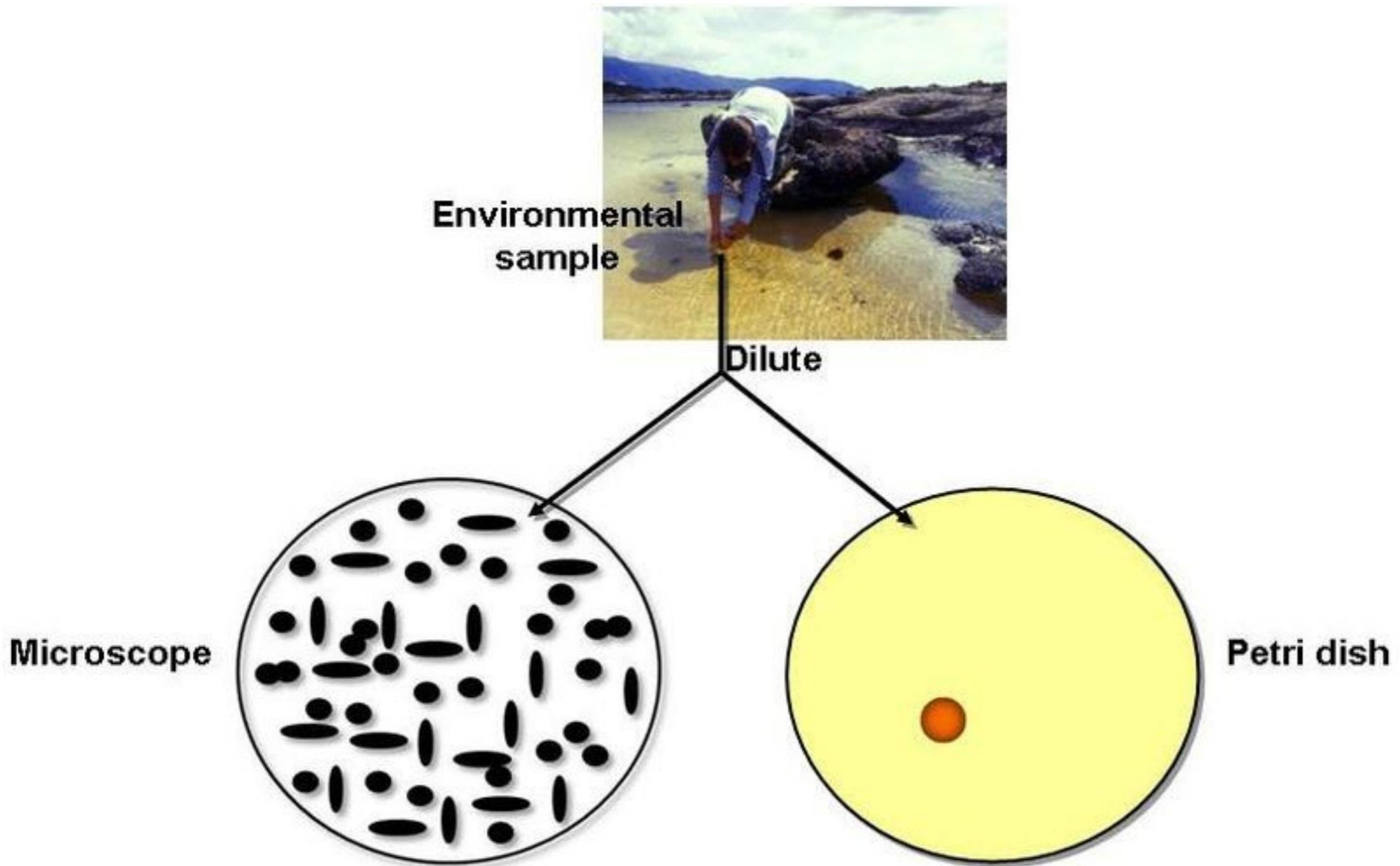


FIG. 2. Phylogenetic stains. Fluorescent in situ hybridization biofilm samples from Iron Mountain Mine, Calif. Nucleic acid probes were labeled with indodicarbocyanine, and DNA was stained nonspecifically with 4',6'-diamidino-2-phenylindole. The nucleic acid probes are specific for (top left) *Sulfobacillus* spp., (top right) Archaea, (bottom left) Archaea on fungal filaments, and (bottom right) Eukarya. Reproduced from reference 5 with permission of the publisher.

Un gran porcentaje de los microorganismos no son fáciles de cultivar



“THE GREAT PLATE COUNT ANOMALY”

~100 times > cells than colonies, 99% unculturable

<http://schaechter.asmblog.org/schaechter/2010/07/the-uncultured-bacteria.html>

...the “Environmental Genomic” approach...



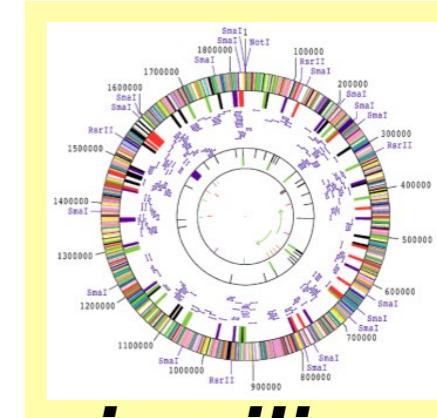
in situ

- Field collection
- Geochemistry
- Sample analysis
- Sample preservation



in vitro

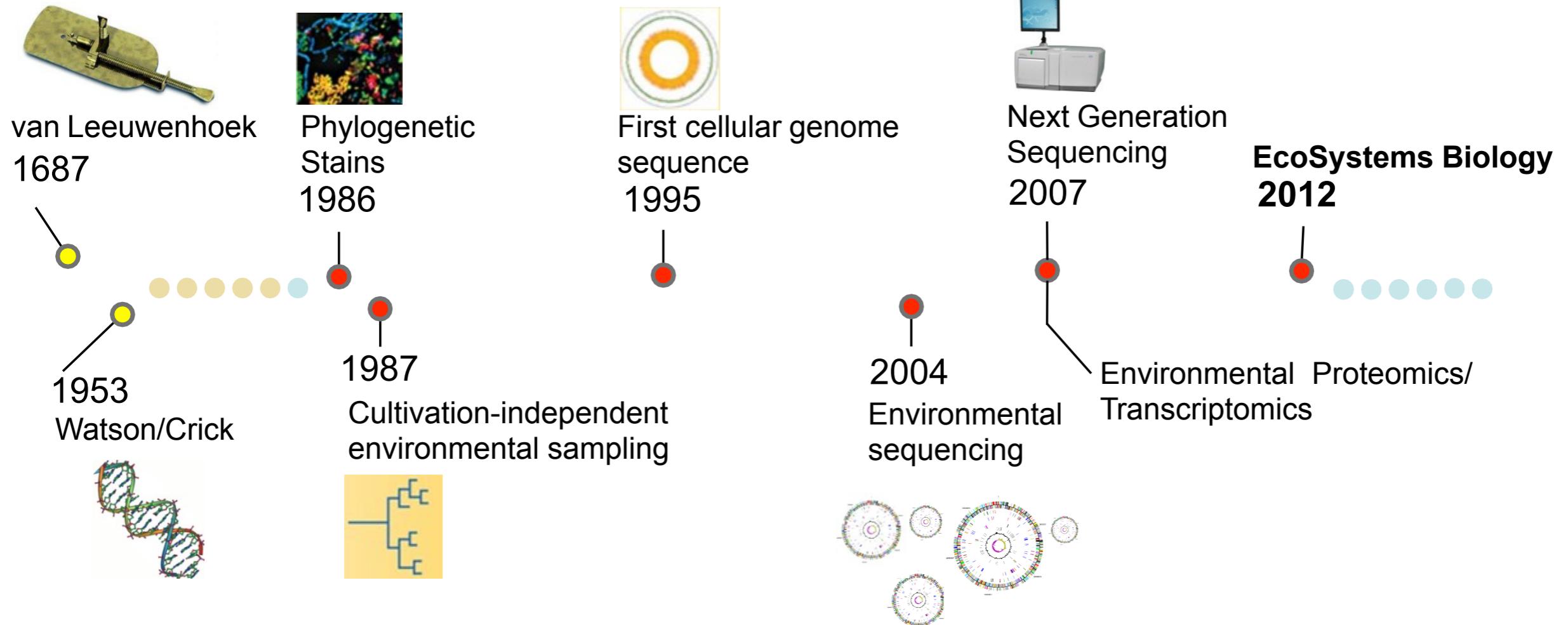
- DNA extraction
- PCR amplification
- Library construction
- RNA extraction
- cDNA synthesis
- DNA sequencing
- Protein extraction
- LC/MS-MS analysis



in silico

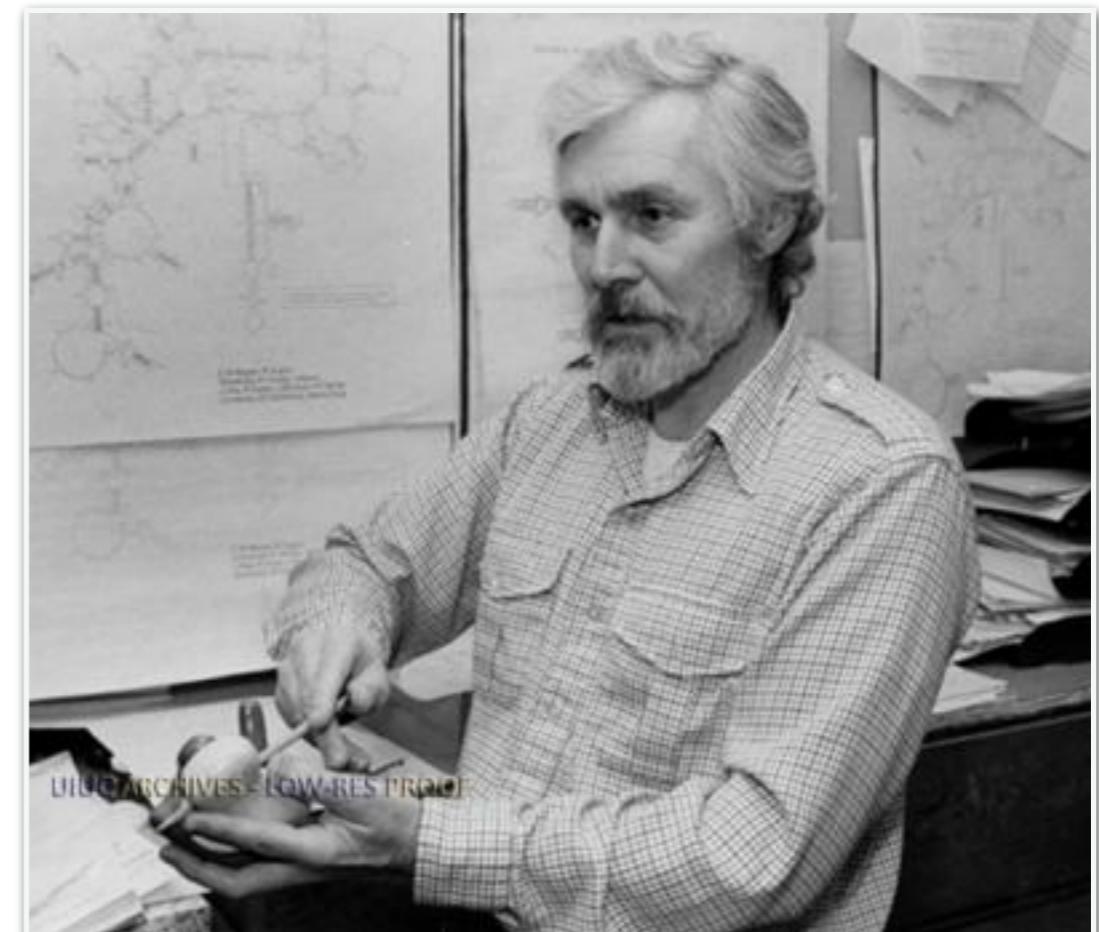
- Sequence processing
- Quality control filtering
- Sequence assembly
- Phylogenetic binning
- Functional annotation
- Comparative genomics
- Metabolism/physiology
- mRNA/protein expression

The\$Rise\$of\$‘Cul.va.on2independent’\$ Molecular\$Microbial\$Ecology\$and\$Genomics\$



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

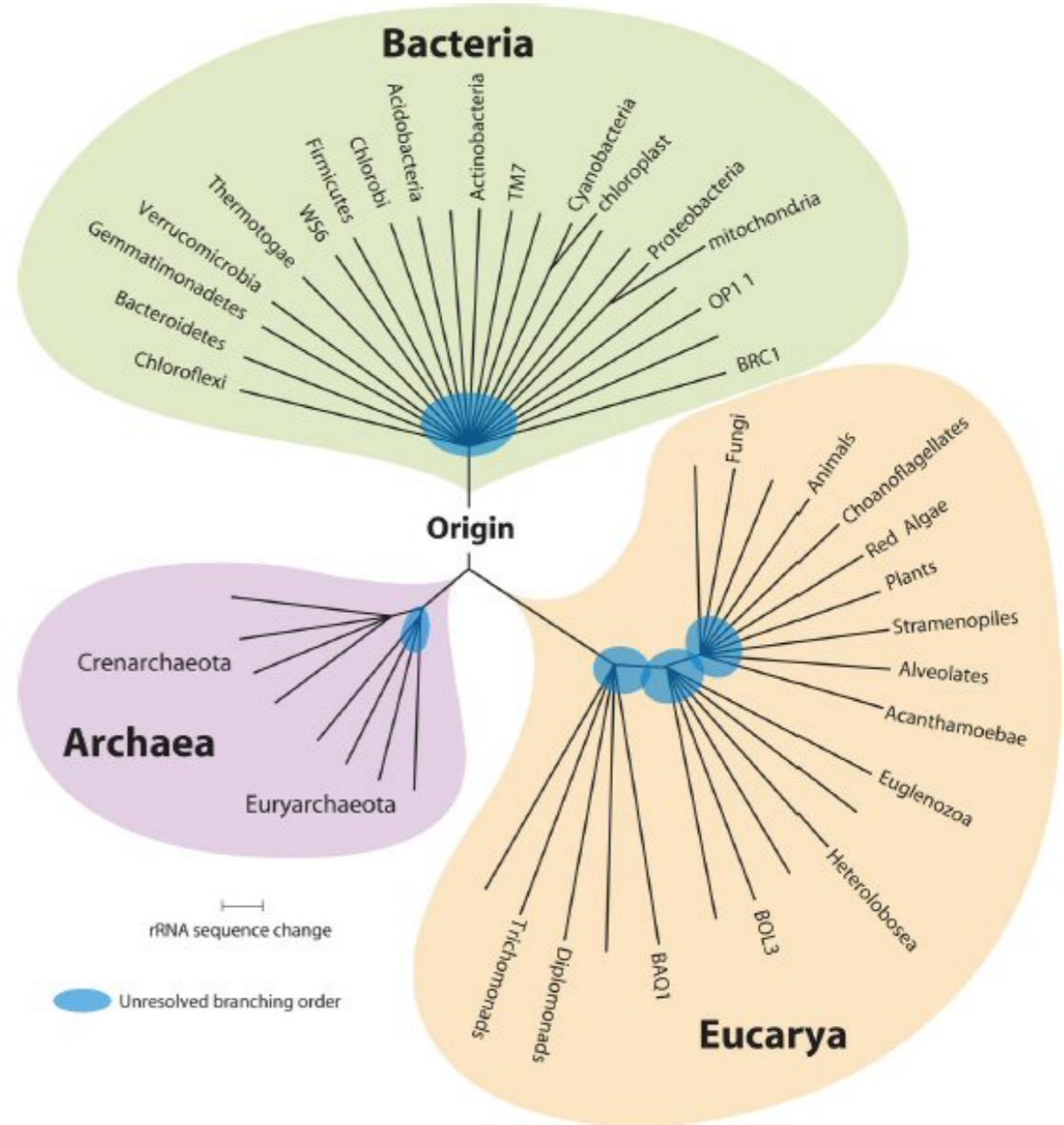
- En 1977 a partir de filogenias de 16/18S
- Arqueas eran consideradas parte de Bacterias

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

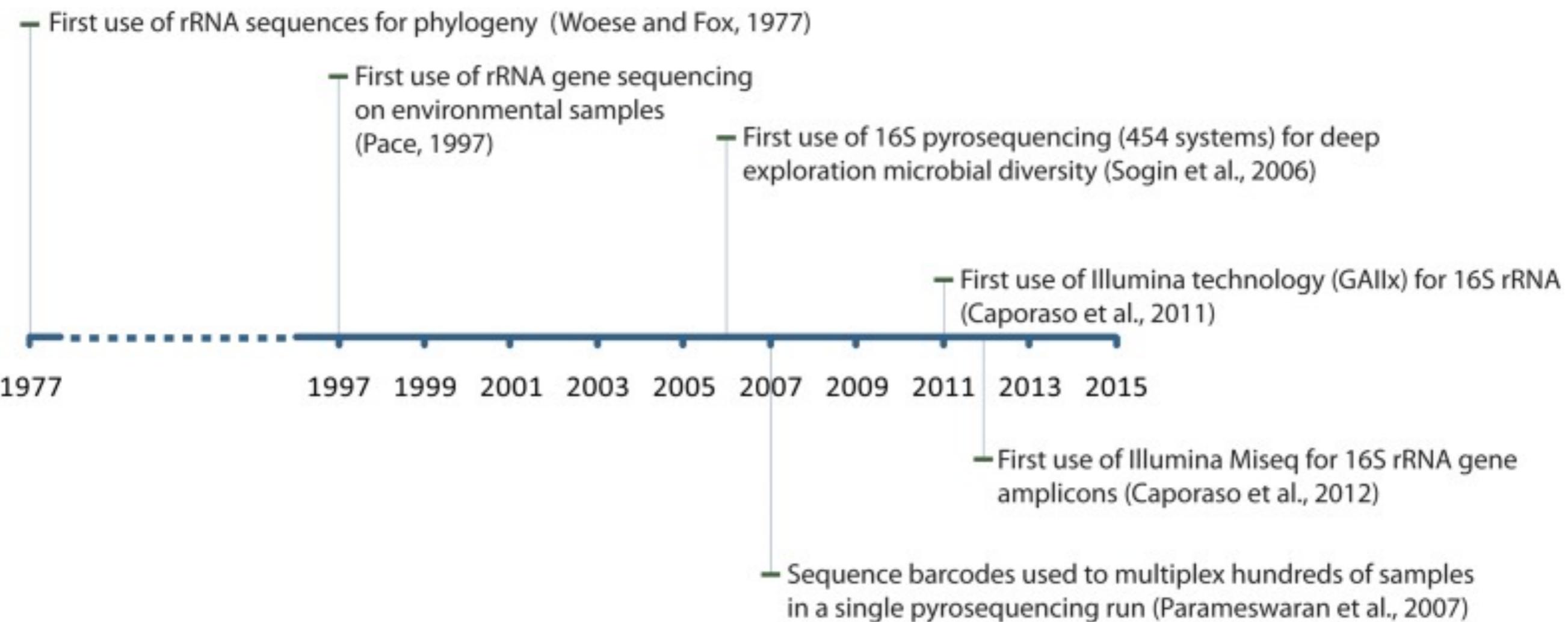
Carl R. Woese and George E. Fox*

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 urkaryotes, now represented in the cytoplasmic component of eukaryotic cells.

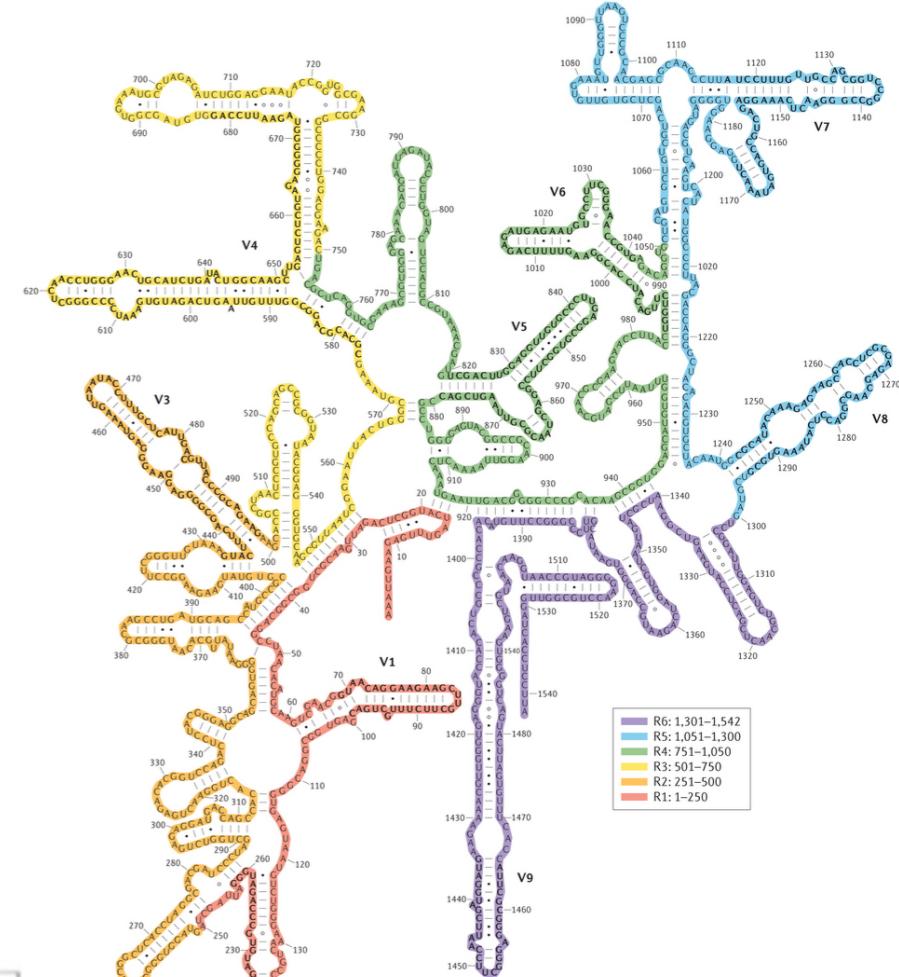
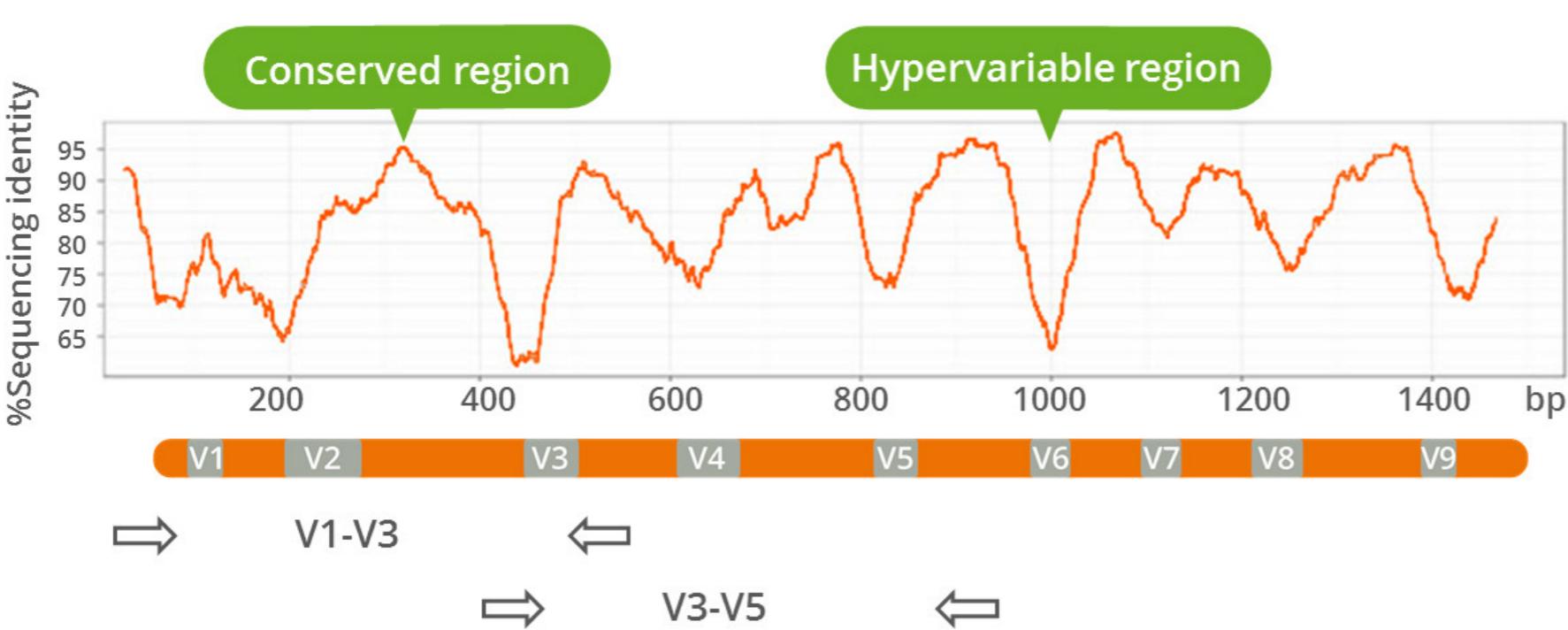


16S rRNA



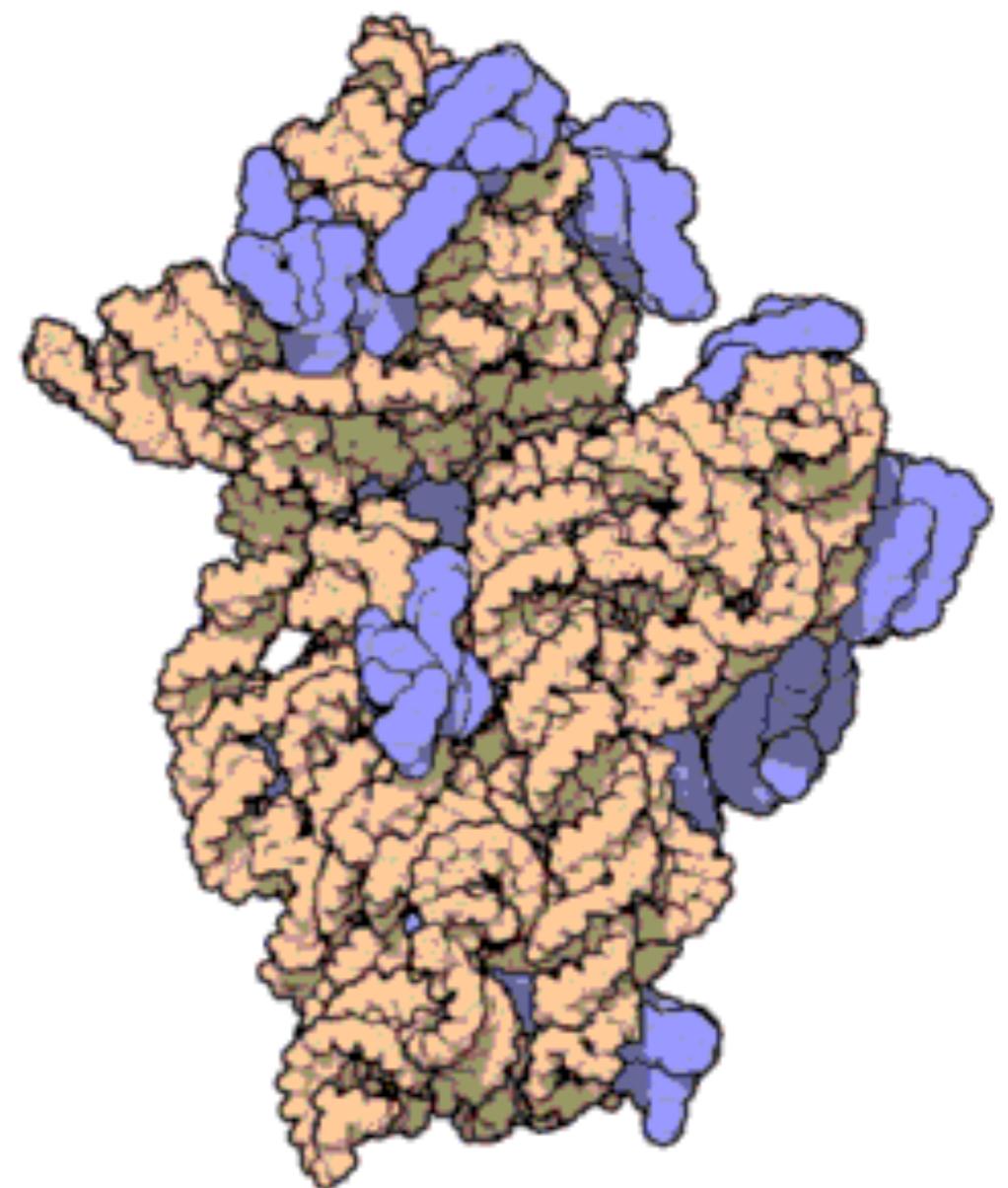
Estructura y ventajas del 16S

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



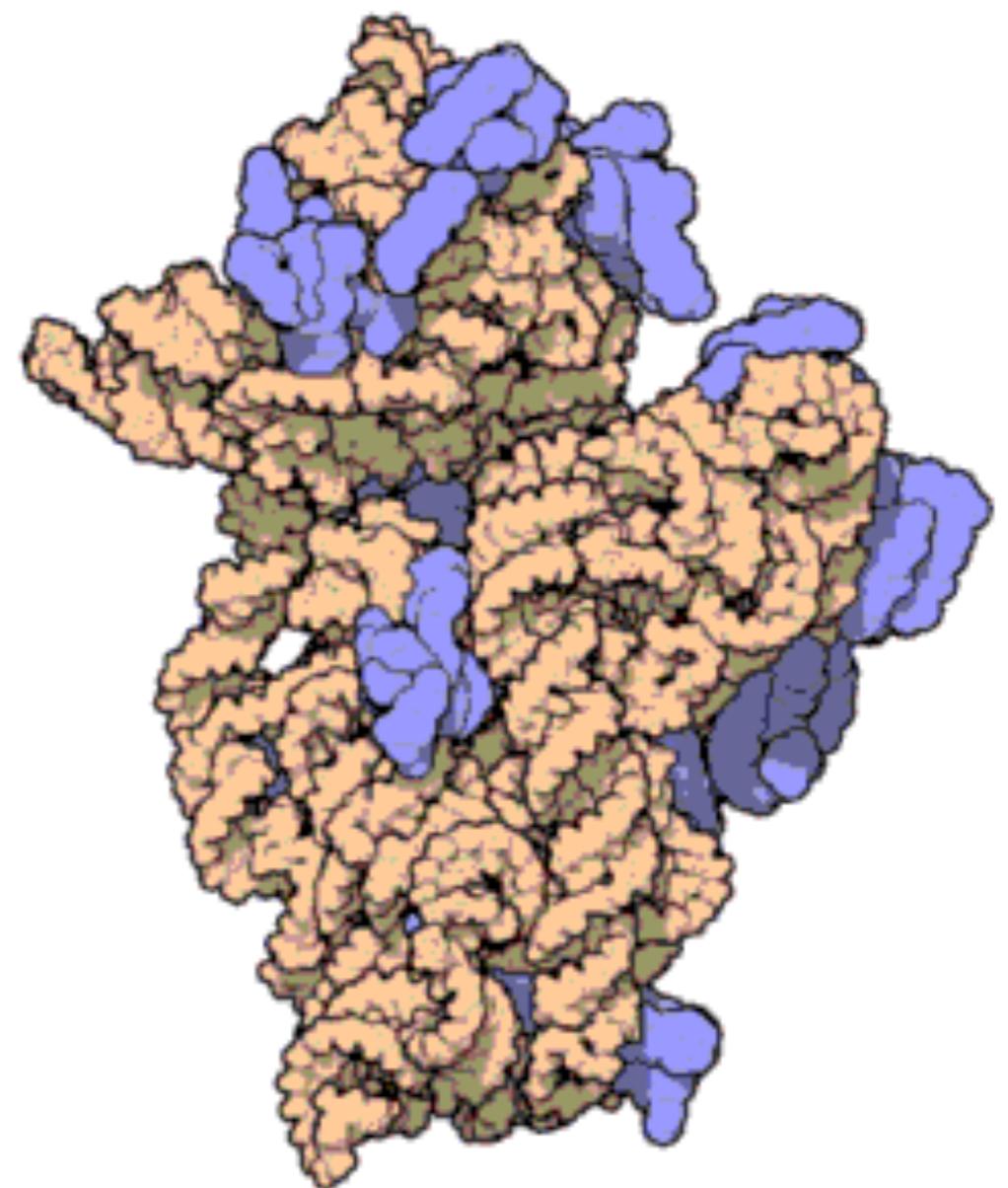
Estructura del 16S

- 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)



Flujo de trabajo típico



- Maximizar el número de muestras biológicas
- Almacenamiento, extracción (bead beating), cantidad de material de entrada

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 ▾

Flujo de trabajo típico - almacenamiento de muestras

- 72 h, diversidad alfa y beta cambian con respecto a muestra control
- Algunos autores reportan efectos poco significativos

Article | OPEN

Sample storage conditions significantly influence faecal microbiome profiles

Jocelyn M Choo, Lex EX Leong & Geraint B Rogers ✉

Article | OPEN

Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data

Ran Blekhman ✉, Karen Tang, Elizabeth A. Archie, Luis B. Barreiro, Zachary P. Johnson, Mark E. Wilson, Jordan Kohn, Michael L. Yuan, Laurence Gesquiere, Laura E. Grieneisen & Jenny Tung ✉

Flujo de trabajo típico - almacenamiento de muestras

- 8 semanas, 95% etanol, FTA, omnigene. Gut presentan diferencias comparables a réplicas técnicas
- T ambiente x 24 h y con RNALater, diferencias inconsistentes entre individuos. Congelar/descongelar lo mismo
- A 20, 4, -20, y -80 no se observan cambios en suelo

Research Article | Applied and Environmental Science

Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies

Se Jin Song, Amnon Amir, Jessica L. Metcalf, Katherine R. Amato, Zhenjiang Zech Xu, Greg Humphrey, Rob Knight

M. Denise Dearing, Editor

DOI: 10.1128/mSystems.00021-16

Storage conditions of intestinal microbiota matter in metagenomic analysis

Silvia Cardona, Anat Eck, Montserrat Cassellas, Milagros Gallart, Carmen Alatrue, Joel Dore, Fernando Azpiroz, Joaquim Roca, Francisco Guarner and Chaysavanh Manichanh 

BMC Microbiology 2012 12:158 | DOI: 10.1186/1471-2180-12-158 | © Cardona et al.; licensee BioMed Central Ltd. 2012

Received: 6 March 2012 | Accepted: 20 July 2012 | Published: 30 July 2012

RESEARCH LETTER

Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples

Christian L. Lauber¹, Nicholas Zhou², Jeffrey I. Gordon³, Rob Knight⁴ & Noah Fierer^{1,2}

¹Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA; ²Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA; ³Center for Genome Sciences, Washington University School of Medicine, St. Louis,

⁴Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA

Flujo de trabajo típico - 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

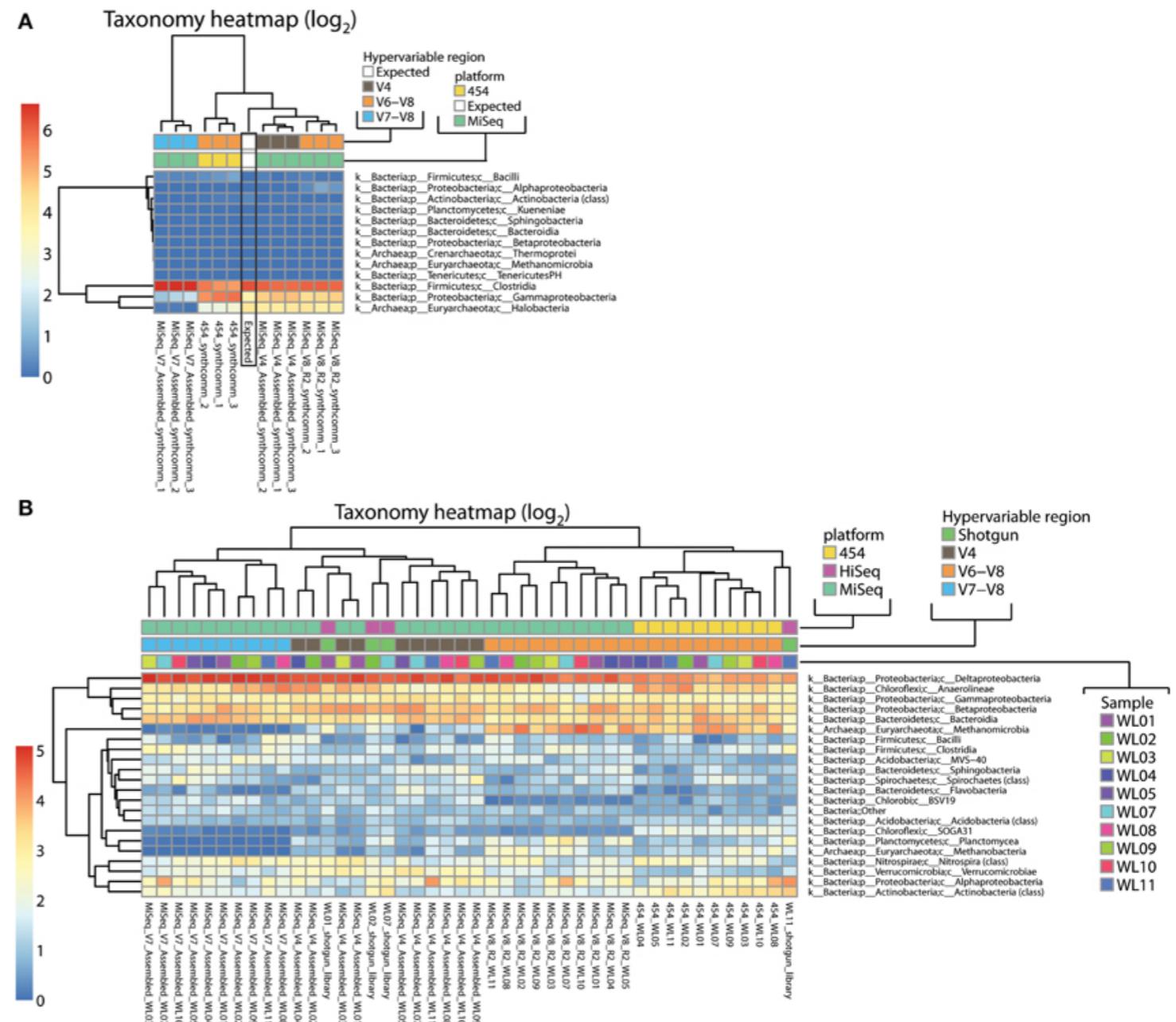
J Paul Brooks , David J Edwards, Michael D Harwich Jr, Maria C Rivera, Jennifer M Fettweis, Myrna G Serrano, Robert A Reris, Nihar U Sheth, Bernice Huang, Philippe Girerd, Vaginal Microbiome Consortium (additional members), Jerome F Strauss III, Kimberly K Jefferson and Gregory A Buck

BMC Microbiology 2015 15:66 | DOI: 10.1186/s12866-015-0351-6 | © Brooks et al.; licensee BioMed Central. 2015

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

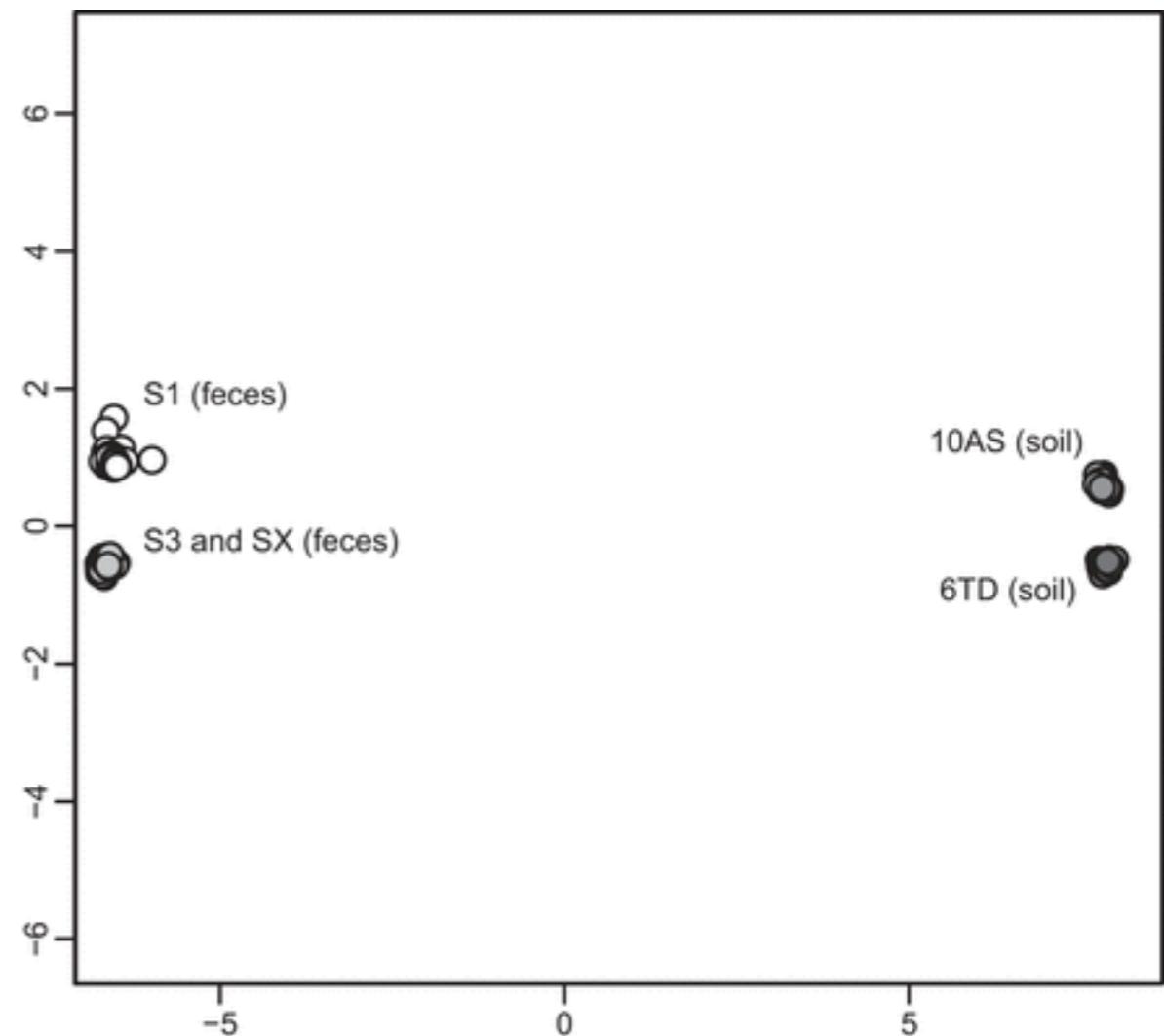
Flujo de trabajo típico - partidores de PCR

- 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



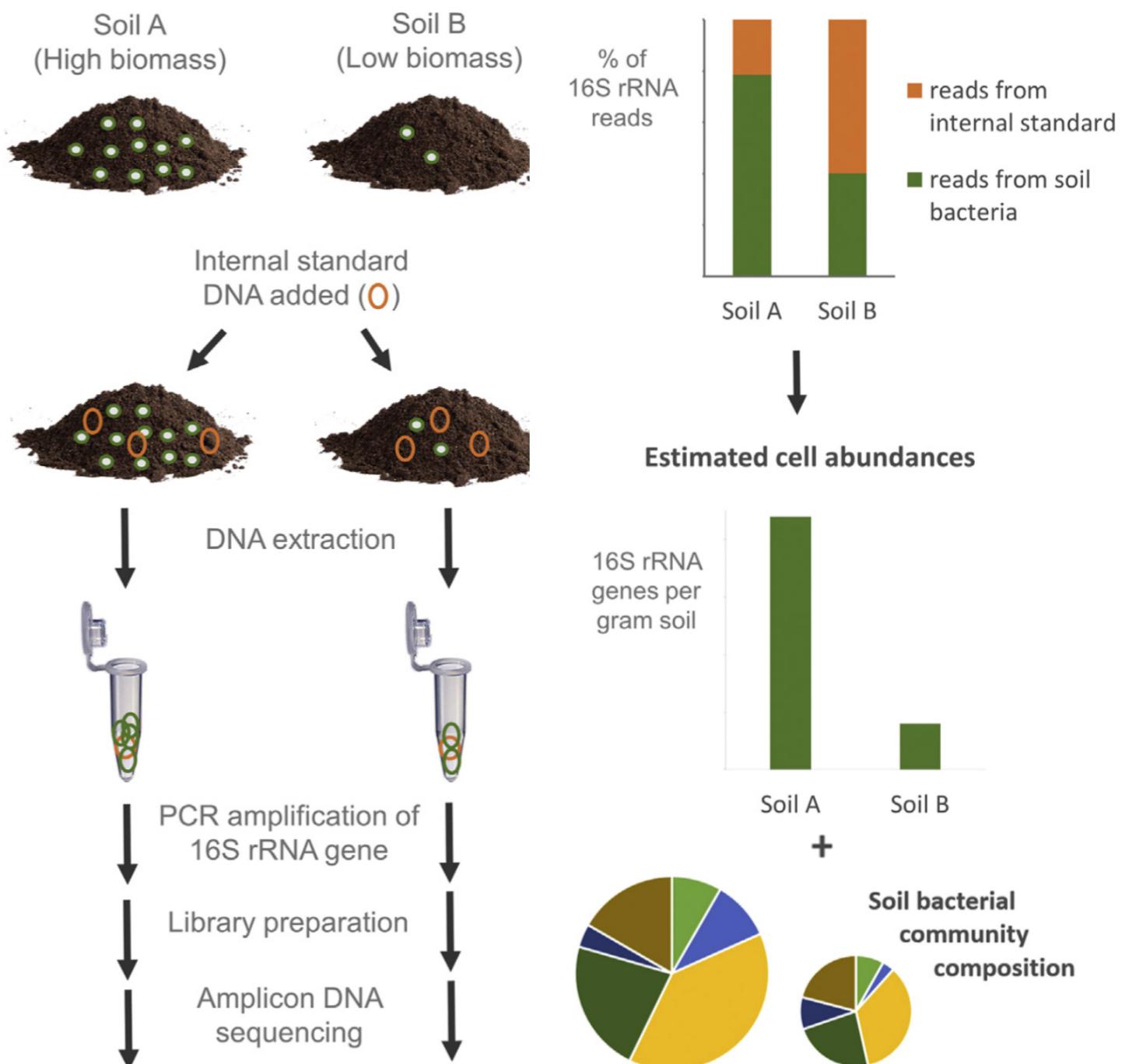
Flujo de trabajo típico - cantidad de templado

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



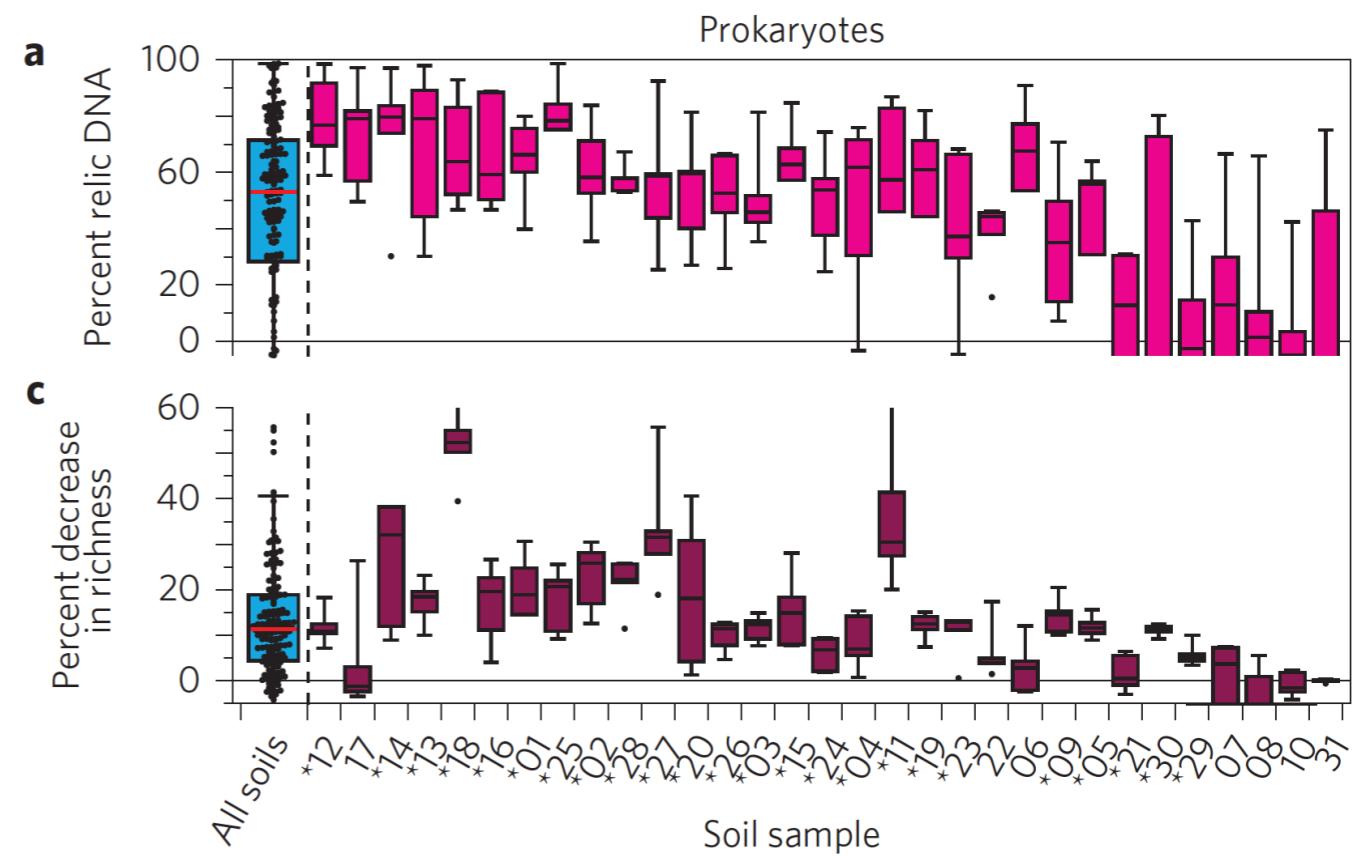
Cuantificación absoluta

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



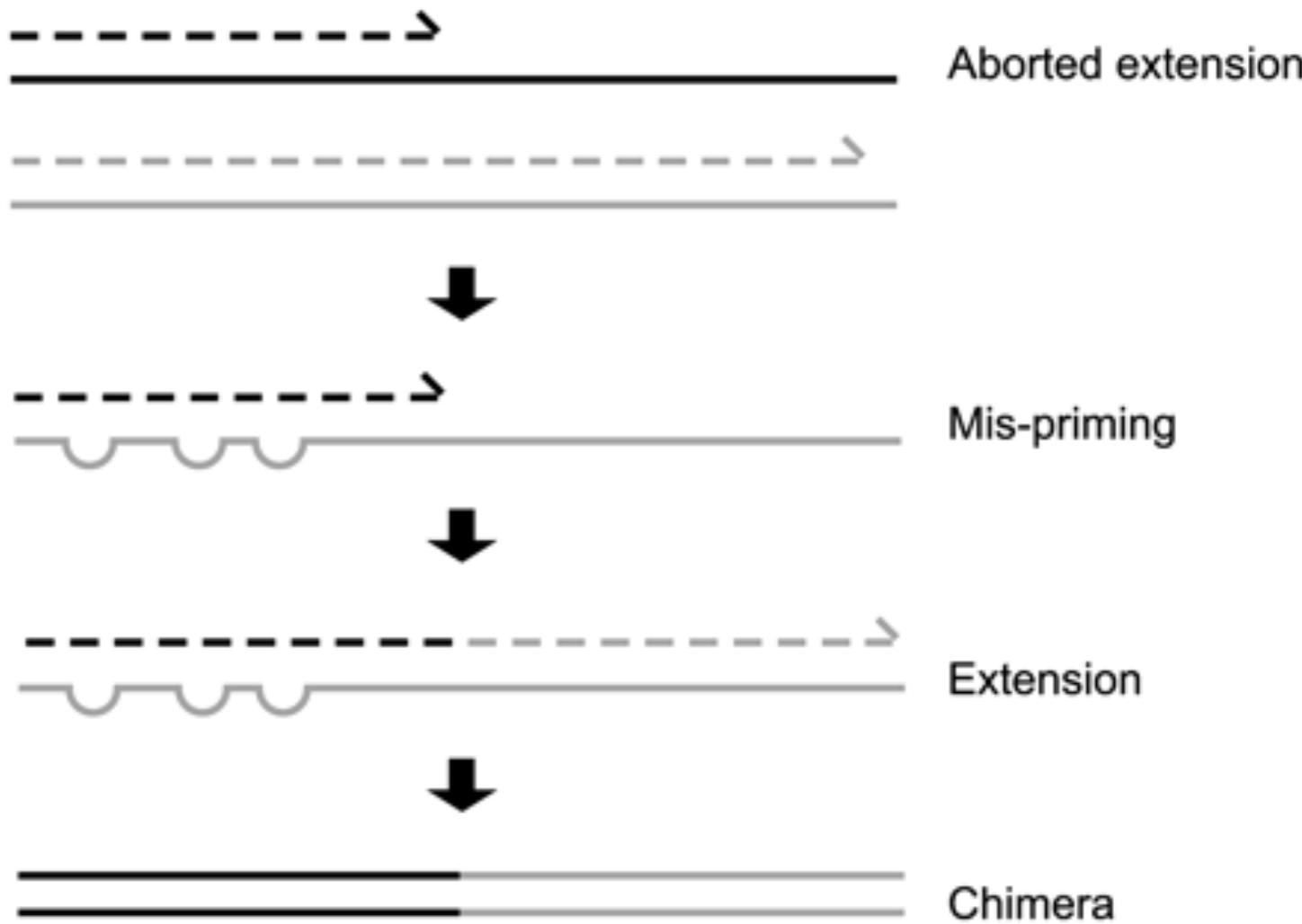
Limitaciones - DNA de células muertas y extracelular

- 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%)



Método popular - Kozich et al. 2013

- Datos del práctico obtenidos con este método



Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform

James J. Kozich,^a Sarah L. Westcott,^a Nielson T. Baxter,^a Sarah K. Highlander,^b Patrick D. Schloss^a

Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA^a; Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, USA^b

Método popular - Kozich et al. 2013

- Datos del práctico obtenidos con este método

V4 Region of the 16S rRNA gene

.....	CTTCCACTTAAATGAGACTT	GTGCCAGCMGCCGCGGTAA.....	ATTAGAWACCCBDGTAGTCC	ATACAGGTGAGCACCTTGTA...	+Strand
.....	GAAGGTGAATTACTCTGAA	CACGGTCGKCGGCCATT.....	TAATCTWTGGGVHCATCAGG	TATGTCCACTCGTGGAACAT...	-Strand

Forward primer construct

.....	GAAGGTGAATTACTCTGAA <u><p5 adapter><i5><padF><linkF></u>	CACGGTCGKCGGCCATT..... GTGCCAGCMGCCGCGGTAA->	TAATCTWTGGGVHCATCAGG	TATGTCCACTCGTGGAACAT...	-Strand
-------	---	---	----------------------	-------------------------	---------

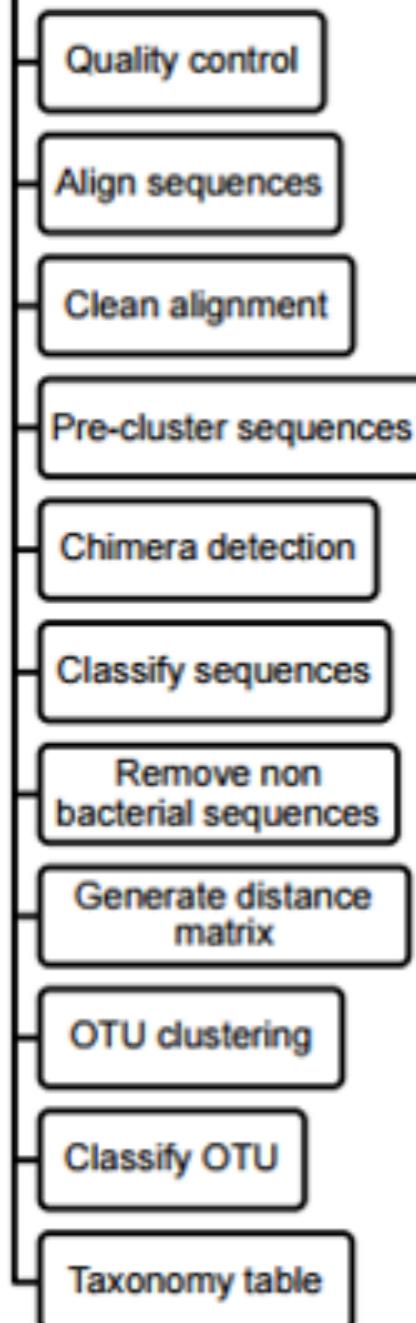
Reverse primer construct

.....	CTTCCACTTAAATGAGACTT	GTGCCAGCMGCCGCGGTAA.....	ATTAGAWACCCBDGTAGTCC <-TAATCTWTGGGVHCATCAGG	ATACAGGTGAGCACCTTGTA...	+Strand
				<linkR><padR><i7><p7 adapter>	

V4 Amplicon

AATGATA	CGGGGACCACCGAGATCTACACNNNNNNNTATGGTAATTGTGTGCCAGCMGCCGCGGTAA.....	ATTAGAWACCCBDGTAGTCCGGCTGACTGACTNNNNNNNNATGACGTATGCCGTCTTCTGCTTG
TTACTATGCCGCTGGCTCTAGATGTGNNNNNNNATACCATTAACACACGGTCGKCGGCCATT.....		TAATCTWTGGGVHCATCAGGCCGACTGACTGANNNNNNNNTAGAGCATACGGCAGAACAGAAC

mothur



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

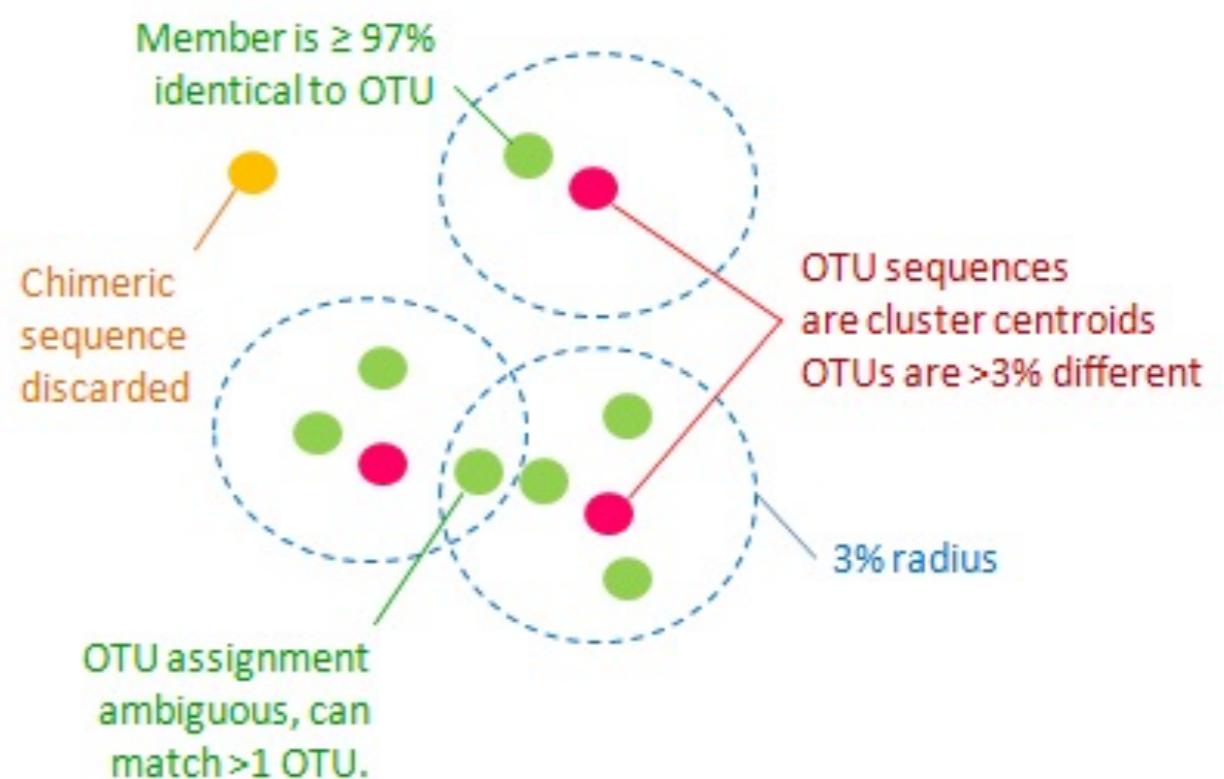
C Quast, E Pruesse, P Yilmaz, J Gerken... - Nucleic acids ..., 2013 - academic.oup.com

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)

PMCID: PMC523561

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

Jill E. Clarridge, III*

[Author information ▶](#) [Copyright and License information ▶](#)

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)

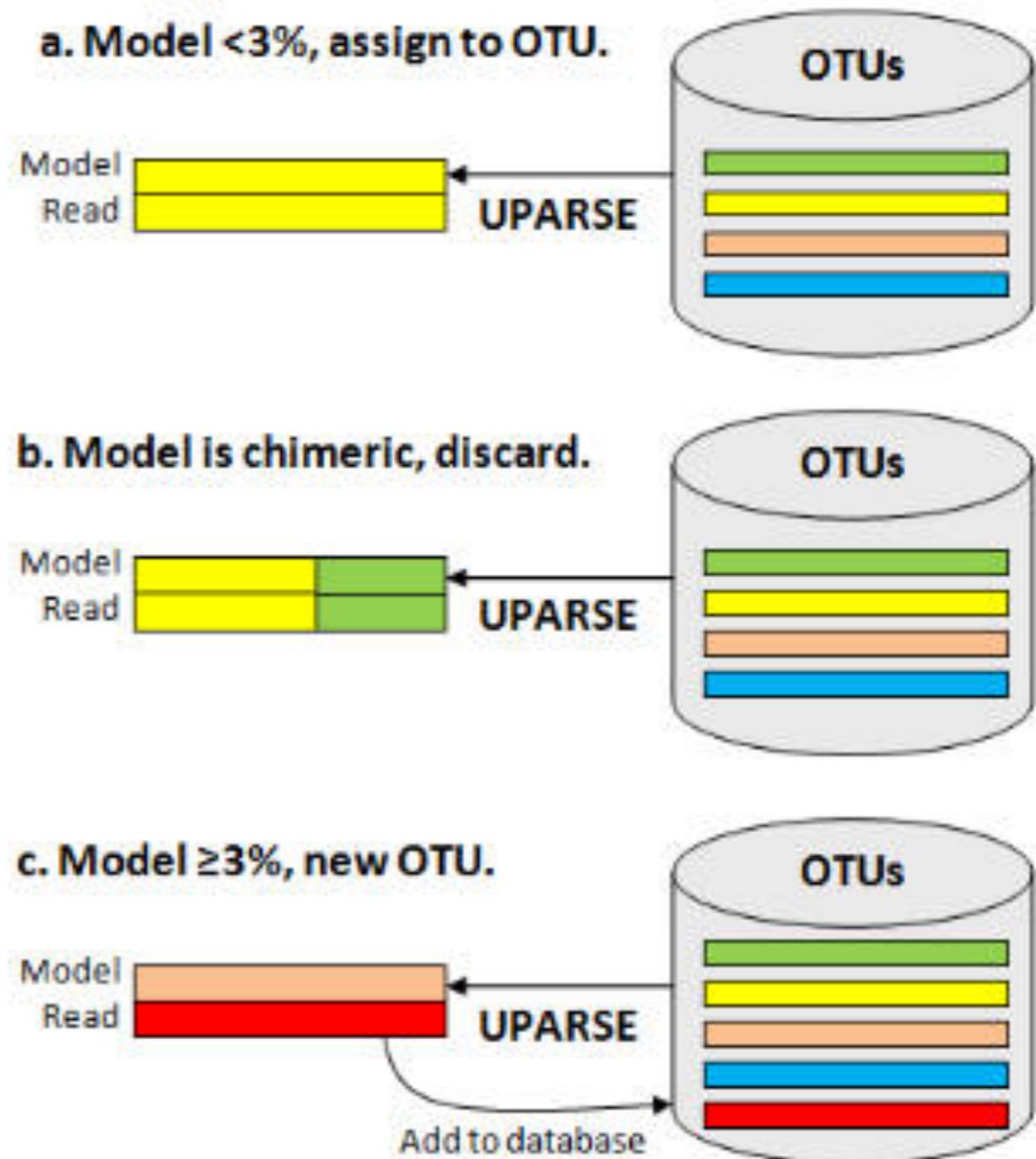
Ecological Significance of Microdiversity: Identical 16S rRNA Gene Sequences Can Be Found in Bacteria with Highly Divergent Genomes and Ecophysiolgies

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](https://doi.org/10.1128/CMR.17.4.840-862.2004)

PMCID: PMC523561

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

Jill E. Clarridge, III*

Author information ▶ Copyright and License information ▶

Elke Jaspers¹ and Jörg Overmann

MC492463

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	Hallella
Otu21163	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	NA
Otu16680	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	Helicobacter
Otu05526	Bacteria	Chlamydiae	Chlamydiae	Chlamydiales	Chlamyiaceae	Chlamydia
Otu24307	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiales_unclassified	NA
Otu19152	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Pseudoflavitractor
Otu20183	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Marvinbryantia
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

Algunos análisis...

Rarefacción

- ¿Hay más diversidad en mi muestra?
- Homologar la profundidad de las muestras

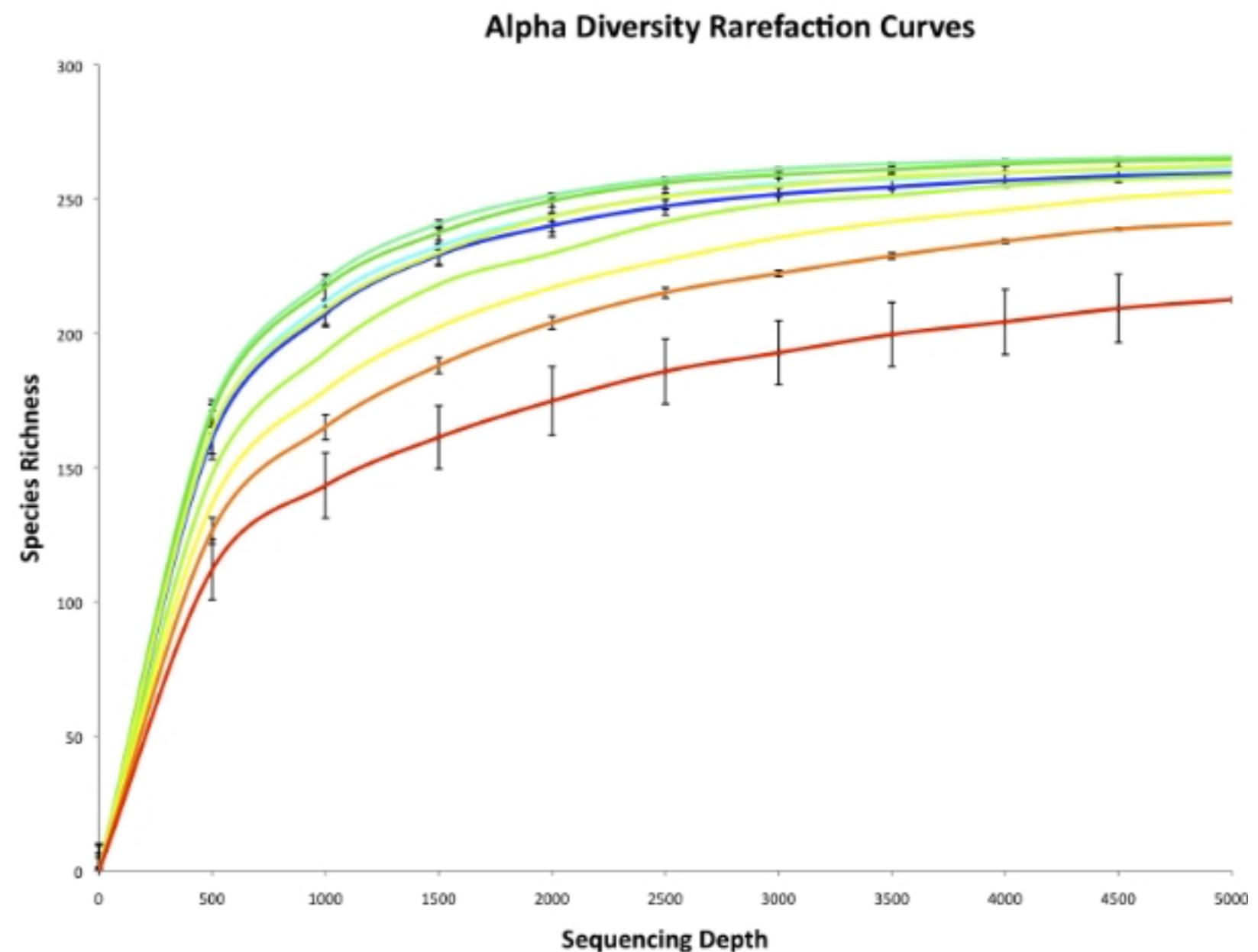
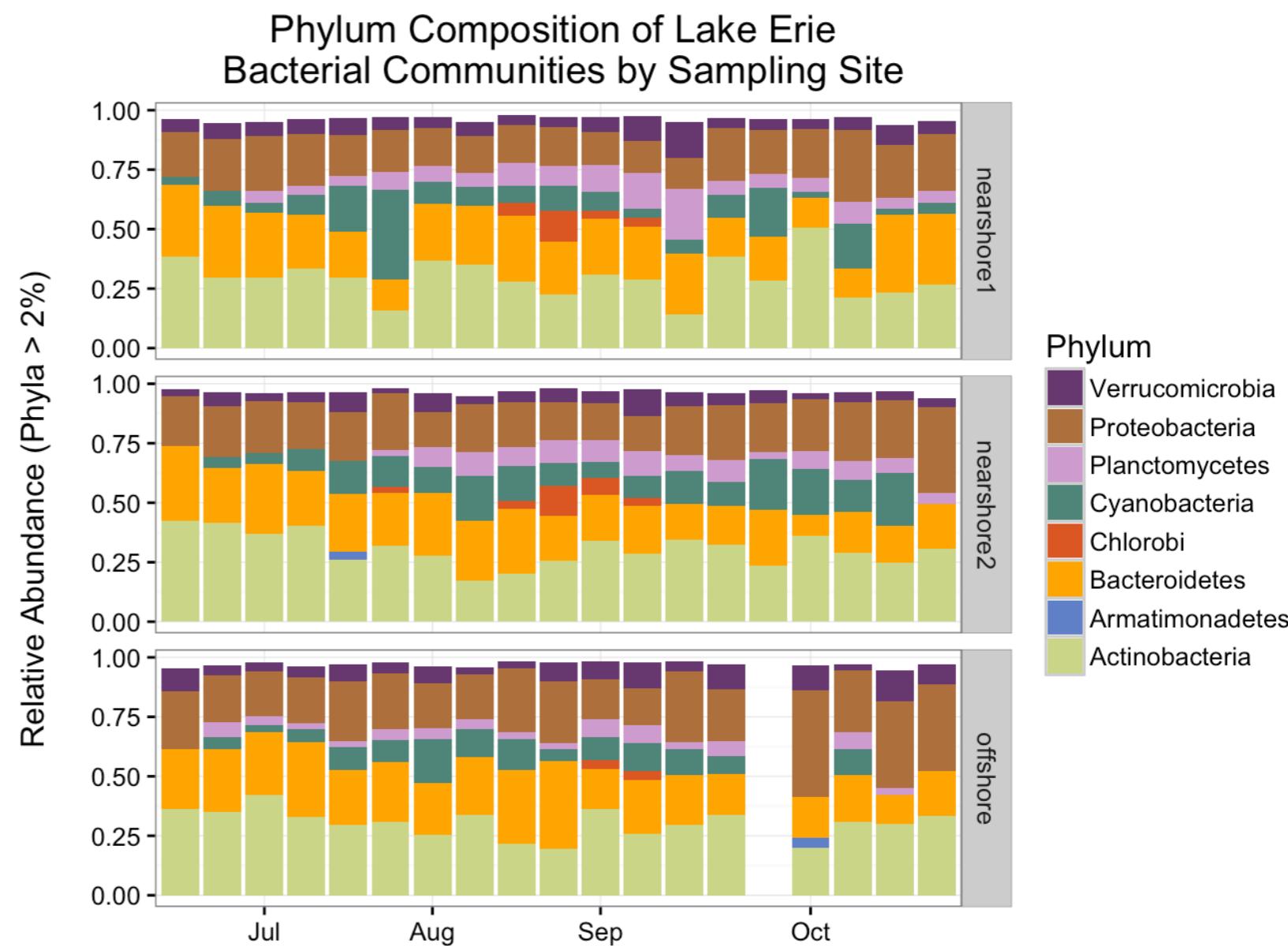


Gráfico de barras apiladas

- Vista general de qué hay y en qué abundancia



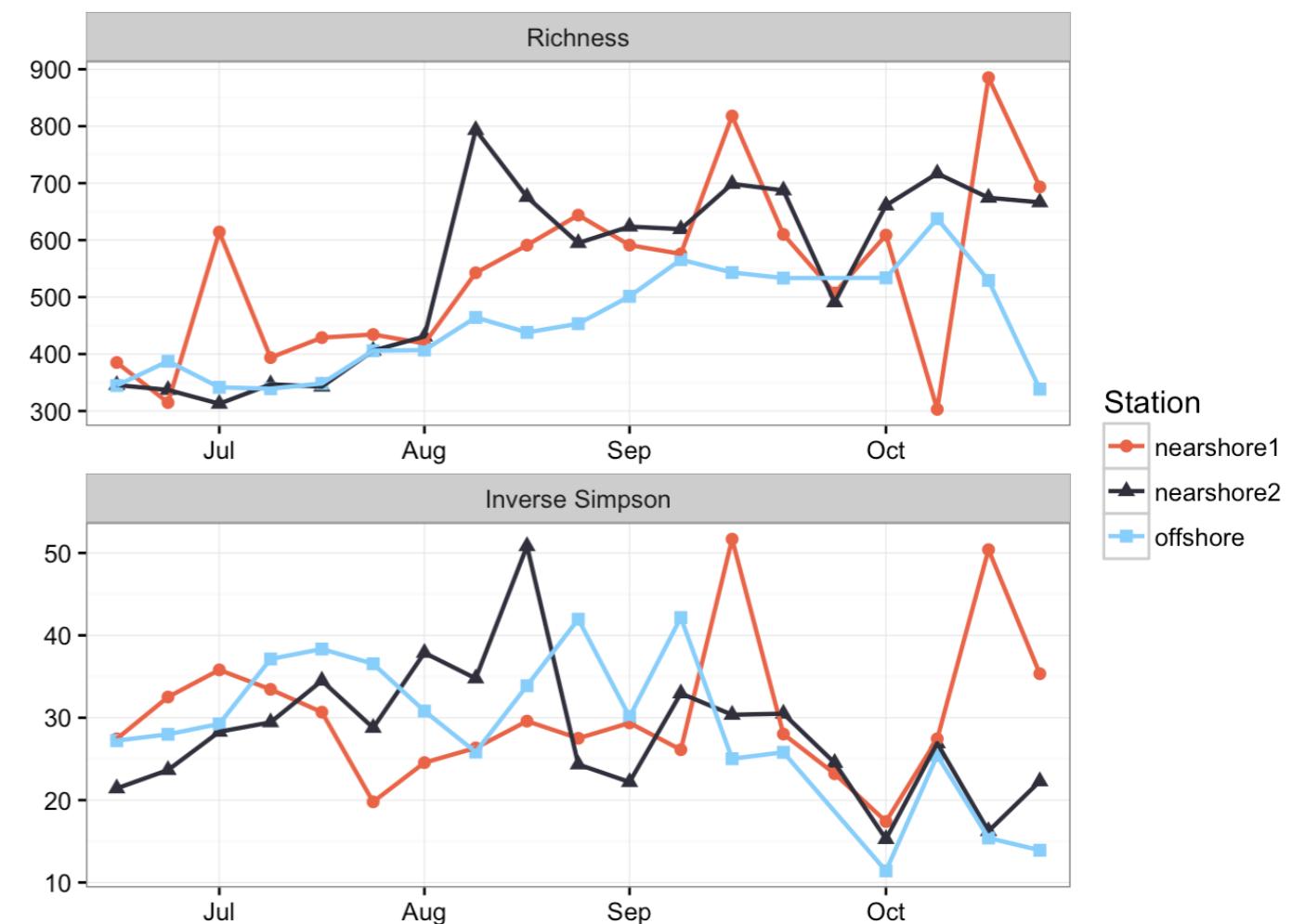
Diversidad Alfa - Species richness vs evenness

- Riqueza = número de especies en un sitio local, no considera abundancia
- Evenness = una comunidad dominada por solo una especie es menos diversa

Flower Species	Numbers of individuals	
	Sample 1	Sample 2
Daisy	300	20
Dandelion	335	49
Buttercup	365	931
Total	1000	1000

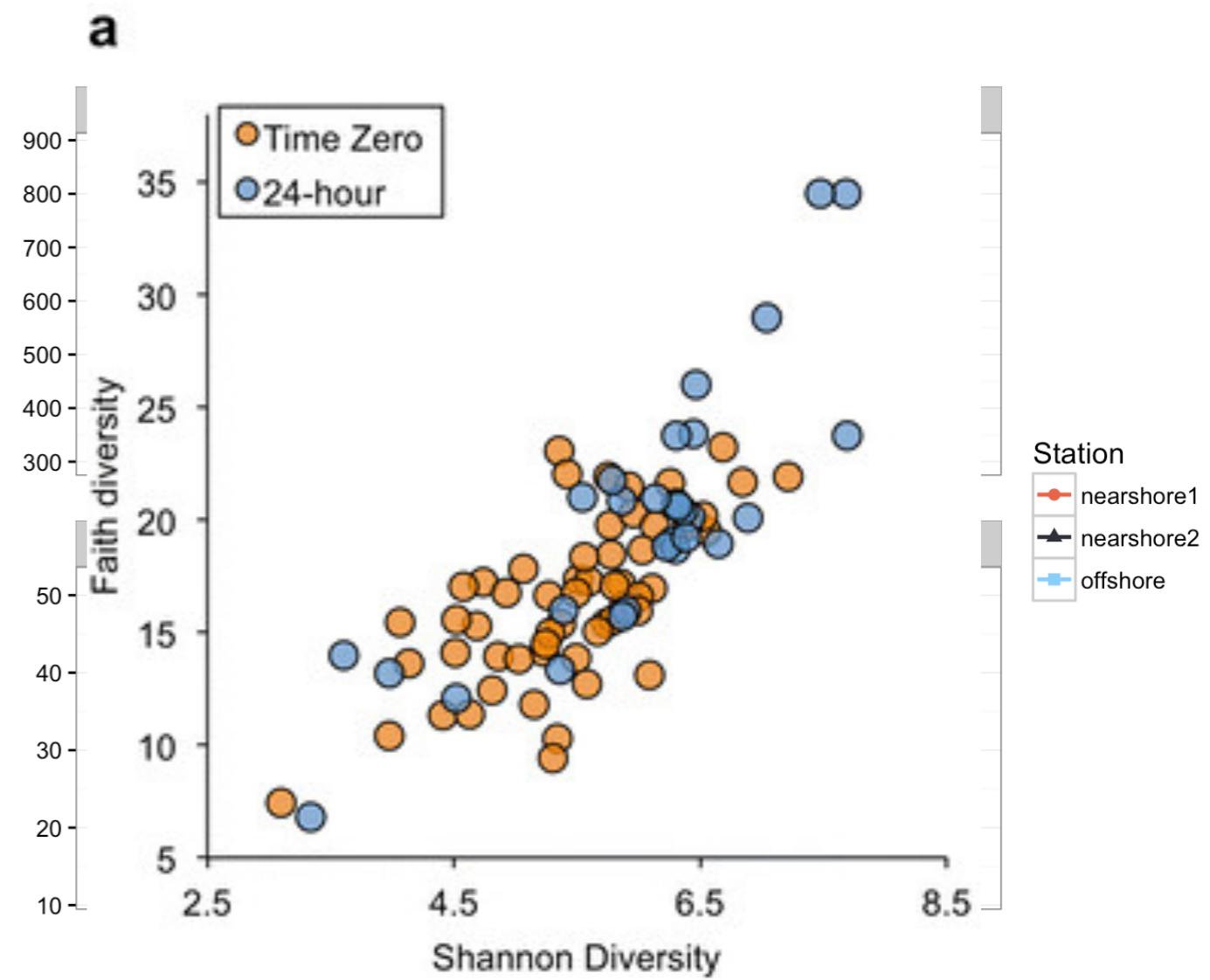
Species richness vs evenness

- Richness y evenness → Shannon; Simpson's
- Richness → Chao1
- Sensibles al tamaño muestral
- Probability of an Interspecific Encounter (PIE) no es sensible
- Diversidad filogenética de Faith's



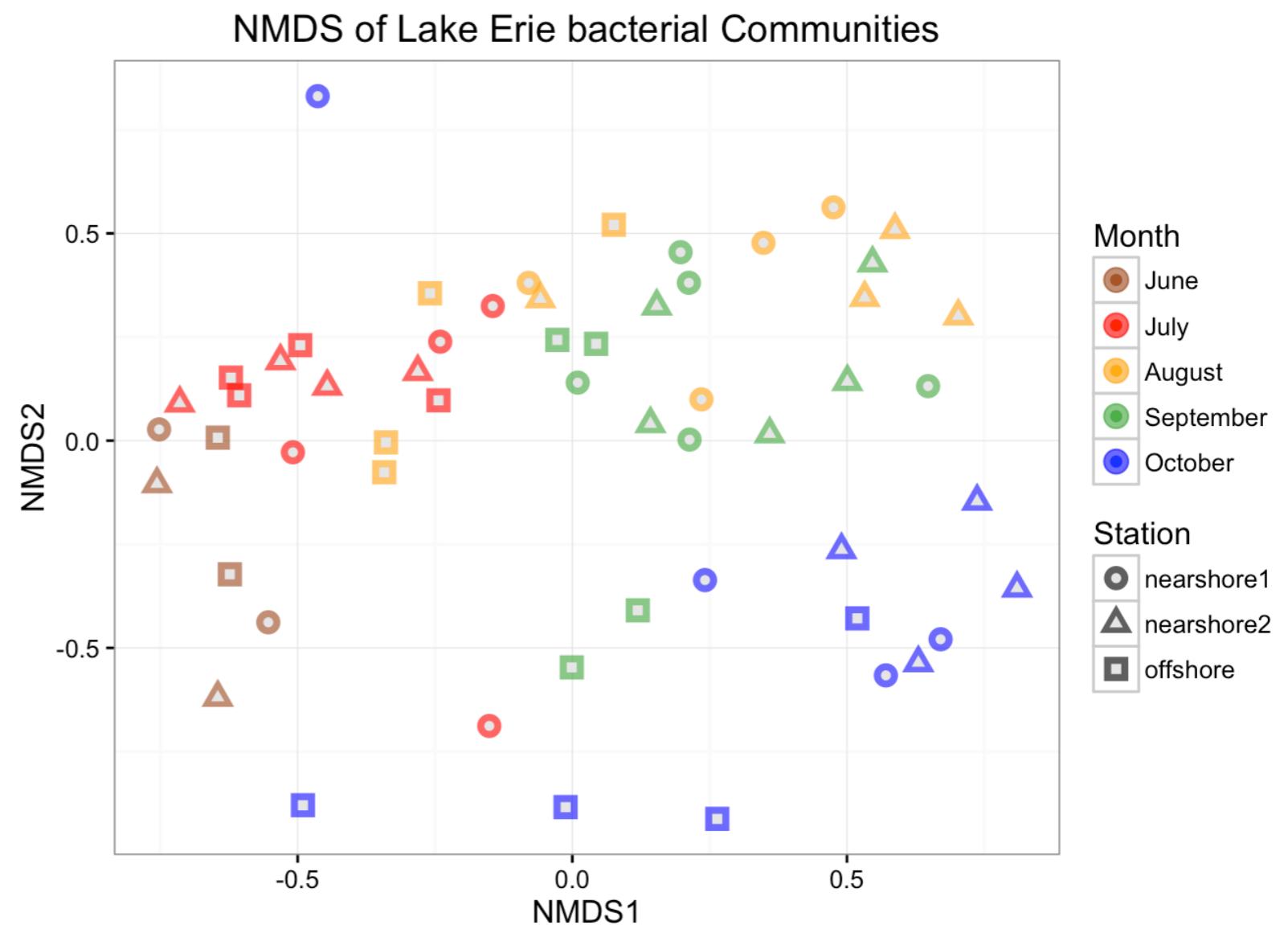
Species richness vs evenness

- Richness y evenness → Shannon; Simpson's
- Richness → Chao1
- Sensibles al tamaño muestral
- Probability of an Interspecific Encounter (PIE) no es sensible
- Diversidad filogenética de Faith's



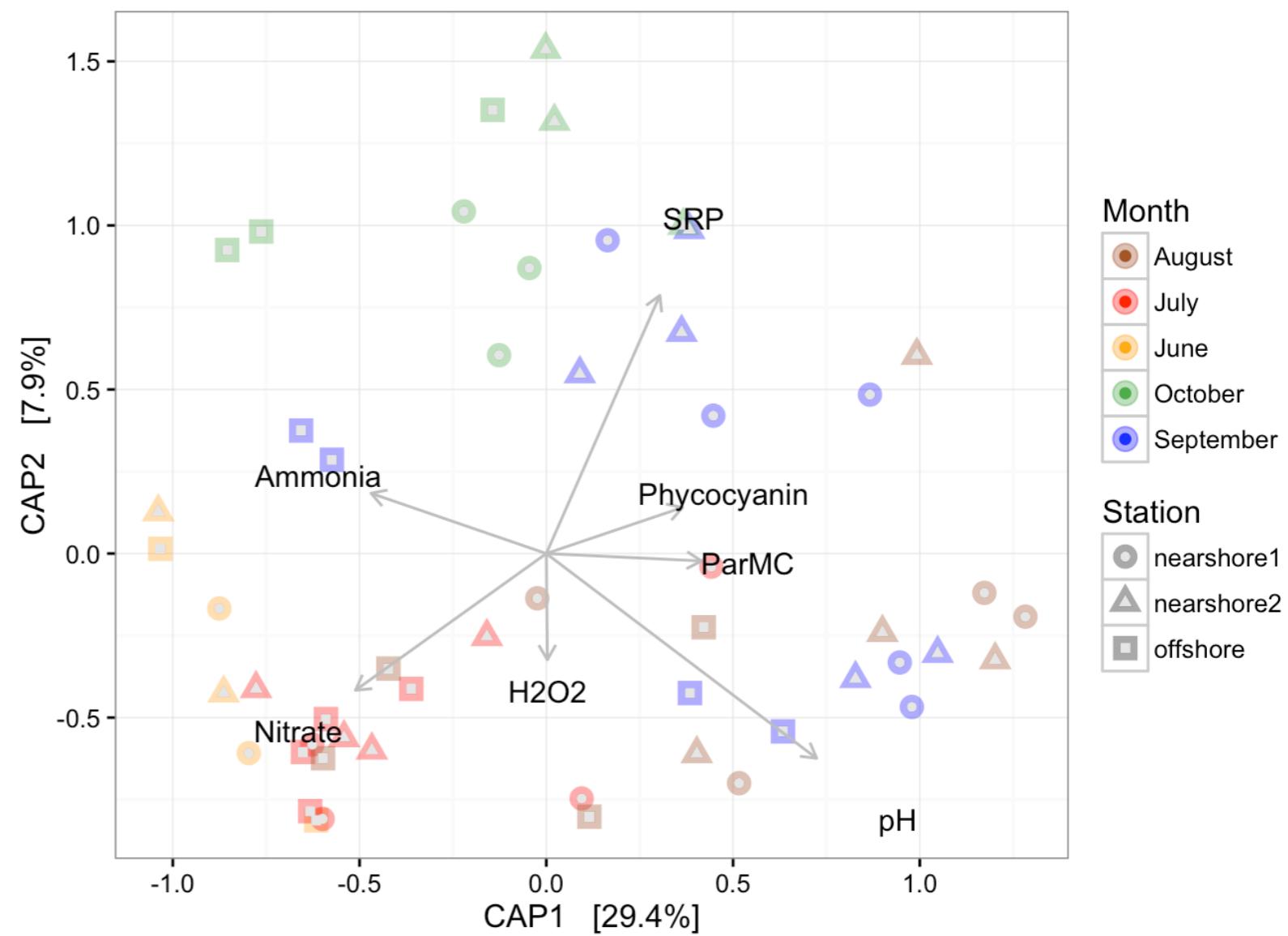
Diversidad beta

- Comparación entre comunidades, e.g., suelo agrícola vs suelo nativo
- Cálculo de distancia (Unifrac; Bray-Curtis) y métodos multivariados, e.g., PCA, CCA, NMDS



Diversidad beta

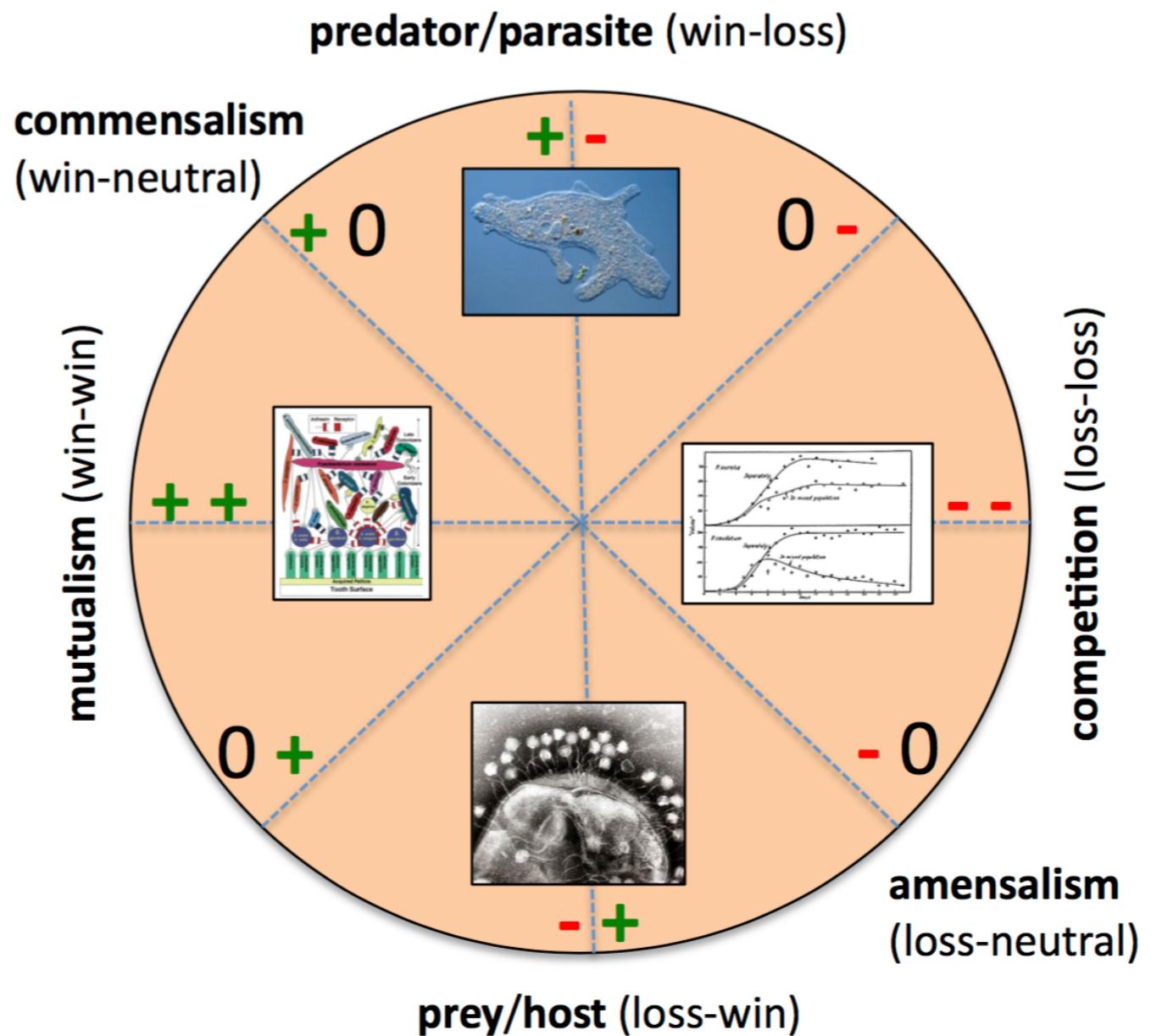
- Comparación entre comunidades, e.g., suelo agrícola vs suelo nativo
- Cálculo de distancia (Unifrac; Bray-Curtis) y métodos multivariados, e.g., PCA, CCA, NMDS



Co-ocurrencia

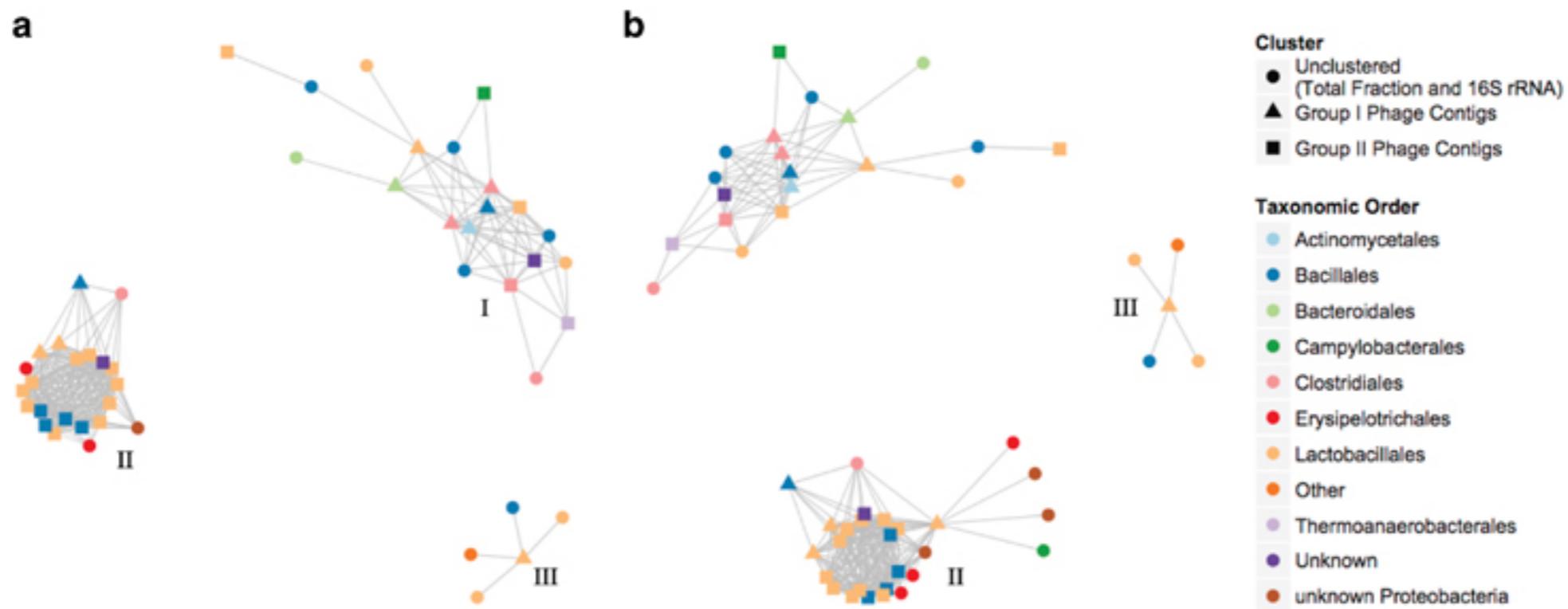
ecological relationships

- OTUs que co-ocurren pueden estar interaccionando funcionalmente
- Evidencia de competencia
- Mutualismo, comensalismo, depredador / presa
- Co-presencia - sobrelapamiento de nicho, comensalismo, mutualismo
- Exclusion mutua - competencia, preferencia de nicho alternativa



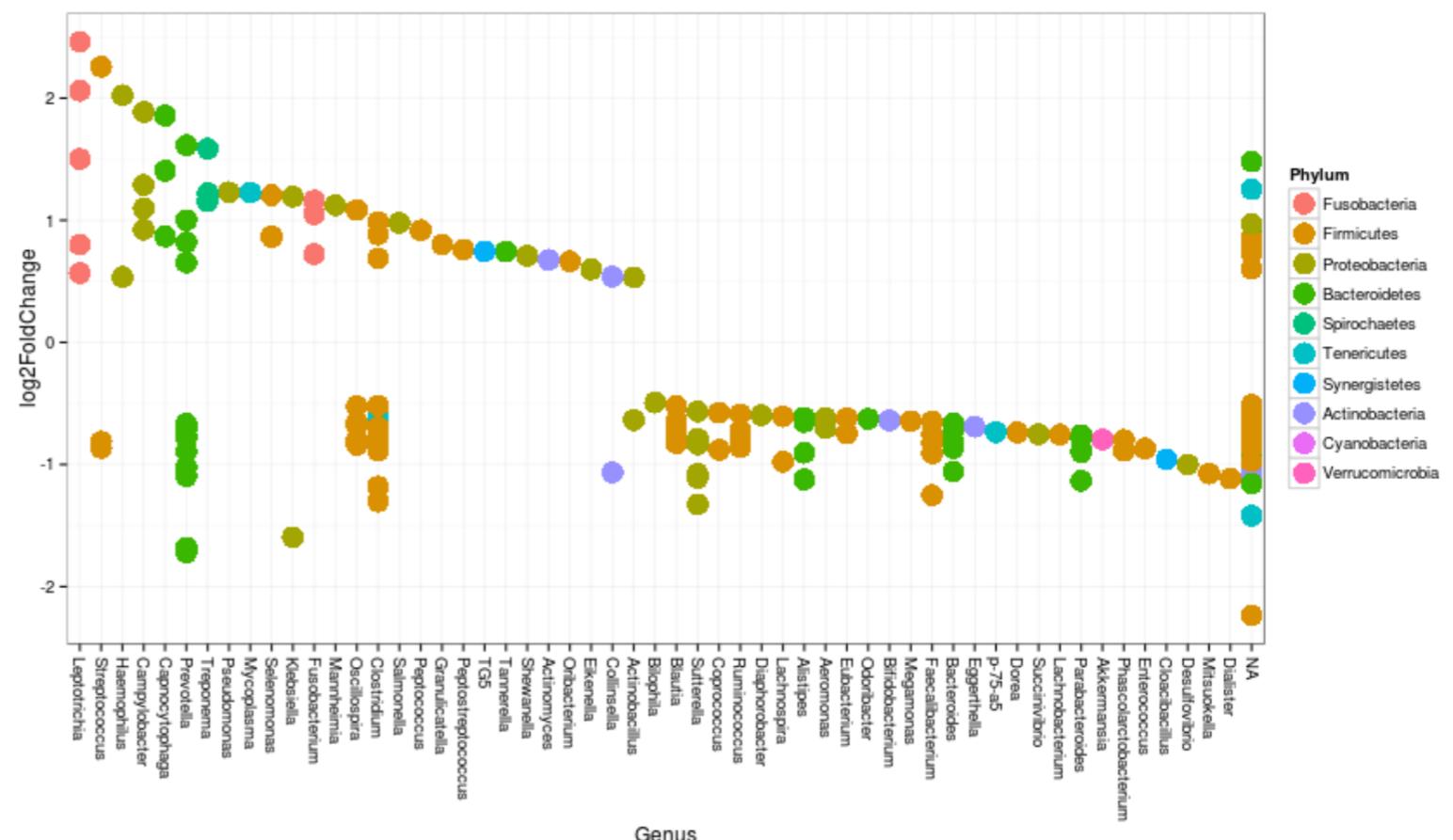
Co-ocurrencia - basada en similitud

- Calcular distancia entre OTUs
- Randomizar para obtener significancia
- Visualizar
- Se puede hacer en R/phyloseq



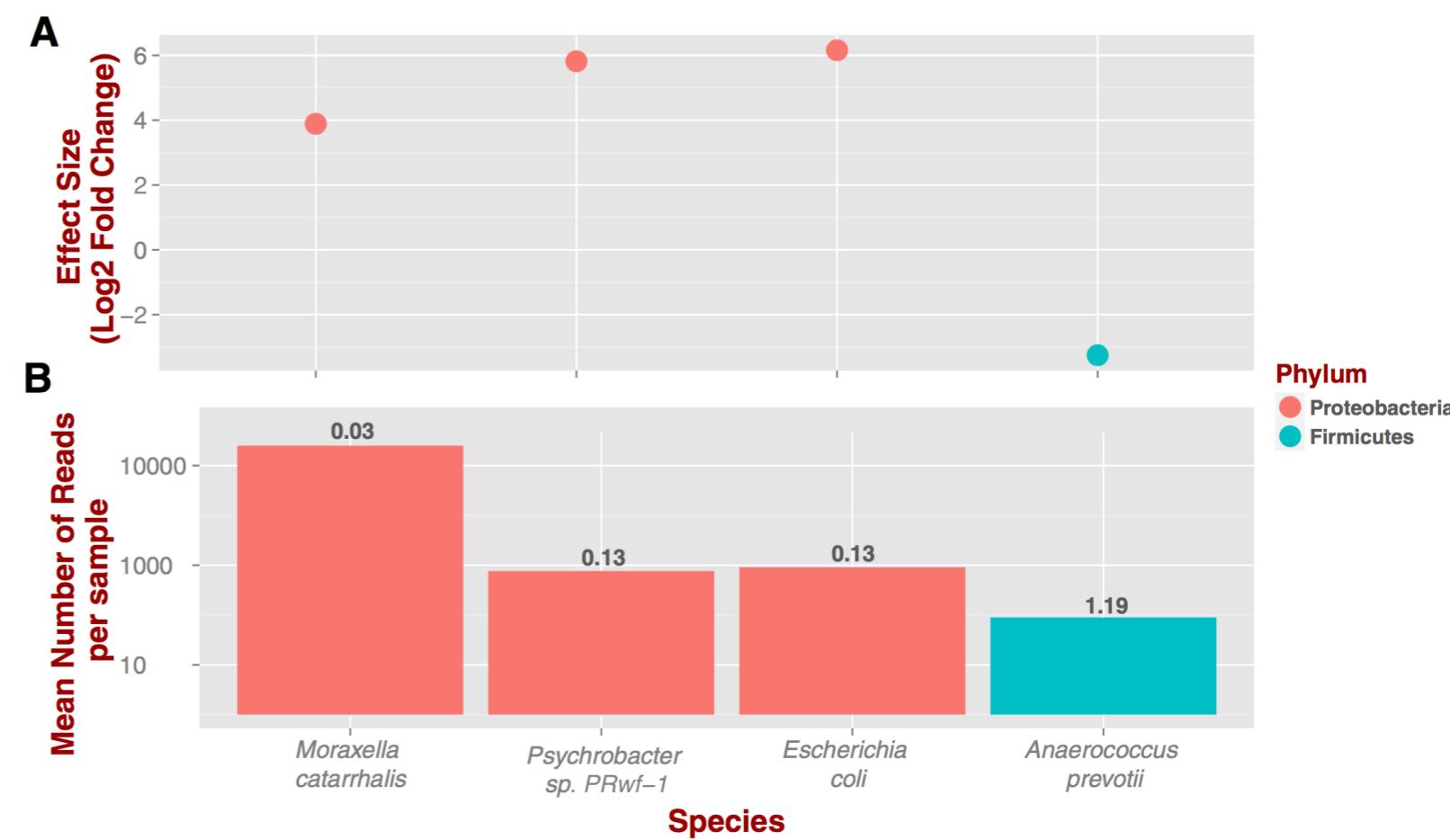
Abundancia diferencial

- Equivalente a expresión diferencial de genes
- Usamos phyloseq y DESeq2 de Bioconductor
- Fold Change en Log2



Abundancia diferencial

- Usamos phyloseq y DESeq2 de Bioconductor
- Normalizamos con una NB y luego hacemos un test de Wald. p-value corregido por BH
- Fold Change en Log2



Práctico para hoy

- DADA2
- Papers: DADA2 y Kozich et al. 2013