

Curso Intensivo Intersemestral  
(Paleo)Bio Indicadores Neotropicales

# Estudio de la biodiversidad del Lago de Chalco usando metagenómica

Últimos 11,600 años

**Dra. Barbara Moguel**

**Instituto de Geología, UNAM**



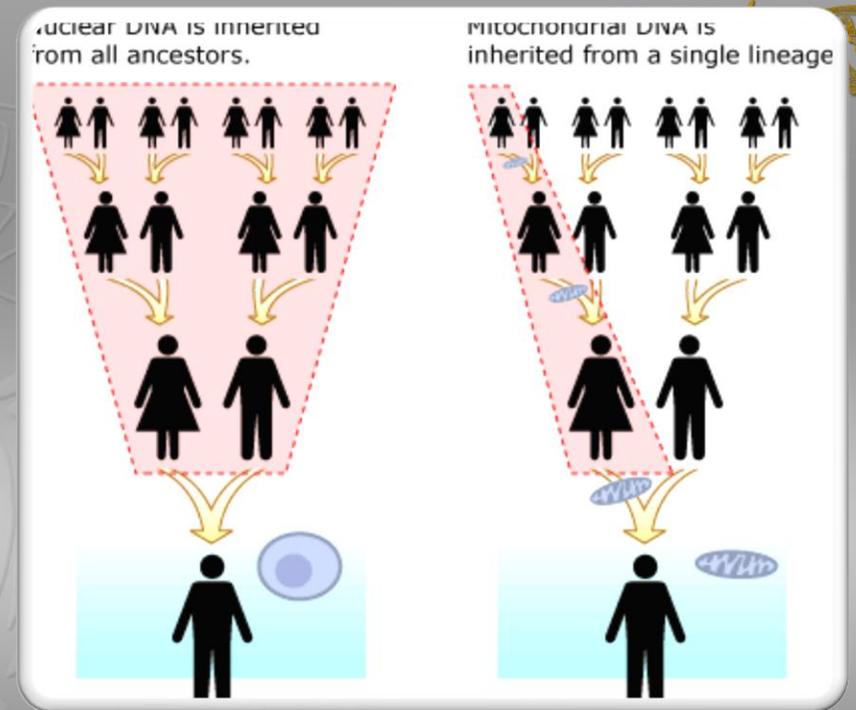
**POSGRADO EN CIENCIAS DEL MAR Y LIMNOLOGÍA**

# Genómica

Es el estudio de la organización, estructura y función de los complementos totales de los genes y otros DNAs dentro de las células de un organismo.

La genómica se puede diferenciar de la genética, ya que toma en cuenta la **función e interacción** de múltiples genes, **DNA no codificante** (INTRONES), **DNA intergénico** (secuencias de DNA localizadas entre agrupaciones de genes) y **la arquitectura genómica**.

- La **genética** se refiere al estudio de un solo gen y sus productos.



# Genómica

Se divide en genómica estructural y genómica funcional

La genómica estructural:

Se refiere a la fase inicial del análisis del genoma, que incluye la **construcción de mapas genéticos y físicos** de un genoma, **la identificación y la anotación** de las características de los genes y **la comparación** de las estructuras del genoma.

La genómica funcional:

Se refiere al análisis de la expresión global y función de los genes de un genoma.



# Procedimiento

- A. Muestreo
- B. Filtrado
- C. Purificación
- D. Construcción de librerías
- E. Secuenciación
- F. Bioinformática
  - A. Control de calidad
  - B. Ensamblado
  - C. Predicción de marcos de lectura
  - D. Anotación
  - E. Análisis estadísticos
- G. Presentación de datos

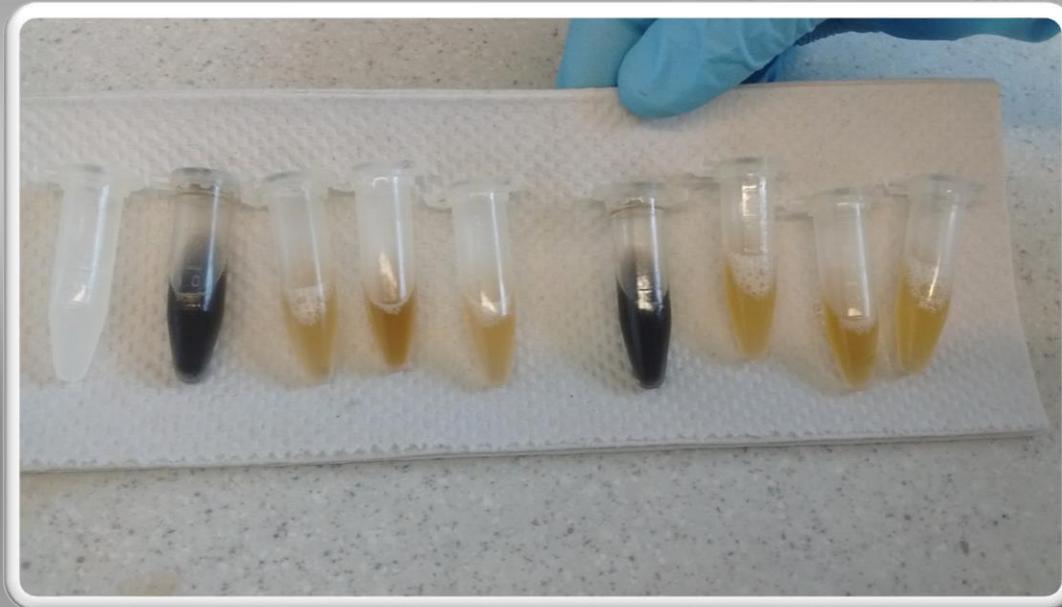


Environmental Shotgun Sequencing (ESS)  
Wooley, 2010

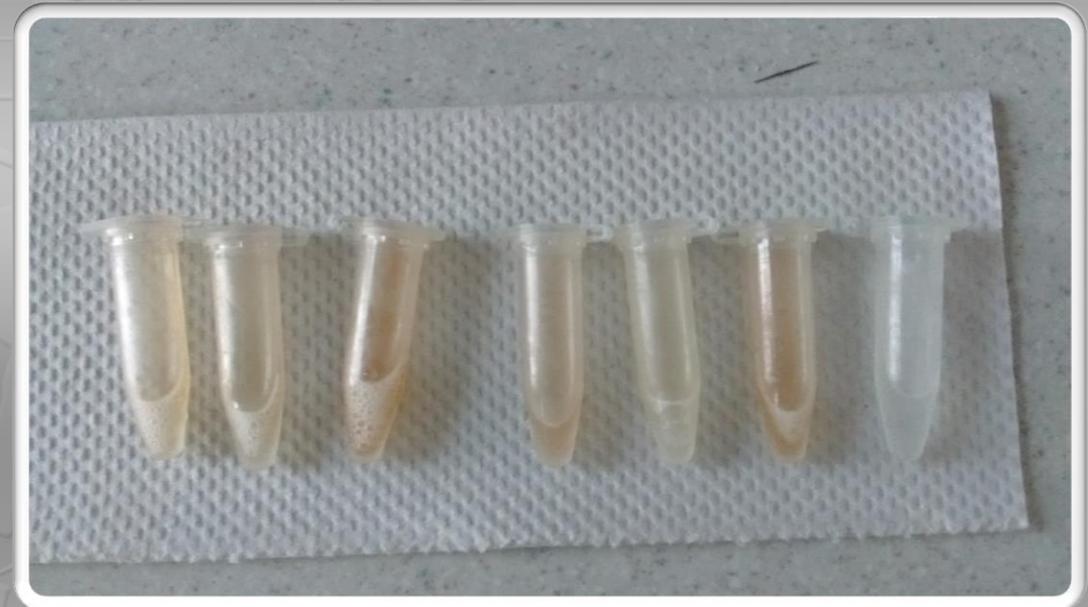
# MUESTREO



# Extracción DNA: lisados de cada muestra

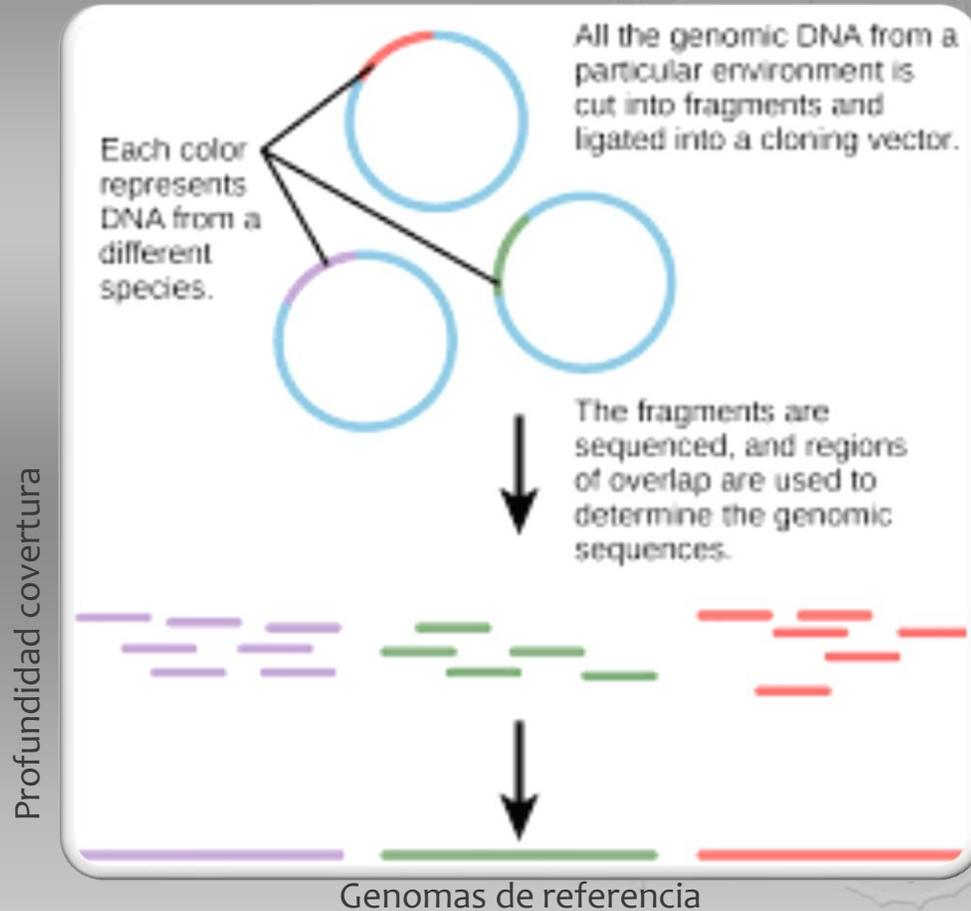


Más superficiales



Más profundas

# Secuenciación



Es un método que nos permite descubrir como es la secuencia del ADN.

# CASO IDEAL

Supongamos que tenemos las siguientes secuencias:

ACCGT  
CGTGC  
TTAC  
TACCGT

Una posible manera de ensamblarlas :

--ACCGT--  
----CGTGC  
TTAC-----  
-TACCGT--

---

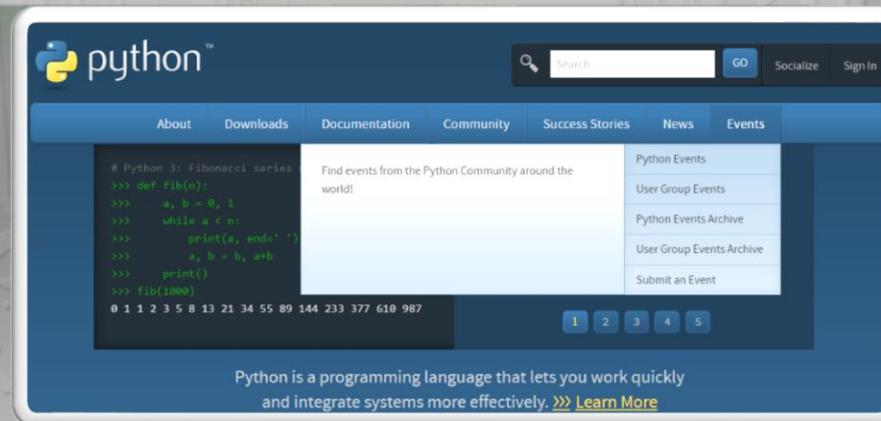
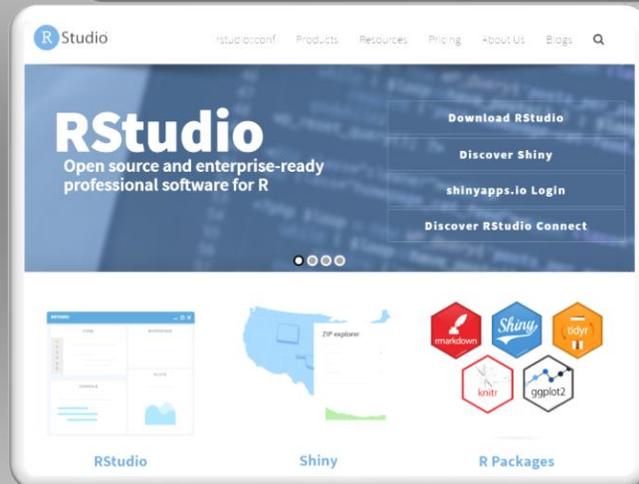
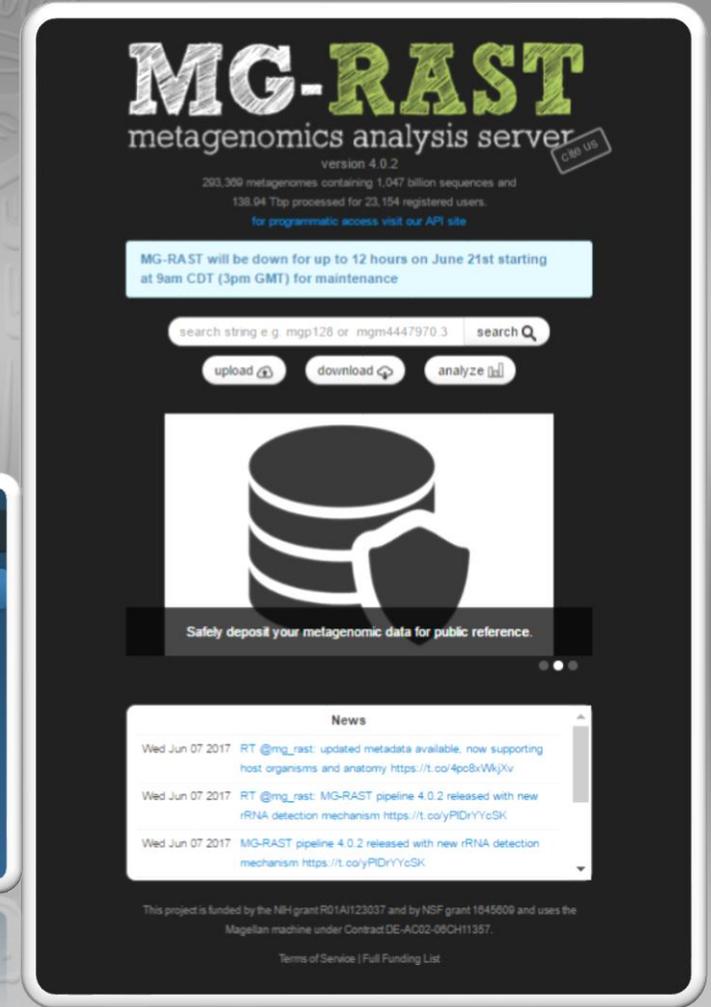
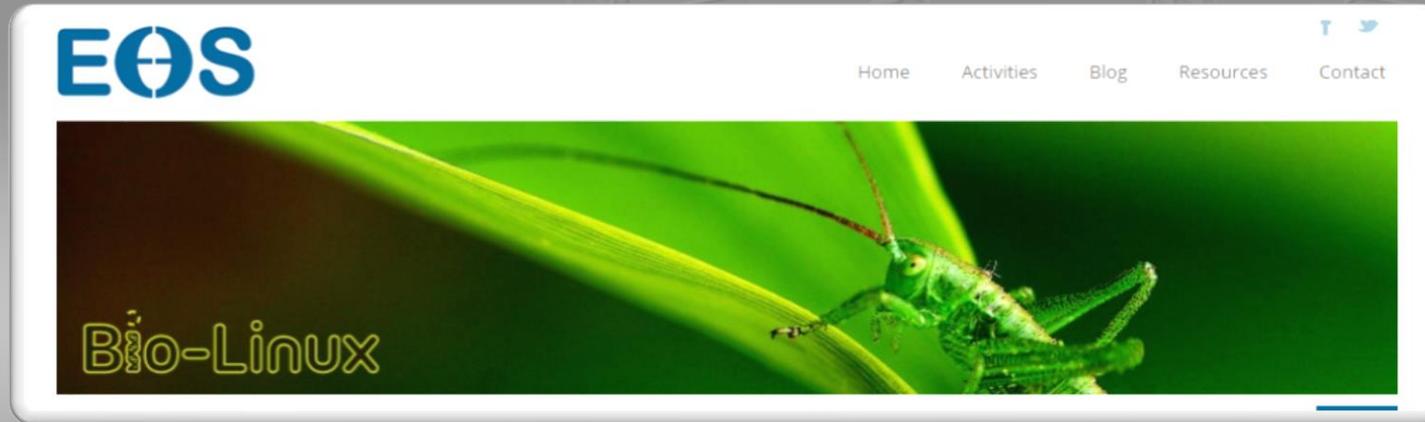
TTACCGTGC

Las secuencias fueron alineadas ignorando los espacios de las extremidades. Se trata de alinear en la misma columna bases que sean iguales.

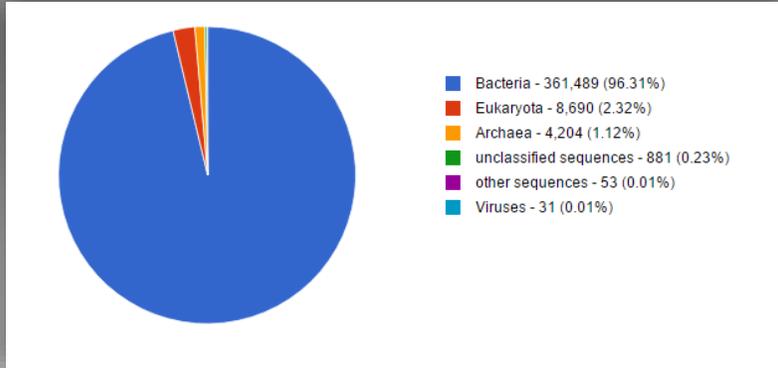
# COMPLICACIONES

1. secuencias reales muy largas
2. Errores
  - Inserciones, deleciones , sustituciones
  - Fragmentos quiméricos
  - Contaminación
  - Regiones repetidas
3. Orientación desconocida
5. Pérdida de cobertura

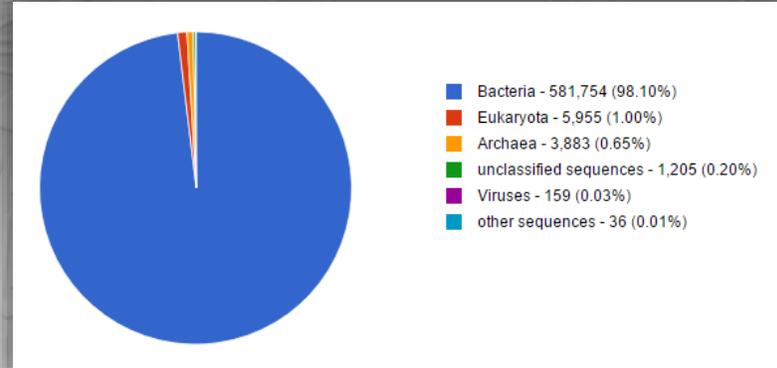
# Análisis bioinformático



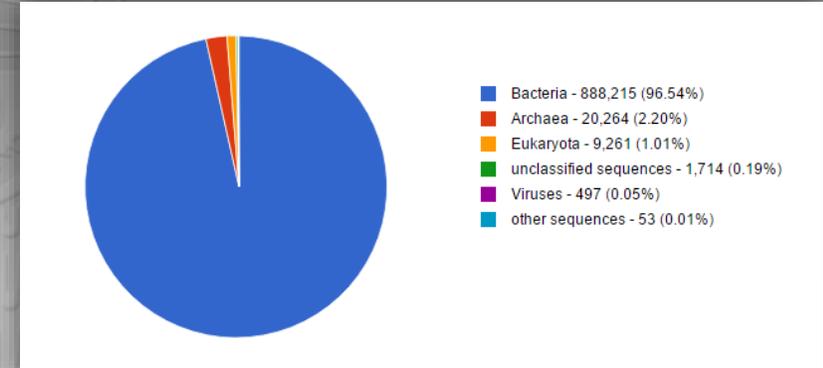
# Primeros Resultados



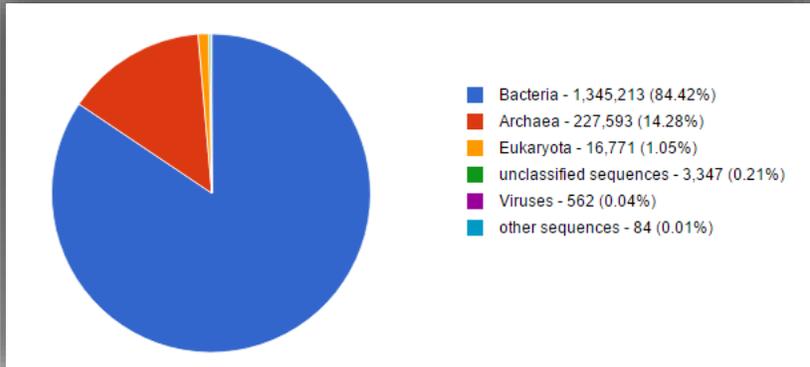
Superficial



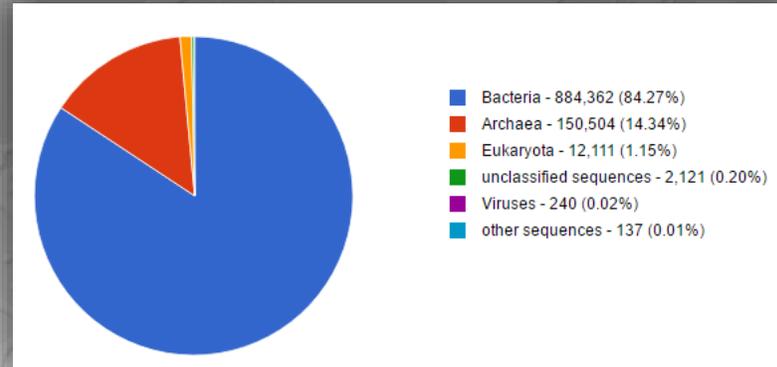
50cm



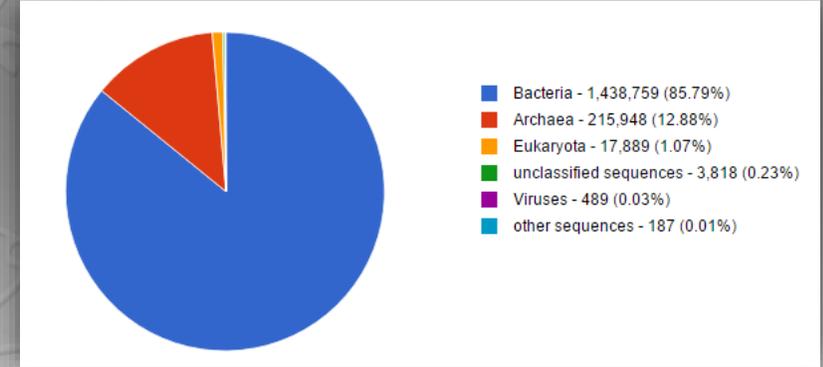
100 cm



150 cm



200 cm



250 cm

# Algunos resultados

# Parámetros utilizados 27.06.2017

## Create a new Analysis

To perform an analysis, you must first load the metagenomic profiles to analyze. A profile holds the abundance values and cutoffs for a list of database sources for a specific dataset. You can select the databases and datasets, as well as a name for your analysis below. Click the ✓-button to load the data from our server.

Profiles are generated on demand. Depending on profile size the initial calculation may take some time. Once computed they will be cached and subsequent requests will download immediately. You can use the 📁-icon in the top menu bar to store profiles on your harddrive and upload them back into your browser cache (without requiring interaction with our server).

Once all required data is loaded you can start the analysis.

selected databases

RefSeq × Subsystems × RDP × Silva LSU × Silva SSU ×

available databases

IMG ▾ add

metagenomes

🛒 add collection ▾

Enter filter

name ▾

analysis name

sequence type **all** shotgun amplicon metatranscriptome status **all** public private

Empty list area for metagenomes.



- s1
- s2
- 501
- 502
- 100\_1
- 100\_2
- 100\_3
- 100\_4
- 200\_1
- 200\_2
- ...



E-value 10  
 % ident 90  
 Largo 15  
 Min. Abundance 1  
 Best hit

# Tabla de metadatos

MGRAST	samp_name	Depth	Age	Temp	Salinity
mgm4736471.3	S1-1	0	0	TEMPERED	FRESH_WATER
mgm4736462.3	S2-2	0	0	TEMPERED	FRESH_WATER
mgm4736465.3	50-1	50	5,000	TEMPERED	FRESH_WATER
mgm4736468.3	50-2	50	5,000	TEMPERED	FRESH_WATER
mgm4736463.3	100-1	60	5,700	TEMPERED	FRESH_WATER
mgm4736464.3	100-2	85	6,500	TEMPERED	FRESH_WATER
mgm4736461.3	100-3	175	7,000	WARM	HYPERSALINE
mgm4736467.3	100-4	135	7,700	WARM	HYPERSALINE
mgm4736469.3	200-1	160	8,700	WARM	HYPERSALINE
mgm4736460.3	200-2	185	9,600	WARM	HYPERSALINE
mgm4736466.3	200-3	210	10,000	WARM	HYPERSALINE
mgm4736470.3	200-4	235	11,600	COLD	SUBSALINE

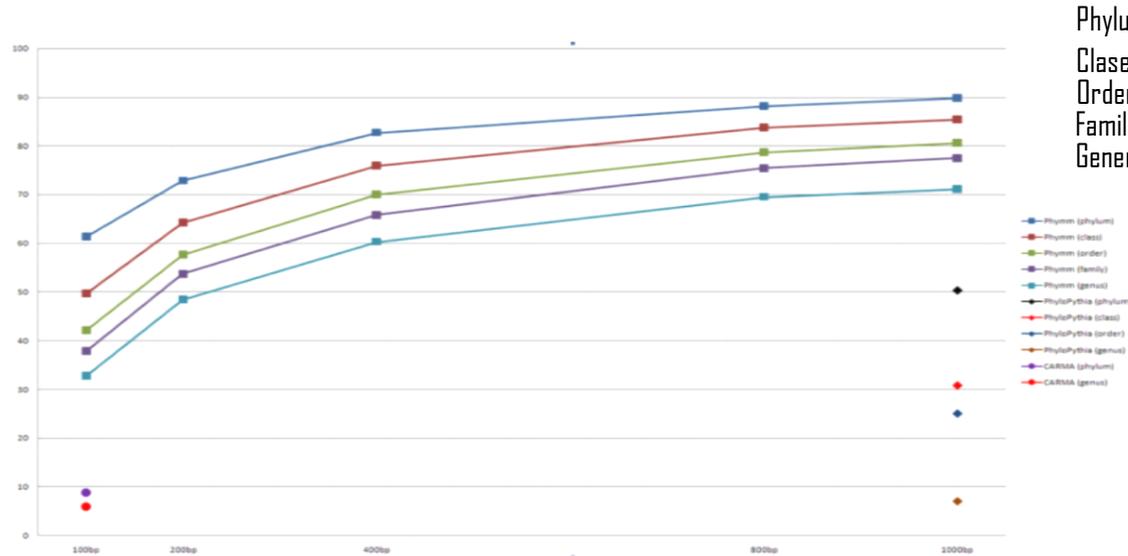
# Tamaño de los fragmentos secuenciados por muestra

filename	total_length	number	mean_length	longest	shortest	N_count	Gaps	N50
100-1_21.fas	168811241	1887417	89.44	4859	41	80	8	85
100-1_31.fas	60107503	486460	123.56	12083	61	0	0	116
100-2_21.fas	308485208	3304698	93.35	3546	41	68	6	93
100-2_31.fas	176113680	1267483	138.95	6647	61	10	1	134
100-3_21.fas	349796177	3911877	89.42	5961	41	60	6	86
100-3_31.fas	188038256	1418823	132.53	14524	61	0	0	127
100-4_21.fas	416479925	4672171	89.14	5162	41	30	3	78
100-4_31.fas	201903793	1510025	133.71	12673	61	0	0	129
200-1_21.fas	240329567	2572442	93.42	6428	41	30	3	86
200-1_31.fas	108903983	768715	141.67	11367	61	10	1	131
200-2_21.fas	476543239	5651099	84.33	4834	41	30	3	76
200-2_31.fas	149098522	1127054	132.29	9402	61	0	0	125
200-3_21.fas	480121605	5559632	86.36	3010	41	28	2	76
200-3_31.fas	179917131	1319946	136.31	6180	61	0	0	131
200-4_21.fas	296495266	3378439	87.76	5034	41	50	5	78
200-4_31.fas	127502956	966729	131.89	4493	61	10	1	126
501_21.fas	120884739	1437643	84.09	1725	41	200	20	77
501_31.fas	27150841	227310	119.44	1235	61	0	0	113
502_21.fas	175809596	1987626	88.45	3117	41	10	1	87
502_31.fas	59608231	476724	125.04	4556	61	0	0	118
S1_21.fas	97446812	1185258	82.22	867	41	270	27	76
S1_31.fas	12711989	115806	109.77	551	61	0	0	108
S2_21.fas	124412015	1524943	81.58	650	41	180	18	76
S2_31.fas	15561783	141477	110	680	61	0	0	108

# Sensibilidad de la predicción taxonómica

Brady and Salzberg

Page 12



**Figure 1. Percent accuracy of Phymm, with species-level matches masked, for read lengths from 100–1000 bp**

Colored dots show classification accuracy reported for PhyloPythia at 1000 bp for genus-through phylum-level predictions, and for CARMA at 100 bp (as a percentage of the entire input data set) for genus-and phylum-level predictions.

## MG-RAST

	#familias	#géneros
Refseq	524	994
RDP	329	757
SSU	458	1010
LSU	519	970

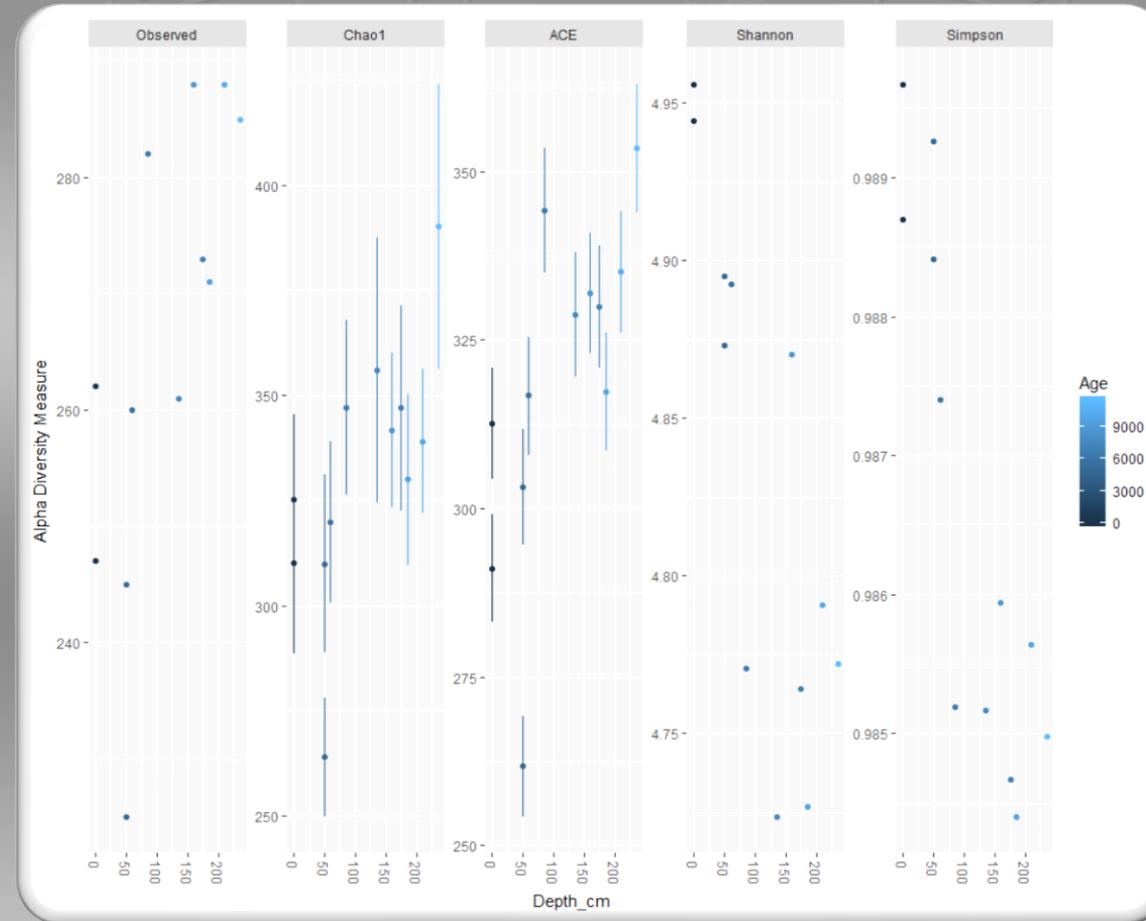
# Bases de datos utilizadas para anotación

- RefSeq: genomas completos
- RDP: basa en 16S para Bacterias y Arqueas y 28S para hongos
- LSU: Sub-unidad grande ribosomal.
- SSU: Sub unidad pequeña ribosomal
- Subsistemas: evalúa funciones de proteínas

# Análisis: base de datos (R, phyloseq)

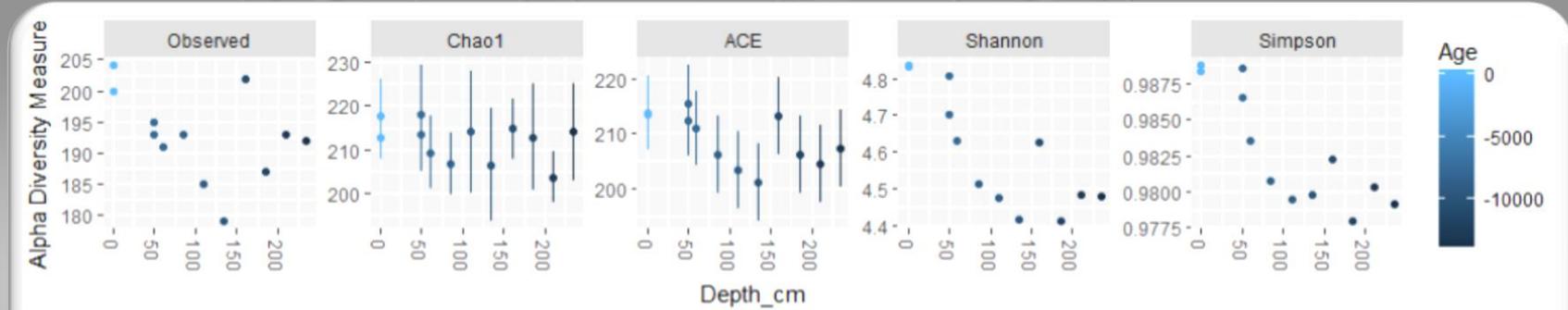
- Riqueza por familia total
- Riqueza familia separado por dominio
- Riqueza género separado por dominio
- Abundancia por phylum separado por dominio (graf. Barras)
- Abundancia por género por dominio (heatmap)
- Agrupación de muestras por 3 índices de diversidad  $\beta$  (Cluster)
- Ordenación de las muestras por la distancia Bray-curtis, por salinidad y temperatura (PCoA)
- Ordenación taxonómica por phylum (PCoA)

# LSU: análisis diversidad/riqueza familia

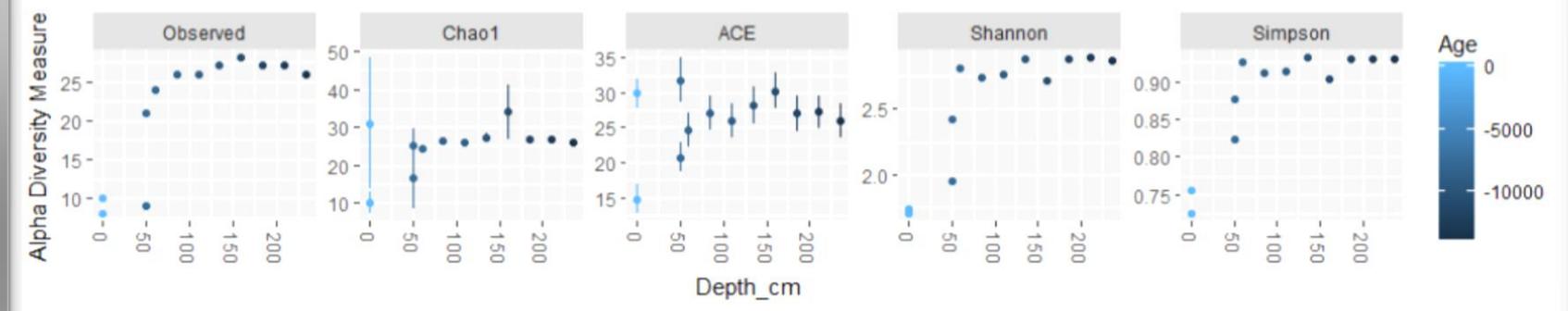


# LSU: análisis diversidad/riqueza familia

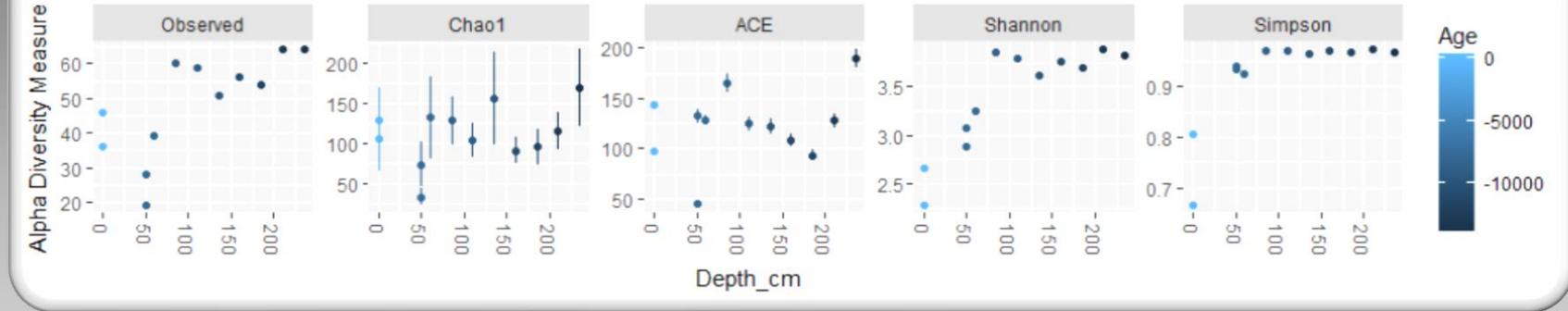
Bacteria



Archea

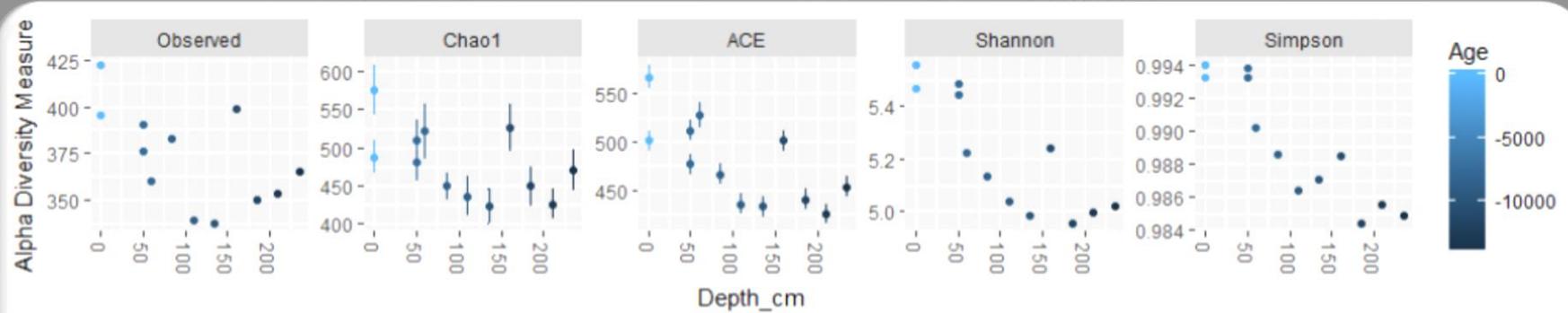


Eukaryota

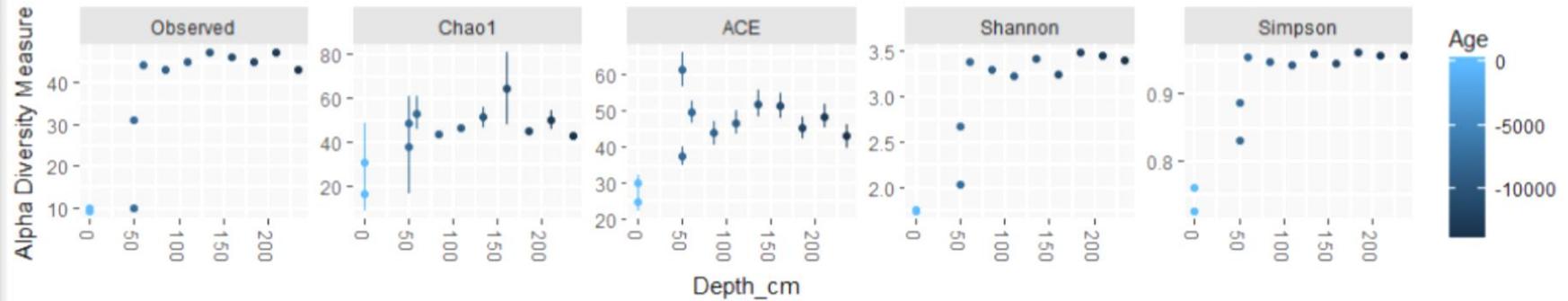


# LSU: análisis diversidad/riqueza género

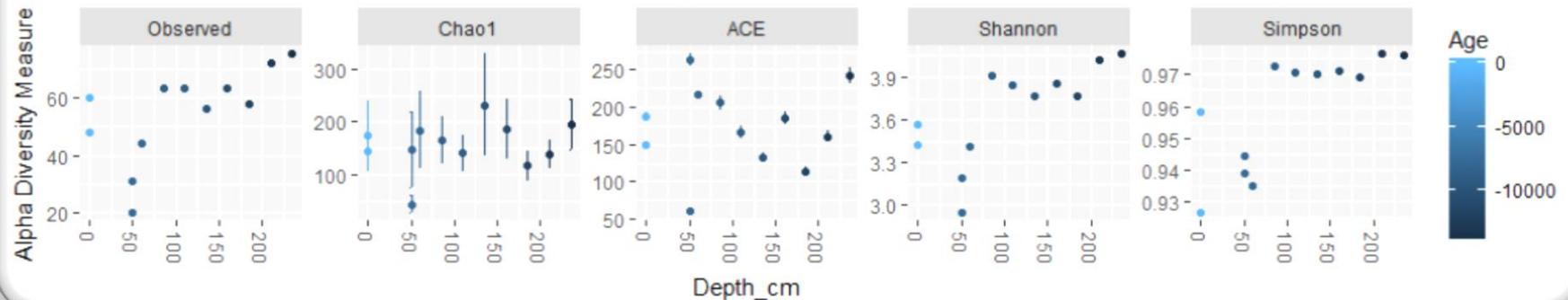
Bacteria



Archea

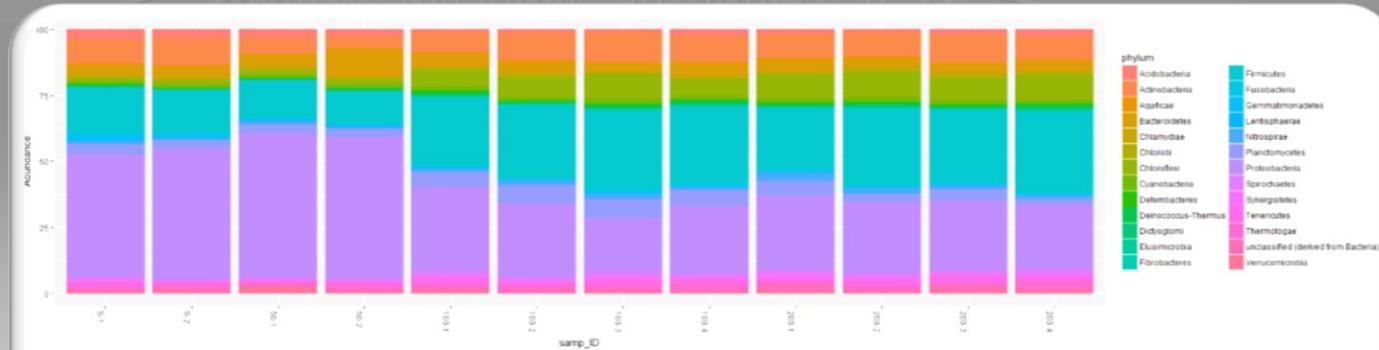


Eukaryota

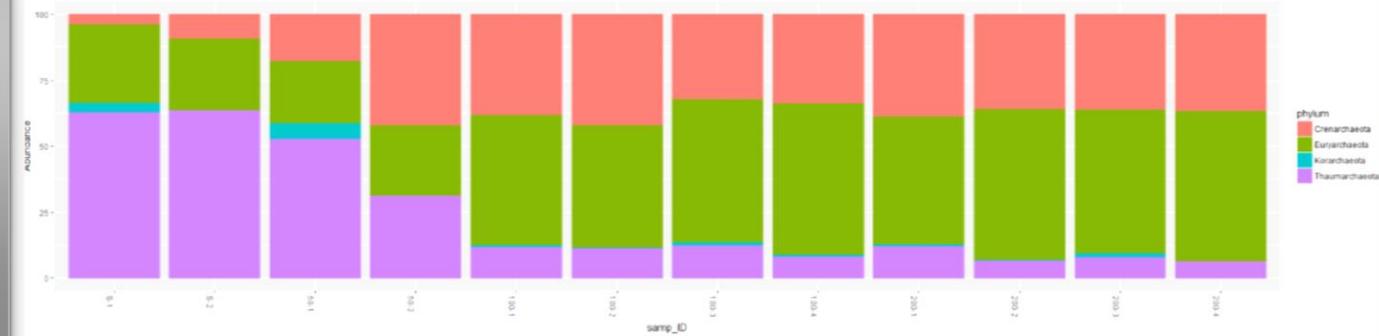


# LSU: análisis abundancia

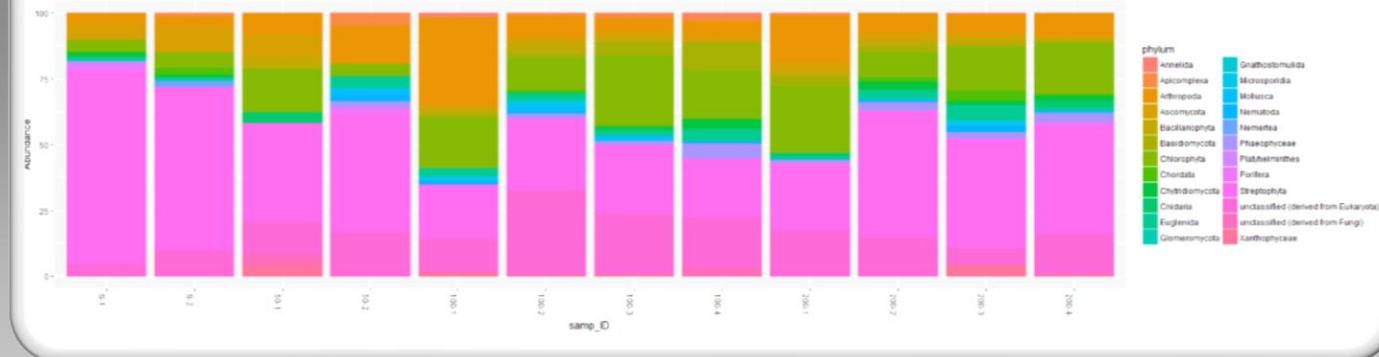
Bacteria



Archea

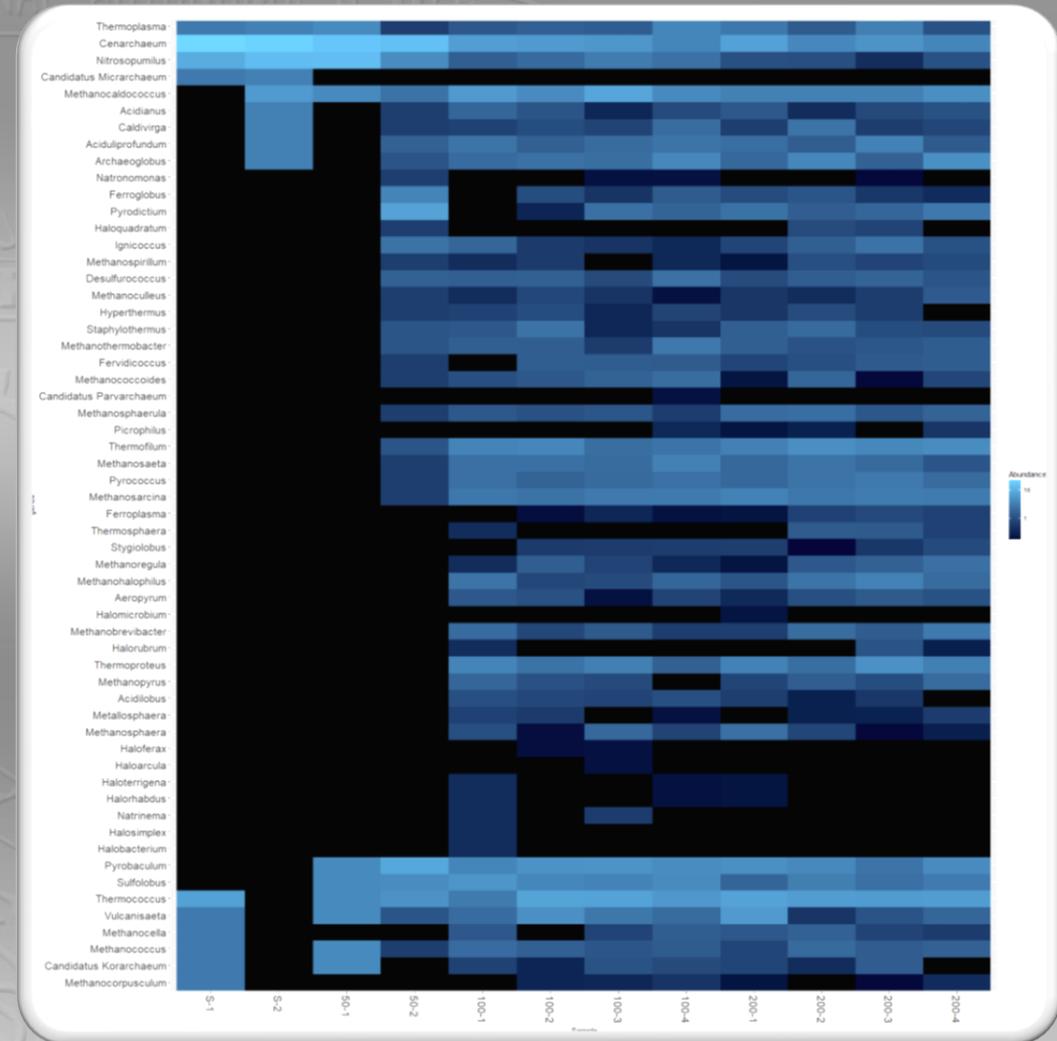


Eukaryota

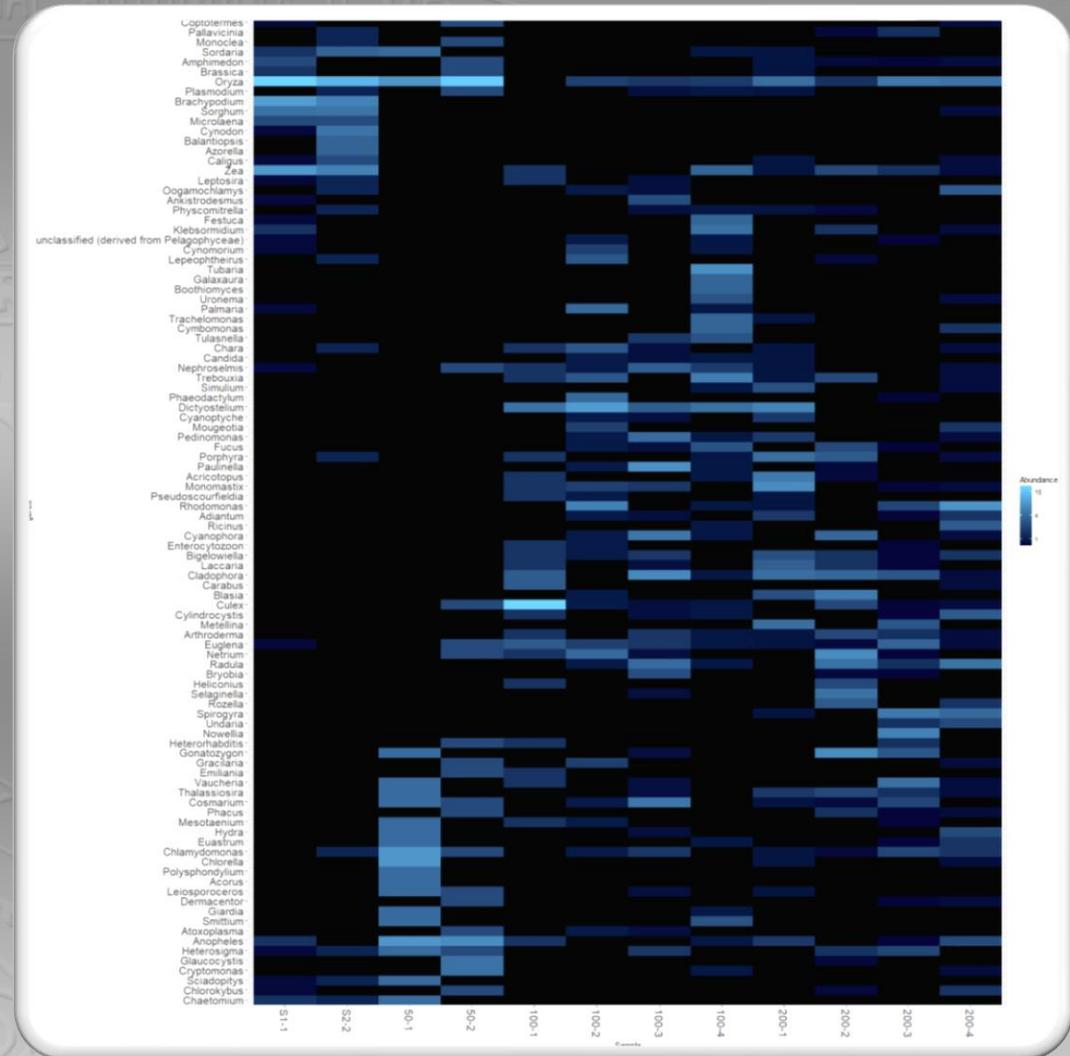




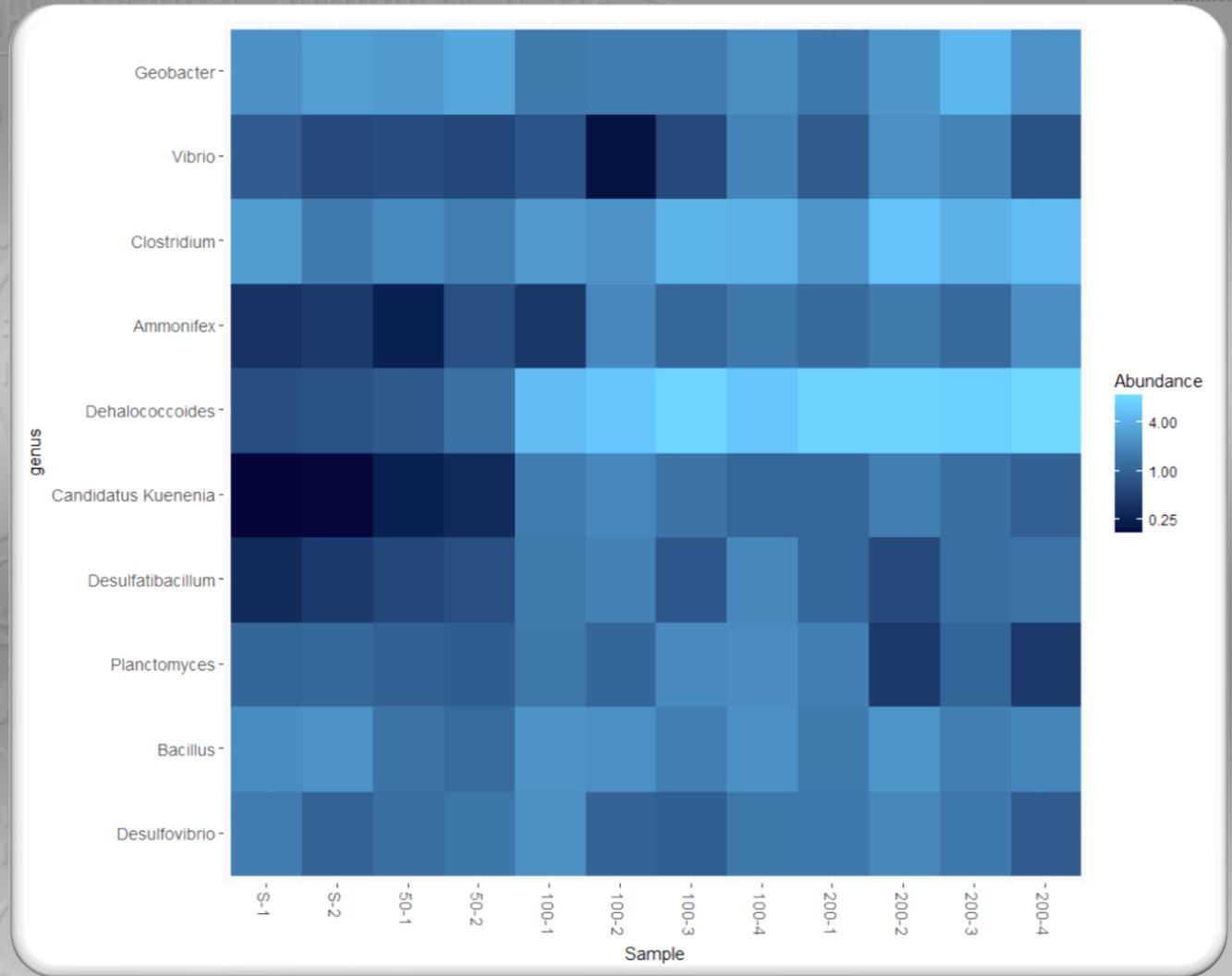
# LSU: Heatmap top\_100 gen\_archea



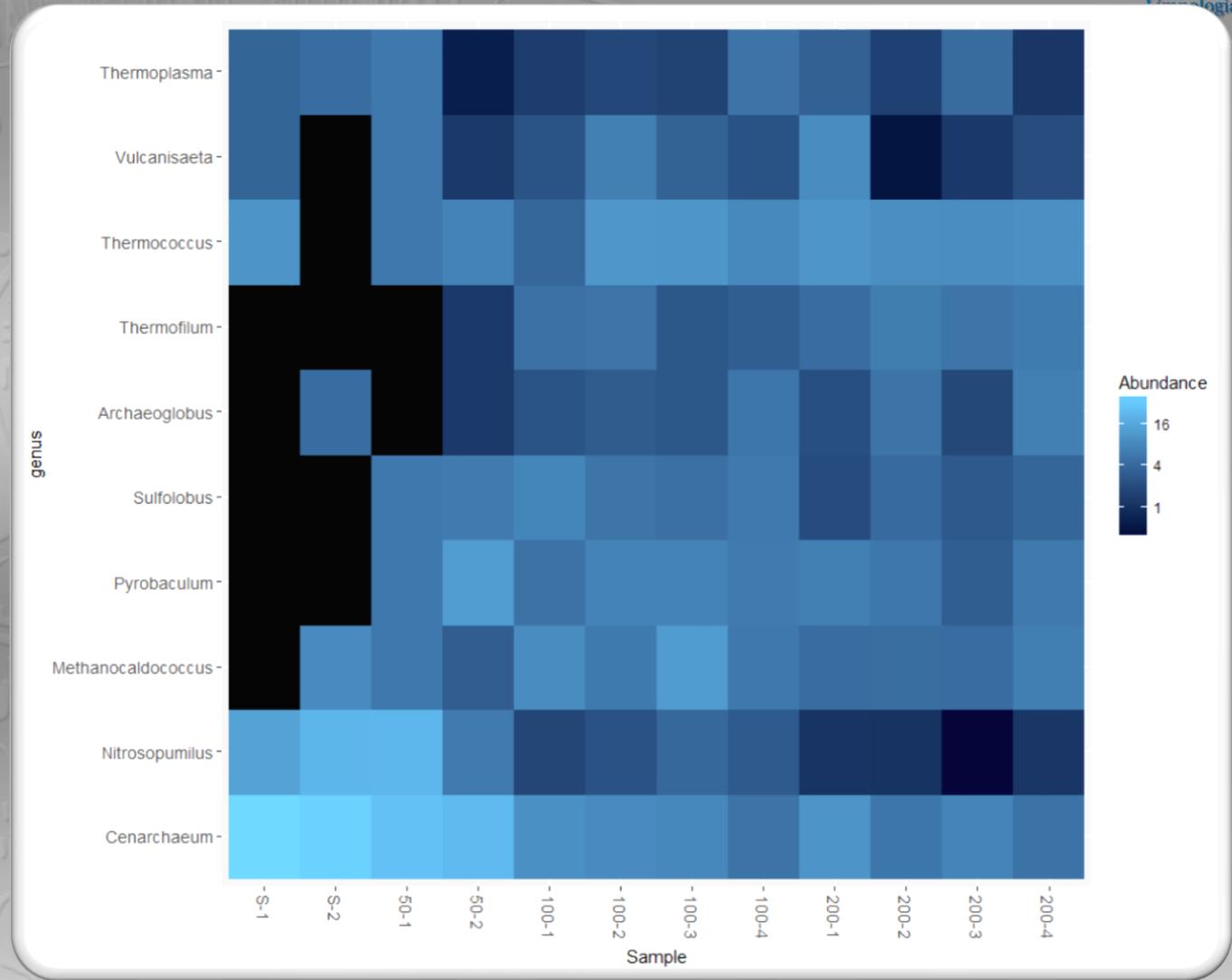
# LSU: Heatmap top\_100 gen\_eukarya



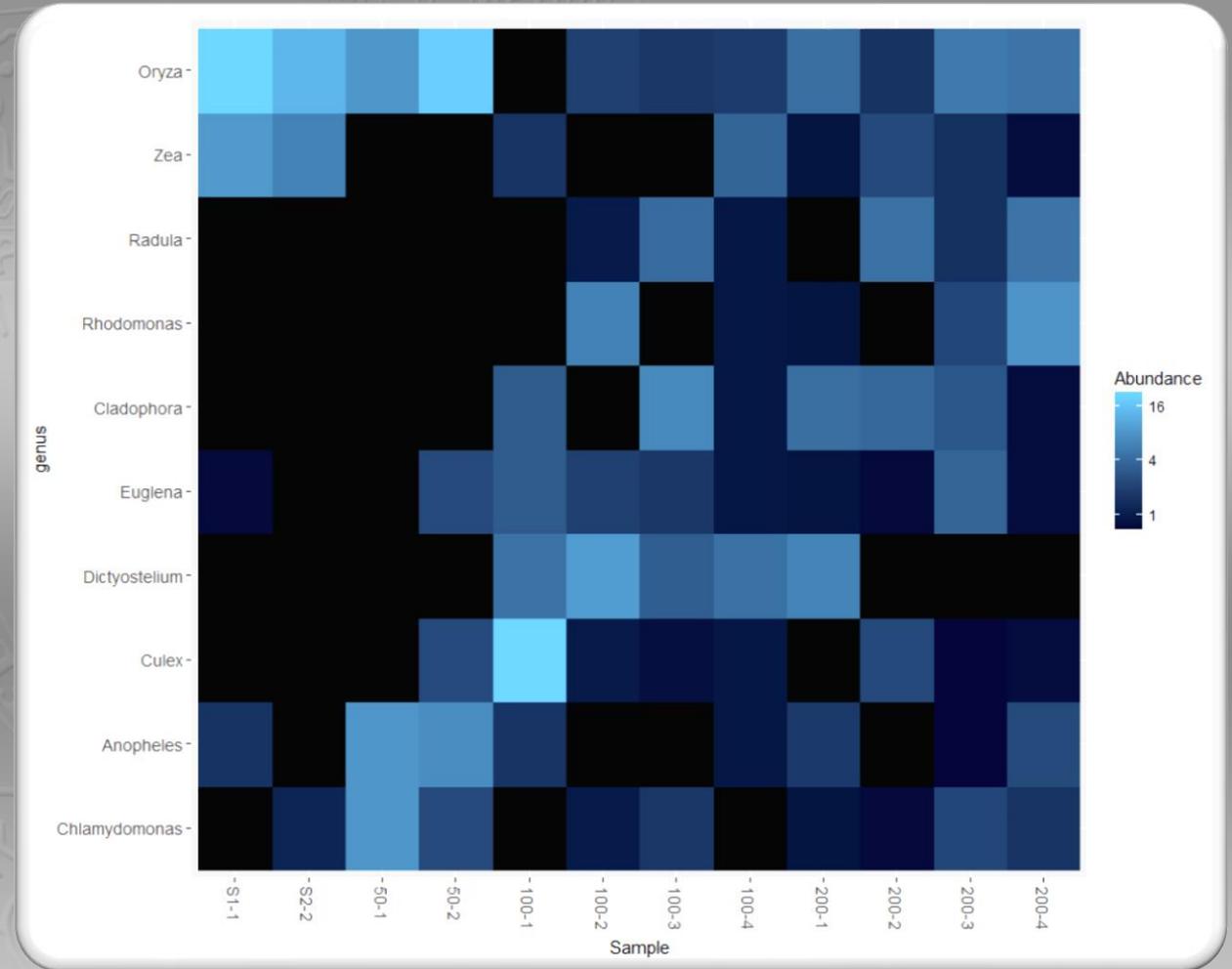
# LSU: Heatmap top\_10 gen\_bacteria

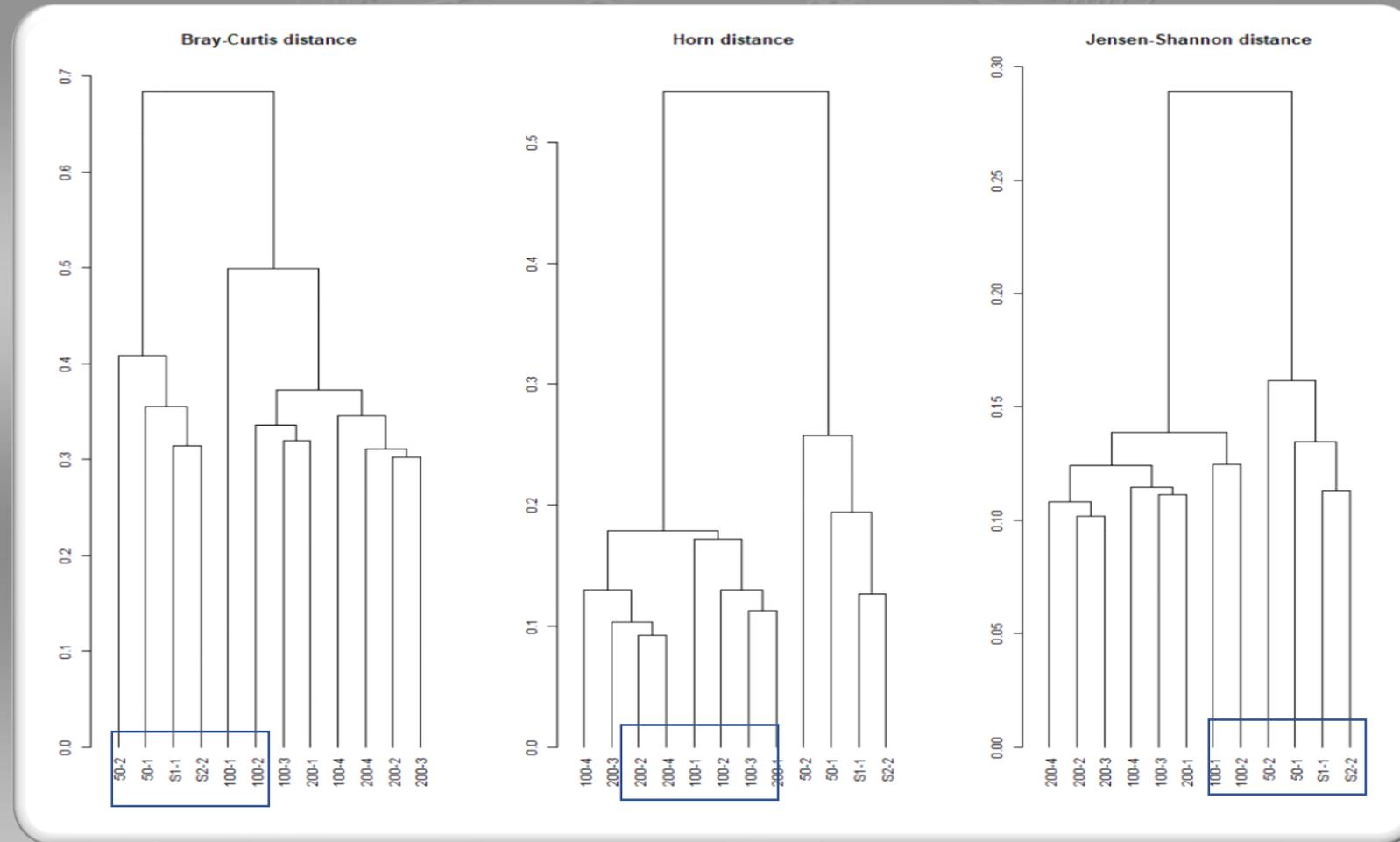


# LSU: Heatmap top\_10 gen\_archea

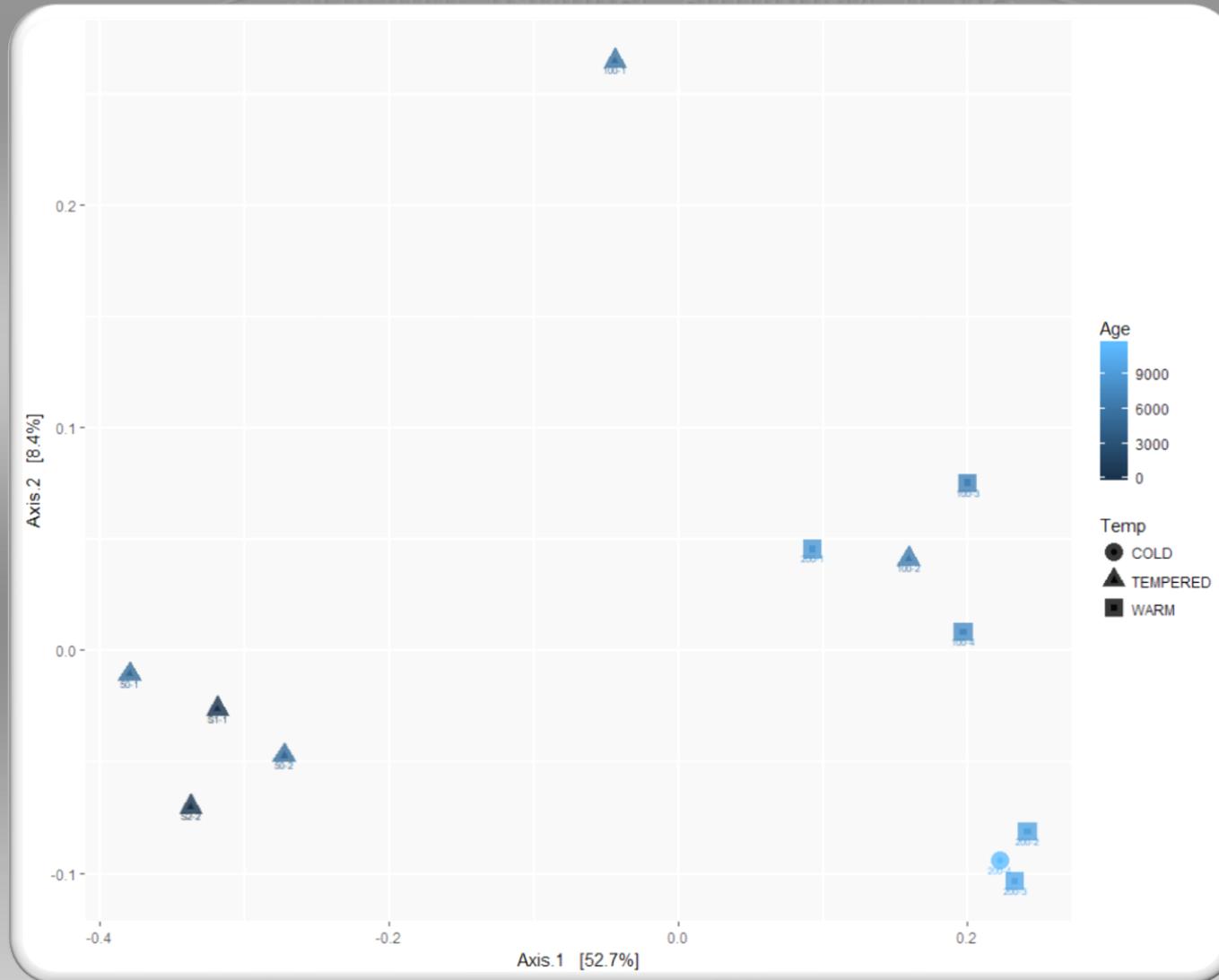


# LSU: Heatmap top\_10 gen\_eukarya





# PCoA análisis Bray-Courtis distance por temp

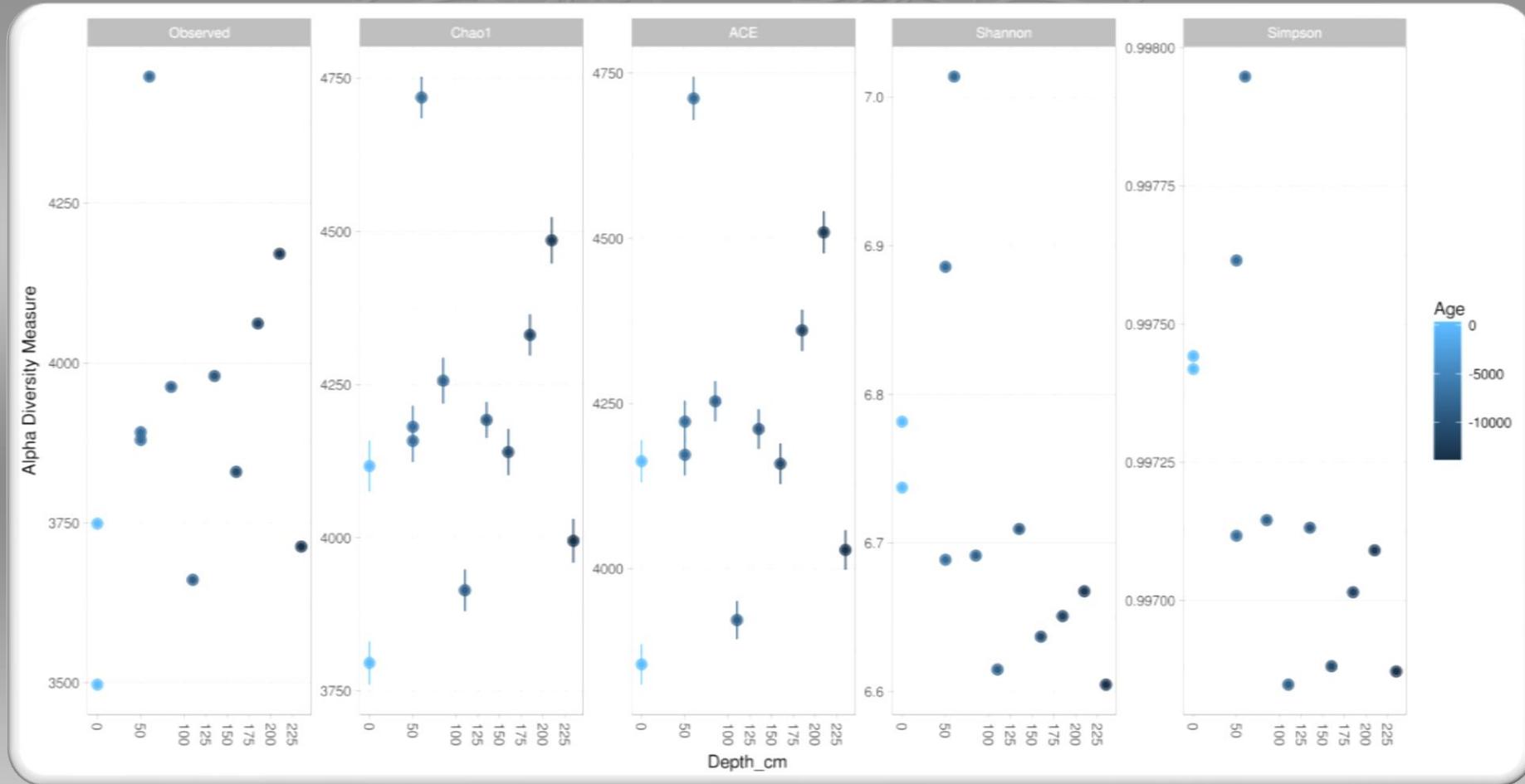




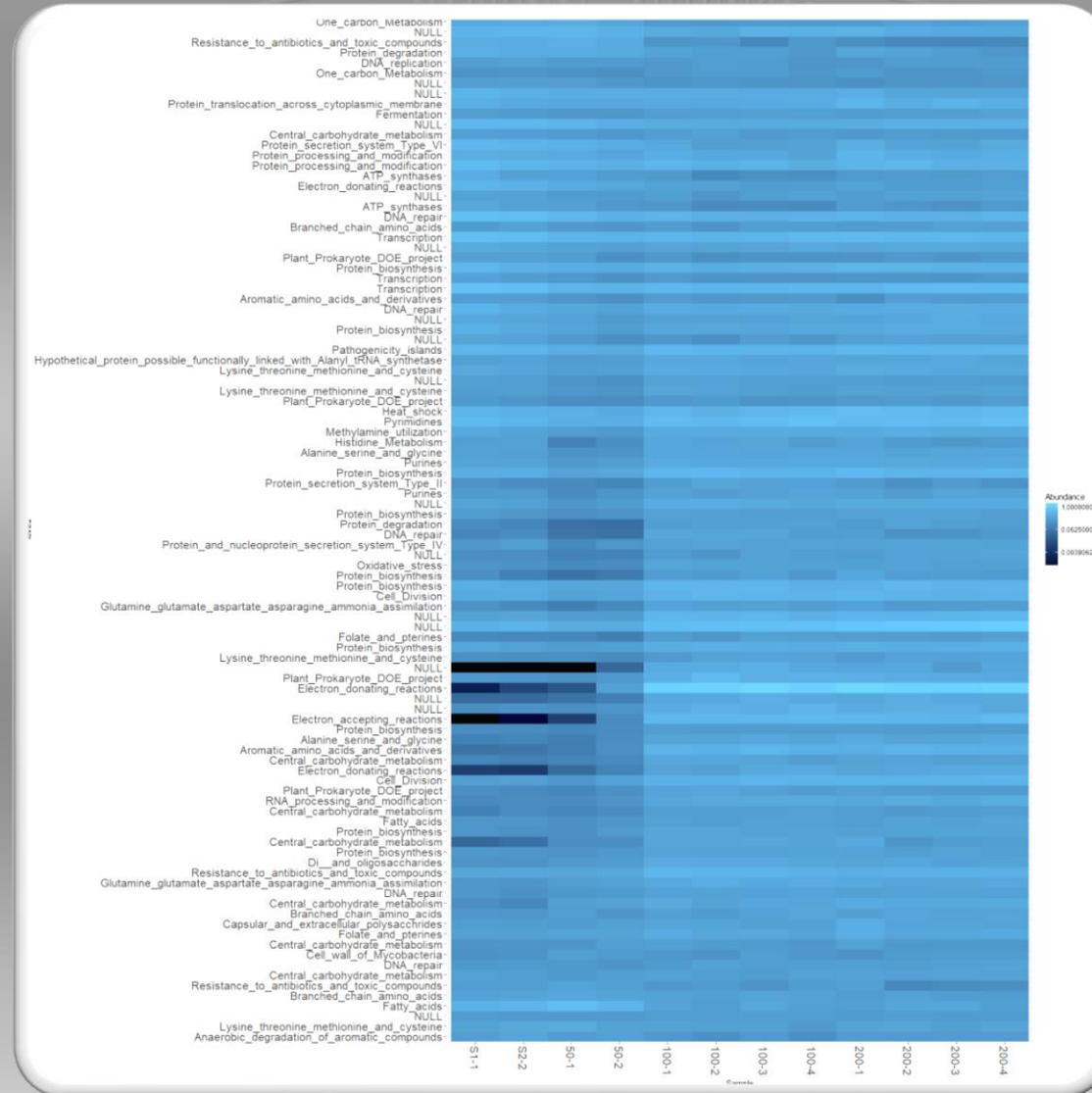
# LSU:PCoA análisis taxa



# Subsistemas: funciones de proteínas



# Heatmap funciones



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

спасибо 谢谢  
GRACIAS

THANK YOU

ありがとうございました MERCI

DANKE धन्यवाद

شُكراً OBRIGADO