



CBIB

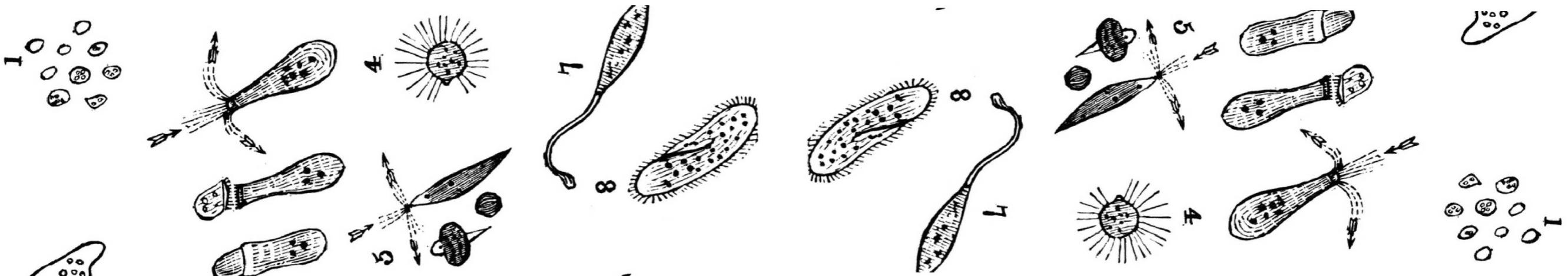
CENTER FOR BIOINFORMATICS
& INTEGRATIVE BIOLOGY

Metagenómica

> parte 2 <

www.castrolab.org
www.cbib.cl

Katterinne N. Mendez, M.Sc.
Eduardo Castro, PhD
Universidad Andrés Bello
30 de octubre de 2018



Antes de comenzar...

Vamos a descargar el material para el **taller de metagenómica**:



- Haz clic en el link a continuación y descarga la carpeta “Taller_Metagenomica”.

https://www.dropbox.com/sh/jgzcea3q1cl9wat/AABODW9DcdhIU_LKU-vrCtwga?dl=0

En la clase pasada...

Metataxonómica

- Secuenciación de un gen marcador que se pueda asociar a taxonomía (16S).
- Qué hay en la muestra? (perfil taxonómico)

Metagenómica

- Secuenciación de todo el DNA presente en una muestra.
- Qué hay en la muestra? (perfil taxonómico)
- Qué potencialmente hacen? (perfil funcional)

Metatranscriptómica

- Secuenciación de todo el RNA presente en una muestra.
- Generar un perfil de expresión de genes en la comunidad microbiana.
- Qué genes se expresan a nivel transcripcional y en qué medida?

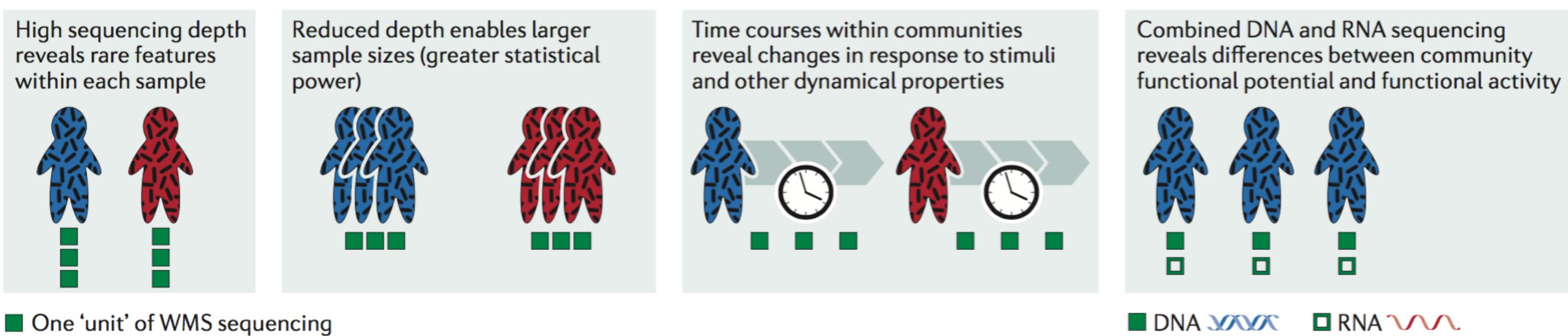
**16S (metataxonómica),
metagenómica, y
metatranscriptómica**

No son excluyentes sino complementarios

Presupuesto fijo

Profundidad vs. tamaño muestral

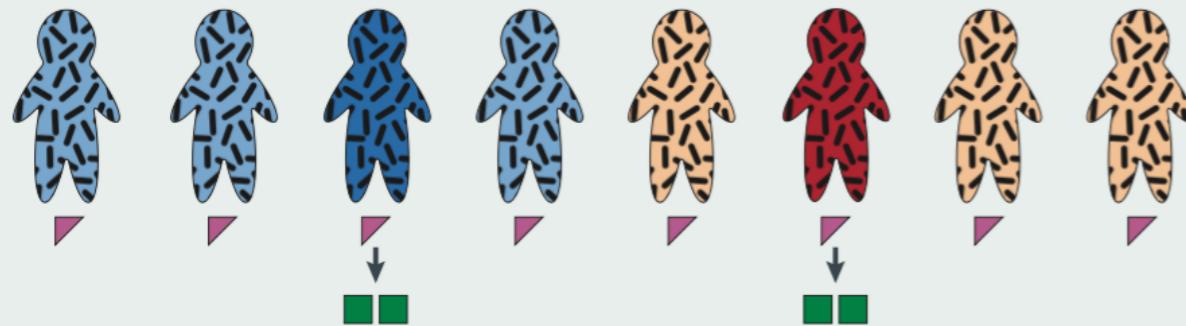
a Fixed sequencing budget



Profundidad = Detección de mutaciones raras
Mayor número de muestras = Mayor poder estadístico
Muestras temporales = Detección de cambios a estímulos
DNA y RNA = Diferencias entre potencial y actividad

Diseño longitudinal

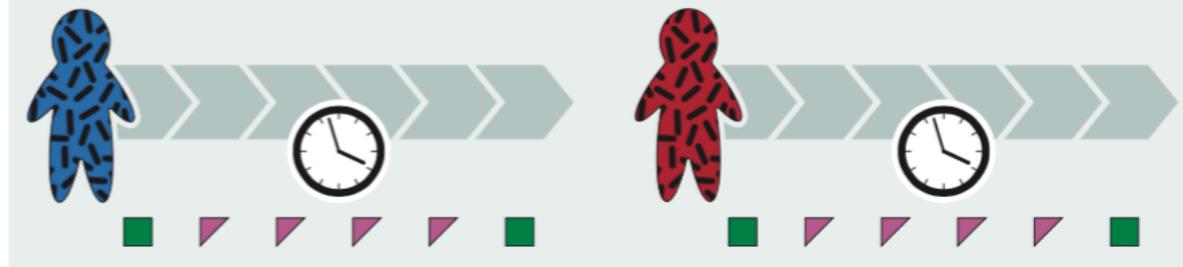
In a tiered study, many samples are initially surveyed by amplicon sequencing; later, a subset of representative or extreme samples are explored in greater detail by WMS sequencing



■ One 'unit' of amplicon sequencing

■ = ■■■■■

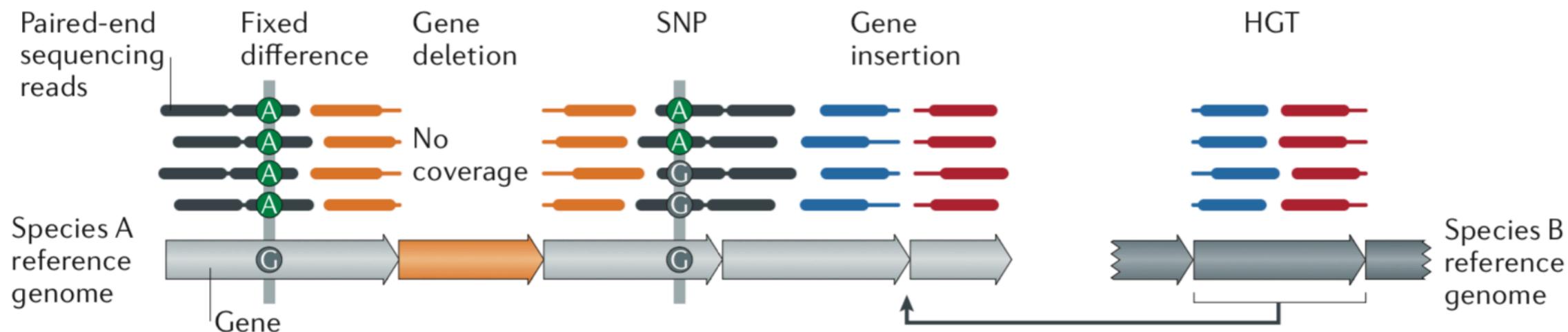
In time-course studies, amplicon sequencing can be applied to survey a large number of internal time points, while WMS sequencing can be used to dissect a subset of time points (e.g. the first and last) in greater detail



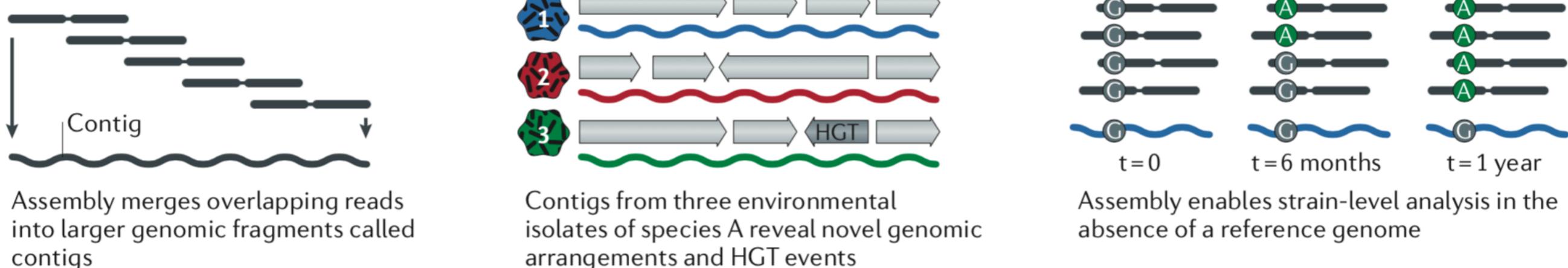
Dos etapas, lo mejor de ambos mundos
16S → Más muestras - primera etapa
Metagenómica → más profundidad - segunda etapa

Variantes

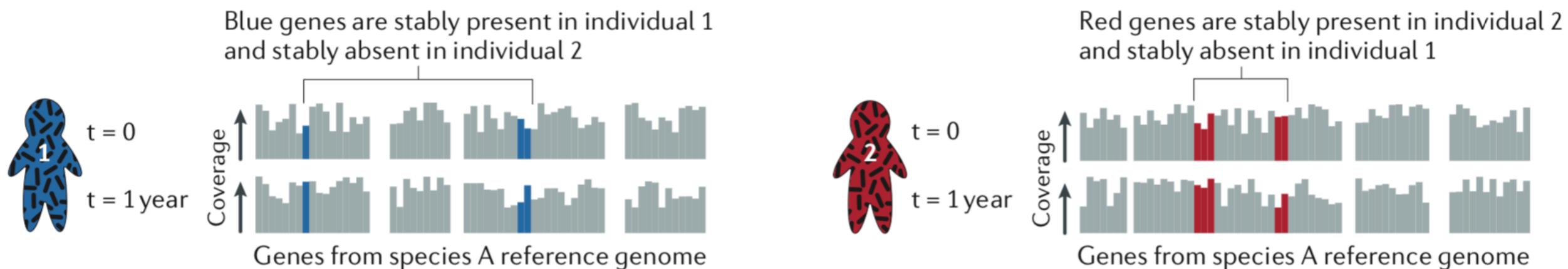
a Detecting strain variation



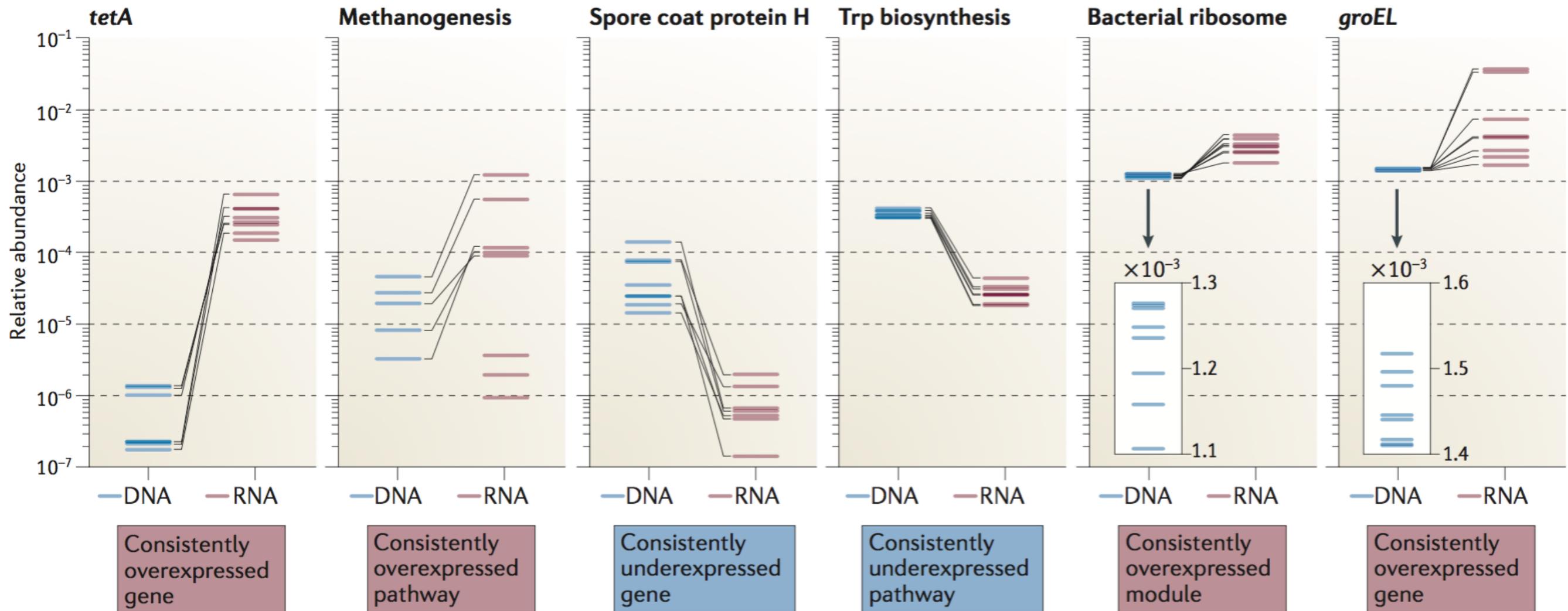
b Using metagenomic assembly



c Longitudinal analysis



Metagenómica y Metatranscriptómica



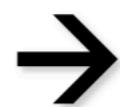
Número de copias de un gen vs. expresión
expresión diferencial de genes, rutas metabólicas y módulos

Taller Metagenómica



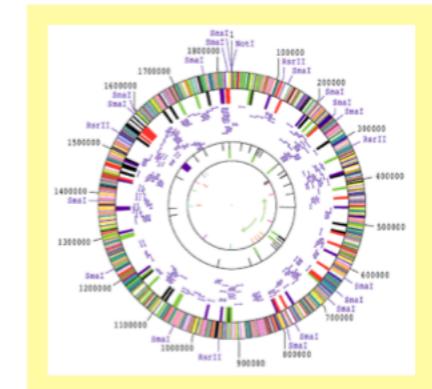
in situ

- Field collection
- Geochemistry
- Sample analysis
- Sample preservation



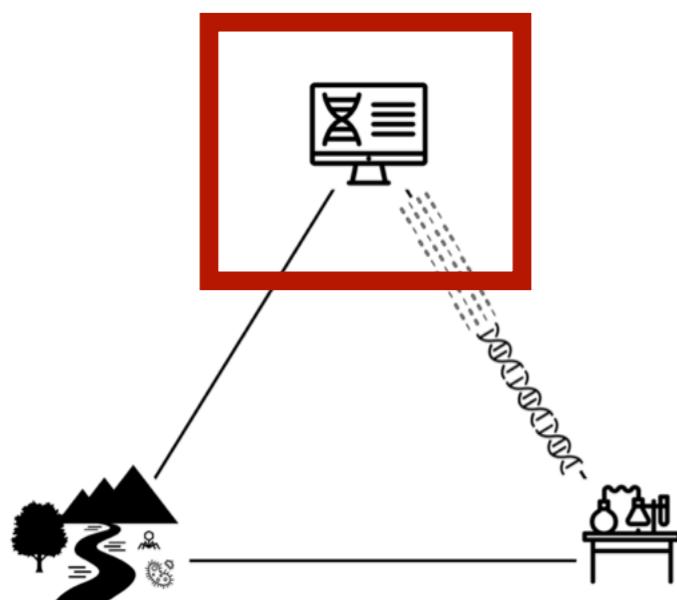
ex vivo

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



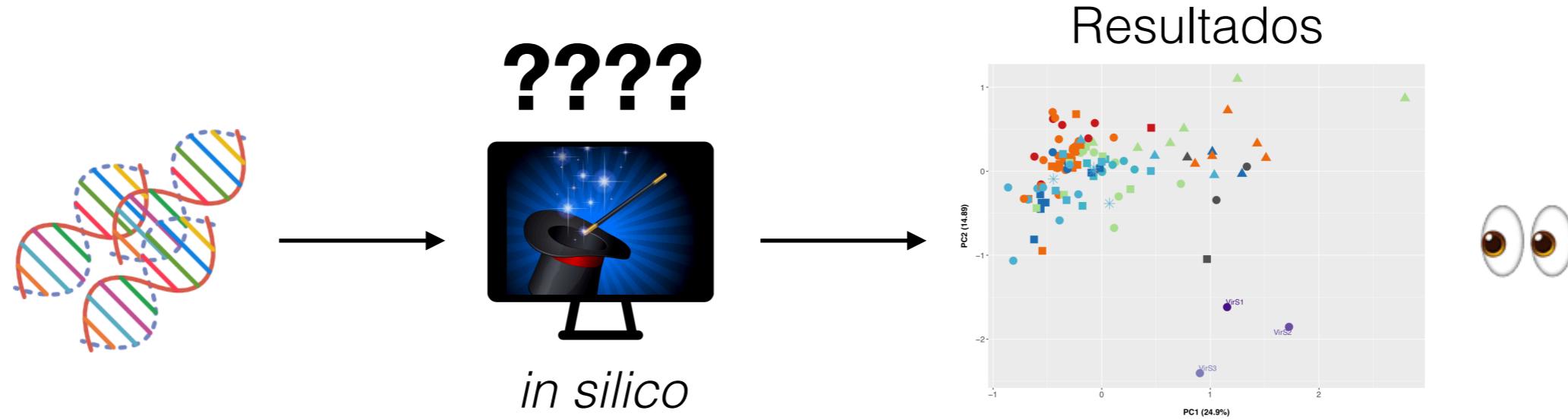
in silico

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



Taller Metagenómica

¡Adiós a la caja negra de los análisis bioinformáticos!



... por qué?

“You wouldn’t go into the laboratory and perform a polymerase chain reaction without a basic understanding of the method. Why would you do the same with a computational analysis?”

So you want to be a computational biologist?

Nick Loman ✉ & Mick Watson ✉

Nature Biotechnology 31, 996–998 (2013) | Download Citation ↴

<https://www.nature.com/articles/nbt.2740>

Taller Metagenómica

1.- Dirígete a la carpeta del taller que descargaste:
“Taller_Metagenomica”

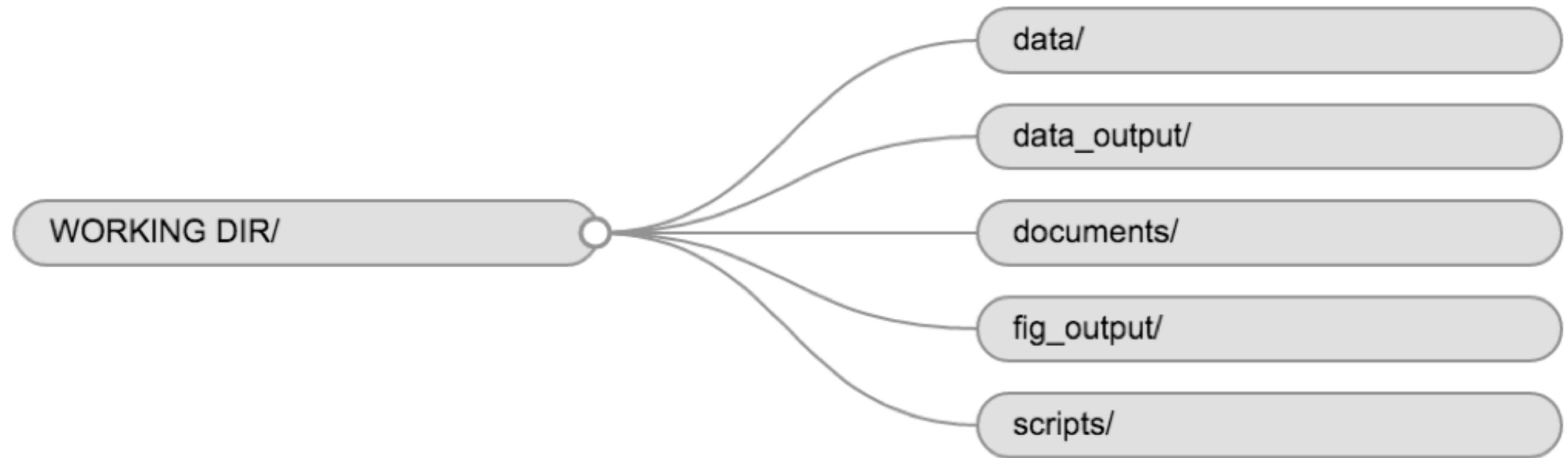


Taller Metagenómica

- Trabajando *in silico*



- Directorio de trabajo (*working directory*)

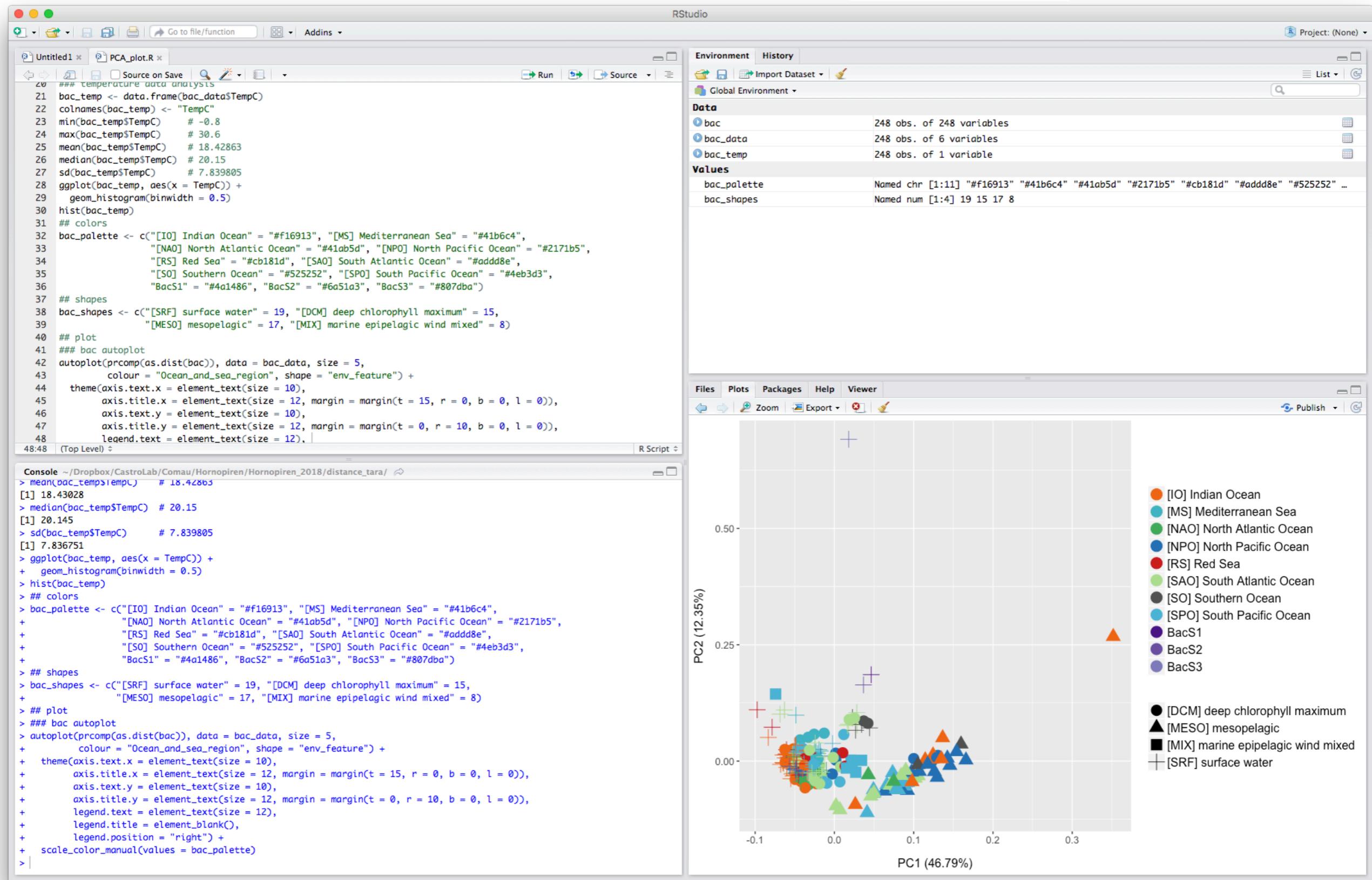


Taller Metagenómica

2.- Haz doble clic en el archivo:
“Taller_Metagenomica/ambiente-de-trabajo.html”.



Taller Metagenómica

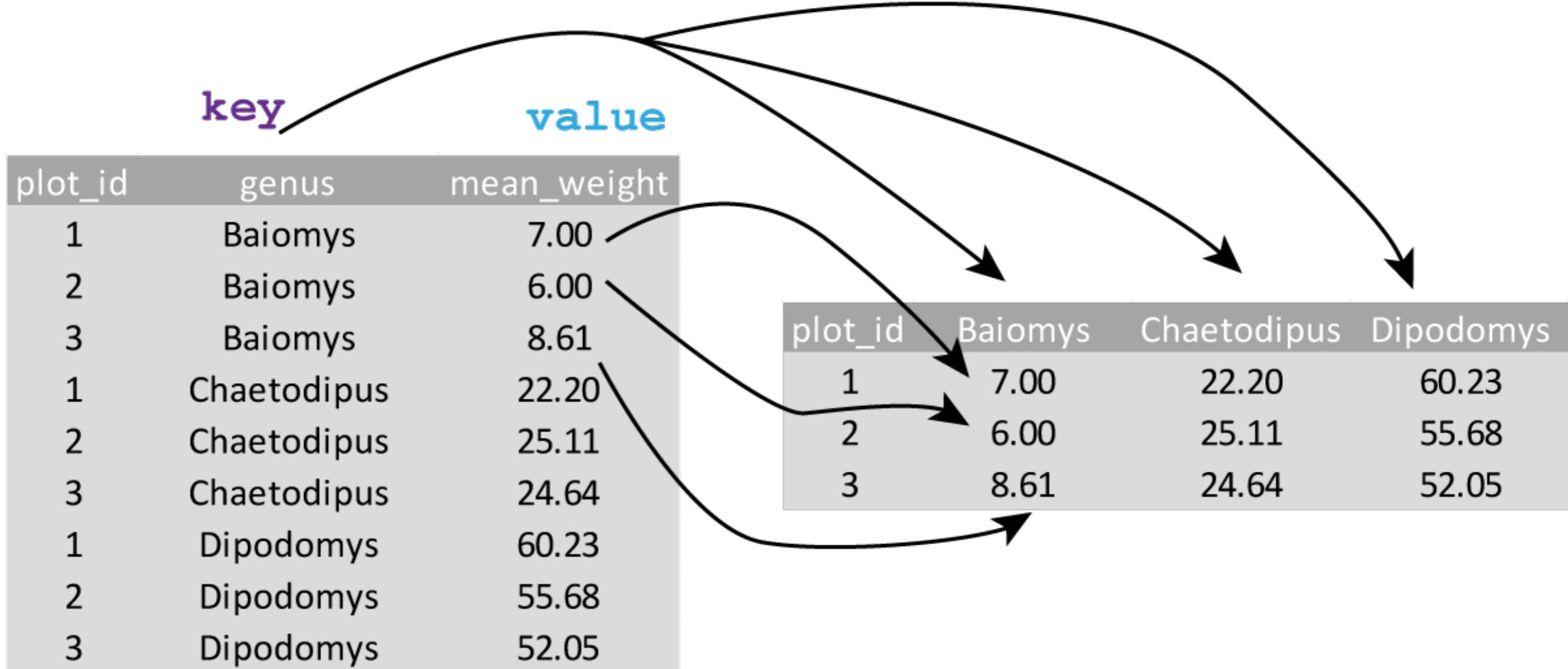


“Taller_Metagenomica/scripts/PaquetesR_configuracion.R”

```
PaquetesR_configuracion.R UNREGISTERED
Taller_paquetes.R PaquetesR_configuracion.R Workshop_UCN.Rmd
1 ### Taller Metagenómica: Configuración ###
2
3 # Definir paquetes a instalar
4 .cran_packages <- c("knitr", "qtl", "bookdown", "magrittr", "plyr", "ggplot2", "grid", "gridExtra", "tidyverse", "c
5 .bioc_packages <- c("phyloseq", "dada2", "DECIPHER", "phangorn", "ggpubr", "BiocInstaller", "DESeq2", "genefilter",
6 .git_packages <- c("btools", "fantaxtic", "ampvis2", "tsnemicrobiota", "microbiome")
7
8 # Instalar paquetes de CRAN
9 .inst <- .cran_packages %in% installed.packages()
10 if(any(!.inst)) {
11   install.packages(.cran_packages[!.inst])
12 }
13
14 # Instalar paquetes de Bioconductor
15 .inst <- .bioc_packages %in% installed.packages()
16 if(any(!.inst)) {
17   source("http://bioconductor.org/biocLite.R")
18   biocLite(.bioc_packages[!.inst], ask = F)
19 }
20
21 # Instalar paquetes de GitHub
22 .inst <- .git_packages %in% installed.packages()
23 if(any(!.inst)) {
24   devtools::install_github('twbattaglia/btools')
25   devtools::install_github("gmteunisse/Fantaxtic")
26   devtools::install_github("microbiome/microbiome")
27   remotes::install_github("MadsAlbertsen/ampvis2")
28   install_github("opisthokonta/tsnemicrobiota")
29 }
30
31 # Cargar paquetes a la actual sesión de R
32 sapply(c(.cran_packages, .bioc_packages, .git_packages), require, character.only = TRUE)
33 |
```

Line 33, Column 1 Tab Size: 4 R

spread



data.frame

column with new
variable names

column of values
for new
variables

```
surveys_gw %>% spread(key = genus, value = mean_weight)
```

gather

The diagram illustrates the process of transforming a wide data frame into a long data frame using the `gather` function. It shows two data frames: a wide data frame on the left and a long data frame on the right.

Wide Data Frame:

plot_id	genus	mean_weight
1	Baiomys	7.00
2	Baiomys	6.00
3	Baiomys	8.61
1	Chaetodipus	22.20
2	Chaetodipus	25.11
3	Chaetodipus	24.64
1	Dipodomys	60.23
2	Dipodomys	55.68
3	Dipodomys	52.05

Long Data Frame:

plot_id	Baiomys	Chaetodipus	Dipodomys
1	7.00	22.20	60.23
2	6.00	25.11	55.68
3	8.61	24.64	52.05

Annotations explain the variables:

- key**: Points to the `genus` column in the wide data frame.
- value**: Points to the `mean_weight` column in the wide data frame.
- variable whose values are column names**: Points to the `genus` column in the long data frame.
- variable whose values are spread over columns**: Points to the `mean_weight` column in the long data frame.
- don't use this value of this variable**: Points to the `plot_id` column in the long data frame.

`data.frame`

```
surveys_spread %>% gather(key = genus, value = mean_weight, -plot_id)
```