

PEC1

cserrano

2025-03-25

En este trabajo se realizará un análisis exploratorio de un estudio metabolómico, escogido desde la base de datos **Metabolomics Workbench**. Por otro lado, usaremos la librería **metabolomicsWorkbenchR** para extraer los datos directamente desde la web y **SummarizedExperiment** para manejar los datos de manera estructurada. Ambas librerías se encuentran en el repositorio **Bioconductor**. Para utilizar estas librerías, primero asegurarse de tener instalado Bioconductor y luego instalar **metabolomicsWorkbenchR** y **SummarizedExperiment** como indica en sus respectivos links.

En esta sección, importamos las librerías necesarias que utilizaremos a lo largo de este trabajo

```
library(metabolomicsWorkbenchR)
library(SummarizedExperiment)
library(knitr)
library(dplyr)
library(tidyr)
library(ggplot2)
```

La librería **metabolomicsWorkbenchR** permite realizar consultas a directamente a su página web. Aquí se utiliza la función `do_query` para solicitar información de estudios, precisamente con la palabra **Bacterial**.

```
estudios = do_query(
  context = 'study',
  input_item = 'study_title',
  input_value = 'Bacterial',
  output_item = 'summary'
)
```

El resultado fue almacenado en la variable `estudios`, la cual se puede visualizar como una tabla. Aquí se muestran los primeros 10 resultados de la consulta.

```
kable(head(estudios))
```

study_id	study_title	species	institute	analysis_type	study_files	publications	first_name	email	submit_date	study_status	species
ST003711	Gpp and DksA play crucial role in reducing the efficacy of -lactam antibiotics by modulating bacterial membrane permeability	Vibrio cholerae	Translational Health Science And Technology Institute (THSTI)	LC-MS	72	NA	NA	NA	NA	NA	NA

study_id	study_title	species	institute	analysis_type	number_of_samples	study_type	department	last_name	first_name	submit_date	study_base	journal	species
ST003521	Metabolic Profiling Unveils Enhanced Antibacterial Synergy of Polymyxin B and Teixobactin against Multi-Drug Resistant Acinetobacter baumannii	Acinetobacter baumannii	Montash University	LC-MS	58	NA	NA	NA	NA	NA	NA	NA	NA
ST003389	Uptake of Glucose Homeostasis by Bacterial Infection Orchestrates Host Innate Immunity Through NAD ⁺ /NADH Balance	Mus musculus	Northwest A&F University	LC-MS	24	NA	NA	NA	NA	NA	NA	NA	NA
ST003187	Targeted plasma metabolomics on bacterial culture supernatants	Escherichia coli	Harvard University	LC-MS	7	NA	NA	NA	NA	NA	NA	NA	NA
ST002973	Define the through-filter recovery of metabolites extracted from a complex bacterial medium		Duke University	LC-MS	20	NA	NA	NA	NA	NA	NA	NA	NA
ST002921	Metabolomic Characteristics of Nontuberculous Mycobacterial Pulmonary Disease	Homo sapiens	Seoul National University College of Medicine and Hospital	LC-MS	418	NA	NA	NA	NA	NA	NA	NA	NA

Teniendo estas alternativas de interés, se decidió seleccionar el estudio ST003521. La justificación de este estudio es mi relación al estudio de microbiología y genes de resistencia. Además, este estudio pudo ser obtenido por los métodos que aquí se muestran y asociarlo con una publicación para mayor información (Se pobraron otros estudios con la misma metodología y varios fallaban). Para mayor información, se puede visitar directamente el estudio **ST003521** en su página web.

A continuación se procede a obtener los datos del estudio. Primero se obtendrá el summary del estudio. En la función `do_query`, se especifica el estudio que al que se quiere acceder, mientras que en `output_item` se indica “summary”

Summary de estudio seleccionado

```
summary <- do_query(
  context = "study",
  input_item = "study_id",
  input_value = "ST003521",
  output_item = "summary"
)
```

Dentro de los campos disponibles en `summary` se encuentran: “study_id”, “study_title”, “species”, “institute”, “analysis_type”, “number_of_samples”, “study_type” “department”, “last_name”, “first_name”,

“email”, “phone”, “submit_date”, “study_summary”, “subject_species”. En este caso en particular, solo están disponible los siguientes campos.

```
cat("ID de estudio: ", summary$study_id,"\n")
```

```
## ID de estudio: ST003521
```

```
cat("Nombre de estudio: ", summary$study_title,"\n")
```

```
## Nombre de estudio: Metabolic Profiling Unveils Enhanced Antibacterial Synergy of Polymyxin B and Te
```

```
cat("Especies de estudio: ", summary$species,"\n")
```

```
## Especies de estudio: Acinetobacter baumannii
```

```
cat("Instituto del estudio: ", summary$institute,"\n")
```

```
## Instituto del estudio: Monash University
```

```
cat("Tipo de análisis: ", summary$analysis_type,"\n")
```

```
## Tipo de análisis: LC-MS
```

```
cat("Número de muestras: ", summary$number_of_samples,"\n")
```

```
## Número de muestras: 58
```

A continuación se obtienen los datos del estudio. Para ello se indica su ID, y en output_item se indicará salida de tipo “SummarizedExperiment”. De esta manera, los datos se obtendrán en formato del objeto **SummarizedExperiment**, el cuál permite contener mucha información del estudio de manera estructurada dentro del mismo objeto.

En este caso, al obtener los datos, se obtiene una lista con 2 elementos SummarizedExperiment. Estos se almacenarán en variable llamada se_list

Carga de Datos en clase SummarizedExperiment

```
se_list = do_query(  
  context = 'study',  
  input_item = 'study_id',  
  input_value = 'ST003521',  
  output_item = 'SummarizedExperiment'  
)
```

```
n=1  
for (se in se_list)  
{cat('Clase de elemento', n , ' : ', class(se), '\n')  
n = n+1}
```

Exploraremos cada una de las matrices que contiene el objeto SummarizedExperiment. A continuación, una imagen representativa obtenida directamente desde la **documentación de la librería**.

Observaremos la matriz ‘assay’. Lo haremos con el primer elemento de los datos, pero se pueden realizar las mismas acciones con ambos. Para ello utilizamos la funcion `assay()` y usamos `head` para visualizar los primeros resultados. Observamos también las dimensiones del dataset, donde obtenemos que este objeto consta de 513 filas y 58 columnas.


```
kable(head(colData(se1)))
```

	local_sample_id	study_id	sample_source	mb_sample_id	raw_data	Sample_source	raw_file_name
Blank_1	Blank_1	ST003521	BLANK	SA386750		Extraction blank	Blank_1
Blank_2	Blank_2	ST003521	BLANK	SA386751		Extraction blank	Blank_2
Blank_3	Blank_3	ST003521	BLANK	SA386752		Extraction blank	Blank_3
Blank_4	Blank_4	ST003521	BLANK	SA386753		Extraction blank	Blank_4
Blank_5	Blank_5	ST003521	BLANK	SA386754		Extraction blank	Blank_5
COM_1h	COM_1h_1	ST003521	Combination_1h	SA386726		Combination_1h	COM_1h_1

```
cat('Dimensiones:\n Filas: ',dim(colData(se1))[1], '\n', 'Columnas: ', dim(colData(se1))[2])
```

```
## Dimensiones:
## Filas: 58
## Columnas: 7
```

Por último, exploramos la metadata del estudio. Para ello usamos la función `metadata()`. Esto nos entrega una lista con información.

```
metadata(se1)
```

```
## $data_source
## [1] "Metabolomics Workbench"
##
## $study_id
## [1] "ST003521"
##
## $analysis_id
## [1] "AN005782"
##
## $analysis_summary
## [1] "HILIC POSITIVE ION MODE"
##
## $units
## [1] "peak height"
##
## $name
## [1] "ST003521:AN005782"
##
## $description
## [1] "Metabolic Profiling Unveils Enhanced Antibacterial Synergy of Polymyxin B and Teixobactin against Gram-negative Bacteria"
##
## $subject_type
## [1] NA
```