PEC1

cserrano

2025-03-25

En este trabajo se realizará un análsis 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))
```

stud y tuidly_title	species institute	$analy \textbf{sis} \underline{\textbf{m}} \textbf{ty} \underline{\textbf{ste}} \underline{\textbf{d}} \underline{\textbf{fd}} \underline{\textbf{e}} \underline{\textbf{spherical}} \underline{\textbf{firstmensail}} \\ \textbf{hose} bm \underline{\textbf{st}} \underline{\textbf{u}} \underline{\textbf{d}} \underline{\textbf{ste}} \underline{\textbf{spectral}} \textbf{spect$	ies
ST0(\$7ppGpp and DksA play	Vibrio Translational	LC- 72 NA NA NA NA NA NA NA NA	
crucial role in reducing the	cholera & Health	MS	
efficacy of -lactam	Science		
antibiotics by modulating	And		
bacterial membrane	Technology		
permeability	Institute		
	(THSTI)		

stud y tuidly_title	species institute		analysismtysteidfdespilustestitstnenaajiloosebmitudisstebjeet								
ST00M52abolic Profiling Unveils Enhanced Antibacterial Synergy of Polymyxin B and Teixobactin against Multi-Drug Resistant	Acinetolokom bau- Univ man- nii		LC- MS	58	NA	NA	NA	NA	NA NA NA	NA	NA
Acinetobacter baumannii ST0C33840ption of Glucose Homeostasis by Bacterial Infection Orchestrates Host Innate Immunity Through NAD+/NADH	mus- A&I		LC- MS	24	NA	NA	NA	NA	NA NA NA	NA	NA
Balance ST0031837geted plasma metabolomics on bacterial culture supernatants	Escheri &hia coli Univ		LC- MS	7	NA	NA	NA	NA	NA NA NA	NA	NA
STO(E2253) ine the through-filter recovery of metabolites extracted from a complex bacterial medium	Duk Univ		LC- MS	20	NA	NA	NA	NA	NA NA NA	NA	NA
ST00M924bolomic Characteristics of Nontuberculous Mycobacterial Pulmonary Disease	ens Univ	ional versity lege of licine	LC- MS	418	NA	NA	NA	NA	NA NA NA	NA	NA

Teniendo estas alternativas de interés, se decidio 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 especifíca 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"
)</pre>
```

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 Summarized Experiment.
Estos se almacenarán en variable llamada se_list

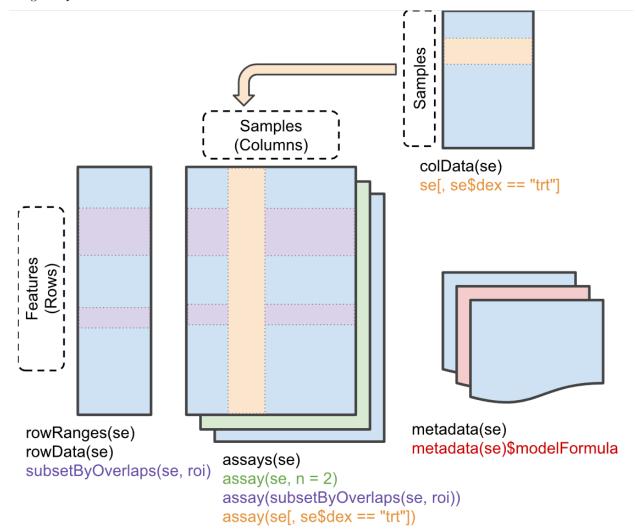
Carga de Datos en clase SummarizedExperiment

```
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}
```

```
## Clase de elemento 1 : SummarizedExperiment
## Clase de elemento 2 : SummarizedExperiment
```

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.

```
se1 <- se_list$AN005782
se2 <- se_list$AN005783

kable(head(assay(se1)))</pre>
```

BIBIBIBIBICACO COLON CONCENSOR COLON COLOR COLOR

N AN AN AN AN AA 8.0670357285415221900201214251293570275542511855570102815749475844641574073895322457355607557448174522874888601881852852555 N AN AN AN AN AA 9.552253611388972545070365454271570571601702081024713663848450311864284275458734728478847850367478847240756 BIBIBIBIRANO CONTROL C

Dimensiones:
Filas: 513
Columnas: 58

Ahora observaremos específicamente la matriz de Features (o Rows). Accedemos a ellas mediante la función rowData(). Al igual que antes, solo se mostrarán los primeros resultados para evitar extender demasiado el documento.

Aquó observamos que las dimensiones de esta matriz es de 513 filas y 3 columnas. Aquí se muestran los valores de metabolitos denotados con ids pero también con su nombre químico. Las columnas que se encuentran son: "metabolite_name", "metabolite_id" y "refmet_name"

kable(head(rowData(se1)))

metabolite_name	$metabolite_$	id refmet_name
ME91751510,11-dihydro-20-trihydroxy-leukotriene B4	ME917515	
ME9174231-(14-methyl-pentadecan oyl)-2-(8-[3]-ladderane-oct anyl)-2-(8-[3]-ladderane-oct anyl)-2-(8-[ME917423	
sn-glycerol		
ME9171231,2-dihexadecanoyl-sn-glycero-3-phosphosulfocholine	ME917123	
ME9172241,3,5-trimethoxybenzene	ME917224	1,3,5-
		Trimethoxybenzene
ME9171201-4-beta-D-Glucan	ME917120	1,4-beta-D-Glucan
ME 9170311, 6-anhydro-N-acetyl muramate	ME917031	

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

Dimensiones:
Filas: 513
Columnas: 3

Ahora revisamos la matriz Samples. Para ello accedemos con la función colData(). En este caso, las dimensiones son 58 filas y 7 columnas. Notar que ahora las filas corresponden a las columnas de la matriz assay. Las columnas de esta matriz corresponden a: "local_sample_id", "study_id", "sample_source", "mb_sample_id", "raw_data", "Sample_source", "raw_file_name".

Los nombres que se observan en assay corresponden a local_sample_id. Se asume que el último sufijo "_N" corresponde al número de replicado de la muestra. En sample_source se muestra el nombre más descriptivo de la muestra. En mb_sample_id se tienen códigos de especies bacterianas. Todas corresponden a Acinetobacter baumannii ATCC 19606.

kable(head(colData(se1)))

	$local_sample_$	_idstudyid	sample_source	mb_sample_ir	aw_data Sample_source	raw_file_name
Blank_1	Blank_1	ST003521	BLANK	SA386750	Extraction	Blank_1
D1 1 0	DI 1 0	CITTO O DE O 1	DI ANIZ	C A 90 6 7 7 1	blank	D1 1 0
Blank_2	Blank_2	ST003521	BLANK	SA386751	Extraction blank	Blank_2
Blank_3	Blank_3	ST003521	BLANK	SA386752	Extraction	Blank_3
					blank	
$Blank_4$	$Blank_4$	ST003521	BLANK	SA386753	Extraction	Blank_4
					blank	
$Blank_5$	$Blank_5$	ST003521	BLANK	SA386754	Extraction	$Blank_5$
					blank	
COM_1h	1h1	ST003521	Combination_	1 S A386726	Combination_11	n COM_1h_1

```
cat('Dimensiones:\n Filas: ',dim(colData(se1))[1], '\n', 'Columnas: ', dim(colData(se1))[2])
## Dimensiones:
## Filas: 58
```

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

metadata(se1)

Columnas: 7

```
## $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 again
## $subject_type
## [1] NA
```