

## Taller 2: BASES DE DATOS Y BÚSQUEDAS DE INFORMACIÓN BIOLÓGICA

Carlos Andrés Díaz - código: 202010343

David León - código: 201615216

Cesar Patiño - código: 201924259

"palabraclave"[field] <operador lógico> "palabraclave"[field] ...

**Nota:** Si no está seguro de qué campos elegir para una búsqueda dada, haga click en "Advanced" debajo de la barra de búsqueda. Esto lo lleva a una interfaz que le permite construir su búsqueda personalizada. Al momento de buscar, en la barra quedará consignado cómo debería escribirse la misma búsqueda usando campos y operadores booleanos directamente. Puede encontrar más ayuda aquí [https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Advanced\\_Search](https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Advanced_Search)

**IMPORTANTE:** Especifique como realizo la búsqueda para cada punto. Si es necesario adjunte pantallazos que acompañen la búsqueda.

### Para entregar:

En el NCBI, haga una búsqueda de literatura en <https://pubmed.ncbi.nlm.nih.gov/>

- Busque artículos con Ruth Ley como autor. Cuantos artículos encuentra?

**Respuesta:** Al utilizar la ecuación de búsqueda "Ley, Ruth [Author] ", Se encontraron 111 artículos publicados por esta autora

NIH National Library of Medicine  
National Center for Biotechnology Information

PubMed.gov

Ley, Ruth[Author] Search

Advanced Create alert Create RSS User Guide

Save Email Send to Sorted by: Best match Display options

MY NCBI FILTERS 111 results

RESULTS BY YEAR

1998 2021

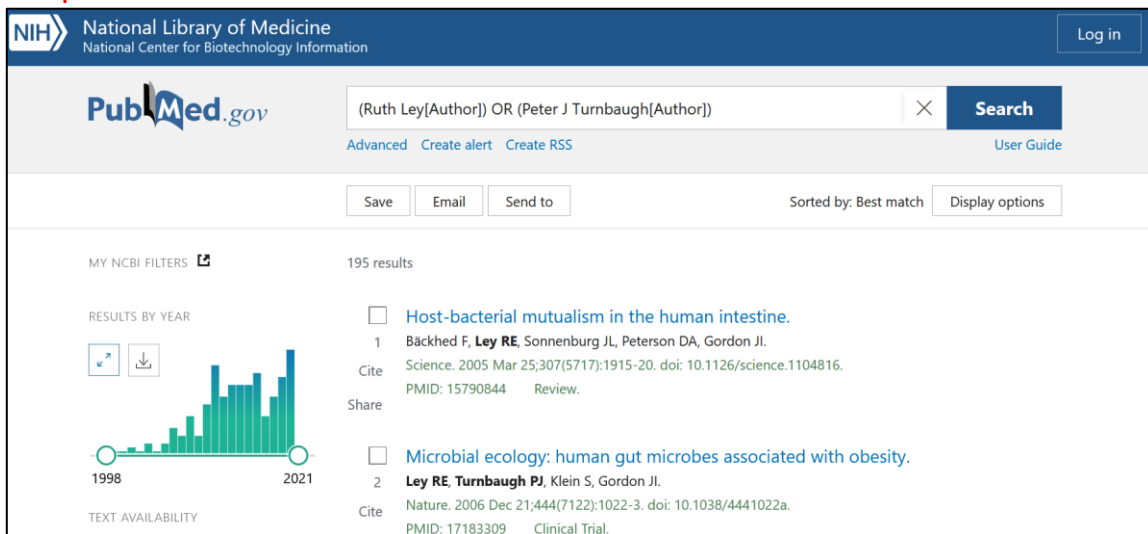
TEXT AVAILABILITY

1 Host-bacterial mutualism in the human intestine.  
Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JL.  
Science. 2005 Mar 25;307(5717):1915-20. doi: 10.1126/science.1104816.  
PMID: 15790844 Review.

2 Microbial ecology: human gut microbes associated with obesity.  
Ley RE, Turnbaugh PJ, Klein S, Gordon JL.  
Nature. 2006 Dec 21;444(7122):1022-3. doi: 10.1038/4441022a.  
PMID: 17183309 Clinical Trial.

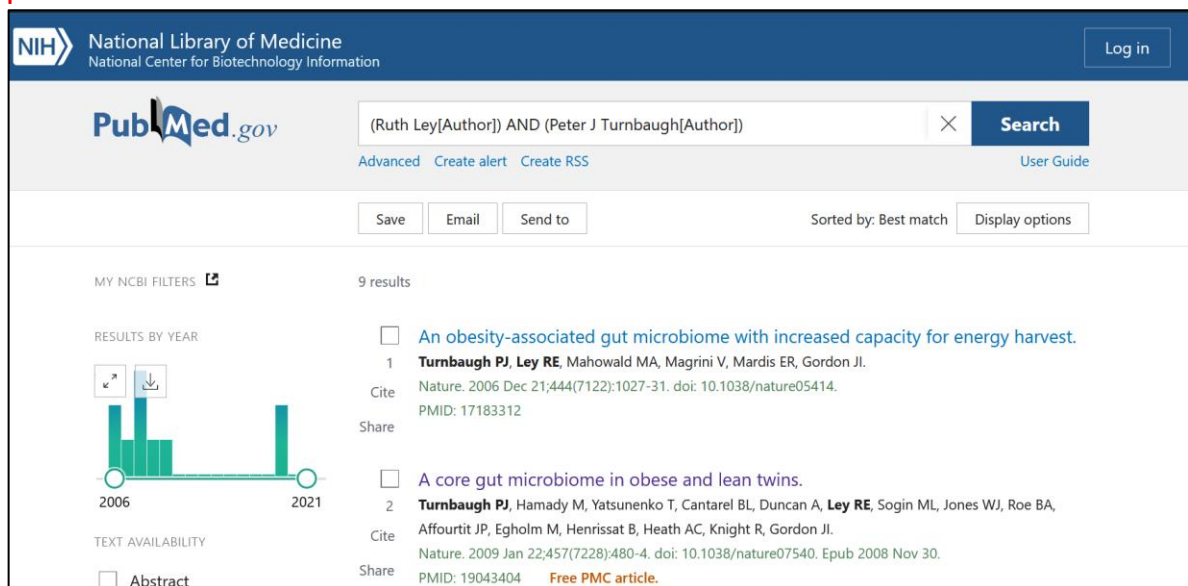
- Ahora busque artículos por Ruth Ley o Peter J Turnbaugh, cuantos obtiene?

**Respuesta:** Al utilizar la ecuación de búsqueda “(Ruth Ley[Author]) OR (Peter J Turnbaugh[Author])”, Se encontraron 195 artículos publicados por ambos autores



- Ahora cuantos artículos fueron escritos por ambos.

**Respuesta:** Al utilizar la ecuación de búsqueda “(Ruth Ley[Author]) AND (Peter J Turnbaugh[Author])”, Se encontraron 9 artículos publicados por ambos autores



- Y si limitamos solo al tiempo en que estuvieron trabajando juntos (2005-2009).  
¿Cuántos se obtienen?

**Respuesta:** Al utilizar la ecuación de búsqueda “(Ruth Ley[Author]) AND (Peter J Turnbaugh[Author]) AND (“2005”[Date - Publication] : “2009”[Date - Publication])”, Se encontraron 6 artículos publicados por ambos autores durante ese periodo de tiempo

NIH National Library of Medicine  
National Center for Biotechnology Information

PubMed.gov

Search: (Ruth Ley[Author]) AND (Peter J Turnbaugh[Author]) AND ("2005"[Date - Publication] : "2009"[Date - Publication])

Advanced Create alert Create RSS User Guide

Save Email Send to Sorted by: Best match Display options

MY NCBI FILTERS

RESULTS BY YEAR

2006 2021

TEXT AVAILABILITY

☐ Abstract

☐ Free full text

6 results

☐ 1 An obesity-associated gut microbiome with increased capacity for energy harvest.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI.

Cite Nature. 2006 Dec 21;444(7122):1027-31. doi: 10.1038/nature05414.

Share PMID: 17183312

☐ 2 A core gut microbiome in obese and lean twins.

Turnbaugh PJ, Hamady M, Yatsunenkov T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI.

Cite Nature. 2009 Jan 22;457(7228):480-4. doi: 10.1038/nature07540. Epub 2008 Nov 30.

Share PMID: 19043404 Free PMC article.

Entre los artículos publicados por ellos uno de los más clásicos es el análisis del microbioma de gemelos delgados y obesos publicado en Nature en 2009. Seleccione ese artículo.

- ¿Cuántos individuos analizaron en ese estudio? Cuántas secuencias de 16S de longitud completa y cuántas parciales? Adicionalmente cuánta información del microbioma se generó?

**Respuesta:** Se analizaron un número total de 154 Individuos, además se encontraron 9,920 secuencias completas y 1,937,461 secuencias parciales de RNA ribosomal 16S. Al realizar la secuenciación se obtuvieron 2.14 Gigabases

Comparative Study
> Nature. 2009 Jan 22;457(7228):480-4. doi: 10.1038/nature07540.
Epub 2008 Nov 30.

FULL TEXT LINKS



## A core gut microbiome in obese and lean twins

Peter J Turnbaugh<sup>1</sup>, Micah Hamady, Tanya Yatsunenko, Brandi L Cantarel, Alexis Duncan, Ruth E Ley, Mitchell L Sogin, William J Jones, Bruce A Roe, Jason P Affourtit, Michael Egholm, Bernard Henrissat, Andrew C Heath, Rob Knight, Jeffrey I Gordon

Affiliations + expand  
PMID: 19043404 PMCID: PMC2677729 DOI: 10.1038/nature07540  
[Free PMC article](#)

ACTIONS

Cite

Favorites

SHARE




Abstract

- ¿Cuántas citas ha obtenido este artículo a la fecha?

**Respuesta:** Al utilizar la seleccionar el artículo, observamos que se encuentra citado en 2,618 artículos científicos, tal y como lo muestra la siguiente figura:

Cited by 2,618 articles

### Understanding Obesity: The Role of Adipose Tissue Microenvironment and the Gut Microbiome.

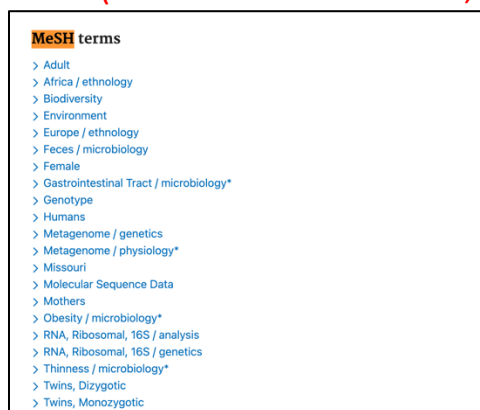
Awan NM, Meurling IJ, O'Shea D.  
Saudi J Med Med Sci. 2021 Jan-Apr;9(1):10-15. doi: 10.4103/sjmms.sjmms\_561\_20. Epub 2020 Dec 26.  
PMID: 33519338 [Free PMC article.](#) [Review.](#)

### Time-Restricted Eating, Intermittent Fasting, and Fasting-Mimicking Diets in Weight Loss.

Fanti M, Mishra A, Longo VD, Brandhorst S.  
Curr Obes Rep. 2021 Jan 29. doi: 10.1007/s13679-021-00424-2. Online ahead of print.  
PMID: 33512641 [Review.](#)

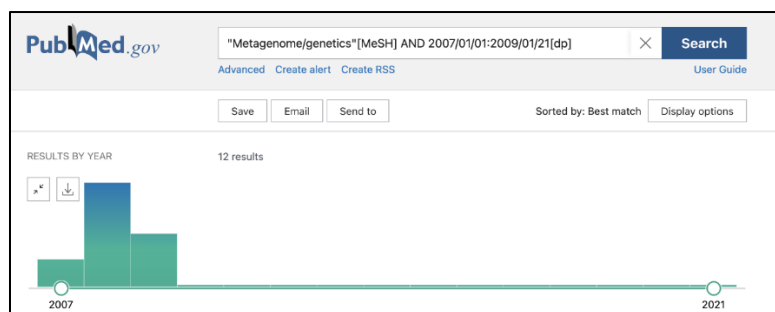
Entre los términos MeSH usados uno de los que se nombra es Metagenome / genetics.  
¿Qué es un término MeSH? Y ¿Cuántos artículos se han publicado con el termino MeSH Metagenome / genetics

**Respuesta:** “El término MeSH hace referencia a los «medical subject headings», también conocidos como «encabezamientos de materia médicos», «títulos de temas médicos» o «descriptores en ciencias de la salud», los cuales son utilizados por la NLM.” (Fernández-Altuna et al., 2016)



- Cuantos artículos se han publicado con el termino MeSH Metagenome / genetics antes de la fecha de publicación de este artículo? Considera que este fue uno de los artículos pioneros en el campo? Por qué?

**Respuesta:** Se encontraron 12 artículos con el término MeSH Metagenome / genetics antes de la fecha de publicación del artículo (22/01/2009). Sí consideramos que fue uno de los artículos pioneros en el campo porque está entre los primeros 15 publicados, además, al revisar los artículos posteriores nos damos cuenta que hay un crecimiento acelerado, por lo que en los últimos años se han añadido más de 1000 artículos con el mismo MeSH



Ahora vamos a buscar los datos.

- En Información relacionada, buscar Proyecto Relacionado. ¿Cuál es el número de acceso de este proyecto?

**Respuesta:** El número de acceso de este proyecto es: PRJNA32089 tal y como se muestra en la siguiente figura

Accession	<a href="#">PRJNA32089</a>
Data Type	Metagenome
Scope	Environment
Organism	<a href="#">human gut metagenome</a> [Taxonomy ID: 408170] unclassified sequences; metagenomes; organismal metagenomes; human gut metagenome
Publications	Turnbaugh PJ <i>et al.</i> , "A core gut microbiome in obese and lean twins.", <i>Nature</i> , 2009 Jan 22;457(7228):480-4
Submission	Registration date: 1-Oct-2008 - Washington University (WashU) - Washington University Center for Genome Sciences
NCBI Links	<ul style="list-style-type: none"> <li>• <a href="#">BLAST</a></li> <li>• <a href="#">Metagenomics Book</a></li> <li>• <a href="#">BLAST against environmental sequences</a></li> </ul>
Locus Tag Prefix	TWGUT

- ¿En cuanto a datos de secuenciación se encuentran tanto en Nucleótidos como en SRA. Cuántas entradas hay en cada uno de estos?

**Respuesta:** Se encuentran 9773 entradas para Nucleótidos y 20 entradas para SRA y 20 entradas para BioSample.

**Project Data:**

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic DNA)	9773
SRA Experiments	20
PUBLICATIONS	
PubMed	1
PMC	1
OTHER DATASETS	
BioSample	20

- Al entrar a los datos en Nucleotide, y dada la información sobre secuenciación reportada en el estudio, ¿a qué información considera que corresponde?

**Respuesta:** La información corresponde a las secuencias parciales del gen ribosomal del ARN 16S de cada una de las bacterias, que no se pueden cultivar que conforman el microbioma de los gemelos y las madres.

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide 32089[BioProject] Search

Create alert Advanced Help

**COVID-19 is an emerging, rapidly evolving situation.**  
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#)

Species: Bacteria (9,773) Customize ...

Molecule types: genomic DNA/RNA (9,773) Customize ...

Source databases: INSDC (GenBank) (9,773) Customize ...

Sequence Type: Nucleotide (9,773)

Sequence length: Custom range...

Release date: Custom range...

Revision date: Custom range...

[Clear all](#)  
[Show additional filters](#)

Summary 20 per page Sort by Default order Send to: Filters: [Manage Filters](#)

Items: 1 to 20 of 9773

<< First < Prev Page 1 of 489 Next > Last >>

- ☐ [Uncultured bacterium clone TS9\\_a04h12 16S ribosomal RNA gene, partial sequence](#)  
 1,047 bp linear DNA  
 Accession: FJ372382.1 GI: 214026763  
[BioProject](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Uncultured bacterium clone TS9\\_a04h11 16S ribosomal RNA gene, partial sequence](#)  
 1,377 bp linear DNA  
 Accession: FJ372381.1 GI: 214026762  
[BioProject](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Uncultured bacterium clone TS9\\_a04h10 16S ribosomal RNA gene, partial sequence](#)  
 1,338 bp linear DNA  
 Accession: FJ372380.1 GI: 214026761  
[BioProject](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Uncultured bacterium clone TS9\\_a04h09 16S ribosomal RNA gene, partial sequence](#)  
 1,361 bp linear DNA  
 Accession: FJ372379.1 GI: 214026760  
[BioProject](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Uncultured bacterium clone TS9\\_a04h08 16S ribosomal RNA gene, partial sequence](#)  
 1,384 bp linear DNA  
 Accession: FJ372378.1 GI: 214026759  
[BioProject](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)
- ☐ [Uncultured bacterium clone TS9\\_a04h07 16S ribosomal RNA gene, partial sequence](#)  
 1,338 bp linear DNA

Find related data  
 Database: Select  
 Find items

Search details  
 32089[BioProject]  
 Search See more...

Recent activity  
 Turn Off Clear  
 32089[BioProject] (9773) Nucleotide  
 human gut metagenome BioProject  
 Project Links for PubMed (Select 19043404) (1) BioProject  
 SRA Links for BioProject (Select 32089) (20) SRA  
 Uncultured bacterium clone TS9\_a04h12 16S ribosomal RNA gene, partial set Nucleotide  
 See more...

- ¿A que corresponden los datos que están en SRA?

**Respuesta:** Los datos que se encuentran almacenados en SRA corresponden a las secuencias obtenidas por la técnica de secuenciación 454 del microbioma intestinal analizado en el estudio.

The screenshot shows the NCBI SRA search results for the query '454 sequencing of human gut metagenome TS51 genomic fragment library'. The page includes a search bar, a 'Send to' dropdown, and a 'Filters: Manage Filters' link. The results are displayed in a table with columns for 'Run', '# of Spots', '# of Bases', 'Size', and 'Published'. The table shows 7 runs, with the first run (SRR006538) having 327,560 spots and 41.2M bases. The page also includes a 'Recent activity' section and a 'Find related data' section.

Hay 2 tipos de datos que diferencia tienen?

**Respuesta:** Al seleccionar los dos tipos de datos observamos que la principal diferencia se encuentra en la estrategia de secuenciación utilizada y en el método de selección empleado y finalmente existe una diferencia entre el número de corridas empleados por cada tipo de dato, esto debido a la diferencia de fecha en la que fueron realizados los experimentos, tal y como se muestra en la figura:

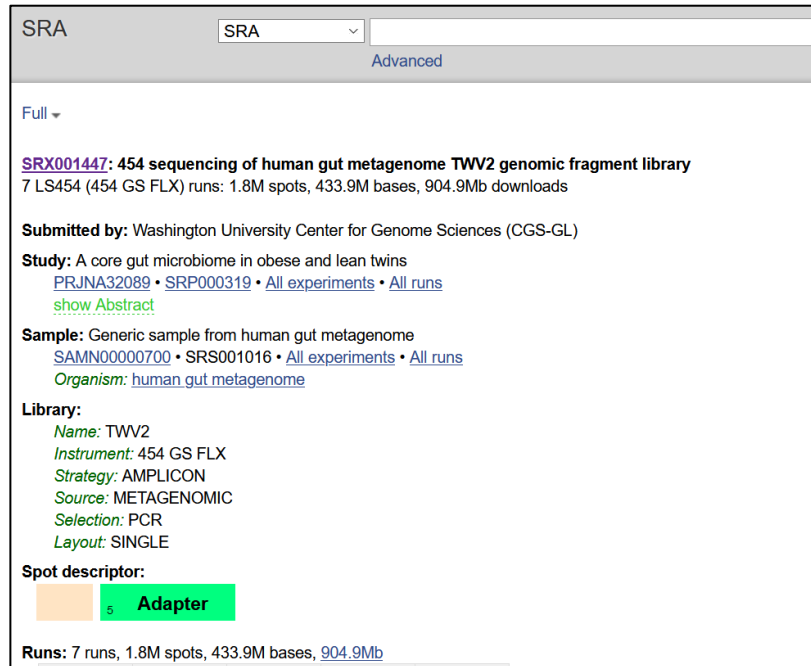
The image shows a side-by-side comparison of two SRA run details. The left panel shows run SRR006538, which was submitted by Washington University Center for Genome Sciences (CGS-GL) and is a 454 GS FLX run. The right panel shows run SRR029698, which was submitted by the same center and is a 454 GS FLX run. The comparison highlights differences in the sequencing strategy (WGS vs. WGS) and the date of publication (2008-12-19 vs. 2009-11-16).

Run	# of Spots	# of Bases	Size	Published
SRR006538	327,560	41.2M	119.6Mb	2008-12-19
SRR006539	373,067	100.1M	225.4Mb	2008-12-19
SRR006540	319,333	86.4M	168.8Mb	2008-12-19
SRR006541	74,031	19.8M	38.9Mb	2008-12-19
SRR006542	65,775	17.5M	35.6Mb	2009-10-27
SRR006543	260,264	69.8M	126.9Mb	2008-12-19
SRR006544	368,840	99.2M	189.7Mb	2008-12-19



- Entrando al proyecto SRX001447. A que tipo de estrategia de secuenciación corresponde? Que instrumento se usó para generar estos datos? Cuantas corridas hacen parte de este experimento? ¿Y cuantas secuencias en total se generaron?

**Respuesta:** Luego de revisar el proyecto se observan las siguientes características:



SRA SRA Advanced

Full ▾

**SRX001447: 454 sequencing of human gut metagenome TWV2 genomic fragment library**  
7 LS454 (454 GS FLX) runs: 1.8M spots, 433.9M bases, 904.9Mb downloads

**Submitted by:** Washington University Center for Genome Sciences (CGS-GL)

**Study:** A core gut microbiome in obese and lean twins  
[PRJNA32089](#) • [SRP000319](#) • [All experiments](#) • [All runs](#)  
[show Abstract](#)

**Sample:** Generic sample from human gut metagenome  
[SAMN00000700](#) • [SRS001016](#) • [All experiments](#) • [All runs](#)  
**Organism:** [human gut metagenome](#)

**Library:**  
**Name:** TWV2  
**Instrument:** 454 GS FLX  
**Strategy:** AMPLICON  
**Source:** METAGENOMIC  
**Selection:** PCR  
**Layout:** SINGLE

**Spot descriptor:**  
5 Adapter

**Runs:** 7 runs, 1.8M spots, 433.9M bases, [904.9Mb](#)

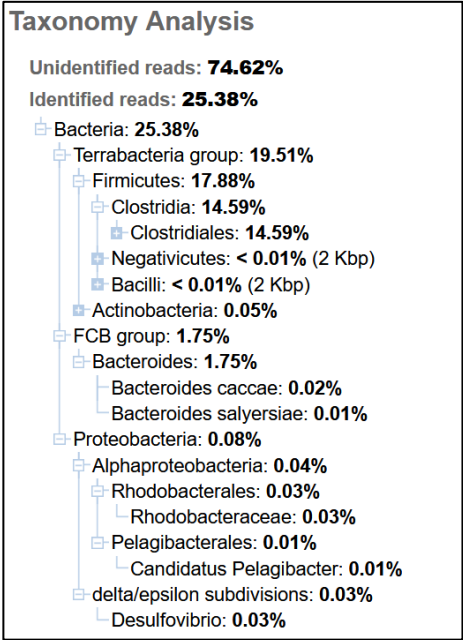
En donde se observa claramente que la plataforma que se utiliza para realizar la secuenciación 454 secuenciación, utilizando la herramienta 454 GS FLX para la obtención de los datos, para la obtención de los datos se realizan 7 corridas, generando 904.9Mb en información de las secuencias del estudio

- ¿Si entra a la corrida SRR006538, en la sección de análisis que información le da? Cuál es el phylum bacteriano más abundante?

**Respuesta:** Luego de revisar la información de la corrida SRR006538 se obtiene la información mostrada en la figura que está adjunta, en ella se realiza una asignación del número conocido y desconocidos de reads que se obtienen durante la secuenciación, y dentro de los reads conocidos se realiza una identificación de los grupos de bacterias más abundantes que son el grupo de las terrabacterias, las esfingobacterias (FCB) y finalmente el grupo de las proteobacterias. Además, presenta una clasificación de las lecturas más fuertes encontrándose que las

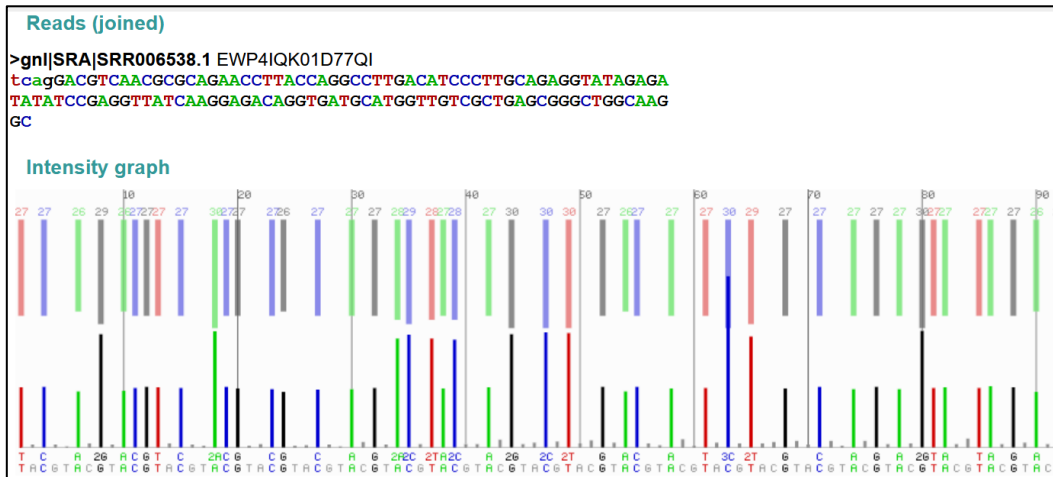
bacterias son los microorganismos más abundantes y que el orden de los Clostridiales tiene una mayor representatividad dentro de estos (23,2%). Por último, resulta evidente que el filum más abundante de todo el análisis es el filum de los Firmicutes con una representatividad del 17.88%

Strong signals					
SuperKingdom	Organism	Rank	%%	Kbp	Coverage
Bacteria	Clostridiales	order	23.2	9,533	
Bacteria	Bacteroides	genus	2.1	855	
Bacteria	Lachnospiraceae	family	1.2	514	



- ¿Que información le da la sección “reads” y la sección “Data Access”?

**Respuesta:** Luego de revisar la información de la sección *reads*, se obtiene una lista con todos los reads obtenidos y además se obtiene una secuencia de los reads juntos, además de esto se muestra un mapa en donde se asigna cada nucleótido



Al revisar la sección *Data Access*, es evidente que esta sección describe el peso que tienen las corridas, en donde se encuentran almacenados así como su nombre, y su tipo de acceso.

54 sequencing of human gut metagenome TWV2 genomic fragment library (SRR006538) [Change accession.](#)

Type	Size	Location	Name	Free Egress	Access Type
run	122,428 Kb	NCBI	<a href="https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos1/sra-pub-run-5/SRR006538/SRR006538.3">https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos1/sra-pub-run-5/SRR006538/SRR006538.3</a>	worldwide	anonymous
		AWS	s3://sra-pub-run-7/SRR006538/SRR006538.3	s3.us-east-1	aws identity
		GCP	gs://sra-pub-run-7/SRR006538/SRR006538.3	gs.US	gcp identity

En el **EBI**, en <http://europepmc.org/>

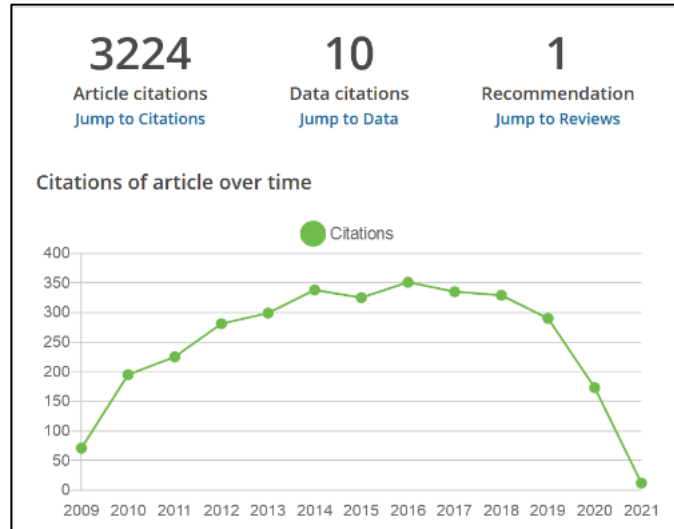
Realice la misma búsqueda de artículos escritos por Ruth Ley y Peter J Turnbaugh entre 2005 y 2009. Como es la búsqueda? Cuantos resultados da? Son los mismos?

**Respuesta:** Al realizar la búsqueda se encontraron 6 artículos científicos entre sí, a la pregunta de si son los mismos que los encontrados en la búsqueda de los puntos anteriores usando la base de datos

The screenshot shows the Europe PMC search interface. At the top, there's a navigation bar with links: About, Tools, Developers, Help, and Europe PMC plus. Below this is a search bar with the text "Search worldwide, life-sciences literature". The search query entered is "(AUTH:"Ley Ruth" AND AUTH:"Turnbaugh Peter J") AND (FIRST\_PDATE:[2005 TO 2009])". There are buttons for "Search" and "Advanced Search". Below the search bar, there are links for "Coronavirus articles and preprints" and "Search examples: 'breast cancer' Smith J". There are also tabs for "Recent history" and "Saved searches". On the left side, under "Search only", there are filters for "Type" (Research articles (5), Reviews (1), Preprints (0)) and "Free full text" (Free to read (4)). On the right side, there's a section for "Suggested authors matching 'AUTH:Ley Ruth' with an ORCID". It lists "Ruth Ley" with 99 Publications and 49,252 Citations, and a link to "View profile". There's also a "Save search" button and an "Export citations" button. At the bottom, it says "1-6 of 6 results" and "Sort by: Date received".

- Seleccione nuevamente el mismo artículo. Cuantas citas obtiene? De un pantallazo de las citas en función del tiempo. ¿Si es diferente a los resultados del NCBI a que cree que se debe la diferencia?

**Respuesta:** Al realizar la búsqueda del artículo, se encontró que este se encuentra citado 3,224 veces, siendo el 2016 el año en el que más citas obtuvo.



Al realizar una comparación con los resultados obtenidos en la búsqueda en el NCBI (2,618 citaciones), resulta evidente que la cantidad de citaciones que presenta cada artículo es diferente, siendo el EBI la base de datos con mayor número de citaciones reportadas (3,224 veces).

Una posible explicación de lo anterior se debe a que la base de datos del EBI utiliza un algoritmo que analiza conceptos de deep learning entrenado para clasificar las citas de acuerdo con tres categorías en específico, tal y como se describe en la siguiente figura:

- Aquellos autores realizan la citación aportando evidencia para soportar su validez
- Aquellos autores realizan la citación aportando evidencia para desaprobar o disputar su validez
- Aquellos autores que no brindan ningún soporte al momento de realizar la citación

## Smart citations by scite.ai ?

Explore citation contexts and check if this article has been supported or disputed.

<https://scite.ai/reports/10.1038/nature07540>

Supporting ✓ 261

Mentioning 4243

Disputing ? 41

- En la sección de datos hay vínculos nuevamente a Bio-estudios, secuencias de nucleótidos y Mgnify. A que corresponde cada uno de estos links?

### BioStudies: supplemental material and supporting data

<http://www.ebi.ac.uk/biostudies/studies/S-EPMC2677729?xr=true>

### Nucleotide Sequences (2)

ENA - FJ372382 (1 citation)

ENA - FJ362604 (1 citation)

### MGnify

<https://www.ebi.ac.uk/metagenomics/projects/SRP000319>

**Respuesta:** Al explorar la sección de datos, el resultado que se obtiene el apartado de cada una de las secciones es el siguiente:

En el apartado de Bio-estudios, el link corresponde a información del estudio la cual incluye su financiación y sus datos de publicación.

The screenshot shows the BioStudies website homepage. At the top, there is a blue header with the BioStudies logo on the left and a search bar on the right. Below the header, there is a navigation bar with links: Home, Browse, Submit, Help, and About BioStudies. The main content area features a large heading "BioStudies – one package for all the data supporting a study" with a background image of a DNA double helix. Below this heading, there is a paragraph describing the database. To the right, there is a "Data Content" box with three statistics: 3,404,677 files, 7,267,926 links, and 1,835,780 studies.

**BioStudies.** Search BioStudies  
Examples: hyperplasia, PMC516016

Home Browse Submit Help About BioStudies Feedback Login

**BioStudies – one package for all the data supporting a study**

The BioStudies database holds descriptions of biological studies, links to data from these studies in other databases at EMBL-EBI or outside, as well as data that do not fit in the structured archives at EMBL-EBI. The database can accept a wide range of types of studies described via a simple format. It also enables manuscript authors to submit supplementary information and link to it from the publication.

**Data Content**

- 3,404,677 files
- 7,267,926 links
- 1,835,780 studies

En el apartado de secuencias de nucleótidos, el link corresponde a información particular de las secuencias, como su topología, el tipo de molécula a la que está asociada la secuencia y a las referencias cruzadas que la contienen.

## Sequence: FJ372382.1

Uncultured bacterium clone TS9\_a04h12 16S ribosomal RNA gene, partial sequence.

<b>Organism:</b>	<a href="#">uncultured bacterium</a>
<b>Accession:</b>	FJ372382
<b>Mol Type:</b>	genomic DNA
<b>Topology:</b>	LINEAR
<b>Base Count:</b>	1047
<b>Dataclass:</b>	STD
<b>Tax Division:</b>	ENV

Show Less

### Navigation & Cross References



- Taxon:  
[Taxon:77133](#)
- Study:  
[PRJNA32089](#)
- EuropePMC:  
[PMC3426808](#), [PMC3516142](#), [PMC4269443](#), [PMC5927471](#)
- MGnify:  
[MGYS00000258](#)
- PubMed:  
[19043404](#)

En el apartado de Mgnify, el link corresponde a una colección de diferentes análisis realizados sobre estudio.

Analyses [-]


Show:  results from a total of 16 (page 1 of 1).

[Download results](#)

Biome	Sample accession	Sample description	Run / Assembly accession	Pipeline version	Analysis accession
	<a href="#">SRS000998</a>	Subject ID - F1T1Le. Sample Characteristics - The MOAFTS twin cohort, comprised of female likesex twin pairs, were identified from Missouri birth records over the period 1994-1999, when the twins had a median age of 15 years. A total of 350 twins from the larger MOAFTS cohort completed screening interviews for the present study. We were able to take advantage of the wave five interview of the MOAFTS twin cohort (which has 90% retention of wave four participants) to identify pairs most likely to meet study criteria. Eligibility was then confirmed at screening interview.	<a href="#">SRR029687</a>	<a href="#">1.0</a>	<a href="#">MGYA00000031</a>
	<a href="#">SRS000999</a>	Subject ID - F1T2Le. Sample Characteristics - The MOAFTS twin cohort, comprised of female likesex twin pairs, were identified from Missouri birth records over the period 1994-1999, when the twins had a median age of 15 years. A total of 350 twins from the larger MOAFTS cohort completed screening interviews for the present study. We were able to take advantage of the wave five interview of the MOAFTS twin cohort (which has 90% retention of wave four participants) to identify pairs most likely to meet study criteria. Eligibility was then confirmed at screening interview.	<a href="#">SRR029692</a>	<a href="#">1.0</a>	<a href="#">MGYA00000032</a>
		Subject ID - F1MOv. Sample Characteristics - The MOAFTS twin cohort, comprised of female likesex twin pairs, were identified from Missouri birth records over the period 1994-1999, when the twins had a median age of 15 years. A total of 350 twins from the larger MOAFTS cohort completed screening interviews for the present study. We were able to take advantage of the wave five interview of the MOAFTS twin cohort (which has 90% retention of wave four participants) to identify pairs most likely to meet study criteria. Eligibility was then confirmed at screening interview.			

- En el link de nucleótidos que lleva al ENA, hay referencia a un estudio. ¿El Accession number de este estudio es el mismo encontrado en el NCBI? A donde lleva ese vínculo?

**Respuesta:** Si, es el mismo *accession number* que encontramos en NCBI. El vínculo nos dirige a un resumen donde encontramos información que varía desde los datos del estudio hasta la publicación del mismo.

**Project: PRJNA32089** 

Total community DNA was isolated from the feces of monozygotic and dizygotic twins and their mothers. We have characterized the fecal microbial communities of adult female monozygotic and dizygotic twin pairs concordant for leanness or obesity, and their mothers. The results demonstrate that a diversity of organismal assemblages can nonetheless yield a core microbiome at a functional level, and that deviations from this core are associated with different physiologic states (obese versus lean).

[FJ362604-FJ372382](#) correspond to 16S ribosomal RNA sequences isolated from this study.

454 sequence data is available from the Short Read Archive (SRA): [SRA002775](#).

---

**Organism:** [human gut metagenome](#)

**Secondary Study Accession:** [SRP000319](#)

**Study Title:** A core gut microbiome in obese and lean twins

**Center Name:** Washington University (WashU)

**Study Name:** human gut metagenome



- Es posible encontrar las mismas corridas? El mismo experimento? Como se reconocen los códigos de acceso de las corridas o los experimentos? Permite ver los análisis o descargar los datos?

**Respuesta:** Si es posible encontrar las mismas corridas y el mismo experimento.

Los códigos de acceso de las corridas se reconocen porque empiezan con SRR, por otra parte, los códigos de acceso de los experimentos se reconocen porque empiezan por SRX. (nota: para facilitar el proceso de búsqueda descargamos los datos en un archivo JSON)

Permite descargar los datos, pero no permite ver los análisis.


Study Accession	Sample Accession	Experiment Accession	Run Accession	Tax Id	Scientific Name	<a href="#">Download All</a>	<a href="#">Download</a>
						FASTQ FTP	Submitte
PRJNA32089	SAMN00000701	SRX001445	SRR005479	408170	human gut metagenome	<input type="checkbox"/> <a href="#">SRR005479.fastq.gz</a>	
PRJNA32089	SAMN00000700	SRX001447	SRR006538	408170	human gut metagenome	<input type="checkbox"/> <a href="#">SRR006538.fastq.gz</a>	
PRJNA32089	SAMN00000700	SRX001447	SRR006539	408170	human gut metagenome	<input type="checkbox"/> <a href="#">SRR006539.fastq.gz</a>	
PRJNA32089	SAMN00000700	SRX001447	SRR006540	408170	human gut metagenome	<input type="checkbox"/> <a href="#">SRR006540.fastq.gz</a>	
PRJNA32089	SAMN00000700	SRX001447	SRR006541	408170	human gut metagenome	<input type="checkbox"/> <a href="#">SRR006541.fastq.gz</a>	

- En el link a Mgnify cuantos análisis están disponibles para este estudio? Corresponden a los datos de secuenciación del gen 16S? o a shotgun metagenómico?

**Respuesta:** En Mgnify hay disponibles 16 análisis para este estudio. Los análisis corresponden a los datos de secuenciación del gen 16S.

Analyses [-]

Show: 25 results from a total of 16 (page 1 of 1).

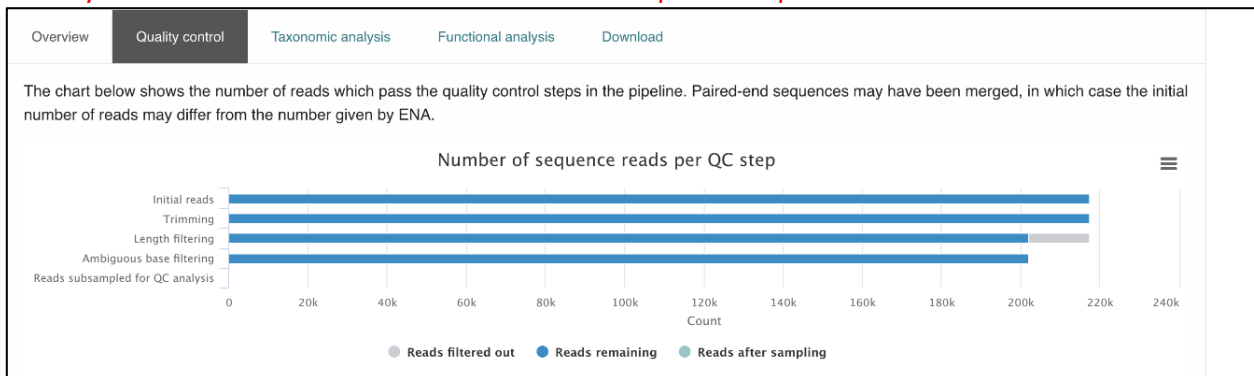
Biome	Sample accession	Sample description	Run / Assembly accession	Pipeline version	Analysis accession
	<a href="#">SRS000998</a>	Subject ID - F1T1Le. Sample Characteristics - The MOAFTS twin cohort, comprised of female likesex twin pairs, were identified from Missouri birth records over the period 1994-1999, when the twins had a median age of 15 years. A total of 350 twins from the larger MOAFTS cohort completed screening interviews for the present study. We were able to take advantage of the wave five interview of the MOAFTS twin cohort (which has 90% retention of wave four participants) to identify pairs most likely to meet study criteria. Eligibility was then confirmed at screening interview.	<a href="#">SRR029687</a>	1.0	<a href="#">MGYA00000031</a>

[Download results](#)

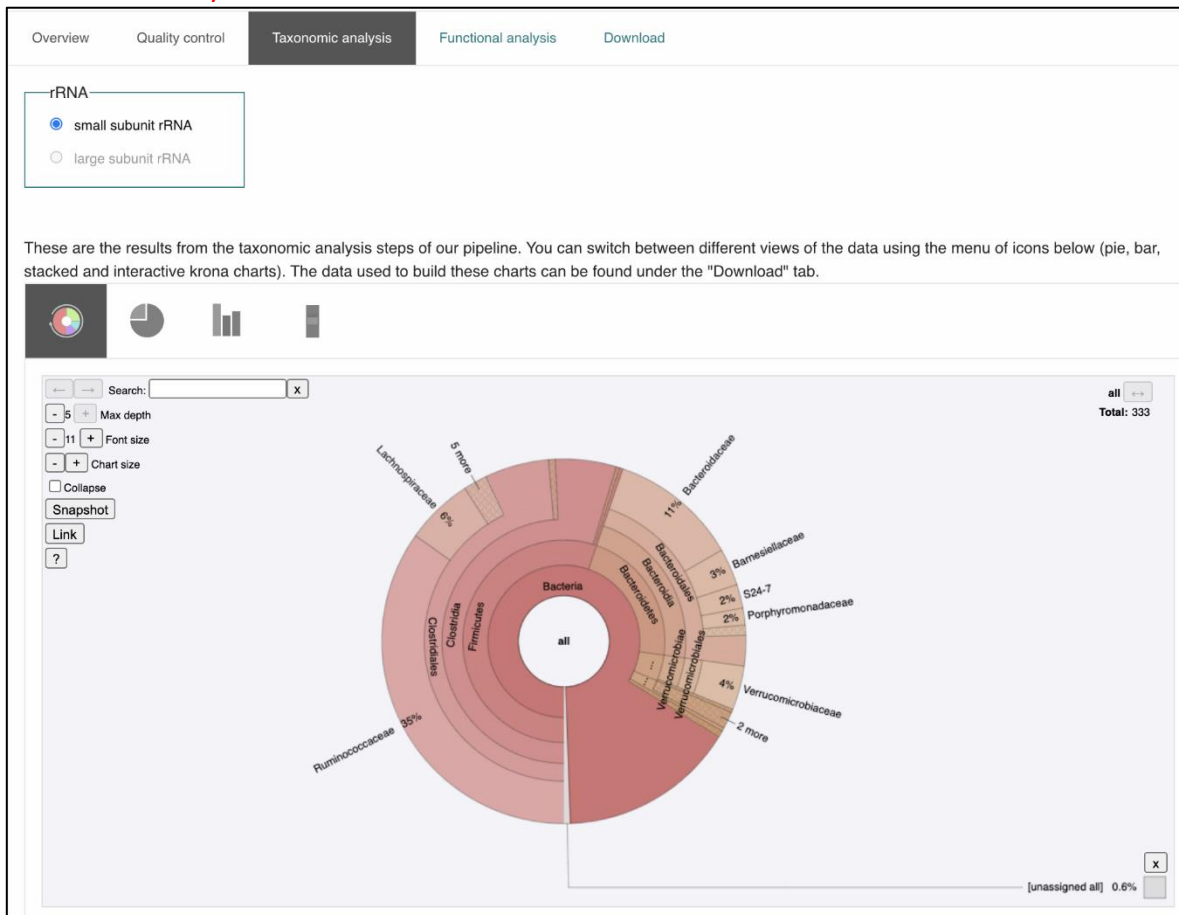
- Entrando al primer análisis, correspondiente al SRR029687 (MGYA00000031). Que información le da las diferentes pestañas? (Quality Control, Taxonomic analysis, Functional analysis, Download).

**Respuesta:** La información que se ve en las diferentes pestañas es la siguiente:

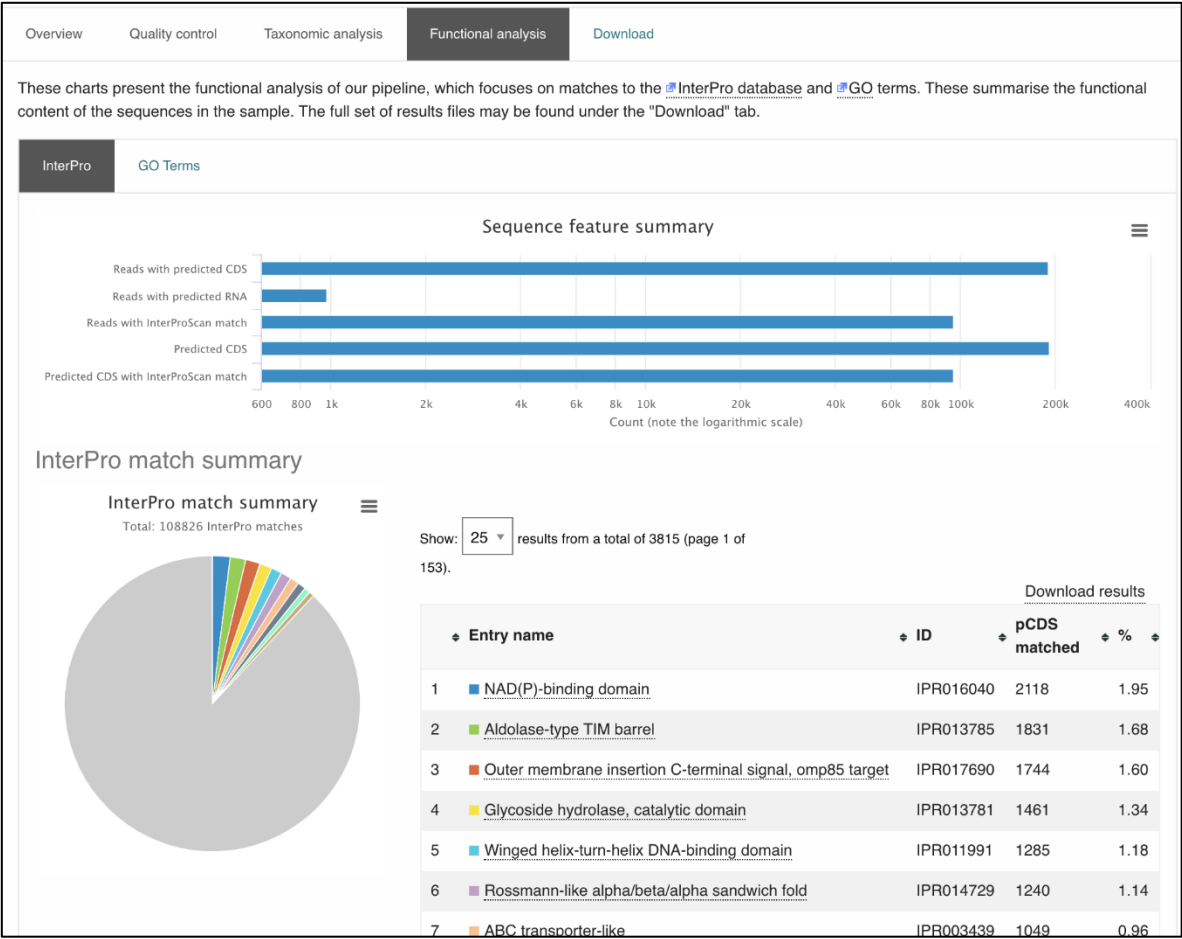
**Quality Control:** Numero de lecturas de secuencia por cada paso QC



Taxonomic analysis:










Functional analysis:






Download: Se presenta la opción de descargar los datos presentados anteriormente (Sequence data, Funcional analysis y Taxonomic analysis) en diferentes formatos, algunos de los formatos son FASTA, CSV, TSV, JSON y Newick format.







## Sequence data

Name	Compression	Format	
Predicted CDS (aa)	gz	FASTA	
Predicted CDS without annotation	gz	FASTA	
Predicted ORF without annotation	gz	FASTA	
Processed reads with annotation	gz	FASTA	
Processed reads without annotation	gz	FASTA	
Processed nucleotide reads	gz	FASTA	
Processed reads with pCDS	gz	FASTA	

## Functional analysis

Name	Compression	Format	
Complete GO annotation	-	CSV	
GO slim annotation	-	CSV	
InterPro matches	gz	TSV	

## Taxonomic analysis

Name	Compression	Format	
Reads encoding 16S rRNA	-	FASTA	
Reads encoding 23S rRNA	-	FASTA	
Reads encoding 5S rRNA	-	FASTA	
Phylogenetic tree	-	Newick format	
OTUs, counts and taxonomic assignments	-	JSON Biom	
OTUs, counts and taxonomic assignments	-	TSV	

- ¿En su opinión personal considera más útiles los resultados obtenidos con el NCBI o con el EBI?

**Respuesta:** Consideramos más útiles los resultados obtenidos con el EBI, pues nos presentaron una perspectiva más completa sobre los diferentes motivos de citación del estudio; por otra parte, nos parecieron útiles las gráficas que podemos observar en los análisis disponibles, la cantidad de información que se puede obtener y la forma en la que esta plataforma permite acceder a los datos organizadamente facilitando a los investigadores lograr la reproducibilidad de los artículos presentados.

## Referencias

Fernández-Altuna, M., Martínez del Prado, A., Arriarán Rodríguez, E., Gutiérrez Rayón, D., Toriz Castillo, H. and Lifshitz Guinzberg, A., 2016. Uso de los MeSH: una guía práctica. *Investigación en Educación Médica*, 5(20), pp.220-229.

scite help desk. 2021. How are citations classified?. [online] Available at: <<https://help.scite.ai/en-us/article/how-are-citations-classified-1a9j78t/>> [Accessed 5 February 2021].

Sequence Read Archive Submissions Staff. Understanding SRA Search Results. In: SRA Knowledge Base [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2011-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK56913/>