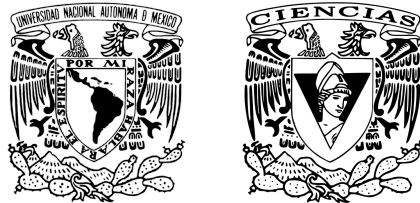


UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS



Genómica Computacional  
**Resistencia a antibióticos en**  
***Mycobacterium tuberculosis***

Contrastando *M. tuberculosis* resistente a  
antibióticos con *M. tuberculosis* no resistente

*Ximena Cervantes López* - 318204930

*Adriana Henández Gasca* - 316161570

*Sebastián Alamina Ramírez* - 318685496

Trabajo presentado como parte del curso de **Genómica Computacional**, impartido por el profesor **Sergio Hernández López** durante el semestre 2023-2 en la Facultad de Ciencias, UNAM.

Fecha de entrega: **martes 6 de junio del 2023.**

# Pregunta de investigación

¿Existen variantes genéticas que le permita a *M. tuberculosis* el desarrollo de resistencia a fármacos durante la primera línea de tratamiento?

## Introducción

La Tuberculosis (TB) es una enfermedad infecciosa transmitida por el aire, prevenible y curable. Suele afectar a los pulmones, aunque también puede afectar otras partes del cuerpo. *Mycobacterium tuberculosis*, la bacteria causante de la TB, es un patógeno extremadamente eficaz para adaptar su supervivencia dentro del huésped donde habita. Además de TB, *M. tuberculosis* también es responsable por otras enfermedades humanas, como ciertas complicaciones respiratorias, enfermedades auto-inmunes, y síndromes metabólicos. [Chai et al., 2018]

Las cepas bacterianas resistentes a múltiples drogas están en aumento alrededor del mundo, esparciendo así TB alrededor del mundo. Su respuesta rigurosa (SR, por sus siglas del inglés *stringent response*, que regula la replicación, transcripción y traducción bacterianas) demuestra que éstas pueden sobrevivir bajo circunstancias hostiles, incluida la exposición a antibióticos. [Arrigoni et al., 2022] Esta resistencia antibiótica se logra mediante una adquisición horizontal de los genes que otorgan la resistencia al recombinar el ADN externo dentro del cromosoma, o por mutaciones en diferentes locus cromosómicos. En biología molecular evolutiva, el término de tasa de mutación se aplica a las estimaciones o conteos por generación de mutaciones por nucleótido o por genoma completo; la frecuencia de mutación mide las mutaciones presentes en una población, sin importar si las mutaciones se presentaron temprana o tardíamente dentro de la población. [Martinez and Baquero, 2000]

La secuenciación de nueva generación es una tecnología moderna utilizada para la secuenciación del ADN y del ARN, y para la detección de variantes y mutaciones. NGS, por sus siglas del inglés, puede secuenciar cientos y miles de genes o genomas completos en poco tiempo. Las mutaciones o variantes detectadas en las secuencias con ayuda de NGS han sido utilizadas ampliamente para el tratamiento de enfermedades; su diagnóstico, su pronóstico, y su seguimiento. [Qin, 2019] Para realizar ésta y otras técnicas, existen diversas herramientas como *UseGalaxy*, que consiste en una plataforma gratuita en línea de código abierto con herramientas necesarias para el análisis de datos masivos que permite realizar investigaciones reproducibles sin necesidad de conocer ni memorizar comandos muy complejos. [Hiltemann et al., 2023]

## Objetivos

1. Identificar el cambio en los nucleótidos dentro de los genomas de las muestras de *M. tuberculosis* con resistencia a antibióticos, respecto al genoma de referencia sin la resistencia.
2. Corroborar con la literatura si la resistencia a antibióticos es debido al cambio de los nucleótidos dentro del genoma.

# Metodología

Para comenzar, reunimos toda la bioinformación requerida para la ejecución del proyecto, para esto fue necesaria la búsqueda del genoma de referencia a trabajar, como en este caso nuestro organismo modelo es *M. tuberculosis*, realizamos una búsqueda en la literatura para encontrar el genoma de referencia más actualizado (Figura 1), el cual es H37Rv [NCBI, 2017], así mismo realizamos una búsqueda de una base de datos pública que tuviera datos de muestras secuenciadas de *M. tuberculosis* resistentes a fármacos (Figura 2), para esto encontramos un estudio realizado en la India que contenía bases de datos de muestras de *M. tuberculosis* que presentaban las características que buscábamos para poder trabajar con ellas. Es importante mencionar que las muestras que trabajamos fueron de SRX26381650 a SRX26381664 [NCBI, 2019].

The screenshot shows the NCBI GenBank entry for the *Mycobacterium tuberculosis* H37Rv complete genome. The page is viewed in Google Chrome. The main title is "Mycobacterium tuberculosis H37Rv, complete genome". Below the title, it says "NCBI Reference Sequence: NC\_000962.3". There are links for "FASTA" and "Graphics". A "Go to:" dropdown menu is visible. The left sidebar contains a table with the following information:

LOCUS	NC_000962	4411532 bp	DNA	linear	CON 14-DEC-2017
DEFINITION	Mycobacterium tuberculosis H37Rv, complete genome.				
ACCESSION	NC_000962				
VERSION	NC_000962.3				
DBLINK	BioProject: <a href="#">PRJNA57777</a>				
KEYWORDS	RefSeq; complete genome.				
SOURCE	Mycobacterium tuberculosis H37Rv				
ORGANISM	<a href="#">Mycobacterium tuberculosis H37Rv</a>				
REFERENCE	1				
AUTHORS	Lew, J.M., Kapopoulou, A., Jones, L.M. and Cole, S.T.				
TITLE	TubercuList--10 years after				
JOURNAL	Tuberculosis (Edinb) 91 (1), 1-7 (2011)				
PUBMED	<a href="#">20980199</a>				
REFERENCE	2				
AUTHORS	Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.				
TITLE	Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv				
JOURNAL	Microbiology (Reading, Engl.) 148 (PT 10), 2967-2973 (2002)				
PUBMED	<a href="#">12368430</a>				
REFERENCE	3				
AUTHORS	Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. III, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S. and				

The right sidebar contains a "Change region shown" dropdown, a "Customize view" dropdown, and a "Send to:" dropdown. Below these are links for "Analyze this sequence", "Run BLAST", and "Pick Primers". There is also a "Related information" section with links for "Assembly", "BioProject", "Protein", "PubMed", "Taxonomy", "Components (Core)", "Full text in PMC", "Gene", "Genome", "Identical GenBank Sequence", "PubMed (Weighted)", "Reference Genome BioProject", and "Representative Genome BioProject". At the bottom, there is a "LinkOut to external resources" section with links for "Dryad Digital Repository" and "Dryad Digital Repository [Dryad Digital Repository]".

Figure 1: Genoma de referencia.

The screenshot shows the NCBI SRA browser interface for the project PRJNA379070. The left sidebar contains filters for Access (Public), Source (DNA), Type (genome), Library Layout (paired), Platform (Illumina), Strategy (Genome), Data in Cloud (GS, S3), and File Type (fastq). The main search results area displays a list of 20 items, with the first five visible. Each item is a WGS of pulmonary tuberculosis isolates: Sputum, generated using Illumina HiSeq 2500. The right sidebar includes a 'Search in related databases' table, a 'Find related data' section, and a 'Recent activity' section.

Database	Access public	Access controlled	all
BioSample			
BioProject		1	1
dbGaP			
GEO Datasets			

Database	Access public	Access controlled	all
BioSample			
BioProject		1	1
dbGaP			
GEO Datasets			

Search results

Items: 1 to 20 of 200

1. WGS of pulmonary tuberculosis isolates: Sputum  
1. ILLUMINA (Illumina HiSeq 2500) run: 1.4M spots, 285.7M bases, 107.3Mb downloads  
Accession: SRX2638164

2. WGS of pulmonary tuberculosis isolates: Sputum  
2. ILLUMINA (Illumina HiSeq 2500) run: 1.7M spots, 341.8M bases, 140.6Mb downloads  
Accession: SRX2638163

3. WGS of pulmonary tuberculosis isolates: Sputum  
3. ILLUMINA (Illumina HiSeq 2500) run: 2.3M spots, 450.5M bases, 181.3Mb downloads  
Accession: SRX2638162

4. WGS of pulmonary tuberculosis isolates: Sputum  
4. ILLUMINA (Illumina HiSeq 2500) run: 4.9M spots, 736.5M bases, 288.5Mb downloads  
Accession: SRX2638161

5. WGS of pulmonary tuberculosis isolates: Sputum  
5. ILLUMINA (Illumina HiSeq 2500) run: 4.9M spots, 736.5M bases, 288.5Mb downloads  
Accession: SRX2638160

Recent activity

PRJNA379070 (200)

Mycobacterium tuberculosis H37Rv, complete genome  
Nucleotide  
gi448814763|ref|NC\_000962.3|g|JN ENOMES|135

Mycobacterium tuberculosis H37Rv, complete genome  
Nucleotide

Mycobacterium tuberculosis H37Rv, complete genome  
Nucleotide

Figure 2: Base de datos de donde se obtuvieron las 14 muestras secuenciadas.

Ahora bien, una vez reunida toda la información necesaria, comenzamos con el análisis bioinformático, para esto el primer paso fue subir todas nuestras secuencias (Figura 3) al servidor de UseGalaxy [van Heusden et al., 2023].

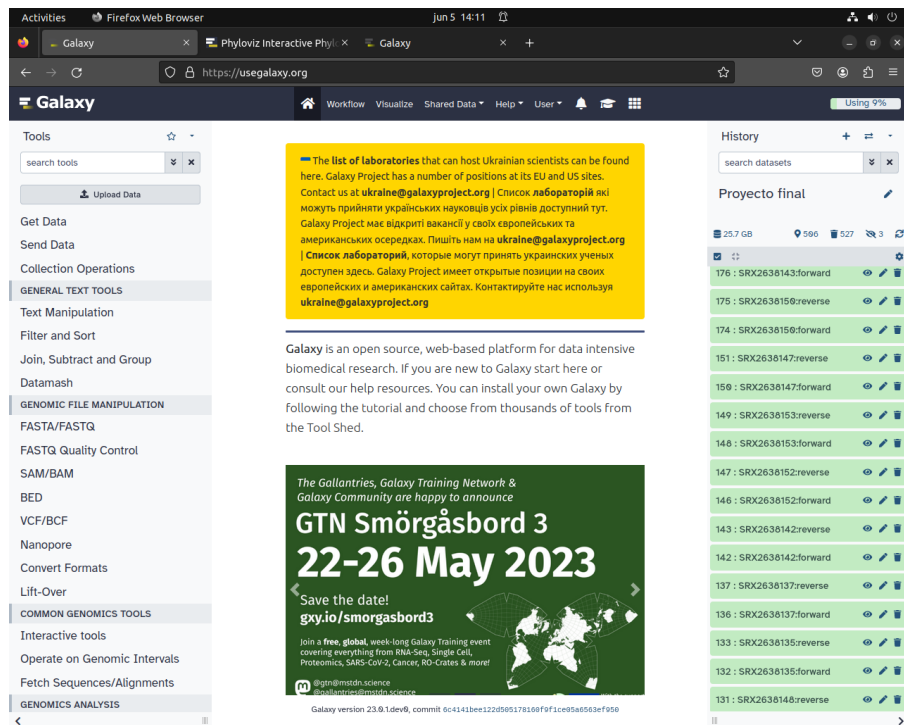


Figure 3: Muestras de *M. tuberculosis* con resistencia a antibióticos en *Galaxy*.

Una vez que subimos las muestras procedimos a realizar un análisis de calidad a cada una (Figura 4), para lo que utilizamos la herramienta FastQC con el fin de poder observar si nuestras secuencias habían sido secuenciadas de manera correcta, para esto ingresamos un archivo `.fastq` y como archivo de salida obtuvimos un `.html` y un `.txt`.

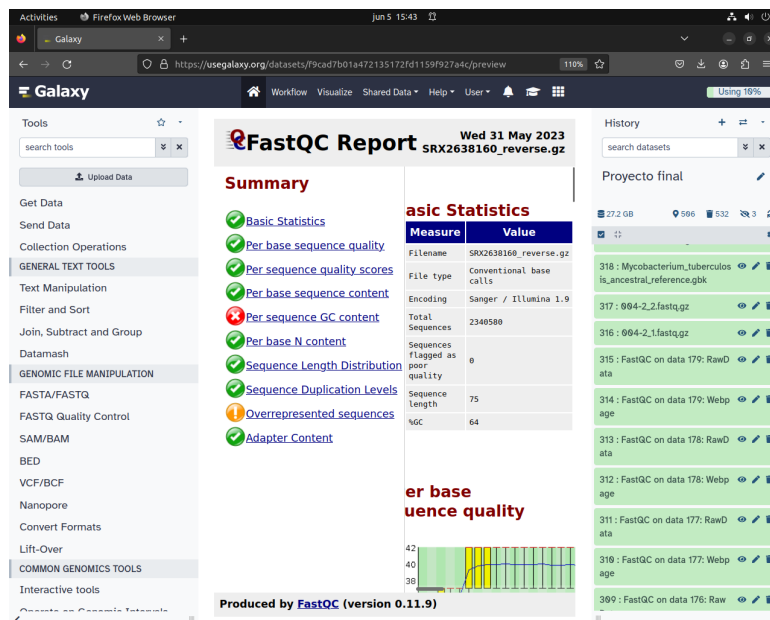


Figure 4: Ejemplo de análisis de calidad de la muestra SRX2638160.

Habiendo concluido el paso anterior, procedimos a realizar un MultiQC (Figura 5) para poder agrupar toda la información de cada uno de los FastQC y así poder observar la presencia de adaptadores, para esto como archivo de entrada utilizamos todos los `.txt` obtenidos anteriormente para así obtener un archivo `.html`.

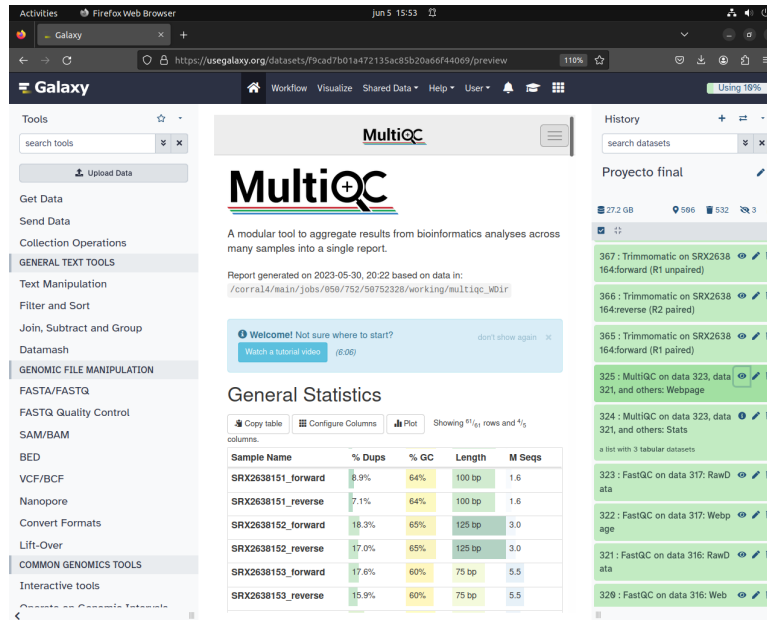


Figure 5: Análisis MultiQC de todas las muestras trabajadas.

Como siguiente paso lo que hicimos fue que, con la herramienta *Trimmomatic*, realizamos el corte de los adaptadores que anteriormente observamos con ayuda del análisis de MultiQC, para esto utilizamos dos archivos de entrada `.fastq`, uno que contenía la cadena *forward* y otro con la cadena *reverse*, esto fue con cada una de las muestras, dentro de las tareas que le pedimos a realizar a la herramienta fue un corte de ventana deslizado con un promedio de calidad de 30 y que dejara las lecturas a un largo específico, el cual fue de un largo mínimo de 20. Al finalizar este procedimiento obtuvimos 4 archivos de salida `.fastq` por cada una de las muestras que ingresamos (Figura 6).

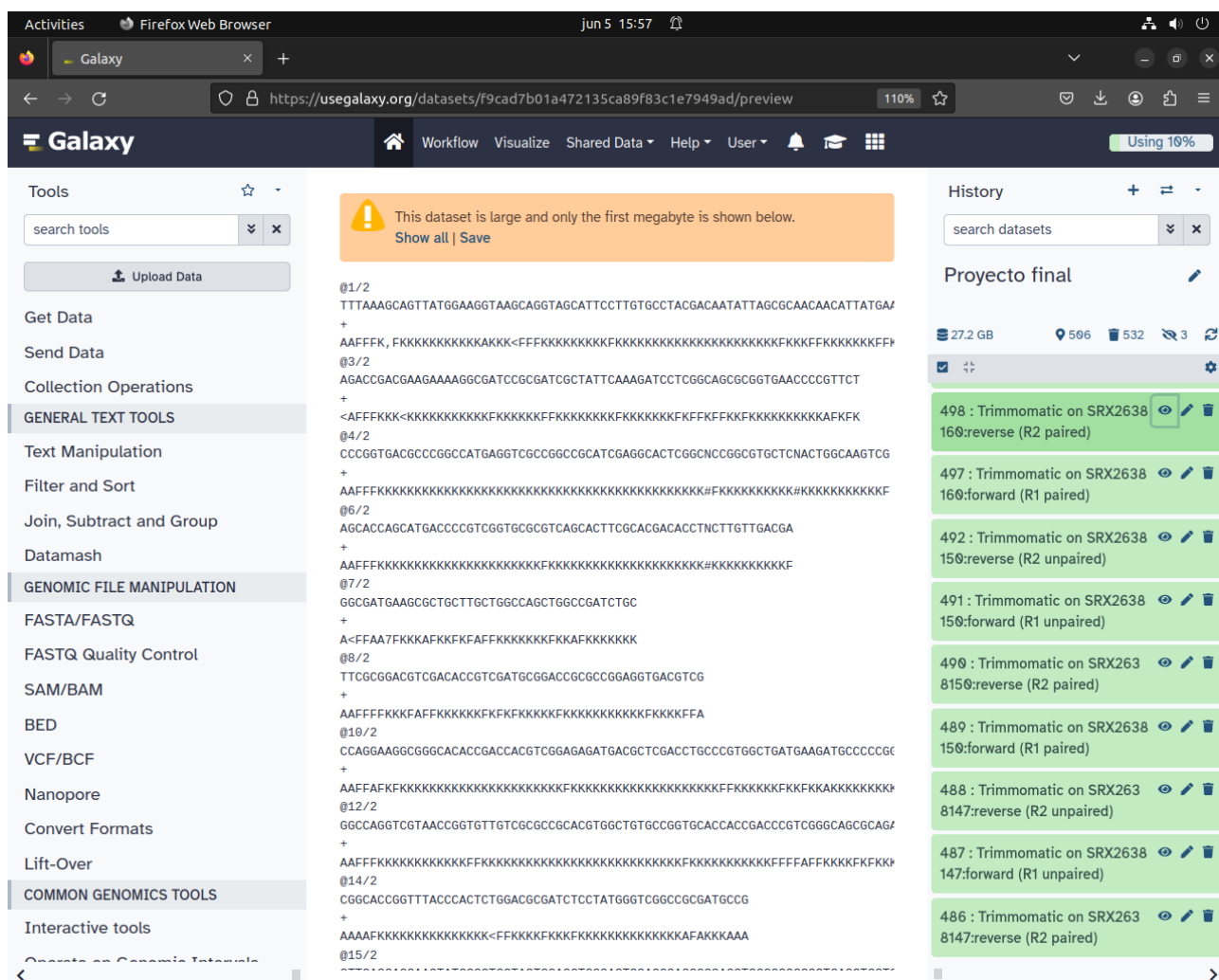


Figure 6: Corte de adaptadores en las secuencias con la herramienta *Trimmomatic*.

Como siguiente paso utilizamos la herramienta Kraken2 para el análisis de contaminación de nuestras muestras (Figura 7), para esto los archivos que ingresamos fueron *.fastq*, ingresamos dos archivos, uno de la cadena *forward* y uno de la cadena *reverse*, así mismo los archivos que seleccionamos fueron los *paired* y le pedimos a la herramienta que el reporte que nos diera de salida fuera con los nombres científicos y con los clados, así mismo la base de datos que utilizamos fue *estándar*.

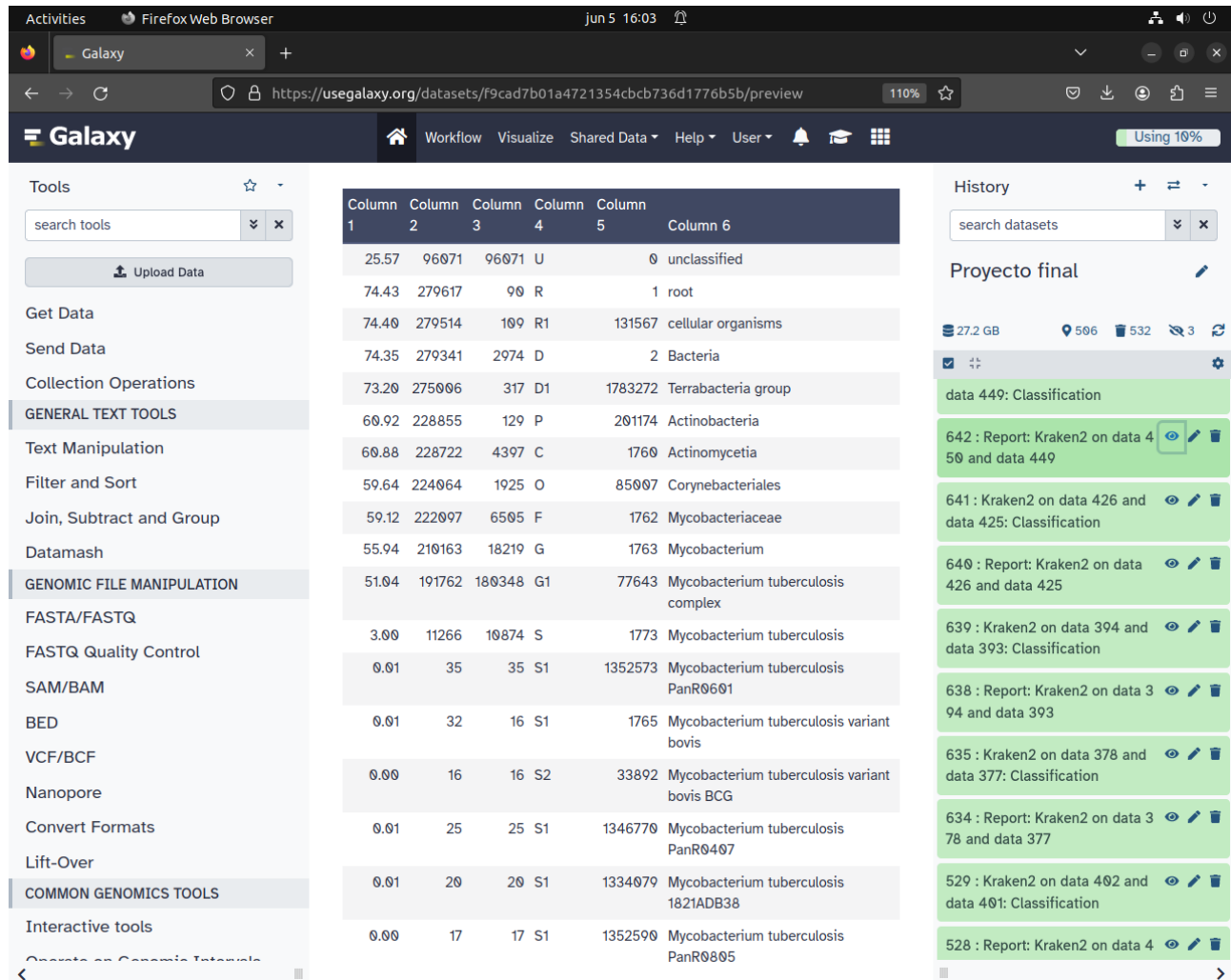


Figure 7: Análisis de calidad de una de las muestras, se observa qué genomas se encontraron dentro de la muestra utilizada



Ahora bien, el siguiente paso que realizamos fue la búsqueda de variantes/polimorfismos (SNPs) de interés en las muestras que estamos trabajando (Figura 8), para esto utilizamos la herramienta *Snippy*, a partir de la cual alineamos nuestras secuencias a un genoma de referencia que como se mencionó anteriormente, nuestro genoma de referencia fue H37Rv, para esto nuestros documentos de entrada fueron de tipo `.fastq`, ingresamos dos archivos, uno correspondiente a la cadena *forward* y otro correspondiente a la cadena *reverse*, una vez concluido el proceso, la herramienta nos dio como resultado tres tipos de archivo, uno de tipo `.bam`, otro `.vcf` y una tabla con toda la información obtenida.

The screenshot shows the Galaxy web interface. The central panel displays a table of variant calls. The table has the following columns: CHROM, POS, TYPE, REF, ALT, EVIDENCE, FTYPE, STRAND, and NT. The table lists several variants across different chromosomes.

CHROM	POS	TYPE	REF	ALT	EVIDENCE	FTYPE	STRAND	NT
Chromosome	157333	snp	C	T	T:14 C:0	CDS	-	26
Chromosome	218599	snp	T	C	C:11 T:0			
Chromosome	309765	snp	T	C	C:11 T:0	CDS	+	67
Chromosome	336005	snp	A	G	G:13 A:1	CDS	-	30
Chromosome	339230	snp	G	C	C:15 G:0			
Chromosome	428921	snp	G	A	A:13 G:0	CDS	-	57
Chromosome	830868	ins	G	GGC	GGC:12 G:0	CDS	+	24
Chromosome	841289	del	GAA	G	G:13 GAA:0	CDS	+	63
Chromosome	919551	snp	C	T	T:10 C:0	CDS	-	3/1
Chromosome	1097220	snp	C	T	T:11 C:0	CDS	+	39
Chromosome	1345895	snp	C	T	T:9 C:1	CDS	+	63

The left sidebar shows the 'Tools' menu with categories like 'GENERAL TEXT TOOLS' and 'GENOMIC FILE MANIPULATION'. The right sidebar shows the 'History' panel with a list of datasets generated by the workflow, including BAM and VCF files.

Figure 8: Llamado de variantes realizado en una de las muestras.

El siguiente paso realizado fue el filtrado de las diferentes variantes (Figura 9) que encontramos anteriormente, esto con el fin de disminuir los probables errores que se hayan presentado durante el camino, esto debido a que dentro del genoma de *M. tuberculosis* hay regiones poco efectivas donde se puedan mapear nuestras secuencias, así mismo pudieron haberse presentado delecciones o inserciones que generen un resultado difuso. Es por esto que con la herramienta *TB-variant-filter* ingresamos un documento tipo *.vcf* y realizamos un filtrado de variantes por región y otro por sitios de profundidad de alineación de lectura, una vez concluido obtuvimos como resultado archivos *.json*, *.vcf* y *.txt*; tres por cada muestra que filtramos.

The screenshot shows the Galaxy web interface in a Firefox browser. The main panel displays a VCF file preview with the following table:

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
Chromosome	1849	.	C	A	2577.47	.	AB=0.0;AO=71;DP=GENE_Rv0002   n.
Chromosome	2532	.	C	T	3604.01	.	AB=0.0;AO=99;DF=ANN=T synonymoi
Chromosome	3446	.	C	T	4007.0	.	AB=0.0;AO=110;DI=ANN=T missense_
Chromosome	9143	.	C	T	3223.11	.	AB=0.0;AO=90;DF=ANN=T synonymoi
Chromosome	11976	.	G	C	3172.34	.	AB=0.0;AO=89;DF=ANN=C synonymoi
Chromosome	12163	.	G	A	2718.59	.	AB=0.0;AO=78;DP=ANN=A missense_
Chromosome	12204	.	G	A	2625.39	.	AB=0.0;AO=73;DP=ANN=A synonymoi
Chromosome	13460	.	G	A	2860.88	.	AB=0.0;AO=79;DP=ANN=A synonymoi
Chromosome	13926	.	C	T	2229.84	.	AB=0.0;AO=61;DP=ANN=T missense_
Chromosome	14401	.	G	A	2795.14	.	AB=0.0;AO=77;DP=ANN=A missense_
Chromosome	21819	.	C	A	2991.41	.	AB=0.0;AO=85;DF=ANN=A missense_
Chromosome	23174	.	A	C	2667.61	.	AB=0.0;AO=75;DP=ANN=C missense_
Chromosome	24007	.	G	T	3784.78	.	AB=0.0;AO=103;D=ANN=T synonymoi
Chromosome	26747	.	GC	G	2988.76	.	AB=0.0;AO=82;DF=

The right panel shows the workflow history for 'Proyecto final'. It lists several steps, including '71', '722 : TB Variant Filter on data 6', '720 : TB Variant Filter on data 6', and '719 : TB Variant Filter on data 6'. The '720 : TB Variant Filter on data 6' step is highlighted, showing its parameters: '1. chrom', '##fileformat=VCFv4.2', '##FILTER=<ID=PASS, Description="All filters"', '##fileDate=20230602', '##source=freeBayes v1.3.2-dirty', and '##reference=reference/ref.fa'.

Figure 9: Filtrado de las variantes en las muestras.

Por último, con la herramienta *TB-profile-profile* lo que hicimos fue una comparación de las variantes que encontramos en nuestras muestras con respecto a las que se tienen registradas en la literatura, esto con el fin de poder conocer a qué fármacos está asociada específicamente su resistencia, al igual que conocer los tipos de mutaciones y en qué parte del cromosoma se encuentran dichas variantes (Figura 10).

**Tools**

search tools

Upload Data

Get Data

Send Data

Collection Operations

**GENERAL TEXT TOOLS**

Text Manipulation

Filter and Sort

Join, Subtract and Group

Datamash

**GENOMIC FILE MANIPULATION**

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

Convert Formats

Lift-Over

**COMMON GENOMICS TOOLS**

Interactive tools

Create a Genomic Interval

**TBProfiler report**

=====

The following report has been generated by TBProfiler.

**Summary**

-----

ID: tbprofiler  
Date: Sun Jun 4 19:31:11 2023  
Strain: lineage1.1.3.1  
Drug-resistance: MDR-TB  
Median Depth: 25

**Lineage report**

-----

Lineage	Estimated Fraction	Spoligotype	Rd
lineage1	1.000	EAI	RD239
lineage1.1	1.000	EAI3;EAI4;EAI5;EAI6	RD239
lineage1.1.3	1.000	EAI6	RD239
lineage1.1.3.1	1.000	NA	RD239

**Resistance report**

-----

Drug	Genotypic Resistance	Mutations
Rifampicin	R	rpoB p.Leu452Pro (1.00)
Isoniazid	R	katG p.Ser315Thr (1.00)
Ethambutol	R	embB p.Asp354Ala (1.00)

**History**

search datasets

**Projecto final**

27.2 GB 506 532 3

1017 : TB-Profiler Profile report on data 697

1016 : TB-Profiler Profile VCF on data 697

1015 : TB-Profiler Profile on data 697: Results.json

1014 : TB-Profiler Profile report on data 694

1013 : TB-Profiler Profile VCF on data 694

1012 : TB-Profiler Profile on data 694: Results.json

1011 : TB-Profiler Profile report on data 691

1010 : TB-Profiler Profile VCF on data 691

1009 : TB-Profiler Profile on data 691: Results.json

1005 : TB-Profiler Profile report

Figure 10: Reporte de comparación con la literatura reportada, la cual nos permite observar a qué fármacos son resistentes

# Resultados

Los pasos descritos en la metodología (el análisis de calidad, el corte de adaptadores, el análisis de contaminación, el alineamiento al genoma, el llamado de variantes y el perfilamiento) nos permitieron obtener muestras de calidad a partir de las cuales pudimos hacer una búsqueda con las mutaciones asociadas a la resistencia de fármacos reportadas en la literatura con respecto a las que encontramos en nuestras muestras analizadas, dándonos como resultado los fármacos a los que se encuentran asociadas nuestras 12 muestras de las 14 ya que 2 no mostraron variantes de resistencia. A continuación se observa un listado de los fármacos que se obtuvieron como resultado en nuestro análisis:

- Estreptomicina (Str)
- Rifampicina (Rif)
- Ofloxacina (Ofi)
- Kanamicina (Kan)
- Moxifloxacino (Mox)
- Isoniazida (Iso)
- Levofloxacino (Lev)
- Bedaquilina (Bed)
- Fluoroquinolonas (Flu)
- Clofazimina (Clo)
- Ciprofloxacina (Cip)
- Etionamida (Eti)
- Etambutol (Etam)
- Capreomicina (Cap)
- Amikacina (Ami)
- Pirazinamida (Pir)
- Capreomicina (Cap)
- Aminoglucósidos (Ami)

Muestra	Tipo de variante	Fármacos a los que es resistente.
SRX26381650	Sensitive	Ninguno
SRX26381651	Missense	Rif, Iso, Etam
SRX26381652	Missense	Mox, Lev, Rif, Etam, Bed, Clo, Str, Eti, Kan
SRX26381653	Frameshift	Str
SRX26381655	Missense	Str, Iso
SRX26381656	Sensitive	Ninguno
SRX26381657	Non coding transcript exon	Str, Kan, Cap, Ami
SRX26381658	Missense	Ofi, Mox, Lev, Flu, Cip, Str, Rif, Iso, Eta, Pir
SRX26381659	Missense	Rif, Iso
SRX26381660	Non coding transcript exon	Str
SRX26381661	Missense	Ofi, Mox, Lev, Flu, Cip
SRX26381662	Non coding transcript exon	Str.
SRX26381663	Missense	Etam
SRX26381664	Missense y Frameshift	Oflo, Mox, Levo, Flu, Cip, Ami, Cap, Ami

Table 1: Descripción del tipo de mutación que presentan las muestras y la resistencia a su fármaco correspondiente.

Una vez que concluimos la comparación de nuestras 14 muestras de *M. tuberculosis* tomadas de pacientes que presentaron una resistencia a antibióticos con respecto al genoma de referencia H37Rv y a partir del uso de herramientas bioinformáticas pudimos observar que las muestras sí presentan mutaciones asociadas a respuesta a antibióticos (Tabla 1), en este caso a una resistencia al tratamiento, en la gran mayoría se observan mutaciones sin sentido (*missense*), 12 muestras mostraron una mutación relacionada a la resistencia a antibióticos mientras 2 muestras no mostraron una mutación asociada a esta respuesta

## Discusión

A partir de lo observado en los análisis obtenidos a través de UseGalaxy, podemos concluir que desde el análisis de FastQC las muestras presentaron una buena calidad de secuenciación y de contaminación, esto nos permitió un mejor manejo de las muestras secuenciadas. Dicho lo anterior podemos concluir que 12 de las 14 muestras presentan mutaciones significativas tales como las mutaciones sin sentido, mutaciones en el marco de lectura, etc, las cuales se han reportado en la literatura que están directamente relacionadas con la respuesta a fármacos, específicamente a su resistencia, a su vez dichas mutaciones demostraron una gran tendencia a la resistencia a estreptomicina, dicha resistencia se presentó en la gran mayoría de las muestras, esto debido a que *M. tuberculosis* con el paso del tiempo ha logrado mantener las mutaciones que le benefician su supervivencia ante la administración de fármacos [Ghosh et al., 2020].

## Conclusión

Para concluir podemos decir que a partir del uso de secuenciación de siguiente generación y sus respectivas herramientas bioinformáticas, se puede realizar un análisis exhaustivo sobre la diversidad de variantes o SNPs que podemos encontrar en muestras con resistencias a fármacos, en este caso, las muestras de *M. tuberculosis* sí presentaron variantes que están directamente relacionadas con una gran diversidad de fármacos.

# Bibliografía

- [Arrigoni et al., 2022] Arrigoni, R., Ballini, A., Topi, S., Bottalico, L., Jirillo, E., and Santacroce, L. (2022). Antibiotic resistance to mycobacterium tuberculosis and potential use of natural and biological products as alternative anti-mycobacterial agents. *Antibiotics (Basel, Switzerland)*, 11(10), 1431. <https://doi.org/10.3390/antibiotics11101431>.
- [Chai et al., 2018] Chai, Q., Zhang, Y., and Liu, C. H. (2018). Mycobacterium tuberculosis: An adaptable pathogen associated with multiple human diseases. *Frontiers in cellular and infection microbiology*, 8, 158. <https://doi.org/10.3389/fcimb.2018.00158>.
- [Ghosh et al., 2020] Ghosh, A., N., S., and Saha, S. (2020). Survey of drug resistance associated gene mutations in mycobacterium tuberculosis, escape and other bacterial species. *Scientific reports*, 10(1), 8957. <https://doi.org/10.1038/s41598-020-65766-8>.
- [Hiltemann et al., 2023] Hiltemann, S., Rasche, H., and Soranzo, N. (2023). A short introduction to galaxy. <https://training.galaxyproject.org/training-material/topics/introduction/tutorials/galaxy-intro-short/slides-plain.html>. Galaxy Training!
- [Martinez and Baquero, 2000] Martinez, J. L. and Baquero, F. (2000). Mutation frequencies and antibiotic resistance. *Antimicrobial Agents and Chemotherapy*, 44(7):1771–1777. DOI: 10.1128/aac.44.7.1771-1777.2000.
- [NCBI, 2017] NCBI (2017). Mycobacterium tuberculosis h37rv, complete genome. [http://www.ncbi.nlm.nih.gov/nuccore/NC\\_000962.3](http://www.ncbi.nlm.nih.gov/nuccore/NC_000962.3). (accessed 2023-06-01).
- [NCBI, 2019] NCBI (2019). WGS of pulmonary tuberculosis isolates: Sputum - sra. [https://www.ncbi.nlm.nih.gov/sra/SRX2638164\[accn\]](https://www.ncbi.nlm.nih.gov/sra/SRX2638164[accn]). (accessed 2023-06-01).
- [Qin, 2019] Qin, D. (2019). Next-generation sequencing and its clinical application. *Cancer biology & medicine*, 16(1), 4–10. <https://doi.org/10.20892/j.issn.2095-3941.2018.0055>.
- [van Heusden et al., 2023] van Heusden, P., Gladman, S., and Lose, T. (2023). M. tuberculosis variant analysis. <https://training.galaxyproject.org/training-material/topics/variant-analysis/tutorials/tb-variant-analysis/tutorial.html>. Galaxy Training!