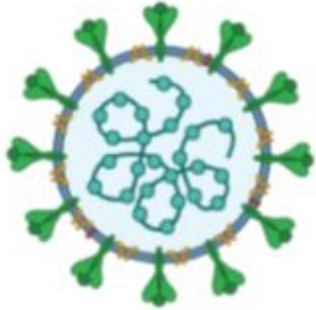


# Alineamiento genómico del SARS-CoV-2



- ❑ Diana Balmaseda Villarreal
- ❑ Gabriel Loya Sanchez
- ❑ Luis Tenorio Hernandez



# Índice

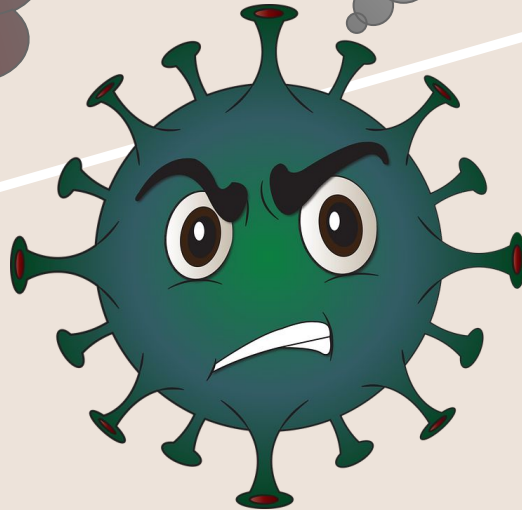
- Introducción
- Objetivos
- Métodos
- Resultados



# Introducción

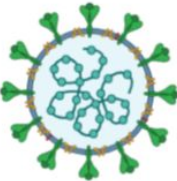
¿Por qué es importante secuenciarlo?

¿Qué es el SARS-CoV2?

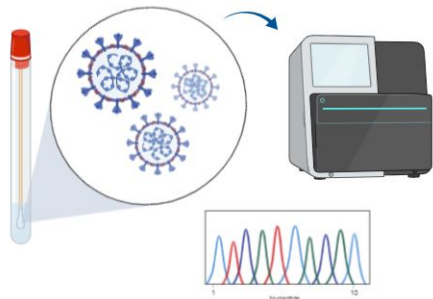


# Objetivos

- ❖ Realizar la lectura de datos crudos referentes del SARS-CoV-2 obtenidos de una base de datos de NCBI.
- ❖ Comparar los resultados obtenidos del alineamiento de lecturas con programas especializados.
- ❖ Realizar el ensamble genómico del SARS-CoV-2 a partir de los datos obtenidos de la secuenciación realizada con programas especializados.



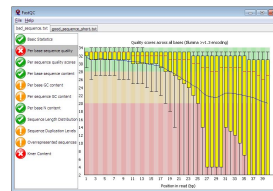
# Métodos



Secuenciación Metatranscriptómica de  
fluido broncoalveolar.  
BioProject accession number:  
**PRJNA603194.**

**sra-toolkit 2.10.5**

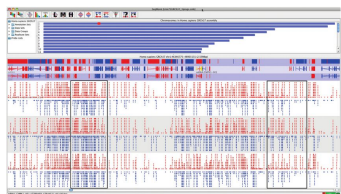
Descarga de dos archivos  
.fastq:  
**SRR10971381\_1.fastq**  
**SRR10971381\_2.fastq**



Análisis de calidad  
de las secuencias

**fastqc 0.11.9**

**SeqMonk 1.47.0**



Visualización

Lecturas  
mapeadas en  
formato SAM o  
BAM

**samtools 1.7**

**BWA 0.7.17**

**Bowtie2 2.4.1**



Mapecto Genoma de  
refrenencia: [MN908947.3](#)

Limpieza de lecturas  
de baja calidad y  
corte de adaptadores

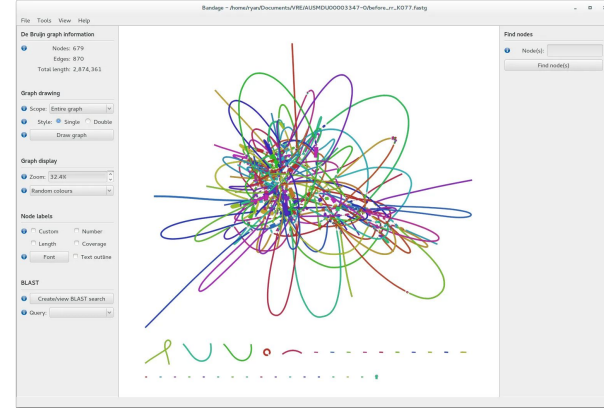
**Trimomatic 0.39**

**ERNE 1.4.4**



Ensamble del  
genoma

**SPAdes-3.13.0**



Visualización de los contigs

**Bandage**



Validación del ensamble

# Resultados

[https://github.com/dbalmaseda/Practicas\\_genomica\\_2020-2.git](https://github.com/dbalmaseda/Practicas_genomica_2020-2.git)

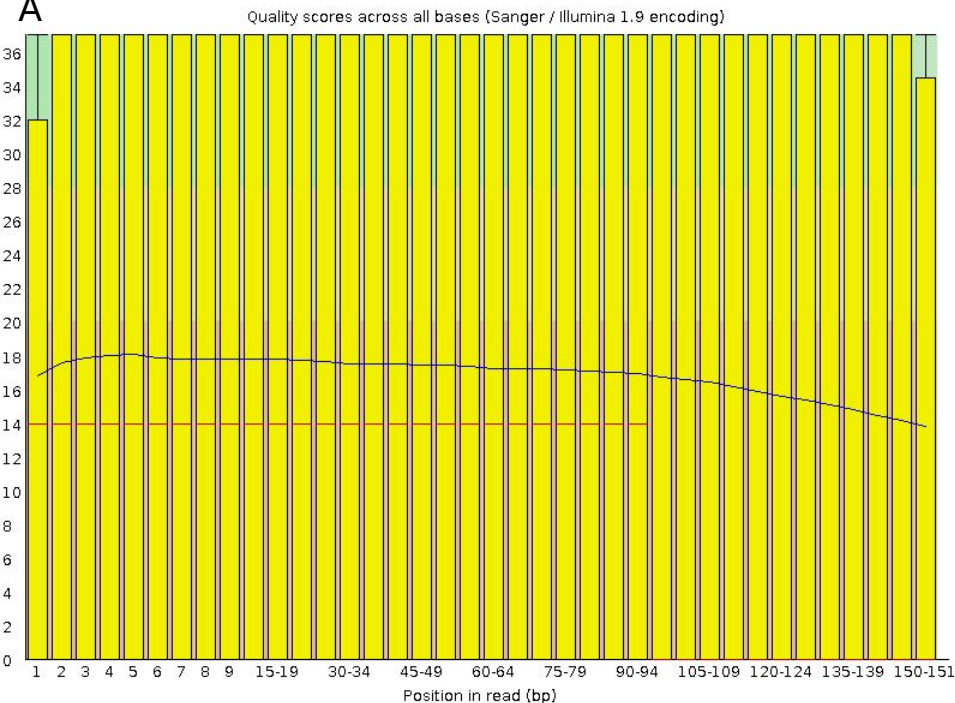




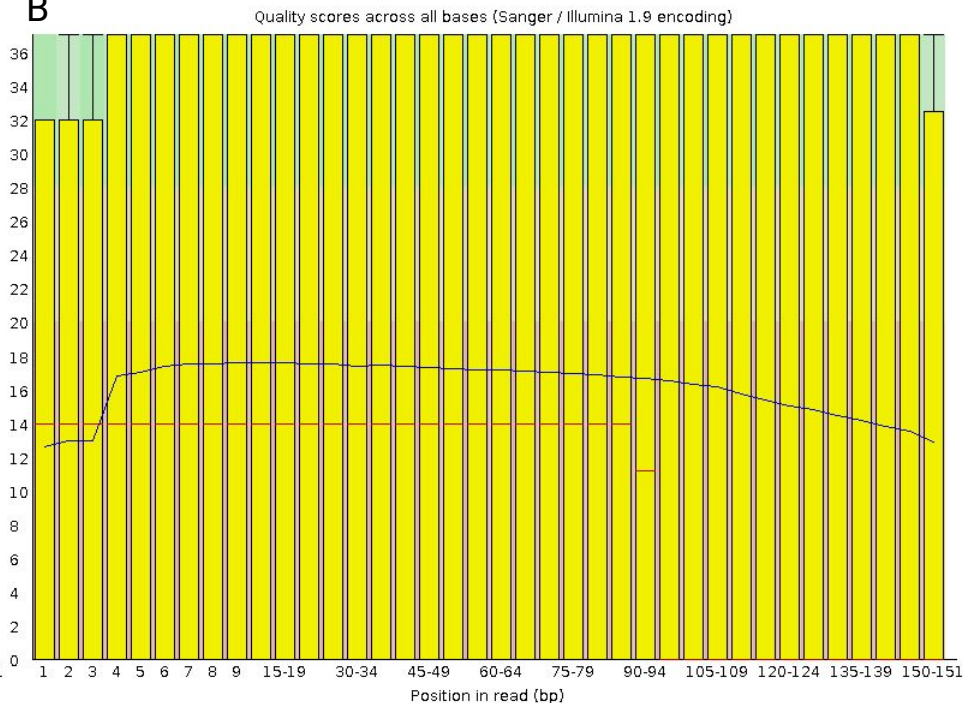
# Análisis de calidad inicial con FastQC sobre las secuencias crudas



A

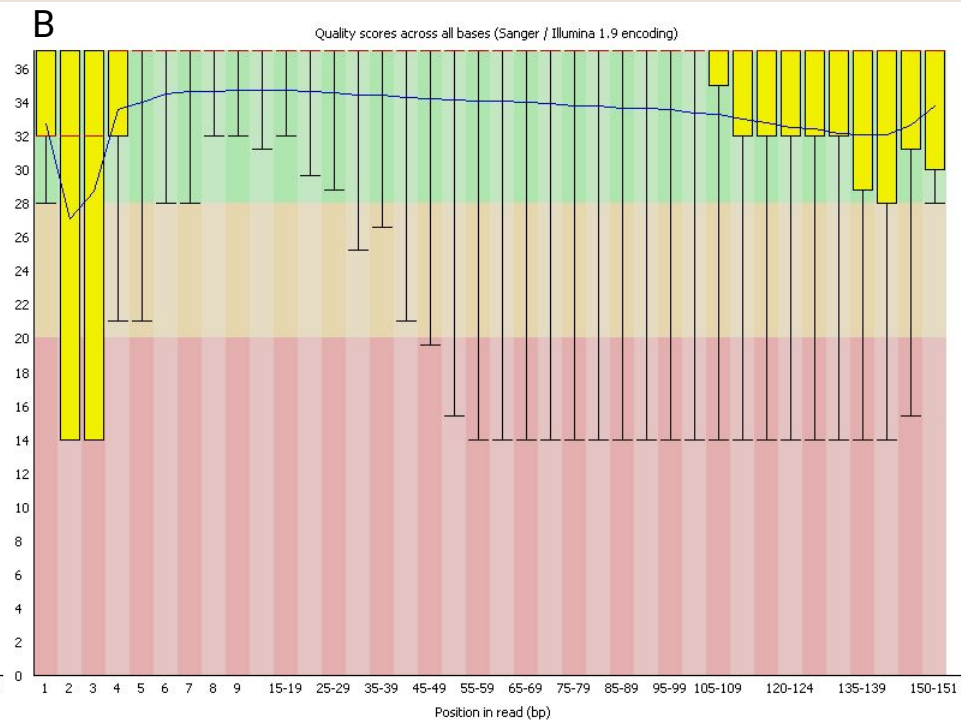
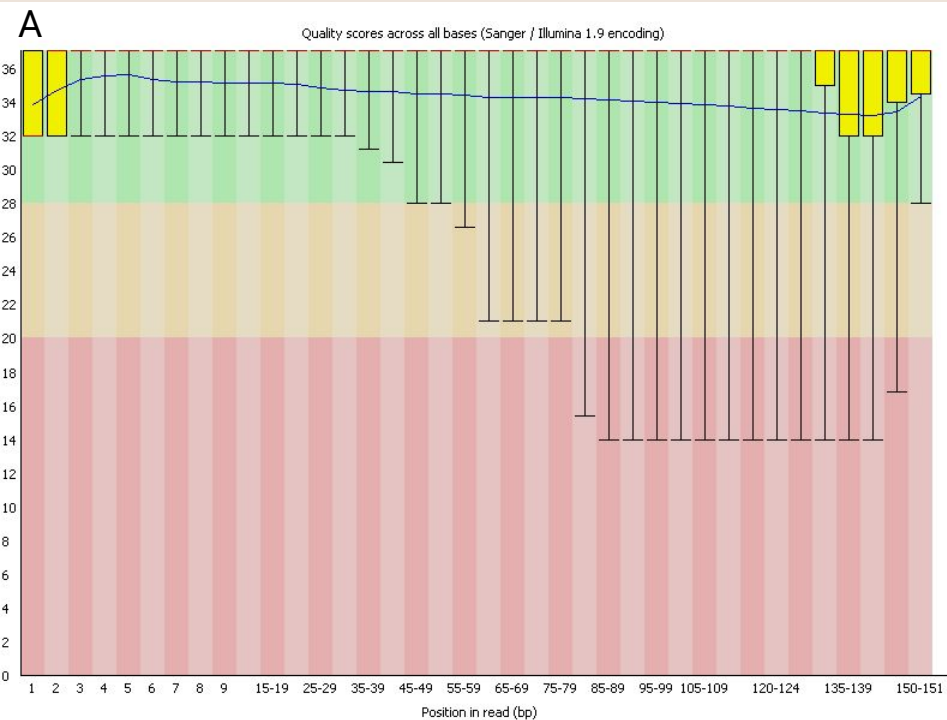


B



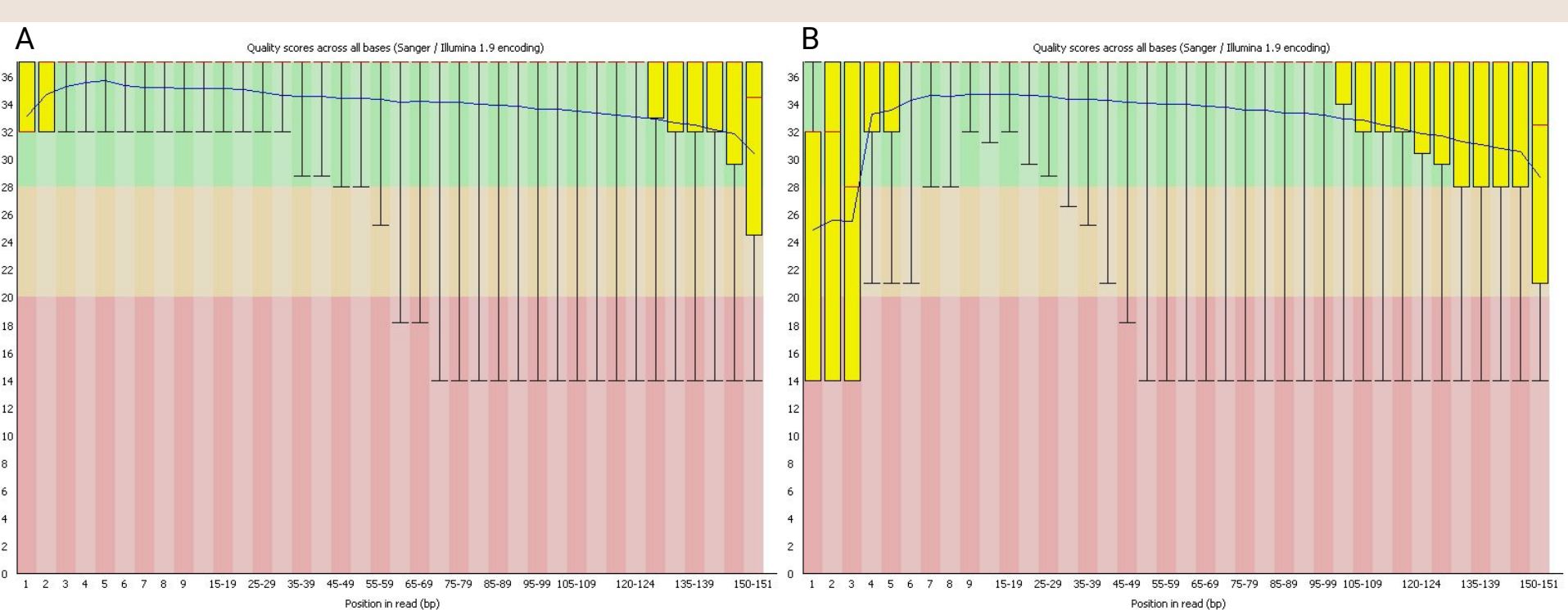


# Análisis de calidad con FastQC sobre las secuencias filtradas con ERNE

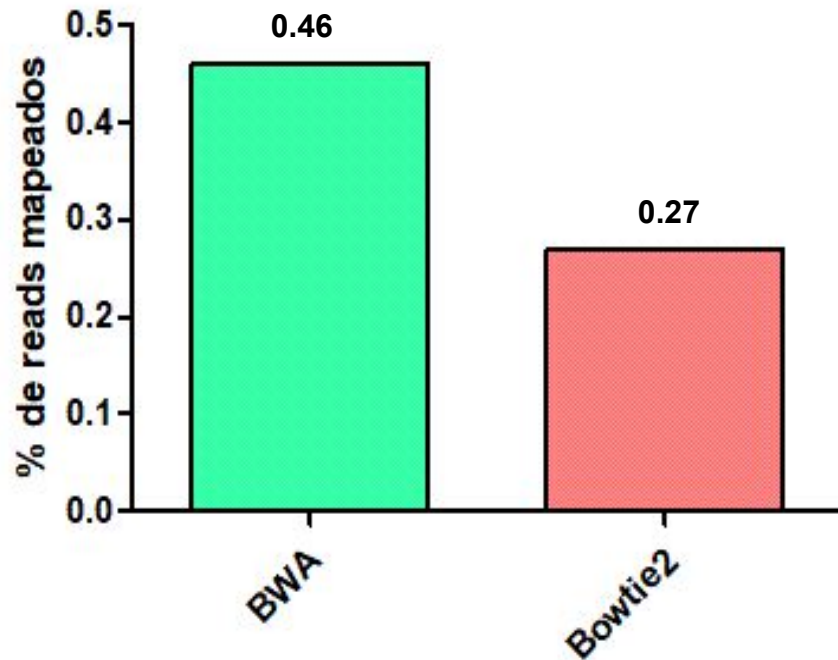




# Análisis de calidad con FastQC sobre las secuencias filtradas con Trimmomatic.



# Comparación del porcentaje de reads mapeados usando los softwares BWA y Bowtie2



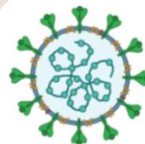
Total de lecturas: **56,624,252**

*BWA*

Lecturas alineadas: **260,042**

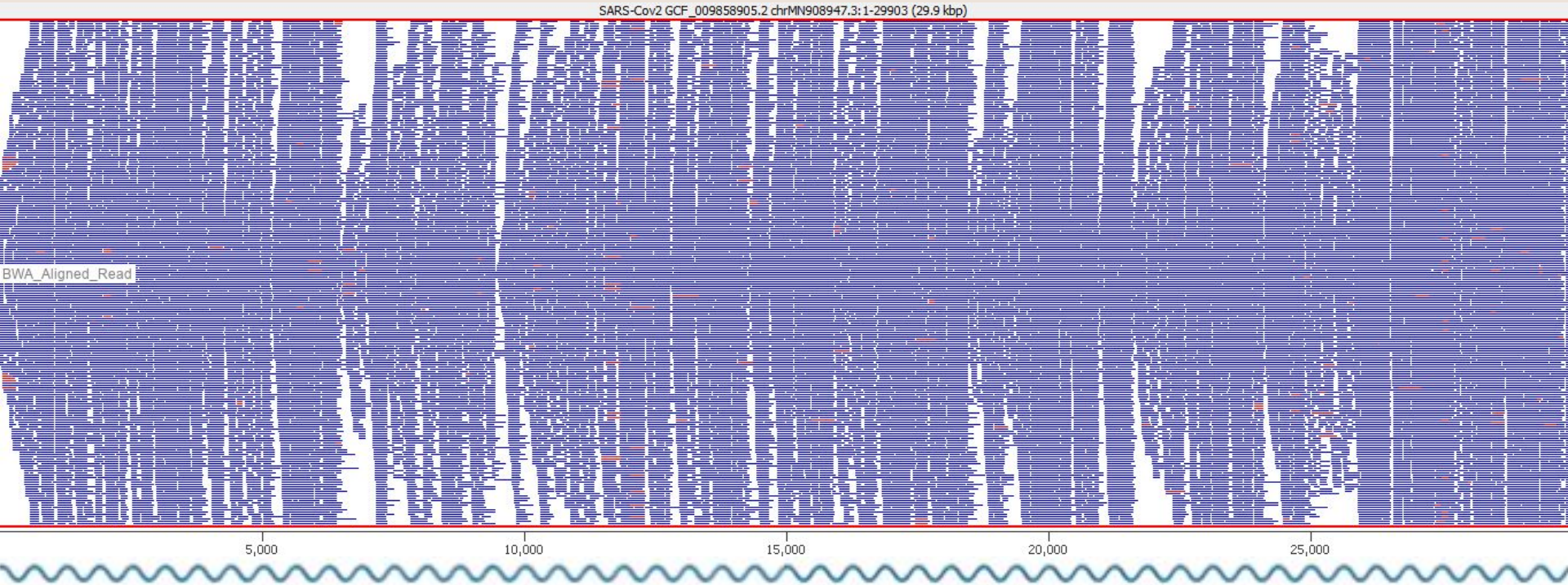
*Bowtie2*

Lecturas alineadas: **154,529**





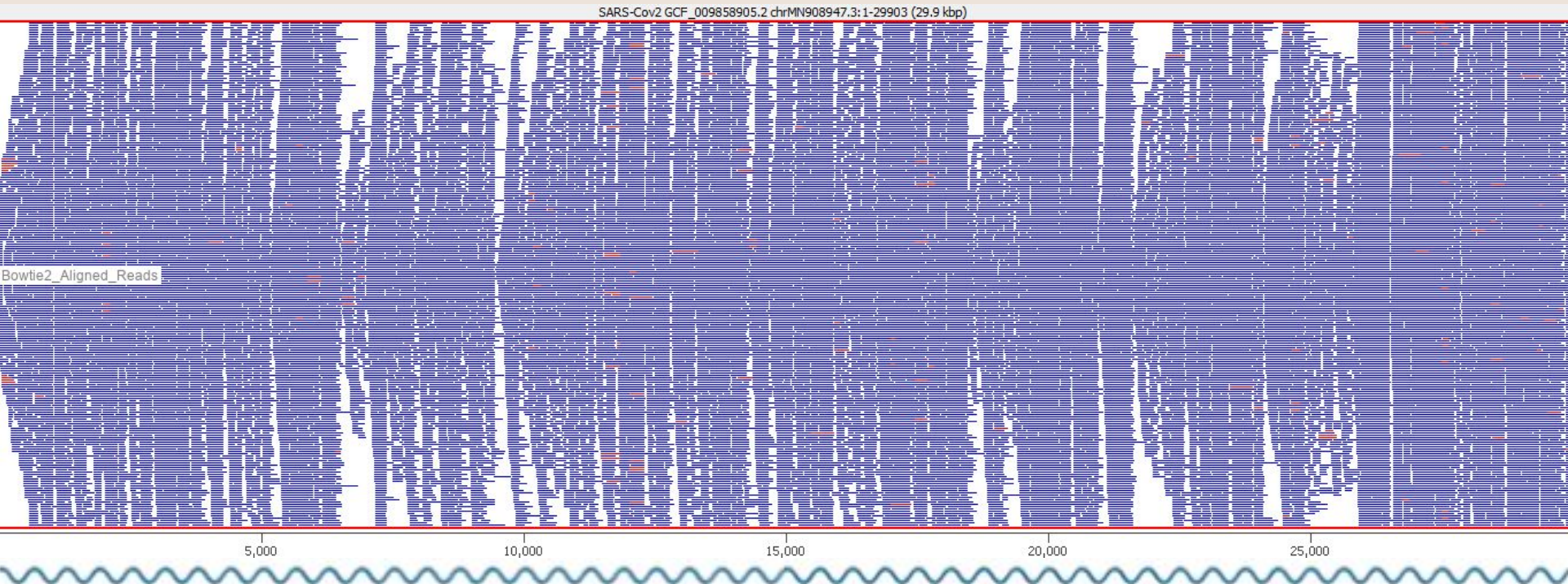
# Visualización en el software SeqMonk de lecturas mapeadas al genoma de referencia obtenidas con el software BWA







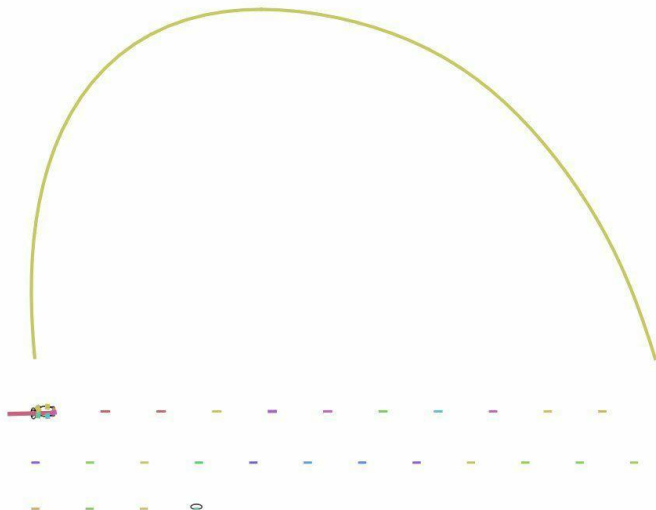
# Visualización en el software SeqMonk de lecturas mapeadas al genoma de referencia obtenidas con el software Bowtie2





# Visualización en el software Bandage del ensamble generado con SPAdes a partir de lecturas mapeadas al genoma de referencia obtenidas con el software BWA

A



**B** » blastn suite » results for RID-E9FV0PEZ016 Home Recent Results Saved Strategies Help

[< Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

**Job Title** NODE\_33272+\_length\_29906\_cov\_267.734  
**RID** [E9FV0PEZ016](#) Search expires on 06-14 19:58 pm [Download All](#) [Citation](#)  
**Program** BLASTN [See details](#)  
**Database** nt [See details](#)  
**Query ID** lcl|Query\_19315  
**Description** NODE\_33272+\_length\_29906\_cov\_267.734  
**Molecule type** dna  
**Query Length** 29906  
**Other reports** [Distance tree of results](#) [MSA viewer](#)

**Filter Results**

**Organism** only top 20 will appear ☐ exclude  
Type common name, binomial, taxid or group name  
[+ Add organism](#)

**Percent Identity**  to  **E value**  to  **Query Coverage**  to   
[Filter](#) [Reset](#)

**Descriptions** [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

**Sequences producing significant alignments** [Download](#) [Manage Columns](#) [Show](#) 100 [?](#)

☒ select all 100 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome</a>	55217	55217	99%	0.0	100.00%	<a href="#">NC_045512.2</a>
<input checked="" type="checkbox"/>	<a href="#">Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDC-0139/2020, complete genome</a>	55212	55212	99%	0.0	100.00%	<a href="#">MT481992.1</a>
<input checked="" type="checkbox"/>	<a href="#">Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/France/10009EE/2020, complete genome</a>	55212	55212	99%	0.0	100.00%	<a href="#">MT470142.1</a>
<input checked="" type="checkbox"/>	<a href="#">Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-1040/2020, complete genome</a>	55212	55212	99%	0.0	100.00%	<a href="#">MT438758.1</a>



# Visualización en el software Bandage del ensamble generado con SPAdes a partir de lecturas mapeadas al genoma de referencia obtenidas con el software bowtie2

A



B

blastn suite » results for RID-EJY3NDF9014

Home Recent Results Saved Strategies Help

[Edit Search](#) [Save Search](#) [Search Summary](#)

Job Title **NODE\_16408+\_length\_29906\_cov\_256.886**

RID [EJY3NDF9014](#) Search expires on 06-18 09:57 am [Download All](#)

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID lcl|Query\_46879

Description NODE\_16408+\_length\_29906\_cov\_256.886

Molecule type dna

Query Length 29906

Other reports [Distance tree of results](#) [MSA viewer](#)

How to read this report? [BLAST Help Videos](#) [Back to Traditional Results Page](#)

### Filter Results

Organism *only top 20 will appear* ☐ exclude

Type common name, binomial, taxid or group name

[Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

[Filter](#) [Reset](#)

**Descriptions** [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

**Sequences producing significant alignments** [Download](#) [Manage Columns](#) [Show](#) 100 [?](#)

☒ select all 100 sequences selected

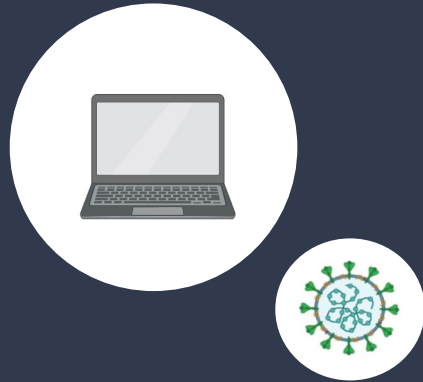
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome	55217	55217	99%	0.0	100.00%	<a href="#">NC_045512.2</a>
<input checked="" type="checkbox"/>	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CDC-0139/2020, complete genome	55212	55212	99%	0.0	100.00%	<a href="#">MT481992.1</a>
<input checked="" type="checkbox"/>	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/France/10009FE/2020, complete genome	55212	55212	99%	0.0	100.00%	<a href="#">MT470142.1</a>
<input checked="" type="checkbox"/>	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-1040/2020, complete genome	55212	55212	99%	0.0	100.00%	<a href="#">MT470142.1</a>
<input checked="" type="checkbox"/>	Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CA-CZB-1004/2020, complete genome	55212	55212	99%	0.0	100.00%	<a href="#">MT470142.1</a>

[Feedback](#)



# Conclusiones

# Resaltar la importancia de:



1. Buena calidad en los datos desde el inicio hasta el final del procesamiento
2. Documentarse acerca de los softwares existentes y tomar la mejor elección para resolver nuestros objetivos particulares
3. Conocer las posibilidades que nos ofrecen las herramientas bioinformáticas y sacar el mayor provecho de ellas



¡GRACIAS POR SU  
ATENCIÓN!

¿PREGUNTAS?