

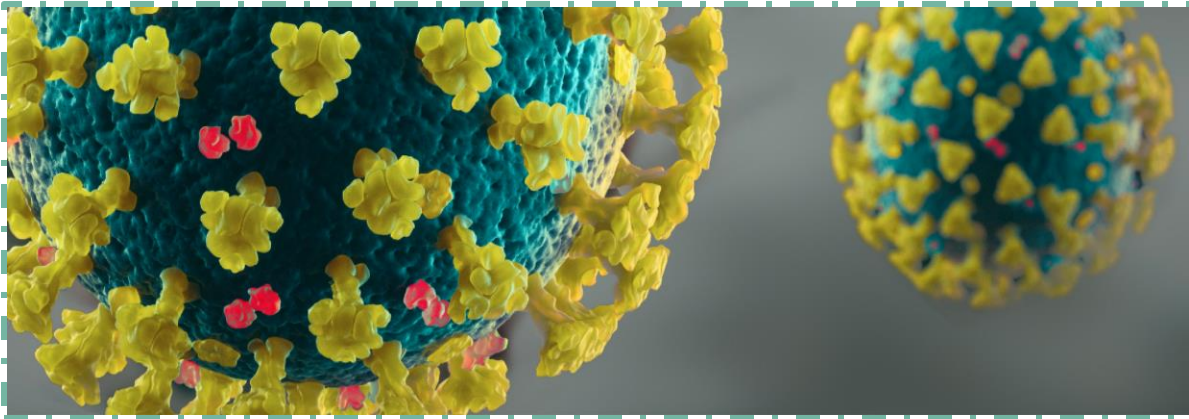
# SARS-CoV-2 vs otros coronavirus y variantes del SARS-CoV-2

Agustín García Doñate, Cristian Morillo Losada, Pedro Sánchez García & Rubén Entenza Pereira.  
Fundamentos de Bioinformática.  
MUBICS | Universidade da Coruña.



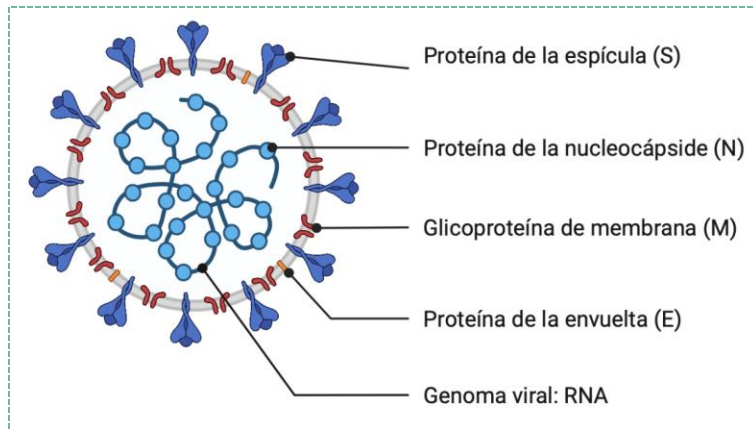
UNIVERSIDADE DA CORUÑA

# INTRODUCCIÓN

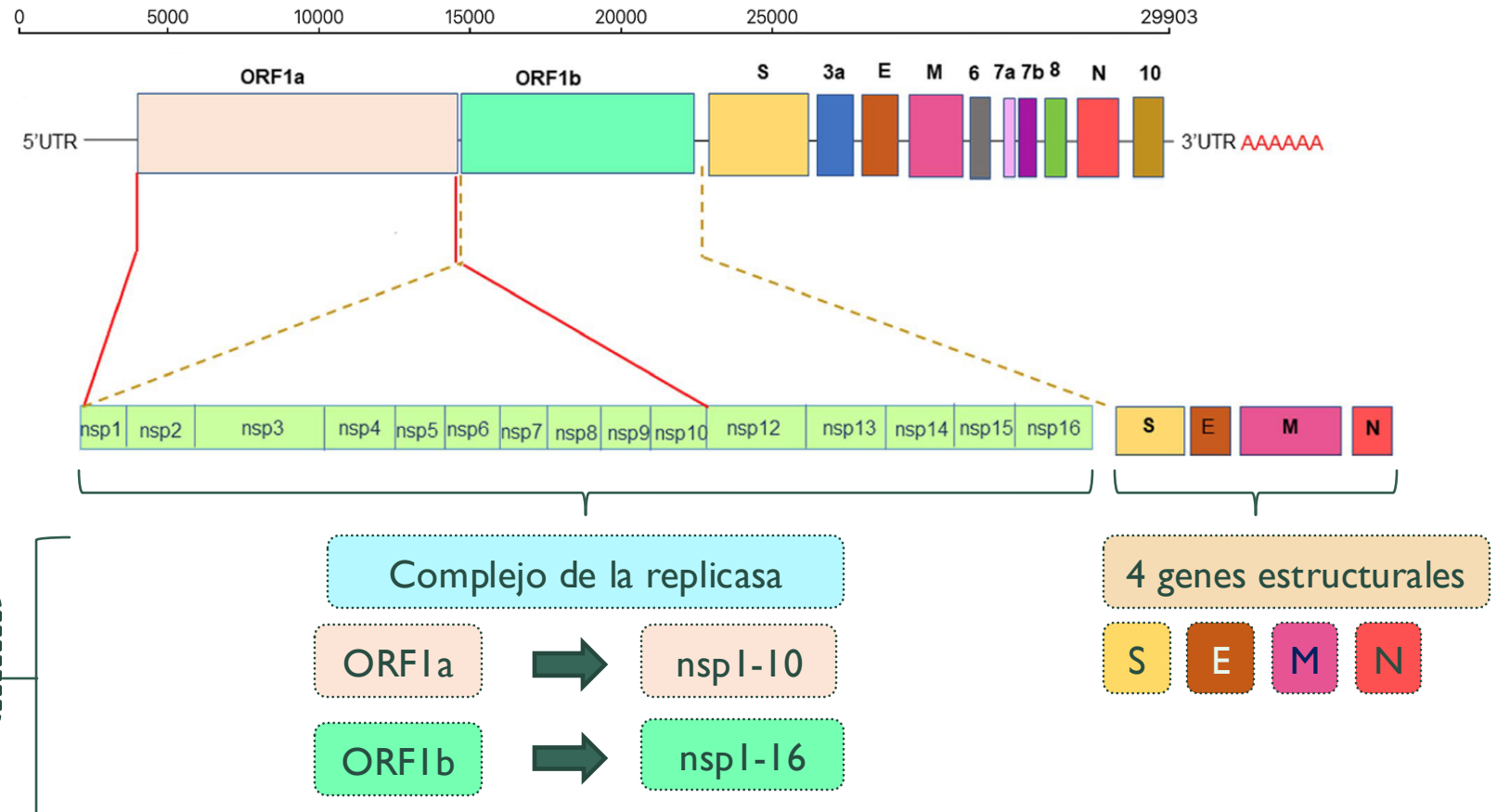


- Pandemia del coronavirus (COVID-19):
  - 290 millones de casos.
  - 5 millones de muertes.
- Género: *Betacoronavirus*.
- Familia: *Coronaviridae*.
- Mayoría de síntomas de la gripe y fiebre, ocasionando el síndrome respiratorio agudo severo y problemas derivados en otros órganos.


# INTRODUCCIÓN



Organización del genoma en SARS-CoV-2  
(aproximadamente 29.903 nucleótidos)



# INTRODUCCIÓN



**TASA DE  
MUTACIÓN  
ELEVADA**

- Los virus de ARN emplean mecanismos de variación genética para perfeccionar y asegurar su expansión.
- Las principales variantes del SARS-CoV-2 están caracterizadas por un conjunto de mutaciones que incrementan su fitness:

 **B.1.1.7 (Alpha)** **B.1.351 (Beta)** **P.1 (Gamma)** **B.1.617 (Delta Plus)** **B.1.617.2 (Delta)** **B.1.1.529 (Omicron)**



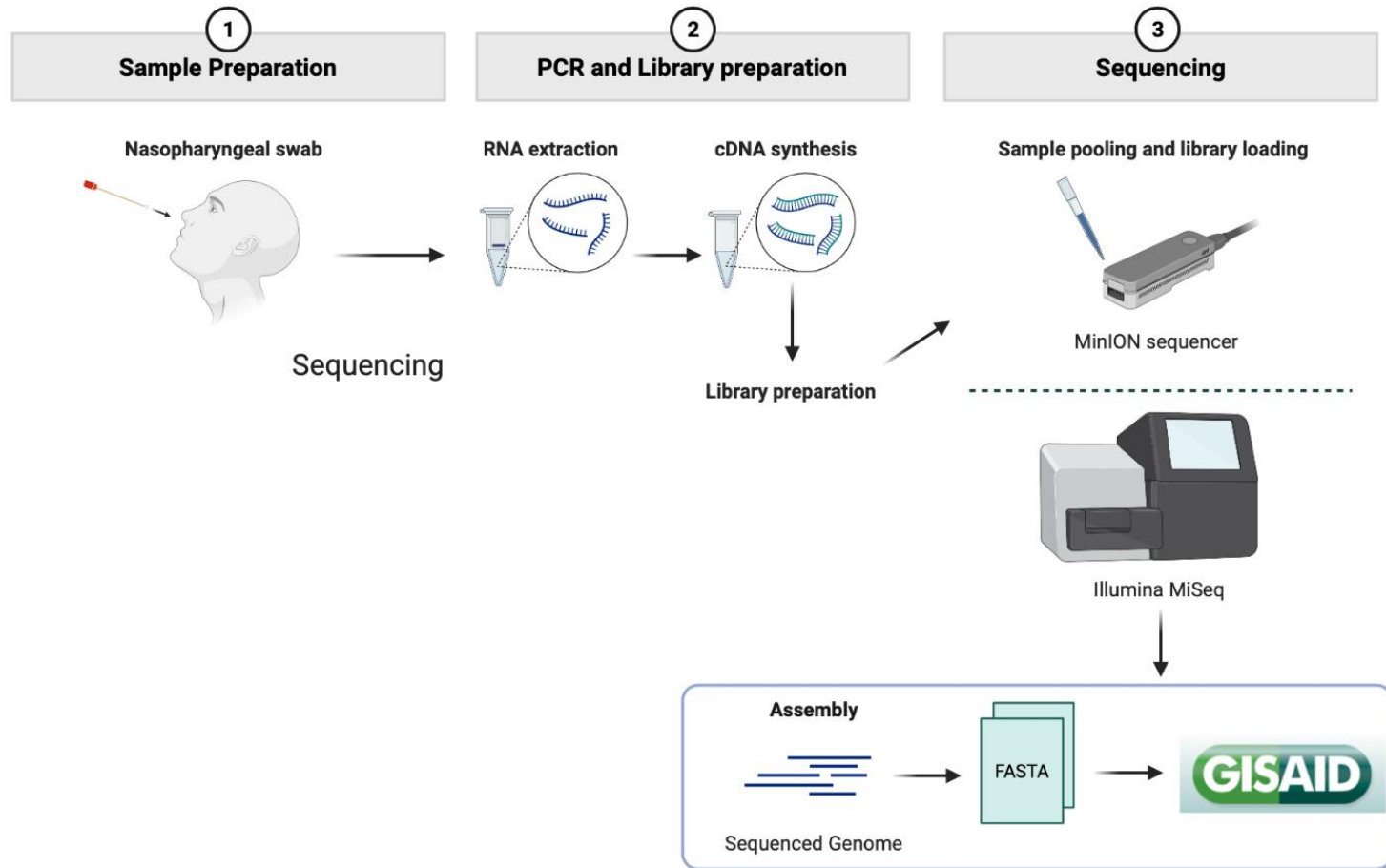
# ANÁLISIS DE DATOS DE SECUENCIACIÓN.

ALINEAMIENTO DE SECUENCIAS Y CLUSTALW.





## SARS-CoV-2 Genome Sequencing using Illumina and Oxford Nanopore Technologies



- Secuencias disponibles en GISAID.
- Genoma de referencia:  
SARS-CoV-2/Wuhan.

# Clustal Omega



**B.1.1.7 (Alpha)**



hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/USA/WA-AG-B.1.1.7/2021|EPI\_ISL\_1164753|2021-02-27

```
CTTGTTCTTACCTTTCTTTTCCAATGTTACTTGTTCCATGCTATACATGTCTCTGGGAC
CTTGTTCTTACCTTTCTTTTCCAATGTTACTTGTTCCATGCTA-----TCTCTGGGAC
*****
```

21780  
21733

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/USA/WA-AG-B.1.1.7/2021|EPI\_ISL\_1164753|2021-02-27

```
CTAAATGGTATATTAGAGTAGGAGCTAGAAAATCAGCACCTTTAATTGAATTGTGCGTGG
CTAAATGGTATATTAGAGTAGGAGCTATAAAATCAGCACCTTTA----AATTGTGCGTGG
*****
```

28080  
28026

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/USA/WA-AG-B.1.1.7/2021|EPI\_ISL\_1164753|2021-02-27

```
TCAATTTTGTAATGATCCATTTTGGGTGTTTATTACCACAAAACAACAAAAGTTGGAT
TCAATTTTGTAATGATCCATTTTGGGTGT---TTACCACAAAACAACAAAAGTTGGAT
*****
```

22020  
21970

➤ Gaps de 3-6 nucleótidos en regiones que codifican para las proteínas S y N.

# Clustal Omega



hCoV-19/Wuhan/WIV04/2019 | EPI\_ISL\_402124 | 2019-12-30  
hCoV-19/Germany/un-FLI-B.1.351/2021 | EPI\_ISL\_2695882 | 2021-02-01

```
TAGTTTGTCTGGTTTTAAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACT
TAGTTTG-----AAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACT
*****
```

11340  
11321

hCoV-19/Wuhan/WIV04/2019 | EPI\_ISL\_402124 | 2019-12-30  
hCoV-19/Germany/un-FLI-B.1.351/2021 | EPI\_ISL\_2695882 | 2021-02-01

```
TAACATCACTAGGTTTCAAACCTTACTTGCTTTACATAGAAGTTATTTGACTCCTGGTGA
TAACATCACTAGGTTTCAA-----CTTTACATAGAAGTTATTTGACTCCTGGTGA
*****
```

22320  
22292

➤ Gaps de 9 nucleótidos en ORF1a y región que codifica para la proteína S.



# Clustal Omega



**P.1 (Gamma)**



hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/France/14\_Gamma/2021|EPI\_ISL\_4536760|2021-03-21

```
TAGTTTGTCTGGTTTTAAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACT
TAGTTTG-----AAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACT
*****
```

11340  
11277

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/France/14\_Gamma/2021|EPI\_ISL\_4536760|2021-03-21

```
CG----AACAACTAAAATGTCTGATAATGGACCCCAAATCAGCGAAATGCACCCCGCA
CGAACAAACAACTAAAATGTCTGATAATGGACCCCAAATCAGCGAAATGCACCCCGCA
**
```

28316  
28257

➤ Gap of 9 nucleótidos en ORF1a e inserción de 4 nucleótidos en región que codifica para la proteína N.

# Clustal Omega



**B.1.617 (Delta Plus)**



hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/Switzerland/ZH-ETHZ-722plus/2020|EPI\_ISL\_483669|2020-04-12

TGAGTTTAAATTGGCTTCACATATGTATTGTTCTTTCTACCCTCCAGATGAGGATGAAGA 3060  
TGAGTTTAAATTGGCTTCACATATGTATTGTTCTTTTACCCTCCAGATGAGGATGAAGA 3030  
\*\*\*\*\*

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/Switzerland/ZH-ETHZ-722plus/2020|EPI\_ISL\_483669|2020-04-12

ATACAAGCCTCACTCCCTTTCGGATGGCTTATTGTTGGCGTTGCACTTCTTGCTGTTTT 25560  
ATACAAGCCTCATTCCCTTTCGGATGGCTTATTGTTGGCGTTGCACTTCTTGCTGTTTT 25530  
\*\*\*\*\*

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/Switzerland/ZH-ETHZ-722plus/2020|EPI\_ISL\_483669|2020-04-12

TAACTTTAGCTTGTTTTGTGCTTGCTGCTGTTTACAGAATAAATTGGATCACCGGTGGAA 26760  
TAACTTTAGCTTGTTTTGTGCTTGCCGCTGTTTACAGAATAAATTGGATCACCGGTGGAA 26730  
\*\*\*\*\*

➤ Mutaciones en regiones que codifican para las proteínas E y M.

# Clustal Omega



**B.1.617.2 (Delta)**



hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/Belgium/AZDelta-2149-16981/2021|EPI\_ISL\_7723671|2021-12-09

```
GGAAAGTGAGTTCAGAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCA
GGAAAGTG-----GAGTTTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCA
*****
```

22080  
22049

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
hCoV-19/Belgium/AZDelta-2149-16981/2021|EPI\_ISL\_7723671|2021-12-09

```
CGTTCATGAAGACTTTTTAGAGTATCATGACGTTTCGTGTTGTTTTAGATTTTCATCTAAA
CGTTCATGAAGACTTTTTAGAGTATCATGACGTTTCGTGTTGTTT-----ATCTAAA
*****
```

28260  
28223

➤ Gaps de 6 y 7 nucleótidos en las regiones que codifican para las proteínas S y N.

# Clustal Omega

 **B.1.1.529 (Omicron)**



hCoV-19/Wuhan/WIV04/2019 | EPI\_ISL\_402124 | 2019-12-30  
OM095411.1

```
AGGAGACATTATACTTAAACCAGCAAATAATAGTTTAAAAATTACAGAAGAGGTTGGCCA
AGGAGACATTATACTTAAACCAGCAAATAATA---TAAAAATTACAGAAGAGGTTGGCCA
*****
```

6540  
6501

hCoV-19/Wuhan/WIV04/2019 | EPI\_ISL\_402124 | 2019-12-30  
OM095411.1

```
TAGTTTGTCTGGTTTTAAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACT
TA-----GTTTTAAGCTAAAAGACTGTGTTATGTATGCATCAGCTGTAGTGTTACT
**                *****
```

11340  
11292

hCoV-19/Wuhan/WIV04/2019 | EPI\_ISL\_402124 | 2019-12-30  
OM095411.1

```
TCAATTTTGTAAATGATCCATTTTGGGTGTTTATTACCACAAAAACAACAAAGTTGGAT
TCAATTTTGTAAATGATCCATTTTGG-----ACCACAAAAACAACAAAGTTGGAT
*****
```

22020  
21957

➤ Gaps en ORF1a y ORF1b

# Clustal Omega

## B.1.1.529 (Omicron)

➤ Numerosas mutaciones clave en  
región que codifica para proteína S.



➤ “Frankenstein” de coronavirus.

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

hCoV-19/Wuhan/WIV04/2019|EPI\_ISL\_402124|2019-12-30  
OM095411.1

TAATATTACAACTTGTGCCCTTTTGGTGAAGTTTTTAACGCCACCAGATTTGCATCTGT 22611  
TAATATTACAACTTGTGCCCTTTTGAAGTTTTTAACGCCACCAATTTGCATCTGT 22554  
\*\*\*\*\*

TTATGCTTGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAA 22671  
TTATGCTTGAACAGGAAGAGAATCAGCAACTGTGTTGCTGATTATTCTGTCCTATATAA 22614  
\*\*\*\*\*

TTCCGCATCATTTTCCACTTTTAAGTGTTATGGAGTGTCTCCTACTAAATTAATGATCT 22731  
TCTCGCACCATTTTCACTTTTAAGTGTTATGGAGTGTCTCCTACTAAATTAATGATCT 22674  
\* \*\*\*\* \*\*\*\*\*

CTGCTTTACTAATGTCTATGCAGATTCATTGTGAATTAGAGGTGATGAAGTCAGACAAAT 22791  
CTGCTTTACTAATGTCTATGCAGATTCATTGTGAATTAGAGGTGATGAAGTCAGACAAAT 22734  
\*\*\*\*\*

CGCTCCAGGGCAAACCTGGAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTAC 22851  
CGCTCCAGGGCAAACCTGGAAGATTGCTGATTATAATTATAAATTACCAGATGATTTTAC 22794  
\*\*\*\*\*

AGGCTGCGTTATAGCTTGAATTCTAACAATCTTGATTCTAAGGTTGGTGAATTATAA 22911  
AGGCTGCGTTATAGCTTGAATTCTAACAAGCTTGATTCTAAGGTTAGGTTGAATTATAA 22854  
\*\*\*\*\*

TTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAAC 22971  
TTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAGATATTTCAAC 22914  
\*\*\*\*\*

TGAAATCTATCAGGCCGGTAGCACACCTTGTAATGGTGTGGAAGGTTTTAATTGTTACTT 23031  
TGAAATCTATCAGGCCGGTAACAAACCTTGTAATGGTGTGCAAGTTTTAATTGTTACTT 22974  
\*\*\*\*\*

TCCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTGTTACCAACCATACAGAGT 23091  
TCCTTTACGATCATATAGTTTCCGACCCACTATGGTGTGTTACCAACCATACAGAGT 23034  
\*\*\*\*\*

AGTAGTACTTTCTTTTGAACCTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTC 23151  
AGTAGTACTTTCTTTTGAACCTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTC 23094  
\*\*\*\*\*

TACTAATTTGGTTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAAAGGCACAGG 23211  
TACTAATTTGGTTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAAAGGCACAGG 23154  
\*\*\*\*\*

# Clustal Omega



Results for job clustalo-l20220104-103916-0215-92048241-p1m

Alignments

Result Summary

Phylogenetic Tree

Results Viewers

Submission Details

Download Phylogenetic Tree Data



- Archivo de árbol filogenético (.dnd) construido en base al alineamiento de secuencias.

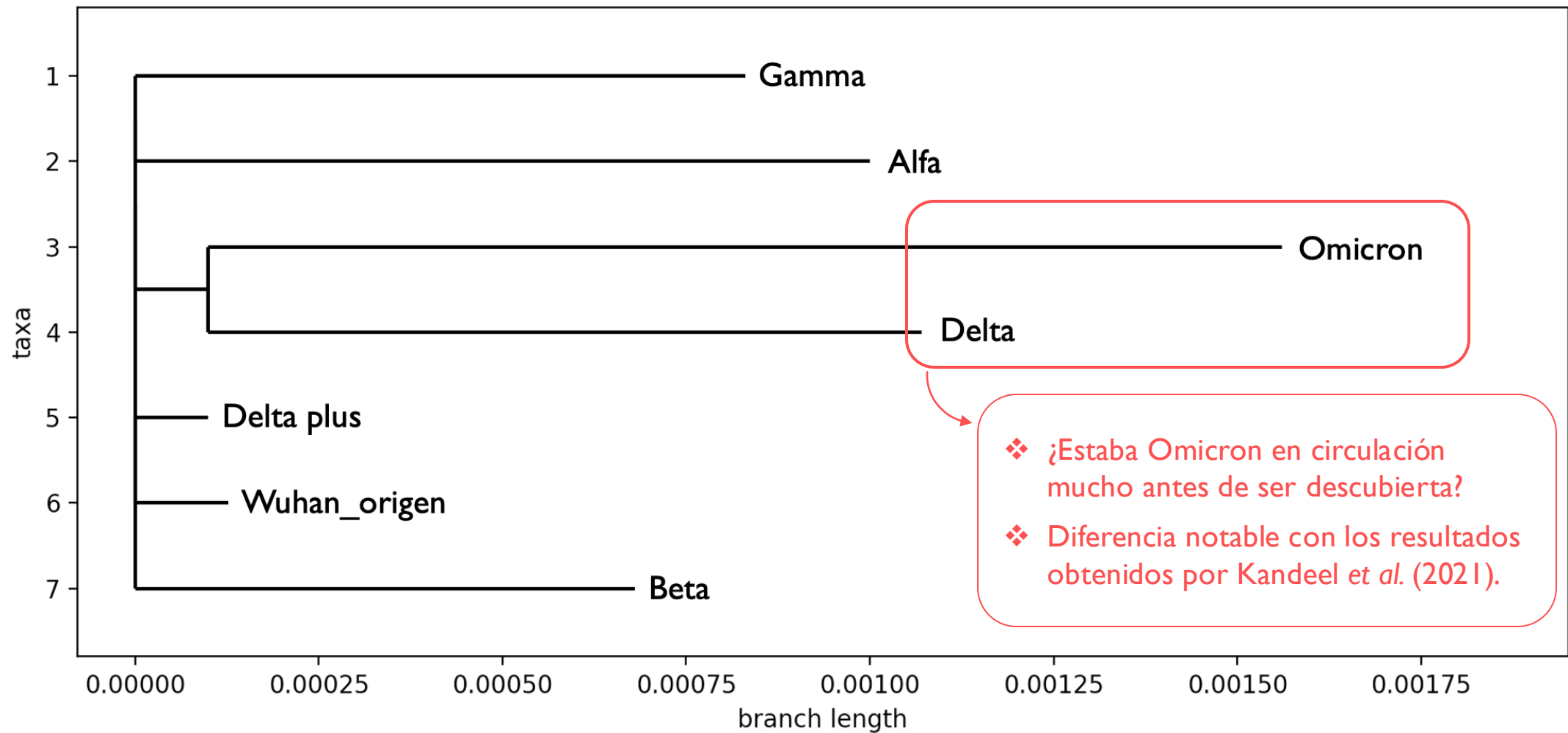


```
# CLUSTALW
```

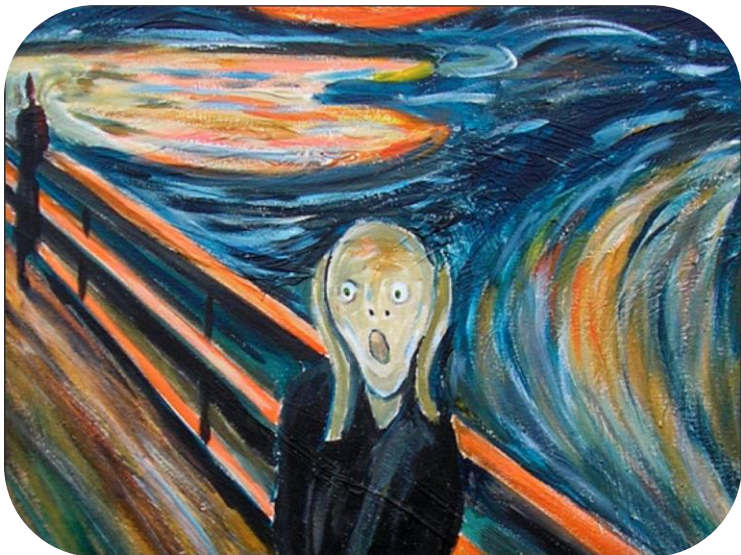
```
# Phylogenetic tree
```

```
import Bio
from Bio import Phylo
```

```
tree = Phylo.read("arbol.dnd", "newick")
tree.rooted = True
Phylo.draw(tree)
```



# ¿QUÉ SUCEDE?



N

NicolaDeMaio

14  May '20

## Issues with SARS-CoV-2 sequencing data

Nicola De Maio<sup>1\*</sup>, Conor Walker<sup>1</sup>, Rui Borges<sup>2</sup>, Lukas Weilguny<sup>1</sup>, Greg Slodkowitz<sup>3</sup>, Nick Goldman<sup>1</sup>

<sup>1</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridgeshire, United Kingdom.

<sup>2</sup>Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, Wien 1210, Austria.

<sup>3</sup>MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, United Kingdom.

\*[demaio@ebi.ac.uk](mailto:demaio@ebi.ac.uk)



- Cobertura de la plataforma de secuenciación.
- Secuencias altamente homólogas que conducen a ruido en los análisis filogenéticos.





## Alineamiento de pares de secuencias (módulo Bio.Align)







```
# PAIRWISE ALIGNER
```

```
import Bio
from Bio import Align
from Bio import SeqIO
```

```
def porcentaje_alineamiento_genoma(fasta1, fasta2='Wuhan.fasta'):
    '''Función planteada con las líneas de comando y parámetros del módulo Bio.Align.
    Se obtiene una puntuación proporcionada por el algoritmo para los alineamientos.
    La función requiere 2 archivos FASTA como parámetros, donde el primero contiene la secuencia de la variante
    , mientras que el segundo es el archivo FASTA con el genoma de referencia del SARS-CoV-2 original de Wuhan.
    Como resultado, la función nos devuelve una puntuación global y la puntuación en porcentaje de similitud en alineamiento'''
    secuencia_variante = SeqIO.read(fasta1, 'fasta')
    secuencia_origen = SeqIO.read(fasta2, 'fasta')
    aligner = Align.PairwiseAligner()
    aligner.match_score = 1
    aligner.mismatch_score = 0
    aligner.gap_score = -2.5
    len_secuencia_origen = len(secuencia_origen.seq)
    score = aligner.score(secuencia_variante.seq, secuencia_origen.seq)
    scoreporc = score * 100 / len_secuencia_origen
    return score, round(scoreporc, 2)
```



## Puntuaciones de los alineamientos - Resultados

seq2	seq1	score	score (%)
Reference genome: SARS-CoV-2/Wuhan	 <b>B.1.1.7 (Alpha)</b>	29.480.5	98.63
	 <b>B.1.351 (Beta)</b>	29.536.0	99.09
	 <b>P.1 (Gamma)</b>	29.174.0	97.51
	 <b>B.1.617 (Delta Plus)</b>	29.704.0	99.37
	 <b>B.1.617.2 (Delta)</b>	29.618.5	98.81
	 <b>B.1.1.529 (Omicron)</b>	29.435.5	98.48



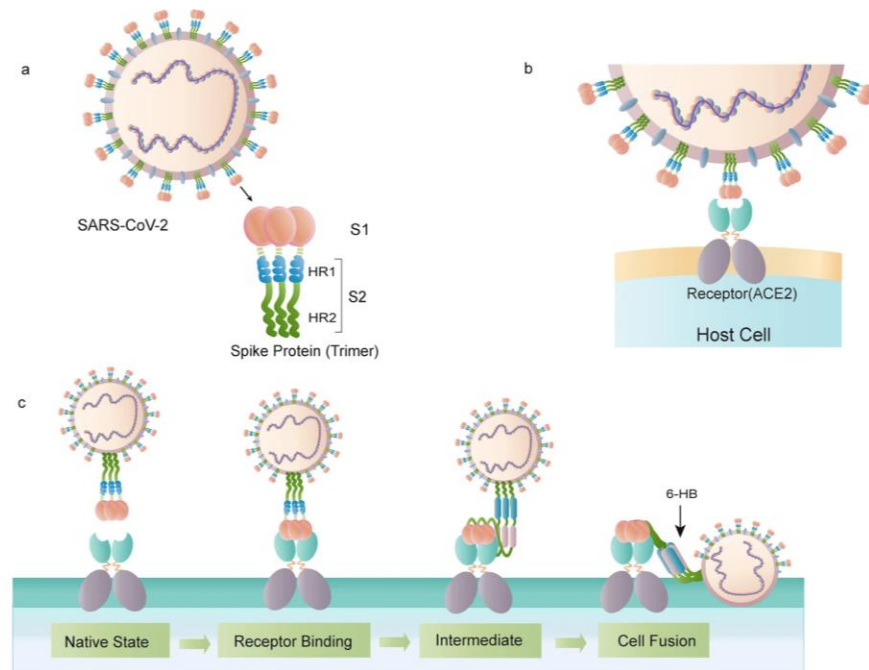
# ANÁLISIS PROTEICO

PAIRWISE SEQUENCE ALIGNMENT & MEGAX.

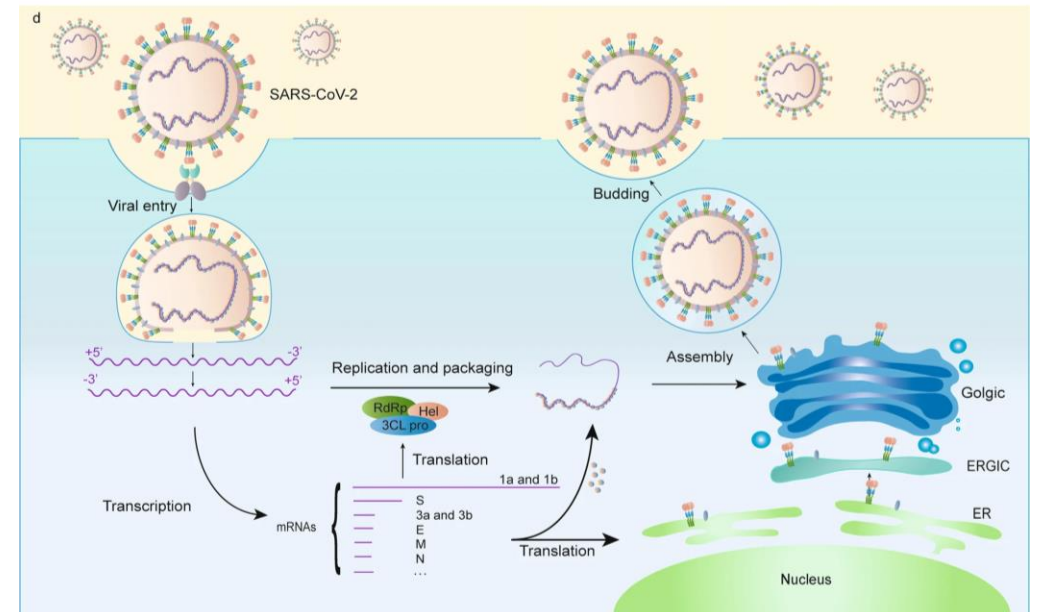


## SPIKE GLYCOPROTEIN (PROTEIN S)

- Proteína S forma un trímero y es responsable de la entrada del virus a las células (unión y fusión).
- Mutaciones en esta proteína varían la forma en la que el virus infecta a la célula



- Los archivos FASTA con la secuencia proteica de la proteína S de las variantes alfa, beta, gamma, delta y omicron fueron obtenidos en PDB





## S protein alignment (Bio.Align module)






```
# PAIRWISE PROTEIN ALIGNER
```

```
import Bio
from Bio import Align
from Bio import SeqIO
```

```
def porcentaje_alineamiento_proteinas(fasta1, fasta2='S_original.fasta'):
    '''Función planteada con las líneas de comando y parámetros del módulo Bio.Align.
    Se obtiene una puntuación proporcionada por el algoritmo para los alineamientos de las proteínas.
    La función requiere 2 archivos FASTA como parámetros, donde el primero contiene la secuencia
    de la proteína S de la variante, mientras que el segundo es el archivo FASTA con la proteína S
    de referencia del SARS-CoV-2 original de Wuhan.
    Como resultado, la función nos devuelve una puntuación global y la puntuación en porcentaje de similitud en alineamiento'''
    secuencia_variante = SeqIO.read(fasta1, 'fasta')
    secuencia_origen = SeqIO.read(fasta2, 'fasta')
    aligner = Align.PairwiseAligner()
    aligner.match_score = 1
    aligner.mismatch_score = 0
    aligner.gap_score = -2.5
    len_secuencia_origen = len(secuencia_origen.seq)
    score = aligner.score(secuencia_variante.seq, secuencia_origen.seq)
    scoreporc = score * 100 / len_secuencia_origen
    return score, round(scoreporc, 2)
```

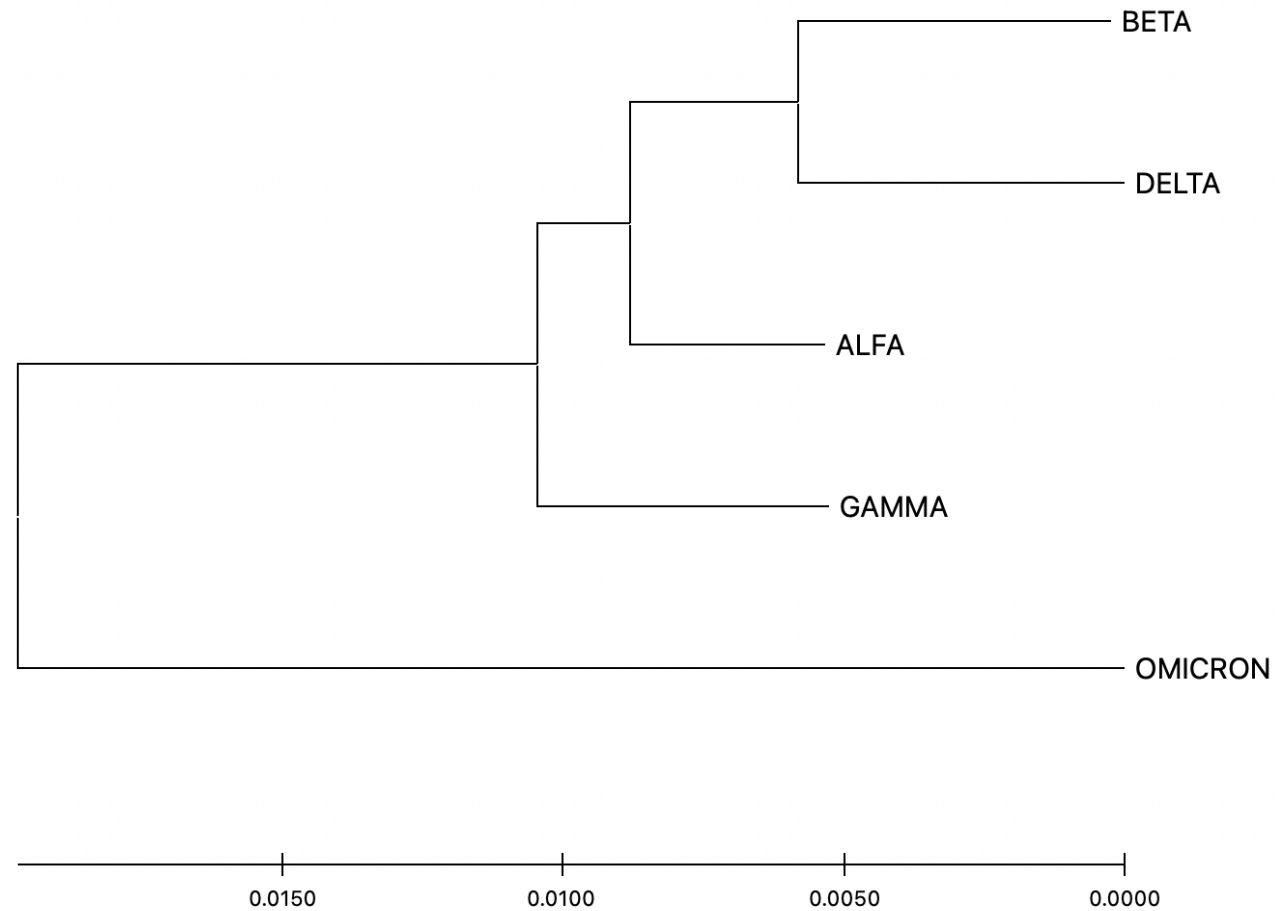


## S protein alignment scores - Results

seq2	seq1	score	score (%)
Reference genome: SARS-CoV-2/Wuhan	 B.1.1.7 (Alpha)	1118	87.82
	 B.1.351 (Beta)	1175	92.30
	 P.1 (Gamma)	974	76.51
	 B.1.617.2 (Delta)	1170	91.95
	 B.1.1.529 (Omicron)	1164	91.44

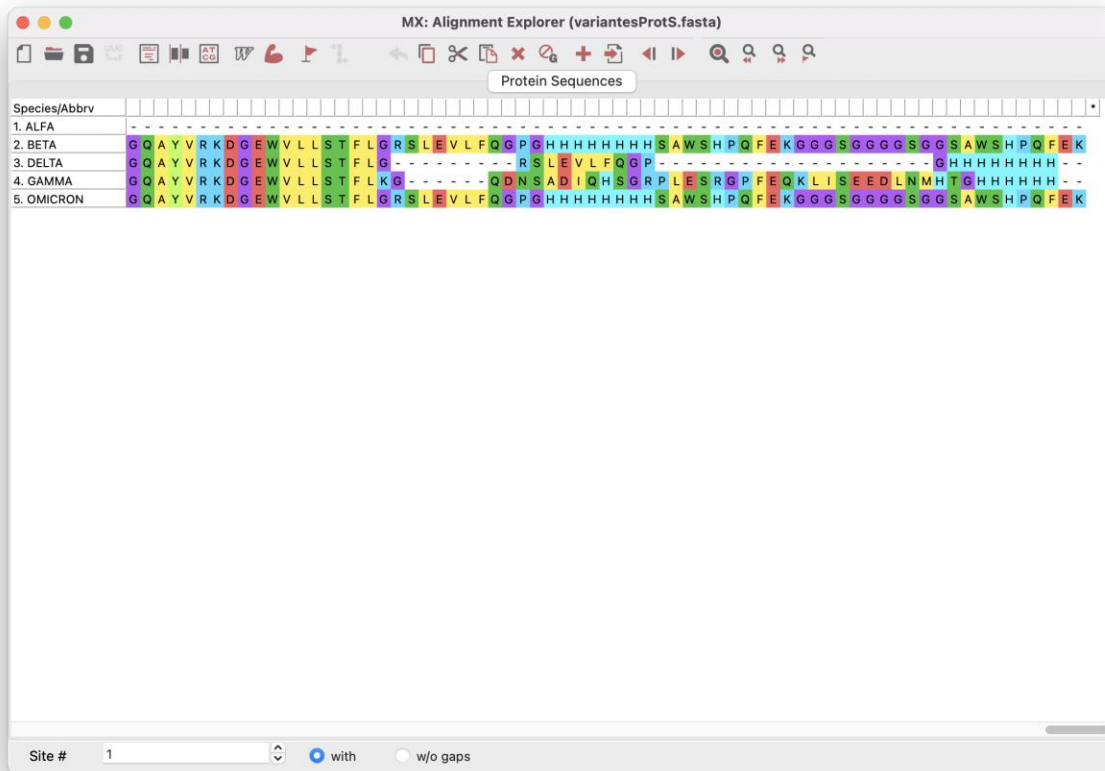


## Alineamiento en MEGAX





# Alineamiento en MEGAX



```
agustingarciadonate — Gblocks — Gblocks — 73x32

5. Allowed Gap Positions: ..... None

r. Restore Defaults
g. Get Blocks

z. Extended Block Options
m. Go To Main Menu

Your Choice: g

/Users/agustingarciadonate/Desktop/variantesProtS.fasta
Original alignment: 1293 positions
Gblocks alignment: 1187 positions (91 %) in 5 selected block(s)

BLOCK PARAMETERS

1. Minimum Number Of Sequences For A Conserved Position: . 3
2. Minimum Number Of Sequences For A Flank Position: ..... 4
3. Maximum Number Of Contiguous Nonconserved Positions: .. 8
4. Minimum Length Of A Block: ..... 10
5. Allowed Gap Positions: ..... None

r. Restore Defaults
g. Get Blocks

z. Extended Block Options
m. Go To Main Menu

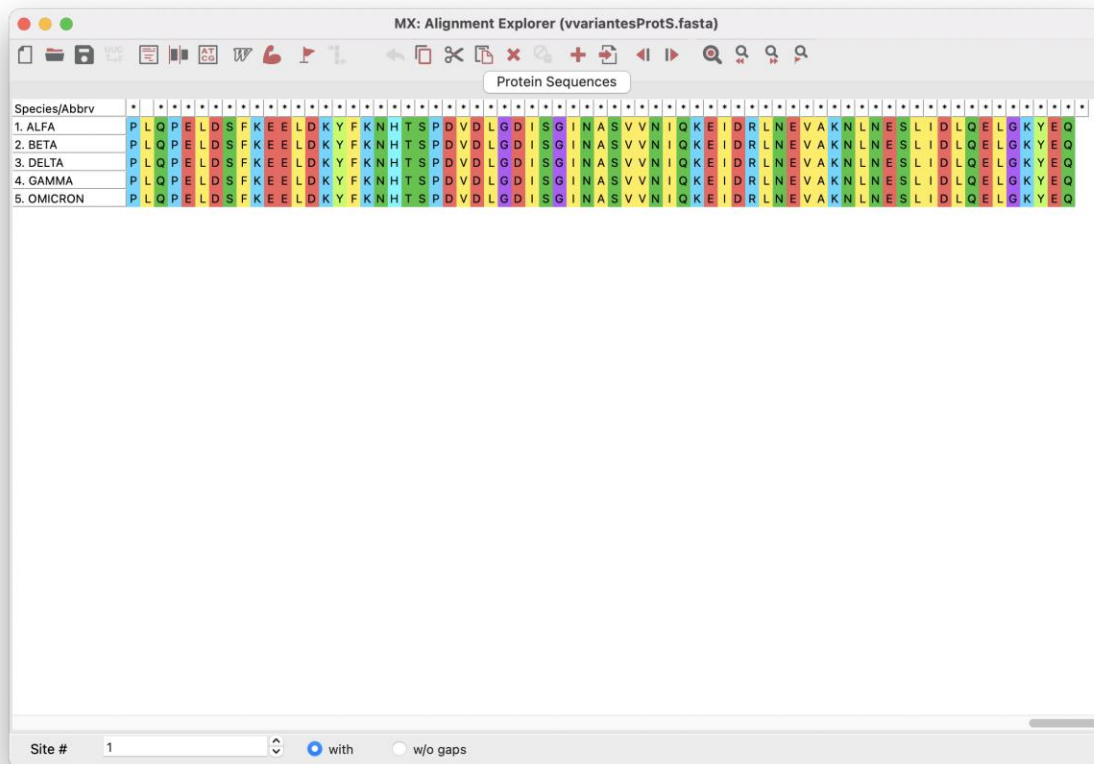
Your Choice: █
```

CLUSTALW





# Alineamiento en MEGAX



MEGA Caption Expert: Find Best-Fit Substitution Model (ML)

Results

**Table. Maximum Likelihood fits of 56 different amino acid substitution models**

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	f(A)	f(R)	f(N)
JTT+G	8	7962.712	7909.227	-3946.601	n/a	0.09	0.077	0.051	0.04
JTT	7	7962.931	7916.130	-3951.056	n/a	n/a	0.077	0.051	0.04
JTT+I	8	7966.627	7913.142	-3948.559	0.48	n/a	0.077	0.051	0.04
JTT+G+I	9	7971.076	7910.909	-3946.439	0.59	0.68	0.077	0.051	0.04
WAG+G	8	7977.892	7924.408	-3954.192	n/a	0.07	0.087	0.044	0.03
WAG	7	7978.508	7931.706	-3958.844	n/a	n/a	0.087	0.044	0.03
WAG+I	8	7982.113	7928.628	-3956.302	0.48	n/a	0.087	0.044	0.03
WAG+G+I	9	7986.258	7926.091	-3954.030	0.60	0.64	0.087	0.044	0.03
cpREV+G	8	8019.811	7966.327	-3975.151	n/a	0.05	0.076	0.062	0.04
Dayhoff+G	8	8022.044	7968.559	-3976.267	n/a	0.05	0.087	0.041	0.04
cpREV	7	8022.744	7975.942	-3980.962	n/a	n/a	0.076	0.062	0.04
JTT+G+F	27	8025.212	7844.875	-3895.310	n/a	0.08	0.064	0.033	0.07
JTT+F	26	8025.606	7851.940	-3899.851	n/a	n/a	0.064	0.033	0.07
cpREV+I	8	8027.073	7973.588	-3978.782	0.48	n/a	0.076	0.062	0.04



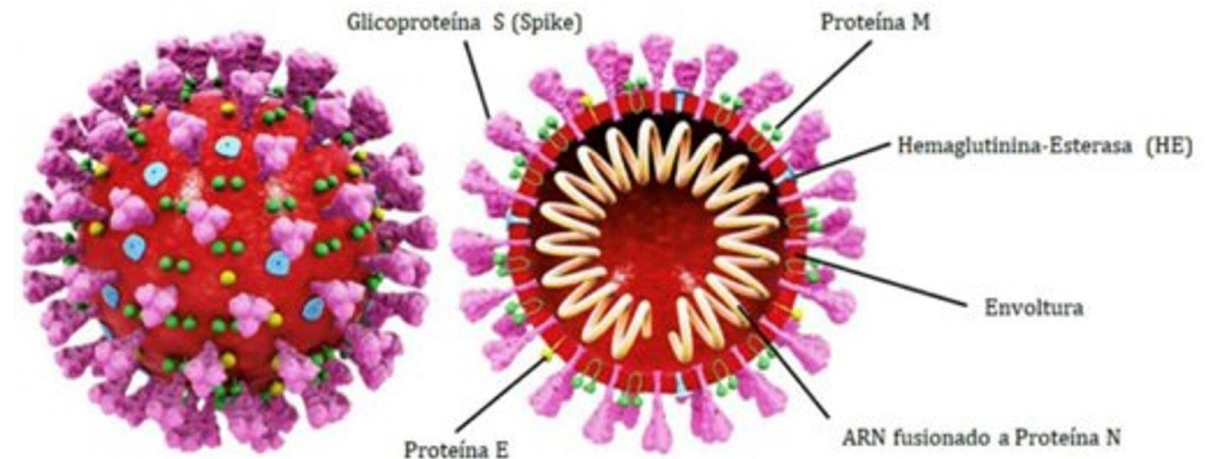
## Alineamiento en MEGAX

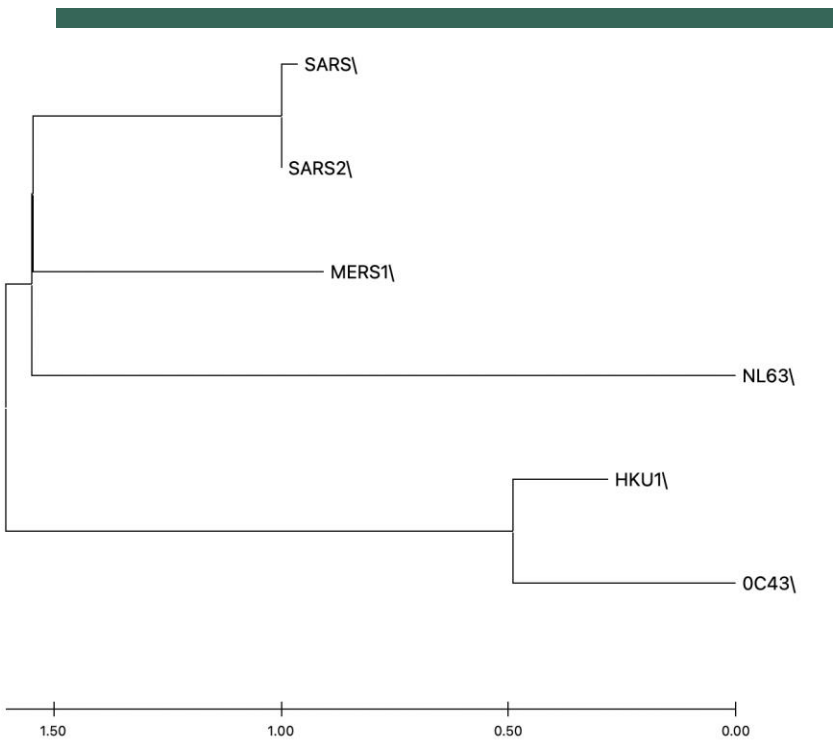
La familia *Coronaviridae* que afecta a humanos

- HCoV-229E
- HCoV-OC43
- HCoV-NL63
- HCoV-HKU1
- SARS-CoV
- SARS-CoV-2
- MERS-CoV

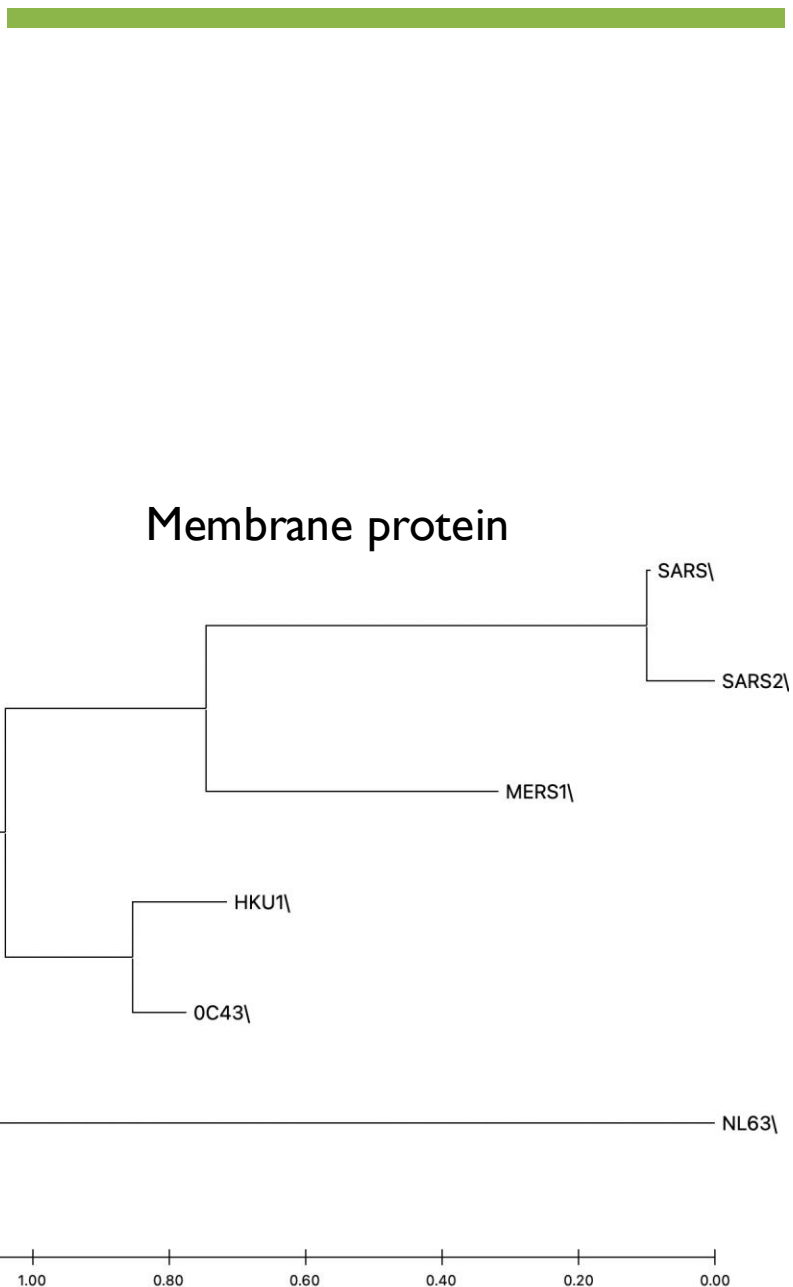
### Proteínas de interés

- Envelope small membrane protein (E)
- Membrane protein (M)
- Nucleoprotein (N)

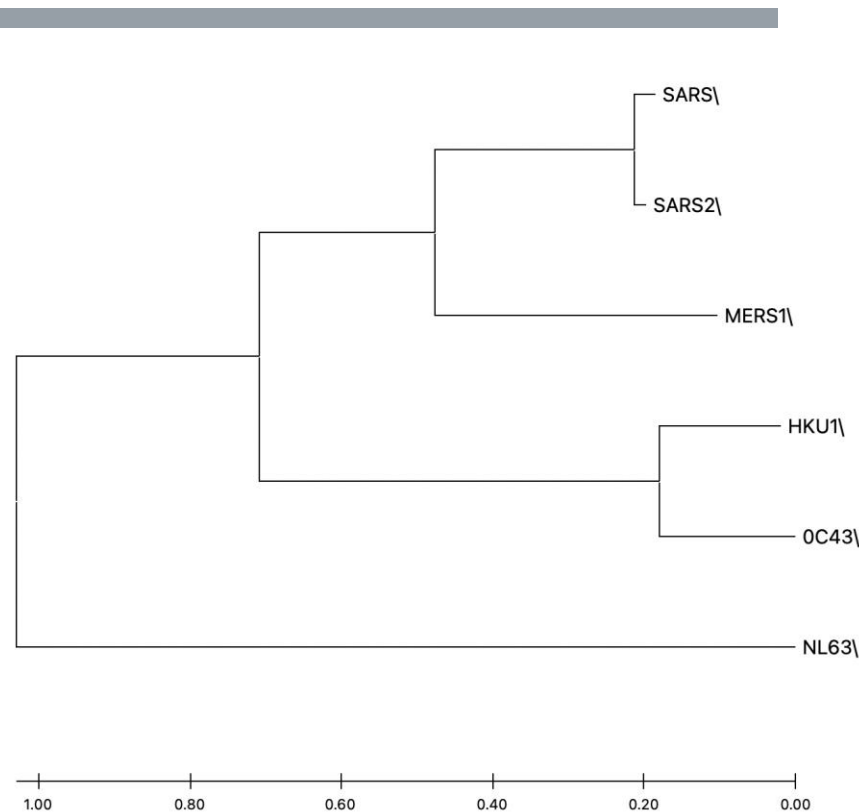




Envelope small membrane protein



Membrane protein



Nucleoprotein

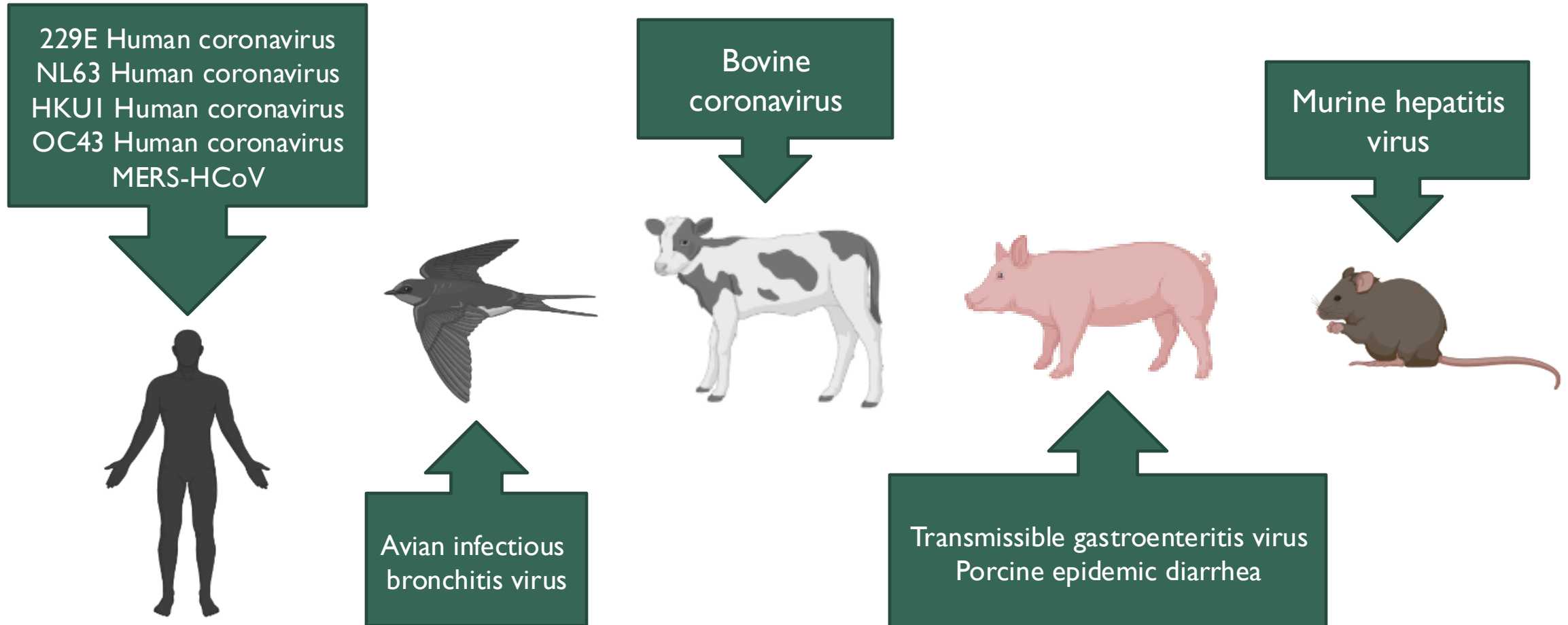


# OTROS CORONAVIRUS

ALINEAMIENTO DE SECUENCIAS Y CLUSTALW.













## Coronavirus que no causan SARS



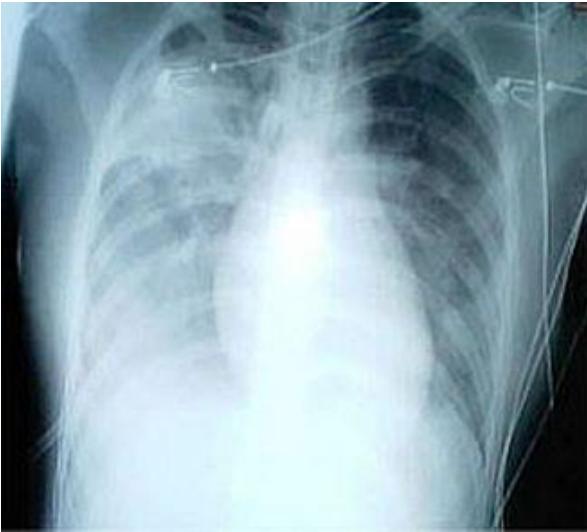


## Alignment scores - Results

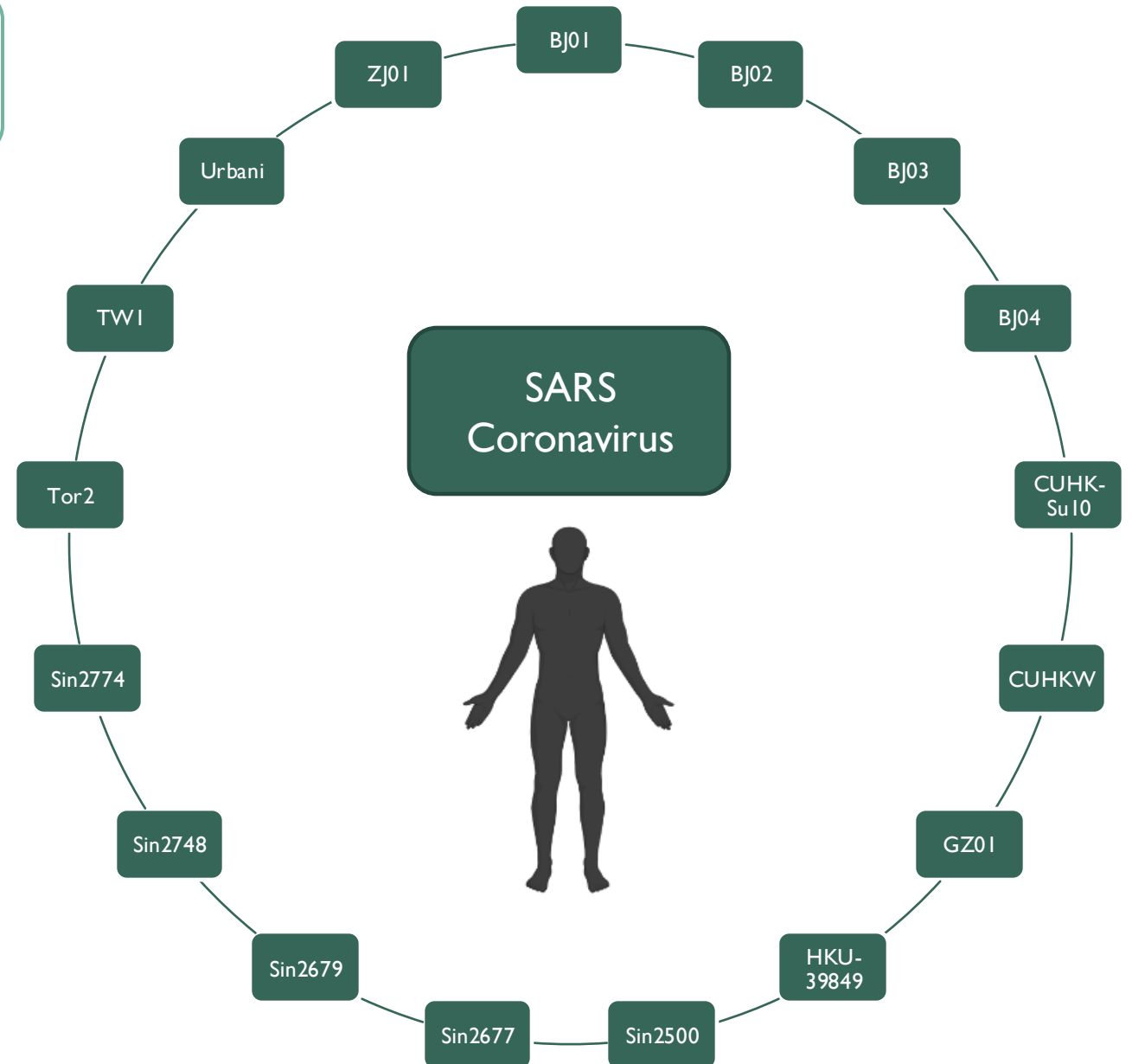
seq2	seq1		score	score (%)	
Reference genome: SARS-CoV-2/Wuhan		229E Human coronavirus	229E	9.047,0	30,27
		OC43 Human coronavirus	OC43	12539.5	41.95
		HKUI Human coronavirus	HKUI	13525.5	45.25
		NL63 Human coronavirus	NL63	9870.0	33.02
		Middle East respiratory syndrome-related coronavirus	MERS-HCoV	13808.0	46.19
		Porcine epidemic diarrhea virus	PEDV	10.509,0	35,16
		Transmissible gastroenteritis virus	TGEV	11.139,5	37,27
		Bovine coronavirus	BCoV	12.069,5	40,38
		Murine hepatitis virus	MHV	11.715,0	39,19
		Avian infectious bronchitis virus	IBV	9.237,5	30,90

# Coronavirus que causan SARS

- Causantes del síndrome respiratorio agudo severo
- Descubierto en 2002 en Foshan la provincia de Guangdong en China




















*Radiografía de pecho de paciente con SARS*





## Alignment scores - Results

seq2	seq1		score	score (%)
Reference genome: SARS-CoV-2/Wuhan	 ARS coronavirus BJ01	BJ01	23.098,0	77,27
	 ARS coronavirus BJ02	BJ02	23.157,0	77,47
	 ARS coronavirus BJ03	BJ03	23.143,5	77,43
	 ARS coronavirus BJ04	BJ04	23.116,5	77,34
	 ARS coronavirus CUHK-Su10	CUHKS	23.121,5	77,35
	 ARS coronavirus CUHKW	CUHKW	23.125,5	77,37
	 ARS coronavirus GZ01	GZ01	23.165,0	77,5
	 ARS coronavirus HKU-39849	HKUN	23.151,5	77,45
	 ARS coronavirus Sin2500	Sin2500	23.052,0	77,12
	 ARS coronavirus Sin2677	Sin2677	23.043,0	77,09
	 ARS coronavirus Sin2679	Sin2679	23.055,0	77,13
	 ARS coronavirus Sin2748	Sin2748	23.043,5	77,09
	 ARS coronavirus Sin2774	Sin2774	23.051,0	77,12
	 ARS coronavirus Tor2	Tor2	23.169,0	77,51
	 ARS coronavirus TW01	TW01	23.112,0	77,32
	 ARS coronavirus Urbani	Urbani	23.105,0	77,3
	 ARS coronavirus ZJ01	ZJ01	23.011,0	76,98





# Alineamiento en MEGAX

MX: Alignment Explorer (allSeq.fasta)

Data Edit Search Alignment Web Sequencer Display Help

DNA Sequences Translated Protein Sequences

Species/Abbrv

1. Wuhan  
2. 229E  
3. BCoV-ENT  
4. BJ02  
5. BJ03  
6. BJ04  
7. BJ01  
8. CUHK-Su10  
9. CUHK-W1  
10. GZ01  
11. HKU-39849  
12. IBV  
13. MHV  
14. PEDV  
15. Sin2500  
16. Sin2677  
17. Sin2679  
18. Sin2748  
19. Sin2774  
20. TGEV  
21. Tor2  
22. TW1  
23. Urbani  
24. ZJ01  
25. MERS-HCoV  
26. NL63  
27. OC43  
28. HKU1

Site # 32751 with w/o gaps



MEGA Caption Expert: Find Best-Fit Substitution Model (ML)

File Edit View Help

Results

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)
GTR+G+I	63	487312.985	486598.725	-243236.356	0.10	2.37	1.22	0.277	0.330	0.183	0.210	0.074	0.056
TN93+G+I	60	488000.825	487320.577	-243600.283	0.11	2.71	1.08	0.277	0.330	0.183	0.210	0.077	0.043
GTR+G	62	488042.050	487339.127	-243607.557	n/a	1.09	1.26	0.277	0.330	0.183	0.210	0.072	0.057
HKY+G+I	59	488044.750	487375.839	-243628.914	0.11	2.71	1.08	0.277	0.330	0.183	0.210	0.077	0.043
T92+G+I	57	488563.980	487917.744	-243901.867	0.11	2.79	1.08	0.303	0.303	0.197	0.197	0.071	0.046
TN93+G	59	488913.392	488244.481	-244063.235	n/a	1.17	1.05	0.277	0.330	0.183	0.210	0.078	0.043
HKY+G	58	489021.076	488363.502	-244123.746	n/a	1.18	1.02	0.277	0.330	0.183	0.210	0.079	0.044
GTR+I	62	489489.942	488787.019	-244331.503	0.14	n/a	1.07	0.277	0.330	0.183	0.210	0.084	0.054
T92+G	56	489556.699	488921.800	-244404.895	n/a	1.20	1.03	0.303	0.303	0.197	0.197	0.073	0.047
TN93+I	59	489977.183	489308.272	-244595.130	0.14	n/a	1.02	0.277	0.330	0.183	0.210	0.079	0.044
HKY+I	58	490037.404	489379.831	-244631.910	0.14	n/a	1.00	0.277	0.330	0.183	0.210	0.080	0.045
T92+I	56	490448.894	489813.995	-244850.992	0.14	n/a	1.00	0.303	0.303	0.197	0.197	0.074	0.048
K2+G+I	56	493994.891	493359.992	-246623.991	0.11	3.17	1.13	0.250	0.250	0.250	0.250	0.059	0.059
K2+G	55	494943.543	494319.982	-247104.986	n/a	1.29	1.10	0.250	0.250	0.250	0.250	0.060	0.060
K2+I	55	495359.559	494735.998	-247312.994	0.14	n/a	1.05	0.250	0.250	0.250	0.250	0.061	0.061
GTR	61	497851.793	497160.208	-248519.098	n/a	n/a	0.99	0.277	0.330	0.183	0.210	0.087	0.056
JC+G+I	55	498031.601	497408.039	-248649.015	0.11	4.09	0.50	0.250	0.250	0.250	0.250	0.083	0.083
TN93	58	498370.584	497713.011	-248798.500	n/a	n/a	0.99	0.277	0.330	0.183	0.210	0.081	0.045

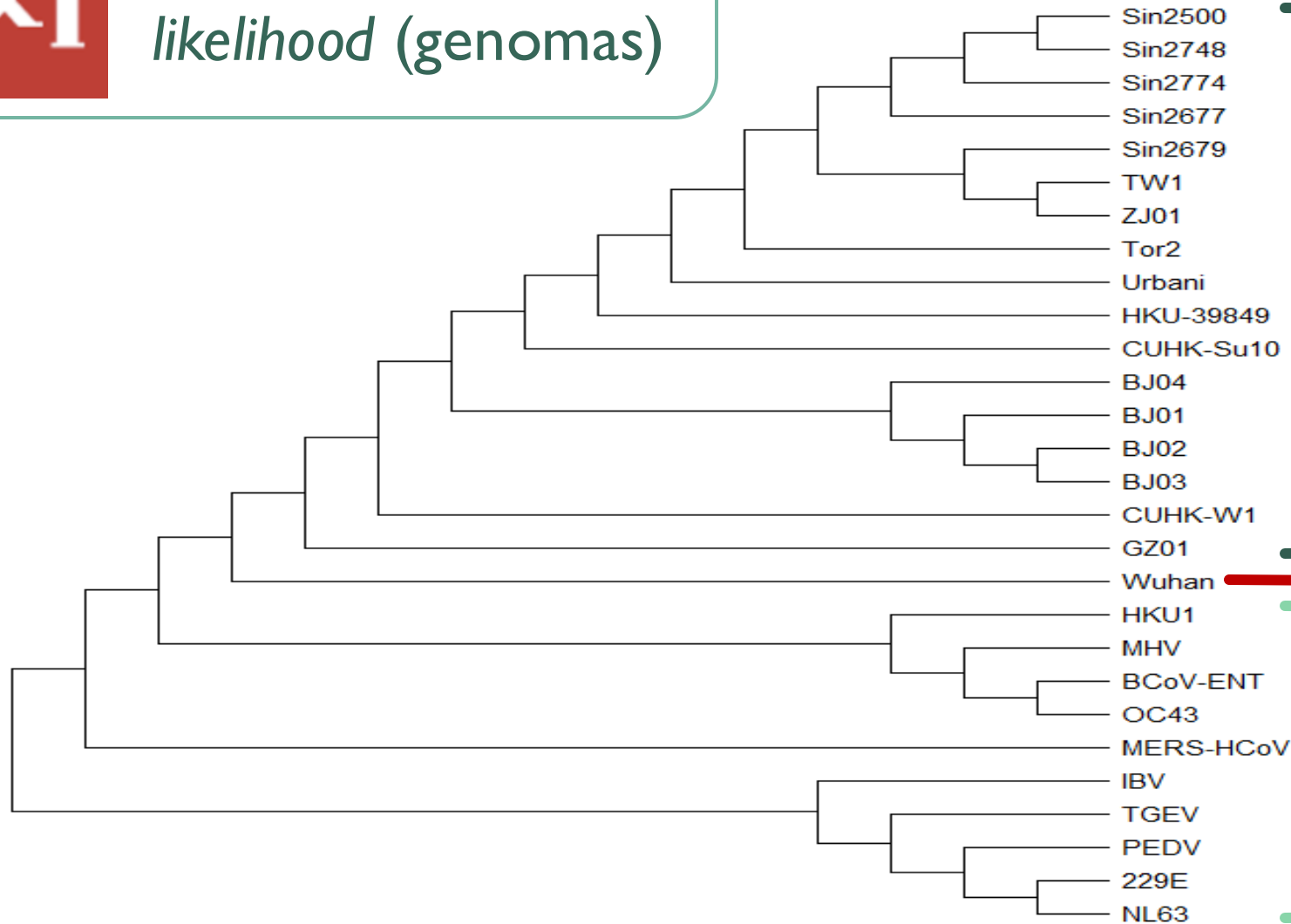


Árbol *maximum likelihood* (genomas)

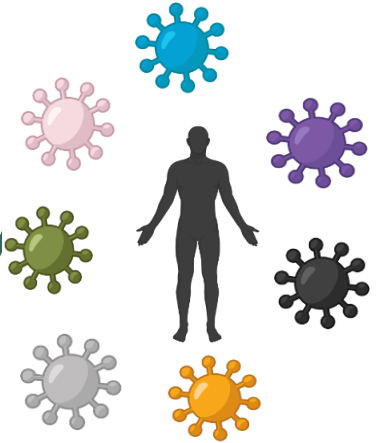




Árbol *maximum likelihood* (genomas)

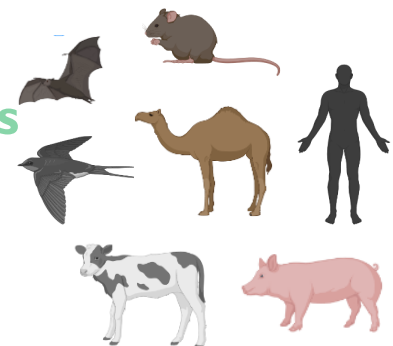


SARS-CoV



SARS-CoV2

Coronavirus  
No SARS





# Alineamiento en MEGAX

MX: Alignment Explorer (allSeq.fasta)

Data Edit Search Alignment Web Sequencer Display Help

DNA Sequences Translated Protein Sequences

Species/Abbrv

1. Wuhan  
2. 229E  
3. BCoV-ENT  
4. BJ02  
5. BJ03  
6. BJ04  
7. BJ01  
8. CUHK-Su10  
9. CUHK-W1  
10. GZ01  
11. HKU-39849  
12. IBV  
13. MHV  
14. PEDV  
15. Sin2500  
16. Sin2677  
17. Sin2679  
18. Sin2748  
19. Sin2774  
20. TGEV  
21. Tor2  
22. TVV1  
23. Urbani  
24. ZJ01  
25. MERS-HCoV  
26. NL63  
27. OC43  
28. HKU1

Site # 1 with w/o gaps Edit disabled for translated protein data.

MEGA Caption Expert: Find Best-Fit Substitution Model (ML)

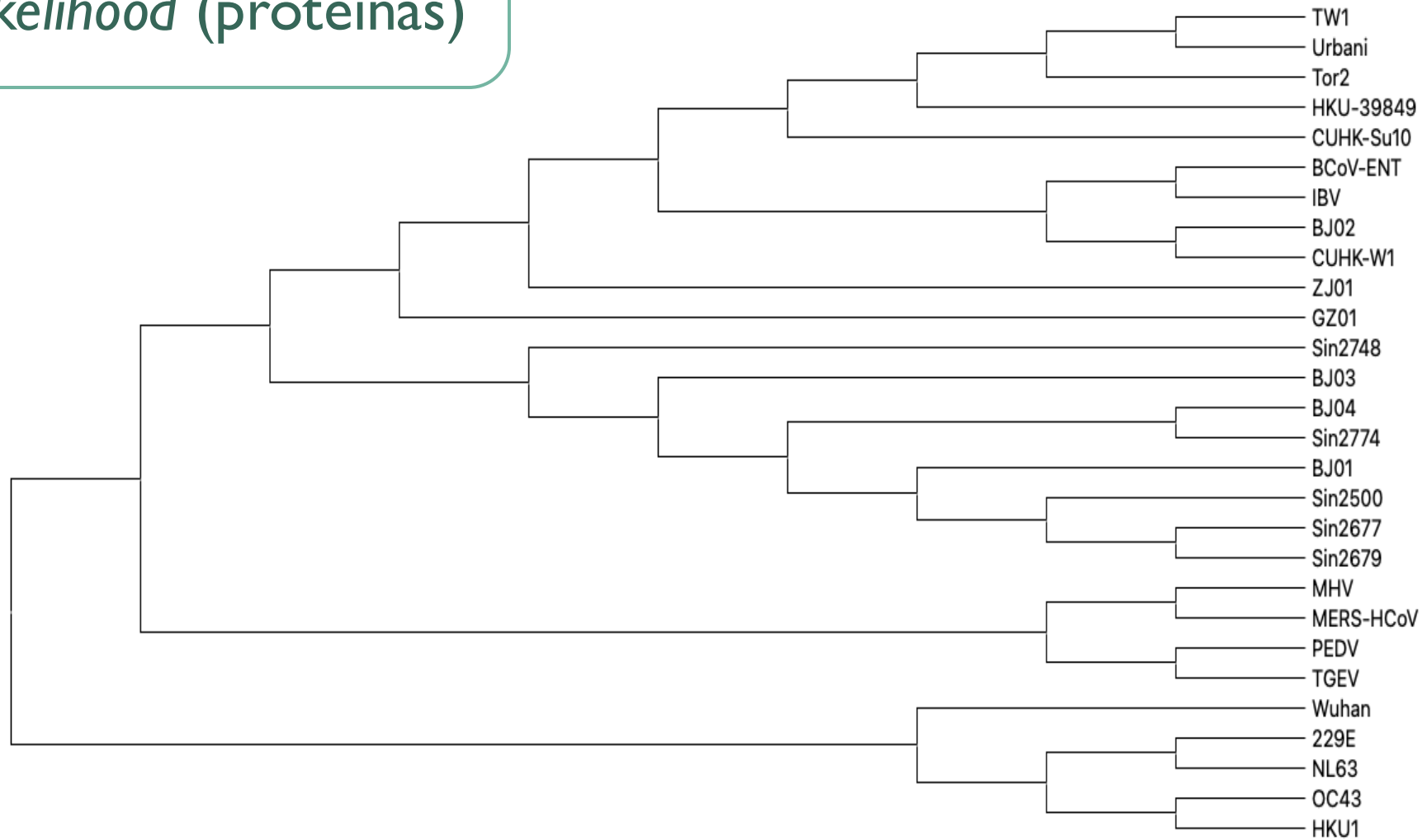
Results

Table. Maximum Likelihood fits of 56 different amino acid substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	f(A)	f(R)
JTT+G+F	73	76609.670	76000.488	-37927.071	n/a	2.10	0.055	0.045
JTT+G+I+F	74	76620.020	76002.497	-37927.071	0.00	2.10	0.055	0.045
WAG+G+F	73	76766.007	76156.825	-38005.239	n/a	2.28	0.055	0.045
WAG+G+I+F	74	76776.357	76158.834	-38005.239	0.00	2.28	0.055	0.045
LG+G+F	73	77405.177	76795.994	-38324.824	n/a	1.75	0.055	0.045
LG+G+I+F	74	77415.526	76798.004	-38324.824	0.00	1.75	0.055	0.045
Dayhoff+G+F	73	77467.003	76857.821	-38355.737	n/a	1.81	0.055	0.045
Dayhoff+G+I+F	74	77477.353	76859.830	-38355.737	0.00	1.81	0.055	0.045
JTT+F	72	77534.148	76933.306	-38394.484	n/a	n/a	0.055	0.045
JTT+I+F	73	77544.478	76935.295	-38394.474	0.00	n/a	0.055	0.045
WAG+F	72	77561.042	76960.200	-38407.932	n/a	n/a	0.055	0.045
WAG+I+F	73	77571.392	76962.210	-38407.931	0.00	n/a	0.055	0.045
rtREV+G+F	73	77674.414	77065.231	-38459.442	n/a	1.79	0.055	0.045
rtREV+G+I+F	74	77684.763	77067.241	-38459.442	0.00	1.79	0.055	0.045

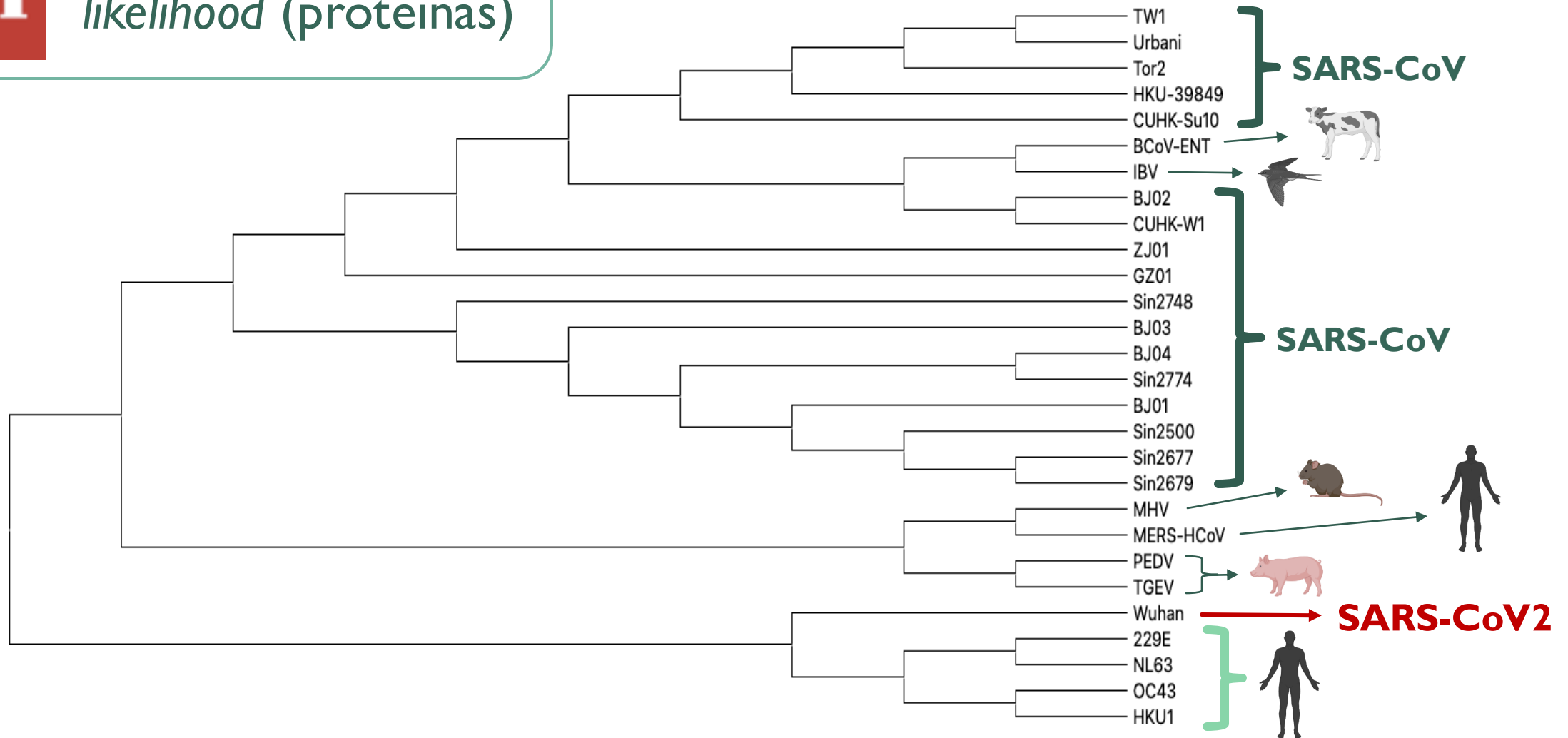


Árbol *maximum likelihood* (proteínas)





Árbol *maximum likelihood* (proteínas)





# CONCLUSIONES



# CONCLUSIONES

- ❑ El análisis comparativo del coronavirus original de Wuhan con las variantes generadas indica, a nivel de genoma, una relación estrecha en el caso de las variantes omicron y delta. En cambio, a nivel de la proteína S, se aprecia un cambio de tendencia en las relaciones filogenéticas.
- ❑ El análisis comparativo a nivel de proteína S muestra diferencias entre las diferentes variantes con respecto al coronavirus original lo que explica las diferencias en la forma de infectar.
- ❑ A nivel nucleotídico, los coronavirus antiguos que causan SARS tienen una mayor similitud con el actual coronavirus de Wuhan que los coronavirus que no causan SARS.
- ❑ No obstante, a nivel proteico los coronavirus que no causan SARS y que afectan a humanos estarían más emparentados con el de Wuhan que los que provocan SARS.



# GRACIAS POR VUESTRA ATENCIÓN

Agustín García Doñate, Cristian Morillo Losada, Pedro Sánchez García & Rubén Entenza Pereira.  
Fundamentos de Bioinformática.  
MUBICS | Universidade da Coruña.



UNIVERSIDADE DA CORUÑA

# BIBLIOGRAFÍA

- Chen, Z., Boon, S.S., Wang, M.H., Chan, W.Y., & Chan, K.S. (2021). Genomic and evolutionary comparison between SARS-CoV-2 and other human coronaviruses. *Journal of Virological Methods*, 289(3), 114032.
- De Maio, N., Walker, C., Weilguny, B.R., & Goldman, S.G. (2020). Issues with SARS-CoV-2 sequencing data. <https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473>. Consulted 27/12/2021.
- J. Castresana (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis, *Molecular Biology and Evolution*, 17 (540–552). <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Kandeel, M., Mohamed, M.E., Abd El-Lateef, H.M., Venugopala, K.N., & El-Beltagi, H.S. (2021). Omicron variant genome evolution and phylogenetics. *Journal of Medical Virology*, 11(12), 1-6.
- Kannan, S.R., Spratt, A.N., Cohen, A.R., Naqvi, S.H., Chand, H.S., Quinn, T.P., Lorson, C.L., Byrareddy, S.N., & Singh, K. (2021). Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *Journal of Autoimmunity*, 124(8), 102715.
- Maguiña Vargas, Ciro, Gastelo Acosta, Rosy, & Tequen Bernilla, Arly. (2020). El nuevo Coronavirus y la pandemia del Covid-19. *Revista Medica Herediana*, 31(2), 125-131. <https://dx.doi.org/10.20453/rmh.v31i2.3776>
- Mohammadi, M., Shayestehpour, M., & Mirzaei, H. (2021). The impact of spike mutated variants of SARS-CoV2 [Alpha, Beta, Gamma, Delta, and Lambda] on the efficacy of subunit recombinant vaccines. *Brazilian Journal of Infectious Diseases*, 25(4), 101606.