

# Evidencia 2

Diego Curiel - A01640372 | Santiago Vera - A | Carlos Palmieri - A | Carlos Magaña - A

## Evidencia 2

### Parte 1

Enlace al video:

<https://youtu.be/FbeJTYp5I>

### Parte 2

En esta evidencia trabajaremos con las secuencias de SARS-CoV-2 reportadas en los 20 países con más casos reportados.

Hide

Library(seqinr)

Para comenzar vamos a leer los archivos FASTA que contienen las secuencias a estudiar:

Hide

```
usa = read.fasta("usa.fasta")
india = read.fasta("india.fasta")
brasil = read.fasta("brasil.fasta")
uk = read.fasta("uk.fasta")
peru = read.fasta("peru.fasta")
sudafrica = read.fasta("sudafrica.fasta")
alemania = read.fasta("alemania.fasta")
rusia = read.fasta("rusia.fasta")
coreasur = read.fasta("coreasur.fasta")
china = read.fasta("china.fasta")
```

A continuación, calcularemos la longitud de las secuencias:

Hide

```
print("Longitud de la secuencia Usa: ")
[1] "Longitud de la secuencia Usa: "
```

Hide

```
length(usa[[1]])
[1] 29862
```

Hide

```
print("Longitud de la secuencia India: ")
[1] "Longitud de la secuencia India: "
```

Hide

```
length(india[[1]])
[1] 29899
```

Hide

```
print("Longitud de la secuencia Brasil: ")
[1] "Longitud de la secuencia Brasil: "
```

Hide

```
length(brasil[[1]])
[1] 29789
```

Hide

```
print("Longitud de la secuencia UK: ")
[1] "Longitud de la secuencia UK: "
```

Hide

```
length(uk[[1]])
[1] 29694
```

Hide

```
print("Longitud de Longitud de la secuencia Peru: ")
[1] "Longitud de Longitud de la secuencia Peru: "
```

Hide

```
length(peru[[1]])
[1] 29782
```

Hide

```
print("Longitud de la secuencia Sud Africa: ")
[1] "Longitud de la secuencia Sud Africa: "
```

Hide

```
length(sudafrica[[1]])
[1] 29885
```

Hide

```
print("Longitud de la secuencia Alemania: ")
[1] "Longitud de la secuencia Alemania: "
```

Hide

```
length(alemania[[1]])
[1] 29878
```

Hide

```
print("Longitud de la secuencia Rusia: ")
[1] "Longitud de la secuencia Rusia: "
```

Hide

```
length(rusia[[1]])
[1] 29903
```

Hide

```
print("Longitud de la secuencia Corea del Sur: ")
[1] "Longitud de la secuencia Corea del Sur: "
```

Hide

```
length(coreasur[[1]])
[1] 29848
```

Hide

```
print("Longitud de la secuencia China: ")
[1] "Longitud de la secuencia China: "
```

Hide

```
length(china[[1]])
[1] 29903
```

Podemos observar que la longitud de las secuencias es prácticamente idéntico entre ellas. La menor cuenta con una longitud de 29694 y la mayor con una longitud de 29903.

Ahora graficaremos el número de bases de ADN de las secuencias:

Hide

```
barplot(seqinr::count(usa[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "Usa", col = c('orange',12,'red',2))
```

Usa

Bar chart showing the count of nucleotide bases (A, C, G, T) for the USA sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(india[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "India", col = c('orange',12,'red',2))
```

India

Bar chart showing the count of nucleotide bases (A, C, G, T) for the India sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(brasil[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "Brasil", col = c('orange',12,'red',2))
```

Brasil

Bar chart showing the count of nucleotide bases (A, C, G, T) for the Brasil sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(uk[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "UK", col = c('orange',12,'red',2))
```

UK

Bar chart showing the count of nucleotide bases (A, C, G, T) for the UK sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(peru[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "Peru", col = c('orange',12,'red',2))
```

Peru

Bar chart showing the count of nucleotide bases (A, C, G, T) for the Peru sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(sudafrica[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "Sud Africa", col = c('orange',12,'red',2))
```

Sud Africa

Bar chart showing the count of nucleotide bases (A, C, G, T) for the Sud Africa sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(alemania[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "Alemania", col = c('orange',12,'red',2))
```

Alemania

Bar chart showing the count of nucleotide bases (A, C, G, T) for the Alemania sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(rusia[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "Rusia", col = c('orange',12,'red',2))
```

Rusia

Bar chart showing the count of nucleotide bases (A, C, G, T) for the Rusia sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(coreasur[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "Corea del Sur", col = c('orange',12,'red',2))
```

Corea del Sur

Bar chart showing the count of nucleotide bases (A, C, G, T) for the Corea del Sur sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Hide

```
barplot(seqinr::count(china[[1]],1), beside = T, legend.text = c('A', 'C', 'G', 'T'), main = "China", col = c('orange',12,'red',2))
```

China

Bar chart showing the count of nucleotide bases (A, C, G, T) for the China sequence. The y-axis represents the count (0 to 8000). The x-axis lists the bases: a (A), c (C), g (G), and t (T). The counts are approximately: A (8000), C (5500), G (5500), and T (8000).

Al igual que en el apartado anterior, podemos observar que el número de bases es prácticamente idéntico entre ellas.

Deducimos que estos cambios tan mínimos se deben a que son variantes relativamente cercanas entre ellas, pues a penas se han desarrollado en un lapso menor a 2 años.