Evidencia 2 | Proyecto integrador

Dael Chávez Ferreyra - A01771963

07/05/2023

## Parte 1 | Video

### Link del video de YouTube:

#### <https://www.youtube.com/watch?v=hfCDRWYOO3o&t=2s>

## Parte 2 | Código

### Investigación

Analizar las secuencias de SARS-CoV-2 reportadas en los 20 países con más casos reportados. Y puedes tratar de responder a la pregunta: ¿Son muy diferentes las variantes entre cada país? ¿Es diferente el SARS-CoV-2 entre las diferentes poblaciones: Asiática, Hispana, Europea o Africana?

El SARS-CoV-2 es un virus que muta constantemente y que ha dado lugar a varias variantes que se han extendido por diferentes países y regiones del mundo. Algunas de estas variantes tienen cambios en la proteína S que pueden afectar a su capacidad de infectar a las células humanas, de escapar a la respuesta inmunitaria o de resistir a las vacunas. Estas variantes se clasifican en variantes de interés (VOI) y variantes de preocupación (VOC), según su impacto potencial o real en la salud pública.

Entre las variantes de interés se encuentran la eta (B.1.525), la iota (B.1.526), la kappa (B.1.617.1) y la lambda (C.37), que se han detectado en varios países de África, Asia, Europa y América. Entre las variantes de preocupación se encuentran la alfa (B.1.1.7), la beta (B.1.351), la gamma (P.1) y la delta (B.1.617.2), que se han asociado con una mayor transmisibilidad, una mayor gravedad o una menor neutralización por los anticuerpos.

El análisis genético de estas variantes ha revelado que existen diferencias en su distribución entre las diferentes poblaciones humanas, que pueden reflejar factores geográficos, demográficos, ecológicos o evolutivos. Por ejemplo, el estudio de Cosar et al.2 encontró que algunas variantes del SARS-CoV-2 tienen una mayor frecuencia en poblaciones africanas (AFR), como la N501Y, la E484K y la K417N, que afectan al RBD de la proteína S y que se encuentran en las variantes beta y gamma. Otras variantes tienen una mayor frecuencia en poblaciones asiáticas orientales (EAS), como la D614G, que aumenta la infectividad del virus y que se encuentra en todas las variantes de preocupación. Otras variantes tienen una mayor frecuencia en poblaciones europeas (EUR), como la N439K, que reduce la susceptibilidad a los anticuerpos neutralizantes y que se encuentra en la variante eta. Otras variantes tienen una mayor frecuencia en poblaciones asiáticas meridionales (SAS), como la P681R, que facilita la entrada del virus a las células y que se encuentra en las variantes kappa y delta.

Estos ejemplos muestran que el SARS-CoV-2 es diferente entre las diferentes poblaciones humanas, lo que puede tener implicaciones para el diagnóstico, el tratamiento y la prevención de la COVID-192. Sin embargo, también hay que tener en cuenta que el virus sigue circulando y mutando globalmente, y que puede haber intercambio genético entre las distintas variantes. Por lo tanto, se necesita una vigilancia continua y una actualización periódica de las vacunas para hacer frente a esta amenaza emergente.

library(Biostrings)

## Loading required package: BiocGenerics

##   
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
## colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
## get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
## match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
## Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,  
## table, tapply, union, unique, unsplit, which.max, which.min

## Loading required package: S4Vectors

## Loading required package: stats4

##   
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:base':  
##   
## expand.grid, I, unname

## Loading required package: IRanges

##   
## Attaching package: 'IRanges'

## The following object is masked from 'package:grDevices':  
##   
## windows

## Loading required package: XVector

## Loading required package: GenomeInfoDb

##   
## Attaching package: 'Biostrings'

## The following object is masked from 'package:base':  
##   
## strsplit

library(seqinr)

##   
## Attaching package: 'seqinr'

## The following object is masked from 'package:Biostrings':  
##   
## translate

library(stringr)  
library(ggplot2)  
library(adegenet)

## Loading required package: ade4

##   
## Attaching package: 'ade4'

## The following object is masked from 'package:Biostrings':  
##   
## score

## The following object is masked from 'package:BiocGenerics':  
##   
## score

##   
## /// adegenet 2.1.10 is loaded ////////////  
##   
## > overview: '?adegenet'  
## > tutorials/doc/questions: 'adegenetWeb()'   
## > bug reports/feature requests: adegenetIssues()

library(ape)

##   
## Attaching package: 'ape'

## The following objects are masked from 'package:seqinr':  
##   
## as.alignment, consensus

## The following object is masked from 'package:Biostrings':  
##   
## complement

library(DECIPHER)

## Loading required package: RSQLite

## Loading required package: parallel

getwd()

## [1] "c:/Users/daeld/OneDrive/Documentos/Programas"

setwd("C:/Users/daeld/OneDrive/Documentos/Programas")

argentina <- read.fasta("sars\_argentina.fasta")  
australia <- read.fasta("sars\_australia.fasta")  
brazil <- read.fasta("sars\_brazil.fasta")  
france <- read.fasta("sars\_france.fasta")  
germany <- read.fasta("sars\_germany.fasta")  
india <- read.fasta("sars\_india.fasta")  
indonesia <- read.fasta("sars\_indonesia.fasta")  
iran <- read.fasta("sars\_iran.fasta")  
italy <- read.fasta("sars\_italy.fasta")  
japan <- read.fasta("sars\_japan.fasta")  
mexico <- read.fasta("sars\_mexico.fasta")  
netherlands <- read.fasta("sars\_netherlands.fasta")  
russia <- read.fasta("sars\_russia.fasta")  
southkorea <- read.fasta("sars\_southkorea.fasta")  
spain <- read.fasta("sars\_spain.fasta")  
taiwan <- read.fasta("sars\_taiwan.fasta")  
turkey <- read.fasta("sars\_turkey.fasta")  
uk<- read.fasta("sars\_uk.fasta")  
usa <- read.fasta("sars\_usa.fasta")  
vietnam <- read.fasta("sars\_vietnam.fasta")

argentinalen <- length(argentina[[1]])  
australialen <- length(australia[[1]])  
brazillen <- length(brazil[[1]])  
francelen <- length(france[[1]])  
germanylen <- length(germany[[1]])  
indialen <- length(india[[1]])  
indonesialen <- length(indonesia[[1]])  
iranlen <- length(iran[[1]])  
italylen <- length(italy[[1]])  
japanlen <- length(japan[[1]])  
mexicolen <- length(mexico[[1]])  
netherlandslen <- length(netherlands[[1]])  
russialen <- length(russia[[1]])  
southkorealen <- length(southkorea[[1]])  
spainlen <- length(spain[[1]])  
taiwanlen <- length(taiwan[[1]])  
turkeylen <- length(turkey[[1]])  
uklen <- length(uk[[1]])  
usalen <- length(usa[[1]])  
vietnamlen <- length(vietnam[[1]])

### Longitud de las secuencias de cada variante.

print(paste0("SARS Argentina: ",argentinalen))

## [1] "SARS Argentina: 29540"

print(paste0("SARS Australia: ",australialen))

## [1] "SARS Australia: 28811"

print(paste0("SARS Brazil: ",brazillen))

## [1] "SARS Brazil: 29813"

print(paste0("SARS France: ",francelen))

## [1] "SARS France: 29701"

print(paste0("SARS Germany: ",germanylen))

## [1] "SARS Germany: 29718"

print(paste0("SARS India: ",indialen))

## [1] "SARS India: 3829"

print(paste0("SARS Indonesia: ",indonesialen))

## [1] "SARS Indonesia: 3728"

print(paste0("SARS Iran: ",iranlen))

## [1] "SARS Iran: 4198"

print(paste0("SARS Italy: ",italylen))

## [1] "SARS Italy: 80"

print(paste0("SARS Japan: ",japanlen))

## [1] "SARS Japan: 29805"

print(paste0("SARS Mexico: ",mexicolen))

## [1] "SARS Mexico: 29540"

print(paste0("SARS Netherlands: ",netherlandslen))

## [1] "SARS Netherlands: 29779"

print(paste0("SARS Russia: ",russialen))

## [1] "SARS Russia: 29701"

print(paste0("SARS South Korea: ",southkorealen))

## [1] "SARS South Korea: 29718"

print(paste0("SARS Spain: ",spainlen))

## [1] "SARS Spain: 29827"

print(paste0("SARS Taiwan: ",taiwanlen))

## [1] "SARS Taiwan: 29778"

print(paste0("SARS Turkey: ",turkeylen))

## [1] "SARS Turkey: 3816"

print(paste0("SARS UK: ",uklen))

## [1] "SARS UK: 29747"

print(paste0("SARS USA: ",usalen))

## [1] "SARS USA: 29804"

print(paste0("SARS Vietnam: " ,vietnamlen))

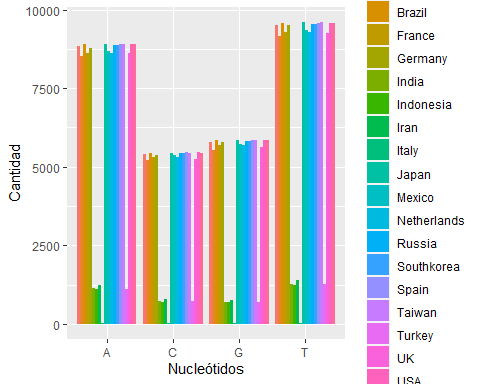
## [1] "SARS Vietnam: 29749"

argentinadf <- as.data.frame(count(argentina[[1]],1))  
colnames(argentinadf) <- c("Nucleotides", "Argentina")  
australiadf <- as.data.frame(count(australia[[1]],1))  
colnames(australiadf) <- c("Nucleotides", "Australia")  
brazildf <- as.data.frame(count(brazil[[1]],1))  
colnames(brazildf) <- c("Nucleotides", "Brazil")  
francedf <- as.data.frame(count(france[[1]],1))  
colnames(francedf) <- c("Nucleotides", "France")  
germanydf <- as.data.frame(count(germany[[1]],1))  
colnames(germanydf) <- c("Nucleotides", "Germany")  
indiadf <- as.data.frame(count(india[[1]],1))  
colnames(indiadf) <- c("Nucleotides", "India")  
indonesiadf <- as.data.frame(count(indonesia[[1]],1))  
colnames(indonesiadf) <- c("Nucleotides", "Indonesia")  
italydf <- as.data.frame(count(italy[[1]],1))  
colnames(italydf) <- c("Nucleotides", "Italy")  
irandf <- as.data.frame(count(iran[[1]],1))  
colnames(irandf) <- c("Nucleotides", "Iran")  
japandf <- as.data.frame(count(japan[[1]],1))  
colnames(japandf) <- c("Nucleotides", "Japan")  
mexicodf <- as.data.frame(count(mexico[[1]],1))  
colnames(mexicodf) <- c("Nucleotides", "Mexico")  
netherlandsdf <- as.data.frame(count(netherlands[[1]],1))  
colnames(netherlandsdf) <- c("Nucleotides", "Netherlands")  
russiadf <- as.data.frame(count(russia[[1]],1))  
colnames(russiadf) <- c("Nucleotides", "Russia")  
southkoreadf <- as.data.frame(count(southkorea[[1]],1))  
colnames(southkoreadf) <- c("Nucleotides", "Southkorea")  
spaindf <- as.data.frame(count(spain[[1]],1))  
colnames(spaindf) <- c("Nucleotides", "Spain")  
taiwandf <- as.data.frame(count(taiwan[[1]],1))  
colnames(taiwandf) <- c("Nucleotides", "Taiwan")  
turkeydf <- as.data.frame(count(turkey[[1]],1))  
colnames(turkeydf) <- c("Nucleotides", "Turkey")  
ukdf <- as.data.frame(count(uk[[1]],1))  
colnames(ukdf) <- c("Nucleotides", "UK")  
usadf <- as.data.frame(count(usa[[1]],1))  
colnames(usadf) <- c("Nucleotides", "USA")  
vietnamdf <- as.data.frame(count(vietnam[[1]],1))  
colnames(vietnamdf) <- c("Nucleotides", "Vietnam")  
  
df1 <- merge(argentinadf,australiadf)  
df2 <- merge(df1,brazildf)  
df3 <- merge(df2,francedf)  
df4 <- merge(df3,germanydf)  
df5 <- merge(df4,indiadf)  
df6 <- merge(df5,indonesiadf)  
df7 <- merge(df6,italydf)  
df8 <- merge(df7,irandf)  
df9 <- merge(df8,japandf)  
df10 <- merge(df9,mexicodf)  
df11 <- merge(df10,netherlandsdf)  
df12 <- merge(df11,russiadf)  
df13 <- merge(df12,southkoreadf)  
df14 <- merge(df13,spaindf)  
df15 <- merge(df14,taiwandf)  
df16 <- merge(df15,turkeydf)  
df17 <- merge(df16,ukdf)  
df18 <- merge(df17,usadf)  
df19 <- merge(df18,vietnamdf)

cantidad1 <- argentinadf[,2]  
cantidad2 <- australiadf[,2]  
cantidad3 <- brazildf[,2]  
cantidad4 <- francedf[,2]  
cantidad5 <- germanydf[,2]  
cantidad6 <- indiadf[,2]  
cantidad7 <- indonesiadf[,2]  
cantidad8 <- italydf[,2]  
cantidad9 <- irandf[,2]  
cantidad10 <- japandf[,2]  
cantidad11 <- mexicodf[,2]  
cantidad12 <- netherlandsdf[,2]  
cantidad13<- russiadf[,2]  
cantidad14 <- southkoreadf[,2]  
cantidad15 <- spaindf[,2]  
cantidad16 <- taiwandf[,2]  
cantidad17 <- turkeydf[,2]  
cantidad18 <- ukdf[,2]  
cantidad19 <- usadf[,2]  
cantidad20 <- vietnamdf[,2]  
  
grafica <- data.frame(  
 "Nucleótidos" = c("A","C", "G", "T"),  
 "Virus" = c(rep("Argentina", 4), rep("Australia", 4), rep("Brazil", 4), rep("France", 4), rep("Germany", 4), rep("India", 4), rep("Indonesia", 4), rep("Italy", 4), rep("Iran", 4), rep("Japan", 4), rep("Mexico", 4), rep("Netherlands", 4), rep("Russia", 4), rep("Southkorea", 4), rep("Spain", 4), rep("Taiwan", 4), rep("Turkey", 4), rep("UK", 4), rep("USA", 4), rep("Vietnam", 4)),  
 "Cantidad" = c(cantidad1, cantidad2, cantidad3, cantidad4,cantidad5,cantidad6,cantidad7,cantidad8,cantidad9,cantidad10,cantidad11,cantidad12,cantidad13,cantidad14,cantidad15,cantidad16,cantidad17,cantidad18,cantidad19,cantidad20)  
)

### Gráfica de comparación de las bases de ADN de cada secuencia de SARS-CoV-2 reportada en los 20 países.

graficar <- ggplot(grafica, aes(x = Nucleótidos, y = Cantidad, fill=Virus)) + geom\_bar(position="dodge", stat="identity")  
graficar



En la gráfica de barras se puede observar la cantidad de nucleótidos que tiene cada una de las secuencias que se investigaron del SARS-CoV-2 reportas en 20 países. Las variantes de cada país están representadas de colores ditintos.En el eje “x” de la grafica se ubican las bases nitrogenadas (A, C, G, T) y en el eje “y” la cantidad de las mismas.

country\_names <- c("sars\_usa", "sars\_india", "sars\_france", "sars\_germany", "sars\_brazil", "sars\_japan", "sars\_southkorea", "sars\_italy", "sars\_uk", "sars\_russia", "sars\_turkey", "sars\_spain", "sars\_vietnam", "sars\_australia", "sars\_taiwan", "sars\_argentina", "sars\_netherlands", "sars\_iran", "sars\_mexico", "sars\_indonesia")  
country\_seq <- c()  
  
for(country in country\_names){  
 cat("\n")  
 file\_path <- paste("C:/Users/daeld/OneDrive/Documentos/Programas/", country, ".fasta", sep="")  
 current <- readDNAStringSet(file\_path)  
 country\_seq <- append(country\_seq, current)  
}

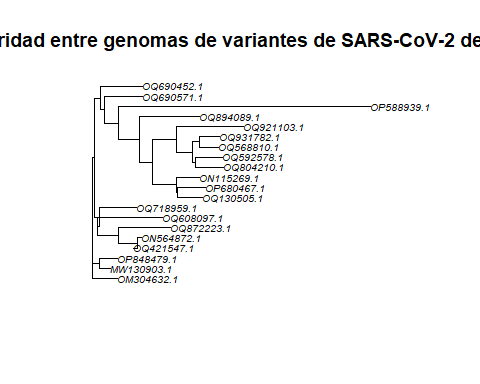
merged <- DNAStringSet(country\_seq)  
  
seqs <- OrientNucleotides(merged)

## ========================================================================================================================================================================================================  
##   
## Time difference of 0.23 secs

aligned <- AlignSeqs(seqs)

## Determining distance matrix based on shared 11-mers:  
## ================================================================================  
##   
## Time difference of 0.04 secs  
##   
## Clustering into groups by similarity:  
## ================================================================================  
##   
## Time difference of 0 secs  
##   
## Aligning Sequences:  
## ================================================================================  
##   
## Time difference of 2.32 secs  
##   
## Iteration 1 of 2:  
##   
## Determining distance matrix based on alignment:  
## ================================================================================  
##   
## Time difference of 0.01 secs  
##   
## Reclustering into groups by similarity:  
## ================================================================================  
##   
## Time difference of 0 secs  
##   
## Realigning Sequences:  
## ================================================================================  
##   
## Time difference of 1.96 secs  
##   
## Iteration 2 of 2:  
##   
## Determining distance matrix based on alignment:  
## ================================================================================  
##   
## Time difference of 0.01 secs  
##   
## Reclustering into groups by similarity:  
## ================================================================================  
##   
## Time difference of 0 secs  
##   
## Realigning Sequences:  
## ================================================================================  
##   
## Time difference of 0.01 secs  
##   
## Refining the alignment:  
## ================================================================================  
##   
## Time difference of 2.81 secs

writeXStringSet(aligned,file="variantes\_alineadas.fasta")  
  
dna\_aligned <- read.alignment("variantes\_alineadas.fasta", format = "fasta")  
  
D <- dist.alignment(dna\_aligned, matrix = "similarity")  
  
arbol\_filogenetico <- nj(D)  
plot(arbol\_filogenetico, cex = 0.6)  
title("Similaridad entre genomas de variantes de SARS-CoV-2 de 20 países")

 El gráfico muestra un árbol filogenético de secuencias de SARS-CoV-2 reportadas en los 20 países con más casos reportados. Cada variante está representada por una rama del árbol, y la longitud de cada rama representa la distancia evolutiva entre las secuencias. Las ramas más largas indican que las secuencias son más divergentes y tienen más diferencias genéticas. Las ramas más cortas, por otro lado, indican que las secuencias son más similares y tienen menos diferencias genéticas.

### REFERENCIAS

Hu, B. Y., Guo, H., Zhou, P., & Shi, Z. (2021). Characteristics of SARS-CoV-2 and COVID-19. Nature Reviews Microbiology, 19(3), 141-154. <https://doi.org/10.1038/s41579-020-00459-7>

NCBI SARS-CoV-2 Resources. (s. f.). <https://www.ncbi.nlm.nih.gov/sars-cov-2/>

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. Journal of advanced research, 24, 91–98. <https://doi.org/10.1016/j.jare.2020.03.005>

Hasöksüz, M., Kiliç, S., & Saraç, F. (2020). Coronaviruses and SARS-COV-2. Turkish journal of medical sciences, 50(SI-1), 549–556. <https://doi.org/10.3906/sag-2004-127>

Nova N. (2021). Cross-Species Transmission of Coronaviruses in Humans and Domestic Mammals, What Are the Ecological Mechanisms Driving Transmission, Spillover, and Disease Emergence?. Frontiers in public health, 9, 717941. <https://doi.org/10.3389/fpubh.2021.717941>

Islam, A., Ferdous, J., Sayeed, M. A., Islam, S., Kaisar Rahman, M., Abedin, J., Saha, O., Hassan, M. M., & Shirin, T. (2021). Spatial epidemiology and genetic diversity of SARS-CoV-2 and related coronaviruses in domestic and wild animals. PloS one, 16(12), e0260635. <https://doi.org/10.1371/journal.pone.0260635>

Krause, P. R., Fleming, T. R., Longini, I. M., Peto, R., Briand, S., Heymann, D. L., Beral, V., Snape, M. D., Rees, H., Ropero, A. M., Balicer, R. D., Cramer, J. P., Muñoz-Fontela, C., Gruber, M., Gaspar, R., Singh, J. A., Subbarao, K., Van Kerkhove, M. D., Swaminathan, S., Ryan, M. J., … Henao-Restrepo, A. M. (2021). SARS-CoV-2 Variants and Vaccines. The New England journal of medicine, 385(2), 179–186. <https://doi.org/10.1056/NEJMsr2105280>

Cosar, B., Karagulleoglu, Z. Y., Unal, S., Ince, A. T., Uncuoglu, D. B., Tuncer, G., Kilinc, B. R., Ozkan, Y. E., Ozkoc, H. C., Demir, I. N., Eker, A., Karagoz, F., Simsek, S. Y., Yasar, B., Pala, M., Demir, A., Atak, I. N., Mendi, A. H., Bengi, V. U., Cengiz Seval, G., … Demir-Dora, D. (2022). SARS-CoV-2 Mutations and their Viral Variants. Cytokine & growth factor reviews, 63, 10–22. <https://doi.org/10.1016/j.cytogfr.2021.06.001>

Fabrizio, C., Termine, A., Caputo, V., Megalizzi, D., Calvino, G., Trastulli, G., Ingrascì, A., Ferrante, S., Peconi, C., Rossini, A., Salvia, A., Caltagirone, C., Strafella, C., Giardina, E., & Cascella, R. (2022). Analysis of Genetic Variants Associated with COVID-19 Outcome Highlights Different Distributions among Populations. Journal of personalized medicine, 12(11), 1851. <https://doi.org/10.3390/jpm12111851>