



Tecnológico de Monterrey

Campus Monterrey

Evidencia 2:

Integrantes:

| | |
|---------------------------|-----------|
| Rodrigo Fajardo Saldaña | A01194935 |
| Francisco Rocha Juárez | A01730560 |
| Ana Carolina Magaña Lemus | A00832042 |
| Alan Fernando Razo Peña | A01703350 |

Profesor: Heriberto García Coronado

Curso: Análisis de biología computacional

Grupo: 502

abril de 2021

PARTE 1:

<https://youtu.be/wtIYiZX5vHM>

PARTE 2.1:

```
virus <- c( "MT856455" , "MT835383" , "MT670017" , "MT470219" , "MT594402" ,  
"MT704818" , "MT577009" , "MW135333" , "MT940490" , "MW134570" , "MT810754" ,  
"MW030194" , "MT919776" , "MT890462" , "MT820472" , "MT324062" , "MW056033" ,  
"MT787487" , "MT873892" , "MW133662" )
```

```
#vector <- c( "Sars_CoV_2_Bangladesh" , "Sars_CoV_2_Brazil" , "Sars_CoV_2_Chile" ,  
"Sars_CoV_2_Colombia" , "Sars_CoV_2_Francia" , "Sars_CoV_2_Alemania" ,  
"Sars_CoV_2_India" , "Sars_CoV_2_Iran" , "Sars_CoV_2_Iraq" , "Sars_CoV_2_Italia" ,  
"Sars_CoV_2_Mexico" , "Sars_CoV_2_Peru" , "Sars_CoV_2_Filipinas" ,  
"Sars_CoV_2_Rusia" , "Sars_CoV_2_Arabia_Saudita" , "Sars_CoV_2_Sudafrica" ,  
"Sars_CoV_2_Espana" , "Sars_CoV_2_Turquia" , "Sars_CoV_2_Reino_Unido" ,  
"Sars_CoV_2_Estados_Unidos" )
```

```
``{r}
```

```
library(Biostrings)
```

```
library(seqinr)
```

```
library(adegenet)
```

```
library(ape)
```

```
library(ggtree)
```

```
library(DECIPHER)
```

```
library(viridis)
```

```
library(ggplot2)
```

```
{r}
```

```
virus_sequences <- read.GenBank(virus)
```

```
{r}
```

```
str(virus_sequences)
```

```
{r}
```

```
attributes(virus_sequences)
```

```
names(virus_sequences) # <-vector
```

```
attr(virus_sequences, "species")
```

```
{r}
```

```
write.dna(virus_sequences, file ="virus_seqs.fasta", format = "fasta", append =  
FALSE, nbcol = 6, colsep = " ", colw = 10)
```

```
{r}  
virus_seq_not_align <- readDNAStringSet("virus_seqs.fasta", format = "fasta")  
virus_seq_not_align
```

```
{r}  
virus_seq_not_align <- OrientNucleotides(virus_seq_not_align)  
  
virus_seq_align <- AlignSeqs(virus_seq_not_align)
```

```
{r}  
BrowseSeqs(virus_seq_align, highlight=0)
```

```
{r}  
writeXStringSet(virus_seq_align, file="virus_seq_align.fasta")
```

```
{r}  
virus_aligned <- read.alignment("virus_seq_align.fasta", format = "fasta")
```

```
{r}  
matriz_distancia <- dist.alignment(virus_aligned, matrix = "similarity")
```

```
{r}  
temp <- as.data.frame(as.matrix(matriz_distancia))  
table.paint(temp, cleg=0, clabel.row=.5, clabel.col=.5) + scale_color_viridis()
```

```
{r}  
"Tamano de Sars Covid 2 en Bangladesh:"  
tam1 = tam(virus_sequences$MT856455)  
print(tam1)  
"Tamano de Sars Covid 2 en Brazil:"  
tam2 = tam(virus_sequences$MT835383)  
print(tam2)  
"Tamano de Sars Covid 2 en Chile:"  
tam3 <- tam(virus_sequences$MT670017)  
print(tam3)  
"Tamano de Sars Covid 2 en Colombia:"  
tam4 <- tam(virus_sequences$MT470219)  
print(tam4)
```

```
"Tamano de Sars Covid 2 en Francia:"
tam5 <- tam(virus_sequences$MT594402)
print(tam5)
"Tamano de Sars Covid 2 en Alemania:"
tam6 <- tam(virus_sequences$MT704818)
print(tam6)
"Tamano de Sars Covid 2 en India:"
tam7 <- tam(virus_sequences$MT577009)
print(tam7)
"Tamano Sars Covid 2 en Iran:"
tam8 <- tam(virus_sequences$MW135333)
print(tam8)
"Tamano de Sars Covid 2 en Iraq:"
tam9 <- tam(virus_sequences$MW134570)
print(tam9)
"Tamano de Sars Covid 2 en Italia:"
tam10 <- tam(virus_sequences$MT940490)
print(tam10)
"Tamano de Sars Covid 2 en Mexico:"
tam11 <- tam(virus_sequences$MT810754)
print(tam11)
"Tamano de Sars Covid 2 en Peru:"
tam12 <- tam(virus_sequences$MW030194)
print(tam12)
"Tamano de Sars Covid 2 en Filipinas:"
tam13 <- tam(virus_sequences$MT919776)
print(tam13)
"Tamano de Sars Covid 2 en Rusia:"
tam14 <- tam(virus_sequences$MT890462)
print(tam14)
"Tamano de Sars Covid 2 en Arabia Saudita:"
tam15 <- tam(virus_sequences$MT820472)
print(tam15)
"Tamano de Sars Covid 2 en Sudafrica:"
tam16 <- tam(virus_sequences$MT324062)
print(tam16)
"Tamano de Sars Covid 2 en Espana:"
tam17 <- tam(virus_sequences$MW056033)
print(tam17)
"Tamano de Sars Covid 2 en Turquía:"
tam18 <- tam(virus_sequences$MT787487)
print(tam18)
"Tamano de Sars Covid 2 en Reino Unido:"
tam19 <- tam(virus_sequences$MT873892)
print(tam19)
"Tamano de Sars Covid 2 en Estados Unidos:"
tam20 <- tam(virus_sequences$MW133662)
print(tam20)
```

```

{r}
fasta1 <- read.fasta("BAN.fasta")
covBAN <- as.character(fasta1[[1]])
fasta2 <- read.fasta("BRA.fasta")
covBRA <- as.character(fasta2[[1]])
fasta3 <- read.fasta("CHI.fasta")
covCHI <- as.character(fasta3[[1]])
fasta4 <- read.fasta("COL.fasta")
covCOL <- as.character(fasta4[[1]])
fasta5 <- read.fasta("FRA.fasta")
covFRA <- as.character(fasta5[[1]])
fasta6 <- read.fasta("GER.fasta")
covGER <- as.character(fasta6[[1]])
fasta7 <- read.fasta("IND.fasta")
covIND <- as.character(fasta7[[1]])
fasta8 <- read.fasta("IRAN.fasta")
covIRAN <- as.character(fasta8[[1]])
fasta9 <- read.fasta("IRAQ.fasta")
covIRAQ <- as.character(fasta9[[1]])
fasta10 <- read.fasta("ITA.fasta")
covITA <- as.character(fasta10[[1]])
fasta11 <- read.fasta("MEX.fasta")
covMEX <- as.character(fasta11[[1]])
fasta12 <- read.fasta("PERU.fasta")
covPERU <- as.character(fasta12[[1]])
fasta13 <- read.fasta("PHI.fasta")
covPHI <- as.character(fasta13[[1]])
fasta14 <- read.fasta("RUS.fasta")
covRUS <- as.character(fasta14[[1]])
fasta15 <- read.fasta("SAUDIARA.fasta")
covSAR <- as.character(fasta15[[1]])
fasta16 <- read.fasta("SOUTHAFR.fasta")
covSAF <- as.character(fasta16[[1]])
fasta17 <- read.fasta("SPA.fasta")
covSPA <- as.character(fasta17[[1]])
fasta18 <- read.fasta("TUR.fasta")
covTUR <- as.character(fasta18[[1]])
fasta19 <- read.fasta("UK.fasta")
covUK <- as.character(fasta19[[1]])
fasta20 <- read.fasta("USA.fasta")
covUSA <- as.character(fasta20[[1]])

```

```

{r}
porcentaje <- function(s) {
  library("seqinr")

```

```

ct<-count(s,wordsize = 1)
p<-ct*100/sum(ct)
return (p)
}

```

```

{r}
"Composicion de nucleotidos de Sars Covid 2 Bangladesh (porcentaje):"
prozen1 <- porcentaje(covBAN)
print(prozen1)
"Composicion de nucleotidos de Sars Covid 2 Brazil (porcentaje):"
prozen2 <- porcentaje(covBRA)
print(prozen2)
"Composicion de nucleotidos de Sars Covid 2 Chile (porcentaje):"
prozen3 <- porcentaje(covCHI)
print(prozen3)
"Composicion de nucleotidos de Sars Covid 2 Colombia (porcentaje):"
prozen4 <- porcentaje(covCOL)
print(prozen4)
"Composicion de nucleotidos de Sars Covid 2 Francia (porcentaje):"
prozen5 <- porcentaje(covFRA)
print(prozen5)
"Composicion de nucleotidos de Sars Covid 2 Alemania (porcentaje):"
prozen6 <- porcentaje(covGER)
print(prozen6)
"Composicion de nucleotidos de Sars Covid 2 India (porcentaje):"
prozen7 <- porcentaje(covIND)
print(prozen7)
"Composicion de nucleotidos de Sars Covid 2 Iran (porcentaje):"
prozen8 <- porcentaje(covIRAN)
print(prozen8)
"Composicion de nucleotidos de Sars Covid 2 Iraq (porcentaje):"
prozen9 <- porcentaje(covIRAQ)
print(prozen9)
"Composicion de nucleotidos de Sars Covid 2 Italia (porcentaje):"
prozen10 <- porcentaje(covITA)
print(prozen10)
"Composicion de nucleotidos de Sars Covid 2 Mexico (porcentaje):"
prozen11 <- porcentaje(covMEX)
print(prozen11)
"Composicion de nucleotidos de Sars Covid 2 Peru (porcentaje):"
prozen12 <- porcentaje(covPERU)
print(prozen12)
"Composicion de nucleotidos de Sars Covid 2 Filipinas (porcentaje):"
prozen13 <- porcentaje(covPHI)
print(prozen13)
"Composicion de nucleotidos de Sars Covid 2 Rusia (porcentaje):"
prozen14 <- porcentaje(covRUS)

```

```

print(prozen14)
"Composicion de nucleotidos de Sars Covid 2 Arabia Saudita (porcentaje):"
prozen15 <- porcentaje(covSAR)
print(prozen15)
"Composicion de nucleotidos de Sars Covid 2 Sudafrica (porcentaje):"
prozen16 <- porcentaje(covSAF)
print(prozen16)
"Composicion de nucleotidos de Sars Covid 2 Espana (porcentaje):"
prozen17 <- porcentaje(covSPA)
print(prozen17)
"Composicion de nucleotidos de Sars Covid 2 Turquía (porcentaje):"
prozen18 <- porcentaje(covTUR)
print(prozen18)
"Composicion de nucleotidos de Sars Covid 2 Reino Unido (porcentaje):"
prozen19 <- porcentaje(covUK)
print(prozen19)
"Composicion de nucleotidos de Sars Covid 2 Estados Unidos (porcentaje):"
prozen20 <- porcentaje(covUSA)
print(prozen20)

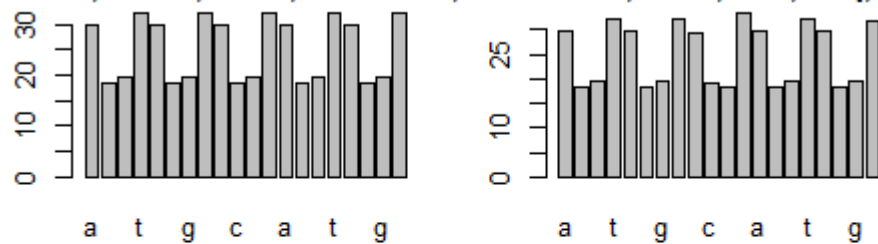
```

```

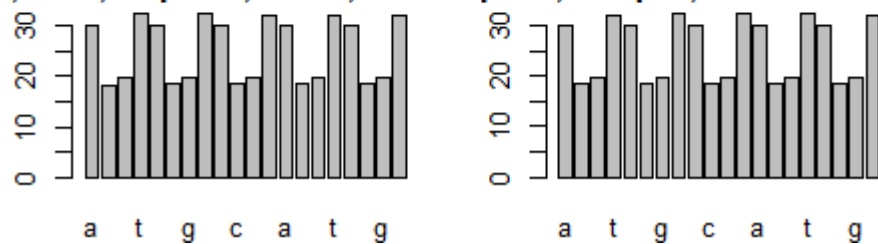
{r}
par(mar=c(2,2,2,2))
par(mfrow=c(2,2))
grafica1 <- c(prozen1 , prozen2 , prozen3 , prozen4 , prozen5)
grafica2 <- c(prozen6 , prozen7 , prozen8 , prozen9 , prozen10)
grafica3 <- c(prozen11 , prozen12 , prozen13 , prozen14 , prozen15)
grafica4 <- c(prozen16 , prozen17 , prozen18 , prozen19 , prozen20)
barplot(grafica1, main = "Bangladesh, Brasil, Chile, Colombia, Francia", xlab =
"Nucleotidos", ylab = "Porcentaje")
barplot(grafica2, main = "Alemania, India, Iran, Iraq, Italia", xlab = "Nucleotidos", ylab =
"Porcentaje")
barplot(grafica3, main = "Mexico, Peru, Filipinas, Rusia, Arabia Saudita", xlab =
"Nucleotidos", ylab = "Porcentaje")
barplot(grafica4, main = "Sudafrica, Espana, Turquía, Reino Unido, Estados Unidos", xlab =
"Nucleotidos", ylab = "Porcentaje")

```

India, Brasil, Chile, Colombia, Alemania, India, Iran, Iraq, Italia



Peru, Filipinas, Rusia, Arabia Saudita, España, Turquía, Reino Unido, Estados Unidos



Como podemos observar en estas gráficas en la parte superior, podemos ver la distribución del contenido de los Nucleótidos en los diferentes países del mundo, esto con relación a los diferentes tipos de Covid-Sars2 que se han desarrollado en cada país. Al igual que en el código pasado podemos observar la longitud de los genomas al igual que la cantidad guanina citosina esto con las funciones ggplot, grid.arrange. Como podemos ver hay un patrón parecido en los países que cuentan con cercanía geográfica, como se puede ver los países latinoamericanos, así como los europeos comparten similitudes en la distribución de los contenidos nucleótidos, tal vez esto se debe por la cercanía entre estos ya se contaba con que el virus iba a estar expuesto a condiciones parecidas.

PARTE 2.2:

```
library(ape)
library(Biostrings)
library(DECIPHER)
library(seqinr)
library(ade4)
library(viridis)
```

```
getwd()
setwd("~/Evidencia2")
```

```
corona_virus <-
c("AY278489", "AY390556", "AY485277", "AY508724", "JX869059", "MN908947", "MN985325", "
```



```
MT292571","MT820489","MT845878","MT873893","MT876433","MT919790","MT959161",
MW000351")
```

```
virus_sequences<-read.GenBank(corona_virus)
```

```
write.dna((virus_sequences),file="coronavirus_seqs.fasta",format="fasta")
```

```
virus_seq_not_align <- readDNAStringSet("coronavirus_seqs.fasta",format="fasta")
```

```
virus_seq_not_align <- OrientNucleotides(virus_seq_not_align)
```

```
virus_seq_align <- AlignSeqs(virus_seq_not_align)
```

```
BrowseSeqs(virus_seq_align)
```

```
writeXStringSet(virus_seq_align, file="coronavirus_seq_align.fasta")
```

```
virus_aligned <- read.alignment("coronavirus_seq_align.fasta", format = "fasta")
```

```
matriz_distancia <- dist.alignment(virus_aligned, matrix = "similarity")
```

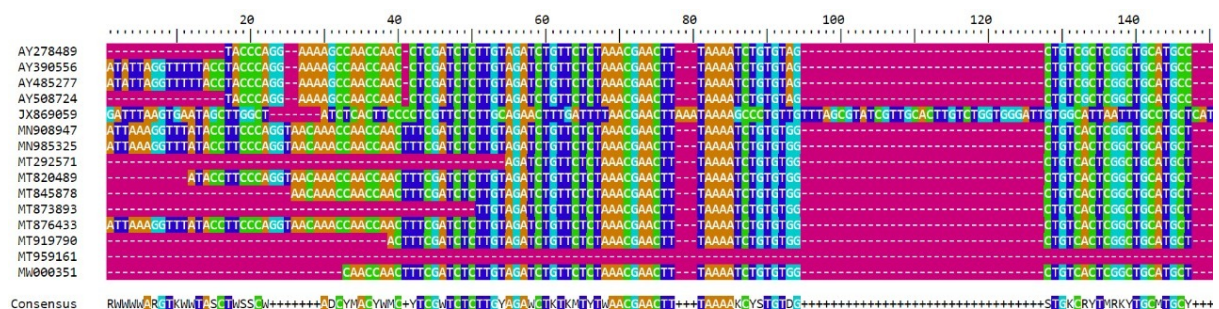
```
temp <- as.data.frame(as.matrix(matriz_distancia))
```

```
table.paint(temp, cleg=0, clabel.row=.5, clabel.col=.5) +scale_color_viridis()
```

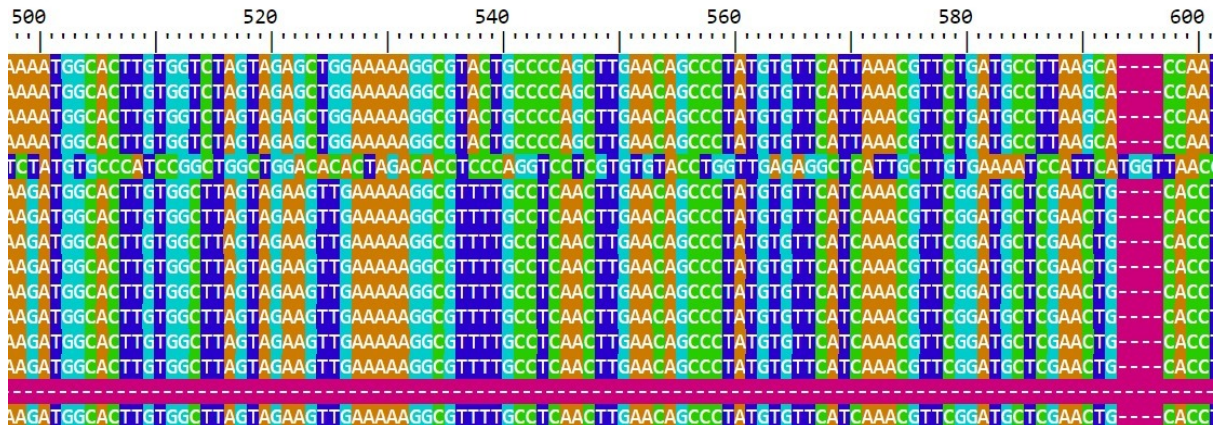
```
virus_tree <- nj(matriz_distancia)
```

```
virus_tree <- ladderize(virus_tree)
```

```
plot(virus_tree)
```



Consensus RNMNMWRRKRWTSWSSN+++++DGYNACYYM+YTGWCTCTEYAGAMEKKKNTVAACGAAGT++TAAACCTYSITDQ+++++STKRYRYRKYTGCTGGY++



VMDATCKSMGYWKYSBYTRFYWRSSWBRIWARRSRCSITHYRSSYQMDKTKWRYNSQYSKNGWGWKSMYYANWSSITSKGAWRMYYBAWBYR+++YMMM



| | AY278489 | AY390556 | AY485277 | AY508724 | JX869059 | MN908947 | MN985325 | MT292571 | MT820489 | MT845878 | MT873893 | MT876433 | MT919790 | MT959161 | MW000351 |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| AY278489 | | | | | | | | | | | | | | | |
| AY390556 | | | | | | | | | | | | | | | |
| AY485277 | | | | | | | | | | | | | | | |
| AY508724 | | | | | | | | | | | | | | | |
| JX869059 | | | | | | | | | | | | | | | |
| MN908947 | | | | | | | | | | | | | | | |
| MN985325 | | | | | | | | | | | | | | | |
| MT292571 | | | | | | | | | | | | | | | |
| MT820489 | | | | | | | | | | | | | | | |
| MT845878 | | | | | | | | | | | | | | | |
| MT873893 | | | | | | | | | | | | | | | |
| MT876433 | | | | | | | | | | | | | | | |
| MT919790 | | | | | | | | | | | | | | | |
| MT959161 | | | | | | | | | | | | | | | |
| MW000351 | | | | | | | | | | | | | | | |

Interpretación:

En este caso se hizo el análisis de 15 variantes del coronavirus, en el primer caso tenemos una comparación de las 15 cadenas de nucleótidos, comparándose una por una, realmente aquí a simple vista no podemos obtener mucha información, ya que es mucha información visual, sin embargo podemos ver que las cadenas se parecen mucho en ciertos segmentos y tienen mucho en común. Después en el árbol filogenético, podemos observar que hay 2 variaciones que se parecen más, la “MT959161” y la “JX869059”, después tenemos otras 4

variaciones que son parecidas, la “AY508724”, “AY485277”, “AY390556”, “AY278489”, finalmente están las restantes que son muy diferentes a estos dos subgrupos. En la tabla de colores que mandamos a graficar, podemos comprobar esto, ya que se observan los subgrupos con un color gris o blanco, mientras que los que son más distintos como el “JX869059” aparecen en un color negro. Podemos concluir en que los virus pueden ir evolucionando de otros y pueden tener variaciones en sus nucleótidos, que al sintetizar diferentes proteínas, se pueden comportar de manera diferente. Comprender esto nos permite también entender cómo tratarlos, ya que hay virus que tienen mucho en común y podemos rastrear su procedencia.

Referencias APA:

Liya G, Yuguang W, Jian L, Huaiping Y, Xue H, Jianwei H, Jiaju M, Youran L, Chen M, Yiqing J. Studies on viral pneumonia related to novel coronavirus SARS-CoV-2, SARS-CoV, and MERS-CoV: a literature review. APMIS. 2020 Jun;128(6):423-432. doi: 10.1111/apm.13047. PMID: 32363707.

Asrani P, Hasan GM, Sohal SS, Hassan MI. Molecular Basis of Pathogenesis of Coronaviruses: A Comparative Genomics Approach to Planetary Health to Prevent Zoonotic Outbreaks in the 21st Century. OMICS. 2020 Nov;24(11):634-644. doi: 10.1089/omi.2020.0131. Epub 2020 Sep 16. PMID: 32940573.