

Instituto Tecnológico y de Estudios Superiores de Monterrey



Evidencia 2 | Proyecto integrador

Análisis de biología computacional

Nombres

Jose Enrique Zaragoza Bravo | A00572939

7 de mayo del 2023

En esta entrega se analizará y se explicara que significa el fenómeno de zoonosis del virus con el cual estaremos trabajando con ecuencias de los coronavirus de otras especies que son similares a SARS-CoV-2

¿Qué es el fenómeno de zoonosis en un virus? pues este fenómeno trata que cuando se produce un virus en un animal se llega a transmitir al ser humano esta capacidad sucede cuando el virus evoluciona y puede infectar a humanos generalmente por el contacto directo con animales infectados o al tener contacto con las secreciones de animales.

En el caso del SARS-CoV-2, se cree que el virus se originó en murciélagos y se transmitió a los humanos a través de un animal intermediario, como podría haber sido un pangolín o un murciélago.

Es importante como sociedad que conozcamos que es la zoonosis ya que muchos casos de enfermedades graves para los humanos provienen de los animales como virus mutados, estos casos cada vez más van en aumento por la evolución del humano ya que al seguir evolucionando van utilizando más recursos y degradando nuestro planeta tierra y todos estos cambios en el clima o en la urbanización hace que se hagan más propensos a la infección de animales a humanos.

Son importantes estas investigaciones ya que le agrega importancia al saber cómo se propagan los virus y concientiza a seguir con medidas para evitar estos casos.

Para trabajar se utilizaron variantes de bovino, de murciélagos y de humanos todos con la característica de pertenecer al grupo de Coronaviridae:

```
~~~{r}
virus <- c( "PA573586", "NC_009019", "NC_009020", "OQ297709", "KU131570", "MH940245", "LV469811", "MN996532", "MG772933",
"AF501235", "KT191142", "MG772921", "JN204180", "KY419113", "S52273", "OQ297709", "NC_009019", "PA550519", "OM189517", "MG991679")
PA573586 <- c("SARSCoV")
NC_009019 <- c("HKU4")
NC_009020 <- c("HKU5")
OQ297709 <- c("HKU9")
KU131570 <- c("HCoV-OC43")
MH940245 <- c("HCoV-HKU1")
LV469811 <- c("MERSCoV")
MN996532 <- c("RaTG13")
MG772933 <- c("batSLCoVZC45")
AF501235 <- c("Influenza")
KT191142 <- c("ebola")
MG772921 <- c("batSLCoVZXC21")
JN204180 <- c("BCoV")
KY419113 <- c("PHEV")
S52273 <- c("MHV")
OQ297709 <- c("batHKU9")
NC_009019 <- c("TYHKU4")
PA550519 <- c("wo")
OM189517 <- c("alpha")
MG991679 <- c("gamma")
~~~
```

Librerías utilizadas:

```
{r}  
library(Biostrings)  
library(sequinr)  
library(adeigenet)  
library(ape)  
library(ggtree)  
library(DECIPHER)  
library(viridis)  
library(ggplot2)
```

Obtenemos las secuencias de dichas variantes:

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

Sacamos las estructuras y podemos ver la longitud de cada una de las variantes.

```
List of 20  
$ PA573586 : raw [1:717] 88 18 48 48 ...  
$ NC_009019: raw [1:30286] 48 88 18 18 ...  
$ NC_009020: raw [1:30482] 48 88 18 18 ...  
$ OQ297709 : raw [1:29085] 28 18 48 28 ...  
$ KU131570 : raw [1:30746] 28 18 28 18 ...  
$ MH940245 : raw [1:29811] 48 88 18 18 ...  
$ LV469811 : raw [1:4062] 88 18 48 88 ...  
$ MN996532 : raw [1:29855] 88 18 18 88 ...  
$ MG772933 : raw [1:29802] 88 18 88 18 ...  
$ AF501235 : raw [1:1729] 88 18 48 48 ...  
$ KT191142 : raw [1:336] 88 18 18 28 ...  
$ MG772921 : raw [1:428] 28 28 48 88 ...  
$ JN204180 : raw [1:248] 48 28 18 88 ...  
$ KY419113 : raw [1:29900] 18 48 88 48 ...  
$ S52273 : raw [1:136] 18 88 18 88 ...  
$ OQ297709 : raw [1:29085] 28 18 48 28 ...  
$ NC_009019: raw [1:30286] 48 88 18 18 ...  
$ PA550519 : raw [1:3822] 88 18 48 18 ...  
$ OM189517 : raw [1:3087] 28 28 18 48 ...  
$ MG991679 : raw [1:469] 18 28 18 88 ...
```

Para comprobar los siguientes datos que sean los correctos y coincidan los códigos de los virus utilizamos el siguiente código:

```

{r}
attributes(virus_sequences)
names(virus_sequences)
attr(virus_sequences, "species")

```

Que nos arroja los datos completos de cada uno de los códigos del virus:

```

$names
[1] "PA573586" "NC_009019" "NC_009020" "OQ297709" "KU131570" "MH940245" "LV469811" "MN996532" "MG772933" "AF501235" "KT191142" "MG772921" "JN204180"
[14] "KY419113" "S52273" "OQ297709" "NC_009019" "PA550519" "OM189517" "MG991679"

$class
[1] "DNABin"

$description
[1] "PA573586.1 WO 2023002944-A/33: ANTIBODY THAT BINDS TO SARS-COV-2 AND SARS-COV"
[2] "NC_009019.1 Tylonycteris bat coronavirus HKU4, complete genome"
[3] "NC_009020.1 Pipistrellus bat coronavirus HKU5, complete genome"
[4] "OQ297709.1 MAG: Rousettus bat coronavirus HKU9 isolate GD/L113.18/2022, partial genome"
[5] "KU131570.1 Human coronavirus OC43 strain HCoV-OC43/UK/London/2011, complete genome"
[6] "MH940245.1 Human coronavirus HKU1 isolate SI17244, complete genome"
[7] "LV469811.1 JP 2016536346-A/1: IMMUNOGENIC MIDDLE EAST RESPIRATORY SYNDROME CORONAVIRUS (MERS-CoV) COMPOSITIONS AND METHODS"
[8] "MN996532.2 Bat coronavirus RaTG13, complete genome"
[9] "MG772933.1 Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome"
[10] "AF501235.1 Influenzavirus A (A/duck/Shanghai/1/2000) hemagglutinin gene, complete cds"
[11] "KT191142.1 Ebola virus isolate Ebola virus H.sapiens-wt/NIV/2015/India/NIV152881 VP40 protein (VP40) gene, partial cds"
[12] "MG772921.1 Bat SARS-like coronavirus isolate bat-SL-CoVZXC21 RNA-dependent RNA polymerase gene, partial cds"
[13] "JN204180.1 Bovine coronavirus isolate BCoV C11/11/Ire RNA-dependent RNA polymerase gene, partial cds"
[14] "KY419113.1 Porcine hemagglutinating encephalomyelitis virus strain PHEV CoV USA-15TOSU1582, complete genome"
[15] "S52273.1 {5' region} [murine hepatitis virus MHV, Genomic RNA Mutant, 136 nt]"
[16] "OQ297709.1 MAG: Rousettus bat coronavirus HKU9 isolate GD/L113.18/2022, partial genome"
[17] "NC_009019.1 Tylonycteris bat coronavirus HKU4, complete genome"
[18] "PA550519.1 WO 2022118975-A/11: Variants of SARS-CoV-2 Spike protein and use thereof"
[19] "OM189517.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/COL/Medellin_ALPHA BC8_20I (Alpha, V1)/2021 surface glycoprotein (S) gene, partial cds"
[20] "MG991679.1 Gammacoronavirus sp. clone SP/22991-23459/G001/AU/2016 spike protein (S) gene, partial cds"

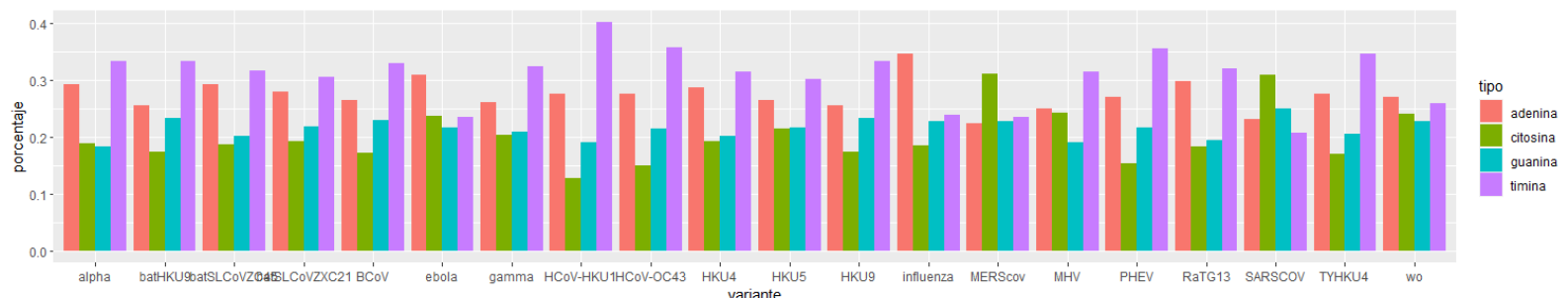
```

Para analizar los resultados obtuvimos dos gráficos considerando las variantes correspondientes obtenemos la gráfica con la que podemos interpretar el porcentaje de La adenina (A), citosina (C), guanina (G) y timina (T) son las cuatro bases nitrogenadas que se encuentran en el ADN.

```

library("ggplot2")
tabla1<-data.frame(variante=rep(c("alpha", "TYHKU4", "batHKU9", "gamma", "MHV", "PHEV", "BCoV", "wo", "batC21", "batC45", "RaTG13", "HCoV-HKU1", "HCoV-OC43", "ebola", "HCoV-HKU4", "HCoV-HKU5", "HCoV-HKU9", "influenza", "MERScov", "MHV", "PHEV", "RaTG13", "SARSCOV", "TYHKU4", "wo"), each=4), tipo = rep(c("Adenina", "Citosina", "Guanina", "Timina"), 6), porcentaje = c(count(alpha[[1]], 1), count(TYHKU4[[1]], 1), count(batHKU9[[1]], 1), count(gamma[[1]], 1), count(MHV[[1]], 1), count(PHEV[[1]], 1), count(BCoV[[1]], 1), count(wo[[1]], 1), count(batC21[[1]], 1), count(batC45[[1]], 1), count(RaTG13[[1]], 1), count(HCoV-HKU1[[1]], 1), count(HCoV-OC43[[1]], 1), count(ebola[[1]], 1), count(HCoV-HKU4[[1]], 1), count(HCoV-HKU5[[1]], 1), count(HCoV-HKU9[[1]], 1), count(influenza[[1]], 1), count(MERScov[[1]], 1), count(MHV[[1]], 1), count(PHEV[[1]], 1), count(RaTG13[[1]], 1), count(SARSCOV[[1]], 1), count(TYHKU4[[1]], 1), count(wo[[1]], 1)))
ggplot(data = tabla1) + geom_bar(aes(x = variante, y = porcentaje, fill = tipo), position = "dodge", stat = "identity")

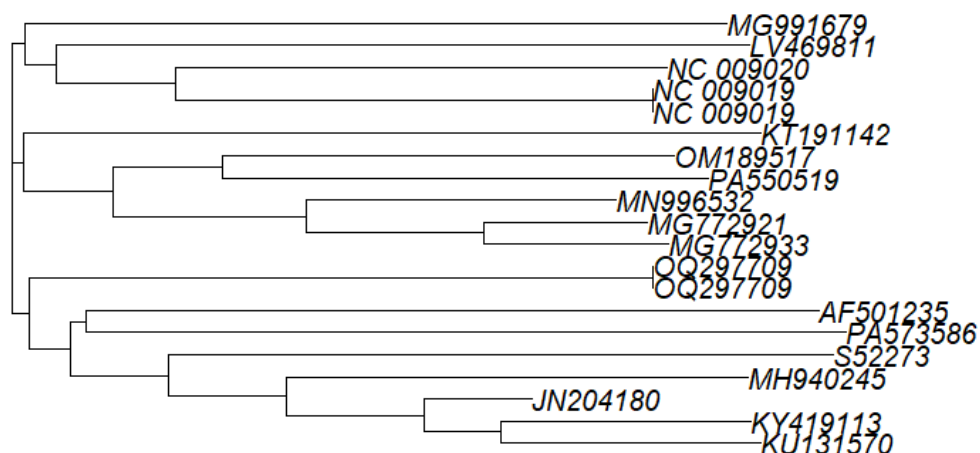
```



Al tener variantes de diferentes especies vivas como humanos y murciélagos, podemos darnos cuenta que en las variantes que afectan principalmente a humanos tienen niveles de ADN más estables entre sí, esto significa que son más estables por lo tanto afectan más fácilmente a los humanos y las variantes de los animales tienen porcentajes de timina muy altos, además podemos apreciar pocas similitudes entre variables, la única que posiblemente comparten es que son de la familia de Coronaviridae y que todos pueden causar enfermedades respiratorias en los seres humanos y/o animales y de diferencias principales son hay coronavirus que son altamente contagiosos (como SARS-CoV-2), mientras que otros no lo son y que la estructura del genoma de ARN y las proteínas virales varían entre diferentes coronavirus.

```
13. Plot del árbol:
```{r}
virus_plot_filogenetico <- ladderize(virus_filogenetico)
plot(virus_plot_filogenetico)
title("Análisis jerárquico global de variantes analizadas con base a SARS_COV2")
```

### Análisis jerárquico global de variantes analizadas con base a SARS\_COV2



La jerarquía que manejamos en esta gráfica es que cada rama señala grupos de cada variante y se basa con los parecidos genéticos de cada una de las variantes, así que señala variantes humanas y de animal de manera separada y juntándose con los parecidos más grandes entre códigos genéticos.

PA573586 = **SARS COV**

NC\_009019 = **HKU4**

NC\_009020 = **HKU5**

MG991679 = **gamma**

LV469811 = **MERScov**

NC\_009020 = **HKU5**

NC\_009019 = **TYHKU4**

NC\_009019 = **HKU4**

KT191142 = **ebola**

OM189517 = **alpha**

PA550519 = **wo**

MN996532 = **RaTG13**

MG772921 = **batSLCoVZXC21**

MG772933 = **batSLCoVZC45**

OQ297709 = **HKU9**

OQ297709 = **batHKU9**

AF501235 = **Influenza**

PA573586 = **SARSCOV**

S52273 = **MHV**

MH940245 = **HCoV-HKU1**

JN204180 = **BCoV**

KY419113 = **PHEV**

KU131570 = **HCoV-OC43**

Como conclusión esta práctica nos ayuda a darnos cuenta de los posibles orígenes de las enfermedades y de las similitudes que tienen entre cada una de ellas, y utilizando las herramientas y el conocimiento biológico necesario podemos expresar todas estas similitudes y diferencias entre los virus, todo esto para seguir dándole la importancia de las enfermedades y de lo fácil es que sigan evolucionando.

Video de la parte 1:

<https://youtu.be/jZ44BCOEWSA>