

## Evidencia 1 | Análisis Inicial

Juan Enrique Ayala Zapata-A01711235

2024-04-26

### Parte 1: Responde justificadamente las siguientes preguntas

**1. Investiga, ¿Cuál es la situación actual de COVID-19 a nivel mundial, en México, en tu estado de origen y en tu municipio, alcaldía o colonia?**

De acuerdo al sitio de la Johns Hopkins University & Medicine, dejó de coleccionar información sobre el COVID-19 el 10 de marzo de 2023, debido a que dejó de ser considerada como pandemia, pero de acuerdo a la última medición se tiene la información de que había 676,609,955 casos activos; con 6,881,955 muertos registrados en el mundo.

En México, a la última fecha de la colección de datos por el Johns Hopkins se tiene un total de 7,483,444 casos; habiendo 333,188 muertes confirmadas hasta la última fecha.

De acuerdo a la Secretaría de Salud de Michoacán, a la última fecha de actualización que es el 29 de junio de 2023, en toda la entidad federativa se contaron un total hasta la fecha de 221,611 casos confirmados; habiendo un total contado hasta la fecha de 8,83 defunciones.

Así mismo, en la entidad de Lázaro Cárdenas, hubo un total de 22,165 casos confirmados hasta la última fecha de corte, sin defunciones reportadas en la misma fecha.

**2. ¿Cuál fue la primera variante del virus que se propagó a todo el mundo?**

De acuerdo al sitio [nextstrain.org/ncov/global](https://nextstrain.org/ncov/global), la primera variante que se propagó a todo el mundo fue la Wuhan/WH04/2024.

**3. ¿Cuáles son las otras variantes del virus que existen en otras regiones del mundo?**

De acuerdo al National Library of Medicine y la OMS (a fecha de 13 de julio de 2021), existen 10 variantes de preocupación y de interés:

Variantes de preocupación (Linaje de Pangolín): Alpha (B.1.1.7), Beta (B.1.351), Gamma, (P.1), Delta (B.1.617.2).

Variantes de interés (Linaje de Pangolín): Épsilon (B.1.427/B.1.429), Zeta (P.2), Eta (B.1.525), Theta (P.3), Iota (B.1.526), Kappa (B.1.617.1).

#### 4. ¿Cómo buscarías información de la variante del virus en tu país?

Para obtener la información necesaria sobre la variante del virus en mi país, tengo que dirigirme a la página oficial de la Secretaría de Salud, para así obtener la información más veraz sobre el virus en México.

#### 5. Imagina que te encuentras en una situación similar a la de Li Wenliang, médico chino que intentó alertar sobre el brote de coronavirus en su país, pero fue detenido por las autoridades y obligado a retractarse, ¿qué harías en su caso? Selecciona un inciso: a) Lo reportas al centro de investigación o la universidad. b) Lo reportas a la prensa. c) Guardas la información.

Como persona en la misma situación que Li Wenliang, yo reportaría el caso a un centro de investigación o a universidades para que trabajaran de manera pronta en encontrar la cura y/o vacuna para la enfermedad, para así prevenir la mayor cantidad de muertes. Así mismo, como médico, me vería en la obligación ética de reportarlo, debido a que con esto se lograría mitigar el riesgo de propagación de la enfermedad.

## Parte 2: Realiza lo siguiente en un script de R y código:

### Cargar librerías

```
library(stringr)
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 object is masked from 'package:utils':
##
##      findMatches

## 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(ggplot2)
```

### Asigna a un vector los nombres de los diferentes archivos

```
variantes <- c("wuhan.fna", "ALPHA.fasta", "beta.fasta", "GAMMA.fasta",
"Omicron.fasta")
```

### Asigna a un vector los nombres de las variantes

```
nombre_variantes <- c("Wuhan", "Alpha", "Beta", "Gamma", "Omicron")
```

### Funcion para contar las bases de ADN

```
contar_bases <- function(ADN) {
  a_count <- str_count(ADN, pattern = "A")
  t_count <- str_count(ADN, pattern = "T")
  g_count <- str_count(ADN, pattern = "G")
  c_count <- str_count(ADN, pattern = "C")
  n_count <- str_count(ADN, pattern = "N")
  data.frame(Base = c("A", "T", "G", "C", "N"), Count = c(a_count,
t_count, g_count, c_count, n_count))
}
```

## Función para calcular %GC

```
adn_info <- function(ADN) {  
  adn_len <- nchar(ADN)  
  gc_count <- str_count(ADN, pattern = "G|C")  
  gc_percentage <- (gc_count / adn_len) * 100  
  cat("Tamaño de las bases:", adn_len, "\n")  
  cat("Porcentaje de GC:", gc_percentage, "%\n\n")  
}
```

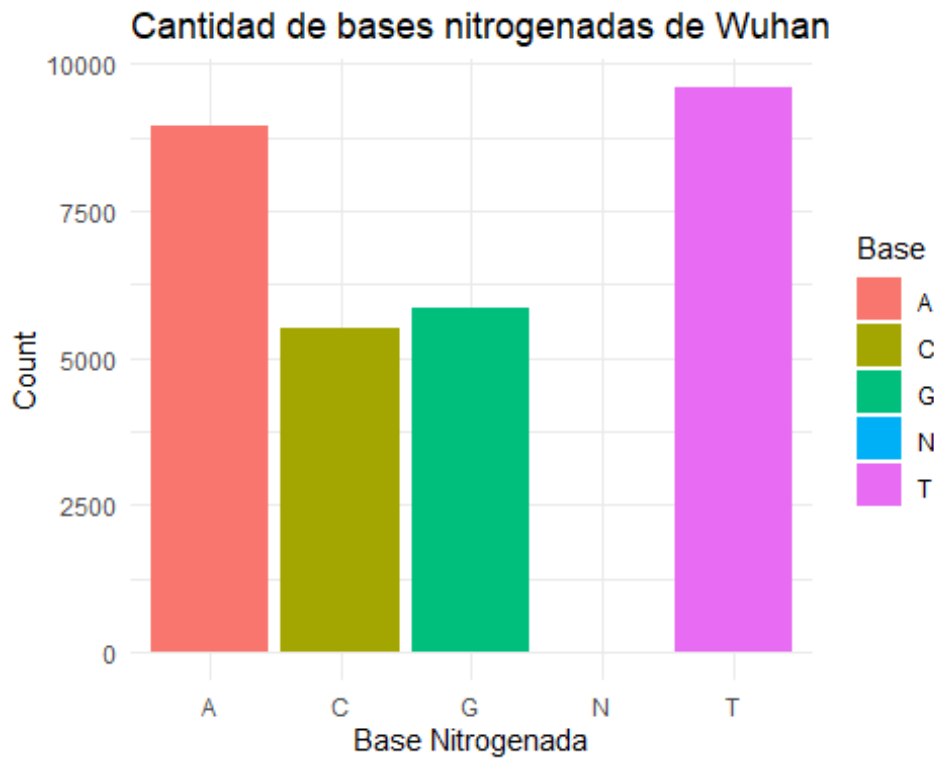
## Funcion para generar secuencia contrasentido

```
generar_contrasentido <- function(ADN_set) {  
  contrasentido <- reverseComplement(ADN_set)  
  return(contrasentido)  
}
```

## Main

```
# Main  
for (i in 1:length(variantes)) {  
  # Imprime el nombre de la variante  
  cat("Nombre de la variante:", nombre_variantes[i], "\n")  
  
  # Lee el archivo de ADN y cuenta las bases  
  ADN_set <- readDNAStringSet(variantes[i])  
  adn <- toString(ADN_set)  
  base_counts <- contar_bases(adn)  
  
  # Gráfico de barras para las bases de ADN  
  ggplot(base_counts, aes(x = Base, y = Count, fill = Base)) +  
    geom_bar(stat = "identity") +  
    labs(title = paste("Cantidad de bases nitrogenadas de",  
      nombre_variantes[i]),  
      x = "Base Nitrogenada", y = "Count") +  
    theme_minimal()  
  
  # Mostrar el gráfico  
  print(ggplot(base_counts, aes(x = Base, y = Count, fill = Base)) +  
    geom_bar(stat = "identity") +  
    labs(title = paste("Cantidad de bases nitrogenadas de",  
      nombre_variantes[i]),  
      x = "Base Nitrogenada", y = "Count") +  
    theme_minimal())  
  
  # Calcula y muestra el %GC  
  adn_info(adn)  
  
  # Genera la secuencia contrasentido y almacénala en una variable  
  contrasentido <- generar_contrasentido(ADN_set)  
}
```

## Nombre de la variante: Wuhan

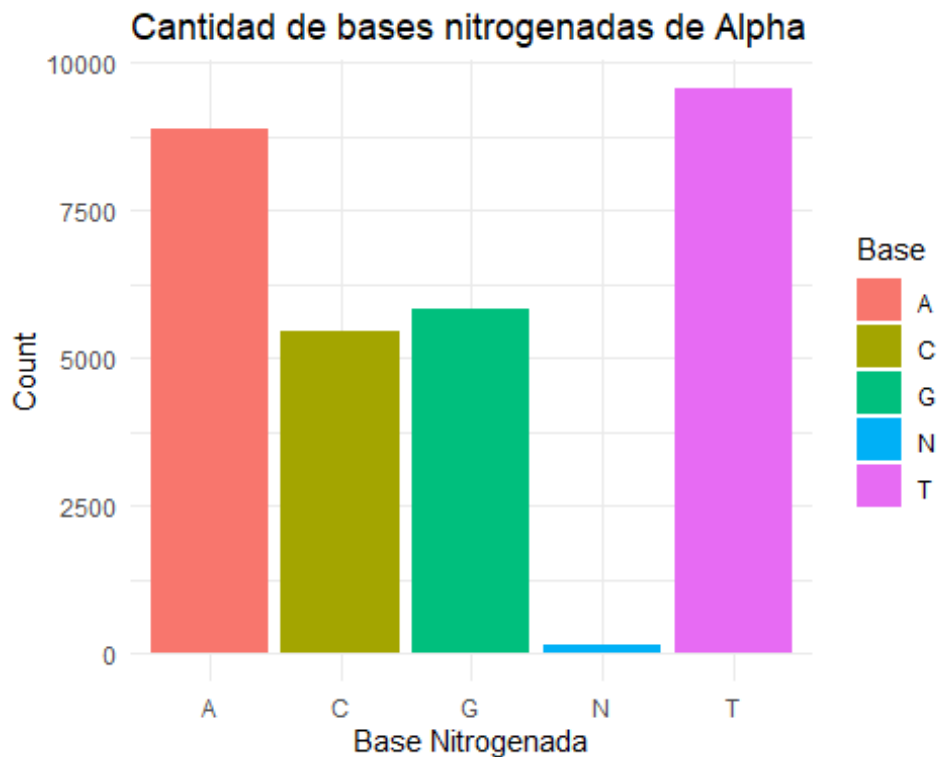


## Tamaño de las bases: 29903

## Porcentaje de GC: 37.97278 %

Esta gráfica muestra la cantidad de bases nitrogenadas en la variante de Covid-19 "Wuhan". Se puede observar que la adenina y timina tienen las cantidades más altas. El porcentaje de %GC es de 37.97278%. Esta información es de suma importancia para comprender la composición genética de la variante del virus.

## Nombre de la variante: Alpha

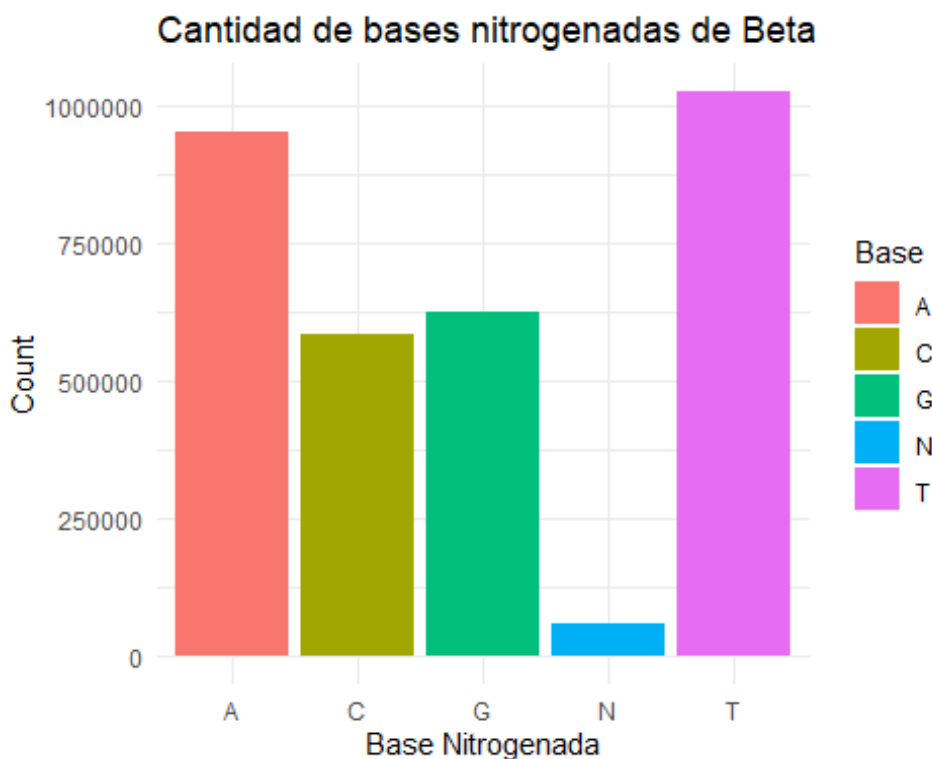


## Tamaño de las bases: 29884

## Porcentaje de GC: 37.83295 %

Esta gráfica muestra la cantidad de bases nitrogenadas en la variante de Covid-19 “Alpha”. Se puede inferir que las bases nitrogenadas con mayor cantidad, son la Adenina y Timina. Así mismo, podemos notar que existe una pequeña cantidad de nucleótidos no identificados, referenciados por la letra N y la barra color Azul. Esto nos ayuda a poder identificar la composición genética de esta variante más, sin embargo, al no tener todos los nucleótidos, es un poco más costoso trabajar sobre esta.

## Nombre de la variante: Beta

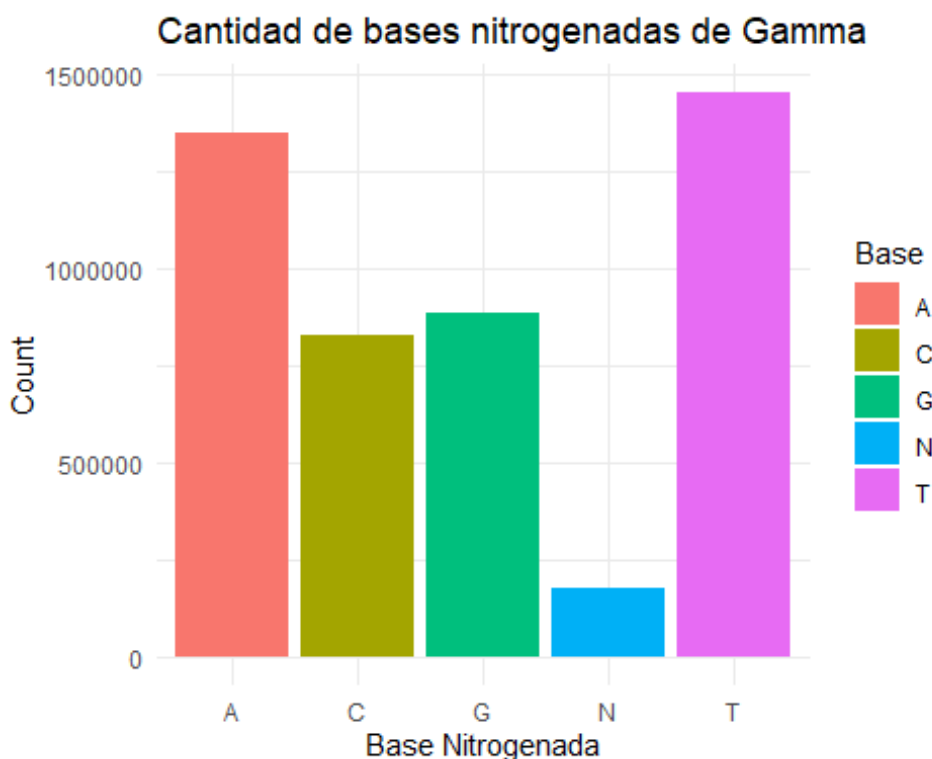


## Tamaño de las bases: 3244622

## Porcentaje de GC: 37.27405 %

Esta gráfica muestra la cantidad de bases nitrogenadas en la variante de Covid-19 “Beta”. Se puede inferir que las bases nitrogenadas con mayor cantidad, son la Adenina y Timina. Así mismo, podemos notar que existe una pequeña cantidad de nucleótidos no identificados, referenciados por la letra N y la barra color Azul. Esto nos ayuda a poder identificar la composición genética de esta variante más, sin embargo, al no tener todos los nucleótidos, es un poco más costoso trabajar sobre esta.

## Nombre de la variante: Gamma



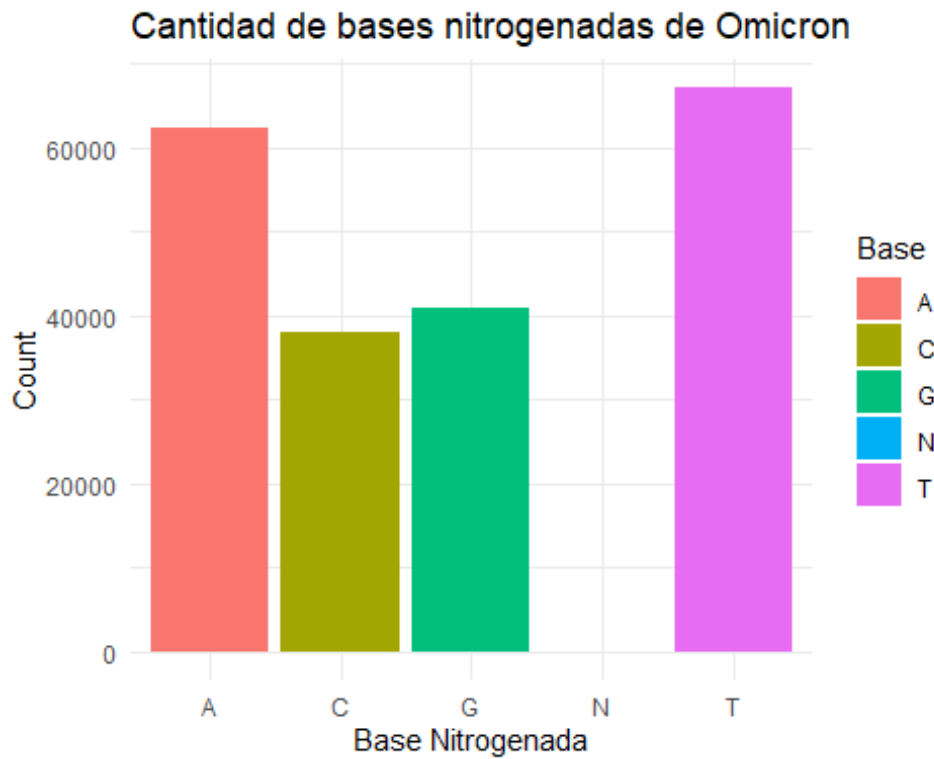
## Tamaño de las bases: 4699287

## Porcentaje de GC: 36.52482 %

Esta gráfica muestra la cantidad de bases nitrogenadas en la variante de Covid-19 "Gamma". Se puede inferir que las bases nitrogenadas con mayor cantidad, son la Adenina y Timina. Así mismo, podemos notar que existe una considerable cantidad de nucleótidos no identificados, referenciados por la letra N y la barra color Azul. Esto nos ayuda a poder identificar la composición genética de esta variante más, sin embargo, al no tener todos los nucleótidos, es un poco más costoso trabajar sobre esta.



## Nombre de la variante: Omicron



## Tamaño de las bases: 208393

## Porcentaje de GC: 37.88371 %

Esta gráfica muestra la cantidad de bases nitrogenadas en la variante de Covid-19 "Omicron". Se puede observar que la adenina y timina tienen las cantidades más altas. Así mismo, podemos observar que no se encuentra ninguna base nitrogenada sin identificar, lo que facilita los esfuerzos para encontrar la composición genética del virus.

## Referencias

COVID-19 Map - Johns Hopkins Coronavirus Resource Center. (s. f.). Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html>

De Comunicación Social, C. G. (s. f.). Informe diario COVID-19 Michoacán. Michoacán. <https://salud.michoacan.gob.mx/informe-diario-covid-19-michoacan-190/>

Variantes del SARS-COV-2, una historia todavía inacabada-PMC. (2021, 13 julio). National Library Of Medicine. Recuperado 24 de abril de 2024, de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8275477/>

Genomic Epidemiology of Novel Coronavirus. (s. f.). Nextstrain.org. <https://web.archive.org/web/20200420123520/https://nextstrain.org/ncov/global>