Evidencia1

Bernardo de la Sierra Rabago y Luis Fernando Perez Robles

29/4/2022

# Parte 2

## Instalación de paquetes y llamado a librerias

# install.packages("seqinr") Ya la tengo instalda por eso se comentó  
library("seqinr")

## Warning: package 'seqinr' was built under R version 4.1.3

## Posicionamiento en el directorio de trabajo

A=getwd()  
A

## [1] "C:/Users/6QV46LA/OneDrive/Documentos"

setwd("/Users/")  
setwd(A)

## 1- Obtén las secuencias de 8 a 10 variantes de SARS-CoV-2 desde el NCBI (Enlaces a un sitio externo.) o el buscador de virus del NCBI (Enlaces a un sitio externo.).Ejemplos:

Mex1 <- read.fasta("SARSCOVMex1.fasta")  
Mex2 <- read.fasta("SARSCOVMex2.fasta")  
Mex3 <- read.fasta("SARSCOVMex3.fasta")  
Usa1 <- read.fasta("SARSCOVUsa1.fasta")  
Usa2 <- read.fasta("SARSCOVUsa2.fasta")  
Usa3 <- read.fasta("SARSCOVUsa3.fasta")  
Can1 <- read.fasta("SARSCOVCan1.fasta")  
Can2 <- read.fasta("SARSCOVCan2.fasta")  
Can3 <- read.fasta("SARSCOVCan3.fasta")

# 2- Obtén mediante código la longitud de las secuencias de cada variante, el contenido de GC, y los primeros nucleótidos de la secuencia complementaria de cada genoma.

size <- function(dna){ # funcion que calcula el tamaño de la secuencia  
 print("Número de bases: ")  
 return (length(dna))  
}  
  
# Este código calcula el porcentaje GC en la secuencia  
cuenta\_GC<- function(dna){ # funcion que calcula el porcentaje de GC y   
 #Contadores   
 g <- 0   
 c <- 0  
 # Bucle de repetición y checa cuantas veces salen las letras  
 for( i in 1:length(dna)){  
 if(dna[i] == "g")  
 g <- g + 1  
 else if(dna[i] == "c")  
 c <- c+1  
 }  
 # Impresión de la información  
 print("Porcentaje de Citosina: ")  
 print((c/length(dna))\*100)  
 print("Porcentaje de Guanina: ")  
 print((g/length(dna))\*100)  
 print("Porcentaje de GC: ")  
 print((c/length(dna))\*100+(g/length(dna))\*100)  
}  
  
complementaria <- function(dna){ # Funcion que crea la hebra complementaria cambiando las letras  
 cadena <- c()# Vector vacío que guarda la nueva hebra  
 # Ciclo de repetición donde se cambian las letras  
 for( i in 1:length(dna)){  
 if(dna[i] == "a")  
 cadena[i] <- "t"  
 else if(dna[i] == "t")  
 cadena[i] <- "a"  
 if(dna[i] == "c")  
 cadena[i] <- "g"  
 else if(dna[i] == "g")  
 cadena[i] <- "c"  
 }  
 print("La secuencia complementaria es: ")  
 return (cadena)   
}  
  
print("Información de Sars covid 2 en México de Oaxaca");

## [1] "Información de Sars covid 2 en México de Oaxaca"

size(Mex1[[1]])

## [1] "Número de bases: "

## [1] 29903

cuenta\_GC(Mex1[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 18.34599  
## [1] "Porcentaje de Guanina: "  
## [1] 19.60673  
## [1] "Porcentaje de GC: "  
## [1] 37.95271

head(complementaria(Mex1[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "t" "a" "a" "t" "t" "t"

print("Información de Sars covid 2 en México de Ciudad de México");

## [1] "Información de Sars covid 2 en México de Ciudad de México"

size(Mex2[[1]])

## [1] "Número de bases: "

## [1] 29903

cuenta\_GC(Mex2[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 17.91459  
## [1] "Porcentaje de Guanina: "  
## [1] 19.17199  
## [1] "Porcentaje de GC: "  
## [1] 37.08658

head(complementaria(Mex2[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "t" "a" "a" "t" "t" "t"

print("Información de Sars covid 2 en México de Baja California");

## [1] "Información de Sars covid 2 en México de Baja California"

size(Mex3[[1]])

## [1] "Número de bases: "

## [1] 29903

cuenta\_GC(Mex3[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 18.34264  
## [1] "Porcentaje de Guanina: "  
## [1] 19.61342  
## [1] "Porcentaje de GC: "  
## [1] 37.95606

head(complementaria(Mex3[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "t" "a" "a" "t" "t" "t"

print("Información de Sars covid 2 en Estados Unidos de Delaware");

## [1] "Información de Sars covid 2 en Estados Unidos de Delaware"

size(Usa1[[1]])

## [1] "Número de bases: "

## [1] 29940

cuenta\_GC(Usa1[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 18.33333  
## [1] "Porcentaje de Guanina: "  
## [1] 19.57582  
## [1] "Porcentaje de GC: "  
## [1] 37.90915

head(complementaria(Usa1[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "t" "a" "a" "t" "t" "t"

print("Información de Sars covid 2 en Estados Unidos de Florida");

## [1] "Información de Sars covid 2 en Estados Unidos de Florida"

size(Usa2[[1]])

## [1] "Número de bases: "

## [1] 29933

cuenta\_GC(Usa2[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 18.08372  
## [1] "Porcentaje de Guanina: "  
## [1] 19.38997  
## [1] "Porcentaje de GC: "  
## [1] 37.47369

head(complementaria(Usa2[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "t" "t" "t" "c" "c" "a"

print("Información de Sars covid 2 en Estados Unidos de Washington");

## [1] "Información de Sars covid 2 en Estados Unidos de Washington"

size(Usa3[[1]])

## [1] "Número de bases: "

## [1] 29921

cuenta\_GC(Usa3[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 18.33495  
## [1] "Porcentaje de Guanina: "  
## [1] 19.6083  
## [1] "Porcentaje de GC: "  
## [1] 37.94325

head(complementaria(Usa3[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "g" "a" "c" "c" "t" "t"

print("Información de Sars covid 2 en Canada de Ontario");

## [1] "Información de Sars covid 2 en Canada de Ontario"

size(Can1[[1]])

## [1] "Número de bases: "

## [1] 29853

cuenta\_GC(Can1[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 17.73691  
## [1] "Porcentaje de Guanina: "  
## [1] 19.02656  
## [1] "Porcentaje de GC: "  
## [1] 36.76347

head(complementaria(Can1[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "g" NA "c" "c" "a" "t"

print("Información de Sars covid 2 en Canada de Alberta");

## [1] "Información de Sars covid 2 en Canada de Alberta"

size(Can2[[1]])

## [1] "Número de bases: "

## [1] 29834

cuenta\_GC(Can2[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 18.16049  
## [1] "Porcentaje de Guanina: "  
## [1] 19.44761  
## [1] "Porcentaje de GC: "  
## [1] 37.6081

head(complementaria(Can2[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "g" "a" "a" "c" "a" "t"

print("Información de Sars covid 2 en Canada de British Columbia");

## [1] "Información de Sars covid 2 en Canada de British Columbia"

size(Can3[[1]])

## [1] "Número de bases: "

## [1] 29782

cuenta\_GC(Can3[[1]])

## [1] "Porcentaje de Citosina: "  
## [1] 18.35001  
## [1] "Porcentaje de Guanina: "  
## [1] 19.63266  
## [1] "Porcentaje de GC: "  
## [1] 37.98267

head(complementaria(Can3[[1]]))

## [1] "La secuencia complementaria es: "

## [1] "t" "c" "t" "a" "g" "a"

# 3- Crea una tabla en donde compares la cantidad las bases de ADN que componen a cada una de las variantes del virus.

print("Tabla de contenido Comparativa de México,Estados Unidos y Canada")

## [1] "Tabla de contenido Comparativa de México,Estados Unidos y Canada"

tabla\_contenido <- data.frame(Adenina = 1:9,Citosina = 1:9, Guanina = 1:9, Timina = 1:9);  
rownames(tabla\_contenido) <- c("Oaxaca de México","Ciudad de México de México","Baja California de México","Delaware de Estados Unidos","Florida de Estados Unidos","Washington de Estados Unidos","Ontario de Canada","Alberta de Canada","British Columbia de Canada")  
tabla\_contenido[1, ] <- count(Mex1[[1]],1)  
tabla\_contenido[2, ] <- count(Mex2[[1]],1)  
tabla\_contenido[3, ] <- count(Mex3[[1]],1)  
tabla\_contenido[4, ] <- count(Usa1[[1]],1)  
tabla\_contenido[5, ] <- count(Usa2[[1]],1)  
tabla\_contenido[6, ] <- count(Usa3[[1]],1)  
tabla\_contenido[7, ] <- count(Can1[[1]],1)  
tabla\_contenido[8, ] <- count(Can2[[1]],1)  
tabla\_contenido[9, ] <- count(Can3[[1]],1)  
tabla\_contenido

## Adenina Citosina Guanina Timina  
## Oaxaca de México 8953 5486 5863 9601  
## Ciudad de México de México 8734 5357 5733 9355  
## Baja California de México 8951 5485 5865 9602  
## Delaware de Estados Unidos 8992 5489 5861 9598  
## Florida de Estados Unidos 8852 5413 5804 9470  
## Washington de Estados Unidos 8969 5486 5867 9599  
## Ontario de Canada 8677 5295 5680 9280  
## Alberta de Canada 8854 5418 5802 9494  
## British Columbia de Canada 8894 5465 5847 9574

# 4- Grafica el contenido de bases de cada genoma usando herramientas básicas de graficación de R.

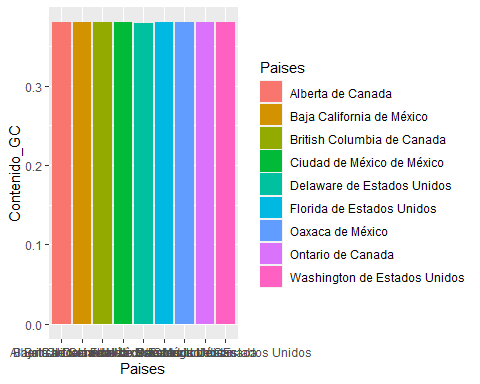
# install.packages("tidyverse") aca se comenta porque ya instale el paquete   
tabla\_ev01<- data.frame(Paises= 1:9,Longitud = 1:9, Contenido\_GC = 1:9);  
rownames(tabla\_ev01) <- c("1","2","3","4","5","6","7","8","9")  
tabla\_ev01[1,1 ] <- "Oaxaca de México"  
tabla\_ev01[2,1] <- "Ciudad de México de México"  
tabla\_ev01[3,1 ] <- "Baja California de México"  
tabla\_ev01[4,1] <- "Delaware de Estados Unidos"  
tabla\_ev01[5,1 ] <- "Florida de Estados Unidos"  
tabla\_ev01[6,1 ] <- "Washington de Estados Unidos"  
tabla\_ev01[7,1 ] <- "Ontario de Canada"  
tabla\_ev01[8,1 ] <- "Alberta de Canada"  
tabla\_ev01[9,1 ] <- "British Columbia de Canada"  
  
tabla\_ev01[1,2 ] <- length(Mex1[[1]])  
tabla\_ev01[2,2] <- length(Mex2[[1]])  
tabla\_ev01[3,2 ] <- length(Mex3[[1]])  
tabla\_ev01[4,2] <- length(Usa1[[1]])  
tabla\_ev01[5,2 ] <- length(Usa2[[1]])  
tabla\_ev01[6,2 ] <- length(Usa3[[1]])  
tabla\_ev01[7,2 ] <- length(Can1[[1]])  
tabla\_ev01[8,2 ] <- length(Can2[[1]])  
tabla\_ev01[9,2 ] <- length(Can3[[1]])  
  
tabla\_ev01[1,3 ] <- GC(Mex1[[1]])  
tabla\_ev01[2,3] <- GC(Mex2[[1]])  
tabla\_ev01[3,3 ] <- GC(Mex3[[1]])  
tabla\_ev01[4,3] <- GC(Usa1[[1]])  
tabla\_ev01[5,3 ] <- GC(Usa2[[1]])  
tabla\_ev01[6,3 ] <- GC(Usa3[[1]])  
tabla\_ev01[7,3 ] <- GC(Can1[[1]])  
tabla\_ev01[8,3 ] <- GC(Can2[[1]])  
tabla\_ev01[9,3 ] <- GC(Can3[[1]])  
  
tabla\_ev01

## Paises Longitud Contenido\_GC  
## 1 Oaxaca de México 29903 0.3795271  
## 2 Ciudad de México de México 29903 0.3800679  
## 3 Baja California de México 29903 0.3795606  
## 4 Delaware de Estados Unidos 29940 0.3790915  
## 5 Florida de Estados Unidos 29933 0.3797353  
## 6 Washington de Estados Unidos 29921 0.3794325  
## 7 Ontario de Canada 29853 0.3793378  
## 8 Alberta de Canada 29834 0.3794643  
## 9 British Columbia de Canada 29782 0.3798522

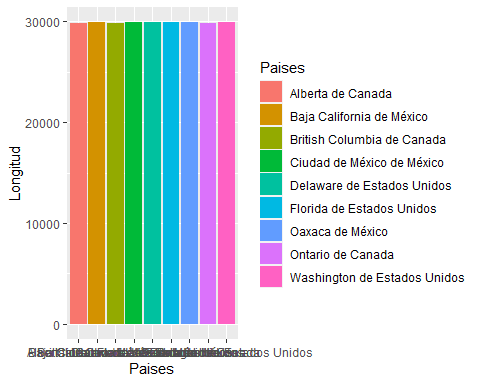
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.1.3

# Graficamos el contenido de Guanina Citosina y la Longitud  
plot\_tabla\_ev01\_GC <-ggplot(tabla\_ev01, aes(x=Paises, y=Contenido\_GC, fill=Paises))+geom\_bar(stat = "identity")  
plot\_tabla\_ev01\_GC



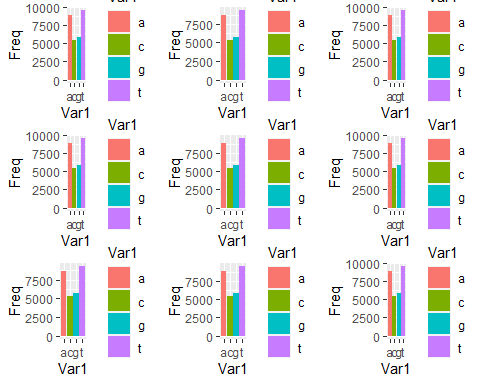
plot\_tabla\_ev01\_length <-ggplot(tabla\_ev01, aes(x=Paises, y=Longitud, fill=Paises))+geom\_bar(stat = "identity")  
plot\_tabla\_ev01\_length



count\_Mex1 <- as.data.frame(count(Mex1[[1]],1))  
count\_Mex2 <- as.data.frame(count(Mex2[[1]],1))  
count\_Mex3 <- as.data.frame(count(Mex3[[1]],1))  
count\_Usa1 <- as.data.frame(count(Usa1[[1]],1))  
count\_Usa2 <- as.data.frame(count(Usa2[[1]],1))  
count\_Usa3 <- as.data.frame(count(Usa3[[1]],1))  
count\_Can1 <- as.data.frame(count(Can1[[1]],1))  
count\_Can2 <- as.data.frame(count(Can2[[1]],1))  
count\_Can3 <- as.data.frame(count(Can3[[1]],1))  
  
plot\_count\_Mex1 <- ggplot(count\_Mex1, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
plot\_count\_Mex2 <- ggplot(count\_Mex2, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
plot\_count\_Mex3 <- ggplot(count\_Mex3, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
plot\_count\_Usa1 <- ggplot(count\_Usa1, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
plot\_count\_Usa2 <- ggplot(count\_Usa2, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
plot\_count\_Usa3 <- ggplot(count\_Usa3, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
plot\_count\_Can1 <- ggplot(count\_Can1, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
plot\_count\_Can2 <- ggplot(count\_Can2, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
plot\_count\_Can3 <- ggplot(count\_Can3, aes(x=Var1,y=Freq,fill=Var1))+geom\_bar(stat = "identity")  
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 4.1.3

grid.arrange(plot\_count\_Mex1,plot\_count\_Mex2,plot\_count\_Mex3,plot\_count\_Usa1,plot\_count\_Usa2,plot\_count\_Usa3,plot\_count\_Can1,plot\_count\_Can2,plot\_count\_Can3,ncol=3)



# 5- Redacta una interpretación desde el punto de vista biológico y con tus propias palabras de las gráficas generadas sobre el contenido de nucleótidos.