library(seqinr)

library(stringr)

library(Biostrings)

leer <- function(archivo) {

secuencia <- readDNAStringSet(archivo)[[1]]

secuencia\_caracter <- as.character(secuencia)

return(secuencia\_caracter)

}

tamaño\_secuencia<-function(txt){

tamaño <- nchar(gsub("[^A-Za-z]", "", txt))

return(tamaño)

}

cant\_nuc <- function(txt, c) {

vec <- gregexpr(c, txt, fixed = TRUE)[[1]]

cant <- length(vec)

cantidad <- cant / (tamaño\_secuencia(txt)) \* 100

return(cantidad)

}

porc\_gc <- function(txt) {

vec\_c <- gregexpr('c', txt, fixed = TRUE)[[1]]

vec\_g <- gregexpr('g', txt, fixed = TRUE)[[1]]

cant\_c <- length(vec\_c)

cant\_g <- length(vec\_g)

cant <- cant\_g + cant\_c

cantidad <- cant / (tamaño\_secuencia(txt)) \* 100

return(cantidad)

}

hebras <- function (txt){

complementaria <- ""

for (base in strsplit(txt, NULL)[[1]]) {

if (base == "A") {

complementaria <- paste0(complementaria, "T")

} else if (base == "T") {

complementaria <- paste0(complementaria, "A")

} else if (base == "C") {

complementaria <- paste0(complementaria, "G")

} else if (base == "G") {

complementaria <- paste0(complementaria, "C")

} else {

complementaria <- paste0(complementaria, base)

}

}

return (complementaria)

}

dengue <- read.fasta(file = "Dengue.fasta")

mec <- read.fasta(file = "M\_corona.fasta")

pneumonia <- read.fasta(file = "Pneu.fasta")

sars <- read.fasta(file = "SARS.fasta")

zika <- read.fasta (file = "zika.fasta")

ha\_h1 <- read.fasta (file = "ha\_h1n1.fasta")

na\_h1 <- read.fasta (file = "na\_h1n1.fasta")

ns1\_h1 <- read.fasta (file = "ns1\_h1n1.fasta")

m1\_h1 <- read.fasta (file = "m1\_h1n1.fasta")

pa\_h1 <- read.fasta (file = "pa\_h1n1.fasta")

np\_h1 <- read.fasta (file = "np\_h1n1.fasta")

pb1\_h1 <- read.fasta (file = "pb1\_h1n1.fasta")

pb2\_h1 <- read.fasta (file = "pb2\_h1n1.fasta")

virus <- list (dengue, mec, pneumonia, sars, zika, ha\_h1, na\_h1,

ns1\_h1, m1\_h1, pa\_h1, np\_h1, pb1\_h1, pb2\_h1)

archivos <- list ("Dengue.fasta", "M\_corona.fasta", "Pneu.fasta",

"SARS.fasta", "zika.fasta", "ha\_h1n1.fasta", "na\_h1n1.fasta",

"ns1\_h1n1.fasta", "m1\_h1n1.fasta", "pa\_h1n1.fasta",

"np\_h1n1.fasta", "pb1\_h1n1.fasta", "pb2\_h1n1.fasta")

nombres <- c("Dengue", "Middle East respiratory syndrome coronavirus",

"Pneumonia", "SARS Coronavirus", "Zika virus", "HA\_H1N1", "NA\_H1N1",

"NSI\_H1N1", "MI\_H1N1", "PA\_H1N1", "NP\_H1N1", "PB1\_H1N1", "PB2\_H1N1")

tamaños <- matrix(nrow = 13, ncol = 1)

i <- 1

for (secuencia in virus) {

tamaños[i,1] <- tamaño\_secuencia(secuencia)

i <- i + 1

}

cant\_nuc\_a <- matrix(nrow = 13, ncol = 1)

i <- 1

for (secuencia in virus) {

cant\_nuc\_a[i, 1] <- cant\_nuc(secuencia, 'a')

i <- i + 1

}

cant\_nuc\_c <- matrix(nrow = 13, ncol = 1)

i <- 1

for (secuencia in virus) {

cant\_nuc\_c[i, 1] <- cant\_nuc(secuencia, 'c')

i <- i + 1

}

cant\_nuc\_g <- matrix(nrow = 13, ncol = 1)

i <- 1

for (secuencia in virus) {

cant\_nuc\_g[i, 1] <- cant\_nuc(secuencia, 'g')

i <- i + 1

}

cant\_nuc\_t <- matrix(nrow = 13, ncol = 1)

i <- 1

for (secuencia in virus) {

cant\_nuc\_t[i, 1] <- cant\_nuc(secuencia, 't')

i <- i + 1

}

p\_gc <- matrix(nrow = 13, ncol = 1)

i <- 1

for (secuencia in virus) {

p\_gc[i, 1] <- porc\_gc(secuencia)

i <- i + 1

}

leer\_ar <- matrix(nrow = 13, ncol = 1)

compl <- matrix(nrow = 13, ncol = 1)

i <- 1

for (secuencia in archivos){

leer\_ar [i,1] <- substr(leer(secuencia),1,30)

i <- i + 1

}

i <- 1

for (secuencia in leer\_ar) {

compl[i, 1] <- hebras(secuencia)

i <- i + 1

}

x <- 1:4

dengue\_pl <- c(cant\_nuc\_a [1],cant\_nuc\_c [1],cant\_nuc\_g [1],cant\_nuc\_t [1])

mec\_pl <- c(cant\_nuc\_a [2],cant\_nuc\_c [2],cant\_nuc\_g [2],cant\_nuc\_t [2])

pneumonia\_pl <- c(cant\_nuc\_a [3],cant\_nuc\_c [3],cant\_nuc\_g [3],cant\_nuc\_t [3])

sars\_pl <- c(cant\_nuc\_a [4],cant\_nuc\_c [4],cant\_nuc\_g [4],cant\_nuc\_t [4])

zika\_pl <- c(cant\_nuc\_a [5],cant\_nuc\_c [5],cant\_nuc\_g [5],cant\_nuc\_t [5])

ha\_h1\_pl <- c(cant\_nuc\_a [6],cant\_nuc\_c [6],cant\_nuc\_g [6],cant\_nuc\_t [6])

na\_h1\_pl <- c(cant\_nuc\_a [7],cant\_nuc\_c [7],cant\_nuc\_g [7],cant\_nuc\_t [7])

ns1\_h1\_pl <- c(cant\_nuc\_a [8],cant\_nuc\_c [8],cant\_nuc\_g [8],cant\_nuc\_t [8])

m1\_h1\_pl <- c(cant\_nuc\_a [9],cant\_nuc\_c [9],cant\_nuc\_g [9],cant\_nuc\_t [9])

pa\_h1\_pl <- c(cant\_nuc\_a [10],cant\_nuc\_c [10],cant\_nuc\_g [10],cant\_nuc\_t [10])

np\_h1\_pl <- c(cant\_nuc\_a [11],cant\_nuc\_c [11],cant\_nuc\_g [11],cant\_nuc\_t [11])

pb1\_h1\_pl <- c(cant\_nuc\_a [12],cant\_nuc\_c [12],cant\_nuc\_g [12],cant\_nuc\_t [12])

pb2\_h1\_pl <- c(cant\_nuc\_a [13],cant\_nuc\_c [13],cant\_nuc\_g [13],cant\_nuc\_t [13])

plot ("Nucleótido","Cantidad de nulceótido", xlim = c(1,4) , ylim = c(0,40) ,

type = "l", main = "composiciones de nucleótidos")

lines (x, dengue\_pl, col = "blue")

lines (x, mec\_pl, col = "purple")

lines (x, pneumonia\_pl, col = "red")

lines (x, sars\_pl, col = "yellow")

lines (x, zika\_pl, col = "green")

lines (x, ha\_h1\_pl, col = "brown")

lines (x, na\_h1\_pl, col = "gray")

legend("bottomleft", legend = c("Dengue", "Middle East respiratory syndrome coronavirus",

"Pneumonia", "SARS Coronavirus", "Zika virus", "HA\_H1N1",

"NA\_H1N1"), col = c("blue", "purple", "red", "yellow",

"green", "brown", "gray"), lwd = 2, bty = "n")

plot ("Nucleótido","Cantidad de nulceótido", xlim = c(1,4) , ylim = c(0,40) ,

type = "l", main = "composiciones de nucleótidos")

lines (x, ns1\_h1\_pl, col = "blue")

lines (x, m1\_h1\_pl, col = "purple")

lines (x, pa\_h1\_pl, col = "red")

lines (x, np\_h1\_pl, col = "yellow")

lines (x, pb1\_h1\_pl, col = "green")

lines (x, pb2\_h1\_pl, col = "brown")

legend("bottomleft", legend = c("NS1\_H1N1", "M1\_H1N1","PA\_H1N1", "NP\_H1N1", "PB1\_H1N1", "PB2\_H1N1",

"NA\_H1N1"), col = c("blue", "purple", "red", "yellow",

"green", "brown"), lwd = 2, bty = "n")

datos\_tama <- data.frame("Nombre" = nombres,"Tamaño de las secuencias." = tamaños, "Porcentaje de la adenina" = cant\_nuc\_a, "Porcentaje de la timina" = cant\_nuc\_t, "Porcentaje de la guanina" = cant\_nuc\_g, "Porcentaje de la citocina" = cant\_nuc\_c, "Porcentaje de GC" = p\_gc, "Hebra directa" = leer\_ar, "Hebra complementaria" = compl)

print (datos\_tama)