

doc proyecto final

Jaime Pérez Y Sofia Peña

2025-06-12

Muestra del código

Este es el código que vamos a utilizar, y solo vamos a utilizar una pequeña muestra de las secuencias de nuestras variables, ya que nuestro poder de procesamiento no es el más óptimo

```
library(seqinr)
library(dplyr)
```

```
##
## Adjuntando el paquete: 'dplyr'

## The following object is masked from 'package:seqinr':
##
##     count

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)

# Función de alineación tipo Needleman-Wunsch simplificada
alineaa = function(gen, gen2){
  m = matrix(data=0, nrow=length(gen2)+1, ncol=length(gen)+1)
  m[1, ] = seq(0, -(length(gen))*2, -2)
  m[, 1] = seq(0, -(length(gen2))*2, -2)

  for (fila in seq(2, nrow(m))) {
    for (col in seq(2, ncol(m))) {
      if (gen[col-1] == gen2[fila-1])
        diag = m[fila-1, col-1] + 1
      else
        diag = m[fila-1, col-1] - 1
      up = m[fila-1, col] - 2
      left = m[fila, col-1] - 2
      m[fila, col] = max(diag, up, left)
    }
  }
}
```

```

    }
}

fila = length(gen2)+1
col = length(gen)+1
solA = solB = c()
while(fila>1 || col>1){
  if (col>1 && fila>1 && gen[col-1]==gen2[fila-1]){
    solA = c(gen[col-1], solA)
    solB = c(gen2[fila-1], solB)
    col = col - 1
    fila = fila - 1
  }else{
    if (col>1 && fila>1 && m[fila-1, col-1]>m[fila, col-1] && m[fila-1, col-1]>m[fila-1, col]){
      solA = c(gen[col-1], solA)
      solB = c(gen2[fila-1], solB)
      col = col - 1
      fila = fila - 1
    }else if (fila==1 || (col>1 && m[fila, col-1] > m[fila-1, col])){
      solA = c(gen[col-1], solA)
      solB = c("_", solB)
      col = col - 1
    }else if (col==1 || (fila>1 && m[fila, col-1] <= m[fila-1, col])){
      solA = c("_", solA)
      solB = c(gen2[fila-1], solB)
      fila = fila - 1
    }
  }
}

sol = matrix(data=c(solA,solB),nrow = 2, ncol = length(solA), byrow = TRUE)
return (sol)
}

# Diccionario de codones
trad = c(UUU="F", UUC="F", UUA="L", UUG="L",
        UCU="S", UCC="S", UCA="S", UCG="S",
        UAU="Y", UAC="Y", UAA="STOP", UAG="STOP",
        UGU="C", UGC="C", UGA="STOP", UGG="W",
        CUU="L", CUC="L", CUA="L", CUG="L",
        CCU="P", CCC="P", CCA="P", CCG="P",
        CAU="H", CAC="H", CAA="Q", CAG="Q",
        CGU="R", CGC="R", CGA="R", CGG="R",
        AUU="I", AUC="I", AUA="I", AUG="M",
        ACU="T", ACC="T", ACA="T", ACG="T",
        AAU="N", AAC="N", AAA="K", AAG="K",
        AGU="S", AGC="S", AGA="R", AGG="R",
        GUU="V", GUC="V", GUA="V", GUG="V",
        GCU="A", GCC="A", GCA="A", GCG="A",
        GAU="D", GAC="D", GAA="E", GAG="E",
        GGU="G", GGC="G", GGA="G", GGG="G")

# DataFrame donde se guardan las mutaciones
datos = data.frame(

```



```

}else{
  if (secuencias_con_indels>10) break
  secuencias_con_indels = secuencias_con_indels + 1
  cat("Mutación de inserción o delección \n")
  cat("Wuhan:", length(gen), "vs:", length(gen2),"\n")
  res = alinea(gen, gen2)
  diff = which(res[1,]!=res[2,])
  if (length(diff)>0) cat("Mutaciones en las posiciones:", diff, "\n")

  prevCod = ""
  align_gen = res[1, ]
  align_gen2 = res[2, ]
  for (k in seq(1, length(align_gen) - 2, by = 3)) {
    cod1 = paste(align_gen[k:(k + 2)], collapse = "")
    cod2 = paste(align_gen2[k:(k + 2)], collapse = "")
    if (grepl("_", cod1) || grepl("_", cod2)) next
    if (!is.na(trad[cod1]) && !is.na(trad[cod2]) && trad[cod1] != trad[cod2]) {
      cambio = paste(trad[cod1], k %/% 3 + 1, trad[cod2])
      if (cambio != prevCod) {
        cat(cod1, "to", cod2, cambio, k %/% 3 + 1, gene, "\n")
        datos[nMut, ] = list(paste(cod1, "to", cod2),
                             paste(cod1, "to", cod2),
                             cambio, k %/% 3 + 1, gene)

        nMut = nMut + 1
        prevCod = cambio
      }
    }
  }
}
}
}
}
}
}
}
}

```

```

## Gen 3 S
## Total de nucleótidos (Wuhan): 3822
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3819
## Mutaciones en las posiciones: 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S

```

```

## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S

```

```

## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3819

```

```

## Mutaciones en las posiciones: 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S

```

```

## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S

```



```

## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S
## Mutación de inserción o delección
## Wuhan: 3822 vs: 3813
## Mutaciones en las posiciones: 200 202 204 205 207 208 284 423 424 426 428 429 430 431 432 433 632 633
## ACU to AUU T 95 I 95 S
## UUU to GUG F 212 V 212 S
## AGU to CGU S 213 R 213 S
## GGU to GAU G 341 D 341 S
## AUC to ACC I 375 T 375 S
## GAU to UAU D 420 Y 420 S
## UCU to GCU S 443 A 443 S
## UGG to UAG W 448 STOP 448 S
## CAC to CAA H 480 Q 480 S
## UGA to UGC STOP 486 C 486 S
## UGG to UAG W 498 STOP 498 S
## UAA to UUA STOP 503 L 503 S
## UUA to UCA L 507 S 507 S
## AAC to AAA N 549 K 549 S
## ACA to AUA T 657 I 657 S
## UUC to GUC F 682 V 682 S
## CCG to ACG P 767 T 767 S
## AGA to AUA R 798 I 798 S
## AAA to UAA K 957 STOP 957 S
## UUU to AUU F 972 I 972 S
## CCU to CUU P 983 L 983 S
## CUC to UUC L 1149 F 1149 S

```

```
cat("Resumen\n")
```

```
## Resumen
```

```
cat("Total de secuencias procesadas:", total_secuencias, "\n")
```

```
## Total de secuencias procesadas: 12
```

```
cat("Secuencias con inserciones/delecciones:", secuencias_con_indels, "\n")
```

```
## Secuencias con inserciones/delecciones: 11
```

```
cat("Porcentaje con INDELs:", round(secuencias_con_indels / total_secuencias * 100, 2), "%\n")
```

```
## Porcentaje con INDELs: 91.67 %
```

```
str(datos)
```

```
## 'data.frame': 240 obs. of 5 variables:
## $ mutacion : chr "ACU to AUU" "UUU to GUG" "AGU to CGU" "GGU to GAU" ...
## $ cambioCodon: chr "ACU to AUU" "UUU to GUG" "AGU to CGU" "GGU to GAU" ...
## $ cambioAmino: chr "T 95 I" "F 212 V" "S 213 R" "G 341 D" ...
## $ pos : num 95 212 213 341 375 420 443 448 480 486 ...
## $ gen : chr "S" "S" "S" "S" ...
```

```
# Frecuencias de cambio de aminoácido
```

```
dfgraph = filter(
  summarise(
    select(
      group_by(datos, cambioAmino),
      mutacion:gen
    ),
    mutacion = first(mutacion),
    cambioCodon = first(cambioCodon),
    pos = first(pos),
    gen = first(gen),
    cuenta = n()
  ),
  cuenta > as.integer(vs * 0.1)
)

cat ("Solo frecuencias superiores al 10%:", as.integer(vs * 0.1), "\n")
```

```
## Solo frecuencias superiores al 10%: 50
```

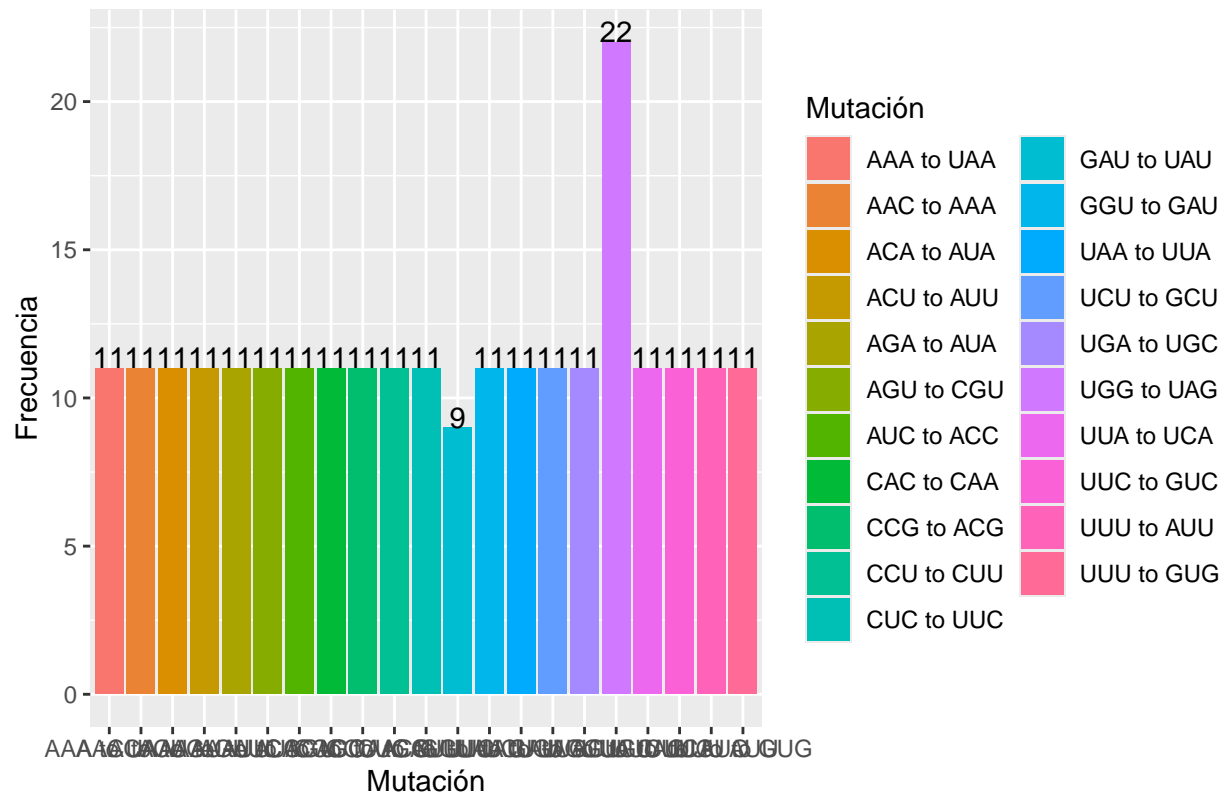
```
print(dfgraph)
```

```
## # A tibble: 0 x 6
## # i 6 variables: cambioAmino <chr>, mutacion <chr>, cambioCodon <chr>,
## # pos <dbl>, gen <chr>, cuenta <int>
```

```
# Gráfica de mutaciones (nucleótidos)
```

```
p = ggplot(datos)
p = p + aes(x=mutacion, fill=mutacion, label=after_stat(count))
p = p + ggtitle(paste("Cambios de nucleótidos Wuhan vs", vs, "secuencias"))
p = p + labs(x="Mutación", y="Frecuencia", fill="Mutación")
p = p + geom_bar(stat="count")
p = p + geom_text(stat="count", vjust=0)
print(p)
```

Cambios de nucleótidos Wuhan vs 500 secuencias



```
# Gráfica de cambios de aminoácidos con validación
if (nrow(dfgraph) > 0) {
  p3 = ggplot(dfgraph)
  p3 = p3 + aes(x=cambioAmino, y=cuenta, fill=cambioAmino, label=cuenta)
  p3 = p3 + ggtitle(paste("Cambio de aminoácidos Wuhan vs", vs, "secuencias"))
  p3 = p3 + labs(x="Amino", y="Frecuencia", fill="Frecuencia")
  p3 = p3 + geom_bar(stat = "identity")
  p3 = p3 + geom_text(stat = "identity", vjust=1.5)
  p3 = p3 + facet_grid(~gen, scales="free", space="free_x")
  print(p3)
} else {
  cat("No hay cambios de aminoácidos con frecuencia mayor al 10% para graficar.\n")
}
```

No hay cambios de aminoácidos con frecuencia mayor al 10% para graficar.