

Actividad 3

Análisis de biología computacional BT1013.525

Bryan Manuel De la O Perea A01246337
Andrés Sarellano Acevedo A01245418
Maximiliano Villegas García A01635825
Víctor Manuel Puga Ruiz A01568636

Dependencies

```
# navigate to your working directory (varies per user)
setwd("/cloud/project/Activities")

install.packages("seqinr")

## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/4.0'
## (as 'lib' is unspecified)

library(seqinr)
```

Genome Files

```
# must upload/move the sequences to the working directory

zika <- read.fasta("Sequences/Zika.fasta")[[1]]
sars.covid <- read.fasta("Sequences/SarsCovid.fasta")[[1]]
wu.han <- read.fasta("Sequences/Wu-Han-1.fasta")[[1]]
mers <- read.fasta("Sequences/MersCovid.fasta")[[1]]
dengue <- read.fasta("Sequences/Dengue.fasta")[[1]]
h1n1 <- read.fasta("Sequences/H1N1.fasta")[[1]]
```

Functions

```
# 1
seq.length <- function(dna.seq) {
  getLength(dna.seq)
}

# 2
seq.composition <- function(dna.seq) {
  count(dna.seq, 1)
}

# 3
seq.gc <- function(dna.seq) {
  GC(dna.seq)
}

#4
seq.translate <- function(dna.seq) {
  comp(dna.seq)
}

# 5
seq.composition.graph <- function(dna.seq) {
  cols = c(
    "a" = "#106BFF",
    "t" = "#FECF0F",
    "g" = "#30D33B",
    "c" = "#FC2B2D"
  )
  pie(seq.composition(dna.seq), col = cols)
}

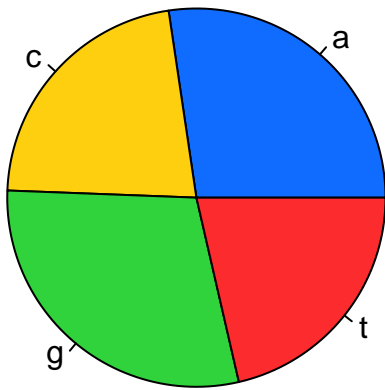
# Stats
stats <- function(dna.seq) {
  annotations <- getAnnot(dna.seq)
  l <- seq.length(dna.seq)
  composition <- seq.composition(dna.seq)
  gc_amount <- seq.gc(dna.seq)
  translation <- seq.translate(dna.seq)

  cat(annotations, "\n\n")
  cat("Length:", l, "bases", "\n\n")
  cat("Composition:")
  print(composition)
  cat("\nCG:", gc_amount, "\n\n")
  cat("Original:  ", dna.seq[1:60], " ... (first 60)\n", sep = "")
  cat("Complement: ", translation[1:60], " ... (first 60)\n\n", sep = "")
  cat("Graph:")
  seq.composition.graph(dna.seq)
}
```

Zika

```
stats(zika)
```

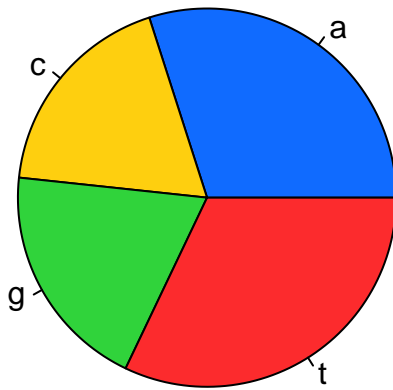
```
## >NC_035889.1 Zika virus isolate ZIKV/H. sapiens/Brazil/Natal/2015, complete genome
##
## Length: 10808 bases
##
## Composition:
##      a      c      g      t
## 2956 2383 3157 2312
##
## CG: 0.5125833
##
## Original:  agttgttgatctgtgtgaatcagactgcgacagttcgagtttgaagcgaaagctagcaac ... (first 60)
## Complement: tcaacaactagacacacttagtctgacgtgtcaagctcaaacttcgctttcgatcgttg ... (first 60)
##
## Graph:
```



SARS Coronavirus

```
stats(sars.covid)
```

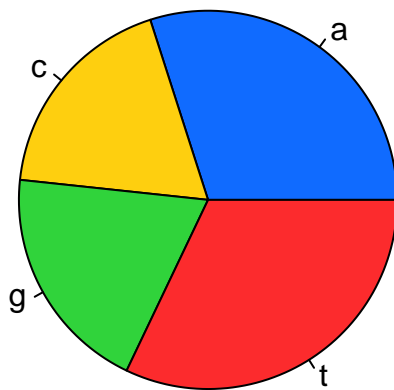
```
## >NC_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome
##
## Length: 29903 bases
##
## Composition:
##   a   c   g   t
## 8954 5492 5863 9594
##
## CG: 0.3797278
##
## Original:  attaaagggtttataccttcccaggtaacaaaccaaccaactttcgatctctttagatct ... (first 60)
## Complement: taatttccaaatatggaagggtccattgtttggttggttgaaagctagagaacatctaga ... (first 60)
##
## Graph:
```



Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1

```
# MN908947.1 was first thought to be pneumonia,  
# but in version MN908947.3 it is now known to be a coronavirus  
stats(wu.han)
```

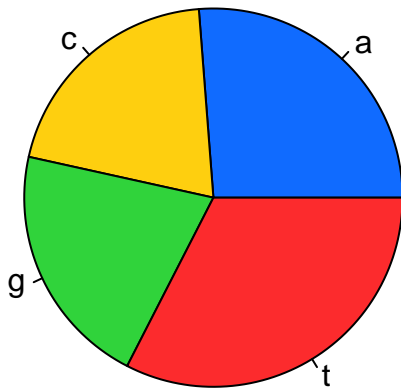
```
## >MN908947.3 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome  
##  
## Length: 29903 bases  
##  
## Composition:  
##   a   c   g   t  
## 8954 5492 5863 9594  
##  
## CG: 0.3797278  
##  
## Original:   attaaaggttttataccttcccaggtaacaaaccaaccaactttcgatctctttagatct ... (first 60)  
## Complement: taatttcctaaatatggaagggtccattgtttggttggttgaaagctagagaacatctaga ... (first 60)  
##  
## Graph:
```



MERS Coronavirus

```
stats(mers)
```

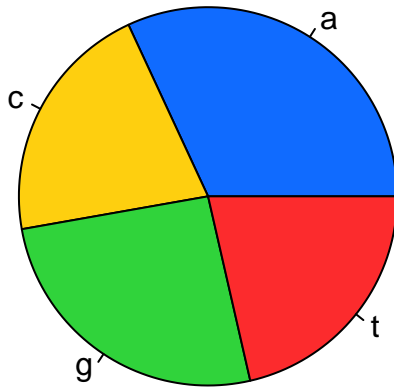
```
## >NC_019843.3 Middle East respiratory syndrome-related coronavirus isolate HCoV-EMC/2012, complete genome
##
## Length: 30119 bases
##
## Composition:
##      a      c      g      t
## 7900 6116 6304 9799
##
## CG: 0.4123643
##
## Original:  gatttaagtgaatagcttggctatctcacttcccctcgttctcttgcagaactttgattt ... (first 60)
## Complement:  ctaaattcacttatcgaaccgatagagtgaaggggagcaagagaacgtcttgaaactaaa ... (first 60)
##
## Graph:
```



Dengue virus 1

```
stats(dengue)
```

```
## >NC_001477.1 Dengue virus 1, complete genome
##
## Length: 10735 bases
##
## Composition:
##      a      c      g      t
## 3426 2240 2770 2299
##
## CG: 0.4666977
##
## Original:  agttgttagtctacgtggaccgacaagaacagtttcgaatcggaagcttgcttaacgtag ... (first 60)
## Complement: tcaacaatcagatgcacctggctgttcttgtcaaagcttagccttcgaacgaattgcatc ... (first 60)
##
## Graph:
```



Influenza H1N1

```
stats(h1n1)
```

```
## >NC_026434.1 Influenza A virus (A/California/07/2009(H1N1)) segment 6 neuraminidase (NA) gene, compl
##
## Length: 1410 bases
##
## Composition:
##   a   c   g   t
## 449 263 331 367
##
## CG: 0.4212766
##
## Original:  atgaatccaaaccaaagataataaccattgggttcggtctgtatgacaattggaatggct ... (first 60)
## Complement: tacttaggttttggttttctattattggtaaccaagccagacatactgttaaccttaccga ... (first 60)
##
## Graph:
```

