Actividad 4

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

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

install <- function(lib) {
  if (!requireNamespace(lib, quietly = TRUE)) {
    install.packages(lib)
  }
}

install("seqinr")
install("ape")
install("phangorn")
install("phytools")
install("geiger")

library(seqinr)
library(ape)
```

```
##
## Attaching package: 'ape'
```

```
## The following objects are masked from 'package:seqinr':
##
##     as.alignment, consensus
```

```r
library(phangorn)
library(phytools)
```

```
## Loading required package: maps
```

```r
library(geiger)
```

Genome Files

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

reference <- read.fasta("SARS-CoV-2 Sequences (Sample)/SARS-CoV-2 (Reference).fasta")[[1]]
```

```r
B.1.1.7   <- read.fasta("SARS-CoV-2 Sequences (Sample)/SARS-CoV-2 (B.1.1.7) Sample.fasta")[[1]]
B.1.351   <- read.fasta("SARS-CoV-2 Sequences (Sample)/SARS-CoV-2 (B.1.351) Sample.fasta")[[1]]
P.1       <- read.fasta("SARS-CoV-2 Sequences (Sample)/SARS-CoV-2 (P.1) Sample.fasta")[[1]]
B.1.427   <- read.fasta("SARS-CoV-2 Sequences (Sample)/SARS-CoV-2 (B.1.427) Sample.fasta")[[1]]
B.1.429   <- read.fasta("SARS-CoV-2 Sequences (Sample)/SARS-CoV-2 (B.1.429) Sample.fasta")[[1]]

all.variants = list(
  "Reference" = reference,
  "B.1.1.7"   = B.1.1.7,
  "B.1.351"   = B.1.351,
  "P.1"       = P.1,
  "B.1.427"   = B.1.427,
  "B.1.429"   = B.1.429
)
```
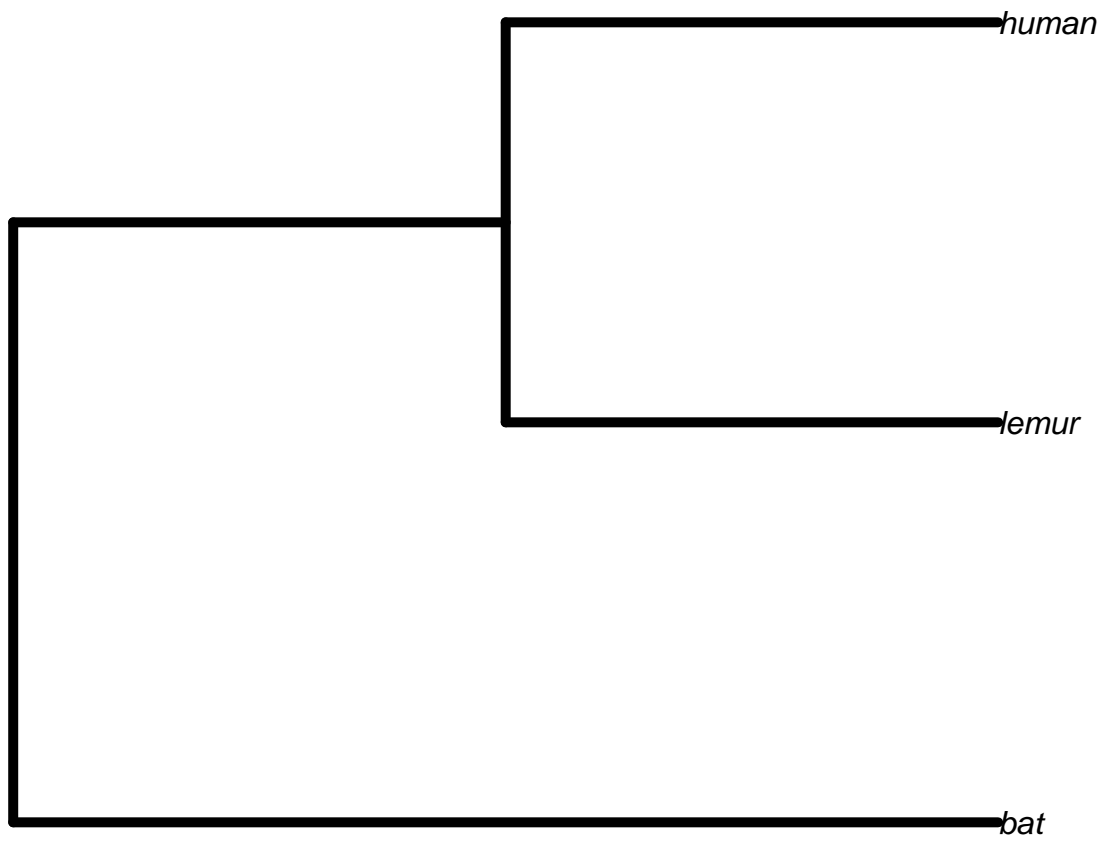
Intro

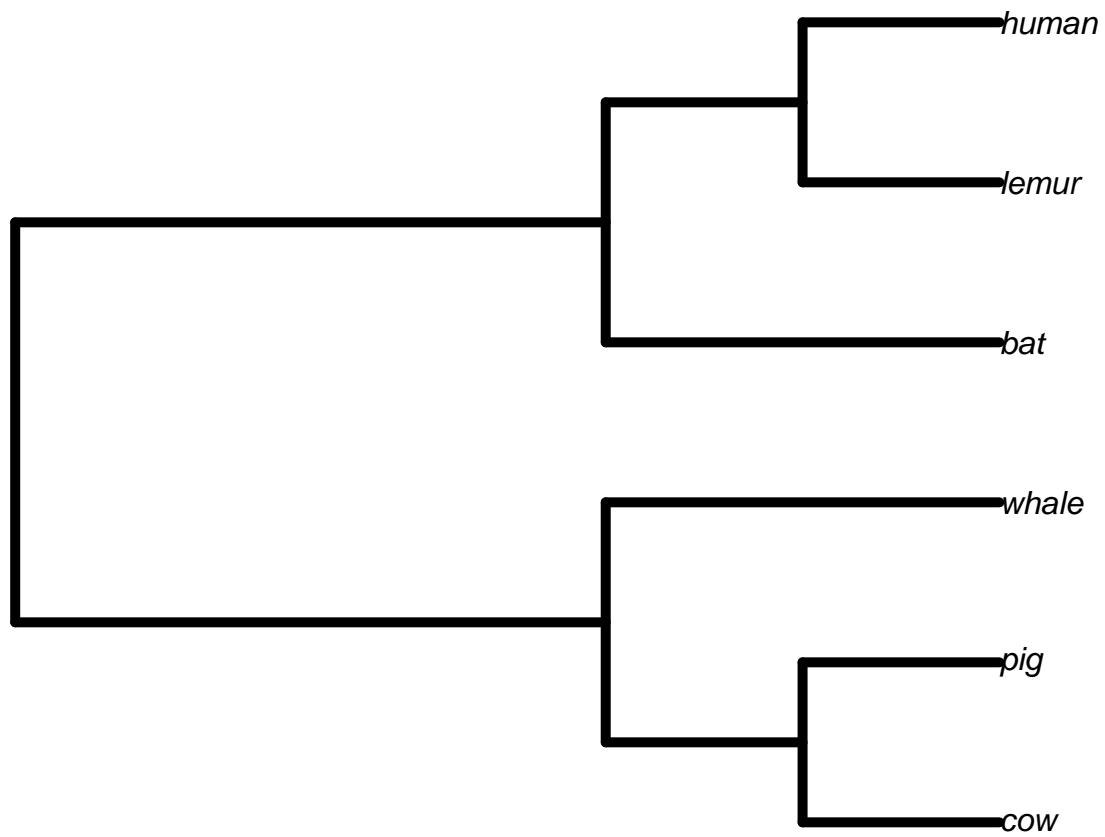## ape: Analysis of Phylogenetics and Evolution

```r
# phylo object
vert.tree <- read.tree(text = "(lemur, human);")
plot(vert.tree, no.margin = TRUE, edge.width = 5)
```
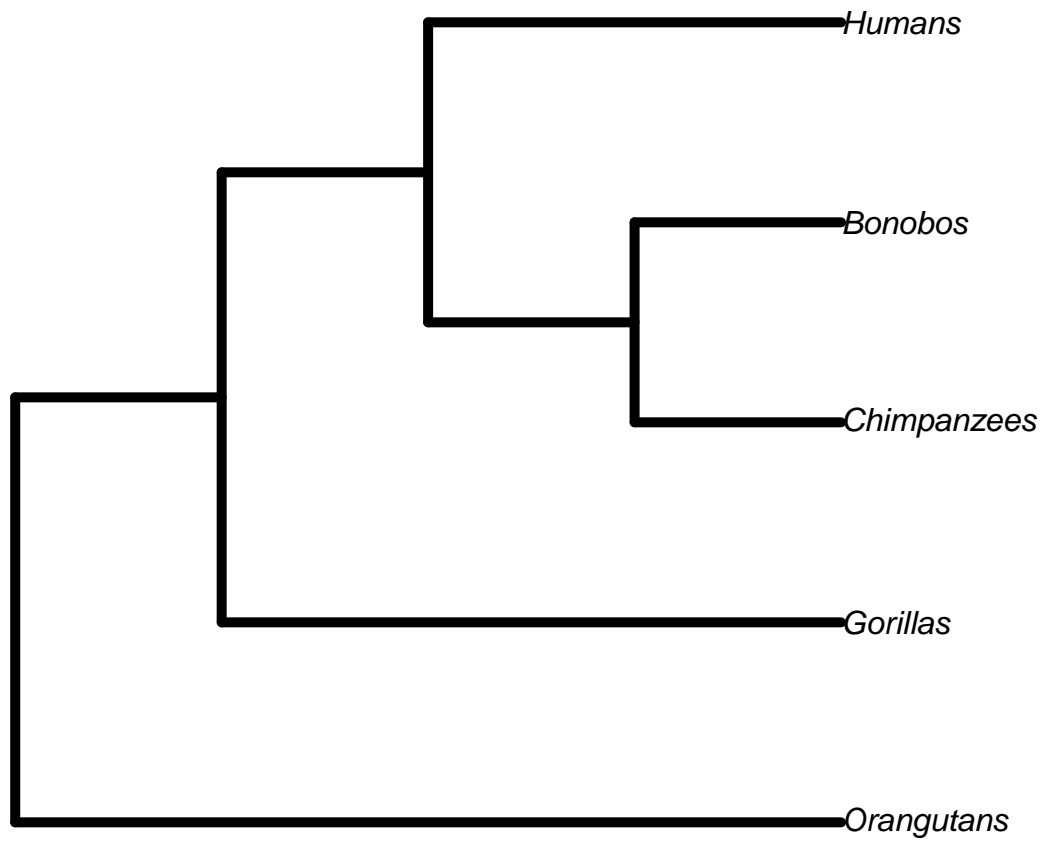


```r
vert.tree <- read.tree(text = "(bat, (lemur, human));")
plot(vert.tree, no.margin = TRUE, edge.width = 5)
```

```
vert.tree <- read.tree(text = "(((cow, pig), whale), (bat, (lemur, human)));")
plot(vert.tree, no.margin = TRUE, edge.width = 5)
```
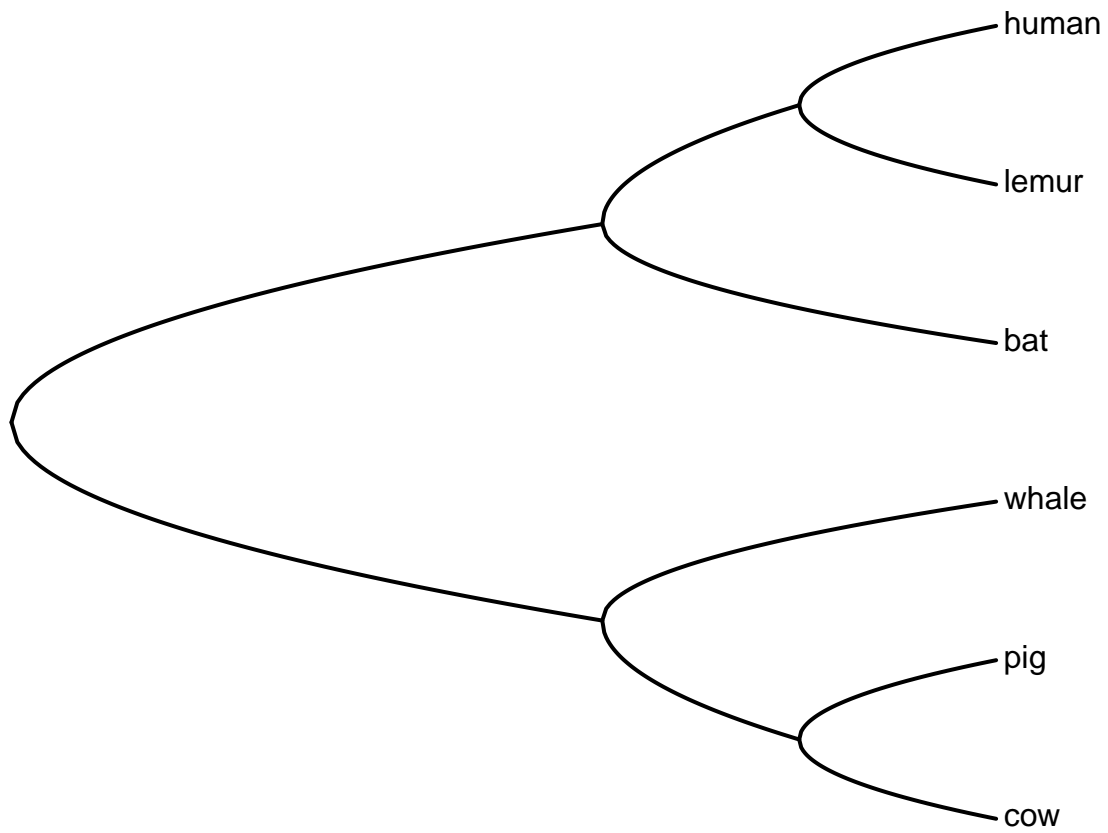
```
vert.tree <- read.tree(text =  "(Orangutans, (Gorillas, ((Chimpanzees, Bonobos), Humans)));")
plot(vert.tree, no.margin = TRUE, edge.width = 5)
```
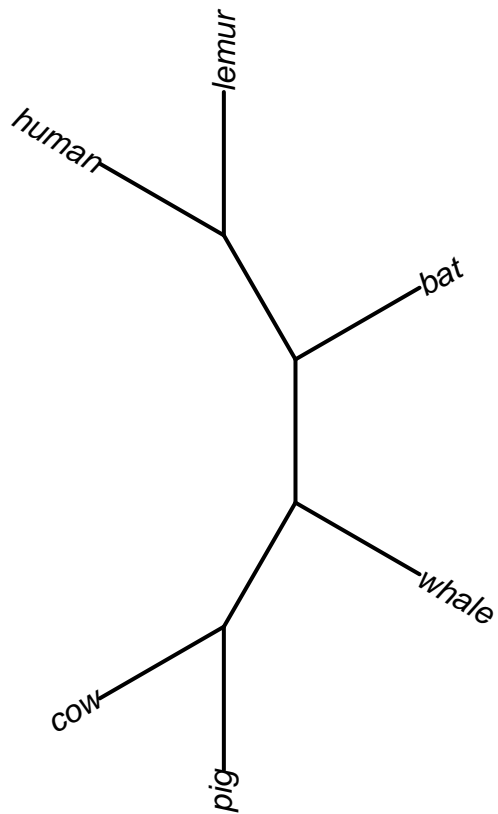
Phylogenetic Tools for Comparative Biology (and Other Things)

```
vert.tree <- read.tree(text = "(((cow, pig), whale), (bat, (lemur, human)));")
roundPhylogram(vert.tree)
```

```
# unrooted
plot(
  unroot(vert.tree),
  type = "unrooted",
  no.margin = TRUE,
  lab4ut = "axial",
  edge.width = 2
)
```
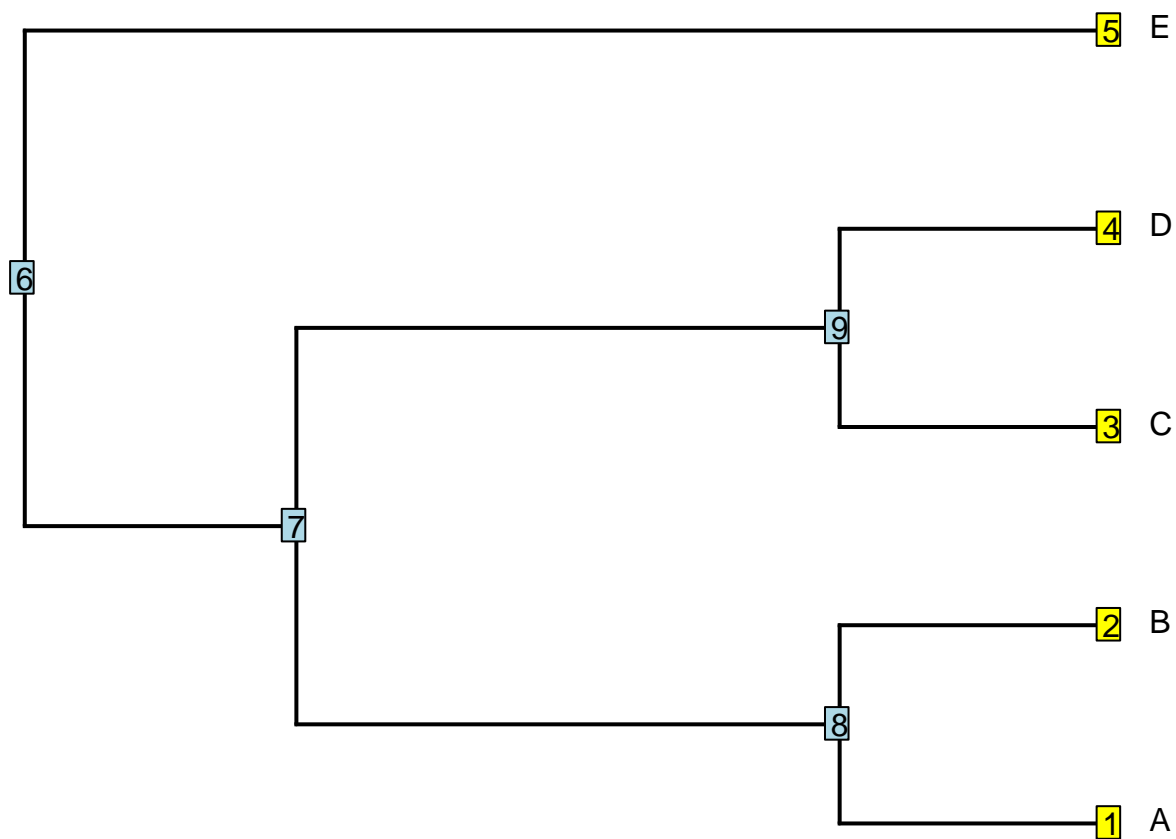
```
vert.tree
```

```
##
## Phylogenetic tree with 6 tips and 5 internal nodes.
##
## Tip labels:
##    cow, pig, whale, bat, lemur, human
##
## Rooted; no branch lengths.
```

```
str(vert.tree)
```

```
## List of 3
##  $ edge     : int [1:10, 1:2] 7 8 9 9 8 7 10 10 11 11 ...
##  $ Nnode    : int 5
##  $ tip.label: chr [1:6] "cow" "pig" "whale" "bat" ...
##  - attr(*, "class")= chr "phylo"
##  - attr(*, "order")= chr "cladewise"
```

```
# with labels
tree <- read.tree(text = "(((A, B),(C, D)),E );")
plotTree(tree, offset = 1)
tiplabels()
nodelabels()
```

Functions

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

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

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

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

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

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(substring(annotations, 1, 79), "[...]\n", substring(annotations, 80), "\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)
}
```
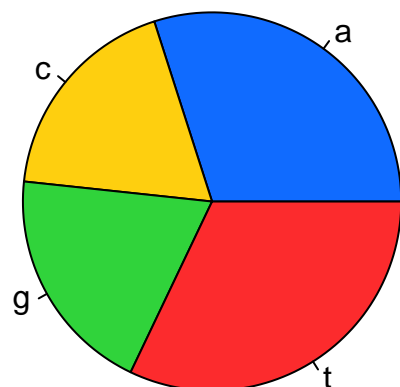
Reference Genome

```
stats(reference)
```

```
## >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:   attaaaggtttataccttcccaggtaacaaaccaaccaactttcgatctcttgtagatct ... (first 60)
## Complement: taatttccaaatatggaagggtccattgtttggttggttgaaagctagagaacatctaga ... (first 60)
##
## Graph:
```
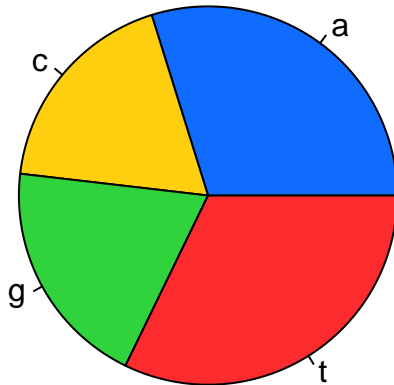
B.1.1.7 Variant

stats(B.1.1.7)

```
## >MW913791.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/ [...]
##  human/USA/IL-CDC-LC0032494/2021, complete genome
##
## Length: 29692 bases
##
## Composition:
##    a    c    g    t
## 8868 5449 5828 9547
##
## CG: 0.3797993
##
## Original:   agatctgttctctaaacgaactttaaaatctgtgtggctgtcactcggctgcatgcttag ... (first 60)
## Complement: tctagacaagagatttgcttgaaattttagacacaccgacagtgagccgacgtacgaatc ... (first 60)
##
## Graph:
```
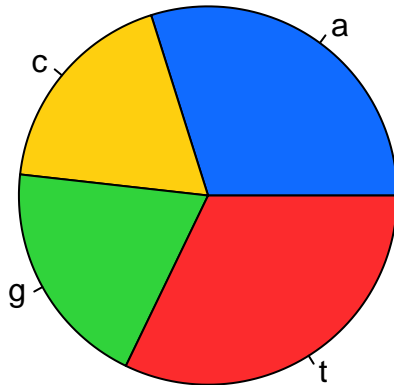
B.1.351 Variant

stats(B.1.351)

```
## >FR990267.1 Severe acute respiratory syndrome coronavirus 2 isolate Switzerland [...]
##  /GE-33136360/2021 genome assembly, chromosome: 1
##
## Length: 29819 bases
##
## Composition:
##    a    c    g    t
## 8838 5437 5824 9543
##
## CG: 0.3799001
##
## Original:   actttcgatctcttgtagatctgttctctaaacgaactttaaaatctgtgtggctgtcac ... (first 60)
## Complement: tgaaagctagagaacatctagacaagagatttgcttgaaattttagacacaccgacagtg ... (first 60)
##
## Graph:
```
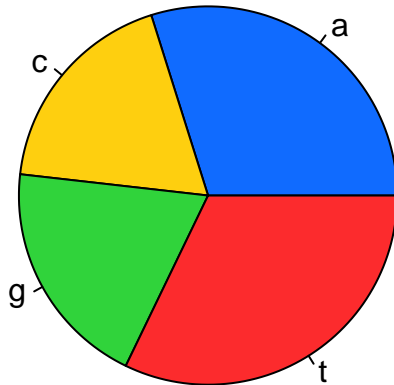
P.1 Variant

```
stats(P.1)
```

```
## >FR990291.1 Severe acute respiratory syndrome coronavirus 2 isolate Switzerland [...]
##  /GE-33154798/2021 genome assembly, chromosome: 1
##
## Length: 29865 bases
##
## Composition:
##    a    c    g    t
## 8919 5483 5855 9608
##
## CG: 0.3796417
##
## Original:   attaaaggtttataccttcccaggtaacaaaccaaccaactttcgatctcttgtagatct ... (first 60)
## Complement: taatttccaaatatggaagggtccattgtttggttggttgaaagctagagaacatctaga ... (first 60)
##
## Graph:
```


14
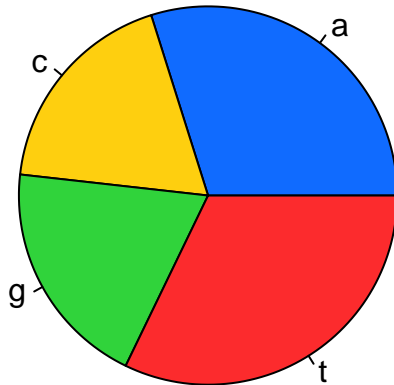
B.1.427 Variant

stats(B.1.427)

```
## >MW453109.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/ [...]
##  human/USA/CA-LACPHL-AF00094/2020, complete genome
##
## Length: 29826 bases
##
## Composition:
##    a    c    g    t
## 8862 5446 5827 9547
##
## CG: 0.3797925
##
## Original:   ccaactttcgatctcttgtagatctgttctctaaacgaactttaaaatctgtgtggctgt ... (first 60)
## Complement: ggttgaaagctagagaacatctagacaagagatttgcttgaaattttagacacaccgaca ... (first 60)
##
## Graph:
```

B.1.429 Variant

stats(B.1.429)

```
## >MW778462.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/ [...]
##  human/USA/AZ-CDC-STM-000020976/2021, complete genome
##
## Length: 29895 bases
##
## Composition:
##    a    c    g    t
## 8946 5482 5856 9609
##
## CG: 0.3792861
##
## Original:   tttataccttcccaggtaacaaaccaaccaactttcgatctcttgtagatctgttctcta ... (first 60)
## Complement: aaatatggaagggtccattgtttggttggttgaaagctagagaacatctagacaagagat ... (first 60)
##
## Graph:
```
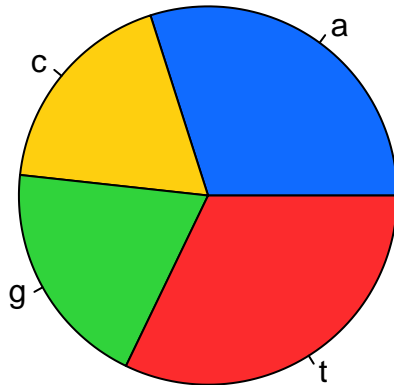
Comparison

```r
data <- data.frame(
  Variant = character(),
  Base = character(),
  Value = integer()
)

for (i in names(all.variants)) {
  comps <- seq.composition(all.variants[i][[1]])
  n <- nrow(data)
  data[n + 1,] <- list(Variant = i, Base = "A", Value = comps["a"])
  data[n + 2,] <- list(Variant = i, Base = "T", Value = comps["t"])
  data[n + 3,] <- list(Variant = i, Base = "G", Value = comps["g"])
  data[n + 4,] <- list(Variant = i, Base = "C", Value = comps["c"])
}

variants <- data["Variant"][[1]]
bases <- data["Base"][[1]]
value <- data["Value"][[1]]


if (FALSE) {

ggplot(data, aes(fill = bases, y = value, x = variants)) +
    geom_bar(position = "stack", stat = "identity") +
    scale_fill_manual(values = c("#106BFF", "#FECF0F", "#30D33B", "#FC2B2D")) +
    geom_text(
      aes(colour = "#000000", label=value),
      position=position_stack(),
      vjust=1.5,
      colour = "white",
      fontface = "bold"
    ) +
    labs(
      title = "Nitrogenous Base Distribution per Variant",
      subtitle = "for SARS-CoV-2",
      caption = "Data source: NCBI",
      x = "Variants",
      y = "Quantity",
      fill = "Bases"
    )
}
```

Interpretación

Todas las variantes del virus tienen diferente cantidad de nucleótidos, y differente composición de los mismos, sin embargo, a la hora de compararlos se pueden observar que las variaciones que hay entre ellos son míminas, y casi desapercibidas.

Esto es interesante, ya que estas minúsculas mutaciones en el genoma del virus pueden llegar a tener rasgos que los vuelvan resistentes a las vacunas que se desarrollen. Otro efecto de las mutaciones puede ser el más facil contagio, o peores síntomas. Este es el motivo por el que se hacen investigaciones para detectar las variantes, sus probables causas de mutación, y mitigar posibles efectos negativos que generen.