

**TECNOLÓGICO DE MONTERREY, CAMPUS MONTERREY**



**Tecnológico  
de Monterrey**

**Evidencia 2 | Proyecto integrador**

**Por**

**Ernesto Andre González Castro, #A01741225**

**Análisis de biología computacional (Gpo 202)**

**Raquel Acuña González**

**6 de mayo del 2023**

## Parte 1

<https://drive.google.com/file/d/1IATUZEOWLEAWogU3nI0VuThimzOGBSTU/view?usp=sharing>

## Parte 2

# Evidencia 2 | Proyecto integrador

Ernesto Andre González Castro, A01741225

2023-05-07

```
variantes <- c("OQ928765", "OQ701260", "OQ921107", "OQ059025", "OQ888915", "OX466324", "OM945722", "ON115271", "OQ748861", "OQ676999", "OP303172", "OQ748862", "OP412818", "OQ457497", "OU327282", "MZ570430", "OQ826413", "OQ120890", "ON545852", "OP013083")
```

```
library(Biostrings)
```

```
## Loading required package: BiocGenerics
```

```
##  
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':  
##  
## IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':  
##  
## anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
## colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
## get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
## match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
## Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,  
## table, tapply, union, unique, unsplit, which.max, which.min
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
##  
## Attaching package: 'S4Vectors'
```

```
## The following objects are masked from 'package:base':  
##  
## expand.grid, I, unname
```

```
## Loading required package: IRanges
```

```
##  
## Attaching package: 'IRanges'
```

```
## The following object is masked from 'package:grDevices':  
##  
## windows
```

```
## Loading required package: XVector
```

```
## Loading required package: GenomeInfoDb
```

```
##  
## Attaching package: 'Biostrings'
```

```
## The following object is masked from 'package:base':  
##  
## strsplit
```

```
library(seqinr)
```

```
##  
## Attaching package: 'seqinr'
```

```
## The following object is masked from 'package:Biostrings':  
##  
## translate
```

```
library(adegenet)
```

```
## Loading required package: ade4
```

```
##  
## Attaching package: 'ade4'
```

```
## The following object is masked from 'package:Biostrings':  
##  
## score
```

```
## The following object is masked from 'package:BiocGenerics':  
##  
## score
```

```
##  
## /// adegenet 2.1.10 is loaded //////////////////////////////////  
##  
## > overview: '?adegenet'  
## > tutorials/doc/questions: 'adegenetWeb()'  
## > bug reports/feature requests: adegenetIssues()
```

```
library(ape)
```

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

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

```
## The following object is masked from 'package:Biostrings':  
##  
##   complement
```

```
library(ggtree)
```

```
## ggtree v3.6.2 For help: https://yulab-smu.top/treedata-book/  
##  
## If you use the ggtree package suite in published research, please cite  
## the appropriate paper(s):  
##  
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.  
## ggtree: an R package for visualization and annotation of phylogenetic  
## trees with their covariates and other associated data. Methods in  
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628  
##  
## Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods  
## for mapping and visualizing associated data on phylogeny using ggtree.  
## Molecular Biology and Evolution. 2018, 35(12):3041-3043.  
## doi:10.1093/molbev/msy194  
##  
## G Yu. Data Integration, Manipulation and Visualization of Phylogenetic  
## Trees (1st ed.). Chapman and Hall/CRC. 2022. ISBN: 9781032233574
```

```
##  
## Attaching package: 'ggtree'
```

```
## The following object is masked from 'package:ape':  
##  
##   rotate
```

```
## The following object is masked from 'package:Biostrings':  
##  
##   collapse
```

```
## The following object is masked from 'package:IRanges':  
##  
##   collapse
```

```
## The following object is masked from 'package:S4Vectors':
```

```
##
```

```
##      expand
```

```
library(DECIPHER)
```

```
## Loading required package: RSQLite
```

```
## Loading required package: parallel
```

```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
library(ggplot2)
```

```
variantes_secuencias <- read.GenBank(variantes)
```

```
str(variantes_secuencias)
```

```
## List of 20
## $ OQ928765: raw [1:29648] 88 48 88 18 ...
## $ OQ701260: raw [1:29764] 88 88 28 88 ...
## $ OQ921107: raw [1:29726] 88 48 88 18 ...
## $ OQ059025: raw [1:29840] 88 88 88 48 ...
## $ OQ888915: raw [1:29747] 18 18 48 18 ...
## $ OX466324: raw [1:29847] f0 f0 f0 f0 ...
## $ OM945722: raw [1:29831] 18 88 18 88 ...
## $ ON115271: raw [1:29596] 88 28 28 18 ...
## $ OQ748861: raw [1:29686] 28 18 18 18 ...
## $ OQ676999: raw [1:600] 88 48 28 88 ...
## $ OP303172: raw [1:29759] 28 18 18 18 ...
## $ OQ748862: raw [1:29685] 18 18 18 28 ...
## $ OP412818: raw [1:668] 88 48 88 48 ...
## $ OQ457497: raw [1:3803] 88 18 48 18 ...
## $ OU327282: raw [1:29884] f0 f0 f0 f0 ...
## $ MZ570430: raw [1:29782] 88 48 88 18 ...
## $ OQ826413: raw [1:29721] 18 18 18 88 ...
## $ OQ120890: raw [1:29324] 48 18 18 28 ...
## $ ON545852: raw [1:29513] 88 48 18 88 ...
## $ OP013083: raw [1:29501] 18 28 48 18 ...
## - attr(*, "class")= chr "DNABin"
## - attr(*, "description")= chr [1:20] "OQ928765.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/AZ00179874/2023 ORF1ab "| __truncated__ "OQ701260.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IND/2892/2023 ORF1ab polypr"| __truncated__ "OQ921107.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/BRA/LACENAL-270247198/2023 "| __truncated__ "OQ059025.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/RUS/6205/2022, complete genome" ...
## - attr(*, "species")= chr [1:20] "Severe_acute_respiratory_syndrome_coronavirus_2" "Severe_acute_respiratory_syndrome_coronavirus_2" "Severe_acute_respiratory_syndrome_coronavirus_2" "Severe_acute_respiratory_syndrome_coronavirus_2" ...
```

```
names(variantes_secuencias) <- c("USA","INDIA","BRAZIL","RUSIA","FRANCIA","REINO UNIDO","TURQUÍA","ESPAÑA",
"ITALIA", "ARGENTINA", "COLOMBIA", "ALEMANIA", "IRÁN", "MÉXICO", "POLONIA", "INDONESIA", "SUDÁFRICA", "PERÚ", "PAÍSES BAJOS", "CHILE")
attributes(variantes_secuencias)
```

```

## $names
## [1] "USA"          "INDIA"          "BRAZIL"          "RUSIA"          "FRANCIA"
## [6] "REINO UNIDO"  "TURQUÍA"        "ESPAÑA"          "ITALIA"          "ARGENTINA"
## [11] "COLOMBIA"     "ALEMANIA"       "IRÁN"            "MÉXICO"          "POLONIA"
## [16] "INDONESIA"     "SUDÁFRICA"      "PERÚ"            "PAÍSES BAJOS"   "CHILE"
##
## $class
## [1] "DNABin"
##
## $description
## [1] "OQ928765.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/AZ0017987
4/2023 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (OR
F3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), and ORF7
b (ORF7b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and ORF10
protein (ORF10) genes, complete cds"
## [2] "OQ701260.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IND/2892/2023
ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), e
nvelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), and ORF7b (ORF7
b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and ORF10 protein
(ORF10) genes, partial cds"
## [3] "OQ921107.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/BRA/LACENAL-27
0247198/2023 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a prote
in (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), and ORF7a protein (ORF7a)
genes, complete cds; ORF7b (ORF7b) and ORF8 protein (ORF8) genes, partial cds; and nucleocapsid phosphoprot
ein (N) and ORF10 protein (ORF10) genes, complete cds"
## [4] "OQ059025.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/RUS/6205/2022,
complete genome"
## [5] "OQ888915.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/FRA/IHUCOVID-1
02580/2023 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein
(ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), and O
RF7b (ORF7b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and ORF
10 protein (ORF10) genes, complete cds"
## [6] "OX466324.1 Severe acute respiratory syndrome coronavirus 2 genome assembly, complete genome: monop
artite"
## [7] "OM945722.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TUR/ERAGEM-OM-
1104/2022, complete genome"
## [8] "ON115271.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/Cat/ESP/Omicron/2022
ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), e
nvelope protein (E), membrane glycoprotein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a gene, co
mplete sequence; and ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein
(ORF10) genes, complete cds"
## [9] "OQ748861.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ITA/FRI-USAFSA
M-S21299/2023 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a prot
ein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), a
nd ORF7b (ORF7b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and
ORF10 protein (ORF10) genes, complete cds"
## [10] "OQ676999.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ARG/CaseM-97L/
2022 surface glycoprotein (S) gene, partial cds"
## [11] "OP303172.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/COL/LIME_2210
5/2022, complete genome"
## [12] "OQ748862.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/DEU/BW-USAFSAM
-S21300/2023 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a prote
in (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), an
d ORF7b (ORF7b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and

```

```

ORF10 protein (ORF10) genes, complete cds"
## [13] "OP412818.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/SARS-COV/IRN/A1/2022
surface glycoprotein (S) gene, partial cds"
## [14] "OQ457497.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/MEX/Omicron/20
22 surface glycoprotein (S) gene, partial cds"
## [15] "OU327282.1 Severe acute respiratory syndrome coronavirus 2 isolate 1 genome assembly, chromosome:
1"
## [16] "MZ570430.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/IDN/RSDS-RCVTD
-UNAIR-54-A/2021 ORF1ab polyprotein (ORF1ab) gene, partial cds; ORF1a polyprotein (ORF1ab) gene, complete c
ds; surface glycoprotein (S) gene, partial cds; ORF3a protein (ORF3a), envelope protein (E), membrane glyco
protein (M), and ORF6 protein (ORF6) genes, complete cds; ORF7a protein (ORF7a) and ORF7b (ORF7b) genes, pa
rtial cds; and ORF8 protein (ORF8), nucleocapsid phosphoprotein (N), and ORF10 protein (ORF10) genes, compl
ete cds"
## [17] "OQ826413.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/ZAF/NHLS-UCT-G
S-CH55/2023 ORF1ab polyprotein (ORF1ab) and ORF1a polyprotein (ORF1ab) genes, partial cds; surface glycopro
tein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7
a protein (ORF7a), and ORF7b (ORF7b) genes, complete cds; ORF8 gene, complete sequence; and nucleocapsid ph
osphoprotein (N) and ORF10 protein (ORF10) genes, complete cds"
## [18] "OQ120890.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/PER/14090AAXA
I/2022 ORF1ab polyprotein (ORF1ab) and ORF1a polyprotein (ORF1ab) genes, partial cds; surface glycoprotein
(S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), and ORF7a
protein (ORF7a) genes, complete cds; ORF7b (ORF7b) and ORF8 protein (ORF8) genes, partial cds; and nucleoca
psid phosphoprotein (N) and ORF10 protein (ORF10) genes, complete cds"
## [19] "ON545852.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/NLD/EMC-BA2-1/
2022 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3
a), envelope protein (E), membrane glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), and ORF7b
(ORF7b) genes, complete cds; ORF8 protein (ORF8) gene, partial cds; and nucleocapsid phosphoprotein (N) and
ORF10 protein (ORF10) genes, complete cds"
## [20] "OP013083.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/Chile/CL-AR-UW
22011771334/2022 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), and surface glycoprotein (S) ge
ne, complete cds; ORF3a gene, complete sequence; and envelope protein (E), membrane glycoprotein (M), ORF6 p
rotein (ORF6), ORF7a protein (ORF7a), ORF7b (ORF7b), ORF8 protein (ORF8), nucleocapsid phosphoprotein (N),
and ORF10 protein (ORF10) genes, complete cds"
##
## $species
## [1] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [2] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [3] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [4] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [5] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [6] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [7] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [8] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [9] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [10] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [11] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [12] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [13] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [14] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [15] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [16] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [17] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [18] "Severe_acute_respiratory_syndrome_coronavirus_2"

```



```
## [19] "Severe_acute_respiratory_syndrome_coronavirus_2"
## [20] "Severe_acute_respiratory_syndrome_coronavirus_2"
```

```
names(variantes_secuencias)
```

```
## [1] "USA"          "INDIA"        "BRAZIL"       "RUSIA"        "FRANCIA"
## [6] "REINO UNIDO"  "TURQUÍA"      "ESPAÑA"      "ITALIA"       "ARGENTINA"
## [11] "COLOMBIA"    "ALEMANIA"     "IRÁN"        "MÉXICO"       "POLONIA"
## [16] "INDONESIA"    "SUDÁFRICA"    "PERÚ"        "PAÍSES BAJOS" "CHILE"
```

```
attr(variantes_secuencias, "Variantes")
```

```
## NULL
```

```
write.dna(variantes_secuencias, file = "variantes_sec.fasta", format = "fasta", append =
FALSE, nbcol = 6, colsep = " ", colw = 10)
```

```
variantes_sec_no_alineadas <- readDNASTringSet("variantes_sec.fasta", format = "fasta")
```

```
## Warning in .Call2("fasta_index", filexp_list, nrec, skip, seek.first.rec, :
## reading FASTA file variantes_sec.fasta: ignored 42493 invalid one-letter
## sequence codes
```

```
variantes_sec_no_alineadas <- OrientNucleotides(variantes_sec_no_alineadas)
```

```
## =====
## =====
##
## Time difference of 0.22 secs
```

```
variantes_sec_alineadas <- AlignSeqs(variantes_sec_no_alineadas)
```

```
## Determining distance matrix based on shared 11-mers:
## =====
##
## Time difference of 0.08 secs
##
## Clustering into groups by similarity:
## =====
##
## Time difference of 0.01 secs
##
## Aligning Sequences:
## =====
##
## Time difference of 5.78 secs
##
## Iteration 1 of 2:
##
## Determining distance matrix based on alignment:
## =====
##
## Time difference of 0.02 secs
##
## Reclustering into groups by similarity:
## =====
##
## Time difference of 0.01 secs
##
## Realigning Sequences:
## =====
##
## Time difference of 5.71 secs
##
## Iteration 2 of 2:
##
## Determining distance matrix based on alignment:
## =====
##
## Time difference of 0.03 secs
##
## Reclustering into groups by similarity:
## =====
##
## Time difference of 0.01 secs
##
## Realigning Sequences:
## =====
##
## Time difference of 5.96 secs
##
## Refining the alignment:
## =====
##
## Time difference of 8.58 secs
```

```
BrowseSeqs(variantes_sec_alineadas, highlight=0)
```

```
writeXStringSet(variantes_sec_alineadas, file = "variantes_sec_alineadas.fasta")
```

```
variantes_alineadas <- read.alignment("variantes_sec_alineadas.fasta", format = "fasta")
```

```
matriz_distancia <- dist.alignment(variantes_alineadas, matrix = "similarity")
```

```
temp <- as.data.frame(as.matrix(matriz_distancia))  
temp
```

##	USA	INDIA	BRAZIL	RUSIA	FRANCIA	REINO
## USA	0.00000000	0.03565044	0.02037601	0.04498606	0.02196950	0.02339021
## INDIA	0.03565044	0.00000000	0.03356844	0.04967848	0.03453715	0.03537438
## BRAZIL	0.02037601	0.03356844	0.00000000	0.04276429	0.01878817	0.02049477
## RUSIA	0.04498606	0.04967848	0.04276429	0.00000000	0.04388456	0.04482480
## FRANCIA	0.02196950	0.03453715	0.01878817	0.04388456	0.00000000	0.02194266
## REINO	0.02339021	0.03537438	0.02049477	0.04482480	0.02194266	0.00000000
## TURQUÍA	0.05163629	0.05430603	0.05054685	0.04777383	0.05046242	0.05188998
## ESPAÑA	0.05117837	0.05420355	0.04999566	0.04653051	0.04997485	0.05143445
## ITALIA	0.02323109	0.03562719	0.01950079	0.04495725	0.02195966	0.02409443
## ARGENTINA	0.12909944	0.13222147	0.13540064	0.05773503	0.12909944	0.13540064
## COLOMBIA	0.04188468	0.04684856	0.03984439	0.02840525	0.04105448	0.04211119
## ALEMANIA	0.02249455	0.03261305	0.01950180	0.04265307	0.02116195	0.02337663
## IRÁN	0.18954721	0.15811388	0.19345581	0.15952775	0.18954721	0.19345581
## MÉXICO	0.08278240	0.08363590	0.08435935	0.07068278	0.08803291	0.08483063
## POLONIA	0.06442218	0.06624137	0.06327392	0.06073752	0.06451530	0.06551553
## INDONESIA	0.05642471	0.05760011	0.05596976	0.05138302	0.05670480	0.05733858
## SUDÁFRICA	0.02027837	0.03493308	0.01774993	0.04414127	0.02047686	0.02122713
## PERÚ	0.02213069	0.03377422	0.02065591	0.04264835	0.02072159	0.02228678
## PAÍSES	0.03518169	0.04057027	0.03410513	0.03211248	0.03455413	0.03540763
## CHILE	0.05119753	0.05522150	0.05083552	0.04738922	0.05108605	0.05221398
##	TURQUÍA	ESPAÑA	ITALIA	ARGENTINA	COLOMBIA	ALEMANIA
## USA	0.05163629	0.05117837	0.02323109	0.12909944	0.04188468	0.02249455
## INDIA	0.05430603	0.05420355	0.03562719	0.13222147	0.04684856	0.03261305
## BRAZIL	0.05054685	0.04999566	0.01950079	0.13540064	0.03984439	0.01950180
## RUSIA	0.04777383	0.04653051	0.04495725	0.05773503	0.02840525	0.04265307
## FRANCIA	0.05046242	0.04997485	0.02195966	0.12909944	0.04105448	0.02116195
## REINO	0.05188998	0.05143445	0.02409443	0.13540064	0.04211119	0.02337663
## TURQUÍA	0.00000000	0.01838192	0.05160234	0.13540064	0.04453755	0.04994603
## ESPAÑA	0.01838192	0.00000000	0.05114446	0.12247449	0.04316927	0.04911475
## ITALIA	0.05160234	0.05114446	0.00000000	0.12909944	0.04185716	0.02248014
## ARGENTINA	0.13540064	0.12247449	0.12909944	0.00000000	0.05773503	0.12909944
## COLOMBIA	0.04453755	0.04316927	0.04185716	0.05773503	0.00000000	0.03937098
## ALEMANIA	0.04994603	0.04911475	0.02248014	0.12909944	0.03937098	0.00000000
## IRÁN	0.13950297	0.14985022	0.18954721	0.17087153	0.14985022	0.18954721
## MÉXICO	0.02293554	0.00000000	0.08278240	0.12247449	0.06693847	0.08278240
## POLONIA	0.05370982	0.05423445	0.06515693	0.16329932	0.05738958	0.06359094
## INDONESIA	0.04431920	0.04474839	0.05706329	0.16832508	0.04761635	0.05538278
## SUDÁFRICA	0.04996921	0.04947579	0.02189302	0.13540064	0.04093077	0.02109773
## PERÚ	0.04842556	0.04862252	0.02290584	0.12909944	0.03923489	0.02132529
## PAÍSES	0.04063410	0.04036931	0.03566453	0.07071068	0.02750232	0.03418987
## CHILE	0.02102943	0.01756640	0.05218181	0.12803688	0.04443196	0.05018942
##	IRÁN	MÉXICO	POLONIA	INDONESIA	SUDÁFRICA	PERÚ
## USA	0.18954721	0.08278240	0.06442218	0.05642471	0.02027837	0.02213069
## INDIA	0.15811388	0.08363590	0.06624137	0.05760011	0.03493308	0.03377422
## BRAZIL	0.19345581	0.08435935	0.06327392	0.05596976	0.01774993	0.02065591
## RUSIA	0.15952775	0.07068278	0.06073752	0.05138302	0.04414127	0.04264835
## FRANCIA	0.18954721	0.08803291	0.06451530	0.05670480	0.02047686	0.02072159
## REINO	0.19345581	0.08483063	0.06551553	0.05733858	0.02122713	0.02228678
## TURQUÍA	0.13950297	0.02293554	0.05370982	0.04431920	0.04996921	0.04842556
## ESPAÑA	0.14985022	0.00000000	0.05423445	0.04474839	0.04947579	0.04862252
## ITALIA	0.18954721	0.08278240	0.06515693	0.05706329	0.02189302	0.02290584
## ARGENTINA	0.17087153	0.12247449	0.16329932	0.16832508	0.13540064	0.12909944
## COLOMBIA	0.14985022	0.06693847	0.05738958	0.04761635	0.04093077	0.03923489

```

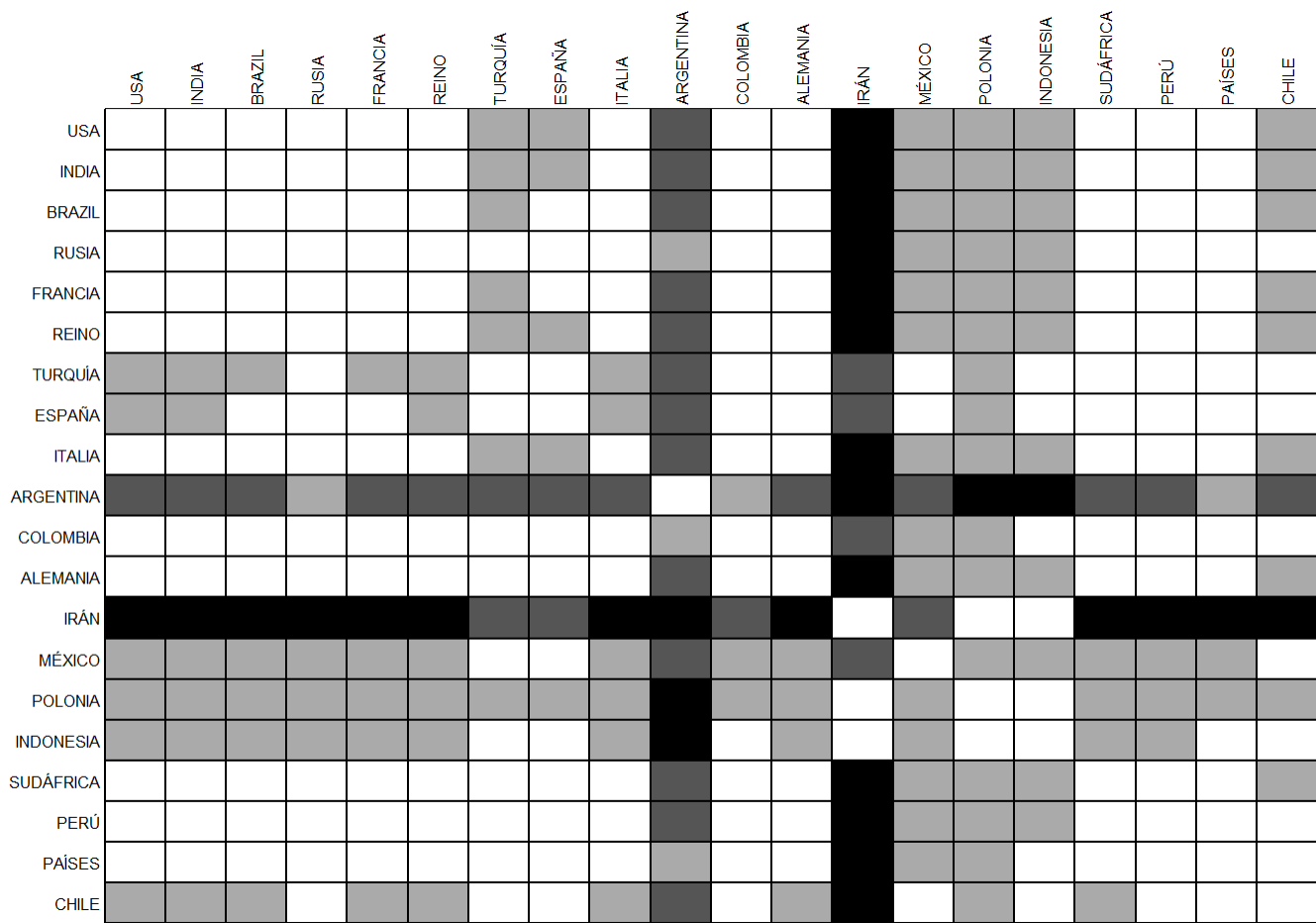
## ALEMANIA 0.18954721 0.08278240 0.06359094 0.05538278 0.02109773 0.02132529
## IRÁN 0.00000000 0.14985022 0.00000000 0.03946685 0.18954721 0.18555629
## MÉXICO 0.14985022 0.00000000 0.09729445 0.09304842 0.08279331 0.08278240
## POLONIA 0.00000000 0.09729445 0.00000000 0.03686188 0.06351852 0.06203689
## INDONESIA 0.03946685 0.09304842 0.03686188 0.00000000 0.05441793 0.05484460
## SUDÁFRICA 0.18954721 0.08279331 0.06351852 0.05441793 0.00000000 0.02145439
## PERÚ 0.18555629 0.08278240 0.06203689 0.05484460 0.02145439 0.00000000
## PAÍSES 0.15476465 0.06287775 0.05624472 0.04609091 0.03393925 0.03260658
## CHILE 0.15592044 0.00000000 0.05534057 0.04630282 0.05056780 0.04992622
## PAÍSES CHILE
## USA 0.03518169 0.05119753
## INDIA 0.04057027 0.05522150
## BRAZIL 0.03410513 0.05083552
## RUSIA 0.03211248 0.04738922
## FRANCIA 0.03455413 0.05108605
## REINO 0.03540763 0.05221398
## TURQUÍA 0.04063410 0.02102943
## ESPAÑA 0.04036931 0.01756640
## ITALIA 0.03566453 0.05218181
## ARGENTINA 0.07071068 0.12803688
## COLOMBIA 0.02750232 0.04443196
## ALEMANIA 0.03418987 0.05018942
## IRÁN 0.15476465 0.15592044
## MÉXICO 0.06287775 0.00000000
## POLONIA 0.05624472 0.05534057
## INDONESIA 0.04609091 0.04630282
## SUDÁFRICA 0.03393925 0.05056780
## PERÚ 0.03260658 0.04992622
## PAÍSES 0.00000000 0.04156002
## CHILE 0.04156002 0.00000000

```

```

table.paint(temp, cleg=0, clabel.row=.5, clabel.col=.5) + scale_color_viridis()

```

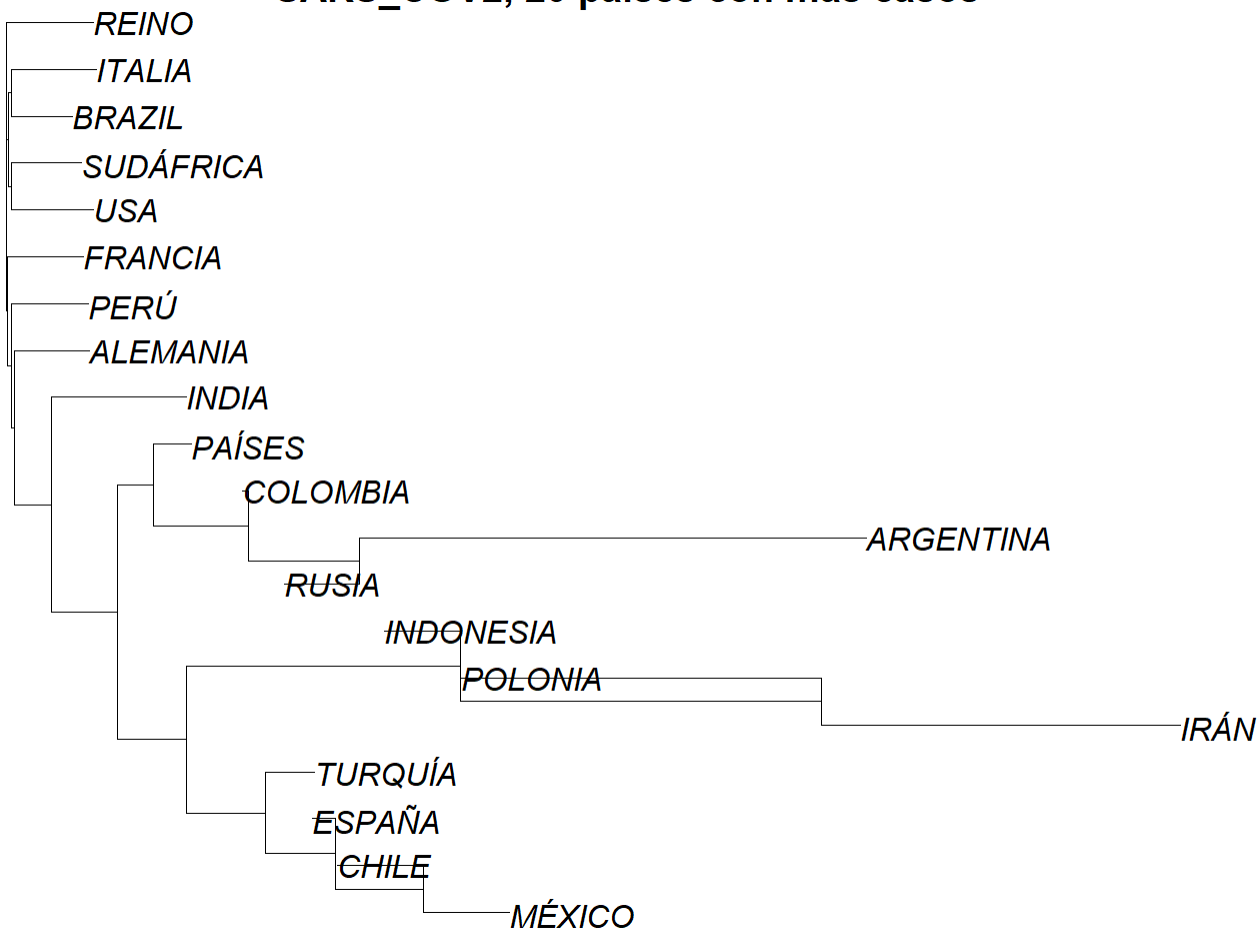


```
## NULL
```

```
variantes_filogenetico <- nj(matriz_distancia)
```

```
variantes_plot_filogenetico <- ladderize(variantes_filogenetico)
plot(variantes_plot_filogenetico,no.margin=TRUE,edge.width=0.5)
title("SARS_COV2, 20 países con más casos")
```

## SARS\_COV2, 20 países con mas casos



```
lista_composicion <- list()
```

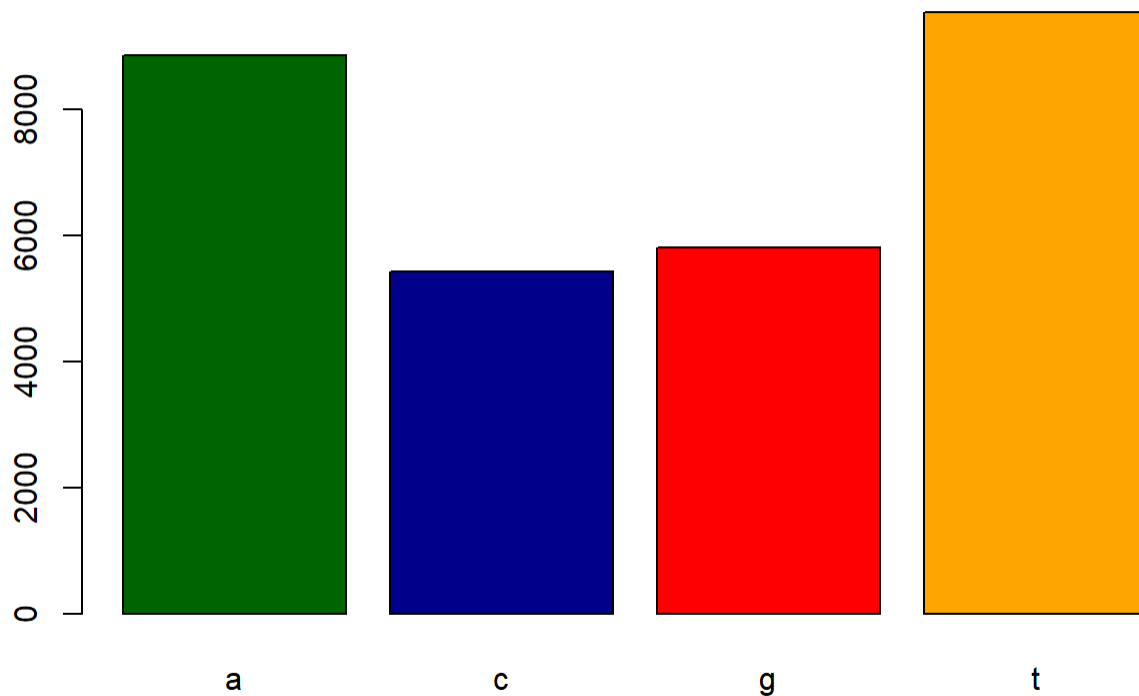
```

for (x in 1:20) {
  variante <- variantes_alineadas[[3]]
  variante <- variante[[x]]
  variante <- gsub("-", "", variante)
  variante <- s2c(variante)
  tamano = length(variante)
  composicion = count(variante, word=1)
  lista_composicion[[x]] <- as.matrix(composicion)
  gc = GC(variante)
  cat((names(variantes_secuencias))[x], "- Ranking:", x, "\n")
  cat("Tamaño:", tamano, "\n")
  print(composicion)
  cat("Porcentaje GC:", "")
  cat(gc, "\n\n")
  barplot(composicion, main=(names(variantes_secuencias))[x], col = c("darkgreen", "darkblue", "red", "orange"))
  contrasentido = rev(variante)
  cat("Secuencia contrasentido\n")
  cat("Primeros 10 nucleótidos:", head(contrasentido, 30), "\n")
  cat("Últimos 10 nucleótidos:", tail(contrasentido, 30))
  cat("\n-----\n")
}

```

```
## USA - Ranking: 1
## Tamaño: 29648
##
##   a   c   g   t
## 8859 5431 5809 9549
## Porcentaje GC: 0.3791149
```

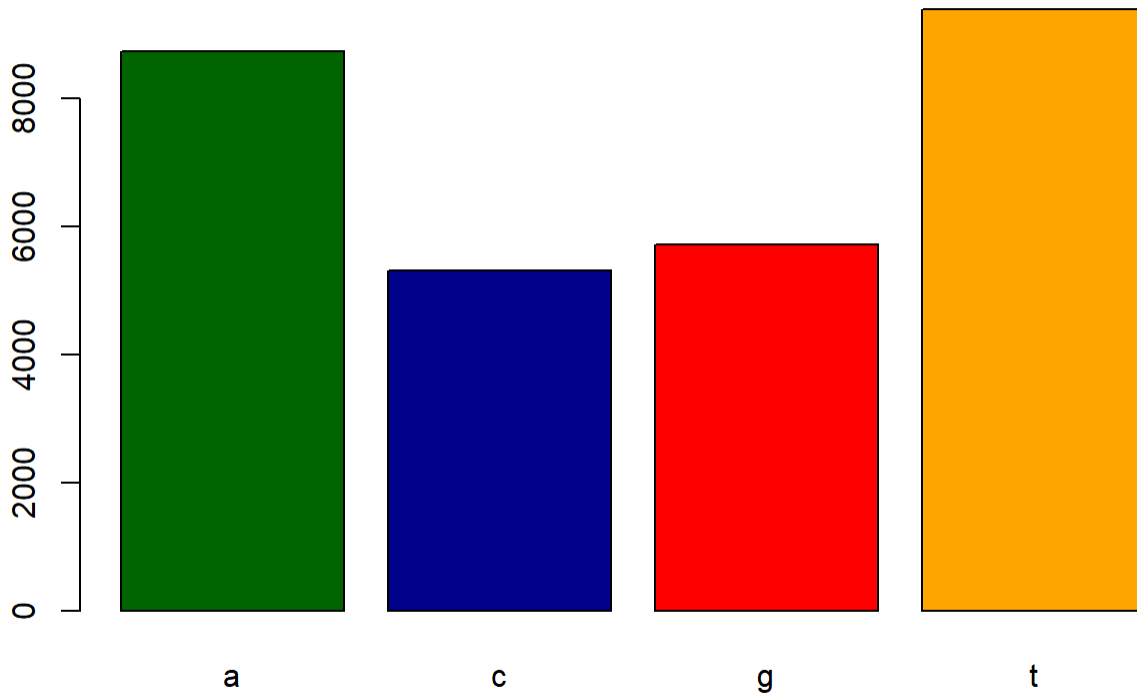
## USA



```
## Secuencia contrasentido
## Primeros 10 nucleótidos: c a c t t t t a c a c c a c c g a g a a a g t t c a g g a g
## Últimos 10 nucleótidos: c t a a a a t t t c a a g c a a a t c t c t t g t c t a g a
## -----
## INDIA - Ranking: 2
## Tamaño: 29764
##
##   a   c   g   t
## 8732 5319 5727 9396
## Porcentaje GC: 0.3786248
```

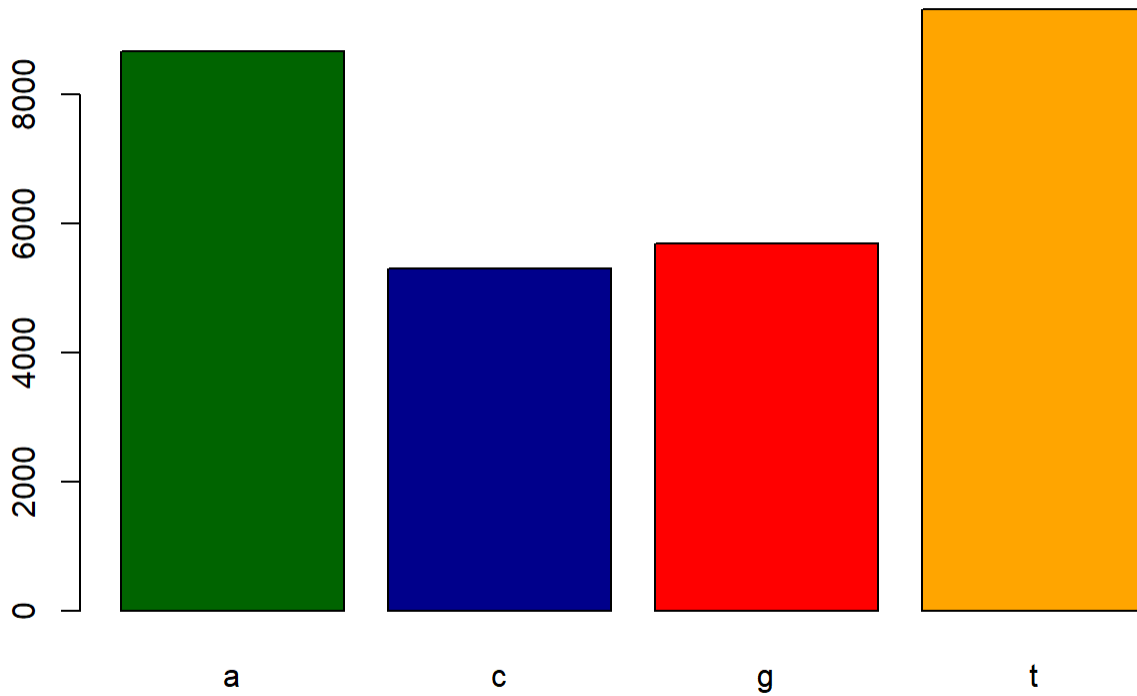


# INDIA



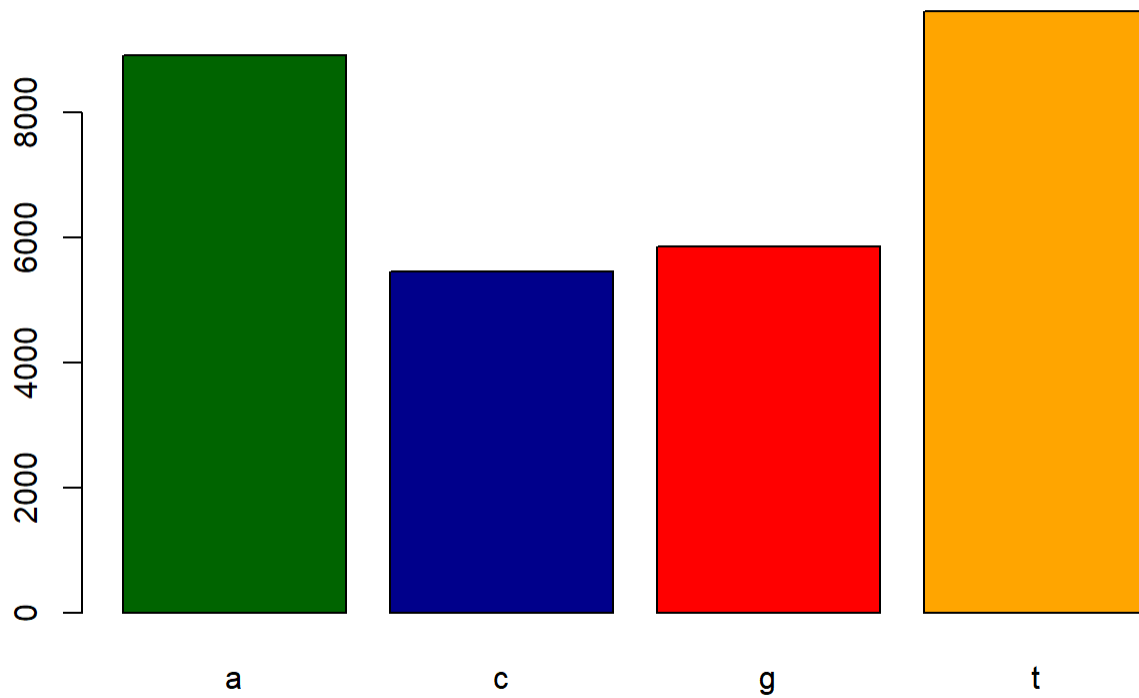
```
## Secuencia contrasentido
## Primeros 10 nucleótidos: t c c g t c g a g a g g g a t c g t a a c a a g t g a c a t
## Últimos 10 nucleótidos: a t g t t c t c t a g t t t t c a a c c a a c c a a a c a a
## -----
## BRAZIL - Ranking: 3
## Tamaño: 29726
##
##   a   c   g   t
## 8667 5296 5695 9323
## Porcentaje GC: 0.3792485
```

## BRAZIL



```
## Secuencia contrasentido
## Primeros 10 nucleótidos: c c t a t c g t g a t g a t t t t a a t t a a a a t g t g t
## Últimos 10 nucleótidos: c t a a a a t t t c a a g c a a a t c t c t t g t c t a g a
## -----
## RUSIA - Ranking: 4
## Tamaño: 29840
##
##   a   c   g   t
## 8915 5457 5855 9613
## Porcentaje GC: 0.3790885
```

## RUSIA



## Secuencia contrasentido

## Primeros 10 nucleótidos: g t g t a c c c c t a t c g t g a t g a t t t t a a t t a a

## Últimos 10 nucleótidos: c c a a a c a a t g g a c c c t t c c a t a t t t g g a a a

## -----

## FRANCIA - Ranking: 5

## Tamaño: 29747

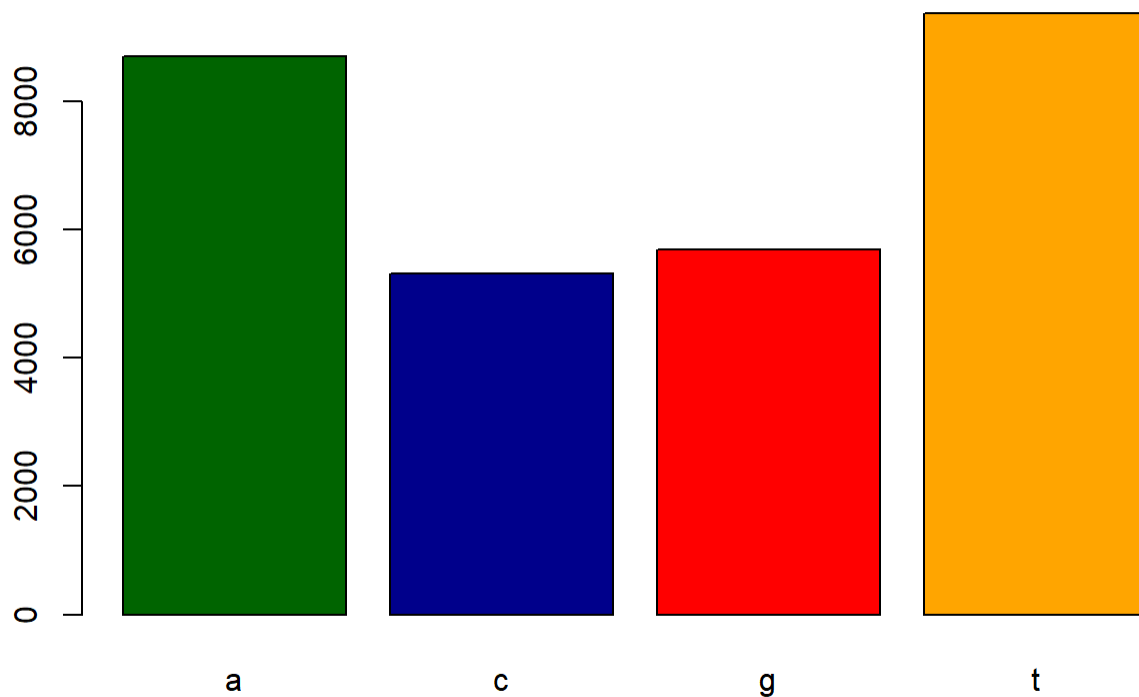
##

## a c g t

## 8698 5311 5696 9373

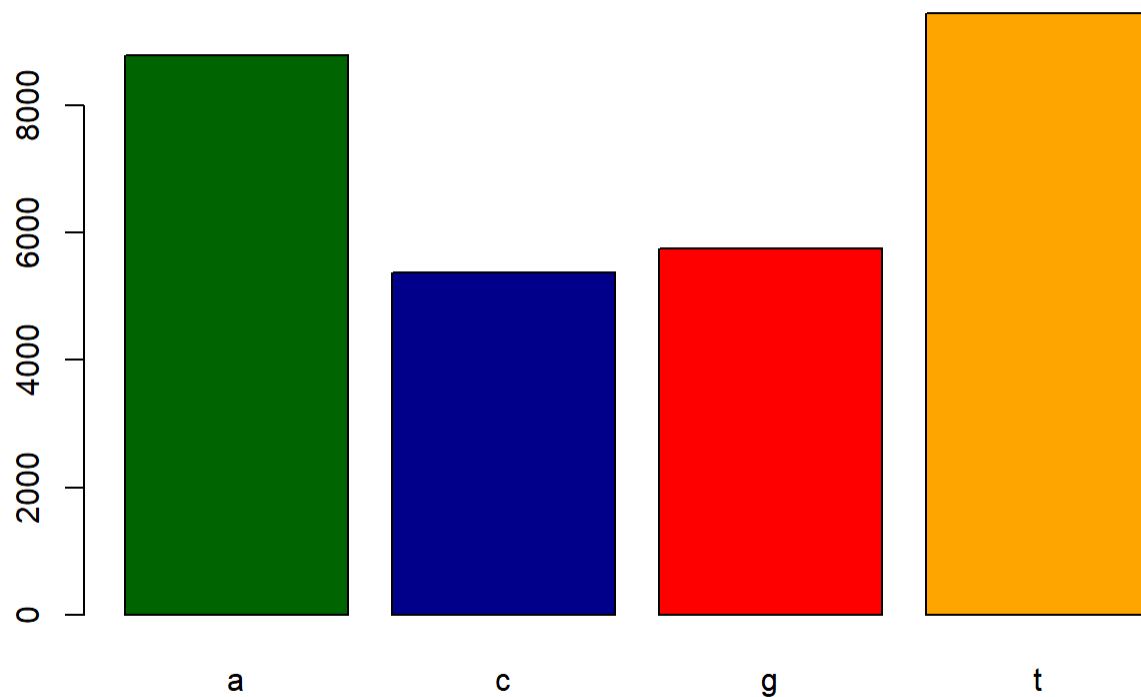
## Porcentaje GC: 0.3785336

## FRANCIA



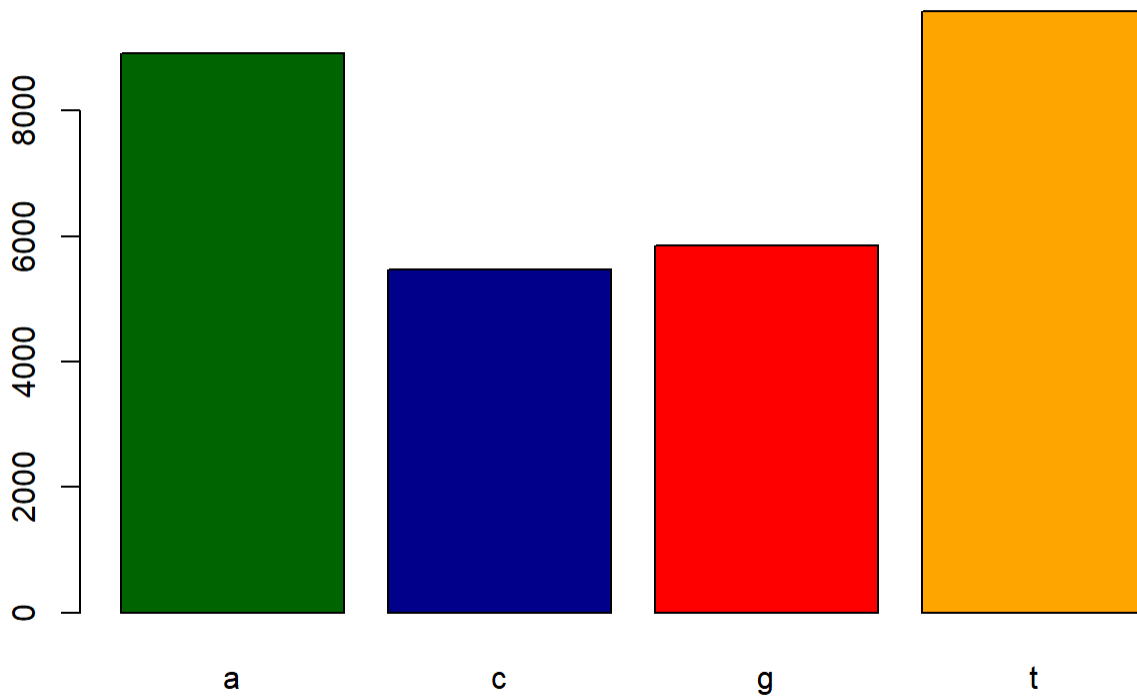
```
## Secuencia contrasentido
## Primeros 10 nucleótidos: a t g a t t t t a a t t a a a a t g t g t a a t c c c g a g
## Últimos 10 nucleótidos: a a t t t c a a g c a a a t c t c t t g t c t a g a t g t t
## -----
## REINO UNIDO - Ranking: 6
## Tamaño: 29847
##
##   a   c   g   t
## 8786 5374 5751 9453
## Porcentaje GC: 0.3788653
```

## REINO UNIDO



```
## Secuencia contrasentido  
## Primeros 10 nucleótidos: n n n n n n n n n n n n n n n n n n n n n n n n n n n n n n  
## Últimos 10 nucleótidos: a a c a a t g g a c c c t t c c a t a t t t g g a a n n n n  
## -----  
## TURQUÍA - Ranking: 7  
## Tamaño: 29831  
##  
##      a      c      g      t  
## 8917 5471 5855 9588  
## Porcentaje GC: 0.3796722
```

## TURQUÍA



## Secuencia contrasentido

## Primeros 10 nucleótidos: a c a g t a a g a g g a t t c t t c g a t a a t t t t a g t

## Últimos 10 nucleótidos: c a a c c a a c c a a a c a a t g g a c c c t t c c a t a t

## -----

## ESPAÑA - Ranking: 8

## Tamaño: 29596

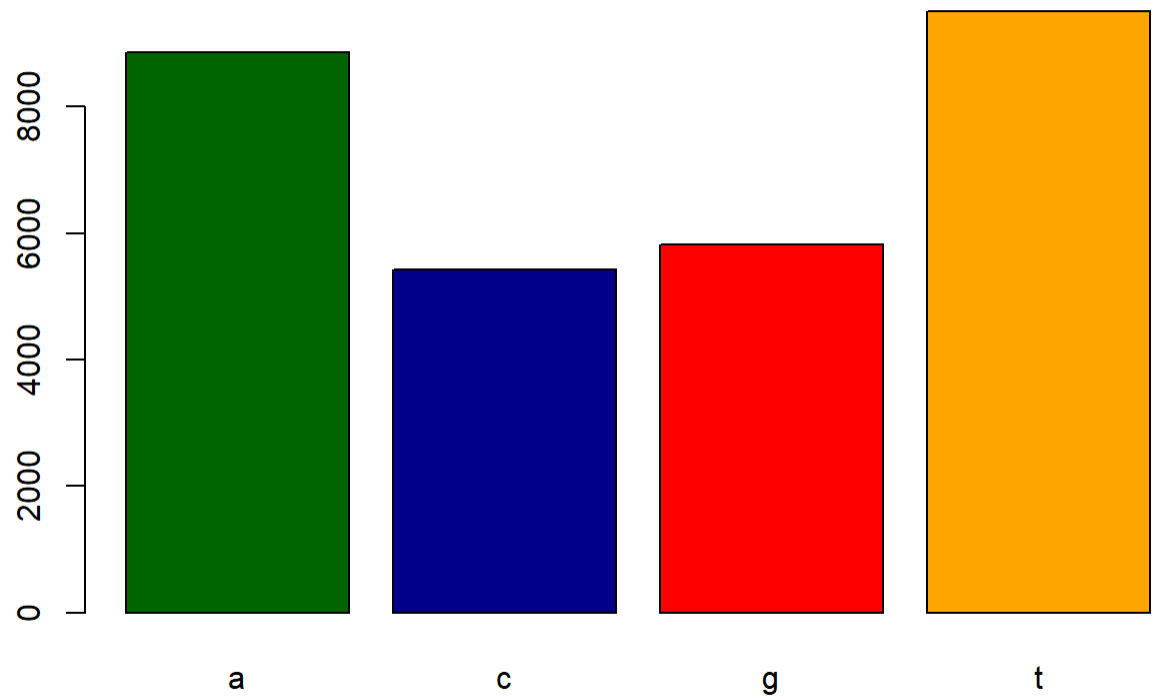
##

## a c g t

## 8851 5421 5817 9507

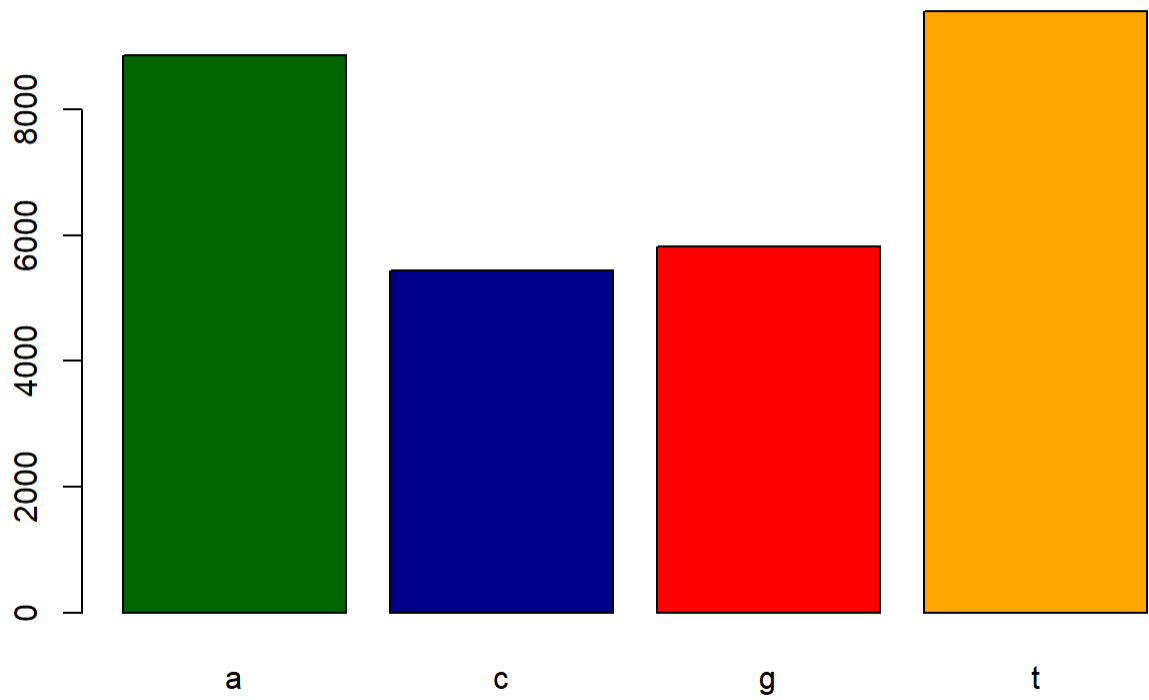
## Porcentaje GC: 0.3797135

ESPAÑA



```
## Secuencia contrasentido
## Primeros 10 nucleótidos: c a g t a a g a g g a t t c t t c g a t a a t t t t a g t g
## Últimos 10 nucleótidos: t t t c a a c c a a c c a a a c a a t g g a c c c t t c c a
## -----
## ITALIA - Ranking: 9
## Tamaño: 29686
##
##   a   c   g   t
## 8868 5434 5819 9565
## Porcentaje GC: 0.3790676
```

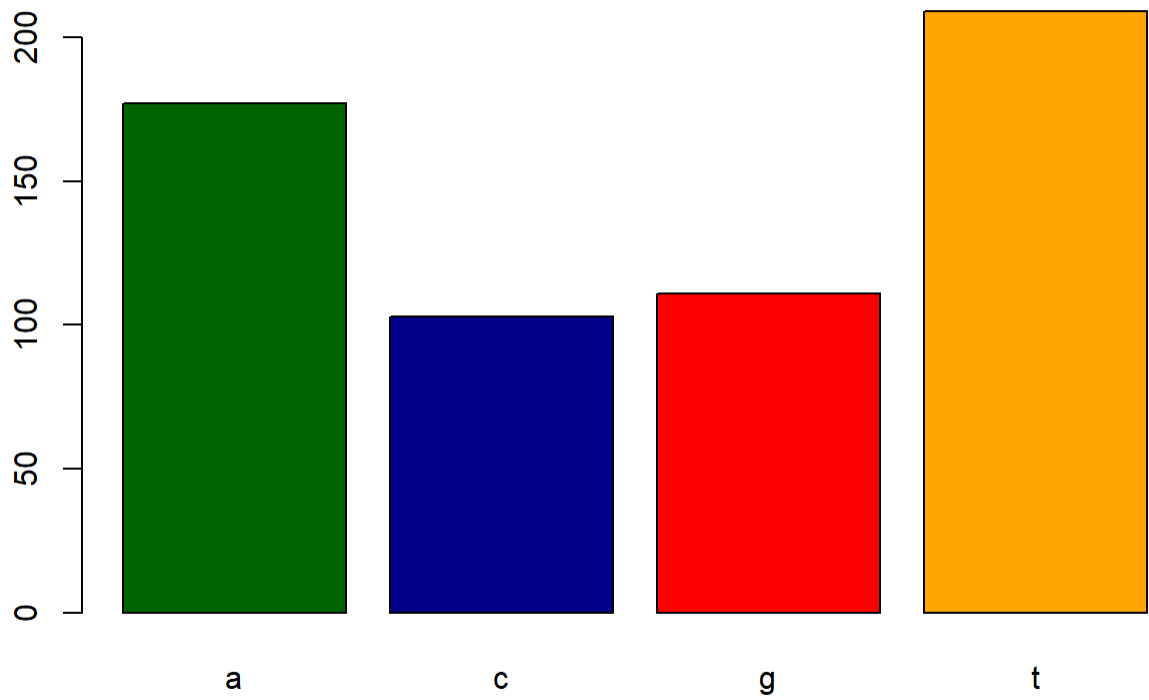
ITALIA



```
## Secuencia contrasentido
## Primeros 10 nucleótidos: g a g g g a t c g t a a c a a g t g a c a t c c a c t t t t
## Últimos 10 nucleótidos: a a t c t c t t g t c t a g a t g t t c t c t a g c t t t c
## -----
## ARGENTINA - Ranking: 10
## Tamaño: 600
##
##   a   c   g   t
## 177 103 111 209
## Porcentaje GC: 0.3566667
```

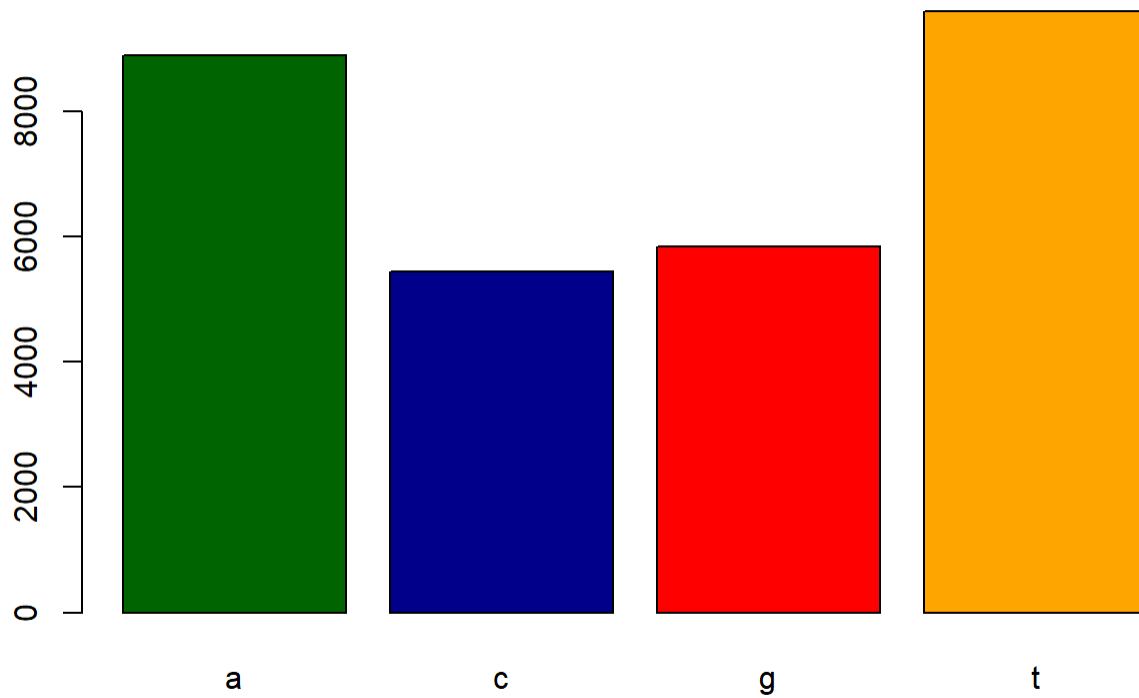


ARGENTINA



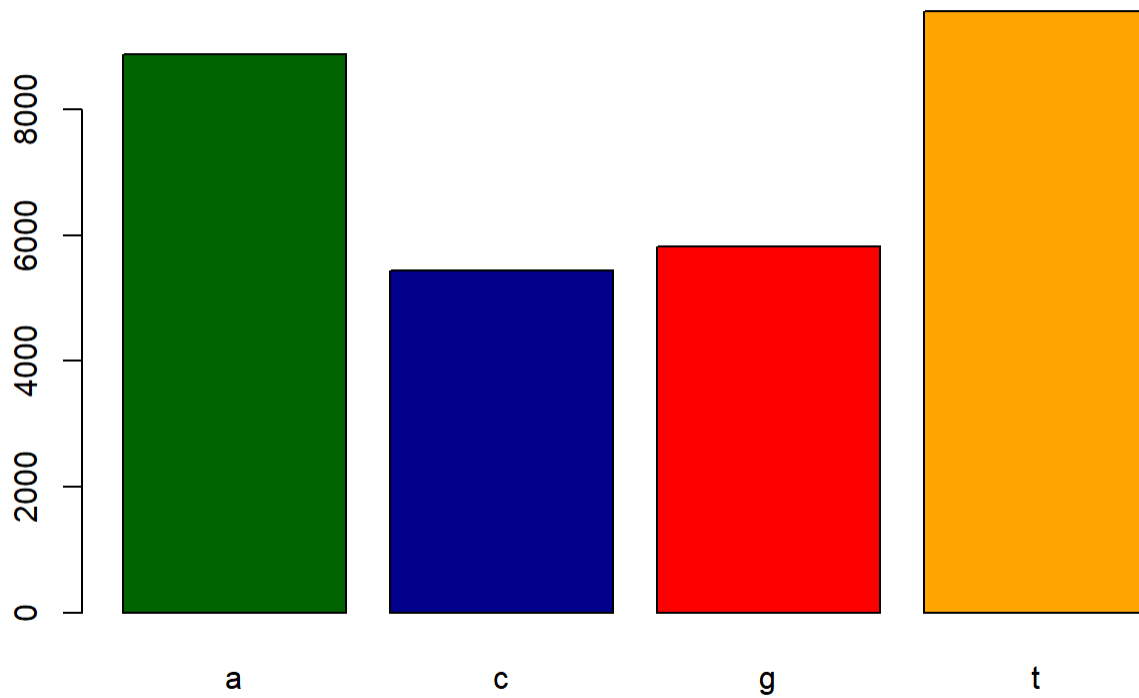
```
## Secuencia contrasentido
## Primeros 10 nucleótidos: g a a a a c a a t c t g a g t c a t t c t t g t g g a c a
## Últimos 10 nucleótidos: a t c c t g t c t t a t t a g t c g t t g t g t c a a c g a
## -----
## COLOMBIA - Ranking: 11
## Tamaño: 29759
##
##   a   c   g   t
## 8891 5441 5834 9593
## Porcentaje GC: 0.378877
```

## COLOMBIA



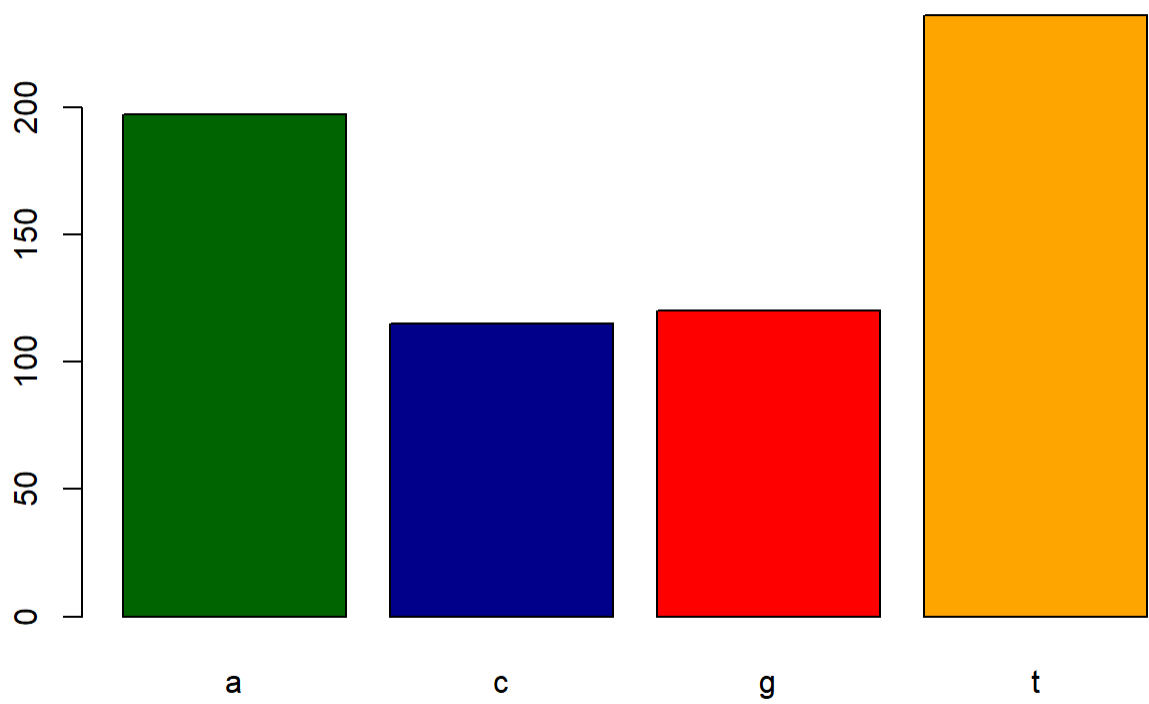
```
## Secuencia contrasentido
## Primeros 10 nucleótidos: c t t c g a t a a t t t t a g t g t a c c c c t a t c g t g
## Últimos 10 nucleótidos: a a t c t c t t g t c t a g a t g t t c t c t a g c t t t c
## -----
## ALEMANIA - Ranking: 12
## Tamaño: 29685
##
##   a   c   g   t
## 8870 5435 5816 9561
## Porcentaje GC: 0.3790513
```

## ALEMANIA



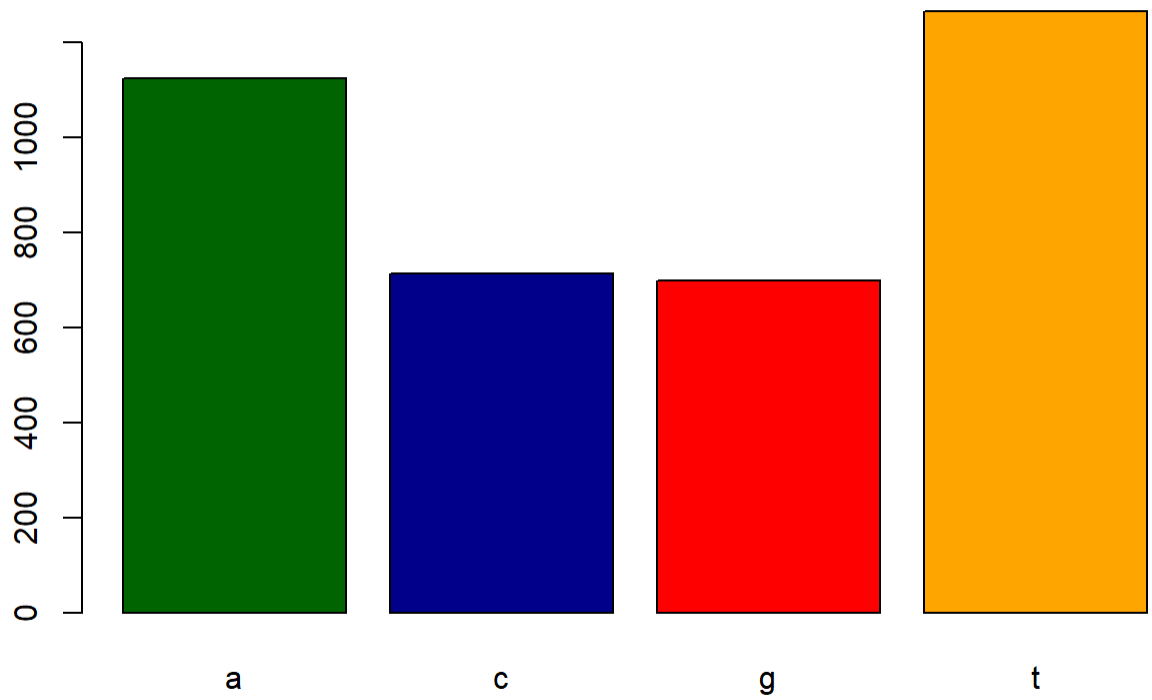
```
## Secuencia contrasentido
## Primeros 10 nucleótidos: g a g g g a t c g t a a c a a g t g a c a t c c a c t t t t
## Últimos 10 nucleótidos: a a a t c t c t t g t c t a g a t g t t c t c t a g c t t t
## -----
## IRÁN - Ranking: 13
## Tamaño: 668
##
##   a   c   g   t
## 197 115 120 236
## Porcentaje GC: 0.3517964
```

IRÁN



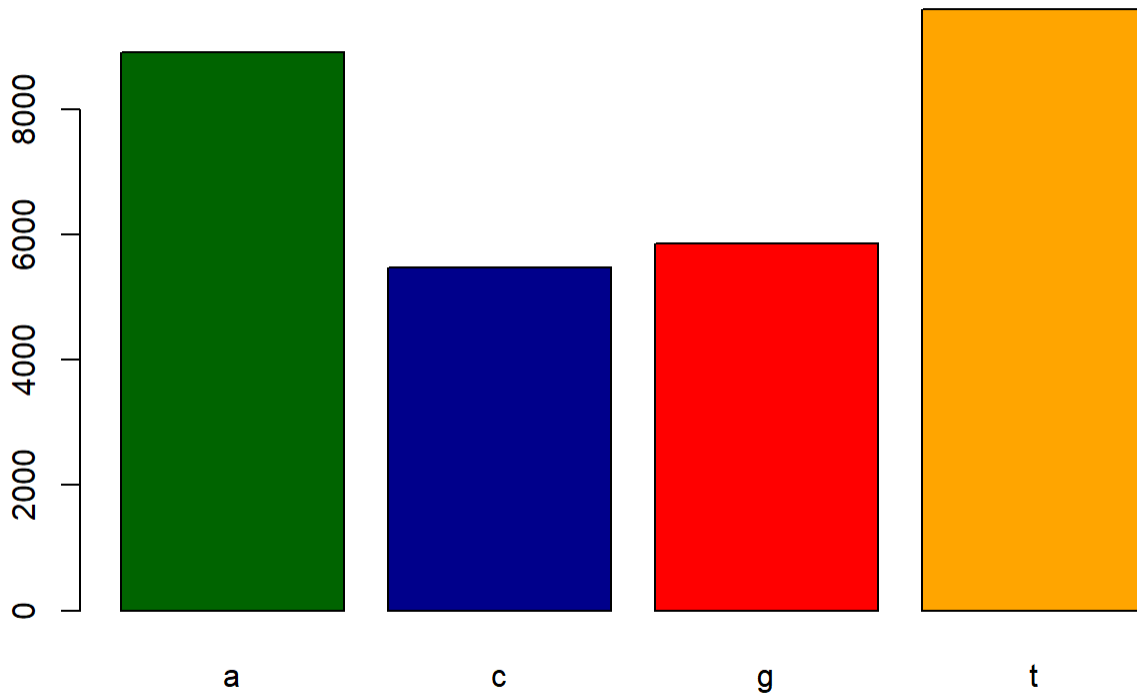
```
## Secuencia contrasentido
## Primeros 10 nucleótidos: t t t a a c t g t g t a a a c a a a a t t g g t t t a a t
## Últimos 10 nucleótidos: a g a t t g t t a t c t a a g a c a a c c a a c c t g a g a
## -----
## MÉXICO - Ranking: 14
## Tamaño: 3803
##
##   a   c   g   t
## 1124 714 699 1266
## Porcentaje GC: 0.3715488
```

MÉXICO



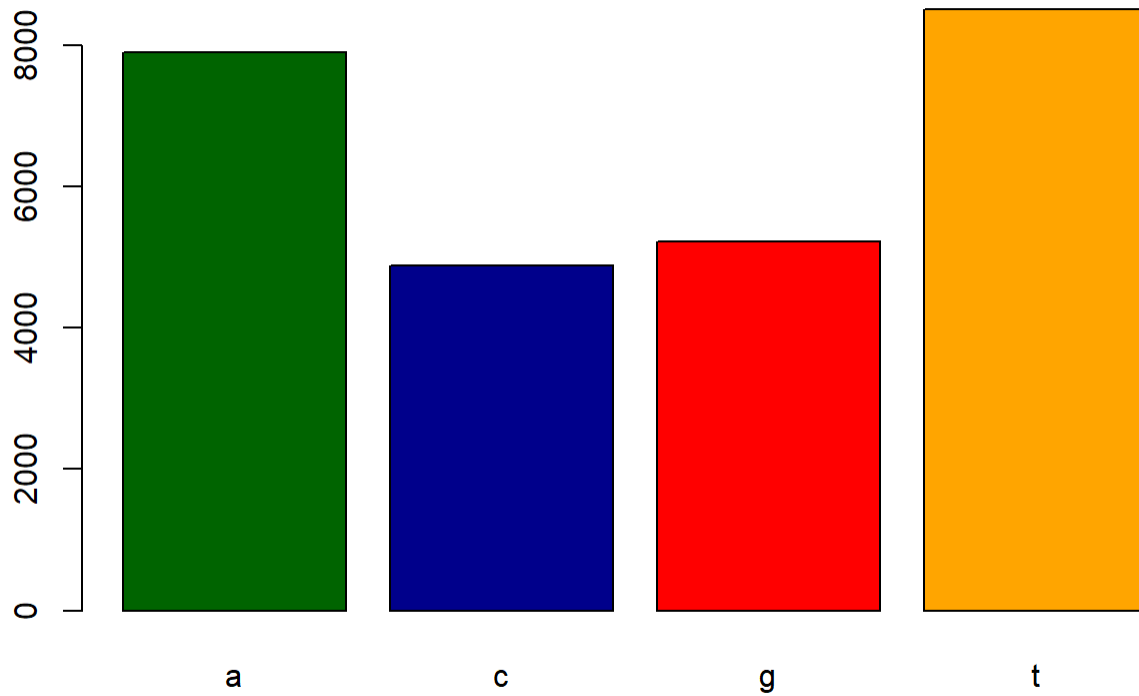
```
## Secuencia contrasentido
## Primeros 10 nucleótidos: a t a c a c a t t a c a t t a a a c t g a g g a a a c t c g
## Últimos 10 nucleótidos: a t c a c c g t t a t t t t g t t c t t t t t g t t t g t a
## -----
## POLONIA - Ranking: 15
## Tamaño: 29884
##
##   a   c   g   t
## 8909 5476 5853 9599
## Porcentaje GC: 0.3796964
```

# POLONIA



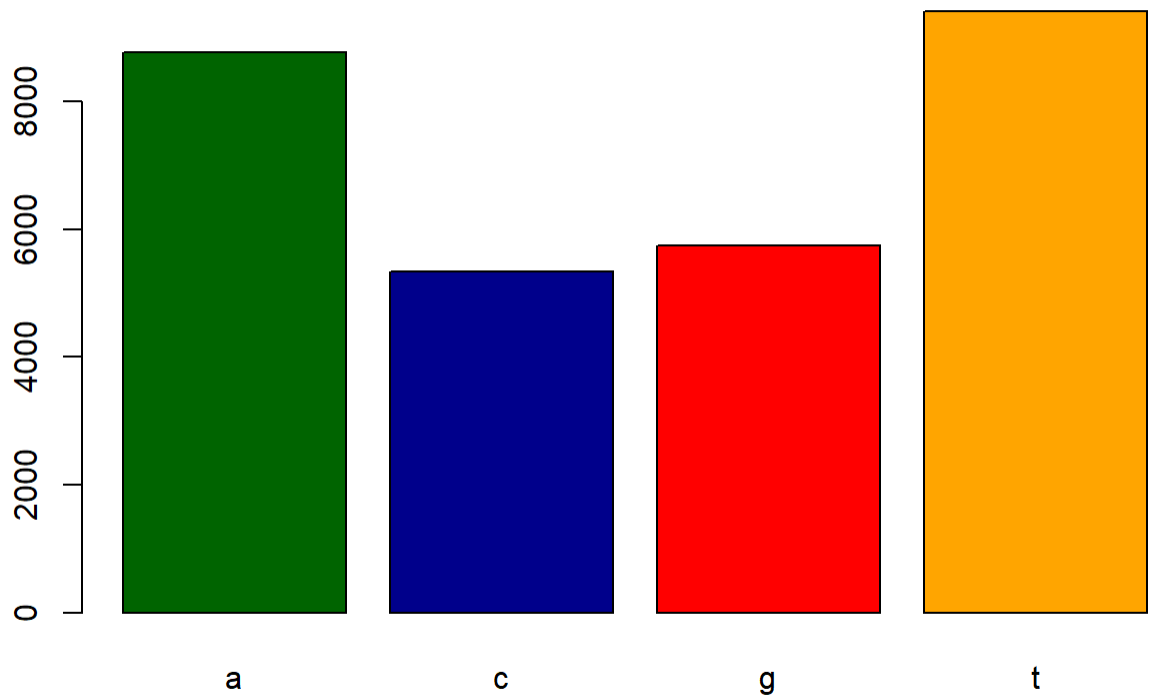
```
## Secuencia contrasentido
## Primeros 10 nucleótidos: n n n n n n n n n n n n n n n n n n n n n n n n n n
## Últimos 10 nucleótidos: a a c a a t g g a c c c t t c c a t a t t t g g n n n n n n
## -----
## INDONESIA - Ranking: 16
## Tamaño: 29782
##
##   a   c   g   t
## 7900 4881 5225 8507
## Porcentaje GC: 0.3811715
```

## INDONESIA



```
## Secuencia contrasentido
## Primeros 10 nucleótidos: c c t a t c g t g a t g a t t t t a a t t a a a a t g t g t
## Últimos 10 nucleótidos: c t a a a a t t t c a a g c a a a t c t c t t g t c t a g a
## -----
## SUDÁFRICA - Ranking: 17
## Tamaño: 29721
##
##   a   c   g   t
## 8764 5339 5739 9412
## Porcentaje GC: 0.3786833
```

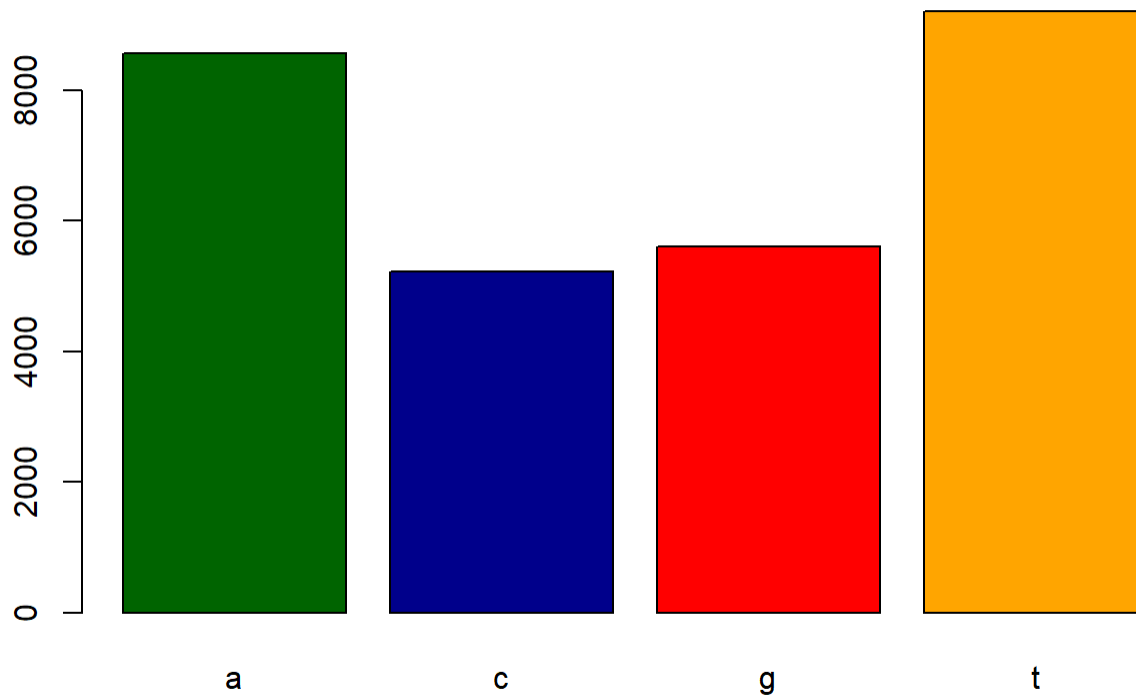
SUDÁFRICA



```
## Secuencia contrasentido
## Primeros 10 nucleótidos: t g a t t t t a a t t a a a a t g t g t a a t c c c g a g a
## Últimos 10 nucleótidos: g t c g g c t c a c t g t c g g t g t g t c t a a a a t t t
## -----
## PERÚ - Ranking: 18
## Tamaño: 29324
##
##   a   c   g   t
## 8556 5218 5607 9208
## Porcentaje GC: 0.3786421
```



## PERÚ



## Secuencia contrasentido

## Primeros 10 nucleótidos: t a a t t t c t a a c g a t a c a c t c t a a t t t c a a t

## Últimos 10 nucleótidos: a g g t t t c g g t g c a t g c t c g t g c a g c g c t t g

## -----

## PAÍSES BAJOS - Ranking: 19

## Tamaño: 29513

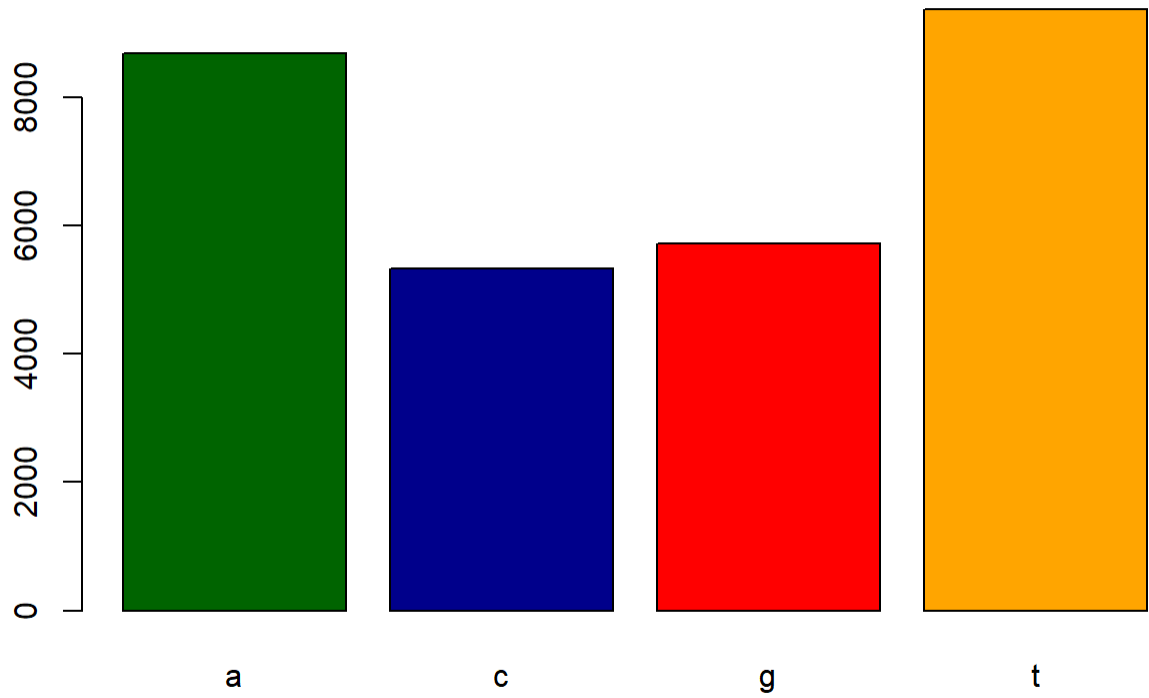
##

## a c g t

## 8680 5328 5715 9369

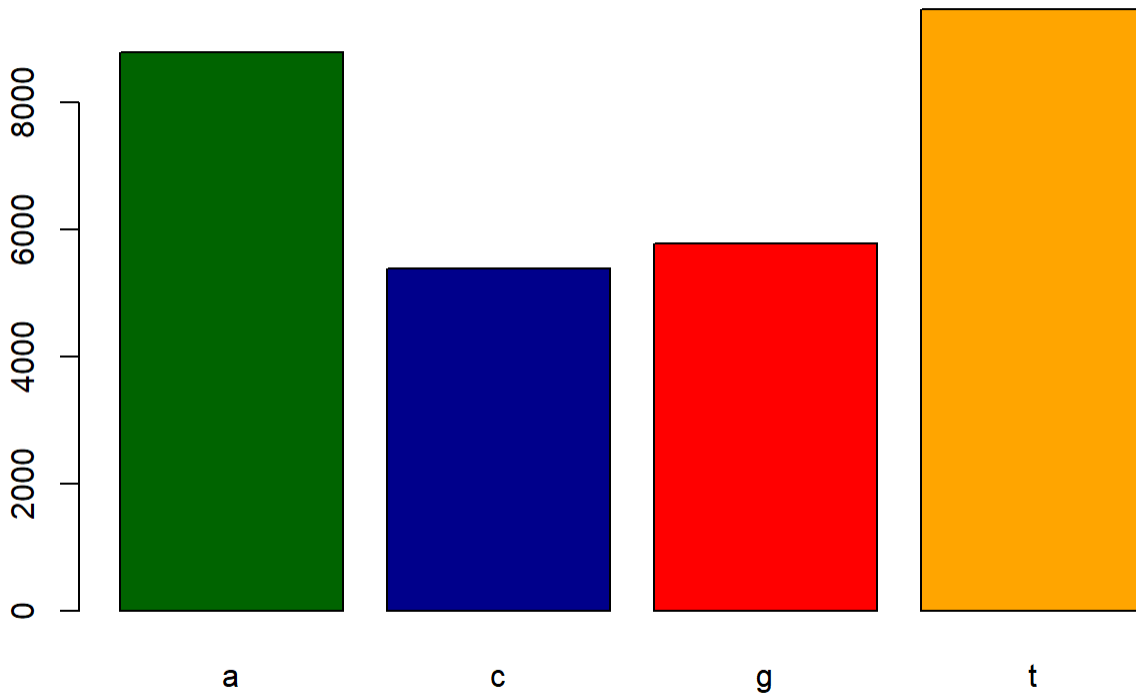
## Porcentaje GC: 0.3795889

PAÍSES BAJOS



```
## Secuencia contrasentido
## Primeros 10 nucleótidos: g a g g g a t t a c a a t g t g t g a c t a a t t t c t a a
## Últimos 10 nucleótidos: t t c g t c g g a c g t c t t c t a t c t g c t c a a t g a
## -----
## CHILE - Ranking: 20
## Tamaño: 29501
##
##   a   c   g   t
## 8786 5386 5776 9466
## Porcentaje GC: 0.3794792
```

## CHILE



```
## Secuencia contrasentido
## Primeros 10 nucleótidos: a g a a a g t t c a g g a g g g a t t a c a a t g t g t g a
## Últimos 10 nucleótidos: a c a c g a c t a c t a g c c g a c g t t g t g c c t g c t
## -----
```

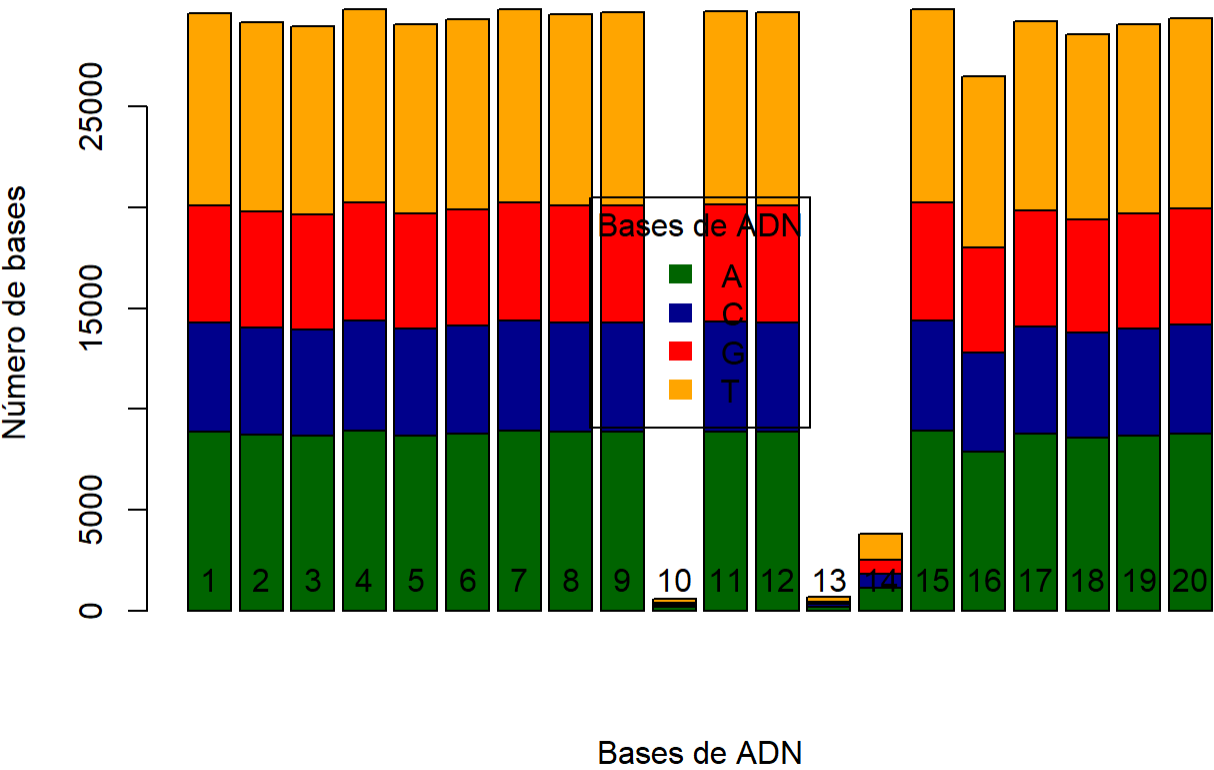
```
matriz_composicion <- do.call(cbind, lista_composicion)

bp <- barplot(matriz_composicion, col = c("darkgreen", "darkblue", "red", "orange"),
  main="Composición de bases de ADN para todas las variantes del virus",
  xlab="Bases de ADN", ylab="Número de bases")

text(bp, 1:ncol(matriz_composicion), labels=1:ncol(matriz_composicion), pos=3)

legend("center", c("A", "C", "G", "T"), fill=c("darkgreen", "darkblue", "red", "orange"), title="Bases de A
DN",bg="transparent", border=NA)
```

Composición de bases de ADN para todas las variantes del virus



Analizar las secuencias de SARS-CoV-2 reportadas en los 20 países con más casos reportados. Y puedes tratar de responder a la pregunta: ¿Son muy diferentes las variantes entre cada país? ¿Es diferente el SARS-CoV-2 entre las diferentes poblaciones: Asiática, Hispana, Europea o Africana?

De acuerdo con lo obtenido del análisis del SARS-CoV-2 por medio de gráficas y recopilación de datos se pueden llevar a cabo ciertas interpretaciones. Primero que nada, a primera instancia las variantes parecen no ser muy diferentes en cada país con respecto a su distribución de nucleótidos. Sin embargo, es notable un aumento de guaninas en los 10 primeros lugares de países con más casos reportados. De igual forma, a simple vista se nota una mayor longitud por parte de estas variaciones. Logrando una interpretación de que entre más guaninas llegar a tener y más larga llega a ser la secuencia, más contagiosa puede ser la variante.

Además, llama la atención como tres de los países con más contagios (México, Argentina e Irán) no cuentan con el genoma completo y actualizado de su variante de SARS-CoV-2, dando indicios de que en estos países la infraestructura y situación no se encontró prevenida para una situación como esta, correlacionando su acceso a la información biológica con el número de casos reportados. Asimismo, gracias al árbol filogenético se puede observar como por más que muchos países se encuentran cerca geográficamente (asiáticos, hispanos, europeos, africanos), sus variantes no llegan a tener relaciones o comparación alguna. Dando pie a la hipótesis de que el

virus se adaptó a las condiciones de cada país y los países con una situación de salud similar compartieron similitudes con la evolución de su variante. Esto apoyado por Global Health Service Monitor para el año 2021 donde se indica a países como Perú, Polonia, Colombia, México y Chile como países con peor sistema de salud en el mundo (IPSOS 2021). Sugiriendo que los lugares con sistemas de salud deficientes o condiciones desfavorables para el virus fueron donde la situación se volvió más riesgosa e inestable.

En conclusión, el análisis de los datos y gráficas del SARS-CoV-2 muestra que la presencia de guaninas y la longitud de la secuencia podrían estar relacionados con la contagiosidad de las variantes. Además, la disponibilidad de información biológica y la infraestructura de salud de cada país parecen ser factores importantes en la cantidad de casos reportados y la peligrosidad del virus en cada región. La adaptación del virus a las condiciones de cada país podría explicar por qué las variantes no tienen relaciones o comparaciones directas entre países cercanos geográficamente. Estos hallazgos subrayan la importancia de una buena infraestructura de salud y acceso a información biológica para prevenir y controlar la propagación del virus.

**Referencias:**

- Organización Mundial de la Salud. (2023). Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. Recuperado de <https://www.who.int/emergencies/disease/novel-coronavirus-2019/situation-reports/>
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270-273. <https://doi.org/10.1038/s41586-020-2012-7>
- Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, 92(4), 418-423. <https://doi.org/10.1002/jmv.25681>
- Li, X., Zai, J., & Wang, X. (2020). Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Science China Life Sciences*, 63(4), 467-470. <https://doi.org/10.1007/s11427-020-1661-4>
- Ipsos. (2021). Global Health Service Monitor 2021. Recuperado el 7 de mayo de 2023, de <https://www.ipsos.com/es-mx/global-health-service-monitor-2021>