

# Proyecto integrador

Cyrce Danae Salinas

May, 6, 2024

```
packageVersion("ape")
packageVersion("phangorn")
packageVersion("phytools")
packageVersion("geiger")
```

Realiza lo siguiente en un script de R y código:

De las siguientes tres opciones, selecciona una opción para investigar: [10 pts] -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?

1. Calcula la longitud de las secuencias que incluyas. [5 pts]

sizes

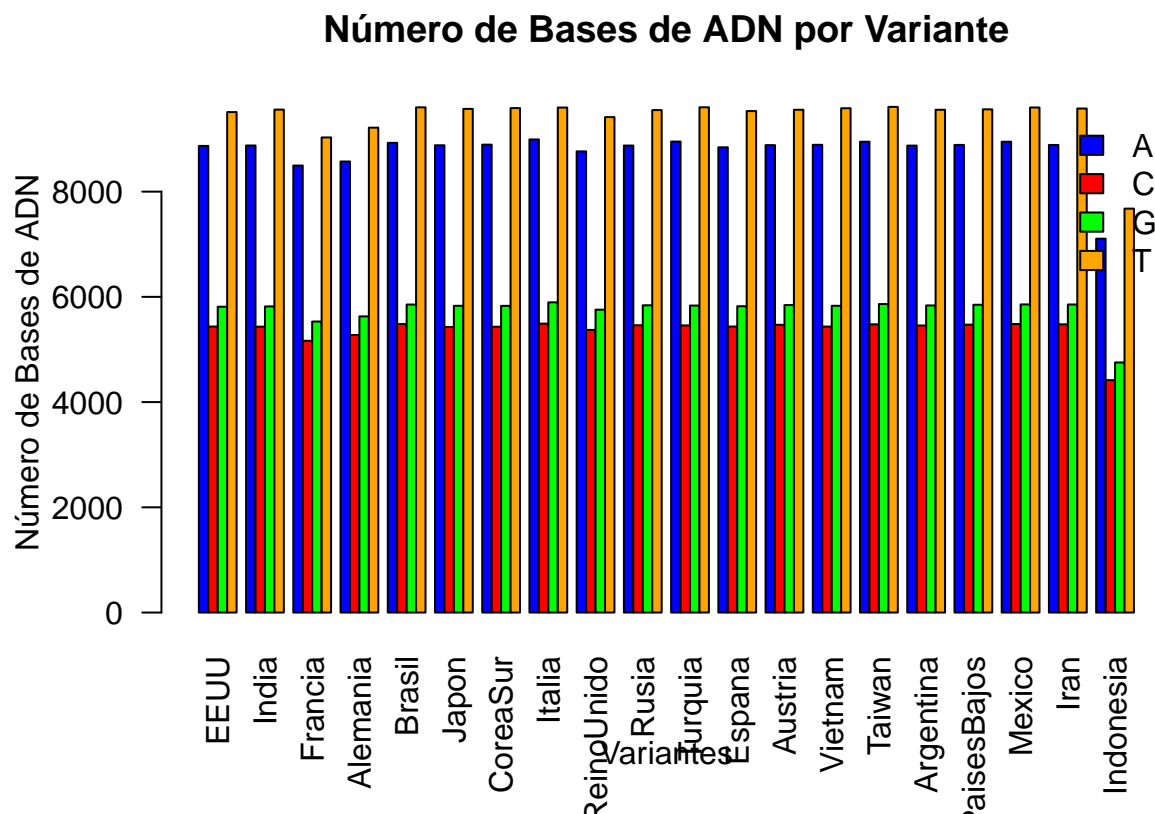
|    |          |           |            |           |             |        |
|----|----------|-----------|------------|-----------|-------------|--------|
| ## | EEUU     | India     | Francia    | Alemania  | Brasil      | Japon  |
| ## | 179542   | 178490    | 178845     | 179055    | 179560      | 178628 |
| ## | CoreaSur | Italia    | ReinoUnido | Rusia     | Turquia     | Espana |
| ## | 178815   | 180227    | 179031     | 178724    | 179446      | 179199 |
| ## | Austria  | Vietnam   | Taiwan     | Argentina | PaísesBajos | Mexico |
| ## | 178887   | 178803    | 179770     | 178706    | 179007      | 179716 |
| ## | Iran     | Indonesia |            |           |             |        |
| ## | 179163   | 179055    |            |           |             |        |

variantes\_seq\_no\_alineadas

```
## DNASTringSet object of length 20:
##      width seq                                     names
## [1] 29863 CCAACTTTCGATCTCTTGTAGAT...AAAAAAAAAAAAAAAAAAAAA EEUU
## [2] 29688 AACAAACCAACCAACTTTTGATC...TGCTAGGGAGAGCTGCCTATAT India
## [3] 29747 ACTTTCGATCTCTTGTAGATCTG...TTTAGTAGTGCTATCCCCATG Francia
## [4] 29782 AGATCTGTTCTCTAAACGAACCT...ATTAANTTTAGTAGTGCTATNC Alemania
## [5] 29866 CCTATAACATCACAGGTAACAAA...GCTTCTTAGGAGAAGCACAAAA Brasil
## ...    ...
## [16] 29724 GATCTGTTCTCTAAACGAACCTT...TGTACAGTGAACAATGCTAGGG Argentina
## [17] 29774 CGATCTCTTGTAGATCTGTTCTC...GTGTAAATTAATTTTAGTAGT PaísesBajos
## [18] 29892 ATACCTTCCCAGGTAACAAACCA...AAAAAAAAAAAAAAAAAAAAA Mexico
## [19] 29800 TCGATCTCTTGTAGATCTGTTCT...TTTAGTAGTGCTATCCCCATGT Iran
## [20] 29782 AGATCTGTTCTCTAAACGAACCT...ATTAATTTTAGTAGTGCTATCC Indonesia
```

2. Crea una primera gráfica donde se comparen el número de bases de ADN que componen todas las variantes del virus. [7 pts]

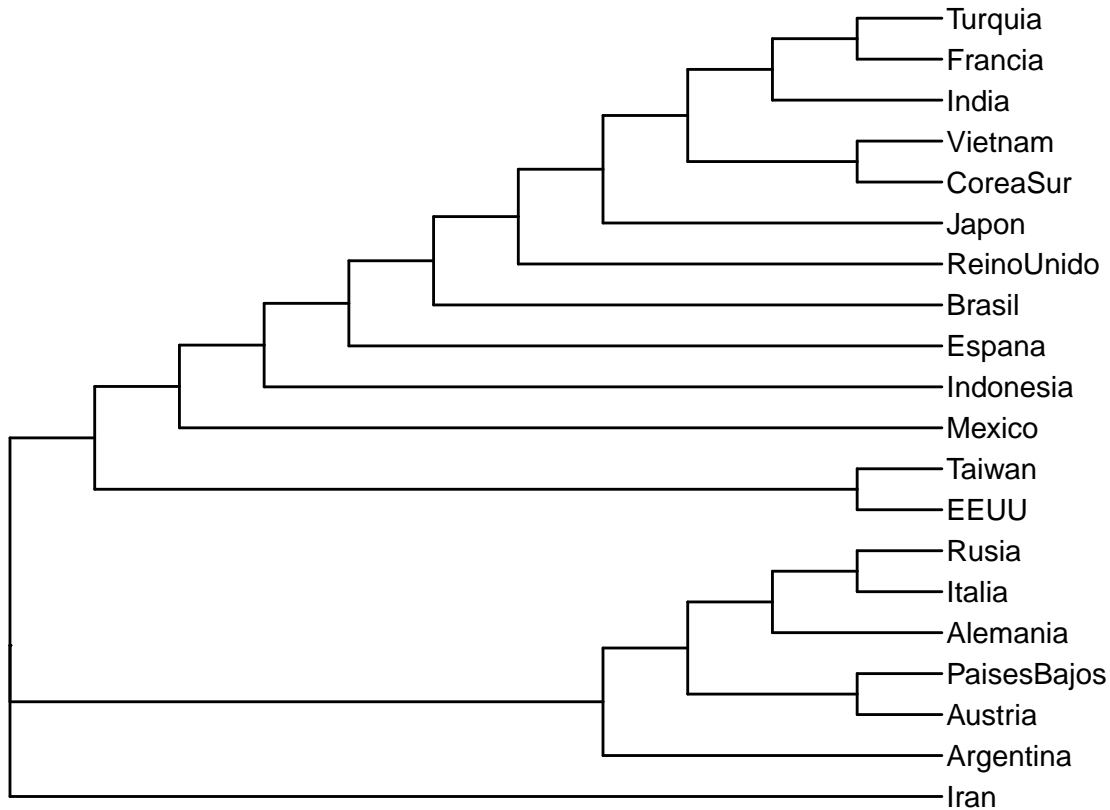
```
barplot(t(base_matrix), beside = TRUE,
      main = "Número de Bases de ADN por Variante",
      xlab = "Variantes",
      ylab = "Número de Bases de ADN",
      col = c("blue", "red", "green", "orange"),
      legend.text = base_names,
      args.legend = list(x = "topright", bty = "n"),
      srt = 45,
      las=2)
```



En este gráfico se nos muestra que, en general, las variantes tienen en promedio el mismo número de bases (a excepción de la variante encontrada en Indonesia). Esto debido a que estamos estudiando el mismo tipo de virus, el SARS-CoV-2. Al comparar la cantidad de bases que componen a los virus, podemos decir que estas secuencias tienen un bajo contenido de citosina y guanina, por lo que las hebras pueden separarse más fácilmente, lo que hace fácil la replicación del virus porque no requieren de tanta energía para la separación de hebras. La visualización de las bases que componen secuencias, ayuda al entendimiento de la similitud entre variantes y permite un mayor entendimiento de estas.

3. Crea una segunda gráfica donde se muestre el árbol filogenético obtenido de las secuencias que se seleccionaron para estudiar. [8 pts]

```
ggtree(varian_plot_filogenetico, branch.length='none') + geom_tiplab(text=element_text(size=3)) + xlim(
```



4. Agrega una interpretación escrita de tus gráficas (¿qué es lo que observas?) y tus conclusiones (¿qué implicaciones tienen tus observaciones?) según el caso de estudio que seleccionaste. No olvides sustentar tus argumentos con las lecturas que realizaste. [20 pts]

En este árbol, se ilustra como todos los virus son descendientes de un mismo ancestro. Puedo ver que en las secuencias que Italia y Alemania están en la misma rama, por lo que puedo concluir que estas son similares. Además, un sustento para mi argumento es que están muy cerca un país de otro, geográficamente hablando. Mismo caso en cuanto a Vietnam y Corea del Sur; sin embargo, en este caso no están tan cerca, se encuentran en una misma rama y sus secuencias presentan similitudes.

Los filogramas permiten visualizar las relaciones evolutivas entre organismos; sin embargo, es importante recalcar que estos son únicamente una hipótesis. Hay que considerar también factores de morfología, registros paleontológicos, etc.

Recuerda incluir tus datos personales y las referencias bibliográficas de tu investigación. 5. Sube tu entregable a la plataforma en el formato indicado. Recuerda incluir la liga de tu video en el script. Liga del video:

```
print("https://youtu.be/b4KTy6hvp6c?si=cmRK376QJeZLH1MQ")
```

```
## [1] "https://youtu.be/b4KTy6hvp6c?si=cmRK376QJeZLH1MQ"
```

Referencias de paquetes

```
citation("dplyr")
```

```
## To cite package 'dplyr' in publications use:
##
## Wickham H, François R, Henry L, Müller K, Vaughan D (2023). _dplyr: A
## Grammar of Data Manipulation_. R package version 1.1.4,
## <https://CRAN.R-project.org/package=dplyr>.
##
## A BibTeX entry for LaTeX users is
##
## @Manual{,
##   title = {dplyr: A Grammar of Data Manipulation},
##   author = {Hadley Wickham and Romain François and Lionel Henry and Kirill Müller and Davis Vaughan},
##   year = {2023},
##   note = {R package version 1.1.4},
##   url = {https://CRAN.R-project.org/package=dplyr},
## }
```

```
citation("ggplot2")
```

```
## To cite ggplot2 in publications, please use
##
## H. Wickham. ggplot2: Elegant Graphics for Data Analysis.
## Springer-Verlag New York, 2016.
##
## A BibTeX entry for LaTeX users is
##
## @Book{,
##   author = {Hadley Wickham},
##   title = {ggplot2: Elegant Graphics for Data Analysis},
##   publisher = {Springer-Verlag New York},
##   year = {2016},
##   isbn = {978-3-319-24277-4},
##   url = {https://ggplot2.tidyverse.org},
## }
```

```
citation("seqinr")
```

```
## To cite seqinr in publications use:
##
## Charif, D. and Lobry, J.R. (2007)
##
## A BibTeX entry for LaTeX users is
##
## @InCollection{,
##   author = {D. Charif and J.R. Lobry},
##   title = {Seqin{R} 1.0-2: a contributed package to the {R} project for statistical computing devot},
##   booktitle = {Structural approaches to sequence evolution: Molecules, networks, populations},
##   year = {2007},
##   editor = {U. Bastolla and M. Porto and H.E. Roman and M. Vendruscolo},
##   series = {Biological and Medical Physics, Biomedical Engineering},
## }
```

```
##   pages = {207-232},
##   address = {New York},
##   publisher = {Springer Verlag},
##   note = {{ISBN :} 978-3-540-35305-8},
## }
```

```
citation("adegenet")
```

```
## To cite the adegenet package:
##
##   Jombart, T. (2008) adegenet: a R package for the multivariate
##   analysis of genetic markers. Bioinformatics 24: 1403-1405. doi:
##   10.1093/bioinformatics/btn129
##
##   Jombart T. and Ahmed I. (2011) adegenet 1.3-1: new tools for the
##   analysis of genome-wide SNP data. Bioinformatics. doi:
##   10.1093/bioinformatics/btr521
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

```
citation("ape")
```

```
## To cite ape in a publication please use:
##
##   Paradis E, Schliep K (2019). "ape 5.0: an environment for modern
##   phylogenetics and evolutionary analyses in R." Bioinformatics,
##   35, 526-528. doi:10.1093/bioinformatics/bty633
##   <https://doi.org/10.1093/bioinformatics/bty633>.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}},
##     author = {Emmanuel Paradis and Klaus Schliep},
##     journal = {Bioinformatics},
##     year = {2019},
##     volume = {35},
##     pages = {526-528},
##     doi = {10.1093/bioinformatics/bty633},
##   }
##
## ape is evolving quickly, so you may want to cite its version number
## (found with 'library(help = ape)' or 'packageVersion("ape")').
```

```
citation("phangorn")
```

```
## Use 2011 to cite phangorn in a publication; 2017 for plotting
## phylogenetic networks. As phangorn is evolving quickly, you may want to
## cite also its version number (phangorn 2.11.1).
##
```

```
## Schliep K.P. 2011. phangorn: phylogenetic analysis in R.
## Bioinformatics, 27(4) 592-593
##
## Schliep, K., Potts, A. J., Morrison, D. A., Grimm, G. W. (2017),
## Intertwining phylogenetic trees and networks. Methods in Ecology and
## Evolution, 8: 1212--1220. doi: 10.1111/2041-210X.12760
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

```
citation("phytools")
```

```
## To cite phytools in a publication please use:
##
## Revell, L. J. (2024) phytools 2.0: an updated R ecosystem for
## phylogenetic comparative methods (and other things). PeerJ, 12,
## e16505.
##
## A BibTeX entry for LaTeX users is
##
## @Article{,
##   title = {{p}hytools 2.0: an updated {R} ecosystem for phylogenetic comparative methods (and other
##   author = {Liam J. Revell},
##   journal = {PeerJ},
##   year = {2024},
##   volume = {12},
##   pages = {e16505},
##   doi = {10.7717/peerj.16505},
## }
##
## As phytools is continually evolving you may also want to cite its
## version number (found with 'library(help = phytools)' or
## 'packageVersion("phytools")').
```

```
citation("geiger")
```

```
## To cite medusa, auteur, MECCA, or geiger in a publication use:
##
## medusa
##
## Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L.
## Rabosky, G. Carnevale, and L. J. Harmon (2009) Nine exceptional
## radiations plus high turnover explain species diversity in jawed
## vertebrates. PNAS, 106, 13410-13414.
##
## auteur
##
## Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon
## (2011) A novel comparative method for identifying shifts in the rate
## of character evolution on trees. Evolution, 65, 3578-3589.
##
## MECCA
```

```

##
## Slater, G. J., L. J. Harmon, D. Wegmann, P. Joyce, L. J. Revell, and
## M. E. Alfaro (2012) Fitting models of continuous trait evolution to
## incompletely sampled comparative data using approximate Bayesian
## computation. Evolution, 66, 752-762.
##
## geiger-orig
##
## Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger
## (2008) GEIGER: investigating evolutionary radiations. Bioinformatics,
## 24, 129-131.
##
## geiger
##
## Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C.
## Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon {2014} geiger
## v2.0: an expanded suite of methods for fitting macroevolutionary
## models to phylogenetic trees. Bioinformatics, 30, 2216-2218.
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.

```