

Evidence 2 | Integrative Project

BT1013.201 Computational Biology Analysis (Group 201)

Carlos Alejandro Galicia Silva A01709890

2024-12-23

From the following three options, select one to investigate:

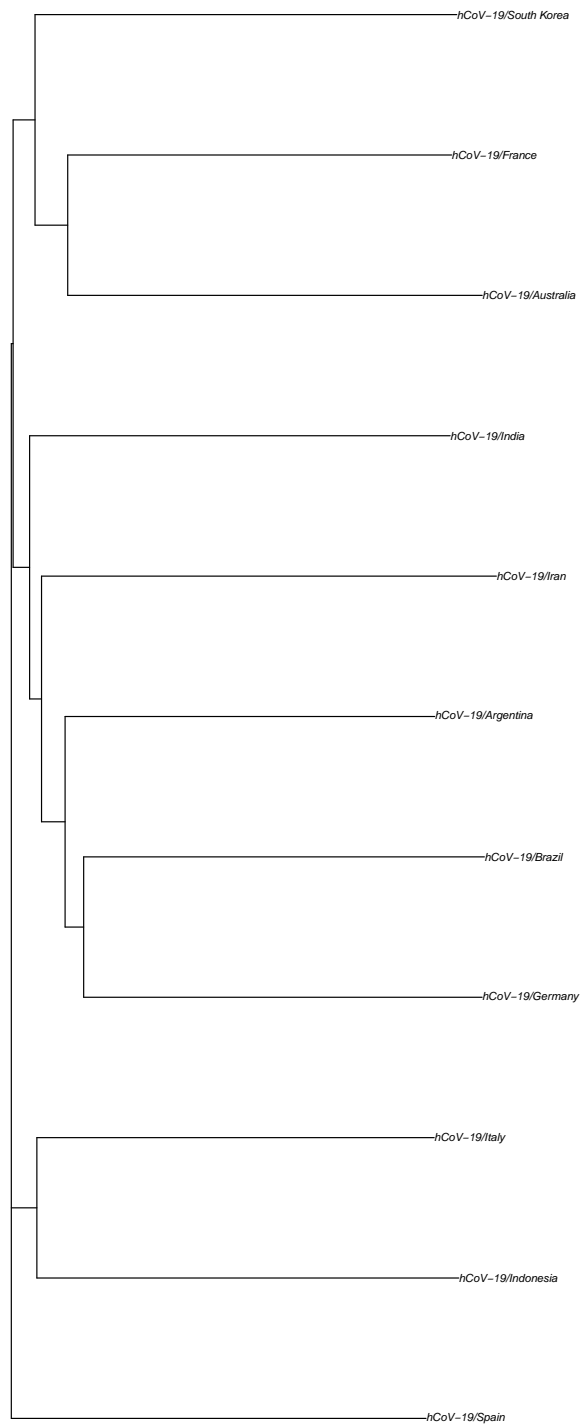
Analyze the SARS-CoV-2 sequences reported in the 20 countries with the highest reported cases. You may try to answer the question: Are the variants very different between countries? Is SARS-CoV-2 different among Asian, Hispanic, European, or African populations?

```
##
## Converting FASTA alignment into a DNABin object...
##
##
## Finding the size of a single genome...
##
##
## genome size is: 29,869 nucleotides
##
## ( 375 lines per genome )
##
## Importing sequences...
## .....

##
## Forming final object...
##
## ...done.
```

Add a global hierarchical analysis obtained from the selected sequences:

Tree of type JC69



In this phylogenetic tree containing the variants from the countries with the highest number of cases world-wide, one can observe the relationships and genetic proximity of each variant based on their genetic similarity. Variants similar to each other include, but are not limited to:

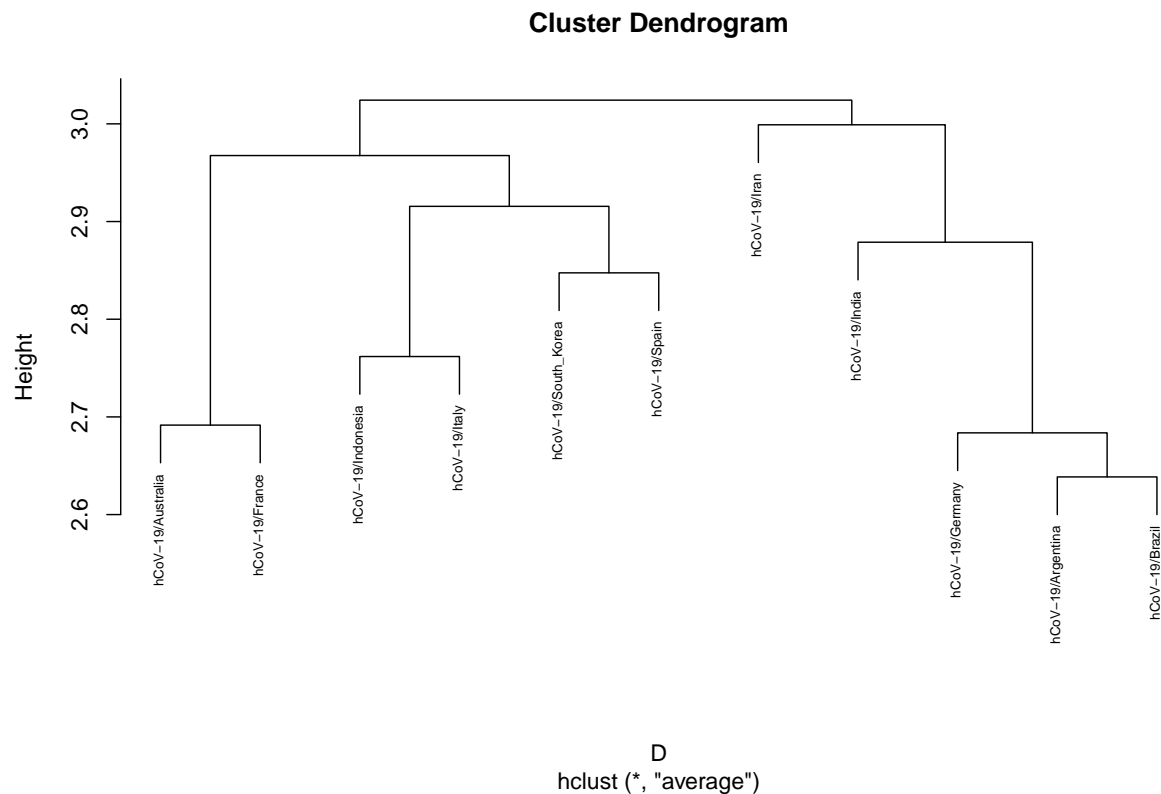
Variants from Iran and India.

Variants from Italy and France.

Variants from Australia and South Korea.

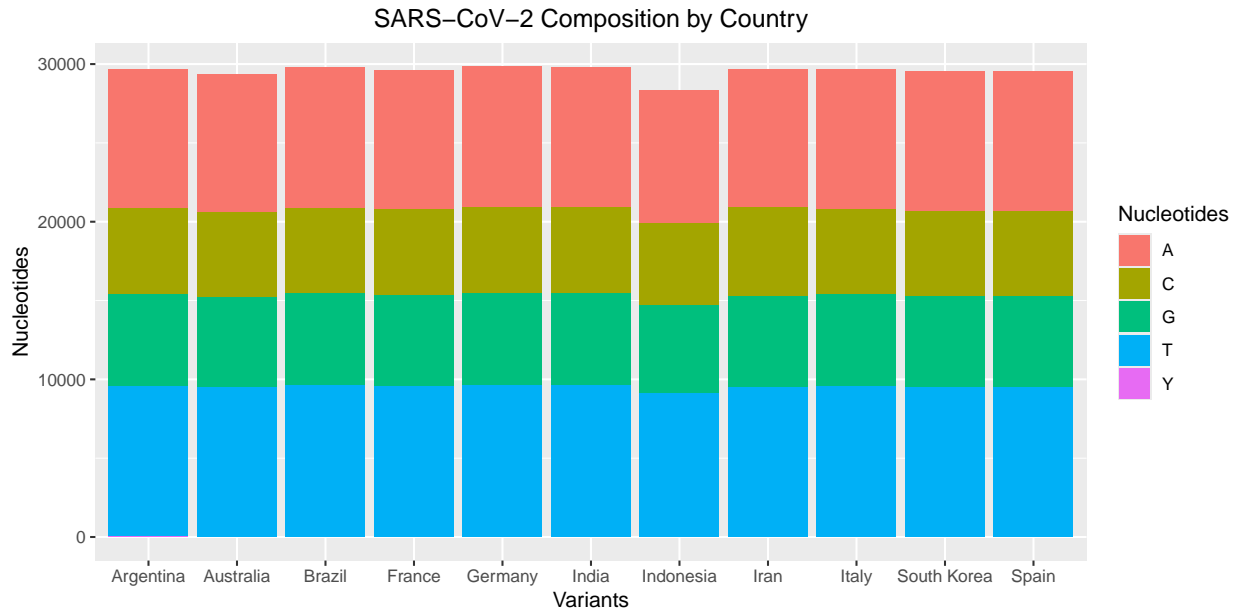
The group of variants from Brazil with Australia and South Korea.

Additionally, the genetic sequences most distant from other groups of country sequences are the variants from Argentina and Germany.



In the dendrogram above, two main groups of sequences can be observed. The first consists of SARS-CoV-2 viruses with a high level of similarity, such as those from Italy, Spain, Brazil, and Australia and South Korea. On the other hand, the second group splits into two subsets: one comprising France, Argentina, and Iran, and another with Indonesia, Germany, and India. This latter subset shows a lower degree of similarity, as indicated by the vertical distance between the country sequences.

Create a single graph comparing the number of DNA bases composing all the virus variants:



In this graph, it is observed that most of the SARS-CoV-2 virus variants from different countries are quite similar in terms of nucleotide count and proportion. At first glance, the variant from Indonesia stands out as the most distant or different compositionally from the other variants shown.

Calculate the length of the included sequences:

```
## [1] "Sequence length:Germany:29869"
## [1] "Sequence length:Argentina:29711"
## [1] "Sequence length:Australia:29378"
## [1] "Sequence length:Brazil:29760"
## [1] "Sequence length:South Korea:29506"
## [1] "Sequence length:France:29632"
## [1] "Sequence length:India:29839"
## [1] "Sequence length:Indonesia:28357"
## [1] "Sequence length:Iran:29699"
## [1] "Sequence length:Spain:29504"
## [1] "Sequence length:Italy:29711"
```

Add a written interpretation of your graphs and conclusions based on the case study you selected. Do not forget to support your arguments with your readings:

Based on the analysis and results obtained from the phylogenetic tree, dendrogram, and the graph of DNA base counts for the variants from the countries with the highest number of SARS-CoV-2 infections, several conclusions can be drawn. One conclusion is that the proportion of bases in each variant is similar, indicating that the level of variation of the virus among countries is relatively low, with the notable exception of the variant from Indonesia. Moreover, upon examining the hierarchical analyses, it can be concluded that SARS-CoV-2 variants from some countries originated or descended from variants of other countries. An example of this is the hCoV-19 variant from South Korea, which is the latest descendant of the hCoV-19 variant in a chain of variants that, in turn, originated from the first virus in Wuhan, China.

References

- COVID-19 Map - Johns Hopkins Coronavirus Resource Center. (n.d.). Johns Hopkins Coronavirus Resource Center. Retrieved from: <https://coronavirus.jhu.edu/map.html>
- SARS-CoV-2 Variants. (2023, April 11). Retrieved from: <https://www.idsociety.org/covid-19-real-time-learning-network/emerging-variants/emerging-covid-19-variants/#:~:text=Ancestral%20strain%3A%20The%20original%20form,%E2%80%9D%20or%20%E2%80%9CWuhan%E2%80%9D%20strain.>
- Aleem A, Akbar Samad AB, Vaqar S. Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). 2023 Feb 5. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. PMID: 34033342. Retrieved from: <https://pubmed.ncbi.nlm.nih.gov/34033342/> <https://coronavirus.gob.mx/variantes-covid-19/>
- Rabaan AA, Al-Ahmed SH, Haque S, Sah R, Tiwari R, Malik YS, Dhama K, Yattoo MI, Bonilla-Aldana DK, Rodriguez-Morales AJ. SARS-CoV-2, SARS-CoV, and MERS-COV: A comparative overview. Infez Med. 2020 Ahead Of Print Jun 1;28(2):174-184. PMID: 32275259. Retrieved from: https://www.infezmed.it/media/journal/Vol_28_2_2020_7.pdf
- Coronavirus: número de casos confirmados por país en 2023 | Statista. (2023, May 5). Statista. <https://es.statista.com/estadisticas/1091192/paises-afectados-por-el-coronavirus-de-wuhan-segun-los-casos-confirmados/>
- Katella, K. (2023). Omicron, Delta, Alpha, and More: What To Know About the Coronavirus Variants. Yale Medicine. <https://www.yalemedicine.org/news/covid-19-variants-of-concern-omicron#:~:text=Alpha%20>