

The *SARS-CoV-2* early infections phylogeny as an example of the misinterpreting of phylogenies in epidemiology

Dear Editor,

Taking into consideration the underlying concept of molecular epidemiology as the axis of epidemiology, uniting insights at the molecular level and the understanding of diseases at the population and geographic levels, it is not surprising that the use of phylogenetic approaches to understand and to determine the geographic origin and traceability of animal cases, including humans, has increased over the years (1). However, methodological misconceptions, arisen because of an indiscriminate use of phylogenetic approaches, in many cases have led to wrong conclusions regarding the ancestors-descendants relationships that can be extrapolated to geographic scales (2).

These types of issues arise when, instead of characterizing monophyletic clades, paraphyletic or polytomies are obtained. In these cases, no ancestry can be inferred within clades and therefore no information about the geographic origin can be assessed (3). For this reason, it is important to take actions to prevent reaching premature or incorrect conclusions from phylogenetic analysis when no monophyly is obtained.

As an example, we evaluated if the geographic origins of the first human cases of COVID-19 from American countries could be recapitulated relative to the official announcements, using a repertoire of 1,622 *SARS-CoV-2* viral genomes available from GenBank as of August 20th, 2020. All analyzes were performed using the same methods reported by Colunga-Salas and Hernández-Canchola (4).

From our analysis, we were not able to reach definitive conclusions since only polytomies were obtained (Figure 1; for complete size phylogenetic reconstruction, see: <https://doi.org/10.5281/zenodo.6533857>). For

Historial del artículo

Recibido: 1 jun 2022

Aceptado: 21 jun 2022

Disponible en línea: 1 sep 2022

Copyright © 2022 por autores y Revista Biomédica.

Este trabajo está licenciado bajo las atribuciones de la *Creative Commons* (CC BY).

<http://creativecommons.org/licenses/by/4.0/>

*Autor para correspondencia:

Pablo Colunga-Salas, Instituto de Biotecnología y Ecología Aplicada, Universidad Veracruzana, Av. de las Culturas Veracruzanas 101, Col. Emiliano Zapata, 91090, Xalapa-Enríquez, Veracruz, México.

E-mail: pcolunga@uv.mx

<https://revistabiomedica.mx>

example, in the case of South America, particularly in Brazil, the official information stated that patient zero corresponded to a 32-year-old male traveling from Italy on February 26th, 2020 (<https://virological.org/t/first-cases-of-coronavirus-disease-covid-19-in-brazil-south-america-2-genomes-3rd-march-2020/409>). However, in our phylogenetic reconstruction the next sequences are from Taiwan, Great Britain, the US and Sri Lanka (Figure 1-A). In the case of Chile, the first case was a 33-year-old individual returning from Singapore (<https://bbc.in/39S9pr5>), and our phylogenetic analysis shows a polytomy where the Chilean sequence is placed near sequences from China, Hong-Kong and the US (Figure 1-B). Analysis of a binational case, including a 19-year-old female (first confirmed case) from Colombia (<https://www.minsalud.gov.co/English/Paginas/Colombia-Confirms-Its-First-Case-of-COVID-19.aspx>) and four individuals from Uruguay (also the first confirmed cases) traveling back from Milan, Italy, showed that sequences from Colombia and Uruguay are placed in what seems to be a

“monophyletic clade” -which is clearly a polytomy when the complete phylogenetic tree is observed-, near sequences from Spain, Germany, the US, Australia, Greece and Jordan (Figure 1-C). In this last example, some correspondence can be observed since they seem to be part of the same infection route. However, no conclusions can be drawn from the phylogeny. In the case of Mexico, a 35-year-old male returning from Italy tested positive by rt-PCR on February 27th, 2020 (<https://www.gob.mx/salud/prensa/077-se-confirma-en-mexico-caso-importado-de-coronavirus-covid-19>), and our phylogenetic reconstruction placed the sequences near to those from France and the US (Figure 1-D). The only case where the phylogenetic results and the official information were in concordance was from Puerto Rico, where the official information of the first cases was from Italian tourists that had visited the US the week prior to the official announcement (<http://hrl.d.us/3L1KIVV>), and viral sequences grouped together with sequences from the US (Figure 1-E).

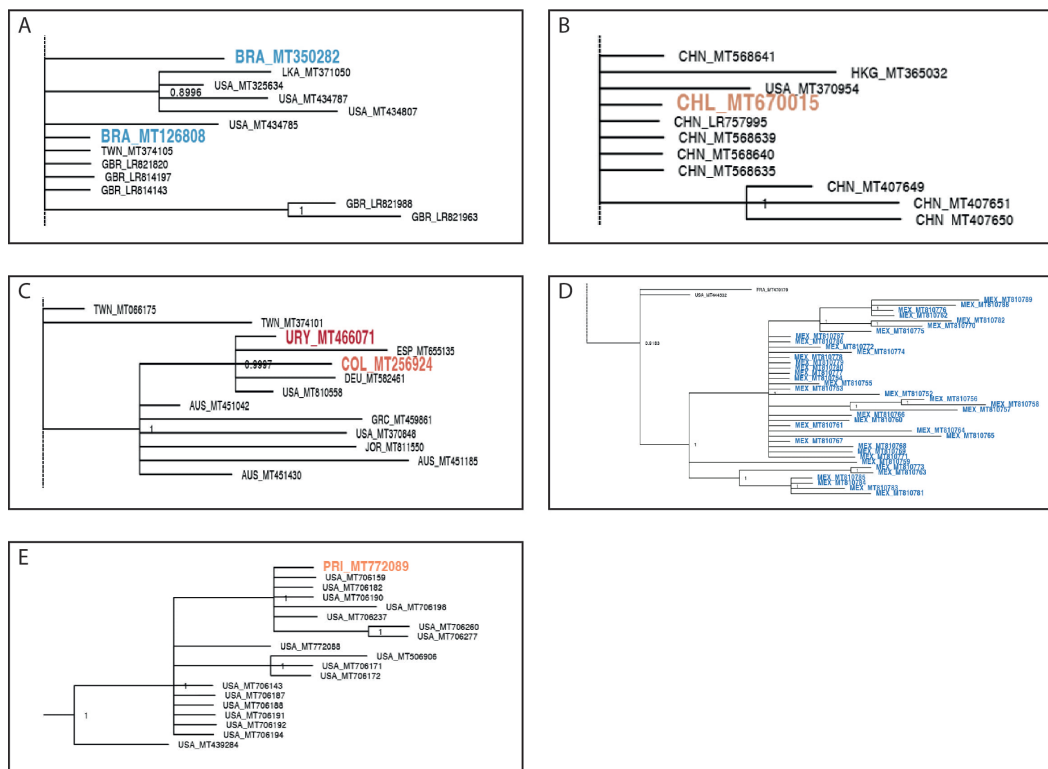


Figure 1. Partial phylogenetic reconstruction of 1,622 SARS-CoV-2 viral genomes from America as of August 20th, 2020. A) BRA – Brazil, B) CHL – Chile, C) COL – Colombia and URY – Uruguay, D) MEX – Mexico, E) PRI – Puerto Rico. For complete phylogenetic reconstruction: <https://doi.org/10.5281/zenodo.6533857>

Taken together, in the case of the *SARS-CoV-2* pandemic and for similar outbreaks where epidemiological information could be scarce, fragmented or in many cases politicized, it is fundamental to have detailed information, including the origin and location of the patient zero, as well as epidemiological data on the infection source from the location which the patient become. We believe that it is extremely important to be cautious with the use of phylogenies when geographic epidemiology studies are the goal.

ACKNOWLEDGMENTS

The sponsorship for the phylogenetic analysis was supported by grants CONACyT 6682 and PAPIIT AG-201221.

REFERENCES

1. Wang X, King Jordan I, Mayer LW. A phylogenetic perspective on molecular epidemiology. Second Edi. In: Tang Y-W, Liu D, Schwartzman J, Sussman M, Poxton I, editors. *Molecular Medical Microbiology*. 2nd ed. Elsevier Ltd; 2014. pp. 517–536.
2. Lam TT-Y, Hon C-C, Tang JW. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. *Crit Rev Clin Lab Sci*. 2010;47: 5–49. <https://doi.org/10.3109/10408361003633318>
3. Henning W. Phylogenetic systematics. *Annu Rev Entomol*. 1965;10: 97–116. <https://doi.org/10.1146/annurev.en.10.010165.000525>
4. Colunga-Salas P, Hernández-Canchola G. Bats and humans during the *SARS-CoV-2* outbreak: The case of bat-coronaviruses from Mexico. *Transbound Emerg Dis*. 2021;68: 987–992. <https://doi.org/10.1111/tbed.13751>

Pablo Colunga-Salas

Instituto de Biotecnología y Ecología Aplicada, Universidad Veracruzana, Xalapa-Enríquez, Veracruz, México. Centro de Medicina Tropical, División de Investigación, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México.

Daniela Segura-Trejo

Centro de Medicina Tropical, División de Investigación, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México.

Carlos Ismar Miranda-Caballero

Centro de Medicina Tropical, División de Investigación, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México.

Sokani Sánchez-Montes

Centro de Medicina Tropical, División de Investigación, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México. Facultad de Ciencias Biológicas y Agropecuarias región Tuxpan, Tuxpam de Rodríguez Cano, Veracruz, México.

Mario Arteaga-Vázquez

Instituto de Biotecnología y Ecología Aplicada, Universidad Veracruzana, Xalapa-Enríquez, Veracruz, México.

Ingeborg Becker

Centro de Medicina Tropical, División de Investigación, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México.