Le Infezioni in Medicina, n. 2, 139-142, 2020

Genomic Epidemiology and its importance in the study of the COVID-19 pandemic

Alfonso J. Rodríguez-Morales[1,2,3,4], Graciela J. Balbin-Ramon[3,4,5], Ali A. Rabaan[6], Ranjit Sah[7], Kuldeep Dhama[8], Alberto Paniz-Mondolfi[3,9,10,11,12], Pasquale Pagliano[13], Silvano Esposito[13]

[1]Public Health and Infection Research Group, Faculty of Health Sciences, Universidad Tecnológica de Pereira, Pereira, Colombia;
[2]Grupo de Investigación Biomedicina, Faculty of Medicine, Fundación Universitaria Autónoma de las Americas, Pereira, Risaralda, Colombia;
[3]Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19), Pereira, Risaralda, Colombia;
[4]Master in Clinical Epidemiology and Biostatistics, Universidad Cientifica del Sur, Lima, Peru;
[5]Hospital de Emergencias Jose Casimiro Ulloa, Lima, Peru;
[6]Molecular Diagnostic Laboratory, Johns Hopkins Aramco Healthcare, Dhahran, Saudi Arabia;
[7]Department of Microbiology, Tribhuvan University Teaching Hospital, Institute of Medicine, Kathmandu, Nepal;
[8]Division of Pathology, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India;
[9]Laboratory of Medical Microbiology, Department of Pathology, Molecular and Cell-based Medicine, The Mount Sinai Hospital-Icahn School of Medicine at Mount Sinai, New York, USA;
[10]Laboratorio de Señalización Celular y Bioquímica de Parásitos, Instituto de Estudios Avanzados (IDEA), Caracas, Caracas, Venezuela;
[11]Academia Nacional de Medicina, Caracas, Venezuela;
[12]Instituto de Investigaciones Biomedicas IDB / Incubadora Venezolana de la Ciencia, Cabudare, Edo. Lara, Venezuela;
[13]Department of Infectious Diseases, University of Salerno, Salerno, Italy

Over the course of three months (December 2019-March 2020) the pandemic of the Coronavirus Disease 2019 (COVID-19), caused by a zoonotic virus, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), has raised multiple concerns and triggered an unprecedented and deep impact across multiple areas of biomedical research, especially the disciplines of molecular biology, virology with particular emphasis on molecular epidemiology, and comparative genomics [1-3].

According to the International Association of Epidemiology, molecular epidemiology is defined as the application of epidemiologic principles to study the molecular, biochemical, cellular and genetic mechanisms that underlie the pathophysiology, etiology and prevention of human diseases and related outcomes, as well as their early detection, treatment, or prognosis.

From an instrumental viewpoint, the use in epidemiological research of molecular and cell biology techniques as well as the integration of genetics and systems biology into the interdisciplinary "omics" approach has revolutionized translational research and greatly contributed to the discovery of novel biomarkers allowing tracing systems through genome navigation, the so-called genomic epidemiology. Molecular epidemiology is making valuable contributions to biomedical, clinical, and population sciences with exceptional impact on the role of gene-environment interactions, etiology of diseases and the complex drivers of disease progression by generating sound evidence about the underlying biological mechanisms and by providing knowledge with to potential primary prevention strategies [4].

Advances in sequencing technology has blossomed into a new era of genomic epidemiology, where traditional molecular diagnostics and gen-

Corresponding author
Alfonso J. Rodriguez-Morales
E-mail: arodriguezm@utp.edu.co

otyping methods are being enhanced and even replaced by novel throughput genomics-based methods to aid in the epidemiologic investigations of communicable diseases. The ability to analyze and compare entire pathogen genomes has allowed for unprecedented resolution into how and why infectious diseases spread. As these genomics-based methods continue to improve in speed, cost, and accuracy, they will be increasingly adopted to inform and guide infection control and public health practices [5].

This emergence in novel genomic tools has been key in understanding the many aspects of the SARS-CoV-2/COVID-19 pandemic. The initial full-genome sequence analysis of SARS-CoV-2 revealed its taxonomic status as a member of the betacoronavirus, with clear divergence from the SARS-CoV and MERS-CoV both players of past epidemics [6-8]. Such tools along with refined methods of phylogenetic analysis provided further evidence that SARS-CoV-2 along with the Bat-SARS-like coronavirus cluster is a distinct lineage within the subgenus of the Sarbecovirus [9, 10].

Virology has embraced the next-generation sequencing revolution, swiftly moving from the time of single genome sequencing to the age of genomic epidemiology [11]. Hundreds and now even thousands of genomes are being processed by massive parallel sequencing for detection of multiple organism species, allowing unprecedented levels of resolution and insight in the evolution and epidemic diffusion of the main pathogens, currently including more than 2,220 full genomes included in the SARS-CoV-2 public database of the Global Initiative on Sharing All

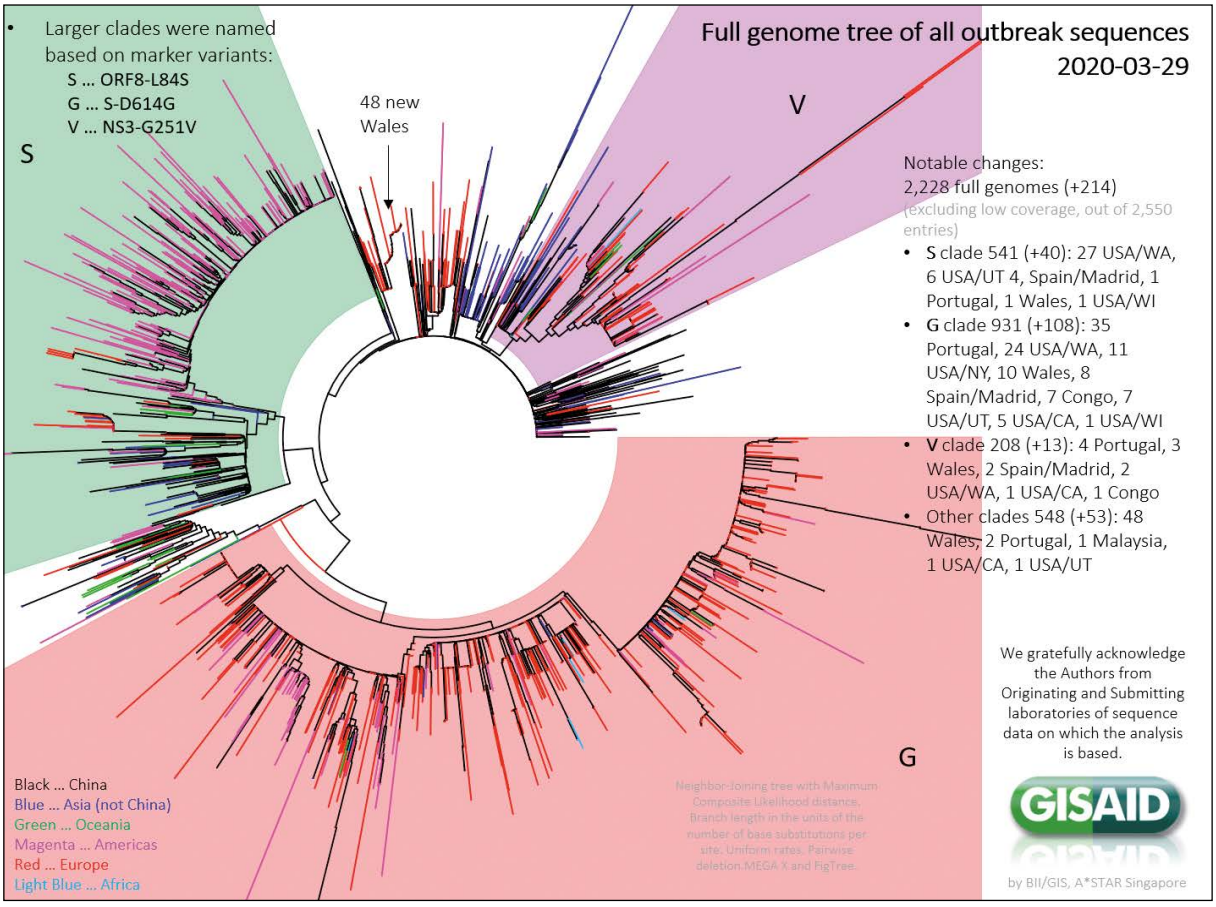

Figure 1 - Full genome tree of all outbreak sequences analyzed by GISAID. Available at: https://www.gisaid.org/hcov-19-analysis-update/

Influenza Data (GISAID) (Figure 1) [12]. GISAID was originally developed for genomic data sharing on influenza, but now has extended its coverage to include a comprehensive, dynamic and constantly updated SARS-CoV-2 database [13].
The analyses show that the SARS-CoV-2 genomes sequenced are located mainly in three clades, in addition to others, S clade (541 genomes), G clade (931), V clade (208), and other additional clades (548) (Figure 1). For example, the GISAID analytical repository on the receptor binding surveillance for high quality genomes, up to March 28, 2020, revealed the presence of four different rare variants near the binding interface: V483A in 16 USA/WA samples, L455I together with F456V in one Brazilian sample and G476S in 10 USA/WA samples [12]. Also, it has provided insights via comparative genomic analysis on potential drug targets based on the degree of similarity of highly conserved between hCoV-19 and SARS. Both, the main protease and polymerase which are potential drug targets are highly conserved between hCoV-19 and SARS with 96% and 97% overall identity, respectively. Inhibitors developed against the SARS-CoV main protease or polymerase have also shown the potential to bind in a similar fashion to hCoV-19 [12].
As predicted, the evolving genomic revolution is already having a profound impact on the practice of epidemiology, virology, infectious diseases, and public health [14-16]. Molecular epidemiology in all its ramifications allows reconstructing the evolutionary history of viruses, analyzing its appearance and global spread such has been the recent experience with SARS-CoV-2. As its genome continues to accumulate mutations (Figure 1), the generation of distinct genetic traces helps defining transmission chains and even rebuild the links in a chain with individuals who ignore their source of transmission, and even assess scenarios such as community transmission. Comparative genomics allows to infer the generation in a transmission chain, *e.g.* from imported case, with full traceability and also allows to estimate the most probable ancestor, and by comparison and geographical location, where it originated. As new genomes of SARS-CoV-2 are being obtained from imported cases, contacts and those of community transmission, in the next days, a better understanding of the genomic epidemiology of SARS-CoV-2 will become available. Genomic epidemiology analyses allow recognition of transmission clusters, its biological evolutionary rate, and finally the possibility of estimating the magnitude of the potential pandemic. In multiple countries of the world, obtaining new SARS-CoV-2 genomes is ongoing, and this will allocate monitoring multiple aspects of this pandemic. These include genetic diversity, association with clinical and epidemiological patterns and profiles, the usefulness of diagnostic methods, and the rational design of therapeutics and vaccine candidates as mentioned above. Most countries in Europe and North America have sequenced their respective SARS-CoV-2 isolate genomes, but in Asia, Middle East, Africa and Latin America, many countries still lack such capabilities.
Fortunately, the ongoing COVID-19 pandemic has opened an era of open science where a number of publically available open-source projects such as GISAID, and their related webs (https://nextstrain.org/ncov), have provided easy an accessible tools to better understand and improve outbreak response by capitalizing in pathogen genomic data [12].

Funding
None.

Conflicts of Interest
None.

REFERENCES

[1] Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, et al. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis*. 2020; 101623. [Online ahead of print]
[2] Rodriguez-Morales AJ, Bonilla-Aldana DK, Balbin-Ramon GJ, et al. History is repeating itself, a probable zoonotic spillover as a cause of an epidemic: the case of 2019 novel Coronavirus. *Infez Med.* 2020; 28 (1), 3-5.
[3] Bracken MB. Genomic epidemiology of complex disease: the need for an electronic evidence-based approach to research synthesis. *Am J Epidemiol.* 2005; 162 (4), 297-301.
[4] Porta M. Molecular epidemiology. In: Porta M, editor. A Dictionary of Epidemiology. Oxford: Oxford University Press; 2014.
[5] Tang P, Croxen MA, Hasan MR, Hsiao WW, Hoang LM. Infection control in the new age of genomic epidemiology. *Am J Infect Control.* 2017; 45 (2), 170-9.

[6] Al-Omari A, Rabaan AA, Salih S, Al-Tawfiq JA, Memish ZA. MERS coronavirus outbreak: Implications for emerging viral infections. *Diagn Microbiol Infect Dis.* 2019; 93 (3), 265-85.
[7] Rabaan AA. Middle East respiratory syndrome coronavirus: five years later. *Expert Rev Respir Med.* 2017; 11 (11), 901-12.
[8] Rizzardini G, Saporito T, Visconti A. What is new in infectious diseases: Nipah virus, MERS-CoV and the Blueprint List of the World Health Organization. *Infez Med*. 2018; 26 (3), 195-8.
[9] Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020 doi 10.1056/NEJMoa2001017.
[10] Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infect Genet Evol.* 2020; 79, 104212.
[11] Gaiarsa S, De Marco L, Comandatore F, Marone P, Bandi C, Sassera D. Bacterial genomic epidemiology, from local outbreak characterization to species-history reconstruction. *Pathog Glob Health*. 2015; 109 (7), 319-27.
[12] GISAID. Genomic epidemiology of novel coronavirus. Available at https://nextstrain.org/ncov 2020 Last accessed 9th March 2020
[13] Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* 2017; 22 (13).
[14] Traynor BJ. The era of genomic epidemiology. *Neuroepidemiol*. 2009; 33 (3): 276-9.
[15] Bastola A, Sah R, Rodriguez-Morales AJ, Lal BK, Jha R, Ojha HC, et al. The first 2019 novel coronavirus case in Nepal. *Lancet Infect Dis.* 2020; 20 (3), 279-80.
[16] Sah R, Rodriguez-Morales AJ, Jha R, et al. Complete genome sequence of a 2019 novel coronavirus (SARS-CoV-2) strain isolated in Nepal. *Microbiol Resour Announc*. 2020; 9 (11).