

Genomic surveillance and sequencing of SARS-CoV-2 across South America

Maria Fernanda Ribeiro Dias,¹ Bruce Veiga Andriolo,² Diego Henrique Silvestre,³ Paula Lopes Cascabulho⁴ and Manuela Leal da Silva⁵

Suggested citation Ribeiro Dias MF, Andriolo BV, Silvestre DH, Cascabulho PL, Leal da Silva M. Genomic surveillance and sequencing of SARS-CoV-2 across South America. *Rev Panam Salud Publica*. 2023;47:e21. <https://doi.org/10.26633/RPSP.2023.21>

ABSTRACT

After 2 years of the COVID-19 pandemic, the protocols used to control infection lack attention and analysis. We present data about deposits of complete genomic sequences of SARS-CoV-2 in the Global Initiative on Sharing All Influenza Data (GISAID) database made between January 2021 and May 31, 2022. We build the distribution profile of SARS-CoV-2 variants across South America, highlighting the contribution and influence of each variant over time. Monitoring the genomic sequences in GISAID illustrates negligence in the follow up of infected patients in South America and also the discrepancies between the number of complete genomes deposited throughout the pandemic by developed and developing countries. While Europe and North America account for more than 9 million of the genomes deposited in GISAID, Africa and South America deposited less than 400 000 genome sequences. Genomic surveillance is important for detecting early warning signs of new circulating viruses, assisting in the discovery of new variants and controlling pandemics.

Keywords

SARS-CoV-2; COVID-19; health surveillance; genome; South America.

In December 2019, cases of severe pneumonia of unknown origin were reported from China. The growing number of cases of this disease led to the identification of the responsible microorganism on January 7, 2020: a new species of coronavirus designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the disease it caused as coronavirus disease 2019 (COVID-19) (1). The virus was identified as belonging to the Coronaviridae family, *Betacoronavirus* genus and *Sarbecovirus* subgenus, also known as lineage B (2-4). Its genome is composed of positive sense RNA, and it is one of the largest viruses, with 30 000 base pairs (or 30 kbp) (2-4).

As time passed and new variants began to appear, public health authorities, such as the United States Centers for Disease Control and Prevention and the World Health Organization (WHO), implemented a classification system. The system divides the main variants into the following groups: variant of concern (VOC), variant of interest and variant under

monitoring, with VOC being the most serious category. They also agreed to use Greek letters to designate variants (5).

VARIANTS AND GENOMIC SURVEILLANCE

In September 2020, the Alpha variant (B.1.1.7) emerged in the south of England with mutations that increased transmission rates and increased severity of disease (4). The Beta variant (B.1.351) was first detected in South Africa and had higher resistance to neutralization by monoclonal antibodies and generated a bigger viral load (4). In Manaus, the capital city of the Amazonas state in Brazil, the Gamma variant (P.1) was discovered. It had high transmissibility, was able to reinfect individuals who had already been cured or vaccinated, and its mutations allowed the virus to evade monoclonal antibodies against SARS-CoV-2 (4). The Delta variant (B.1.617.2) was found in India in late 2020, and it was 40% to 60% more transmissible

¹ Secretaria de Educação do Espírito Santo (SEDU), Vila Velha, Espírito Santo, Brazil

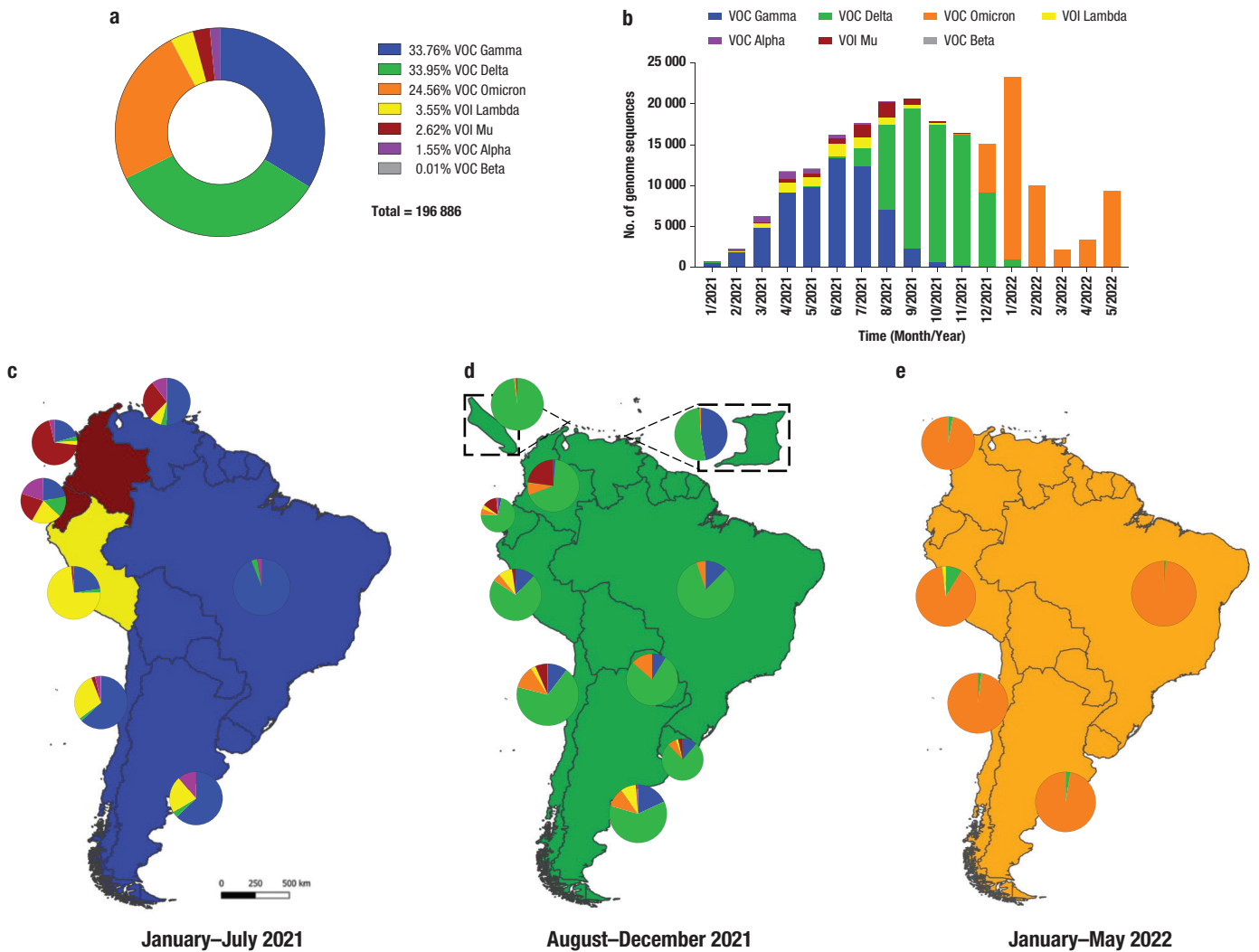
² Programa de Pós-graduação em Biotecnologia, Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO), Duque de Caxias, Rio de Janeiro, Brazil

³ Instituto de Nutrição Josué de Castro, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil ✉ Diego Henrique Silvestre, diegosilvestre@ufrj.br

⁴ Universidade Católica de Petrópolis, Petrópolis, Rio de Janeiro, Brazil

⁵ Instituto de Biodiversidade e Sustentabilidade (NUPEM), Universidade Federal do Rio de Janeiro (UFRJ), Macaé, Rio de Janeiro, Brazil

FIGURE 1. SARS-CoV-2 variants in South America with complete genomes sequenced and deposited in the Global Initiative on Sharing All Influenza Data (GISAID) database, January 2021 to May 2022. The panels show (a) the total number of sequences classified, by variant; (b) the prevalence of variants, by month and year; (c–e) the prevalence of the Gamma (blue), Delta (green) and Omicron (orange) variants in South America during the COVID-19 pandemic, by time frame.



VOC: variant of concern; VOI: variant of interest.

Source: Figure prepared by the authors from the results of their study.

than the Alpha variant (6). In late 2021, the Omicron variant (B.1.1.529) was detected in Botswana and South Africa and identified by WHO as a VOC. It has the most mutations yet observed, with 50 in total, 32 of which occur only in the spike protein, and 15 of those occur in the receptor-binding domain. These mutations confer high transmissibility and risk of reinfection, and are associated with lower efficacy of vaccines and milder symptoms when compared with the parental lineages of SARS-CoV-2 (6–8).

It is essential to analyze the dispersion of variants over time and geography so that predictions can be made and genomic surveillance can be performed. In this paper, we present an overview of the distribution of variants in South America so that patterns can be identified and, consequently, better strategies can be created to fight the pandemic. The countries included are Argentina, Aruba, Bolivia (Plurinational State of), Bonaire, Brazil, Chile, Colombia, Curaçao, Ecuador, French Guiana,

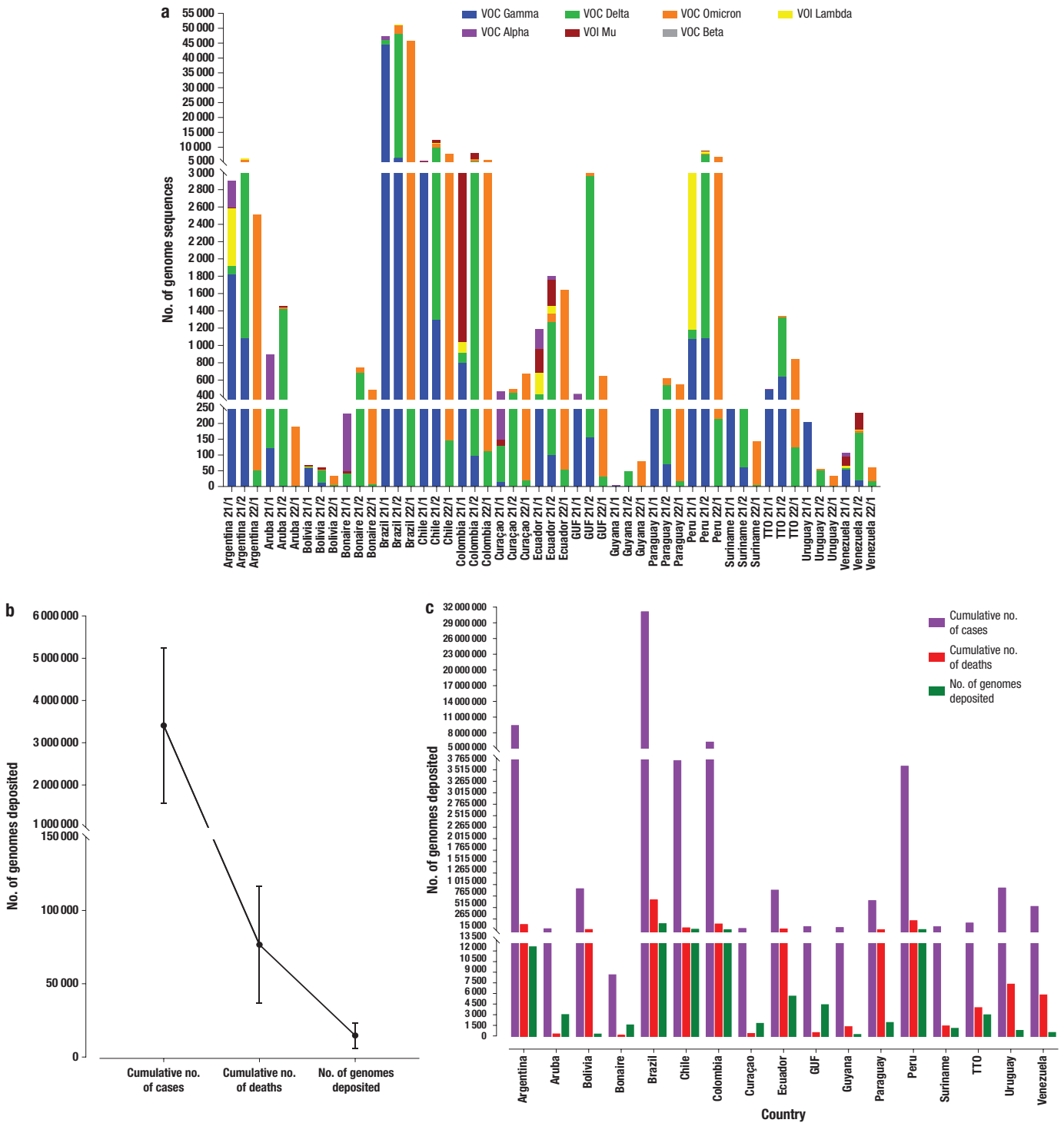
Guyana, Paraguay, Peru, Suriname, Trinidad and Tobago, Uruguay and Venezuela (Bolivarian Republic of).

IDENTIFICATION AND DISPERSION ANALYSIS OF VARIANTS

The identification and dispersion analysis of new variants of SARS-CoV-2 are key to controlling the pandemic (9). The more sequenced genomes that are uploaded to international databases, the easier it becomes to diagnose new COVID-19 cases and to identify new variants of the virus. More complete databases also contribute to the development of effective treatments for the disease (10).

As of May 2022, more than 10 671 179 genome sequences of SARS-CoV-2 had been made available in public databases (11). As of October 2022, there are more than 10 999 305 deposits in the Global Initiative on Sharing All Influenza Data (GISAID)

FIGURE 2. Number of SARS-CoV-2 variants sequenced and deposited, with cumulative cases and deaths in South America. The panels show (a) the numbers of each variant sequenced and the prevailing variants, by country, with 21/1 indicating the period January–July 2021, 21/2 indicating August–December 2021 and 22/1 indicating January–May 2022; (b) the relationship between cumulative cases and deaths, according to WHO, and deposits in the Global Initiative on Sharing All Influenza Data (GISAID) database from South America, expressed as the mean number with the standard error; (c) the relationship between cumulative cases and deaths, according to WHO, and the genome sequences deposited in GISAID, by country.



GUF: French Guiana; TTO: Trinidad and Tobago; VOC: variant of concern; VOI: variant of interest. Source: Figure prepared by the authors from the results of their study.

database, an open access platform available to researchers throughout the world (12), recommended and used by WHO to follow the spread of the virus and its mutations. These data highlight the positive effects of investments in computational biology.

This paper reports on the genetic sequencing of SARS-CoV-2 variants detected in South American countries and islands. The analysis used data deposited up until May 31, 2022, in the GISAID database and other data collected by WHO.

Analysis showed that Brazil has uploaded the most genomic data (148 226 genomes), followed by Chile and Peru (25 677 and 20 336 genomes, respectively), and Guyana has uploaded the fewest sequences (145 sequences).

Among the most noteworthy variants, the highest number of genomes uploaded were from the Delta variant (75 318), followed by the Gamma (64 615) and Omicron (63 140) variants; the lowest number of genomes uploaded were for the Beta variant (27 genomes). The data reflect the high rate of infection with the Omicron variant, the latest variant to be discovered: the Omicron variant became the dominant circulating variant in only 5 months and thus accounted for more deposits in a shorter amount of time than the Delta variant, which was circulating for more than 6 months.

Figure 1 shows the number and variety of variant genomes deposited in GISAID between January 2021 and May 2022 (Figure 1a,b). Figure 1c–e shows the prevalence of the Gamma (January–July 2021), Delta (August–December 2021) and Omicron (January–May 2022) variants in different South American countries. Figure 2a shows the number of SARS-CoV-2 genome sequences by variant for each country in each of the analyzed periods (January–July 2021, August–December 2021, January–May 2022). Figure 2b shows the number of genomes uploaded by the cumulative number of cases, deaths and genomes deposited in GISAID. Figure 2c shows the number of deposits by country.

The data deposited in GISAID were compared with the infection rates and numbers of deaths registered by WHO. The connections between these data show the low number of genomes uploaded by most countries in WHO's Region of the Americas when compared with their high infection rates. Some countries stand out for the discrepancies between the numbers of genome sequences deposited in GISAID, the cumulative cases reported and the cumulative deaths from COVID-19 reported by WHO, which demonstrates the lack of synchronization between health agencies (Figure 2b). The following countries stand out in this regard: Argentina (9 230 573 cumulative cases; 128 889 cumulative deaths; 12 516 deposited sequences), Brazil (30 953 579 cumulative cases; 666 453 cumulative deaths; 148 226 deposited sequences), Chile (3 698 667 cumulative cases; 57 887 cumulative deaths; 26 476 deposited sequences), Colombia (6 103 455 cumulative cases; 139 854 cumulative deaths; 18 734 deposited sequences) and Peru (3 580 051 cumulative cases; 213 173 cumulative deaths; 21 060 deposited sequences) (Figure 2c).

GENOMIC SURVEILLANCE AND SEQUENCING ACROSS SOUTH AMERICA

Obtaining genomic data is a long process that includes collecting the sample, sequencing the material using trusted methods and depositing the sequence in a public database (13). This process demands public funding, specific materials and skilled labor, and inconsistencies in the availability of these

help explain the differences in population sampling for SARS-CoV-2 between countries.

The genomic sequence from the first reported case associated with the Omicron variant in South America was deposited in October 2021, during the rise of the Delta variant infection rate. The recombination of Delta and Omicron can result in mutations with different viral properties and characteristics. Bolze et al. identified cases of Delta and Omicron co-infection in the United States of America between November 2021 and February 2022 (14). An analysis of the first 87 695 SARS-CoV-2 genomes shared on GISAID in 2021 identified 225 sequences of likely recombinant origin (15).

Africa and South America account for only 349 680 of all genomes deposited in GISAID, while Europe accounts for 5 703 112 and North America for 3 758 763. These data highlight differences in public investment in genomic and epidemiological surveillance between developed countries and developing countries. Up until the end of May 2022, only 242 304 sequences had been uploaded by the 17 countries of South America (Figure 2b), representing approximately 2.2% of the total sequences deposited in the database since January 2022. This discrepancy allows new variants to emerge and disseminate without immediate detection and monitoring. The need for permanent, active genomic surveillance is illustrated by the experience in countries such as New Zealand, which had a low index of cumulative deaths (1 128) compared with the number of cumulative cases (1 157 227) and the number of genomes uploaded (12 849). In New Zealand, higher GISAID deposit rates, along with other measures, resulted in a decrease in infection rates and the number of deaths.

WHO's report from June 22, 2022, (16) notes the increase in cases since April 22, 2022. During this period, the Region of the Americas registered more than 1.1 million new cases, with the United States leading the ranking with 652 217 new cases, followed by Brazil with 256 034 and Chile with 73 455. Despite the rise in the number of cases, only 14 732 genomes were uploaded to GISAID during May 2022 from South American countries, representing only 2.7% of the total number of deposits made worldwide during the same period, which was 542 740.

It is important to highlight the increase in vaccination rates since June 2021 that worked to slow infection rates and reduce the number of reported deaths. This increase in vaccination may have affected the number of genome sequences collected.

The lack of strong genomic surveillance systems is not the only hardship faced by health care systems in the Region as there are also the difficulties in monitoring public funding allocated to primary health care, with its lack of accountability and lack of standardization of the definition of primary health care, as highlighted by the Organization for Economic Co-operation and Development and WHO. These difficulties were found in an analysis of primary health care in three South American countries (17), showing that troubles measuring investments in health care can interfere with the diagnosis, management and treatment of patients in addition to monitoring public or other health interventions made during the COVID-19 pandemic. This reveals that inadequacies remain in the health surveillance process that could, and should, have been rectified after 2 years of the pandemic, and that these inadequacies allowed for the spread of new variants that may render humanity's efforts to defeat this virus obsolete, hence the emergence of the fourth wave of the pandemic that has affected some countries.

In contrast, there are positive examples from countries such as China and New Zealand that proved to be extremely capable of stopping the spread of the virus. Although New Zealand is tiny compared with most South American countries, which hampers the spread of the virus, China has a large population and managed to have fewer cumulative cases (2 938 534) and deaths (16 769) than Brazil. Both countries employed effective and rigorous forms of lockdown to protect their population, while providing help so that people could survive despite not being allowed out of their homes. Meanwhile, Sweden and the USA, which adopted the so-called herd immunity strategy, had high rates of infection (at 2 508 736 and 83 248 755 cases, respectively) and death (at 19 040 and 998 547 deaths, respectively), and did not achieve such immunity.

For countries to become effective in combating new pandemics, actions must be taken by health agencies to develop strategies among themselves and with the Pan American Health Organization and WHO in parallel to the work of political leaders to ensure there are strong plans of action for surveillance and response to stop the spread of diseases as part of a group effort to stabilize the Region.

Authors' contributions. MLS and MFRD conceived the original idea for the study. PLC, MLS and MFRD planned the data collection. PLC and BVA collected the data, and PLC, BVA and DHS analyzed the data. BVA, DHS, MLS and MFRD interpreted the results. BVA, DHS and MFRD wrote the paper. BVA, DHS, MLS and MFRD reviewed the paper. All authors reviewed and approved the final version of the paper.

Conflicts of interest. None declared.

Funding. This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Disclaimer. Authors hold sole responsibility for the views expressed in the manuscript, which may not necessarily reflect the opinion or policy of the *Revista Panamericana de Salud Pública/Pan American Journal of Public Health* or those of the Pan American Health Organization.

REFERENCES

- Fagbule OF. 2019 novel coronavirus. *Ann Ib Postgrad Med.* 2019;17(2):108–10.
- Esakandari H, Nabi-Afjadi M, Fakkari-Afjadi J, Farahmandian N, Miresmaeili S, Bahreini E. A comprehensive review of COVID-19 characteristics. *Biol Proced Online.* 2020;22(1):19.
- Ludwig S, Zarbock A. Coronaviruses and SARS-CoV-2: a brief overview. *Anesth Analg.* 2020;131(1):93–6.
- Naji H. Notable variants of SARS-CoV-2. *Eur J Med Health Sci.* 2021;3(2):44–7.
- World Health Organization. Tracking SARS-CoV-2 variants [Internet]. Geneva: World Health Organization; 2022 [cited 2022 June 24]. Available from: <https://www.who.int/activities/tracking-SARS-CoV-2-variants>
- Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: a comparative computational study of spike protein. *J Med Virol.* 2022;94(4):1641–9.
- Fernandes Q, Inchakalody VP, Merhi M, Mestiri S, Taib N, Moustafa Abo El-Ella D, et al. Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics and vaccines. *Ann Med.* 2022;54(1):524–40.
- Meo SA, Meo AS, Al-Jassir FF, Klonoff DC. Omicron SARS-CoV-2 new variant: global prevalence and biological and clinical characteristics. *Eur Rev Med Pharmacol Sci.* 2021;25(24):8012–8.
- Slavov SN, Patané JSL, Bezerra RDS, Giovanetti M, Fonseca V, Martins AJ, et al. Genomic monitoring unveil the early detection of the SARS-CoV-2 B.1.351 (beta) variant (20H/501Y.V2) in Brazil. *J Med Virol.* 2021;93(12):6782–7.
- Genomic sequencing in pandemics. *Lancet.* 2021;397:445.
- Romano CM, Melo FL. Genomic surveillance of SARS-CoV-2: a race against time. *Lancet Reg Health Am.* 2021;1:100029.
- Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance.* 2017;22(13):130494.
- Chiara M, D'Erchia AM, Gissi C, Manzari C, Parisi A, Resta N, et al. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Brief Bioinform.* 2021;22(2):616–30.
- Bolze A, Basler T, White S, Rossi AD, Wyman D, Dai H. Evidence for SARS-CoV-2 Delta and Omicron co-infections and recombination. *Med (N.Y.).* 2022;2022:S2666-6340(22)00449-4.
- Varabyou A, Pockrandt C, Salzberg SL, Peretea M. Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. *Genetics.* 2021;218(3):iyab074.
- World Health Organization. COVID-19 weekly epidemiological update: 22 June 2022. Geneva: World Health Organization; 2022. <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---22-june-2022>
- Rathe M, Hernández-Peña P, Pescetto C, Van Mosseveld C, Borges dos Santos MA, Rivas L. Gasto em atenção primária à saúde nas Américas: medir o que importa [Primary health care expenditure in the Americas: measuring what matters]. *Rev Panam Salud Pública.* 2022;46:e13. Portuguese. <https://iris.paho.org/handle/10665.2/56016>

Manuscript submitted 25 June 2022. Revised version accepted for publication on 16 October 2022.

Vigilancia genómica y secuenciación del SARS-CoV-2 en América del Sur

RESUMEN

Tras dos años de pandemia del COVID-19, los protocolos empleados para controlar la infección carecen de atención y análisis. En este artículo se presentan datos sobre depósitos de secuencias genómicas completas del SARS-CoV-2 en la base de datos de secuenciación GISAID, la Iniciativa mundial para intercambiar todos los datos sobre la gripe aviar, realizadas entre enero del 2021 y el 31 de mayo del 2022. Se creó el perfil de distribución de las variantes del SARS-CoV-2 en América del Sur, en el que se destacaron la contribución y la influencia de cada variante a lo largo del tiempo. El monitoreo de las secuencias genómicas en GISAID ilustra la negligencia en el seguimiento de los pacientes infectados en América del Sur, así como las discrepancias entre el número de genomas completos depositados a lo largo de la pandemia por parte de los países desarrollados y los países en desarrollo. Mientras que Europa y América del Norte han depositado más de 9 millones de genomas en GISAID, África y América del Sur han aportado menos de 400 000 secuencias genómicas. La vigilancia genómica es importante para detectar los primeros signos de alerta de virus nuevos en circulación, ayudar en el descubrimiento de nuevas variantes y controlar las pandemias.

Palabras clave SARS-CoV-2; COVID-19; vigilancia sanitaria; genoma; América del Sur.

Vigilância genômica e sequenciamento do SARS-CoV-2 na América do Sul

RESUMO

Após 2 anos da pandemia de covid-19, os protocolos usados para controlar a infecção necessitam maior atenção e análise. Apresentamos dados sobre as sequências genômicas completas do SARS-CoV-2 depositadas no banco de dados da iniciativa internacional para o intercâmbio de dados sobre os vírus da influenza (GISAID) entre janeiro de 2021 e 31 de maio de 2022. Construímos o perfil de distribuição das variantes do SARS-CoV-2 na América do Sul, destacando a contribuição e a influência de cada variante ao longo do tempo. O monitoramento das sequências genômicas do GISAID ilustra a negligência no acompanhamento de pacientes infectados na América do Sul e as discrepâncias entre os países desenvolvidos e em desenvolvimento com relação ao número de genomas completos depositados ao longo da pandemia. Enquanto a Europa e a América do Norte respondem por mais de 9 milhões dos genomas depositados no GISAID, a África e a América do Sul depositaram menos de 400 000 sequências genômicas. A vigilância genômica é importante para detectar sinais de alerta precoces de novos vírus circulantes, auxiliar na descoberta de novas variantes e controlar pandemias.

Palavras-chave SARS-CoV-2; COVID-19; vigilância sanitária; genoma; América do Sul.
