

Introduction

RNA viruses, like Ebola, Influenza, Dengue, Zika and SARS-CoV, rapidly evolved with a constant accumulation of mutations in their genomes (1,2). Since December 2019, a large number of SARS-CoV-2 genomes has been generated worldwide and deposited in public repositories (i.e., GISAID, Genbank), allowing us to track, almost in real time, how this virus was evolving (2,4–6). Nevertheless, we only find studies about transmission patterns in their local populations from a few European and North-American countries (7,8). Latin America has also generated more than 100,000 genomes until now, data especially from Brazil, Chile, Peru, Colombia, Ecuador and Uruguay (2,5). For instance, the Peruvian National Institute of Health (INS-Peru) in collaboration with the National Center of Epidemiology, Prevention and Control of Disease of this country (CDC MINSA) has generated more than 40,000 SARS-CoV-2 genomes by 2023 (<https://web.ins.gob.pe/es/covid19/secuenciamiento-sars-cov2>), confirming the circulation of many variants from the beginning of the COVID-19 pandemic (i.e., lambda, gamma, alpha, delta, mu, zeta, epsilon, etc. and others; https://nextstrain.org/community/quipu/Peru_lambda) (9–12). Complementary to human detection of the virus, environmental surveillance data of varied viruses, especially those related to viability and potential infectivity like SARS-CoV, could act as a useful tool to predict timely disease outbreaks and issue rapid warnings to health authorities (13,14). **In this study, we aimed at detecting by next-generation sequencing the variants of SARS-CoV-2 coming from different hospital effluents in Peru during the period between March and September of 2022.** In addition, the data retrieved were correlated to the circulating variants monitored by the surveillance systems from INS-Peru, in order to assess the usefulness of complementing information from both sources.

Methodology

Wastewater samples

The samples were collected between March and September of 2022, from nine hospitals in Peru (departments of Lima, San Martín, Puno, Cuzco, Cajamarca, and La Libertad), each sampling point was georeferenced (Figure 1). Residual water samples were collected in sterile glass bottles (1000 mL) and then labeled, transported, and stored at 4°C at the molecular biology laboratory of the Universidad Peruana Unión, following the CDC recommendations (<https://www.cdc.gov/healthywater/surveillance/wastewater-surveillance.html>).

Samples preparation and pre-treatment

A total of 40 mL of homogenized wastewater samples was treated with proteinase K and incubated for 30 minutes. Then, we added the binding buffers following the protocol of the Wizard® Enviro Wastewater TNA kit (Promega Corp), and isopropanol Sigma Aldrich Co. (St. Louis, MO, United States). Columns were submitted to a vacuum system, and the total content finally eluted with RNase/DNase-free water.

RNA purification and electrophoresis analysis

The eluate obtained from the previous step was purified to obtain total RNA using silica columns. The total content of the eluate was passed throughout the column, washed and finally recovered with 70 uL of RNase/DNase-free water.

RNAs were quantified by absorbance at 260 nm using a Nanodrop and the integrity was visualized on 1% agarose gels (Sigma Aldrich Co., St. Louis, MO, United States) stained with sybr gold.

cDNA and Library preparation and NGS

Total RNA was retrotranscribed to cDNA using random primers, and the library preparation performed using the Illumina COVID Seq Kit RUO. MiSeq equipment (Illumina®) was used to sequence according to the manufacturer's instructions.

Bioinformatic analysis and epidemiological comparison to INS-Peru data

The quality of reads was assessed by the software implemented in Illumina® package of tools, with an additional analysis performed by the softwares Nextclade and Pangolin. These two softwares allowed us to identify which variants and lineages were circulating in the country.

Additionally, we assessed the findings (variants and lineages) with those from patient samples tracked by Peruvian National Institute of Health (INS-Peru) during the same period and geographical area (<https://web.ins.gob.pe/es/covid19/secuenciamiento-sars-cov2>).

Results

Figure 1. Healthcare facilities sampled according to location and level of care.



Sequences from each sample were obtained, gathering 20 sequences in total. However, Pangolin and Nextclades softwares firstly assessed their quality, discarding all of them with the first program (labeled as ambiguous content or fail), and retrieving six results with the latter (Table 2). As depicted, all the sequences with quality enough were allocated to Omicron variant according to World Health Organization (WHO) label whereas the lineages were diverse within the three clades found: clade 21K with lineage BA.1.1 (n=1); clade 21L with lineage BA.2 (n=2); and clade 22B with the lineages BA.5.1 (n=2) and BA.5.5 (n=1). In addition, when we compared the results from wastewater to those from the same period monitored by INS-Peru, we mostly found a good concordance either in variant (all of them belonged to Omicron, the same circulating variant in the Peruvian population) and lineages (Table 2). All the 14 sequences except one (WW20) discarded by Nextclade software because of their low quality were analyzed by Illumina tools anyway, and they seemed to belong to the clade 19A of Omicron. This result could not be considered robust though (the other tools had discarded them), but somehow also showed the underlying relation with Omicron variant as well.

Table 1. Phylogenetic analysis of SARS-CoV-2 sequences from wastewater samples from nine hospitals in Peru.

Sampling locations (Peruvian department)	ID sample	National Health Institute COVID Data Tracker			Wastewater sample		
		Date	Epidemiological week	Predominant SARS-CoV-2 variants (WHO label) and lineages ¹	SARS-CoV-2 variant (Nextclade assignment)	Lineage	Clade
Hospital Carlos Monge Medrano (Puno, PE)	WW1	16/08/2022	24	Omicron, lineages BA.2.12.1, BA.4.1, BA.5.1, BA.5.2, BA.4.6, BA.5.1.8	NA		
	WW2	2/09/2022	35	Omicron, lineages BA.5, BA.5.1, BA.5.2.1, BA.5.2, BA.5.6	Omicron	BA.5.1	22B
Hospital Regional Docente Trujillo (La Libertad, PE)	WW3	13/08/2022	24	Omicron, lineages BA.2, BA.2.5, BA.4, BA.5, BA.2.12.1, BA.4.1, BA.5.1, BA.2.9	NA		
	WW4	13/07/2022	28	Omicron, lineages BA.2, BA.4, BA.2.12.1, BA.4.1, BA.5.1, BA.5.2, BA.4.6	NA		
Hospital Emergencias Ate Vitarte (Lima, PE)	WW5	25/08/2022			NA		
	WW6	25/08/2022	26	Omicron, lineages BA.2, BA.4, BA.5, BA.2.12.1	NA		
	WW7	25/08/2022			NA		
Hospital Huaycán (Lima, PE)	WW8	14/05/2022	19	Omicron, lineages BA.1, BA.1.1, BA.2, BA.4, BA.5, BA.2.12.1	NA		
	WW9	14/05/2022			NA		
Hospital Regional de Cuzco (Cuzco, PE)	WW10	9/08/2022	32	Omicron, lineages BA.2, BA.4, BA.5, BA.2.12.1, BA.4.1, BA.5.1, BA.5.1.1, BA.5.2.1, BE.1, BA.4.6, BA.5.1, BA.2.9	Omicron	BA.2	21L
	WW11	9/08/2022			Omicron	BA.2	21L
	WW12	2/09/2022	35	Omicron, lineages BA.5, BA.5.1, BA.5.2.1, BA.5.2, BA.5.6	Omicron	BA.5.5	22B
Clínica Americana Juliaca (Puno, PE)	WW13	2/09/2022			Omicron	BA.3.1	22B
	WW14	23/08/2022			Omicron	BA.1.1	21K
Hospital Regional de Cajamarca (Cajamarca, PE)	WW15	23/08/2022	34	Omicron, lineages BA.4, BA.5, BA.4.1, BA.5.1, BA.5.2.1, BA.5.2, BA.4.6, BA.5.1.3, BA.5.6, BA.5.1.3, BA.5.8.1	NA		
	WW16	7/03/2022	10	Omicron, lineages BA.1, BA.1.1, BA.2	NA		
	WW17	7/03/2022			NA		
Hospital Tarapoto II (San Martín, PE)	WW18	26/08/2022			NA		
	WW19	26/08/2022	34	Omicron, lineages BA.4.1, BA.5.1, BA.5.1.1, BA.5.2.1, BA.5.6.1	NA		
	WW20	26/08/2022			NA		

Conclusion

-Eighteen out of 20 hospital wastewater samples (30%) provided sequences with quality enough to be classified as Omicron variant according to WHO label.

-Among them, six could also be assigned by Nextclade to clades 21K lineage BA.1.1 (n=1), 21L lineage BA.2 (n=2), and 22B lineages BA.5.1 (n=2) and BA.5.5 (n=1).

-These findings show a good correlation between SARS-CoV-2 variants detected by Peruvian surveillance programs and those from wastewater samples. Wastewater systems monitoring can provide information not only about incidence, but also regarding seasonality and genotype distribution in the environment of viruses like SARS-CoV-2.

References

- Grubaugh ND. Translating virus evolution into epidemiology. *Cell Host Microbe*. 2022 Apr 13;30(4):444–8.
- 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* [Internet]. 2021 Mar 22 [cited 2023 Jun 13];22(2):616–30. Available from: <https://doi.org/10.1093/bib/bbaa297>
- Quer J, Colomer-Castell S, Campos C, Andrés C, Piñana M, Cortese MF, et al. Next-Generation Sequencing for Confronting Virus Pandemics. Vol. 14, *Viruses*. MDPI; 2022.
- GISAID. GISAID - gisaid.org [Internet]. [cited 2023 Jun 13]. Available from: <https://gisaid.org/>
- Chen Z, Azman AS, Chen X, Zou J, Tian Y, Sun R, et al. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. [cited 2023 Jun 13]; Available from: <https://doi.org/10.1038/s41588-022-01033-y>
- Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* [Internet]. 2017 Jan [cited 2023 Jun 13];1(1):33–46. Available from: <https://pubmed.ncbi.nlm.nih.gov/31565258/>
- Islam MR, Hoque MN, Rahman MS, Alam ASMUR, Akther M, Puspo JA, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Scientific Reports* 2020 10:1 [Internet]. 2020 Aug 19 [cited 2023 Jun 13];10(1):1–9. Available from: <https://www.nature.com/articles/s41598-020-70812-6>