Monitoring SARS-CoV-2 prevalence and variants in wastewater in Valencia

A wastewater-based epidemiology approach

Rubén Cañas Cañas[1], Rosa Bermejo Ramirez[1], Ester Méndez Belinchón[1]

1. General de Análisis, Materiales y Servicios, SL (GAMASER SL – Global Omnium), Carrer del Corretger 51, 46988, Paterna, Valencia, Spain.

Action co-financed by the European Union through the European Regional Development Fund (FEDER) Operational Program for the *Comunitat Valenciana* 2014-2020.

1. Summary

COVID-19 pandemic has dealt an immense health and socioeconomic impact worldwide. SARS-CoV-2 surveillance programmes have allowed following the evolution of an emergent virus for the first time in history. Currently, prevalence determination is conducted via individual diagnostic tests, which are limited and lead to other problems such as not covering the whole population.

Wastewater-Based Epidemiology (WBE) has shown to overcome these limitations, monitoring the viral spread on the whole population and surveillance of its circulant variants which has served as an extremely powerful epidemiologic surveillance tool for public health.

2. Methods

① Wastewater collection

Four key wastewater points from Valencia manhole were collected weekly since April 2020 to November 2022. Those key points were selected because they gather a representative sample of the whole population. Sample concentration and RNA extraction was performed according to IATA-CEBAS-CSIC protocol[1].

③ Bioinformatic analysis

SARS-CoV-2 reads were preprocessed using fastp v0.20.1[2] and mapped to the virus genome sequence (NC_045512.2) using bwa mem v0.7.17[3]. Variant calling was performed with the iVar software v1.3.1[4]. Mutations of interest for each sample were parsed using an in-house python script, including mutation frequency and depth.

Reference mutations of interest for each SARS-CoV-2 variant were retrieved from various databases (i.e. NCBI ACTIV-TRACE Lineage Definitions, NCBI Variant Overview and Outbreak.info). In order to simplify posterior data analysis, variants were assigned to their parent lineage, excluding those of interest having a high prevalence or interest in the region (e.g. BA.2.9, BA.2.75, BA.4.6, BE.1, BF.1, BQ.1,...).

Clinical SARS-COV-2 sequencing data from the region of Valencia was retrieved from GISAID[5] with the purpose of comparing wastewater and clinical results.

② SARS-CoV-2 quantification and sequencing

Quantification of SARS-CoV-2 RNA was realized by RT-qPCR. Genome Units per liter (GU/L) were assessed interpolating the cycle threshold obtained in a standard curve with known concentrations of the virus. Whole genome of positive samples were sequenced by NGS (next generation sequencing) using the ARTIC v4 primers in a MySeq (Illumina) equipment.

④ Deconvolution algorithm

Variant deconvolution is performed using a constrained Generalized Additive Model (GAM) with a 0 and 1 inflated beta linker function following the equation $\boldsymbol{y = \Sigma\beta x}$; where **y** represents the mutation frequencies, **x** represents the matrix presence or absence of the mutations in the variants and the regression coefficient **β** represents the estimated abundance of the variant. Prior to calculating the model, only variants with 10 or more mutations and at least 1 of them exclusive of the lineage are classified to calculate the model. The depth of each mutation (DP_{alt}) is used as weights for the regression $\boldsymbol{w = \ln(DP_{alt} + 1)}$.

If the sum of the estimated frequencies exceed 1, the frequencies are scaled down proportionally, and if the sum didn't reach 1, the remaining until 1 is determined as Uncertain lineage. Variants with negative coefficients are discarded. Calculations for the model are iterated until the model converge.

3. Results

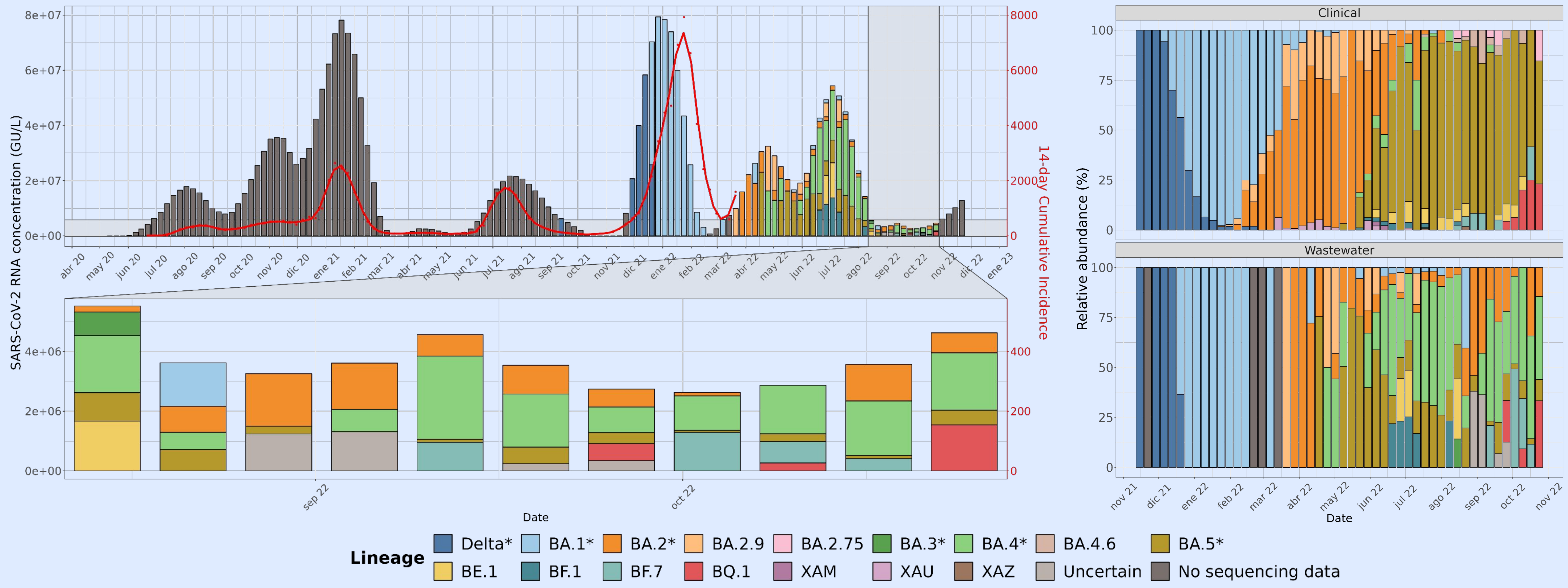

Figure 1. SARS-CoV-2 evolution through time (**Left**) and circulating variants (**Right**). SARS-CoV-2 concentrations (bars) are compared with COVID-19 14-day CI (red line). Current situation of circulating variants in wastewater are zoomed. Sequenced clinical cases are compared with wastewater estimated abundance of variants as relative abundances.

4. Discussion

SARS-CoV-2 concentration in wastewater overlaps with 14-day CI, especially when COVID-19 testing was not limited (from the fourth wave onwards). Moreover, an advance (aprox. 7 days) in SARS-CoV-2 concentration can be noticed in the ascending and descending trends of the waves.

Concerning SARS-CoV-2 whole genome sequencing, it can be observed that the variant relative abundance approximately coincides with clinical sequencing data. Interestingly, BA.4 and BA.2 variants show a high prevalence from May 2022 and onwards, in opposition to clinical data where these variants don't appear anymore. This could indicate that these variants are provoking a high number of asymptomatic cases that go unnoticed by the clinical surveillance system.

Furthermore, we were able to detect some variants (i.e BA.5 and BA.4) in advance in comparison with clinical samples.

5. Conclusions and Future Perspectives

- Combining viral quantification by RT-qPCR and identification of different lineages by NGS has established a SARS-CoV-2 surveillance program in wastewater that allows following the virus circulating variants.
- WBE is a powerful prediction tool that has helped to provide support to health authorities in crucial decision making moments.
- WBE can be used to monitor a great number of markers and environmental factors, such as virus and antibiotic resistance pathogens via massive sequencing; chronic diseases biomarkers and human metabolites via proteomics/metabolomics; or pharmaceutical compounds and illicit drugs via HPLC-MS/MS. This could contribute to evaluate the population habits and health situation from a broader point of view.

References:

1. Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., & Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Research*, *181*. https://doi.org/10.1016/j.watres.2020.115942
2. Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560
3. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
4. Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., de Jesus, J. G., Main, B. J., Tan, A. L., Paul, L. M., Brackney, D. E., Grewal, S., Gurfield, N., van Rompay, K. K. A., Isern, Michael, S. F., Coffey, L. L., Loman, N. J., & Andersen, K. G. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-018-1618-7
5. Khare, S., et al (2021) GISAID's Role in Pandemic Response. *China CDC Weekly*, 3(49): 1049-1051. doi: 10.46234/ccdcw2021.255 PMCID: 8668406