

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

A pan-European study of SARS-CoV-2 variants in wastewater under the EU Sewage Sentinel System

Shelesh Agrawal¹, Laura Orschler¹, Selina Schubert¹, Kira Zachmann¹, Leo Heijnen², Simona Tavazzi³, Bernd Manfred Gawlik³, Miranda de Graaf⁴, Gertjan Medema², Susanne Lackner^{1*}

¹ Technical University of Darmstadt, Department of Civil and Environmental Engineering Sciences, Institute IWAR, Chair of Wastewater Engineering; Darmstadt, Germany.

² KWR Water Research Institute; Nieuwegein, The Netherlands.

³ European Commission, Joint Research Centre; Ispra (Va), Italy.

⁴ Department of Viroscience, Erasmus Medical Center; Rotterdam, The Netherlands

*Corresponding author. Email: susanne.lackner@tu-darmstadt.de

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1 **Abstract:**

2 Wastewater based surveillance employing qPCR has already shown its utility for monitoring
3 SARS-CoV-2 at community level, and consequently the European Commission has
4 recommended the implementation of an EU Sewage Sentinel System. However, using
5 sequencing for the determination of genomic variants in wastewater is not fully established
6 yet. Therefore, we focused on the sequencing analysis of SARS-CoV-2 RNA in wastewater
7 samples collected across 20 European countries including 54 municipalities. Our results
8 provide unprecedented insight into the abundance and the profile of the mutations associated
9 with the variants of concerns: B.1.1.7, P.1, B.1.351 and B.1.617.2, which were present in
10 various wastewater samples. This study shows that integrating genomic and wastewater-
11 based epidemiology (WBE) can support the identification of variants circulating in a city at
12 community level.

1 Introduction

2 Undeniably, the sudden emergence of SARS-CoV-2, which has caused a global pandemic, is
3 a significant and in many regards unprecedented threat to public health. SARS-CoV-2 rapidly
4 resulted in a high number of people requiring hospitalization, casualties and major socio-
5 economic disruptions, with consequences which we still do not fully oversee. Consequently,
6 most countries have been forced to implement severe lockdown measures to ensure the
7 physical distance between people and interrupt virus transmission (1). Overall, the high
8 transmission rate and the rapidly evolving nature of the virus, leading to the emergence of
9 new variants that may transmit more readily and evade the immune response, raise broad
10 concerns about SARS-CoV-2 (2–4).

11 The current phase of the COVID-19 pandemic is shaping into an era of genomic surveillance
12 to track the genomic changes in the SARS-CoV-2 virus, which belongs to the family
13 *Coronaviridae*, genus *Betacoronavirus*. According to PANGO lineages as of now, 1503
14 SARS-CoV-2 variants are known since the initial detection of SARS-CoV-2 by sequencing
15 (5). In the last few months, genomic epidemiology, the analysis of genome sequences, has
16 revealed some fast-spreading and highly virulent SARS-CoV-2 variants (3, 6, 7), making
17 them variants of concern (VOC) (8, 9). This development underlines the importance of
18 sequencing analyses. Although the European Commission recommended to sequence 5-10%
19 of the SARS-CoV-2 positive patient samples, as of 22 January 2021 (10), most Member
20 States are below this recommended sequencing target (11).

21 Wastewater-based epidemiology (WBE) is an emerging paradigm for monitoring the
22 circulation of SARS-CoV-2 in a community. Several research groups across the globe have
23 shown that WBE provides additional information about the dynamics of SARS-CoV-2 at
24 community level (12–17). WBE efforts have primarily focused on quantitative polymerase
25 chain reaction (qPCR), determining the titers of SARS-CoV-2 in the sewage and its
26 correlation to the reported number of SARS-CoV-2 positive cases. At present, very few
27 studies are available that looked for the combined potential of genomic epidemiology and
28 WBE to determine the SARS-CoV-2 genomic variants circulating in a specific region (18–
29 21). This pan-European study, which has been conducted under the umbrella of the EU
30 Sewage Sentinel System for SARS-CoV-2 (22), a direct result of what is called “The HERA
31 Incubator” (23), assesses whether next-generation sequencing (NGS) of wastewater samples
32 can provide information about the diversity of the SARS-CoV-2 variants and associated
33 mutations at community level.

SARS-CoV-2 Variants and Mutations of Concern in Europe

During March 2021, most countries in Europe reported an increase in the COVID-19 positive cases. To get an overview of the clinical situation, we determined the count of sequences associated with current VOCs (i.e., B.1.1.7, B.1.351, P.1, and B.1.617.2) (8, 9) in the 20 European countries that had provided wastewater samples (Fig. 1). The increased COVID-19 incidence rate corresponded with the increase in sequences of B.1.1.7 among the clinical samples sequenced in these countries. For P.1 and B.1.351, no general pattern appeared and they were only present to a small fraction. During this time, also very few B.1.617.2 sequences were reported (Fig. 1). While looking at the emergence of the mutations during March 2021, we found that among the most abundant mutations across all countries, the D614G (spike protein) and P323L (non-structural protein, Nsp12) mutations were detected in all the countries (S.fig.2).

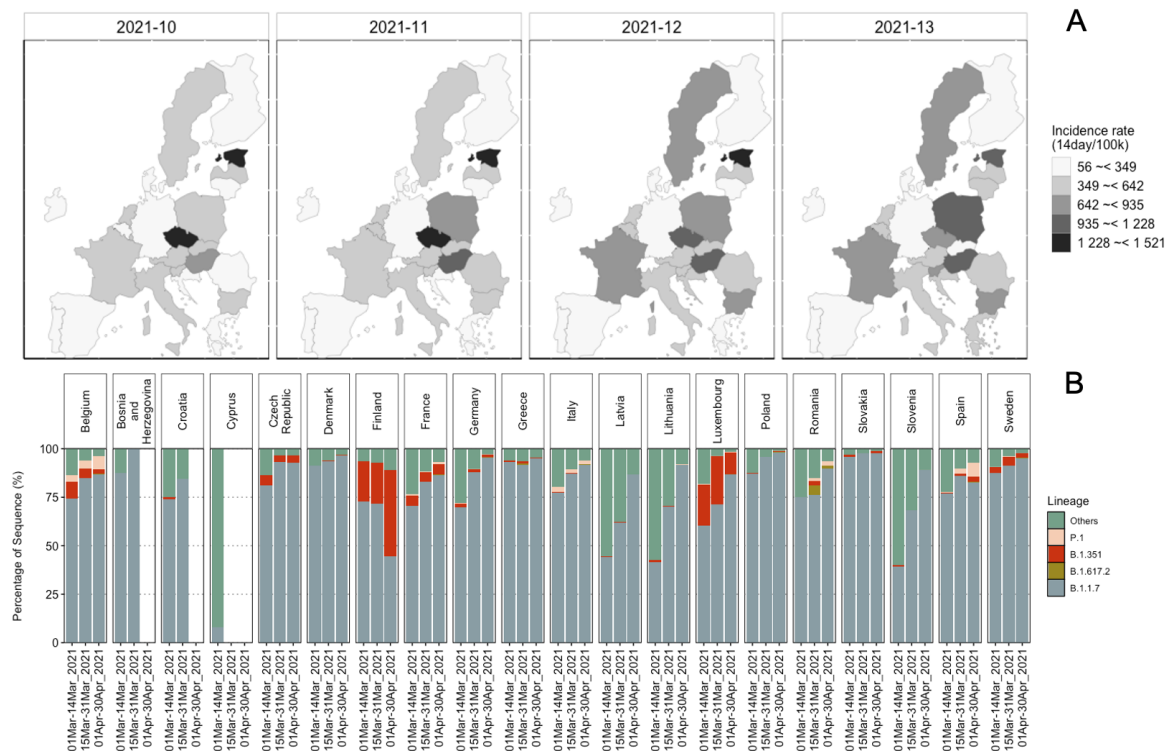


Fig 1: COVID-19 situation in 20 European countries from clinical sequencing data. (A) Maps showing the incidence rate of the COVID-19 positive cases reported in the countries, during weeks 10 to 13, from when the wastewater samples originated. (B) Relative abundance of all sequences for the countries from where wastewater samples were analyzed, available in GISAID (<https://www.gisaid.org>) on 31-05-2021. All the sequences except B.1.1.7, B.1.351, P.1, and B.1.617 were categorized as “Others”.

1 Co-occurrences of D614G and P323L have often been reported (24), however, some samples
2 from Germany and China lacked P323L (25). The spike protein N501Y and H69del
3 mutations were also highly abundant in most of the countries, earlier associated with B.1.1.7
4 and later with all VOCs except B.1.617.2 (S.fig.1, S.fig.2). In some countries the S106del,
5 G107del, F108del mutations, which are ORF1b signature mutations of the VOCs P.1, B.1.1.7
6 and B.1.351 (26), were also among the abundant mutations (S.fig.2). Across the twenty
7 European countries twenty-six ORF1ab mutations, fourteen spike protein, eight nucleocapsid
8 (N) protein, six ORF8, and three ORF3 mutations were among the dominant mutations,
9 exhibiting spatial and temporal variation (S.fig.2).

11 **Mutations associated with Variants of Concern in Wastewater Samples**

12 For this study, wastewater influent 24h composite samples from 54 wastewater treatment
13 plants (WWTPs) across 20 European countries were collected between weeks 10 – 13 of
14 2021 (i.e. 10th to 30th March 2021). We identified the mutations, which vary in their
15 association with the VOCs, in each wastewater sample, by mapping the reads generated to the
16 SARS-CoV-2 reference genome (Wuhan-Hu-1 [GenBank accession numbers NC_045512
17 and MN908947.3]). In total, 711 different mutations were identified across all the samples,
18 out of which 633 mutations were associated with the VOCs. These 633 VOC associated
19 mutations include mutations that are also associated with other SARS-CoV-2 lineages. Out of
20 these mutations, 619 mutations were observed at >2.5%, 311 mutations at >5% and 23
21 mutations at >50% allele frequency (Fig.2). Most of the 23 mutations were present in all but
22 a few samples. For example, the W131C mutation was only detected in wastewater samples
23 from Denmark (Fig.2). W131C is one of the important mutations in ORF3a, which is found
24 to assist the ion channel formation and thereby supports the virus in its infectivity (27). The
25 A220V mutation was only identified in samples from Lithuania (Fig.2), which corresponds to
26 the high count of A220V reported in clinical patient samples (S.fig.2).

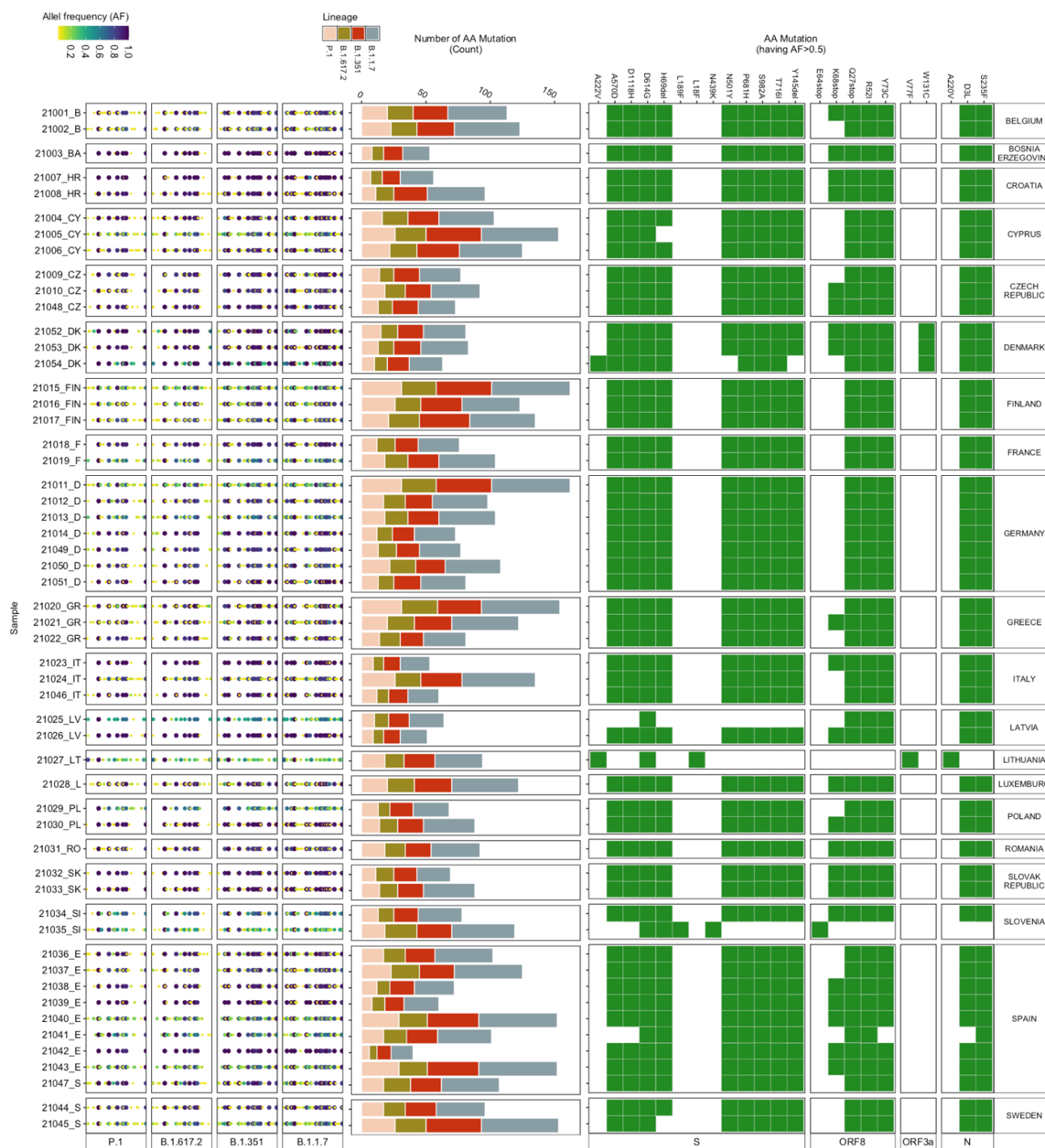


Fig 2: Amino Acid Mutations detected in wastewater samples. (Left) Allele frequency of the mutations (relative to the reference genome Wuhan-Hu-1 [GenBank accession numbers NC_045512 and MN908947.3]), associated with the VOCs, in each wastewater sample. (Middle) Number of mutations (count) detected in each sample corresponding to each VOC. (Right) Heatmap showing the presence (green) / absence (white) of mutations having more than 50 % allele frequency.

Although many low-frequency mutations were observed, the read abundance of low-frequency mutations was mostly similar to the abundance of the high-frequency mutations (S.fig.4). The highest count (ranging between 16-60) was observed for mutations associated

1 with B.1.1.7 in all the samples (Fig.3), followed by B.1.351. P.1 and B.1.617.2, which had a
2 low count of associated mutations but included signature mutations. For example, signature
3 spike protein mutations (i.e. L452R, T478K, P681R, D950N) (9, 28) of B.1.617.2 were
4 identified in some of the wastewater samples.

6 **Abundance of Spike protein AA mutations**

7 As the spike protein AA mutations have been associated with changes to characteristics of
8 SARS-CoV-2, leading to an increase in transmissibility and reduced efficacy of treatments, a
9 particular list of spike protein AA mutations is used by the European Centre for Disease
10 Prevention and Control (ECDC) for characterizing the VOCs (9, 29). Therefore, we also
11 assessed the read abundance of these spike protein AA mutations in our wastewater samples
12 (Fig. 3).

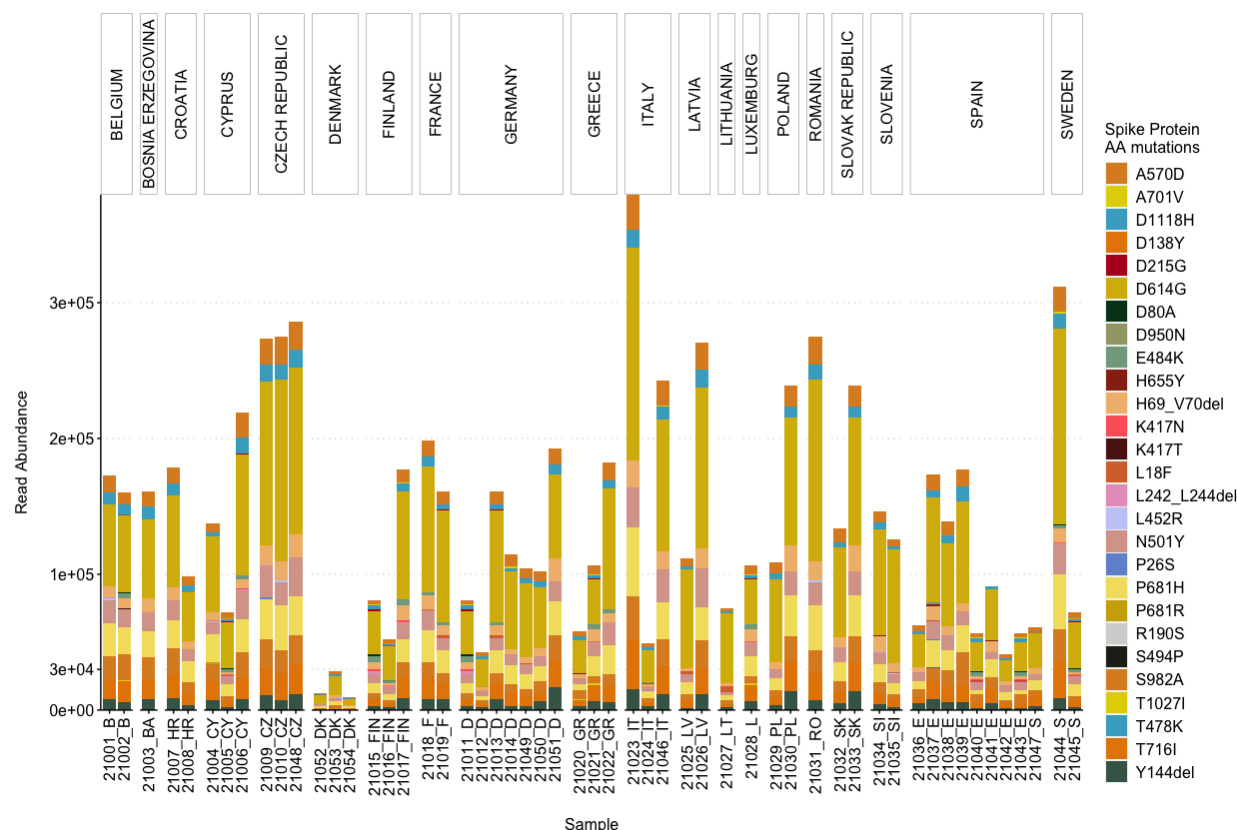
13 Overall, D614G was most abundant, followed by: P681H, T716I, A570D, S982A, H69del,
14 Y144del, D1118H, N501Y, K417N, E484K and others, in decreasing order. Only six out of
15 the 27 AA mutations (i.e. D1118H, D614G, H69del, N501Y, P681H, S982A, and T716I)
16 were present in all the samples (Fig.3). A570D and Y144del were identified in 53 samples.
17 D1118H, S982A, T716I, H69del, P681H and A570D have been mainly found in the B.1.1.7
18 variant, whereas N501Y has been associated with B.1.1.7, B.1.351, and P.1 (30). For 31
19 samples, the total read abundance of these AA mutations was above $1e+05$ reads, the lowest
20 values in all three sample were detected in the samples from Denmark ranging between
21 $1e+04$ to $3e+04$ reads (Fig.3).

23 **Abundance and prevalence of dominant AA mutations**

24 Most of the attention is given to spike protein AA mutations, especially since the emergence
25 of B.1.1.7, because the spike protein mediates virus to host cell-surface attachment, and it is
26 also the principal target of neutralizing antibodies (4). However, mutations in other regions of
27 the SARS-CoV-2 genome are also relevant (30, 31). Therefore, we analysed the most
28 abundant AA mutations found across all the samples for a comprehensive insight into the
29 mutations detected in the wastewater samples (Fig.4). Three mutations (Q27stop, R52I, and
30 Y73C) in ORF8; D3L and S235F in the N protein; P681H, D614G, H69del, S982A, T716I,
31 D1118H, and N501Y in the spike protein; P4804P, T5304T, F3677del, A1708D, S216S,
32 F924F, I2230T, T1001I, F1907F, and H5005H in ORF1ab appeared to be most dominant and
33 prevalent in the samples (Fig.4). The presence of Q27stop and R52I mutations, along with the
34 spike protein mutations, is reported to likely increase the transmissibility of SARS-CoV-2

1 (32). Another two ORF8 protein mutations, Y73C and K68stop, were present in most of the
 2 samples. The Y73C mutation is known to be unique in B.1.1.7 (33). Although no clear
 3 relevance of K68stop is known, it is reported to be present in SARS-CoV-2 genomes with the
 4 highest number of spike protein mutations (32). The P681H spike protein mutation was the
 5 third most abundant across the samples. In total, 13 spike protein mutations were amongst the
 6 dominant mutations (Fig. 4).

7



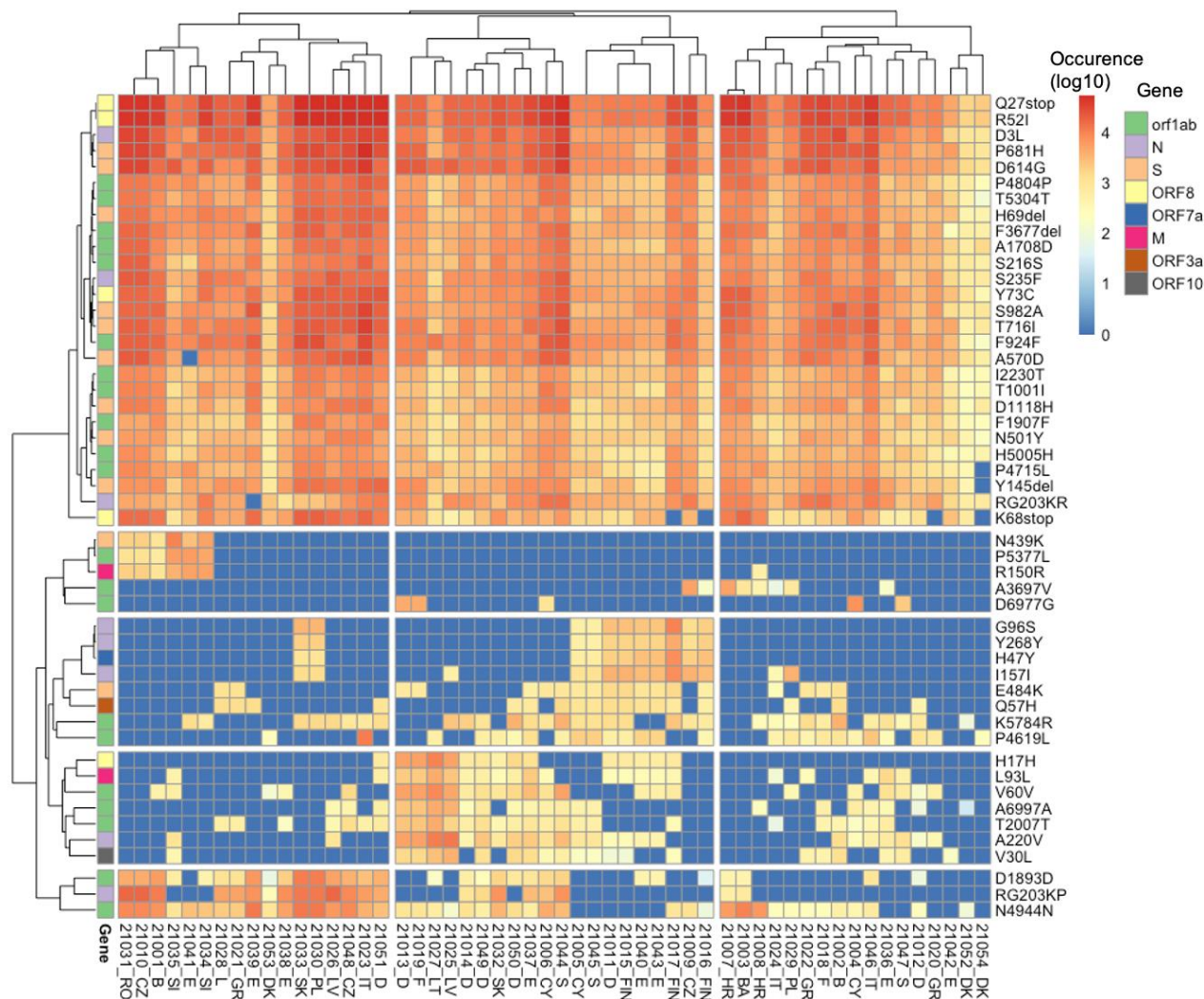
8

9 **Fig 3: Abundance of the reads associated with Spike protein AA mutations used by the**
 10 **ECDC for the characterization of the VOCs.**

11

12 All spike protein mutations, along with the earlier mentioned ORF8 mutations, have been
 13 found in the B.1.1.7 variant in clinical samples (32). E484K was observed in all wastewater
 14 samples from France and Sweden. This E484K mutation has been found in B.1.1.7, B.1.351,
 15 and P.1(34). Eight mutations of the nucleocapsid protein were also dominant in the samples,
 16 especially D3L and S235F. Both of these mutations are known to likely alter the stability and
 17 immunogenic properties of the N protein (35) and are signature mutations of B.1.1.7 (36). In
 18 the ORF1ab region, 21 mutations were abundant in the samples. The F924F and P4715L
 19 ORF1ab mutations have been reported to have a strong allelic association with D614G in
 20 variants dominant in Europe (37, 38).

1



2

3 **Fig 4: Heatmap** showing the read abundance of the top 50 AA mutations found across all the
4 samples.

5

6 **Wastewater based Epidemiology reveals Genomic Variants**

7 The pattern of genomic variants and the abundance of VOCs were consistent between the
8 clinical and the wastewater sequencing data, especially for dominant mutations (S.fig.2,
9 Fig.4). For example, both data-sets show a high read abundance of the D614G mutation.
10 Similarly, P681H was also dominant in clinical and wastewater samples (S.fig.2, Fig.4). On
11 the other hand, wastewater sequencing data can also reveal genomic variants which are not
12 reported as dominant in clinical data. For example, Q27stop was one of the dominant
13 mutations in wastewater samples but not in the clinical samples of all countries (S.fig.2,
14 Fig.4). The ORF8 mutations (such as Q27stop, R52I, K68stop) have been emphasized to be
15 relevant and require closure attention (32, 39), especially because they occur recurrently (30,
16 32). However, at present the count of the ORF

1 8 mutations in genomes deposited in GISAID (<https://www.gisaid.org>) is low, which
2 according to a previous study is due to bias in samples sequenced (32). Across most of the
3 wastewater samples, we detected a high occurrence of ORF8 mutations (i.e. Q27stop, R52I)
4 (Fig. 4), which provides evidence for the circulation of SARS-COV-2 variants containing
5 these mutations in the sampled regions. Also, wastewater sequencing data can reveal spatial
6 prevalence of mutations. For example, the Q57H mutation was dominant in all three samples
7 from Finland, which is consistent with clinical genomic data of Finland (40). This
8 emphasizes that sequencing for SARS-CoV-2 in wastewater can provide additional
9 information about the prevalence of mutations (such as Q27stop), which might not currently
10 appear abundant based on clinical data. Also relative abundance data grouping the mutations
11 that are associated with a particular variant revealed a clear dominance of B.1.1.7 among the
12 samples, followed by mutations associated with B.1.351, P.1, and B.1.617 (S.fig.5), which is
13 similar to the clinical sequencing data.

14 The results obtained in this study clearly show that surveillance of SARS-CoV-2 mutation
15 profiles associated with VOCs in wastewater samples is possible using NGS. The data
16 generated also presents the possibility to attain a sufficient coverage of the SARS-CoV-2
17 genome from the wastewater samples. In wastewater samples, a mixture of genomic material
18 of multiple SARS-CoV-2 variants may be present. Nevertheless, it is still possible to obtain
19 information about variants based on the prevalence of key mutations of the respective variant.
20 However, it is essential to note that sequencing surveillance of wastewater samples should be
21 considered complementary information to whole-genome sequencing of clinical samples.

References

1. J. Martin, D. Klapsa, T. Wilton, M. Zambon, E. Bentley, E. Bujaki, M. Fritzsche, R. Mate, M. Majumdar, Tracking SARS-CoV-2 in Sewage: Evidence of Changes in Virus Variant Predominance during COVID-19 Pandemic. *Viruses*. 12, 1144 (2020).
2. M. Worobey, J. Pekar, B. B. Larsen, M. I. Nelson, V. Hill, J. B. Joy, A. Rambaut, M. A. Suchard, J. O. Wertheim, P. Lemey, The emergence of SARS-CoV-2 in Europe and North America. *Science*. 370, 564–570 (2020).
3. K. D. McCormick, J. L. Jacobs, J. W. Mellors, The emerging plasticity of SARS-CoV-2. *Science*. 371, 1306–1308 (2021).
4. W. T. Harvey, A. M. Carabelli, B. Jackson, R. K. Gupta, E. C. Thomson, E. M. Harrison, C. Ludden, R. Reeve, A. Rambaut, S. J. Peacock, D. L. Robertson, SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*, 1–16 (2021).
5. A. Rambaut, E. C. Holmes, V. Hill, Á. O’Toole, J. McCrone, C. Ruis, L. du Plessis, O. G. Pybus, “A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology” (preprint, *Microbiology*, 2020), , doi:10.1101/2020.04.17.046086.
6. V. Priesemann, R. Balling, M. M. Brinkmann, S. Ciesek, T. Czypionka, I. Eckerle, G. Giordano, C. Hanson, Z. Hel, P. Hotulainen, P. Klimek, A. Nassehi, A. Peichl, M. Perc, E. Petelos, B. Prainsack, E. Szczurek, An action plan for pan-European defence against new SARS-CoV-2 variants. *The Lancet*. 397, 469–470 (2021).
7. D. Cyranoski, Alarming COVID variants show vital role of genomic surveillance. *Nature*. 589, 337–338 (2021).
8. World Health Organization, Weekly epidemiological update on COVID-19 - 1 June 2021, (available at <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---1-june-2021>).
9. European Centre for Disease Prevention and Control, SARS-CoV-2 variants of concern as of 3 June 2021. European Centre for Disease Prevention and Control, (available at <https://www.ecdc.europa.eu/en/covid-19/variants-concern>).
10. European Commission, “Commission Recommendation (EU) 2021/472 on a common approach to establish a systematic surveillance of SARS-CoV-2 and its variants in wastewaters in the EU” (C/2021/1925), (available at <https://www.europeansources.info/record/commission-recommendation-eu-2021-472-on-a-common-approach-to-establish-a-systematic-surveillance-of-sars-cov-2-and-its-variants-in-wastewaters-in-the-eu/>).
11. Detection and characterisation capability and capacity for SARS-CoV-2 variants within the EU/EEA. European Centre for Disease Prevention and Control (2021), (available at <https://www.ecdc.europa.eu/en/publications-data/detection-and-characterisation-capability-and-capacity-sars-cov-2-variants>).
12. S. Agrawal, L. Orschler, S. Lackner, Long-term monitoring of SARS-CoV-2 RNA in wastewater of the Frankfurt metropolitan area in Southern Germany. *Sci Rep*. 11, 5372 (2021).
13. G. Medema, L. Heijnen, G. Elsinga, R. Italiaander, A. Brouwer, “Presence of SARS-Coronavirus-2 in sewage” (preprint, *Occupational and Environmental Health*, 2020), , doi:10.1101/2020.03.29.20045880.
14. G. Medema, L. Heijnen, G. Elsinga, R. Italiaander, A. Brouwer, Presence of SARS-

- 1 Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in
2 the Early Stage of the Epidemic in The Netherlands. *Environ. Sci. Technol. Lett.* (2020),
3 doi:10.1021/acs.estlett.0c00357.
- 4 15. W. Ahmed, N. Angel, J. Edson, K. Bibby, A. Bivins, J. W. O'Brien, P. M. Choi, M.
5 Kitajima, S. L. Simpson, J. Li, B. Tschärke, R. Verhagen, W. J. M. Smith, J. Zaugg, L.
6 Dierens, P. Hugenholtz, K. V. Thomas, J. F. Mueller, First confirmed detection of
7 SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the
8 wastewater surveillance of COVID-19 in the community. *Science of The Total
9 Environment*. 728, 138764 (2020).
- 10 16. M. Kitajima, W. Ahmed, K. Bibby, A. Carducci, C. P. Gerba, K. A. Hamilton, E.
11 Haramoto, J. B. Rose, SARS-CoV-2 in wastewater: State of the knowledge and research
12 needs. *Science of The Total Environment*, 139076 (2020).
- 13 17. G. La Rosa, M. Iaconelli, P. Mancini, G. Bonanno Ferraro, C. Veneri, L. Bonadonna, L.
14 Lucentini, E. Suffredini, First detection of SARS-CoV-2 in untreated wastewaters in
15 Italy. *Science of The Total Environment*. 736, 139652 (2020).
- 16 18. S. Agrawal, L. Orschler, S. Lackner, Metatranscriptomic Analysis Reveals SARS-CoV-2
17 Mutations in Wastewater of the Frankfurt Metropolitan Area in Southern Germany.
18 *Microbiol Resour Announc*. 10 (2021), doi:10.1128/MRA.00280-21.
- 19 19. A. Nemudryi, A. Nemudraia, T. Wiegand, K. Surya, M. Buyukyoruk, C. Cicha, K. K.
20 Vanderwood, R. Wilkinson, B. Wiedenheft, Temporal Detection and Phylogenetic
21 Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Reports Medicine*, 100098
22 (2020).
- 23 20. K. Jahn, D. Dreifuss, I. Topolsky, A. Kull, P. Ganesanandamoorthy, X. Fernandez-Cassi,
24 C. Bänziger, E. Stachler, L. Fuhrmann, K. P. Jablonski, C. Chen, C. Aquino, T. Stadler,
25 C. Ort, T. Kohn, T. R. Julian, N. Beerenwinkel, "Detection of SARS-CoV-2 variants in
26 Switzerland by genomic analysis of wastewater samples" (preprint, *Infectious Diseases
27 (except HIV/AIDS)*, 2021), , doi:10.1101/2021.01.08.21249379.
- 28 21. R. W. I. Lara, G. Elsinga, L. Heijnen, B. B. O. Munnink, C. M. E. Schapendonk, D.
29 Nieuwenhuijse, M. Kon, L. Lu, F. M. Aarestrup, S. Lycett, G. Medema, M. P. G.
30 Koopmans, M. de Graaf, medRxiv, in press, doi:10.1101/2020.09.21.20198838.
- 31 22. B. Gawlik, S. Tavazzi, G. Mariani, H. Skejo, M. Sponar, T. Higgins, G. Medema, T.
32 Wintgens, SARS-CoV-2 surveillance employing sewage : towards a sentinel system.
33 (Publications Office of the European Union, 2021; [http://op.europa.eu/en/publication-
34 detail/-/publication/816cec6d-abbd-11eb-927e-01aa75ed71a1/language-en](http://op.europa.eu/en/publication-detail/-/publication/816cec6d-abbd-11eb-927e-01aa75ed71a1/language-en)).
- 35 23. European Commission, "Communication from the commission to the European
36 parliament, the European council and the council HERA Incubator: Anticipating together
37 the threat of COVID-19 variants" (COM/2021/78 final), (available at [https://eur-
38 lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021DC0078&from=en](https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021DC0078&from=en)).
- 39 24. S. Vilar, D. G. Isom, One Year of SARS-CoV-2: How Much Has the Virus Changed?
40 *Biology*. 10, 91 (2021).
- 41 25. B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N.
42 Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, K. M. Hastie, M. D. Parker, D. G.
43 Partridge, C. M. Evans, T. M. Freeman, T. I. de Silva, C. McDanal, L. G. Perez, H. Tang,
44 A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Sapphire, D. C. Montefiori, A.
45 Angyal, R. L. Brown, L. Carrilero, L. R. Green, D. C. Groves, K. J. Johnson, A. J.

- 1 Keeley, B. B. Lindsey, P. J. Parsons, M. Raza, S. Rowland-Jones, N. Smith, R. M.
2 Tucker, D. Wang, M. D. Wyles, Tracking Changes in SARS-CoV-2 Spike: Evidence that
3 D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 182, 812-827.e19 (2020).
- 4 26. F. G. Naveca, V. Nascimento, V. C. de Souza, A. de L. Corado, F. Nascimento, G. Silva,
5 Á. Costa, D. Duarte, K. Pessoa, M. Mejía, M. J. Brandão, M. Jesus, L. Gonçalves, C. F.
6 da Costa, V. Sampaio, D. Barros, M. Silva, T. Mattos, G. Pontes, L. Abdalla, J. H.
7 Santos, I. Arantes, F. Z. Dezordi, M. M. Siqueira, G. L. Wallau, P. C. Resende, E.
8 Delatorre, T. Gräf, G. Bello, COVID-19 in Amazonas, Brazil, was driven by the
9 persistence of endemic lineages and P.1 emergence. *Nat Med*, 1–9 (2021).
- 10 27. Sk. S. Hassan, D. Attrish, S. Ghosh, P. P. Choudhury, B. Roy, Pathogenic perspective of
11 missense mutations of ORF3a protein of SARS-CoV-2. *Virus Res*. 300, 198441 (2021).
- 12 28. A. Winger, T. Caspari, The Spike of Concern—The Novel Variants of SARS-CoV-2.
13 *Viruses*. 13, 1002 (2021).
- 14 29. European Centre for Disease Prevention and Control, “Risk realted to the spread of new
15 SARS-CoV-2 variants of concern in the EU/EEA - first update” (Stockholm, 2021), p.
16 29.
- 17 30. T. P. Peacock, R. Penrice-Randal, J. A. Hiscox, W. S. Barclay, SARS-CoV-2 one year
18 on: evidence for ongoing viral adaptation. *Journal of General Virology*. 102 (2021),
19 doi:10.1099/jgv.0.001584.
- 20 31. C. W. Nelson, Z. Ardern, T. L. Goldberg, C. Meng, C.-H. Kuo, C. Ludwig, S.-O.
21 Kolokotronis, X. Wei, A previously uncharacterized gene in SARS-CoV-2 illuminates
22 the functional dynamics and evolutionary origins of the COVID-19 pandemic, 32.
- 23 32. F. Pereira, SARS-CoV-2 variants combining spike mutations and the absence of ORF8
24 may be more transmissible and require close monitoring. *Biochem Biophys Res*
25 *Commun*. 550, 8–14 (2021).
- 26 33. J. Singh, J. Samal, V. Kumar, J. Sharma, U. Agrawal, N. Z. Ehtesham, D. Sundar, S. A.
27 Rahman, S. Hira, S. E. Hasnain, Structure-Function Analyses of New SARS-CoV-2
28 Variants B.1.1.7, B.1.351 and B.1.1.28.1: Clinical, Diagnostic, Therapeutic and Public
29 Health Implications. *Viruses*. 13, 439 (2021).
- 30 34. H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D.
31 Doolabh, S. Pillay, E. J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M.
32 Allam, A. Ismail, T. Mohale, A. J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, F.
33 Petruccione, A. Sigal, D. Hardie, G. Marais, M. Hsiao, S. Korsman, M.-A. Davies, L.
34 Tyers, I. Mudau, D. York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A.
35 Alisoltani-Dehkordi, A. Godzik, C. K. Wibmer, B. T. Sewell, J. Lourenço, L. C. J.
36 Alcantara, S. L. K. Pond, S. Weaver, D. Martin, R. J. Lessells, J. N. Bhiman, C.
37 Williamson, T. de Oliveira, “Emergence and rapid spread of a new severe acute
38 respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike
39 mutations in South Africa” (preprint, *Epidemiology*, 2020), ,
40 doi:10.1101/2020.12.21.20248640.
- 41 35. G. K. Azad, Identification and molecular characterization of mutations in nucleocapsid
42 phosphoprotein of SARS-CoV-2. *PeerJ*. 9, e10666 (2021).
- 43 36. S. E. Galloway, Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States,
44 December 29, 2020–January 12, 2021. *MMWR Morb Mortal Wkly Rep*. 70 (2021),
45 doi:10.15585/mmwr.mm7003e2.

- 1 37. H.-C. Yang, C. Chen, J.-H. Wang, H.-C. Liao, C.-T. Yang, C.-W. Chen, Y.-C. Lin, C.-H.
2 Kao, M.-Y. J. Lu, J. C. Liao, Analysis of genomic distributions of SARS-CoV-2 reveals
3 a dominant strain type with strong allelic associations. *PNAS*. 117, 30679–30686 (2020).
4 38. T. Koyama, D. Platt, L. Parida, Variant analysis of SARS-CoV-2 genomes. *Bull World*
5 *Health Organ*. 98, 495–504 (2020).
6 39. I. Jungreis, R. Sealfon, M. Kellis, SARS-CoV-2 gene content and COVID-19 mutation
7 impact by comparing 44 Sarbecovirus genomes. *Nature Communications*. 12, 2642
8 (2021).
9 40. J.-Y. Lam, C.-K. Yuen, J. D. Ip, W.-M. Wong, K. K.-W. To, K.-Y. Yuen, K.-H. Kok,
10 Loss of orf3b in the circulating SARS-CoV-2 strains. *Emerging Microbes & Infections*.
11 9, 2685–2696 (2020).

12 13 **Acknowledgement**

14 We gratefully acknowledge the contribution from the originating laboratories responsible for
15 obtaining the specimens and the submitting laboratories where genetic sequence data were
16 generated and shared via the GISAID Initiative (<https://www.gisaid.org>). We thank all
17 WWTP operators for providing wastewater samples, We also thank Ray Izquierdo-Lara for
18 critically reading the manuscript.

Supplementary Material

A pan-European study of SARS-CoV-2 variants in wastewater under the EU Sewage Sentinel System

Shelesh Agrawal¹, Laura Orschler¹, Selina Schubert¹, Kira Zachmann¹, Leo Heijnen², Simona Tavazzi³, Bernd Manfred Gawlik³, Miranda de Graaf⁴, Gertjan Medema², Susanne Lackner^{1*}

Affiliations:

¹ Technical University of Darmstadt, Department of Civil and Environmental Engineering Sciences, Institute IWAR, Chair of Wastewater Engineering; Darmstadt, Germany.

² KWR Water Research Institute; Nieuwegein, The Netherlands.

³ European Commission, Joint Research Centre; Ispra (Va), Italy.

⁴ Department of Viroscience, Erasmus Medical Center; Rotterdam, The Netherlands

*Corresponding author. Email: susanne.lackner@tu-darmstadt.de

Material and Methods

Sequencing

For this study, 24 h composite wastewater samples were collected between weeks 10 – 13 of 2021 (i.e. 10th to 30th March 2021) and shipped TU Darmstadt (Darmstadt, Germany), packed with icepacks (approximately at 6°C), for sequencing analysis. In Darmstadt, one litre of the untreated wastewater was filtered through a 0.45 µm electronegative membrane filter to concentrate the SARS-CoV-2 RNA, followed by extraction using the Fast RNA Blue Kit (MP Biomedicals) according to the manufacturer’s protocol. Another 500 ml of the untreated wastewater was concentrated by ultrafiltration in 100 kDa Centricon® Plus-70 centrifugal ultrafilters (Merck) and RNA was extracted using the Ultra Microbiome kit (ThermoFisher Scientific) according to the manufacturer’s protocol. Both RNA extracts were pooled together for downstream analysis. From the pooled RNA, cDNA was synthesized using SuperScript™ VILO™ Master Mix (ThermoFisher Scientific), followed by library preparation using the Ion AmpliSeq SARS-CoV-2 Research Panel (ThermoFisher Scientific) according to manufacturer’s instructions. This panel consists of 237 primer pairs, resulting in an amplicon length range of 125–275 bp, which cover the near-full genome of SARS-CoV-2. We performed multiple sequencing runs to achieve a high number of reads per sample. For each sequencing run, eight libraries were multiplexed and sequenced using an Ion Torrent 530 chip on an Ion S5 sequencer (ThermoFisher Scientific) according to manufacturer’s instructions.

We used the SARS-CoV-2 Research Plug-in Package, which we installed in our Ion Torrent Suite software (v5.12.2) of Ion S5 sequence. We used the SARS_CoV_2_coverageAnalysis (v5.16) plugin, which maps the generated reads to a SARS-CoV-2 reference genome (Wuhan-Hu-1-NC_045512/MN908947.3), using TMAP software included in the Torrent Suite. The summary of mapping of each sample is provided in S.Table 1. For mutation calls, additional Ion Torrent plugins were used, similar to our previous study (1). First, all single nucleotide variants (SNVs) were called using Variant Caller (v5.12.0.4) with “Generic - S5/S5XL (510/520/530) - Somatic - Low Stringency” default parameters. Then, for annotation and determination of the base substitution effect, COVID19AnnotateSnEff (v1.3.0.2), a plugin developed explicitly for SARS-CoV-2, was used. The raw metagenomic sequence data were uploaded to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under Submission ID SUB9829162, BioProject number PRJNA736964.

S.Table 1: Summary of sequencing data for all samples.

S.No.	Country	SampleID	Total number of reads	No. of mapped reads	Avg read identity to target (%)*	S.No.	Country	SampleID	Total number of reads	No. of mapped reads	Avg read identity to target (%)*
1	BELGIUM	21001_B	2,808,905	2,088,647	99.55	28	GREECE	21022_GR	2,786,455	1,788,408	99.48
2	BELGIUM	21002_B	2,680,409	1,893,060	99.53	29	GREECE	21020_GR	2,468,367	1,110,096	99.32
3	BOSNIA ERZEGOVINA	21003_BA	2,508,512	1,766,499	99.51	30	ITALY	21023_IT	3,496,150	2,969,772	99.52
4	CROATIA	21007_HR	2,552,070	1,870,713	99.64	31	ITALY	21024_IT	1,725,352	818,506	99.48
5	CROATIA	21008_HR	2,111,859	1,273,558	99.55	32	ITALY	21046_IT	3,260,879	2,256,762	99.52
6	CYPRUS	21004_CY	2,686,062	1,607,809	99.57	33	LATVIA	21026_LV	3,546,039	2,496,946	99.54
7	CYPRUS	21005_CY	2,044,674	829,904	98.07	34	LATVIA	21025_LV	2,063,000	1,247,974	99.39
8	CYPRUS	21006_CY	3,150,002	2,428,422	99.53	35	LITHUNIA	21027_LT	2,596,024	1,391,166	99.63
9	CZECH REPUBLIC	21009_CZ	3,290,470	2,840,707	99.21	36	LUXEMBOURG	21028_L	2,554,722	1,565,165	99.72
10	CZECH REPUBLIC	21010_CZ	3,754,585	2,804,456	99.55	37	POLAND	21030_PL	3,538,732	3,215,972	99.32
11	CZECH REPUBLIC	21048_CZ	3,436,877	2,688,684	99.57	38	POLAND	21029_PL	2,209,073	1,375,405	99.57
12	DENMARK	21052_DK	3,273,309	763,257	99.37	39	ROMANIA	21031_RO	2,784,858	2,119,141	99.64
13	DENMARK	21053_DK	2,547,875	817,493	99.58	40	SLOVAK REPLUBLIC	21033_SK	3,345,138	2,507,501	99.58
14	DENMARK	21054_DK	2,608,811	809,437	98.17	41	SLOVAK REPUBLIC	21032_SK	2,589,970	1,469,732	98.66
15	FINLAND	21015_FIN	2,539,762	1,135,620	99.49	42	SLOVENIA	21034_SI	2,838,240	1,941,543	99.52
16	FINLAND	21017_FIN	2,831,785	1,852,350	99.60	43	SLOVENIA	21035_SI	2,470,338	1,544,585	99.63
17	FINLAND	21016_FIN	1,918,714	776,343	99.55	44	SPAIN	21039_E	2,683,669	1,964,724	99.52
18	FRANCE	21018_F	2,731,094	1,892,165	99.48	45	SPAIN	21040_E	2,284,844	1,014,470	99.54
19	FRANCE	21019_F	2,485,475	1,568,069	99.52	46	SPAIN	21043_E	1,935,880	695,817	98.89
20	GERMANY	21011_D	2,421,764	1,150,371	99.51	47	SPAIN	21047_S	2,280,905	1,133,647	99.45
21	GERMANY	21013_D	2,204,439	1,523,263	99.51	48	SPAIN	21036_E	2,252,726	1,019,145	99.52
22	GERMANY	21051_D	2,862,135	2,144,701	99.54	49	SPAIN	21041_E	2,484,336	1,410,983	99.21
23	GERMANY	21012_D	1,788,643	810,402	99.52	50	SPAIN	21042_E	1,897,633	601,931	99.02
24	GERMANY	21014_D	2,914,973	1,378,621	99.58	51	SPAIN	21037_E	2,913,114	1,747,719	99.50
25	GERMANY	21050_D	3,253,048	1,638,447	99.52	52	SPAIN	21038_E	2,426,135	1,386,068	99.53
26	GERMANY	21049_D	2,532,868	1,200,315	99.52	53	SWEDEN	21044_S	3,462,193	2,738,247	99.60
27	GREECE	21021_GR	2,022,878	818,984	99.45	54	SWEDEN	21045_S	2,932,821	1,063,215	99.63

*The target sequence was the SARS-CoV-2 reference genome (Wuhan-Hu-1 [GenBank accession numbers NC_045512 and MN908947.3]).

qPCR Methods

Samples were received at the KWR laboratory and processed as previously described (2). The N2 assay targeting a fragment of the nucleocapsid gene, as published by US CDC (US-CDC 2020), was used to quantify SARS-CoV-2 RNA in the sewage samples. All RT-PCR's were run as technical duplicates on 5 µl extracted nucleic acid. RT-qPCR reactions on serial dilutions containing RT-ddPCR calibrated EURM-019 single stranded RNA (provided by the Joint Research Centre) were used to construct calibration curves that subsequently were used to quantify N2 in RNA extracted from the sewage samples. Reactions were considered positive if the cycle threshold was below 40 cycles.

CrAssphage CPQ_064 specific PCR (3) was used to quantify this DNA-virus that is ubiquitously present exclusively occurs in human intestinal tracts in high concentrations. Assays were performed in duplicate on 5 µl 1:10 diluted extracted nucleic acid. Quantification was performed using PCR assays on dilution series of a synthetic quantified gBlock (obtained from IDT, Leuven, Belgium) containing the CPQ_064 gene fragment.

Data Analysis

We downloaded the variant surveillance data package from GISAID on 31th May 2021. This data package consists of information about the identified variants, the corresponding amino acid (AA) mutations and sample location. We filtered the dataset, limiting it to human samples with complete coverage. This dataset was used to determine associations between the amino acid (AA) mutations detected in wastewater samples and AA mutations, with their corresponding pangolin lineage, reported from clinical samples. From the GISAID data package, we also determined the fraction of clinical samples reporting the current variants of concern (VOC): (1) B.1.1.7, (2) P.1, (3) B.1.351, (4) and B.1.617.2 (4, 5). Data analysis was performed in R(v3.6.2) using the ggplot (v3.3.3) package for data visualization, and pheatmap (v1.0.12) for hierarchy clustering and heatmap construction.

Mutations in clinical samples

We determined the abundance spike protein mutations, which are suggested by the European Centre for Disease Prevention and Control (ECDC) for the characterization of the current VOCs (5), reported for clinical human samples. We found that the abundance of the mutations varied for the respective country, though associated with the same VOC (S.fig.1). For example, four out of eight spike protein mutations for P.1 for sequences from Slovakia were not reported.

SARS-CoV-2 RNA in wastewater samples

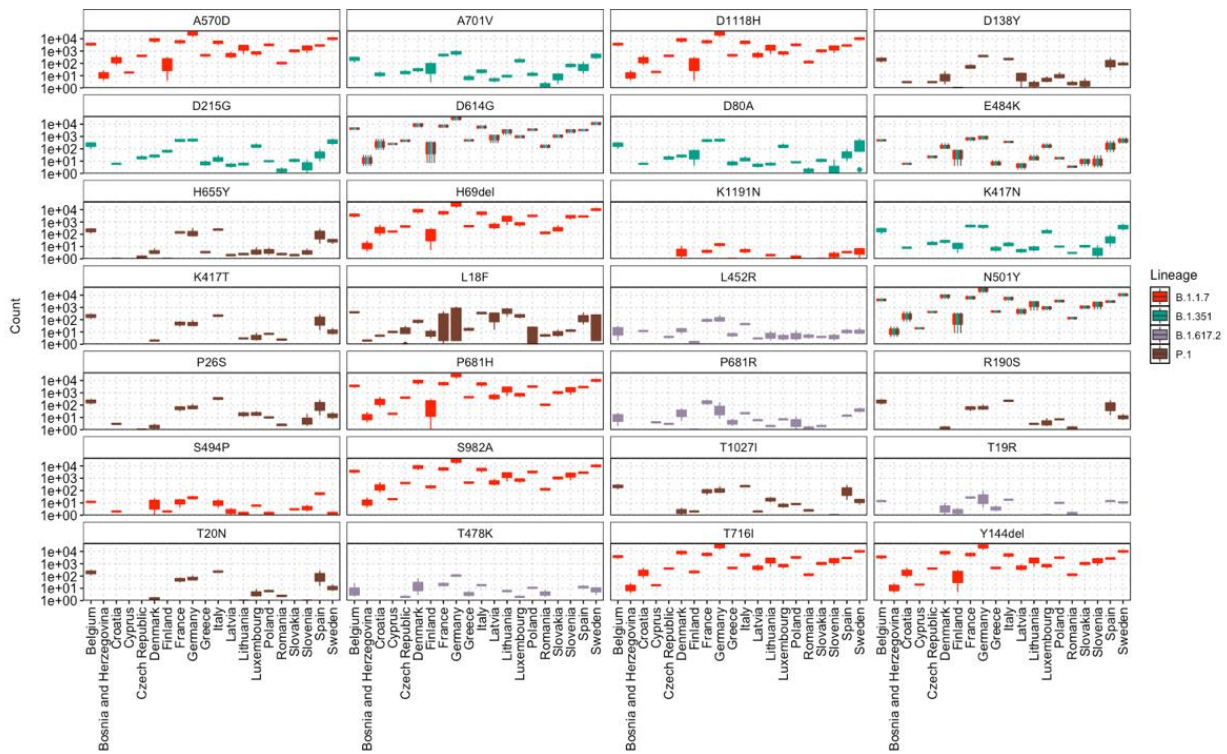
qPCR results

SARS-CoV-2 N2-gene RNA was detected in all 54 samples in concentrations ranging from 0.4 - 735 gene copies/ml (S. fig 3). The concentration protocol for NGS provided sufficient read depth (S. Table 1), even with the low SARS-CoV-2 concentration samples. The crassphage concentration (S. fig 3) is an index of the dilution of human fecal input in wastewater. This varied between the cities; in INF_21002_B and INF_21003_BA the Crassphage concentrations were markedly low, indicating high dilution of human fecal input in these wastewaters. Normalizing the SARS-CoV-2 N2 concentration for this dilution would markedly change the ranking of the wastewaters, which implies that normalization of 'raw' SARS-CoV-2 concentrations in wastewater for the level of dilution of human fecal input is necessary when linking these data to COVID-19 prevalence data.

Fraction of variants of concern in wastewater samples

We determined the relative abundance of the VOCs based on the abundance of reads associated with certain AA mutations. As there is quite an overlap between the AA mutations of different SARS-COV-2 variants, for better determination of the relative abundance of the VOCs, we specifically looked for the abundance of unique and shared AA mutations corresponding to each VOC (S.fig.5). Categorization of mutations as unique or shared was based on the percentage of sequences for associated mutations submitted in GISAID. We looked for percentage of sequences for each mutation for every lineage. Then for each VOC, mutations reported in more than 0.5% of total number of sequences for each VOC were selected. Among these selected mutations, mutations that are associated with more than one lineage were categorized as shared mutation otherwise they were associated with the respective VOC.

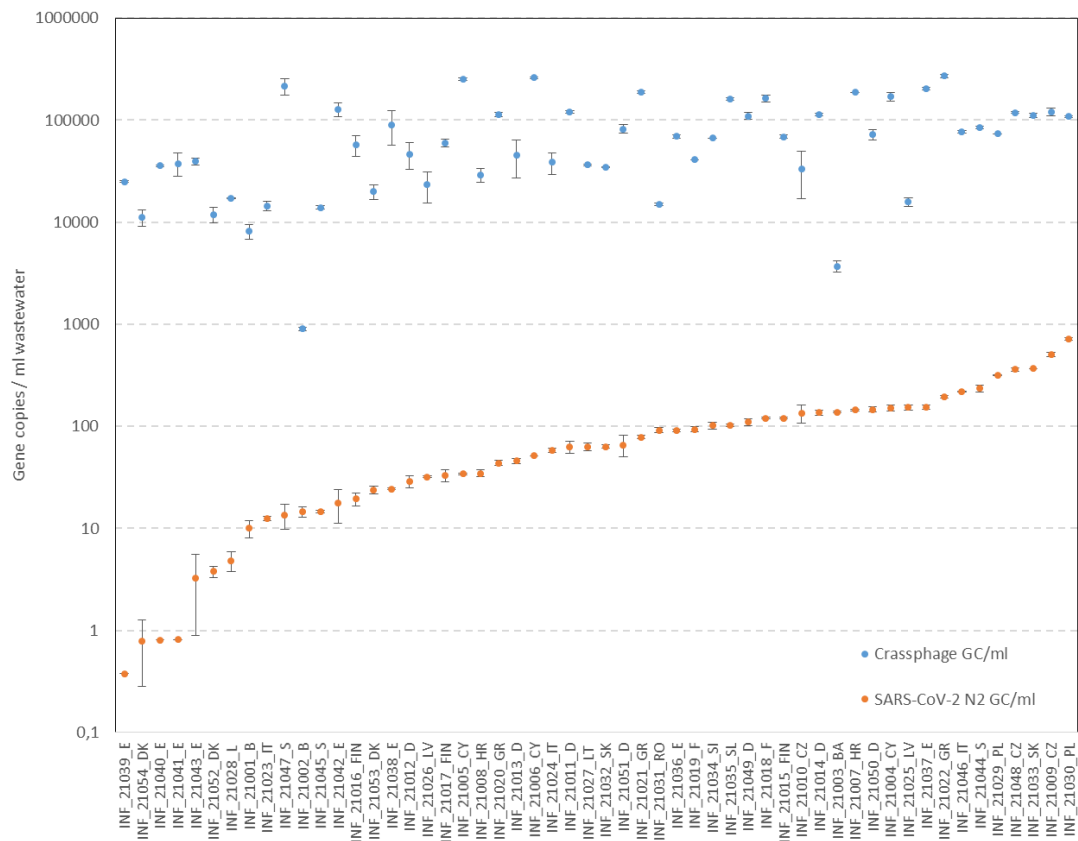
The abundance of mutations associated with B.1.1.7 was highest among the samples ranging from 15 to 40%, followed by abundance of mutations associated with B.1.351, P.1, and B.1.617 (S.fig.5), which is similar to the clinical sequencing data. The mutations associated with B.1.351 were detected in 33 samples, whereas for B.1.617 were detected in 21 samples and for P.1 in 15 samples. B.1.351 mutations were detected in all samples from Finland, Germany, and Sweden; however, the total relative abundance of these mutations varied from 2 to 8%. The relative abundance of the shared AA mutations accounted for 55 to 70% across all samples (S.fig.5).



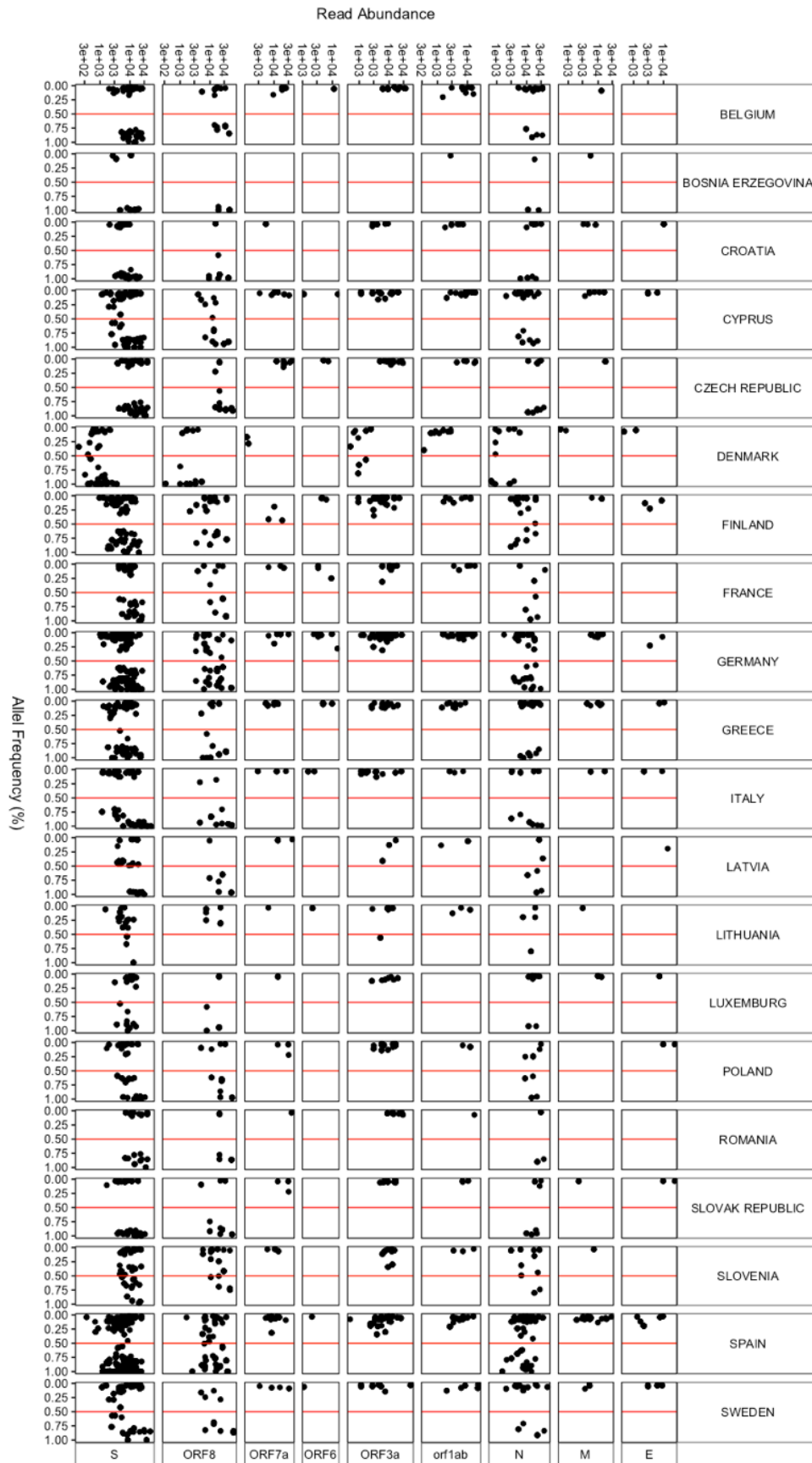
S.fig. 1: Count (occurrence of mutation in number of genome sequences submitted in GISAID) of spike protein mutations, which are considered for characterization of VOC by ECDC (5,6), found in the variant surveillance data for clinical samples of GISAID dated 31st May 2021. The counts are presented in log₁₀ scale.



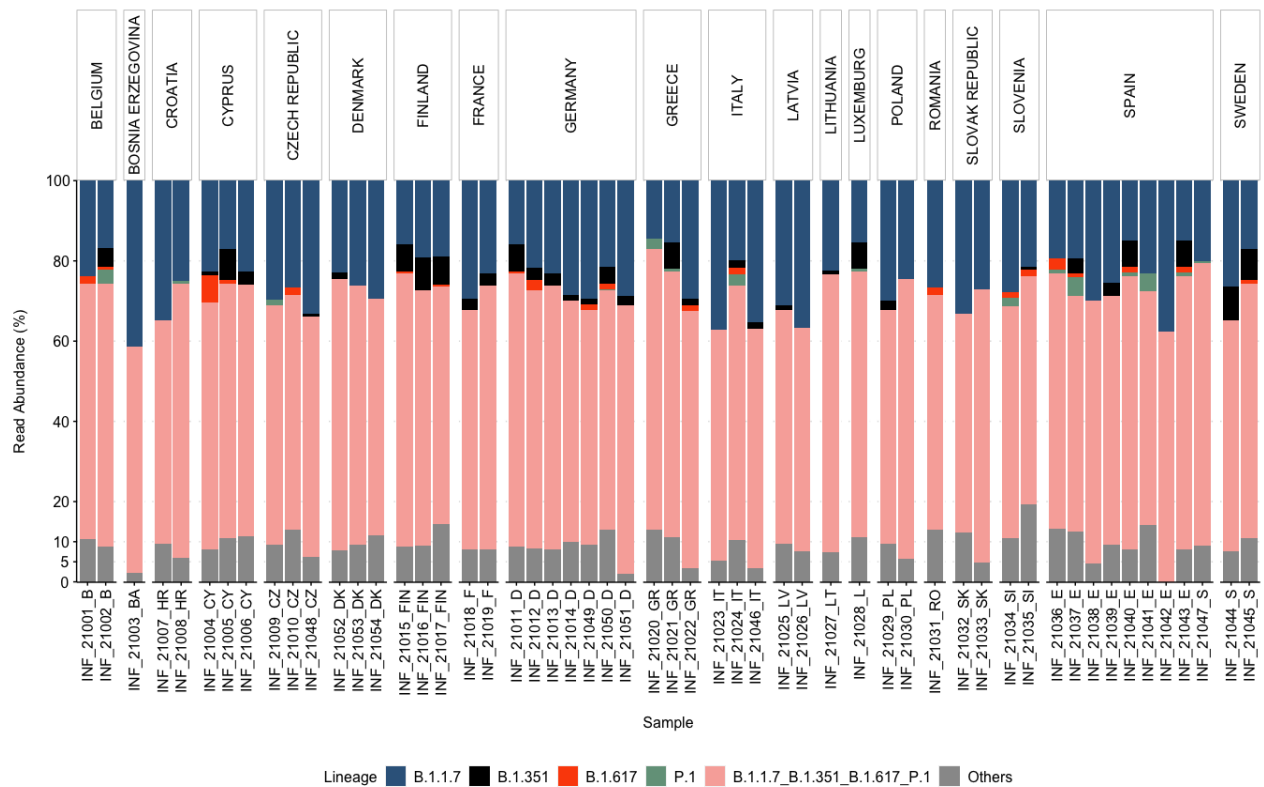
S.fig. 2: Heatmap representing the top 20 abundant mutations in each country during different time period, based on the count (occurrence of mutation in genome sequences submitted in GISAID) of the mutations.



S.fig. 3: qPCR based analysis showing the concentration of SARS-CoV-2 N2 gene copies detected in each sample.



S.fig. 4: Scatter plot showing the distribution of the read abundance of each mutation against the allele frequency of each mutation.



S.fig. 5: Relative abundance of the variants of concern (VOC) and other variants, based on the abundance of the reads associated with each SNP, respectively. AA mutations shared among different SARS-CoV-2 VOCs are represented as “B.1.1.7_B.1.351_B.1.617_P.1”.

Reference

1. S. Agrawal, L. Orschler, S. Lackner, Metatranscriptomic Analysis Reveals SARS-CoV-2 Mutations in Wastewater of the Frankfurt Metropolitan Area in Southern Germany. *Microbiol Resour Announc.* 10 (2021), doi:10.1128/MRA.00280-21.
2. G. Medema, L. Heijnen, G. Elsinga, R. Italiaander, A. Brouwer, Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environ. Sci. Technol. Lett.* (2020), doi:10.1021/acs.estlett.0c00357.
3. E. Stachler, C. Kelty, M. Sivaganesan, X. Li, K. Bibby, O. C. Shanks, Quantitative CrAssphage PCR Assays for Human Fecal Pollution Measurement. *Environ Sci Technol.* 51, 9146–9154 (2017).
4. World Health Organization, Weekly epidemiological update on COVID-19 - 1 June 2021, (available at <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---1-june-2021>).
5. European Centre for Disease Prevention and Control, SARS-CoV-2 variants of concern as of 3 June 2021. European Centre for Disease Prevention and Control, (available at <https://www.ecdc.europa.eu/en/covid-19/variants-concern>).
6. European Centre for Disease Prevention and Control, “Risk related to the spread of new SARS-CoV-2 variants of concern in the EU/EEA - first update” (Stockholm, 2021), p. 29.