**Supplementary Material**

**Title**

# Prevalence and circulation patterns of SARS-CoV-2 variants in European sewage mirror clinical data of 54 European cities

**Authors**

Shelesh Agrawal[1]*, Laura Orschler[1], Selina Schubert[1], Kira Zachmann[1], Leo Heijnen[2], Simona Tavazzi[3], Bernd Manfred Manfred, Miranda de Graaf[4], Gertjan Medema[2], Susanne Lackner[1]

**Affiliations**

[1] Technical University of Darmstadt, Department of Civil and Environmental Engineering Sciences, Institute IWAR, Chair of Water and Environmental Biotechnology; 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: s.agrawal@iwar.tu-darmstadt.de

## 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 (European Centre for Disease Prevention and Control, 2021a), reported for clinical human samples. We found that the abundance of the mutations varied for the respective country, though associated with the same VOC (Fig. S1). 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 (Fig. S3). The concentration protocol for NGS provided sufficient read depth (S. Table 1), even with the low SARS-CoV-2 concentration samples. The crassphage concentration (Fig. S3) 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 (Fig. S5). 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 (Fig. S5), 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 (Fig. S5).

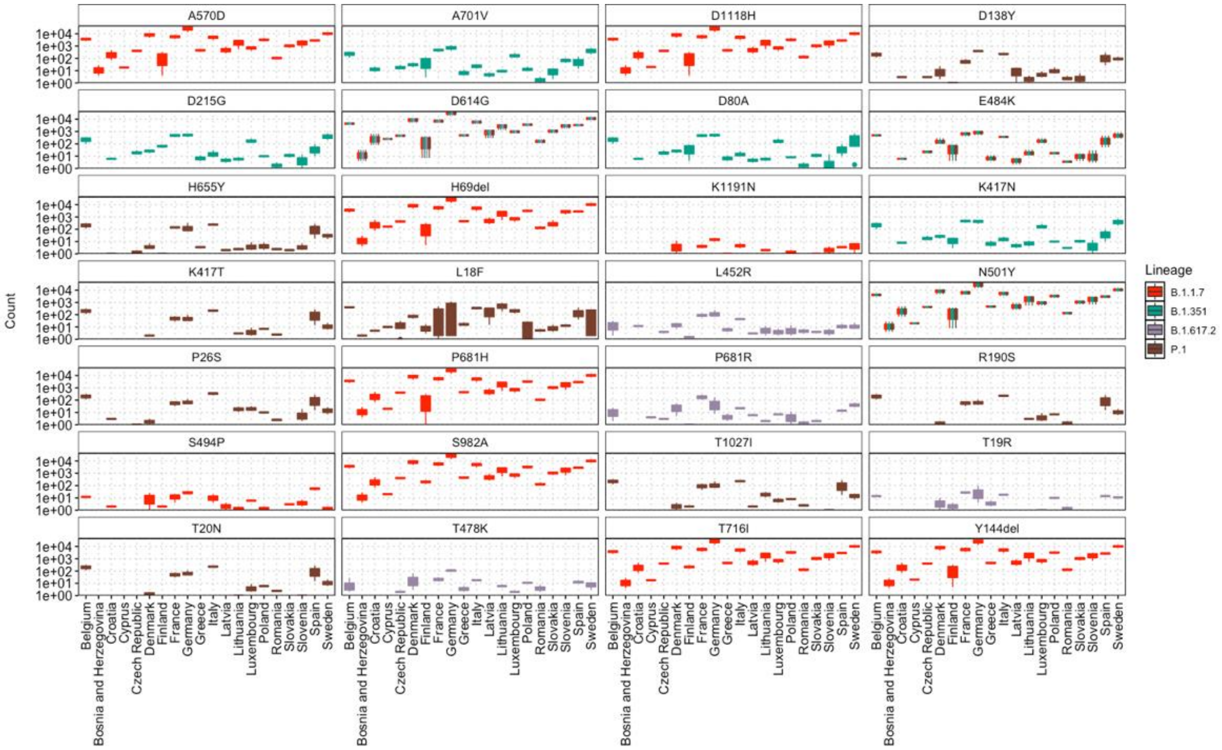


**Fig. S1.** Count (occurrence of mutation in number of genome sequences submitted in GISAID) of spike protein mutations, which are considered for characterization of VOC by ECDC (European Centre for Disease Prevention and Control, 2021b, 2021c), found in the variant surveillance data for clinical samples of GISAID dated 31st May 2021. The counts are presented in log10 scale.

**Fig. S2.** 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.

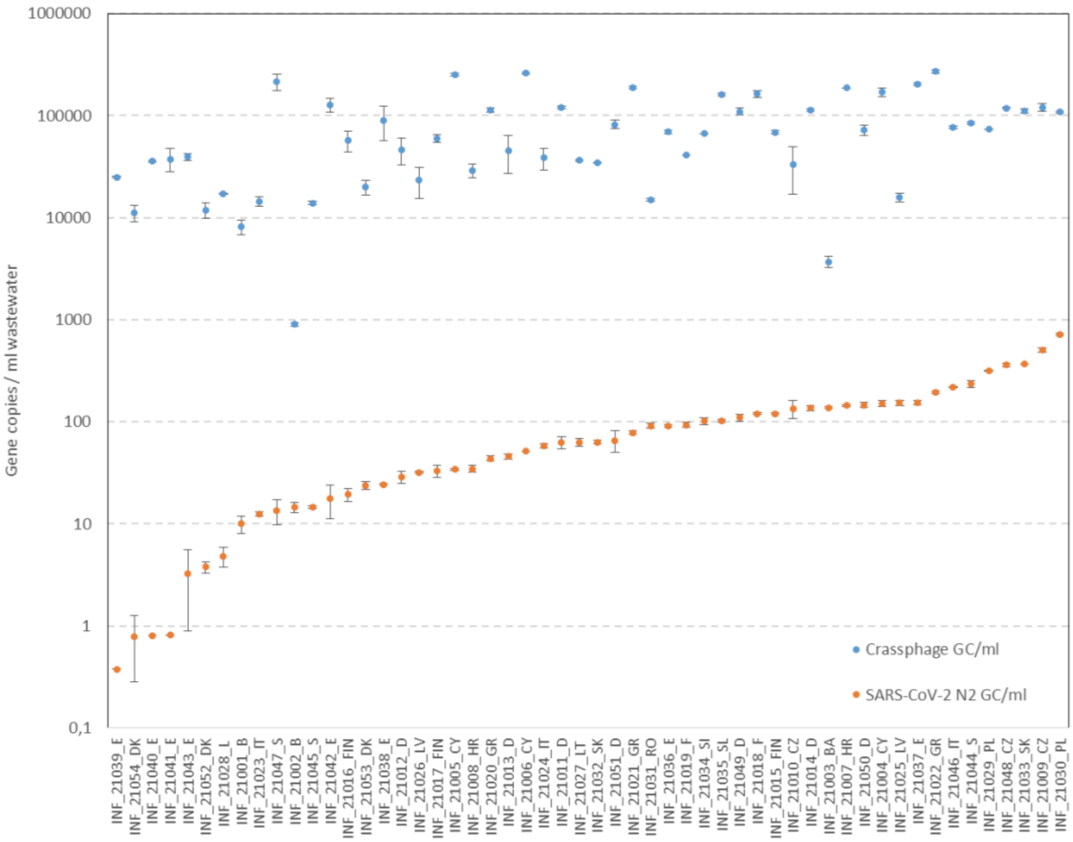


**Fig. S3.** qPCR based analysis showing the concentration of SARS-CoV-2 N2 gene copies detected in each sample.

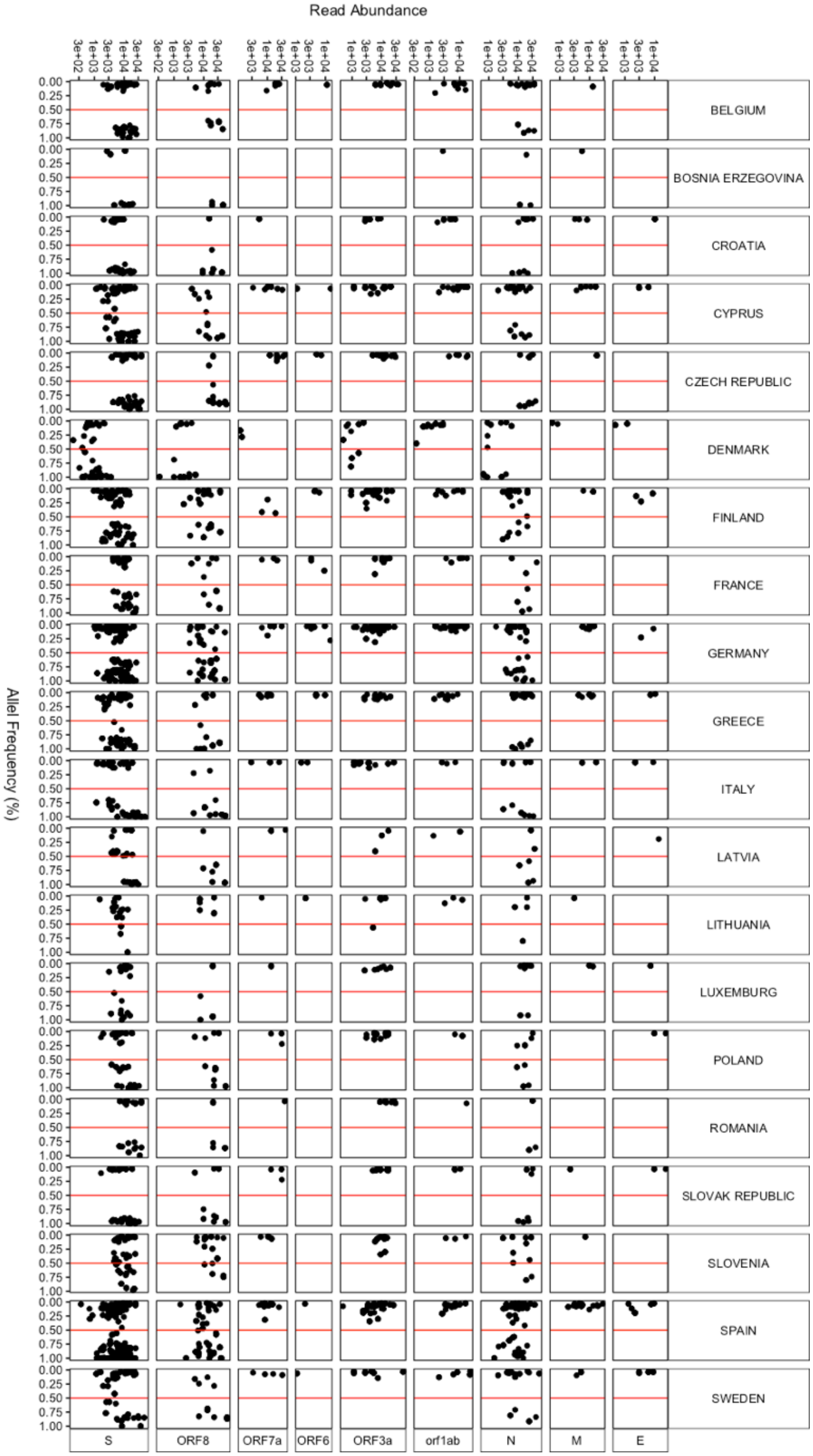


**Fig. S4.** Scatter plot showing the distribution of the read abundance of each mutation against the allele frequency of each mutation.

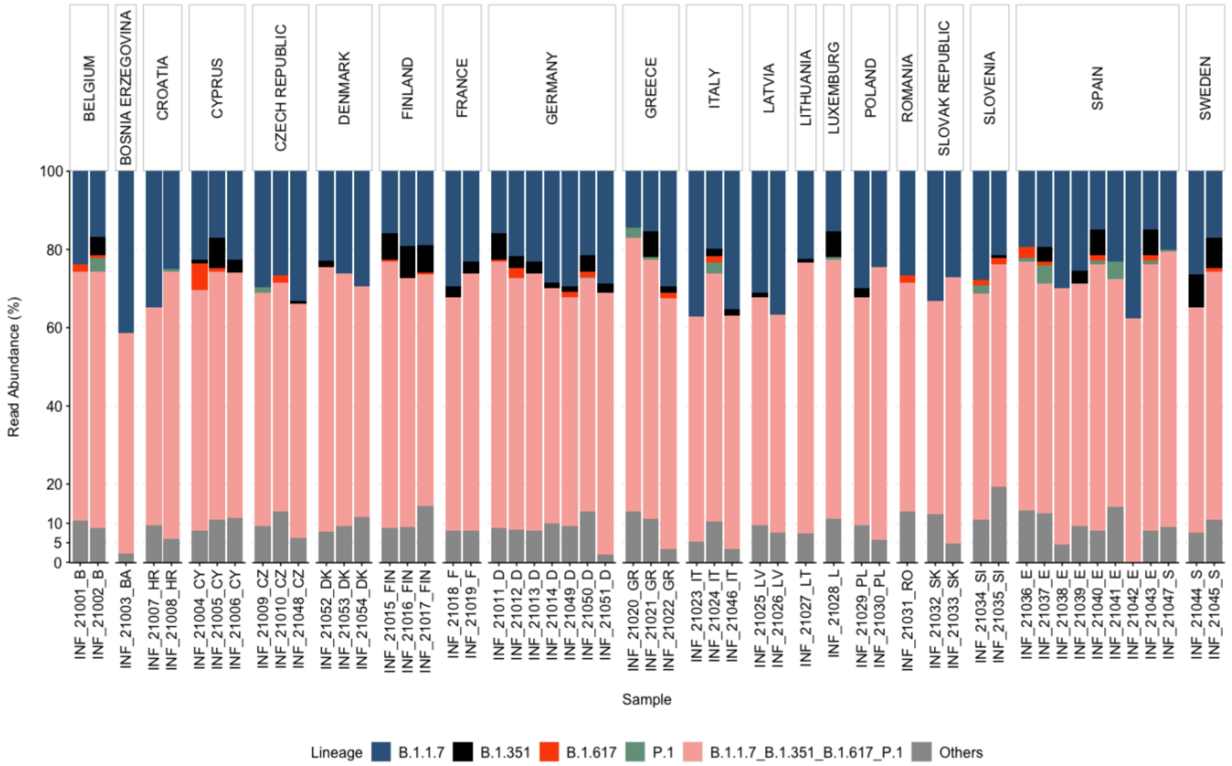


**Fig. S5.** Relative abundance of the variants of concern (VOC) and other variants, based on the abundance of the reads associated with each SNP, respectively. AA mutations shared among different SARS-CoV-2 VOCs are represented as "B.1.1.7_B.1.351_B.1.617_P.1".

**Table S1.** *Summary of sequencing data for all samples.*

| 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 |
| 2 | BELGIUM | 21002_B | 2,680,409 | 1,893,060 | 99.53 |
| 3 | BOSNIA ERZEGOVINA | 21003_BA | 2,508,512 | 1,766,499 | 99.51 |
| 4 | CROATIA | 21007_HR | 2,552,070 | 1,870,713 | 99.64 |
| 5 | CROATIA | 21008_HR | 2,111,859 | 1,273,558 | 99.55 |
| 6 | CYPRUS | 21004_CY | 2,686,062 | 1,607,809 | 99.57 |
| 7 | CYPRUS | 21005_CY | 2,044,674 | 829,904 | 98.07 |
| 8 | CYPRUS | 21006_CY | 3,150,002 | 2,428,422 | 99.53 |
| 9 | CZECH REPUBLIC | 21009_CZ | 3,290,470 | 2,840,707 | 99.21 |
| 10 | CZECH REPUBLIC | 21010_CZ | 3,754,585 | 2,804,456 | 99.55 |
| 11 | CZECH REPUBLIC | 21048_CZ | 3,436,877 | 2,688,684 | 99.57 |
| 12 | DENMARK | 21052_DK | 3,273,309 | 763,257 | 99.37 |
| 13 | DENMARK | 21053_DK | 2,547,875 | 817,493 | 99.58 |
| 14 | DENMARK | 21054_DK | 2,608,811 | 809,437 | 98.17 |
| 15 | FINLAND | 21015_FIN | 2,539,762 | 1,135,620 | 99.49 |
| 16 | FINLAND | 21017_FIN | 2,831,785 | 1,852,350 | 99.60 |
| 17 | FINLAND | 21016_FIN | 1,918,714 | 776,343 | 99.55 |
| 18 | FRANCE | 21018_F | 2,731,094 | 1,892,165 | 99.48 |
| 19 | FRANCE | 21019_F | 2,485,475 | 1,568,069 | 99.52 |
| 20 | GERMANY | 21011_D | 2,421,764 | 1,150,371 | 99.51 |
| 21 | GERMANY | 21013_D | 2,204,439 | 1,523,263 | 99.51 |
| 22 | GERMANY | 21051_D | 2,862,135 | 2,144,701 | 99.54 |
| 23 | GERMANY | 21012_D | 1,788,643 | 810,402 | 99.52 |
| 24 | GERMANY | 21014_D | 2,914,973 | 1,378,621 | 99.58 |
| 25 | GERMANY | 21050_D | 3,253,048 | 1,638,447 | 99.52 |
| 26 | GERMANY | 21049_D | 2,532,868 | 1,200,315 | 99.52 |
| 27 | GREECE | 21021_GR | 2,022,878 | 818,984 | 99.45 |
| 28 | GREECE | 21022_GR | 2,786,455 | 1,788,408 | 99.48 |
| 29 | GREECE | 21020_GR | 2,468,367 | 1,110,096 | 99.32 |
| 30 | ITALY | 21023_IT | 3,496,150 | 2,969,772 | 99.52 |
| 31 | ITALY | 21024_IT | 1,725,352 | 818,506 | 99.48 |
| 32 | ITALY | 21046_IT | 3,260,879 | 2,256,762 | 99.52 |
| 33 | LATVIA | 21026_LV | 3,546,039 | 2,496,946 | 99.54 |
| 34 | LATVIA | 21025_LV | 2,063,000 | 1,247,974 | 99.39 |
| 35 | LITHUNIA | 21027_LT | 2,596,024 | 1,391,166 | 99.63 |
| 36 | LUXEMBOURG | 21028_L | 2,554,722 | 1,565,165 | 99.72 |
| 37 | POLAND | 21030_PL | 3,538,732 | 3,215,972 | 99.32 |
| 38 | POLAND | 21029_PL | 2,209,073 | 1,375,405 | 99.57 |
| 39 | ROMANIA | 21031_RO | 2,784,858 | 2,119,141 | 99.64 |
| 40 | SLOVAK REPLUBLIC | 21033_SK | 3,345,138 | 2,507,501 | 99.58 |
| 41 | SLOVAK REPUBLIC | 21032_SK | 2,589,970 | 1,469,732 | 98.66 |
| 42 | SLOVENIA | 21034_SI | 2,838,240 | 1,941,543 | 99.52 |
| 43 | SLOVENIA | 21035_SI | 2,470,338 | 1,544,585 | 99.63 |
| 44 | SPAIN | 21039_E | 2,683,669 | 1,964,724 | 99.52 |
| 45 | SPAIN | 21040_E | 2,284,844 | 1,014,470 | 99.54 |
| 46 | SPAIN | 21043_E | 1,935,880 | 695,817 | 98.89 |
| 47 | SPAIN | 21047_S | 2,280,905 | 1,133,647 | 99.45 |
| 48 | SPAIN | 21036_E | 2,252,726 | 1,019,145 | 99.52 |
| 49 | SPAIN | 21041_E | 2,484,336 | 1,410,983 | 99.21 |
| 50 | SPAIN | 21042_E | 1,897,633 | 601,931 | 99.02 |
| 51 | SPAIN | 21037_E | 2,913,114 | 1,747,719 | 99.50 |
| 52 | SPAIN | 21038_E | 2,426,135 | 1,386,068 | 99.53 |
| 53 | SWEDEN | 21044_S | 3,462,193 | 2,738,247 | 99.60 |
| 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]).

## SI References

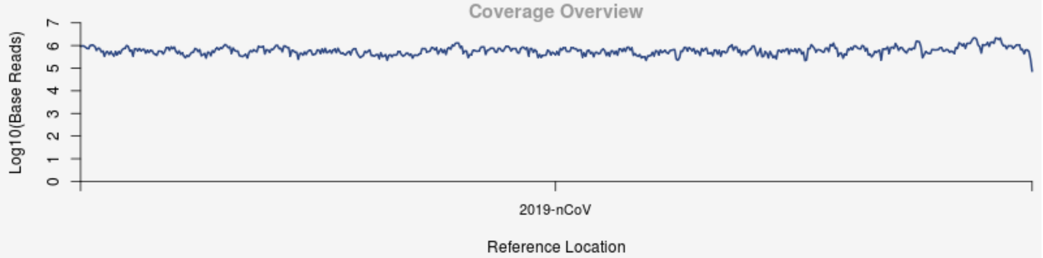
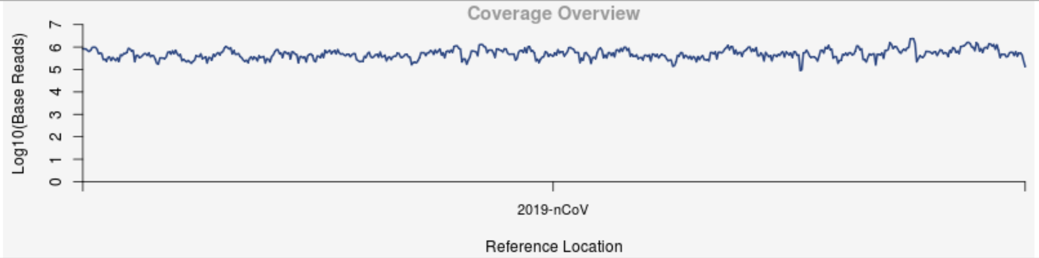
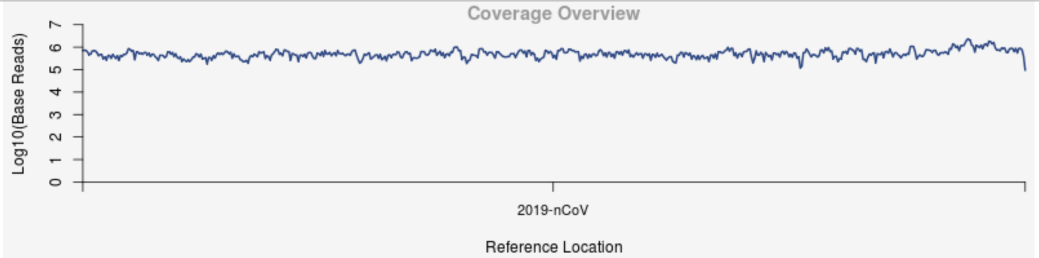
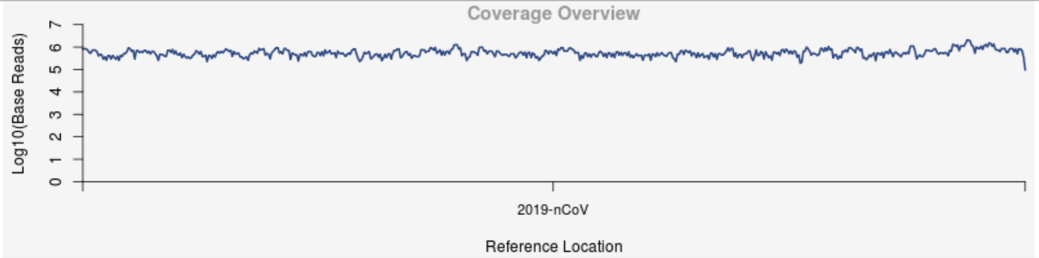
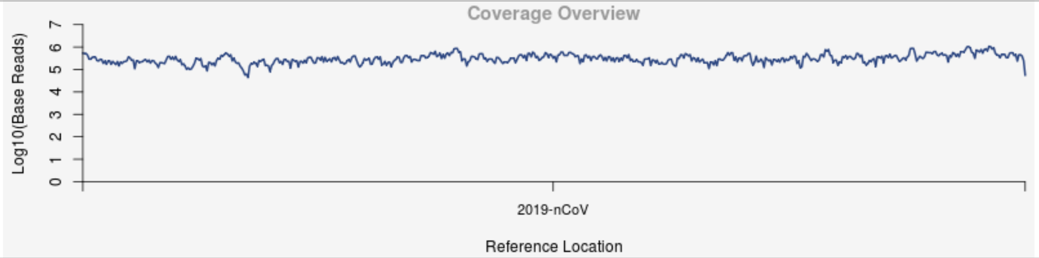
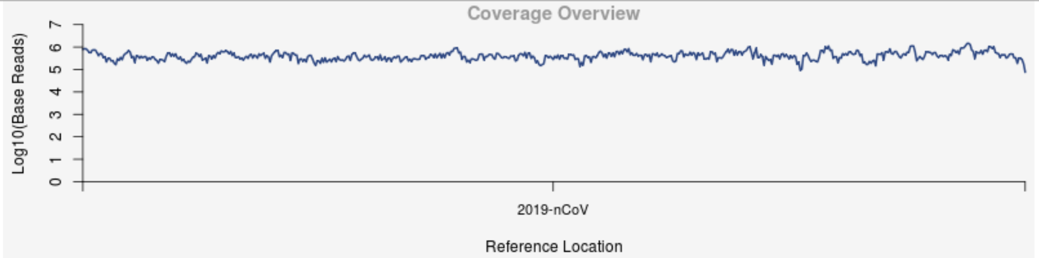
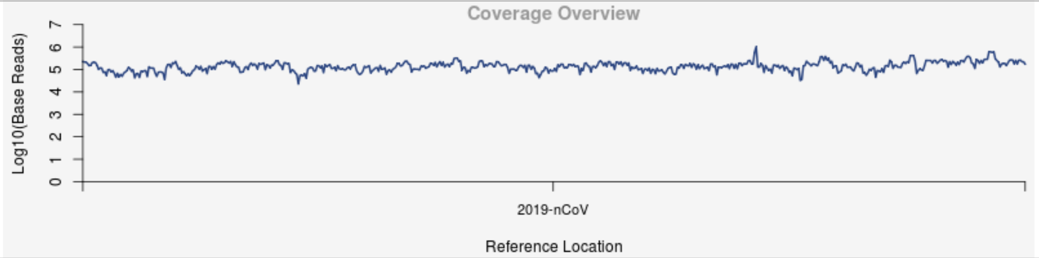
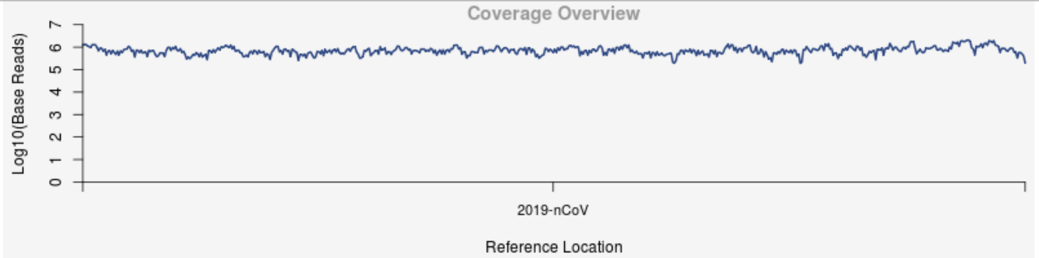
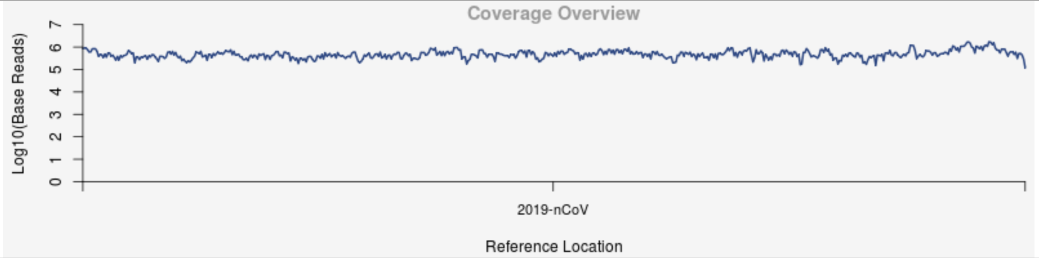
European Centre for Disease Prevention and Control, 2021a. SARS-CoV-2 variants of concern as of 3 June 2021 [WWW Document]. European Centre for Disease Prevention and Control. URL https://www.ecdc.europa.eu/en/covid-19/variants-concern (accessed 6.6.21).

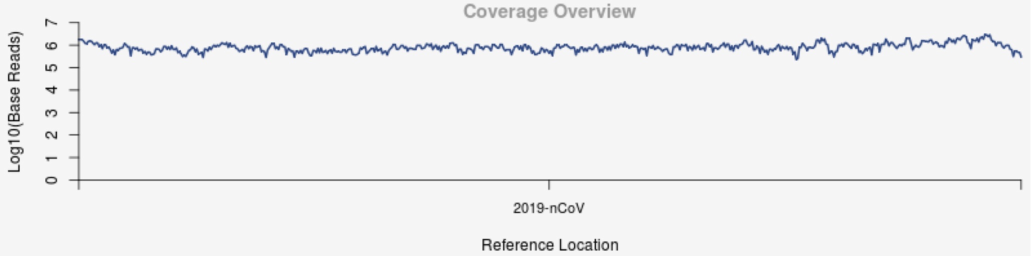
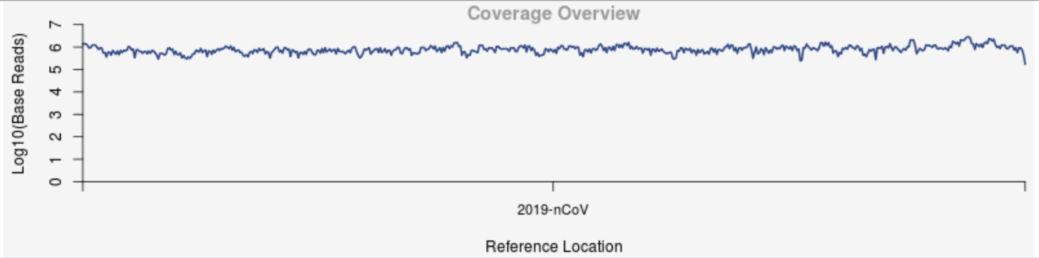
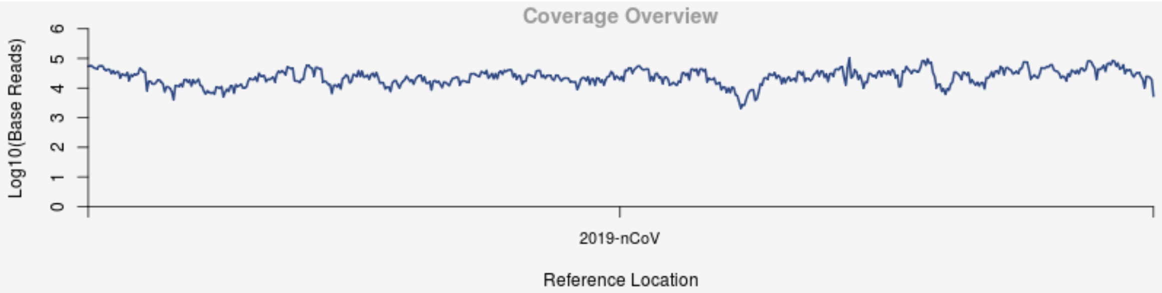
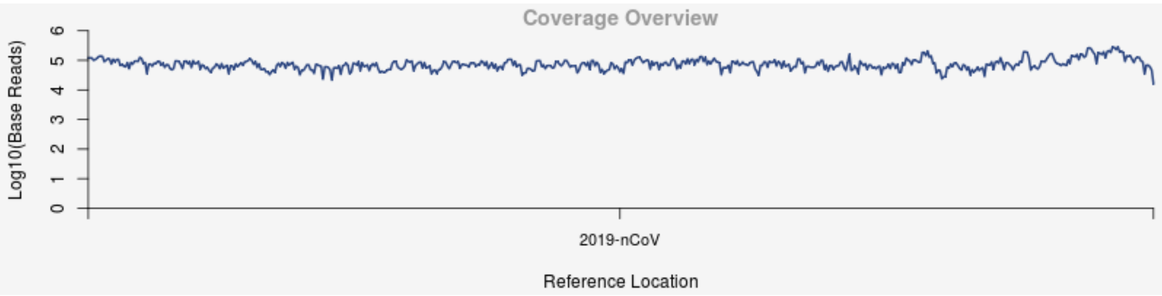
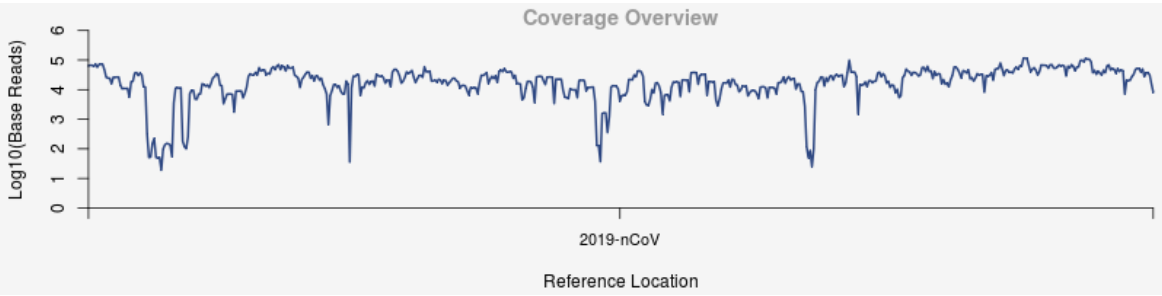
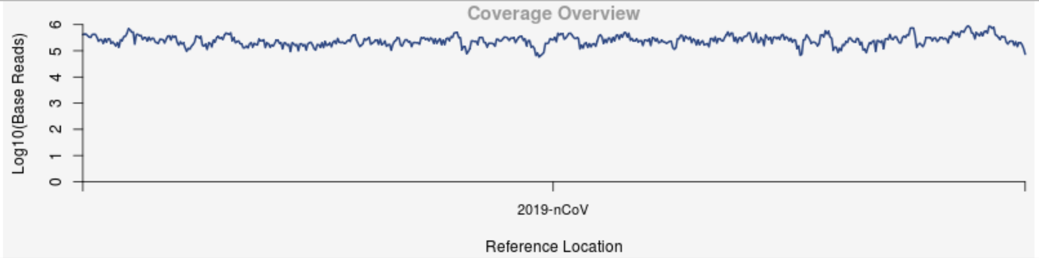
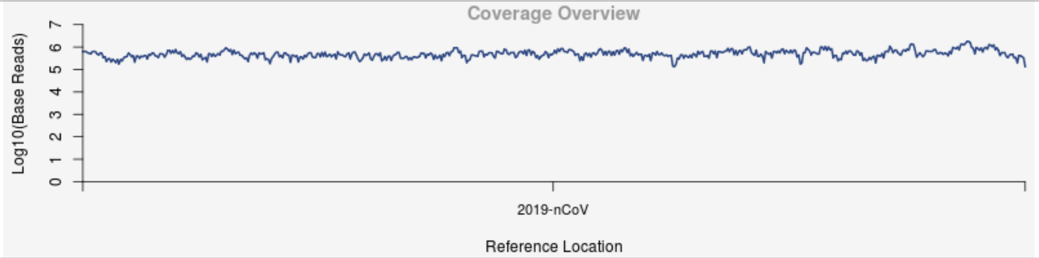
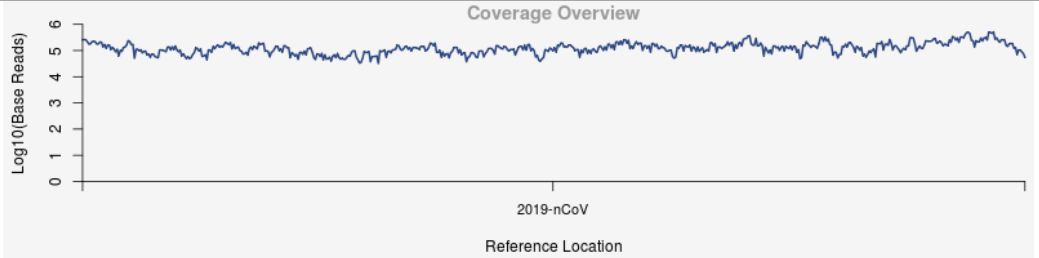
European Centre for Disease Prevention and Control, 2021b. Detection and characterisation capability and capacity for SARS-CoV-2 variants within the EU/EEA [WWW Document]. European Centre for Disease Prevention and Control. URL https://www.ecdc.europa.eu/en/publications-data/detection-and-characterisation-capability-and-capacity-sars-cov-2-variants (accessed 4.5.21).

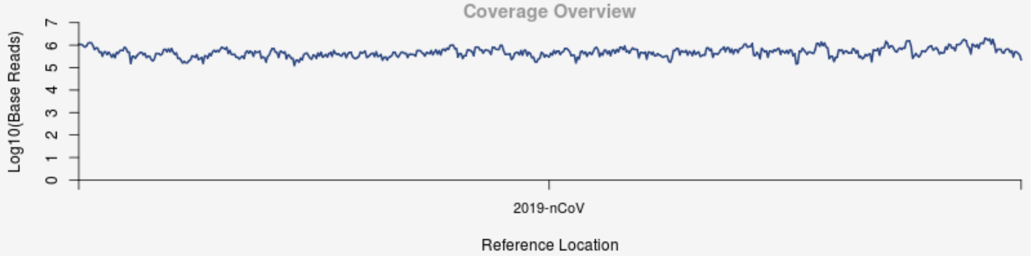
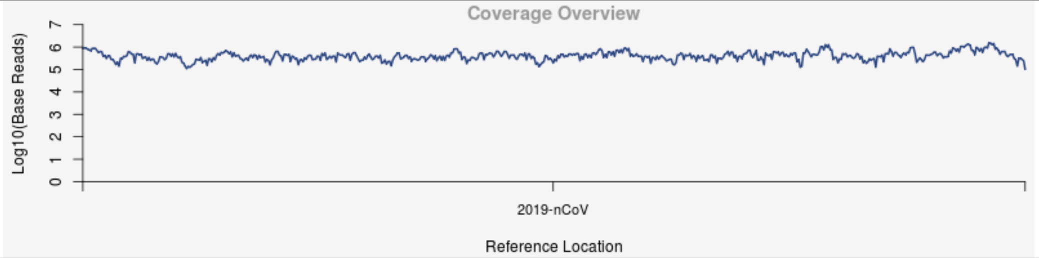
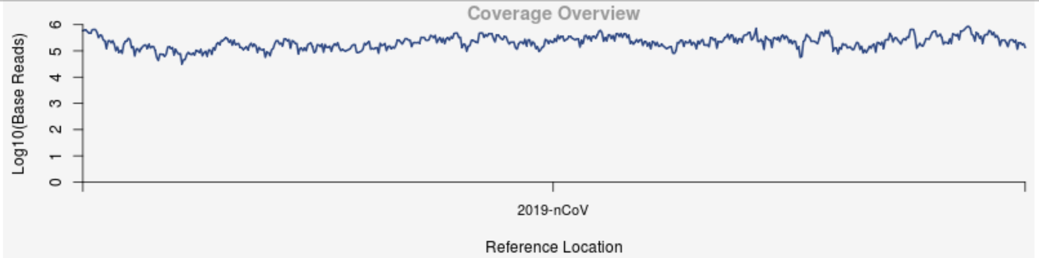
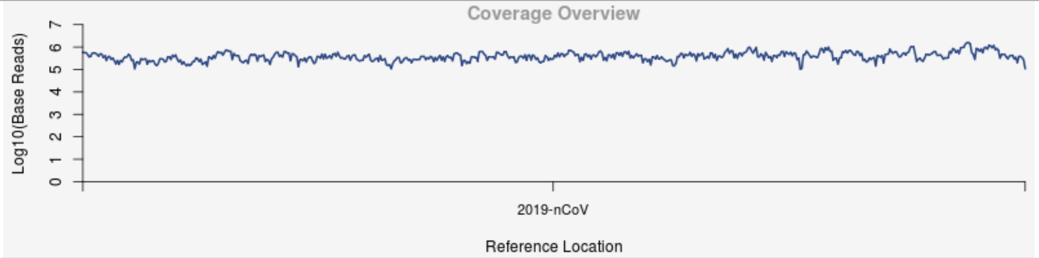
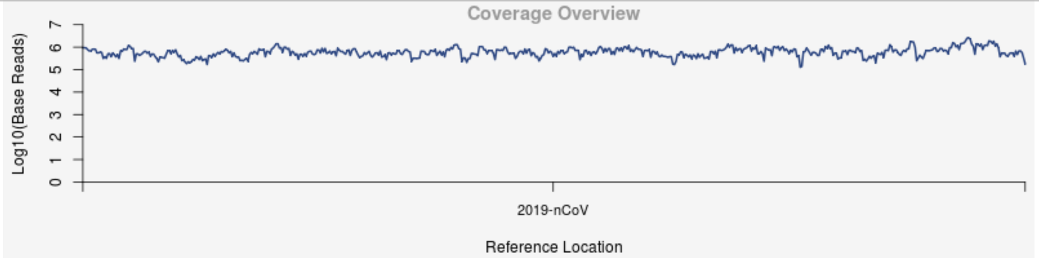
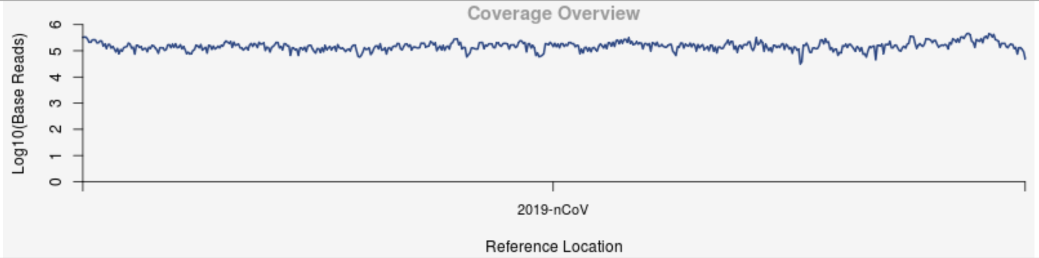
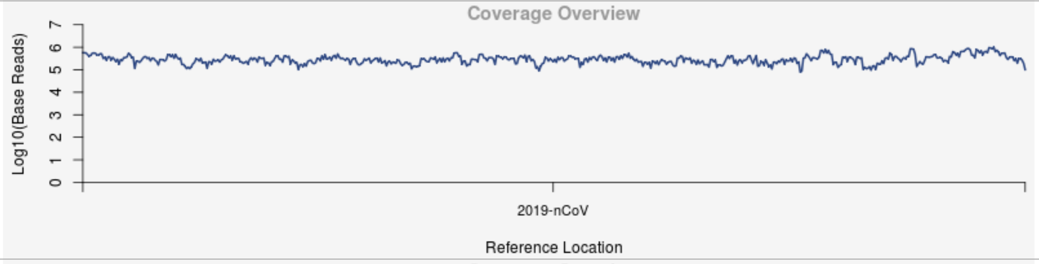
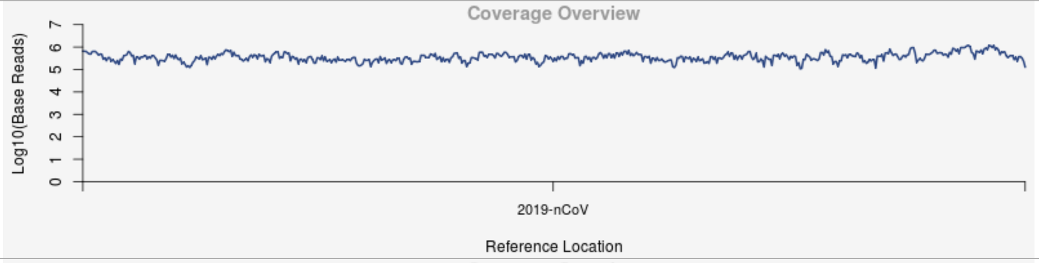
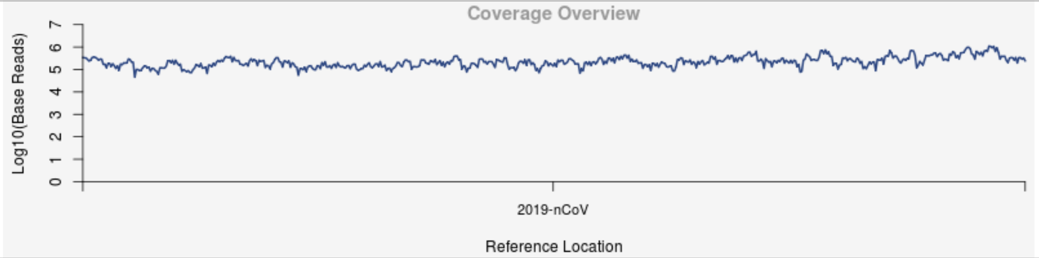
European Centre for Disease Prevention and Control, 2021c. Risk realted to the spread of new SARS-CoV-2 variants of concern in the EU/EEA - first update. Stockholm.

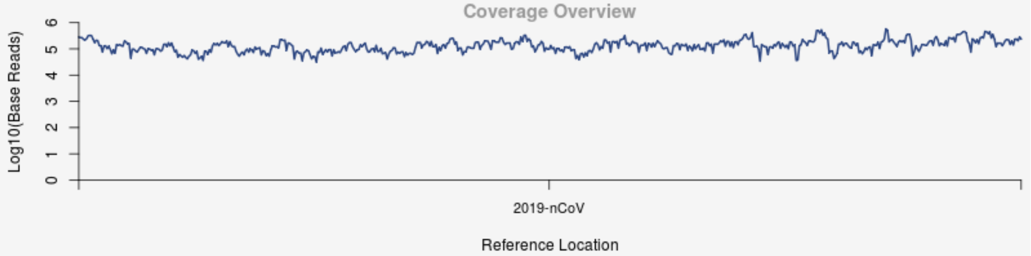
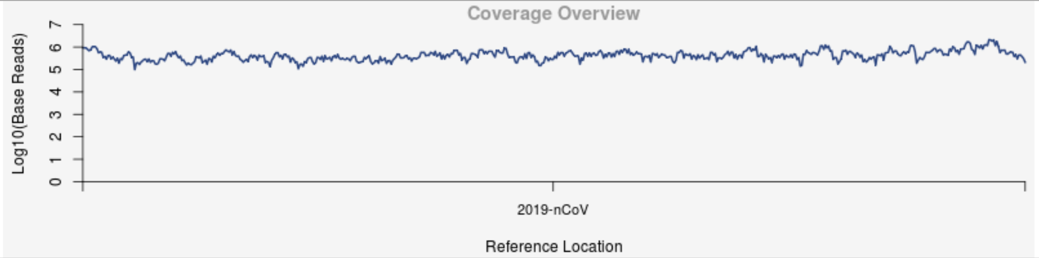
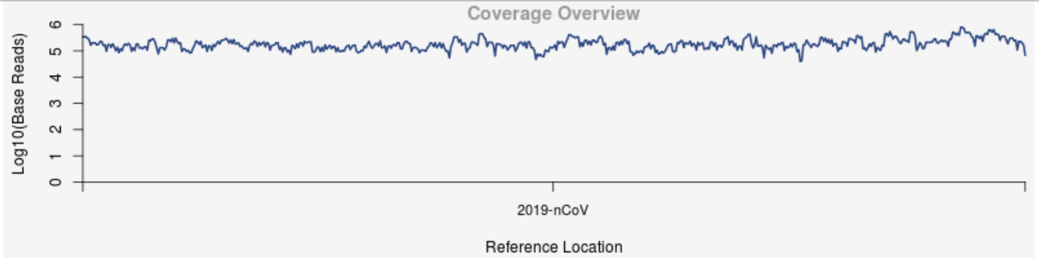
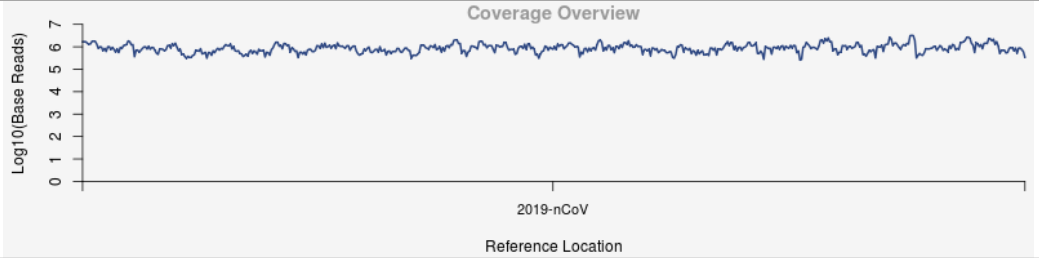
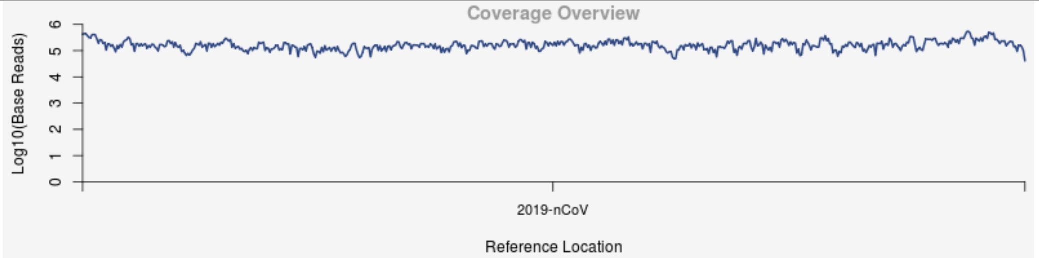
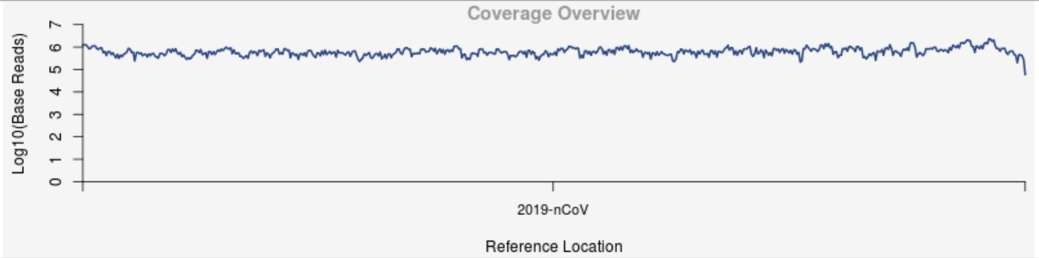
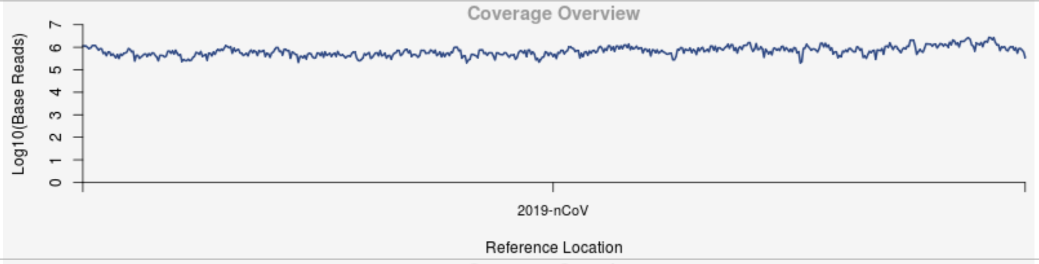
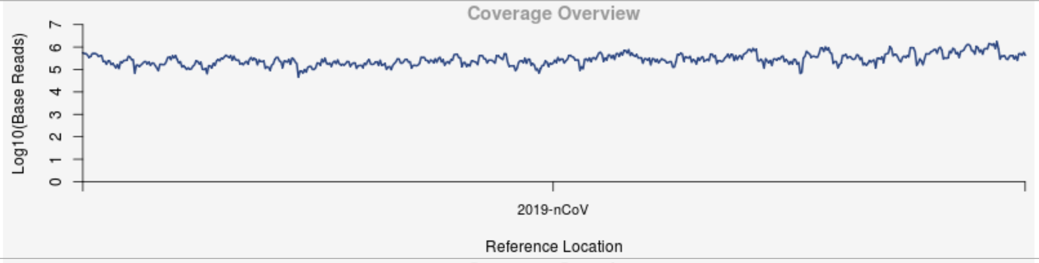
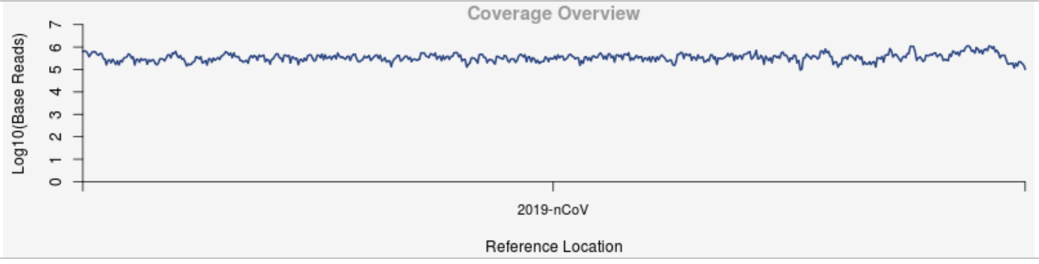
## Data S1. Coverage overview

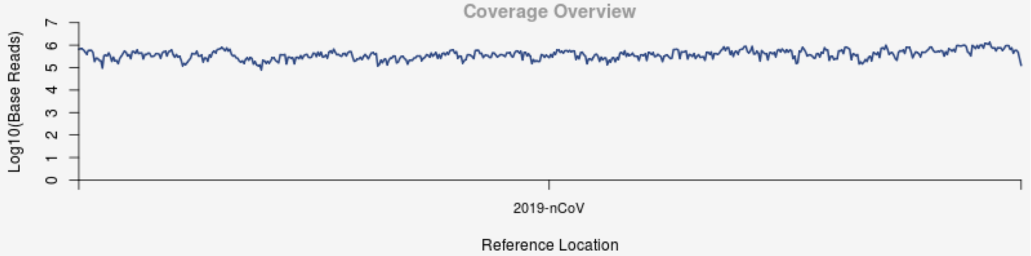
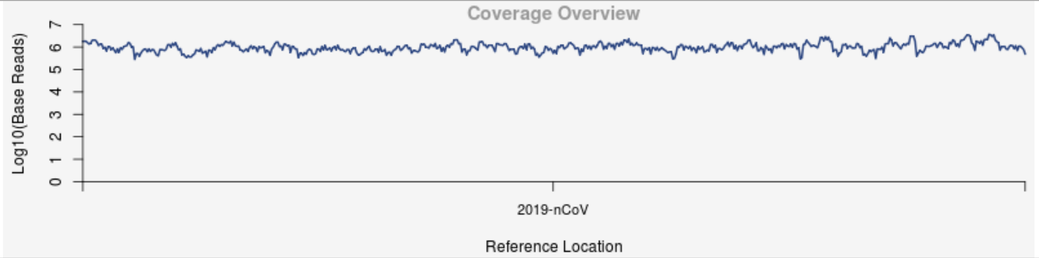
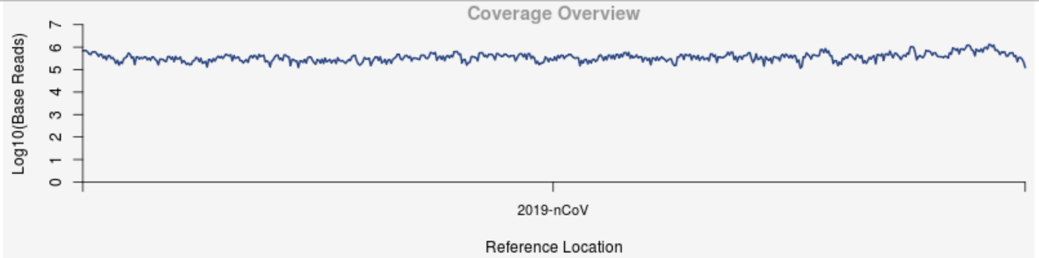
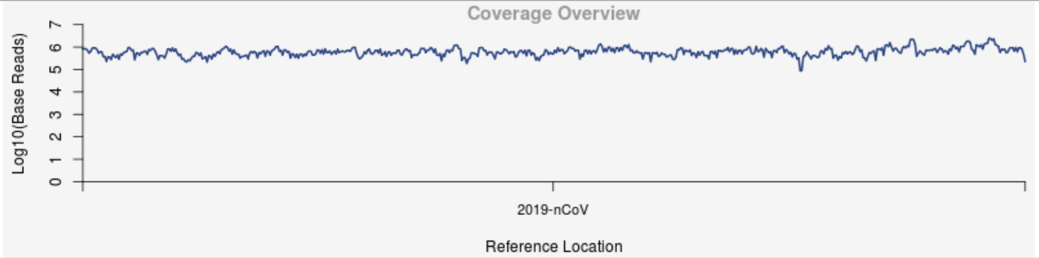
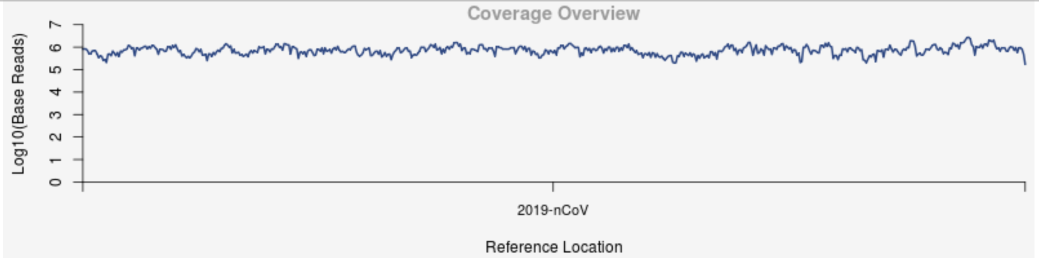
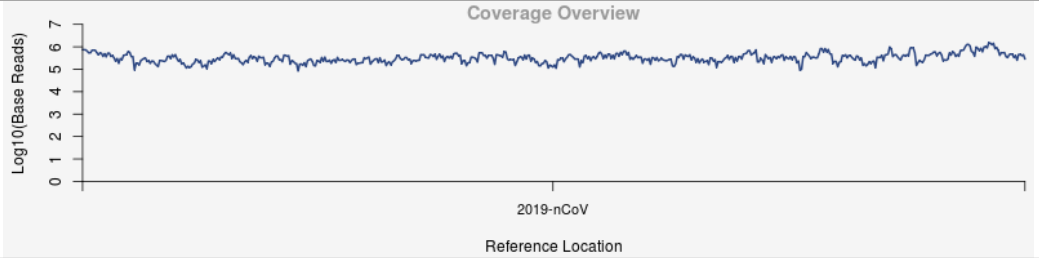
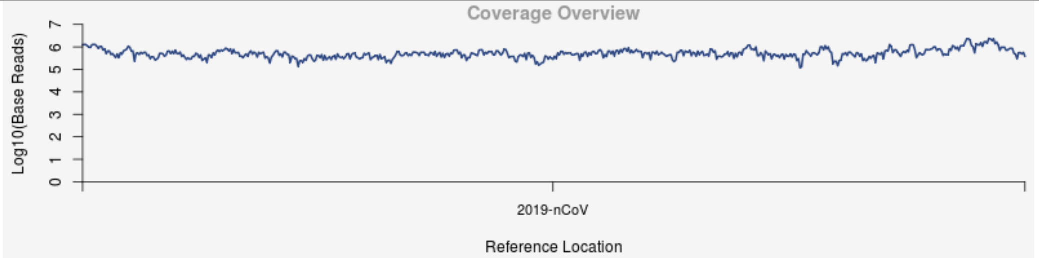
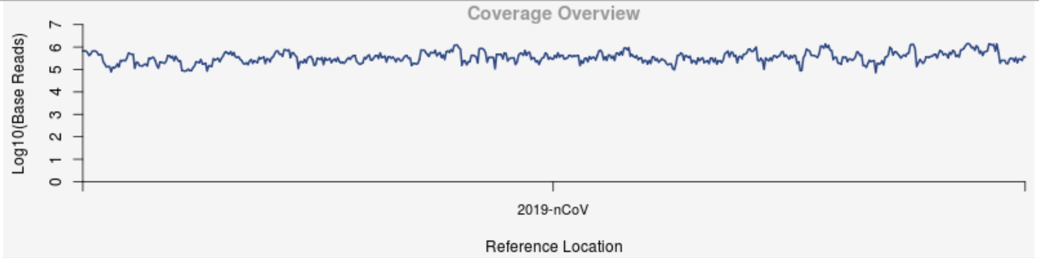
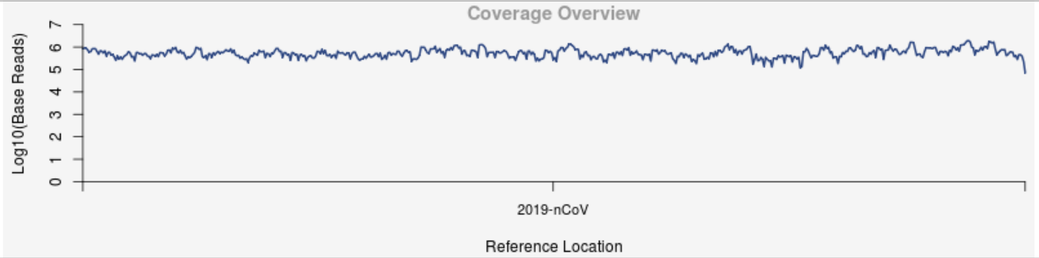
Complete list of all coverage overview graphs for all samples that were analyzed in this study.

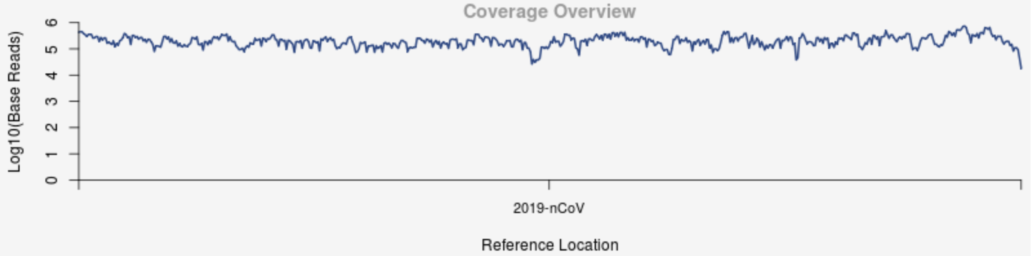
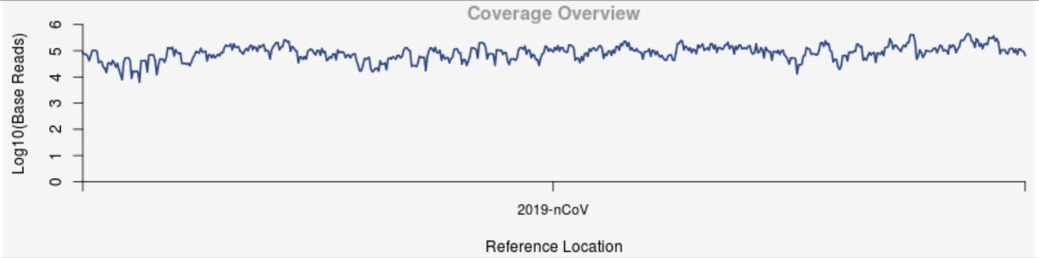
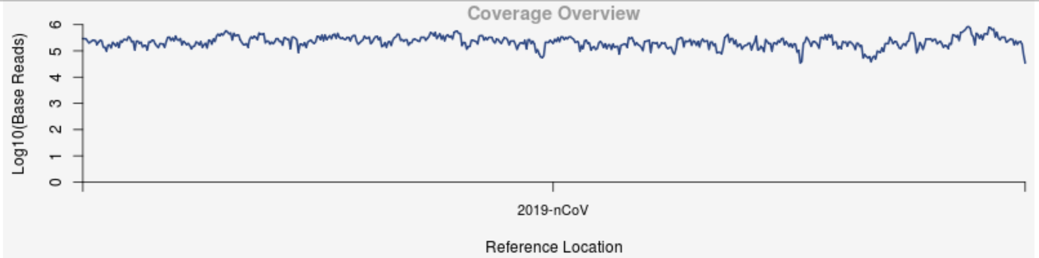
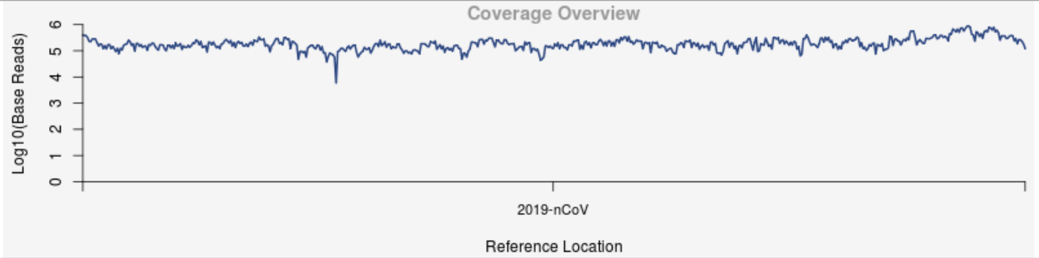
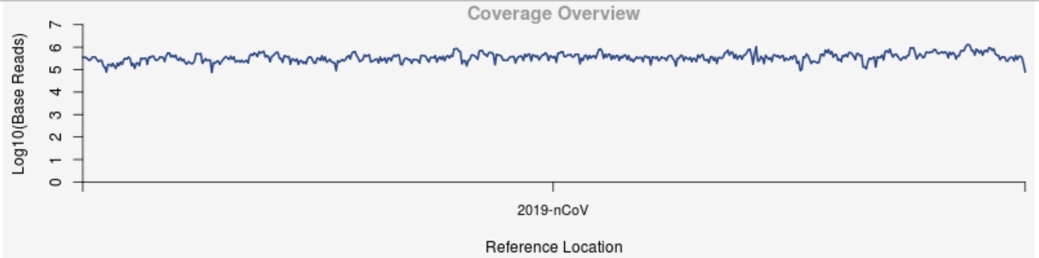
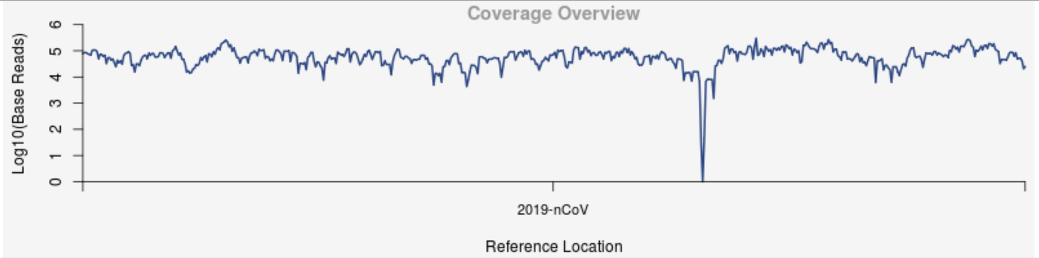
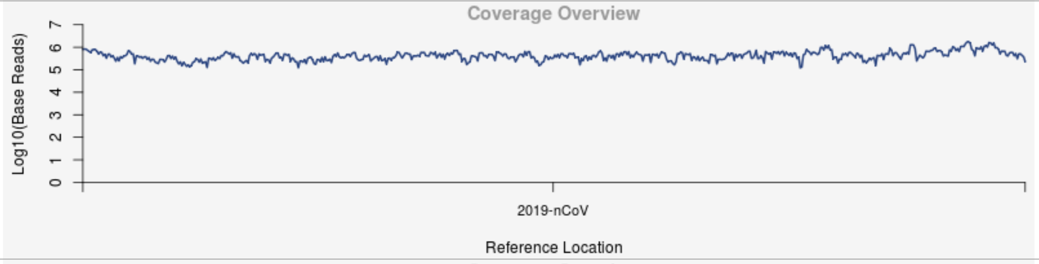
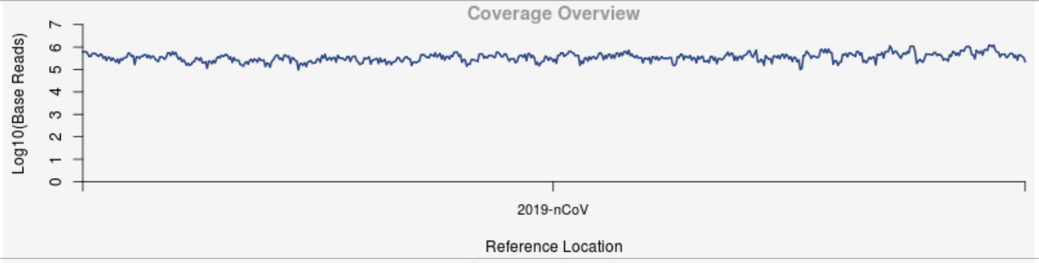
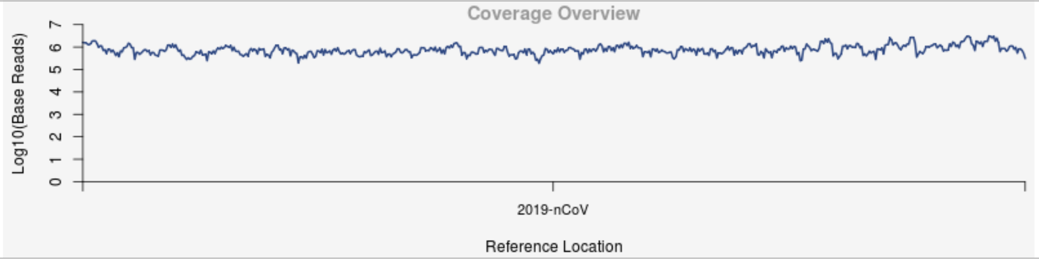
| S.No. | Country | SampleID | Coverage overview |
|---|---|---|---|
| 1 | BELGIUM | 21001_B |  |
| 2 | BELGIUM | 21002_B |  |
| 3 | BOSNIA ERZEGOVINA | 21003_BA |  |
| 4 | CROATIA | 21007_HR |  |
| 5 | CROATIA | 21008_HR |  |
| 6 | CYPRUS | 21004_CY |  |
| 7 | CYPRUS | 21005_CY |  |
| 8 | CYPRUS | 21006_CY |  |
| 9 | CZECH REPUBLIC | 21009_CZ |  |

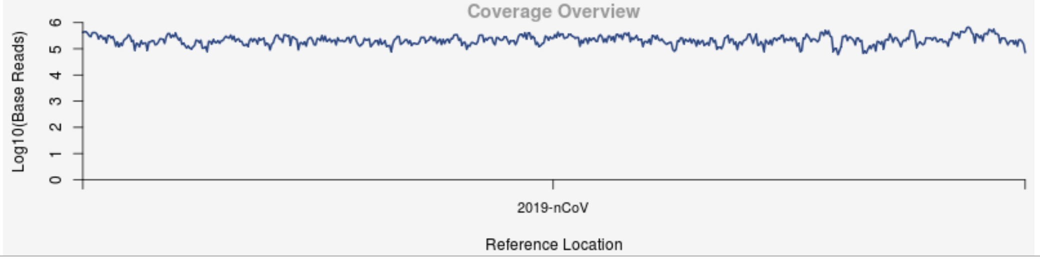
<table>
<tr><th>S.No.</th><th>Country</th><th>SampleID</th><th>Coverage overview</th></tr>
<tr><td>10</td><td>CZECH REPUBLIC</td><td>21010_CZ</td><td>
</td></tr>
<tr><td>11</td><td>CZECH REPUBLIC</td><td>21048_CZ</td><td>
</td></tr>
<tr><td>12</td><td>DENMARK</td><td>21052_DK</td><td>
</td></tr>
<tr><td>13</td><td>DENMARK</td><td>21053_DK</td><td>
</td></tr>
<tr><td>14</td><td>DENMARK</td><td>21054_DK</td><td>
</td></tr>
<tr><td>15</td><td>FINLAND</td><td>21015_FIN</td><td>
</td></tr>
<tr><td>16</td><td>FINLAND</td><td>21017_FIN</td><td>
</td></tr>
<tr><td>17</td><td>FINLAND</td><td>21016_FIN</td><td>
</td></tr>
</table>

| S.No. | Country | SampleID | Coverage overview |
|---|---|---|---|
| 18 | FRANCE | 21018_F |  |
| 19 | FRANCE | 21019_F |  |
| 20 | GERMANY | 21011_D |  |
| 21 | GERMANY | 21013_D |  |
| 22 | GERMANY | 21051_D |  |
| 23 | GERMANY | 21012_D |  |
| 24 | GERMANY | 21014_D |  |
| 25 | GERMANY | 21050_D |  |
| 26 | GERMANY | 21049_D |  |

| S.No. | Country | SampleID | Coverage overview |
|---|---|---|---|
| 27 | GREECE | 21021_GR |  |
| 28 | GREECE | 21022_GR |  |
| 29 | GREECE | 21020_GR |  |
| 30 | ITALY | 21023_IT |  |
| 31 | ITALY | 21024_IT |  |
| 32 | ITALY | 21046_IT |  |
| 33 | LATVIA | 21026_LV |  |
| 34 | LATVIA | 21025_LV |  |
| 35 | LITHUNIA | 21027_LT |  |

| S.No. | Country | SampleID | Coverage overview |
|---|---|---|---|
| 36 | LUXEMBOURG | 21028_L |  |
| 37 | POLAND | 21030_PL |  |
| 38 | POLAND | 21029_PL |  |
| 39 | ROMANIA | 21031_RO |  |
| 40 | SLOVAK REPLUBLIC | 21033_SK |  |
| 41 | SLOVAK REPUBLIC | 21032_SK |  |
| 42 | SLOVENIA | 21034_SI |  |
| 43 | SLOVENIA | 21035_SI |  |
| 44 | SPAIN | 21039_E |  |

| S.No. | Country | SampleID | Coverage overview |
|---|---|---|---|
| 45 | SPAIN | 21040_E |  |
| 46 | SPAIN | 21043_E |  |
| 47 | SPAIN | 21047_S |  |
| 48 | SPAIN | 21036_E |  |
| 49 | SPAIN | 21041_E |  |
| 50 | SPAIN | 21042_E |  |
| 51 | SPAIN | 21037_E |  |
| 52 | SPAIN | 21038_E |  |
| 53 | SWEDEN | 21044_S |  |

| S.No. | Country | SampleID | Coverage overview |
|---|---|---|---|
| 54 | SWEDEN | 21045_S |  |