Wastewater Monitoring of SARS-CoV-2 Variants in England: Demonstration Case Study for Bristol (Dec 2020 - March 2021)
Summary for SAGE 08/04/21

Mathew R. Brown[1,2], Matthew J. Wade[1,2], Shannon McIntyre-Nolan[1,3], Irene Bassano[1,5], Hubert Denise[1], David Bass[4], John Bentley[1], Joshua T. Bunce[1,2,3], Jasmine Grimsley[1], Alwyn Hart[6], Till Hoffmann[1,7], Aaron Jeffries[8], Steve Paterson[9], Mark Pollock[1], Jonathan Porter[6], David Smith[4] Ronny van Aerle[4], Glenn Watts[6], Andrew Engeli[1], Gideon Henderson[3]

[1] Joint Biosecurity Centre, Department of Health and Social Care, London WC1B 4DA, UK
[2] School of Engineering, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK
[3] Department for Environment, Food and Rural Affairs, London SW1P 4DF, UK
[4] International Centre of Excellence for Aquatic Animal Health, Cefas, Barrack Road, Weymouth, DT 8UB, UK
[5] Department of Infectious Disease, Imperial College London, London SW7 2AZ, UK
[6] Environment Agency, Research, Horizon House, Deanery Road, Bristol BS1 5AH, UK
[7] Department of Mathematics, Imperial College London, London SW7 2AZ, UK
[8] Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK
[9] Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool CH64 7TE UK

Contents

SAGE Discussion

This paper demonstrates the ability to assess presence of SARS-CoV-2 variants of concern at population level using wastewater measurements. How should recently expanded laboratory capability for such variant monitoring be best deployed in the short to medium term to help detect and contain SARS-CoV-2?

Abstract

This paper presents a case study of an operational pilot demonstrating the use of wastewater genomic sequencing to rapidly assess community transmission of SARS-CoV-2 variants in Bristol. Tracking the presence of novel variants in a population is key to controlling the spread of SARS-CoV-2 and ensuring appropriate and resilient public health response. Confidence in this information is predicated on capturing a true representation of SARS-CoV-2 prevalence in a population through clinical testing. However, this is challenging. Some communities are opposed to testing, while a large analytical effort and cost is associated with mass testing of individuals. Wastewater sampling and PCR analysis of virus RNA has been demonstrated as a relatively inexpensive method to monitor SARS-CoV-2 in an unbiased manner, capturing the dynamics of the virus at discrete population level to augment knowledge of the state of the disease from other sources. In this report, an assessment of the utilisation of wastewater sampling for detecting SARS-CoV-2 variants and mutations of concern is presented, focusing on 33 wastewater samples collected from 11 sub-catchments in Bristol (average catchment population: 26,869). A multi-mutational approach is pursued, aimed at identifying 118 discrete mutations (e.g., E484K, N501Y), which are signature mutations for known Variants of Concern (VOC) and Variants Under Investigation (VUI). The B.1.1.7 lineage (VOC-20DEC-01) was identified in all 11 sub-catchments, with signature mutations increasing in tandem with clinically confirmed cases over the same time period**.** The E484K mutation was observed in 8 of the 11 sub-catchments, likely indicative of the presence of the B.1.1.7 lineage with E484K (VOC-20REF-02; the 'Bristol Variant'). This information was relayed to local and national response teams, complementing existing clinical data. This multi-mutational genomic approach has been expanded and is sequencing 200 - 400 wastewater samples per week to assist in local, regional and national VOC/VUI response. Samples are prioritised based on known areas of concern and/or high wastewater SARS-CoV-2 RNA concentrations in each locality.

Background

Significant national and international attention is now focused on the detection and containment of VUIs and VOCs due to the threat they pose to public health. In September 2020, the emergence of variant B.1.1.7 (VOC-20DEC-01) - and its subsequent rapid spread throughout the UK - demonstrated the need to have robust national response systems to protect the effectiveness of non-pharmaceutical interventions (NPIs). In future, the release of NPIs will partly depend on an adequate and cost-effective national detection and containment capability within the UK. This will be required to combat the continued risk of emergence of mutations, especially as travel restrictions ease and the effectiveness of vaccines against mutations remains unknown.

Wastewater monitoring

Wastewater-based epidemiology (WBE) is one such capability. This nascent scientific field - complementary to clinical epidemiology - can be used to explore SARS-CoV-2 genetic diversity within a population. This includes the detection and spread of new and emerging variants or mutations of concern (Prado *et al.*, 2021). By screening populations at a community level, WBE has the power to provide insight rapidly and efficiently for the presence of VOCs/VUIs, and then guide the deployment of cost and resource intensive response systems - such as mass testing.

Since July 2020, wastewater (WW) monitoring for SARS-CoV-2 has been operational in England - under the Environmental Monitoring for Health Protection programme (EMHP). Routine collection and analysis of WW samples for total SARS-CoV-2 RNA concentrations - taken from sewage treatment works (STWs), in-network sites and near-to-source infrastructure - have provided key insights to aid both national and local responses, and helped to guide the deployment of valuable clinical resources. Led by the Joint Biosecurity Centre (JBC), with Defra Group as the core strategic infrastructure partner, the EMHP programme has been rapidly scaled since inception. It now monitors virus RNA in WW four times a week in catchments, representing two-thirds of the English population.

The case study is a real-life example of the WW monitoring capabilities for SARS-CoV-2 variant detection, as well as its current limitations (which are being addressed through intensive development of methods and standards). Following this demonstration, EMHP continues to work with local and national stakeholders on developing a clear response pathway for when VOCs, VUIs and mutations of concern are detected in WW during routine surveillance, and operationalising this process on a national scale.

Operational Case Study: Bristol

Summary

- Overall, this case study demonstrated WBE is an effective tool for detecting VOCs, VUIs and mutations of interest within a population (Bristol and South Gloucestershire.).
- The tool provides timely, non-invasive, and unbiased community-level insights at lower budget and resource expense compared with mass clinical testing.
- When this approach is used continuously - across time and space - it has the potential to identify outbreaks and clusters of known VOCs and VUIs and elucidate their transmission and spread across England.
- It can also aid in targeting resourcing intensive clinical testing; and assessing the success of containment and the continuing effectiveness of NPIs.

Methods

Sample Collection, Concentration and RNA Extraction

WW grab samples (1L per sample) were collected from 11 locations across the Bristol sewer network on the 5th, 7th, 8th and 11th of January and the 5th, 6th and 7th February 2021, as part of the ongoing EMHP programme in England. Samples were transported to the Environment Agency lab and subsequently stored at 4 - 6°C until analysis, minimising RNA degradation. 200mL sub-samples were then purified via centrifugation (10,000 x *g* for 30 minutes at 4°C). 150mL of supernatants were retained, pH adjusted (7.0 - 7.6 using 1M NaOH) and then concentrated to 2mL via polyethylene glycol

precipitation (PEG, 40% PEG 8000, 8% NaCl) overnight at 4°C and further centrifugation (10,000 x *g* for 30 minutes at 4°C). RNA was then extracted using the NucliSENS® MiniMag® Nucleic Acid Purification System (BioMérieux SA, Marcy-l'Étoile, France) according to manufacturer instructions, generating RNA extracts of 50 - 100 µL in volume. Extracts were stored at -80°C until further processing.

SARS-CoV-2 RNA Amplicon Sequencing

WW RNA extracts, with negative controls, were treated with DNase to reduce non-specific amplification and then used for cDNA generation (NEB Luna Script). cDNA was then processed using the ARTIC v3 pipeline, analogous to clinical SARS-CoV-2 samples (https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye), generating approximately 400 bp amplicons tiling the entire SARS-CoV-2 genome, which were subsequently used for library construction. For the January samples amplicons were diluted (1/2), pooled and size selected to 450-850bp using a Pippin Prep, libraries were then sequenced on an Illumina MiSeq platform generating 2x250bp paired end reads. For the February samples three sets of 11 barcoded samples, plus the negative control, were pooled, and run on Oxford Nanopore (ONT) GridionX5 flowcells. Following demultiplexing, base calling was performed using Guppy (V 3.2.10+aabd4ec) in high-accuracy mode.

Bioinformatics Analysis

Raw reads were processed using the *ncov2019-artic-nf v3* pipeline (Tyson *et al.*, 2020) using default parameters. Briefly, reads were aligned to the SARS-CoV-2 reference genome (Wu *et al.*, 2020), using minimap v2.17 (Li, 2018). Single Nucleotide Polymorphisms (SNPs) and insertions/deletions (Indels) were then identified from BAM files using samtools (v0.1.18-r580, Li *et al.*, 2009) and VarScan (v2.3, with default settings, Koboldt *et al.*, 2012) on 100,000 sequencing reads with an alignment score > 10. Identified SNPs and Indels were then filtered against signature mutations of known VOCs and VUIs, as defined by Public Health England (PHE) at the time of writing (https://github.com/phe-genomics/variant_definitions). These were then visualised for each VOC/VUI using RStudio (v. 1.2.1335, R Core Team, 2017) and R version 4.0.2 (R Core Team, 2017). Custom scripts were developed to extract sequencing quality indicators, such as genome coverage, number of mapped reads and average read length.

To increase confidence in the detection of low frequency mutations, due to the innate error rate of ONT sequencing, mutations were considered to be true mutations when they occurred on both strands of the cDNA (a mate approach). In some instances, two mutations of interest were present and quantified on the same sequencing reads, facilitating localised co-occurrence analysis.

A genome-wide co-occurrence approach was also adopted to aid VOCs and VUIs identification in WW samples. Briefly, co-occurring events were called from BAM files using CoOccurrence adJusted Analysis and Calling (COJAC) (Jahn *et al.*, 2021), facilitating the identification of signature mutations (GitHub - phe-genomics/variant_definitions) co-occurring on the same sequencing read, that is, a read or paired read coming from the same amplicon, thus one SARS-CoV-2 virion. Co-occurrence analysis was undertaken for signature mutations of the B.1.1.7 (VOC-20DEC-01), B.1.351 (VOC-20DEC-02), P.1 (VOC-21JAN-02), B.1.1.7 with E484K (VUI-21JAN-02), A.23.1 (VUI-21FEB-01), B.1.318 (VUI-21FEB-04) lineages. Since a proportion of signature mutations are shared amongst these VOCs/VUIs, amplicons with co-occurring mutations can be listed for multiple variants.

Interpretation and limitations

Analysing WW for SARS-CoV-2 is challenging. The low concentration of SARS-CoV-2 in WW, the genomes of which may be highly degraded and fragmented, requires enrichment through concentration procedures. While a necessity, this can also enrich for PCR inhibitors and contaminating

bacterial, viral, and human nucleic acids (Peccia *et al.*, 2020). Poor and inconsistent amplification of target amplicons (400 - 450 bp here, *q.v.*) can consequently arise, resulting in patchy coverage of SARS-CoV-2 genomes. Even if amplified and successfully sequenced, WW harbours a mixed population of SARS-CoV-2 variants. This makes data interpretation difficult as the genome linkage between SNPs/Indels used to assign phylogeny and subsequently lineage is lost. Nevertheless, lineage has been assigned to the predominant SARS-CoV-2 genotype in WW samples with exceptional genome coverage (Crits-Christoph *et al.*, 2021).

One solution to garner valuable population-level information on circulating SARS-CoV-2 variants from WW is a multi-mutational approach, focusing on signature SNPs and Indels of VOCs/VUIs (Jahn *et al.*, 2021). SNPs or Indels identified in WW can be filtered against these signature mutations, where their presence, particularly multiple, in combination, is indicative of a specific SARS-CoV-2 variant in a population. Greater confidence in presence can be inferred by co-occurrence analysis, the identification of two or more of these signature SNPs or Indels on the same read, i.e. from the same virion. However, sequencing errors, amplification biases and contamination can still complicate this multi-mutational approach (Jahn *et al.*, 2021), contributing to false positives. This can be controlled by using suitable read-depth thresholds and sequencing negative controls or evaluated through detection of mutations on reads in both orientations.

Nevertheless, false negatives are much more likely to occur, where a variant or mutation present at low frequency will often not be observed because there are too few virions present in the initial sample. Poor amplification of target amplicons and patchy genome coverage, as aforementioned, could also contribute to false negatives given that the genomic information needed for identification is lacking. The use of multiple mutations to determine a variants presence may mitigate against this, where signature SNPs/Indels will mostly be sequenced independently on different amplicons and, thus, are collectively less likely to remain undetected. Similarly, the temporal and spatial tracking of SARS-CoV-2 genomic information in WW increases our ability to make reliable calls on the presence or absence of a known VOCs or VUIs. The generation of consistently high-quality SARS-CoV-2 genomic data would also aid in reducing false negative signals.

Limitations in sampling strategy may also contribute to false negatives in areas where clinically confirmed cases of a VOCs/VUIs exist. As discussed in the programme's previous update to SAGE (Wade *et al.* 2020) two methods, grab and composite (using autosampler devices), are routinely employed for collecting WW samples. Autosamplers are programmed to collect WW at preselected intervals (e.g. hourly) over a set period, while grab samples are taken at a single point in time. Grab samples are therefore more influenced by fluctuations in WW composition and may be less representative of the population.

Results and Discussion

Detection of B.1.1.7 (VOC-20DEC-01)

We observed all signature SNPs (alleles) of the B.1.1.7 lineage (VOC-20DEC-01) in WW samples collected across Bristol (Fig. 1). In general, across all catchments, we observed an increase in the number of B.1.1.7 signature SNPs and the frequency at which they were observed through time (Fig. 1). Co-occurrence analysis gives us even greater confidence of the presence of B.1.1.7, given that we observed co-occurrence of P681H and T716I, Q27*, R52I and Y73C and R52I, Y73C and D3L across most sites (Table 1). In catchments where signature SNPs were not detected, e.g. CLF (Fig. 1), poor genome coverage was observed (coverage < 20X), impeding our ability to detect signature SNPs.

Detection of the E484K mutation

The E484K mutation was observed at low frequency (1.2 - 2.1 %) in eight of 11 catchments across Bristol and temporally within two catchments (Fig. 2 and Table 2). Low-frequency detection of E484K was confirmed in six of 10 samples (mate reads, Table 2). E484k is present on several known VOCs and VUIs (Table 1), yet evidence of other signature SNPs defining these variants, except for the B.1.1.7 lineage with E484K (VOC-21FEB-02) was missing. One other signature SNP of VOC-21FEB-02 was observed temporally at moderate to high frequency in one catchment (HOR, Fig. 2). Co-occurrence analysis also identified several reads from each site containing both E484K and N501Y (Table 2), providing further evidence of the possible presence of VOC-21FEB-02. Whilst this combination of mutations is also found on both the B.1.351 and P.1 lineages, co-occurrence analysis in other amplicons found no evidence of their presence, corroborating a lack of individual signature SNPs defining these lineages in WW samples.

Comparison of WW E484K detection with clinical cases of B.1.1.7 with E484K (VOC-21FEB-02)

In the eight Bristol WW catchments where E484K was detected and where evidence of VOC-21FEB-02 exists, seven had clinically confirmed or probable, possible, and suspected cases of VOC-21FEB-02 (Fig.3). One catchment where E484K was detected in WW, BRE, had no confirmed or probable, possible, or suspected cases of VOC-21FEB-02. E484K was not detected in the CLF catchment despite confirmed or suspected cases of VOC-21FEB-02 (Fig. 3). Inadequate genome coverage was observed for samples collected from the CLF catchment, impeding our ability to detect E484K and other signature SNPs (as described for B.1.1.7).

The detection of multiple B.1.1.7 signature SNPs across time and space in Bristol, some of which co-occurred on the same sequencing read, gives us confidence in the use of a multi-mutational approach to track the spread of SARS-CoV-2 variants in WW. Thus, while E484K was detected at very low frequency, potentially close to the error rate of the sequencing technology (ONT), it was observed through time and at geographically distinct locations. In most cases, it could be confirmed via the presence of these SNPs on both sense and anti-sense reads, providing greater confidence in the signal. A contributing factor was that E484K had only previously been detected in 8 of 286 samples sequenced on ONT runs (~2.8%), while here E484K was detected in 10 of 33 samples (~30%).

Monitoring response to detection

This detection of E484K and its potential affiliation with emerging variants prompted expanded WW surveillance across two additional catchments of interest (AVO and RED, Fig. 3), identified based on their suspected association with movement patterns of known cases and proximity to local 'hotspots'.

WW samples collected on 18th – 20th February 2021 and 26th – 28th February 2021 from these and 11 other operational sites across Bristol (Fig. 3) were subsequently sequenced to provide assurances the variant had not spread beyond the geography covered by surge testing implemented in response to known clinical cases. An additional eight catchments (Fig. 3, sampling commenced 8th March 2021) have since been incorporated into routine EMHP WW monitoring in Bristol to facilitate surveillance of SARS-CoV-2 presence in the community and potential VOC/VUI outbreaks, providing insight beyond clinical data on local variant spread.

Given the added value in the area and upon the emergence of clinically confirmed P.1 (VOC-21JAN-02) cases, EMHP were asked to contribute to the local, regional and national responses to VOCs/VUIs. In Bristol specifically, EMHP has continued WW surveillance on all 13 operational catchments and set up four temporary sampling sites in response to P.1 (Fig. 3), based on the location and epidemiology of clinically confirmed cases. Samples collected between 1st – 4th March 2021 from 17 catchments (13 operational and 4 temporary sites, ~68 samples) and between the 11th - 13th of March 2021 from 21 catchments (13 operational and 8 new sites, ~63 samples) were subsequently sequenced. Results

were relayed directly to local, regional and national response teams 10 - 14 days after sample collection to validate clinical findings and provide some level of assurance that the spread of the variant had been contained. No VOC, VUI or mutation of concern (e.g. E484K), bar B.1.1.7, was detected in any WW sample collected in March. Thus, EMHP contributed to the comprehensive genomic surveillance of the whole of Bristol and South Gloucestershire.

Additional use cases in brief

Beyond the detailed Bristol use-case, EMHP are actively contributing to the national VOC/VUI response across England and have provided insight across several cities and regions to date. Noteworthy examples include the detection of all 13 signature SNPs of the B.1.351 lineage (VOC-20DEC-02) from a sewer network site in Nottingham on the 19th March (Fig. 4), as well as the temporal detection of five signature SNPs of the P.2 lineage (VUI-21JAN-01) at a sewer network site in Manchester (Fig. 4). In both cases the majority of signature SNPs of the B.1.1.7 lineage (VOC-DEC20-01) were also observed (Fig. 4), highlighting simultaneous detection of multiple VOC/VUIs from one sample. EMHP are working with local response teams to link virus detection in WW with clinical findings and to aid in monitoring the spread and containment of these localised VOC/VUI outbreaks.

References

Crits-Christoph A., Kantor R.S., Olm M.R., et al. (2021) Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio*, **12**(1):e02703-20. doi:10.1128/mBio.02703-20.

Fontenele R.S., Kraberger S., Hadfield J., et al. (2021) High-throughput sequencing of SARS-CoV-2 in wastewater provides insights into circulating variants. *medRxiv* [Preprint] 2021.01.22.21250320. doi: 10.1101/2021.01.22.21250320.

Jahn K., Dreifuss D., Topolsky I., et al. (2021) Detection of SARS-CoV-2 variants in Switzerland by genomic analysis of wastewater samples. *medRxiv* [Preprint] 2021.01.08.21249379. doi:10.1101/2021.01.08.21249379.

Kazuharu M., Kazutaka K., Kuma K-i. and Miyata T. (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**(14):3059-3066. doi:10.1093/nar/gkf436.

Koboldt D.C., Zhang Q., Larson D.E., et al. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.*, **22**(3):568-576. doi:10.1101/gr.129684.111.

Li H., Handsaker B., Wysoker A., et al. (2009) 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools, *Bioinformatics*, **25**(16):2078-2079. doi:10.1093/bioinformatics/btp352.

Li H. (2020) Minimap2: pairwise alignment for nucleotide sequences, Bioinformatics. 15;34(18):3094-3100. doi: 10.1093/bioinformatics/bty191.

Peccia J., Zulli A., Brackney D.E., et al. (2020) Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat. Biotechnol.* **38**:1164-1167. doi:10.1038/s41587-020-0684-z.

Prado T., Fumian T.M., Mannarino C.F., et al. (2021) Wastewater-based epidemiology as a useful tool to track SARS-CoV-2 and support public health policies at municipal level in Brazil. *Water Res.*, **191**:116810. doi: 10.1016/j.watres.2021.116810.

Rambaut A., Holmes E.C., O'Toole Á., et al. (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.,* **5**:1403-1407. doi:10.1038/s41564-020-0770-5.

R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Shen W., Le S., Li Y. and Hu F. (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PloS ONE*, **11**(10):e0163962. doi: 10.1371/journal.pone.0163962.

Tyson J.R., James P., Stoddart, D., et al. (2020) Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* [Preprint] 2020.09.04.283077. doi: 10.1101/2020.09.04.283077.

Wade M.J., Jones D., Singer A., et al. (2020). Wastewater COVID-19 monitoring in the UK. JBC/DEFRA, 19th November 2020. Report to the Scientific Advisory Group for Emergencies.

Wu F., Zhao S., Yu B., et al. (2020) A new coronavirus associated with human respiratory disease in China. *Nature*, **579**(7798):265-269. doi:10.1038/s41586-020-2008-3.

Appendix

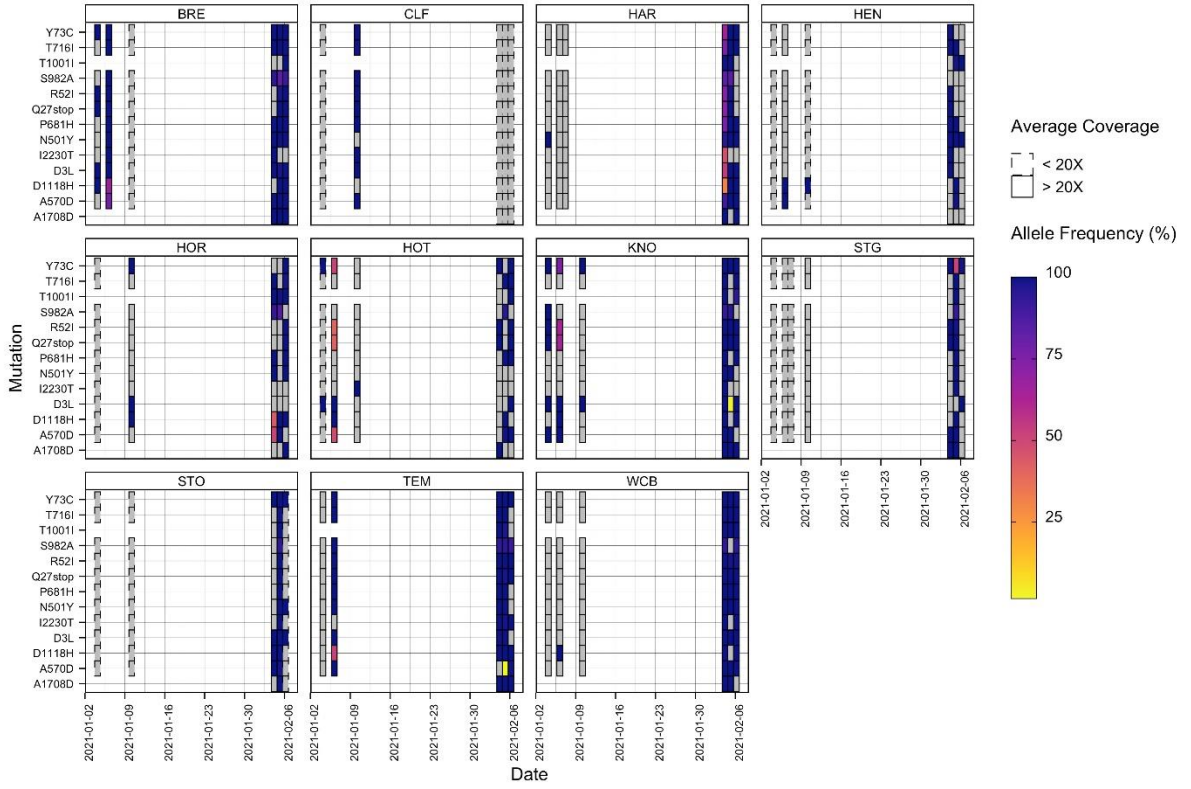

Figure 1. Allele frequencies of the 13 signature SNPs of the B.1.1.7 lineage (VOC-20DEC-01) across Bristol, 4th January to 7th February. Note only samples from the 5th, 7th, 8th and 11th January and 5th, 6th, and 7th February have been sequenced for most sites. Grey or missing boxes highlight the SNP was not detected.

Table 1 – Co-occurrence analysis of the signature mutations of known VOCs and VUIs among the 33 Bristol samples.

Lineage (Amplicon)	B.1.1.7 (78)	B.1.1.7 (92)	B.1.1.7 (93)	B.1.1.7 + E484K (78)	B.1.1.7 + E484K (92)	B.1.1.7 + E484K (93)	B.1.1.7 +E484K (76)	P1 (76)	B.1.351 (76)
Codons within Amplicon	P681H,T716I	Q27*,R52I,Y73C	R52I,Y73C,D3L	P681H,T716I	Q27*,R52I,Y73C	Y73C,D3L	E484K-N501Y	E484K, N501Y	E484K,N501Y
Amplicon ref position	23466 -23822	27808 -28145	28104- 28442	23466 -23822	27808 -28145	28104- 28442	22821-23189	22821-23189	22821 -23189
Variant called (Varscan)	YES/YES	YES/YES/YES	YES/YES/YES	YES/YES	YES/YES/YES	YES/YES/YES	YES/YES	YES/YES	YES/YES
BRE (5th Feb)	173 / 190	0 / 0	508 / 658	173 / 190	0 / 0	508 / 658	1 / 71	1 / 71	1 / 71
	91.05%	NA	77.20%	91.05%	NA	77.20%	1.41%	1.41%	1.41%
BRE (6th Feb)	22 / 25	59 / 80	71 / 105	22 / 25	59 / 80	71 / 105	4 / 390	4 / 390	4 / 390
	88.00%	73.75%	67.62%	88.00%	73.75%	67.62%	1.03%	1.03%	1.03%
BRE (7th Feb)	266 / 306	132 / 199	163 / 226	266 / 306	132 / 199	163 / 226	6 / 1000	6 / 1000	6 / 1000
	86.93%	66.33%	72.12%	86.93%	66.33%	72.12%	0.60%	0.60%	0.60%
CLF (5th Feb)	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
	NA	NA	NA	NA	NA	NA	NA	NA	NA
CLF (6th Feb)	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
	NA	NA	NA	NA	NA	NA	NA	NA	NA
CLF (7th Feb)	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
	NA	NA	NA	NA	NA	NA	NA	NA	NA
HAR (5th)	313 / 465	465 / 997	451 / 1000	313 / 465	465 / 997	451 / 1000	1 / 255	1 / 255	1 / 255
	67.31%	46.64%	45.10%	67.31%	46.64%	45.10%	0.39%	0.39%	0.39%
HAR (6th)	68 / 78	77 / 119	262 / 347	68 / 78	77 / 119	262 / 347	8 / 605	8 / 605	8 / 605
	87.18%	64.71%	75.50%	87.18%	64.71%	75.50%	1.32%	1.32%	1.32%
HAR (7th)	48 / 55	0 / 0	113 / 142	48 / 55	0 / 0	113 / 142	1 / 289	1 / 289	1 / 289
	87.27%	NA	79.58%	87.27%	NA	79.58%	0.35%	0.35%	0.35%
HEN (5th)	651 / 730	444 / 645	764 / 1000	651 / 730	444 / 645	764 / 1000	1 / 173	1 / 173	1 / 173
	89.18%	68.84%	76.40%	89.18%	68.84%	76.40%	0.58%	0.58%	0.58%
HEN (6th)	33 / 37	0 / 0	0 / 0	33 / 37	0 / 0	0 / 0	2 / 134	2 / 134	2 / 134
	89.19%	NA	NA	89.19%	NA	NA	1.49%	1.49%	1.49%
HEN (7th)	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	2 / 211	2 / 211	2 / 211
	NA	NA	NA	NA	NA	NA	0.95%	0.95%	0.95%
HOR (5th)	203 / 221	0 / 684	0 / 683	203 / 221	0 / 684	0 / 683	5 / 644	5 / 644	5 / 644
	91.86%	0.00%	0.00%	91.86%	0.00%	0.00%	0.78%	0.78%	0.78%
HOR (6th)	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
	NA	NA	NA	NA	NA	NA	NA	NA	NA
HOR (7th)	62 / 68	291 / 433	0 / 0	62 / 68	291 / 433	0 / 0	4 / 291	4 / 291	4 / 291
	91.18%	67.21%	NA	91.18%	67.21%	NA	1.37%	1.37%	1.37%
HOT (5th)	0 / 0	118 / 196	0 / 0	0 / 0	118 / 196	0 / 0	0 / 0	0 / 0	0 / 0
	NA	60.20%	NA	NA	60.20%	NA	NA	NA	NA
HOT (6th)	27 / 34	0 / 0	3 /7	27 / 34	0 / 0	3/ 7	0 / 0	0 / 0	0 / 0
	79.41%	NA	42.86%	79.41%	NA	42.86%	NA	NA	NA
HOT (7th)	10/13	55 / 90	17 / 25	10 /13	55 / 90	17 / 25	0 / 0	0 / 0	0 / 0
	76.92%	61.11%	68.00%	76.92%	61.11%	68.00%	NA	NA	NA
KNO (5th)	270 / 291	528 / 817	767 / 1000	270 / 291	528 / 817	767 / 1000	2 / 246	2 / 246	2 / 246
	92.78%	64.63%	76.70%	92.78%	64.63%	76.70%	0.81%	0.81%	0.81%
KNO (6th)	0 / 0	74 / 125	80 / 116	0 / 0	74 / 125	80 / 116	4 / 422	4 / 422	4 / 422
	NA	59.20%	68.97%	NA	59.20%	68.97%	0.95%	0.95%	0.95%
KNO (7th)	36 / 38	77 / 118	41 / 68	36 / 38	77 / 118	41 / 68	0 / 0	0 / 0	0 / 0
	94.74%	65.25%	60.29%	94.74%	65.25%	60.29%	NA	NA	NA
STG (5th)	0 / 0	333 / 469	0 / 0	0 / 0	333 / 469	0 / 0	0 / 0	0 / 0	0 / 0
	NA	71.00%	NA	NA	71.00%	NA	NA	NA	NA
STG (6th)	57 / 66	77 / 133	0 / 177	57 / 66	77 / 133	0 / 177	2 / 167	2 / 167	2 / 167
	86.36%	57.89%	0.00%	86.36%	57.89%	0.00%	1.20%	1.20%	1.20%
STG (7th)	0 / 0	0 / 0	65 / 88	0 / 0	0 / 0	65 / 88	0 / 0	0 / 0	0 / 0
	NA	NA	73.86%	NA	NA	73.86%	NA	NA	NA
STO (5th)	0 / 0	0 / 0	489 / 641	0 / 0	0 / 0	489 / 641	0 / 0	0 / 0	0 / 0
	NA	NA	76.29%	NA	NA	76.29%	NA	NA	NA
STO (6th)	106 / 117	89 / 131	232 / 317	106 / 117	89 / 131	232 / 317	4 / 782	4 / 782	4 / 782
	90.60%	67.94%	73.19%	90.60%	67.94%	73.19%	0.51%	0.51%	0.51%
STO (7th)	0 / 0	0 / 0	23 / 28	0 / 0	0 / 0	23 / 28	2 / 163	2 / 163	2 / 163
	NA	NA	82.14%	NA	NA	82.14%	1.23%	1.23%	1.23%
TEM (5th)	280 / 318	395 / 616	774 / 1000	280 / 318	395 / 616	774 / 1000	7 / 651	7 / 651	7 / 651
	88.05%	64.12%	77.40%	88.05%	64.12%	77.40%	1.08%	1.08%	1.08%
TEM (7th)	58 / 64	57 / 96	251 / 353	58 / 64	57 / 96	251 / 353	0 / 606	0 / 606	0 / 606
	90.62%	59.38%	71.10%	90.62%	59.38%	71.10%	0.00%	0.00%	0.00%
TEM (7th)	0 / 0	36 / 59	0 / 0	0 / 0	36 / 59	0 / 0	0 / 1	0 / 1	0 / 1
	NA	61.02%	NA	NA	61.02%	NA	0.00%	0.00%	0.00%
WCB (5th)	398 / 437	643 / 997	806 / 1000	398 / 437	643 / 997	806 / 1000	1 / 265	1 / 265	1 / 265
	91.08%	64.49%	80.60%	91.08%	64.49%	80.60%	0.38%	0.38%	0.38%
WCB (6th)	28 / 32	113 / 177	120 / 162	28 / 32	113 / 177	120 / 162	3 / 396	3 / 396	3 / 396
	87.50%	63.84%	74.07%	87.50%	63.84%	74.07%	0.76%	0.76%	0.76%
WCB (7th)	154 / 173	82 / 139	419 / 587	154 / 173	82 / 139	419 / 587	9 / 999	9 / 999	9 / 999
	89.02%	58.99%	71.38%	89.02%	58.99%	71.38%	0.90%	0.90%	0.90%
Negative Control	*0 / 0*	*0 / 0*	*0 / 0*	*0 / 0*	*0 / 0*	*0 / 0*	*0 / 0*	*0 / 0*	*0 / 0*
	NA	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*

Table 2 – Observed total and mapped reads and reads associated with specific genomic locations in each sample

Site	Date	Total Reads	Mapped Reads	Total Reads at position 23,012	Reads with E484K mutation (G>A) at position 23,012 (sense strand)	Reads with E484K mutation (A) at position 23,012 (anti-sense strand)	Reads with N501Y mutation (A>T) at position 23,063 (both strands)	Reads with both E484K +N501Y mutations
BRE	05/02/2021	17190732	49824	60				
HOR	05/02/2021	17958886	47271	568		4	488	4
HAR	05/02/2021	35551979	98829	212	3	2	178	
HOT	05/02/2021	1039209	2781					
STG	05/02/2021	11949711	31128					
WCB	05/02/2021	42946519	122882	225			194	
HEN	05/02/2021	10681222	28126	142	2	2	90	2
STO	05/02/2021	6854167	17701					
TEM	05/02/2021	12528409	32917	563		4	472	4
KNO	05/02/2021	12332196	32423	216	3		160	
CLF	05/02/2021	5673	21					
BRE	06/02/2021	2648930	6959	329	4		240	
HOR	06/02/2021	1527638	4061					
HAR	06/02/2021	3520002	9277	513		2	302	2
HOT	06/02/2021	1255097	3331					
STG	06/02/2021	5830684	15516	142	3		2	
WCB	06/02/2021	2898009	7699	314		2	222	2
HEN	06/02/2021	1635476	4336	116			62	
STO	06/02/2021	7677285	20289	641	8	4	446	2
TEM	06/02/2021	6682890	17565	497	7		320	
KNO	06/02/2021	4516686	11821	349			244	
CLF	06/02/2021	95182	253					
BRE	07/02/2021	7189133	19393	858			280	
HOR	07/02/2021	2167716	5838	468	6	2	140	2
HAR	07/02/2021	1944276	5193	240	3	2	184	
HOT	07/02/2021	897836	2490					
STG	07/02/2021	622276	1700					
WCB	07/02/2021	19085137	50750	870				
HEN	07/02/2021	1138957	3158	338	4	2	94	2
STO	07/02/2021	164009	454	138				
TEM	07/02/2021	1285743	3361	1				
KNO	07/02/2021	1105875	3084					
CLF	07/02/2021	51821	146				2	

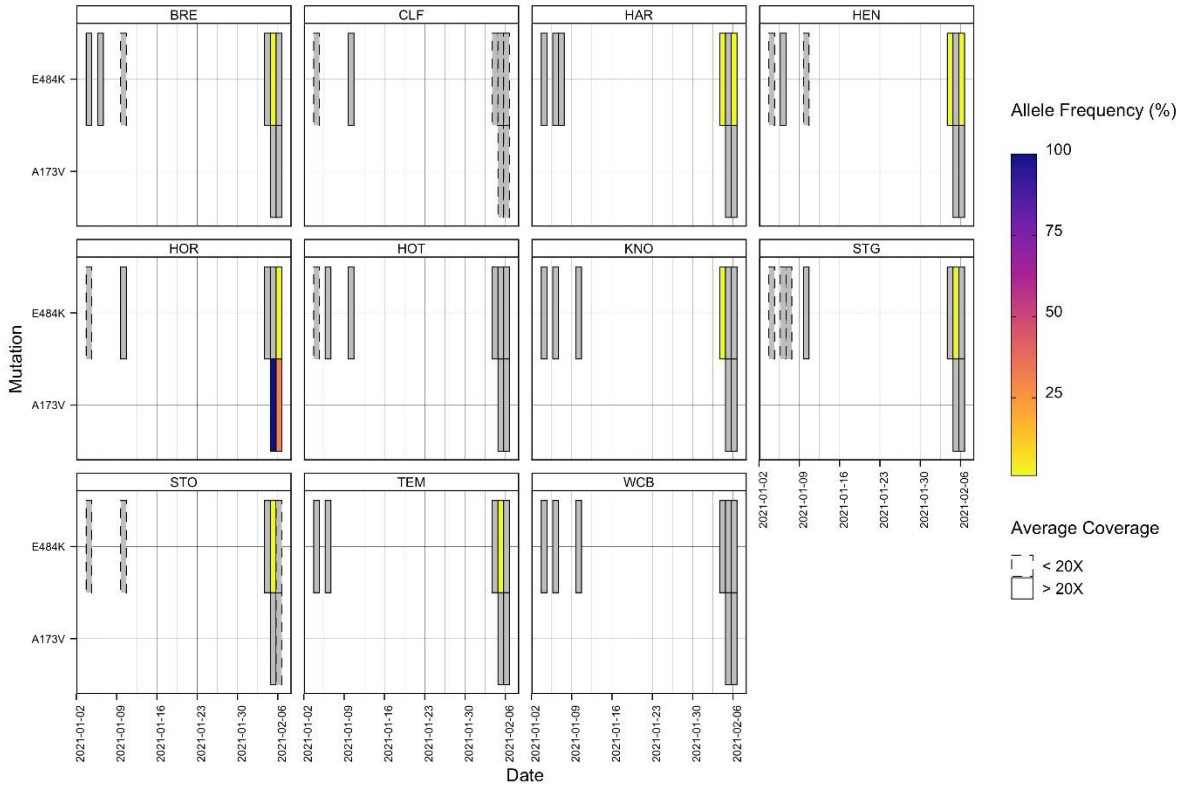

Figure 2. Allele frequencies of two of the additional signature SNPs of the B.1.1.7 lineage with the E484K mutation (VOC-21FEB-02) across Bristol, 4th January to 7th February. Note the other three SNPs were not detected. Only samples from the 5th, 7th, 8th and 11th January and 5th, 6th, and 7th February have been sequenced for most sites. Grey or missing boxes highlight the SNP was not detected.

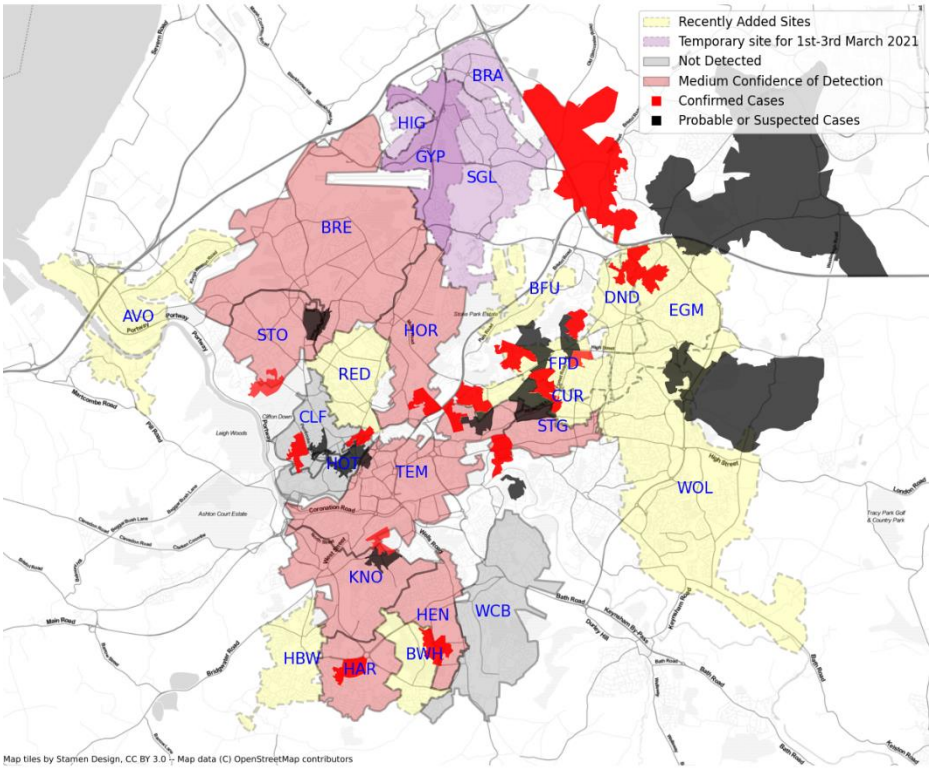

Figure 3. E484k detection (red shaded) and non-detection (grey shaded) in Bristol WW sampling catchments between the 5th, 6th and 7th of February overlaid with confirmed (red areas) and probable, possible and suspected (black areas) clinical cases of VOC-21FEB-02 between the 17th December 2020 and the 2nd February 2021. New sampling catchments (yellow) introduced in response to WW E484K detection and to support surge testing, AVO and RED introduced 09.02.21 with all others added 08.03.21. Temporary sampling catchments (Purple) introduced to support P.1 response, 01.03.21 – 03.03.21.

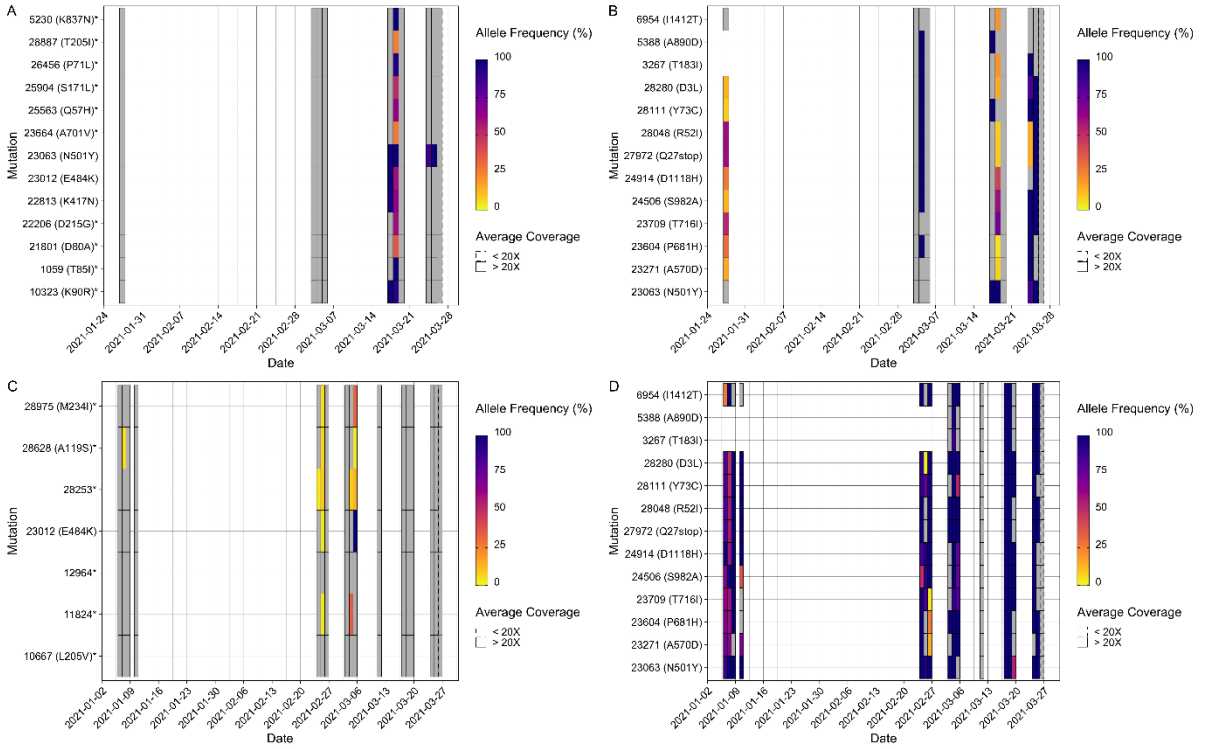

Figure 4. Allele frequencies of the 13 signature SNPs of the B.1.351 lineage (**A**, VOC-20DEC-02) and the 13 signature SNPs of the B.1.1.7 lineage (**B**, VOC-20DEC-01) in a Nottingham sewer network site, as well as allele frequencies of the 7 signature SNPs of the P.2 lineage (**C**, VUI-21JAN-01) and the 13 signature SNPs of the B.1.1.7 lineage (**D**, VOC-20DEC-01) in a Manchester sewer network site. Grey or missing boxes highlight the SNP was not detected. *Mutation unique to VOC/VUI when compared to other known VOCs and VUIs.