

TECHNICAL BULLETIN: SARS CoV-2 Test Variant – VOC Omicron

Background:

Variant of Concern (VOC) Omicron (Pango lineage B.1.1.529), emerged during the week of November 22, 2021. The origin of this variant is currently unclear as there is not yet enough data to determine the origin. It was first discovered in South Africa and has now been detected in several other countries, including Botswana, Hong Kong, United Kingdom, and the United States of America. VOC Omicron is characterized by an unusually large number of amino acid substitutions in the Spike protein, including several in the receptor-binding domain and a two amino deletion at positions 69/70 also observed in VOC Alpha, as well as a three amino acid insertion at position 214. *In vitro* studies suggest the combinations of mutations increase transmissibility of the virus, though it is unknown how they impact disease severity.

DTX COVID-19 Assay Targets

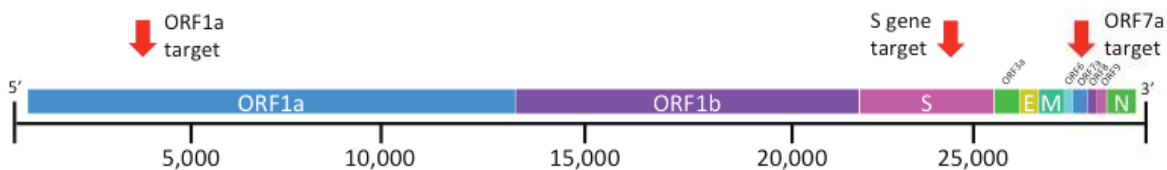


Figure 1: Diagram of SARS-CoV-2 genome and COVID-19 panel target loci

The Eurofins Diatherix EUA approved SARS-CoV-2 assay uses a target in the S gene for detection. Two additional targets in the ORF1a and ORF7a genes have been added to the assay and an amendment submitted to the FDA for approval (Figure 1). If any one of these targets results in a positive result, the sample is reported as “Detected” for SARS-CoV-2. Each target is amplified by two sets of nested primers, and products are hybridized to a fluorescently labeled detection probe. This defines mutations for VOC Omicron as determined by the CoVariants Project within the assay target regions as shown in Figure 2.

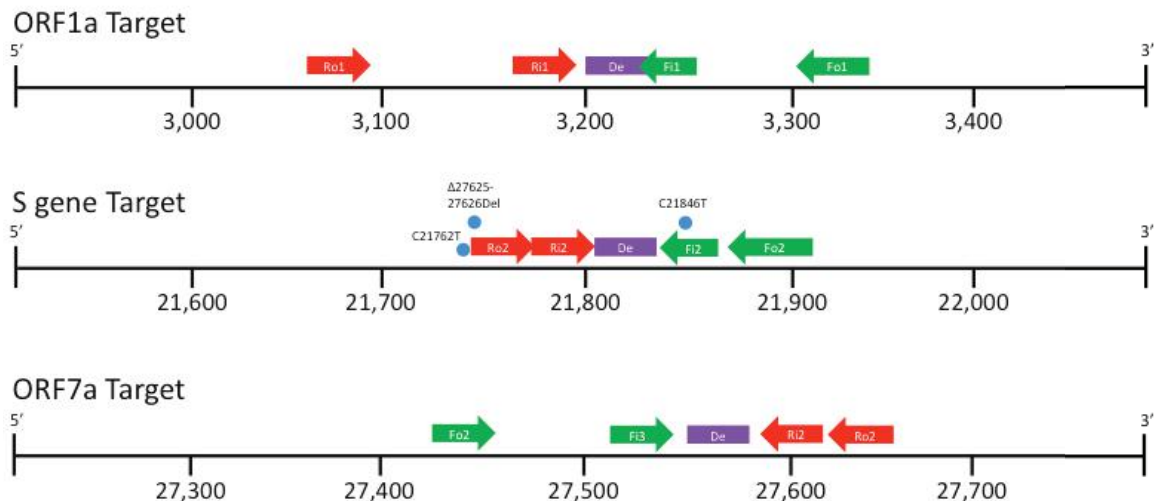


Figure 2: Diagrams of primer design for the three target regions in the assay. Axes denote absolute position at the genome level in nucleotides. Green and red arrows represent forward and reverse primers, respectively. Purple boxes represent detection probes. Blue circles and their captions represent Omicron variant-defining mutations identified by the CoVariants Project (covariants.org).

Mutations within some elements of the assay will have more significant effects on results than others. Outer primer sets generally serve to increase target template abundance in a PCR reaction, so a mutation that impacts their binding affinity will likely only have a minor impact on the assay's sensitivity for that target. However, mutations that impact binding affinity for the inner primers or the detection probe may have a more significant effect. Additionally, mutations within a few bases of the 3' end of a primer will be more likely to prevent extension during PCR and thus prevent detection.

As shown in Figure 2, there are two defining mutations in VOC Omicron within target regions of our assay, both of which are located in the S gene. One of these is a 2-base deletion in the S gene also observed in VOC Alpha, which sits right at the 5' end of the forward outer primer, Fo2. While this deletion is not in a particularly deleterious region of our assay, we were concerned that the deletion may impact our assay. As a result, we performed our assay using synthetic VOC Alpha S gene template to determine the effect on assay LOD and found no change to the assay's sensitivity. The other is a C – T transition at locus 21,846, which lies at position 14 of S gene forward inner primer, S_Fi2. While it is currently unclear how this will impact the sensitivity of the S gene assay, the two remaining targets are unaffected by mutations in VOC Omicron, so overall assay inclusivity should not be compromised.

Database Results

As the defining mutations do not represent all mutations present in VOC Omicron, we performed a BLAST analysis using primer sequences to query a database of VOC Omicron genomes, the results of which are listed in the supplementary files included with this report. For a detailed description of the methods used, refer to the methods section below. All sequences containing SNPs within the primer binding sites are provided in full in the "total_snps" folder, some of which are broken down into 4 categories that are relevant to the sensitivity of the assay:

- 1.) 3' end SNPs occur within 5 bases of the 3' end of the PCR oligo, which could significantly reduce the sensitivity of the assay to detect SARS-CoV-2 genomes that possess this mutation.
- 2.) Abundant SNPs occur in greater than 1% of the SARS-CoV-2 genome database as a whole and represent mutations that are being maintained by virus populations, potentially reducing the sensitivity of our assay to an emerging clade.
- 3.) Gaps are instances of insertions or deletions in the PCR oligo binding region, and may significantly reduce the sensitivity of our assay.
- 4.) Truncations occur when 3' end of the primer do not match the binding region of the genome due to strings of mutations, which would prevent extension during amplification and potentially prevent detection.

Of the 158 genomes in the coronavirus genome database remaining after quality filtering, none were perfect matches to all primers and probes from all three assays presented here (S gene, ORF7a, and ORF1ab). This is due to the two amino acid deletion in the S gene, as described above, which intersects with the 5' end of the reverse outer primer. Excluding these, SNPs were detected only in one other alignment between the ORF1a_Fo1 primer and a single reference genome (T – C transition at position 12). No alignments had SNPs within 5 bases of the 3' end of the S gene primers/probe, none contained gaps, and none were truncated at the 3' end of primers. The genomes that were not 100% matches to all primers/detection elements for a given assay were cross-referenced against the other 2 assays to see if they would catch any missed strains. No genomes were found that did not match to at least one entire assay.

Methods

The most recent SARS-CoV-2 genome database was downloaded from the GISAID website on 10/12/2021, containing 4,064,860 sequences. All genomes not identified as VOC Omicron were filtered from the database, leaving 1,119,374 sequences. This database was then filtered to remove genomes with excessive ambiguous bases, generated during sequencing, within the target regions for our assay. Genomes with more than one “N” ambiguous base pairs within 20 bases upstream and downstream of at least one primer annealing site were removed from the database prior to analysis. Eurofins Diatherix TEM-PCR primers were then aligned against the genome database using blastn-short with a maximum e-value threshold of 100 and max_target_seqs set to the size of the reference database to capture as many alignments as possible. BLAST results were then parsed using a custom Python pipeline to build this report and the supporting data files, which show lists of reference sequences that are either 100% identical to Eurofins Diatherix PCR oligonucleotides or contain mutations.

Data Organization

The code used to generate this report is stored in the “code” folder included with this report. Also included are the results of the analysis presented here, which are stored in the “data” folder. The data folder contains a number of sub-directories where the results are organized. The “matches” folder contains a list of all IDs corresponding to genomes that had a 100% match to each primer or primer set. The “total_snps” folder contains alignments showing every mutation detected in a primer binding region. The “abundant_snps” folder contains files for specific SNPs detected in more than 1% of the filtered genome database for a given primer. The “3_prime_snps” folder contains alignments for all SNPs detected within 5 bases of the 3’ end of a given primer. The “gaps” folder contains alignments for all gaps detected for each primer. The “truncations” folder contains alignments for all instances where the 3’ end of a primer did not align to a genome. The “misses” folder contains a list of any genomes that were not covered by at least one of the 3 targets included in this assay.

Note

Prior to official designation by the World Health Organization, initial reporting prematurely designated the variant by the next letter in the Greek alphabet “Nu” (ν). The WHO skipped Nu and Xi in accordance with established best practices for naming new diseases, designating Pango lineage B.1.1.529 as VOC Omicron.