TITLE: *In Silico* Design of Specific Primer Sets for the Detection of B.1.1.529 SARS-CoV-2 Variant of Concern (Omicron)

AUTHORS: Mauro Petrillo[1], Maddalena Querci[2], Philippe Corbisier[3], Antonio Marchini[3], Gerhard Buttinger[3] and Guy Van den Eede[3#]***.***

Author details:

1. Seidor Italy srl, Milano, Italy. *Past affiliation (until 15/06/2021) European Commission, Joint Research Centre (JRC), Ispra, Italy*.
2. European Commission, Joint Research Centre (JRC), Ispra, Italy.
3. European Commission, Joint Research Centre (JRC), Geel, Belgium.

[#]Corresponding author: Guy.VAN-DEN-EEDE@ec.europa.eu

ABSTRACT

On 26 November 2021, the World Health Organization (WHO) classified variant B.1.1.529, reported earlier by the South African authorities as "Variant of Concern" with the name Omicron. The decision was taken because Omicron is bearing a number of mutations with a potential impact on its transmissibility, severity of disease following infection, as well as on the effectiveness of immune protection resulting from both vaccines and natural infection. In this paper, we present an *in silico* developed RT-PCR based detection method (named OmMet) that was designed to be highly specific for the detection of the Omicron variant. Bioinformatic prediction tests demonstrate that the sequences in the primer sets are highly accurate, and do not match with genetic sequences of other viruses, including other coronaviruses or SARS-CoV-2 variants. The methodology presented does not rely on S-gene target failure (SGTF) of existing RT-PCR assays, widely used currently, but allows the direct specific identification of the variant B.1.1.529 (Omicron).

INTRODUCTION

The B.1.1.529 variant was first reported to the World Health Organization (WHO) from Republic of South Africa on 24 November 2021. First identified in countries in the southern part of Africa, the variant was then quickly detected in a growing number of countries in the rest of the world. On 26 November 2021, WHO classified variant B.1.1.529 as Variant of Concern (VOC) and named it "Omicron", after the 15th letter of the Greek alphabet [1].

Omicron is characterised by more than 50 nucleic acid changes (including deletions) as compared to the reference SARS-CoV-2 sequence. The majority of these are found in the gene encoding for the spike glycoprotein [2]. The spike protein is a key structure that facilitates the viral entry to human cells through interacting with the human ACE2 receptor [3,4], and is also the target of mRNA- and viral vector-based vaccines [5-8].

At the time of writing (2 December 2021), 329 SARS-CoV-2 genomic sequences have been deposited in GISAID [9], 305 of them were flagged as complete and with less than 5% undetermined bases. By inspecting the spike protein encoding regions of all Omicron-flagged complete sequences, we identified a target region with a unique and Omicron-specific cluster of nucleic acid sequence. Using this region, we designed a variant-specific set of primers to be used as real-time polymerase chain reaction (RT-PCR) Omicron detection method, called OmMet that we propose here.

METHODOLOGY

1. SARS-CoV-2 genomic sequences deposited in GISAID, flagged "complete" and with less than 5% undetermined bases have been downloaded and analysed with Nextclade [10]. Only mutations present in at least 80% of the analysed sequences were retained to define a set of commonly shared mutations.
2. The *S* gene, in which most of these mutations occur, was manually inspected to look for genomic regions not longer than 250 bp and with at least 6 mutations. The National Center for Biotechnology Information (NCBI) Reference Sequence [11] NC_045512.2 was used as reference [12]. A region (NC_045512.2:22950-23150) was identified with these features and fed into *Primer3Plus* [13] to design *in silico* primers and probes methods (program run with pre-loaded qPCR settings).
3. A candidate method, specific for the genomic region NC_045512.2:22991-23128, was selected. We call this method OMicron METhod (OmMet).
4. OmMet has been tested *in silico* by using the *thermonucleotideBLAST* software [14] with the following parameters: *-e 20 -E 20 -l 200* on the selected set of SARS-CoV-2 genomic sequences.
5. A set of Pango lineages [15] consensus sequences was obtained by running an in-house developed script that parses data retrieved from the Broad Institute COVID CG [16] application programming interface (used *consensus_threshold=0.9*). This set was used to check *in silico* the OmMet specificity.
6. Sequence alignments were produced by using MAFFT online service [17].
7. Alignments' representations were obtained by running the *showseq* tool of EMBOSS package [18].

RESULTS

OmMet specifications

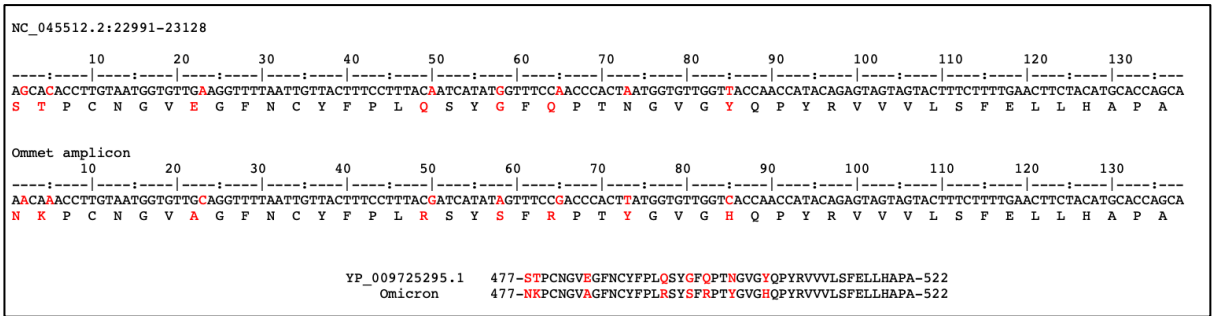
The proposed sequences for the forward primer, the reverse primer and the probe are summarised in Table 1, where Omicron-specific nucleic acid base changes are highlighted in red.

Table 1. Forward primer, the reverse primer and the probe sequences.

Name	Code	Oligo	Length (bp)	GC (%)	Tm (°C)	Position (NC_045512.2)
OmMet	Omt-F	5′-AACAAACCTTGTAATGGTGTTGC-3′	23	39.1	58.6	22991-23128 size: 138bp
	Omt-R	5′-TGCTGGTGCATGTAGAAGTTC-3′	21	47.6	58.8	
	Omt-P	5′-FAM-GATCATATAGTTTCCGACCCACTTATGGTGTTGGTC-QSY-3′	28	50	60.6	
Amplicon	AACAAACCTTGTAATGGTGTTGCAGGTTTTAATTGTTACTTTCCTTTACGATCATATAGTTTCCGACCCA CTTATGGTGTTGGTCACCAACCATACAGAGTAGTAGTACTTTCTTTTGAACTTCTACATGCACCAGCA					

The overview of the OmMet target amplicon compared to the SARS-CoV-2 homologous sequence (from NC_045512.2) is shown in Figure 1, together with the alignment of the corresponding protein regions (477-522) of the spike Omicron and of reference proteins.

Figure 1. Visual comparison of SARS-CoV-2 reference and Omicron amplicons targeted by OmMet, together with the alignment of the two corresponding translated regions. Differences are highlighted in red.



In silico validation

OmMet was tested *in silico* and found with perfect matching on the selected set of Omicron sequences retrieved from GISAID (data not shown). Additionally, OmMet was tested *in silico* on the consensus sequences of all known Pango lineages. In all of the analysed sequences, differences were found in the regions of annealing of both the forward primer and the hybridisation oligo sequences (see Supplementary file).

These results confirm *in silico* OmMet's capacity to detect the Omicron sequences with very high specificity.

DISCUSSION AND CONCLUSIONS

The epidemic potential of the new SARS-CoV-2 B.1.1.529 variant is of concern due to its apparent higher transmissibility and the likely negative effect that many of the mutations it features might have upon natural or vaccine-induced immunity and on the ability to detect the virus. To our knowledge, no other sequence-specific RT-PCR tests which are capable to distinguish this new VOC from other SARS-CoV-2 variants have been published to date.

Currently, public health officials mostly rely on the so-called S-gene target failure (SGTF) phenomenon to quickly identify Omicron in patient samples. SGTF occurs in commercial assays designed to detect sequences from three SARS-CoV-2 genes – S, N and ORF1ab. Due

to the large number of mutations in the S gene of Omicron, in the presence of Omicron, these assays give positive signal only for N and ORF1ab targets, but negative for S. This differential result is interpreted as corresponding to a SARS-CoV-2 positive sample containing the Omicron variant. Although much more rapid than sequencing identification, SGTF suffers from lower specificity. For example, the S gene of a number of rare SARS-CoV-2 variants features a genetic sequence that is not recognised by the assays. In addition, swab samples with lower virus concentration may fail to detect S. Furthermore, the S gene of some actual Omicron subvariants can in fact be detected by the tests, resulting in a lack of discrimination with the SGTF approach. The method we propose allows for a very high accuracy, typical for the sequencing approaches, while allowing much shorter turn over times needed for public health response and decision making.

This study describes the design of primers and probe to identify the Omicron variant by a RT-PCR. RT-PCR is simple to set up in any biotech laboratory that can conduct PCR assays and should aid in globally monitoring the spread of the Omicron variant. The method was validated *in silico* and is recommended for the specific the detection of the Omicron variant. We feel its immediate sharing with the scientific community is critical and invite control laboratories in the world to validate it *in vivo* on clinical samples and to report results back to authors in order to improve the test if needed.

Additional information

The authors declare no competing interests.

ACKNOWLEDGEMENTS

We gratefully acknowledge the Authors and the Originating laboratories where the clinical specimen or virus isolate was first obtained and the Submitting laboratories, where sequence data have been generated and submitted to GISAID, on which this research is based.
We gratefully acknowledge the COVID CG team of the BROAD INSTITUTE for helping us in setting up the query to the COVID CG application programming interface (API) and Peter Markov (JRC) for the critical revision of the manuscript.

REFERENCES

1. https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern (last access: 01-12-2021).
2. Callaway E. Heavily mutated Omicron variant puts scientists on alert. Nature. 2021. doi: 10.1038/d41586-021-03552-w.
3. Shang J et al. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci USA. 2020. 117(21):11727-11734. doi:10.1073/pnas.2003138117.
4. Yang J et al. Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor. Nat Commun. 2020. 11(1):4541. doi: 10.1038/s41467-020-18319-6.
5. Voysey M et al. Lancet. 2021. 397(10269):99-111. doi: 10.1016/S0140-6736(20)32661-1.
6. Baden LR et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. N Engl J Med. 2021. 384(5):403-416. doi: 10.1056/NEJMoa2035389.

7. Skowronski DM et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med. 2021. 384(16):1576-1577. doi: 10.1056/NEJMc2036242.
8. Sadoff J et al. Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. N Engl J Med. 2021. 384(23):2187-2201. doi: 10.1056/NEJMoa2101544.
9. Elbe S et al. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017. 1(1):33-46. doi: 10.1002/gch2.1018.
10. Aksamentov I et al. Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software. 2021. 6(67):3773. doi: 10.21105/joss.03773.
11. O'Leary NA et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016. 44(D1):D733-45. doi: 10.1093/nar/gkv1189.
12. Wu F et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020. 579(7798):265-269. doi: 10.1038/s41586-020-2008-3.
13. Untergasser A et al. Primer3--new capabilities and interfaces. Nucleic Acids Res. 2012. doi: 10.1093/nar/gks596.
14. Gans JD, Wolinsky M. Improved assay-dependent searching of nucleic acid sequence databases. Nucleic Acids Res. 2008. 36(12):e74. doi: 10.1093/nar/gkn301.
15. Rambaut A et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020. 5(11):1403-1407. doi: 10.1038/s41564-020-0770-5.
16. Chen AT et al. COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. Elife. 2021. 10:e63409. doi: 10.7554/eLife.63409.
17. Katoh K et al. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019. 20(4):1160-1166. doi: 10.1093/bib/bbx108.
18. Rice P et al. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000. 16(6):276-7. doi: 10.1016/s0168-9525(00)02024-2.