Technical bulletin

SARS-CoV-2 variants and mutations

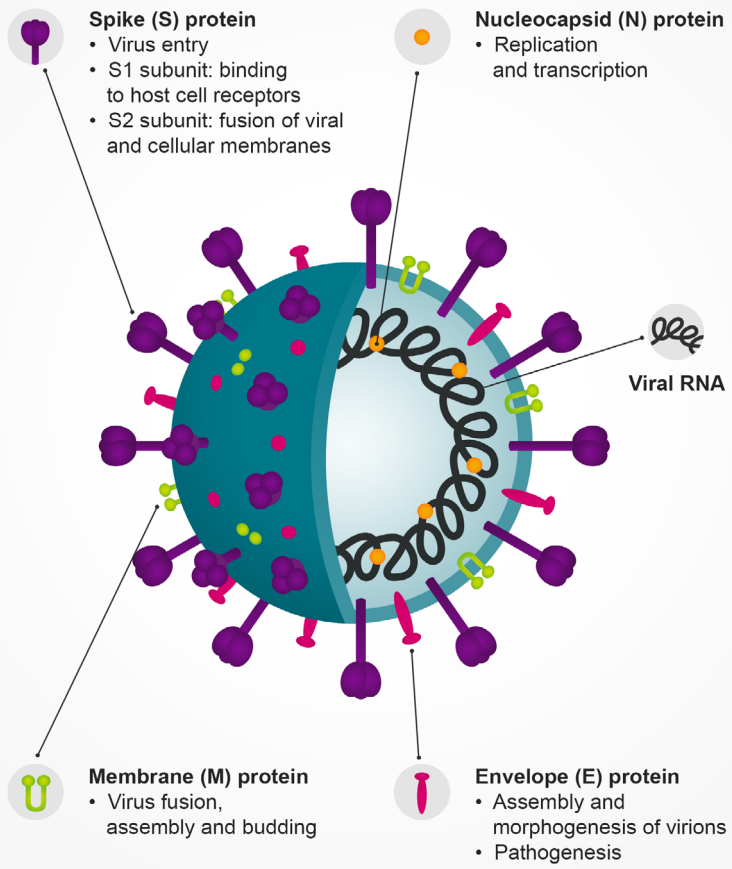

Figure 1. SARS-CoV-2 viral structure

Introduction

Coronaviruses are enveloped, single-stranded RNA viruses that have an outer crown-like appearance, from which they derive their name. They are grouped into four genera of which alphacoronavirus and betacoronavirus are found in bats and rodents. SARS-CoV-2, responsible for the current COVID-19 pandemic, is a betacoronavirus thought to have transferred from bat to human. The virus contains four structural proteins: the envelope (E), spike (S), membrane (M) and nucleocapsid (N).[1, 2]

The viral envelope is formed from the E, S and M proteins. The N protein and genomic RNA form the nucleocapsid within the envelope (Figure 1).

BIOSEARCH™ TECHNOLOGIES
GENOMIC ANALYSIS BY LGC

Technical bulletin

SARS-CoV-2 variants and mutations

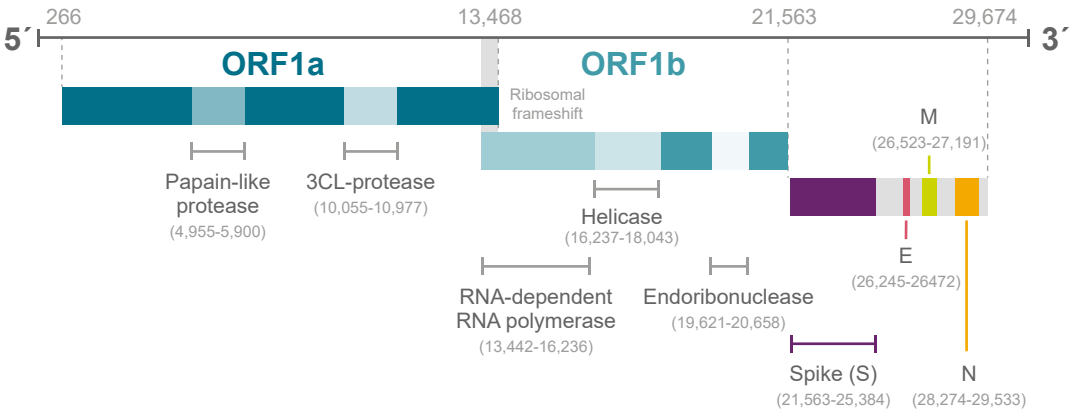

Figure 2. Arrangement of genes in the SARS-CoV-2 genome

SARS-CoV-2 genome

The SARS-CoV-2 genome is a positive-sense, single-stranded RNA of 29,891 nucleotides, encoding 9,860 amino acids. The genome is arranged as: 5′-leader sequence>UTR> replicase>S>E>M>N>3′UTR>poly A tail, with accessory genes located towards the 3′ end of the genome (Figure 2).

SARS-CoV-2 variants

Following the publication of the original SARS-CoV-2 sequence[3] isolated from clinical samples in Wuhan (China), a number of variant strains have been described. Some of these variants contain defining mutations of clinical significance. Of these, four variants have been associated with higher transmission rates, or the potential to evade an immune response originally raised to a different strain.[4]

Gene	Length (nt)	% genome	Approx. rate of mutation
N	1,259	4.2	1/1,000
S	3,821	12.7	2.3/1,000
ORF1ab	21,289	71.2	3/10,000
ORF8	365	1.2	7/1,000
ORF3a	827	2.8	6/10,000
E	227	<1	1/1,000

Table 1. Mutation rate in SARS-CoV-2 genome

Mutations may be a single base change, insertions, deletions or recombination events. These are defined according to the RNA or amino acid location.

RNA base notation:
The base notation for a significant mutation in South African (B1.351): G23012A. This describes the original base at position 23012 as G but has now mutated to A.

Amino acid notation:
At the amino acid level, this mutation is described as E484K – the amino acid at 484 was E (glutamic acid) in the original strain but has mutated to K (lysine).

E (glutamic acid) is encoded by RNA bases GAA or GAG.
K (lysine) is encoded by AAA or AAG.

484	Codon number	481	482	483	484	485	486	487
484E	Nucleotide sequence	AAT	GGT	GTT	GAA	GGT	TTT	AAT
	Protein sequence	N	G	V	E	G	F	N
484K	Nucleotide sequence	AAT	GGT	GTT	AAA	GGT	TTT	AAT
	Protein sequence	N	G	V	K	G	F	N

Box 1. Describing a SARS-CoV-2 mutation

Four current variants of concern are referred to as the UK (B1.1.7), B1.1.7 + E484K, South African (B1.351, 501v2) and Brazilian (B1.1.28, P1). Each of these strains have mutations in the spike gene which code for novel spike proteins that are potentially not recognised by previously trained immune systems.

Technical bulletin

SARS-CoV-2 variants and mutations

Four variants of concern have been identified.[4] These share some mutations but differ in others. The mutations in these variants are listed in Table 2 (x indicates the mutant base is present).

N gene mutation	Nucleotides	B1.1.7 UK	B1.1.7 + E484K	B1.351 S. Africa	B1.1.28, Brazil
D3L	28280-2 2delinsCTA	x	x		
A173V			x		
A398T			x		
T205I				x	
P80R	C28512G				x
	A28877T				x
	A28878T				x
SNP, N-mid[11]	C29200T				
SNP, N-Nterm[11]	C28858T				
SNP, N-Cterm[11]	C29451T				
S gene mutation	**Nucleotides**	**B1.1.7 UK**	**B1.1.7 + E484K**	**B1.351 S. Africa**	**B1.1.28, Brazil**
del 69/70	21765-21770	x	x		
del Y144	21991-21993	x	x		
N501Y	A23063T	x	x	x	x
A570D	C23271A	x	x		
P681H	C23604A	x	x		
T716I	C23709T	x	x		
S982A	T24506G	x	x		
D1118H	G24914G	x	x		
E484K	G23012A		x	x	x
L18F	C21614T			x	x
D80A	A21801C			x	
D215G	A22206G			x	
R246I	G22299T			x	
K417N	G22813T			x	
K417T	A22813C				x
A701V	C23664T			x	
del 242-244				x	
T20N	C21621A				x
P26S	C21638T				x
D138Y	G21974T				x
R190S	G22132T				x
H655Y	C23525T				x
T1027I	C24642T				x

Table 2. Mutations associated with variants of concern.

Technical bulletin

SARS-CoV-2 variants and mutations

ORF1ab	Nucleotides	B1.1.7 UK	B1.1.7 + E484K	B1.351 S. Africa	B1.1.28, Brazil
T1001I	C3267T	x	x		
A1708D	C5388A	x	x		
I2230T	T6954C	x	x		
del 3675-3677	11288-96 del	x	x	x	x
L730F			x		
T265I	C1059T			x	
K1655N	G5230T			x	
K3352R	A10323G			x	
	T733C				x
	C2749T				x
S1188L	C3828T				x
K1795Q	A5648C				x
	C12778T				x
	C13860T				x
E5665D	G17259T				x
ORF8	**Nucleotides**	**B1.1.7 UK**	**B1.1.7 + E484K**	**B1.351 S. Africa**	**B1.1.28, Brazil**
Q27*	C27972T	x	x		
R521	G28048T	x	x		
Y73C	A28111G	x	x		
E92K	G28167A				x
	28263insAACA				x
ORF3a	**Nucleotides**	**B1.1.7 UK**	**B1.1.7 + E484K**	**B1.351 S. Africa**	**B1.1.28, Brazil**
Q57H	G25563T			x	
S171L	C25904T			x	
E gene	**Nucleotides**	**B1.1.7 UK**	**B1.1.7 + E484K**	**B1.351 S. Africa**	**B1.1.28, Brazil**
P71L	C26456T			x	

Table 2. Mutations associated with variants of concern. (continued)

On average, a typical SARS-CoV-2 virus appears to be acquiring two single-letter mutations per month in its genome.[5] This is slower than other RNA viruses, such as HIV and influenza, possibly due to the proofreading RNA polymerase, with identical mutations arising independently. Mutations have occurred along the length of the genome, although approximately 50% of the reported mutations have been observed in the spike gene (Table 1). The S gene represents only approximately 13% of the entire genome and encodes the protein associated with entry into host cells and a target for immune response. Orf8 is also predicted to encode a rapidly evolving protein, also as a result of immune pressure.[6]

Technical bulletin

SARS-CoV-2 variants and mutations

The majority of these are single base changes, as exemplified by the E484K mutation (Box 1). Notable exceptions include two sites of deletion evident in the B.1.1.7 and B1.1.7+ E484K variants (Table 2); ΔH69/70V (deletions of amino acids 69 and 70 of the spike protein) which results from deletion of bases 21,765-21,770; and ΔY144 (deletion of amino acid 144 from the spike protein) resulting from deletion of bases 21,991-21,993. Although mutations continuously arise, variants such as Brazilian (P1) demonstrate accumulation of events and significant differences from the parent strain, exhibiting viral escape.[4, 7]

Implications of mutations on testing programmes

The most immediate impact of arising mutations in new variants is in the effect on SARS-CoV-2 testing protocols. PCR technologies are one of the most widely-used techniques for SARS-CoV-2 detection. These either monitor amplification of a product in real time (RT-qPCR) and report the number of cycles required to reach a defined quantification parameter (Cq)[8] or complete the PCR and take measurements at the end-point of the reaction (ePCR). [12] In each case, the reaction contains primers specific to each target of viral nucleotide sequence and a fluorescent probe that is located between the primers. During the reaction, the probe is hydrolysed, releasing a fluorescent signal that is proportional to the number of molecules synthesised, until a reaction plateau is reached.

A mutation arising in the target sequence for the primers or probes could have adverse effects on the detection assay. The degree to which an assay is affected depends upon the location and type of mutation. The catastrophic failure of one RT-qPCR assay to detect the S gene due to the delH69V70 mutation in the variant B1.1.7 has been widely described.[9]

Events such as this have resulted in stringent sequence screening to ensure that currently applied assays do not target regions containing mutations.

Target	Amplicon amino acids		Gene
	Start	End	
N1	5	29	N
N2	297	319	N

Table 3.a. CDC N1 and N2 assay locations by N-gene amino acid

	Nucleotide sequence	
	N1	N2
Forward	GACCCCAAAATCAGCGAAAT	TTACAAACATTGGCCGCAAA
Reverse	TCTGGTTACTGCCAGTTGAATCTG	GCGCGACATTCCGAAGAA
Probe	ACCCCGCATTACGTTTGGTGGACC	ACAATTTGCCCCCAGCGCTTCAG

Table 3.b. CDC N1 and N2 assay locations by N-gene amino acid. Oligonucleotide sequences for the N1 and N2 assays. Location of the C29200T mutation in the N2 probe is highlighted in orange.

Technical bulletin

SARS-CoV-2 variants and mutations

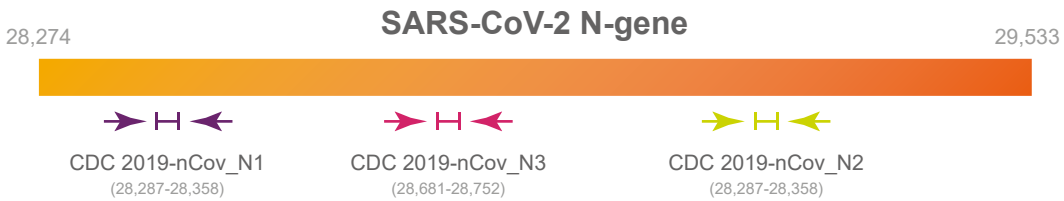

Figure 3. Location of Lu *et al* 2020 (CDC) assays[9] relative to NC_045512.2 (Wuhan)

It is, therefore, critical to have knowledge of the nucleotide regions targeted by any SARS-CoV-2 diagnostic assay. Many of the most widely used assays are listed by WHO and those targeting N1 and N2 are being used worldwide.[10]

Both of the N1 and N2 assays target regions of the N gene (Figure 3; Table 3.a., 3.b.) for which fewer mutations have been described, when compared to S gene.

A comparison between Table 3 and Table 2 shows that none of the mutations described for the four major variants of concern lie within the target regions of the N1 or N2 assays.

However, a mutation within these regions of the N gene has been described: C29200T. The mutation was reported in 0.22% of the SARS-CoV-2 sequences submitted to GISAID (up until the end of January 2021). This mutation site is within the probe sequence for the N2 assay (Figure 4).[11] The site is marked in orange on Table 3.b. and sits at the 3′ end of a run of C bases. Ziegler *et al* (2020) demonstrated that a commercial assay targeting the N gene failed to detect mutant viral material,[10] whereas the CDC 2019_nCOV_N2 probe detected viral RNA containing the mutation with almost no loss of sensitivity. This example illustrates that the nature and position of arising mutations have different impacts on the assay.

As mutations in the SARS-CoV-2 genome continue to regularly arise, it remains critical to screen all SARS-CoV-2 assays against emerging mutations and assess the impact on diagnostic assays. Health agencies such as FDA (USA) and PHE (UK) require assay providers to monitor and report any mutations that may affect assay performance. The practices that LGC, Biosearch Technologies™

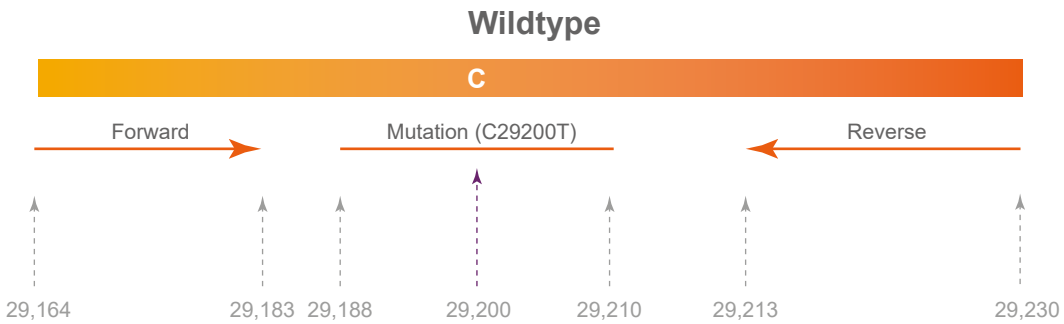

Figure 4. Location of C29200T mutation in N2 probe

Technical bulletin

SARS-CoV-2 variants and mutations

have implemented include fortnightly screens of the N1 and N2 assays against the sequences held in GISAID. These are reviewed at both the international and national levels for potential impact on the prevalence of the mutation and location within the assay. Should the incidence for a variant of concern arise, there is a protocol in place to facilitate rigorous testing to ensure ongoing functionality. Meanwhile, alternative diagnostic assay combinations have been assessed and are being prepared for deployment, should replacements ever be needed.

Resources

microbenotes.com
https://covariants.org/
covidreference.com/variants
https://www.gisaid.org/

FAQ

1. Do the four prevailing variants of concern, UK (B1.1.7), B1.1.7 + E484K, South African (B1.351, 501v2) and Brazilian (B1.1.28, P1), impact detection by the commonly used SARS-CoV-2 detection assays?

Many of the most widely used assays across the world listed by WHO target the N1 and N2 regions and are therefore not impacted by emerging mutations in the S gene.

2. How does Biosearch Technologies screen all emerging SARS-CoV-2 assays against emerging sequence variations to assess impact on assay detection?

Biosearch Technologies regularly screens N1 and N2 assays against the sequences held in GISAID for potential impact assessments. There is a protocol in place for ready deployment of replacement testing assays, if the current protocol fails.

3. What approach does Biosearch Technologies use to standardise efforts to screen for and sequence new, emerging SARS-CoV-2 variants?

Biosearch Technologies is developing a range of SARS-CoV-2 Variant ValuPanels of probes and primers to allow for qualitative detection of emerging variants as well as identifying emerging variants by offering a whole genome sequencing (WGS) based service.

References

1. Lu R, Zhao X, Li J et al., 2020, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet, 395:565-574. DOI: 10.1016/S0140-6736(20)30251-8

2. Chan JF, Kok KH, Zhu Z et al., 2020, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, Emerg Microbes Infect. 9:221-236. DOI: 10.1080/22221751.2020.1719902

3. Zhang Y-Z, 2020, Novel 2019 coronavirus genome. Virological Jan 10 [accessed 2020 Mar 1]. http://virological.org/t/novel-2019-coronavirus-genome/319

4. PHE: Investigation of novel SARS-CoV-2 variants of concern (England) - technical briefing 6, 11 February 2021 [accessed 2020 Mar 1]. https://www.gov.uk/government/publications/phe-investigation-of-novel-sars-cov-2-variants-of-concern-england-technical-briefing-6-11-february-2021

5: Callaway, E. Making Sense of Coronavirus Mutations. Nature. 585(7824):174-177. 2020 Sep 10 [accessed 2021 Apr 1]. https://media.nature.com/original/magazine-assets/d41586-020-02544-6/d41586-020-02544-6.pdf

6. Flower TG, Buffalo CZ, Hooy RM et al., 2021, Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. Proc Natl Acad Sci U S A, 118 (2) e2021785118 DOI: 10.1073/pnas.2021785118

7. Faria NR, Claro IM, Candido D et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Virological 2021 Jan 12 [accessed 2020 Mar 1]. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586

8. Bustin SA, Benes V, Garson JA et al., 2009, The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem Apr;55(4):611-622. DOI: 10.1373/clinchem.2008.112797

9. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01 Technical briefing 4. [accessed 2020 Mar 1]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959359/Variant_of_Concern_VOC_202012_01_Technical_Briefing_4.pdf

10. Lu X, Wang L, Sakthivel SK et al., 2020, US CDC Real-Time Reverse Transcription PCR Panel for Detection of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis 26(8), 1654-1665. DOI: 10.3201/eid2608.201246

11. Ziegler K, Steininger P, Ziegler R et al., 2020, SARS-CoV-2 samples may escape detection because of a single point mutation in the N gene. Euro Surveill. 25(39):pii=2001650. DOI: 10.2807/1560-7917.ES.2020.25.39.2001650

12. Evaluation of endpoint PCR (EPCR) as a central laboratory based diagnostic test technology for SARS-CoV-2 Published 28 January 2021 [accessed 2020 Mar 1]. https://www.gov.uk/government/publications/evaluation-of-endpoint-pcr-epcr-as-a-diagnostic-test-technology-for-sars-cov-2/evaluation-of-endpoint-pcr-epcr-as-a-central-laboratory-based-diagnostic-test-technology-for-sars-cov-2