

SARS-CoV-2 genome surveillance in Mainz, Germany, reveals convergent origin of the N501Y spike mutation in a hospital setting

Lemmermann NA^{1,2#}§, Lieb B^{3#}, Laufs T³, Renzaho A², Runkel S^{1,4}, Kohnen W^{1,5}, Linke M^{1,6}, Gerber S^{1,6}, Schweiger S^{1,6}, Michel A^{1,7}, Bikar S-E^{1,3*}, Plachter B^{1,2*§}, Hankeln T^{1,8*}

¹ SARS-CoV-2 Sequencing Consortium Mainz, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

² Institute for Virology and Research Center for Immunotherapy (FZI), University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

³ StarSEQ GmbH, Mainz, Germany

⁴ Transfusion Unit & Test Center, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

⁵ Department of Hygiene and Infection Prevention, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

⁶ Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

⁷ Medical Management Department, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

⁸ Faculty of Biology, Institute of Organismal and Molecular Evolution, Molecular Genetics & Genome Analysis, Johannes Gutenberg University Mainz, Mainz, Germany

* joint senior authors

equal contribution

§ correspondence to NL: lemmermann@uni-mainz.de

Niels A. Lemmermann
University Medical Center Mainz
Institute for Virology
Obere Zahlbacher Str. 67
55131 Mainz, Germany

Key words: SARS-CoV-2, spike protein, N501Y mutation, full genome sequencing, by convergent evolution

Abstract

While establishing a regional SARS-CoV-2 variant surveillance by genome sequencing, we have identified three infected individuals in a clinical setting (two long-term hospitalized patients and a nurse) that shared the spike N501Y mutation within a genotype background distinct from the current viral variants of concern. We suggest that the adaptive N501Y mutation, known to increase SARS-CoV-2 transmissibility, arose by convergent evolution around December in Mainz, Germany. Hospitalized patients with a compromised immune system may be a potential source of novel viral variants, which calls for monitoring viral evolution by genome sequencing in clinical settings.

Introduction

Coronaviruses are (+)ssRNA viruses with a moderate to high mutation rate compared to other ssRNA viruses and a calculated substitution rate of about 10^{-3} to 10^{-4} per nucleotide and year (Su et al. 2016; Zhao et al. 2004). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the beta-coronavirus that causes the ongoing pandemic with substantial morbidity and mortality worldwide, displays a mutation rate in the same order of magnitude (Coronavirus Study Group, 2020; Wu et al. 2020, Zhou et al. 2020). However, probably due to the very large population size of this virus with phases of exponential growth (Moya et al. 2004; Vasilariou et al. 2020), thousands of changes in the genome have already been described. While most of them will not affect viral properties, natural selection is highly efficient in such large populations, and adaptive mutations that increase viral fitness are likely to occur (Moya et al. 2004). In fact, several SARS-CoV-2 variants have already emerged during the pandemic, showing enhanced host adaptation associated with increased infectivity and transmission. All of these variants show specific mutations in the S gene which encodes the spike (S) protein. The S protein mediates binding of the virus to the cellular angiotensin-converting enzyme 2 (ACE2) receptor and subsequent fusion with the host cell. Therefore, the S protein is the target for vaccines as well as therapeutic antibody development. The first evolutionarily successful mutation of the S protein, D614G, was initially observed in China in March 2020 and rapidly spread worldwide until the end of 2020 (Korber et al. 2020; Yurkovetskiy et al. 2020). In September and December 2020, additional virus variants with conspicuous S mutations have been reported (the latter three have also been designated as *variants of concern*, VOC): The “cluster 5/mink” variant isolated in Denmark is marked by the S gene Δ H69/ Δ V70 deletion in association with the receptor binding domain (RBD) mutation Y453F (ECDC 2020). The B.1.1.7 VOC (*syn.* 20I/501Y.V1) possibly emerged in the UK and is now rapidly spreading across the country and into continental Europe (Rambaut et al. 2020). It carries multiple changes in addition to Δ H69/ Δ V70, among which the RBD mutations N501Y and P681H are the most notable ones. In the VOC B.1.351 (*syn.* 20H/501Y.V2), spreading recently in South Africa, the N501Y mutation is combined with the replacements K417N and E484K, all three affecting the RBD (Tegally et al. 2020). Another rapidly spreading mutant virus lineage possibly originating in Brazil, B.1.1.28 (P1; *syn.* 20J/501Y.V3), also carries these three changes, but with the variation K417T (Vasques Nonaka et al. 2020).

Structural analyses of the RBD provided evidence for the importance of N501 and K417 in the interaction of the S protein with ACE2 (Lan et al. 2020, Shang et al. 2020). N501 is part of the receptor binding motif (RBM), which includes most of the direct interacting residues, and forms a hydrogen bond to the Y41 of ACE2 that is further stabilized by the E484 residue. K417 is a unique ACE2-interacting residue outside of the RBM. Amino acid changes N501Y and E484K enhance the binding affinity of the S protein to ACE2 which might be correlated to the epidemiologically observed higher infectivity and transmission (Zahradnik et al. 2021). Furthermore, both exchanges are discussed as a potential risk for immune escape (Nelson et al. 2021, Wibmer et al. 2021).

Monitoring the mutational spectrum that emerges within the S gene of SARS-CoV-2 is thus most critical for the implementation of interventional methods and combatting the pandemic. The origin and evolution of specific mutations also deserve attention. While variant D614G clearly originated once and then spread worldwide, the divergent geographic distribution of the three new virus lineages which are jointly characterized by the N501Y mutation suggests an independent origin, implying convergent evolution. Will N501Y mutations thus independently arise more often in the future, and under which circumstances? The mutation harbors the potential for virus escape, especially in immunocompromised patients presenting with prolonged viral replication. Also, therapeutic intervention with neutralizing antibodies could be a scenario that enhances S protein-mediated evasion and the outgrowth of escape variants (Williams & Burgers, 2021).

Initiating coordinated efforts for regional SARS-CoV-2 variant surveillance by genome sequencing, we detected a micro-cluster of three individuals in a hospital setting, where N501Y was reproducibly found, but without any other typical mutations of the known variants of concern. The scenario strongly suggests the possibility of a convergently recurring emergence of critical SARS-CoV-2 spike mutations.

Materials and Methods

Sample acquisition, RNA isolation and PCR testing

Nasopharyngeal swabs were taken routinely for SARS-CoV-2 diagnostics by health care professionals, eluted in Virocult transport medium (MWE, Corsham, GB) and analyzed by RT-qPCR for SARS-CoV-2 RNA within 18h. RNA isolation and RT-qPCR were performed for the N- and Nsp-2 gene on a NeuMoDx 288 Molecular System (Molecular Systems, Ann Arbor, USA). A full list of the studied samples including sample type and origin, PCR Ct values and genome sequencing parameters is provided in Tab. 1.

PCR validation for N501Y

The presence of the N501Y mutation was identified in the isolates by melting curve analysis, using the RTVirSNiP SARS-CoV-2 Spike N501Y Kit (TibMolbiol, No. 53-0780-96, Berlin, Germany) according to the supplier's instructions. Briefly, viral RNA was isolated from stored nasopharyngeal swabs using the EMAG system (Biomerieux, Nürtingen, Germany). Purified viral RNAs were subsequently deployed to identify the nucleotide substitution A23063T (corresponding to N501Y) in the S gene. To this end, a 130 Bp fragment was amplified and analyzed by melting curve assay on a 7500 Real-Time PCR System (Applied Biosystems, Darmstadt, Germany).

Whole-genome sequencing and bioinformatic processing

Sars-CoV-2 genome sequences were inferred using a targeted amplicon-based strategy. First, purified viral RNA was converted to cDNA by random hexamer priming using the Lunascript RT kit (New England Biolabs, Frankfurt, Germany). Typically, 5-10 ul of the purified RNA extract were used for reverse-transcription. Additionally, a negative control reaction was set up, containing water instead of RNA extract. We then applied the EasySeq RC-PCR SARS CoV-2 Whole Genome Sequencing Kit (NimaGen, Nijmegen, NL) for amplifying the whole viral genome by tiled PCR primer sets, which are based on the ARTIC standard protocol (<https://artic.network/ncov-2019>). Libraries were inspected on a QIAxcel capillary electrophoresis (Qiagen, Hilden, Germany), normalized to about 9 pM and run on a MiSeq sequencer (Illumina, Berlin, Germany; 2x250 bp v2 chemistry; 2x150 bp for samples #19, #20, #21, #22, #24, #25). Between 500 k and 1.5 mio reads were received per sample.

For bench-marking purposes, we applied two bioinformatic analysis pipelines in parallel, the virSEAK pipeline (JSI medical systems, Ettenheim, Germany) and the Geneious prime platform (Biomatters, Auckland, NZ). Sequence reads were processed by quality trimming, removal of Illumina adapters, and amplicon-specific PCR primers. Thereafter, amplicon

reads were mapped to the SARS-CoV-2 reference sequence Wuhan-Hu-1 (NC_045512), followed by calling of the sequence variants. Read depth typically ranged between 3000 and 6000 per nucleotide position. Both pipelines essentially produced the identical set of single nucleotide variants (SNVs) for the studied genomes. In addition to automatic calling, SNVs were inspected visually by the SeqNext (JSI medical systems) and Geneious prime tools, respectively. Variants were typically called using a coverage threshold of 20 reads and a frequency of at least 90%. Thereafter, variants were integrated into the reference sequence to produce consensus sequence files in FASTA format for each individual genome. FASTA genome consensus sequences were submitted electronically to the GISAID repository (www.gisaid.org) and to the Robert Koch Institute, Germany (www.rki.de).

Initial lineage assignment for all genomes was performed by applying the NextClade v0.12.0 online tool (<https://clades.nextstrain.org>), virSEAK (JSI medical systems) and Pangolin (<https://pangolin.cog-uk.io/>). For phylogenetic tree reconstruction, individual FASTA genome files were combined into a single multi-sequence FASTA file and aligned by MAFFT v7.450 (Katoh and Standley 2013, Katoh et al. 2002) using default criteria. Maximum likelihood trees were calculated using RAxML (Stamatakis 2014) under the GTR substitution model and gamma distribution with 1000 bootstraps. Trees were visualized in Geneious prime, showing bootstrap support values at the respective nodes.

Mapping and visualization of individual amino acid replacements in the SARS-CoV-2 spike protein (data not shown) were produced using the CoVsurver online tool (<https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/>).

Results

While establishing a regional SARS-CoV-2 surveillance routine for the State of Rhineland-Palatinate, the South-West of Germany and the city of Mainz, we sequenced the genomes of 25 samples from 20 individuals, obtained from clinical settings at the University Medical Center Mainz during December 2020 and early January 2021 (Tab. 1). The samples were tested positive for SARS-CoV-2 RNA by routine RT-PCR for the N and NSP2 genes. PCR-positive samples with Ct values ranging from 10 to 27 yielded complete genome sequences with a mean read depth (coverage) ranging between 2,000 and 13,700. Between 16 and 29 variant positions were identified per genome in comparison to the reference sequence (Fig. 1, Tab. 1). According to the Nextstrain classification (<https://nextstrain.org/blog/2021-01-06-updated-SARS-CoV-2-clade-naming>), samples could be assigned to different viral clades, i. e. to the globally distributed clades 20B (n=7) and 20A.EU2 (n=2), to clade 20E.EU1 (n=14), probably of European origin, or to clade 20A (n=1) (Tab. 1; Fig. 2). All genomes shared the spike protein mutation D614G, which has spread worldwide since springtime 2020. To infer relationships and possible transmission routes, the Mainz genome sequences were analyzed by phylogenetic tree reconstruction (Fig. 3). Results indicated one major cluster (bootstrap support 100%) containing 14 genomes, which in the clinical setting derived from the same ward. Within this cluster, some sequences were even almost identical, arguing in favour of potential direct transmission events. On a separate branch (bootstrap support 100%), we identified a micro-cluster of replicate samples from two patients who displayed an identical SARS-CoV-2 genomic sequence (samples #4, #23). These two individuals had a negative PCR test result upon hospital admission, were hospitalized for ca. 2 weeks and shared the same room, but resided on a different ward than the other patients in the cohort. Their sequences were characterized by a long branch in the phylogenetic tree, indicative of fast sequence evolution (Fig. 3).

The subsequent detailed inspection of the identified genomic sequence variants from the full study cohort showed that neither sample belonged to the three virus lineages currently creating the most concern internationally, i. e. VOC B.1.1.7, B.1.351 and B.1.1.248 (P.1). Interestingly, however, in the samples collected from the two room-neighbor patients in December 2020 (#4, #23) we could unambiguously identify the spike receptor binding

domain (RBD) mutation N501Y in combination with another change in the S protein, A1070S. The latter is relatively rare with a frequency of 0.034 in the GISAID genome database (www.gisaid.org) and resides in the spike S2 subunit which is involved in viral membrane fusion and cell entry. None of the other spike changes characteristic for VOC lineages B.1.1.7, B.1.351, and B.1.1.248 could be detected. This result strongly suggests that the N501Y change present in samples #4 and #23 has evolved independently by convergence in a clade 20B genotype background in December in Mainz, Germany. Both individuals did not have a travel history.

We confirmed the genotypes of both, #4 and #23, by sequencing additional RNA samples taken from the very same patients at earlier or later dates in December. We did neither observe fixed genetic differences between the samples, nor sequence heterogeneity within samples, possibly indicative of ongoing viral evolution. In addition, the N501Y mutation was confirmed by RT-PCR and subsequent melting curve analysis (Fig. 4). Looking for additional contact individuals of the two patients, material was sequenced from a nurse working on the same ward (#25), who had tested PCR-positive around the same time. Due to a high PCR cycle threshold (Ct of 33) and thus very little RNA material, only a few thousand reads were obtained for this person during sequencing and no full-length genome sequence could be reconstructed (Tab. 1). However, positions N501Y and A1070S could also be confirmed in this individual.

Discussion

In early 2021, mutant viral lineages designated B.1.1.7, B.1.351 and B.1.1.248 (P.1), each having a unique combination of mutations in their spike (S) gene, are creating much concern due to their presumed higher transmissibility and potential for immune system escape, illustrating the key role of variant surveillance by genome sequencing (Cyranoski 2021, Mahase 2021). A mutation which is common to the S genes of these three virus variants is resulting in the N501Y amino acid change within the RBD of the S protein. N501Y substantially increases the binding affinity of S to its cellular receptor ACE2, thus ameliorating virus entry (Starr et al. 2020; Zahradnik et al. 2021). In the viral VOCs mentioned above, N501Y is accompanied by additional lineage-specific mutations that may render the viral phenotype even more problematic (Zahradnik et al. 2021).

The divergent geographic origin of those VOCs makes it likely that they emerged in separate events, possibly in late 2020, suggesting that the mutual N501Y change evolved in parallel by convergence. Here we report the detection of the same N501Y mutation in a small cluster of individuals, who were tested positive for SARS-CoV-2 in a clinical setting in Mainz, Germany. We found N501Y in a genomic background distinct from the known VOCs, thus occurring out of the context of their additional S mutations. This lends further evidence to the hypothesis of convergent evolution of the N501Y change. In fact, a genome surveillance group based in Ohio, USA, very recently also reported such local emergence of N501Y from a still different viral clade, 20C (Tu et al. 2021). This scenario suggests that the N501Y mutation is an initial key event in driving the evolutionary success of SARS-CoV-2, and its emergence should be monitored world-wide with attention. Preliminary data now indicate that such an initial N501Y mutation can – probably anytime and anywhere - be followed by a second, phenotype-aggravating event: in mid-January 2021, the E484K spike mutation, previously reported to enhance infectivity specifically in the “South-African” B1.351 and the “Brazilian” B.1.1.248 (P1) VOCs (Tegally et al. 2020; Faria et al. 2021), was first-time detected as a novel feature in several samples of the “British” VOC B.1.17 and thus probably also emerged within this lineage by convergence (PHE 2021).

But under which circumstances do these spike gene mutations arise in humans, and which conditions favor their adaptive selection? In our particular hospital micro-cluster with the two bed neighbour patients and one infected nurse, the definite origin of the N501Y spike mutation cannot be inferred unambiguously. Further cases carrying this exact genotype have

not been identified so far, and the SARS-CoV-2 cases in the neighboring ward had different roots (comp. Fig. 1). However, we wish to point out that the two Mainz patients both were immunocompromised and hospitalized for ca. 2 weeks. In fact, Kemp et al. (2020) and Chio et al. (2020) recently reported individual cases of immunocompromised patients in whom SARS-CoV-2 variants evolved rapidly during long-term infection, showing elevated nucleotide substitution rates and the emergence of conspicuous S amino acid changes (including N501Y in one case).

In consequence, we suggest to focus some attention during the future SARS-CoV-2 surveillance by genome sequencing on clinical settings which might favor viral immune escape and the selection of mutations of clinical relevance. While PCR testing (e. g. specifically for the N501Y or E484K mutations) can readily be applied for rapid pre-screening, only whole-genome sequencing will uncover the full repertoire of changes that determine the evolution and clinical phenotype of this virus.

Declaration of interests:

BL, TL and SEB are employees of StarSEQ GmbH, Mainz.

References

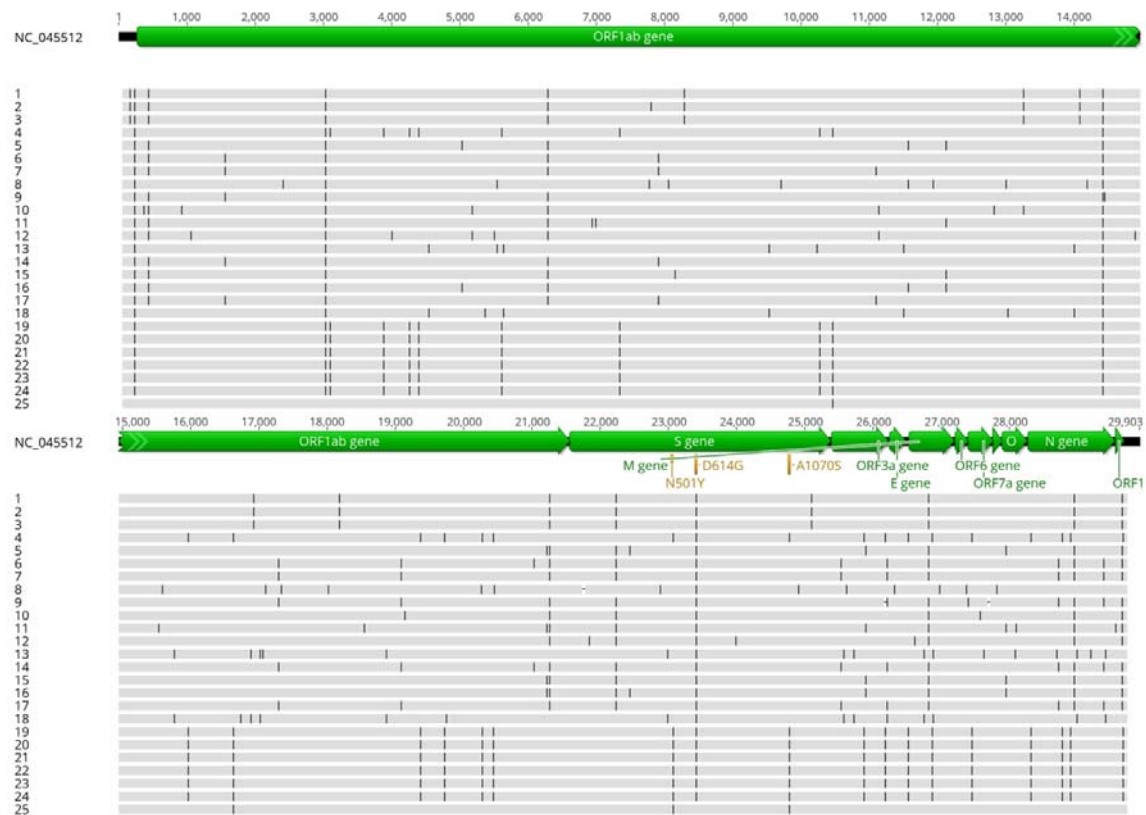
1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 2020 5:536-544. doi:10.1038/s41564-020-0695-z
2. Cyranoski D. Alarming COVID variants show vital role of genomic surveillance. *Nature.* 2021 589:337-338. doi:10.1038/d41586-021-00065-4
3. European Centre for Disease Prevention and Control. Detection of new SARS-CoV-2 variants related to mink - 12 November 2020. ECDC: Stockholm; 2020.
4. Faria NR, Morales Claro I, Candido D, Moyses Franco LA, Andrade PS, Coletti, TM, et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus. 2021 <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>
5. Katoh K, et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002 30:3059-66. doi:10.1093/nar/gkf436
6. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013 30:772-80. doi:10.1093/molbev/mst010
7. Kemp SA, Collier DA, Datir R, Ferreira IATM, Gayed S, Jahun A, et al Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation. medRxiv 2020 2020.12.05.20241927; doi:10.1101/2020.12.05.20241927
8. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G Increases infectivity of the COVID-19 Virus. *Cell.* 2020 182:812-827.e19. doi:10.1016/j.cell.2020.06.043
9. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020 581:215-220. doi:10.1038/s41586-020-2180-5
10. Mahase E. Covid-19: What new variants are emerging and how are they being investigated? *BMJ.* 2021 372:n158. doi:10.1136/bmj.n158
11. Moya A, Holmes EC, González-Candelas F. The population genetics and evolutionary epidemiology of RNA viruses. *Nat Rev Microbiol.* 2004 2:279-88. doi:10.1038/nrmicro863
12. Nelson G, Buzko O, Spilman P, Niazi K, Rabizadeh S, Soon-Shiong P. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. bioRxiv 2021.01.13.426558; doi:10.1101/2021.01.13.426558
13. Public Health England. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01. Technical briefing 5. 2021 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957504/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5_England.pdf
14. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. 2020 <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>
15. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature.* 2020 581:221-224. doi:10.1038/s41586-020-2179-y
16. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014 30:1312-3. doi:10.1093/bioinformatics/btu033.

17. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KH, Dingens AS, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ace2 binding. *Cell*. 2020 182:1295-1310.e20. doi:10.1016/j.cell.2020.08.012
18. Su S, Wong G, Shi W, Liu J, Lai AC, Zhou J, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol*. 2016 24:490-502. doi:10.1016/j.tim.2016.03.003
19. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. 2020 medRxiv 2020.12.21.20248640 doi:10.1101/2020.12.21.20248640
20. Tu H, Avenarius MR, Kubatko L, Hunt M, Pan X, Ru P, et al. Distinct patterns of emergence of SARS-CoV-2 spike variants including N501Y in clinical samples in Columbus Ohio. bioRxiv 2021 2021.01.12.426407; doi:10.1101/2021.01.12.426407
21. Vasilariou M, Alachiotis N, Garefalaki J, Beloukas A, Pavlidis O Population genomics insight into the recent evolution of SARS-CoV-2. bioRxiv 2020 2020.04.21.054122; doi:10.1101/2020.04.21.054122
22. Vasques Nonaka CK, Miranda Franco M, Gräf T, Almeida Mendes AV, Santana de Aguiar R, Giovanetti M, et al. Genomic evidence of a SARS-CoV-2 reinfection case with E484K spike mutation in Brazil. Preprints. 2021, 2021010132
23. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Lambson BE, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. bioRxiv 2021.01.18.427166; doi:10.1101/2021.01.18.427166
24. Williams TC, Burgers WA. SARS-CoV-2 evolution and vaccines: cause for concern? *Lancet Respir Med*. 2021 S2213-2600(21)00075-8. doi:10.1016/S2213-2600(21)00075-8
25. Wu F, Zhao S, Yu B, Chen YM, Wang W, 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. Erratum in: *Nature*. 2020 580(7803):E7
26. Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell*. 2020 183:739-751.e8. doi:10.1016/j.cell.2020.09.032
27. Zahradnik J, Marciano S, Shemesh M, Zoler E, Chiaravalli J, Meyer B, et al. SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. bioRxiv 2021.01.06.425392; doi:10.1101/2021.01.06.425392.
28. Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang YP, et al. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol Biol*. 2004 4:21. doi:10.1186/1471-2148-4-21
29. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020 579:270-273. doi:10.1038/s41586-020-2012-7

# sample	comment	type	Ct value	no of mapped reads	coverage (mean)	% covered sequence (with 1 or more reads)	no of variant positions to reference sequence NC_0245512.2	Spike AA changes	Nextstrain lineage	Pangolin lineage	GISAI clade
1		swab	10	1,423,455	10,86	99.2	18	A222V, D614G, V1176F	20E (EU1)	B.1.117	GV
2		swab	14	1,101,218	8,872	99.8	19	A222V, D614G, V1176F	20E (EU1)	B.1.117	GV
3		swab	12	1,790,033	13,706	99.8	18	A222V, D614G, V1176F	20E (EU1)	B.1.117	GV
4	hospitalized patient	swab	16	1,465,572	11,319	99.8	29	N501Y, D614G, A1070S	20B	B.1.1.70	GR
5		swab	17	942,174	7,283	99.8	18	A222V, D614G	20E (EU1)	B.1.117	GV
6		swab	16	1,311,378	9,961	99.8	20	A222V, D614G	20E (EU1)	B.1.117	GV
7		swab	18	906,38	6,987	99.8	20	A222V, D614G	20E (EU1)	B.1.117	GV
8		swab	18	1,290,366	10,067	99.8	26	HV69-70del, N439K, D614G	20A	B.1.258	G
9		swab	25	589,841	4,786	99.8	21	A222V, D614G	20E (EU1)	B.1.117	GV
10		swab	22	487,671	3,846	99.8	19	A222V, D614G	20E (EU1)	B.1.117	GV
11		swab	25	514,001	4,093	99.8	21	A222V, D614G	20E (EU1)	B.1.117	GV
12		swab	17	1,385,310	10,617	99.8	20	T95I, A222V, D614G, P809S	20E (EU1)	B.1.117	GV
13		swab	18	1,276,485	9,884	99.8	27	S477N, D614G	20A.EU2	B.1.1.60	GH
14		swab	21	307,779	2,353	99.8	20	A222V, D614G	20E (EU1)	B.1.117	GV
15		swab	23	740,518	5,952	99.8	16	A222V, D614G	20E (EU1)	B.1.117	GV
16		swab	20	658,459	5,198	99.8	18	A222V, D614G	20E (EU1)	B.1.117	GV
17		swab	18	1,172,488	8,991	99.8	20	A222V, D614G	20E (EU1)	B.1.117	GV
18		swab	23	537,788	4,315	99.8	26	S477N, D614G	20A.EU2	B.1.1.60	GH
19	replicate sample from #4	swab	27	476,65	2,034	97.9	29	N501Y , D614G, A1070S	20B	B.1.1.70	GR
20	replicate sample from #4	sputum	13	766,929	3,202	99.8	29	N501Y , D614G, A1070S	20B	B.1.1.70	GR
21	replicate sample from #4	swab	n.a.	1,352,960	5,768	99.8	29	N501Y , D614G, A1070S	20B	B.1.1.70	GR
22	replicate sample from #4	swab	21	509,597	2,163	98.0	29	N501Y , D614G, A1070S	20B	B.1.1.70	GR
23	hospitalized patient, #4 contact	swab	22	937,811	3,973	99.8	29	N501Y , D614G, A1070S	20B	B.1.1.70	GR
24	replicate sample from #23	swab	14	1,404,371	5,927	99.8	29	N501Y , D614G, A1070S	20B	B.1.1.70	GR
25	contact person of #4 and #23	swab	33	3,889	15	76.7	16	N501Y , A1070S	(19A)	B	(L)

Tab.1 Study sample overview

A.



B.

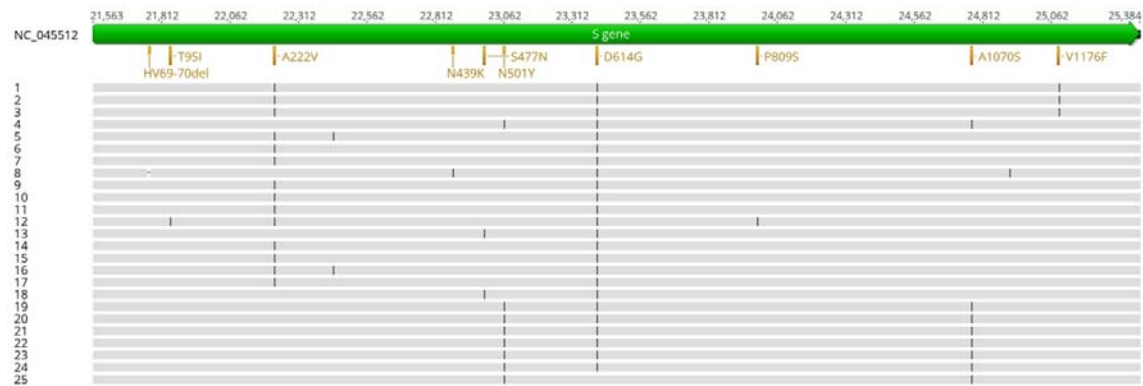


Fig. 1 Position of mutations observed in the sequenced SARS-CoV-2 samples, shown for the full genomes (A) and enlarged for the S gene (B). Sample names given at the right side correspond to Tab. 1.

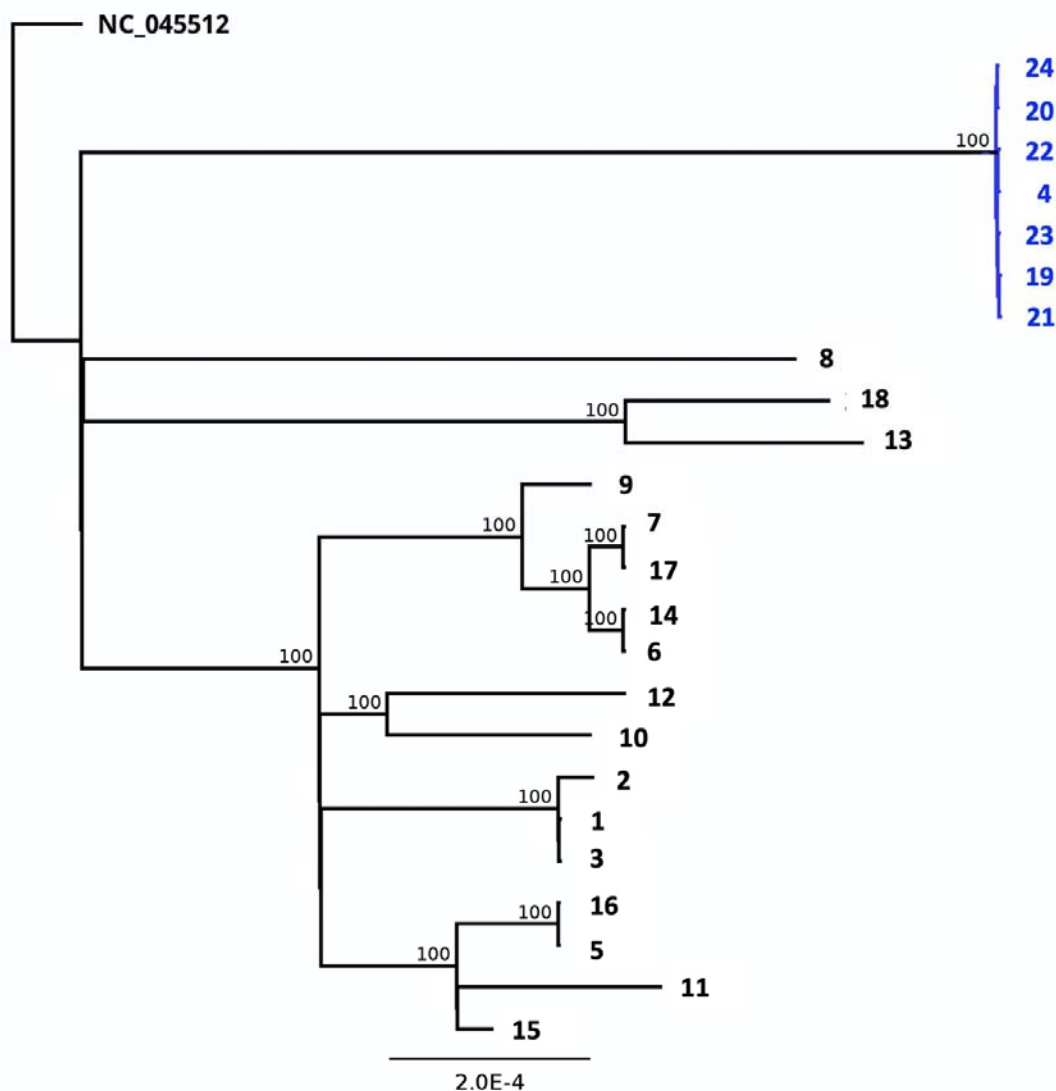


Fig. 2 Maximum likelihood phylogenetic tree based on full genome sequences reveals the relationships of the study samples. Bootstrap values are indicated at the respective nodes. The cluster of sequences labeled in blue originates from the samples of the two patients and contains the N501Y spike mutation. Sample names at the branch tips correspond to Tab. 1.

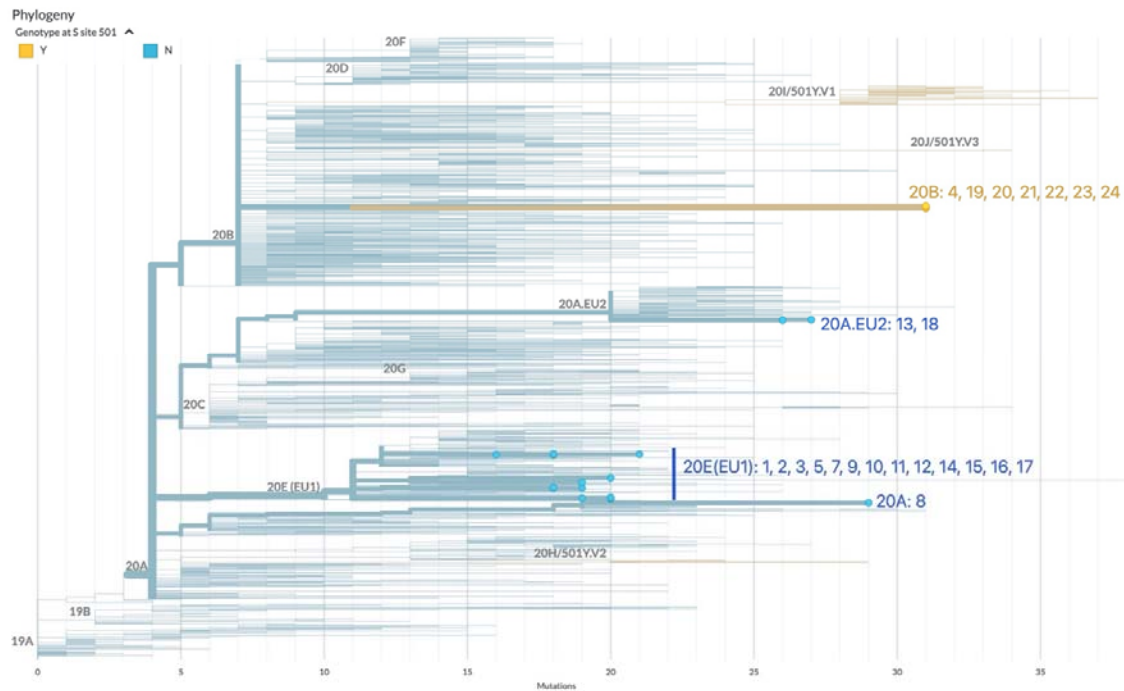


Fig. 3 Clade classification of the study sample according to “Nextstrain”. Major clades are indicated along the tree branches. Study samples are depicted with coloured tips (blue, yellow), their designations correspond to Tab. 1. The position of the currently known VOCs is also indicated.

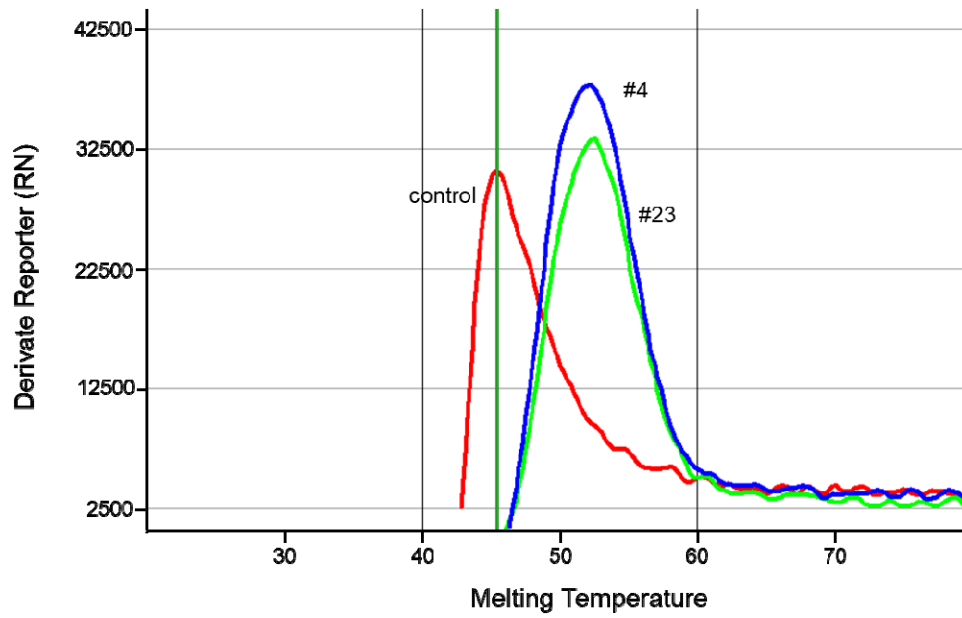


Fig. 4 Confirmation of the N501Y mutation in samples #4 and #23 by PCR and subsequent melting curve analysis.