medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Failure of the cobas® SARS-CoV-2 (Roche) E-gene assay is associated with a C-to-T transition at position 26340 of the SARS-CoV-2 genome

Maria Artesi[1,#], Sébastien Bontems[2,#], Paul Göbbels[3], Marc Franckh[3], Raphaël Boreux[2], Cécile Meex[2], Pierrette Melin[2], Marie-Pierre Hayette[2,§], Vincent Bours[1,4,§], Keith Durkin[1,§]

[1]Laboratory of Human Genetics, GIGA Research Institute, Liège, Belgium,
[2]Department of Clinical Microbiology, University Hospital of Liège, Liège, Belgium
[3]St-Nikolaus Hospital, Eupen, Belgium
[4]Department of Human Genetics, University Hospital of Liège, Liège, Belgium

[#]These authors contributed equally
[§]These authors jointly supervised the work

Correspondence: kdurkin@uliege.be; vbours@uliege.be; mphayette@chuliege.be

Abstract

Control of the ongoing severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic requires accurate laboratory testing to identify infected individuals, while also clearing essential staff to continue work. At the current time a number of RT-PCR tests have been developed to identify SARS-CoV-2, targeting multiple regions in the viral genome. In comparison to other RNA viruses the mutation rate of SARS-CoV-2 is moderate, however given the large number of transmission chains it is prudent to monitor circulating viruses for mutations that might compromise these tests. Here we report the identification of a C-to-T transition at position 26340 of the SARS-CoV-2 genome which is associated with failure of the cobas® SARS-CoV-2 E-gene assay. This variant was detected in four health care workers from the same team. Whole genome sequencing of SARS-CoV-2 showed all four to carry genetically identical viruses. Examination of viral genomes deposited on GISAID showed this mutation has arisen independently on three occasions. This work highlights the necessity of monitoring SARS-CoV-2 for the emergence of SNPs which might adversely affect the RT-PCRs used in diagnostics. Additionally, it argues that two regions in the SARS-CoV-2 should be targeted in RT-PCRs to avoid false negatives.

Introduction

Coronavirus disease of 2019 (COVID-19) originated in Wuhan, China in late 2019, [1,2] and has generated a global pandemic [3]. As of the 18th of April 2020, over 2 million confirmed cases and more than 148,000 deaths have been reported worldwide [4]. Metagenomic RNA sequencing revealed that COVID-19 is caused by a novel coronavirus, severe acute respiratory syndrome coronavirus (SARS-CoV-2). SARS-CoV-2 is a close relative of SARS-CoV and MERS-CoV [2], coronaviruses which have both been responsible for large outbreaks of

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

respiratory illness within the last two decades [5,6]. The release of the first SARS-CoV-2 genome sequence on the 10th of January spurred the development of RT-PCR assays [7–9] and thereby enabled reliable laboratory diagnosis of infections. In addition, protocols were also developed to allow for rapid sequencing of the SARS-CoV-2 genome [10], sharing of the resultant data [11] and phylogenetic analysis [12].

Laboratory testing for SARS-CoV-2 is a cornerstone of the strategy to mitigate the spread of SARS-CoV-2 [13] as it facilitates the identification and isolation of infected individuals, while negative tests can allow essential personnel to continue work. In the context of SARS-CoV-2, with its high transmissibility[14], false negatives could have particularly adverse effects on efforts to control its spread. As RT-PCR oligos rely on binding to small ~20bp sequences, mutations in these regions have the potential to generate false negative results through impaired amplification or probe binding. In contrast to other RNA viruses, coronaviruses have a moderate mutation rate due their ability to carry out RNA proofreading [15]. Nevertheless, given the large number of ongoing transmission chains, it is prudent to monitor the integrity of RT-PCR assays.

Here we report the identification of a SNP in the E-gene of SARS-CoV-2 that is associated with failure of the E-gene assay used in the cobas® SARS-CoV-2 test (Roche). This observation highlights the necessity of targeting two regions in SARS-CoV-2 RT-PCR assays and shows the role sequencing can play in resolving and anticipating problems with the RT-PCR assays in use.

Methods

RNA extraction and Real Time PCR

The study was approved by the Comité d'Ethique Hospitalo-Facultaire Universitaire de Liège (Reference number: CE 2020/137). The COVID-19 detection was routinely performed using the cobas® 6800 platform (Roche). For this, 400 µL of nasopharyngeal swabs in preservative medium (AMIES or UTM) were first incubated at room temperature for 30 minutes with 400 µL of cobas® PCR Media kit (Roche) for viral inactivation. Samples were then loaded on the cobas® 6800 platform using the cobas® SARS-CoV-2 assay for the detection of ORF1ab and E genes.

For RT-PCR control and sequencing analysis, RNA was extracted from clinical samples (300µL) on a Maxwell 48 device using the Maxwell RSC Viral RNA kit (Promega) following a viral inactivation step using Proteinase K according to manufacturer's instructions. RNA elution occurred in 50µL RNAse free water and 5 µL were used for the RT-PCR. Reverse transcription and RT-PCR were performed on a LC480 thermocycler (Roche) based on Corman et al. [9] protocol for the detection of RdRP and E genes using the Taqman Fast Virus 1-Step Master Mix (Thermo Fisher). Primers and probes (Eurogentec, Belgium) were used as described by the authors [9].

SARS-CoV-2 whole genome sequencing

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Reverse transcription was carried out using SuperScript IV VILO™ Master Mix, 3.3 µL of RNA was combined with 1.2 µL of master mix and 1.5 µL of H_2O. This was incubated at 25°C for 10 min, 50°C for 10 min and 85°C for 5 min. PCR reactions used the primers and conditions recommended in the nCoV-2019 sequencing protocol [16]. Samples were multiplexed using the Oxford Nanopore Native Barcoding Expansion kits 1-12 and 13-24, in conjunction with Ligation Sequencing Kit. Sequencing was carried out on a Minion using R9.4.1 flow cells. Data analysis followed the nCoV-2019 novel coronavirus bioinformatics protocol of the Artic network [16]**.** The resulting consensus viral genomes have been deposited at the Global Initiative on Sharing All Influenza Data (GISAID)[11]

Sanger sequencing

Reverse transcription was carried out as above. The primers nCoV-2019_87_LEFT and nCoV-2019_87_RIGHT from the Artic network nCoV-2019 amplicon set [16] were used to amplify the regions between positions 26198-26590. The resultant PCR product was purified using Ampure XP beads (Beckman Coulter), sequenced using Big Dye terminator cycle-sequencing kit (Applied Biosystems) and run on a ABI PRISM 3730 DNA analyser (Applied Biosystems).

Phylogeny

The Phylogenetic trees shown were produced via the Nextstain website (https://nextstrain.org/ncov/), which utilises the viral genomes deposited on GISAID (https://www.gisaid.org/). By selecting the appropriate SNP in the diversity panel, the viruses carrying that variant are highlighted. The graphs shown were generated on 20th April 2020.

Results

During routine testing using the SARS-CoV-2 test of the cobas® system (Roche), it was noted that a group of four samples were negative for the E-gene assay, but positive for the ORF1ab assay (Table 1). The four samples were retested using the Corman et al. [9] SARS-CoV-2 assay that targets the RdRP and E genes. In this instance both assays were positive (Table 1). All four samples came from Belgian health care workers in the same team, suggesting a common source of infection.

We next carried out whole genome sequencing of the viruses from the four patients using the Artic Network protocol [16]. The consensus genomes generated showed all four to be infected with a genetically identical virus. The virus differed from the MN908947.3 reference isolated in Wuhan at only three positions (Figure1). The first two SNPs were towards the 5' end of the virus at positions 1440 and 2891 respectively. The third SNP, a C-to-T transition at position 26340 is within the E gene of the virus and was validated by Sanger sequencing in the four samples (Supplementary Figure 1). This SNP overlaps with the E gene probe used in the Corman et al. [9] RT-PCR assay, however, as was mentioned above, it does not appear to affect the performance of this assay in our hands. Unfortunately, the position of primers and probes utilised in the cobas® E-gene assay (Roche) are not publicly available, nevertheless it is parsimonious to assume that this SNP is the cause of the failure of the E-gene assay implemented in the cobas® system.

Out of the 186 SARS-CoV-2 genomes we have sequenced, only these four samples carry a SNP at position 26340. We then checked the sequences deposited in GISAID via the

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Nextstain website for a variant at the same position. We found that a SNP had arisen at this position in two additional viral genomes, one isolated in England and the second in Australia (Figure 2). As the English, Australian and Belgian viruses do not cluster together in the tree generated by Nextstain, it appears that this SNP has arisen independently on three different occasions (Figure 2).

Finally we also checked the regions encompassed by the primers and probes reported by Corman et al.[9], the Chinese CDC [7] and US CDC [8] in the 186 SARS-CoV-2 genomes we have sequenced to date at the GIGA Institute (Supplementary Table 1). We found mutations in binding sites for three of the oligos from the US CDC tests and two oligos from Corman et al. [9]. All were present in a small number of the viruses sequenced by us, as well as in a small number of viruses deposited in GISAID. In contrast, in the Nucleoprotein-protein assay from the Chinese CDC, we saw that 50 of our samples contain a cluster of three SNP which change GGG-to-AAC at the 3' end of the binding site of the 2019-nCoV-NFP oligo. This 3 base change is present in >600 viruses on Nextstrain and all cluster together (Figure 2). Vogels et al [17] also identified this 3 base change as well as other SNPs in primer/probe binding sites from a number of RT-PCR assays for SARS-CoV-2.

Discussion

As the positions of the primers and probes used in the cobas® E-gene assay(Roche) have not been disclosed to us upon request, we cannot definitively conclude that the C-to-T transition at position 26340 of the SARS-CoV-2 genome causes the failure in the E-gene assay in our four patients. However, given the available data, causality appears likely. The cobas® E-gene assay may use an alternate primer probe combination that is more sensitive to the presence of the SNP, alternatively it may target the same positions as the Corman et al.[9] E-gene assay, but differences in reagents used and cycling conditions may prevent binding of the probe in the presence of the SNP.

It should be stressed out that despite the failure of the cobas® E-gene assay in these four patients the cobas® ORF1ab assay was positive. This highlights the prudence of targeting more than one position in the viral genome in a RT-PCR assay. The Corman et al.[9] protocol recommends the use of its E-gene assay as a first-line screening tool, with confirmatory testing using the RdRp gene assay [9]. This SNP does not affect the Corman et al E-gene in our hands, however our results highlight how mutations in the virus can generate false negative results. In most cases such mutations will be rare, however as the viral sequencing being carried out across the world and shared on GISAID [11] has shown, such mutations have the potential to arise independently in separate transmission chains.

As regards the other mutations identified affecting primer or probe binding sites in widely used assays, we have no evidence that they affect the respective assays and in the case of the single base pair changes they are unlikely to have a major impact. However, the three base pair changes observed in a large number of viruses for the annealing site of the 2019-nCoV-NFP oligo from the Chinese CDC is likely to have a more dramatic effect on the performance of the assay and warrants further investigation.

The fact that all four individuals we identified with the SNP at position 26340 were health care professionals and worked in the same team, highlights the risks these individuals face. Unfortunately, we could not identify the index patient responsible for the contamination.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

However, our ability to show that each individual carries a genetically identical virus demonstrates the potential whole genome sequencing has for tracking chains of transmission.

This work shows the danger of relying on an assay that only targets a single position in the viral genome. It also highlights the utility of combining testing with rapid sequencing of a subset of the positive samples, especially in cases where one of the assays fails. Finally, it is an example of the need for sharing assay's specifics between manufacturers and users, in order to enable correct data analysis and interpretation.

Acknowledgements

This work was supported by the Région Wallonne project WALGEMED(convention n° 1710180). We would like to acknowledge and thank the laboratories who submitted and shared their sequences to GISAID.

	cobas® SARS-CoV-2 (Roche)		**Corman et al. 2020**	
Patient ID	Ct ORF1ab	Ct E gene	Ct RdRp	Ct E gene
10088	26.65	0	28.66	26.51
10089	22.64	0	25.02	22.85
10095	24.62	0	23.73	22
10096	20.45	0	22.89	20.33

Table 1 Real-time PCR Ct values observed in the four patients for the two respective tests

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

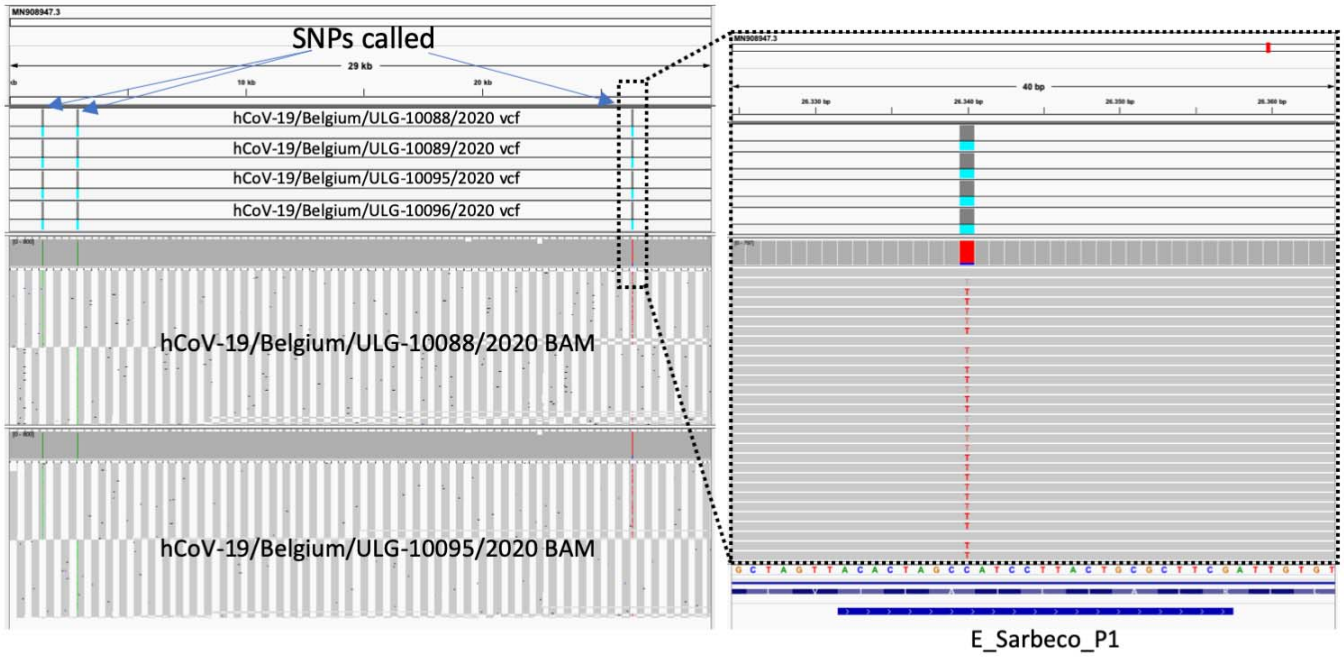
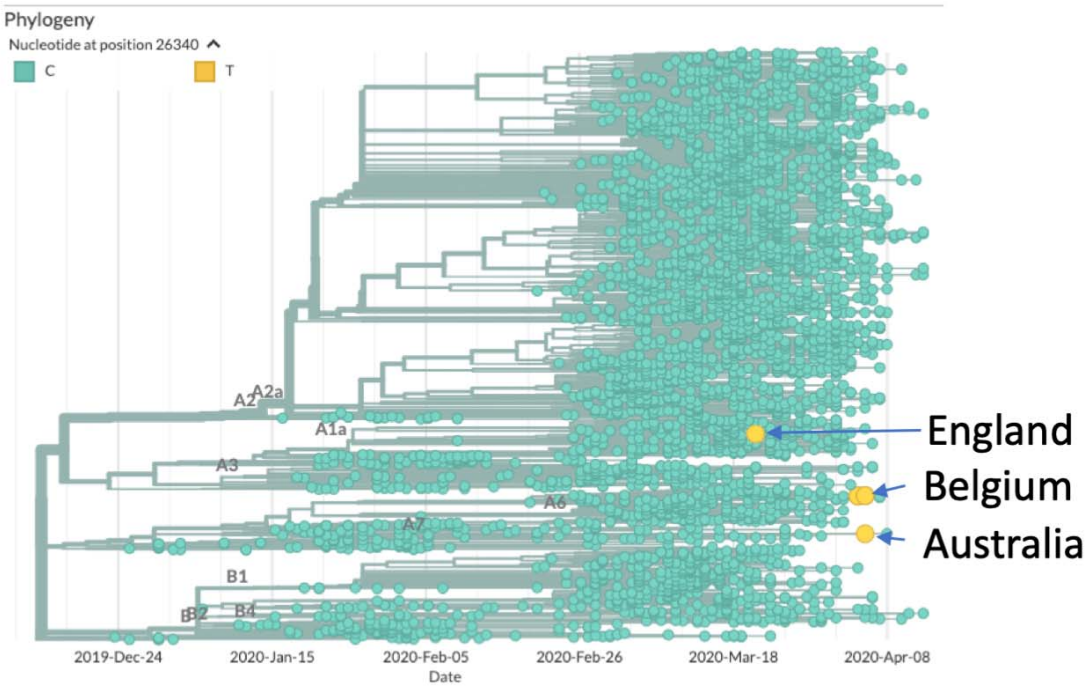


Figure 1 Screen shot form IGV (Integrative Genomics Viewer: http://software.broadinstitute.org/software/igv/home) shows the VCFs for the four viruses with the variants identified as well as the BAM files showing the raw reads and coverage for two viruses. The zoomed section shows the SNP at 26340, the blue rectangle labeled E_Sarbeco_P1 corresponds to the region the E oligo anneals to in Corman et al. [9].

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

A



B

```
MN908947.3                        ATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCC    26340
hCoV-19/Belgium/ULG-10088/2020    ATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCT    26340
hCoV-19/Belgium/ULG-10089/2020    ATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCT    26340
hCoV-19/Belgium/ULG-10095/2020    ATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCT    26340
hCoV-19/Belgium/ULG-10096/2020    ATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCT    26340
hCoV-19/Australia/VIC872/2020     ATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCT    26337
hCoV-19/England/SHEF-C088C/2020   ATAGTTAATAGCGTACTTCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCT    26340
                                  ************************************************************
```

Figure 2 A: Phylogenetic tree generated by Nextstrain [12] with the viruses carrying the T allele at 26340 highlighted in yellow (The four Belgian viruses overlap). **B:** Partial alignment of the four Belgian, one English and one Australian viral genomes, along with the SARS-CoV-2 (MN908947.3) reference genome, the SNP at position 26340 is highlighted.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

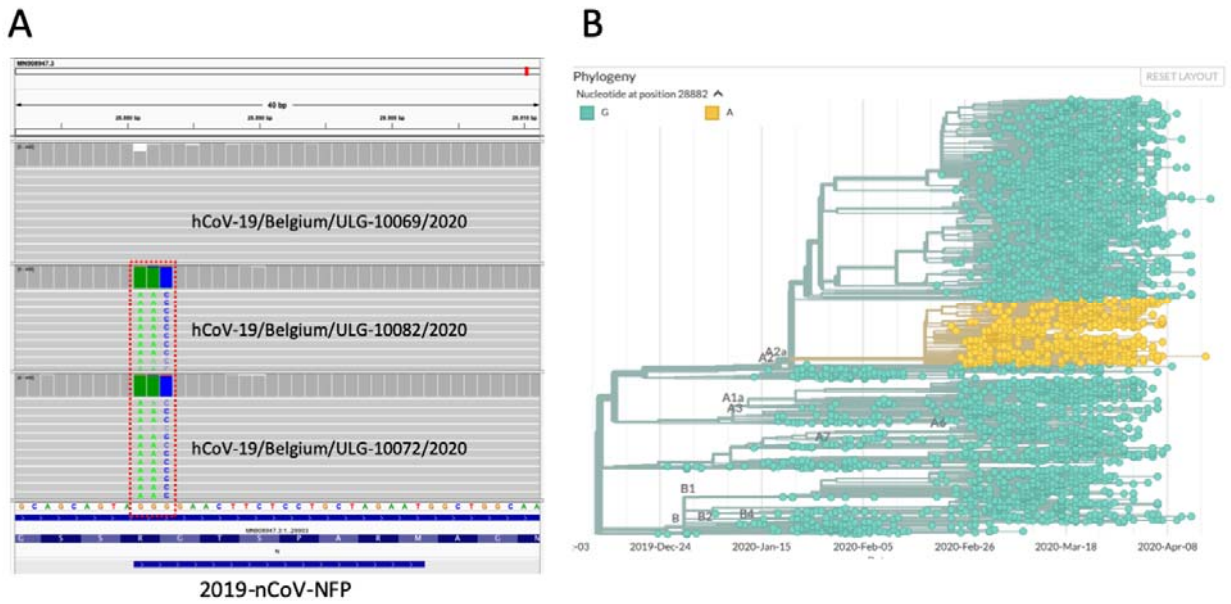


Figure 3 A Screenshot from IGV shows the annealing location of the 2019-nCoV-NFP oligo from the Chinese CDC (blue rectangle), a viral genome with the wildtype sequence and two with the 3bp mutation are shown. This 3 base mutation affects the last three bases in the 3' end of the oligo binding site. **B** Phylogenetic tree generated by nextstrain [12] where the viruses carrying the 3bp mutation are highlighted in yellow.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020;382: 727–733.

2. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579: 265–269.

3. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 51. 11 Mar 2020. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10

4. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 89. 18 Apr 2020. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200418-sitrep-89-covid-19.pdf?sfvrsn=3643dd38_2

5. Drosten C, Günther S, Preiser W, van der Werf S, Brodt H-R, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med. 2003;348: 1967–1976.

6. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med. 2012;367: 1814–1820.

7. Chinese CDC, Specific primers and probes for detection 2019 novel coronavirus. [cited 20

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Apr 2020]. Available: http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html

8. CDC. US CDC Coronavirus Disease 2019 (COVID-19) Real-time RT-PCR Primer and Probe. In: Centers for Disease Control and Prevention [Internet]. 10 Apr 2020 [cited 20 Apr 2020]. Available: https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html

9. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. 2020;25. doi:10.2807/1560-7917.ES.2020.25.3.2000045

10. SARS-CoV-2_Sequencing. Github; Available: https://github.com/CDCgov/SARS-CoV-2_Sequencing

11. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017;22. doi:10.2807/1560-7917.ES.2017.22.13.30494

12. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34: 4121–4123.

13. Centers for Disease Control and Prevention. Implementation of mitigation strategies for communities with local covid-19 transmission. 2020. Available: https://www.cdc.gov/coronavirus/2019-ncov/downloads/community-mitigation-strategy.pdf

14. Steven Sanche, Yen Ting Lin, Chonggang Xu, Ethan Romero-Severson, Nick Hengartner, Ruian Ke. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerging Infectious Disease journal. 2020;26. doi:10.3201/eid2607.200282

15. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol. 2011;8: 270–279.

16. Artic Network. [cited 19 Apr 2020]. Available: https://artic.network/ncov-2019

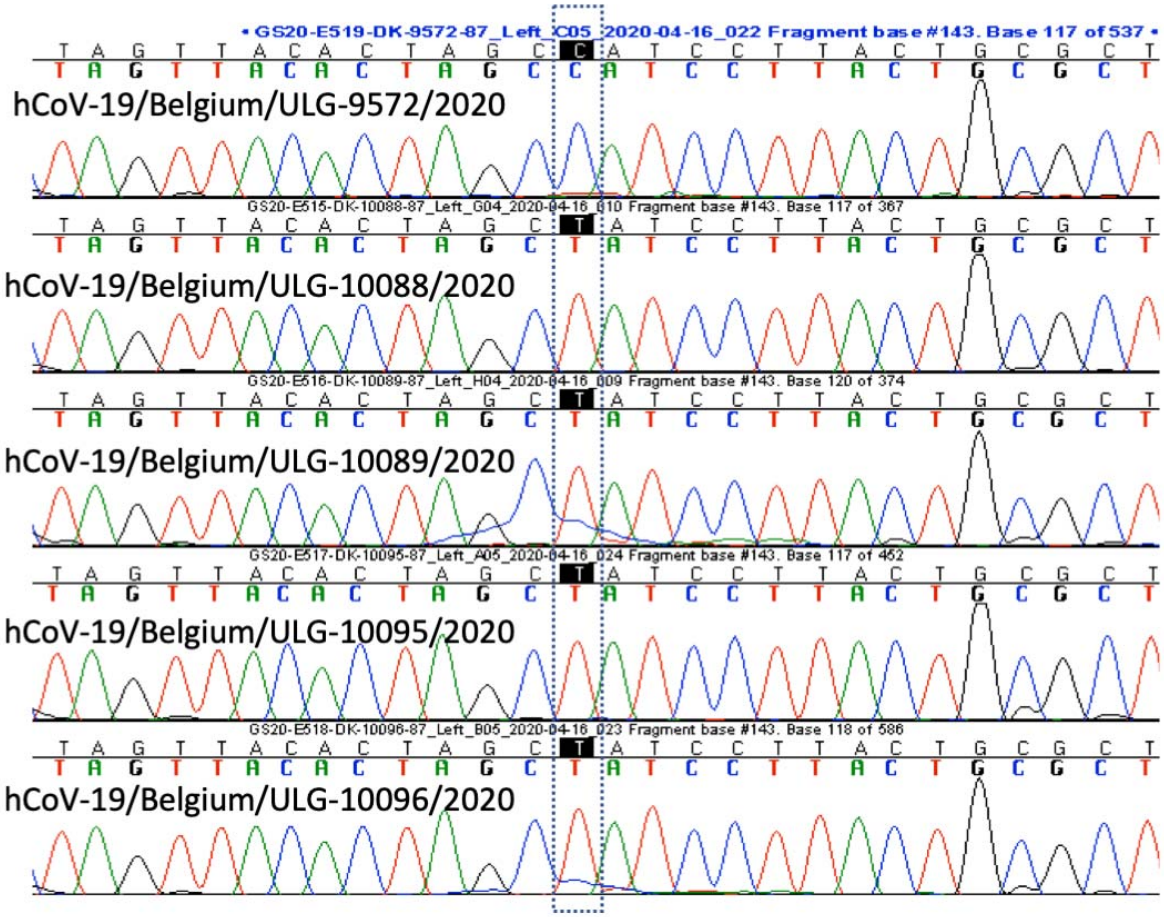
17. Vogels CBF, Brito AF, Wyllie AL, Fauver JR, Ott IM, Kalinich CC, et al. Analytical sensitivity and efficiency comparisons of SARS-COV-2 qRT-PCR assays. medRxiv. 2020; 2020.03.30.20048108.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Oligo Name	Origin	position	# GIGA viruses with variant	# other GISAID viruses with variant	# independent mutations
E_Sarbeco_P1	Corman et al.	26340	4	2	3
RdRp_SARSr-F	Corman et al.	15448	1	0	1
N_Sarbeco_P	Corman et al.	28770	1	0	1
2019-nCoV_N3-F	US CDC	28690	1	1	2
2019-nCoV_N2-F	US CDC	29167	3	1	2
2019-nCoV_N2-F	US CDC	29179	1	1	2
2019-nCoV-NFP	Chinese CDC	28881-28883	50	>600	1

Supplementary Table 1 SNPs in oligo binding sites observed in the 186 SARS-CoV-2 genomes sequenced at the GIGA, Liège. SNPs found in the GIGA data were also checked in the GISAID database via the nextstrain website. Regarding the number of independent mutations estimated by nextstrain, the C-to-T transition at 26340 is found at three different positions in the phylogenetic tree indicating three independent mutations. The 3bp at 28881-2888 is observed in many more viral genomes but as these all cluster together this indicates descent from the same virus and therefore a single mutational event.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.28.20083337; this version posted May 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .



Supplementary Figure 1 Sanger sequencing of the SNP at position at 26340. The top chromatogram shows a virus carrying the wildtype base as well as the four viruses with the C-to-T transition.