Current Genetic Medicine Reports
https://doi.org/10.1007/s40142-020-00197-5

BIOINFORMATICS (A PITTMAN, SECTION EDITOR)

Decoding Covid-19 with the SARS-CoV-2 Genome

Phoebe Ellis[1] · Ferenc Somogyvári[2] · Dezső P. Virok[2] · Michela Noseda[3] · Gary R. McLean[1,3,4]

Accepted: 29 December 2020

Abstract

Purpose of Review SARS-CoV-2, the recently emerged coronavirus (CoV) that is responsible for the current global pandemic Covid-19, first appeared in late 2019 in Wuhan, China. Here, we summarise details of the SARS-CoV-2 genome to assist understanding of the emergence, evolution and diagnosis of this deadly new virus.

Recent Findings Based on high similarities in the genome sequences, the virus is thought to have arisen from SARS-like CoVs in bats but the lack of an intermediate species containing a CoV with even greater similarity has so far eluded discovery. The critical determinant of the SARS-CoV-2 genome is the spike (S) gene encoding the viral structural protein that interacts with the host cell entry receptor ACE2. The S protein is sufficiently adapted to bind human ACE2 much more readily than SARS-CoV, the most closely related human CoV.

Summary Although the SARS-CoV-2 genome is undergoing subtle evolution in humans through mutation that may enhance transmission, there is limited evidence for attenuation that might weaken the virus. It is also still unclear as to the events that led to the virus' emergence from bats. Importantly, current diagnosis requires specific recognition and amplification of the SARS-CoV-2 RNA genome by qPCR, despite these ongoing viral genome changes. Alternative diagnostic procedures relying on immunoassay are becoming more prevalent.

Keywords SARS-CoV-2 · Covid-19 · Pandemic · Genome analysis · Diagnosis · qPCR

Introduction

In December 2019, a pneumonia illness of unknown aetiology surfaced in Wuhan, China, with an animal meat market at the epicentre. The causative agent was quickly discovered and initially termed by the World Health Organization (WHO) as the 2019 novel coronavirus (2019-nCoV). Genetic analysis identified strong similarities between the severe acute respiratory syndrome (SARS) coronavirus (CoV) that was discovered in 2003 and it was renamed SARS-CoV-2 [1, 2]. Infection with SARS-CoV-2 produces a clinical syndrome known as 2019 novel coronavirus disease (Covid-19). The most notable differences in SARS-CoV-2 from its predecessor can be seen in the speed and ease with which it spreads and the disease severity which demonstrates a reduced case fatality rate [3]. However, unlike SARS, just 1 month after the initial discovery, the WHO declared the outbreak of coronavirus disease, or Covid-19, an international public health emergency and, at 11 months into the outbreak, SARS-CoV-2 has spread worldwide and infected > 72 million people, causing > 1.6 million deaths.[1] SARS-CoV-2, like other CoVs, is known to transmit between individuals by direct contact and airborne mechanisms via droplets or aerosols [4]. Measures to control such transmission such as social distancing and personal hygiene alongside a robust testing and contact tracing system remain important to control the surging pandemic in the absence of pharmaceutical interventions.

It is believed that genetic events occurring in animal CoVs led to changes to the structure of the SARS-CoV-2 spike (S)

This article is part of the Topical Collection on *Bioinformatics*

✉ Gary R. McLean
g.mclean@londonmet.ac.uk

[1] School of Human Sciences, London Metropolitan University, London, UK

[2] Department of Medical Microbiology and Immunobiology, University of Szeged, Szeged, Hungary

[3] National Heart and Lung Institute, Imperial College London, London, UK

[4] Cellular and Molecular Immunology Research Centre, London Metropolitan University, London, UK

[1] According to data obtained from the Covid-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)

Published online: 09 January 2021

gene that altered the S protein sequence [5]. These changes have enabled it to have a stronger affinity for the human host cell receptor (angiotensin-converting enzyme 2, ACE2) than the previous SARS-CoV, increasing its ability to enter cells and replicate, ultimately facilitating improved human to human transmission [6]. ACE2 is a carboxypeptidase that converts angiotensin II into angiotensin, with a counterbalancing role in the renin-angiotensin-aldosterone system required for cardiovascular homeostasis [7]. Its tissue distribution dictates the infection course and subsequent pathology of SARS-CoV-2. This review will summarise the knowledge and features of the SARS-CoV-2 genome and how this information has been used to understand the virus' emergence, cell and tissue tropism leading to disease, the diagnosis of infection and evolution of the virus.

Human CoVs Are Animal RNA Viruses Grouped in the Coronaviridae Family

Human CoVs are zoonotic pathogens derived from animal CoVs and mainly cause respiratory diseases [8]. All CoVs have non-segmented, large single-stranded, positive-sense RNA genomes that have a similar organisation of non-coding untranslated regions (UTRs) and coding regions or open reading frames (ORFs). They can be categorised into four separate genera according to their genome and protein sequences: alpha, beta, gamma and delta [9]. Alpha and beta-CoVs affect humans and other mammals whereas the gamma and delta strains affect mainly birds. There are now seven known human CoVs: HCoV-229E (alpha), HCoV-NL63 (alpha), HCoV-OC43 (beta), HCoV-HKU1 (beta), MERS-CoV (beta), SARS-CoV (beta) and the most recently discovered SARS-CoV-2 (beta). All seven human CoVs have zoonotic origins linked to bats, mice or domestic animals [10]. Most human CoV infections are not life-threatening as infections with the 229E, NL63, OC43 and HKU1 variants produce symptoms ranging from a mild common cold to pneumonia, or are even completely asymptomatic. The 2003 SARS-CoV outbreak, beginning in Guangdong, China, however, caused severe lower respiratory tract illness, had a death rate of approximately 9% and lasted around 8 months, lingering cases appearing in several countries worldwide until 2004, with China, Singapore and Canada amongst the worst hit [11]. SARS is now thought of as a near-miss, from the perspective of an emerging infectious disease that causes widespread turmoil. The danger of emerging human CoVs, however, was not fully realised until the subsequent Middle East respiratory syndrome (MERS) outbreak years later, which more than tripled the mortality rate of SARS to approximately 34%. The 2012 MERS outbreak began in Saudi Arabia through human contact with infected camels. Despite a higher death rate, MERS-CoV does not transmit easily between humans and most infections are seen from direct camel to human contact [12]. Since 2012, there have been sporadic outbreaks of MERS primarily in the Middle East and Korea, which have resulted in over 2000 confirmed cases and approximately 700 deaths [13]. Most recently, the newest CoV to infect humans, SARS-CoV-2, is known to transmit easily between humans in an airborne manner then causes a range of upper and lower respiratory tract symptoms similar to the endemic CoVs. In addition, symptoms such as loss of sense of smell and taste, severe cough and fever characterise infection with SARS-CoV-2 [14]. The range of symptoms is variable however and when severe, Covid-19 manifests as a systemic illness characterised by hyperinflammation and cytokine storm affecting multiple organs outside of the respiratory system including the heart, kidney, liver and brain [15]. Disease severity is also variable but can be fatal for individuals with underlying health conditions and so-called long Covid, where following recovery from acute disease, many ongoing effects are prevalent [16]. While the true death rate of SARS-CoV-2 cannot be calculated during an ongoing pandemic, it is estimated to be approximately 1%, with the majority of fatalities occurring in individuals over 60 years old, those who are immune-compromised or those with pre-existing conditions that are worsened by the effects of Covid-19 disease [17].

SARS-CoV-2 Derives from Bat CoVs and Is Closely Related to SARS-CoV

New CoVs are thought to arise from complex recombination events when two related viral genomes are found within the same cell—these events most often occur in non-human mammal species such as bats, resulting in progeny viruses that acquire the ability to infect human cells [18, 19]. Thus, CoV genomes often retain key features of their ancestral virus but also include new features that allow for the species jump. Not surprisingly, the genome organisation of SARS-CoV-2 is largely similar to that of the existing human CoVs and, in particular, to SARS-CoV which is the human CoV with the most identity at a nucleotide level. The SARS-CoV-2 genome, a single-stranded positive-sense RNA molecule of approximately 29,800 nucleotides, is arranged into 14 open reading frames (ORFs) encoding 27 proteins and is shown schematically in Fig. 1 [20]. The majority of the genome contains the ORF1a and ORF1b that encodes 16 different non-structural proteins (nsp1-nsp16) involved in the 'replicase' complex although many have very diverse but critical functions. The final one-third of the genome houses several ORFs encoding 4 structural (spike, S; envelope, E; membrane, M; nucleocapsid, N) and 10 accessory proteins (ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8a, ORF8b, ORF9b, ORF9c, ORF10). Some of these ORFs are overlapping or found within a larger ORF (Fig. 1a). At either end of the genome are non-coding or

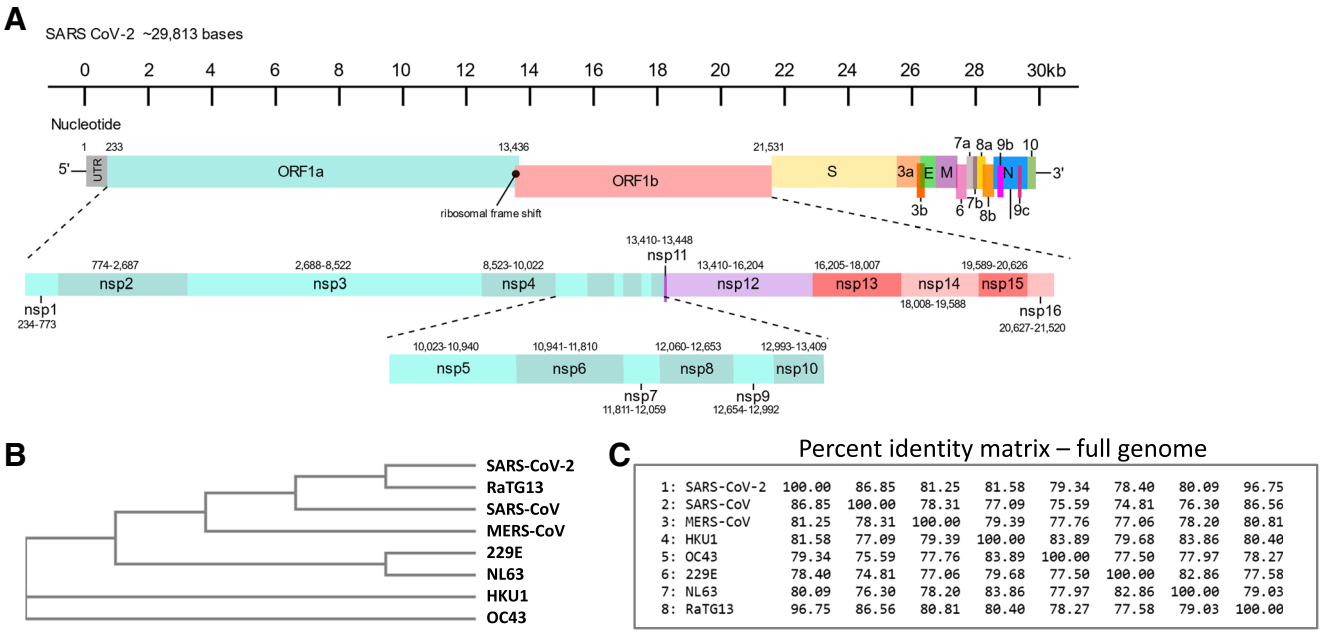


Fig. 1 Human CoV genome organisation and relationships. Schematic of SARS-CoV-2 genome based on NCBI sequence MT786327 (severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TUR/Kafkas-SARSCoV2-001/2020, complete genome, 2020) (**a**). The genome is shown with ORFs boxed in colour below the nucleotide numbering. The region from ORF1a to ORF1b (nucleotides ~233–~21,531) is expanded below to resolve nsp1–nsp16. Phylogenetic tree of human CoVs and closest bat CoV (RaTG13) created by Kalign (EMBL-EBI) (https://www.ebi.ac.uk/Tools/msa/kalign/) (**b**). NCBI reference sequences accession numbers: SARS-CoV-2 (MT786327); SARS-CoV (NC_004718); MERS-CoV (NC_019843); HKU1 (NC_006577); OC43 (NC_006213); 229E (NC_002645); NL63 (NC_005831). Percent identity matrix of full-genome sequences produced by Clustal2.1 alignment of EMBL-EBI Kalign output (**c**)

untranslated regions (UTRs) known as the 5′UTR and 3′UTR. These UTRs are relatively short, being approximately 230 bases, but have important regulatory functions [21]. The 5′ UTR is highly structured and thought to contain 5 stem-loop (SL1–SL5) structures, with the SL3 structure also housing the transcriptional regulation sequence (TRS-L) important for generation of subgenomic mRNAs [22]. Each gene in the genome also has its own upstream TRS-B sequence. The 3′ UTR is equally important in CoV replication containing conserved and hypervariable bulged stem-loop (BSL) structures that may function as molecular switches [23]. Another critical RNA structure between ORF1a and 1b is predicted to exist that allows ribosomal frameshifting [23]. This stem-loop structure may form a pseudoknot that creates a so-called slippery sequence allowing a translational shift of frame, thought to be necessary for the correct translation of the nsp12 or RNA-dependent RNA polymerase (RdRp)—the viral enzyme critical for the genesis of new genomic material.

We performed phylogenetic analyses of all seven human CoV full-genome reference sequences that revealed that the endemic CoVs (229E, NL63, HKU1, OC43) are more closely related to each other and diverge from the newest three CoVs (MERS, SARS1, SARS2), which also cluster together (Fig. 1b). At the genome level, SARS-CoV-2 is 96.75% similar to a bat CoV, known as RaTG13, but its nearest human CoV relative is SARS-CoV at just 86.85% similar (Fig. 1c). A complete genome identity matrix for all seven human CoVs revealed that SARS-CoV-2 is less similar to the remaining CoVs, including MERS, at approximately 80% identity (Fig. 1c).

The closest relative of SARS-CoV-2 has long been thought to be the bat CoV RaTG13 (reported to be discovered in 2013), but the genome sequence was made publicly available only in March 2020, well after the Covid-19 pandemic beginnings [24]. Although several bat CoVs were discovered in Wuhan, China, and reported in 2017, the name RaTG13 does not feature in the literature, until 2020 [25]. For optimal zoonotic transfer to humans, an intermediate species of CoV, more closely related to SARS-CoV-2, should exist, although the pangolin has been suggested as one possibility due to genetic similarity to SARS-CoV-2 of a specific CoV, isolated from sick animals [26]. Here, SARS-CoV-2 would require selection and adaptation to bind a human-like ACE2 and indeed both RaTG13 and a recently identified pangolin coronavirus (Pangolin-CoV-2020) have been demonstrated to bind human ACE2 [27]. Chinese horseshoe bats are natural reservoirs of SARS-like CoVs that enter cells using ACE2 and these would usually require an intermediate species before adapting to infect humans [28]. Surprisingly, the S protein of the RaTG13 strain was found to be unable to bind to the ACE2 receptor of two different types of horseshoe bats, despite being the suspected natural host species [29]. RaTG13, discovered from a faecal sample of a horseshoe bat *Rhinolophus affinis*, has 96.75% homology with SARS-CoV-2 and there is no other reported CoV of higher similarity.

The S ORF of RaTG13 is 94.9% identical to that of SARS-CoV-2 (Fig. 2a). Speculation of a non-zoonotic origin has been spurred by the inability to find an intermediate strain of higher homology and the reported appearance of the RaTG13 sequence after the SARS-CoV-2 outbreak had already started. Thus, arguments exist that challenge the zoonotic origin of SARS-CoV-2 and, instead, suggest that the virus may not have a natural origin. A report investigating features of the virus genome that do not align with common aspects of CoVs evolution speculates that the bat coronaviruses ZC45 and ZXC21 (ZC45 GenBank accession MG772933.1; ZXC21 GenBank accession MG772934.1; both sequences submitted January 5, 2018) could have been used as a backbone for the genetic engineering of SARS-CoV-2 [30]. Here, it is noted that ZC45 and ZXC21[2] share approximately 86% similarity with SARS-CoV-2 and would require minimal modifications to create the novel CoV [31]. Unusual features highlighted in support of this are the high conservation of ORF8 and E between ZC45/ZXC21 and SARS-CoV-2. ZC45/ZXC21 shares 94.2% identity with the SARS-CoV-2 ORF8 and no other CoVs share more than 58% identity with SARS-CoV-2 on this particular protein [31]. A 2017 analysis of various SARS-related bat CoVs not only acknowledges the high variability of ORF8 but also reports that ORF8 is highly conserved between the human SARS-CoV and a closely related bat CoV [25]. Therefore, this high conservation is not entirely unheard of in natural settings. The E sequence is 100% identical between ZC45/ZXC21 and SARS-CoV-2, even though this small structural protein is tolerant of amino acid substitutions [31]. Evidence also suggests that the SARS-CoV-2 E sequence is changing throughout the pandemic [31] albeit at a low frequency [32]. Finally, it has been speculated that the S protein of SARS-CoV-2 may have been manipulated. The successful engineered swapping of the S receptor-binding motif (RBM) of SARS-CoV, a region of S that is critical for ACE2 recognition, with several closely related bat coronaviruses in 2008, allowed modification of the S protein to interact with the ACE2 entry receptor [33]. Interestingly, the SARS-CoV-2 S ORF contains two unique restriction sites EcoRI (gaattc) and BstEII (ggttacc) flanking the RBM sequence that are not present in other human CoVs (HKU1, OC43, MERS, SARS-CoV) and could facilitate such engineering. It is striking that the RaTG13 S ORF contains the identical EcoRI and BstEII restriction sites. However, the full SARS-CoV-2 genome sequence contains 9 EcoRI and 4 BstEII sites, rendering these not unique.

The origin of SARS-CoV-2 remains an area of intense investigation considering the scope of the pandemic. Efforts to understand its evolution via an intermediate species remain at the forefront and would require recombination events between two CoVs, whereby a S RBM with improved affinity for human ACE2 was present. Andersen et al. [5] argue that the RBM of SARS-CoV-2 contains substituted amino acids diverging from SARS-CoV that indicate 'the high-affinity binding of the SARS-CoV-2 spike protein to human ACE2 is most likely the result of natural selection on a human or human-like ACE2 that permits another optimal binding solution to arise'. The authors propose the origin of SARS-CoV-2 as either (1) natural selection in an animal host before zoonotic transfer and (2) natural selection in humans following zoonotic transfer [5]. Until an intermediate CoV with the key features of the S gene found in SARS-CoV-2 is discovered, these proposals and the unnatural origin theory are impossible to prove or disprove.

The SARS-CoV-2 Spike Protein Determines Cell and Tissue Tropism

The S ORF of human CoVs encodes the critical S protein that covers the surface of the viral particle and facilitates entry into cells. The S protein sequence has high homology to that of RaTG13 (94.9%) and also to SARS-CoV (84.74%); therefore, the SARS-CoV-2 S is well-adapted to bind human ACE2. SARS-CoV-2 has high affinity for human ACE2 which is the entry receptor [6]. Homology with the remaining human CoVs that do not bind ACE2 and use a variety of different entry receptors is not surprisingly lower (Fig. 2a). S is a large protein of over 1200 amino acids that can be broken down into two subunits S1 and S2 by host cell enzymatic cleavage. The S1 subunit is the most external region and determines interaction with ACE2 through its RBM (Fig. 2b). Alignment of the SARS-CoV-2, SARS-CoV and RaTG13 RBM demonstrate homologous regions that specify interaction with ACE2 (Fig. 2c). To infect cells, SARS-like CoVs must first bind with ACE2 on the surface of cells followed conformational changes which precedes membrane fusion events and uptake of the virus into cells [34]. This interaction between SARS-CoV-2 and ACE2 defines the cells and tissues that the virus can infect and subsequent pathology.

Cellular entry of SARS-CoV-2 depends not only on the binding of the S protein to ACE2 but also on the subsequent S protein priming by cellular proteases, in particular TMPRSS2 and cathepsin B/L activity which in vitro can substitute for TMPRSS2 [35]. ACE2 and TMPRSS2 have been detected in both nasal and bronchial epithelia by immunohistochemistry [36]. Single-cell transcriptomics datasets from different tissues, including those of the cornea, retina, respiratory tract, oesophagus, ileum, colon, heart, skeletal muscle, spleen, liver, placenta/decidua, kidney, testis, pancreas, prostate gland, brain and skin, have been used to map the expression of *ACE2* and *TMPRSS2* genes [37, 38]. Notably,

[2] An NCBI Blast search of ZXC21 reveals 100% coverage and 97.48% identity to ZC45 with the remainder of hits being SARS-CoV-2 clinical isolates with 91–95% coverage and 88.67% identity.

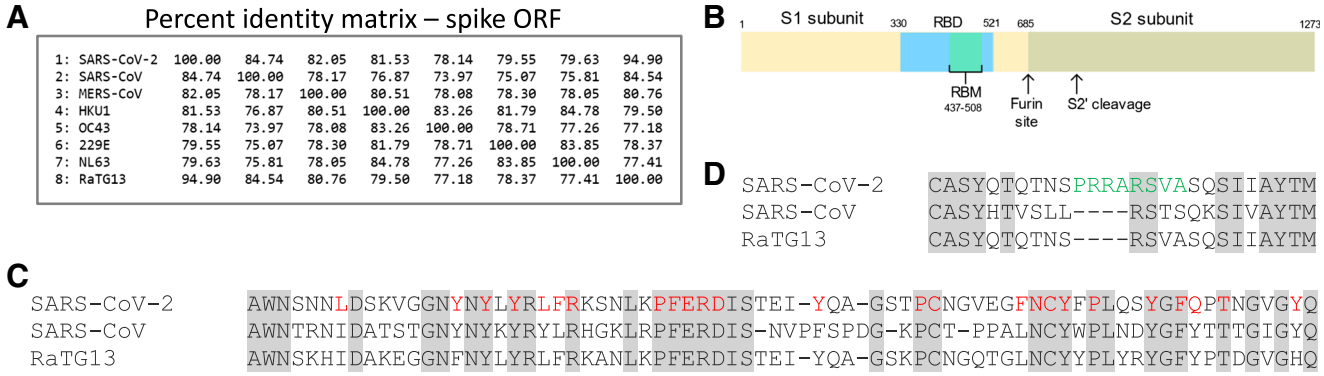


Fig. 2 CoV spike protein ORF and sequence details comparisons. Percent identity matrix of spike ORF sequences produced by Clustal2.1 alignment (**a**). Based on NCBI sequence MT786327 (severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/TUR/Kafkas-SARSCoV2-001/2020, complete genome, 2020). CoV spike sequences based on information within Fig. 1. Schematic view of SARS-CoV-2 spike protein domains (**b**). Numbering according to amino acid position. Receptor-binding domain (RBD, blue), receptor-binding motif (RBM, green), spike subunits (S1 and S2) generated by enzymatic cleavage (furin site and S2' cleavage). Alignments of the spike RBM (**c**) and S1/S2 junction amino acids showing furin polybasic recognition site (**d**) (green residues) of SARS-CoV-2, SARS-CoV and RaTG13 as created by Kalign (EMBL-EBI) (https://www.ebi.ac.uk/Tools/msa/kalign/). Critical residues for ACE-2 receptor interaction are displayed in red and conserved residues boxed in grey

although the absence of gene expression should be interpreted with caution as analysis of single-cell transcriptomics may lack specific cell types due to their rarity, to the limitations of cell isolation protocols, methods of analysis and dropout effects, detection of *ACE2* and *TMPRSS2* is highly reliable. *ACE2* is detected in cells of many tissues, including the cornea, respiratory tract, heart, kidney and testis, along the digestive tract in the oesophagus, ileum, colon and in accessory digestive organs—liver and gallbladder. Some cells co-express both ACE2 and TMPRSS2 such as cells of the cornea and of the airways. Interestingly, *TMPRSS2* expression shows a broader distribution suggesting that ACE2 may be a limiting factor for viral entry [38]. Specifically, co-expression of *ACE2/TMPRSS2* in nasal goblet and ciliated cells implicates their potential importance as entry sites and reservoirs for SARS-CoV-2. Notably, although gene expression data clearly show nasal *ACE2* mRNA expression, detection of its protein is less clear [36, 38, 39]. The co-expression in conjunctival cells could also imply a putative spread through the nasolacrimal duct and it explains ocular symptoms observed in some Covid-19 patients [40]. The concomitant presence of *ACE2* and *TMPRSS2* transcripts in cells of the oesophagus, ileum and colon could explain viral faecal shedding previously reported [41] and suggests a potential faecal-oral transmission of SARS-CoV-2, which so far does not have other support.

SARS-CoV-2 Spike ORF Contains a Unique Insertion Encoding a Furin Cleavage Site

After receptor interaction through the S glycoprotein, CoVs utilise diverse host cell proteases for cleavage activation of virus-host cell membrane fusion and subsequent genome delivery. Inspection of the SARS-CoV-2 genome reveals the presence of a furin cleavage site at the S1/S2 junction of the S protein that is not found in other beta-CoVs of any species. This is achieved by a 12-base insertion (cctcggcgggca) encoding PRRA amino acids (Fig. 2d). Protease cleavage of the S protein is necessary for initiating host cell invasion after ACE2 attachment and is usually achieved by the S2' cleavage site that is catalyzed by the host cell serine protease TMPRSS [35]. Dual-protease cleavage of SARS-CoV-2 spike would therefore provide a fusion and entry advantage over that of related human CoVs, potentially priming S for optimal conformation and entry receptor interaction that could enhance replication and transmission. This concept is not without precedent, as manipulation of a porcine CoV enabled researchers to alter the trypsin-dependent protease cleavage site to be activated instead by furin, which enhanced infectivity of target cells [42]. Although this suggests furin cleavage can be synthetically engineered, a separate study of proteolytic cleavage in the MERS-CoV spike protein emphasises that MERS-CoV is capable of adapting to various conditions and cleavage can be activated by either trypsin or furin proteases. Furin cleavage motifs were identified at S1/S2 and S2' in MERS both by sequence alignment and a furin cleavage prediction algorithm [43]. This adaptability contradicts theories of genetic manipulation of the spike protein cleavage site, as SARS-like CoVs may have already been readily capable of naturally switching to furin-activated cleavage.

Diagnosis of SARS-CoV-2 Infection

Detection of SARS-CoV-2 infection can be performed by two complementary methods: (1) detection of the virus itself (protein or RNA) in mostly upper respiratory tract samples or (2) detection of the immune response to the virus, preferentially the

IgM and/or IgG antibody responses. In the acute phase of the infection, due to the time required to induce the adaptive immune response, the serological methods have lower sensitivity. Therefore, after the first SARS-CoV-2 genome sequence (Wuhan-Hu-1, GenBank accession number MN908947) was made available in January 2020 (https://virological.org/t/novel-2019-coronavirus-genome/319), a variety of molecular diagnostic tests have been developed to detect the viral genome. RT-PCR is therefore routinely used to diagnose current infection by detecting parts of the SARS-CoV-2 RNA genome. Swabs are taken from the nose or throat, followed by extraction of RNA and downstream processing using commercially available reagents, eventually making use of specific primer-probe sets to quantitate the viral RNA [44, 45]. According to the FIND database which collects the descriptions of commercially available or in development SARS-CoV-2 diagnostic assays, as of 29 September 2020, there were 853 assays in the database with 377 described to target the SARS-CoV-2 genome (https://www.finddx.org/covid-19/pipeline/). Besides the most commonly available PCR-based and quantitative PCR (qPCR)–based methods, other methods such as loop-mediated isothermal amplification (LAMP), transcription-mediated amplification (TMA), clustered regularly interspaced short palindromic repeats (CRISPR)–based assays, rolling circle amplification, microarray and metagenome sequencing are theoretically available [46]. Despite the availability of various detection methods, a review of 112 molecular detection assays showed that 90% are based on PCR, 6% are based on isothermal amplification technologies (e.g. LAMP and TMA), 2% are based on hybridisation technologies and 2% utilise CRISPR-based technologies [46]. Amongst the PCR methods, qPCR is the most commonly used due to the shorter assay time (no need for gel electrophoresis), higher sensitivity and higher specificity (two primers and a probe compared to two primers in a conventional PCR). Several problems arise regarding PCR-based detection of SARS-CoV-2 genome. First, SARS-CoV-2 is an RNA virus; therefore, the RNA should be reverse transcribed into a cDNA that would serve as a template in PCR. The reverse transcription (RT) step can be performed before the PCR, or there is a one-step possibility when the RT step and the PCR step are performed in the same tube. The one-step method requires no transfer between the RT tube and the PCR tube decreasing the chance of contamination. The second problem is the fact that most of the PCR-based detection methods detect more than one SARS-CoV-2 gene. The US Centers for Disease Control and Prevention (CDC) recommended RT-qPCR method containing primers and probes for two regions of the SARS-CoV-2 nucleocapsid gene (N1 and N2) and also for one region of the human RNase P as a positive control for RNA extraction and lack of inhibition in the qPCR. The CDC assay is considered positive when both N1 and N2 singleplex qPCR assays are positive; otherwise, the result is not conclusive. The World Health Organization (WHO) recommends primer and probe sets for the E and the RdRp genes. First, the E gene is tested, and positivity is confirmed by a separate RdRp qPCR. Additional confirmation is based on the detection of the N gene [47].

The most streamlined approach would be a one-step multiplex RT-qPCR assay that performs the RT and identifies several genes in the same reaction. Examples of multiplex RT-qPCRs include the QIAstat-Dx Respiratory SARS-CoV-2 panel targeting the E and RdRp genes, the TaqPath Covid-19 combo targeting the ORF1b and N and S genes, and the Allplex 2019-nCoV assay targeting the E, N and RdRP genes [46]. The need to detect more than one viral gene sequence is related to the mutation of the viral genome that potentially renders these sequence-based detection methods less usable (lower sensitivity) or unusable (lack of inclusivity) due to primer and/or probe mismatches. The genomic mutation rates for RNA viruses are generally higher than for the DNA viruses [48]; however, the RNA polymerase of CoVs in cooperation with the non-structural protein 14 (nsp14) contains proofreading activity; therefore, the mutation rate of CoVs is lower. The estimated evolutionary rate of the SARS-CoV-2 is 2.24×10^{-3} substitutions/site/year [49] and analysis of 28,726 SARS-CoV-2 whole genome sequences found 7823 SNP profiles with 4968 single mutations in SARS-CoV-2 isolates originating within the USA [50]. Several studies exist linking genomic changes to diagnostic sensitivity. In one such study, 39 primer and probe sequences used in SARS-CoV-2 diagnostic qPCRs were mapped to 30 Colombian SARS-CoV-2 sequences [51]. 5 nucleotides (primers or probes) showed mismatches with at least one Colombian SARS-CoV-2 sequence. Examples include the Corman-Berlin (2020) RdRP SARSr assay forward primer, which had a mismatch in one Colombian SARS-CoV-2 sequence located in a 3′ region and strongly influenced the binding of the primer and subsequent DNA extension. The reverse primer of the same assay also had a mismatch in all the included Colombian SARS-CoV-2 sequences, but it was located at an internal site of the primer and may not influence the binding significantly. The 20-nucleotide Hong Kong (2020) HKU-NP probe had 4 nucleotide mismatches with all of the Colombian sequences potentially leading to inefficient binding and false-negative results. The forward primer of the Zhu 2020 CDC-China Set II had a significant GGG➔AAC mismatch in three Colombian SARS-CoV-2 sequences, while the 3′ end of the same primer contained a single nucleotide mismatch in 2 genome sequences. In another study, primer-probe sequences of 8 commonly used SARS-CoV-2 qPCR tests were compared to 15,001 SARS-CoV-2 genome sequences [52]. 12 primer-probe sets covered over 98% of SARS-CoV-2 genomes without mismatches. Reverse primers of two primer-probe sets contained a single mismatch in over 99% of genomes. Forward primer of the China CDC assay showed mismatches against 23 SARS-CoV-2 genomes with up to 8 nucleotide mismatches. Similarly, to the previously mentioned Colombian data, the forward primer of the China CDC assay showed a trinucleotide mismatch

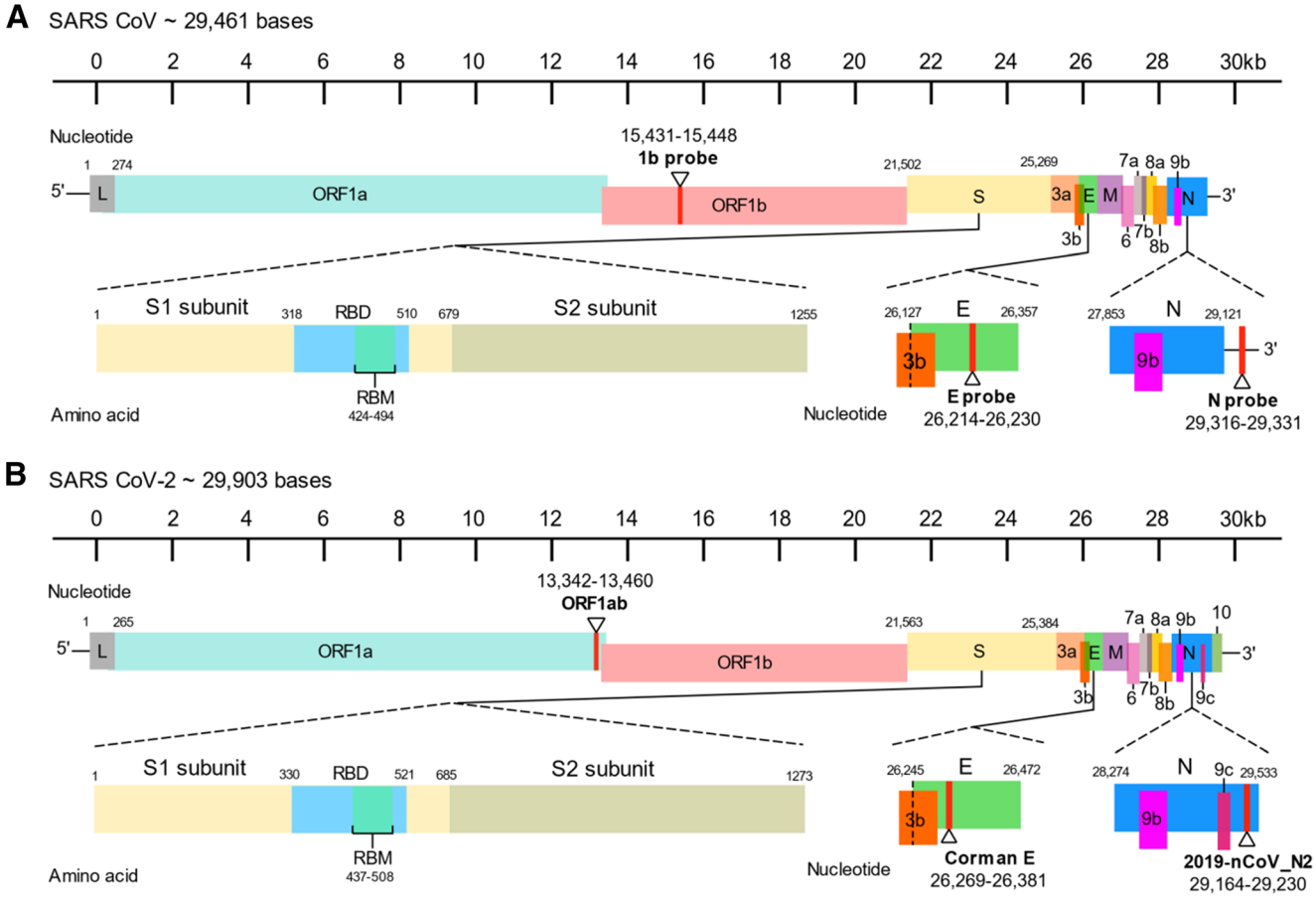
GGG➔AAC in one of the SARS-CoV-2 genome variants. This variant was detected first in February 2020 and by June 8, 2020, was presented in 18.8% of the analyzed SARS-CoV-2 genomes. Altogether, the dynamic changes in the SARS-CoV-2 genome and the emergence of new virus clades direct us to assess genomic detection results with caution. Inconclusive results or negative results with the presence of typical clinical symptoms require repeating the diagnostic qPCR with another sample or performing the diagnostic qPCR with another assay targeting different regions of the SARS-CoV-2 genome.

Genetic Divergence of SARS-CoV-2 from SARS-CoV and Diagnostic Considerations

As outlined earlier, the diagnosis of current SARS-CoV-2 infection is largely performed by qRT-PCR amplification of virus-specific sequences although serological testing can confirm evidence of current and past infection [53]. The challenge in qRT-PCR diagnostic testing of SARS-CoV-2 is therefore to target a region of the genome that is both different enough from other CoVs to be distinguishable, yet reliably stable despite mutation accumulation. Critically, the diagnostic test must be compatible with the vast majority of Covid-19 cases, regardless of geographical mutation differences. The primer-probes used to distinguish SARS-CoV and SARS-CoV-2 are important considering the genome similarity between the two CoVs [1]. As described above, many primer-probe mixtures are currently in use for diagnosis of SARS-CoV-2 by qRT-PCR. While the sensitivity and selectivity of each mixture can vary, there are more than a dozen combinations with proven specificity and sensitivity in laboratory settings. A comparison of ten primer-probe sets by a Korean research team identified 2019-nCoV_N2, N3, developed in the USA, and NIID_2019-nCOV_N, developed in Japan, as the most sensitive primer-probe sets for targeting the N gene. Another set, 'ORF1ab' from China, was determined to be the most sensitive primer-probe mixture for targeting the RdRp region of the SARS-CoV-2 genome [54]. Researchers inferred that a combination of these primer-probe sets would yield the best selectivity and sensitivity for diagnostic testing. The reliability and variety of primer-probe sets that target the N gene are in agreement with the use of the antibody tests which detect anti-N antibodies as a means of identifying if an individual has been previously exposed to SARS-CoV-2 [55]. This further supports the concept that the N region of the genome is both unique and reliably stable despite the evolution of the virus as it spreads. Another analysis of seven different primer-probe sets at the University of Washington Clinical Virology Lab found that assays using N2 and a set targeting the E gene, known as 'Corman E gene' primer-probe set, were the most sensitive for SARS-CoV-2, although all primer-probe sets being analyzed showed high specificity and none exhibited cross-reactivity or false positives [56]. The location of the 2019-nCoV_N2, ORF1ab and Corman E gene probes in the SARS-CoV-2 reference genome is indicated in Fig. 3. For comparison purposes, the probes illustrated for SARS-CoV correspond to the same proteins as those illustrated for SARS-CoV-2. These are the N probe, 1b probe and E probe. However, there are also several reliably sensitive primer-probe sets for the detection of SARS-CoV, including those that can target regions of the spike ORF and other structural proteins ORFs. Several SARS-CoV-2 primer-probe sets whose nucleotide location could be identified at the time of this report have also been listed in Fig. 3.

SARS-CoV-2 Genome Mutation Analysis

RNA viruses routinely accumulate mutations and changes in their genome sequence as a result of polymerase infidelity [48]. SARS-CoV-2 is no different but incorporates proofreading activity as outlined above to minimise nucleotide mutation. A regularly updated phylogenetic tree of SARS-CoV-2 is publicly available at Nextstrain.org [57] and is displayed in Fig. 4. Transmission of the virus over time is represented by branching, with interactive nodes indicating reported sample genomes and mutational events. Branching between nodes is hypothesised based on genomic similarities and region. The changes in colour indicate the spread to different continents as genetic drift causes a gradual divergence of the viral gene pools based on an infected individual's geographic location. A comparison of the original SARS-CoV-2 Wuhan reference genome with genomes compiled from the worldwide sample datasets depicts areas of highest variability. Tracking mutations and variable regions is critical for the development of effective therapeutics as genetic diversity can have implications towards drug resistance. It can be readily seen in Fig. 4a that the SARS-CoV-2 genome accumulates numerous mutations over time and that these genotypes are distributed in a unique fashion globally (Fig. 4a). Some individuals may contract SARS-CoV-2 in a geographical region that has different accumulated mutations from other parts of the world, and these mutations may occur in critical parts of the genome. Other mutations will have little to no effect on viral replication and transmission but will accumulate over time and become a molecular marker of virus spread geographically. Given that the pandemic is still ongoing, and mutations will continue to accumulate and branch off from the original reference genome, this is a dynamic and variable process that will determine features of the virus pandemic. Figure 4c, also obtained from Nextstrain.org, demonstrates that mutations have accumulated across the SARS-CoV-2 genome with hotspots frequently found within ORF1b, S and N genes. The accumulated mutations have led to geographically different viral gene pools, seen by the differences in colour shown in Fig. 4a, b. It is possible that Covid-19 could become a seasonal disease,



C

Additional SARS CoV-2 primer-probe sets

Name	Nucleotide Range
RdRp_SARSr	15,431-15,530
HKU-ORF1b_nsp14	18,778-18,909
2019-nCoV_N1	28,287-28,358
WH-NIC N	28,320-28,376
2019-nCoV_N3	28,681-28,752
N	28,881-28,979
NIID_2019-nCOV_N	29,125-29,282
HKU_N	29,145-29,254

Fig. 3 SARS-CoV and SARS-CoV-2 genomes schematic displaying diagnostic RT-PCR primer-probe sets. The SARS-CoV genome (**a**) and SARS-CoV-2 genome (**b**) schematics showing positions of important RT-PCR primer-probe sets for diagnostic analysis (red lines). Additional SARS-CoV-2 primer-probe sets and their genome location are displayed in (**c**)

with peaks in cases occurring annually. This effect would imply that either the virus mutates quickly and attenuates itself or could evade our immunity by genetic drift similarly to influenza virus. Neither outcome can be determined during the ongoing pandemic, but they are important concepts to consider given that the former path would herald an end to the pandemic and the latter would require constant adaptation of diagnostic tests and updated vaccine preparations to be generated.

Another concept, considering the prevalence of virus variants, is that re-infection following recovery could be possible. One such event has been documented and showed that the genome sequence of the virus strain in the first episode of Covid-19 infection was clearly different from that of the virus strain found during the second episode of infection [58]. Here, a young and healthy patient had a second episode of Covid-19 infection diagnosed 4.5 months after the first episode. Viral genomes from first and second episodes belonged to different

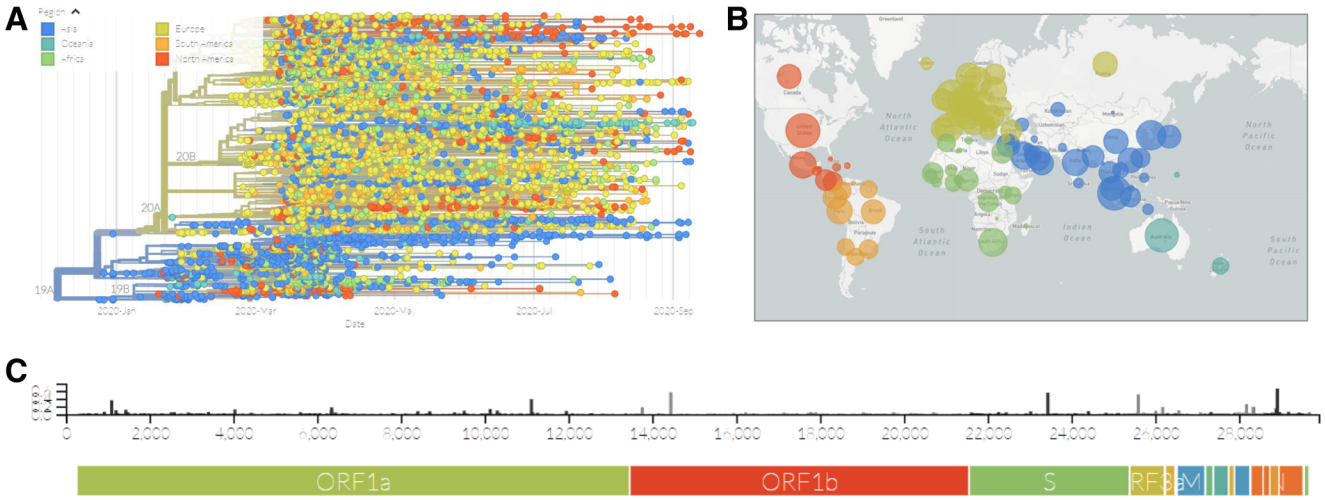


Fig. 4 Genomic epidemiology of SARS-CoV-2—global subsampling. Screenshots of SARS-CoV-2 genome mutation (**a**), variation globally (**b**) and regions of frequent mutation are demonstrated as bars within the SARS-CoV-2 genome schematic (**C**). Images were downloaded from https:nextstrain.org/ncov/global on 24 September 2020 showing 4640 genomes sampled between December 2019 and September 2020 (last updated 18 September 2020)

lineages and 24 nucleotides were different between the viruses. Amino acid changes were identified in 9 viral proteins and included a large truncation of ORF8 protein that was present only in the virus from the first infection. This case demonstrates that not only is re-infection with SARS-CoV-2 possible but also that virus mutation could drive this process to avoid pre-existing protective immunity.

As transmission and global spread of SARS-CoV-2 has increased, many accumulated mutations have prevailed (Fig. 4a). Mutations that either strengthen the virus' infectivity or replication, or those that do not detrimentally affect the function of the virus, are most likely to be passed on. One such mutation at position 614 of the spike protein that changes the encoded amino acid from aspartic acid to glycine, known as mutation D614G, has been found to increase SARS-CoV-2 infectivity of ACE2 expressing cells [59] and has become the prevalent virus variant in the pandemic [60, 61]. The mechanism behind this enhancement is understood to be a result of decreased S1 shedding when S protein is cleaved at the S1/S2 junction. A single virion with more S proteins that have not shed the S1 subunit can, therefore, result in a greater likelihood for functional S proteins to encounter ACE2 for attachment. This concept has been proved experimentally in vitro using a leukaemia virus pseudotyped with G614 SARS-CoV-2 S genes, resulting in 9 times greater infectivity and decreased S1 shedding than of those containing D614 S proteins [62].

A common mutation at position 314 of ORF1b affects the RdRp complex, which is encoded by nsp12. Mutations to the RdRp complex are rarely viable since it is responsible for transcription processes and its structure is highly conserved across both SARS-CoVs. The mutation identified at codon 314 results in an amino acid change from proline to leucine and has been observed to increase mutation rate incidences [63]. Proline is associated with flexibility and leucine with stabilisation in proteins; thus, increased rigidity in the overall RdRp structure due to the leucine introduction may result in poorer interactions with unwinding RNA, leading to a higher rate of transcription error, an effect confirmed by structural analysis of the molecular flexibility of a mutated RdRp molecule [64]. As can be expected, the emergence of the RdRp mutation in February 2020 led to increased subsequent mutations, including those in spike and the nucleocapsid, such that these alterations tend to exist simultaneously [65].

Mutations in N resulting in two amino acid substitutions, arginine to lysine, and glycine to arginine at positions 203 and 204, have also been documented [65]. These genomic changes have been associated with decreased microRNA (miRNA) binding, which can result in a higher susceptibility to infection. Host miRNA binding at sites of infection can limit invading viral pathogens before a successful infection has been established [66].

In addition to these point mutations accumulating in the SARS-CoV-2 genome, there are several documented larger genomic deletions that have occurred during the pandemic. Since CoV genome evolution is driven by a series of recombination events and incremental adaptive mutations, the appearance of more dramatic genomic variants is not surprising. Deletions within the spike gene resulting in short losses of amino acids that can attenuate SARS-CoV-2 have been found [67–69] and suggest that the S gene allowing human infection may be under strong selective pressure. Considering the large proportion of asymptomatic cases that are observed in the pandemic, screening such individuals for the presence of S deletion mutants would be of interest. Additionally, it is likely that S deletion mutants might make useful attenuated vaccine

candidates. Further deletion mutants have been described in nsp1 [70], ORF7b [71] and ORF8 [72], but overall, these are much rarer events than single point mutations.

Concluding Remarks

SARS-CoV-2, a recently emerged CoV responsible for the current global pandemic Covid-19, first appeared in late 2019 in Wuhan, China. The virus is thought to have arisen from SARS-like CoVs in bats due to high similarities in genome sequence which are also shared with the prior SARS-CoV. The SARS-CoV-2 genome retains many features of endemic human CoVs but the critical determinant, the S protein, is sufficiently adapted to bind the human entry receptor ACE2 much more readily than SARS-CoV which is the most closely related human CoV. There is evidence that the viral genome is undergoing subtle evolution through mutation to enhance transmission and there is evidence for limited attenuation that might weaken the virus. The scientific and medical community has mobilised in an unprecedented fashion to understand the virus at molecular and epidemiological levels, determine its pathological consequences, understand protective immunity and develop specific antivirals, vaccines and other treatments. However, as of writing, the pandemic is yet to be fully under control although some territories have had success. Currently, the only effective measures to restrict viral transmission are limiting social interactions, mass diagnostic testing and contact tracing applications. Further understanding of the genetics behind the virus' emergence and mechanisms of replication will be critical to generate specific therapeutics and protective vaccines to halt the continued spread. This knowledge will also assist and limit the future potential for new CoVs to emerge.

Authors' Contributions PE prepared figures, researched and wrote sections of the manuscript; FS, DPV and MN researched and wrote sections of the manuscript; GRM planned the manuscript, prepared figures, researched and wrote sections of the manuscript, and assembled the final version.

Funding DPV was supported by the Hungarian–European Union Grant EFOP-3.6.1-16-2016-00008. MN was a recipient of the Imperial College Covid-19 research fund. GRM received Research and Knowledge Exchange funds from London Metropolitan University.

Data Availability All data contained within the manuscript are freely available by searching relevant databases and references cited or by contacting the corresponding author.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent The authors declare that no human or animal studies were performed during this research.

Consent for Publication All authors have approved this submission and agree to publish this original manuscript.

References

1. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. Viruses. 2020;12(2). https://doi.org/10.3390/v12020244.
2. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270–3. https://doi.org/10.1038/s41586-020-2012-7.
3. Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. Lancet Infect Dis. 2020;20(9):e238–e44. https://doi.org/10.1016/s1473-3099(20)30484-9.
4. Mukhra R, Krishan K, Kanchan T. Possible modes of transmission of novel coronavirus SARS-CoV-2: a review. Acta Bio-med. 2020;91(3):e2020036. https://doi.org/10.23750/abm.v91i3.10039.
5. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020;26(4):450–2. https://doi.org/10.1038/s41591-020-0820-9.
6. Wang Y, Liu M, Gao J. Enhanced receptor binding of SARS-CoV-2 through networks of hydrogen-bonding and hydrophobic interactions. Proc Natl Acad Sci U S A. 2020;117(25):13967–74. https://doi.org/10.1073/pnas.2008209117.
7. Turner AJ, Hiscox JA, Hooper NM. ACE2: from vasopeptidase to SARS virus receptor. Trends Pharmacol Sci. 2004;25(6):291–4. https://doi.org/10.1016/j.tips.2004.04.001.
8. Fung TS, Liu DX. Human coronavirus: host-pathogen interaction. Annu Rev Microbiol. 2019;73:529–57. https://doi.org/10.1146/annurev-micro-020518-115759.
9. de Wilde AH, Snijder EJ, Kikkert M, van Hemert MJ. Host factors in coronavirus replication. Curr Top Microbiol Immunol. 2018;419:1–42. https://doi.org/10.1007/82_2017_25.
10. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. Int J Biol Sci. 2020;16(10):1686–97. https://doi.org/10.7150/ijbs.45472.
11. Hung LS. The SARS epidemic in Hong Kong: what lessons have we learned? J R Soc Med. 2003;96(8):374–8. https://doi.org/10.1258/jrsm.96.8.374.
12. Reusken CB, Messadi L, Feyisa A, Ularamu H, Godeke GJ, Danmarwa A, et al. Geographic distribution of MERS coronavirus among dromedary camels, Africa. Emerg Infect Dis. 2014;20(8):1370–4. https://doi.org/10.3201/eid2008.140590.
13. Aleanizy FS, Mohmed N, Alqahtani FY, El Hadi Mohamed RA. Outbreak of Middle East respiratory syndrome coronavirus in Saudi Arabia: a retrospective study. BMC Infect Dis. 2017;17(1):23. https://doi.org/10.1186/s12879-016-2137-3.
14. Lovato A, de Filippis C. Clinical presentation of COVID-19: a systematic review focusing on upper airway symptoms. Ear Nose Throat J. 2020:145561320920762. https://doi.org/10.1177/0145561320920762.
15. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. 2020;382(18):1708–20. https://doi.org/10.1056/NEJMoa2002032.
16. Palmer K, Monaco A, Kivipelto M, Onder G, Maggi S, Michel JP, et al. The potential long-term impact of the COVID-19 outbreak on patients with non-communicable diseases in Europe: consequences

for healthy ageing. Aging Clin Exp Res. 2020;32(7):1189–94. https://doi.org/10.1007/s40520-020-01601-4.
17. Borges do Nascimento IJ, von Groote TC, O'Mathúna DP, Abdulazeem HM, Henderson C, Jayarajah U, et al. Clinical, laboratory and radiological characteristics and outcomes of novel coronavirus (SARS-CoV-2) infection in humans: a systematic review and series of meta-analyses. PLoS One. 2020;15(9):e0239235. https://doi.org/10.1371/journal.pone.0239235.
18. Chan JF, To KK, Tse H, Jin DY, Yuen KY. Interspecies transmission and emergence of novel viruses: lessons from bats and birds. Trends Microbiol. 2013;21(10):544–55. https://doi.org/10.1016/j.tim.2013.05.005.
19. Hu B, Ge X, Wang LF, Shi Z. Bat origin of human coronaviruses. Virol J. 2015;12:221. https://doi.org/10.1186/s12985-015-0422-1.
20. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265–9. https://doi.org/10.1038/s41586-020-2008-3.
21. Yang D, Leibowitz JL. The structure and functions of coronavirus genomic 3′ and 5′ ends. Virus Res. 2015;206:120–33. https://doi.org/10.1016/j.virusres.2015.02.025.
22. Miao Z, Tidu A, Eriani G, Martin F. Secondary structure of the SARS-CoV-2 5′-UTR. RNA Biol. 2020:1–10. https://doi.org/10.1080/15476286.2020.1814556.
23. Rangan R, Zheludev IN, Das R. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses. bioRxiv. 2020. https://doi.org/10.1101/2020.03.27.012906.
24. Boni MF, Lemey P, Jiang X, Lam TT, Perry BW, Castoe TA, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol. 2020;5:1408–17. https://doi.org/10.1038/s41564-020-0771-4.
25. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog. 2017;13(11):e1006698. https://doi.org/10.1371/journal.ppat.1006698.
26. Liu P, Jiang JZ, Wan XF, Hua Y, Li L, Zhou J, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog. 2020;16(5):e1008421. https://doi.org/10.1371/journal.ppat.1008421.
27. Li Y, Wang H, Tang X, Fang S, Ma D, Du C, et al. SARS-CoV-2 and three related coronaviruses utilize multiple ACE2 orthologs and are potently blocked by an improved ACE2-Ig. J Virol. 2020. https://doi.org/10.1128/jvi.01283-20.
28. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature. 2013;503(7477):535–8. https://doi.org/10.1038/nature12711.
29. Mou H, Quinlan BD, Peng H, Guo Y, Peng S, Zhang L, et al. Mutations from bat ACE2 orthologs markedly enhance ACE2-Fc neutralization of SARS-CoV-2. bioRxiv. 2020. https://doi.org/10.1101/2020.06.29.178459.
30. Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerg Microbes Infect. 2018;7(1):154. https://doi.org/10.1038/s41426-018-0155-5.
31. Yan L-M, Kang S, Guan J, Hu S. Unusual features of the SARS-CoV-2 genome suggesting sophisticated laboratory modification rather than natural evolution and delineation of its probable synthetic route. Zenodo. 2020. https://doi.org/10.5281/zenodo.4028830.
32. Sarkar M, Saha S. Structural insight into the role of novel SARS-CoV-2 E protein: a potential target for vaccine development and other therapeutic strategies. PLoS One. 2020;15(8):e0237300. https://doi.org/10.1371/journal.pone.0237300.
33. Menachery VD, Yount BL Jr, Debbink K, Agnihothram S, Gralinski LE, Plante JA, et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat Med. 2015;21(12):1508–13. https://doi.org/10.1038/nm.3985.
34. Chen J, Subbarao K. The immunobiology of SARS*. Annu Rev Immunol. 2007;25:443–72. https://doi.org/10.1146/annurev.immunol.25.022106.141706.
35. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020;181(2):271–80.e8. https://doi.org/10.1016/j.cell.2020.02.052.
36. Bertram S, Heurich A, Lavender H, Gierer S, Danisch S, Perin P, et al. Influenza and SARS-coronavirus activating proteases TMPRSS2 and HAT are expressed at multiple sites in human respiratory and gastrointestinal tracts. PLoS One. 2012;7(4):e35876. https://doi.org/10.1371/journal.pone.0035876.
37. Zou X, Chen K, Zou J, Han P, Hao J, Han Z. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. Front Med. 2020;14(2):185–92. https://doi.org/10.1007/s11684-020-0754-0.
38. Sungnak W, Huang N, Bécavin C, Berg M, Queen R, Litvinukova M, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat Med. 2020;26(5):681–7. https://doi.org/10.1038/s41591-020-0868-6.
39. Hamming I, Timens W, Bulthuis ML, Lely AT, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol. 2004;203(2):631–7. https://doi.org/10.1002/path.1570.
40. Hong N, Yu W, Xia J, Shen Y, Yap M, Han W. Evaluation of ocular symptoms and tropism of SARS-CoV-2 in patients confirmed with COVID-19. Acta Ophthalmol. 2020;98(5):e649–e55. https://doi.org/10.1111/aos.14445.
41. Xu Y, Li X, Zhu B, Liang H, Fang C, Gong Y, et al. Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. Nat Med. 2020;26(4):502–5. https://doi.org/10.1038/s41591-020-0817-4.
42. Li W, Wicht O, van Kuppeveld FJ, He Q, Rottier PJ, Bosch BJ. A single point mutation creating a furin cleavage site in the spike protein renders porcine epidemic diarrhea coronavirus trypsin independent for cell entry and fusion. J Virol. 2015;89(15):8077–81. https://doi.org/10.1128/jvi.00356-15.
43. Millet JK, Whittaker GR. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. Proc Natl Acad Sci U S A. 2014;111(42):15214–9. https://doi.org/10.1073/pnas.1407087111.
44. van Kasteren PB, van der Veer B, van den Brink S, Wijsman L, de Jonge J, van den Brandt A, et al. Comparison of seven commercial RT-PCR diagnostic kits for COVID-19. J Clin Virol. 2020;128:104412. https://doi.org/10.1016/j.jcv.2020.104412.
45. Yan Y, Chang L, Wang L. Laboratory testing of SARS-CoV, MERS-CoV, and SARS-CoV-2 (2019-nCoV): current status, challenges, and countermeasures. Rev Med Virol. 2020;30(3):e2106. https://doi.org/10.1002/rmv.2106.
46. Carter LJ, Garner LV, Smoot JW, Li Y, Zhou Q, Saveson CJ, et al. Assay techniques and test development for COVID-19 diagnosis. ACS Central Sci. 2020;6(5):591–605. https://doi.org/10.1021/acscentsci.0c00501.
47. Peñarrubia L, Ruiz M, Porco R, Rao SN, Juanola-Falgarona M, Manissero D, et al. Multiple assays in a real-time RT-PCR SARS-CoV-2 panel can mitigate the risk of loss of sensitivity by new genomic variants during the COVID-19 outbreak. In J Infect Dis. 2020;97:225–9. https://doi.org/10.1016/j.ijid.2020.06.027.
48. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010;84(19):9733–48. https://doi.org/10.1128/jvi.00694-10.

49. Li J, Li Z, Cui X, Wu C. Bayesian phylodynamic inference on the temporal evolution and global transmission of SARS-CoV-2. J infect. 2020;81(2):318–56. https://doi.org/10.1016/j.jinf.2020.04.016.
50. Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei G. Characterizing SARS-CoV-2 mutations in the United States. Res Square. 2020. https://doi.org/10.21203/rs.3.rs-49671/v1.
51. Álvarez-Díaz DA, Franco-Muñoz C, Laiton-Donato K, Usme-Ciro JA, Franco-Sierra ND, Flórez-Sánchez AC, et al. Molecular analysis of several in-house rRT-PCR protocols for SARS-CoV-2 detection in the context of genetic variability of the virus in Colombia. Infect Genet Evol. 2020;84:104390. https://doi.org/10.1016/j.meegid.2020.104390.
52. Kuchinski KS, Jassem AN, Prystajecky NA. Assessing oligonucleotide designs from early lab developed PCR diagnostic tests for SARS-CoV-2 using the PCR_strainer pipeline. J Clin Virol. 2020;131:104581. https://doi.org/10.1016/j.jcv.2020.104581.
53. Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. Cochrane Database Syst Rev. 2020;6(6):Cd013652. https://doi.org/10.1002/14651858.cd013652.
54. Jung YJ, Park G-S, Moon JH, Ku K, Beak S-H, Kim S, et al. Comparative analysis of primer-probe sets for the laboratory confirmation of SARS-CoV-2. bioRxiv. 2020:2020.02.25.964775. https://doi.org/10.1101/2020.02.25.964775.
55. Kontou PI, Braliou GG, Dimou NL, Nikolopoulos G, Bagos PG. Antibody tests in detecting SARS-CoV-2 infection: a meta-analysis. Diagnostics (Basel, Switzerland). 2020;10(5). https://doi.org/10.3390/diagnostics10050319.
56. Nalla AK, Casto AM, Huang M-LW, Perchetti GA, Sampoleo R, Shrestha L, et al. Comparative performance of SARS-CoV-2 detection assays using seven different primer-probe sets and one assay kit. J Clin Microbiol. 2020;58(6):e00557–20. https://doi.org/10.1128/jcm.00557-20.
57. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics (Oxford, England). 2018;34(23):4121–3. https://doi.org/10.1093/bioinformatics/bty407.
58. To KK-W, Hung IF-N, Ip JD, Chu AW-H, Chan W-M, Tam AR, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin Infect Dis. 2020. https://doi.org/10.1093/cid/ciaa1275.
59. 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. https://doi.org/10.1016/j.cell.2020.09.032.
60. 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(4):812–27.e19. https://doi.org/10.1016/j.cell.2020.06.043.
61. Isabel S, Graña-Miraglia L, Gutierrez JM, Bundalovic-Torma C, Groves HE, Isabel MR, et al. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. Sci Rep. 2020;10(1):14031. https://doi.org/10.1038/s41598-020-70827-z.
62. Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv. 2020. https://doi.org/10.1101/2020.06.12.148726.
63. Eskier D, Karakülah G, Suner A, Oktay Y. RdRp mutations are associated with SARS-CoV-2 genome evolution. PeerJ. 2020;8:e9587. https://doi.org/10.7717/peerj.9587.
64. Begum F, Mukherjee D, Das S, Thagriki D, Tripathi PP, Banerjee AK, et al. Specific mutations in SARS-CoV2 RNA dependent RNA polymerase and helicase alter protein structure, dynamics and thus function: effect on viral RNA replication. bioRxiv. 2020:2020.04.26.063024. https://doi.org/10.1101/2020.04.26.063024.
65. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med. 2020;18(1):179. https://doi.org/10.1186/s12967-020-02344-6.
66. Girardi E, López P, Pfeffer S. On the importance of host microRNAs during viral infection. Front Genet. 2018;9:439. https://doi.org/10.3389/fgene.2018.00439.
67. Liu Z, Zheng H, Lin H, Li M, Yuan R, Peng J, et al. Identification of common deletions in the spike protein of severe acute respiratory syndrome coronavirus 2. J Virol. 2020;94(17). https://doi.org/10.1128/jvi.00790-20.
68. Lau SY, Wang P, Mok BW, Zhang AJ, Chu H, Lee AC, et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. Emerg Microbes Infect. 2020;9(1):837–42. https://doi.org/10.1080/22221751.2020.1756700.
69. Andrés C, Garcia-Cehic D, Gregori J, Piñana M, Rodriguez-Frias F, Guerrero-Murillo M, et al. Naturally occurring SARS-CoV-2 gene deletions close to the spike S1/S2 cleavage site in the viral quasispecies of COVID19 patients. Emerg Microbes Infect. 2020;9(1):1900–11. https://doi.org/10.1080/22221751.2020.1806735.
70. Benedetti F, Snyder GA, Giovanetti M, Angeletti S, Gallo RC, Ciccozzi M, et al. Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. J Transl Med. 2020;18(1):329. https://doi.org/10.1186/s12967-020-02507-5.
71. Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, et al. Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. mBio. 2020;11(4). https://doi.org/10.1128/mBio.01610-20.
72. Young BE, Fong SW, Chan YH, Mak TM, Ang LW, Anderson DE, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet (London, England). 2020;396(10251):603–11. https://doi.org/10.1016/s0140-6736(20)31757-8.