Computational Inference of Selection Underlying the Evolution of the Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2

Rachele Cagliani,[a] Diego Forni,[a] Mario Clerici,[b,c] Manuela Sironi[a]

[a]Scientific Institute IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Italy
[b]Department of Physiopathology and Transplantation, University of Milan, Milan, Italy
[c]Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

Rachele Cagliani and Diego Forni contributed equally to this article. Author order was determined alphabetically.

ABSTRACT The novel coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that recently emerged in China is thought to have a bat origin, as its closest known relative (BatCoV RaTG13) was described previously in horseshoe bats. We analyzed the selective events that accompanied the divergence of SARS-CoV-2 from BatCoV RaTG13. To this end, we applied a population genetics-phylogenetics approach, which leverages within-population variation and divergence from an out-group. Results indicated that most sites in the viral open reading frames (ORFs) evolved under conditions of strong to moderate purifying selection. The most highly constrained sequences corresponded to some nonstructural proteins (nsps) and to the M protein. Conversely, nsp1 and accessory ORFs, particularly ORF8, had a non-negligible proportion of codons evolving under conditions of very weak purifying selection or close to selective neutrality. Overall, limited evidence of positive selection was detected. The 6 bona fide positively selected sites were located in the N protein, in ORF8, and in nsp1. A signal of positive selection was also detected in the receptor-binding motif (RBM) of the spike protein but most likely resulted from a recombination event that involved the BatCoV RaTG13 sequence. In line with previous data, we suggest that the common ancestor of SARS-CoV-2 and BatCoV RaTG13 encoded/encodes an RBM similar to that observed in SARS-CoV-2 itself and in some pangolin viruses. It is presently unknown whether the common ancestor still exists and, if so, which animals it infects. Our data, however, indicate that divergence of SARS-CoV-2 from BatCoV RaTG13 was accompanied by limited episodes of positive selection, suggesting that the common ancestor of the two viruses was poised for human infection.

IMPORTANCE Coronaviruses are dangerous zoonotic pathogens; in the last 2 decades, three coronaviruses have crossed the species barrier and caused human epidemics. One of these is the recently emerged SARS-CoV-2. We investigated how, since its divergence from a closely related bat virus, natural selection shaped the genome of SARS-CoV-2. We found that distinct coding regions in the SARS-CoV-2 genome evolved under conditions of different degrees of constraint and are consequently more or less prone to tolerate amino acid substitutions. In practical terms, the level of constraint provides indications about which proteins/protein regions are better suited as possible targets for the development of antivirals or vaccines. We also detected limited signals of positive selection in three viral ORFs. However, we warn that, in the absence of knowledge about the chain of events that determined the human spillover, these signals should not be necessarily interpreted as evidence of an adaptation to our species.

In December 2019, a human-infecting coronavirus, now referred to as coronavirus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (1), emerged in Wuhan, China, causing respiratory disease in a large number of people and being responsible for thousands of deaths (https://www.who.int/emergencies/diseases/novel-coronavirus-2019) (2). After SARS-CoV (severe acute respiratory syndrome coronavirus) and MERS-CoV (Middle East respiratory syndrome coronavirus), SARS-CoV-2 is the third coronavirus to cause a human epidemic in the last 2 decades (3, 4).

Coronaviruses (family *Coronaviridae*, order *Nidovirales*) have positive-sense, single-stranded RNA genomes which are unusually long and complex compared to those of other RNA viruses. Two-thirds of the coronavirus genome is occupied by two large overlapping open reading frames (ORFs), ORF1a and ORF1b, that are translated into the pp1a and pp1ab polyproteins. These are processed to generate 16 nonstructural proteins (nsp1 to nsp16) (5). The remaining portion of the genome includes ORFs for the structural proteins spike (S), envelope (E), membrane (M), and nucleoprotein (N), as well as a variable number of accessory proteins (3–5).

Several coronavirus genera and subgenera are recognized (https://talk.ictvonline.org/ictv-reports/) (1, 6, 7). Whereas MERS-CoV is a member of the *Merbecovirus* subgenus, phylogenetic analyses indicated that SARS-CoV-2 clusters with SARS-CoV and other bat-derived viruses in the *Sarbecovirus* subgenus (genus *Betacoronavirus*) (1, 8, 9). A recent report by the Coronavirus Study Group of the International Committee on Taxonomy of Viruses (ICTV) indicated that SARS-CoV-2 can be assigned to the species *Severe acute respiratory syndrome-related coronavirus* (1).

Bats host a large diversity of coronaviruses related to SARS-CoV (5, 10, 11), and, in general, these animals are believed to represent the original reservoir of several human-infecting coronaviruses (3, 4). This also seems to be the case for SARS-CoV-2, as analysis of the viral genome indicated that its closest known relative, with an average identity of ~96%, is a virus (BatCoV RaTG13) identified in horseshoe bats (*Rhinolophus affinis*) (8). Two other bat-derived coronaviruses (bat-SL-CoVZC45 and bat-SL-CoVZXC21) display high levels of similarity (>70%) to SARS-CoV-2, with various levels of identity along the genome (9, 12, 13). However, because both SARS-CoV and MERS-CoV were transmitted to humans via intermediate hosts (3, 4), it remains unclear whether the Wuhan epidemic was initiated by a spillover from bats or from other animals. Recent data suggested that viruses related to SARS-CoV-2 are found in pangolins (*Manis javanica*) (14–17), but the role of these animals in fueling the human epidemic remains unclear.

A major determinant of coronavirus host range is represented by the binding affinity between the spike protein and the cognate cellular receptor (18–22). Notably, this was previously shown to be the case for SARS-CoV, which, in analogy to SARS-CoV-2, uses ACE2 (angiotensin-converting enzyme 2) to enter host cells (8, 23). A limited number of amino acid changes in the receptor binding domain (RBD) of SARS-CoV were shown to modulate the binding efficiency to ACE2 from different mammalian species and to contribute to the adaptation of the virus to human cells (24–26). However, the SARS-CoV epidemic was characterized by another signature change in the viral genome; relatively early during the human-to-human transmission chain, SARS-CoV strains acquired a 29-nucleotide deletion which split ORF8, encoding an accessory protein, into two functional ORFs (27). Together with the observation that ORF8 is evolving quickly in SARS-CoV strains, this finding was taken to imply adaptation to our species (28). The evidence for adaptation was subsequently questioned, and recent data indicated that the 29-nucleotide deletion most likely represents a founder effect, which causes fitness loss irrespective of the host species (4, 29). These data underscore the relevance (and possible pitfalls) of evolutionary analyses in the study of viral species emergence and host shifts.

Here, we used available SARS-CoV-2 strains to describe the selective events that accompanied the divergence of this novel human pathogen from its closest known relative (BatCoV RaTG13) (8).

RESULTS AND DISCUSSION

As mentioned above, the closest relative (BatCoV RaTG13) of the novel human-infecting SARS-CoV-2 was identified in bats (8). It is presently unknown whether BatCoV RaTG13 can be transmitted in human populations and if it can infect human cells. Likewise, the reservoir and the animal host that fueled the human transmission of SARS-CoV-2 are presently uncertain. It is certain that ample data now indicate that human-to-human transmission has a role in spreading the SARS-CoV-2 epidemic (30–33) and that, in addition to humans, the virus can infect cells from bats, small carnivores, and pigs (8). We thus set out to determine the selective events that accompanied the divergence of the SARS-CoV-2 lineage from BatCoV RaTG13. In doing so, we do not imply that any such event was primarily responsible for human adaptation, as high efficiency of human infection might instead represent an incidental by-product of adaptation to another host.

Based on the alignment of 44 SARS-CoV-2 genomes and the BatCoV RaTG13 sequence, 147 amino acid replacements, unevenly distributed along the genome, were found to separate SARS-CoV-2 from its closest relative. A total of 41 amino acid changes are polymorphic in the SARS-CoV-2 population (Fig. 1A).

To investigate the selection patterns acting on SARS-CoV-2 genomes, we applied a method that combines analysis of within-population variation (i.e., variation among SARS-CoV-2 strains) and divergence from an outgroup (BatCoV RaTG13). Specifically, nucleotide alignments were analyzed using gammaMap (34), which estimates selection coefficients (γ) along coding regions and allows the detection of fine-scale differences in selective pressures at specific codons. In practical terms, γ values can be considered a measure of the fitness consequences of new nonsynonymous mutations. The method categorizes selection coefficients into 12 predefined classes ranging from −500 (inviable) to 100 (strongly beneficial). For gammaMap analysis, we divided the ORF1a and ORF1b alignments into the 16 nsps; because nsp3 is a long, multidomain protein, it was split into domains also. Likewise, the coronavirus S protein includes two functionally distinct units (S1 and S2), which were separately analyzed. Alignments of more than 80 codons were analyzed with gammaMap (Fig. 1A).

As previously shown for several other viruses (35–37), we found that most sites evolved under conditions of strong to moderate purifying selection (γ value less than −5). However, the strength of purifying selection varied depending on the region. The strongest constraints were observed for nsp6 to nsp10, for nsp16, and for the M ORF (Fig. 1B). Whereas nsp6 is involved in the formation of the reticulovesicular membrane network where viral RNA replication occurs, nsp7 to nsp10 are small proteins that function as cofactors for viral replicative enzymes, including nsp16, a 2'-O-methyl transferase (38). Conversely, the M ORF encodes a structural protein which is highly abundant in the virion of coronaviruses (39). The M protein interacts with other structural viral proteins and plays an important role in virion morphogenesis (40). Importantly, the M protein is a dominant immunogen for both the humoral and the cellular immune responses (41, 42). The latter features and its high level of constraint suggest that the M protein represents an excellent target for vaccine design.

Among the nonaccessory ORFs, the lowest levels of constraint were observed for nsp1 and the acidic domain of nsp3 (Fig. 1B and C). This is in line with data indicating that these regions are quickly evolving in coronaviruses at large (see below) (43, 44). Accessory ORFs, and, in particular, ORF8, had a nonnegligible proportion of codons evolving under conditions of very weak purifying selection or close to selective neutrality. On one hand, this is in line with the idea that genetic variation in accessory ORFs causes limited fitness consequences, as the above-mentioned case of SARS-CoV ORF8 indicates (4, 29). In fact, gains and losses of accessory proteins have been common during the evolutionary history of coronaviruses and accessory ORFs differ in
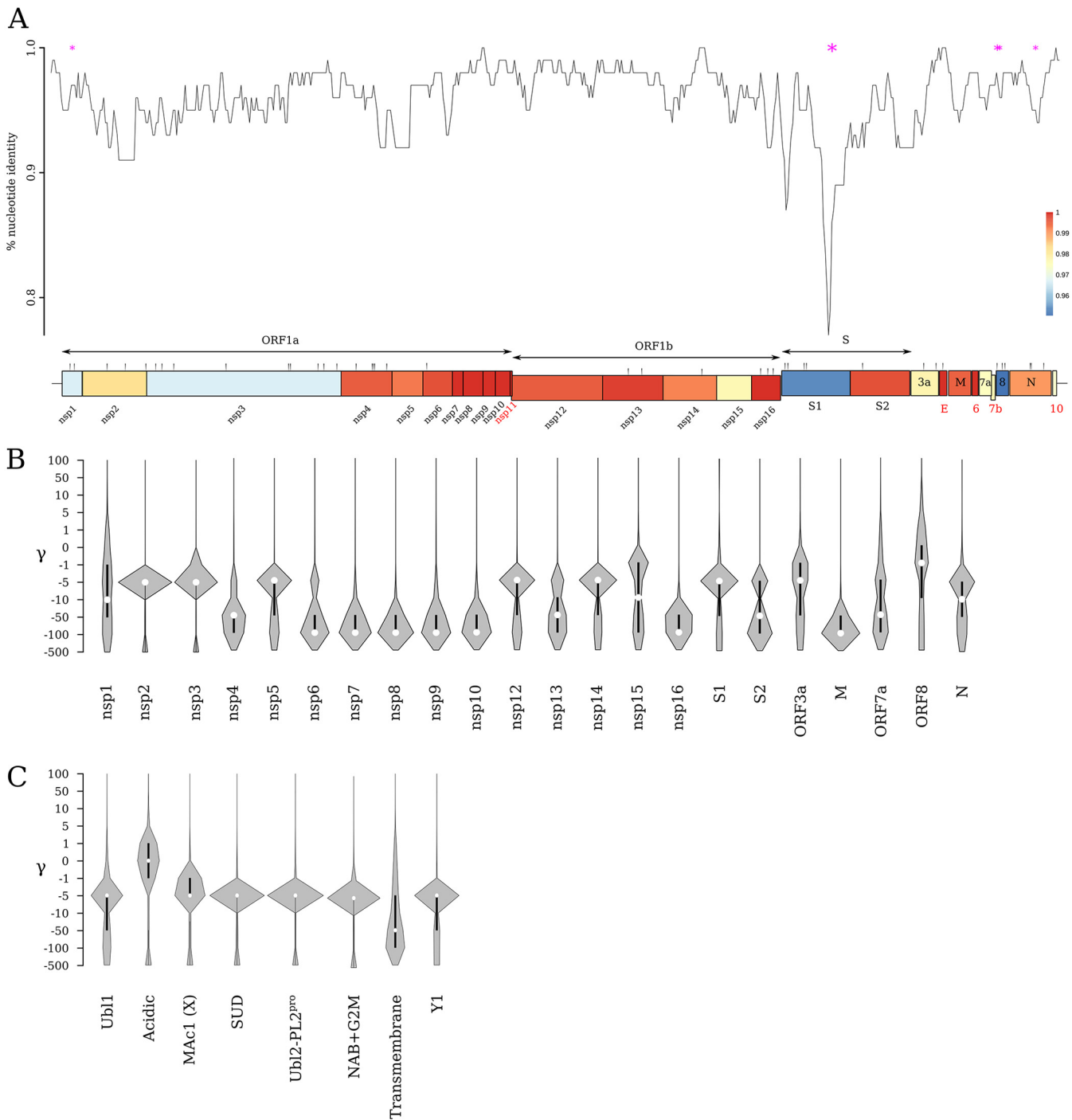
FIG 1 Selective patterns of SARS-CoV-2. (A) Similarity plot (generated with SimPlot) of BatCoV RaTG13 relative to SARS-CoV-2 (Wuhan-Hu-1 reference strain, NC_045512.2). Similarity (Kimura distance) was calculated within sliding windows of 250 bp moving with steps of 50 bp. A schematic representation of the SARS-CoV-2 genome is also shown. ORF and nsp (nonstructural protein) names, lengths, and relative positions are in accordance with the annotation for the reference Wuhan-Hu-1 sequence. Box colors indicate the level of amino acid identity between the SARS-CoV-2 and BatCoV RaTG13 sequences. Black triangles indicate amino acid changes that are polymorphic in the analyzed SARS-CoV-2 genomes. Asterisks denote positively selected sites, and their sizes are proportional to the number of selected sites/region. Short ORFs with names in red were not analyzed with gammaMap. (B and C) Violin plots (median, white dot; interquartile range, black bar) of selection coefficients (γ) for the longest (more that 80 codons) ORFs (B) and nsp3 subdomains (C) are shown. Nsp3 domains were retrieved from the SARS-CoV annotation (68).

number and sequence even among coronaviruses belonging to the same genus or subgenus (4). On the other hand, accessory proteins have often been shown to contribute to the modulation of immune responses, as well as to virulence (3, 4). It is thus conceivable that their limited constraint maintains variability in coronavirus
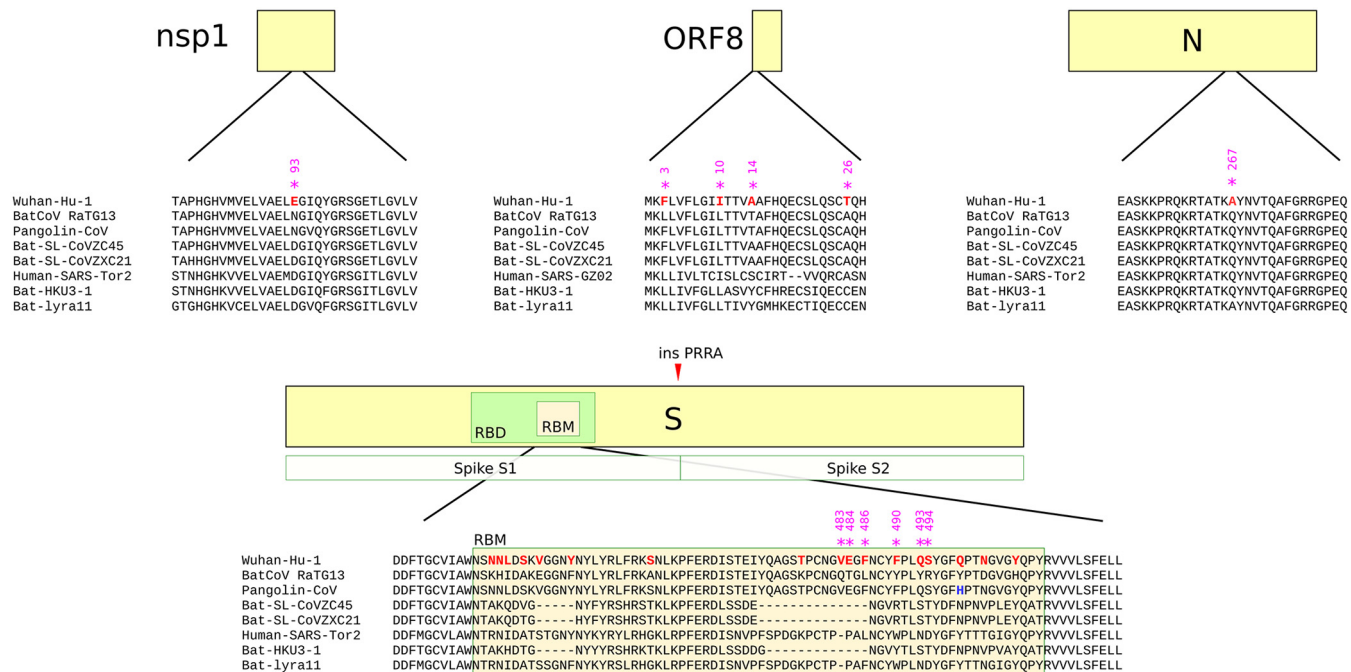
FIG 2 SARS-CoV-2 positively selected sites. A schematic representation of the nsp1, ORF8, spike (S), and nucleocapsid (N) proteins is presented. Positively selected sites (magenta) and amino acid substitutions between SARS-CoV-2 and BatCoV RaTG13 (red) and between SARS-CoV-2 and pangolin-CoV MP789 (blue) are indicated in the alignments. The location of an insertion (insPRRA) in the spike glycoprotein is also shown. This insertion is predicted to occur in the S1/S2 furin-like cleavage site (69, 70).

accessory ORFs, eventually facilitating rapid adaptation when the environment (e.g., host) changes.

We next wished to determine whether positive selection at specific sites also drove the evolution of SARS-CoV-2. We thus estimated codon-wise posterior probabilities for each selection coefficient. Very strong evidence (defined as a posterior probability > 0.80 of $\gamma \geq 1$) of positive selection was detected for seven sites, including six in the S1 region of the spike protein and one in N (Fig. 2). When the posterior probability cutoff was lowered to a less stringent value of 0.50, five additional sites in ORF8 ($n = 4$) and in nsp1 ($n = 1$) were identified (Fig. 2). It should be noted that this P value cutoff represents reasonably strong evidence of positive selection. Using these criteria, positively selected sites were estimated to account for the 0.12% of analyzed codons seen using 0.5 as the cutoff (0.07% for a 0.8 cutoff) (34, 45, 46).

The S1 region contains the RBD, and the crystal structure of the SARS-CoV S protein in complex with human ACE2 showed that, in turn, the RBD is formed by two subdomains, a core structure and the receptor-binding motif (RBM, which directly contacts ACE2) (47, 48). The S2 region includes the fusion machinery (49). We performed homology modeling of the SARS-CoV-2 S protein onto the SARS-CoV structure, and we analyzed the distribution of selection coefficients (Fig. 3A). The S2 subunit was characterized by stronger constraint than the S1 portion, and five of six putative positively selected sites were found to be located in the RBM, at the binding interface with ACE2 (Fig. 3A).

Comparing SARS-CoV-2 and BatCoV RaTG13, the RBM stands out as the single most divergent region (Fig. 1A) (8, 16). Very recent evidence indicated that, although the average level of genome similarity is lower than that seen with BatCoV RaTG13, coronaviruses isolated from pangolins have RBMs almost identical to that of SARS-CoV (14–17). This clearly implies that recombination might have inflated the estimation of positive selection in the S1 region. A pangolin virus available in GenBank (isolate MP789) has an RBM with high identity to SARS-CoV-2. Thus, using
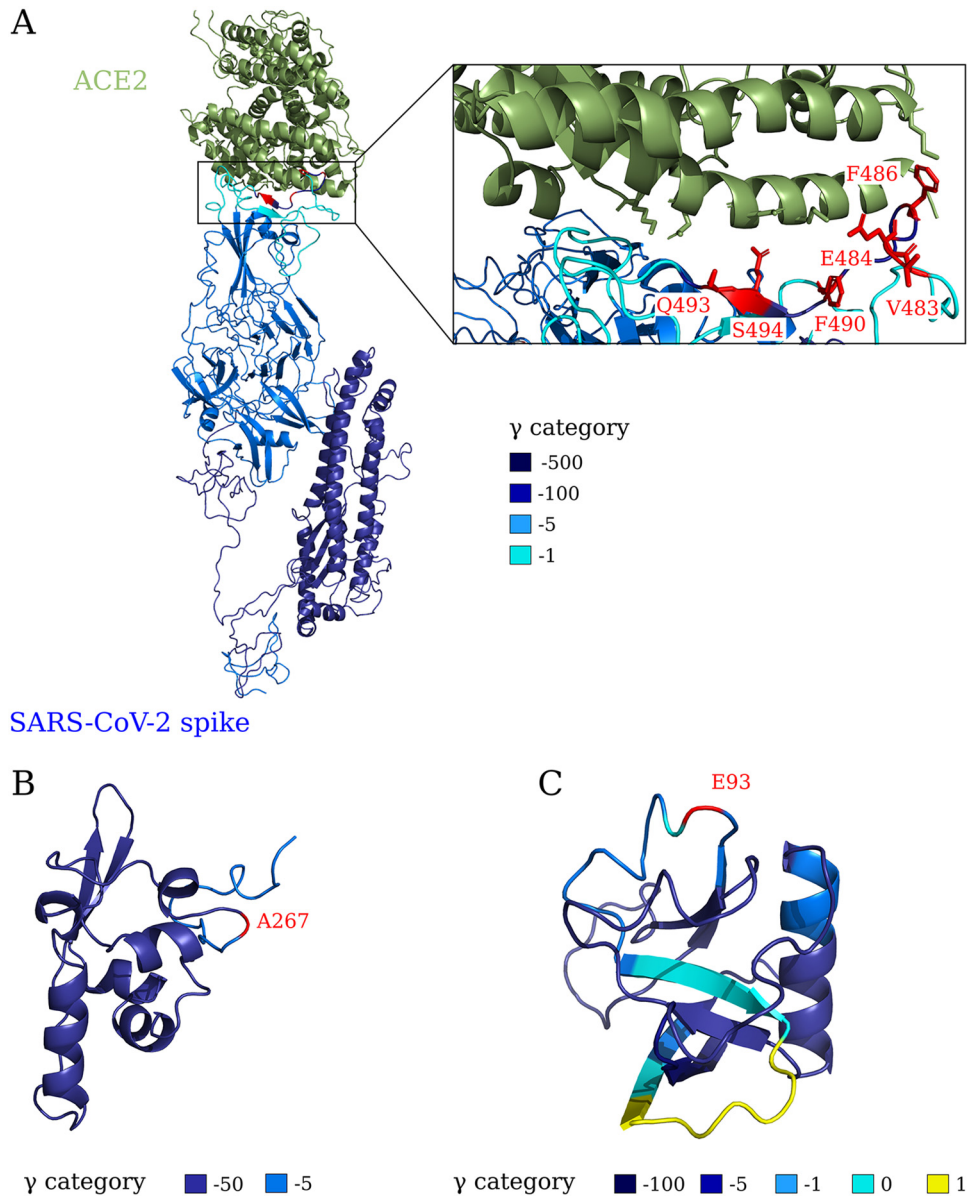
FIG 3 Homology modeling of positively selected SARS-CoV-2 proteins. Selected sites are mapped onto the 3D structure of models obtained using SARS-CoV proteins as a templates (PDB ID: 6ACG for panel A, 2CJR for panel B, 2HSX for panel C). Coronavirus proteins are colored in hues of blue based on the most likely selection coefficient. Positively selected sites are marked in red. (A) Ribbon representation of the spike glycoprotein model (one monomer is shown) in complex with human ACE2 (green) (48). The binding interface is shown in the enlargement. (B) Ribbon representation of the C-terminal domain of the nucleocapsid protein. (C) Ribbon representation of the N-terminal portion of nsp1. Note that although some sites had the highest posterior probability for $\gamma = 1$ (yellow), they were not called as positively selected because the 0.5 threshold was not reached.

the genome sequences of isolate MP789, SARS-CoV-2, and BatCoV RaTG13, we searched for recombination events using RDP4 (50). No evidence of recombination was detected, but that finding might have been due to the fact that the parental sequence with which BatCoV RaTG13 recombined is presently unsampled. We thus analyzed synonymous substitutions in the RBM alignment for these viruses, and we found that 41% ($n = 37$) of such substitutions are shared between SARS-CoV-2 and isolate MP789, whereas only 27% ($n = 10$) are shared between SARS-CoV-2 and BatCoV RaTG13. Overall, these findings strongly suggest that recombination rather than positive selection shaped the genetic diversity at the RBM, as previously

suggested (16). Recombination is known to affect evolutionary inference (51). In this case, because we used the BatCoV RaTG13 as an outgroup, the spurious signals were generated by considering the selected sites to represent amino acid replacements that arose and became fixed in the SARS-CoV-2 population, whereas they might represent changes that occurred in the outgroup through recombination. We consider that this was not the case for the other signals that we detected, as all of them were located in regions of high overall similarity between BatCoV RaTG13 and SARS-CoV-2, indicating no evidence of recombination (Fig. 1A).

The positively selected site (A267) in the nucleocapsid protein is located in the C-terminal domain. Homology modeling using the SARS-CoV N protein as a template indicated that A267 is located on an exposed loop on the protein surface (Fig. 3B) (52). The N protein is the most abundant protein in coronavirus-infected cells (53, 54). Its primary function is to package the viral genome into a ribonucleoprotein complex. In addition, the N protein performs nonstructural functions, as it regulates the host cell cycle and the stress response, it acts as a molecular chaperone, and it interferes with the host immune response (53, 54). Because these activities are mediated by interactions with different cellular proteins, the positively selected site might be evolving to establish, maintain, or avoid the binding of different host molecules.

Another positively selected site was detected in the nsp1 region, which also displayed relatively weak selective constraint. In SARS-CoV and other betacoronaviruses, nsp1 is a virulence factor and is essential for viral replication at least in the presence of an intact host interferon (IFN) response (55–57). Despite their relevant role for viral fitness *in vivo*, nsp1 proteins tend to be variable in sequence both within and among coronavirus genera. Detailed analysis of SARS-CoV nsp1 indicated that the protein plays multiple roles during viral infection, including inhibition of host protein synthesis, antagonism of IFN responses, modulation of the calcineurin/NFAT (nuclear factor of activated T cells) pathway, and induction of chemokine secretion (43). Homology modeling using the SARS-CoV nsp1 structure indicated that the positively selected site (E93) is exposed on the protein surface (Fig. 3C). Extensive mutagenesis of SARS-CoV nsp1 showed that exposed charged residues, including the positively selected site, mediate inhibition of gene expression and antiviral signaling (58). Moreover, the N-terminal half of SARS-CoV nsp1 interacts with immunophilins and calcipressins to modulate the calcineurin/NFAT pathway (59). Overall, these observations suggest that the diversity of coronavirus nsp1 proteins is driven by the need to establish interactions with multiple cellular partners and to evade immune surveillance. This is also likely to explain the positive selection signal that we detected. In general, a better understanding of the evolutionary constraints and forces acting on coronavirus nsp1 proteins may be extremely relevant, as the generation of viruses carrying nsp1 mutations was previously reported to be regarded as a potential strategy to generate attenuated vaccine strains (57, 60), and inhibitors of cyclophilins were previously reported to be potential antivirals for coronavirus treatment (59).

Finally, all of the selected sites that we identified in ORF8 (F3, I10, A14, are T26) are located in the N-terminal portion of the protein (Fig. 2). The SARS-CoV-2 ORF8 protein displays 30% identity to the intact ORF8 from the SARS-CoV GZ02 stain. It is presently unclear whether the SARS-CoV ORF8 N terminus is cleaved as a signal peptide or inserted into the endoplasmic reticulum membrane (61, 62). Using computational methods to predict signal peptides and transmembrane helices, we found evidence for both in the case of the N terminus of SARS-CoV-2 ORF8 (not shown). Clearly, experimental analyses will be required to determine the function of the N-terminal region of ORF8 and, more generally, the relevance of the selected sites for virus fitness or pathogenicity.

Overall, our analyses indicate that distinct coding regions in the SARS-CoV-2 genome evolve under conditions of different degrees of constraint and are consequently more or less prone to tolerate amino acid substitutions. In practical terms, the level of

constraint can provide indications concerning which specific proteins or protein regions are better suited to being possible targets for the development of antivirals or vaccines. Conversely, the current available knowledge and the analyses reported here allow no inference on the selective events (or lack thereof) that turned SARS-CoV-2 into a human pathogen. Recent analyses paid much attention to changes in the RBM. This is indeed expected to represent a major determinant of host range, and its sequence is highly variable among SARS-CoV-related viruses (as is also evident in the data presented in Fig. 2). Albeit preliminary and necessarily limited to currently sampled genomes, our analyses suggest that recombination had a role in shaping the diversity of the RBMs in these viruses. Our data also indicate that divergence of SARS-CoV-2 from BatCoV RaTG13 was accompanied by limited episodes of positive selection, suggesting that the common ancestor of the two viruses was poised for human infection. We also emphasize that lack of knowledge about the reservoir host and the chain of events that determined the human spillover prevent us from drawing any conclusion on the selective pressure underlying the limited positive selection events tha twe detected. These will need to be interpreted in the future, by incorporating epidemiological, biochemical, and additional genetic data.

Clearly, a caveat of our analyses lies in the quality and paucity of SARS-CoV-2 genomes, as well as in the limited availability of genomes of other coronaviruses closely related to SARS-CoV-2. Available sequences were obtained using different methods and most likely contain errors. This is unlikely to strongly affect inference of positive selection, as the frequency of all selected sites is high in the SARS-CoV-2 population. Also, the SARS-CoV-2 sequences that we analyzed display limited diversity (with only 41 nonsynonymous polymorphisms, most of them present in one or a few sequences). Thus, although the availability of additional genomes may increase the power to detect selective events and the confidence with which evolutionary patterns are inferred, simply increasing the number of genomes is unlikely to change the bulk of our results. However, sustained viral spread in the human population will necessarily introduce new mutations in the viral population. Thus, data reported here can depict only the situation of the early phases of the human epidemic. Follow-up analyses of the SARS-CoV-2 population will be required to determine the evolutionary trajectories of new mutations and to assess whether and how they affect viral fitness in the human hots.

MATERIALS AND METHODS

Sequences and alignments. Genome sequences were retrieved from the National Center for Biotechnology Information database (NCBI; https://www.ncbi.nlm.nih.gov/). Only complete or almost-complete genome sequences were included in the analysis (Table 1).

Alignments were generated using MAFFT (63), setting sequence type as codons.

Population genetics—phylogenetic analysis. Analyses were performed with gammaMap, which uses intraspecies variation and interspecies diversity to estimate, along coding regions, the distribution of selection coefficients (γ). In this framework, γ is defined as $2PN_es$, where P is the ploidy, N_e is effective population size, and s is the fitness advantage of any amino acid-replacing derived allele (34).

For the eight longest ORFs in the SARS-CoV-2 genome, the corresponding coding sequence of BatCoV RaTG13 was used as the outgroup.

We assumed θ (neutral mutation rate per site), k (transitions/transversions ratio), and T (branch length) to vary within genes following log-normal distributions, whereas p (probability of adjacent codons to share the same selection coefficient) was assumed to follow a log-uniform distribution. For each ORF, we set the neutral frequencies of non-STOP codons (1/61). For selection coefficients, we considered a uniform Dirichlet distribution with the same prior weight for each selection class. For each ORF, we performed 2 runs with 100,000 iterations each and with a thinning interval of 10 iterations. Runs were merged after checking for convergence.

The similarity plot was computed using a Kimura (two-parameter) distance model with SimPlot version 3.5.1 (64). The strip gap option was set at the 50% default value. Similarity scores were calculated in sliding windows of 250 bp moving with a step of 50 bp.

Protein 3D structures and homology modeling. The three-dimensional (3D) structures of SARS-CoV N (PDB identifier [ID]: 2CJR) (65) and S (PDB ID: 6ACG) (48) proteins were obtained from the Protein Data Bank (PDB).

Homology modeling analysis was performed through the SWISS-MODEL server (66). The accuracy of the models was examined through the GMQE (Global Model Quality Estimation) and QMEAN (Qualitative Model Energy ANalysis) scores (67).

TABLE 1 List of analyzed strains

Strain name	GenBank ID
Wuhan-Hu-1	NC_045512.2
2019-nCoV WHU01	MN988668.1
2019-nCoV WHU02	MN988669.1
2019-nCoV_HKU-SZ-005b_2020	MN975262.1
2019-nCoV_HKU-SZ-002a_2020	MN938384.1
SARS-CoV-2/WH-09/human/2020/CHN	MT093631.1
SARS-CoV-2/IQTC01/human/2020/CHN	MT123290.1
HZ-1	MT039873.1
BetaCoV/Wuhan/IPBCAMS-WH-01/2019	MT019529.1
BetaCoV/Wuhan/IPBCAMS-WH-03/2019	MT019531.1
BetaCoV/Wuhan/IPBCAMS-WH-02/2019	MT019530.1
BetaCoV/Wuhan/IPBCAMS-WH-04/2019	MT019532.1
BetaCoV/Wuhan/IPBCAMS-WH-05/2020	MT019533.1
WIV02	MN996527.1
WIV04	MN996528.1
WIV05	MN996529.1
WIV06	MN996530.1
WIV07	MN996531.1
SARS-CoV-2/Yunnan-01/human/2020/CHN	MT049951.1
nCoV-FIN-29-Jan-2020	MT020781.1
SARS0CoV-2/61-TW/human/2020/NPL	MT072688.1
SNU01	MT039890.1
SARS-CoV-2/01/human/2020/SWE	MT093571.1
SARS-CoV-2/NTU01/2020/TWN	MT066175.1
SARS-CoV-2/NTU02/2020/TWN	MT066176.1
2019-nCoV/USA-WA1/2020	MN985325.1
2019-nCoV/USA-AZ1/2020	MN997409.1
2019-nCoV/USA-CA1/2020	MN994467.1
2019-nCoV/USA-CA2/2020	MN994468.1
2019-nCoV/USA-CA3/2020	MT027062.1
2019-nCoV/USA-CA4/2020	MT027063.1
2019-nCoV/USA-CA5/2020	MT027064.1
2019-nCoV/USA-CA6/2020	MT044258.1
2019-nCoV/USA-CA7/2020	MT106052.1
2019-nCoV/USA-CA8/2020	MT106053.1
2019-nCoV/USA-CA9/2020	MT118835.1
2019-nCoV/USA-IL2/2020	MT044257.1
2019-nCoV/USA-IL1/2020	MN988713.1
2019-nCoV/USA-MA1/2020	MT039888.1
2019-nCoV/USA-TX1/2020	MT106054.1
2019-nCoV/USA-WA1-A12/2020	MT020880.1
2019-nCoV/USA-WA1-F6/2020	MT020881.1
2019-nCoV/USA-WI1/2020	MT039887.1
Australia/VIC01/2020	MT007544.1
Bat coronavirus RaTG13	MN996532.1
Pangolin coronavirus isolate MP789	MT084071.1
Bat SARS-like coronavirus isolate bat-SL-CoVZC45	MG772933.1
Bat SARS-like coronavirus isolate bat-SL-CoVZXC21	MG772934.1
SARS-CoV tor2	NC_004718.3
SARS-CoV GZ02	AY390556.1
Bat SARS coronavirus HKU3-1	DQ022305.2
Rhinolophus affinis coronavirus isolate LYRa11	KF569996.1

3D structures were rendered using PyMOL (The PyMOL Molecular Graphics System, Version 1.8.4.0; Schrödinger, LLC).

REFERENCES

1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. 2020. The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5:536–544. https://doi.org/10.1038/s41564-020-0695-z.

2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W; China Novel Coronavirus Investigating and Research Team. 2020. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017.

3. Cui J, Li F, Shi ZL. 2019. Origin and evolution of pathogenic coronavi-

ruses. Nat Rev Microbiol 17:181–192. https://doi.org/10.1038/s41579-018-0118-9.

4. Forni D, Cagliani R, Clerici M, Sironi M. 2017. Molecular evolution of human coronavirus genomes. Trends Microbiol 25:35–48. https://doi.org/10.1016/j.tim.2016.09.001.

5. Luk HKH, Li X, Fung J, Lau SKP, Woo PCY. 2019. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. Infect Genet Evol 71:21–30. https://doi.org/10.1016/j.meegid.2019.03.001.

6. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, Galiano M, Gorbalenya AE, Memish ZA, Perlman S, Poon LL, Snijder EJ, Stephens GM, Woo PC, Zaki AM, Zambon M, Ziebuhr J. 2013. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. J Virol 87:7790–7792. https://doi.org/10.1128/JVI.01244-13.

7. Gorbalenya AE, Snijder EJ, Spaan WJ. 2004. Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. J Virol 78:7863–7866. https://doi.org/10.1128/JVI.78.15.7863-7866.2004.

8. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273. https://doi.org/10.1038/s41586-020-2012-7.

9. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in China. Nature 579:265–269. https://doi.org/10.1038/s41586-020-2008-3.

10. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, Xie JZ, Shen XR, Zhang YZ, Wang N, Luo DS, Zheng XS, Wang MN, Daszak P, Wang LF, Cui J, Shi ZL. 2017. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog 13:e1006698. https://doi.org/10.1371/journal.ppat.1006698.

11. Wang L, Fu S, Cao Y, Zhang H, Feng Y, Yang W, Nie K, Ma X, Liang G. 2017. Discovery and genetic analysis of novel coronaviruses in least horseshoe bats in southwestern China. Emerg Microbes Infect 6:e14–e18. https://doi.org/10.1038/emi.2016.140.

12. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. 2020. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395:565–574. https://doi.org/10.1016/S0140-6736(20)30251-8.

13. Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. 2020. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect Genet Evol 79:104212. https://doi.org/10.1016/j.meegid.2020.104212.

14. Lam TT, Shum MH, Zhu H, Tong Y, Ni X, Liao Y, Wei W, Cheung WY, Li W, Li L, Leung GM, Holmes EC, Hu Y, Guan Y. 2020. Identification of 2019-nCoV related coronaviruses in Malayan pangolins in southern China. Biorxiv https://doi.org/10.1101/2020.02.13.945485.

15. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou JJ, Li N, Guo Y, Li X, Shen X, Zhang Z, Shu F, Huang W, Li Y, Zhang Z, Chen R, Wu Y, Peng S, Huang M, Xie W, Cai Q, Hou F, Liu Y, Chen W, Xiao L, Shen Y. 2020. Isolation and characterization of 2019-nCoV-like coronavirus from Malayan pangolins. Biorxiv https://doi.org/10.1101/2020.02.17.951335.

16. Wong MC, Javornik Cregeen SJ, Ajami NJ, Petrosino JF. 2020. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. Biorxiv https://doi.org/10.1101/2020.02.07.939207.

17. Liu P, Jiang J, Wan X, Hua Y, Wang X, Hou F, Chen J, Zou J, Chen J. 2020. Are pangolins the intermediate host of the 2019 novel coronavirus (2019-nCoV)? Biorxiv https://doi.org/10.1101/2020.02.18.954628.

18. Haijema BJ, Volders H, Rottier PJ. 2003. Switching species tropism: an effective way to manipulate the feline coronavirus genome. J Virol 77:4528–4538. https://doi.org/10.1128/JVI.77.8.4528-4538.2003.

19. Kuo L, Godeke GJ, Raamsman MJ, Masters PS, Rottier PJ. 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. J Virol 74:1393–1406. https://doi.org/10.1128/JVI.74.3.1393-1406.2000.

20. McCray PB, Jr, Pewe L, Wohlford-Lenane C, Hickey M, Manzel L, Shi L, Netland J, Jia HP, Halabi C, Sigmund CD, Meyerholz DK, Kirby P, Look DC, Perlman S. 2007. Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. J Virol 81:813–821. https://doi.org/10.1128/JVI.02012-06.

21. Moore MJ, Dorfman T, Li W, Wong SK, Li Y, Kuhn JH, Coderre J, Vasilieva N, Han Z, Greenough TC, Farzan M, Choe H. 2004. Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2. J Virol 78:10628–10635. https://doi.org/10.1128/JVI.78.19.10628-10635.2004.

22. Schickli JH, Thackray LB, Sawicki SG, Holmes KV. 2004. The N-terminal region of the murine coronavirus spike glycoprotein is associated with the extended host range of viruses from persistently infected murine cells. J Virol 78:9073–9083. https://doi.org/10.1128/JVI.78.17.9073-9083.2004.

23. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M. 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426:450–454. https://doi.org/10.1038/nature02145.

24. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong SK, Huang IC, Xu K, Vasilieva N, Murakami A, He Y, Marasco WA, Guan Y, Choe H, Farzan M. 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J 24:1634–1643. https://doi.org/10.1038/sj.emboj.7600640.

25. Wu K, Peng G, Wilken M, Geraghty RJ, Li F. 2012. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. J Biol Chem 287:8904–8911. https://doi.org/10.1074/jbc.M111.325803.

26. Qu XX, Hao P, Song XJ, Jiang SM, Liu YX, Wang PG, Rao X, Song HD, Wang SY, Zuo Y, Zheng AH, Luo M, Wang HL, Deng F, Wang HZ, Hu ZH, Ding MX, Zhao GP, Deng HK. 2005. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. J Biol Chem 280:29588–29595. https://doi.org/10.1074/jbc.M500662200.

27. Chinese SARS Molecular Epidemiology Consortium. 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303:1666–1669. https://doi.org/10.1126/science.1092002.

28. Lau SK, Feng Y, Chen H, Luk HK, Yang WH, Li KS, Zhang YZ, Huang Y, Song ZZ, Chow WN, Fan RY, Ahmed SS, Yeung HC, Lam CS, Cai JP, Wong SS, Chan JF, Yuen KY, Zhang HL, Woo PC. 2015. Severe acute respiratory syndrome (SARS) coronavirus ORF8 protein is acquired from SARS-related coronavirus from greater horseshoe bats through recombination. J Virol 89:10532–10547. https://doi.org/10.1128/JVI.01048-15.

29. Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, Gloza-Rausch F, Balboni A, Battilani M, Rihtaric D, Toplak I, Ameneiros RS, Pfeifer A, Thiel V, Drexler JF, Muller MA, Drosten C. 2018. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. Sci Rep 8:15177. https://doi.org/10.1038/s41598-018-33487-8.

30. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW, Lo SK, Chan KH, Poon VK, Chan WM, Ip JD, Cai JP, Cheng VC, Chen H, Hui CK, Yuen KY. 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395:514–523. https://doi.org/10.1016/S0140-6736(20)30154-9.

31. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Li M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JTK, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z. 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med https://doi.org/10.1056/NEJMoa2001316.

32. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, Nguyen TT, Cao TM, Pham QD. 2020. Importation and human-to-human transmission of a novel coronavirus in Vietnam. N Engl J Med 382:872–874. https://doi.org/10.1056/NEJMc2001272.

33. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Pastore Y, Piontti A, Mu K, Rossi L, Sun K, Viboud C, Xiong X, Yu H, Halloran ME, Longini IM, Jr, Vespignani A. 6 March 2020, posting date. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science https://doi.org/10.1126/science.aba9757.

34. Wilson DJ, Hernandez RD, Andolfatto P, Przeworski M. 2011. A population genetics-phylogenetics approach to inferring natural selection in coding sequences. PLoS Genet 7:e1002395. https://doi.org/10.1371/journal.pgen.1002395.

35. Ho SY, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper

A. 2011. Time-dependent rates of molecular evolution. Mol Ecol 20:3087–3101. https://doi.org/10.1111/j.1365-294X.2011.05178.x.

36. Wertheim JO, Kosakovsky Pond SL. 2011. Purifying selection can obscure the ancient age of viral lineages. Mol Biol Evol 28:3355–3365. https://doi.org/10.1093/molbev/msr170.

37. Wertheim JO, Chu DK, Peiris JS, Kosakovsky Pond SL, Poon LL. 2013. A case for the ancient origin of coronaviruses. J Virol 87:7039–7045. https://doi.org/10.1128/JVI.03273-12.

38. Snijder EJ, Decroly E, Ziebuhr J. 2016. The nonstructural proteins directing coronavirus RNA synthesis and processing. Adv Virus Res 96:59–126. https://doi.org/10.1016/bs.aivir.2016.08.008.

39. Armstrong J, Niemann H, Smeekens S, Rottier P, Warren G. 1984. Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus. Nature 308:751–752. https://doi.org/10.1038/308751a0.

40. Siu YL, Teoh KT, Lo J, Chan CM, Kien F, Escriou N, Tsao SW, Nicholls JM, Altmeyer R, Peiris JSM, Bruzzone R, Nal B. 2008. The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. J Virol 82:11318–11330. https://doi.org/10.1128/JVI.01052-08.

41. Liu J, Sun Y, Qi J, Chu F, Wu H, Gao F, Li T, Yan J, Gao GF. 2010. The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. J Infect Dis 202:1171–1180. https://doi.org/10.1086/656315.

42. Pang H, Liu Y, Han X, Xu Y, Jiang F, Wu D, Kong X, Bartlam M, Rao Z. 2004. Protective humoral responses to severe acute respiratory syndrome-associated coronavirus: implications for the design of an effective protein-based vaccine. J Gen Virol 85:3109–3113. https://doi.org/10.1099/vir.0.80111-0.

43. Narayanan K, Ramirez SI, Lokugamage KG, Makino S. 2015. Coronavirus nonstructural protein 1: common and distinct functions in the regulation of host and viral gene expression. Virus Res 202:89–100. https://doi.org/10.1016/j.virusres.2014.11.019.

44. Neuman BW. 2016. Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. Antiviral Res 135:97–107. https://doi.org/10.1016/j.antiviral.2016.10.005.

45. Brand CL, Cattani MV, Kingan SB, Landeen EL, Presgraves DC. 2018. Molecular evolution at a meiosis gene mediates species differences in the rate and patterning of recombination. Curr Biol 28:1289–1295.e4. https://doi.org/10.1016/j.cub.2018.02.056.

46. Hemmer LW, Blumenstiel JP. 2016. Holding it together: rapid evolution and positive selection in the synaptonemal complex of Drosophila. BMC Evol Biol 16:91. https://doi.org/10.1186/s12862-016-0670-8.

47. Li F, Li W, Farzan M, Harrison SC. 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309:1864–1868. https://doi.org/10.1126/science.1116480.

48. Song W, Gui M, Wang X, Xiang Y. 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog 14:e1007236. https://doi.org/10.1371/journal.ppat.1007236.

49. Graham RL, Baric RS. 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J Virol 84:3134–3146. https://doi.org/10.1128/JVI.01394-09.

50. Martin DP, Murrell B, Khoosal A, Muhire B. 2017. Detecting and analyzing genetic recombination using RDP4. Methods Mol Biol 1525:433–460. https://doi.org/10.1007/978-1-4939-6622-6_17.

51. Martin DP, Lemey P, Posada D. 2011. Analysing recombination in nucleotide sequences. Mol Ecol Resour 11:943–955. https://doi.org/10.1111/j.1755-0998.2011.03026.x.

52. Takeda M, Chang CK, Ikeya T, Guntert P, Chang YH, Hsu YL, Huang TH, Kainosho M. 2008. Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. J Mol Biol 380:608–622. https://doi.org/10.1016/j.jmb.2007.11.093.

53. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. 2014. The SARS coronavirus nucleocapsid protein–forms and functions. Antiviral Res 103:39–50. https://doi.org/10.1016/j.antiviral.2013.12.009.

54. Surjit M, Lal SK. 2008. The SARS-CoV nucleocapsid protein: a protein with

55. Wathelet MG, Orr M, Frieman MB, Baric RS. 2007. Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. J Virol 81:11620–11633. https://doi.org/10.1128/JVI.00702-07.

56. Brockway SM, Denison MR. 2005. Mutagenesis of the murine hepatitis virus nsp1-coding region identifies residues important for protein processing, viral RNA synthesis, and viral replication. Virology 340:209–223. https://doi.org/10.1016/j.virol.2005.06.035.

57. Zust R, Cervantes-Barragan L, Kuri T, Blakqori G, Weber F, Ludewig B, Thiel V. 2007. Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. PLoS Pathog 3:e109. https://doi.org/10.1371/journal.ppat.0030109.

58. Jauregui AR, Savalia D, Lowry VK, Farrell CM, Wathelet MG. 2013. Identification of residues of SARS-CoV nsp1 that differentially affect inhibition of gene expression and antiviral signaling. PLoS One 8:e62416. https://doi.org/10.1371/journal.pone.0062416.

59. Pfefferle S, Schöpf J, Kögl M, Friedel CC, Müller MA, Carbajo-Lozoya J, Stellberger T, von Dall'Armi E, Herzog P, Kallies S, Niemeyer D, Ditt V, Kuri T, Züst R, Pumpor K, Hilgenfeld R, Schwarz F, Zimmer R, Steffen I, Weber F, Thiel V, Herrler G, Thiel H-J, Schwegmann-Weßels C, Pöhlmann S, Haas J, Drosten C, von Brunn A. 2011. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. PLoS Pathog 7:e1002331. https://doi.org/10.1371/journal.ppat.1002331.

60. Jimenez-Guardeño JM, Regla-Nava JA, Nieto-Torres JL, DeDiego ML, Castaño-Rodriguez C, Fernandez-Delgado R, Perlman S, Enjuanes L. 2015. Identification of the mechanisms causing reversion to virulence in an attenuated SARS-CoV for the design of a genetically stable vaccine. PLoS Pathog 11:e1005215. https://doi.org/10.1371/journal.ppat.1005215.

61. Oostra M, de Haan CA, Rottier PJ. 2007. The 29-nucleotide deletion present in human but not in animal severe acute respiratory syndrome coronaviruses disrupts the functional expression of open reading frame 8. J Virol 81:13876–13888. https://doi.org/10.1128/JVI.01631-07.

62. Sung SC, Chao CY, Jeng KS, Yang JY, Lai MM. 2009. The 8ab protein of SARS-CoV is a luminal ER membrane-associated protein and induces the activation of ATF6. Virology 387:402–413. https://doi.org/10.1016/j.virol.2009.02.021.

63. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.

64. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol 73:152–160. https://doi.org/10.1128/JVI.73.1.152-160.1999.

65. Chen C, Chang C, Chang Y, Sue S, Bai H, Riang L, Hsiao C, Huang T. 2007. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. J Mol Biol 368:1075–1086. https://doi.org/10.1016/j.jmb.2007.02.069.

66. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258. https://doi.org/10.1093/nar/gku340.

67. Benkert P, Biasini M, Schwede T. 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27:343–350. https://doi.org/10.1093/bioinformatics/btq662.

68. Lei J, Kusov Y, Hilgenfeld R. 2018. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. Antiviral Res 149:58–74. https://doi.org/10.1016/j.antiviral.2017.11.001.

69. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 9:221–236. https://doi.org/10.1080/22221751.2020.1719902.

70. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. 2020. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res 176:104742. https://doi.org/10.1016/j.antiviral.2020.104742.