Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

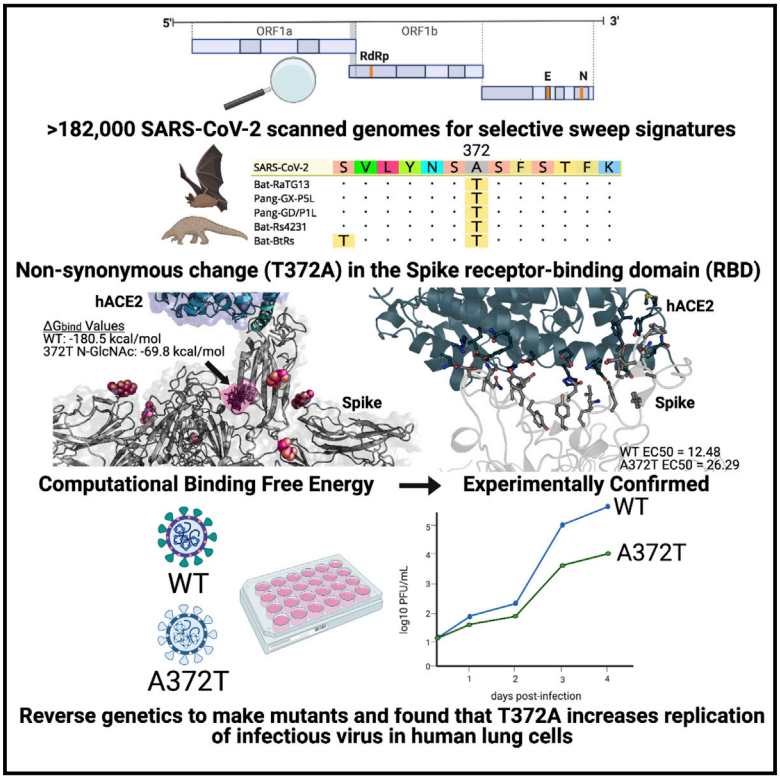
Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Cell

Article

A selective sweep in the *Spike* gene has driven SARS-CoV-2 human adaptation

Graphical abstract



Highlights

- Over 182,000 SARS-CoV-2 genomes were screened for selective sweep signatures
- An adaptive change within the spike protein receptor-binding domain was identified
- This change was predicted and experimentally confirmed to increase affinity to hACE2
- As a result, viral replication is enhanced relative to the putative ancestral variant

Authors

Lin Kang, Guijuan He, Amanda K. Sharp, Xiaofeng Wang, Anne M. Brown, Pawel Michalak, James Weger-Lucarelli

Correspondence

pmichalak@vcom.edu (P.M.), weger@vt.edu (J.W.-L.)

In brief

A non-synonymous change (T372A) within the spike protein RBD of human SARS-CoV-2 shows higher binding affinity to hACE2 and enhanced replication in human lung cells compared with its putative ancestral variant (T372), providing evidence of a viral mutation that is likely to have been necessary to enable human-to-human transmission.

Kang et al., 2021, Cell *184*, 4392–4400
August 19, 2021 © 2021 Elsevier Inc.
https://doi.org/10.1016/j.cell.2021.07.007

CellPress

CellPress

Cell

Article

A selective sweep in the *Spike* gene has driven SARS-CoV-2 human adaptation

Lin Kang,[1,2] Guijuan He,[3] Amanda K. Sharp,[4] Xiaofeng Wang,[3] Anne M. Brown,[4,5,6] Pawel Michalak,[1,7,8,*] and James Weger-Lucarelli[2,9,*]

[1]Edward Via College of Osteopathic Medicine, Monroe, LA 71203, USA
[2]Department of Biomedical Sciences and Pathobiology, VA-MD Regional College of Veterinary Medicine, Virginia Tech, Blacksburg, VA 24060, USA
[3]School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA 24061, USA
[4]Program in Genetics, Bioinformatics, and Computational Biology (GBCB), Virginia Tech, Blacksburg, VA 24061, USA
[5]Research and Informatics, University Libraries, Virginia Tech, Blacksburg, VA 24061, USA
[6]Department of Biochemistry, Virginia Tech, Blacksburg, VA 24061, USA
[7]Center for One Health Research, VA-MD Regional College of Veterinary Medicine, Blacksburg, VA 24060, USA
[8]Institute of Evolution, University of Haifa, Haifa 3498838, Israel
[9]Lead contact
*Correspondence: pmichalak@vcom.edu (P.M.), weger@vt.edu (J.W.-L.)
https://doi.org/10.1016/j.cell.2021.07.007

SUMMARY

The coronavirus disease 2019 (COVID-19) pandemic underscores the need to better understand animal-to-human transmission of coronaviruses and adaptive evolution within new hosts. We scanned more than 182,000 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomes for selective sweep signatures and found a distinct footprint of positive selection located around a non-synonymous change (A1114G; T372A) within the spike protein receptor-binding domain (RBD), predicted to remove glycosylation and increase binding to human ACE2 (hACE2), the cellular receptor. This change is present in all human SARS-CoV-2 sequences but not in closely related viruses from bats and pangolins. As predicted, T372A RBD bound hACE2 with higher affinity in experimental binding assays. We engineered the reversion mutant (A372T) and found that A372 (wild-type [WT]-SARS-CoV-2) enhanced replication in human lung cells relative to its putative ancestral variant (T372), an effect that was 20 times greater than the well-known D614G mutation. Our findings suggest that this mutation likely contributed to SARS-CoV-2 emergence from animal reservoirs or enabled sustained human-to-human transmission.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), has caused over 155 million infections with at least 3.2 million deaths worldwide as of early May 2021. The virus was first described in late 2019 in Wuhan, China, and quickly spread globally (Zhou et al., 2020b). SARS-CoV-2 is closely related to SARS-CoV, which caused a more limited outbreak in several countries in 2003 (Peiris et al., 2003; Rota et al., 2003); however, several bat- and pangolin-derived viruses are even more closely related to SARS-CoV-2, indicative of a zoonotic origin (Lam et al., 2020; Xiao et al., 2020; Zhou et al., 2020a). Bat CoV RaTG13—originally isolated in China from *Rhinolophus affinis* bats in 2013—shares 96% nucleotide identity with SARS-CoV-2 across the genome and ~97% amino acid identity in the Spike (S) protein, which mediates receptor binding and membrane fusion and is the key CoV determinant of host tropism (Graham and Baric, 2010). Similarly, several viruses found in Malayan pangolins (*Manis javanica*) are closely related to SARS-CoV-2, with up to 97.4% amino acid concordance in the receptor-binding domain (RBD) of the S protein (Lam et al., 2020; Xiao et al., 2020). However, the exact origin and mechanism of cross-species transmission of the SARS-CoV-2 progenitor are still unknown.

In the past two decades, the emergence of SARS-CoV (Drosten et al., 2003; Ksiazek et al., 2003; Peiris et al., 2003; Rota et al., 2003) and Middle East respiratory syndrome CoV (MERS-CoV) (Zaki et al., 2012) in humans and swine acute diarrhea syndrome CoV (SADS-CoV) in pigs has highlighted the epidemic potential of CoVs (Zhou et al., 2018). Typically, only modest changes to a virus are required to initiate adaptation to a new host; for example, only two amino acid changes were necessary to produce a dramatic difference in human adaptation in SARS-CoV and MERS-CoV S proteins (Li et al., 2005; Yang et al., 2015). This phenomenon is readily observed in other viruses. Ebola viruses' human adaptation following spillover from bats was at least partly mediated by a single alanine-to-valine mutation at position 82 in the glycoprotein (Diehl et al., 2016; Urbanowicz

 © 2021 Elsevier Inc.

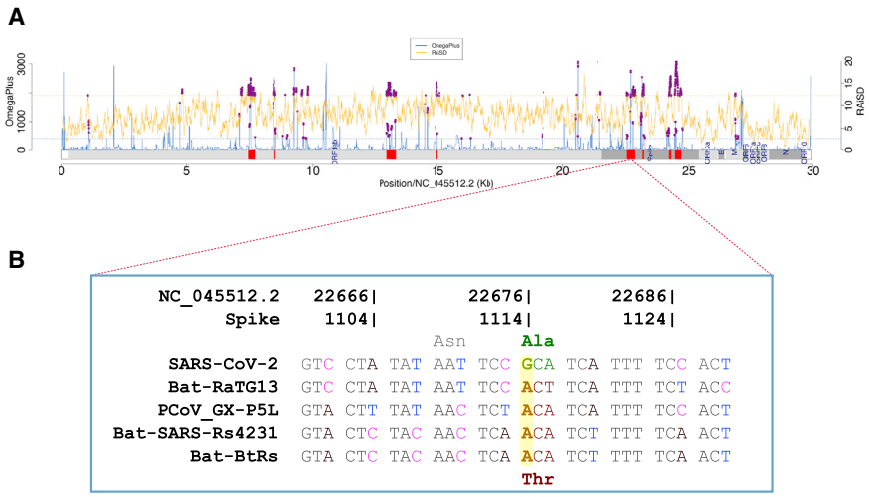


Figure 1. Selective sweeps analysis

(A) Selective sweep regions (shown as red blocks) identified in 182,792 SARS-CoV-2 genomes using OmegaPlus (blue lines) and RAiSD (yellow lines). The common outliers (0.05 cutoff, purple dots) from the two methods were used to define selective sweep regions.

(B) Non-synonymous difference (Thr372Ala) between SARS-CoV-2 and four other Sarbecovirus members found in the putative selective sweep region (22,529–22,862).

et al., 2016). Similarly, individual amino acid changes have been associated with recent outbreaks of several RNA viruses: chikungunya virus (Vazeille et al., 2007), West Nile virus (Ebel et al., 2004; Moudy et al., 2007), and Zika virus (Liu et al., 2017). Although an individual mutation that likely increases replication of SARS-CoV-2 in humans has been identified—a single aspartic acid-to-glycine change at position 614 in the S protein (Korber et al., 2020; Plante et al., 2021)—this occurred after emergence into humans, and the genetic determinants of SARS-CoV-2 expansion from an animal reservoir into humans remain entirely unknown.

For a virus acquired recently through cross-species transmission, rapid evolution and a strong signature of positive selection are expected. For example, several rounds of adaptive changes have been demonstrated in SARS-CoV genomes during the short SARS epidemic in 2002–2003 (Chinese SARS Molecular Epidemiology Consortium, 2004; Yeh et al., 2004). However, during its brief epidemic, SARS-CoV-2 has been characterized by relatively low genetic variation, concealing signals of positive selection and leading to contradictory reports of limited positive selection (Cagliani et al., 2020), "relaxed" selection (Chaw et al., 2020), or even negative (purifying) selection (Li et al., 2020; Lv et al., 2020). However, these results are based on dN/dS tests that are traditionally designed for eukaryotic interspecies comparisons and, thus, ill equipped to detect hallmark signatures of positive selection in viral lineages with limited sequence divergence (Kryazhimskiy and Plotkin, 2008). Here we employ highly sensitive methods enabling detection of selective sweeps in which a selectively favorable mutation spreads all or part of the way through the population, causing a reduction in the level of sequence variability at nearby genomic sites (Smith and Haigh, 2007). With high statistical power that leverages information from more than 182,000 SARS-CoV-2 genomes, we demonstrate that positive selection, manifested as selective sweeps in *Spike* and several other regions, has likely played a critical role in the adaptive evolution of SARS-CoV-2. Within one of the selective sweep regions, we identify an amino acid change that was fixed within the RBD of the S protein of all SARS-CoV-2 sequences (S A372) but is different in closely related viruses from animal reservoirs (S T372). Given the S protein's role in CoV host tropism, we hypothesized and experimentally validated that this change involves an adaptive mutation enhancing replication in human lung cells and increasing binding to human ACE2, which, in turn, could facilitate more efficient human-to-human transmission. Our findings have important implications for the origins of the COVID-19 pandemic and identify regions under positive selection that could be targeted for further analysis and interventions.

RESULTS

Selective sweeps analysis identified an S region with high confidence from 182,792 sequences

OmegaPlus (Alachiotis et al., 2012) and RAiSD (Alachiotis and Pavlidis, 2018) were used to find putative selective sweep regions in 182,792 SARS-CoV-2 genomes downloaded from the publicly available GISAID EpiCov database (https://www.gisaid.org). Eight selective sweep regions were detected, including four in ORF1ab and four in the S region (Figure 1; Table 1). The S protein plays an important role in the receptor recognition and cell membrane fusion process during viral infection. Next we screened genomic sites in the *Spike* region that may be involved in adaptive evolution of SARS-CoV-2 in the new host by comparing the non-synonymous differences between SARS-CoV-2 and four other Sarbecovirus members (one pangolin CoV and three bat CoVs; STAR Methods). A total of six such sites were identified (Table S1); notably, only a single site (A1114G, genomic position 22,676; Figure 1B) was centrally located in one of the sweep regions; this is within the codon position 372 of the S protein. The amino acid threonine in this position of the four Sarbecovirus members was substituted with alanine (Thr372Ala) in human SARS-CoV-2. Of the 182,792 SARS-CoV-2 genomes, no sequence polymorphism was found in this position (G1114), suggesting rapid fixation of this mutation via sweep. The alternative, putatively ancestral CoV variant (A1114) was perfectly conserved in Sarbecovirus members from bats and pangolins.

Structure-based analysis of SARS-CoV-2 S protein variants

Comparative molecular modeling of wild-type (WT) (A372, SARS-CoV-2), T372, and G614 S protein was performed to

Table 1. Putative sweep regions (the region containing the S G1114A position is bolded).

Start	End	Codon position in gene	Annotation	Score
7,445	7,711	2,394–2,482	ORF1ab/NSP3	0.98
8,426	8,542	2,721–2,759	ORF1ab/NSP3	1
12,978	13,350	4,238–4,362	ORF1ab/NSP10	0.99
14,907	15,027	4,881–4,921	ORF1ab/NSP12	0.96
22,529	**22,862**	**323–434**	**S**	**0.97**
23,132	23,196	524–545	S	1
24,225	24,319	888–919	S	0.69
24,456	24,712	965–1,050	S	1

connect the selective sweep G1114A mutation to structural data (Figure 2). The S D614G mutant was included because it now predominates worldwide and has been associated with higher titers in nasopharyngeal swabs in humans and increased replication in human cells and hamsters (Korber et al., 2020; Plante et al., 2021). Structures were energy minimized after mutation and analyzed for changes in ACE2 binding and probability of N-linked glycosylation sites as a result of mutation. An increased probability of N-linked glycosylation at N370 was observed in the T372 variant (Figures 2A–2C), given that mutation to a threonine provided a standard N-linked glycosylation site motif (NXT/S) and the solvent-accessible surface area of N370 (Figures 2D and 2E). This site has been identified previously to be glycosylated in SARS-CoV (N357) with a complex glycan (Watanabe et al., 2020a) but not in SARS-CoV-2 (Watanabe et al., 2020b). No glycosylation site was predicted at N370 in the WT S protein. To further probe, in a simple model, the effect of the predicted glycosylation site at N370 as a result of the presence of a threonine at position 372 of S protein, N370 was glycosylated with an N-acetylglucosamine (GlcNAc) glycan and energy minimized on the T372 S protein model to observe any minor, local side chain readjustment as a result of a simple N-glycan presence. A simple glycosylation (GlcNAc), not a complex multi-unit mannose and GlcNAc glycan, was used in this work to mimic the resolved structure glycosylation and bound state to ACE2. N370 glycosylation of T372 S protein occurs in close structural proximity to the essential glycosylation site of N343 (Casalino et al., 2020), further providing additional N-glycan shielding of the RBD (Figure 2A). Surface maps also reveal an additional space-filling and polar surface that is now occupied by the N370 N-glycan (Figures 2B and 2C). Additionally, molecular mechanics generalized born surface area (MM/GBSA) free energy of binding of ACE2 to S protein was calculated for WT, T372, and N370-glycosylated T372 S protein. Free energy of binding of ACE2 to WT S protein showed a very negative, favorable relative binding affinity (−180.503 kcal/mol), whereas free energy of binding of ACE2 to the putatively ancestral T372 variant and N370-glycosylated T372 variant was less negative and favorable (−95.7685 kcal/mol and −76.401 kcal/mol, respectively), highlighting that, although the glycosylation at N370 is not in close proximity (>10 Å) of the receptor-binding motif (RBM), it influences the RBD of S protein and its potentiality of binding ACE2. Structural analysis of G614 did not indicate any major local modifications to the structure of S protein nor its proximity to the RBD. Additionally, residue 614 is in close proximity to a glycosylation site (N616), but G614 did not change the probability of glycosylation or general surface properties compared with D614.

SARS-CoV-2 S A372T reduces binding to human ACE2

We next sought to experimentally validate our molecular modeling data using functional ELISA. We probed hACE2 with RBDs from WT (A372), A372T, and N501Y as a positive control; D614G could not be used because it is not within the RBD. The N501Y mutation is present in several variants of concern and has been shown previously to increase binding to hACE2 (Collier et al., 2021; Laffeber et al., 2021; Liu et al., 2021). As expected, 50% effective dose (EC_{50}) values were lower for the N510Y mutant (5.83 ± 0.94 ng/mL; Figure 3A) than the WT (12.48 ± 1.26 ng/mL), indicating a stronger binding affinity for hACE2. In contrast, EC_{50} values were markedly higher for the A372T mutant (26.29 ± 0.08 ng/mL), consistent with our molecular modeling data suggesting a weaker interaction with hACE2 compared with the WT. EC_{50} values compared directly showed robust differences between the WT and A372T or N501Y (Figure 3B; both $p < 0.0001$ by one-way ANOVA with Dunnett's multiple comparisons test). These results suggest that the S T372A mutation that occurred in the SARS-CoV-2 ancestral virus enhanced affinity to hACE2.

SARS-CoV-2 S A372 enhances replication in human lung cells

Here we sought to define the effect of the S T372A mutation on viral replication in human cells. We used an infectious clone of WT SARS-CoV-2 (A372) to revert to the ancestral residue (T372) using a bacterium-free cloning approach we developed previously to prevent bacterial toxicity associated with manipulating unstable viral genomes in bacteria (Bates et al., 2021; Weger-Lucarelli et al., 2018). For clarity, we will refer to the mutant as A372T because we reverted WT SARS-CoV-2 (A372) to its ancestral form (T372). Concurrently, we generated the S D614G mutant, which increases replication in human cells (Plante et al., 2021). Both mutants were constructed in an infectious clone originally produced in yeast (Thi Nhu Thao et al., 2020) of an early SARS-CoV-2 strain, 2019-nCoV BetaCoV/Wuhan/WIV04/2019 (Zhou et al., 2020b). A schematic of the A372T mutant is presented in Figure 4A; although not depicted, the D614G mutant was made by replacing the WT codon (GAT) with the glycine-encoding codon (GGC). Following virus rescue, viral plaque morphology on Vero E6 cells was similar for all three viruses, although the A372T mutant plaques appear slightly smaller (Figure 4B).

We next evaluated the replication kinetics of each virus—WT, S A372T, and S D614G—in Vero E6 and Calu-3 cell lines, monkey kidney and human lung epithelial cell lines, respectively. Following infection in Vero E6 cells, viral titers rose rapidly for all three viruses, and only minor differences in peak titers were observed among the viruses (Figure 4C). In Calu-3 cells, the D614G mutant produced significantly higher titers than WT 1 day post-infection (dpi), but levels were similar for the

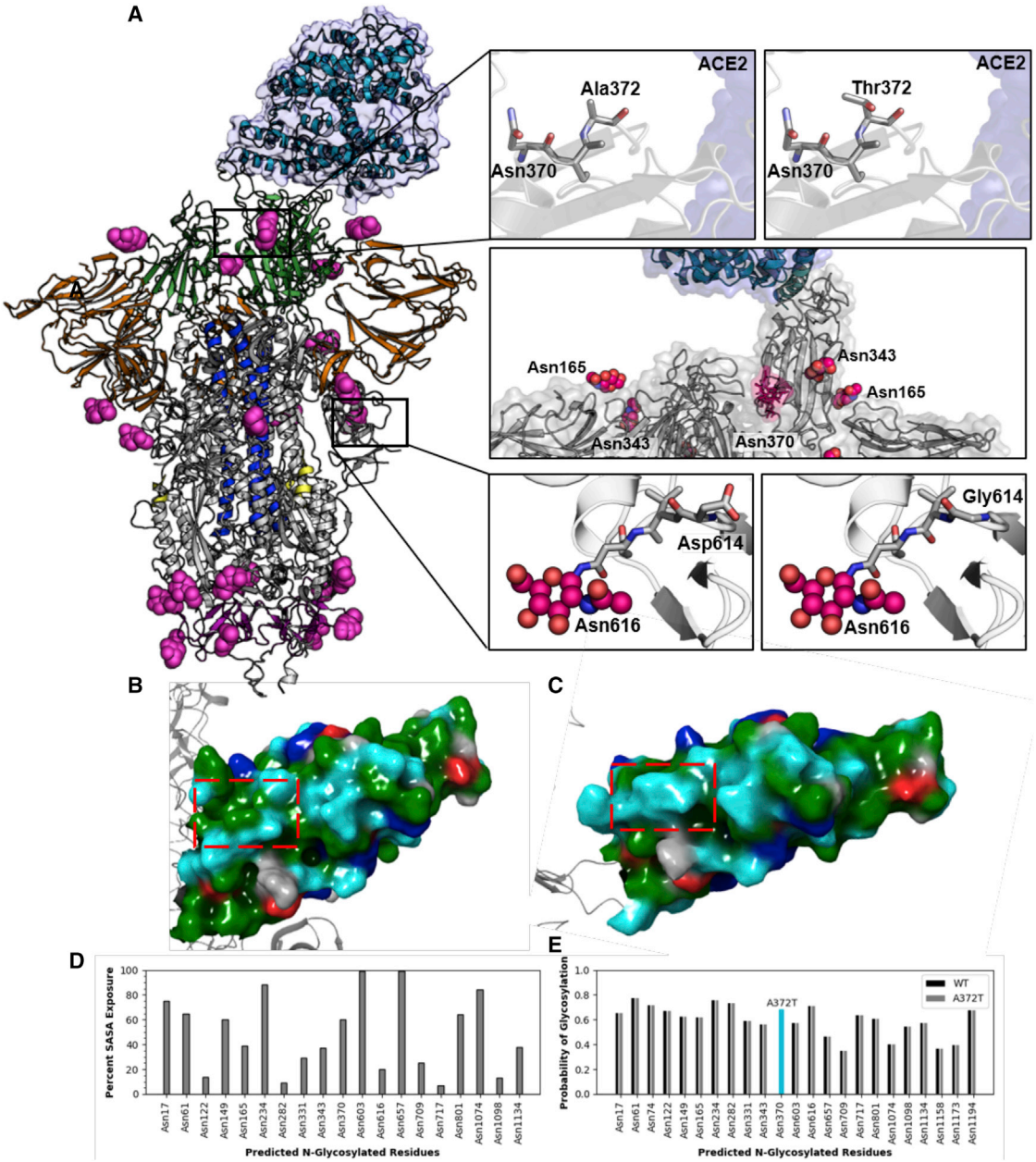


Figure 2. Structure-based analysis of SARS-CoV-2 S protein variants

(A) Visualization of the T372 and D614G mutants. The structure of S protein (PDB: 7A94) is displayed as a cartoon and colored by RBD (green), N-terminal domain (NTD; orange), central helix (CH;blue), FP (yellow), and connector domain (CD;pink). Glycans are displayed as spheres colored hot pink. The top panel shows the WT (A372) and T372 mutant, the center panel displays a glycosylated N370 T372 S protein with various rotamers of the GlcNAc-glycosylated N370, and the bottom panel shows the WT and G614 mutant.

(B and C) Surface map of the WT S protein (B) and the N370-glycosylated T372 S protein (C), colored by the residue side-chain properties: green for hydrophobic, blue for positively charged, red for negatively charged, teal for polar uncharged, and gray for neutral.

(D) Predicted N-glycosylated residues identified by Schrödinger-Maestro's BioLuminate (v.2020-2) Reactive Residue package with percent solvent-accessible surface area (SASA) exposure of each residue.

(E) Predicted N-glycosylated residues identified by the NetNGlyc 1.0 server with the probability of being glycosylated.

remaining time points (Figure 4D; $p = 0.0066$ by 2-way ANOVA with Dunnett's correction at 1 dpi). No differences were observed 24 h after infection between the WT and A372T mutant, but later time points showed a marked reduction in replication for the A372T mutant ($p = 0.0033$, $p < 0.0001$, and $p < 0.0001$ for 2, 3, and 4 dpi, respectively). Compared with the WT, D614G had modest differences of 2.9-, 2.9-, 1.3-, and 0.8-fold in viral titers 1, 2, 3, and 4 dpi, respectively; in contrast, compared with the

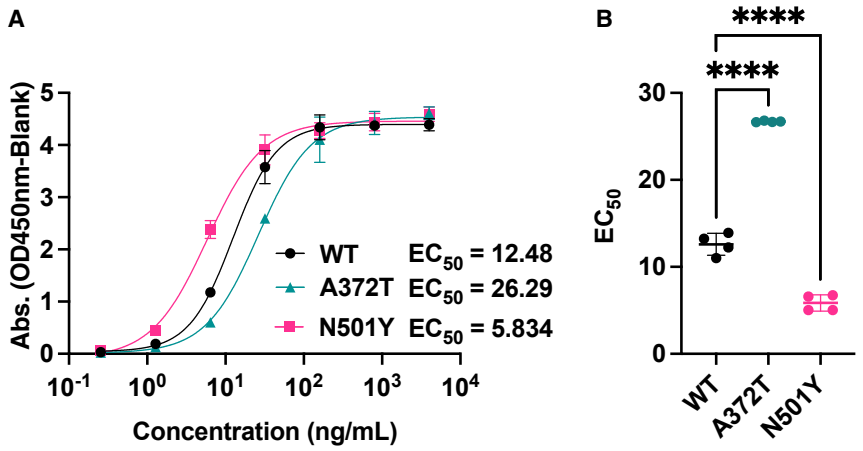


Figure 3. Decreased binding of the A372T mutant to human ACE2

(a) Functional ELISA was used to determine the binding affinity of different S protein receptor-binding domains (RBDs). Plates were coated with recombinant human ACE2 receptor (2 μg/mL at 100 μL/well) and then probed with varying concentrations (0.256–4000 ng/mL) of purified RBDs from WT SARS-CoV-2 (S A372), A372T, and N501Y (positive control). To determine EC_{50} values, the absorbance values (450 nM) were fit to a sigmoidal, 4PL nonlinear model using Prism 9 (GraphPad). The experiment was repeated in two independent replicates with four total technical replicates per sample. Error bars represent standard deviation of the mean.

(B) The EC_{50} values were compared by one-way ANOVA with Dunnett's multiple comparisons test. ****$p < 0.0001$ compared with WT SARS-CoV-2 (A372). Error bars represent standard deviation of the mean.

WT, A372T titers were 1.8-, 5.5-, 31.1-, and 64.1-fold lower 1, 2, 3, and 4 dpi, respectively (Figure 4D). These data indicate that an alanine at S position 372 confers a robust fitness advantage over several time points in human lung cells and that this effect is considerably more substantial than the change at position 614.

Based on structural analysis, others have postulated that the SARS-CoV-2 S trimer would have higher thermal stability than the S of bat virus RaTG13 (Wrobel et al., 2020). To determine whether A372T altered SARS-CoV-2 thermal stability, we incubated 10^5 plaque-forming units (PFUs) of WT SARS-CoV-2, D614G, or A372T at room temperature (~25°C) or 37°C to mimic environmental and human body temperature, respectively. A372T titers did not differ significantly from the WT at any time point for temperature (Figures 4E and 4F). Following 48-h incubation at room temperature, the titer of D614G was higher than that of WT SARS-CoV-2 ($p = 0.0303$), which is consistent with a previous report (Plante et al., 2021).

DISCUSSION

COVID-19 has now claimed the lives of more than 3.2 million people worldwide, dwarfing the number of deaths caused by SARS-CoV (774; Cherry, 2004) and MERS-CoV (858; Memish et al., 2020). Although phylogenetic and epidemiological data suggest a zoonotic origin for SARS-CoV-2, little is known about the viral mutations that likely occurred to adapt the virus to human transmission. The SARS-CoV-2 progenitor would have likely required new adaptations to sustain human-to-human transmission—a process that likely included a strong positive selection event, favoring the viruses with the greatest replication in the human respiratory tract. Here we identified a region in the *Spike* gene with a strong signal of such an event—a selective sweep—from over 180,000 SARS-CoV-2 genomes. Within this region, present in the RBD, we identified a non-synonymous single-nucleotide polymorphism (SNP) that is fixed in all SARS-CoV-2 genomes sequenced to date, whereas an alternative, and presumably ancestral SNP, is fixed in the other members of the Sarbecovirus lineage.

Residue 372 lies within the RBD (Figure 2A), which mediates viral entry through the human ACE2 receptor (Zhou et al., 2020b). Although positioned adjacent to the ACE2 interface of the RBD, the presence of an alanine at position 372 (A372) is predicted to remove a glycosylation site present at the asparagine at position 370 (Figures 2D and 2E; Wrobel et al., 2020). Indeed, molecular modeling of GlcNAc at N370 in an open conformation of T372 S protein shows a highly solvent accessible glycan site (Figure 2). In the closed conformation of T372, the N370 glycan site becomes less solvent exposed and further fills a solvent-accessible region on the outer edge of the RBD. Additionally, N-Glycans are known to modulate the RBD of S protein, with glycans at position N165 and N234 influencing the open/closed metastable conformation states of the RBD and N-glycans at N331 and N343 having more of a shielding role of the RBD itself, regardless of state (Casalino et al., 2020). N370 glycosylation is in close structural proximity of the N-glycan site at N343 and is in relative distance to the RBM and RBD/ACE2 interface (Figures 2A–2C). Free energy of binding of ACE2 to S protein indicates a decrease in relative binding affinity of ACE2 to S protein in the N370-glycosylated T372 variant compared with the WT (−76.401 kcal/mol versus −180.503 kcal/mol, respectively). Molecular dynamics (MD) simulations of the S protein glycan shield have shown to be a key influence on the transition between the open and closed states in the RBD (Casalino et al., 2020). The effect of glycosylation and, more importantly, complex glycosylation of N370 on the structural morphology and dynamics of S protein will need to be investigated further to determine whether it influences the structural state in a similar manner. Recent work confirms the simplistic models proposed here by using MD simulation (MDS) to probe the influence of N370 glycosylation on the open/close confirmation of the RBD. Harbison et al. (2021) show that, although glycosylation at N370 stabilized the open S protein conformation, N370 glycosylation promoted increased interactions between adjacent RBDs that ultimately improved and strengthened the closed state of the RBDs, proposing that the presence of the N370 glycan favors the closed state as opposed to the WT non-glycosylated N370. We determined experimentally that the RBD from the T372 variant

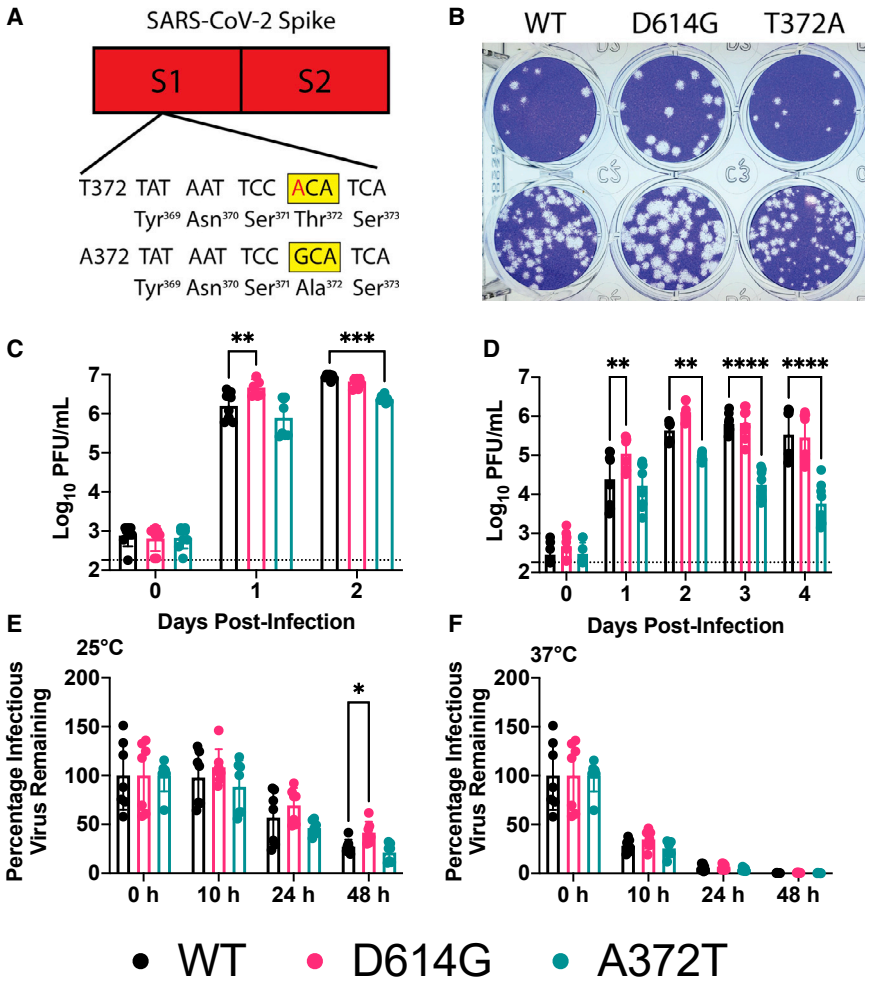


Figure 4. A372T substitution decreases SARS-CoV-2 replication on human lung epithelial cells

(A) The S T372 SARS-CoV-2 mutant was generated by making a single G-to-A substitution. The mutant nucleotide is presented in red, and the altered codon is highlighted in a yellow box.

(B) Plaque morphology of WT and mutant viruses. Plaques were visualized 2 days post-infection (dpi) on Vero E6 cells.

(C and D) Viral replication on Vero E6 (C) and Calu-3 (D) cells following infection at an MOI of 0.05. The sample at 0 dpi was collected immediately after infection to ensure cells were exposed to similar levels of virus, and then samples were collected at 24-h intervals.

(E and F) Kinetics of thermal stability. A solution of 10^5 PFU of each virus was incubated at the indicated temperature for different lengths of time. Infectious virus was measured by plaque assay on Vero E6 cells.

Statistical comparisons were made using two-way ANOVA with Dunnett's multiple comparisons test. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. Error bars represent standard deviation of the mean.

bound hACE2 with lower affinity than WT SARS-CoV-2 RBD (A372; Figure 3); thus, open-close conformational dynamics cannot fully account for the difference in binding strength between WT and T372 S. Interestingly, SARS-CoV S has T372 in its S protein, suggesting that other residues may perform similar functions for other CoVs (Harbison et al., 2021).

Using a reverse genetics system to generate a SARS-CoV-2 mutant containing the putative ancestral SNP, we show that the A372T S mutant virus replicates over 60-fold less efficiently than WT SARS-CoV-2 in Calu-3 human lung epithelial cells (Figure 4d). Further, growth of the A372T S mutant was reduced greatly for multiple days, which may be indicative of an effect on viral shedding kinetics in humans. We also generated the D614G S mutant here—reported widely to increase SARS-CoV-2 infectivity (Korber et al., 2020)—which only increased viral titers by a maximum of 2.9-fold in Calu-3 cells compared with the WT, a finding that is consistent with previous results (Plante et al., 2021). We also observed slight attenuation for the A372T S mutant in Vero E6 cells (3.8-fold lower titers compared with the WT 2 dpi). The large replication differences between the two cell lines suggest a cell-specific mechanism of attenuation. In fact, besides their species of origin, Calu-3 and Vero E6 cells differ in several important aspects. First, Vero E6 cells are deficient in type 1 interferon signaling (Desmyter et al., 1968), which inhibits SARS-CoV-2 replication (Felgenhauer et al., 2020; Mantlo et al., 2020). However, the S protein is not known to antagonize interferon (IFN) production, and, therefore, IFN is unlikely to drive the differences observed here. Additionally, the S protein requires host-mediated proteolytic cleavage to undergo fusion, which can be driven by several proteases, including TMPRSS2 at the cell surface and cathepsins B and L (CatB/L) in endosomes (Hoffmann et al., 2020). Calu-3 cells express low levels of cathepsins but high levels of TMPRSS2, suggesting a TMPRSS2-dependent entry mechanism in Calu-3 cells (González-Hernández et al., 2019). In contrast, SARS-CoV-2 infection of Vero E6 cells is CatB/L dependent (Hoffmann et al., 2020). Clinical isolates of CoVs prefer entry through TMPRSS2 as opposed to CatB/L (Shirato et al., 2016, 2018); accordingly, Calu-3 cells mimic the human environment closely in terms of S protein priming. These data hint that, along with decreased receptor-binding, inefficient TMPRSS2 cleavage of A372T S could at least partially mediate the attenuation we observed. Host proteases have been implicated in cross-species transmission of MERS-CoV from bats to humans (Yang et al., 2014). Hence, it will be important for future studies to define the importance of TMPRSS2-mediated cleavage of the S protein in the context of these mutations.

We did not observe large temperature stability differences between viruses here. A previous report predicted that the SARS-CoV-2 S protein would have higher thermal stability than that of bat CoV RaTG13 (Wrobel et al., 2020); however, it does not appear that the residue difference at position 372 dictates this difference.

It may also be that differences would have been observed at different time points or temperatures; nonetheless, these data suggest that thermal stability is not a likely driving factor in emergence of the variants at position 372 or 614 in the S protein.

Our data supply solid evidence that S protein residue 372 is critical for replication in human cells. The fact that this site is not polymorphic in more than 180,000 SARS-CoV-2 sequences further underscores its importance. The threonine-to-alanine change may have enabled the putative ancestral virus to replicate more efficiently in human cells, possibly enabling efficient human-to-human transmission. Although other studies have identified evidence of positive selection in SARS-CoV-2 (Cagliani et al., 2020; Korber et al., 2020; Velazquez-Salinas et al., 2020), these studies are entirely computational or use pseudotyped viruses. Although useful information can be obtained using pseudotyped viruses, they typically express only the S protein; consequently, they do not fully recapitulate the viral life cycle, including interactions between different viral proteins and the host, and cannot complete an entire viral replication cycle. Plante et al. (2021) used a reverse genetics system to generate the D614G S protein mutant and showed increased replication in cell culture and hamsters, highlighting the utility of using a live virus to characterize critical viral mutations. Our use of a live virus enables future studies in hamster or ferret models that recapitulate human-to-human transmission (Chan et al., 2020; Kim et al., 2020; Richard et al., 2020; Sia et al., 2020).

Limitations of the study

The OmegaPlus and RAiSD programs we used to identify selective sweeps have not been optimized on viral genomes; therefore, caution should be exercised when considering experimentally unvalidated candidate sites. Although the experimental data presented here clearly demonstrate the dramatic effect of the S protein A372T mutation on SARS-CoV-2 replication in human lung cells, we cannot definitively conclude that it enabled efficient human-to-human transmission or that it was necessary for cross-species transmission. Our findings suggest, though, that efficient replication of SARS-CoV-2 in a human would be unlikely with a threonine at S protein position 372, from which we could infer that transmission would be equally unlikely. Because the true putative SARS-CoV-2 ancestor has not been isolated, it is impossible to know when this mutation may have arisen. Phylogenetic estimates suggest that SARS-CoV-2 emerged in late November 2019 to early December 2019 (Rambaut, 2020), but the first known case was not detected until December 1, 2019 (Huang et al., 2020). However, this case had no connection to the Huanan seafood market, indicating that transmission was ongoing before early December or that the seafood market is not the origin of the pandemic but, rather, a spreading point. Although it is impossible to know SARS-CoV-2's exact emergence date, it seems likely that transmission occurred unnoticed for some period of time, providing a window for SARS-CoV-2's ancestor to adapt to human replication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines and growing conditions
  - Virus strains
  - Authentication
  - Ethics and biosafety
- METHOD DETAILS
  - Putative Selective Sweep Region Detection
  - Molecular Modeling and Free Energy of Binding Calculations
  - Nucleoprotein expression construct
  - Bacteria-free cloning (BFC) and site-directed mutagenesis (SDM)
  - Virus rescue
  - Plaque assays and growth curves
  - Temperature stability
  - Functional ELISA
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.07.007.

ACKNOWLEDGMENTS

We thank Dr. Volker Thiel for sharing the yeast SARS-CoV-2 reverse genetics platform. We also thank Stephen DiGiuseppe, David Veesler, and Samantha Zepeda for input. This work was supported by a VCOM One Health Research Seed Program grant (no. 10360 to P.M. and J.W.-L.), a National Science Foundation grant (no. 2032166 (to J.W.-L. and L.K.), and a Virginia Tech Institute for Critical Technology and Applied Science Junior Faculty Award (to J.W.-L.).

AUTHOR CONTRIBUTIONS

Conceptualization, L.K., P.M., and J.W.-L.; investigation, L.K., G.H., A.K.S., A.M.B., P.M., and J.W.-L.; writing – original draft, L.K., A.M.B., P.M., and J.W.-L.; writing - review and editing, L.K., X.W., A.M.B., P.M., and J.W.-L.; supervision, X.W., A.M.B., P.M., and J.W.-L.; funding acquisition, P.M. and J.W.-L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 22, 2020
Revised: May 7, 2021
Accepted: July 2, 2021
Published: July 7, 2021

REFERENCES

Alachiotis, N., and Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Commun. Biol. *1*, 79.

Alachiotis, N., Stamatakis, A., and Pavlidis, P. (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics *28*, 2274–2275.

Bates, T.A., Chuong, C., Hawks, S.A., Rai, P., Duggal, N.K., and Weger-Lucarelli, J. (2021). Development and characterization of infectious clones of two strains of Usutu virus. Virology *554*, 28–36.

Beard, H., Cholleti, A., Pearlman, D., Sherman, W., and Loving, K.A. (2013). Applying physics-based scoring to calculate free energies of binding for single amino acid mutations in protein-protein complexes. PLoS ONE *8*, e82849.

Benton, D.J., Wrobel, A.G., Xu, P., Roustan, C., Martin, S.R., Rosenthal, P.B., Skehel, J.J., and Gamblin, S.J. (2020). Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature *588*, 327–330.

Cagliani, R., Forni, D., Clerici, M., and Sironi, M. (2020). Computational Inference of Selection Underlying the Evolution of the Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2. J. Virol. *94*.

Casalino, L., Gaieb, Z., Goldsmith, J.A., Hjorth, C.K., Dommer, A.C., Harbison, A.M., Fogarty, C.A., Barros, E.P., Taylor, B.C., McLellan, J.S., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent. Sci. *6*, 1722–1734.

Chan, J.F.-W., Zhang, A.J., Yuan, S., Poon, V.K.-M., Chan, C.C.-S., Lee, A.C.-Y., Chan, W.-M., Fan, Z., Tsoi, H.-W., Wen, L., et al. (2020). Simulation of the Clinical and Pathological Manifestations of Coronavirus Disease 2019 (COVID-19) in a Golden Syrian Hamster Model: Implications for Disease Pathogenesis and Transmissibility. Clin. Infect. Dis. *71*, 2428–2446.

Chattopadhyay, A., Schmidt, M.C., and Khan, S.A. (2005). Identification of a 450-bp region of human papillomavirus type 1 that promotes episomal replication in Saccharomyces cerevisiae. Virology *340*, 133–142.

Chaw, S.-M., Tai, J.-H., Chen, S.-L., Hsieh, C.-H., Chang, S.-Y., Yeh, S.-H., Yang, W.-S., Chen, P.-J., and Wang, H.-Y. (2020). The origin and underlying driving forces of the SARS-CoV-2 outbreak. J. Biomed. Sci. *27*, 73.

Cherry, J.D. (2004). The chronology of the 2002-2003 SARS mini pandemic. Paediatr. Respir. Rev. *5*, 262–269.

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.

Collier, D.A., De Marco, A., Ferreira, I.A.T.M., Meng, B., Datir, R.P., Walls, A.C., Kemp, S.A., Bassi, J., Pinto, D., Silacci-Fregni, C., et al.; CITIID-NIHR BioResource COVID-19 Collaboration; COVID-19 Genomics UK (COG-UK) Consortium (2021). Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. Nature *593*, 136–141.

Desmyter, J., Melnick, J.L., and Rawls, W.E. (1968). Defectiveness of interferon production and of rubella virus interference in a line of African green monkey kidney cells (Vero). J. Virol. *2*, 955–961.

Diehl, W.E., Lin, A.E., Grubaugh, N.D., Carvalho, L.M., Kim, K., Kyawe, P.P., McCauley, S.M., Donnard, E., Kucukural, A., McDonel, P., et al. (2016). Ebola Virus Glycoprotein with Increased Infectivity Dominated the 2013-2016 Epidemic. Cell *167*, 1088–1098.e6.

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H.-R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A.M., et al. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. *348*, 1967–1976.

Ebel, G.D., Carricaburu, J., Young, D., Bernard, K.A., and Kramer, L.D. (2004). Genetic and phenotypic variation of West Nile virus in New York, 2000-2003. Am. J. Trop. Med. Hyg. *71*, 493–500.

Felgenhauer, U., Schoen, A., Gad, H.H., Hartmann, R., Schaubmar, A.R., Failing, K., Drosten, C., and Weber, F. (2020). Inhibition of SARS-CoV-2 by type I and type III interferons. J. Biol. Chem. *295*, 13958–13964.

González-Hernández, M., Müller, A., Hoenen, T., Hoffmann, M., and Pöhlmann, S. (2019). Calu-3 cells are largely resistant to entry driven by filovirus glycoproteins and the entry defect can be rescued by directed expression of DC-SIGN or cathepsin L. Virology *532*, 22–29.

Graham, R.L., and Baric, R.S. (2010). Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. *84*, 3134–3146.

Harbison, A.M., Fogarty, C.A., Phung, T.K., Satheesan, A., Schulz, B.L., and Fadda, E. (2021). Fine-tuning the Spike: Role of the nature and topology of the glycan shield in the structure and dynamics of SARS-CoV-2 S. bioRxiv. https://doi.org/10.1101/2021.04.01.438036.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell *181*, 271–280.e8.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395*, 497–506.

Jacobson, M.P., Friesner, R.A., Xiang, Z., and Honig, B. (2002). On the role of the crystal environment in determining protein side-chain conformations. J. Mol. Biol. *320*, 597–608.

Jacobson, M.P., Pincus, D.L., Rapp, C.S., Day, T.J.F., Honig, B., Shaw, D.E., and Friesner, R.A. (2004). A hierarchical approach to all-atom protein loop prediction. Proteins *55*, 351–367.

Julenius, K. (2007). NetCGlyc 1.0: prediction of mammalian C-mannosylation sites. Glycobiology *17*, 868–876.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kim, Y.-I., Kim, S.-G., Kim, S.-M., Kim, E.-H., Park, S.-J., Yu, K.-M., Chang, J.-H., Kim, E.J., Lee, S., Casel, M.A.B., et al. (2020). Infection and Rapid Transmission of SARS-CoV-2 in Ferrets. Cell Host Microbe *27*, 704–709.e2.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., et al.; Sheffield COVID-19 Genomics Group (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell *182*, 812–827.e19.

Krieg, P.A. (1990). Improved synthesis of full-length RNA probe at reduced incubation temperatures. Nucleic Acids Res. *18*, 6463.

Kryazhimskiy, S., and Plotkin, J.B. (2008). The population genetics of dN/dS. PLoS Genet. *4*, e1000304.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., et al.; SARS Working Group (2003). A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. *348*, 1953–1966.

Laffeber, C., de Koning, K., Kanaar, R., and Lebbink, J.H.G. (2021). Experimental evidence for enhanced receptor binding by rapidly spreading SARS-CoV-2 variants. J. Mol. Biol. *433*, 167058.

Lam, T.T.-Y., Jia, N., Zhang, Y.W., Shum, M.H., Jiang, J.F., Zhu, H.C., Tong, Y.G., Shi, Y.X., Ni, X.B., Liao, Y.S., et al. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature *583*, 282–285.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.-K., Huang, I.-C., Xu, K., Vasilieva, N., et al. (2005). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. *24*, 1634–1643.

Li, X., Giorgi, E.E., Marichannegowda, M.H., Foley, B., Xiao, C., Kong, X.-P., Chen, Y., Gnanakaran, S., Korber, B., and Gao, F. (2020). Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci. Adv. *6*, eabb9153.

Liu, Y., Liu, J., Du, S., Shan, C., Nie, K., Zhang, R., Li, X.-F., Zhang, R., Wang, T., Qin, C.-F., et al. (2017). Evolutionary enhancement of Zika virus infectivity in Aedes aegypti mosquitoes. Nature *545*, 482–486.

Liu, H., Zhang, Q., Wei, P., Chen, Z., Aviszus, K., Yang, J., Downing, W., Jiang, C., Liang, B., Reynoso, L., et al. (2021). The basis of a more contagious 501Y.V1 variant of SARS-CoV-2. Cell Res. *31*, 720–722.

Lv, L., Li, G., Chen, J., Liang, X., and Li, Y. (2020). Comparative Genomic Analyses Reveal a Specific Mutation Pattern Between Human Coronavirus SARS-CoV-2 and Bat-CoV RaTG13. Front. Microbiol. *11*, 584717.

Schrödinger, LLC (2020). Maestro (Schrödinger, LLC).

Mantlo, E., Bukreyeva, N., Maruyama, J., Paessler, S., and Huang, C. (2020). Antiviral activities of type I interferons to SARS-CoV-2 infection. Antiviral Res. *179*, 104811.

Memish, Z.A., Perlman, S., Van Kerkhove, M.D., and Zumla, A. (2020). Middle East respiratory syndrome. Lancet *395*, 1063–1077.

Moudy, R.M., Meola, M.A., Morin, L.-L.L., Ebel, G.D., and Kramer, L.D. (2007). A newly emergent genotype of West Nile virus is transmitted earlier and more efficiently by Culex mosquitoes. Am. J. Trop. Med. Hyg. *77*, 365–370.

Peiris, J.S.M., Lai, S.T., Poon, L.L.M., Guan, Y., Yam, L.Y.C., Lim, W., Nicholls, J., Yee, W.K.S., Yan, W.W., Cheung, M.T., et al.; SARS study group (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet *361*, 1319–1325.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., et al. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. Nature *592*, 116–121.

Rambaut, A. (2020). Phylodynamic Analysis. https://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356.

Richard, M., Kok, A., de Meulder, D., Bestebroer, T.M., Lamers, M.M., Okba, N.M.A., Fentener van Vlissingen, M., Rockx, B., Haagmans, B.L., Koopmans, M.P.G., et al. (2020). SARS-CoV-2 is transmitted via contact and via the air between ferrets. Nat. Commun. *11*, 3496.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Peñaranda, S., Bankamp, B., Maher, K., Chen, M.-H., et al. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science *300*, 1394–1399.

Salam, N.K., Adzhigirey, M., Sherman, W., and Pearlman, D.A. (2014). Structure-based approach to the prediction of disulfide bonds in proteins. Protein Eng. Des. Sel. *27*, 365–374.

Schrödinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8 (Schrödinger, LLC).

Shirato, K., Kanou, K., Kawase, M., and Matsuyama, S. (2016). Clinical Isolates of Human Coronavirus 229E Bypass the Endosome for Cell Entry. J. Virol. *91*, 91.

Shirato, K., Kawase, M., and Matsuyama, S. (2018). Wild-type human coronaviruses prefer cell-surface TMPRSS2 to endosomal cathepsins for cell entry. Virology *517*, 9–15.

Sia, S.F., Yan, L.-M., Chin, A.W.H., Fung, K., Choy, K.-T., Wong, A.Y.L., Kaewpreedee, P., Perera, R.A.P.M., Poon, L.L.M., Nicholls, J.M., et al. (2020). Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. Nature *583*, 834–838.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19–27.

Smith, J.M., and Haigh, J. (2007). The hitch-hiking effect of a favourable gene. Genet. Res. *89*, 391–403.

Thi Nhu Thao, T., Labroussaa, F., Ebert, N., V'kovski, P., Stalder, H., Portmann, J., Kelly, J., Steiner, S., Holwerda, M., Kratzel, A., et al. (2020). Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. Nature *582*, 561–565.

Urbanowicz, R.A., McClure, C.P., Sakuntabhai, A., Sall, A.A., Kobinger, G., Müller, M.A., Holmes, E.C., Rey, F.A., Simon-Loriere, E., and Ball, J.K. (2016). Human Adaptation of Ebola Virus during the West African Outbreak. Cell *167*, 1079–1087.e5.

Vazeille, M., Moutailler, S., Coudrier, D., Rousseaux, C., Khun, H., Huerre, M., Thiria, J., Dehecq, J.-S., Fontenille, D., Schuffenecker, I., et al. (2007). Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, Aedes albopictus. PLoS ONE *2*, e1168.

Velazquez-Salinas, L., Zarate, S., Eberl, S., Gladue, D.P., Novella, I., and Borca, M.V. (2020). Positive Selection of ORF1ab, ORF3a, and ORF8 Genes Drives the Early Evolutionary Trends of SARS-CoV-2 During the 2020 COVID-19 Pandemic. Front. Microbiol. *11*, 550674.

Watanabe, Y., Berndsen, Z.T., Raghwani, J., Seabright, G.E., Allen, J.D., Pybus, O.G., McLellan, J.S., Wilson, I.A., Bowden, T.A., Ward, A.B., and Crispin, M. (2020a). Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nat. Commun. *11*, 2688.

Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S., and Crispin, M. (2020b). Site-specific glycan analysis of the SARS-CoV-2 spike. Science *369*, 330–333.

Weger-Lucarelli, J., Garcia, S.M., Rückert, C., Byas, A., O'Connor, S.L., Aliota, M.T., Friedrich, T.C., O'Connor, D.H., and Ebel, G.D. (2018). Using barcoded Zika virus to assess virus population structure in vitro and in Aedes aegypti mosquitoes. Virology *521*, 138–148.

Wrobel, A.G., Benton, D.J., Xu, P., Roustan, C., Martin, S.R., Rosenthal, P.B., Skehel, J.J., and Gamblin, S.J. (2020). SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. Nat. Struct. Mol. Biol. *27*, 763–767.

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.-J., Li, N., Guo, Y., Li, X., Shen, X., et al. (2020). Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature *583*, 286–289.

Xie, X., Muruato, A., Lokugamage, K.G., Narayanan, K., Zhang, X., Zou, J., Liu, J., Schindewolf, C., Bopp, N.E., Aguilar, P.V., et al. (2020). An Infectious cDNA Clone of SARS-CoV-2. Cell Host Microbe *27*, 841–848.e3.

Yang, Y., Du, L., Liu, C., Wang, L., Ma, C., Tang, J., Baric, R.S., Jiang, S., and Li, F. (2014). Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. Proc. Natl. Acad. Sci. USA *111*, 12516–12521.

Yang, Y., Liu, C., Du, L., Jiang, S., Shi, Z., Baric, R.S., and Li, F. (2015). Two Mutations Were Critical for Bat-to-Human Transmission of Middle East Respiratory Syndrome Coronavirus. J. Virol. *89*, 9119–9123.

Yeh, S.-H., Wang, H.-Y., Tsai, C.-Y., Kao, C.-L., Yang, J.-Y., Liu, H.-W., Su, I.-J., Tsai, S.-F., Chen, D.-S., and Chen, P.-J.; National Taiwan University SARS Research Team (2004). Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution. Proc. Natl. Acad. Sci. USA *101*, 2542–2547.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D.M.E., and Fouchier, R.A.M. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. *367*, 1814–1820.

Zhou, P., Fan, H., Lan, T., Yang, X.-L., Shi, W.-F., Zhang, W., Zhu, Y., Zhang, Y.-W., Xie, Q.-M., Mani, S., et al. (2018). Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. Nature *556*, 255–258.

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C., et al. (2020a). A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. Curr. Biol. *30*, 2196–2203.e3.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020b). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270–273.

Zhu, K., Day, T., Warshaviak, D., Murrett, C., Friesner, R., and Pearlman, D. (2014). Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction. Proteins *82*, 1646–1655.

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
BetaCoV/Wuhan/WIV04/2019	Yeast infectious clone-derived (Thi Nhu Thao et al., 2020)	WT SARS-CoV-2
BetaCoV/Wuhan/WIV04/2019 D614G	This paper	SARS-CoV-2 D614G
BetaCoV/Wuhan/WIV04/2019 A372T	This paper	SARS-CoV-2 A372T
Chemicals, peptides, and recombinant proteins		
WT SARS-CoV-2 RBD	Sino Biological	40592-V08H
A372T SARS-CoV-2 RBD	Sino Biological	40592-V08H36
N501Y SARS-CoV-2 RBD	Sino Biological	40592-V08H82
hACE2 protein	Sino Biological	10108-H05H
Experimental models: Cell lines		
Vero E6	ATCC	CRL1586
Calu-3	ATCC	HTB-55
BHK-21	ATCC	CCL-10
Experimental models: Organisms/strains		
Saccharomyces cerevisiae strain YPH500	Sikorski and Hieter, 1989	YPH500
Oligonucleotides		
Nucleoprotein cloning Forward: gtaaaacgacggccagtgaattgtaatacgactc actatagATGTCTGATAATGGACCCC	This paper	N/A
Nucleoprotein cloning Reverse: ctcgaggtcgacggtatcgataagcttgatatc gaattcTTAGGCCTGAGTTGAGTCAG	This paper	N/A
JW475-SARS2-Spike-A372T-SweepMut-For: GTCCTATATAATTCCACATCATTTTCCAC	Integrated DNA technologies (IDT)	N/A
JW476-SARS2-Spike-A372T-SweepMut-Rev: GTGGAAAATGATGTGGAATTATATAGGAC	IDT	N/A
JW457-SARS2-Spike-D614G-For: CTTTATCAGGGCGTTAACTGCAC	IDT	N/A
JW458-SARS2-Spike-D614G-Rev: GTGCAGTTAACGCCCTGATAAAG	IDT	N/A
Recombinant DNA		
SARS-CoV-2 strain 2019-nCoV BetaCoV/Wuhan/WIV04/2019	Thi Nhu Thao et al., 2020	WT SARS-CoV-2 yeast plasmid
D614G Bacteria-free clone	This paper	D614G SARS-CoV-2 clone
A372T bacteria-free clone	This paper	A372T SARS-CoV-2 clone
pRS313: yeast cloning vector	Sikorski and Hieter, 1989	pRS313
pRS313-T7-N	This paper	N/A
Software and algorithms		
Minimap 2	Li, 2018	https://github.com/lh3/minimap2
MAFFT	Katoh and Standley, 2013	https://mafft.cbrc.jp/alignment/software/
OmegaPlus	Alachiotis et al., 2012	https://cme.h-its.org/exelixis/web/software/omegaplus/index.html
RAiSD	Alachiotis and Pavlidis, 2018	https://github.com/alachins/raisd
Schrödinger-Maestro (v. 2020-2) software	N/A	https://www.schrodinger.com/products/maestro

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PyMOL	Schrödinger, LLC, 2015	https://www.schrodinger.com/products/pymol
NetNGlyc 1.0 Server	Julenius 2007	http://www.cbs.dtu.dk/services/NetNGlyc/
Graphpad Prism version 9	GraphPad	https://www.graphpad.com/
Other		
SARS-CoV-2 sequences	GISAID Database	https://www.gisaid.org/

RESOURCE AVAILABILITY

Lead contact

Requests for information or reagents and resources should be directed to and will be fulfilled by the Lead Contact, James Weger-Lucarelli (weger@vt.edu).

Materials availability

SARS-CoV-2 mutants generated here are available on request. No materials transfer agreement is necessary.

Data and code availability

All data and code can be requested by contacting the lead author or the co-corresponding author (Pawel Michalak; pmichalak@vcom.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and growing conditions

Vero E6, monkey kidney cells (ATCC CRL1586), and Calu-3, human lung epithelial cells (ATCC HTB-55) were purchased from ATCC. Vero E6 were maintained in Dulbecco's Modified Eagle's medium (DMEM) containing 5% fetal bovine serum (FBS; R&D Systems), gentamicin (50 μg/mL), 10 mM HEPES, and 1x nonessential amino acids (NEAA). Calu-3 cells were grown in DMEM with the same additives except with 20% FBS. All cell lines were held in a humidified incubator at 37°C with 5% CO_2.

Virus strains

Infectious SARS-CoV-2 strain 2019-nCoV BetaCoV/Wuhan/WIV04/2019 was recovered from a previously described infectious clone (Thi Nhu Thao et al., 2020). The viral rescue procedure is described under

Authentication

Sequences were confirmed by Sanger sequencing of virus stocks. Only virus direct from transfection (p0 stock) was used for further characterization. Virus titers were assessed by plaque assay on Vero E6 cells.

Ethics and biosafety

The generation of recombinant SARS-CoV-2 was approved by the Institutional Biosafety Committee at Virginia Tech. All studies with live infectious SARS-CoV-2 or mutant viruses were performed in an approved BSL3 facility following CDC and NIH guidelines. Researchers manipulating live virus wore an N95 respirator or Powered Air Purifying Respirators (PAPR) as approved by the IBC.

METHOD DETAILS

Putative Selective Sweep Region Detection

A total of 182,792 complete SARS-CoV-2 genomes from the human host (low coverage genomes with N's > 5% were excluded) were downloaded from the GISAID EpiCov database (https://www.gisaid.org/) as of Nov. 11, 2020. Sequences were first aligned to SARS-CoV-2 reference (NCBI Reference Sequence/NC_045512.2) using Minimap2 (Li, 2018) (with default parameters other than '-ax asm5'). Sequences with aligned lengths less than 20,000 were excluded from the analysis. The 136,114 remaining sequences were then aligned by using MAFFT (Katoh and Standley, 2013). OmegaPlus (Alachiotis et al., 2012) and RAiSD (Alachiotis and Pavlidis, 2018) were used for sweep region detection, and the SARS-CoV-2 isolate Wuhan-Hu-1 genome (NC_045512.2) was used as an outgroup. OmegaPlus was performed with the following parameters: the minimum and maximum windows to be used for computing linkage disequilibrium values between SNPs were set to be 100bp and 1,000bp, respectively (-minwin 100 -maxwin 1000); the number of omegas to be computed in the alignment was set to be approximately the number of SNPs found among SARS-CoV-2 genomes (-grid 20000). RAiSD was executed with the following parameters: ploidy was set to 1 (-y 1); the total number of evaluation

points across the data was set to be approximately the number of SNPs found among SARS-CoV-2 genomes (-G 20000); imputation of missing data was enabled (-M 1; per SNP); the sliding window size for μ statistic was set to be 50bp (-w 50). The common-outlier method integrated into RAiSD was used to identify the overlapped positions reported by both methods, setting the cut-off threshold of 0.05 (-COT 0.05) and the maximum distance between outliers of 100 (-COD 100). Finally, the common outliers were manually grouped into eight regions with the size of each region greater than 50 bp. The scores of identified putative sweep regions were obtained by the resampling process. For each resample, 60% of SARS-CoV-2 genome sequences were randomly selected from the original pool, followed by the sweep detection process via OmegaPlus and RAiSD described above. The resampling was repeated 100 times, and the proportion of the 100 resamples supporting a sweep region was assigned as the score of that sweep region. To test if the UK sample dominance (about 40% of samples in the GISAID database were from the UK) introduces bias to the results, we ran our pipeline on a subset excluding samples collected from the UK, and the identified selective sweep regions of this subset largely overlapped with that of using the full dataset (Table S2). Four genome sequences (Pangolin coronavirus isolate PCoV_GX-P5L: GenBank/MT040335.1; Bat coronavirus RaTG13: GenBank/ MN996532.2; Bat SARS-like coronavirus isolate Rs4231: GenBank/ KY417146.1; Bat coronavirus BtRs-BetaCoV: GenBank/MK211376.1) were used to assess the nucleotide changes among different *Sarbecovirus* members.

Molecular Modeling and Free Energy of Binding Calculations

Glycosylated S protein structure was downloaded from the RCSB Protein Data Bank (PDB ID: 7A94 (Benton et al., 2020)) and was energy minimized using Schrödinger-Maestro (v. 2020-2) software (Schrödinger, LLC, 2020). The S protein was mutated using PyMOL (Schrödinger, LLC, 2015) to the D614G and A372T S protein variants. After mutation, energy minimization was performed using the OPLS3e force field. To identify glycosylation propensity and predicted glycosylated residues of the WT S protein and the A372T mutant, the NetNGlyc 1.0 Server (Julenius, 2007) and Schrödinger-Maestro's BioLuminate (v. 2020-2) Reactive Residue package was used (Beard et al., 2013; Salam et al., 2014; Zhu et al., 2014). Schrödinger-Maestro's (v. 2020-2) Workspace Operations was used for glycosylation of the Asn370 with N-Acetylglucosamine to identify various Asn370-glycan rotamers and to analyze the surface residue properties of the WT S Protein and A372T mutant. Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) binding free energy was calculated using Schrödinger-Maestro's (v. 2020-1) Prime Package (Jacobson et al., 2002, 2004). Structures were visualized using PyMOL.

Nucleoprotein expression construct

Homologous recombination was used to make the N gene expression plasmid in yeast. The *Saccharomyces cerevisiae* strain YPH500 (MATα ura3-52, lys2-801, ade2-101, trp1-Δ63, his3-Δ200, leu2-Δ1) was used for homologous recombination. All yeast cells were grown at 30°C in a synthetic defined (SD) medium containing 2% glucose as the carbon source. Histidine was omitted from the growth medium to maintain plasmid selection. To construct the nucleoprotein expression vector, pRS313 was linearized by digestion with BamHI and XbaI to serve as the backbone. The PCR-amplified N gene product was cloned into pRS313 by homologous recombination in yeast cells. Plasmid DNA was extracted from the yeast colonies that grew on SD-His plate and then transformed into *E. coli* for amplification. The construct pRS313-T7-N was sequencing-confirmed with the correct coding sequence of the N gene and expected junction sites where the N gene is inserted downstream of the T7 promoter and upstream of the EcoRI site. The primers used for cloning were: Forward: 5′ gtaaaacgacggccagtgaattgtaatacgactcactatagATGTCTGATAATGGACCCC 3′ and Reverse: 5′ cctcgaggtcgacggtatcgataagcttgatatcgaattcTTAGGCCTGAGTTGAGTCAG 3′ where uppercase letters represent the sequences of N gene, and lowercase letters are from vector sequences.

Bacteria-free cloning (BFC) and site-directed mutagenesis (SDM)

SDM was performed using BFC, starting with the yeast clone as a template (Thi Nhu Thao et al., 2020). Primers for mutagenesis were obtained from Integrated DNA Technologies (IDT). PCRs were performed using Platinum SuperFi PCR Master Mix (Invitrogen) or repliQa HiFi ToughMix (Quantabio). Amplicons were purified from a GelGreen nucleic acid stained-gel (Biotium) using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel). Gel-purified amplicons were then assembled using NEBuilder HiFi DNA Assembly Master Mix at a 1:1 molar ratio for each DNA fragment and incubated at 50°C for two hours. To confirm that no parental yeast clone was carried through the process, we included a control containing the DNA fragments but no assembly mix; this was then treated identically to the other samples for the remainder of the process. The assembly was then digested with Exonuclease I, Exonuclease III, and DpnI (all from NEB) to remove single-stranded DNA, double-stranded DNA, and bacterial-derived plasmid DNA, respectively; note, in this case, DpnI was not strictly necessary because yeast-derived plasmids are resistant to DpnI-cleavage (Chattopadhyay et al., 2005); however, it was included for consistency with our previous studies (Weger-Lucarelli et al., 2018). We then amplified the circular product using the FemtoPhi DNA Amplification (RCA) Kit with Random Primers (Evomic Science) at 30°C for 16 hours.

Virus rescue

RCA reactions were linearized with EagI-HF (NEB) and then column purified (Macherey-Nagel). The N expression plasmid was linearized using EcoRV-HF (NEB). Capped-RNA was produced using the mMESSAGE mMACHINE T7 Transcription Kit (Invitrogen) by overnight incubation (~16 h) at 20°C using 2-3 μg of DNA. We used this lower temperature to obtain more full-length transcripts (Krieg, 1990). Reactions for full-length viral transcripts were supplemented with an additional 4.5 mM of GTP. We electroporated

the RNA transcripts into a mixture of Vero E6 (75%) and BHK-21 (25%) cells containing a total of $2x10^7$ cells per electroporation (Thi Nhu Thao et al., 2020). The Bio-Rad Gene Pulser Xcell Electroporation System was used with the following conditions: 270 V, resistance set to infinity, and capacitance of 950 μF (Xie et al., 2020). Before pulsing, the cells were washed thoroughly and then resuspended in Opti-Mem (Invitrogen). Following a single pulse, cells were allowed to incubate at room temperature for 5 minutes, and we then added fresh growth media before seeding a T-75 flask and placing it at 37°C with 5% CO_2. The cells were monitored daily, and the supernatant was harvested at 25% CPE.

Plaque assays and growth curves

Viral titration was performed on Vero E6 cells by plaque assay. Briefly, serial ten-fold dilutions of each sample were made and then added to confluent monolayers of Vero E6 cells. An overlay containing 0.6% tragacanth gum (Millipore Cat# 104792) was then added; plaques were visualized following formalin fixation and staining with crystal violet. For growth curves, Vero E6 and Calu-3 were infected at a multiplicity of infection (MOI) of 0.05 with each virus. Following one hour of infection, we removed the virus inoculum, washed once with 1x PBS, and added fresh growth media. We then collected supernatant as the 0-day time point and daily after that until 50% cytopathic effect (CPE) was observed, each time replacing the volume taken with fresh growth media. Infectious virus was measured by plaque assay on Vero E6 cells.

Temperature stability

A virus stock containing 10^5 PFU of each virus was prepared in RPMI-1640 containing 2% FBS and 10 mM HEPES. The virus stock was aliquoted into tubes in triplicate or quadruplicate for each time point; a 0-hour time point was collected immediately and stored at −80°C for normalization. At each time point, we placed a subset of the tubes at −80°C for storage until virus titration by plaque assay. The remaining virus was calculated by dividing the individual titers at each time point by the average of the viral titer at the 0-hour time point.

Functional ELISA

Functional ELISA was performed by Sino Biological (Wayne, PA) using purified RBD from WT (Cat: 40592-V08H), A372T (Cat: 40592-V08H36), and N501Y (Cat: 40592-V08H82). RBDs were expressed in HEK293 cells and purified using the polyhistidine tag at the C terminus. Purity was > 85% as measured by SDS-PAGE. The functional ELISA was performed by immobilizing hACE2 protein (Cat # −10108-H05H) at 2 μg/mL (100 μL/well) in PBS, pH 7 at 4°C overnight. The wells were then blocked for one hour in 2% bovine serum albumin in PBS containing 0.1% Tween-20 (PBST). We then probed with varying concentrations of RBD (0.256-4000 ng/mL) diluted in PBST containing 0.1% BSA for one hour. Next, we added goat anti-His tag mAb/HRP diluted to 0.2 μg/mL in PBST containing 0.5% BSA for one hour. Finally, TMB substrate was added, incubated for 20 minutes at room temperature, and then the reaction was stopped with 50 μL of 2 M H_2SO_4. Absorbance values were measured at 450 nm. EC_{50} values were determined by fitting the absorbance values to a Sigmoidal, 4PL nonlinear model using Prism 9 (GraphPad). The experiments were performed in two independent replicates with a total of four technical replicates per group. Statistical comparisons were made using a one-way ANOVA with Dunnett's multiple comparisons test compared to WT SARS-CoV-2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed in Prism version 9 (GraphPad). Viral titers were compared to WT using a two-way ANOVA with Dunnett's multiple comparisons test; a *p value* of less than 0.05 was considered significant. The detection limit for our plaque assays is 2.3 $\log_{10}$ PFU/mL; however, negative values were given an arbitrary value of 0.9 plaques for a ten-fold diluted sample, which corresponds to 2.26 log10 PFU/mL.