

The N501Y spike substitution enhances SARS-CoV-2 infection and transmission

<https://doi.org/10.1038/s41586-021-04245-0>

Received: 9 March 2021

Accepted: 12 November 2021

Published online: 24 November 2021

 Check for updates

Yang Liu^{1,2,6}, Jianying Liu^{2,3,4,6}, Kenneth S. Plante^{2,3,4,6}, Jessica A. Plante^{2,3,4}, Xuping Xie¹, Xianwen Zhang¹, Zhiqiang Ku⁵, Zhiqiang An⁵, Dionna Scharton^{2,3,4}, Craig Schindewolf^{2,4}, Steven G. Widen¹, Vineet D. Menachery^{2,4}, Pei-Yong Shi^{1,2,3,6} & Scott C. Weaver^{2,3,4,6}

The B.1.1.7 variant (also known as Alpha) of SARS-CoV-2, the cause of the COVID-19 pandemic, emerged in the UK in the summer of 2020. The prevalence of this variant increased rapidly owing to an increase in infection and/or transmission efficiency¹. The Alpha variant contains 19 nonsynonymous mutations across its viral genome, including 8 substitutions or deletions in the spike protein that interacts with cellular receptors to mediate infection and tropism. Here, using a reverse genetics approach, we show that of the 8 individual spike protein substitutions, only N501Y resulted in consistent fitness gains for replication in the upper airway in a hamster model as well as in primary human airway epithelial cells. The N501Y substitution recapitulated the enhanced viral transmission phenotype of the eight mutations in the Alpha spike protein, suggesting that it is a major determinant of the increased transmission of the Alpha variant. Mechanistically, the N501Y substitution increased the affinity of the viral spike protein for cellular receptors. As suggested by its convergent evolution in Brazil, South Africa and elsewhere^{2,3}, our results indicate that N501Y substitution is an adaptive spike mutation of major concern.

Since its emergence in 2019⁴, more than 227 million SARS-CoV-2 infections have occurred worldwide, with more than 4.6 million fatalities. Despite exhibiting proofreading activity⁵, coronaviruses mutate frequently⁶, leading to thousands of mutations among the different strains. The most crucial mutations have involved the spike protein, which mediates binding to the angiotensin-converting enzyme 2 (ACE2) receptor via its S1 domain and membrane fusion via the S2 domain⁷. Spike substitutions influence host range, tissue tropism, transmission and pathogenesis⁸.

The first dominant spike substitution, D614G, enhances replication in human airway epithelial (HAE) cells and transmission in animal models^{9–12}. Additional spike mutations, including the N501Y mutation that occurred convergently in the UK, South Africa, Brazil and elsewhere, have been identified. The Alpha variant, which contains seven additional spike substitutions, emerged in September 2020 in the UK. It was estimated to be 70–80% more transmissible than its ancestor¹ and quickly spread worldwide^{2,3}. In vitro studies show enhanced replication of Alpha in human airway cells¹³. Although it is effectively neutralized by vaccine-induced antibodies^{14–16}, and is not consistently associated with more severe COVID-19¹⁷ disease, increased Alpha transmissibility could counteract increasing herd immunity, exacerbating the pandemic.

To investigate mechanisms of increased Alpha transmission, we used reverse genetics, a hamster model and HAE cultures to probe the effects of spike mutations. We determined the phenotypes of individual mutations as well as of the entire spike gene using competition fitness assays. Our results suggest that N501Y is a critical determinant of enhanced infection of the upper airway and transmission.

Results

To investigate phenotypes of individual Alpha spike substitutions, we infected hamsters intranasally with each mutant virus, as well as with chimeric viruses containing the entire spike gene (Extended Data Fig. 1a–c). These mutations were engineered into the USA_WA1/2020 cDNA clone containing the dominant D614G spike substitution¹⁰ using site-directed mutagenesis¹⁸ (resulting in USA_WA1/2020-G614, hereafter referred to as the wild-type (WT) virus). These mutations did not affect Vero cell plaque phenotypes (Extended Data Fig. 1d) and the RNA:plaque-forming units (PFU) ratios were similar to those for the virus expressing wild-type spike (Extended Data Fig. 1e). We tested fitness of viruses expressing the mutant and Alpha spike proteins in an in vivo competition assay by mixing WT viruses expressing mutant and wild-type spike at a PFU ratio of 1:1 for intranasal infection of hamsters (Fig. 1a). When assessed by quantification of viral RNA, these ratios were in a range of 0.72–1.54, presumably owing to minor variations in RNA:PFU ratios (Supplementary Table 1). Although the enrichment of N501Y mutants after competition overlapped with this range in some cases, the increases in viral RNA and typically larger changes in PFU ratio indicated robust fitness advantages. This competition approach has been used extensively to study microbial fitness^{19–22} including for SARS-CoV-2^{10,13}, owing to its precision and reproducibility²³. Major advantages of this approach include the internally controlled replication of each virus, which eliminates host-to-host variation that can reduce experimental power, and the ability to assay strain ratios with more precision than when using individual virus titres. To verify the accuracy of estimates across a range of

¹Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA. ²Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, TX, USA. ³World Reference Center for Emerging Viruses and Arboviruses, University of Texas Medical Branch, Galveston, TX, USA. ⁴Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, TX, USA. ⁵Texas Therapeutics Institute, Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA. ⁶These authors contributed equally: Yang Liu, Jianying Liu, Kenneth S. Plante. ✉e-mail: peshi@utmb.edu; sweaver@utmb.edu

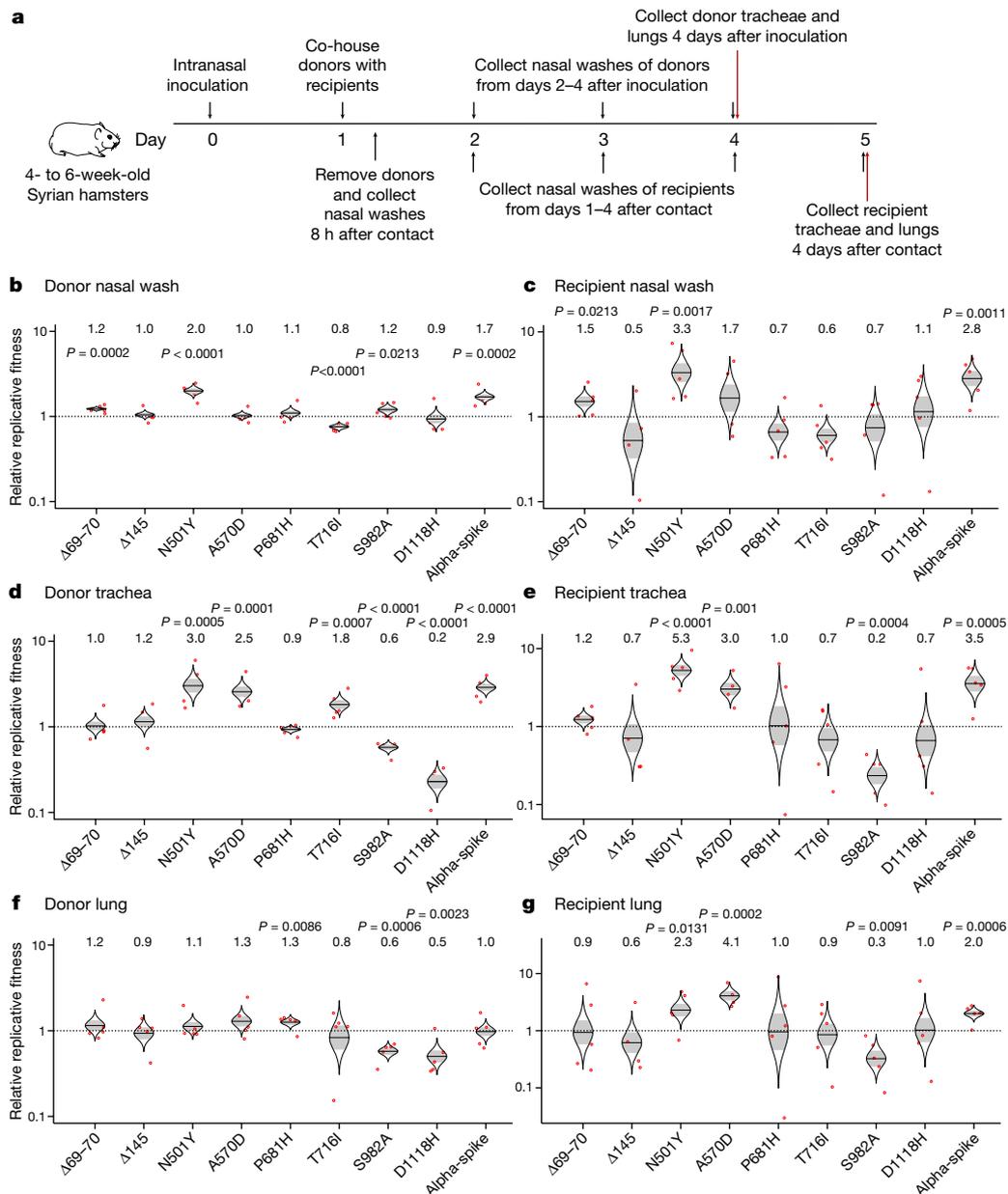


Fig. 1 | The screening of the SARS-CoV-2 Alpha variant spike substitutions in hamsters by competition assay. **a**, Design of the hamster competition fitness studies. The mutant viruses were mixed 1:1 (PFU ratio) with WT virus and inoculated into donor hamsters intranasally at a total titre of 10^5 PFU per hamster. The donor hamsters were co-housed with recipient hamsters 1 day after infection. After 8 h of contact, the donors were removed. All hamsters were subjected to nasal washes daily until 4 days after infection and organs were collected 4 days after inoculation or contact. **b–g**, Competition of different Alpha mutants with WT virus. **b, d, f**, Final:inoculum PFU ratio of WT

virus expressing 8 individual spike mutations or the Alpha spike protein that includes all 8 mutations: nasal washes (**b**), tracheae (**d**) and lungs (**f**) in donor hamsters 4 days after inoculation. **c, e, g**, Final:donor inoculum ratios of WT virus expressing 8 individual spike mutations or the Alpha spike protein: nasal washes (**c**), tracheae (**e**) and lungs (**g**) in recipient hamsters 4 days after contact. In **b–g**, red dots represent individual hamsters ($n = 5$). In catseye plots, the horizontal line shows the mean and the shaded region represents s.e.m. Numbers along the top indicate the relative fitness estimates. *P* values were calculated for the group (strain) coefficient for each linear regression model.

strain ratios, we performed experiments with each mixture using Sanger sequencing, and found a high degree of consistency between PFU- and genome-based estimates (Extended Data Fig. 2). We also evaluated all virus stocks used for experiments by next-generation sequencing to confirm that there were no single-nucleotide polymorphisms (SNPs) with a frequency above 3%; sequencing after two serial passages in hamsters did not detect additional SNPs. These results indicate that the mutant mixtures were genetically stable and that the PFU ratios were robust.

Hamsters were sampled up to four days after infection, following peak viral loads¹⁰, using nasal washes and necropsied tracheae and lungs

(Fig. 1a). One day after infection, hamsters were co-housed with recipients for 8 h to assess transmission; recipients were maintained for an additional 4 days for equivalent sampling. Inocula and hamster samples were analysed by PCR with reverse transcription (RT-PCR) spanning the mutations, and the amplicons were analysed by Sanger sequencing and electropherogram peak analyses to determine mutant:wild-type ratios. Ratio changes during infection reflected the relative fitness of the mutants.

Competition results from donor nasal washes indicated that of the 8 mutations examined, only the deletion of codons 69 and 70 (Δ69–70),

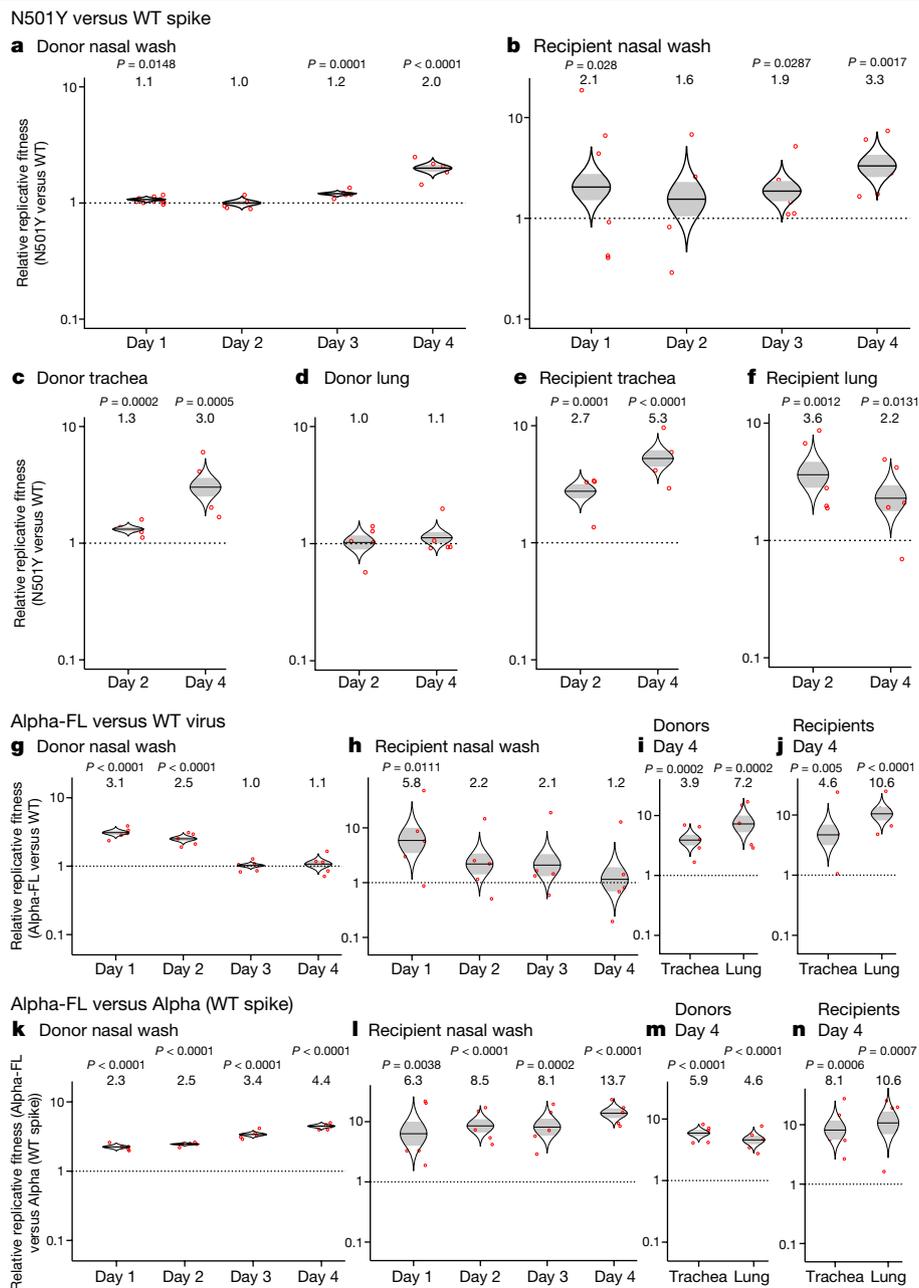


Fig. 2 | The SARS-CoV-2 spike N501Y mutant has a consistent advantage over wild-type spike in upper airway replication in hamsters.

a, b, Competition assays between N501Y and wild-type (WT) spike assessed by sampling nasal washes of donor (**a**) and recipient (**b**) hamsters on days 1 to 4 after inoculation (donors) or contact (recipients). **c–f**, Competition assays between N501Y and wild-type spike in the tracheae and lungs of donor (**c, d**) and recipient hamsters (**e, f**) on days 2 and 4 after inoculation (donors) or contact (recipients). **g–j**, Competition assays between Alpha-FL virus and WT virus inoculated intranasally into hamsters. Nasal washes were taken on days 1 to 4 after inoculation (donors) (**g**) or contact (recipients) (**h**) and tracheae and lungs of donor (**i**) and recipient hamsters (**j**) 4 days after inoculation (donors) or contact (recipients). **k–n**, Competition between Alpha-FL and Alpha (WT spike)

assessed by sampling nasal washes on days 1 to 4 after inoculation (donors) (**k**) or contact (recipients) (**l**) and tracheae and lungs of donor (**m**) and recipient hamsters (**n**) 4 days after inoculation (donors) or contact (recipients). The fitness advantages of the N501Y substitution and Alpha-FL compared with the wild-type both during infection (in donors) and after transmission of the virus to recipients are shown by the changes in ratios between the collected samples and inocula. Red dots represent individual hamsters (all $n = 5$, except $n = 10$ for day-1 nasal wash samples of N501Y group). In catseye plots, the horizontal line represents the mean, the shaded region represents s.e.m. Numbers along the top indicate the relative fitness estimates. P values were calculated for the group (strain) coefficient for each linear regression model.

N501Y and the combined Alpha spike mutations conferred significant fitness advantages, with other substitutions showing inconsistently reduced fitness or no difference from the wild-type spike (Fig. 1b). The same fitness trends were observed in recipient hamsters (Fig. 1c). Examination of donor tracheae showed similar results, except that $\Delta 69-70$ had no effect on fitness (Fig. 1d), whereas A570D increased fitness and

S982A and D1118H decreased fitness. Following transmission, the same overall effects on fitness were maintained in recipient tracheae (Fig. 1e), consistent with results from nasal washes (Fig. 1c). Results from day-4 lungs were distinct from those in the upper airway (nasal washes and tracheae), with only slight fitness effects of the spike mutations; even the combined Alpha-spike mutations did not increase fitness significantly

with respect to replication in donor lungs (Fig. 1f). However, there were significant fitness gains for N501Y, A570D and Alpha-spike mutations in recipients, suggesting more efficient transmission, whereas S982A decreased fitness in recipients (Fig. 1g). We also found that N501Y conferred an advantage across days 1–4 of nasal shedding, beginning on day 3 in donors, that was maintained throughout days 1–4 in recipients (Fig. 2a, b). Among the other mutations, only $\Delta 69-70$ and the Alpha-spike mutations showed consistent fitness gains (Extended Data Fig. 3).

Because it conferred the clearest phenotype among spike mutations, we examined N501Y in trachea and lungs sampled on days 2 and 4. Similar to nasal washes, recipient samples showed a consistent fitness advantage for N501Y at both timepoints, but no difference was detected in donor lungs (Fig. 2c–f). Together, these results demonstrated a clear fitness advantage for N501Y in terms of upper airway shedding during days 1–4, including after intra-cage transmission.

To further ensure that the Sanger sequencing method did not introduce errors in our estimates of mutant ratios, all experiments with N501Y were repeated using next-generation sequencing. As shown in Extended Data Fig. 4, results using next-generation sequencing were very similar and statistically indistinguishable from those obtained with Sanger sequencing.

To examine whether the spike protein was the main determinant of Alpha replication and transmission, we generated a full-length Alpha infectious clone (Alpha-FL) (EPI_ISL_999340) incorporating all Alpha mutations, as well as chimeric viruses in which the spike was swapped (Extended Data Fig. 1c). In competition assays with WT virus, Alpha-FL had a marked fitness advantage in nasal washes from donor hamsters on days 1 and 2, when titres were maximal (Extended Data Fig. 5); similar results were observed in recipient hamsters (Fig. 2g, h). Alpha-FL showed increased fitness in the trachea and lungs of both donors and recipients on day 4, after titres had peaked (Fig. 2i, j, Extended Data Fig. 5). This difference in fitness of Alpha at different sites over time could reflect the complex effects of innate host defences and/or lower stability in the nasal cavity. In competition against Alpha expressing wild-type spike (Alpha (WT spike)), Alpha-FL showed a consistent advantage on days 1–4 in nasal washes of both donor and recipient hamsters (Fig. 2k, l) and on day 4 in tracheae and lungs (Fig. 2m, n). This again indicated that the Alpha spike was a major fitness determinant during infection. Alpha-FL showed consistently reduced fitness in competition assays with Alpha (WT spike) in nasal washes of donors and recipients, and in trachea and lungs (Extended Data Fig. 6), again suggesting that some non-spike Alpha genome regions confer a fitness disadvantage *in vivo*.

We also assayed relative fitness of WT virus expressing N501Y or Alpha spike in recipient hamsters using the PFU ratio in donor hamster day-1 nasal wash (representing the transmitted population) as the starting PFU ratio rather than that of the original donor inoculum, with little effect on fitness estimates (Extended Data Fig. 7).

To examine the fitness effect of the Alpha spike substitutions in human cells, we used HAE and other human cell cultures (Fig. 3a). Replication of WT virus expressing N501Y and complete Alpha-spike mutants was significantly faster than that of WT virus at early stages of infection, as measured by PFU (Fig. 3b), but there was little difference in the number of RNA copies (Fig. 3c). The RNA:PFU ratios were significantly lower for virus expressing N501Y and Alpha spike compared with wild-type spike on days 1–3 (Fig. 3d), suggesting that viruses with these mutant spike proteins show increased specific infectivity in HAE cultures. In competition infection assays, N501Y showed significantly higher fitness at nearly all time points in Vero E6 and Calu-3 cells (Fig. 3e, f). In HAE cells, N501Y had a fitness advantage over wild-type spike and Alpha-FL spike had a fitness advantage over Alpha (WT spike) (Fig. 3g, h, Extended Data Fig. 8). Alpha-FL also showed a consistent fitness advantage over Alpha with wild-type spike, suggesting that, *in vitro*, the Alpha spike is the major fitness determinant of the Alpha variant (Fig. 3i). Overall, these *in vivo* and *in vitro* results suggest that N501Y is the major, but not the only, determinant of fitness of the Alpha variant.

To assess virulence, we examined hamsters infected with each mutant virus individually for weight loss; we found no significant differences between wild-type, N501Y and Alpha spike proteins on weight loss (Extended Data Fig. 9a, b). However, the trend of less weight loss in hamsters infected with WT virus suggested that differences may emerge in experiments with a larger cohort. Nasal washes showed consistently higher titres on days 1–3 and 5 for both N501Y and Alpha spike, but only when measured in PFU. These differences were significant only on day 1 (Extended Data Fig. 9c, d), reflecting the lower sensitivity of individual infections compared with competition assays. Genomic RNA:PFU ratios were higher for virus with wild-type spike than for mutant spike (Extended Data Fig. 9e), again suggesting that the viruses with mutant spike have greater specific infectivity. Additional studies are needed to determine whether these specific infectivity differences are biologically relevant. Trachea and lung samples showed no significant differences in titre when measured as PFU or number of copies of viral RNA (Extended Data Fig. 9f–h). Together, these results again suggest that N501Y results in increased viral fitness in the hamster upper airway.

As the SARS-CoV-2 pandemic has progressed, several mutations have occurred convergently in multiple variants; of the 8 Alpha spike substitutions, only $\Delta 69-70$ and N501Y have evolved convergently, consistent with our fitness findings (Extended Data Table 1). The frequency of N501Y worldwide increased markedly after October 2020 (Fig. 4a). These data corroborate our results indicating that the N501Y substitution confers a major fitness advantage.

We next examined the mechanism behind the fitness advantage of N501Y, focusing on its effect on receptor binding. N501Y is located in the receptor-binding domain (RBD) of spike (Fig. 4b) and is proposed to result in increased interactions with the human ACE2 receptor at K353, increasing the affinity for the receptor. To test this, we performed binding assays using recombinant spike RBD and human ACE2 proteins on a bio-layer interferometry system (Fig. 4c, d). The N501Y substitution increased binding, as indicated by the more than 530-fold improvement in K_D and more than 819-fold improvement in k_{on} . These results are consistent with earlier work demonstrating that spike residue 501 is a critical residue for ACE2 receptor binding²⁴ and suggest that N501Y increases viral fitness for replication in the upper airway, resulting in enhanced transmission, via enhanced spike–receptor interactions.

Discussion

To corroborate epidemiological findings of more efficient transmission of the Alpha variant¹ of SARS-CoV-2, and to identify genetic determinants of this phenotype, we performed experimental infections of hamsters and HAE cells with each of the eight individual spike gene mutations found in Alpha. Of these mutations, only N501Y and the deletion of codons 69 and 70 resulted in consistent fitness advantages for replication in the upper airway of hamsters (with increased shedding in nasal secretions) as well as in human cells. In most experiments, N501Y alone had a phenotype similar to that of the chimeric WT virus containing the complete Alpha spike or the complete Alpha-FL, suggesting that N501Y is the major determinant driving increased transmission. Finally, our studies of chimeric SARS-CoV-2 with swapped spike genes suggest—as in another recent study²⁵—that non-spike genomic regions may interact *in vivo* with innate immune responses to both positively and negatively affect replication in a site- and time-dependent manner.

Although the data are not directly comparable, the typical relative fitness values of 2–3 that we observed with the N501Y substitution in the hamster or HAE models are consistent with reproduction numbers between 1.4 and 1.8 estimated by epidemiological modelling²⁶. This substitution, which is also present in other variants as a result of convergent evolution, is therefore a major adaptive spike mutation of concern. The N501Y mutation also enables SARS-CoV-2 infection of mice²⁷. Its continued evolution in human viral strains only in combination with other spike mutations suggests epistatic interactions. The mechanism of

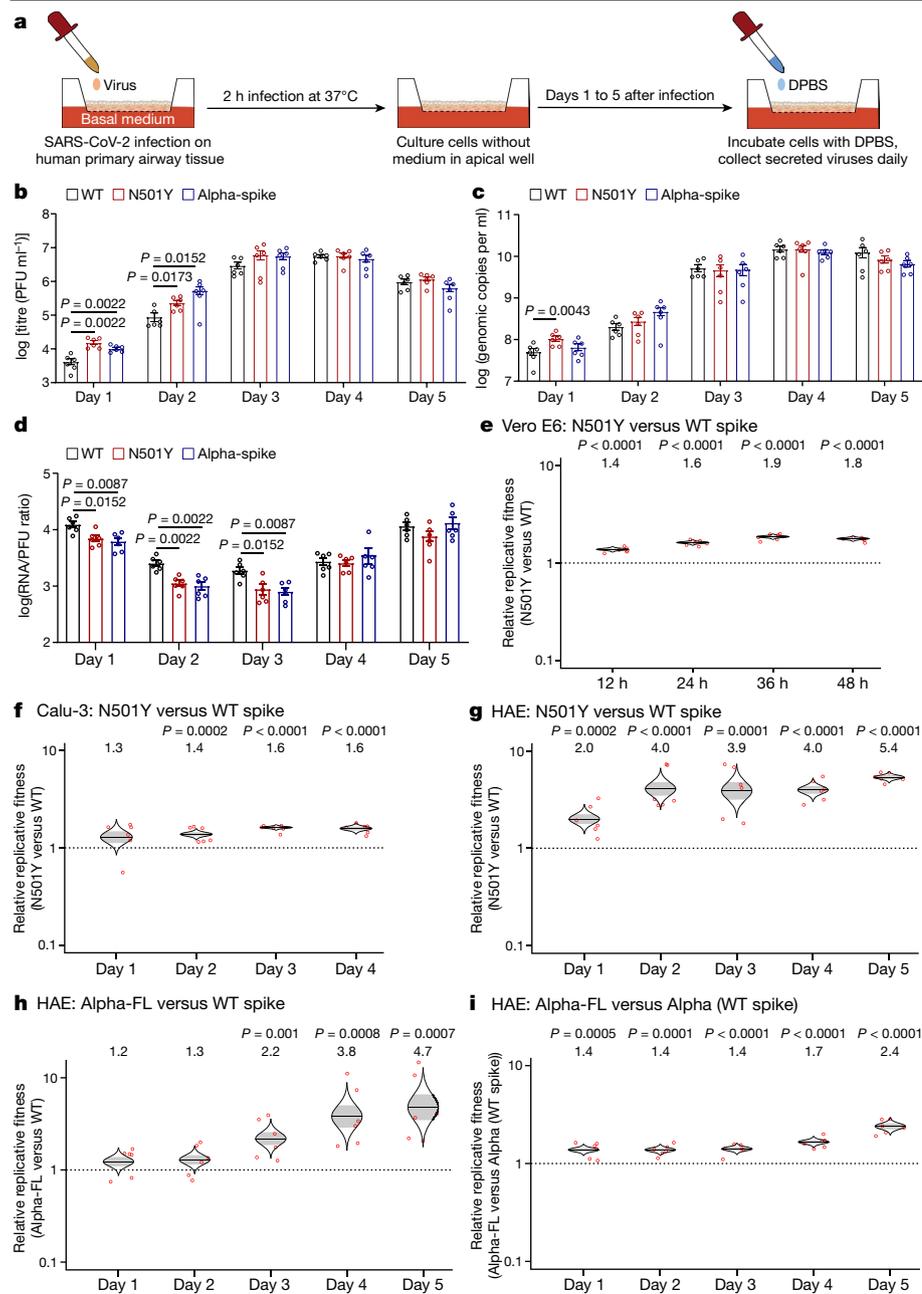


Fig. 3 The spike N501Y mutation enhances viral replication in primary human airway cells and confers an advantage in competition with wild-type spike *in vitro*. **a**, Experimental scheme. Viruses expressing wild-type spike, and N501Y and Alpha spike mutants were inoculated onto HAE cells at a multiplicity of infection (MOI) of 5 PFU per cell. After a 2-h incubation, the culture was washed with Dulbecco's PBS 3 times and maintained for 5 days. The secreted viruses were collected in Dulbecco's phosphate-buffered saline (DPBS) by 30 min incubation at 37 °C every day. **b–d**, Viral replication kinetics and genomic RNA:PFU ratios. The amounts of infectious virus (**b**) and genomic RNA (**c**) were quantified by plaque assay and quantitative PCR with reverse transcription (RT-qPCR), respectively. The genomic RNA:PFU ratio (**d**) was calculated as an indication of virion infectivity. Dots represent individual biological replicates ($n = 6$) pooled from 2 independent experiments. Data are mean \pm s.e.m. Non-parametric two-tailed Mann-Whitney test with Bonferroni correction to account for multiple comparisons. Differences were considered significant if $P < 0.025$. **e–g**, Competition assay between WT virus expressing spike N501Y and WT virus on Vero E6 (**e**), Calu-3 (**f**) and HAE (**g**) cells. **h, i**, Competition assay between Alpha-FL and WT virus (**h**) and Alpha-FL and Alpha virus expressing wild-type spike protein (**i**) on HAE cells. **e–i**, The Vero E6, Calu-3 and HAE cells were infected with mixed viruses at MOI of 0.01, 0.1 and 5, respectively. Red dots represent individual cell cultures ($n = 6$), pooled from two independent experiments. In catseye plots, the horizontal line represents the mean, the shaded region represents s.e.m. Numbers along the top indicate the relative fitness estimates. P values were calculated for the group (strain) coefficient for each linear regression model.

enhanced N501Y fitness is likely to be related to increased replication and shedding in the nasal cavity, leading to more efficient airborne and/or fomite-mediated transmission, both of which are important for the spread of SARS-CoV-2. More detailed transmission experiments are needed to discriminate between these two modes. In addition to upper airway shedding, increased transmission could involve differences in virus stability and infectivity (for example, the 50% infectious dose).

Tests to date with N501Y indicate that this mutation results in minimal changes in susceptibility to immunity, reflected in neutralizing antibody titre (PRNT₅₀) values^{14–16}, suggesting that it does not increase resistance to vaccine protection. These data along with ours suggest that selection of Alpha in nature has occurred owing to more efficient receptor binding rather than immune escape.

Although $\Delta 69-70$ has also been found in a mink variant as well as in wild-type SARS-CoV-2 strains (Extended Data Table 1), it has not evolved convergently in major variants. Furthermore, in the USA, despite its presence since October, 2020, it has not increased substantially in frequency²⁸. These findings corroborate our results indicating that

$\Delta 69-70$ has less effect on transmission than N501Y. However, the frequent co-occurrence of this deletion²⁹ with N501Y suggests possible epistasis. Furthermore, initial epidemiologic estimates from the UK have indicated that N501Y alone increased transmissibility by about 10%, whereas strains with $\Delta 69-70$ showed 70–80% higher transmissibility¹, suggesting that the deletion is also important for Alpha transmission.

Considering that only two of the eight spike substitutions conferred consistent fitness effects on infection of the upper airway and transmission, the origins and roles of the other six mutations deserve consideration. The first of these to appear were deletions of codons 69–70 and 145, followed by all the other except for S982A by March 2020 (Supplementary Table 2). However, from September 2020, all eight mutations—and especially N501Y—began increasing in prevalence (Fig. 4a, Supplementary Table 2). Our data indicate that of these only $\Delta 69-70$ and N501Y have a consistent phenotype, and some of the other mutations reduced fitness in certain experiments. This suggests that these other mutations occurred through drift mechanisms such as founder and hitchhiking effects, and were maintained by linkage to $\Delta 69-70$ and N501Y, or by

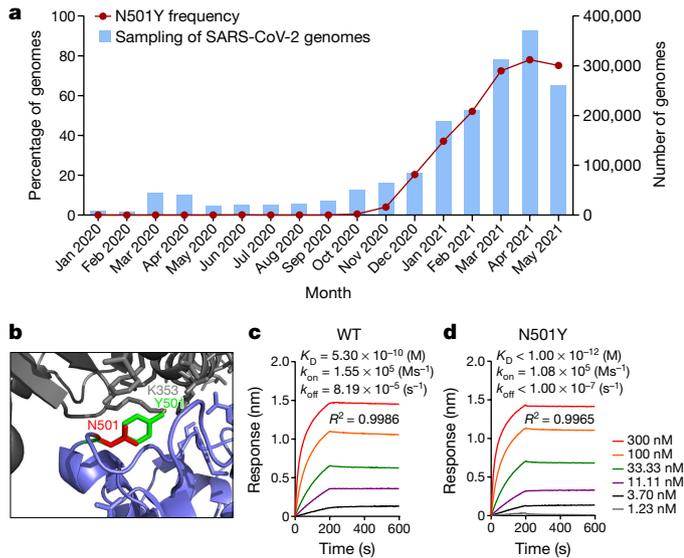


Fig. 4 | The spike N501Y substitution spread quickly and increases spike protein binding affinity for the human ACE2 receptor. **a**, The frequency of the N501Y substitution over time in all genomic SARS-CoV-2 sequences available from the GISAID database worldwide up to May 2021. The blue bars represent the total numbers of SARS-CoV-2 genomes sequenced worldwide. The red line indicates the percentage of N501Y variant in total SARS-CoV-2 genomes. **b**, The predicted binding site of spike N501 and Y501 residues on the human ACE2 receptor. **c**, **d**, Binding affinities of wild-type spike (**c**) and spike (N501Y) (**d**) to the human ACE2 receptor. K_D , dissociation constant; k_{off} , dissociation rate constant; k_{on} , association rate constant. The affinity of ACE2 to the N501Y mutant RBD is below the detection limit and is plotted as $<10^{-12}$. Data are derived from a single experiment.

recombination, which occurs frequently during SARS-CoV-2 replication⁶. Other possibilities include selection by convalescent plasma or antiviral treatments during chronic infections. It is also likely that mutations in other parts of the genome contribute to the fitness of the Alpha variant, as indicated by our results from chimeric virus.

The hamster is a useful model for SARS-CoV-2 infection in many respects¹⁰, but its ACE2 receptors differ slightly the human receptors³⁰. The spike binding studies should be repeated with the hamster receptors to confirm our affinity findings. Although we used infection of primary HAE cells to address this limitation, the use of a primate model would further substantiate our in vivo findings.

In conclusion, we used authentic SARS-CoV-2 to demonstrate that only two out of eight spike substitutions in the Alpha variant are major contributors its enhanced transmission, and the N501Y substitution appears to be the major determinant. This mutation is present in several global regions and should be closely monitored to anticipate public health measures needed to control the spread of SARS-CoV-2.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-04245-0>.

- Leung, K., Shum, M. H., Leung, G. M., Lam, T. T. & Wu, J. T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill* **26**, 2002106 (2021).
- Galloway, S. E. et al. Emergence of SARS-CoV-2 B.1.1.7 lineage—United States, December 29, 2020–January 12, 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 95–99 (2021).
- Claro, I. M. et al. Local transmission of SARS-CoV-2 lineage B.1.1.7, Brazil, December 2020. *Emerg. Infect. Dis.* **27**, 970–972 (2021).
- Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- Smith, E. C., Blanc, H., Surdel, M. C., Vignuzzi, M. & Denison, M. R. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog.* **9**, e1003565 (2013).
- Gribble, J. et al. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* **17**, e1009226 (2021).
- Huang, Y., Yang, C., Xu, X. F., Xu, W. & Liu, S. W. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* **41**, 1141–1149 (2020).
- Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310 e1220 (2020).
- Hou, Y. J. et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).
- Plante, J. A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
- Chan, J. F. et al. 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 (2020).
- Zhou, B. et al. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* **592**, 122–127 (2021).
- Touret, F. et al. Replicative fitness of a SARS-CoV-2 2019/501Y.V1 variant from lineage B.1.1.7 in human reconstituted bronchial epithelium. *mBio* **12**, e0085021 (2021).
- Chen, R. E. et al. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nat. Med.* **27**, 717–726 (2021).
- Xie, X. et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* **27**, 620–621 (2021).
- Liu, Y. et al. Neutralizing activity of BNT162b2-elicited serum—preliminary report. *N. Engl. J. Med.* **384**, 1466–1468 (2021).
- Frampton, D. et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet Infect. Dis.* **21**, 1246–1256 (2021).
- Xie, X. et al. An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe* **27**, 841–848.e3 (2020).
- Wiser, M. J. & Lenski, R. E. A comparison of methods to measure fitness in *Escherichia coli*. *PLoS ONE* **10**, e0126210 (2015).
- Grubaugh, N. D. et al. Genetic drift during systemic arbovirus infection of mosquito vectors leads to decreased relative fitness during host switching. *Cell Host Microbe* **19**, 481–492 (2016).
- Bergren, N. A. et al. ‘Submergence’ of Western equine encephalitis virus: evidence of positive selection argues against genetic drift and fitness reductions. *PLoS Pathog.* **16**, e1008102 (2020).
- Coffey, L. L. & Vignuzzi, M. Host alternation of chikungunya virus increases fitness while restricting population diversity and adaptability to novel selective pressures. *J. Virol.* **85**, 1025–1035 (2011).
- Liu, J. et al. Role of mutational reversions and fitness restoration in Zika virus spread to the Americas. *Nat. Commun.* **12**, 595 (2021).
- Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *J. Virol.* **94**, e00127-20 (2020).
- Thorne, L. G. et al. Evolution of enhanced innate immune evasion by the SARS-CoV-2 B.1.1.7 UK variant. Preprint at <https://doi.org/10.1101/2021.06.06.446826> (2021).
- Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).
- Gu, H. et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **369**, 1603–1607 (2020).
- Larsen, B. B. & Worobey, M. Identification of a novel SARS-CoV-2 spike 69–70 deletion lineage circulating in the United States. *virological.org* <https://virological.org/t/identification-of-a-novel-sars-cov-2-spike-69-70-deletion-lineage-circulating-in-the-united-states/577> (2020).
- Kemp, S. A. et al. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep.* **35**, 109292 (2021).
- Chan, J. F. et al. Simulation of the clinical and pathological manifestations of coronavirus disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility. *Clin. Infect. Dis.* **71**, 2428–2446 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Article

Methods

Ethics statement

Hamster studies were performed in accordance with the guidance for the Care and Use of Laboratory Animals of the University of Texas Medical Branch (UTMB). The protocol was approved by the Institutional Animal Care and Use Committee at UTMB. All the hamster operations were performed under anaesthesia by isoflurane to minimize animal suffering.

Animals and Cells

The Syrian golden hamsters (HsdHan:AURA strain) were purchased from Envigo. African green monkey kidney epithelial Vero E6 cells (ATCC) were grown in Dulbecco's modified Eagle's medium (DMEM; Gibco/Thermo Fisher) with 10% fetal bovine serum (FBS; HyClone Laboratories) and 1% antibiotic/streptomycin (Gibco). Human lung adenocarcinoma epithelial Calu-3 cells (ATCC) were maintained in a high-glucose DMEM containing sodium pyruvate and GlutaMAX (Gibco) with 10% FBS and 1% penicillin/streptomycin at 37 °C with 5% CO₂. The EpiAirway system is a primary human airway 3D tissue model purchased from MatTek Life Science (Ashland). This EpiAirway system was maintained with the provided culture medium at 37 °C with 5% CO₂ following the manufacturer's instruction. All other culture medium and supplements were purchased from Thermo Fisher Scientific. All cell lines were verified and tested negative for mycoplasma.

Generation of SARS-CoV-2 mutant viruses

Generation of individual point mutant viruses and WT virus expressing Alpha spike. Individual point mutations in the spike gene (Δ 69–70, Δ 145, N501Y, A570D, P681H, T716I, S982A and D1118H) or 8 spike mutations were introduced into a subclone pcc1-CoV-2-F5–7 containing the spike gene of SARS-CoV-2-D614G¹⁷ by overlap fusion PCR.

Generation of Alpha-FL and Alpha-FL (WT spike). The sequence for constructing Alpha-FL and Alpha-non-spike was downloaded from GISAID under accession ID EPI_ISL_999340. In brief, individual point mutations (NSP3: P153L, T183I, A890D, I1412T; NSP6: SGF106-108del; NSP12: P323L; Spike: HV69-70del, Y145del, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H; ORF8: Q27stop, R52I, Y73C, S84L; N: D3L, R203K, G204R, S235F) were introduced into the three subclones pcc1-Alpha-F1–3, puc57-Alpha-F4 and pcc1-Alpha-F5–7 by overlap fusion PCR. To generate the Alpha-FL (WT spike), the spike gene of Alpha-FL was replaced with the spike gene of SARS-CoV-2-D614G. All the primers used for the construction are listed in Supplementary Table 3.

The full-length infectious clones of the variant SARS-CoV-2 viruses were assembled by *in vitro* ligation of contiguous DNA fragments the protocol as previously described³¹. *In vitro* transcription was then performed to synthesize full-length genomic RNA. For recovery of the mutant viruses, the RNA transcripts were electroporated into Vero E6 cells. The viruses from electroporated cells were collected at 40 h after electroporation and served as P0 stocks. All viruses were passaged once on Vero E6 cells for subsequent experiments and subjected to next-generation sequencing after RNA extraction to confirm the introduction and stability of substitutions. Viral titres were determined by plaque assay on Vero E6 cells. All virus preparation and experiments were performed in a biosafety level 3 facility. Viruses and plasmids are available from the World Reference Center for Emerging Viruses and Arboviruses (WRCEVA) at the UTMB.

RNA extraction, RT-PCR, RT-qPCR and Sanger sequencing

Cell culture supernatants or clarified tissue homogenates were mixed with a fivefold excess of TRIzol LS Reagent (Thermo Fisher Scientific). Viral RNAs were extracted according to the manufacturer's instructions. The extracted RNAs were dissolved in 20 μ l nuclease-free water. For the validation of mutant viruses, two microlitres of RNA samples were

used for reverse transcription by using the SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific) with random hexamer primers. Nine DNA fragments flanking the entire viral genome were amplified by PCR. The resulting DNAs were cleaned up by the QIAquick PCR Purification Kit, and the genome sequences were determined by Sanger sequencing at GENEWIZ (South Plainfield).

To quantify mutant ratios for competition assays, RT-PCR products were amplified from extracted RNA using a SuperScript III One-Step RT-PCR kit (Invitrogen). A 20- μ l reaction was assembled in PCR 8-tube strips through the addition of 10 μ l 2 \times reaction mix, 0.4 μ l SuperScript III RT/Platinum Taq Mix, 0.8 μ l Forward Primer (10 μ M), 0.8 μ l reverse primer (10 μ M), 2 μ l RNA, and 6 μ l RNase-free water. Primers are listed in Supplementary Table 3. Reverse transcription and amplification were completed using the following protocol: (1) 55 °C, 30 min; 94 °C, 2 min; (2) 94 °C, 15 s; 60 °C, 30 s; 68 °C, 1 min; 40 cycles; (3) 68 °C, 5 min; (4) indefinite hold at 4 °C. The presence and size of the desired amplicon was verified with 2 μ l of PCR product on an agarose gel. The remaining 18 μ l were purified by a QIAquick PCR Purification kit (Qiagen) according to the manufacturer's protocol.

Sequences of the purified RT-PCR products were generated using a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems). The sequencing reactions were purified using a 96-well plate format (EdgeBio) and analysed on a 3500 Genetic Analyzer (Applied Biosystems). The peak electropherogram height representing each mutation site and the proportion of each competitor was analysed using the QSVAnalysr program³².

Next generation sequencing

The competition results between N501Y and WT in hamsters and HAE cells derived via Sanger sequencing were confirmed using next generation sequencing methods. In brief, the viral RNAs from the N501Y–WT competition groups were used for a specific one-step RT-PCR that containing the A23063T mutation site, primers for the RT-PCR are listed in Supplementary Table 3. The PCR products were then purified by a QIAquick PCR Purification kit (Qiagen) according to the manufacturer's protocol. Dual-indexed adapter sequences (New England Biolabs) were added with 5 cycles of PCR. Samples were pooled and sequenced by the UTMB Next Generation Sequencing Core on an Illumina MiSeq Mid-Output flow cell with the paired-end 150 base protocol. The reads were filtered for *Q*-scores of 37 at the A23063T mutation site and adjacent bases and counted. The input ratios and output ratios of the two viruses were obtained and used for the relative replicative fitness analysis.

Plaque assay

Approximately 1.2×10^6 Vero E6 cells were seeded to each well of 6-well plates and cultured at 37 °C, 5% CO₂ for 16 h. Virus was serially diluted in DMEM with 2% FBS and 200 μ l diluted viruses were transferred onto the monolayers. The viruses were incubated with the cells at 37 °C with 5% CO₂ for 1 h. After the incubation, overlay medium was added to the infected cells per well. The overlay medium contained DMEM with 2% FBS, 1% penicillin/streptomycin and 1% sea-plaque agarose (Lonza). After a 2-day incubation, plates were stained with neutral red (Sigma-Aldrich) and plaques were counted on a light box. The detection limitation of the plaque assay was 10 PFU ml⁻¹.

Quantitative real-time RT-PCR assays

RNA copies of SARS-CoV-2 samples were detected by quantitative real-time RT-PCR (RT-qPCR) assays were performed using the iTaq SYBR Green One-Step Kit (Bio-Rad) on the QuantStudio 7 Flex system (Applied Biosystems) following the manufacturers' protocols. Primers are listed in Supplementary Table 3. The absolute quantification of viral RNA was determined by a standard curve method using an RNA standard (in vitro transcribed 3,480 bp containing genomic nucleotide positions 26,044 to 29,883 of SARS-CoV-2 genome).

Viral infection of cell lines

Approximately 3×10^5 Vero E6 or Calu-3 cells were seeded onto each well of 12-well plates and cultured at 37 °C, 5% CO₂ for 16 h. N501Y or Alpha-spike mutants were mixed with wild-type viruses and inoculated onto Vero E6 and Calu-3 cells at a MOI of 0.01 and 0.1, respectively. The mixed virus was incubated with the cells at 37 °C for 2 h. After infection, the cells were washed with DPBS 3 times to remove any un-attached virus. One millilitre of culture medium was added into each well for the maintenance of the cells. At each time point, 100 µl of culture supernatants were lysed in TRIzol LS reagent for the detection of competition assay, and 100 µl of fresh medium was added into each well to replenish the culture volume. The cells were infected in triplicate for each group of viruses. All samples were stored at -80 °C until analysis.

Viral infection in a primary human airway cell culture model

The EpiAirway system is a primary human airway 3D mucociliary tissue model consisting of normal, human-derived tracheal/bronchial epithelial cells. For viral replication kinetics, wild-type, N501Y or Alpha-spike mutant viruses were inoculated onto the culture at a MOI of 5, respectively. After 2 h infection at 37 °C with 5% CO₂, the inoculum was removed, and the culture was washed 3 times with DPBS. The infected epithelial cells were maintained without any medium in the apical well, and medium was provided to the culture through the basal well. The infected cells were incubated at 37 °C, 5% CO₂. From 1–5 days, 300 µl of DPBS were added onto the apical side of the airway culture and incubated at 37 °C for 30 min to elute the released viruses. Titres of samples taken immediately after washing the inoculum were not plotted because initial experiments showed the presence of very little or no detectable virus. All virus samples in DPBS were stored at -80 °C.

Spike RBD and ACE2 binding

The human ACE2 protein was purchased from Sino Biological (catalogue (cat.) no. 10108-H08H) and the human IgG1 Fc-tagged RBD proteins were made in-house as previously described³³. The affinity measurement was performed on the ForteBio Octet RED 96 system (Sartorius). In brief, wild-type or N501Y mutant RBD proteins (20 µg ml⁻¹) were captured onto protein A biosensors for 300 s. The loaded biosensors were then dipped into the kinetics buffer for 10 s for adjustment of baselines. Subsequently, the biosensors were dipped into serially diluted (1.23–300 nM) human ACE2 protein for 200 s to record association kinetics and then dipped into kinetics buffer for 400 s to record dissociation kinetics. Kinetic buffer without ACE2 was used to correct the background. The Octet Data Acquisition 9.0 software was used to collect affinity data. For fitting of K_D values, Octet Data Analysis software V11.1 was used to fit the curve by a 1:1 binding model and use of the global fitting method.

Hamster infections

Four- to six-week-old male golden Syrian hamsters, strain HsdHan:AURA (Envigo), were randomly selected (although the investigators were not blinded to the assignments) and inoculated intranasally with 100 µl SARS-CoV-2. For transmission competition assays, 5 donor hamsters received a mixture containing 5×10^4 PFU of WT virus and an equal amount of variant virus. One day later, 1 infected donor hamster was co-housed with one naive hamster for 8 h (10 pairs for N501Y group, 5 pairs for the other groups) and the donors were returned to their cages. For virus replication assays, the hamsters received DMEM with 2% FBS and 1% penicillin/streptomycin (Mock, $n = 4$), WT virus ($n = 9$), N501Y mutant virus ($n = 9$), Alpha-spike mutant virus ($n = 9$) at a dose of 10^4 PFU per hamster. The infected hamsters were weighed and monitored for signs of illness daily. Nasal washes were collected in 400 µl sterile DPBS at indicated time points. For the transmission competition study, hamsters were humanely euthanized for organ collections at 2 or 4 days after inoculation (donors) or after contact (recipients). For the virus replication study, 4 hamsters in each group were euthanized for organ collection

2 days after infection and others were euthanized 7 days after infection. The collected tracheae and lungs were placed in a 2-ml homogenizer tube containing 1 ml of maintenance medium (DMEM supplemented with 2% FBS and 1% penicillin/streptomycin) and stored at -80 °C. Samples were subsequently thawed, lung or tracheae were homogenized for 1 min at 26 s⁻¹, and debris was pelleted by centrifugation for 5 min at 16,100g. Infectious titres were determined by plaque assay. Genomic RNAs were quantified by RT-qPCR. Ratios of mutant:wild-type RNA were determined by RT-PCR with quantification of Sanger peak heights.

Competition assays

For the competition assay on Vero E6, Calu-3 and primary human epithelial airway (HAE) cells, WT and mutant viruses were mixed and inoculated onto the cells at a final MOI of 0.01, 0.01 and 5, respectively. For the competition in hamsters, 100 µl mixtures of WT and variant viruses (total 1×10^5 PFU per hamster) were inoculated intranasally into 4–6-week-old Syrian hamsters. On 1–4 days after inoculation (donors) or after contact (recipients), the infected hamsters were sampled for competition detection. An aliquot of the inoculum used for both hamster and cell infections was back-titrated to confirm the initial ratio of viruses. All samples were stored in -80 °C freezer prior to analysis.

Validation of competition assay by Sanger sequencing

To validate the consistency and accuracy of competition assay by Sanger sequencing, the WT and mutant viruses were mixed at ratios of 10:1, 5:1, 3:1, 1:1, 1:3, 1:5 and 1:10 based on their PFU titres (total 10^6 PFU viruses) or mixed with 10^6 , 10^5 , 10^4 , 10^3 and 10^2 PFU of the two viruses at a ratio of 1:1. The total RNA of these mixed viruses was isolated and amplified by RT-PCR. The ratios of WT:mutant were calculated by the peak heights of Sanger sequencing. Data were analysed by linear regression with correlation coefficients (r) and significance (P).

Statistics

Male hamsters were randomly allocated into different groups. The investigators were not blinded to allocation during the experiments or to the outcome assessment. No statistical methods were used to predetermine sample size. Dead hamster were excluded from sample collections and data analysis. Descriptive statistics have been provided in the figure legends. For in vitro replication kinetics, Kruskal–Wallis analysis of variance was conducted to detect any significant variation among replicates. If no significant variation was detected, the results were pooled for further comparison. Differences between continuous variables were assessed with a two-tailed non-parametric Mann–Whitney test. The PFU, genomic copies and RNA:PFU ratios were analysed using non-transformed values. The weight loss data are shown as mean \pm s.d. and statistically analysed using two-way ANOVA with Turkey's multiple comparison. Analyses were performed in Prism version 9.0 (GraphPad).

For virus competition experiments, relative replicative fitness values for mutant virus compared to G614 WT virus were analysed according to $w = (f_0/i_0)$, where w is relative fitness, i_0 is the initial mutant:WT ratio and f_0 is the final mutant:WT ratio after competition. Sanger sequencing (initial timepoint T_0) counts for each virus strain being compared were based on average counts over three replicate samples of inocula per experiment, and post-infection (timepoint T_1) counts were taken from samples of individual subjects. For cell culture samples, multiple experiments were performed, so that f_0/i_0 was clustered by experiment. To model f_0/i_0 , the ratio T_0/T_1 was found separately for each subject in each strain group, log (base 10) transformed to an improved approximation of normality and modelled by analysis of variance with relation to group, adjusting by experiment when appropriate to control for clustering within experiment. Specifically, the model was of the form $\log_{10}(\text{Ratio at } T_1/\text{Ratio at } T_0) - \text{Experiment} + \text{Group}$. Fitness ratios between the two groups (the model's estimate of $w = (f_0/i_0)$) were assessed per the coefficient of the model's Group term, which was transformed to the original scale as $10^{\text{coefficient}}$. This modelling approach compensates for any correlation due to

Article

clustering within experiment similarly to that of corresponding mixed effect models and is effective since the number of experiments was small. Statistical analyses were performed using R statistical software (R Core Team, version 3.6.1). In all statistical tests, two-sided $\alpha = 0.05$. Catseye plots³⁴, which illustrate the normal distribution of the model-adjusted means, were produced using the catseyes package³⁵.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The sequences, strain dates and locations of collection of SARS-CoV-2 variants include USA_WA1/2020 (MT020880 from the NCBI database) and the Alpha variant (EPI_ISL_999340 from the GISAID database). All other information is available upon request. Source data are provided with this paper.

31. Xie, X. et al. An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe* **27**, 841–848. e843 (2020).
32. Carr, I. M. et al. Inferring relative proportions of DNA variants from sequencing electropherograms. *Bioinformatics* **25**, 3244–3250 (2009).
33. Ku, Z. et al. Molecular determinants and mechanism for antibody cocktail preventing SARS-CoV-2 escape. *Nat. Commun.* **12**, 469 (2021).
34. Cumming, G. The new statistics: why and how. *Psychol. Sci.* **25**, 7–29 (2014).

35. Andersen, C. Catseyes: create catseye plots illustrating the normal distribution of the means. R package version 0.2.3 (2019).

Acknowledgements This research was supported by grants from the NIA and NIAID of the NIH (AI153602 and AG049042 to V.D.M., and R24AI120942 to S.C.W.); a STARs award provided by the University of Texas System to V.D.M.; a Welch Foundation grant AU-0042-20030616; and Cancer Prevention and Research Institute of Texas (CPRIT) grants RP150551 and RP190561 to Z.A. P.-Y.S. was supported by NIH grant AI134907 and UL1TR001439, and by awards from the Sealy & Smith, Kleberg, John S. Dunn, Amon G. Carter, Gilson Longenbaugh and Summerfield Robert foundations. J.L. was supported by a James W. McLaughlin Fellowship Fund.

Author contributions Conceptualization: Y.L., J.L., K.S.P., J.A.P., X.X., V.D.M., P.-Y.S. and S.C.W. Methodology: Y.L., J.L., K.S.P., J.A.P., X.X., X.Z., Z.K., Z.A., D.S., C.S., S.G.W., V.D.M., P.-Y.S. and S.C.W. Investigation: Y.L., J.L., K.S.P., J.A.P., V.D.M., P.-Y.S. and S.C.W. Resources: X.X., Z.K. and Z.A. Data curation: Y.L., J.L., K.S.P., J.A.P., V.D.M., P.-Y.S. and S.C.W. Writing, original draft: Y.L., P.-Y.S. and S.C.W. Writing, review and editing: Y.L., J.L., K.S.P., J.A.P., V.D.M., P.-Y.S. and S.C.W. Supervision: P.-Y.S. and S.C.W. Funding acquisition: Z.A., V.D.M., P.-Y.S. and S.C.W.

Competing interests X.X., V.D.M. and P.-Y.S. have filed a US provisional patent 63/000,713 'Reverse genetic system of SARS-CoV-2', filed on 27 March 2020, on the reverse genetic system and reporter SARS-CoV-2. The other authors declare no competing interests.

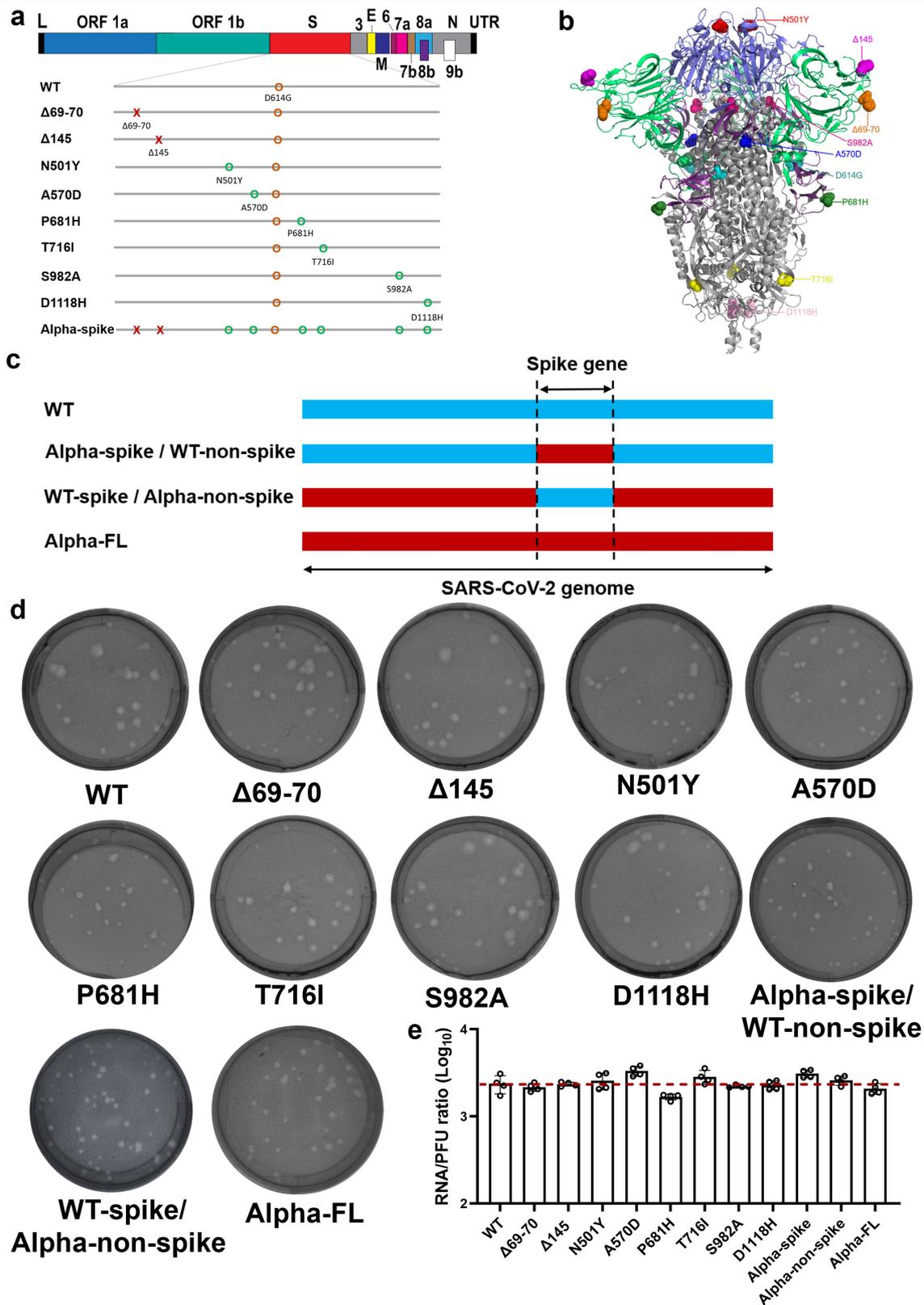
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-04245-0>.

Correspondence and requests for materials should be addressed to Pei-Yong Shi or Scott C. Weaver.

Peer review information Nature thanks Stanley Perlman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

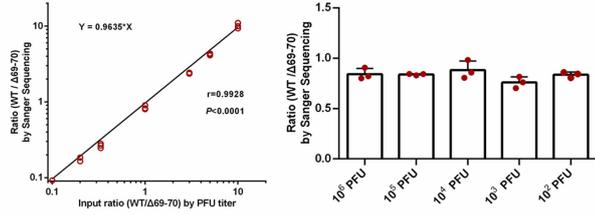
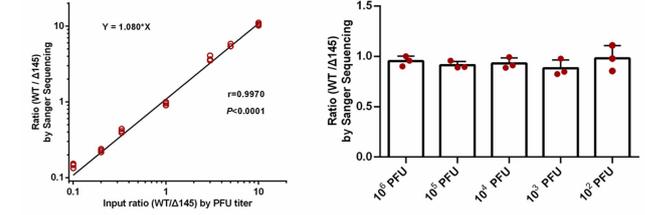
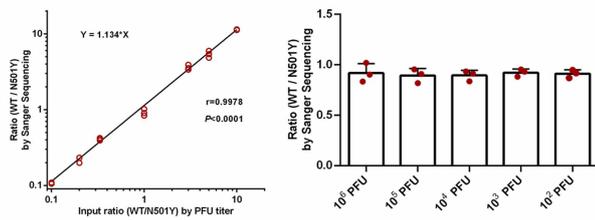
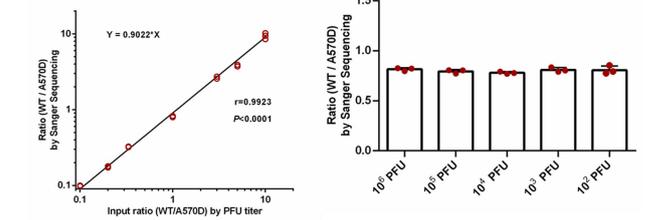
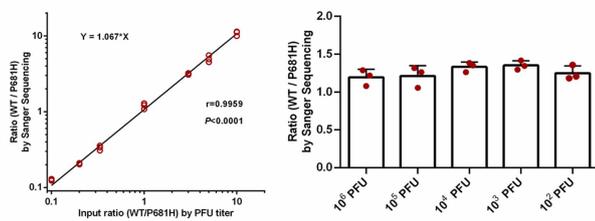
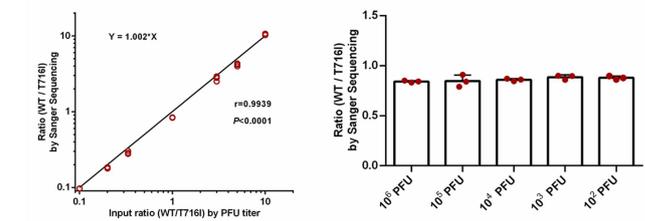
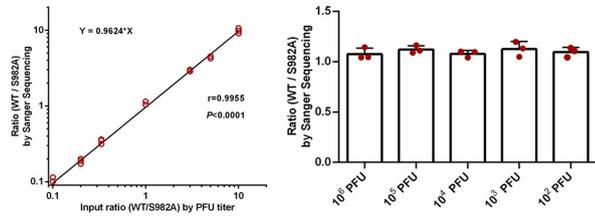
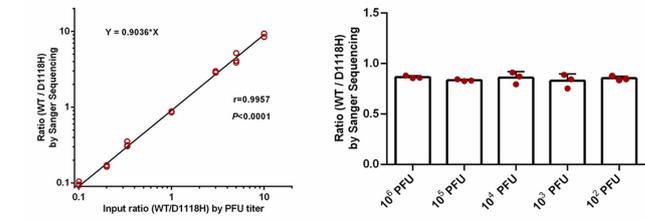
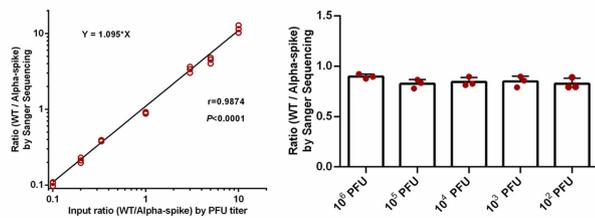
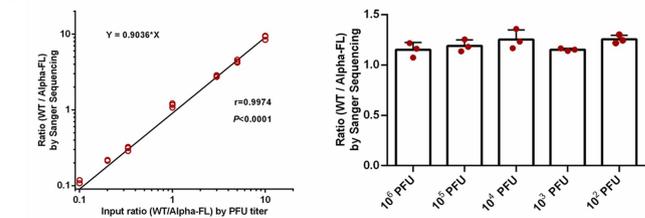
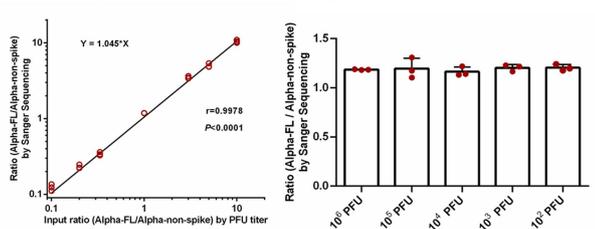
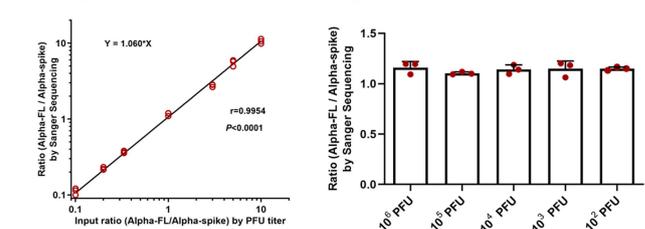


Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | The construction and morphology of SARS-CoV-2 mutants. a, The reverse genetic construction design of all the individual and combined mutations on the wt backbone (USA_WA1/2020 spike D614G mutant). L, leader sequence; S, Spike gene; Open reading frames, ORFs; E, envelope glycoprotein gene; M, membrane glycoprotein gene; N, nucleocapsid gene; UTR, untranslated region. **b,** The location of all 8 Alpha B.1.1.7 substitutions and D614G on the SARS-CoV-2 spike protein trimer. **c,** The schematic designs of the chimeric SARS-CoV-2 viruses. The wt is the USA_WA1/2020 strain with the spike D614G substitution. The Alpha-spike/WT-non-spike contains the spike gene from the Alpha variant in the WT backbone. The WT-spike/Alpha-non-spike has the WT spike gene with all non-spike gene regions from the Alpha variant. The Alpha-FL is the infectious clone of full-length Alpha variant. **d,** The morphologies of all the rescued mutant SARS-

CoV-2 variants. The plaques were stained 2.5 days post-infection of Vero E6 cells. **e,** The virus stocks of individual recombinant SARS-CoV-2 mutants, as well as Alpha-FL were quantified for their genomic RNA and infectious plaque-forming units by RT-qPCR and plaque assay, respectively. The genomes/PFU ratio was calculated to determine specific infectivity. Dots represent individual biological replicates from 4 different aliquots of viruses. The red dotted line indicates the average genome:PFU ratio of WT. The values in the graph represent the mean \pm standard deviation. A two-tailed non-parametric Mann-Whitney test was used to determine significant differences between the variants and wt. *P* values were adjusted using the Bonferroni correction to account for multiple comparisons. No statistical differences were detected among groups.

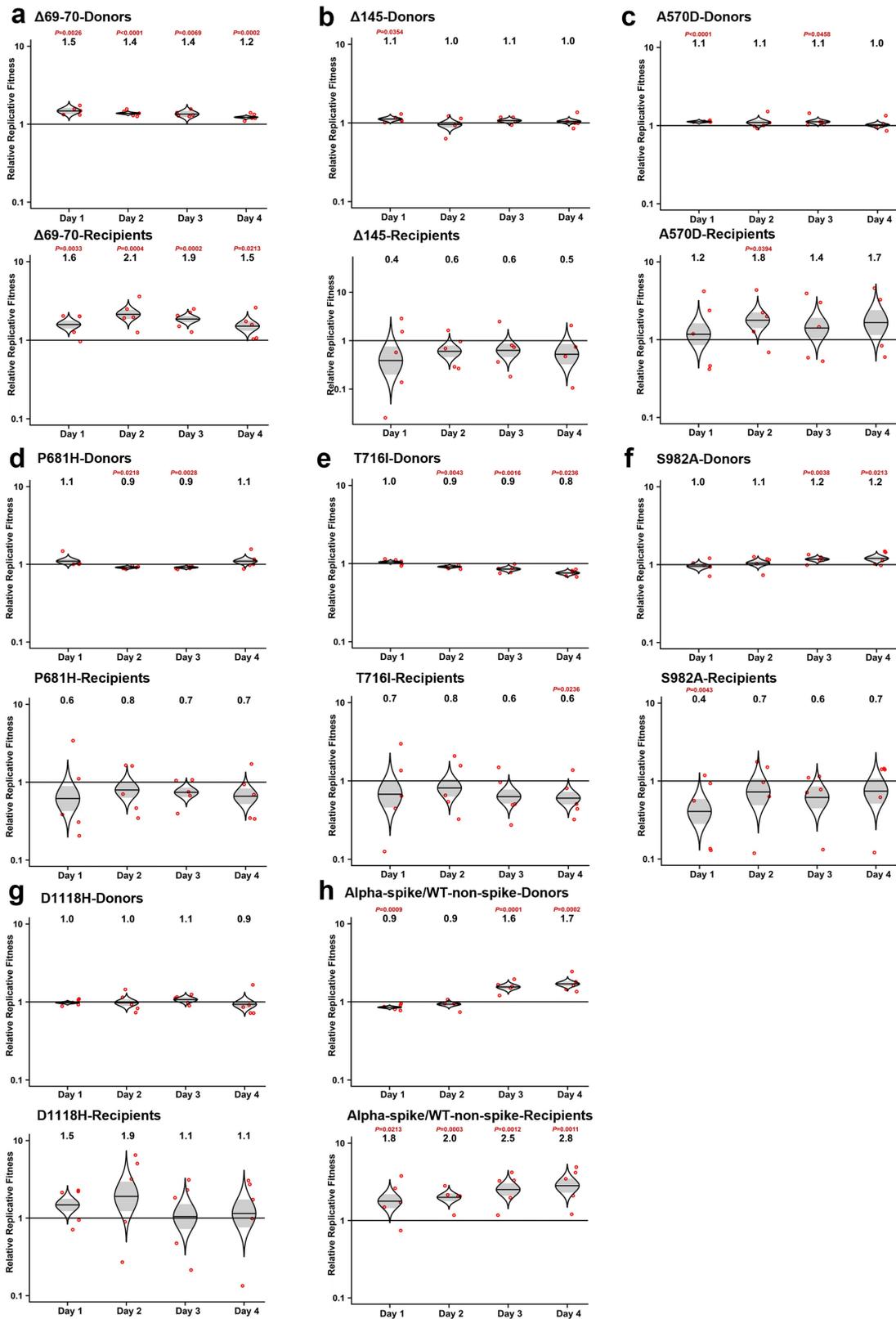
a WT vs Δ69-70**b WT vs Δ145****c WT vs N501Y****d WT vs A570D****e WT vs P681H****f WT vs T716I****g WT vs S982A****h WT vs D1118H****i WT vs Alpha-spike/WT-non-spike****j WT vs Alpha-FL****k Alpha-FL vs WT-spike/Alpha-non-spike****l Alpha-FL vs Alpha-spike/WT-non-spike**

Extended Data Fig. 2 | See next page for caption.

Article

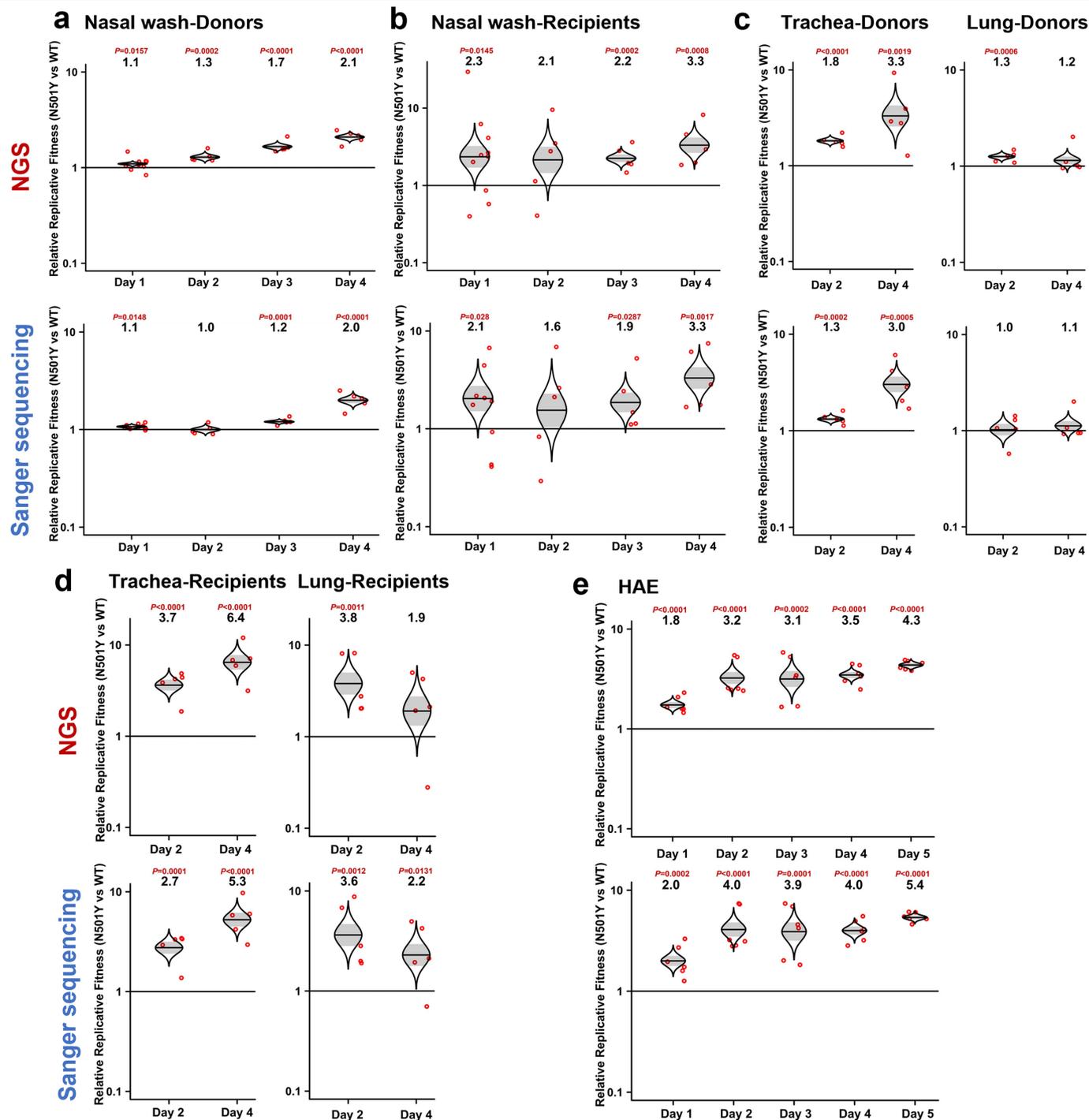
Extended Data Fig. 2 | Validation of SARS-CoV-2 ratio determination by Sanger sequencing. a-I (Left panels), The correlation between input PFU ratios and output RT-PCR amplicon ratios determined by Sanger sequencing. WT and Mutant viruses were mixed at PFU ratios of 10:1, 5:1, 3:1, 1:1, 1:3, 1:5, or 1:10. Total RNAs of the virus mixtures were extracted and amplified by RT-PCR. The WT/Mutant ratios were calculated by the peak heights of Sanger sequencing. Data were analyzed by linear regression with correlation coefficients (r) and significance (P). Symbols represent individual replicates, and the solid line represents the fitted line. **a-I (Right panels)**, Assay range

evaluation. The ratio of wt/Mutant virus mixtures calculated from Sanger sequencing were consistent when using a wide range of virus amounts. The WT/Mutant viruses were mixed at 1:1 PFU ratio. The total titers of the mixed viruses were 10^6 , 10^5 , 10^4 , 10^3 , and 10^2 PFU. The total RNA of virus mixture was isolated and amplified by RT-PCR. The wt/Mutant ratios were calculated by the peak heights from Sanger sequencing. Symbols represent individual replicates, bar heights represent the mean, and error bars represent the standard deviation. **a-I**, Data are derived from a single experiment conducted in triplicate ($n = 3$).



Extended Data Fig. 3 | The competition of other SARS-CoV-2 mutants with wt in hamsters. a–h, Eight SARS-CoV-2 spike mutants: $\Delta 69-70$ (a), $\Delta 145$ (b), A570D (c), P681H (d), T716I (e), S982A (f), D1118H (g) and Alpha-spike/WT-non-spike (h) were mixed with wt virus at a PFU ratio of 1:1. The mixture was then inoculated intranasally into donor hamsters and transmitted to the recipient hamsters following the scheme in Figure 1a. The total titer for infection was 10^5 PFU per hamster. The ratios of the mutant:wt in the nasal

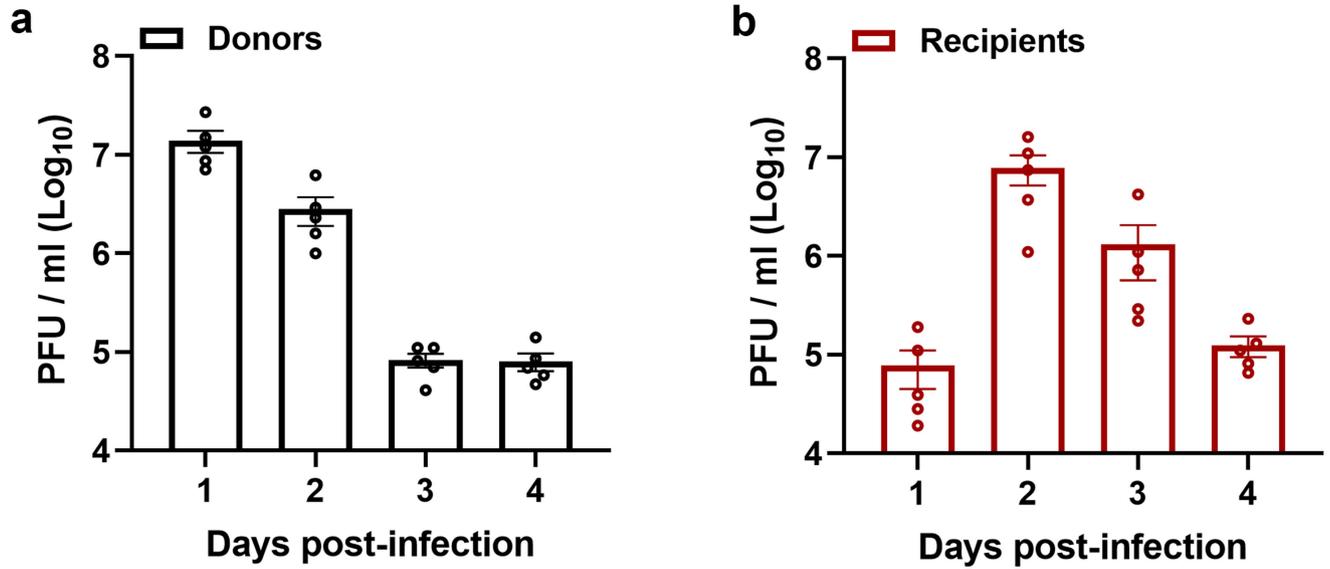
washes of hamsters sampled 1–4 days after infection were measured by Sanger sequencing. Red dots represent individual animals ($n = 5$), the horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a \log_{10} scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. *P* values were calculated for the group (strain) coefficient for each linear regression model.



Extended Data Fig. 4 | Validation of NS01Y competition assay by next generation sequencing, a-e. The same RNA samples from NS01Y competition with wt that were initially assessed using Sanger sequencing (Fig. 2a-f, Fig. 3g; lower panels here) were retested using next generation sequencing (NGS, upper panels). The samples were collected from the nasal washes (a, b) and tissues (c, d) from both donor and recipient hamsters and HAE cells (e). Red dots represent individual animals (n = 5, n = 10 in day 1 nasal wash of NS01Y

group) or biological repeats (n = 6 for HAE cells pooled from two independent experiments), the horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a log₁₀ scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values are calculated for the group (strain) coefficient for each linear regression model.

Nasal wash-growth kinetics-Alpha FL vs WT

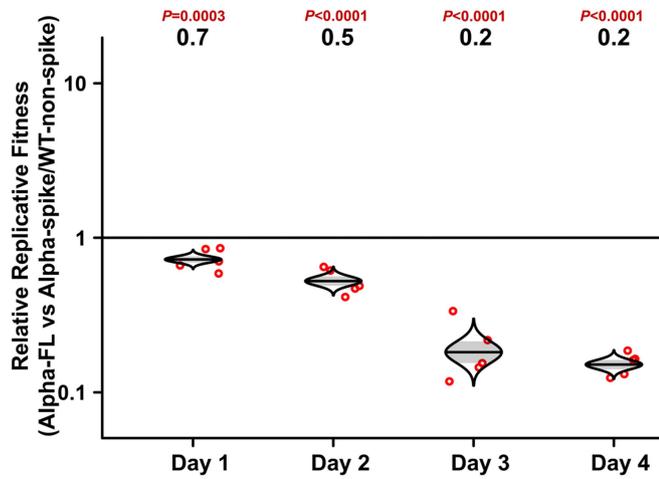


Extended Data Fig. 5 | Growth kinetics of mixed viruses in nasal wash of hamsters. The replication kinetics of the mixed viruses (Alpha-FL with wt) in the nasal washes from both donors (**left panel**) and recipients (**right panel**) hamsters were measured by plaque assay. The nasal wash samples were

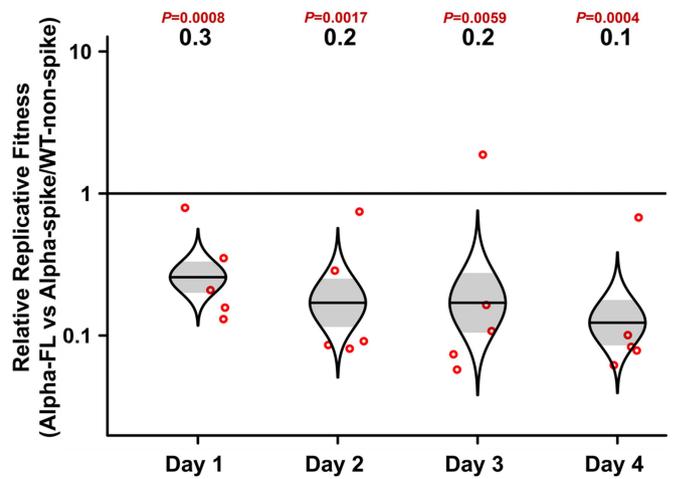
collected from 1–4 days post-inoculation (donors) or post-contact (recipients). Dots represent individual hamsters (n = 5). The values in the graph represent the mean ± standard error of the mean.

Alpha-FL vs Alpha-spike/WT-non-spike

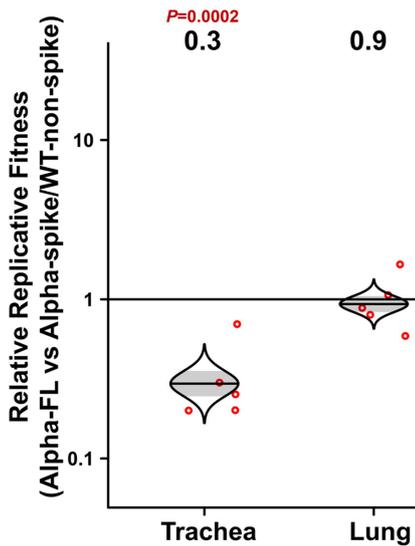
a Nasal wash-Donors



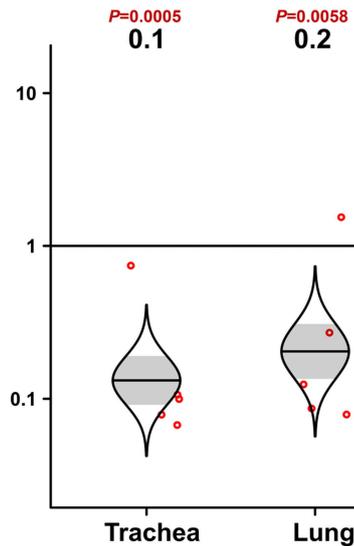
b Nasal wash-Recipients



c Tissues-donors



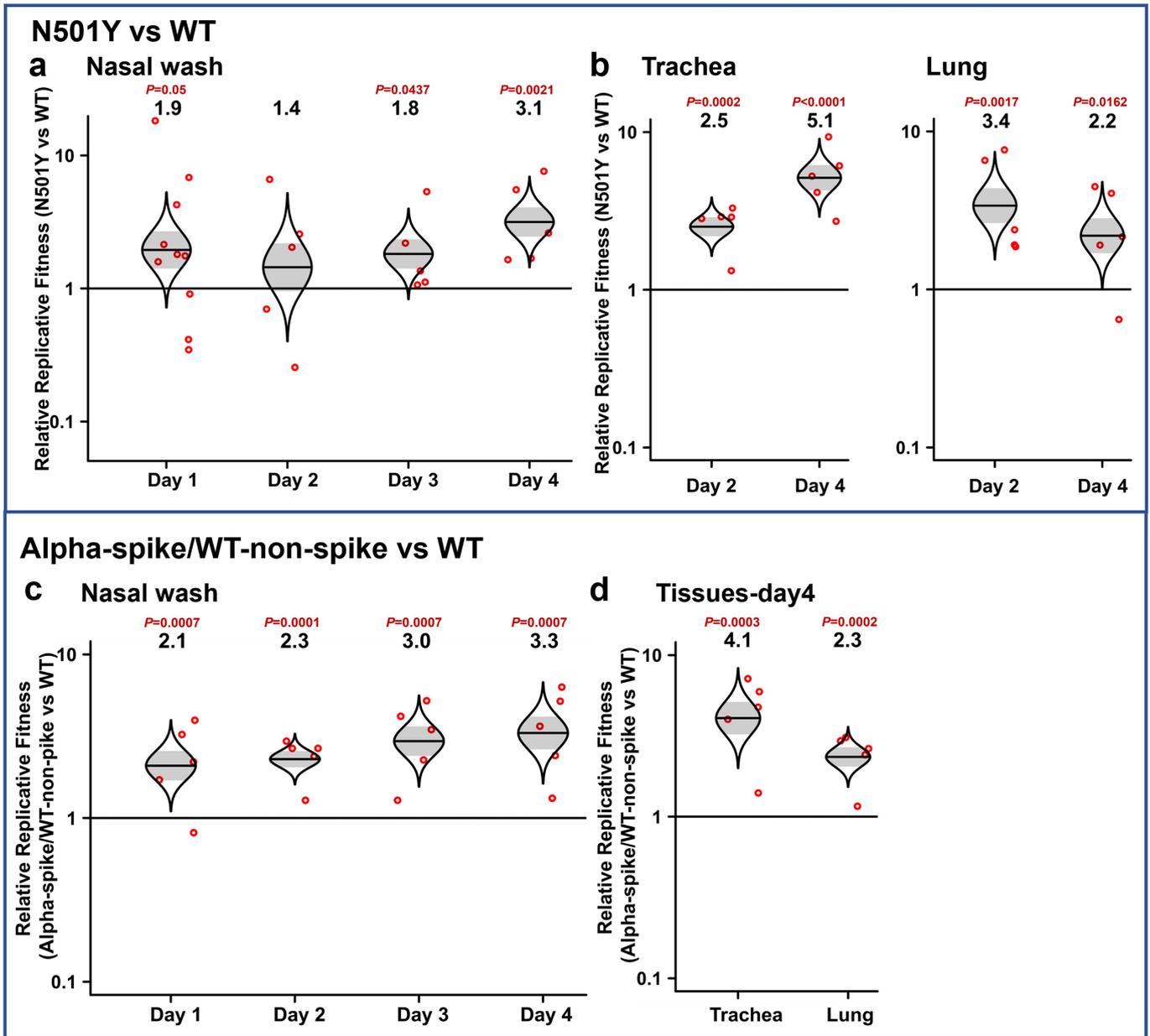
d Tissues-Recipients



Extended Data Fig. 6 | Competitions between Alpha-FL and Alpha-spike/WT-non-spike in hamsters. **a, b,** The Alpha-FL was mixed with Alpha-spike/WT-non-spike virus and inoculated intranasally into hamsters. Results of the competition were assessed by sampling nasal washes of both donor (**a**) and recipient hamsters (**b**) from 1–4 days post-inoculation (donors) or post-contact (recipients). **c, d,** Results of the competition between the Alpha-FL and the Alpha-spike/WT-non-spike in the tracheae and lungs of both donor (**c**) and

recipient hamsters (**d**) at 4 days post-inoculation or post-contact. **a–d,** Red dots represent individual animals ($n = 5$), the horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a \log_{10} scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values are calculated for the group (strain) coefficient for each linear regression model.

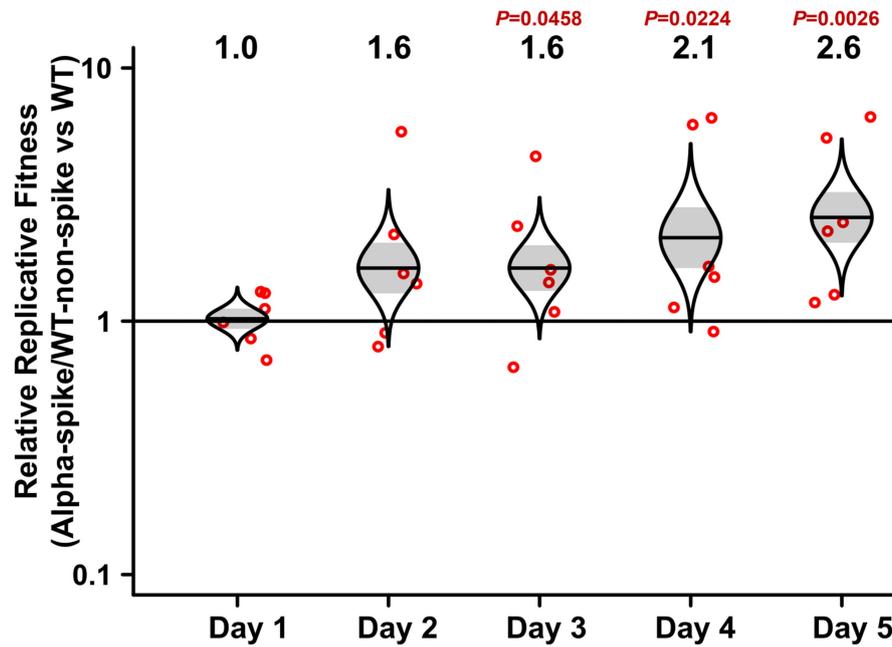
Transmission to recipients



Extended Data Fig. 7 | The advantage of the N501Y substitution and Alpha-spike/WT-non-spike variant during the transmission from donor to recipient hamsters. a-d. The ratios of mixed viruses in the nasal washes (a, c), tracheae and lungs (b, d) of recipient hamsters were compared to the ratios of N501Y:wt or Alpha-spike/WT-non-spike:wt measured on the day 1 nasal wash of donor hamsters to assess fitness for transmission to and early replication in the recipient hamsters, respectively. The total infection titer of the mixed viruses

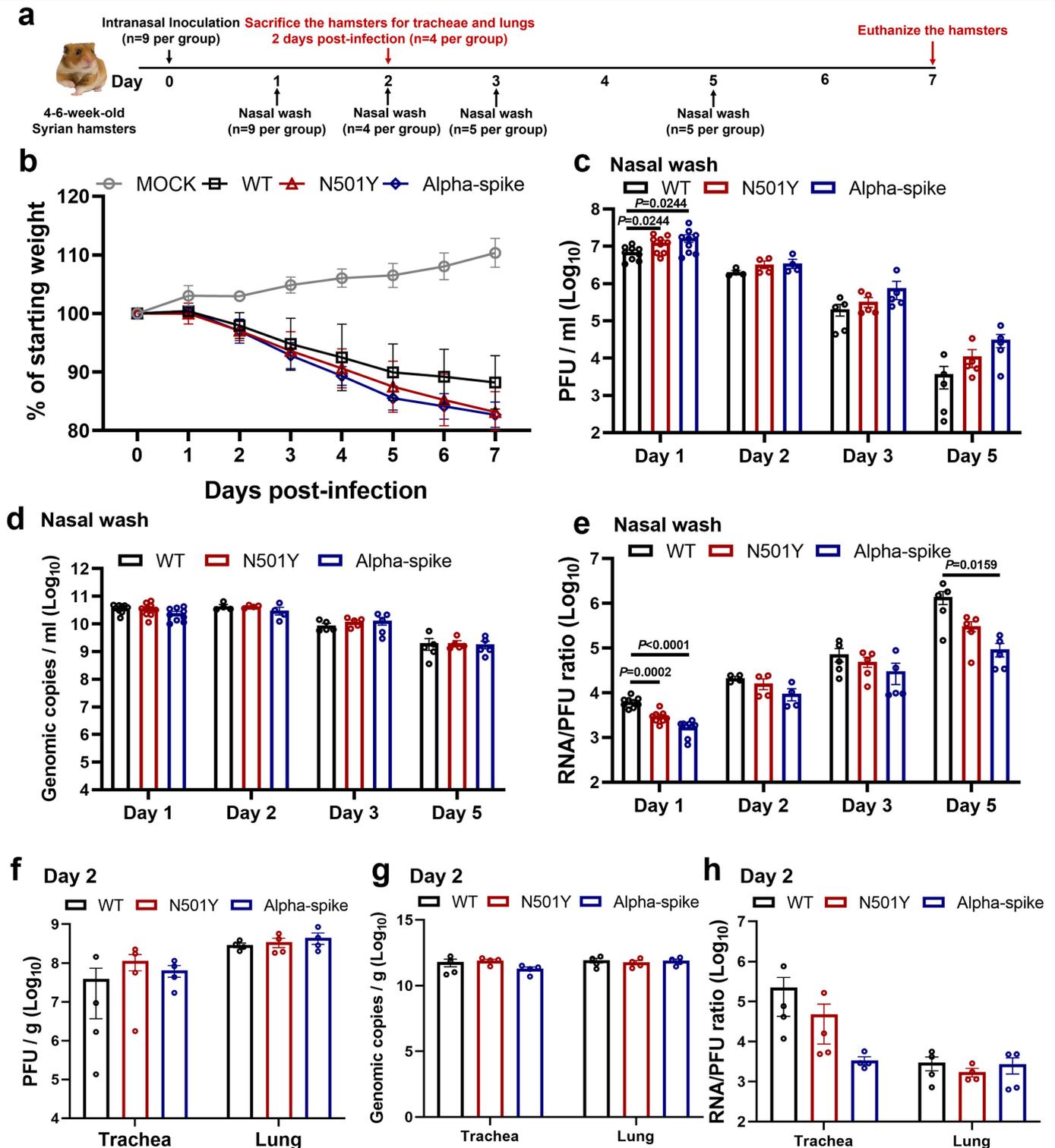
was 10^5 PFU per hamster. Red dots represent individual animals ($n = 5$; $n = 10$ in day 1 nasal wash of N501Y group), the horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a \log_{10} scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values are calculated for the group (strain) coefficient for each linear regression model.

HAE-Alpha-spike/WT-non-spike vs WT



Extended Data Fig. 8 | Competition assay between Alpha-spike/WT-non-spike and WT on primary human airway epithelial cells. The Alpha-spike/WT-non-spike and WT were mixed in 1:1 pfu ratio and inoculated onto human airway epithelial (HAE) cells at a total MOI of 5. The ratios of Alpha-spike/WT-non-spike to the wt virus were measured by Sanger sequencing. Red dots represent individual biological replicates (n = 6), pooled

from 2 independent experiments. The horizontal lines in each catseye represent the mean, shaded regions represent standard error of the mean; y-axes use a log₁₀ scale. Black numbers above each set of values (catseye) indicate the relative fitness estimates. P values are calculated for the group (strain) coefficient for each linear regression model.



Extended Data Fig. 9 | The spike N501Y substitution benefits viral infection of hamster upper airways. **a**, Design of the hamster infection kinetic studies. The wt, N501Y and Alpha-spike viruses were intranasally inoculated into hamsters at a titer of 10^4 PFU per hamster. Nine hamsters were utilized for the initial infection in each group. At 2 days post-infection, 4 hamsters were sacrificed for organ collections. The nasal washes of the hamsters were collected on days 1, 2, 3 and 5 post-infection or before sacrifice. **b**, Weight change in hamsters following infection by the N501Y ($n = 5$) and Alpha-spike ($n = 5$) mutants compared to the wt ($n = 5$). MOCK group ($n = 4$) served as the negative (uninfected) control. The body weights were measured from 1–7 days post-infection. The weight loss data are shown as mean \pm standard deviation and statistically analyzed using two-factor analysis of variance (ANOVA) with

Tukey's post hoc test. No significant differences were seen between the N501Y/Alpha-spike and wt groups. **c–h**, The infection of N501Y and Alpha-spike mutants compared to the wt in the nasal washes (**c–e**) collected 1 ($n = 9$), 2 ($n = 4$), 3 ($n = 5$), or 5 ($n = 5$) days post-infection and in the organs (**f–h**) 2 days ($n = 4$) post-infection. The amounts of infectious virus (**c, f**) and genomic RNA (**d, g**) were quantified by plaque assay and RT-qPCR, respectively. The genomic RNA:PFU ratio (**e, h**) was calculated as an indication of virion infectivity. The values in the graph represent the mean \pm standard error of the mean. A non-parametric two-tailed Mann-Whitney test was used to determine significant differences. *P* values were adjusted using the Bonferroni correction to account for multiple comparisons. Differences were considered significant if $p < 0.025$.

Article

Extended Data Table 1 | The frequencies of the spike protein amino acid substitutions in SARS-CoV-2 strains

Amino acid position	Amino acid change	Presence in SARS-CoV-2 isolates
69	H→del	Denmark B.1.1.298 (86.5%; n=1270)
70	V→del	Denmark B.1.1.298 (86.5%; n=1270)
145	Y→del	N/A
501	N→Y	South Africa B.1.351 (89.7%; n=49), Brazil P.1 (54.5%; n=11)
570	A→D	N/A
614	D→G	South Africa B.1.351 (99.4%; n=49), Denmark B.1.1.298 (99.9%; n=1270), Brazil P.1 (100%; n=11), California B.1.429 (100%; n=100)
681	P→H	N/A
716	T→I	N/A
982	S→A	N/A
1118	D→H	N/A

The amino acid substitutions in the Alpha B.1.1.7 variant refers to the USA_WA1/2020 SARS-CoV-2 sequence (GenBank accession No. MT020880). N/A=Not applicable.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection The peak electropherogram height was collected using the QSVanalyser program(version 20121206).The Octet Data Acquisition 9.0 software was used to collect affinity data.

Data analysis Statistical analyses were performed using R statistical software (version 3.6.1),catseyes (version 0.2.3) and Graphad Prism (version 9.0).T For fitting of KD values, Octet Data Analysis software V11.1 was used to fit the curve by a 1:1 binding model and use of the global fitting method.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Extended Data and source data for generating all figures are available in the online version of the paper. The sequences and epidemics of SARS-CoV-2 variants are from GISAID database and NCBI database, the accession code for USA_WA1/2020 and Alpha-FL are MT020880 and EPI_ISL_999340, respectively.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine the sample size. Sample size was chosen based on previous experience and standards in the field.
Data exclusions	Animals that died before measurement were excluded from analysis. The criteria was pre-established.
Replication	Competition and growth curve experiments in Vero E6, calu-3 and HAE cells were performed with two different experiments each in triplicate; competition and replication assays in hamsters were performed once using 5-10 animals and were confirmed using different methods. All attempts at replication were successful.
Randomization	Hamsters were received from Envigo by dedicated animal research personnel at UTMB, who randomly assigned the animals to cages of 1 or 5 hamsters with no additional knowledge of study design. No further randomization was performed by research personnel.
Blinding	The investigators were not blinded to the allocation during the experiments or to the outcome assessment. Blinding is not necessary because the results are quantitative and did not require subjective judgment or interpretation. Blinding is not typically used in the field.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	The Vero E6 cells and calu-3 cells were purchased from ATCC. The HAE cells were purchased from MatTek Life Science.
Authentication	The Vero and calu-3 cells have been authenticated by ATCC through STR profiling, the HAE cells have been authenticated by MatTek life Science through QC testing.
Mycoplasma contamination	All the cell lines we used were negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell line was used.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Four- to six-week-old male golden Syrian hamsters, strain HsdHan:AURA (Envigo, Indianapolis, IN), were used for animal study
Wild animals	No wild animals were used in this study.

Field-collected samples

No field-collected samples were used in this study.

Ethics oversight

Hamster studies were performed in accordance with the guidance for the Care and Use of Laboratory Animals of the University of Texas Medical Branch (UTMB). The protocol was approved by the Institutional Animal Care and Use Committee (IACUC) at UTMB. All the hamster manipulations were performed under anesthesia by isoflurane to minimize suffering.

Note that full information on the approval of the study protocol must also be provided in the manuscript.