Comparison of SARS-CoV-2 evolution in vitro during passage in Vero and primary human airway cells

Collaboration Between Host and Viral Factors Shape SARS-CoV-2 Evolution

Connor G. G. Bamford[1*], Lindsay Broadbent[1], Elihu Aranday-Cortes[2], Mary McCabe[1], James McKenna[3], David Courtney[1], Olivier Touzelet[1], Ahlam Ali[1,4], Grace Roberts[1], Guillermo Lopez Campos[1], David Simpson[1], Conall McCaughey[3], Derek Fairley[3], Ken Mills[4], Ultan F. Power[1*], and the Breathing Together Investigators[±]

[1] Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, UK.
[2] Medical Research Council-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, UK.
[3] Regional Virus Laboratory, Belfast Health and Social Care Trust, Belfast, UK
[4] Patrick. G. Johnston Centre for Cancer Research, Queen's University Belfast, Belfast, UK.

[*]co-corresponding authors

Connor G. G. Bamford email: c.bamford@qub.ac.uk
Ultan F. Power email: u.power@qub.ac.uk

± = Breathing Together investigators: Professor Andrew Bush, Professor Adnan Custovic, Professor Peter Ghazal, Dr. Mindy Gore, Professor Jonathan Grigg, Professor Clare M. Lloyd, Professor Benjamin Marsland, Professor Ultan Power, Professor Graham Roberts, Professor Sejal Saglani, Professor Jurgen Schwarze, Professor Mike Shields, and Professor Steve Turner.

Comparison of SARS-CoV-2 evolution in vitro during passage in Vero and primary human airway cells

Abstract

SARS-CoV-2 continues to evolve, resulting in several 'variants of concern' with novel properties. The factors driving SARS-CoV-2 fitness and evolution in the human respiratory tract remain poorly defined. Here, we provide evidence that both viral and host factors co-operate to shape SARS-CoV-2 genotypic and phenotypic change. Through viral whole-genome sequencing, we explored the evolution of two clinical isolates of SARS-CoV-2 during passage in unmodified Vero-derived cell lines and in parallel, in well-differentiated primary nasal epithelial cell (WD-PNEC) cultures. We identify a consistent, rich genetic diversity arising in vitro, variants of which could rapidly rise to near-fixation with 2 passages. Within isolates, SARS-CoV-2 evolution was dependent on host cells, with Vero-derived cells facilitating more profound genetic changes. However, most mutations were not shared between strains. Furthermore, comparison of both Vero-grown isolates on WD-PNECs disclosed marked growth attenuation mapping to the loss of the polybasic cleavage site (PBCS) in Spike while the strain with mutations in NSP12 (T293I), Spike (P812R) and a truncation of ORF7a remained viable in WD-PNECs. Our work highlights the significant genetic plasticity of SARS-CoV-2 while uncovering an influential role for collaboration between viral and host cell factors in shaping viral evolution and fitness in human respiratory epithelium.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Family: *Coronaviridae*; Genus: *Betacoronavirus*) has a positive-sense, non-segmented, single-stranded RNA genome of ~30,000 nucleotides in length (**Lu et al., 2020; Wu et al., 2020**). The SARS-CoV-2 genome encodes at least 29 proteins, expressed from translation of a 5' major open reading frame (ORF1ab), including Nsp3 and Nsp12 (viral RNA-dependent RNA polymerase), and a series of nested transcripts at the 3' terminus, including Spike (S; the viral attachment and fusion glycoprotein) and ORF7a. SARS-CoV-2 emerged into the human population in late 2019, causing coronavirus virus disease 2019 (COVID-19) (**Wang et al., 2020**). Reflecting its likely zoonotic origins, SARS-CoV-2-like and other SARS-related viruses have been detected and isolated from horseshoe bats and pangolins from Asia (Boni et al., 2020). SARS-CoV-2 is a highly transmissible virus with an R0 of up to ~5, and has a relatively high case mortality rate (~1%), especially pathogenic in elderly or individuals with co-morbidities (**Cevik et al., 2020**). While safe and effective vaccines were recently developed (**Krammer, 2020**), there is a dearth of highly-effective therapeutic interventions, with notable exceptions such as dexamethasone (**Recovery Group, 2021**).

SARS-CoV-2 productively infects the epithelial cells lining the upper and lower respiratory tract, including those in the nasal cavity and the alveoli of the lung (**Hou et al., 2020**). By virtue of interaction with Spike, SARS-CoV-2 exploits host cell protein angiotensin-converting enzyme 2 (ACE2) as its receptor (**Shang et al., 2020**). Additionally, for entry to occur Spike requires activation by two host proteases, furin and transmembrane protease serine 2 (TMPRSS2)-like proteases, which cleave Spike at the S1/S2 boundary between its two subunits (S1 and S2) and the S2' site in S2 allowing release of the fusion peptide (**Hoffmann et al., 2020**). Following binding to ACE2, a proteolytically-activated Spike

can fuse the viral envelope with the host cell membrane releasing the infectious genome into the cytoplasm.

Since its initial emergence, SARS-CoV-2 has continued to evolve and adapt to the human population with several putatively beneficial mutations arising in Spike, such as D614G and N501Y, that affect Spike stability and binding to ACE2, and antibody-escape mutations in the amino-terminal domain (NTD) (**Harvey et al., 2021**). Additionally, loss of the polybasic cleavage site (PBCS), which is a unique feature of SARS-CoV-2 and facilitates furin cleavage at the S1/S2 boundary, has been demonstrated to reduce transmission and virulence of SARS-CoV-2 in animal models (**Johnson et al., 2021; Peacock et al., 2021**). Together, these mutations of interest are found in constellations in so-called 'variants of concern' (VOC), which are strains of SARS-CoV-2 with evident phenotypic differences, such as enhanced transmissibility, pathogenicity, or reduced sensitivity to antibody-mediated neutralization in humans (**Harvey et al., 2021**).

As SARS-CoV-2 continues to spread, and interventions and vaccines are being rolled out, there remain significant unknowns as to how SARS-CoV-2 may adapt further to humans. Knowledge of the genetic and molecular correlates of this difference in transmissibility is crucial for understanding of coronavirus pandemic preparedness and inform strategies for surveillance and control. In vitro models can help disentangle the factors affecting evolution, identify new ones, and highlight mutational tolerance. Here, we undertook a side-by-side comparison of SARS-CoV-2 evolution by whole genome sequencing of two isolates, grown in parallel in standard Vero-derived cells and human 'well-differentiated primary nasal epithelial cells' (WD-PNECs), which are a useful model for probing virus-host interactions in the respiratory tract (**Guo-Parke et al., 2013; Hou et al., 2020; Villenave et al., 2012**). Our data demonstrate clear roles of both viral and host cell factors in

shaping SARS-CoV-2 genetic and functional changes, identifying genetic features required for efficient infection of primary cells.

Results

Isolation and passage of SARS-CoV-2 in unmodified Vero-derived cells

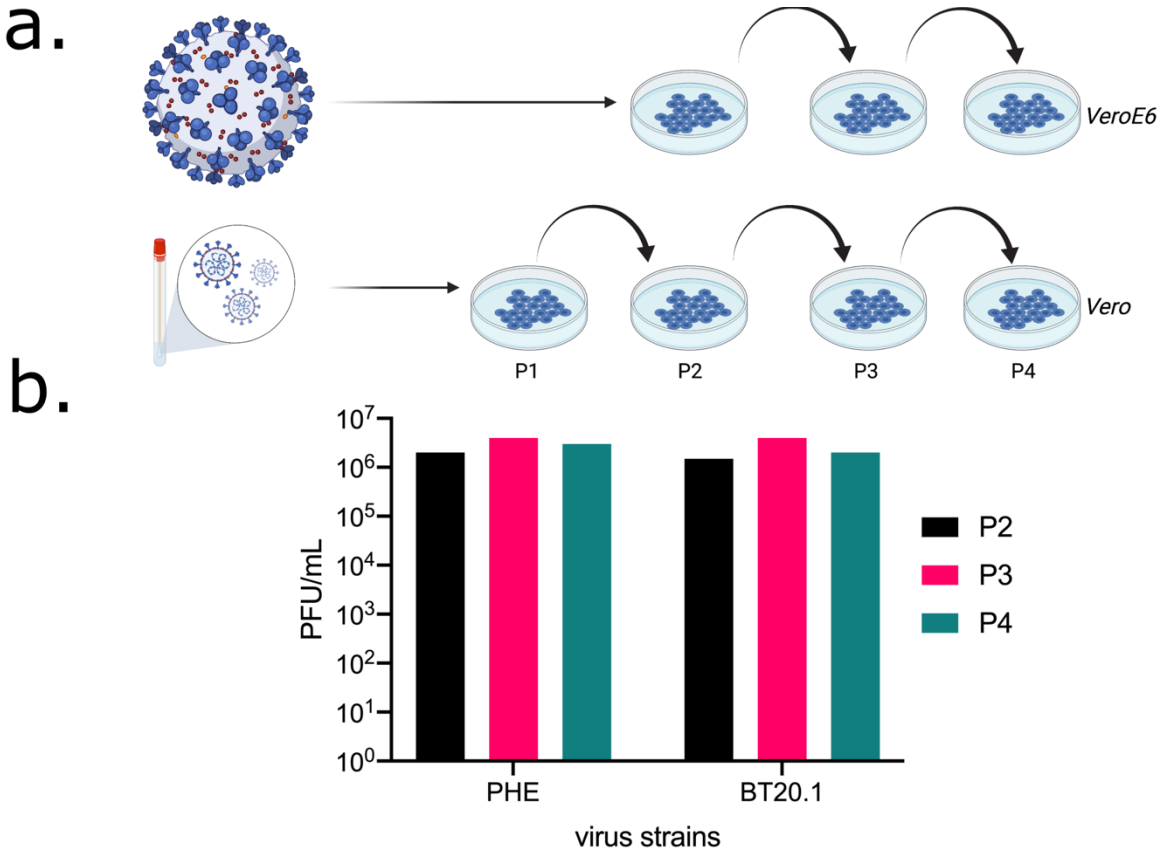
To begin to understand the evolution of SARS-CoV-2 we first needed to generate characterised stocks of virus (**Fig 1A**). In the first instance, a low passage isolate (passage 1, P1) of SARS-CoV-2 (England 02/20) was obtained from Public Health England (PHE) and is referred to as 'PHE'. This stock was from a sample isolated on VeroE6 cells and represents one of the earliest isolates of SARS-CoV-2 in the UK during the pandemic. PHE is from clade A and does not contain the D614G substitution in Spike (**supplementary table 1**)(**Davidson et al., 2020; Holden et al., 2020**). Upon receipt, we carried out a further three passages on VeroE6 cells passaging at an MOI of 0.001, harvesting stocks at 96 hpi when extensive cytopathic effect was observed. The PHE strain grew efficiently, reaching titres of $>10^6$ pfu/mL (**Fig 1B**) and was cytopathic, inducing 'webbing' and cell rounding, consistent with previous reports (data not shown).

As we wanted to understand the viral factors that may drive evolution and results obtained from only one isolate may be non-representative, we next obtained an independent – but comparable - clinical nasal/pharyngeal swab sample containing SARS-CoV-2, which we termed BT20.1 (Belfast 06/20). This strain represents an isolate from the UK's 'first wave' and is a representative of clade B that contains the D614G mutation in Spike (table 1). Unlike PHE, BT20.1 was isolated on standard Vero cells (CCL-81) and passaged to P4 (multiplicity of infection

[MOI]~0.001 and passaged every 3 days). Like PHE, BT20.1 grew efficiently, reaching titres of $>10^6$ pfu/mL (**Fig 1B**).

Both PHE and BT20.1 formed plaques on standard Vero cells in all passages (**Fig 1C and D**). Comparison of plaque sizes between P2 and P4 identified differences in plaque size composition following Vero cell passage. This was most evident for PHE, which became predominantly large plaques (**Fig 1 E and F**). As observed in the plaque edge, BT20.1 induced consistent cell-to-cell fusion, unlike PHE (**sFig 1 A and B**).

Comparison of SARS-CoV-2 evolution in vitro during passage in Vero and primary human airway cells



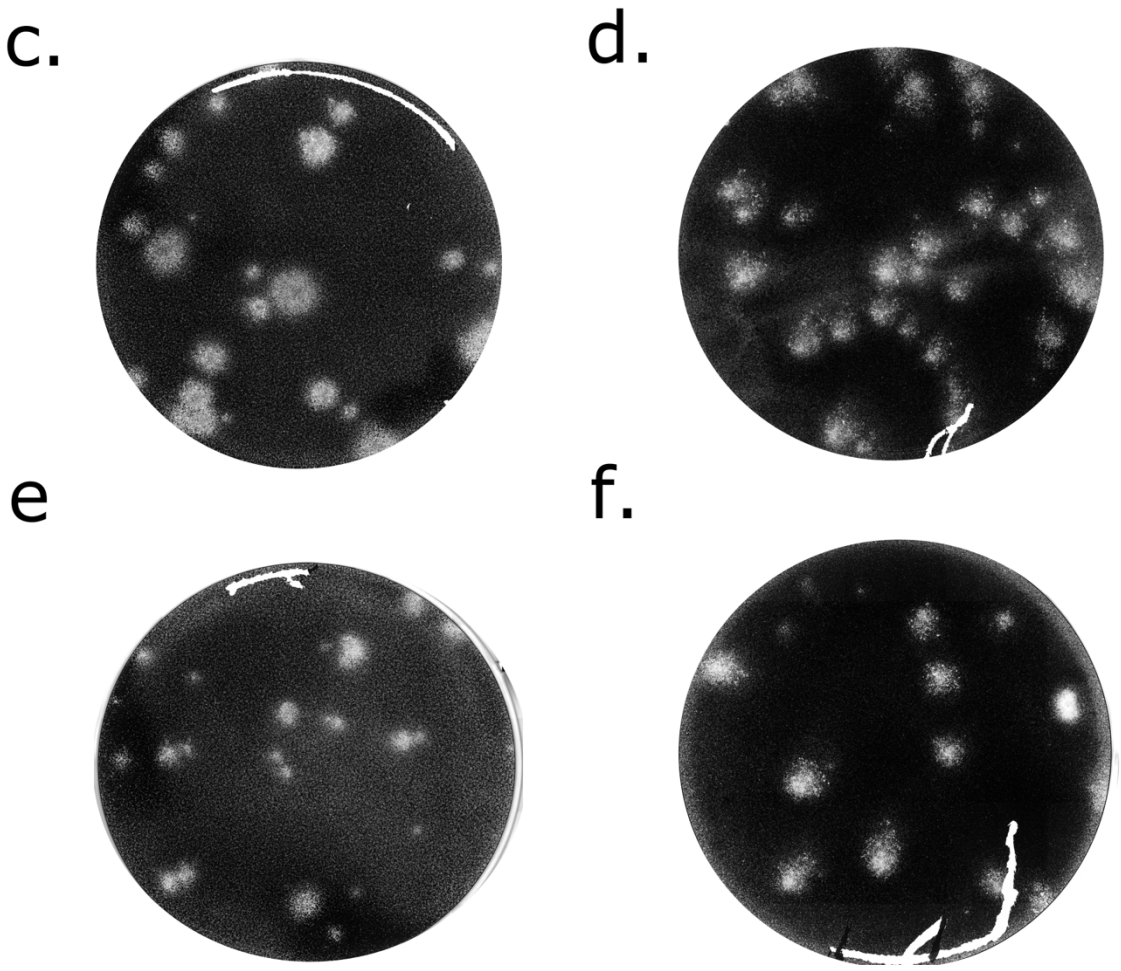


Figure 1. Isolation and passage of PHE and BT20.1 in Vero-derived cells. Schematic of SARS-CoV-2 isolation/serial passage series on VeroE6 or Vero cells for PHE and BT20.1 from isolation to P4 (a). Extracellular infectivity titres for stocks generated from P2-P4 VeroE6/Vero passage for PHE and BT20.1 using plaque assay protocol on Vero cells (b). Plaque visualisation of PHE (c) and BT20.1 (d) P4, and P2 (e and f) on Vero cells. Figures were generated with the aid of Biorender (https://biorender.com/).

Sequencing of SARS-CoV-2 passage series in unmodified Vero-derived cells

As we had successfully generated comparable in vitro passage series for two relevant isolates of SARS-CoV-2, we next determined what genetic changes, if any, occurred during passage in unmodified Vero-derived cells. Whole genome sequences of our SARS-CoV-2 stocks at each passage were generated and minor sequence analysis (>5% minor allele frequency) was carried out, comparing variation arising to the Wuhan-Hu-1 reference (NC_045512.2) genome sequence for SARS-CoV-2 (**supplementary table 1**). Unfortunately, the sequence depth and quality were not sufficient to reconstruct a whole genome sequences for BT20.1 P1 isolate material, likely due to insufficient viral material resulting from the initial isolation. Therefore, we focused our analysis on PHE P1-4 and BT20.1 P2-P4.

Analysing mutations in the PHE passage series we identified 4 changes (C8782T; T18488T; T28144C; A29596G) relative to Wuhan-Hu-1 consistently at ~100% at all passages, likely reflecting fixation in the original virus stock (**Fig 2A**). These changes were considered intrinsic to that particular strain and were not analysed any further herein as we wished to focus on variants arising during passage. Sequencing confirmed the presence of D614 in Spike, consistent with it being an early SARS-CoV-2 isolate.

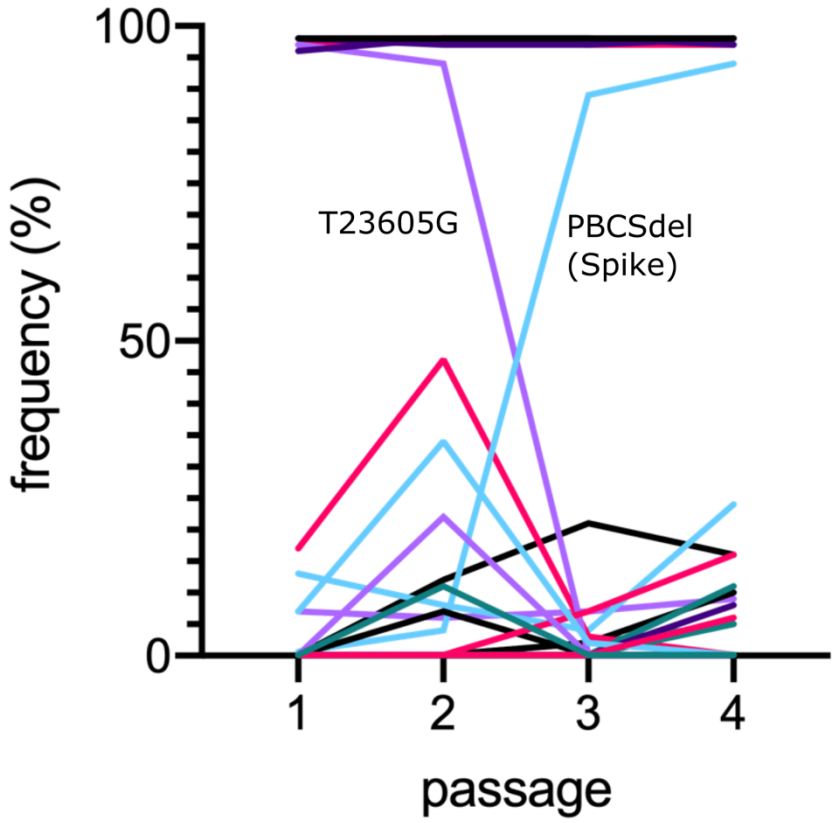
Outwith the core changes described above, two major mutations were observed: a synonymous (T23605G) and non-synonymous out-of-frame deletion (deletion of 24 nucleotides AATTCTCCTCGGCGGGCACGTAGTG 23597A; resulting in the replacement

of of 9 amino acids [679-687; NSPRRARSV] in Spike with an isoleucine [I]) mapping to the polybasic cleavage site (PBCS) (**Fig 2A**). Deletion of the PBCS ablated the T23605G synonymous variant in the process. This occurred at P3, although the deletion was observed in the original P1 material from PHE. Furthermore, we detected 15 minor variants (non-consensus) that had an allele frequency (AF) of >5% in at least one sample of the passage series. These changes mapped to several genes and proteins of SARS-CoV-2, including ORF1AB, Spike, E, N, and ORF10. (**supplementary table 1**). Interestingly, we observed a cluster of three mutations occurring in the amino terminal domain (NTD) of Spike, appearing at P3 and rising in frequency at P4. Two of these Spike NTD mutations were similar to mutations occurring in VOCs: D215G and a deletion of 24 nucleotides (GCTATACATGTCTCTGGGACCAATGGTA21761G) resulting in a loss of 9 amino acids IHVSGTNGT (aa68-77). Additionally, we noticed a convergent mutation of L37 in E, detecting two mutations resulting in L37F and L37R. To determine the reproducibility of passage sequencing, an independent P4 PHE (P4B) was generated from P3 and sequenced, with very high levels of similarity between the two (**supplementary table 1**).

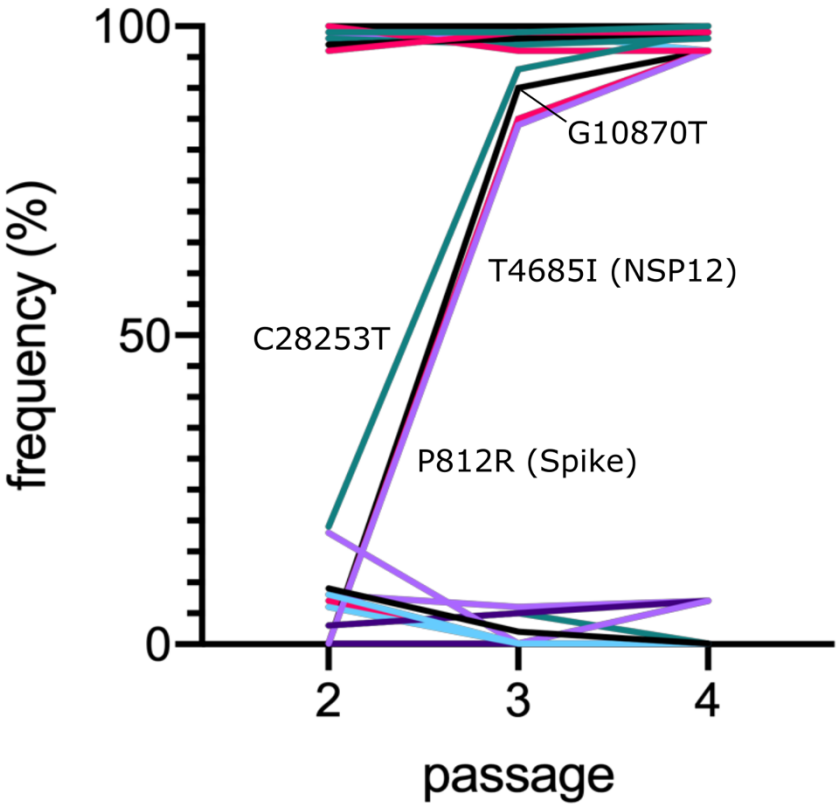
Like PHE, we identified core changes inherent to BT20.1 (**Fig 2B**), which were greater in number than PHE (10 vs 4), consistent with its later isolation (February 2020 versus June 2020) (**supplementary table 1**). These changes included, but were not limited to, D614G in Spike; R203K & G204K in N; and an out-of-frame deletion of 5 nucleotides in ORF7A leading to its premature truncation. Like PHE, we identified mutations arising rapidly upon consecutive passage in Vero cells (i.e., were not detected at P2), including the non-synonymous mutations T293I in NSP12 and P812R in Spike. Both mutations had similar patterns of change in frequency and constituted the majority of sequences by P3. Like PHE, we also detected minor variants (9), including G1251V and S1252C in Spike.

Comparison of SARS-CoV-2 evolution in vitro during passage in Vero and primary human airway cells

a.



b.



223

Figure 2. Analysis of PHE and BT20.1 whole genome sequences during Vero cell passage. Frequency of mutations detected for PHE (a) and BT20.1 (b) passage series on VeroE6 or Vero cells, respectively, relative to the reference sequence (Wuhan-Hu-1). Only sequences from P1-P4 (PHE) and P2-P4 (BT20.1) were analysed to facilitate adequate comparisons. Core changes are found consistently at high frequency and minor variants found at consistently low frequency (e.g. <50%). Only variants that significantly changed in frequency are marked on the graph. Colours do not reflect relationships between variants.

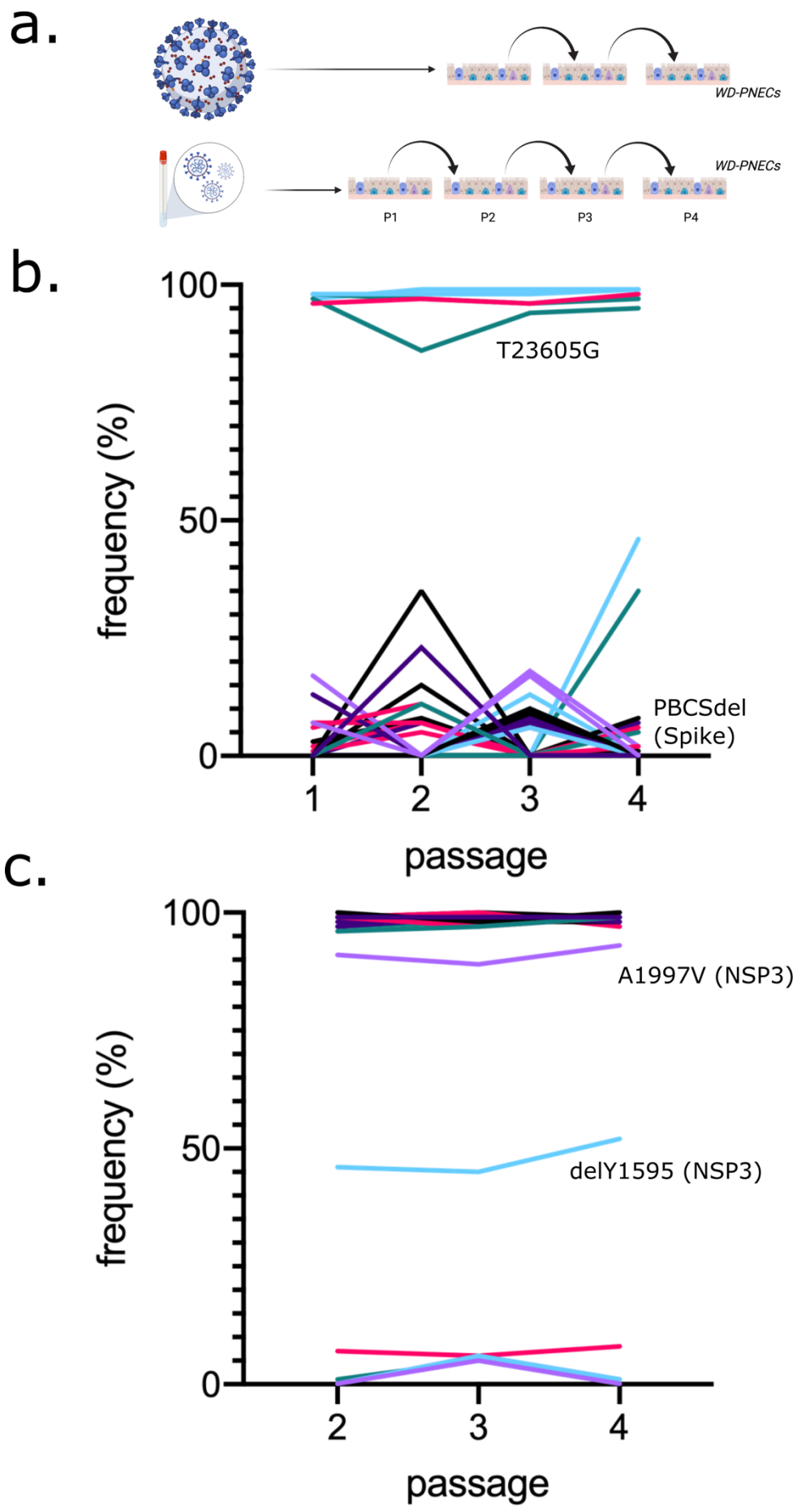
Passage of SARS-CoV-2 on primary human airway cultures

We next sought to investigate the effect of host cell type on subsequent viral evolution, as our previous analysis assessed the contribution of viral background to viral evolution in Vero-derived cells. To this end, in parallel, we passaged SARS-CoV-2 samples on well-differentiated primary human airway epithelial cell cultures until P4, in a similar protocol as was carried out in Vero cells (**Fig 3A**). Primary cultures included WD-PNECs derived from two paediatric donors. For both PHE and BT20.1 robust infection and passage on WD-PNECs was established. For PHE, WD-PNECs were initially infected at MOI of 0.1 and virus harvested at 2-3 dpi, using the original P1 virus material. This was repeated for BT20.1, except unlike PHE, BT20.1 was directly isolated on primary cultures from the obtained clinical material. SARS-CoV-2 grew well in the primary cultures, reaching titres of ~10^6 pfu/mL in 2-3 days in the apical compartment. Samples at each passage were subjected to sequencing as outlined above and analysed in a similar manner to those from the Vero cell passage series. For BT20.1 only P2, P3 and P4 were sequenced to compare with the data available for the equivalent Vero passage series.

In contrast to what was observed in VeroE6 cells, we did not detect any major genetic changes in PHE following passage in WD-PNECs (**Fig 3B**). However, we did identify the PBCS deletion at low levels in minor variant analysis, but never reaching majority. Together with PBCS we found 34 changes as minor variants. From passage to passage, these mutations appeared, and disappeared, stochastically. Similarly, in BT20.1, unlike the Vero cell passage, we did not find corresponding mutations in

Nsp12 or Spike (**Fig 3C**). However, we identified a single amino acid deletion, Y1595, in NSP3. Intriguingly, this variant was maintained throughout the passage series at a moderate frequency of ~45%. At each passage, where possible, SARS-CoV-2 was titrated by plaque assay on Vero cells (**sFig 2**). However, we were unable to obtain titres for PHE and BT20.1 passage P3 and P1, respectively. We noticed slightly reduced titres of BT20.1 in primary cells at P4 compared to earlier passages, which was not observed for PHE (**sFig2**). WD-PNEC-grown viruses had less obvious plaques (**sFig 3 A and B)** and no evidence of cell-to-cell fusion was identified, even for BT20.1 (**sFig 3 C**). Similar to passage in Vero cells we identified two mutations in Spike (G1251V and S1252C), which appeared at low frequencies (<10%) and never increased (**table 1**).

Comparison of SARS-CoV-2 evolution in vitro during passage in Vero and primary human airway cells



273

Figure 3. Analysis of PHE and BT20.1 whole genome sequences during WD-PNECs passage. Schematic of SARS-CoV-2 isolation/passage series on WD-PNECs for PHE and BT20.1 (a). Frequency of mutations detected for PHE (b) and BT20.1 (c) passage series on WD-PNECs, respectively, relative to the reference sequence (Wuhan-Hu-1). Only sequences from P1-P4 (PHE) and P2-P4 (BT20.1) were analysed. PHE P1 is the original stock material obtained and hence is the same sequence as PHE P1 in figure 2. Core changes were found consistently at high frequency and minor variants found at consistently low frequency (e.g. <50%). Only variants that significantly changed in frequency are marked on the graph. Colours do not reflect relationships between variants. Figures were generated with the aid of Biorender (https://biorender.com/).

Phenotypic differences between SARS-CoV-2 'PHE' and 'BT20.1' P4

Our data showing host cell dependency in viral evolution suggested differential fitness for specific viral genotypes (e.g., Vero cell-derived mutations that were not observed in WD-PNECs were less fit in primary cells). To test this hypothesis, we focused subsequent analysis on PHE and BT20.1 Vero P4 stocks with clear genetic differences between them, including the PHE PBCS deletion in Spike, and the P812R (Spike) and NSP12 mutations in BT20.1. To this end, we wished to directly compare the growth and multi-cycle replication kinetics of both strains in cell culture models of infection. To achieve this, we carried out a comparison of growth kinetics in several cell culture models, including Vero cells, VeroE6 cells modified to express human ACE2 and TMPRSS2 (VAT) (Rihn et al., 2021), and WD-PNECs (adult nasal) (**Fig 4A-C**). Of note, unmodified Vero and VeroE6 cells do not express human ACE2 and have very low levels of TMPRSS2 (**Matsuyama et al., 2020**). In Vero cells, SARS-CoV-2 grew to peak extracellular infectivity titres by ~48 hpi with titres of ~10^6 pfu/mL. We noticed a growth attenuation of BT20.1 in Vero cells compared to PHE (**Fig 4A**). Comparing virus growth in VAT cells (**Fig 4B**), both viruses grew better but the relative attenuation of BT20.1 was not observed in VAT cells. In contrast to previous Vero cell experiments, we observed a prominent growth defect of PHE compared to BT20.1 at early time points during infection (24/48 h) in WD-PNECs. However, both viruses reached similar titres by 72 hpi (**Fig 4C**). Together these data

clearly demonstrate phenotypic differences between our Vero cell-passaged viruses, demonstrating a critical role for the PBCS for efficient replication in primary cells.

Comparison of SARS-CoV-2 evolution in vitro during passage in Vero and primary human airway cells

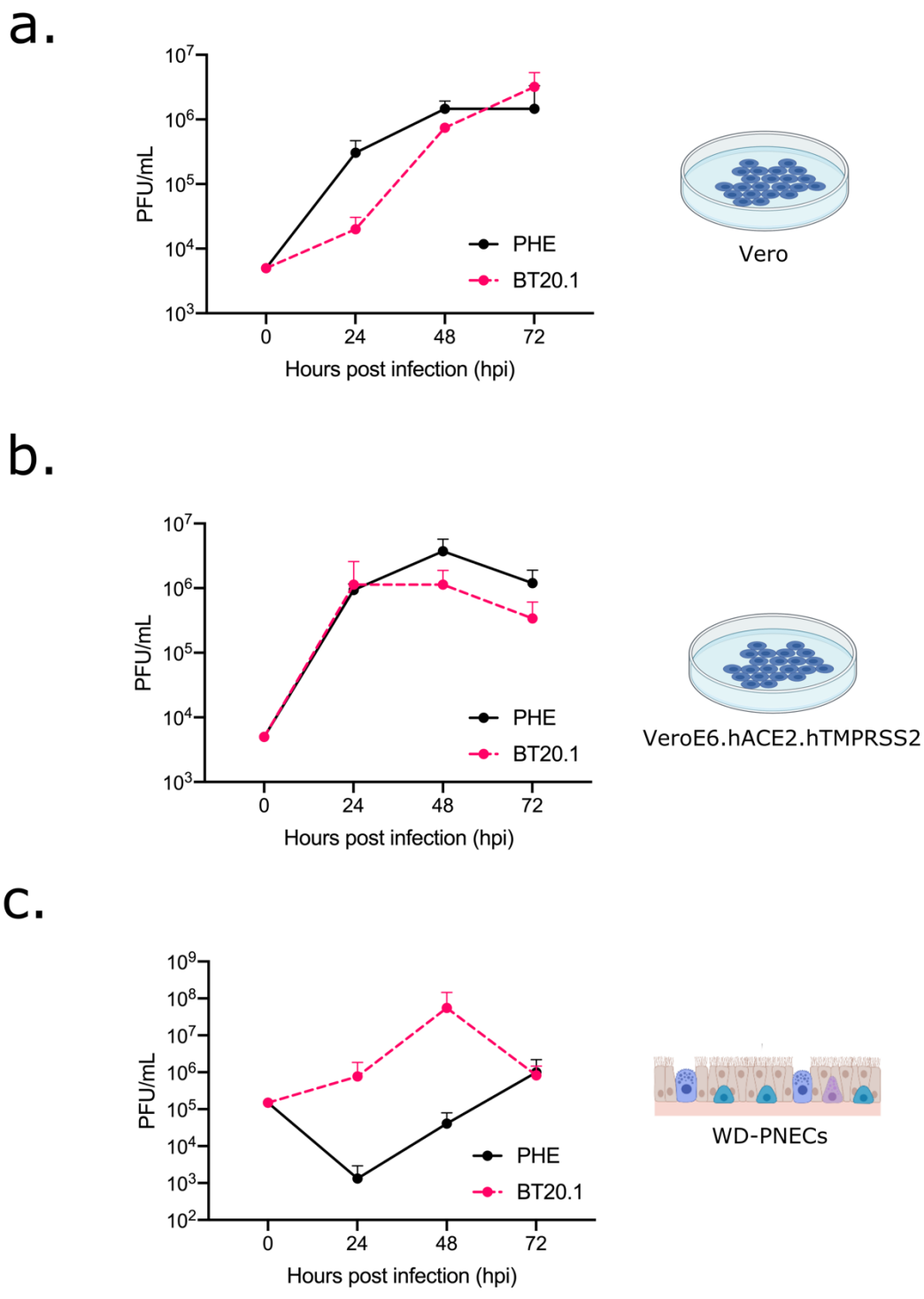


Figure 4. Comparison of PHE P4 (Vero) and BT20.1 P4 (Vero) growth on different cell substrates. Multicycle growth curves (MOI 0.01 for Vero or 0.1 for WD-

315 PNECs) for PHE P4 (VeroE6) and BT20.1 P4 (Vero) on Vero (a), VeroE6 cells expressing
316 human ACE2 and human TMPRSS2 (b), and adult WD-PNECs from 3 donors (c). Titres
317 for Vero-derived cells are shown as means +/- SEM for triplicate wells and are
318 representative of two independent experiments. Titres for WD-PNECs are shown as
319 means +/- SEM for single wells from 3 donors. Data using BT20.1 are presented here as
320 averages from 3 donors but have also been incorporated into a sister paper using
321 separated, individual donor data (Broadbent et al., 2021 in submission). Figures were
322 generated with the aid of Biorender (https://biorender.com/).
323

Discussion

Investigation of the patterns of SARS-CoV-2 genetic diversity worldwide during outbreaks has already facilitated a genetic-based nomenclature of lineages and has also highlighted the emergence of functionally-relevant mutations, such as D614G in Spike, and those contained in extant VOCs (**Harvey et al., 2021**). Complementary to this, in vitro systems are an incredibly useful and tractable means to understand the forces influencing this viral evolution, in particular those that mimic in vivo-relevant conditions, such as WD-PNECs (**Hou et al., 2020**).

Our data and that of others demonstrate significant standing genetic diversity in viral populations in vitro that can be acted upon by rapid evolutionary processes. Although also observed in our work, from early in the study of SARS-CoV-2 evolution in Vero-like cells, it was revealed that the virus readily diversifies during culture with the most evident being mutations mapping to the PBCS of Spike (**Davidson et al., 2020; Klimstra et al., 2020; Lamers et al., 2021; Ogando et al., 2020; Pohl et al., 2021**). Consistent with our work, other studies have identified enhanced genetic stability, in particular of the PBCS only, during passage of one strain in Calu3 or primary airway organoids (**Lamers et al., 2021**). One striking finding of our work, which builds on previous studies, is that on several occasions for both isolates we observed a rapid increase in frequency of specific mutations in the PBCS and independently of it (PBCS deletion & P812R in Spike, and T293I in NSP12) to near fixation over the course of a couple of passages in Vero cells. These patterns suggest a selective phenotypic advantage in that particular cell culture system. Similar changes (including P812R and the NTD deletion) were identified in other studies (**Dieterle et al., 2020; Ramirez et al., 2021**). The fact that identical mutations arise independently (e.g., loss of the PBCS and P812R) is highly suggestive of convergent evolution, perhaps toward a similar phenotype. Our work on the loss of the PBCS in

Vero cells and its association with attenuation in WD-PNECs is consistent with previous reports. However, it is noteworthy that we did not observe deletion or mutations in and around the PBCS during passage of BT20.1 in Vero cells.

In addition to the loss of the PBCS, we observed P812R in Spike and T293I in NSP12, although we were not able to associate them with changes in virus growth in WD-PNECs due to a lack of an additional comparable 'wild-type' isolates. However, the fact that parallel passage in WD-PNECs did not result in their increased frequency suggests that they confer a hitherto unrecognised disadvantage in the primary epithelial cell system. P812R is a non-conservative change and rapidly rose to near-fixation alongside NSP12 in BT20.1 in Vero cells. P812 sits near the S2' cleavage site and is a highly conserved position in SARS-CoV-2. However, non-P residues (e.g. serine) were occasionally found in nature but are rare (https://nextstrain.org/), which suggests a functional defect in vivo. Interestingly, P812R was observed before, in at least two other studies, associated with a change in Spike activity using full-length SARS-CoV-2 and one using chimeric vesicular stomatitis virus encoding SARS-CoV-2 Spike (**Dieterle et al., 2020; Ramirez et al., 2021**). Like previous work, we noted an association of P812R with enhanced cell-to-cell fusion when BT20.1 grown on Vero cells is compared to that grown in WD-PNECs (**sFig 3**). It was suggested that P812R generated a novel PBCS at the S2' site (**Ramirez et al., 2021**). Cleavage by furin-like proteases could thus compensate for lack of TMPRSS2-mediated proteolysis and activation in Vero cells. Although it is possible that P812R confers a similar phenotypic change as the PBCS deletion, it is not likely to be identical, given the clear differences in growth between PHE and BT20.1 in Vero cells and WD-PNECs. Along with P812R in S, BT20.1 carried a mutation in NSP12 (T4685I/T293I), which is the viral RNA-dependent RNA polymerase. The mutation sits on the surface in close proximity to a zinc-binding site of the interface domain that mediates intra-NSP12 interactions and interactions

between NSP12 and other polymerase co-factors, such as NSP8 (**Hillen et al., 2020**). Given the linkage between P812R and T4685I, further molecular virological work using isogenic viruses generated through reverse genetics is required to ascertain the impact of this mutation in relevant cell models. The mutations T4685I arose with P812R possibly suggesting genetic linkage, although this remains to be determined.

It is of interest that BT20.1 carries a deletion in ORF7A that results in a frameshift and C-terminal truncation of the protein, likely ablating the transmembrane domain and tail. ORF7A is a type 1 transmembrane protein and it has numerous putative functions involved in host-pathogen interactions and immune evasion (**Nemudryi et al., 2021**). ORF7A truncations in SARS-CoV-2 isolates have been discovered before, possibly associated with reduced capacity to subvert the innate immune response (**Nemudryi et al., 2021**). However, the previous work was carried out using non-clinically relevant cell models, such as Vero or HEK-derived lines. Our work suggests that full-length ORF7A is not required for replication in Vero or WD-PNECs and likely serves an accessory function that may affect replication and/or transmission in particular circumstances.

Not only did we observe changes reaching near-fixation in our dataset we also identified several lower frequency mutations in our viral populations. Consistent with this variation within a stock, we also noticed plaque size variation in passage stocks suggestive of functional differences between viral sub-clones (**Fig 1 E and F**). We detected an in-frame deletion of a single codon in the C-terminus of NSP3, located in the Y1 domain, which is located on the cytoplasmic face of the virus-remodelled ER membrane, where it may regulate replication complex stability by interacting with NSP4 (**Lei et al., 2018**). NSP3 itself is a multifunctional protein involved in numerous viral processes. The fact that the deletion did not rise to fixation suggests that it is at a competitive disadvantage compared to wild-type. The mutation in NSP3 is also

interesting because it is maintained at a moderate frequency. Of considerable interest is the overlap between variations observed in Vero cells and that of VOCs, especially in the NTD of Spike. We observed three mutations in the NTD in PHE P3 and P4: E180K, D215G, and a deletion resulting in the loss of 9 amino acids. Variants identified in this study mapping to the ectodomain of Spike are marked on a structural model (**sFig 4**). For D215G and the deletion, these mutations are similar to those in VOCs, such as alpha and beta variants. Regarding mutations in NTD loops, several VOCs have convergently modified the amino acid identity of the loop. While in vivo this may be the result of antibody selection, in our system there are no antibodies, which suggests a role for NTD mutations independent of antibody selection. The rise in frequency is suggestive of a fitness advantage of these mutations. Further work is required to determine the function of the NTD of Spike and the impact of these mutations on the virus life cycle.

While general trends were similar between our two isolates in Vero cells (i.e., mutations rising to high frequency), specific mutations observed were not. It is likely that evolution of key mutations reflect inherent biological differences in viruses and not subtle changes in passaging conditions. In PHE, loss of the PBCS occurred, which was not observed in BT20.1, and vice versa regarding P812R and NSP12. This is consistent with an effect dependent on viral input or strain or genetic background through epistatic interactions between mutations, such as D614G in Spike. However, in numerous reports, isolation and passage of SARS-CoV-2 on Vero cells selected for a loss of the PBCS, which was not observed in our BT20.1 passage series. Alternatively, the mutation P812R could functionally achieve the same phenotype as the deletion of the PBCS, although our primary cell infection model where PHE was attenuated compared to BT20.1, would suggest that this is not the case.

By comparing evolution of the same isolates in two distinct cell culture systems, we observed a dependence on host cell substrate on

downstream virus evolution. Namely, passage in WD-PNECs resulted in enhanced stability of SARS-CoV-2 genetic diversity at the consensus level. While the PBCS, P812R and NSP12 changes were identified in PHE and BT20.1 when grown in Vero cells, these changes did not rise to high frequencies in WD-PNECs. Differential accumulation of mutations may reflect distinct host cellular environments encountered upon passage in Vero or WD-PNECs. This reflects major differences in these cell substrates, including i) species and tissue differences; ii) reduced levels of TMPRSS2 in Vero cells; iii) reduced innate immune response in Vero cells as they are deficient in type 1 interferon production (**Emeny & Morgan, 1979**). However, what affects the rise in P812R/NSP12 mutation remains unknown. Future work will assess the effect of these changes in BT20.1 upon replication in WD-PNECS. Additionally, during passage of BT20.1 in WD-PNECs we identified a deletion in Nsp3, although the relevance and mechanism of this change is unknown.

In conclusion, by studying the evolution of SARS-CoV-2 during passage in distinct cellular substrates we shed light on the forces that shape viral fitness, unveiling a collaboration between both viral and host factors in driving SARS-CoV-2 genetic diversity, which helps define the molecular correlates of fitness in the natural target cells. Finally, on a practical note, our results support close characterisation of virus stocks for experimentation in vitro and in vivo and suggest ways to mitigate unwanted cell culture artefacts, critical for understanding host-pathogen interactions and identification of antiviral interventions.

475

476 # Figure legends

477

478 **Figure 1. Isolation and passage of PHE and BT20.1 in Vero-derived**
479 **cells.** Schematic of SARS-CoV-2 isolation/serial passage series on VeroE6
480 or Vero cells for PHE and BT20.1 from isolation to P4 (a). Extracellular
481 infectivity titres for stocks generated from P2-P4 VeroE6/Vero passage for
482 PHE and BT20.1 using plaque assay protocol on Vero cells (b). Plaque
483 visualisation of PHE (c) and BT20.1 (d) P4, and P2 (e and f) on Vero cells.
484 Figures were generated with the aid of Biorender
485 (https://biorender.com/).

486

487 **Figure 2. Analysis of PHE and BT20.1 whole genome sequences**
488 **during Vero cell passage.** Frequency of mutations detected for PHE (a)
489 and BT20.1 (b) passage series on VeroE6 or Vero cells, respectively,
490 relative to the reference sequence (Wuhan-Hu-1). Only sequences from
491 P1-P4 (PHE) and P2-P4 (BT20.1) were analysed to facilitate adequate
492 comparisons. Core changes are found consistently at high frequency and
493 minor variants found at consistently low frequency (e.g. <50%). Only
494 variants that significantly changed in frequency are marked on the graph.
495 Colours do not reflect relationships between variants.

496

497 **Figure 3. Analysis of PHE and BT20.1 whole genome sequences**
498 **during WD-PNECs passage.** Schematic of SARS-CoV-2
499 isolation/passage series on WD-PNECs for PHE and BT20.1 (a). Frequency
500 of mutations detected for PHE (b) and BT20.1 (c) passage series on WD-
501 PNECs, respectively, relative to the reference sequence (Wuhan-Hu-1).
502 Only sequences from P1-P4 (PHE) and P2-P4 (BT20.1) were analysed.
503 PHE P1 is the original stock material obtained and is hence the same
504 sequence as PHE P1 in figure 2. Core changes are found consistently at

high frequency and minor variants found at consistently low frequency (e.g., <50%). Only variants that significantly changed in frequency are marked on the graph. Colours do not reflect relationships between variants. Figures were generated with the aid of Biorender (https://biorender.com/).

Figure 4. Comparison of PHE P4 (Vero) and BT20.1 P4 (Vero) growth on different cell substrates. Multicycle growth curves (MOI 0.01 for Vero or 0.1 for WD-PNECs) for PHE P4 (VeroE6) and BT20.1 P4 (Vero) on Vero (a), VeroE6 cells expressing human ACE2 and human TMPRSS2 (b), and adult WD-PNECs from 3 donors (c). Titres for Vero-derived cells are shown as means +/- SEM for triplicate wells and are representative of two independent experiments. Titres for WD-PNECs are shown as means +/- SEM for single wells from 3 donors. Data using BT20.1 have been reproduced here as averages from 3 donors but have also been incorporated into a sister paper using separated, individual donor data (Broadbent et al., 2021 in submission). Figures were generated with the aid of Biorender (https://biorender.com/).

sTable 1. Frequency of variants in reference to Wuhan-Hu-1 identified in this study. Variants only shown where there was at least one instance of frequency >5%. Where undetectable an arbitrary value of 0.1 was assigned. Frequency data is highlighted by colour (green for higher, yellow for lower). Mutations have been assigned status of core, variant or minor. Additionally, for each variant, data for nucleotide location, reference and variant nucleotides, gene & protein location, and consequence (e.g., synonymous [S] or non-synonymous [NS]) are shown.

sFig 1. Fusogenicity of PHE and BT20.1 P4 on Vero cells. Higher magnification images of plaque visualisation of PHE (a) and BT20.1 (b) P4 on Vero cells from the same images shown in Figure 1.

sFig 2. Growth kinetics of PHE and BT20.1 during passage in WD-PNECs. Infectivity titres for material generated from isolation/passage of PHE and BT20.1 on WD-PNECs.

sFig 3. Plaque morphology of SARS-CoV-2 grown in WD-PNECs. Plaque visualisation of PHE (a) and BT20.1 (b) P4 on Vero cells. Higher magnification images of plaque visualisation of BT20.1 (c) P4 on Vero cells from the same images shown (b).

sFig 4. Location of Spike mutant variants observed in this study on model structure of a single Spike monomer in the pre-fusion state (PDB 7C2L

from (Chi et al., 2020)). Variants identified in the Spike cytoplasmic tail (G1251V and S1252C) are not shown.

555

556 # Materials and methods

557

558 ## Continuous cell line culture

559 In this study, 3 continuous cell lines were used: Vero wildtype (number),
560 Vero E6, and Vero E6 expressing human ACE2 and TMPRSS2 (VAT) (Rihn
561 et al., 2021). All cells were grown in DMEM (5% FCS v/v) with antibiotics.
562 VAT cells were maintained in the presence of additional antibiotics to
563 select of cells carrying transgenes. Cell lines were routinely tested for
564 mycoplasma contamination and no evidence of contamination was
565 detected.

566

567 ## WD-PNECs

568 Nasal epithelial cells from preschool age children with recurrent wheeze
569 (for initial passaging) and from healthy adults (for final comparison of PHE
570 and BT20.1 P4 viruses) were obtained by brushing of the nasal turbinates
571 with an interdental brush (DentoCare). Cells were cultured in monolayer
572 until passage 3 then seeded onto collagen-coated Transwells (6 mm, 0.4
573 µm pore size; Corning). Once confluent the apical medium was removed
574 to create an air-liquid interface which, together with specialised media
575 (Pneumacult ALI, Stemcell Technologies), triggered differentiation
576 (Broadbent et al. 2016. Broadbent et al. 2020). Complete differentiation
577 (after a minimum of 21 d) was confirmed by an intact culture, extensive
578 cilia coverage and mucus production.

579

580 ## Viruses

581 Two SARS-CoV-2 isolates were used throughout this study, including
582 'PHE' and 'BT20.1'. PHE was provided as an early passage isolate on
583 VeroE6 cells while BT20.1 was provided directly as a nasopharyngeal
584 swab in virus transport media clinical material from a positive case from

Belfast in June 2020. Stocks were prepared in Vero or VeroE6 cells in DMEM containing 2.5% FCS (v/v) infected at a low MOI (~0.001). Infections were harvested when maximal cytopathic effect was noted, usually between 3-4 days post infection. Infected culture supernatant was harvested, clarified by centrifugation and stored at -80°C. WD-PNECs were apically infected with SARS-CoV-2 for 1 h, after which the inoculum was removed and the apical surface gently rinsed with DMEM. Virus was harvested from WD-PNECs by incubation of the apical surface with DMEM for 5 min at room temperature in the absence of serum. Harvested virus was immediately stored at -80. All SARS-CoV-2 work was carried out under BSL3 conditions in a dedicated facility in QUB.

Plaque assays

Our plaque assay protocol is based on the methodology available here: https://www.protocols.io/view/viral-titration-of-sars-cov-2-by-plaque-assay-semi-be4zjgx6. Near confluent monolayers of Vero cells in 24 or 6 well plates were infected. On the day of titration growth media was replaced with DMEM (0% FCS) (250 µL). Virus dilutions were prepared in plate and incubated for 30 min after which the 2x overlay medium (containing 2% agarose) was added. Plates were incubated for 3 days at 37°C. At 3 dpi PFA (8%) was added to the cultures and cells fixed/inactivated for at least 20 min. Following fixation, the PFA was removed and monolayers stained for 10 min using crystal violet (1% w/v in ethanol 20%). Following staining, residual crystal violet solution was removed, plates were rinsed in water and submerged in Chemgene prior to drying and removing from the hood for visualisation and quantification. To calculate PFU/mL, plaques at a dilution were quantified, the precipice of this number used, and multiplied by the dilution factor (4). For visualisation of plaque assays, whole plates were scanned using a Celigo imaging cytometer (Nexcelom Bioscience).

Virus whole genome sequencing

Virus whole genome sequencing used methods developed by the ARTIC network (https://artic.network;(**Tyson et al., 2020**)) and the COG-UK Consortium. Culture supernatants were inactivated by addition of Triton X-100 to 1.5% (v/v). Viral RNA (total nucleic acid) was extracted from inactivated samples (200 µL) using the MagNA Pure Compact instrument and MagNA Pure Compact Nucleic Acid Isolation Kit I (Roche Molecular Systems Inc, Burgess Hill, UK). Purified nucleic acid was eluted into 100 µL and used immediately or stored at -80°C. For first-strand cDNA synthesis, nucleic acid (5 µL) was used as template for reverse transcription using LunaScript® RT SuperMix Kit (New England Biolabs, Hitchin, UK) in 20 µL reaction volume. Primers were annealed (65°C, 5min, snap-cool on ice) prior to addition of reverse transcriptase. Reactions were incubated at 42°C (50 min) then stopped at 70°C (10 min). The resulting cDNA was used immediately for PCR or stored at -20°C. In brief, these were run as two separate multiplex PCR "pools" (A & B) using the ARTIC version 3 primer set (ARTIC nCoV-2019 V3 Panel, IDT DNA Inc, Leuven, Belgium; https://github.com/artic-network/primer-schemes/tree/master/nCoV-2019) and Q5 DNA polymerase mastermix (New England Biolabs). Following PCR, the amplicons from pools A & B were combined, and the resulting pooled amplicons (98 x 450 bp overlapping tiled amplicons, spanning the SARS-CoV-2 genome) were purified using Kapa HyperPure beads (Roche Molecular Systems Inc) and quantified using a Qubit fluorometer and dsDNA HS Assay Kit (Thermo Fisher Inc, Manchester, UK). Amplicon sequencing libraries were prepared using the Nextera DNA Flex library kit according to the manufacturer's instructions (Illumina Ltd., Cambridge, UK). Libraries were sequenced on a MiSeq (Illumina) using a MiSeq Reagent Kit v2 and 2 x 151 bp paired-end sequencing protocol (Illumina).

Sequence analysis

The FASTQ files were uploaded into the Galaxy web platform, and we used the public server at usegalaxy.eu to analyse the data (**Afgan et al., 2018).** The workflow used was specially optimized for Illumina-sequenced based ARTIC pair end data with the intention to detect allelic variants (AVs) in SARS-CoV-2 genomes (**Maier et al., 2021**). This analysis converted FASTQ data to annotated AVs through a series of steps that include QC, trimming ARTIC primer sequences off reads with the iVar package, mapping using bwa-mem, deduplication, AV calling using lofreq, and filtering AVs that both occurred at an allele frequency (AF) ≥5%, and were supported by ≥10 reads. As we could not determine the background frequency of mutations we focused on those variants with a minor allele frequency ≥5%, and were supported by ≥10 reads in at least one passage of the series. Furthermore, we focused our greater analysis on those found in more than one passage and those that substantially rise in frequency. Raw data and consensus sequences will be uploaded during review and before publishing.

Comparison of SARS-CoV-2 evolution in vitro during passage in Vero and primary human airway cells

677

PHE VEROE6												
POS	REFERENCE	VARIANT	P1	P2	P3	P4A	P4B	CATEGORY	GENE	S/NS?	CONSEQUENCE	PROTEIN
7749	C	T	0.1	12	21	16	19	MINOR	ORF1AB	NS	T2495I	NSP3
8782	C	T	97	98	97	97	98	CORE	ORF1AB	S		
9534	C	T	0.1	0.1	2	5	5	MINOR	ORF1AB	NS	T3090I	NSP4
18488	T	C	98	97	97	98	96	CORE	ORF1AB	NS	I6075T	NSP14
19983	C	CT	7	6	7	9	8	MINOR	ORF1AB	NS	FRAMESHIFT. D6576*	NSP15
21697	C	T	13	8	4	24	15	MINOR	ORF1AB	S		
21761	GCTATACATGTCTCTGGGACCAATGGTA	G	0.1	0.1	2	10	6	MINOR	SPIKE	NS	IHVSGTNGT68-77del	SPIKE
21846	C	T	17	47	3	0.1	0.1	MINOR	SPIKE	NS	T95I	SPIKE
22100	G	A	0.1	0.1	0.1	11	15	MINOR	SPIKE	NS	E180K	SPIKE
22206	A	G	0.1	0.1	0.1	8	11	MINOR	SPIKE	NS	D215G	SPIKE
23605	**T**	**G**	**97**	**94**	0.1	0.1	0.1	**VARIANT**	**SPIKE**	**S**		
23597	**AATTCTCCTCGGCGGGGCACGTAGTG**	**A**	**0.6**	**4**	**89**	**94**	**94**	**VARIANT**	**SPIKE**	**NS**	**NSPRRARSV679I**	**SPIKE**
25339	C	T	0.1	7	0.1	0.1	0.1	MINOR	SPIKE	S		SPIKE
26353	C	T	0.1	0.1	7	16	19	MINOR	E	NS	L37F	E
26354	T	G	0.1	0.1	0.1	5	5	MINOR	E	NS	L37R	E
28144	T	C	96	98	98	97	97	CORE	ORF8	NS	L84S	ORF8
28833	C	T	0.1	22	0.1	0.1	0.1	MINOR	N	NS	S187L	N
29366	C	T	7	34	2	0.1	0.1	MINOR	N	NS	P365S	N
29596	A	G	98	98	98	98	98	CORE	ORF10	NS	I13M	ORF10
29637	T	C	0.1	0.1	0.1	6	5	MINOR	ORF10	NS	I27T	ORF10
29844	AT	A	0.1	11	0.1	0.1	0.1	MINOR	3' UTR	S		NC

PHE WD-PNECS										
POS	REFERENCE	VARIANT	P1	P2	P3	P4	CATEGORY	GENE	S/NS?	CONSEQUENCE	PROTEIN
241	C	T	0.1	0.1	6	0.1	MINOR	5'UTR	S		
514	TGTTATG	T	2	7	0.1	0.1	MINOR	ORF1AB	NS	MV85DEL	NSP1
1593	C	T	0.1	0.1	0.1	35	MINOR	ORF1AB	NS	S443F	NSP2
2994	A	C	0.1	0.1	0.1	7	MINOR	ORF1AB	NS	E910A	NSP3
4455	C	T	0.1	15	0.1	0.1	MINOR	ORF1AB	NS	A1397V	NSP3
4928	A	T	0.1	0.1	0.1	46	MINOR	ORF1AB	NS	N1555Y	NSP3
6696	C	CT	3	8	0.1	1	MINOR	ORF1AB	NS	L2146S*	NSP3
6750	A	G	0.1	7	0.1	0.1	MINOR	ORF1AB	NS	N2162S	NSP3
7444	A	G	0.1	0.1	10	0.1	MINOR	ORF1AB	S		
7749	C	T	0.1	7	0.1	0.1	MINOR	ORF1AB	NS	T2495I	NSP3
7866	G	T	0.1	0.1	6	0.1	MINOR	ORF1AB	NS	G2534V	NSP3
8782	C	T	97	99	99	99	CORE	ORF1AB	S		
9532	C	T	0.1	15	0.1	0.1	MINOR	ORF1AB	S		
11074	C	CT	1	5	0.1	0.1	MINOR	ORF1AB	NS	L3606F*	NSP6
12809	C	T	0.1	0.1	0.1	5	MINOR	ORF1AB	NS	L4182F	NSP9
12860	A	G	0.1	0.1	10	0.1	MINOR	ORF1AB	NS	S4199G	NSP9
13604	G	A	0.1	0.1	7	0.1	MINOR	ORF1AB	NS	R4447H	NSP12
16949	C	T	0.1	0.1	13	0.1	MINOR	ORF1AB	NS	P5562L	NSP13
17440	C	T	0.1	35	0.1	0.1	MINOR	ORF1AB	NS	P5726S	NSP13
18063	TA	T	0.1	0.1	9	0.1	MINOR	ORF1AB	NS	frameshift	
18488	T	C	98	97	96	97	CORE	ORF1AB	NS	I6075T	NSP14
18508	C	T	0.1	0.1	0.1	8	MINOR	ORF1AB	NS	L6082F	NSP14
19983	C	CT	7	7	0.1	6	MINOR	ORF1AB	NS	FRAMESHIFT. D6576*	NSP15
20178	C	T	0.1	0.1	0.1	5	MINOR	ORF1AB	S		
21697	C	T	13	0.1	8	0.1	MINOR	SPIKE	S		
21846	C	T	17	0.1	18	2	MINOR	SPIKE	NS	T95I	Spike
23277	C	T	0.1	0.1	10	0.1	MINOR	SPIKE	NS	T572I	Spike
23582	T	C	0.1	0.1	9	0.1	MINOR	SPIKE	NS	Y674H	Spike
23597	**AATTCTCCTCGGCGGGCACGTAGTG**	**A**	**6**	**11**	**0**	**2**	**MINOR**	**SPIKE**	**NS**	**NSPRRARSV679I**	**Spike**
23605	**T**	**G**	**97**	**86**	**94**	**95**	**CORE**	SPIKE	**S**		
26681	C	T	0.1	0.1	7	0.1	MINOR	M	S		M
27434	C	T	0.1	0.1	6	0.1	MINOR	ORF7A	NS	T14I	7A
27509	C	T	0.1	0.1	6	0.1	MINOR	ORF7A	NS	T39I	7A
27814	TTG	T	0.1	0.1	10	0.1	MINOR	ORF7B	NS	FRAMESHIFT	7B
28144	T	C	96	97	96	98	CORE	ORF8	NS	L84S	8
28393	T	C	0.1	11	0.1	0.1	MINOR	ORF8	S		
29274	C	T	0.1	23	0.1	0.1	MINOR	N	NS	T334I	N
29366	C	T	7	0.1	17	0.1	MINOR	ORF10	NS	P365S	N
29596	A	G	98	98	98	99	CORE	10	NS	I13M	10

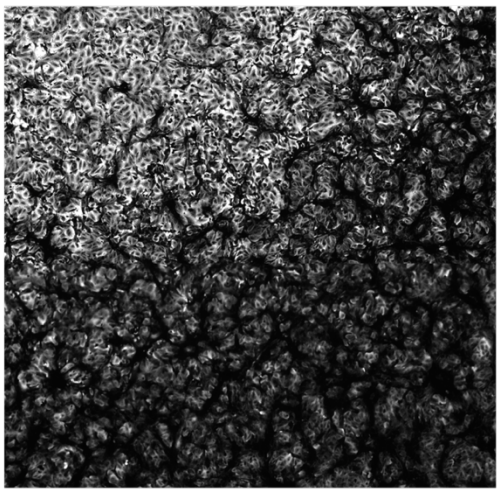
BT20.1 VERO										
POS	REFERENCE	VARIANT	P1	P2	P3	P4	CATEGORY	GENE	S/NS?	CONSEQUENCE	PROTEIN
241	C	T		100	100	100	CORE	5' UTR	S		
635	C	T		7	0.1	0.1	MINOR	ORF1AB	NS	R124C	NSP1
1420	C	T		98	99	99	CORE	ORF1AB	S		
1681	G	A		99	99	99	CORE	ORF1AB	S		
3037	C	T		98	99	99	CORE	ORF1AB	S		
6255	C	T		96	99	96	CORE	ORF1AB	NS	A1997V	NSP3
10870	**G**	**T**		0.1	**90**	**96**	**VARIANT**	ORF1AB	**S**		
14318	**C**	**T**		0.1	**85**	**96**	**VARIANT**	ORF1AB	**NS**	**T4685I**	**NSP12**
14408	C	T		98	97	98	CORE	ORF1AB	NS	P4715L	NSP12
16293	CATACGT	C		0.1	0.1	7	MINOR	ORF1AB	NS	CHANGE	
19983	C	CT		8	6	7	MINOR	ORF1AB	NS	FRAMESHIFT. D6576*	NSP15
22311	C	T		6	0.1	0.1	MINOR	SPIKE	NS	T205I	SPIKE
23403	A	G		97	98	99	CORE	SPIKE	NS	D614G	SPIKE
23997	**C**	**G**		0.1	**84**	**96**	**VARIANT**	SPIKE	**NS**	**P812R**	SPIKE
25314	G	T		3	5	0.1	MINOR	SPIKE	NS	G1251V	SPIKE
25317	C	G		3	5	7	MINOR	SPIKE	NS	S1252C	SPIKE
25521	C	T		18	0.1	7	MINOR	ORF3A	S		
27208	C	T		8	0.1	0.1	MINOR	ORF6	NS	H3Y	ORF6
27213	C	T		9	2	0.1	MINOR	ORF6	S		
27671	TTCAAG	T		100	96	96	CORE	ORF7A	NS	FRAMESHIFT and truncation	ORF7A
28253	**C**	**T**		**19**	**93**	**99**	**VARIANT**	**ORF8**	**S**		
28881	G	A		97	98	99	CORE	N	NS	R203K	N
28882	G	A		96	99	99	CORE	N	S		
28883	G	C		99	99	100	CORE	N	NS	G204R	N

BT20.1 WD-PNECS										
POS	REFERENCE	VARIANT	P1	P2	P3	P4	CATEGORY	GENE	S/NS?	CONSEQUENCE	PROTEIN
241	C	T		99	100	99	CORE	5' UTR	S	.	
1420	C	T		99	100	97	CORE	ORF1AB	S		
1681	G	A		97	98	99	CORE	ORF1AB	S		
3037	C	T		98	99	99	CORE	ORF1AB	S		
6255	C	T		91	89	93	CORE	ORF1AB	NS	A1997V	NSP3
6683	**AATT**	**A**		**46**	**45**	**52**	**VARIANT**	**ORF1AB**	**NS**		
14408	C	T		99	98	98	CORE	ORF1AB	NS	P4715L	NSP12
19983	C	CT		7	6	8	MINOR	ORF1AB	NS	FRAMESHIFT. D6576	NSP15
21101	G	GT		1	5	0.1	MINOR	ORF1AB	NS	FRAMESHIFT. F6948	NSP16
23403	A	G		97	99	98	CORE	SPIKE	NS	D614G	SPIKE
25314	G	T		0.1	6	1	MINOR	SPIKE	NS	G1251V	SPIKE
25317	C	G		0.1	6	1	MINOR	SPIKE	NS	S1252C	SPIKE
27671	TTCAAG	T		100	98	100	CORE	ORF7A	NS		
28881	G	A		99	97	99	CORE	N	NS	R203K	N
28882	G	A		96	97	99	CORE	N	S		
28883	G	C		99	99	99	CORE	N	NS	G204R	N
29418	T	A		0.1	5	0.1	MINOR	N	NS	L382*	N

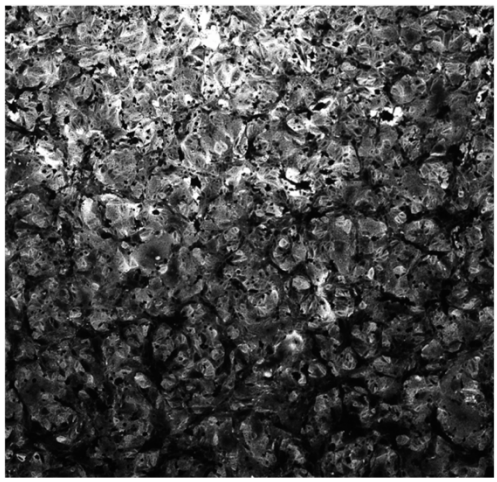
678

679 **Supplementary table 1.**

a.



b.

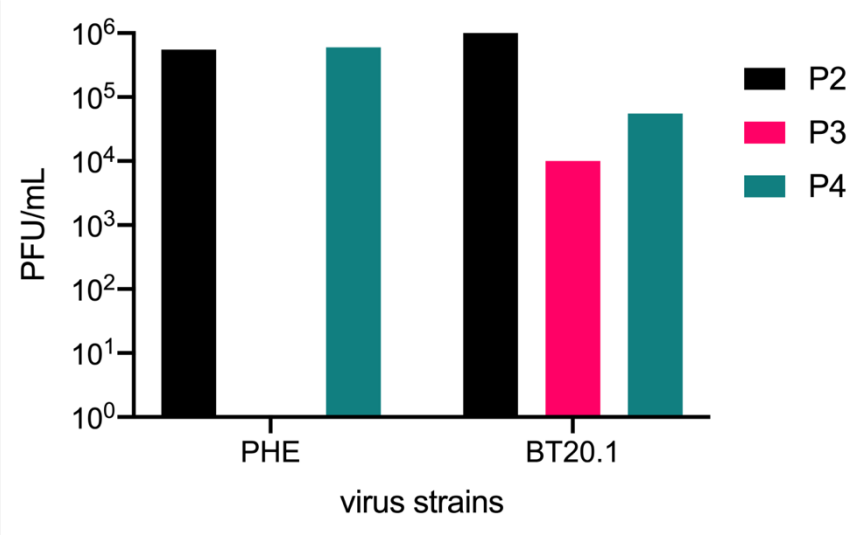


680

681 **Supplementary figure 1.**

Comparison of SARS-CoV-2 evolution in vitro during passage in Vero and primary human airway cells

a.



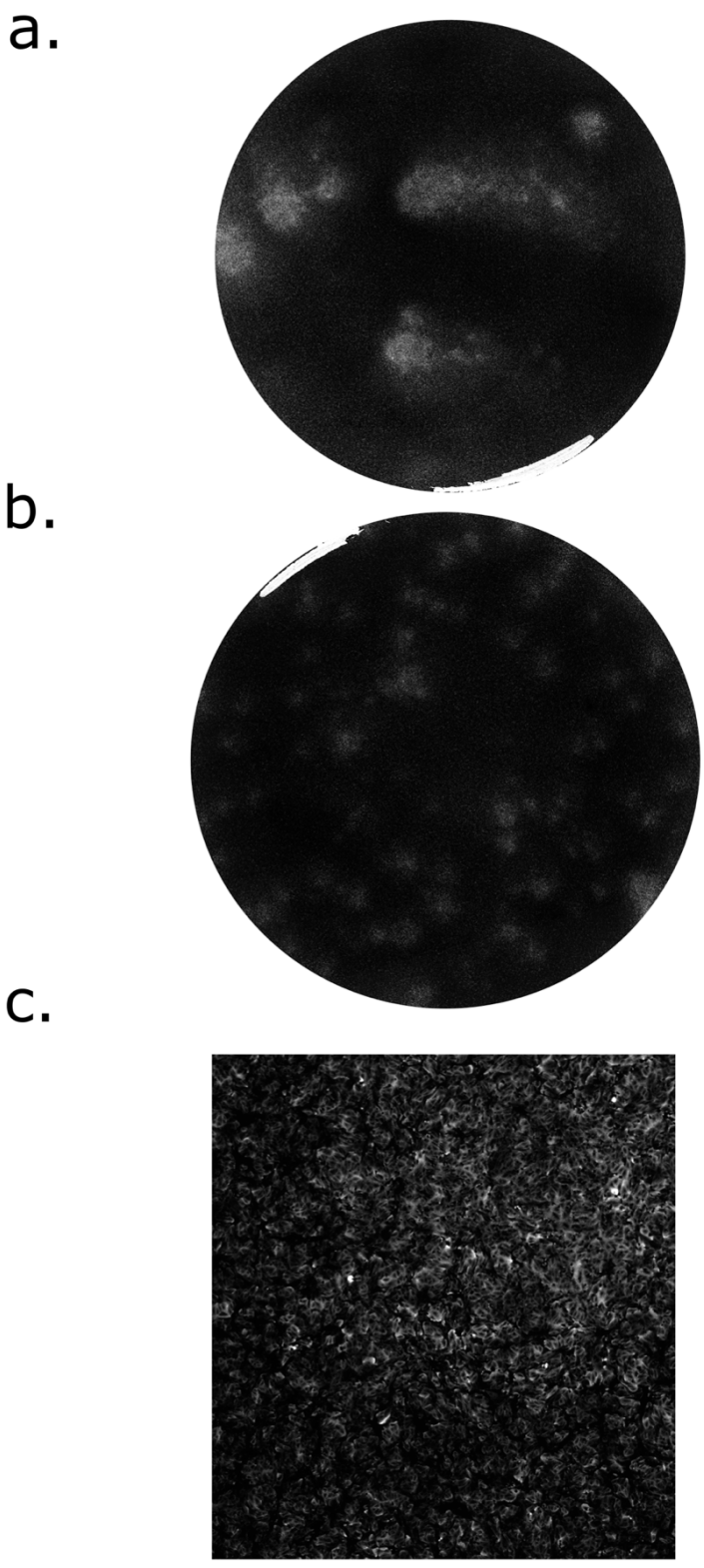
682

683 **Supplementary figure 2.**

684

685

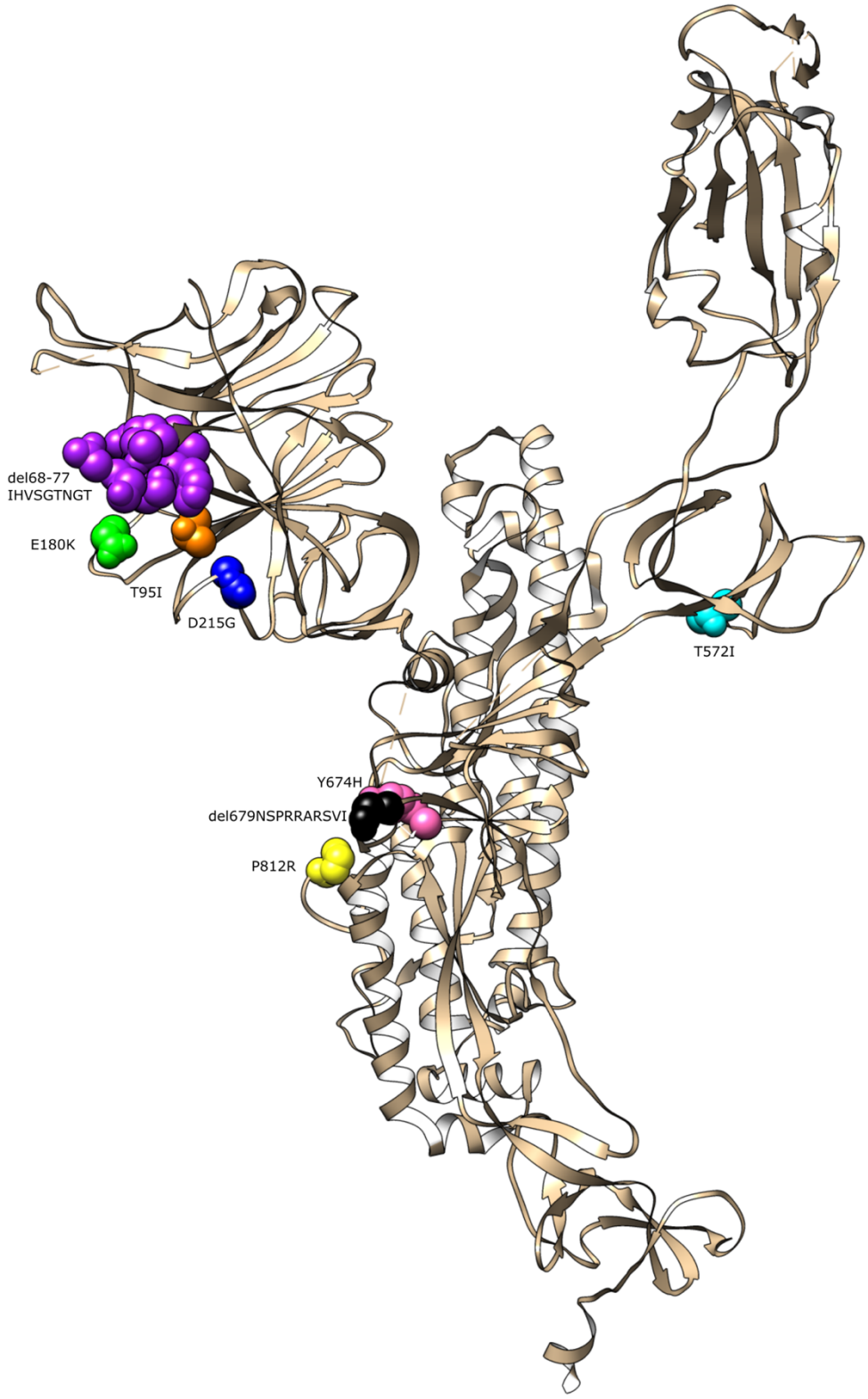
686



687

688 **Supplementary figure 3.**

689



690

691 **Supplementary figure 4.**

692

693

694 ## Acknowledgements

695 The authors would like to thank the following individuals and
696 organisations for provision of critical reagents, and technical assistance:
697 Public Health England and Professor Maria Zambon for the original SARS-
698 CoV-2 England 02/20 isolate; Belfast Health and Social Care Trust for the
699 original SARS-CoV-2 clinical material from which BT20.1 was derived;
700 Professor Alain Kohl (MRC-University of Glasgow Centre for Virus
701 Research) for the VeroE6 cells; and Dr Suzannah Rihn (MRC-University of
702 Glasgow Centre for Virus Research) for the VeroE6 cells expressing
703 human ACE2 and human TMPRSS2; QUB staff Mervyn McCaigue, Cathy
704 Fenning, David Norwood, Nuala McCann, Paul Crowe and Zoe Hunter for
705 their assistance in establishing and maintaining the QUB SARS-CoV-2
706 BSL3 facility; and finally, Professor Jose Bengoechea for unwavering
707 support for QUB virology at this critical time. We also wish to thank the
708 QUB Genomics Core Technology Unit for their help with sequencing.
709 Funding from UKRI/NIHR (MC_PC_19057) to UP, KM; PHA HSCNI R&D
710 Division (COM/5613/20) to UP, KM, LB, CGGB, and GLC; and generous
711 donations from the public to the Queen's University of Belfast Foundation
712 was used for the study.

713

714

715

References

Boni, M. F., Lemey, P., Jiang, X., Lam, T. T. Y., Perry, B. W., Castoe, T. A., Rambaut, A., & Robertson, D. L. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, *5*(11), 1408–1417. https://doi.org/10.1038/s41564-020-0771-4

Cevik, M., Bamford, C. G. G., & Ho, A. (2020). COVID-19 pandemic—a focused review for clinicians. In *Clinical Microbiology and Infection* (Vol. 26, Issue 7, pp. 842–847). Elsevier B.V. https://doi.org/10.1016/j.cmi.2020.04.023

Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., Zhang, Z., Fan, P., Dong, Y., Yang, Y., Chen, Z., Guo, Y., Zhang, J., Li, Y., Song, X., Chen, Y., Xia, L., Fu, L., Hou, L., … Chen, W. (2020). A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science*, *369*(6504), 650–655. https://doi.org/10.1126/SCIENCE.ABC6952

Davidson, A. D., Williamson, M. K., Lewis, S., Shoemark, D., Carroll, M. W., Heesom, K. J., Zambon, M., Ellis, J., Lewis, P. A., Hiscox, J. A., & Matthews, D. A. (2020). Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Medicine 2020 12:1*, *12*(1), 1–15. https://doi.org/10.1186/S13073-020-00763-0

Dieterle, M. E., Haslwanter, D., Bortz, R. H., Wirchnianski, A. S., Lasso, G., Vergnolle, O., Abbasi, S. A., Fels, J. M., Laudermilch, E., Florez, C., Mengotto, A., Kimmel, D., Malonis, R. J., Georgiev, G., Quiroz, J., Barnhill, J., Pirofski, L. anne, Daily, J. P., Dye, J. M., … Jangra, R. K. (2020). A Replication-Competent Vesicular Stomatitis Virus for Studies of SARS-CoV-2 Spike-Mediated Cell Entry and Its Inhibition. *Cell Host and Microbe*, *28*(3), 486-496.e6. https://doi.org/10.1016/j.chom.2020.06.020

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Gruning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. https://doi.org/10.1093/NAR/GKY379

Group, R. collaborative. (2021). Dexamethasone in Hospitalized Patients with Covid-19. *New England Journal of Medicine*, *384*(8), 693–704. https://doi.org/10.1056/nejmoa2021436

Guo-Parke, H., Canning, P., Douglas, I., Villenave, R., Heaney, L. G., Coyle, P. V., Lyons, J. D., Shields, M. D., & Power, U. F. (2013). Relative Respiratory Syncytial Virus Cytopathogenesis in Upper and Lower Respiratory Tract Epithelium. *Https://Doi.Org/10.1164/Rccm.201304-0750OC*, *188*(7), 842–851. https://doi.org/10.1164/RCCM.201304-0750OC

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., COVID-19 Genomics UK (COG-UK) Consortium, Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews. Microbiology*, *19*(7), 409–424. https://doi.org/10.1038/s41579-021-00573-0

Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., & Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. *Nature*, *584*(7819), 154–156. https://doi.org/10.1038/s41586-020-2368-8

763 Hoffmann, M., Kleine-Weber, H., & Pöhlmann, S. (2020). A Multibasic Cleavage Site in
764 the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells.
765 *Molecular Cell*, *78*(4), 779-784.e5. https://doi.org/10.1016/j.molcel.2020.04.022

766 Holden, B., Quinney, A., Padfield, S., Morton, W., Coles, S., Manley, P., Wensley, A.,
767 Hutchinson, C., Lillie, P. J., Duncan, C. J. A., Schmid, M. L., Li, A., Foster, K.,
768 Anaraki, S., Dabrera, G., Zambon, M., Hughes, G. J., & Gent, M. (2020). COVID-19:
769 public health management of the first two confirmed cases identified in the UK.
770 *Epidemiology and Infection*, *148*. https://doi.org/10.1017/S0950268820001922

771 Hou, Y. J., Okuda, K., Edwards, C. E., Martinez, D. R., Asakura, T., Dinnon, K. H., Kato,
772 T., Lee, R. E., Yount, B. L., Mascenik, T. M., Chen, G., Olivier, K. N., Ghio, A., Tse,
773 L. V., Leist, S. R., Gralinski, L. E., Schäfer, A., Dang, H., Gilmore, R., … Baric, R. S.
774 (2020). SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the
775 Respiratory Tract. *Cell*, *182*(2), 429-446.e14.
776 https://doi.org/10.1016/j.cell.2020.05.042

777 Emeny, J.M., & Morgan, M.J. (1979). Regulation of the interferon system: evidence that
778 Vero cells have a genetic defect in interferon production. *The Journal of General*
779 *Virology*, *43*(1), 247–252. https://doi.org/10.1099/0022-1317-43-1-247

780 Johnson, B. A., Xie, X., Bailey, A. L., Kalveram, B., Lokugamage, K. G., Muruato, A.,
781 Zou, J., Zhang, X., Juelich, T., Smith, J. K., Zhang, L., Bopp, N., Schindewolf, C.,
782 Vu, M., Vanderheiden, A., Winkler, E. S., Swetnam, D., Plante, J. A., Aguilar, P., …
783 Menachery, V. D. (2021). Loss of furin cleavage site attenuates SARS-CoV-2
784 pathogenesis. *Nature*, *591*(7849), 293–299. https://doi.org/10.1038/s41586-021-
785 03237-4

786 Klimstra, W. B., Tilston-Lunel, N. L., Nambulli, S., Boslett, J., Mcmillen, C. M., Gilliland,
787 T., Dunn, M. D., Sun, C., Wheeler, S. E., Wells, A., Hartman, A. L., Mcelroy, A. K.,
788 Reed, D. S., Rennick, L. J., & Duprex, W. P. (2020). SARS-CoV-2 growth, furin-
789 cleavage-site adaptation and neutralization using serum from acutely infected
790 hospitalized COVID-19 patients. *Journal of General Virology*, *101*, 1156–1169.
791 https://doi.org/10.1099/jgv.0.001481

792 Krammer, F. (2020). SARS-CoV-2 vaccines in development. In *Nature* (Vol. 586, Issue
793 7830, pp. 516–527). Nature Research. https://doi.org/10.1038/s41586-020-2798-3

794 Lamers, M. M., Mykytyn, A. Z., Breugem, T. I., Wang, Y., Wu, D. C., Riesebosch, S., van
795 den Doel, P. B., Schipper, D., Bestebroer, T., Wu, N. C., & Haagmans, B. L. (2021).
796 Human airway cells prevent sars-cov-2 multibasic cleavage site cell culture
797 adaptation. *ELife*, *10*. https://doi.org/10.7554/ELIFE.66815

798 Lei, J., Kusov, Y., & Hilgenfeld, R. (2018). Nsp3 of coronaviruses: Structures and
799 functions of a large multi-domain protein. In *Antiviral Research* (Vol. 149, pp. 58–
800 74). Elsevier B.V. https://doi.org/10.1016/j.antiviral.2017.11.001

801 Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu,
802 N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L.,
803 … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel
804 coronavirus: implications for virus origins and receptor binding. *The Lancet*,
805 *395*(10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

806 Maier, W., Bray, S., Beek, M. van den, Bouvier, D., Coraor, N., Miladi, M., Singh, B.,
807 Argila, J. R. De, Baker, D., Roach, N., Gladman, S., Coppens, F., Martin, D. P.,
808 Lonie, A., Grüning, B., Pond, S. L. K., & Nekrutenko, A. (2021). Freely accessible
809 ready to use global infrastructure for SARS-CoV-2 monitoring. *BioRxiv*,
810 2021.03.25.437046. https://doi.org/10.1101/2021.03.25.437046

Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., Nagata, N., Sekizuka, T., Katoh, H., Kato, F., Sakata, M., Tahara, M., Kutsuna, S., Ohmagari, N., Kuroda, M., Suzuki, T., Kageyama, T., & Takeda, M. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proceedings of the National Academy of Sciences*, *117*(13), 7001–7003. https://doi.org/10.1073/PNAS.2002589117

Nemudryi, A., Nemudraia, A., Wiegand, T., Nichols, J., Snyder, D. T., Hedges, J. F., Cicha, C., Lee, H., Vanderwood, K. K., Bimczok, D., Jutila, M. A., & Wiedenheft, B. (2021). SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. *Cell Reports*, *35*(9), 109197. https://doi.org/10.1016/j.celrep.2021.109197

Ogando, N. S., Dalebout, T. J., Zevenhoven-Dobbe, J. C., Limpens, R. W. A. L., van der Meer, Y., Caly, L., Druce, J., de Vries, J. J. C., Kikkert, M., Barcena, M., Sidorov, I., & Snijder, E. J. (2020). SARS-coronavirus-2 replication in Vero E6 cells: Replication kinetics, rapid adaptation and cytopathology. *Journal of General Virology*, *101*(9), 925–940. https://doi.org/10.1099/jgv.0.001453

Peacock, T. P., Goldhill, D. H., Zhou, J., Baillon, L., Frise, R., Swann, O. C., Kugathasan, R., Penn, R., Brown, J. C., Sanchez-David, R. Y., Braga, L., Williamson, M. K., Hassard, J. A., Staller, E., Hanley, B., Osborn, M., Giacca, M., Davidson, A. D., Matthews, D. A., & Barclay, W. S. (2021). The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nature Microbiology*, 1–11. https://doi.org/10.1038/s41564-021-00908-w

Pohl, M. O., Busnadiego, I., Kufner, V., Glas, I., Karakus, U., Schmutz, S., Zaheri, M., Abela, I., Trkola, A., Huber, M., Stertz, S., & Hale, B. G. (2021). SARS-CoV-2 variants reveal features critical for replication in primary human cells. *PLOS Biology*, *19*(3), e3001006. https://doi.org/10.1371/journal.pbio.3001006

Ramirez, S., Fernandez-Antunez, C., Galli, A., Underwood, A., Pham, L. V., Ryberg, L. A., Feng, S., Pedersen, M. S., Mikkelsen, L. S., Belouzard, S., Dubuisson, J., Sølund, C., Weis, N., Gottwein, J. M., Fahnøe, U., & Bukh, J. (2021). Overcoming culture restriction for SARS-CoV-2 in human cells facilitates the screening of compounds inhibiting viral replication. *Antimicrobial Agents and Chemotherapy*, *65*(7). https://doi.org/10.1128/aac.00097-21

Rihn, S. J., Merits, A., Bakshi, S., Turnbull, M. L., Wickenhagen, A., Alexander, A. J. T., Baillie, C., Brennan, B., Brown, F., Brunker, K., Bryden, S. R., Burness, K. A., Carmichael, S., Cole, S. J., Cowton, V. M., Davies, P., Davis, C., Lorenzo, G. De, Donald, C. L., … Mahalingam, S. (2021). A plasmid DNA-launched SARS-CoV-2 reverse genetics system and coronavirus toolkit for COVID-19 research. *PLOS Biology*, *19*(2), e3001091. https://doi.org/10.1371/JOURNAL.PBIO.3001091

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221–224. https://doi.org/10.1038/s41586-020-2179-y

Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J. H., Lapointe, H., Kamelian, K., Smith, A. D., Prystajecky, N., Goodfellow, I., Wilson, S. J., Harrigan, R., Snutch, T. P., Loman, N. J., & Quick, J. (2020). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *BioRxiv*, *3*, 2020.09.04.283077. https://doi.org/10.1101/2020.09.04.283077

Villenave, R., Thavagnanam, S., Sarlang, S., Parker, J., Douglas, I., Skibinski, G., Heaney, L. G., McKaigue, J. P., Coyle, P. V, Shields, M. D., & Power, U. F. (2012). In vitro modeling of respiratory syncytial virus infection of pediatric bronchial

epithelium, the primary target of infection in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 1–6. https://doi.org/10.1073/pnas.1110203109

Wang, Z., Yang, B., Li, Q., Wen, L., & Zhang, R. (2020). Clinical Features of 69 Cases with Coronavirus Disease 2019 in Wuhan, China. *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciaa272

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3