Check for updates

Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness

Ruian Ke[1], Pamela P. Martinez[2,3,4], Rebecca L. Smith[4,5,6], Laura L. Gibson[7], Agha Mirza[8], Madison Conte[8], Nicholas Gallagher[9], Chun Huai Luo[9], Junko Jarrett[9], Ruifeng Zhou[10], Abigail Conte[10], Tongyu Liu[2], Mireille Farjo[2], Kimberly K. O. Walden[11], Gloria Rendon[11], Christopher J. Fields[11], Leyi Wang[12], Richard Fredrickson[12], Darci C. Edmonson[4], Melinda E. Baughman[4], Karen K. Chiu[4], Hannah Choi[4], Kevin R. Scardina[4], Shannon Bradley[4], Stacy L. Gloss[4], Crystal Reinhart[4], Jagadeesh Yedetore[4], Jessica Quicksall[4], Alyssa N. Owens[13], John Broach[14,15], Bruce Barton[16,17], Peter Lazar[16], William J. Heetderks[18], Matthew L. Robinson[8], Heba H. Mostafa[9], Yukari C. Manabe[8,10], Andrew Pekosz[10], David D. McManus[19] and Christopher B. Brooke[2,4] ✉

The dynamics of SARS-CoV-2 replication and shedding in humans remain poorly understood. We captured the dynamics of infectious virus and viral RNA shedding during acute infection through daily longitudinal sampling of 60 individuals for up to 14 days. By fitting mechanistic models, we directly estimated viral expansion and clearance rates and overall infectiousness for each individual. Significant person-to-person variation in infectious virus shedding suggests that individual-level heterogeneity in viral dynamics contributes to 'superspreading'. Viral genome loads often peaked days earlier in saliva than in nasal swabs, indicating strong tissue compartmentalization and suggesting that saliva may serve as a superior sampling site for early detection of infection. Viral loads and clearance kinetics of Alpha (B.1.1.7) and previously circulating non-variant-of-concern viruses were mostly indistinguishable, indicating that the enhanced transmissibility of this variant cannot be explained simply by higher viral loads or delayed clearance. These results provide a high-resolution portrait of SARS-CoV-2 infection dynamics and implicate individual-level heterogeneity in infectiousness in superspreading.

Transmission of SARS-CoV-2 by both presymptomatic and asymptomatic individuals has been a major contributor to the explosive spread of this virus[1–5]. Recent epidemiological investigations of community outbreaks have indicated that transmission of SARS-CoV-2 is highly heterogeneous, with a small fraction of infected individuals (often referred to as superspreaders) contributing a disproportionate share of forward transmission[6–8]. Transmission heterogeneity has also been implicated in the epidemic spread of several other important viral pathogens, including measles and smallpox[9]. Numerous behavioural and environmental explanations have been offered to explain transmission heterogeneity,

but the extent to which the underlying features of the infection process within individual hosts contribute towards the superspreading phenomenon remains unclear. Addressing this gap in knowledge will inform the design of more targeted and effective strategies for controlling community spread.

Viral infection is a highly complex process in which viral replication and shedding dynamics are shaped by the complex interplay between host and viral factors. Recent studies have suggested that the magnitude and/or duration of viral shedding in both nasal and saliva samples correlate with disease severity, highlighting the potential importance of viral dynamics in influencing infection

[1]T-6, Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM, USA. [2]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [3]Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [4]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [5]Department of Pathobiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [6]Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [7]Division of Infectious Diseases and Immunology, Departments of Medicine and Pediatrics, University of Massachusetts Medical School, Worcester, MA, USA. [8]Division of Infectious Diseases, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA. [9]Division of Medical Microbiology, Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [10]W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [11]High-Performance Biological Computing at the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [12]Veterinary Diagnostic Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [13]Center for Clinical and Translational Research, University of Massachusetts Medical School, Worcester, MA, USA. [14]UMass Memorial Medical Center, Worcester, MA, USA. [15]Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, MA, USA. [16]Division of Biostatistics and Health Services Research, University of Massachusetts Medical School, Worcester, MA, USA. [17]Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA. [18]National Institute for Biomedical Imaging and Bioengineering, Bethesda, MD, USA. [19]Division of Cardiology, University of Massachusetts Medical School, Worcester, MA, USA. ✉e-mail: cbrooke@illinois.edu

outcomes[10–13]. Variation in viral load has also been suggested to correlate with transmission risk[14]. In addition to implications for pathogenesis and transmission, defining the contours of viral shedding dynamics is also critical for designing effective surveillance, screening and testing strategies[15]. To date, studies aimed at describing the longitudinal dynamics of SARS-CoV-2 shedding have been limited by (1) sparse sampling frequency, (2) failure to capture the early stages of infection when transmission is most likely, (3) absence of individual-level data on infectious virus shedding kinetics and (4) biasing towards the most severe clinical outcomes[16–21]. This is also true for viruses beyond SARS-CoV-2, because the dynamics of natural infection in humans have not been described in detail for any acute viral pathogen.

Here we capture the longitudinal viral dynamics of mild and asymptomatic early acute SARS-CoV-2 infection in 60 people by recording daily measurements of both viral RNA shedding (from mid-turbinate nasal swabs and saliva samples) and infectious virus shedding (from mid-turbinate nasal swabs) for up to 14 days. We reveal a striking degree of individual-level heterogeneity in infectious virus shedding between individuals, thus providing a partial explanation for the central role of superspreaders in community transmission of SARS-CoV-2. We also directly compare the shedding dynamics of Alpha (B.1.1.7) and previously circulating non-Alpha viruses, revealing no substantial differences in nasal or saliva shedding. Altogether, these results provide a high-resolution, multiparameter empirical profile of acute SARS-CoV-2 infection in humans and implicate person-to-person variation in infectious virus shedding in driving patterns of epidemiological spread of the pandemic.

Description of cohort and study design
During the fall of 2020 and spring of 2021, all faculty, staff and students at the University of Illinois at Urbana-Champaign were required to undergo at least twice weekly quantitative PCR with reverse transcription (RT–qPCR) testing for SARS-CoV-2 (ref. [22]). We leveraged this large-scale, high-frequency screening programme to enrol symptomatic, presymptomatic and asymptomatic SARS-CoV-2-infected individuals. We enroled university faculty, staff and students who reported a negative RT–qPCR test result in the past 7 days and were either (1) within 24 h of a positive RT–qPCR result or (2) within 5 days of exposure to someone with a confirmed positive RT–qPCR result. These criteria ensured that we enroled people within the first days of infection.

We collected both nasal and saliva samples daily for up to 14 days to generate a high-resolution portrait of viral dynamics during the early stages of SARS-CoV-2 infection. Participants also completed a daily online symptom survey. Our study cohort was primarily young (median age, 28 years; range, 19–73 years), non-Hispanic white and skewed slightly towards males (Supplementary Table 1). All infections were either mild or asymptomatic, and none of the participants were ever hospitalized for COVID-19. All participants in this cohort reported that they had never been previously infected with SARS-CoV-2, and none were vaccinated against SARS-CoV-2 at the time of enrolment.

Early SARS-CoV-2 viral dynamics vary significantly between individuals
To examine viral dynamics at the individual level, we plotted cycle theshold (Ct)/cycle number (CN) values from both saliva and nasal swab samples (the RT–qPCR assay used for nasal swab samples reports CN values, an objective measure of the cycle number of the maximal rate of PCR signal increase, rather than Ct values. CN and Ct values are equivalent in suitability for quantitative estimates[23], Quidel SARS Sofia 2 antigen fluorescent immunoassay (FIA) results and viral culture data from nasal swabs, as a function of time relative to the lowest observed CN values (Fig. 1a and Extended Data Fig. 1).

In many cases we captured both the rise and fall of viral genome shedding in nasal and/or saliva samples. A comparison between individuals revealed substantial heterogeneity in shedding dynamics, with obvious differences in the duration of detectable infectious virus shedding, clearance kinetics and the temporal relationship between shedding in nasal and saliva compartments. Further, nine out of 60 individuals had no detectable infectious virus in nasal samples (Fig. 1a and Extended Data Fig. 1).

Generally, earlier positivity results in the viral culture assay (which suggests higher infectious viral loads) were associated with lower CN values in nasal samples (Fig. 1b). This is unsurprising, as both nasal viral genome load and viral infectivity were assayed using the same sample. Saliva Ct values tended to be higher than matched nasal samples, probably due in part to the lower molecular sensitivity of the specific saliva RT–qPCR assay used, which does not include an RNA extraction step[24]. For both sample types the relationship between viral culture results and Ct/CN values was not absolute, because several nasal swab samples with CN values >30 also tested positive for infectious virus. These data indicate that caution must be exercised when using a simple Ct/CN value cutoff as a surrogate for infectious status.

We also assessed the relationship between antigen FIA and viral culture results, and found that participants tested positive by antigen FIA on 93% of the days on which they also tested positive by viral culture (Fig. 1c). This finding is consistent with earlier cross-sectional studies examining the relationship between antigen test positivity and infectious virus shedding[25,26].

While the symptom profiles self-reported by study participants varied widely across individuals, all cases were mild and did not require medical treatment (Extended Data Fig. 2). To determine whether any specific symptoms correlated with viral culture positivity, we compared the reported frequencies for each symptom on days where individuals tested viral culture positive or negative (Extended Data Fig. 3). Muscle aches, runny nose and scratchy throat were significantly more likely to be reported on days when participants were viral culture positive, suggesting these specific symptoms as potential indicators of infectious status. No other symptoms examined exhibited a clear association with viral culture status. Self-reported symptom data from this study may be partially skewed by having been collected after participants were notified of their initial positive test result or potential exposure.

Within-host mechanistic models capture viral dynamics in nasal and saliva samples
To better quantify the specific features of viral dynamics within individuals, we implemented five within-host mechanistic models based on models developed previously for SARS-CoV-2 and influenza infection (Methods, Fig. 2a and Extended Data Fig. 4)[27–29]. We fit these models to viral genome loads derived from the observed Ct/CN values using a population mixed-effect modelling approach (Methods). The viral dynamics in nasal and saliva samples were distinct from each other in most individuals, indicating strong compartmentalization of the oral and nasal cavities. We thus fit the models to data from nasal and saliva samples separately. For each sample type, viral genome loads from four individuals remained very low or undetectable throughout the sampling period (Extended Data Fig. 1), suggesting that these individuals either (1) were enroled late during infection despite having a recent negative test result or (2) exhibited highly irregular shedding dynamics. Because we were primarily interested in early infection dynamics, data from these individuals were excluded. Altogether, we selected data from 56 out of 60 individuals for each sample type for model fitting. Addition of the excluded individuals did not change the main conclusions (analysis not shown).

To identify factors that might partially explain the observed variation in individual-level dynamics, for each model we tested
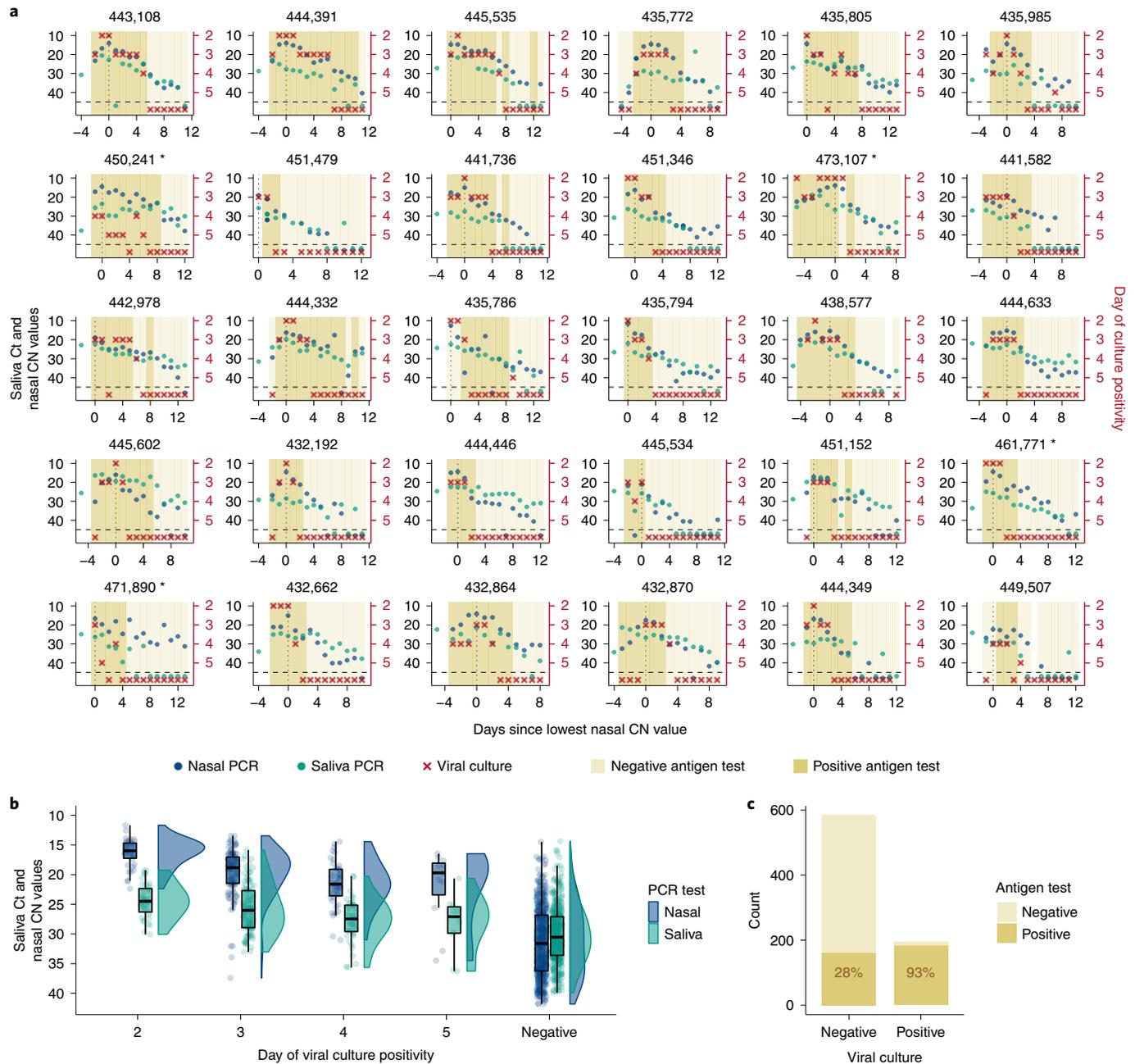
Fig. 1 | SARS-CoV-2 viral dynamics captured through daily sampling. a, Temporal trends for saliva RT–qPCR (teal dots), nasal swab RT–qPCR (navy blue dots), nasal swab viral culture (red crosses) and positive nasal swab antigen test results (dark mustard shaded area). The left-hand y axis indicates Ct values for saliva RT–qPCR assay (covidSHIELD) and CN values for nasal swab RT–qPCR assay (Abbott Alinity); the right-hand y axis indicates results of viral culture assays, where day of culture positivity indicates the day of incubation at which >50% of Vero-TMPRSS2 cells infected with the sample were positive for cytopathic effect. The vertical dotted line shows the day at which the lowest nasal CN value was observed while the horizontal dashed line indicates the limit of detection of RT–qPCR and viral culture assays. The title of each plot corresponds to the participant ID for the top 30 individuals with the most data points (the remaining 30 participants are shown in Supplementary Fig. 1). Asterisks next to participant ID indicate variant B.1.1.7. **b**, Individual Ct (for saliva) and CN (for nasal swabs) values from samples plotted based on concurrent results from viral culture assay. Negative indicates samples for which viral culture assay showed no viral growth after 5 days. The boxplot shows first and third quartiles (interquartile range, IQR), where the horizontal line is the median value and whiskers are 1.5× IQR. **c**, Plot showing antigen FIA results from day where participants tested either positive or negative by viral culture. Text within bars indicates the percentage of antigen FIA results that were positive when concurrent viral culture sample was positive or negative.

whether the age of participants or the infecting viral genotype (that is, non-B.1.1.7 versus B.1.1.7) covaried with any of the estimated model parameters in the model fitting. A total of 114 model variations were tested (see Methods). We compared the relative abilities of these model variations to capture RT–qPCR data using the

corrected Akaike information criterion (AICc) and found that, in general, the refractory and effector cell models best describe data from nasal and saliva samples, respectively (Supplementary Tables 2 and 3). In the refractory model (Fig. 2a), we assumed that target cells can be rendered refractory to infection through the activity

of soluble immune mediators released by infected cells such as interferon[30]. In the best-fit immune effector cell model (Fig. 2a), we assumed that innate and adaptive immune cells are activated and recruited to eliminate infected cells, leading to increased viral clearance[28]. See Supplementary Tables 4–6 for estimated values of the population and individual parameters and the fixed parameter values, respectively. Overall, these models described the observed Ct/CN values in both nasal and saliva samples very well (Fig. 2b).

The frequent longitudinal sampling of participants during early infection provided a unique opportunity for precise quantification of viral load kinetics during the viral expansion phase, before the peak in genome shedding. We estimated the mean early exponential expansion rate, r, before peak viral load (growth rate, for short) to be $4.4\,d^{-1}$ (s.d. $\pm 0.5\,d^{-1}$) in the nasal compartment. The growth rate is $8.8\,d^{-1}$ (s.d. $\pm 1.8\,d^{-1}$) in the saliva compartment, much higher than in the nasal compartment (Fig. 2c,d).

Viral clearance kinetics clearly differed between nasal and saliva samples (Fig. 2b–d). For nasal samples, viral genome loads decreased relatively quickly after peak, mostly driven by loss of productively infected cells, and we estimated an average death rate of productively infected cells at $2.5\,d^{-1}$ (s.d. $\pm 0.4\,d^{-1}$); however, viral decline slowed over time. In saliva, post-peak viral genome loads declined initially at a slower rate than that in nasal samples. Consequently, we estimated a much smaller average death rate of productively infected cells in saliva during this phase, at $0.4\,d^{-1}$ (s.d. $\pm 0.3\,d^{-1}$). However, our model suggested the existence of a second clearance phase with a more rapid decline occurring 1–2 weeks after infection, potentially due to the onset of effector cell and/or neutralizing antibody responses. Overall, we estimate that it takes on average $4.9\,d$ (s.d. $\pm 0.5\,d$) and $3.9\,d$ (s.d. $\pm 0.8\,d$) from infection to peak viral loads in the nasal and the saliva compartments, respectively (Fig. 2c,d). The average period from peak to undetectable genome viral load was $22.3\,d$ (s.d. $\pm 8.3\,d$) and $14.9\,d$ (s.d. $\pm 3.2\,d$) in the nasal and saliva compartments, respectively.

Interestingly, the model predicts a significant correlation ($P < 0.01$) in nasal samples between age and the Φ parameter, which describes the effectiveness of the antiviral immune response in rendering target cells refractory to infection (Fig. 2e). This suggests that innate immune responses are less effective at limiting SARS-CoV-2 in the nasal compartment of older individuals within our cohort, consistent with previous studies describing dysregulation of innate immunity to viral infection in aged individuals[31–33]. There was no significant correlation between age and either growth rate or clearance rate in nasal samples (Extended Data Fig. 5).

Overall, we noted a surprising degree of discordance in viral dynamics between nasal and saliva samples for many participants. In most individuals (46 out of 54 analysed), viral genome shedding peaked at least 1 day earlier in saliva than in nasal samples (Fig. 2f).

In contrast, the peak in nasal shedding preceded the saliva peak by at least 1 day in four individuals.

Significant heterogeneity in the infectious potential of individuals
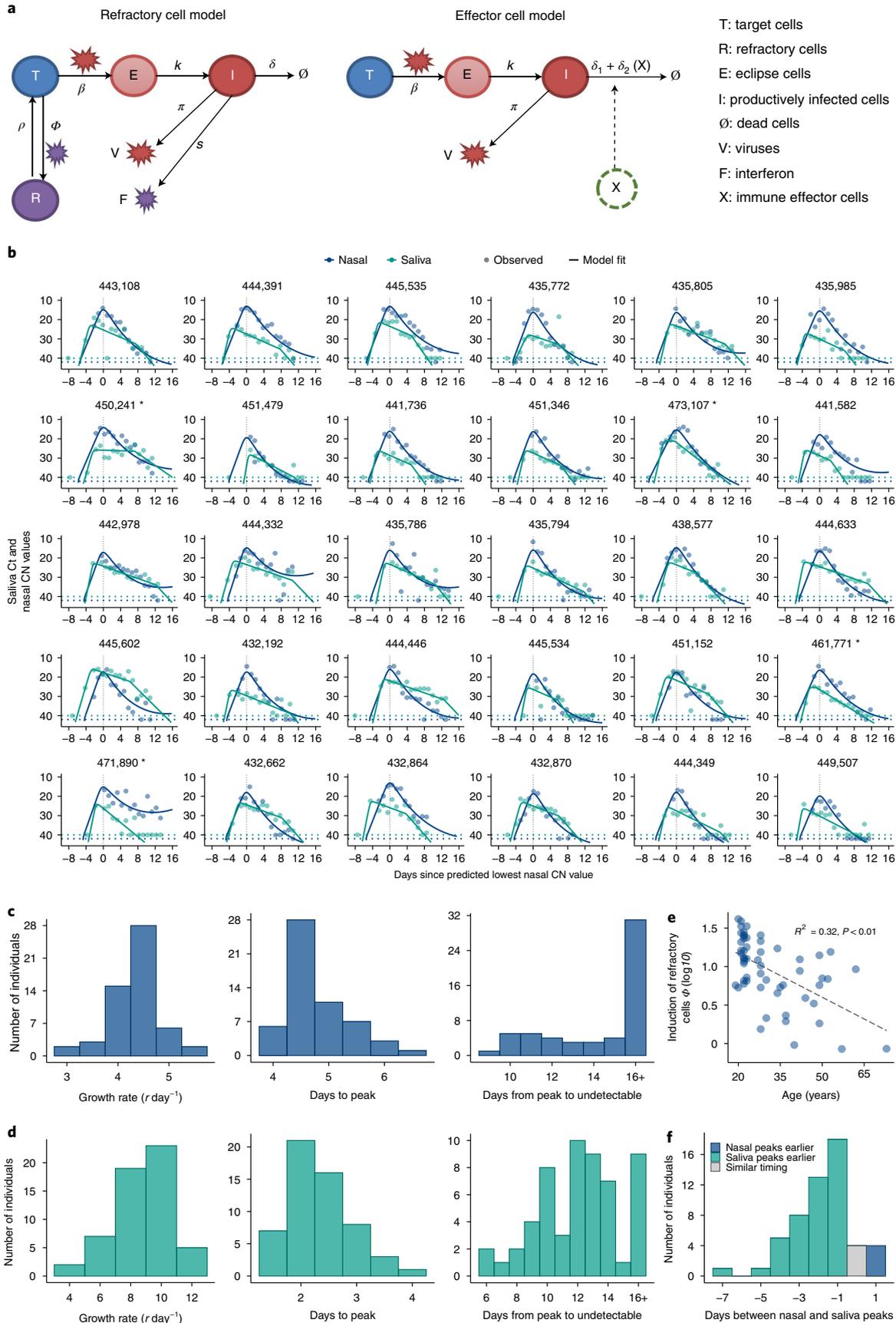
We next examined the duration of infectious virus shedding in nasal samples, as a surrogate for the infectious potential of an individual. There exists a large variation in the number of days for which an individual tested positive for cell culture on nasal swabs (Fig. 3a). Nine out of 60 individuals tested negative by viral culture throughout the sampling period, whereas one individual tested positive for 9 days (Fig. 3a). We found a weak positive correlation between the duration of viral culture positivity and participant age (Fig. 3b). Of note, many study participants were viral culture positive on the first day of sample collection, suggesting that we failed to capture the onset of viral culture positivity for these individuals and thus may be underestimating the duration of infectious virus shedding for a subset of study participants.

To better quantify the infectious potential of each individual, we first used viral culture data as a measure for intrinsic infectiousness (infectiousness for short, below) to characterize how infectiousness depends on viral genome load. We fitted three alternative models as previously proposed[27] to paired nasal RT–qPCR and viral culture data collected from each individual using a non-linear mixed-effect modelling approach (see Extended Data Fig. 6 for workflow and Methods for details). Comparing models using AICc scores, we found that the relationship is best described by a saturation model where the infectious virus load is a Hill-type function of viral genome load (Fig. 3c, Extended Data Fig. 7 and Supplementary Table 7). See Supplementary Table 8 for the best-fit parameter values.

Using the best-fit models, we estimated the infectiousness of each individual over the course of infection from their predicted genome viral loads and infectious viral loads (Extended Data Fig. 8). Note that the dataset allows us to estimate only a quantity that is a constant proportion of the infectious virus load (rather than its absolute value) across time and between individuals, and thus we report the predicted values in arbitrary units (a.u.) as a relative measure of infectiousness. Our model predicts that infectious virus shedding increases sharply when nasal CN values fall <22, and that the average amount of infectious virus shed is zero for CN values >29 (Fig. 3d). Importantly, there exists a high level of heterogeneity in infectiousness across different individuals that is not fully explained by differences in viral genome load (Fig. 3d). For example, at nasal CN values around 13, infectious virus shedding reached values >20 a.u. in three individuals while in 11 individuals it was <4 a.u. This suggests that viral Ct/CN values are not precisely predictive of infectiousness.

We next estimated the total infectiousness of each individual by integrating the area under the infectious virus load curve over the

Fig. 2 | Model fits quantify heterogeneity in viral dynamics and discordance in genome shedding between nasal and saliva samples. a, Diagrams outlining structures of the refractory cell and immune effector cell models that best fit nasal swab and saliva RT–qPCR data, respectively. In the refractory cell model, target cells (T) are infected by viruses (V) at rate β. Infected cells first become eclipse cells (E) and do not produce viruses; at rate k, eclipse cells become productively infected cells (I) producing both viruses and interferon (F) at rates π and s, respectively; they die at rate δ. Binding of interferons with target cells induces an antiviral response that converts target cells into cells refractory to infection (R). The rate of induction of the antiviral response is Φ. Refractory cells can revert to target cells at rate ρ. In the effector cell model we assume that, over the course of infection, immune effector cells (X) that clear infected cells are activated and recruited, leading to an increase in infected cell death rate from δ_1 to $\delta_1 + \delta_2$. **b**, Model fits to nasal sample (navy blue) and saliva (teal) RT–qPCR results from the same subset of individuals shown in Fig. 1a; includes last recorded negative saliva RT–qPCR result before study enrolment. Dotted lines represent the limit of detection (LoD) for RT–qPCR assays; dots on dotted lines denote measurements under LoD. **c,d**, Distributions of exponential viral growth rates, days from infection to peak viral genome load and days from peak to undetectable viral genome loads predicted by the refractory cell model (nasal data, **c**) and the immune effector model (saliva data, **d**) across 56 individuals in this cohort. **e**, Association between age and the estimated strength of innate immune response (Φ) based on nasal sample data. The y axis is shown on a $\log_{10}$ scale. Associations were examined using standard linear regression analysis, with R^2 and P values reported. **f**, Distribution of differences in estimated times of peak viral genome loads between saliva and nasal samples. Bars coloured teal and navy blue represent estimated saliva peaks that occurred at least 0.5 day earlier or later than nasal samples, respectively; grey bar indicates the number of individuals with similar timing in peaks.
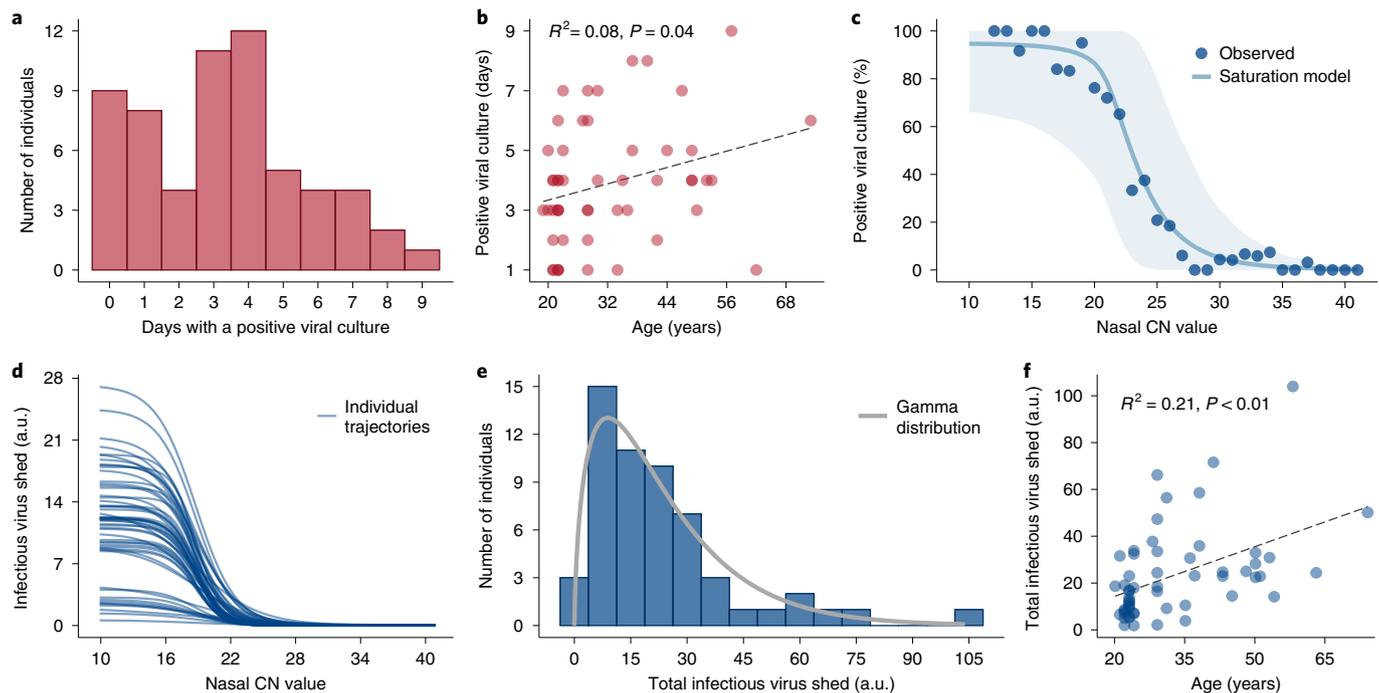
a

Refractory cell model

Effector cell model

T: target cells
R: refractory cells
E: eclipse cells
I: productively infected cells
Ø: dead cells
V: viruses
F: interferon
X: immune effector cells

b

Nasal — Saliva — Observed ● Model fit —

Days since predicted lowest nasal CN value

c

d

e

f

Fig. 3 | Substantial heterogeneity in infectious virus shedding between individuals. a, Distribution in the numbers of days in which participants tested positive by viral culture on nasal swabs following study enrolment. **b**, Association between age of study participants and number of days of positive viral culture using standard linear regression analysis. R^2 and P values are reported. **c**, Relationship between CN value in nasal samples and probability of the sample being viral culture positive, summarized across all individuals. Individual samples were binned based on their CN values. Dots indicate the observed percentage of positive samples within a bin that were viral culture positive. Solid line and shaded area represent mean and 90% CIs, respectively, of trajectories generated using the best-fit parameters of the saturation model (see Extended Data Fig. 7 for individual fits). **d**, Relationship between infectious virus shed and CN values by the saturation model for 56 individuals in our analysis. **e**, Distribution of estimated total cumulative level of infectious virus shed from nasal passages by each participant over the course of infection. Solid line shows the best-fit gamma distribution with a shape parameter of 1.6. **f**, Association between age and estimated total infectious virus shed by each individual. **b,f**, R^2 and the P values from linear regression (dashed line) are shown.

course of infection. This approach again revealed a large degree of heterogeneity in individual-level infectiousness, with >57-fold difference between the highest and lowest estimated infectiousness (104.0 and 1.8 a.u., respectively; Fig. 3e). We found that a gamma distribution with a shape parameter of 1.6 describes the distribution of individual infectiousness well (Fig. 3e). These data suggest that the previously reported heterogeneity in secondary transmission rates[6,7] is likely to arise from a combination of heterogeneity in contact structure and heterogeneity in intrinsic infectiousness[34]. This emphasizes the potential for a small subset of individuals that exhibit high intrinsic infectiousness to function as superspreaders if they have frequent and/or high-risk contacts during the infectious period. Finally, we observed a significant correlation between age and total infectiousness ($P < 0.01$, $R^2 = 0.21$; Fig. 3f).

Analysis of B.1.1.7 viral dynamics

Finally, we asked whether infection with the B.1.1.7 (Alpha) variant of concern (VOC) is associated with any significant differences in viral dynamics that could potentially explain the enhanced transmissibility of this genotype[35–37]. Previous studies have suggested that B.1.1.7 infection may result in higher peak viral loads or prolonged shedding compared with previously circulating genotypes[38–40]. Within our cohort, 16 out of 60 individuals were infected with B.1.1.7.

Both the empirical data and our model analysis (Fig. 4a,c) suggest that the overall viral genome shedding dynamics in both nasal and saliva samples are indistinguishable between B.1.1.7 and non-B.1.1.7 infections (none of the latter were VOC genotypes except for a single P.1 (Gamma) infection; Supplementary Table 9).

Although comparison of parameter estimates in nasal samples suggested a slightly slower growth rate and time to peak for B.1.1.7 versus non-B.1.1.7 (Fig. 4b), it is not clear whether this difference is biologically meaningful (Fig. 4a). Most importantly, we estimate that there is no significant difference between B.1.1.7 and non-B.1.1.7 viruses in total infectiousness in the nasal compartment (Fig. 4b). Previously, we have shown that the area under the logarithm of genome viral loads, denoted as AUC(log), may serve as a surrogate for infectiousness[27]. Here we calculated AUC(log) from predicted viral load trajectories in the saliva compartment in each individual and found no difference between B.1.1.7 and non-B.1.1.7 viruses (Fig. 4d). These data indicate that other mechanisms not reflected in viral shedding dynamics drive the increased transmissibility of the B.1.1.7 (Alpha) variant.

Discussion

This study describes the results of daily multicompartment sampling of viral dynamics within dozens of individuals newly infected with SARS-CoV-2 and provides a comprehensive, high-resolution description of viral shedding and clearance dynamics in humans.

Superspreading, in which a small subset of infected individuals are responsible for a disproportionately large share of transmission events, has been identified as a major driver of community spread of SARS-CoV-2, SARS-CoV and many other acute viral pathogens[6,7,9]. Superspreading is believed to arise from heterogeneity in both (1) contact structure between individuals arising from behavioural and environmental factors and (2) the intrinsic infectiousness of individuals[9,34,41]. While heterogeneity in contact structure has been studied extensively[42–45], the extent of heterogeneity in infectiousness
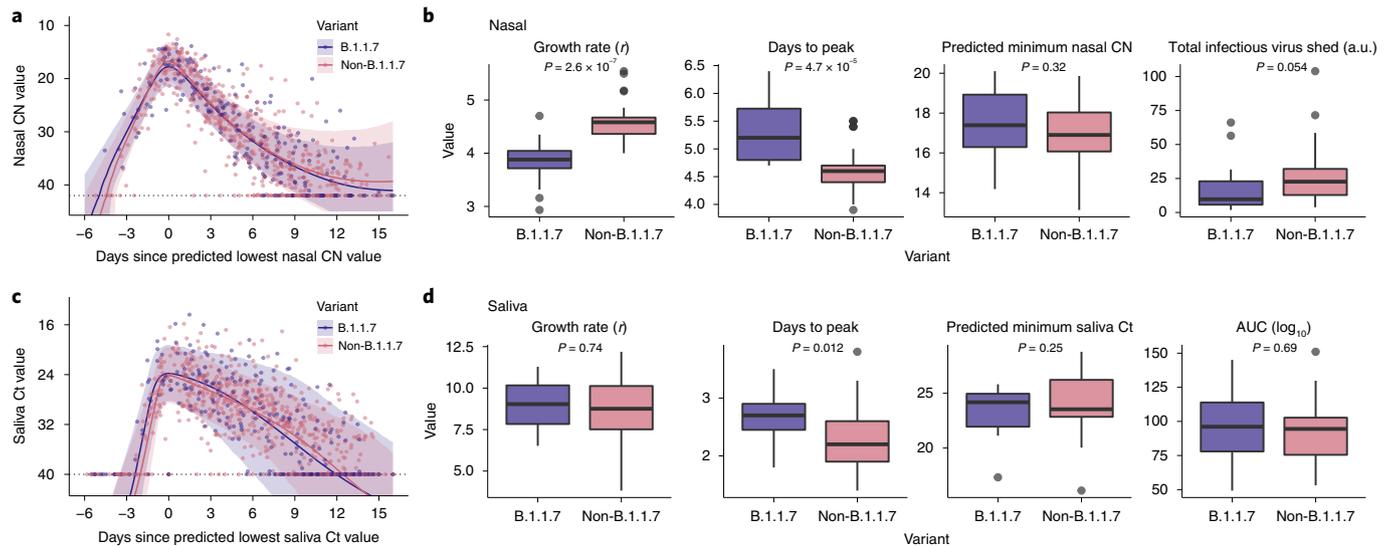
Fig. 4 | Comparison of viral dynamics between B.1.1.7 and non-B.1.1.7 viruses. a–d, Comparison of viral dynamics between B.1.1.7 and non-B.1.1.7 viruses based on nasal (**a,b**) and saliva samples (**c,d**). **a,c,** Viral genome load of B.1.1.7 infections (purple) and non-B.1.1.7 infections (pink) over time, as measured by RT–qPCR (dots). Ribbons indicate 90% CIs of predicted CN and Ct value trajectories, respectively, using population parameters estimated from modelling analysis (Supplementary Table 4). **b,d,** Comparison of estimated values for indicated summary statistics of viral dynamics between individual B.1.1.7 and non-B.1.1.7 infections ($n = 14$ and 42 for B.1.1.7 and non-B.1.1.7 infections, respectively). Boxes in boxplots start in the first quartile and end in the third quartile of the data (for example, IQR), and lines within boxes represent the median; whiskers represent 1.5× IQR and filled circles represent outliers. P values for Wilcoxon rank-sum test are reported. **b,** Note that, because age covaries with total infectious virus shed (Fig. 3f), comparison of total infectious virus shed after adjusting for age is shown (Methods).

arising from individual-level viral dynamics remains unknown. Although several studies have attempted to quantify this[20,34], the lack of empirical measurement of viral genome load and infectious virus shedding dynamics during early infection, which is a critical period for SARS-CoV-2 transmission, prevents precise estimation.

To address this question, we empirically quantified infectious virus shedding through daily longitudinal sampling of individuals infected with SARS-CoV-2. The substantial heterogeneity in infectious virus shedding that we observed among individuals indicates that superspreading is probably driven by individual-level variation in specific features of the infection process, in addition to behavioural and environmental factors. We also found that heterogeneity in infectious virus shedding is only partly explained by individual-level heterogeneity in viral genome load dynamics, suggesting that additional factors such as variation in the timing and magnitude of the neutralizing antibody response might contribute[46]. Our results here suggest caution in assessing the infectiousness of an individual using viral genome load data alone. Further, the absence of clear viral genetic correlates of infectiousness within this dataset suggests the existence of specific host determinants of superspreading. While we identified age as a significant correlate of infectiousness, additional determinants probably exist. Defining these correlates could aid future efforts to mitigate community spread of the virus by helping identify individuals with elevated risk of becoming superspreaders.

Our finding that viral shedding often peaks earlier in saliva versus the nasal compartment, sometimes by several days, corroborates a recent study of four individuals[47] and has several important implications. First, saliva screening may be a more effective sample type than nasal swabs for detection of infected individuals before or early in the infectious period[48]. Early detection and isolation of infected individuals is absolutely critical for breaking transmission chains[15]. Moreover, early viral shedding from the oral cavity may contribute to the high prevalence of presymptomatic SARS-CoV-2 transmission. We were unable to directly assess viral infectivity in saliva, so it remains unclear whether the earlier peaks in viral RNA shedding that we observed in saliva reflect earlier shedding of transmission-competent virus. The earlier detection of virus in saliva also raises questions about the initial site of SARS-CoV-2 infection. A recent study demonstrated that both salivary glands and oral mucosal epithelium can support SARS-CoV-2 replication, suggesting that infection could be initiated within the oral cavity[49]. Alternatively, if infection is initiated in the nasopharynx or soft palate, viral RNA might be detectable in saliva before detection in the mid-turbinate swabs used in this study. The discordance in shedding dynamics between oral and nasal samples that we observed in many participants is consistent with a significant degree of compartmentalization between these adjacent but distinct tissue sites, as has been observed in animal models of influenza virus infection[50,51].

The specific mechanisms driving the enhanced transmissibility of the B.1.1.7 variant remain poorly understood. Recent studies have identified alterations in the structural conformation of the spike protein and enhanced antagonism of innate immunity by B.1.1.7 as potential contributors[52,53]. Contrary to previous clinical studies, we observed no significant differences in either peak viral loads or clearance kinetics between B.1.1.7 and non-B.1.1.7 viruses as measured in either nasal swabs or saliva. Our results are consistent with studies demonstrating the absence of a growth advantage for B.1.1.7 in primary human respiratory epithelial cells[54,55]. Similarly, a recent longitudinal study of RNA shedding observed no significant differences in mean peak viral RNA loads, clearance kinetics or infection duration of the Alpha and Delta variants compared with non-VOCs[39]. If the timing of symptom onset differs between B.1.1.7 and non-B.1.1.7 infections, it could potentially explain why cross-sectional analyses of viral loads might register lower Ct values for B.1.1.7 samples. These data suggest that the enhanced transmissibility of the B.1.1.7 variant may also be driven by features not reflected in shedding dynamics—for example, enhanced environmental stability or a lower infectious dose threshold.

This study has several limitations that must be considered. First, our study cohort was limited to faculty, students and staff of the University of Illinois at Urbana-Champaign and did not include anyone who was hospitalized for COVID-19. The limited demographic and clinical profile of this cohort means that our results may not reflect the dynamics that occur during severe and lethal infections and/or in populations not well represented in our study. Second, there are multiple potential sources of technical variation that could contribute to noise in our experimental measurements. These include variability in sample collection quality and the potential for detection of subgenomic viral RNA in our RT–qPCR assays. While we took steps to minimize variation in sample collection quality, including having all sample collections remotely observed by trained study staff, it is possible that some of the sample-to-sample variation we observed is due to differences in sample quality. Finally, it must be noted that the results of viral culture assays performed on nasal swabs may not perfectly correlate with the actual transmission potential of an individual.

Altogether, our data provide a high-resolution view of the longitudinal viral dynamics of SARS-CoV-2 infection in humans and implicate individual-level heterogeneity in viral shedding as playing a critical role in community spread of this virus.

Methods

This study was approved by the Western Institutional Review Board, and all participants provided informed consent.

Participants. All on-campus students and employees of the University of Illinois at Urbana-Champaign are required to submit saliva for RT–qPCR testing every 2–4 days as part of the SHIELD campus surveillance testing programme. Individuals testing positive were instructed to isolate and were eligible to enrol in this study for a period of 24 h following receipt of their positive test result. Close contacts of individuals who test positive (particularly those co-housed with them) are instructed to quarantine and were eligible to enrol for up to 5 days after their last known exposure to an infected individual. All participants were also required to have received a negative saliva RT–qPCR result 7 days before enrolment.

Individuals were recruited via either a link shared in an automated text message providing isolation information sent within 30 min of a positive test result, a call from a study recruiter or a link shared by an enroled study participant or included in information provided to all quarantining close contacts. In addition, signs were used at each testing location and a website was available to inform the community about the study.

Participants were required to be at least 18 years of age, have a valid university ID, speak English, have Internet access and live within 8 miles of the university campus. After enrolment and consent, participants completed an initial survey to collect information on demographics and health history and were provided with sample collection supplies. Participants who tested positive before enrolment or during quarantine were followed for up to 14 days. Quarantining participants who continued to test negative by saliva RT–qPCR were followed for up to 7 days after their last exposure. All participants' data and survey responses were collected in the Eureka digital study platform. All study participants were asked whether they had previously tested positive for SARS-CoV-2 or been vaccinated against SARS-CoV-2. All participants included in this cohort reported no previous SARS-CoV-2 infection and were unvaccinated at the time of enrolment.

Sample collection. Each day, participants were remotely observed by trained study staff, who collected the following samples.

(1) Saliva (2 ml), into a 50-ml conical tube
(2) One nasal swab from a single nostril using a foam-tipped swab that was placed within a dry collection tube
(3) One nasal swab from the other nostril using a flocked swab that was subsequently placed in a collection vial containing 3 ml of viral transport medium (VTM). Swab and VTM manufacturer were not changed throughout the study.

The order of nostrils (left versus right) used for the two different swabs was randomized. For nasal swabs, participants were instructed to insert the soft tip of the swab at least 1 cm into the indicated nostril until they encountered mild resistance, rotate the swab around the nostril five times and leave it in place for 10–15 s. After daily sample collection, participants completed a symptom survey. A courier collected all participant samples within 1 h of sampling using a no-contact pickup protocol designed to minimize courier exposure to infected participants.

Saliva RT–qPCR. After collection, saliva samples were stored at room temperature and RT–qPCR was run within 12 h of initial collection in a Clinical Laboratory

Improvement Amendments (CLIA)-certified diagnostic laboratory. The protocol for the covidSHIELD direct saliva-to-RT–qPCR assay used has been detailed previously[24]. In brief, saliva samples were heated at 95 °C for 30 min followed by the addition of 2× Tris/Borate/EDTA buffer (TBE) at a 1:1 ratio (final concentration 1× TBE) and Tween-20 to a final concentration of 0.5%. Samples were assayed using the Thermo Taqpath COVID-19 assay.

Antigen testing. Foam-tipped nasal swabs were placed in collection tubes, transported in cold packs and stored at 4 °C overnight based on guidance from the manufacturer. The morning after collection, swabs were run through the Sofia SARS antigen FIA on Sofia devices according to the manufacturer's protocol.

Nasal swab RT–qPCR. Collection tubes containing VTM and flocked nasal swabs were stored at −80 °C after collection and were subsequently shipped to Johns Hopkins University for RT–qPCR and virus culture testing. After thawing, VTM was aliquoted for RT–qPCR and infectivity assays. One millilitre of VTM from the nasal swab was assayed on the Abbott Alinity, according to the manufacturer's instructions, in a College of American Pathologist and CLIA-certified laboratory.

Calibration curve for nasal swab RT–qPCR assay. Calibration curves for Alinity assay were determined using digital droplet PCR (ddPCR) as previously described[56]. Nasal swab samples previously quantified using the Alinity assay were stored in a freezer at −80 °C between initial quantification and extraction for calibration curves. Samples were extracted simultaneously using the Perkin Elmer Chemagic 360 automated extraction platform, with sample input and eluate volumes of 300 and 60 μl, respectively. RNA eluates were stored at −80 °C. Digital droplet RT–PCR was performed following the Bio-Rad EUA assay package insert (https://www.fda.gov/media/137579/download). A master mix was prepared per sample using the reagents provided in the ddPCR Supermix for Probes kit as follows: 5.5 μl of SuperMix (Bio-Rad), 2.2 μl of reverse transcriptase (Bio-Rad), 1.1 μl of dithiothreitol (Bio-Rad), 1.1 μl of CDC triplex SARS-CoV-2 primer and probe mix (IDT) and 7.1 μl of nuclease-free water; 17 μl of master mix was then transferred to a 96-well PCR plate and combined with 5 μl of RNA in eluate, and the plate was then loaded on to a QX-200 automated droplet generator (Bio-Rad). The droplet-containing plate was then heat sealed with foil in a plate sealer (Bio-Rad) and placed on a C1000 Touch thermal cycler (Bio-Rad) to perform reverse transcription and amplification. Droplets were read using the QX-200 droplet reader (Bio-Rad). Data were analysed with QuantaSoft Analysis Pro 1.0 software.

Virus culture from nasal swabs. Vero-TMPRSS2 cells were grown in complete medium (CM) consisting of DMEM with 10% foetal bovine serum (Gibco), 1 mM glutamine (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 100 U ml⁻¹ penicillin (Invitrogen) and 100 μg ml⁻¹ streptomycin (Invitrogen)[57]. Viral infectivity was assessed on Vero-TMPRSS2 cells as previously described using infection medium (identical to CM except that FBS is reduced to 2.5%)[26]. When a cytopathic effect was visible in >50% of cells in a given well, the supernatant was harvested. The presence of SARS-CoV-2 was confirmed through RT–qPCR, as described previously, by extracting RNA from the cell culture supernatant using the Qiagen viral RNA isolation kit and performing RT–qPCR using N1 and N2 SARS-CoV-2-specific primers and probes, in addition to primers and probes for the human RNaseP gene with the CDC research-use-only 2019-Novel Coronavirus (2019-nCoV) Real-time RT–PCR primer and probes sequences, and utilizing synthetic RNA target sequences to establish a standard curve[58].

Viral genome sequencing and analysis. Viral RNA was extracted from 140 μl of heat-inactivated (30 min at 95 °C, as part of the protocol detailed in ref. [24]) saliva samples using the QIAamp viral RNA mini kit (Qiagen); 100 ng of viral RNA was used to generate complementary DNA using the SuperScript IV first strand synthesis kit (Invitrogen). Viral cDNA was then used to generate sequencing libraries utilizing the Swift SNAP Amplicon SARS CoV2 kit with additional coverage panel and unique dual indexing (Swift Biosciences), which were sequenced on an Illumina Novaseq SP lane. Data were run through the nf-core/viralrecon workflow (https://nf-co.re/viralrecon/1.1.0) using the Wuhan-Hu-1 reference genome (NCBI accession NC_045512.2). Swift v.2 primer sequences were trimmed before variant analysis from iVar v.1.3.1 (https://doi.org/10.1186/s13059-018-1618-7), retaining all calls with a minimum allele frequency of 0.01 and higher. Viral lineages were called using the Pangolin tool (https://github.com/cov-lineages/pangolin) v.2.4.2, pango v.1.2.6 and the 5/19/21 version of the pangoLEARN model based on the nomenclature system described in ref. [59].

Statistics and reproducibility. Details of statistical analysis methods are given below. No statistical method was used to predetermine sample size. For some analyses, a small number of individuals were excluded for reasons detailed above, where relevant. Experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Statistical analyses. The difference in the distribution of a parameter of interest between the non-B.1.1.7 and B.1.1.7 infection groups was assessed using univariate

analysis, and *P* values calculated using the Wilcoxon rank-sum test. Comparison of infectious virus shedding between the two groups was performed using multivariate analysis with age as an additional variate. Levels of infectious viral shedding, after adjusting for age, were predicted by assuming an age of 28 years—that, is the median age of the cohort (Fig. 4c).

Generation of figures. All figures, except for Fig. 2a, were generated using RStudio. Figure 2a was generated using Microsoft Powerpoint.

Overview of model construction and parameter estimation. The goal of quantitative analyses is to use mathematical models to characterize viral shedding dynamics based on both viral genome loads (as measured by RT–qPCR) and the presence or absence of infectious virus (as measured by viral culture assay). Analysing the model results, we quantify individual-level heterogeneity in both viral genome shedding dynamics and individual infectiousness. See Extended Data Fig. 6 for an overview of the analysis workflow.

First, we performed experiments to derive the calibration curves for transformation of Ct/CN values from RT–qPCR to viral genome loads (Viral genome load calibration from Ct/CN values). Note that, due to the nature of RT–qPCR assays and sampling noise, viral genome loads derived using calibration curves represent a proxy for the actual quantities. Nonetheless, this approach is the best available to derive viral genome loads for the purpose of viral dynamic modelling, and is widely used in understanding SARS-CoV-2 dynamics[21,60].

Second, we constructed viral dynamic models and fit these to viral genome loads (Viral dynamics models). We estimated key parameters governing infection processes in the nasal- and the saliva-associated compartments, such as viral exponential growth rate before peak viral genome load and viral clearance rate. This allows us to characterize individual-level heterogeneity in infection kinetics.

Third, we constructed mathematical models to describe how the amount of infectious virus shed relates to changes in viral genome load, as measured by RT–qPCR (Modelling infectiousness of an individual). We fit the models to viral culture assay data. Using the best model and predicted viral genome load kinetics from the viral dynamics model, we predicted the extent of infectious virus shedding—that is the infectiousness, for each individual—and thus quantified the individual-level heterogeneity in infectiousness.

Viral genome load calibration from Ct/CN values. *Viral genome load calibration: nasal samples.* To calculate viral genome loads from CN values reported for nasal samples, we performed calibration curve experiments to empirically define the relationship between CN values obtained from the RT–qPCR assay used on nasal swab samples, and absolute viral genome loads within samples, as quantified by ddPCR. We quantified viral genome loads for 62 nasal samples with CN values ranging between 17 and 38. For each sample, absolute copy numbers of viral genomes were measured using two different N-gene-specific primer sets (N1 and N2). To account for technical noise between samples, we also determined the concentration of the host RNAse P (RP) transcript as a control (Supplementary Table 10). We then normalized copy numbers of N1 and N2 targets by dividing by their corresponding RP target numbers, then multiplied the mean of RP concentration across all samples. Note that the unit of these measurements is per millilitre: this is because nasal swab samples were each collected in 3 ml of VTM.

Plotting the logarithm of normalized viral genome loads against the associated CN values shows a clear linear relationship, justifying the use of linear regression below. Linear regression lines with similar coefficients were used as calibration curves in other studies[21,60]. We also note that the noise in genome viral loads is high when CN values are high (for example, >33), probably a reflection of increased noise when the signal is low[26]. However, this high level of variation at high CN values will not impact on the conclusion of our study, because the range of viral loads relevant to transmission is much higher (>10^6 copies ml⁻¹; Fig. 3d).

We then performed linear regression on measured CN values and log₁₀ viral genome loads (Extended Data Fig. 9). This led to the following formula for the relationship between CN values and viral genome load:

$$\log_{10} V = 11.35 - 0.25\text{CN}$$

where *V* and CN denote the viral genome load and CN value, respectively. Note that, because of the high number of data points measured, the level of uncertainty in the regression line is minimal (Extended Data Fig. 9).

Viral genome load calibration: saliva samples. Unlike for nasal samples, we were unable to measure the calibration curve using saliva samples taken from participants. To quantify the efficiency of the RT–qPCR assay used on saliva samples, we used data from calibration experiments in which saliva samples obtained from healthy donors were spiked with SARS-CoV-2 genomic RNA. More specifically, 0.9 ml of saliva from a healthy donor was spiked with 0.1 ml of 1.8×10^8, 5.4×10^5 or 6.0×10^4 RNA copies ml⁻¹. For samples spiked with 1.8×10^8 RNA copies ml⁻¹, tenfold serial dilutions were performed to a final concentration of 1.8×10^4 RNA copies ml⁻¹. A total of 24 samples were collected and Ct values of the N gene then measured (Supplementary Table 11).

As above, we plotted the logarithm of viral loads against Ct values (Extended Data Fig. 10). The plot shows a clear linear relationship, justifying the use of linear regression below. We then performed linear regression on measured CN values and log₁₀ viral genome loads (Extended Data Fig. 10). This led to the following formula for the relationship between CN values and viral genome load:

$$\log_{10} V = 14.24 - 0.28\text{Ct}$$

where *V* and Ct denote viral genome load and Ct value, respectively. In regard to the nasal calibration curve, the level of uncertainties in the regression line is minimal (Extended Data Fig. 10).

Note that a major difference between samples spiked with viral genomes and those taken from infected individuals is that the latter are likely to be noisier because of variation in the sample collection process. However, the two approaches should not differ substantially in assessing the efficiency of the RT–PCR protocol. The impact of noise in the nasal sample can be minimized by taking a large number of samples over a wide range of CN values, as we did for the nasal samples. Therefore, the calibration curves derived above represent an accurate translation of Ct/CN values to viral load.

Viral dynamics models. We constructed viral dynamics models to describe the dynamic changes in viral genome load. The viral genome load patterns in nasal and saliva samples are distinct from each other in many individuals, suggesting compartmentalization of infection dynamics in these two sample sites. Therefore, we use the models below to describe data collected from these two compartments separately. See Fig. 2a and Extended Data Fig. 4 for schematics of these models.

The target-cell-limited model. We first constructed a within-host model based on the target-cell-limited (TCL) model used for other respiratory viruses such as influenza[61] and, more recently, SARS-CoV-2 (refs. [27,29,62]). We keep track of the total numbers of target cells (*T*), cells in the eclipse phase of infection (*E*)—that is, infected cells not yet producing virus, productively infected cells (*I*) and viruses (*V*). The ordinary differential equations are:

$$\begin{aligned}
\frac{dT}{dt} &= -\beta VT \\
\frac{dE}{dt} &= \beta VT - kE \\
\frac{dI}{dt} &= kE - \delta I \\
\frac{dV}{dt} &= \pi I - cV
\end{aligned} \tag{1}$$

In this model, target cells are infected by virus with rate constant β, cells in the eclipse phase become productively infected cells at per-capita rate *k* and productively infected cells die at per-capita rate δ. We use *V* to describe viruses measured in nasal or saliva samples, representing a proportion of the total virus in the compartment under consideration. Therefore, rate π is the product of viral production rate per infected cell and the proportion of virus that is sampled (see Ke et al.[27] for a detailed derivation). Viruses are cleared at per-capita rate *c*.

Refractory cell model. We extend the TCL model by including an early innate response—that is the type-I/III interferon response, where interferons are secreted from infected cells and bind to receptors on uninfected target cells, stimulating an antiviral response that renders them refractory to viral infection. Note that this is the best model to describe the viral genome load dynamics as measured by RT–qPCR from nasal samples.

We keep track of interferon (*F*) and cells refractory to infection (*R*), in addition to other quantities in the TCL model. The full ordinary differential equations (ODEs) for target cells, refractory cells and interferon are

$$\begin{aligned}
\frac{dT}{dt} &= -\beta VT - \phi FT + \rho R \\
\frac{dR}{dt} &= \phi FT - \rho R \\
\frac{dE}{dt} &= \beta VT - kE \\
\frac{dI}{dt} &= kE - \delta I \\
\frac{dV}{dt} &= \pi I - cV \\
\frac{dF}{dt} &= sI - \mu F
\end{aligned} \tag{2}$$

In this model, the impact of the innate immune response is to convert target cells into refractory cells at rate ϕFT where ϕ is a rate constant. Refractory cells can become target cells again at rate ρ. Interferon is produced and cleared at rates *s* and μ, respectively.

For simplicity, and due to a lack of empirical data on interferon responses in our study, we simplify the model by making the quasi-steady-state assumption that the interferon dynamics are much faster than the dynamics of infected cells and assume that $\frac{dF}{dt} = 0$. Thus $sI = \mu F$ or $F = \frac{s}{\mu}I$.

Let $\Phi = \phi\frac{s}{\mu}$, so that the ODEs for the innate immunity model become:

$$\frac{dT}{dt} = -\beta VT - \Phi IT + \rho R$$

$$\frac{dR}{dt} = \Phi IT - \rho R$$

$$\frac{dE}{dt} = \beta VT - kE \qquad (3)$$

$$\frac{dI}{dt} = kE - \delta I$$

$$\frac{dV}{dt} = \pi I - cV$$

Viral production reduction model. In addition to making target cells refractory to infection, the impact of interferons may include reducing virus production from infected cells. We include this action of interferons in the viral production reduction model. As above, we make the quasi-steady-state assumption that interferon dynamics are much faster than those of infected cells and assume that F is proportional to I. The ODEs for the model are:

$$\frac{dT}{dt} = -\beta VT$$

$$\frac{dE}{dt} = \beta VT - kE$$

$$\frac{dI}{dt} = kE - \delta I \qquad (4)$$

$$\frac{dV}{dt} = \frac{\pi}{1+\gamma I} I - cV$$

where γ is a constant representing the effect of interferon in reducing viral production.

Immune effector cell model. Over the course of infection, immune effector cells are activated and recruited to kill infected cells. These immune effector cells include innate immune cells such as macrophages and natural killer cells, as well as cells developed during the adaptive immune response such as cytotoxic T lymphocytes and antibody-secreting B cells. To consider the impact of these immune effector cells, we develop a model—the effector cell model—based on a previous model for influenza infection[28]. In this model, we assume that the death rate of infected cells is δ_1 at the beginning of the infection. This may reflect the cytotoxic effects of viral infection. After time t_1, the death rate of infected cells increases by δ_2, where δ_2 models the killing of infected cells by immune effector cells. The ODEs for the model are:

$$\frac{dT}{dt} = -\beta VT$$

$$\frac{dE}{dt} = \beta VT - kE$$

$$\frac{dI}{dt} = kE - \delta(t)I \qquad (5)$$

$$\frac{dV}{dt} = \pi I - cV$$

$$\delta(t) = \begin{cases} \delta_1 & t < t_1 \\ \delta_1 + \delta_2 & t \geq t_1 \end{cases}$$

Note that this is the best model to describe the viral genome load dynamics as measured by RT–qPCR from saliva samples.

Combined model. In the full model, we combine the refractory cell model and immune effector cell model to consider both the immediate interferon response and immune effector response. The ODEs for the model are:

$$\frac{dT}{dt} = -\beta VT - \Phi IT + \rho R$$

$$\frac{dR}{dt} = \Phi IT - \rho R$$

$$\frac{dE}{dt} = \beta VT - kE$$

$$\frac{dI}{dt} = kE - \delta(t)I \qquad (6)$$

$$\frac{dV}{dt} = \pi I - cV$$

$$\delta(t) = \begin{cases} \delta_1 & t < t_1 \\ \delta_1 + \delta_2 & t \geq t_1 \end{cases}$$

Choice parameter values. <u>Total target cell numbers.</u> We calculate the total numbers of target cells in the nasal and saliva compartments by multiplying the total number of epithelial cells in these two compartments by the fraction of epithelial cells expected to be targets for SARS-CoV-2 infection.

For the total number of epithelial cells in the nasal compartment, we use the estimate from Baccam et al.[61], 4×10^8 cells. This is calculated from the estimate that the surface area of the nasal turbinates is 160 cm^2 (ref. [63]) and the surface area per epithelial cell is 2×10^{-11} to 4×10^{-11} m^2 per cell (ref. [61]). For the saliva compartment, the total surface area of the mouth was estimated to be 214.7 cm^2 (ref. [64]). Therefore, we estimate that the total number of epithelial cells in the mouth is approximately $4 \times 10^8 \times 214.7/160 = 5.4 \times 10^8$.

Hou et al. estimated that the fraction of cells expressing angiotensin-converting enzyme 2—that is, the receptor for SARS-CoV-2 entry—on the cell surface is approximately 20% in the upper respiratory tract[65]. Therefore, in our model, the initial numbers of target cells in the nasal and saliva compartments are calculated as $4 \times 10^8 \times 20\% = 8 \times 10^7$ and $5.4 \times 10^8 \times 20\% = 1.08 \times 10^8$, respectively.

Note that these estimates are approximations using available best estimates in the literature. For a standard viral dynamics model, the number of initial target cells and virus production rate are unidentifiable and only their product is identifiable[66]. Thus, if the actual number of target cells differs from that estimated here, an increase in the initial number of target cells will lead to a corresponding decrease in the estimate of virus production rate, and vice versa.

<u>Initial number of infected cells.</u> We assume that one cell in the compartment of interest is infected at the start of infection, $E_0 =$ one cell, consistent with refs. [27,67]. The small number of infected cells is also consistent with a recent work which estimated from sequencing data that the transmission bottleneck is small for SARS-CoV-2 and that there are probably between one and three infected cells at the initiation of infection[68–70]. Note that, in an earlier work, we showed that changes in the number of initially infected cells of between one and five in the model do not substaintially change the inference results[27].

Initial viral growth rate, r. For all models above, the initial growth of the viral population before peak viral genome load is dominated by viral infection. This means that the immune responses considered in our models act to change the viral growth trajectory substantially only at later time points[71]. Thus, we derive an approximation to the initial viral growth rate using the TCL model only (equation (1)). This approximation also represents a good approximation for other models.

We first make two simplifying assumptions commonly used in analysis of the initial dynamics of viral dynamic models[72,73]. First, because at the initial stage of infection the number of infected cells is orders of magnitude lower than the number of target cells, we assume that the number of target cells is at a constant level, T_0. Second, the dynamics of viruses are much faster than those of infected cells. For example, the rate of viral clearance is in the time scale of minutes and hours whereas the death of productively infected cells is in days. Therefore, we make the quasi-steady-state assumption, $\frac{dV}{dt} \approx 0$, such that the concentrations of viruses are always in proportion to the concentration of productively infected cells—that is, $\pi I \approx cV$. This gives $V \approx \frac{\pi}{c} I$.

With these two assumptions, equation (1) becomes a system of linear ODEs with two variables, E and I:

$$\frac{dE}{dt} = \beta \frac{\pi}{c} IT_0 - kE$$

$$\frac{dI}{dt} = kE - \delta I \qquad (7)$$

The Jacobian matrix, J, for this system of ODEs is:

$$J = \begin{bmatrix} -k & \beta\frac{\pi}{c}T_0 \\ k & -\delta \end{bmatrix}$$

The initial growth rate, r, is the leading eigenvalue of the Jacobian matrix of the ODE system. We calculate the eigenvalues, λ, for the Jacobian matrix above from $|J - \lambda I| = 0$, where I is the identity matrix, and get:

$\lambda = \frac{1}{2}\left[-(k+\delta) \pm \sqrt{(k+\delta)^2 + 4k\delta(R_0 - 1)} \right]$, where $R_0 = \frac{\beta\pi}{\delta c} T_0$.

Then, the leading eigenvalue—that is, the initial growth rate r—is:

$$r = \frac{1}{2}\left[-(k+\delta) + \sqrt{(k+\delta)^2 + 4k\delta(R_0 - 1)} \right]. \qquad (8)$$

Model fitting strategy. <u>Fitting viral dynamic models to viral genome load data.</u> We took a non-linear mixed-effect modelling approach to fit the viral dynamic models to viral genome load data from all individuals simultaneously. All estimations were performed using Monolix (Monolix Suite 2019R2, Lixoft: https://lixoft.com/products/monolix/). We allowed random effects on the fitted parameters (unless specified otherwise). All population parameters, except for the starting time of simulation, t_0, are positive and therefore we assume that they follow log-normal distributions. For t_0 we assume a normal distribution because t_0 can be positive or negative.

The parameters β and π in the viral dynamic models strongly correlate with each other when the models are fitted to viral genome load data[66]. We tested three choices in handling this correlation in fitting all five viral dynamic models: (1) a correlation is assumed between parameter β and π in Monolix; (2) parameter β has a fixed effect only (that is, its value is set to be the same across all individuals); and (3) parameter π has a fixed effect only.

To test whether the age of the individuals and/or the infecting viral genotype (categorized as either non-B.1.1.7 or B.1.1.7) explains the heterogenous patterns in viral genome load trajectories across the cohort, we tested whether they covary with any of the fitted parameters in the model by setting the two variables as a continuous and a categorical covariate, respectively, in Monolix.

The assumptions on parameters β and π and the choice of parameters that covariate with age or viral strain of infection led to a large number of model choices for fitting. Therefore, we took the following strategy to ensure that we identified the best model and parameter combinations to describe the data.

- First, we tested the three assumptions about parameters β and π in the five viral dynamic models without any covariate and selected the best assumption for further analysis based on their corrected Akaike information criterion (AICc) scores.
- Second, using the best assumption, we tested the model by including the age of the individuals as a continuous covariate of all fitted parameter values with a random effect first. We then took an iterative approach to test whether the covariate should be removed from any of the parameters in the model using Pearson's correlation test in Monolix. The parameter(s) that has a non-significant P value ($P > 0.05$) or with the lowest P value is removed from next round of parameter fitting. We iterated the process until all parameters were removed.
- The best model variant with the lowest AICc score was then selected for analysis on whether parameter estimates differed in individuals infected by different viral strains. As before, we took an iterative approach. We first set the viral strain—that is, non-B.1.1.7 or B.1.1.7—as a categorical covariate of all fitted parameter values with a random effect in the model. We then tested whether the covariate should be removed from any of the parameters in the model using the analysis of variance in Monolix. The parameter(s) that has a non-significant P value ($P > 0.05$) or with the lowest P value is removed from the next round of parameter fitting. We iterated the process until all parameters were removed.
- Finally, the model variant with the lowest AICc score was selected as the best model.

Prediction of viral genome load trajectories for non-B.1.1.7 and B.1.1.7 strains. We randomly sampled 5,000 sets of parameter combinations from the distribution specified by the best-fit population parameters (Supplementary Table 4). For the effector cell model for the saliva compartment, β and π are strongly correlated. We thus applied formulations such that correlations between the two parameter values are preserved in the random sampling in accordance with the estimated correlation coefficient. We simulated the best-fit model using the 5,000 sets of parameter combinations for each of the strain. The median and the fifth and 95th quantilse of viral genome loads at each time points are reported.

Modelling infectiousness of an individual. We model how infectiousness depends on the viral genome load in an individual, similarly to the framework proposed in Ke et al.[27]. Specifically, we first use the viral culture data collected in this study to infer how the level of infectious virus shed relates to viral genome loads as measured by RT–qPCR. From this model, we predict how the level of infectious virus shedding changes over time in each individual and how the overall infectiousness of the infection varies among participants.

Relationship between viral genome load and infectious viruses. We first consider three alternative models describing how the amount of infectious virus in a sample is related to viral genome load (derived from the CN values): the 'linear' model, 'power-law' model and 'saturation' model. In these models, due to the nature of stochasticity in sampling, we assume the number of infectious viruses that was in the sample for cell culture experiment to be a random variable, Y, that follows a Poisson distribution, with $V_{\inf}$ representing the expected number of infectious viruses—that is, $V_{\inf} = E(Y)$.

(1) The linear model:
We assume that $V_{\inf}$, is proportional to the viral genome load, V, in the sample:

$$V_{\inf} = E(Y) = AV \qquad (9)$$

where A is a constant.

(2) The power-law model:
We assume that $V_{\inf}$ is related to the viral genome load, V, by a power function:

$$V_{\inf} = E(Y) = BV^h \qquad (10)$$

where B and h are constants.

(3) The saturation model:

We assume that $V_{\inf}$ is related to the viral genome load, V, by a Hill function:

$$V_{\inf} = E(Y) = V_m \frac{V^h}{V^h + K_m^h} \qquad (11)$$

where V_m and K_m are constants and h is the Hill coefficient.

Probability of cell culture being positive. If each infectious virus has a probability ϱ to establish infection such that the cell culture becomes positive, the number of

viruses that successfully establish an infection in cell culture is Poisson distributed with parameter $\lambda = E(Y)\varrho = V_{\inf}\varrho$. Thus, the probability of one or more viruses successfully infecting the culture so that it tests positive is

$$p_{\text{positive}} = 1 - \exp(-\lambda) = 1 - \exp(-V_{\inf}\varrho) \qquad (12)$$

Substituting the expressions of $V_{\inf}$ from the three models above, we get the following expressions for p_{positive} from the three models (note that we use the subscripts '1', '2' and '3' to denote the three models for $V_{\inf}$):

$$p_{\text{positive},1} = 1 - \exp(-V_{\inf}\varrho) = 1 - \exp(-DV) \qquad (13)$$

where $D = A\varrho$.

$$p_{\text{positive},2} = 1 - \exp(-V_{\inf}\varrho) = 1 - \exp\left(-GV^h\right) \qquad (14)$$

where $G = B\varrho$.

$$p_{\text{positive},3} = 1 - \exp(-V_{\inf}\varrho) = 1 - \exp\left(-J\frac{V^h}{V^h + K_m^h}\right) \qquad (15)$$

where $J = V_m\varrho$.

Note that, from the expressions above, it becomes clear that we will not be able to estimate parameters A, B and V_m in the three models because they appear as products with the unknown parameter ϱ in the equations. This means that the viral culture data do not allow us to estimate the absolute number of infectious viruses in a sample or provide a viral genome load; instead, we are able to estimate a quantity that is a constant proportion of the actual number of infectious viruses over time and across individuals. Therefore, we report estimations of infectious viruses in arbitrary units. These estimates represent a relative measure of infectiousness. Two estimates measured at different time points and/or from different individuals can be compared using this method.

Model fitting using a population effect modelling approach. For each sample, viral genome load and cell culture positivity were measured. Using these data, we estimate parameter values in the three models by minimizing the negative log-likelihood of the data.

More specifically, the likelihood of the mth observation being positive or negative in cell culture is calculated as:

$$p_{i,m} = \begin{cases} p_{\text{positive},i}(V_m), & \text{if the } k\text{th observation is positive} \\ 1 - p_{\text{positive},i}(V_m), & \text{if the } k\text{th observation is negative} \end{cases} \qquad (16)$$

where V_m is the viral genome load of the mth observation.

Because we have the paired nasal RT–qPCR and viral culture data for each individual, we fit the three mathematical models using a nonlinear mixed-effect modelling approach. Again, all estimations were performed using Monolix. We allowed random effects on the fitted parameters (unless specified otherwise). All population parameters with a random effect are assumed to follow log-normal distributions.

To find the best model explaining the data, we tested models with different combinations of parameters either with or without a random effect (Supplementary Table 7). The model with the lowest AIC score was selected as the best model.

Note that, for each of the three models, we tested a model variation where all parameters in the models have fixed effects only—that is, a single set of parameters is used to explain viral culture data from every individual. In this case, there is no heterogeneity in parameter values across individuals. The resulting AIC scores are significantly worse than the best-fit model assuming random effects on parameters (Supplementary Table 7). This indicates that there is a substantial level of individual heterogeneity in the relationship between infectious virus shedding and viral genome loads (as shown in Fig. 3d).

Calculation of CIs of the cell culture positivity curve (Fig. 3c). Similar to the procedures performed for prediction of CIs of viral genome load trajectories, we randomly sampled 5,000 sets of parameter combinations from the distribution specified by the best-fit population parameters of the best model—that is, the saturation model assuming that K_m has only a fixed effect (Supplementary Table 8). More specifically, we sampled parameters from a log-normal distribution for J and h, with their means and standard deviations at the best-fit values. Using the parameter combinations, we generated curves of probability of cell culture positivity at CN values ranging between 10 and 40. The median and the fifth and 95th quantiles of viral genome loads at each CN values are reported.

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

References

1. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
2. Ferretti, L. et al. The timing of COVID-19 transmission. Preprint at *medRxiv* https://www.medrxiv.org/content/10.1101/2020.09.04.20188516v2 (2020).
3. Szablewski, C. M. et al. SARS-CoV-2 transmission and infection among attendees of an overnight camp — Georgia, June 2020. *MMWR Morb. Mortal. Wkly Rep.* **69**, 1023–1025 (2020).
4. Long, Q.-X. et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020).
5. Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
6. Sun, K. et al. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science* **371**, eabe2424 (2021).
7. Adam, D. C. et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26**, 1714–1719 (2020).
8. Bi, Q. et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919 (2020).
9. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
10. Néant, N. et al. Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. *Proc. Natl Acad. Sci. USA* **118**, e2017962118 (2021).
11. Silva, J. et al. Saliva viral load is a dynamic unifying correlate of COVID-19 severity and mortality. Preprint at *medRxiv* https://www.medrxiv.org/content/10.1101/2021.01.04.21249236v2 (2021).
12. The Massachusetts Consortium for Pathogen Readiness et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat. Commun.* **11**, 5493 (2020).
13. Zheng, S. et al. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study. *BMJ* https://doi.org/10.1136/bmj.m1443 (2020).
14. Marks, M. et al. Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study. *Lancet Infect. Dis.* **21**, P629–P636 (2021).
15. Larremore, D. B. et al. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci. Adv.* **7**, eabd5393 (2021).
16. Kim, J. Y. et al. Viral load kinetics of SARS-CoV-2 infection in first two patients in Korea. *J. Korean Med. Sci.* **35**, e86 (2020).
17. Wölfel, R. et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020).
18. Lescure, F.-X. et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. *Lancet Infect. Dis.* **20**, 697–706 (2020).
19. Young, B. E. et al. Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. *JAMA* **323**, 1488 (2020).
20. Jones, T. C. et al. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science* **373**, 6551 (2021).
21. Kissler, S. M. et al. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. *PLoS Biol.* **19**, e3001333 (2021).
22. Ranoa, D. R. E. et al. Mitigation of SARS-CoV-2 transmission at a large public university. Preprint at *medRxiv* https://www.medrxiv.org/content/10.1101/2021.08.03.21261548v1 (2021).
23. Shain, E. B. & Clemens, J. M. A new method for robust quantitative and qualitative analysis of real-time PCR. *Nucleic Acids Res.* **36**, e91 (2008).
24. Ranoa, D. R. E. et al. Saliva-based molecular testing for SARS-CoV-2 that bypasses RNA extraction. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2020.06.18.159434v1 (2020).
25. Monel, B. et al. Release of infectious virus and cytokines in nasopharyngeal swabs from individuals infected with non-alpha or alpha SARS-CoV-2 variants: an observational retrospective study. *eBioMedicine* **73**, 1036372021 (2021).
26. Pekosz, A. et al. Antigen-based testing but not real-time polymerase chain reaction correlates with severe acute respiratory syndrome coronavirus 2 viral culture. *Clin. Infect. Dis.* **76**, e2861–e2866 (2021).
27. Ke, R., Zitzmann, C., Ho, D. D., Ribeiro, R. M. & Perelson, A. S. In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. *Proc. Natl Acad. Sci. USA* **118**, e2111477118 (2021).
28. Pawelek, K. A. et al. Modeling within-host dynamics of influenza virus infection including immune responses. *PLoS Comput. Biol.* **8**, e1002588 (2012).
29. Goyal, A., Cardozo-Ojeda, E. F. & Schiffer, J. T. Potency and timing of antiviral therapy as determinants of duration of SARS-CoV-2 shedding and intensity of inflammatory response. *Sci. Adv.* **6**, eabc7112 (2020).
30. Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045 (2020).
31. Shaw, A. C., Goldstein, D. R. & Montgomery, R. R. Age-dependent dysregulation of innate immunity. *Nat. Rev. Immunol.* **13**, 875–887 (2013).
32. Molony, R. D. et al. Aging impairs both primary and secondary RIG-I signaling for interferon induction in human monocytes. *Sci. Signal.* **10**, eaan2392 (2017).
33. Angioni, R. et al. Age-severity matched cytokine profiling reveals specific signatures in Covid-19 patients. *Cell Death Dis.* **11**, 957 (2020).
34. Goyal, A., Reeves, D. B., Cardozo-Ojeda, E. F., Schiffer, J. T. & Mayer, B. T. Viral load and contact heterogeneity predict SARS-CoV-2 transmission and super-spreading events. *eLife* **10**, e63537 (2021).
35. Washington, N. L. et al. Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* **184**, P2587–2594 (2021).
36. Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).
37. Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).
38. Calistri, P. et al. Infection sustained by lineage B.1.1.7 of SARS-CoV-2 is characterised by longer persistence and higher viral RNA loads in nasopharyngeal swabs. *Int. J. Infect. Dis.* **105**, 753–755 (2021).
39. Kissler, S. M. et al. Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated persons. *N. Engl. J. Med.* **385**, 2489–2491 (2021).
40. Althaus, C. L. et al. A tale of two variants: spread of SARS-CoV-2 variants Alpha in Geneva, Switzerland, and Beta in South Africa. Preprint at *medRxiv* https://www.medrxiv.org/content/10.1101/2021.06.10.21258468v1 (2021).
41. Lakdawala, S. S. & Menachery, V. D. Catch me if you can: superspreading of COVID-19. *Trends Microbiol.* **29**, P919–P929 (2021).
42. Mossong, J. et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
43. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**, 1481–1486 (2020).
44. Mistry, D. et al. Inferring high-resolution human mixing patterns for disease modeling. *Nat. Commun.* **12**, 323 (2021).
45. Wallinga, J., Teunis, P. & Kretzschmar, M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am. J. Epidemiol.* **164**, 936–944 (2006).
46. van Kampen, J. J. A. et al. Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). *Nat. Commun.* **12**, 267 (2021).
47. Savela, E. S. et al. SARS-CoV-2 is detectable using sensitive RNA saliva testing days before viral load reaches detection range of low-sensitivity nasal swab tests. https://doi.org/10.1101/2021.04.02.21254771 (2021).
48. Smith, R. L. et al. Longitudinal assessment of diagnostic test performance over the course of acute SARS-CoV-2 infection. *J. Infect. Dis.* **224**, 976–982 (2021).
49. Huang, N. et al. SARS-CoV-2 infection of the oral cavity and saliva. *Nat. Med.* **27**, 892–903 (2021).
50. Lakdawala, S. S. et al. The soft palate is an important site of adaptation for transmissible influenza viruses. *Nature* **526**, 122–125 (2015).
51. Amato, K. A. et al. Influenza A virus undergoes compartmentalized replication in vivo dominated by stochastic bottlenecks. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2021.09.28.462198v2 (2021).
52. Cai, Y. et al. Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. *Science* **373**, 6555 (2021).
53. Thorne, L. G. et al. Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature* **602**, 487–495 (2022).
54. Brown, J. C. et al. Increased transmission of SARS-CoV-2 lineage B.1.1.7 (VOC 2020212/01) is not accounted for by a replicative advantage in primary airway cells or antibody escape. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/2021.02.24.432576v2 (2021).
55. Ulrich, L. et al. Enhanced fitness of SARS-CoV-2 variant of concern Alpha but not Beta. *Nature* **602**, 307–331 (2022).
56. Gniazdowski, V. et al. Repeated COVID-19 molecular testing: correlation of SARS-CoV-2 culture with molecular assays and cycle thresholds. *Clin. Infect. Dis.* **73**, e860–e869 (2021).
57. Matsuyama, S. et al. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proc. Natl Acad. Sci. USA* **117**, 7001–7003 (2020).
58. Waggoner, J. J. et al. Triplex real-time RT-PCR for severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**, 1633–1635 (2020).

59. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).

60. Han, M. S., Byun, J.-H., Cho, Y. & Rim, J. H. RT-PCR for SARS-CoV-2: quantitative versus qualitative. *Lancet Infect. Dis.* **21**, P165 (2021).

61. Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G. & Perelson, A. S. Kinetics of influenza A virus infection in humans. *J. Virol.* **80**, 7590–7599 (2006).

62. Gonçalves, A. et al. Timing of antiviral treatment initiation is critical to reduce SARS-CoV-2 viral load. *Clin. Pharmacol. Ther.* **9**, 509–514 (2020).

63. Ménache, M. G. et al. Upper respiratory tract surface areas and volumes of laboratory animals and humans: considerations for dosimetry models. *J. Toxicol. Environ. Health* **50**, 475–506 (1997).

64. Collins, L. M. C. & Dawes, C. The surface area of the adult human mouth and thickness of the salivary film covering the teeth and oral mucosa. *J. Dent. Res.* **66**, 1300–1302 (1987).

65. Hou, Y. J. et al. SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* **182**, 429–446 (2020).

66. Miao, H., Xia, X., Perelson, A. S. & Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev. Soc. Ind. Appl. Math.* **53**, 3–39 (2011).

67. Smith, A. P., Moquin, D. J., Bernhauerova, V. & Smith, A. M. Influenza virus infection model with density dependence supports biphasic viral decay. *Front. Microbiol.* **9**, 1554 (2018).

68. Martin, M. A. & Koelle, K. Comment on 'Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2'. *Sci. Transl. Med.* **13**, eabh1803 (2021).

69. Braun, K. M. et al. Acute SARS-CoV-2 infections harbor limited within-host diversity and transmit via tight transmission bottlenecks. *PLoS Pathog.* **17**, e1009849 (2021).

70. Valesano, A. L. et al. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. *PLoS Pathog.* **17**, e1009499 (2021).

71. Michael Lavigne, G., Russell, H., Sherry, B. & Ke, R. Autocrine and paracrine interferon signalling as 'ring vaccination' and 'contact tracing' strategies to suppress virus infection in a host. *Proc. R. Soc. Lond. B Biol. Sci.* **288**, 20203002 (2021).

72. Perelson, A. S. & Nelson, P. W. Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Rev. Soc. Ind. Appl. Math.* **41**, 3–44 (1999).

73. Perelson, A. S. & Ke, R. Mechanistic modeling of SARS-CoV-2 and other infectious diseases and the effects of therapeutics. *Clin. Pharmacol. Ther.* **109**, 829–840 (2021).

Author contributions

Conceptualization was provided by R.K., R.L.S., W.J.H., Y.C.M., A.P., L.L.G. and C.B.B. Data curation was performed by R.L.S., B.B. and P.L. Formal analysis was carried out by R.K., P.P.M. and R.L.S. Funding acquisition was the responsibility of D.D.M., L.L.G. and C.B.B. Investigation was carried out by R.K., P.P.M., R.L.S., A.M., M.C., N.G., C.H.L., J.J., A.C., T.L., M.F., K.K.O.W., C.J.F., L.W., R.F., M.E.B., K.K.C., H.C., K.R.S., A.N.O., J.B. and M.L.R. Methodology was performed by R.K., R.L.S. and C.B.B. Project administration was overseen by D.C.E., K.R.S., S.B., S.L.G., C.R., J.Y. and J.Q. Software was provided by R.K., P.P.M. and R.L.S. H.H.M., Y.C.M., A.P., L.L.G. and C.B.B. supervised the study. Visualization was undertaken by R.K., P.P.M., R.L.S. and C.B.B. Writing of the original draft was done by R.K. and C.B.B., with writing, review and editing by R.K., P.P.M., R.L.S., Y.C.M., A.P., L.L.G. and C.B.B.

Competing interests

C.B.B. and L.W. are listed as inventors on a patent application for the saliva RT–qPCR test used in this study. The remaining authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-022-01105-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-022-01105-z.

Correspondence and requests for materials should be addressed to Christopher B. Brooke.

Peer review information *Nature Microbiology* thanks David Eyre and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
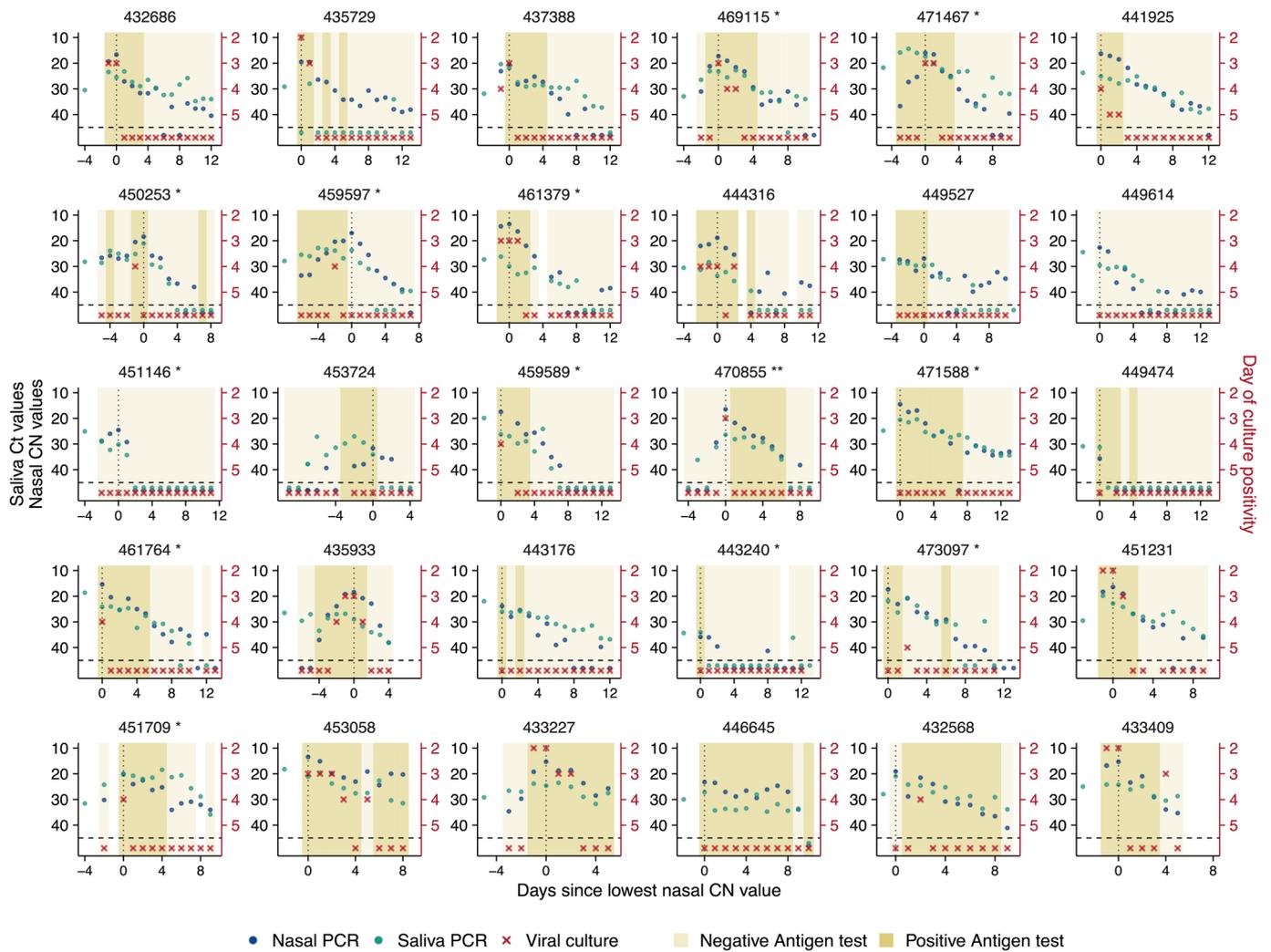
Extended Data Fig. 1 | Remainder of individual plots. Plots of longitudinal assay results from study participants not shown in Fig. 1a. Single asterisk next to the participant ID indicates B.1.1.7 variant infection, while double asterisks indicate P1 variant infection.

Extended Data Fig. 2 | Individual-level symptom data. Self-reported symptom data from study participants, overlaid with viral culture status. Participants were asked to complete a survey through the Eureka digital study platform inquiring about the presence or absence of the indicated set of symptoms each day after sample collection.
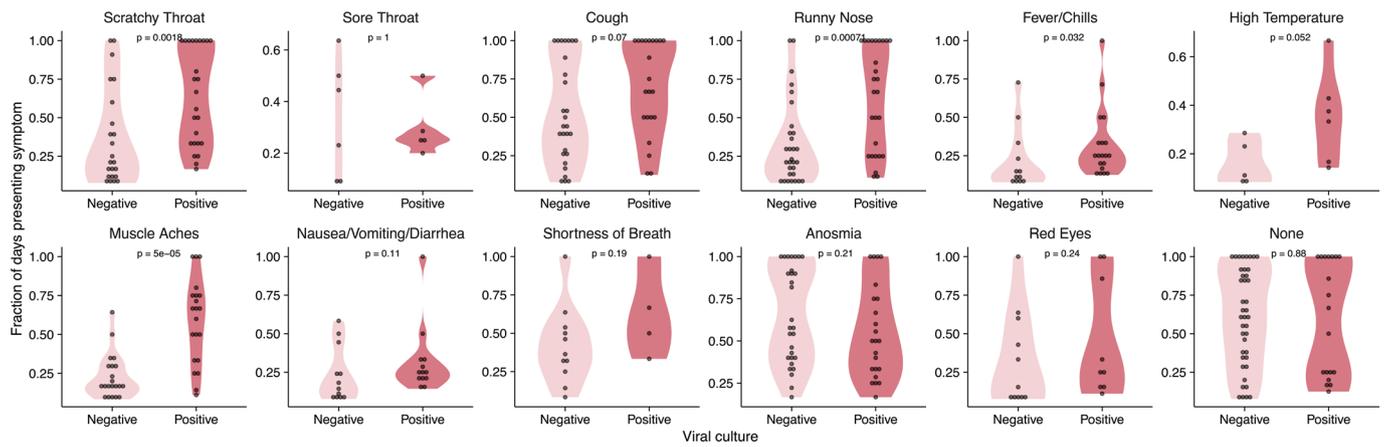
Extended Data Fig. 3 | Comparison of symptoms and viral culture status. Plots show the proportions of either viral culture negative or viral culture positive days for which participants reported the indicated symptoms. The p-values for the Wilcoxon rank-sum test are reported. Data are only shown for individuals who reported the indicated symptom at least once.
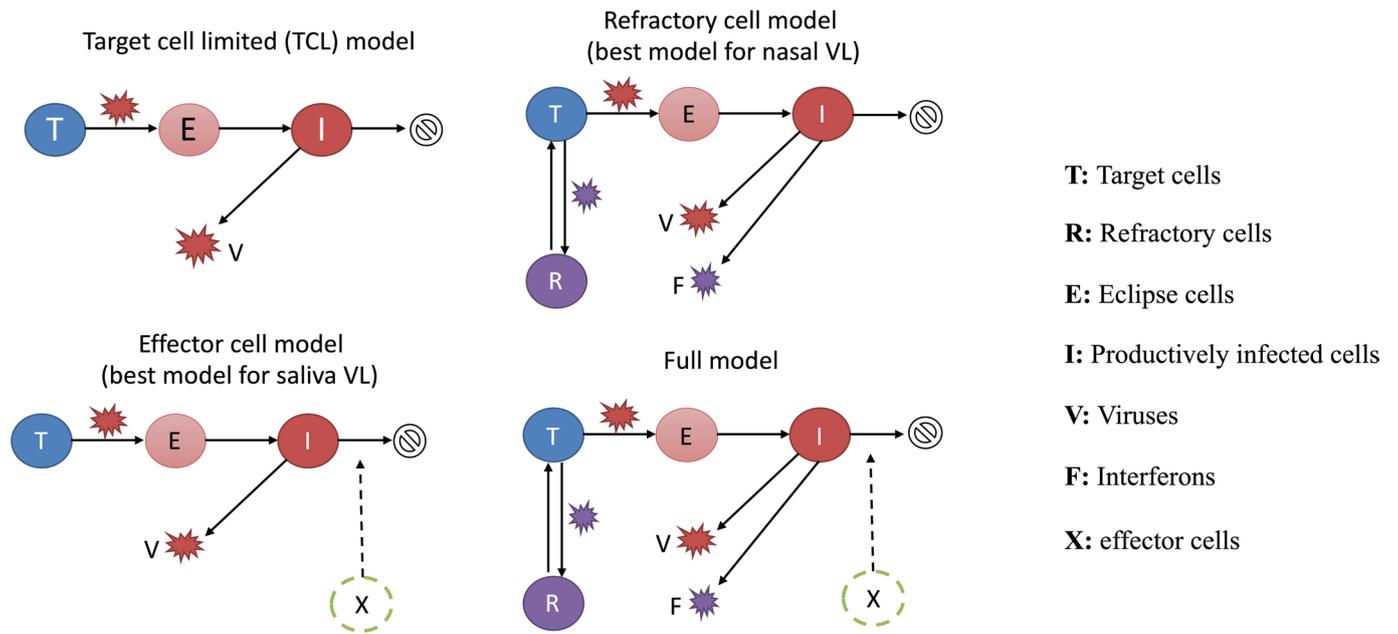
Extended Data Fig. 4 | Model structures. Diagrams showing the structures of the additional three models (not shown in Fig. 2a) considered for describing viral load data. See Supporting Text for descriptions of the models.
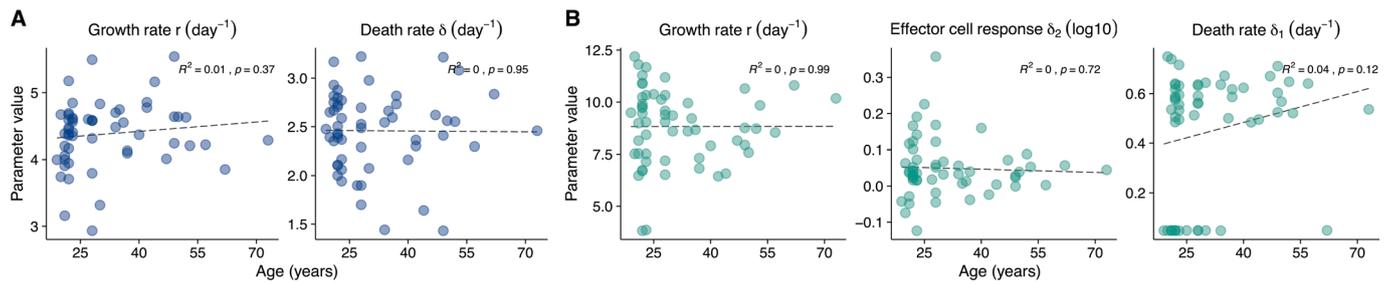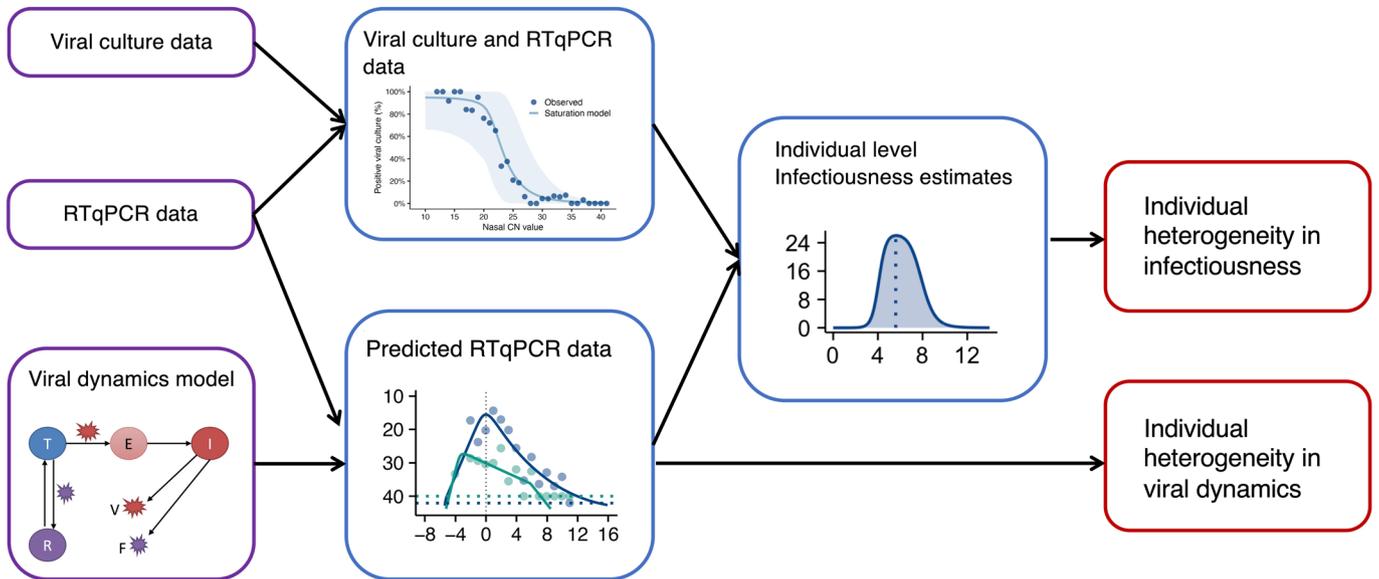
Extended Data Fig. 5 | Model parameter estimates as a function of age. Plots showing the relationship between age and the indicated model parameter estimates for **(A)** the refractory cell model (nasal data) and the **(B)** the immune effector cell model (saliva data). Linear regressions were performed on the data. R^2 values and p-values are shown.
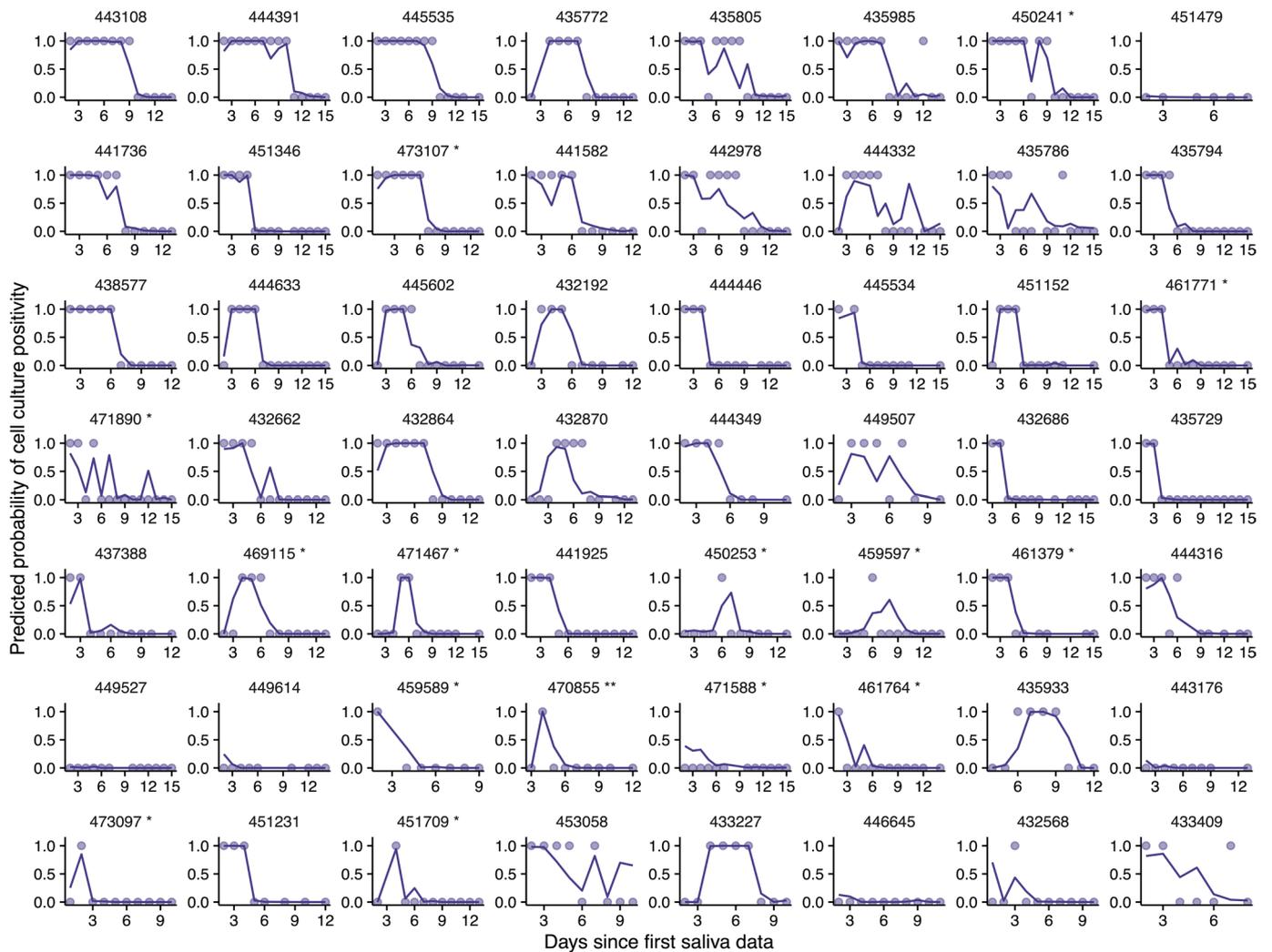
Extended Data Fig. 6 | Analysis workflow. Diagram indicating how empirical RTqPCR and viral culture data were used to generate estimations of individual level viral dynamics and infectiousness.

Extended Data Fig. 7 | The saturation model accurately predicts the cell culture positivity data. Lines denote the predicted probability of cell culture being positive. Dots denotes cell culture positivity data, where a dot is at 1 or 0 when the cell culture is positive or negative, respectively.
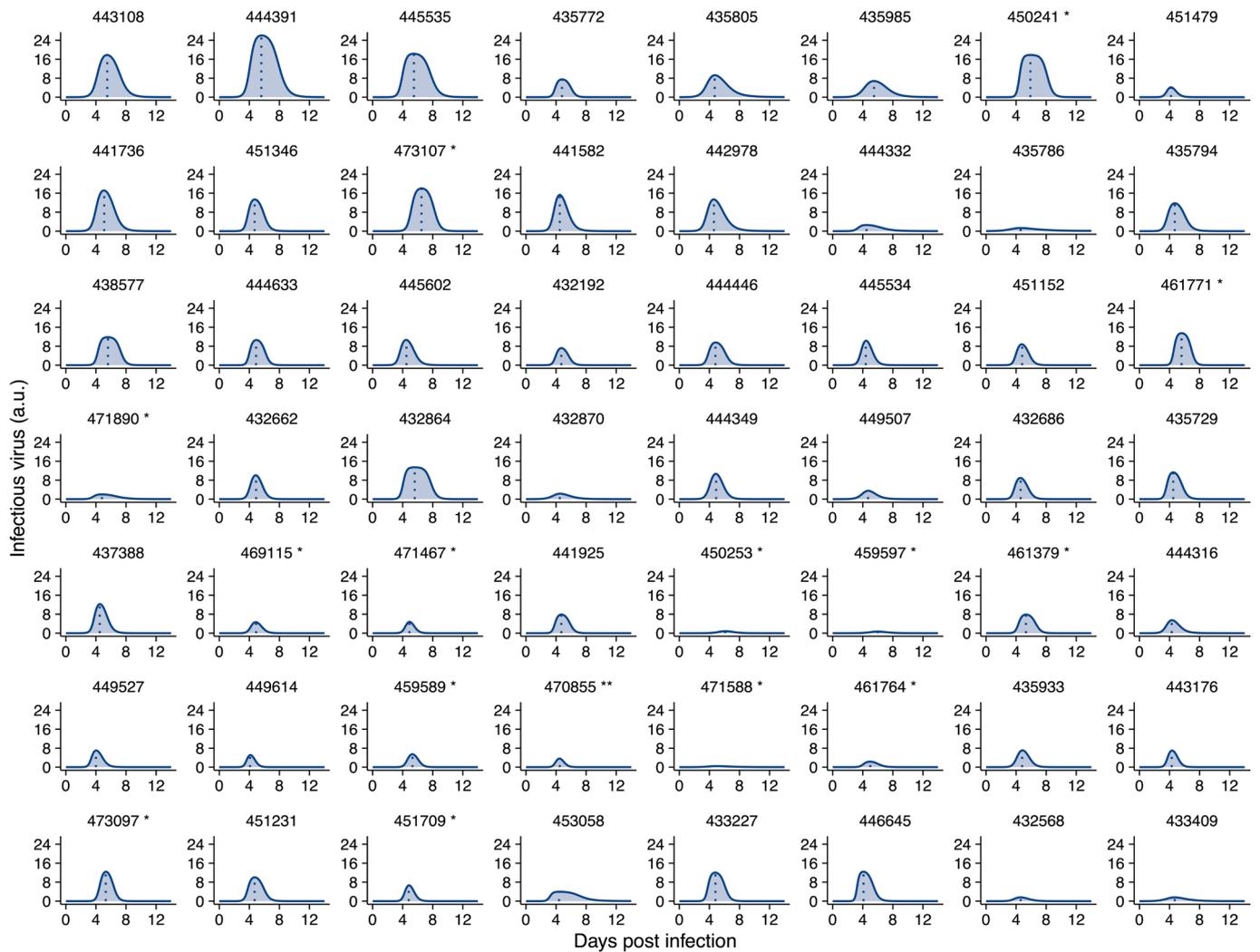
Extended Data Fig. 8 | Individual infectiousness plots. Estimated infectiousness over time plotted for individual study participants. Dashed lines indicate inferred peak in infectiousness.
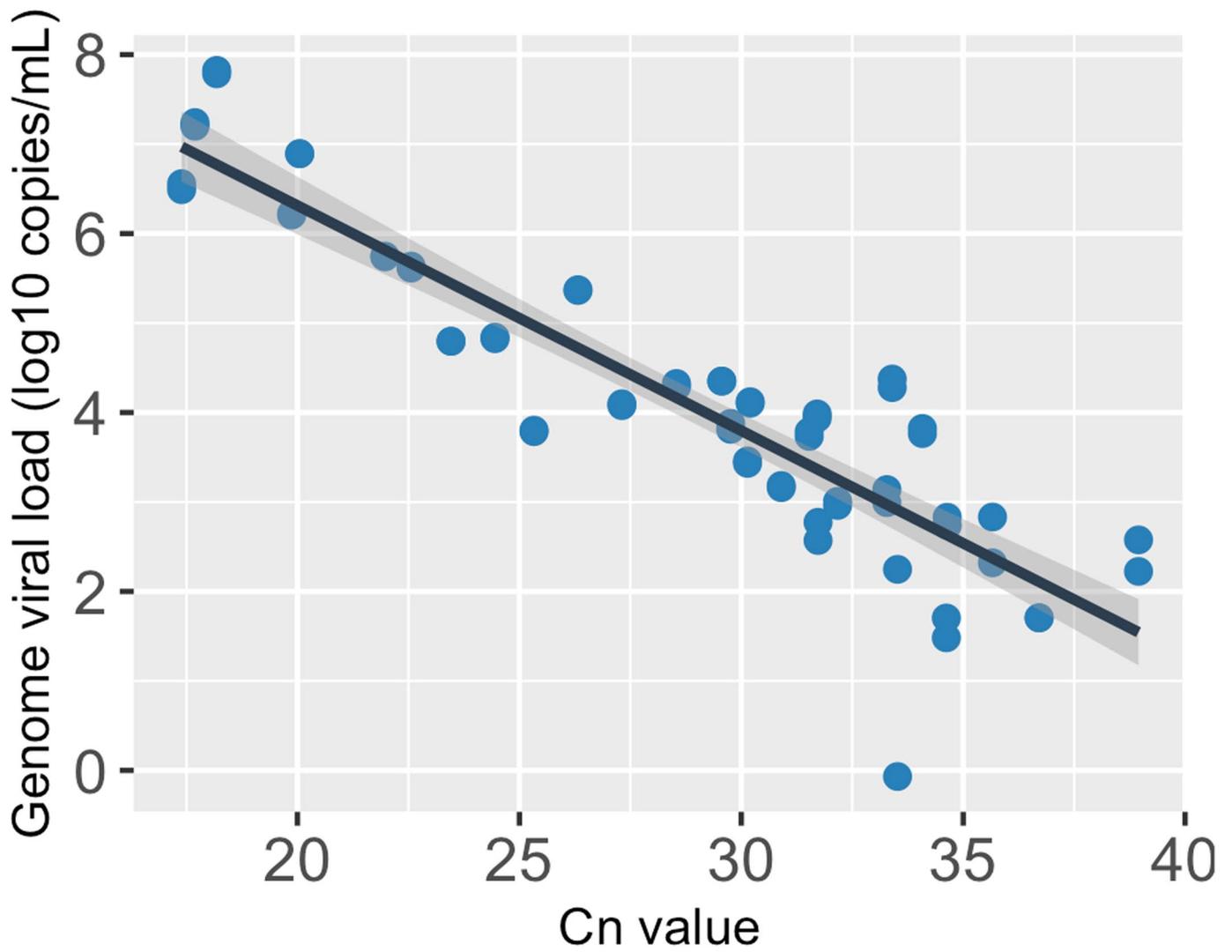
Extended Data Fig. 9 | The relationship between genome viral load (y-axis; on a log$_{10}$ scale) and CN value of the nasal samples. The black line, that is the center of the error band, represents the linear regression calibration curve. The shading around the black line shows the standard error for the regression.
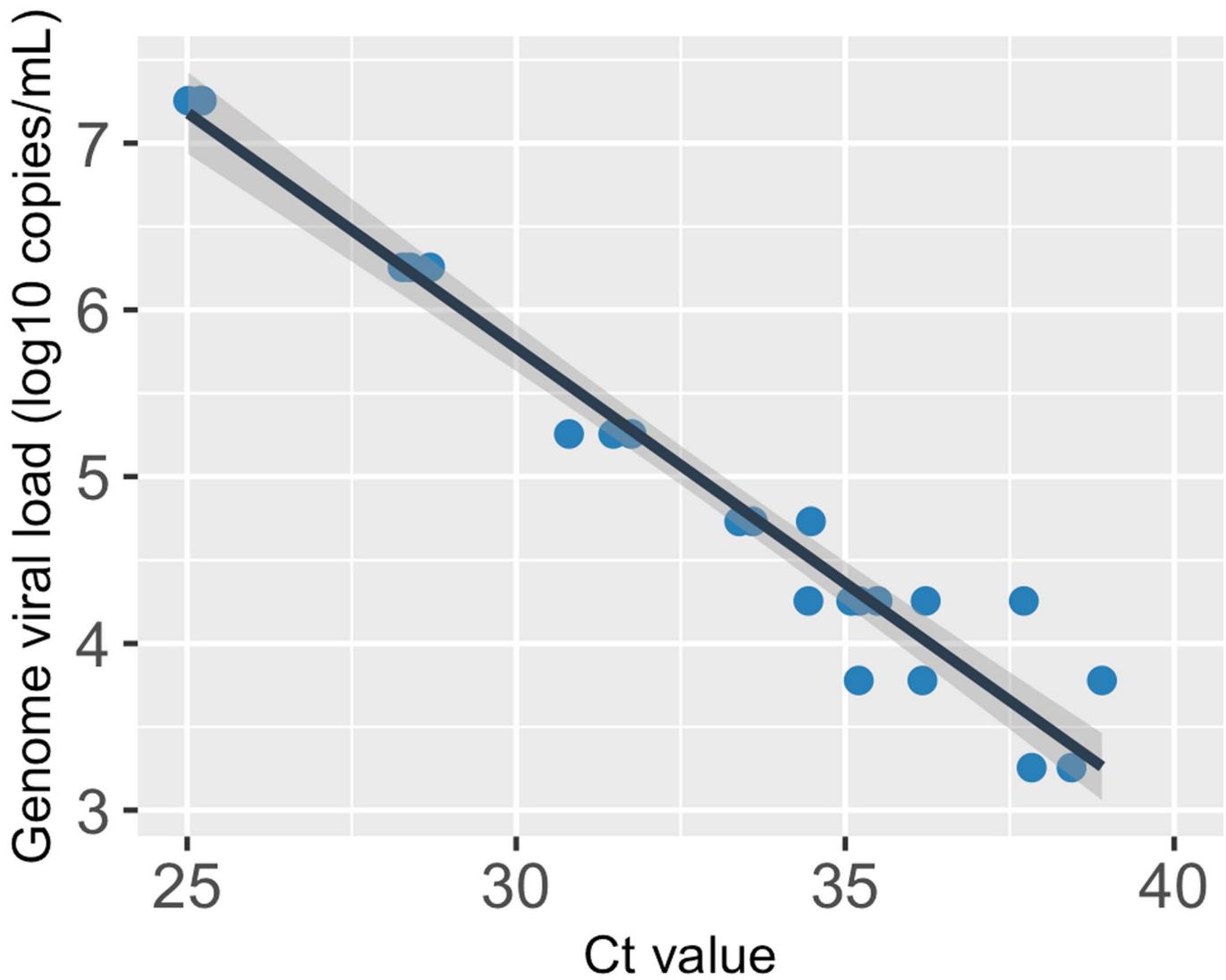
Extended Data Fig. 10 | The relationship between genome viral load (y-axis; on a log$_{10}$ scale) and Ct value of the saliva samples. The black line, that is the center of the error band, represents the linear regression calibration curve. The shading around the black line shows the standard error for the regression.

nature portfolio

Corresponding author(s): Christopher Brooke

Last updated by author(s): 03/03/2022

Reporting Summary

Nature Portfolio 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 Portfolio 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	All participants' data and survey responses were collected in the Eureka digital study platform.
Data analysis	All figures were generated using r studio. Viral dynamics models were generated using custom codes in the R programing language and environment. The codes cannot be shared due to institutional policies of Los Alamos National Laboratory. All simulation results of the models can be fully reproduced from the equations in the Supporting Text and the parameter values listed in Supplementary Tables. All model fitting estimations were performed using Monolix. Viral genome sequence data was analyzed using the nf-core/viralrecon workflow. Viral lineages were called using Pangolin version 2.4.2, pango version 1.2.6, and the 5/19/21 version of the pangoLEARN model.

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

> The computer codes for the mathematical analyses in this paper are available at https://github.com/BROOKELAB/Viral-dynamics-modeling and are available at doi: 10.5281/zenodo.6311388. All raw data used is included as a supplemental table. Raw sequence data files can be found under BioProject ID PRJNA809434.

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](#)

Life sciences study design

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

Sample size	Study cohort size was determined by overall study enrollment criterion. No sample size calculation for the analyses included in this manuscript was performed.
Data exclusions	For some analyses, data from a small number of participants were excluded due to limited sample numbers. In all cases this is indicated in the text.
Replication	Replication is not applicable to this study because samples were collected under the non-replicatable conditions of a global pandemic. All analyses can be replicated using the raw data and code provided.
Randomization	Randomization was not required due to the absence of clinical interventions.
Blinding	Investigators only saw a random participant ID number during the study, with no other identifying information.

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	Involved in the study
☒	Antibodies
☐	☒ Eukaryotic cell lines
☒	Palaeontology and archaeology
☒	Animals and other organisms
☐	☒ Human research participants
☒	Clinical data
☒	Dual use research of concern

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	Vero-TMPRSS2 cells: National Institute of Infectious Diseases, Japan
Authentication	Representative aliquots of the cell stock used in this study were evaluated by morphological observation and testing for SARS-CoV-2 susceptibility
Mycoplasma contamination	Absence of mycoplasma was tested using MycoSeq from AppliedBiosystems when the cell line aliquots were initially frozen down.

Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Human research participants

Policy information about studies involving human research participants

Population characteristics	Participants were required to be at least 18 years of age, have a valid University of Illinois ID, speak English, have internet access, and live within 8 miles of the university campus. Participants were required to be within 24 hours of a positive on-campus COVID test or within 5 days of exposure to a COVID+ individual. Participants must have a documented on-campus negative COVID result within the previous 7 days. All demographic information on the study cohort, including covariate-relevant characteristics such as age and health status are detailed in supplemental table 1. Where relevant, the potential of age to act as a co-variate was accounted for in the analyses, as detailed in the methods section
Recruitment	Individuals were recruited via either a link shared in an automated text message providing isolation information sent within 30 minutes of a positive test result, a call from a study recruiter, or a link shared by an enrolled study participant or included in information provided to all quarantining close contacts. In addition, signs were used at each testing location and a website was available to inform the community about the study.
Ethics oversight	This study was approved by the Western Institutional Review Board, and all participants provided informed consent.

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