eAppendix

SARS-CoV-2 outbreak dynamics in an isolated US military recruit training center with rigorous prevention measures

Supplementary Methods

Cohort Enrollment and Subject Sampling in context of USMC Training (includes methods for questionnaire and symptom ascertainment)

The observation period for this prospective cohort study began when Marine recruits arrived at Marine Corps Recruit Depot – Parris Island (MCRDPI) to commence basic training. Prior to transferring to MCRDPI, the United States Marine Corps (USMC) implemented two separate quarantine protocols. The first was a two-week home quarantine. After that, the recruits traveled, while masked and socially distanced, to a second USMC-supervised two-week quarantine situated at either a college campus between May-July 2020, or at a hotel between August-October 2020.

Within 48 hours of arriving at the supervised quarantine location, recruits were offered the opportunity to volunteer for CHARM. Recruits were eligible if they were ≥18 years of age. Institutional Review Board approval was obtained from the Naval Medical Research Center in compliance with all applicable U.S. federal regulations governing the protection of human subjects. All participants provided written informed consent.

The supervised quarantine employed extensive public health measures that were strictly enforced by US Marine instructors at all times. Recruits and staff were forbidden to leave, and no visitors other than deliveries of supplies and food along with local essential workers and the study staff were allowed onto the premises. At the end of this quarantine period, the USMC required all recruits to test negative for SARS-CoV-2 by qPCR before proceeding to MCRDPI to initiate basic training.

At enrollment, participants completed a questionnaire consisting of demographic information, risk-factors, reporting of 14 specific COVID-19 related symptoms (subjective fever, chills, muscle aches, fatigue, runny nose, sore throat, cough, shortness of breath, nausea or vomiting, headache, decreased taste or smell, abdominal pain, diarrhea, other) or any other unspecified symptom, and brief medical-history. At quarantine weeks 0, 1 and 2, mid-turbinate nasal swab specimens were obtained for SARS-CoV-2 qPCR testing and questionnaires were administered. The follow-up questionnaire inquired about the same COVID-19 related symptoms since the last study visit.

Recruits were assigned a company upon entry to training at MCRDPI; barring extenuating circumstances such as illness or training failure, recruits complete the 13-week training with their company consisting of 400-500 recruits. Although only one male and potentially one female company simultaneously complete the activities of a particular training week, there are multiple companies at MCRDPI that overlap in different phases of training. Recruits participating in CHARM came from 20 different companies (13 male and 7 female) over the study period graduating every 13 weeks. Companies 1, 4, 5, 7, 8, 10, 11, 13, 14, and 16 through 19 were male, and companies 2, 3, 6, 9, 12, 15, and 20 were female. The percent enrolled ranged from 26% to 97% within individual companies with a mean enrollment of 64.4% of recruits throughout the study. Company sizes vary over time with female companies in general being smaller than male companies.

In response to the COVID-19 pandemic and prior to the beginning of the CHARM study, military public health officials instituted non-pharmaceutical preventive measures at MCRDPI, including masking of recruits and staff except during long runs, increased spacing during formations, head-to-toe sleeping arrangements, increased hand hygiene and surface cleaning, controlled movement of recruits, and reduced company size. Additionally, travel to and from the base was limited, visitors were no longer allowed on base, and base amenities were closed or had limited access. While most recruits slept in two-bedded rooms during the supervised quarantine, during basic training at MCRDPI companies slept in a large barracks.

After arrival at MCRDPI, Study participants were sampled at approximately 14 days, 28 days, and 42 days (+/- 3 days based on accommodations for training exercises). When clinically indicated due to the development of symptoms, participants were evaluated at the MCRDPI clinic and diagnosed by rapid testing. If positive, they went to the isolation barracks, where the study team was able to follow up and repeat testing outside of the scheduled longitudinal follow up encounters.

Nine de-identified MCRDPI staff specimens collected through routine surveillance and sequenced at the NMRC as nonhuman subject research were also included in this analysis for comparison. Long-term passive surveillance data are not specific to CHARM. The long-term passive surveillance data are from the preceding three years of passive surveillance data for all recruits at MCRDPI, as pulled from the HL7 data.

Incidence of Non-SARS-CoV-2 Infections among CHARM Subjects

Military Public Health officials collect data regarding communicable disease that commonly affect recruits as part of their regular Disease and Injury Surveillance (1). Surveillance data compiled by the Navy and Marine Corps Public Health Center utilize the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) encoded medical encounter data from January 1, 2017 to December 31, 2020 to obtain incident cases of non-

SARS-CoV-2 Acute Respiratory Infections (ARI) and pneumonia. They report incident cases per 1,000 recruits per week.

SARS-CoV-2 qPCR Testing

The qPCR testing of mid-turbinate nasal swab specimens for SARS-CoV-2 was performed within 48 hours of sample collection by Lab24 (Boca Raton, FL) and the Naval Infectious Diseases Diagnostic Laboratory (Naval Medical Research Center, Silver Spring, MD). Swab specimens in viral transport media were kept at 4°C. Assays were carried out at high complexity Clinical Laboratory Improvement Amendments-certified laboratories using the US Food and Drug Administration-authorised Thermo Fisher TaqPath COVID-19 Combo Kit (Thermo Fisher Scientific, Waltham, MA, USA).

Sequencing

RNA was extracted from 0.25 mL of VTM using 0.75 mL of TRIzol LS reagent (Invitrogen) according to manufacturer's protocol. RNA concentration was measured using Qubit RNA High Sensitivity assay (ThermoFisher Scientific) prior to use in the YouSeq SARS-CoV-2 Coronavirus NGS Library prep kit (YouSeq). Approximately 100 ng of RNA was reverse-transcribed as in the protocol except one modification where the YouSeq reverse transcriptase was replaced with SuperScript IV (ThermoFisher Scientific). cDNA was amplified using multiplex qPCR and samples were cleaned using 1x AMPure XP beads (Beckman Coulter) and re-suspended in nuclease-free molecular grade water. The samples were then processed following the QiaSeq FX DNA library protocol (Qiagen). Completed libraries were quality-checked using an Agilent Bioanalyzer High sensitivity kit (Agilent) and quantitated using the

Qubit DNA High Sensitivity assay (ThermoFisher Scientific) prior to sequencing using MiSeq v3 2x300 chemistry (Illumina).

SARS-CoV-2 genome amplification and sequencing performed at Mount Sinai were done with custom primers using Nextera XT and MiSeq 2x150 chemistry (Illumina).

Bioinformatic Analyses

Consensus genomes were obtained using steps described in BDRD Genomics Viral Amplicon Illumina Workflow on Docker Hub (2). Briefly, the Illumina MiSeq raw reads were processed using Viral Amplicon Illumina Workflow. Reads were trimmed using BBDuk (Q20) and the resulting paired reads were merged using BBMerge and aligned to the Wuhan reference genome (NC_045512.2) using BBMap (3, 4). YouSeq primer sequences were trimmed from sequence ends using align_trim (ARTIC pipeline). Consensus genomes were generated and Single Nucleotide Variants (SNVs) were determined using SAMtools mpileup (3, 5) and iVar (Intrahost variant analysis of replicates) (6). The primer trimmed alignment was visualized and the final genome and SNVs were verified using CLC Genomics Workbench (2020.0.3). Genome assembly for sequencing performed at Mount Sinai was done with a custom reference-based assembly pipeline (https://github.com/mjsull/COVID_pipe), as previously reported (7).

PopART version 1.7 (http://popart.otago.ac.nz) was used to build the median joining haplotype networks (8, 9). For building these networks, an alignment was constructed of all positions with variants relative to the reference, with insertions and deletions of any length represented by a single base in this alignment. Singlet isolates with unique mutational profiles but within four variants of another isolate were merged into the nearest multiplet node for visualization purposes.

The phylogenetic relationships of the SARS-CoV-2 isolates from study participants were analyzed in a South Carolina-focused background GISAID (2021-05-04 download).

The time-calibrated tree was built using Nextstrain v1(10) for SARS-CoV-2 (https://github.com/nextstrain/ncov) with default parameters, using a division-focus subsampling scheme and maximum sampling date of 2021-11-30. The final tree contained a total of 5886 sequences. A complete list of included sequences and authors is provided as an Appendix.

Cell culture

The Vero E6 cell line (ATCC #CRL-1586) and Normal human bronchial epithelial (NHBE) cells (Lonza CC-2540) were both used in SARS-CoV-2 infections isolated from study participants. Vero E6 cells were maintained in Dulbecco's Modified Eagles Medium (DMEM, Gibco), which was supplemented with 10% fetal bovine serum (FBS, Corning) and penicillin/streptomycin (Gibco). Primary NHBE cells from a female donor were purchased from Lonza and allowed to differentiate in the air-liquid interface (ALI) on collagen-coated porous transwell inserts for 4-6 week differentiation process (following the manufacturer's recommendations). Both Vero E6 and NHBE cells were cultured in a humidified 37 °C incubator in an atmosphere of 5% CO2.

Viral isolation and propagation for in vitro assays

Specimen samples were selected from individuals confirmed with infection of variants within the three subclusters (1A, 1B, and 1C) of Cluster 1. To assess possible changes in viral replication based on the presence of mutation of interest (S3883A, found in subclusters 1B and 1C), viral isolates were cultured from nasal swab specimens stored in viral transport media (VTM). A 50 µl aliquot of selected VTM samples was mixed with 50 µl of 2XMEM containing 2x antibiotics/antimycotics (penicillin/streptomycin and amphotericin B) medium and then serially diluted to inoculate Vero E6 cells seeded in 96-well plates at a concentration of 25,000 cells/well. Inoculated cultures grew in a humidified 37 °C incubator in an atmosphere of 5% CO2, and cytopathic effects (CPE) were observed each day post inoculation. Standard plaques assays for SARS-CoV2 were used to titer viral stocks (11). Viral stocks were sequenced to confirm the mutation of interest before use in subsequent in-vitro assays. All experiments involving SARS-CoV-2 infections were performed in a biosafety level 3 (BSL3) facility.

qPCR

During a 72-h time course experiment, cells were harvested every 24 hours and lysed to extract the RNA using the RNAdvance Cell v2 (Beckmann). RNA samples were then reverse transcribed to prepare cDNA, which was used in qPCR assays quantifying the SARS-CoV2 E gene and the ribosomal 18S gene as a control. The qPCR protocol was as previously described in (12).

Transmission Dynamics and estimation of R_0

Epidemiological modeling

In order to model point prevalence data across four two-week periods (i.e. at day 14, 28, 42, and 56 after the start of the 2-week supervised quarantine), we extended the standard Susceptible-Exposed-Infectious-Recovered (SEIR) model to include an additional post-infection compartment (P compartment) before recovery, during which a person has detectible viral load but can no longer transmit. Then, the proportion of population in each compartment can be described as follows:

$dS/dt = -\beta SI$

$dE/dt = \beta SI - \sigma E$

$dI/dt = \sigma E - \gamma I$

$dP/dt = \gamma I - \nu P$

$dR/dt = \nu P$

where β represents the transmission rate, $1/\sigma$ represents the mean latent period, $1/\gamma$ represents the mean infectious period, and $1/\nu$ represents the mean post-infection period. For this model, the basic reproduction number is given by $R_0 = \beta/\gamma$. A similar model was used to infer the spread of SARS-CoV-2 from point prevalence data earlier (13).

The time series of point prevalence data typically begins with one or two zeroes, but the deterministic model, which initially predicts exponential growth, would have to start at an unrealistically low prevalence in order to match multiple zeroes. Instead, we only consider the last zero before the first non-zero observation. We can then model the observed number of positive cases at time t (C(t)) using a binomial likelihood:

$C(t) \sim \text{Binomial}(T(t), I(t) + P(t))$

where T(t) represents the total number of qPCR tests performed at time t. Simulations are run from one day before the first observation in order to incorporate all data points that we considered into the likelihood. The initial conditions are assumed to be $S(t_0) = 1 - I_0$, $E(t_0) = I_0/3$, $I(t_0) = I_0/3$, $P(t_0) = I_0/3$, and $R(t_0) = 0$ with $t_0 = 13 \; or \; 27$, depending on the initial number of zero observations as explained earlier. We assume that all infected compartments E, I, and P are at a low prevalence initially in order to avoid numerical issues during the initial integration step. By allowing the initial conditions to vary, we implicitly account for the variation in the timing of the pathogen introduction. For example, a company with earlier will have a higher proportion of infected individuals at t_0.

Finally, we impose weakly informative priors in order to reflect our prior knowledge on the disease progression of SARS-CoV-2 infection and to constrain the parameter space:

$R_0 \sim \text{Gamma}(3, 1)$

$1/\sigma \sim \text{Gamma}(4, 2)$

$1/\gamma \sim \text{Gamma}(4, 1)$

$1/\nu \sim \text{Gamma}(4, 2)$

$I_0 \sim \text{Beta}(1, 399)$

These assumptions correspond to following prior mean and 95% quantiles: $R_0 = 3$ (0.62– 7.89), $1/\sigma = 2$ days (0.54–4.74 days), $1/\gamma = 4$ days (1.09–9.49 days), $1/\nu = 2$ days (0.54–4.74 days), and $I_0 = 2.50 \times 10^{-3}$ (6.35×10^{-5}–9.20×10^{-3}). Parameters are estimated using Hamiltonian Monte Carlo in Stan (14). We ran 4 independent chains with 1000 iterations after 1000 warm up iterations. Convergence is assessed via a lack of warning messages from Stan, indicating sufficiently low Gelman-Rubin statistics (“R-hat”), sufficiently high effective sample sizes, and no divergent chains.

Testing sources of variability in reproduction number estimates

In order to test whether variability in the inferred R_0 across companies can be explained by stochasticity alone, we generated synthetic data using a stochastic model that accounts for superspreading events and fitted the same deterministic model we used previously under the same procedure. To do so, we first modeled the number, instead of proportions, of individuals in each compartment using a stochastic model, discretized at a time step of one day (Δt = 1 day) using a binomial Euler scheme (15):

$$\bar{\imath}(t) = S(t - \Delta t)(1 - \exp(-\beta I(t - \Delta t)\Delta t/N)\ )$$

$$i(t) \sim NegativeBinomial(\bar{\imath}(t), k)$$

$$\Delta N_{S\to E}(t) = \begin{cases} i(t) & \text{if } i(t) \leq S(t - \Delta t) \\ S(t - \Delta t) & otherwise \end{cases}$$

$$\Delta N_{E\to I}(t) \sim Binomial(E(t - \Delta t), 1 - \exp(-\sigma\Delta t))$$

$$\Delta N_{I\to P}(t) \sim Binomial(I(t - \Delta t), 1 - \exp(-\gamma\Delta t))$$

$$\Delta N_{P\to R}(t) \sim Binomial(P(t - \Delta t), 1 - \exp(-\nu\Delta t))$$

$$S(t) = S(t - \Delta t) - \Delta N_{S\to E}(t)$$

$$E(t) = E(t - \Delta t) - \Delta N_{E\to I}(t) + \Delta N_{S\to E}(t)$$

$$I(t) = I(t - \Delta t) - \Delta N_{I\to P}(t) + \Delta N_{E\to I}(t)$$

$$P(t) = P(t - \Delta t) - \Delta N_{P\to R}(t) + \Delta N_{I\to P}(t)$$

$$R(t) = R(t - \Delta t) + \Delta N_{P\to R}(t)$$

where N represents the population size, and the Negative Binomial distribution is characterized by the mean parameter and the over-dispersion parameter k (capturing the degree of over-dispersion).

Model parameters were assumed to be equal to the median of the mean estimates across companies from the previous analysis: R_0 = 5.51, 1/σ = 1.65 days, 1/γ = 7.26 days, and 1/ν = 3.59 days. We assumed k = 0.1 to account for superspreading events of SARS-CoV-2 (16). Neglecting under-sampling, we assumed C(t) = I(t)+P(t) and T(t) = N and considered their values across four biweekly periods to generate synthetic data (days 14, 28, 42, and 56); including observation error would generate even greater degrees of uncertainty as well as variability in R_0

estimates. We simulated 20 outbreaks in a population of 400 with a single exposed individual introduced on day 15; accounting for high over-dispersion (k = 0.1) caused many simulations to fade out before taking off, so we excluded simulations where no positive cases were detected over the four biweekly periods. For each of the 20 synthetic datasets, we fitted the deterministic model based on the same procedure that we used to analyze real outbreak data. We then compared the mean and variance of R_0 estimates across 20 synthetic datasets.

Quantifying relationships between the initial conditions and the frequency of fade-out events

Running stochastic simulations with one initially exposed individual resulted in frequent fade-out events. In order to understand the impact of the initial conditions on the persistence of the epidemic, we simulated the stochastic model while varying the initial number of exposed individuals between 1 and 10. For a given value of the initial number of exposed individuals, we simulated the model 10,000 times (using the same simulation conditions that we used to generate synthetic data) and calculated the proportion of simulations that results in which no positive cases were observed across four biweekly periods (days 14, 28, 42, and 56).

Matching distributions of final sizes

To further assess whether stochasticity alone can explain the variability observed in the data, we tried to match distributions of the predicted final sizes (i.e., cumulative proportion of infections) using stochastic simulations. First, we predicted the cumulative proportion of infections for each company on days 28, 42, and 56 from the deterministic model fits across each posterior distribution and calculated the median of the cumulative proportions for each company. Then, for a given value of R_0 and k, we simulated the stochastic model 2,000 times while holding all other parameters constant as before; the initial number of exposed individuals were drawn from a Poisson distribution with a mean of 4 in order to account for random initial seeding events and prevent frequent fade-out events. We compared the distributions of cumulative proportions of infections at days 28, 42, and 56 predicted from stochastic simulations with those predicted from fitted deterministic models using two sample Kolmogorov-Smirnov statistic, which measures a distance between two probability distributions (17). We then found the combination of R_0 and k that minimizes the sum of Kolmogorov-Smirnov statistics across three periods. We quantified parameter uncertainty region by considering parameter combinations whose sum of Kolmogorov-Smirnov statistics is within a 20% error of the minimum value. We did not fit stochastic models to individual outbreak data directly due to sparsity of the data. Kolmogorov-Smirnov statistic was calculated using the ks.test() function in R (18).

Quantifying the impact of frequent testing in prevention onward transmission

Finally, we tested whether frequent testing and isolation alone can sufficiently reduce onward transmission. In particular, we ask what proportion of transmission between infector-infectee pairs can be prevented by testing the infector. Note that testing the infectee has no impact on this particular chain of transmission although it will prevent onward transmission from the infectee to their own infectees.

To do so, we first considered 50,000 pairs of infector-infectee pairs and sampled a generation interval Gi (i.e., the time between when the infector becomes infected and when the infectee becomes infected) for each pair i = 1, . . . , 50, 000. Generation intervals G_i are modeled as the sum of latent period L_i—during which infected individuals will not test positive or transmit infection—and transmission interval X_i, which we define as time between onset of infectiousness and transmission. Latent periods are drawn from a gamma distribution with a mean of 2 days and a squared coefficient of variation of 1/2. Transmission intervals are drawn from a gamma distribution with a mean of 3, 4, 5, or 6 days, corresponding to the mean generation interval of 5, 6, 7, and 8 days, respectively; the squared coefficient of variation in the transmission-interval distribution is set so that the resulting generation-interval distribution has the squared coefficient of variation of 1/5 (19).

Given frequency of testing f (ranging from 1–7 days), we can determine when each infector will be tested before they transmit to their infectee. During this step, the time between infection and their first test after infection is sampled uniformly between 0 and f. Then, given sensitivity of a qPCR test (ranging from 0.5 to 0.95), we can determine when infectors will test positive; for simplicity, we assumed specificity of 1. Infectors are then isolated after a fixed time of positive-to-isolation delay. If the time of isolation occurs before the transmission, we are able to prevent the transmission. For given values of qPCR sensitivity, testing frequency, length of positive-to-isolation delay, and mean generation interval, we calculate the proportion p of transmission we prevent across 50,000 infector-infectee pairs. These estimates then correspond to reduction in reproduction number, and therefore, we can prevent an outbreak if $p > 1 - 1/R_0$.

eTables

Variable	Infected group (n=1107)*	Non-infected group (n=1362)†	Total (n=2469)
Mean age, years	19.0 (1.8)	19.2 (1.9)	19.1 (1.9)
Age group			
[18,20]	955 (86.3%)	1153 (84.7%)	2108 (85.4%)
[21,31]	152 (13.7%)	209 (15.3%)	361 (14.6%)
Sex			
Female	117 (10.6%)	102 (7.5%)	219 (8.9%)
Male	990 (89.4%)	1260 (92.5%)	2250 (91.1%)
Race			
Non-Hispanic White	393 (35.5%)	433 (31.8%)	826 (33.5%)
Non-Hispanic Black	68 (6.1%)	115 (8.4%)	183 (7.4%)
Non-Hispanic Other	111 (10.0%)	128 (9.4%)	239 (9.7%)
Hispanic	535 (48.3%)	686 (50.4%)	1221 (49.5%)

eTable 1. Participant demographics and SARS-CoV-2 positivity.

Data are mean (SD) or n (%). Table includes all 2469 participants tested during their first 6 weeks of basic training at MCRDPI between May 25, 2020 and Nov 5, 2020.

*: SARS-CoV-2 infection was determined using biweekly qPCR tests.

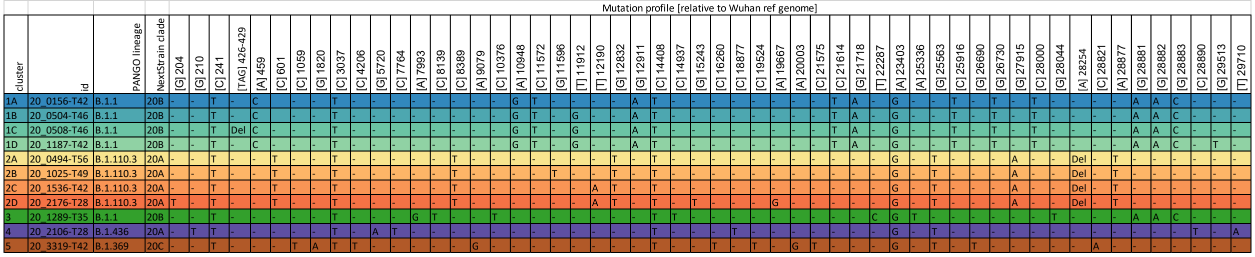
†: Includes the 5 participants who had no conclusive qPCR test results.

infection	baseline			year 2020			The change of difference
	before	after	difference	before	after	difference	
ARI	19.6 (3.6)	19.4 (4.0)	-0.13 (-2.66 to 2.40, p=0.918)	38.3 (6.2)	9.3 (10.0)	-29.0 (-34.9 to -23.1, p<0.001)	-28.9 (-35.3 to -22.5, p<0.001)
PNA	8.0 (1.5)	7.0 (1.5)	-0.99 (-1.95 to -0.03, p=0.043)	12.1 (3.5)	1.7 (3.4)	-10.4 (-12.7 to -8.2, p<0.001)	-9.5 (-11.8 to -7.1, p<0.001)

eTable 2. Changes in other infections during implementation of SARS-CoV-2 mitigation measures.

The incidence (cases per 1000 per week) before week 13 of the year and after week 13 (when the final mitigation measures were introduced in 2020) were compared in the previous three years combined (baseline) and in 2020 (year 2020) using two-subject t-test. These differences in 2020 and in the three earlier years combined were compared by ANOVA. ARI, acute respiratory infection. PNA, pneumonia.

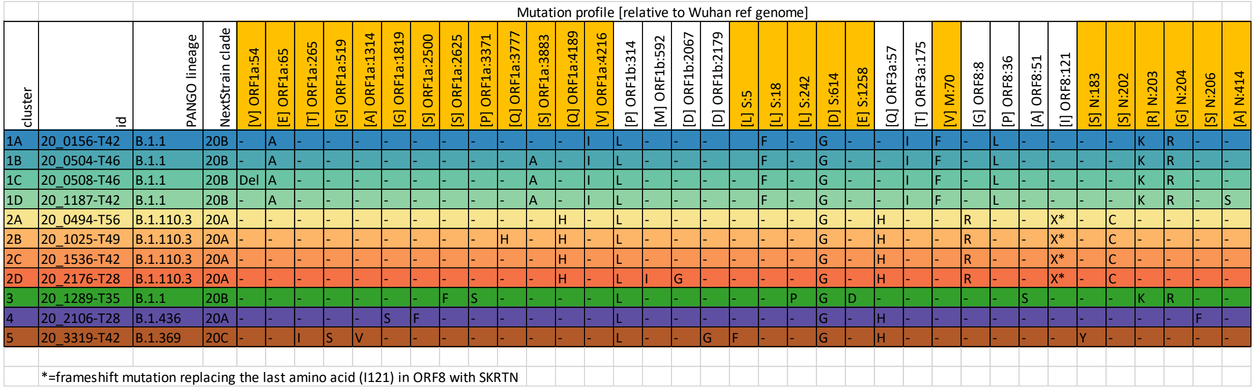
A.



Mutation profile [relative to Wuhan ref genome]

cluster	id	PANGO lineage	NextStrain clade	[G] 204	[G] 210	[C] 241	[TAG] 426-429	[A] 459	[C] 601	[C] 1059	[G] 1820	[C] 3037	[C] 4206	[G] 5720	[C] 7764	[A] 7993	[C] 8139	[C] 8389	[A] 9079	[C] 10376	[A] 10948	[C] 11572	[G] 11596	[T] 11912	[T] 12190	[G] 12832	[G] 12911	[C] 14408	[C] 14937	[G] 15243	[C] 16260	[C] 18877	[C] 19524	[A] 19667	[A] 20003	[C] 21575	[C] 21614	[G] 21718	[T] 22287	[A] 23403	[A] 25336	[G] 25563	[C] 25916	[G] 26690	[G] 26730	[G] 27915	[C] 28000	[G] 28044	[A] 28254	[C] 28821	[A] 28877	[G] 28881	[G] 28882	[G] 28883	[C] 28890	[G] 29513	[T] 29710
1A	20_0156-T42	B.1.1	20B	-	-	T	-	C		-	-	T	-	-	-	-	-	-	-	-	G	T	-	-	-	-	A	T	-	-	-	-	-	-	-	-	T	A	-	G	-	-	T	-	T	-	T	-	-	-	-	A	A	C	-	-	-
1B	20_0504-T46	B.1.1	20B	-	-	T	-	C		-	-	T	-	-	-	-	-	-	-	-	G	T	-	G	-	-	A	T	-	-	-	-	-	-	-	-	T	A	-	G	-	-	T	-	T	-	T	-	-	-	-	A	A	C	-	-	-
1C	20_0508-T46	B.1.1	20B	-	-	T	Del	C	-	-	-	T	-	-	-	-	-	-	-	-	G	T	-	G	-	-	A	T	-	-	-	-	-	-	-	-	T	A	-	G	-	-	T	-	T	-	T	-	-	-	-	A	A	C	-	-	-
1D	20_1187-T42	B.1.1	20B	-	-	T	-	C	-	-	-	T	-	-	-	-	-	-	-	-	G	T	-	G	-	-	A	T	-	-	-	-	-	-	-	-	T	A	-	G	-	-	T	-	T	-	T	-	-	-	-	A	A	C	-	T	-
2A	20_0494-T56	B.1.110.3	20A	-	-	T	-	-	T	-	-	T	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	T	-	-	-	-	-	-	-	-	-	-	-	G	-	T	-	-	-	A	-	-	Del	-	T	-	-	-	-	-	-
2B	20_1025-T49	B.1.110.3	20A	-	-	T	-	-	T	-	-	T	-	-	-	-	-	T	-	-	-	-	T	-	-	T	-	T	-	-	-	-	-	-	-	-	-	-	-	G	-	T	-	-	-	A	-	-	Del	-	T	-	-	-	-	-	-
2C	20_1536-T42	B.1.110.3	20A	-	-	T	-	-	T	-	-	T	-	-	-	-	-	T	-	-	-	-	-	-	A	T	-	T	-	-	-	-	-	-	-	-	-	-	-	G	-	T	-	-	-	A	-	-	Del	-	T	-	-	-	-	-	-
2D	20_2176-T28	B.1.110.3	20A	T	-	T	-	-	T	-	-	T	-	-	-	-	-	T	-	-	-	-	-	-	A	T	-	T	-	T	-	-	-	G	-	-	-	-	-	G	-	T	-	-	-	A	-	-	Del	-	T	-	-	-	-	-	-
3	20_1289-T35	B.1.1	20B	-	-	T	-	-	-	-	-	T	-	-	-	G	T	-	-	T	-	-	-	-	-	-	-	T	T	-	-	-	-	-	-	-	-	-	C	G	T	-	-	-	-	-	-	T	-	-	-	A	A	C	-	-	-
4	20_2106-T28	B.1.436	20A	-	T	T	-	-	-	-	-	T	-	A	T	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	T	-	-	-	-	-	-	-	G	-	T	-	-	-	-	-	-	-	-	-	-	-	-	T	-	A
5	20_3319-T42	B.1.369	20C	-	-	T	-	-	-	T	A	T	T	-	-	-	-	-	G	-	-	-	-	-	-	-	-	T	-	-	T	-	T	-	G	T	-	-	-	G	-	T	-	T	-	-	-	-	-	A	-	-	-	-	-	-	-

B.



Mutation profile [relative to Wuhan ref genome]

cluster	id	PANGO lineage	NextStrain clade	[V] ORF1a:54	[E] ORF1a:65	[T] ORF1a:265	[G] ORF1a:519	[A] ORF1a:1314	[G] ORF1a:1819	[S] ORF1a:2500	[S] ORF1a:2625	[P] ORF1a:3371	[Q] ORF1a:3777	[S] ORF1a:3883	[Q] ORF1a:4189	[V] ORF1a:4216	[P] ORF1b:314	[M] ORF1b:592	[D] ORF1b:2067	[D] ORF1b:2179	[L] S:5	[L] S:18	[L] S:242	[D] S:614	[E] S:1258	[Q] ORF3a:57	[T] ORF3a:175	[V] M:70	[G] ORF8:8	[P] ORF8:36	[A] ORF8:51	[I] ORF8:121	[S] N:183	[S] N:202	[R] N:203	[G] N:204	[S] N:206	[A] N:414
1A	20_0156-T42	B.1.1	20B	-	A	-	-	-	-	-	-	-	-	-	-	I	L	-	-	-		F	-	G	-	-	I	F	-	L	-	-	-	-	K	R	-	-
1B	20_0504-T46	B.1.1	20B	-	A	-	-	-	-	-	-	-	-	A	-	I	L	-	-	-		F	-	G	-	-	I	F	-	L	-	-	-	-	K	R	-	-
1C	20_0508-T46	B.1.1	20B	Del	A	-	-	-	-	-	-	-	-	A	-	I	L	-	-	-	-	F	-	G	-	-	I	F	-	L	-	-	-	-	K	R	-	-
1D	20_1187-T42	B.1.1	20B	-	A	-	-	-	-	-	-	-	-	A	-	I	L	-	-	-	-	F	-	G	-	-	I	F	-	L	-	-	-	-	K	R	-	S
2A	20_0494-T56	B.1.110.3	20A	-	-	-	-	-	-	-	-	-	-	-	H	-	L	-	-	-	-	-	-	G	-	H	-	-	R	-	-	X*	-	C	-	-	-	-
2B	20_1025-T49	B.1.110.3	20A	-	-	-	-	-	-	-	-	-	H	-	H	-	L	-	-	-	-	-	-	G	-	H	-	-	R	-	-	X*	-	C	-	-	-	-
2C	20_1536-T42	B.1.110.3	20A	-	-	-	-	-	-	-	-	-	-	-	H	-	L	-	-	-	-	-	-	G	-	H	-	-	R	-	-	X*	-	C	-	-	-	-
2D	20_2176-T28	B.1.110.3	20A	-	-	-	-	-	-	-	-	-	-	-	H	-	L	I	G	-	-	-	-	G	-	H	-	-	R	-	-	X*	-	C	-	-	-	-
3	20_1289-T35	B.1.1	20B	-	-	-	-	-	-	-	F	S	-	-	-	-	L	-	-	-	-	-	P	G	D	-	-	-	-	-	S	-	-	-	K	R	-	-
4	20_2106-T28	B.1.436	20A	-	-	-	-	-	S	F	-	-	-	-	-	-	L	-	-	-	-	-	-	G	-	H	-	-	-	-	-	-	-	-	-	-	F	-
5	20_3319-T42	B.1.369	20C	-	-	I	S	V	-	-	-	-	-	-	-	-	L	-	-	G	F	-	-	G	-	H	-	-	-	-	-	-	Y	-	-	-	-	-

*=frameshift mutation replacing the last amino acid (I121) in ORF8 with SKRTN

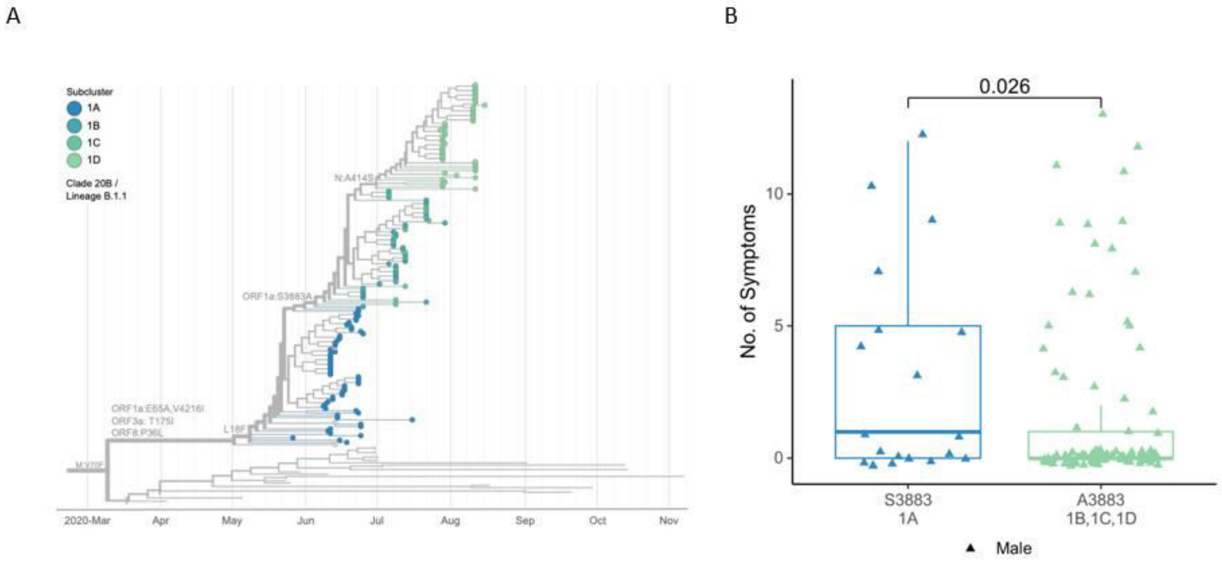
eTable 3. Complete mutation profile of the transmission clusters/subclusters, including clade-defining variants.

(**A**) Base substitutions. (**B**) Amino acid substitutions.

	Group	Mutation	N	Median	IQR
All participants	S3883	No	31	3.0	[0.0,6.5]
	A3883	Yes	98	0.0	[0.0,1.0]
Male participants	S3883	No	19	1.0	[0.0,5.0]
	A3883	Yes	97	0.0	[0.0,1.0]

eTable 4. Summary of the number of symptoms stratified on the S3883A mutation. ‘All participants’ designate men and women combined (see also **eFigure 4**). ‘Male participants’ are men only (see also **eFigure 1B**). N, population size. IQR, inter quartile range.

eFigures



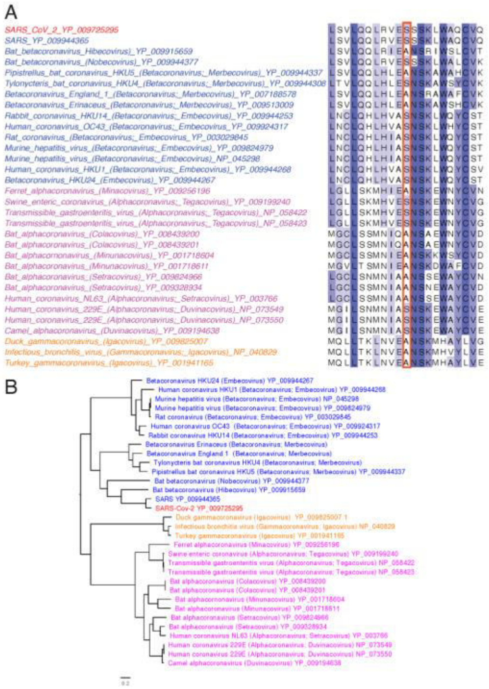
eFigure 1. SARS-CoV-2 mutation was associated with fewer symptoms.

(**A**) Time-calibrated maximum-likelihood phylogenetic subtree of the SARS-CoV-2 sequences constituting subclusters 1A-D. A global tree with background sequences from GISAID with a subsampling scheme focused on South Carolina was inferred using Nextstrain (https://nextstrain.org). NextStrain clade and PANGO lineage are indicated in the legend and amino acid substitutions are shown at the nodes. (**B**) Boxplots of the number of symptoms reported for the period from 2 weeks preceding to 2 weeks following initial diagnosis, in SARS-CoV-2-positive male participants with the viral mutation S3883A (‘A3883’) compared to male participants without the mutation (‘S3883’). A Wilcoxon test was used to calculate the p value.

NC_045512_SARS_CoV_2/1-21	LSVLQQLRVESSSKLWAQCVQ
MG772933_Bat_SARS_like_coronavirus_CoVZC45/1-21	LSVLQQLRVESSSKLWAQCVQ
MG772934_Bat_SARS_like_coronavirus_CoVZXC21/1-21	LSVLQQLRVESSSKLWAQCVQ
MN996532_Bat_coronavirus_RaTG13/1-21	LSVLQQLRVESSSKLWAQCVQ
MK211374_Coronavirus_BtRl_BetaCoV_SC2018/1-21	LSVLQQLRVESSSKLWAQCVQ
DQ412043_Bat_SARS_coronavirus_Rm1/1-21	LSVLQQLRVESSSKLWAQCVQ
KY938558_Bat_coronavirus_strain_16BO133/1-21	LSVLQQLRVESSSKLWAQCVQ
KY770860_Bat_coronavirus_Jiyuan_84/1-21	LSVLQQLRVESSSKLWAQCVQ
KJ473812_BtRf_BetaCoV_HeB2013/1-21	LSVLQQLRVESSSKLWAQCVQ
DQ412042_Bat_SARS_coronavirus_Rf1/1-21	LSVLQQLRVESSSKLWAQCVQ
DQ648856_Bat_coronavirus_BtCoV_273_2005/1-21	LSVLQQLRVESSSKLWAQCVQ
GQ153542_Bat_SARS_coronavirus_HKU3_7/1-21	LSVLQQLRVESSSKLWAQCVQ
DQ022305_Bat_SARS_coronavirus_HKU3_1/1-21	LSVLQQLRVESSSKLWAQCVQ
GQ153547_Bat_SARS_coronavirus_HKU3_12/1-21	LSVLQQLRVESSSKLWAQCVQ
KJ473814_BtRs_BetaCoV_HuB2013/1-21	LSVLQQLRVESSSKLWAQCVQ
JX993987_Bat_coronavirus_Rp_Shaanxi2011/1-21	LSVLQQLRVESSSKLWAQCVQ
JX993988_Bat_coronavirus_Cp_Yunnan2011/1-21	LSVLQQLRVESSSKLWAQCVQ
KU973692_UNVERIFIED_SARS_related_coronavirus_F46/1-21	LSVLQQLRVESSSKLWAQCVQ
KF569996_Rhinolophus_affinis_coronavirus_LYRa11/1-21	LSVLQQLRVESSSKLWAQCVQ
KJ473815_BtRs_BetaCoV_GX2013/1-21	LSVLQQLRVESSSKLWAQCVQ
KP886808_Bat_SARS_like_coronavirus_YNLF_31C/1-21	LSVLQQLRVESSSKLWAQCVQ
DQ071615_Bat_SARS_coronavirus_Rp3/1-21	LSVLQQLRVESSSKLWAQCVQ
NC_004718_SARS_coronavirus/1-21	LSVLQQLRVESSSKLWAQCVQ
FJ588686_Bat_SARS_CoV_Rs672_2006/1-21	LSVLQQLRVESSSKLWAQCVQ
KJ473816_BtRs_BetaCoV_YN2013/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417145_Bat_SARS_like_coronavirus_Rf4092/1-21	LSVLQQLRVESSSKLWAQCVQ
MK211375_Coronavirus_BtRs_BetaCoV_YN2018A/1-21	LSVLQQLRVESSSKLWAQCVQ
KY770858_Bat_coronavirus_Anlong_103/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417144_Bat_SARS_like_coronavirus_Rs4084/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417149_Bat_SARS_like_coronavirus_Rs4255/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417148_Bat_SARS_like_coronavirus_Rs4247/1-21	LSVLQQLRVESSSKLWAQCVQ
KT444582_SARS_like_coronavirus_WIV16/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417143_Bat_SARS_like_coronavirus_Rs4081/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417152_Bat_SARS_like_coronavirus_Rs9401/1-21	LSVLQQLRVESSSKLWAQCVQ
MK211377_Coronavirus_BtRs_BetaCoV_YN2018C/1-21	LSVLQQLRVESSSKLWAQCVQ
MK211376_Coronavirus_BtRs_BetaCoV_YN2018B/1-21	LSVLQQLRVESSSKLWAQCVQ
MK211378_Coronavirus_BtRs_BetaCoV_YN2018D/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417146_Bat_SARS_like_coronavirus_Rs4231/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417151_Bat_SARS_like_coronavirus_Rs7327/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417142_Bat_SARS_like_coronavirus_As6526/1-21	LSVLQQLRVESSSKLWAQCVQ
KF367457_Bat_SARS_like_coronavirus_WIV1/1-21	LSVLQQLRVESSSKLWAQCVQ
KY417147_Bat_SARS_like_coronavirus_Rs4237/1-21	LSVLQQLRVESSSKLWAQCVQ
KY352407_SARS_related_coronavirus_strain_BtKY72/1-21	LSVLQQLRIESSSKLWTQCVQ
NC_014470_Bat_coronavirus_BM48_31_BGR_2008/1-21	LSVLQQLRIESSSKLWAQCVQ

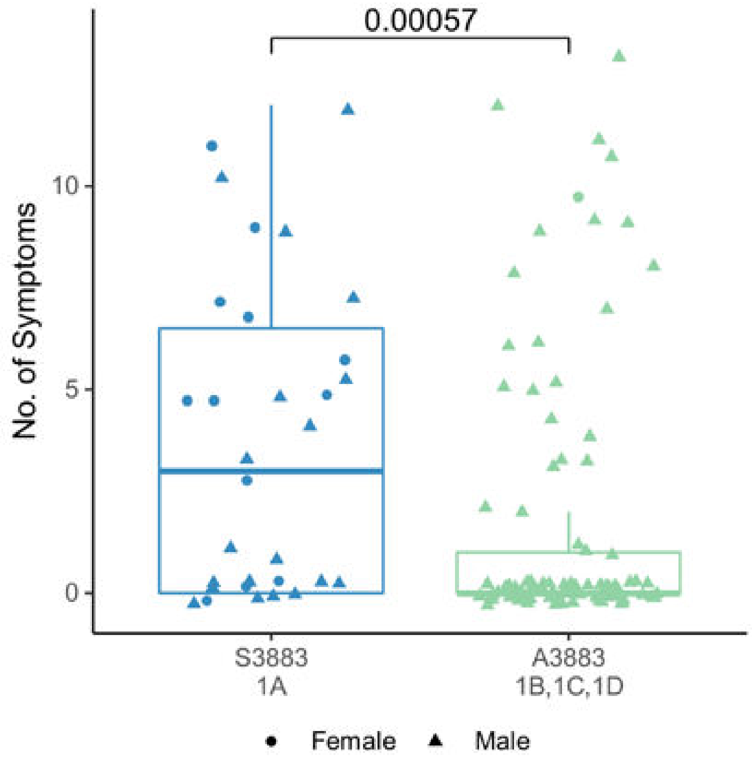
eFigure 2. Alignment of a portion of the SARS-CoV-2 nsp7amino acid sequence with other members of the subgenus Sarbecovirus.

Sarbecovirus genomes are from (20). Boxed in red is the Serine 3883 residue.


A
SARS_CoV_2_YP_009725295 LSVLQQLRVESSSKLWAQCVQ
SARS_YP_009944365 LSVLQQLRVESSSKLWAQCVQ
Bat_betacoronavirus_Hibecovirus)_YP_009915659 LSVLQQLRIEANSRIWSLCVT
Bat_betacoronavirus_(Nobecovirus)_YP_009944377 LSVLQQLRVESSSKLWALCVK
Pipistrellus_bat_coronavirus_HKU5_(Betacoronavirus;_Merbecovirus)_YP_009944337 LSVLQQLHLEANSKAWAHCVK
Tylonycteris_bat_coronavirus_HKU4_(Betacoronavirus;_Merbecovirus)_YP_009944308 LTVLQQLHLESNSKAWSYCVK
Betacoronavirus_England_1_(Betacoronavirus;_Merbecovirus)_YP_007188578 LSVLQQLHLEANSRAWAFCVK
Betacoronavirus_Erinaceus_(Betacoronavirus;_Merbecovirus)_YP_009513009 LSVLQQLHLEANSKAWSHCVK
Rabbit_coronavirus_HKU14_(Betacoronavirus;_Embecovirus)_YP_009944253 LNCLQHLHVASNSKLWQYCST
Human_coronavirus_OC43_(Betacoronavirus;_Embecovirus_YP_009924317 LNCLQHLHVASNSKLWHYCST
Rat_coronavirus_(Betacoronavirus;_Embecovirus_YP_003029845 LNCLQHLHIASNSKLWQYCST
Murine_hepatitis_virus_(Betacoronavirus;_Embecovirus)_YP_009824979 LNCLQHLHIASNSKLWQYCST
Murine_hepatitis_virus_(Betacoronavirus;_Embecovirus)_NP_045298 LNCLQHLHIASNSKLWQYCST
Human_coronavirus_HKU1_(Betacoronavirus;_Embecovirus)_YP_009944268 LNCLQHLHVASNSKLWQYCSV
Betacoronavirus_HKU24_(Embecovirus)_YP_009944267 LNCLQHLHVASNSKLWQYCST
Ferret_alphacoronavirus_(Minacovirus)_YP_009256196 LGLLSKMHIEANSKEWNYCVN
Swine_enteric_coronavirus_(Alphacoronavirus;_Tegacovirus)_YP_009199240 LGLLSKMHVESNSKEWNYCVG
Transmissible_gastroenteritis_virus_(Alphacoronavirus;_Tegacovirus)_NP_058422 LGLLSKMHVESNSKEWNYCVG
Transmissible_gastroenteritis_virus_(Alphacoronavirus;_Tegacovirus)_NP_058423 LGLLSKMHVESNSKEWNYCVG
Bat_alphacoronavirus_(Colacovirus)_YP_008439200 MGCLSSMNIQANSAEWNYCVD
Bat_alphacoronavirus_(Colacovirus)_YP_008439201 MGCLSSMNIQANSAEWNYCVD
Bat_alphacomonavirus_(Minunacovirus)_YP_001718604 MGCLSSMNIEANSKKWSYCVD
Bat_alphacoronavirus_(Minunacovirus)_YP_001718611 MGVLTSMNIEANSKDWAFCVD
Bat_alphacoronavirus_(Setracovirus)_YP_009824966 LGCLSNMNIAANSSEWAYCVN
Bat_alphacoronavirus_(Setracovirus)_YP_009328934 LGCLSSMNIAANSNEWAYCVD
Human_coronavirus_NL63_(Alphacoronavirus;_Setracovirus)_YP_003766 LGCLSSMNIAANSSEWAYCVD
Human_coronavirus_229E_(Alphacoronavirus;_Duvinacovirus)_NP_073549 MGILSNMNIASNSKEWAYCVE
Human_coronavirus_229E_(Alphacoronavirus;_Duvinacovirus)_NP_073550 MGILSNMNIASNSKEWAYCVE
Camel_alphacoronavirus_(Duvinacovirus)_YP_009194638 MGILSNMNIASNSKEWAYCVE
Duck_gammacoronavirus_(Igacovirus)_YP_009825007 MQLLTKLNVEANSKMHAYLVG
Infectious_bronchitis_virus_(Gammacoronavirus;_Igacovirus)_NP_040829 MQLLTKLNVEANSKMHVYLVE
Turkey_gammacoronavirus_(Igacovirus)_YP_001941165 MQLLTKLNVEANSKMHAYLVE
B
Betacoronavirus HKU24 (Embecovirus) YP_009944267
Human coronavirus HKU1 (Betacoronavirus; Embecovirus) YP_009944268
Murine hepatitis virus (Betacoronavirus; Embecovirus) NP_045298
Murine hepatitis virus (Betacoronavirus; Embecovirus) YP_009824979
Rat coronavirus (Betacoronavirus; Embecovirus) YP_003029845
Human coronavirus OC43 (Betacoronavirus; Embecovirus) YP_009924317
Rabbit coronavirus HKU14 (Betacoronavirus; Embecovirus) YP_009944253
Betacoronavirus Erinaceus (Betacoronavirus; Merbecovirus)
Betacoronavirus England 1 (Betacoronavirus; Merbecovirus)
Tylonycteris bat coronavirus HKU4 (Betacoronavirus; Merbecovirus)
Pipistrellus bat coronavirus HKU5 (Betacoronavirus; Merbecovirus) YP_009944337
Bat betacoronavirus (Nobecovirus) YP_009944377
Bat betacoronavirus (Hibecovirus) YP_009915659
SARS YP_009944365
SARS-Cov-2 YP_009725295
Duck gammacoronavirus (Igacovirus) YP_009825007.1
Infectious bronchitis virus (Gammacoronavirus; Igacovirus) NP_040829
Turkey gammacoronavirus (Igacovirus) YP_001941165
Ferret alphacoronavirus (Minacovirus) YP_009256196
Swine enteric coronavirus (Alphacoronavirus; Tegacovirus) YP_009199240
Transmissible gastroenteritis virus (Alphacoronavirus; Tegacovirus) NP_058422
Transmissible gastroenteritis virus (Alphacoronavirus; Tegacovirus) NP_058423
Bat alphacoronavirus (Colacovirus) YP_008439200
Bat alphacoronavirus (Colacovirus) YP_008439201
Bat alphacornonavirus (Minunacovirus) YP_001718604
Bat alphacoronavirus (Minunacovirus) YP_001718611
Bat alphacoronavirus (Setracovirus) YP_009824966
Bat alphacoronavirus (Setracovirus) YP_009328934
Human coronavirus NL63 (Alphacoronavirus; Setracovirus) YP_003766
Human coronavirus 229E (Alphacoronavirus; Duvinacovirus) NP_073549
Human coronavirus 229E (Alphacoronavirus; Duvinacovirus) NP_073550
Camel alphacoronavirus (Duvinacovirus) YP_009194638
0.2

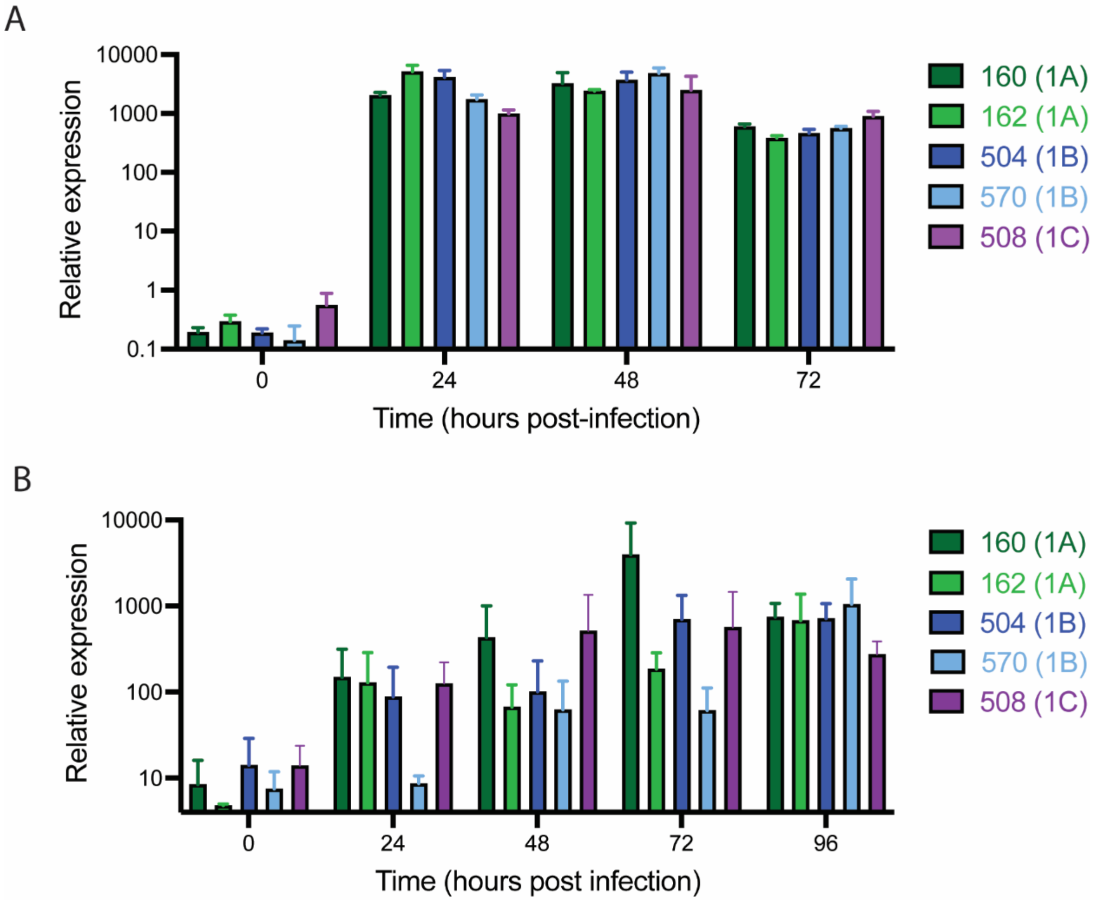
eFigure 3. (A) Alignment of a portion of the SARS-CoV-2 amino acid sequence with RefSeq sequences for other members of the main Coronaviridae subfamily, Orthocoronavirinae.

Boxed in red is the Serine 3883 residue. Isolate names are colored by genus and the alignment is visualized using Jalview. (**B**) Maximum-likelihood phylogenetic tree for the ORF1A protein sequences for the isolates shown above, constructed with RAxML using the PROTCATLG model.



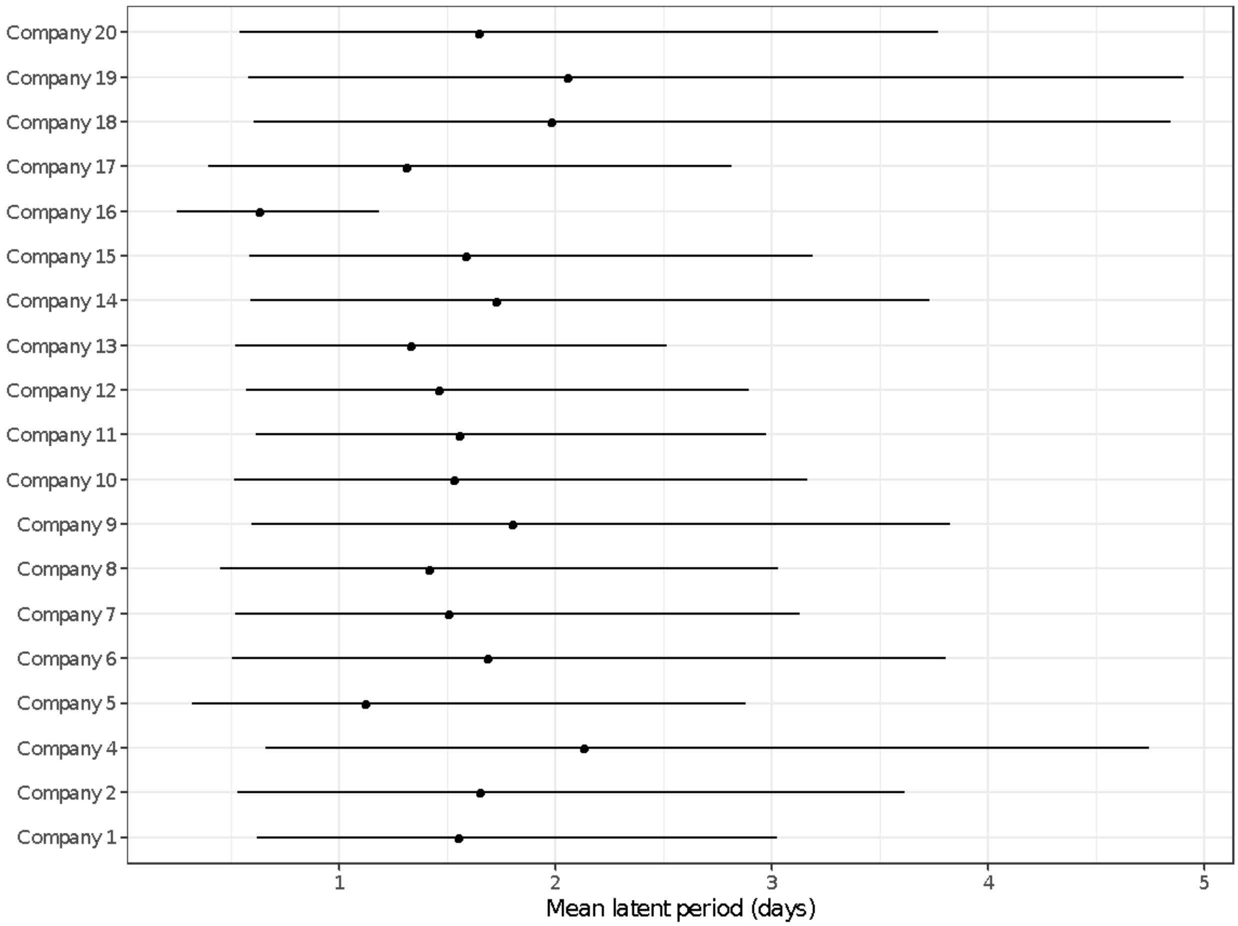
eFigure 4. Boxplots of the number of total symptoms during the two weeks preceding plus the two weeks following initial infection, in SARS-CoV-2-positive participants with the viral mutation S3883A (‘A3883’) *vs.* participants without the mutation (‘S3883’).

Circles denote female participants, while triangles denote male participants. A Wilcoxon test was used to calculate the p value.



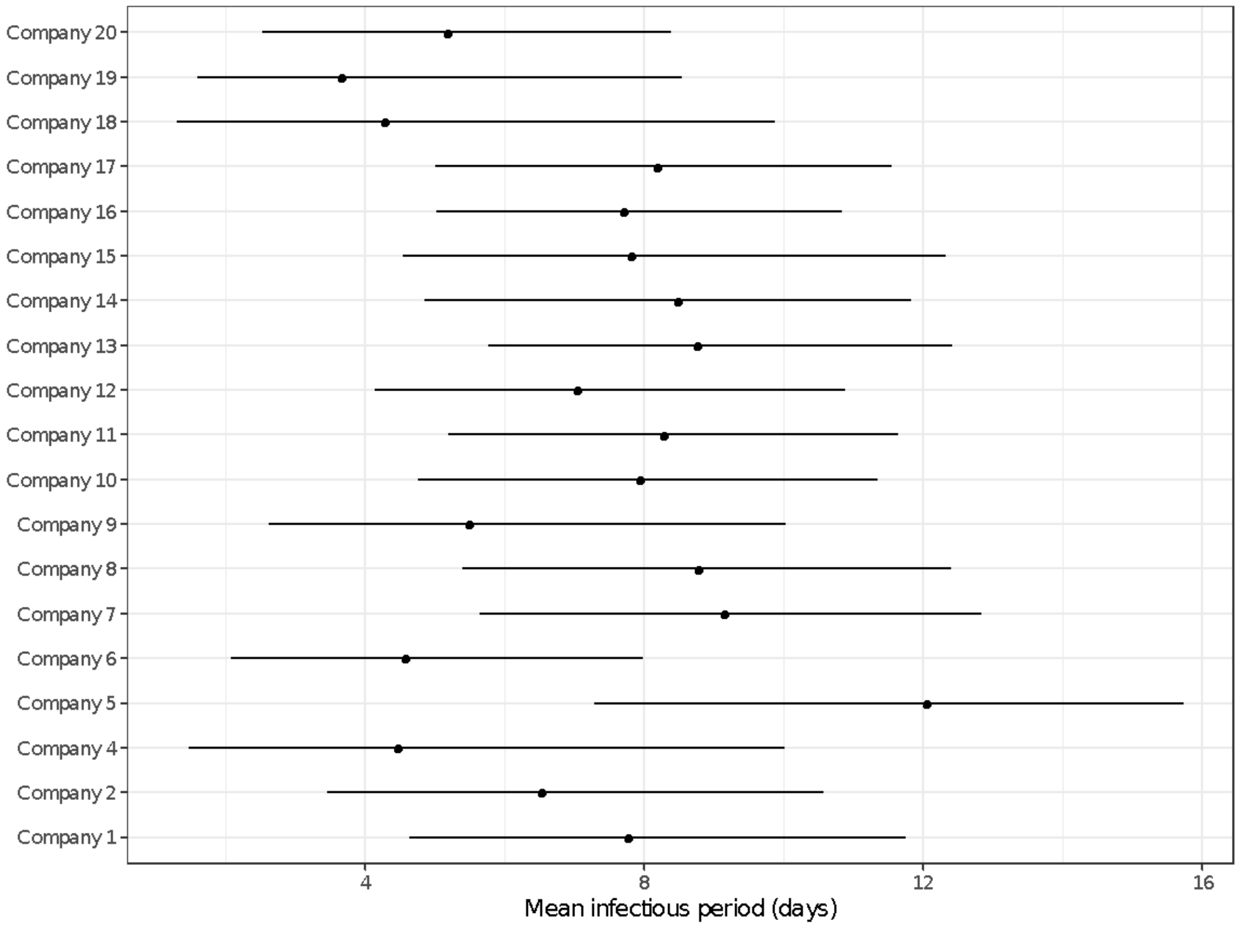
eFigure 5. Virus isolates carrying S3883 (subcluster 1A) or A3883 (subclusters 1B, 1C) show no difference in replication efficiency in cultured epithelial cells.

The Vero E6 cell line (A) and primary Normal Human Bronchial Epithelial (NHBE) cells (B) were both infected with cultured SARS-CoV-2 isolates from subclusters 1A, 1B, and 1C of Cluster 1, and virus replication was assessed by quantifying the SARS-CoV2 E gene using qPCR. Results were normalized to the 18S ribosomal gene. Similar levels of replication were observed across the isolates from the different subclusters. The numbers to the right of the bar graph indicate the virus isolate number and in parenthesis, the subcluster number.



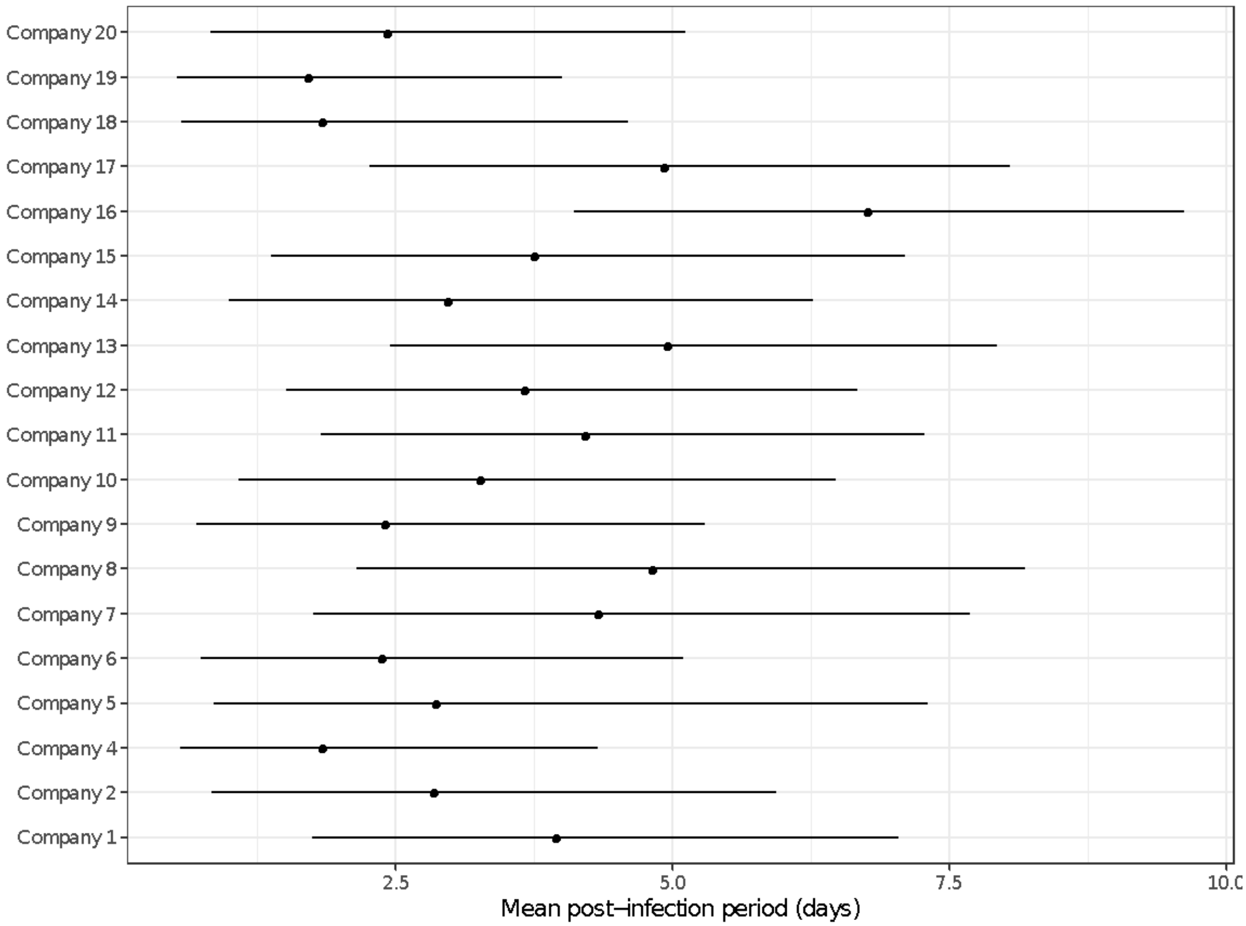
eFigure 6. Estimates of mean latent periods across 20 companies.

This figure is based on the mathematical modeling of SARS-CoV-2 transmission dynamics within companies. Points represent posterior medians. Error bars represent 95% credible intervals.



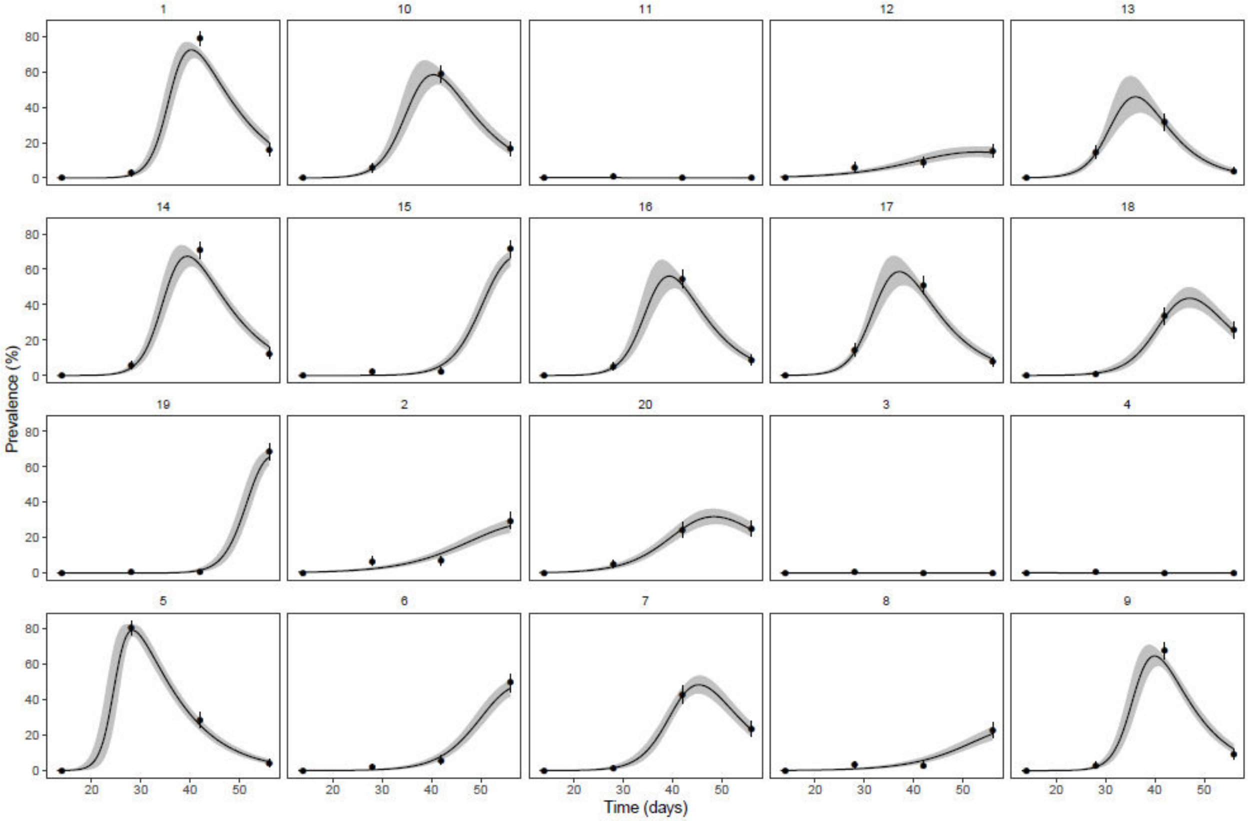
eFigure 7. Estimates of mean infectious periods across 20 companies.

Points represent posterior medians. Error bars represent 95% credible intervals.



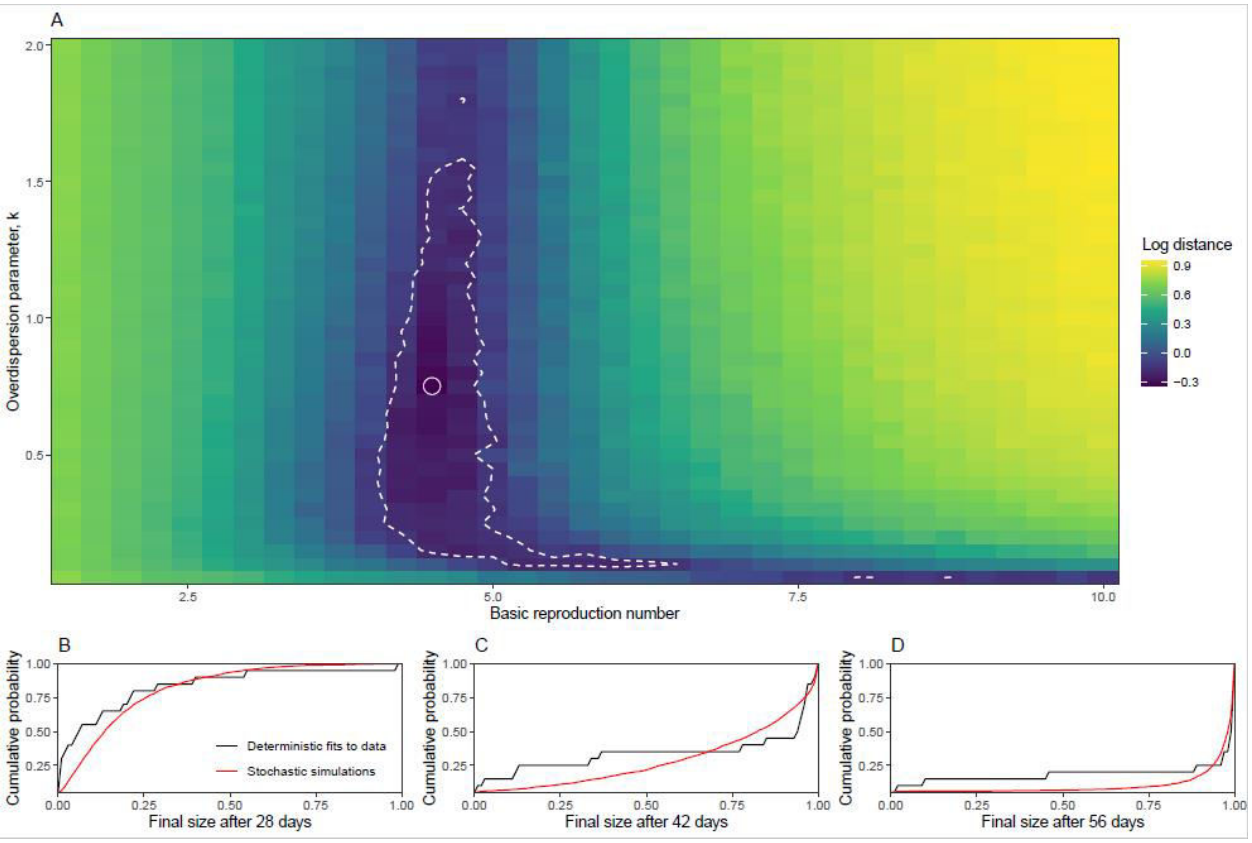
eFigure 8. Estimates of mean post-infection periods across 20 companies.

Points represent posterior medians. Error bars represent 95% credible intervals.



eFigure 9. Simulated outbreaks using stochastic model and deterministic model fits.

Model fits to synthetic prevalence data from 20 simulations. Points represent the prevalence from stochastic simulations. Solid lines and shaded regions represent the posterior median and 95% credible intervals of the predicted prevalence using the deterministic model fitted to simulated data.



eFigure 10. Parameter estimates from stochastic model fits to predicted cumulative proportions of infected in each company.

For a given value of R_0 and k, we compared the distribution of cumulative proportion of infections after 28, 42, and 56 days simulated from stochastic simulations and those predicted from fitted deterministic models to data using Kolmogorov-Smirnov statistic. (**A**) Heat map of log-distance between simulated and predicted distributions, defined as the sum of Kolmogorov-Smirnov statistic. White point represents the parameter set that minimizes the distance. White dashed lines represent contour lines for parameters whose distance is 20% away from the minimum distance. (**B**-**D**) Comparison of simulated and predicted distributions of cumulative proportions of infections after 28, 42, and 56 days.

Supplementary References

1. Stewart JN. DOD INSTRUCTION 6490.03 DEPLOYMENT HEALTH. In: Defense o, editor. Directives Division Website at http://www.esd.whs.mil/DD 2019.

2. Viral Amplicon Illumina Workflow (VAIW): A custom pipeline to analyze the SARS-CoV-2 genomes prepared with an amplicon (ARTIC (v3) and YouSeq (v2)) based library protocols [Internet]. 2020. Available from: https://hub.docker.com/r/bdrdgenomics/viral_amplicon_illumina_workflow.

3. BBMap short read aligner, and other bioinformatic tools. [Internet]. 2014. Available from: https://sourceforge.net/projects/bbmap.

4. Bushnell B, Rood J, Singer E. BBMerge - Accurate paired shotgun read merging via overlap. PLoS One. 2017;12(10):e0185056.

5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

6. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20(1):8.

7. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. Science. 2020;369(6501):297-301.

8. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16(1):37-48.

9. Leigh JW, Bryant D. POPART: full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6(9):1110-6.

10. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121-3.

11. Harcourt J, Tamin A, Lu X, Kamili S, Sakthivel SK, Murray J, et al. Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. Emerg Infect Dis. 2020;26(6):1266-73.

12. El Jamal SM, Pujadas E, Ramos I, Bryce C, Grimes ZM, Amanat F, et al. Tissue-based SARS-CoV-2 detection in fatal COVID-19 infections: Sustained direct viral-induced damage is not necessary to drive disease progression. Hum Pathol. 2021;114:110-9.

13. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. Nature. 2020;584(7821):425-9.

14. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A Probabilistic Programming Language. 2017. 2017;76(1):32.

15. He D, Ionides EL, King AA. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. J R Soc Interface. 2010;7(43):271-83.

16. Endo A, Centre for the Mathematical Modelling of Infectious Diseases C-WG, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res. 2020;5:67.

17. Smirnov N. Table for Estimating the Goodness of Fit of Empirical Distributions. Ann Math Stat. 1948;19(2):279-.

18. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

19. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dorner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368(6491).

20. Jungreis I, Sealfon R, Kellis M. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. Nat Commun. 2021;12(1):2642.