# SUPPLEMENTAL MATERIAL

## SUPPLEMENTAL METHODS

*cDNA Synthesis and Viral Genome Amplification*

cDNA synthesis was performed with SuperScript IV First Strand Synthesis Kit (Thermo) using 11 µl of extracted viral nucleic acids and random hexamers according to manufacturer's specifications. Direct amplification of the viral genome cDNA was performed in multiplexed PCR reactions to generate ~400 bp amplicons tiled across the genome. The multiplex primer set, comprised of two non-overlapping primer pools, was created using Primal Scheme and provided by the Artic Network (version 3 release).[1] PCR amplification was carried out using Q5 Hot Start HF Taq Polymerase (NEB) with 5 µl of cDNA in a 25 µl reaction volume. A two-step PCR program was used with an initial step of 98°C for 30 seconds, then 35 cycles of 98°C for 15 seconds followed by five minutes at 65°C. Separate reactions were carried out for each primer pool and validated by agarose gel electrophoresis.

*Sequencing Library Preparation and Nanopore Sequencing*

Sequencing library preparation approach was adapted from the ARTIC Network protocol and Oxford Nanopore protocol "PCR tiling of COVID-19 virus".[1-3] Amplicons from both primer pools were combined and purified with a 1x volume of AmpureXP beads (Beckman Coulter). A total of 50 ng of DNA was treated with NEB Ultra II End Prep Enzyme mix (NEB). Up to 24 specimen libraries were barcoded using Nanopore Native Barcoding Expansion kits (Oxford Nanopore) and NEBNext Ultra II Ligase (NEB) for simultaneous sequencing. Uniquely barcoded samples were pooled and cleaned with a 0.4x volume of AmpureXP beads. Adapter ligation and cleanup was performed using the Oxford Nanopore Genomic DNA by Ligation kit according to manufacturer specifications. Libraries were sequenced on the Nanopore MinION device using FLO-MIN106D Type R9.4.1 flow cells.

*Sequence data analysis*

Base calling of nanopore reads was performed on the Northwestern University's Quest High Performance Computing Cluster for Genomics using Guppy v3.4.5 with the DNA R9.4.1 450bps High Accuracy configuration and default quality score filtering (--qscore_filtering). Read quality filtering, reference alignment, and consensus sequence generation was performed using ARTIC Network software and protocols.[3,4] Briefly, demultiplexed and quality filtered sequence reads were aligned to a SARS-CoV-2 reference genome sequence (NCBI accession number MN908947.3) using minimap2 v2.17.[5] Alignments were filtered to remove low-quality alignments, and variants relative to the reference were determined from raw signal data using nanopolish v0.12.5.[6] Regions with less than 10-fold read coverage were masked. We chose to use a lower read coverage cutoff than the default of 20 in the ARTIC Network software to allow for consensus sequence calls in regions with relatively poor amplification due to inefficient primer binding. Manual validation of variant calls in regions with 10 – 19x coverage was performed using Tablet software.[7] Isolate lineages were determined with PANGOLIN using the 2020-05-19 definitions.[8,9] Consensus genome sequences were deposited in the GISAID database. Accession numbers are given in **Supplemental Table 1**. Publicly-available SARS-CoV-2 genome sequences were downloaded from the GISAID database on 04/10/2020.[10] Only full genome sequences in the database with <1% Ns were selected for inclusion in the analyses.

*Phylogenetic analysis and phylogeographic inference*

Genome sequences were aligned using MAFFT v7.453 software and manually edited using MEGA v6.06.[11,12] The US analyses included 901 SARS-CoV-2 sequences from the US deposited in GISAID as of April 4, 2020 and the global analysis included 3099 SARS-CoV-2 sequences deposited in GISAID from 64 different countries up to April 4, 2020. All Maximum Likelihood (ML) phylogenies were inferred with IQ-Tree v 2.0.5 using its ModelFinder before each analysis to estimate the nucleotide substitution model best-fitted for each dataset by means of Bayesian information criterion (BIC).[13,14] We assessed the tree topology for each phylogeny both with the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) and with ultrafast bootstrap (UFboot) with 1000 replicates each.[15,16] TreeTime v0.7.4 was used for the assessment of root-to-tip correlation, the estimation of time scaled phylogenies and ancestral reconstruction of most likely sequences of internal nodes of the tree and transitions between geographical locations along branches.[17] We used the sampling dates of the sequences to estimate the evolutionary rates and determine the best rooting of the tree. We used the population proportion per location in each of the datasets as weights in the analysis to correct for strong sampling biases in the datasets.

Bayesian time-scaled phylogenetic analyses were only performed for the phylogenies of each clade separately due to the high complexity of the full US and global phylogenies. We used BEAST v2.5.2 to estimate the date and location of the most recent common ancestors (MRCA) as well as to estimate the rate of evolution of the virus.[18] BEAST priors were introduced with BEAUTI v2.5.2 including an uncorrelated relaxed molecular clock model with a lognormal distribution and the TreeTime estimate as the prior for the mean, a GTR substitution model with invariant sites, as the best-fitted model obtained with ModelFinder, and a Coalescence Bayesian Skyline to model the population size changes through time. Markov chain Monte Carlo (MCMC) runs of at least 100 million states with sampling every 5,000 steps were computed.[19] The convergence of MCMC chains was monitored using Tracer v.1.7.1, ensuring that the effective sample size (ESS) values were greater than 200 for each parameter estimated.

Phylogeographic patterns were estimated using a discrete-state continuous time Markov chain to reconstruct the spatial dynamics between geographical locations,[20] assuming an asymmetric transition model with separated rate parameters for each possible transition. MCMC was run for over 100 million steps with a burn-in of 20%. Parameters were sampled every 5,000 steps and trees sampled every 10,000 steps. Pairwise migration rate estimates had an ESS of more than 1000 in every case. Statistics from the trees were extracted using PACT v0.9.5 and SpreaD3 v0.9.6 was used to visualize the phylogeographic patterns associated with the phylogenies.[21,22] In each case, the maximum clade credibility (MCC) trees were obtained from the tree posterior distribution using TreeAnnotator v2.5.2 after 20% burn-in.

*Statistical Analysis*

All statistical analyses were performed in R version 4.0.0. The R package *lme4* was used for the linear and logistic regressions, *fitdistrplus* was used to test the best fitted model for the data, *lsmeans* was used to test for all possible contrasts and perform multiple comparison correction, *nnet* was used for the multinomial logistic regression, *finalfit* package was used to generate the descriptive table, and *ggplot2* and *ggpubr* were used to generate the statistical figures.

To evaluate for differences in Ct values between clades, specimens were grouped according to the estimated clade of the detected virus. All specimens that did not cluster in any of the 3 major clades were placed into a separate group. A linear model was fitted after testing for normal distribution of the data. Clade membership was included in the model, controlling for all possible confounders such as the date of the qPCR, age, sex, race, maximal severity of illness, and the specimen collection source. All possible contrasts within the model were performed and corrected for multiple comparisons using Tukey's test. Corrected p-values of less than 0.05 were considered statistically significant.

We also tested for differences in viral load by specimen source (nasopharyngeal swab vs. bronchoalveolar lavage) by fitting a linear model that included the same confounders as above, extended to contain the interaction between clade and specimen source. For this analysis we only used data from patients with viruses from the two most abundant clades, Clades 1 and 2 (n = 76 patients total). Within this model, we tested for all possible combinations of clade, source, and their interaction and corrected for multiple comparisons as before.

To determine if the difference in Ct values between clades was driven by differences in time since symptom onset or exposure, the number of days between symptom onset or exposure and specimen collection was determined by chart review. We tested for differences between Clade 1 and Clade 2 in the number of days between symptom onset and specimen collection (n = 76) or possible virus exposure (n = 29 patients) using a Wilcoxon Signed-Rank Test.

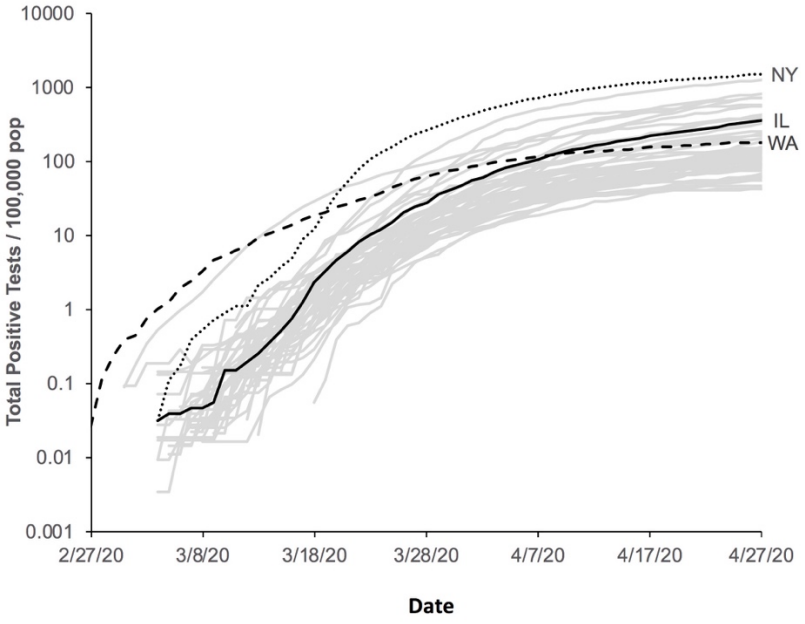
Finally, to test for association between disease severity and viral clade, we employed multinomial logistic regression, using maximal severity of COVID-19 per patient as the outcome variable [binned as mild (outpatient or ED only), moderate (inpatient hospitalization), or severe (ICU admission and/or death)], and clade, Ct values, and other demographic variables as the predictors. To analyze the association between Clade 1 and 2 and all available clinical and demographic variables, we fitted a logistic regression model including age, sex, race, severity of illness, time since symptoms onset, and specimen source to test for any possible association between these variables and the clade observed.

**SUPPLEMENTAL REFERENCES**


1. Artic Network. https://artic.network/ncov-2019 (accessed March 23, 2020).
2. Oxford Nanopore Technologies. https://community.nanoporetech.com (accessed April 27, 2020).

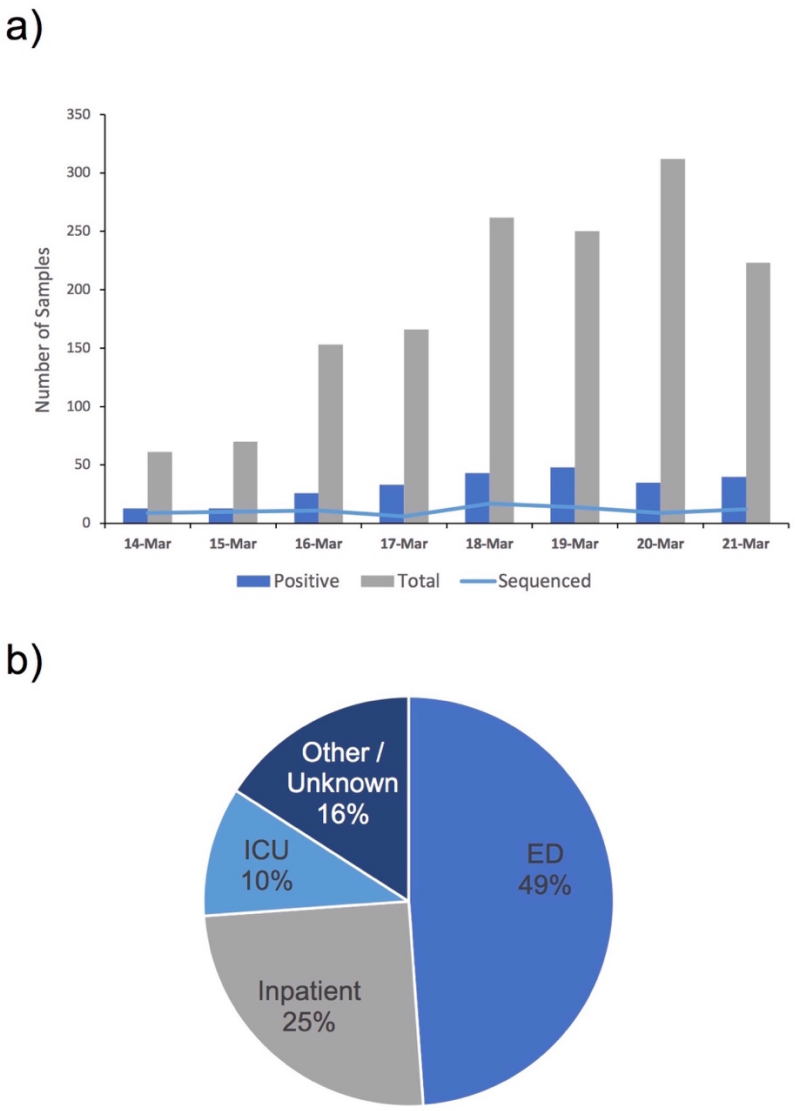
3. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017; **12**(6): 1261-76.
4. nCoV-2019 novel coronavirus bioinformatics protocol. https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html (accessed March 23, 2020).
5. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018; **34**(18): 3094-100.
6. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat Methods* 2015; **12**(8): 733-5.
7. Milne I, Bayer M, Stephen G, Cardle L, Marshall D. Tablet: Visualizing Next-Generation Sequence Assemblies and Mappings. *Methods Mol Biol* 2016; **1374**: 253-68.
8. Rambaut A, Holmes EC, Hill V, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* 2020: 2020.04.17.046086.
9. O'Toole Á, McCrone JT. pangolin (Phylogenetic Assignment of Named Global Outbreak LINeages). https://github.com/hCoV-2019/pangolin (accessed June 4, 2020).
10. GISAID: Over 10,000 viral genome sequences of hCoV-19 shared with unprecedented speed via GISAID. 2020. gisaid.org (accessed 4/22/2020 2020).
11. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; **30**(4): 772-80.
12. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 2013; **30**(12): 2725-9.
13. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020; **37**(5): 1530-4.
14. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017; **14**(6): 587-9.
15. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010; **59**(3): 307-21.
16. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 2018; **35**(2): 518-22.
17. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol* 2018; **4**(1): vex042.
18. Bouckaert R, Heled J, Kuhnert D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2014; **10**(4): e1003537.
19. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 2005; **22**(5): 1185-92.
20. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 2009; **5**(9): e1000520.
21. Hueck CJ. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* 1998; **62**(2): 379-433.
22. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Mol Biol Evol* 2016; **33**(8): 2167-9.

**Supplemental Figure 1**



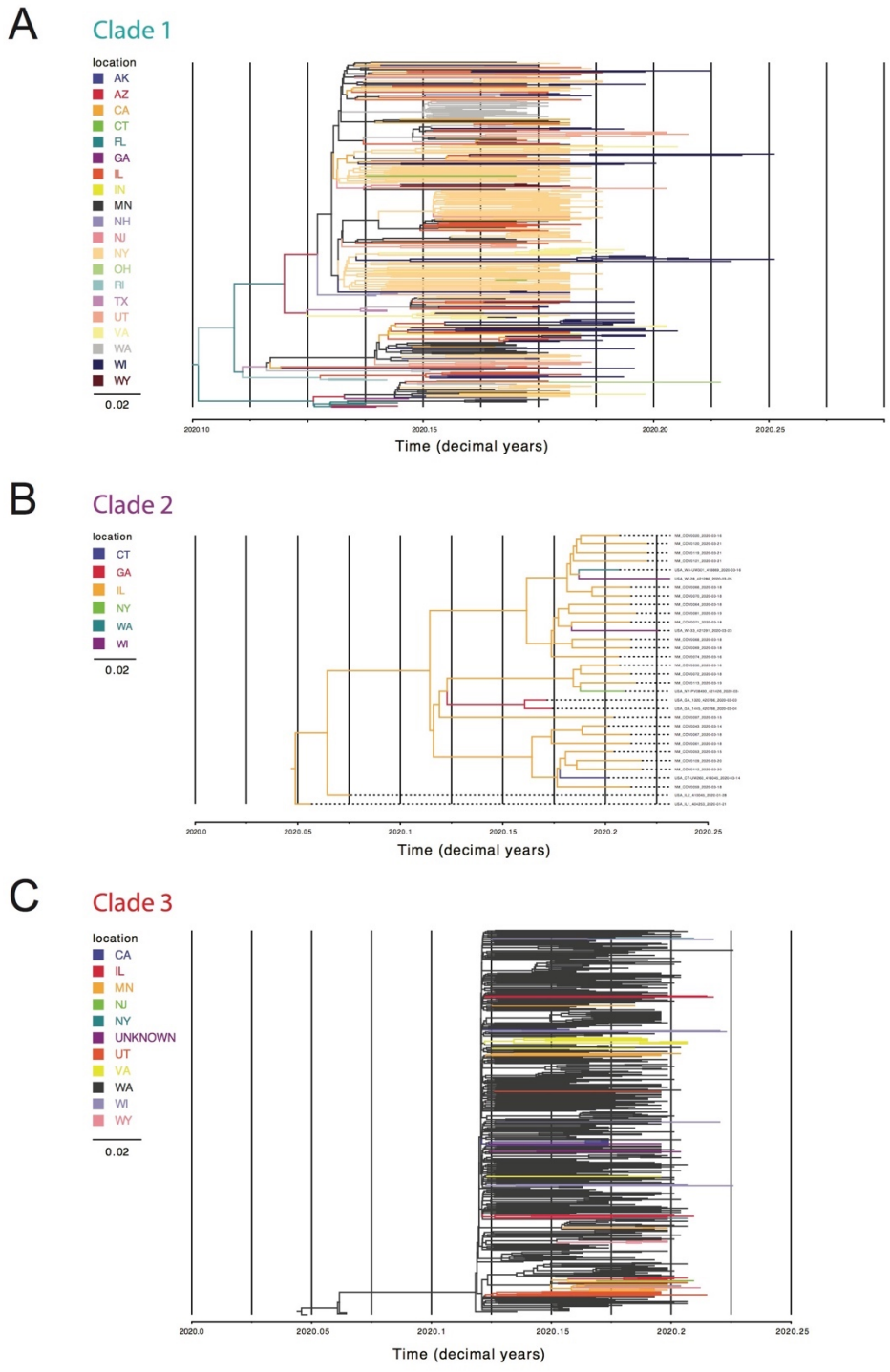

Per-state SARS-CoV-2 infections in the United States from Feb 27 through April 27, 2020. Lines represent total infection volume over time in number of positive tests per 100,000 residents in each of the 50 US states and the District of Columbia. States highlighted in black are Illinois (IL, solid line), New York (NY, dotted line), and Washington (WA, dashed line).

**Supplemental Figure 2**



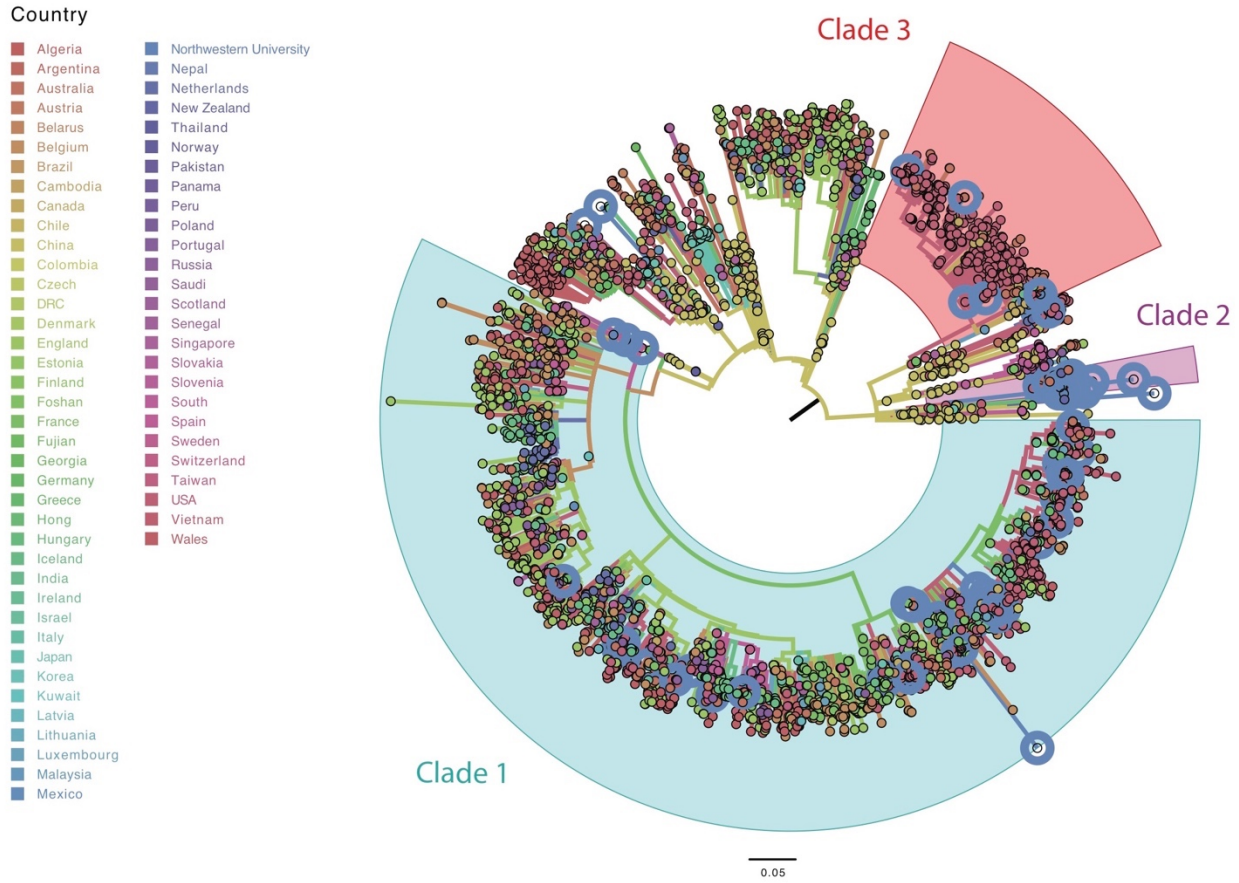

Collection characteristics of sequenced SARS-CoV-2 specimens. **a)** Numbers of specimens collected per day. **b)** Clinical site of specimen collection. ED = Emergency department, Inpatient = Non-ICU hospital ward, ICU = Intensive Care Unit, Other / Unknown = Outpatient, corporate health, or source undetermined.

**Supplemental Figure 3**



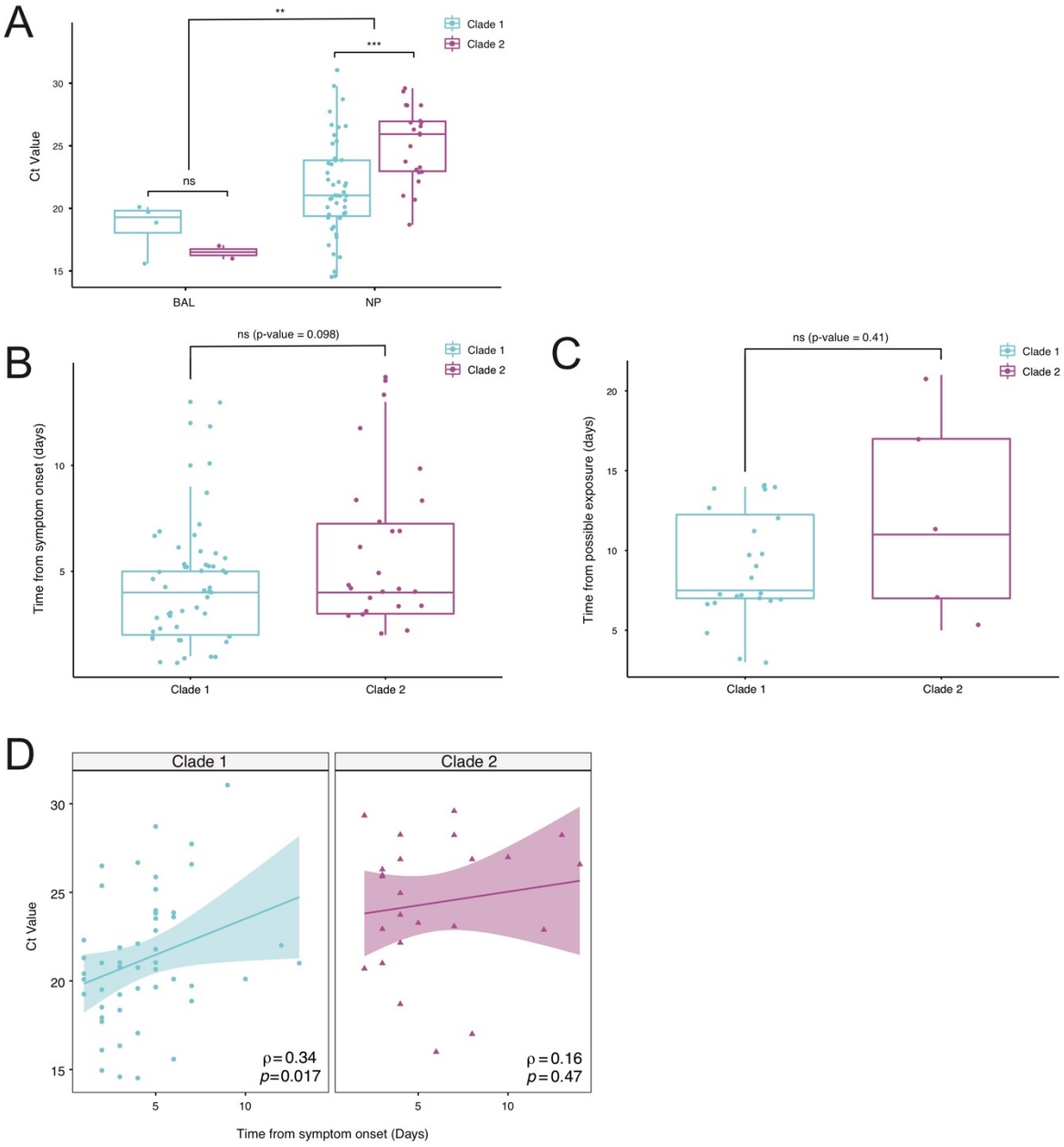

Phylodynamic trees of US isolates in three major clades represented in the Chicago collection. Maximum clade credibility tree where branch colors represent the most probable geographical location of their descendent node inferred through Bayesian reconstruction of the ancestral state. The width of the branches represents their posterior probability. X-axis corresponds to the date in decimal years. **a)** Clade 1, **b)** Clade 2, **c)** Clade 3.

**Supplemental Figure 4**




ML phylogenetic tree of all available global SARS-CoV-2 genomes. Clades most closely related to the three major clades in the Chicago collection are highlighted. We included sequences from Northwestern and global sequences from GISAID. Branches are colored by location and tips corresponding to Northwestern sequences are highlighted. Well-supported clades of the tree that include our defined Chicago clades are indicated.

**Supplemental Figure 5**




Comparisons of sample source and time to symptom onset by clade. **a)** PCR Cycle threshold (Ct) values per sample source grouped by Clades 1 and 2. Significance is indicated for the comparisons performed within each fitted model (* = q-value<0.05; ** = q-value <0.01). **b)** Patient-reported time from onset of COVID-19 symptoms to date of specimen collection, measured in days, among patients infected with Clade 1 or Clade 2 viruses. P-values for Wilcoxon tests between the groups are shown. **c)** Time from self-reported known or potential exposure to SARS-CoV-2-infected individual to date of specimen collection, measured in days, among patients infected with Clade 1 (n = 24) or Clade 2 (n = 5) virus. P-values for Wilcoxon tests between the groups are shown. For **a)** through **c)**, the horizontal line in each box represents the median value and lower and upper error bars are the interquartile ranges. **d)** Correlation of times from onset of COVID-19 symptoms to Ct values at time of specimen collection. Lines represent the best fit regression lines for the values for patients infected with each Clade (Clade 1: blue; Clade 2: pink). Rho and p-values are indicated in the figure.

**Supplemental Table 1** Accession numbers for the consensus SARS-CoV-2 genome sequences deposited in this study in the GISAID database with Pangolin lineage assignments.

| ID | GISAID Accession | GISAID ID | Clade | Lineage |
|---|---|---|---|---|
| NM-nCoV-006 | EPI_ISL_444520 | hCoV-19/USA/IL-NM06/2020 | Clade 1 | B.1 |
| NM-nCoV-008 | EPI_ISL_444521 | hCoV-19/USA/IL-NM08/2020 | Clade 1 | B.1 |
| NM-nCoV-009 | EPI_ISL_444522 | hCoV-19/USA/IL-NM09/2020 | Clade 3 | A.1 |
| NM-nCoV-010 | EPI_ISL_444523 | hCoV-19/USA/IL-NM010/2020 | Clade 1 | B.1 |
| NM-nCoV-012 | EPI_ISL_444524 | hCoV-19/USA/IL-NM012/2020 | Clade 1 | B.1 |
| NM-nCoV-014 | EPI_ISL_444525 | hCoV-19/USA/IL-NM014/2020 | Clade 1 | B.1.5 |
| NM-nCoV-016 | EPI_ISL_444526 | hCoV-19/USA/IL-NM016/2020 | Clade 3 | A.1 |
| NM-nCoV-017 | EPI_ISL_444527 | hCoV-19/USA/IL-NM017/2020 | Clade 1 | B.1 |
| NM-nCoV-018 | EPI_ISL_444528 | hCoV-19/USA/IL-NM018/2020 | Clade 1 | B.1.1 |
| NM-nCoV-019 | EPI_ISL_444529 | hCoV-19/USA/IL-NM019/2020 | Clade 1 | B.1.5 |
| NM-nCoV-020 | EPI_ISL_444530 | hCoV-19/USA/IL-NM020/2020 | Clade 2 | A.3 |
| NM-nCoV-021 | EPI_ISL_444531 | hCoV-19/USA/IL-NM021/2020 | Clade 1 | B.1 |
| NM-nCoV-022 | EPI_ISL_444532 | hCoV-19/USA/IL-NM022/2020 | Clade 3 | A.1 |
| NM-nCoV-023 | EPI_ISL_444533 | hCoV-19/USA/IL-NM023/2020 | Clade 1 | B.1 |
| NM-nCoV-024 | EPI_ISL_444534 | hCoV-19/USA/IL-NM024/2020 | Clade 1 | B.1 |
| NM-nCoV-025 | EPI_ISL_444535 | hCoV-19/USA/IL-NM025/2020 | Clade 1 | B.1 |
| NM-nCoV-026 | EPI_ISL_444536 | hCoV-19/USA/IL-NM026/2020 | Clade 1 | B.1 |
| NM-nCoV-027 | EPI_ISL_444537 | hCoV-19/USA/IL-NM027/2020 | Clade 1 | B.1 |
| NM-nCoV-029 | EPI_ISL_444538 | hCoV-19/USA/IL-NM029/2020 | Clade 1 | B.1 |
| NM-nCoV-030 | EPI_ISL_444539 | hCoV-19/USA/IL-NM030/2020 | Clade 2 | A.3 |
| NM-nCoV-032 | EPI_ISL_444540 | hCoV-19/USA/IL-NM032/2020 | Clade 1 | B.1 |
| NM-nCoV-033 | EPI_ISL_444541 | hCoV-19/USA/IL-NM033/2020 | Clade 1 | B.1 |
| NM-nCoV-038 | EPI_ISL_444542 | hCoV-19/USA/IL-NM038/2020 | Clade 1 | B.1 |
| NM-nCoV-043 | EPI_ISL_444543 | hCoV-19/USA/IL-NM043/2020 | Clade 2 | A.3 |
| NM-nCoV-045 | EPI_ISL_444544 | hCoV-19/USA/IL-NM045/2020 | Clade 1 | B.1 |
| NM-nCoV-047 | EPI_ISL_444545 | hCoV-19/USA/IL-NM047/2020 | Clade 1 | B.1 |
| NM-nCoV-049 | EPI_ISL_444546 | hCoV-19/USA/IL-NM049/2020 | Clade 1 | B.1 |
| NM-nCoV-050 | EPI_ISL_444547 | hCoV-19/USA/IL-NM050/2020 | Clade 1 | B.1 |
| NM-nCoV-051 | EPI_ISL_444548 | hCoV-19/USA/IL-NM051/2020 | Clade 1 | B.1 |
| NM-nCoV-053 | EPI_ISL_444549 | hCoV-19/USA/IL-NM053/2020 | Clade 2 | A.3 |
| NM-nCoV-054 | EPI_ISL_444550 | hCoV-19/USA/IL-NM054/2020 | Clade 1 | B.1 |
| NM-nCoV-055 | EPI_ISL_444551 | hCoV-19/USA/IL-NM055/2020 | Clade 3 | A |
| NM-nCoV-056 | EPI_ISL_444552 | hCoV-19/USA/IL-NM056/2020 | Clade 1 | B.1 |
| NM-nCoV-057 | EPI_ISL_444553 | hCoV-19/USA/IL-NM057/2020 | Clade 1 | B.1 |
| NM-nCoV-058 | EPI_ISL_444554 | hCoV-19/USA/IL-NM058/2020 | Clade 3 | A.1 |
| NM-nCoV-059 | EPI_ISL_444555 | hCoV-19/USA/IL-NM059/2020 | Clade 2 | A.3 |
| NM-nCoV-060 | EPI_ISL_444556 | hCoV-19/USA/IL-NM060/2020 | Clade 1 | B.1 |
| NM-nCoV-061 | EPI_ISL_444557 | hCoV-19/USA/IL-NM061/2020 | Clade 2 | A.3 |
| NM-nCoV-062 | EPI_ISL_444558 | hCoV-19/USA/IL-NM062/2020 | Clade 1 | B.1 |
| NM-nCoV-064 | EPI_ISL_444559 | hCoV-19/USA/IL-NM064/2020 | Clade 2 | A.3 |
| NM-nCoV-065 | EPI_ISL_444560 | hCoV-19/USA/IL-NM065/2020 | Other | B |
| NM-nCoV-066 | EPI_ISL_444561 | hCoV-19/USA/IL-NM066/2020 | Clade 2 | A.3 |
| NM-nCoV-067 | EPI_ISL_444562 | hCoV-19/USA/IL-NM067/2020 | Clade 2 | A.3 |
| NM-nCoV-068 | EPI_ISL_444563 | hCoV-19/USA/IL-NM068/2020 | Clade 2 | A.3 |
| NM-nCoV-069 | EPI_ISL_444564 | hCoV-19/USA/IL-NM069/2020 | Clade 2 | A.3 |
| NM-nCoV-070 | EPI_ISL_444565 | hCoV-19/USA/IL-NM070/2020 | Clade 2 | A.3 |
| NM-nCoV-071 | EPI_ISL_444566 | hCoV-19/USA/IL-NM071/2020 | Clade 2 | A.3 |
| NM-nCoV-072 | EPI_ISL_444567 | hCoV-19/USA/IL-NM072/2020 | Clade 2 | A.3 |
| NM-nCoV-073 | EPI_ISL_444568 | hCoV-19/USA/IL-NM073/2020 | Other | A |
| NM-nCoV-074 | EPI_ISL_444569 | hCoV-19/USA/IL-NM074/2020 | Clade 2 | A.3 |
| NM-nCoV-075 | EPI_ISL_444570 | hCoV-19/USA/IL-NM075/2020 | Clade 1 | B.1 |
| NM-nCoV-079 | EPI_ISL_444571 | hCoV-19/USA/IL-NM079/2020 | Clade 1 | B.1 |
| NM-nCoV-081 | EPI_ISL_444572 | hCoV-19/USA/IL-NM081/2020 | Clade 2 | A.3 |
| NM-nCoV-083 | EPI_ISL_444574 | hCoV-19/USA/IL-NM083/2020 | Clade 1 | B.1 |
| NM-nCoV-084 | EPI_ISL_444575 | hCoV-19/USA/IL-NM084/2020 | Other | B |
| NM-nCoV-085 | EPI_ISL_444576 | hCoV-19/USA/IL-NM085/2020 | Clade 1 | B.1 |
| NM-nCoV-088 | EPI_ISL_444578 | hCoV-19/USA/IL-NM088/2020 | Clade 1 | B.1 |
| NM-nCoV-089 | EPI_ISL_444579 | hCoV-19/USA/IL-NM089/2020 | Clade 1 | B.1 |
| NM-nCoV-090 | EPI_ISL_444580 | hCoV-19/USA/IL-NM090/2020 | Clade 1 | B.1.5 |
| NM-nCoV-091 | EPI_ISL_444581 | hCoV-19/USA/IL-NM091/2020 | Clade 1 | B.1 |
| NM-nCoV-093 | EPI_ISL_444582 | hCoV-19/USA/IL-NM093/2020 | Clade 1 | B.1 |
| NM-nCoV-095 | EPI_ISL_444583 | hCoV-19/USA/IL-NM095/2020 | Other | A |

| | | | | |
|---|---|---|---|---|
| NM-nCoV-096 | EPI_ISL_444584 | hCoV-19/USA/IL-NM096/2020 | Clade 1 | B.1 |
| NM-nCoV-097 | EPI_ISL_444585 | hCoV-19/USA/IL-NM097/2020 | Clade 2 | A |
| NM-nCoV-098 | EPI_ISL_444586 | hCoV-19/USA/IL-NM098/2020 | Clade 1 | B.1 |
| NM-nCoV-101 | EPI_ISL_444587 | hCoV-19/USA/IL-NM0101/2020 | Clade 1 | B.1 |
| NM-nCoV-103 | EPI_ISL_444588 | hCoV-19/USA/IL-NM0103/2020 | Clade 2 | A |
| NM-nCoV-104 | EPI_ISL_444589 | hCoV-19/USA/IL-NM0104/2020 | Clade 1 | B.1 |
| NM-nCoV-105 | EPI_ISL_444590 | hCoV-19/USA/IL-NM0105/2020 | Other | A |
| NM-nCoV-106 | EPI_ISL_444591 | hCoV-19/USA/IL-NM0106/2020 | Clade 1 | B.1 |
| NM-nCoV-107 | EPI_ISL_444592 | hCoV-19/USA/IL-NM0107/2020 | Clade 1 | B.1 |
| NM-nCoV-108 | EPI_ISL_444593 | hCoV-19/USA/IL-NM0108/2020 | Clade 3 | A.1 |
| NM-nCoV-109 | EPI_ISL_444594 | hCoV-19/USA/IL-NM0109/2020 | Clade 2 | A.3 |
| NM-nCoV-111 | EPI_ISL_444595 | hCoV-19/USA/IL-NM0111/2020 | Clade 1 | B.1 |
| NM-nCoV-112 | EPI_ISL_444596 | hCoV-19/USA/IL-NM0112/2020 | Clade 2 | A.3 |
| NM-nCoV-113 | EPI_ISL_444597 | hCoV-19/USA/IL-NM0113/2020 | Clade 2 | A.3 |
| NM-nCoV-114 | EPI_ISL_444598 | hCoV-19/USA/IL-NM0114/2020 | Clade 1 | B.1 |
| NM-nCoV-116 | EPI_ISL_444599 | hCoV-19/USA/IL-NM0116/2020 | Clade 1 | B.1 |
| NM-nCoV-117 | EPI_ISL_444600 | hCoV-19/USA/IL-NM0117/2020 | Clade 1 | B.1 |
| NM-nCoV-118 | EPI_ISL_444601 | hCoV-19/USA/IL-NM0118/2020 | Clade 3 | A.1 |
| NM-nCoV-119 | EPI_ISL_444602 | hCoV-19/USA/IL-NM0119/2020 | Clade 2 | A.3 |
| NM-nCoV-120 | EPI_ISL_444603 | hCoV-19/USA/IL-NM0120/2020 | Clade 2 | A.3 |
| NM-nCoV-121 | EPI_ISL_444604 | hCoV-19/USA/IL-NM0121/2020 | Clade 2 | A.3 |
| NM-nCoV-122 | EPI_ISL_444605 | hCoV-19/USA/IL-NM0122/2020 | Clade 1 | B.1 |
| NM-nCoV-124 | EPI_ISL_444606 | hCoV-19/USA/IL-NM0124/2020 | Other | A.4 |
| NM-nCoV-125 | EPI_ISL_444607 | hCoV-19/USA/IL-NM0125/2020 | Clade 1 | B.1 |
| NM-nCoV-126 | EPI_ISL_444608 | hCoV-19/USA/IL-NM0126/2020 | Clade 1 | B.1 |
| NM-nCoV-127 | EPI_ISL_444609 | hCoV-19/USA/IL-NM0127/2020 | Clade 1 | B.1 |

**Supplemental Table 2** Summary of the clinical and demographic data for all patient samples analyzed in this study. P-values represent the results of Chi-squared tests for categorical explanatory variables and F-tests for continuous variables.

| Dependent: Clade | | Clade 1 | Clade 2 | Clade 3 | Others | p |
|---|---|---|---|---|---|---|
| Total N (%) | | 51 (58.0) | 24 (27.3) | 7 (8.0) | 6 (6.8) | |
| Ct Value | Mean (SD) | 21.4 (3.9) | 24.4 (3.8) | 25.6 (5.7) | 23.1 (4.2) | 0.006 |
| Isolation Date | 2020-03-14 | 7 (13.7) | 1 (4.2) | 1 (14.3) | | 0.172 |
| | 2020-03-15 | 6 (11.8) | 2 (8.3) | 2 (28.6) | | |
| | 2020-03-16 | 8 (15.7) | 3 (12.5) | | | |
| | 2020-03-17 | 4 (7.8) | | 1 (14.3) | 1 (16.7) | |
| | 2020-03-18 | 5 (9.8) | 10 (41.7) | 1 (14.3) | 1 (16.7) | |
| | 2020-03-19 | 10 (19.6) | 2 (8.3) | | 2 (33.3) | |
| | 2020-03-20 | 4 (7.8) | 2 (8.3) | 1 (14.3) | 2 (33.3) | |
| | 2020-03-21 | 7 (13.7) | 4 (16.7) | 1 (14.3) | | |
| Isolation Source | ED | 26 (51.0) | 12 (50.0) | 3 (42.9) | 2 (33.3) | 0.852 |
| | Inpatient | 12 (23.5) | 5 (20.8) | 3 (42.9) | 2 (33.3) | |
| | ICU | 5 (9.8) | 3 (12.5) | 1 (14.3) | | |
| | Other | 8 (15.7) | 4 (16.7) | | 2 (33.3) | |
| Severity Group | Mild | 25 (49.0) | 10 (41.7) | 3 (42.9) | 3 (50.0) | 0.715 |
| | Moderate | 14 (27.5) | 10 (41.7) | 1 (14.3) | 2 (33.3) | |
| | Severe | 12 (23.5) | 4 (16.7) | 3 (42.9) | 1 (16.7) | |
| Source | BAL | 4 (7.8) | 2 (8.3) | 1 (14.3) | 1 (16.7) | 0.859 |
| | NP | 47 (92.2) | 22 (91.7) | 6 (85.7) | 5 (83.3) | |
| Race | White | 26 (54.2) | 10 (41.7) | 4 (57.1) | 4 (66.7) | 0.591 |
| | Black or African American | 10 (20.8) | 9 (37.5) | 2 (28.6) | 2 (33.3) | |
| | Asian | 4 (8.3) | 3 (12.5) | | | |
| | Other | 7 (14.6) | 1 (4.2) | | | |
| | Declined | 1 (2.1) | 1 (4.2) | 1 (14.3) | | |
| Sex | F | 22 (45.8) | 11 (45.8) | 3 (42.9) | 3 (50.0) | 0.995 |
| | M | 26 (54.2) | 13 (54.2) | 4 (57.1) | 3 (50.0) | |
| Age | Mean (SD) | 52.2 (18.6) | 57.2 (16.4) | 58.1 (14.8) | 39.4 (14.9) | 0.137 |

**Supplemental Table 3** Logistic regression model of clinical, demographic, and virologic characteristics for Clade 1 and Clade 2 viruses.

| Characteristic | OR[1] | 95% CI[1] | p-value |
|---|---|---|---|
| **Severity Group** | | | |
| *Mild* | — | — | |
| *Moderate* | 1.13 | 0.22, 5.63 | 0.9 |
| *Severe* | 0.12 | 0.01, 1.13 | 0.092 |
| **Ct Value** | 1.37 | 1.12, 1.73 | **0.005** |
| **Sex** | | | |
| *F* | — | — | |
| *M* | 3.30 | 0.84, 15.5 | 0.10 |
| **Age** | 1.01 | 0.97, 1.06 | 0.6 |
| **Race** | | | |
| *White* | — | — | |
| *Asian* | 0.93 | 0.10, 8.38 | >0.9 |
| *Black or African American* | 3.12 | 0.68, 16.1 | 0.2 |
| *Declined* | 7.12 | 0.10, 891 | 0.4 |
| *Other* | 0.20 | 0.00, 2.73 | 0.3 |
| **Source** | | | |
| *BAL* | — | — | |
| *NP* | 0.08 | 0.00, 1.60 | 0.10 |
| **Days from symptom onset** | 1.21 | 0.92, 1.69 | 0.2 |

[1] OR = Odds Ratio, CI = Confidence Interval