science.sciencemag.org/cgi/content/full/science.abb9263/DC1

Supplementary Materials for

Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California

Xianding Deng, Wei Gu, Scot Federman, Louis du Plessis, Oliver G. Pybus, Nuno Faria, Candace Wang, Guixia Yu, Brian Bushnell, Chao-Yang Pan, Hugo Guevara, Alicia Sotomayor-Gonzalez, Kelsey Zorn, Allan Gopez, Venice Servellita, Elaine Hsu, Steve Miller, Trevor Bedford, Alexander L. Greninger, Pavitra Roychoudhury, Lea M. Starita, Michael Famulare, Helen Y. Chu, Jay Shendure, Keith R. Jerome, Catie Anderson, Karthik Gangavarapu, Mark Zeller, Emily Spencer, Kristian G. Andersen, Duncan MacCannell, Clinton R. Paden, Yan Li, Jing Zhang, Suxiang Tong, Gregory Armstrong, Scott Morrow, Matthew Willis, Bela T. Matyas, Sundari Mase, Olivia Kasirye, Maggie Park, Godfred Masinde, Curtis Chan, Alexander T. Yu, Shua J. Chai, Elsa Villarino, Brandon Bonin, Debra A. Wadford, Charles Y. Chiu*

*Corresponding author. Email: charles.chiu@ucsf.edu

Published 8 June 2020 on *Science* First Release
DOI: 10.1126/science.abb9263

This PDF file includes:

Materials and Methods
Figs. S1 and S2
Captions for tables S1 to S5
Caption for data S1
References

Other supplementary material for this manuscript includes:

Tables S1 to S5 (Excel format)
Data S1 (.tree format)

Materials and Methods

Quantitative RT-PCR testing

Quantitative reverse transcription-polymerase chain reaction (RT-PCR) testing for SARS-CoV2 was performed by the US Centers for Disease Control and Prevention (CDC), California Department of Public Health (CDPH), University of California, San Francisco Clinical Microbiology Laboratory (UCSF), or Santa Clara County Public Health Department (SCCPHD). At UCSF, samples were extracted on a Magna Pure 24 Viral Kit (Roche Diagnostics, Indianapolis, USA) or an EZ1 Viral Mini Kit (Qiagen, Hilden, Germany). At the CDC, CDPH, or SCCPHD, samples were extracted using Qiagen DSP Viral RNA Mini Kit with carrier RNA added (Qiagen). Samples were tested for SARS-CoV-2 using the FDA EUA-approved 2019-nCoV CDC Real-Time RT-PCR Diagnostic Panel assay (*16*), which targets the N1 and N2 regions of the nucleoprotein gene. The cutoff for a confirmed positive sample was determined to be a cycle threshold (C_t) value of 40 cycles, with a positive result requiring detection of both the N1 and N2 targeted regions of the nucleoprotein gene.

Clinical SARS-CoV-2 samples

Nasopharyngeal and/or oropharyngeal swabs in universal transport media (Copan Diagnostics, Murrieta, CA, USA) from RT-PCR positive COVID-19 patients were obtained from the UCSF Clinical Microbiology Laboratory, CDPH, and SCCPHD. Patient samples were randomly selected for broad representation of the 9 counties in Northern California and the Grand Princess cruise ship.

Next-generation sequencing library preparation and PCR confirmation of the G29711T SNV

Extracted RNA from nasopharyngeal (NP) swab samples in universal transport medium was reverse transcribed to complementary DNA (cDNA) using the MSSPE method as previously described (*15*). The custom-designed 13-nucleotide (nt) SARS-CoV-2 primers (IDT Technologies) were constructed using an alignment of 30 SARS-CoV-2 genome references available in the NCBI GenBank database as of end of February 2020 **(Table S3).** Barcoded sequencing libraries were constructed from cDNA by Nextera tagmentation (Ilumina), followed by 12 cycles of amplification per the manufacturer's protocol. A second round of amplification (14 cycles) was used to boost the yield of cDNA library due to low RNA yield in general from low-input samples such as NP swabs (*33*). The final libraries were pooled and sequenced on the MiSeq, NextSeq, or HiSeq 1500 (Illumina Inc., San Diego, USA) as 1x150 single-end or 2x150 paired-end reads. For some genomes with low coverage (<50%), tiling amplicon PCR using SARS-CoV-2 primers generated from the Primal algorithm **(Table S4)** was used to increase genome coverage according to the published protocol (*17*). To confirm that UC15 and UC25 belong to the SCC1 lineage, we performed specific PCR using designed primers to amplify a region containing the lineage defining G29711T SNV, followed by Sanger sequencing to confirm the presence of this SNV **(Table S5).**

Phylogenetic analysis

Raw reads were first screened via BLASTn (BLAST+ package 2.9.0) (*34*) for alignment to SARS-CoV-2 reference genome NC_045512. They were then aligned to the reference genome with LASTZ version 1.04.03. For libraries generated using the MSSPE protocol (*15*), single-end

reads were trimmed using Geneious version 11.1.3 by removal of 13 nucleotides (nt) (the length of the MSSPE primer) and low-quality reads from the ends, followed by removal of duplicate reads. Trimmed reads were then mapped to reference genome NC_045512 in Geneious with no gaps allowed and a maximum of 5% mismatches per read. The assembled contig was then manually annotated, and a consensus genome was generated using a majority threshold criterion. For libraries generated using tiling multiplex PCR (Primal) protocol (*17*), paired-end reads were processed using the BBTools suite, version 38.82 (https://sourceforge.net/projects/bbmap/). Quality scores were recalibrated, and reads were trimmed for removal of adapters and low-quality sequences using BBDuk. Subsequently, mate pairs were aligned with BBMap, and duplicate pairs removed on the basis of mapping to the same nucleotide positions. Mapped reads containing deletion events or multiple (>2) substitution events unsupported by any other read were removed to reduce overall noise, and soft clipping was applied to reduce the impact specifically of artifactual noise near read ends. Variants were called with CallVariants and applied to the genome using a depth cutoff of 3 to yield the final assembly. Regions with depth of less than 3 were represented by N.

We downloaded all 762 complete (>29,000 bp), high-coverage SARS-CoV-2 genomes that had been deposited into GISAID as of March 20, 2020 (*31, 32*), and then added the 36 genomes in the current study to generate an initial dataset of 798 genomes. This dataset was then trimmed by removal of low-quality genomes and known duplicate genomes from the same patient, yielding a final dataset of 789 genomes. Sequences were aligned using MAFFT v7.427 (*35*) using default settings and multiple sequence alignments were manually curated for accuracy. Phylogenetic maximum likelihood trees were constructed in PhyML v3.3 (*36*) under an HKY+Γ_4 substitution model (*37, 38*), after trimming the ends of the alignment. To compute

branch support values we used the aLRT (approximate likelihood ratio test) method, implemented in PhyML (*39*), which is better suited than bootstrapping to data sets with few informative sites per taxon.

For the separate WA1 lineage phylogenetic analysis in **Figure 2B**, we discarded all genomes of the WA1 lineage, leaving 88 genomes. We further removed 4773 nucleotide sites from the alignment that were either ambiguous or unknown in more than one genome. As above, the phylogenetic WA1 lineage subtree was constructed in PhyML v3.3 under an HKY+Γ_4 substitution model. The locations of SNVs and gaps shown in **Figure 3** were extracted from an alignment of 154 sequences from the US, aligned to the reference sequence (NC_045512), using custom scripts (*30*).

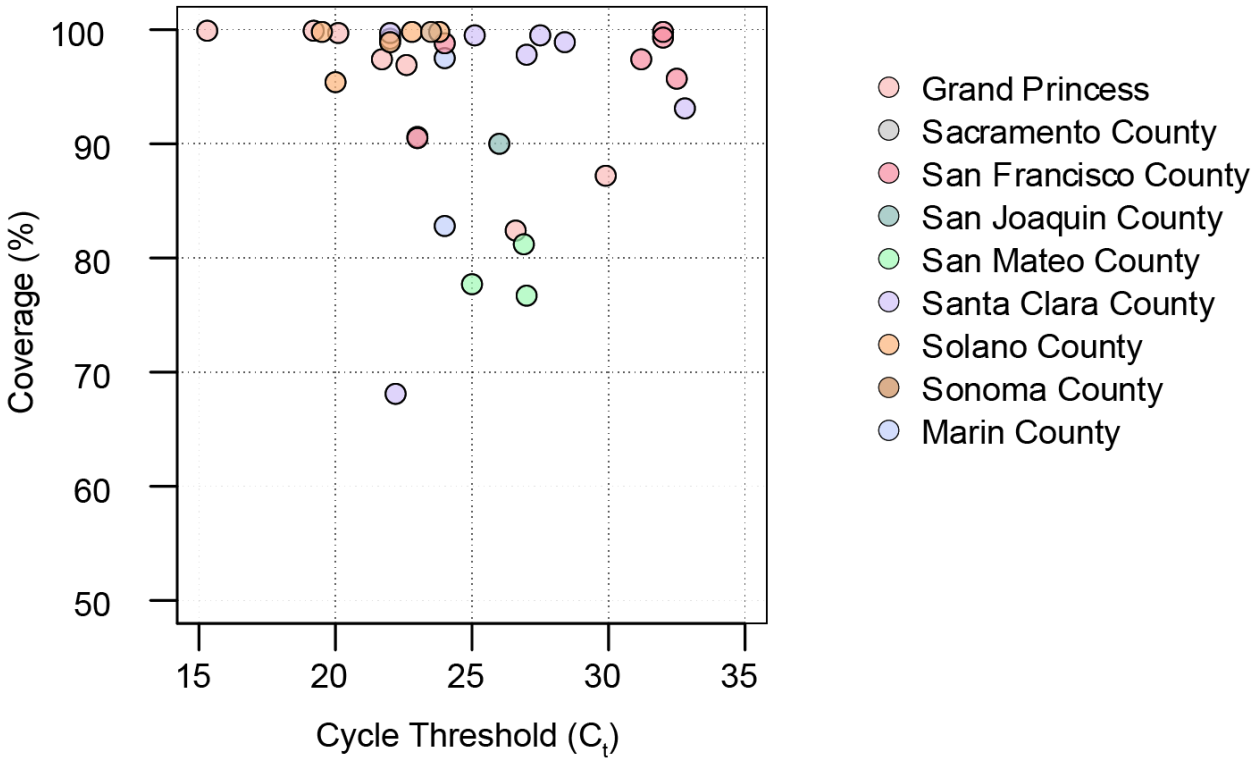

Fig. S1. SARS-CoV2 genome coverage versus cycle threshold for respiratory swab samples from COVID-19 patients (n=34) in this study. The x-axis designates the virus titer based on the reported real-time RT-PCR C_t value (higher values correspond to lower viral loads), while the y-axis designates genome coverage in percentage. The data points are color coded to designate either the patient's county of residence in Northern California or whether the patient was a passenger or crew member aboard the Grand Princess cruise ship.

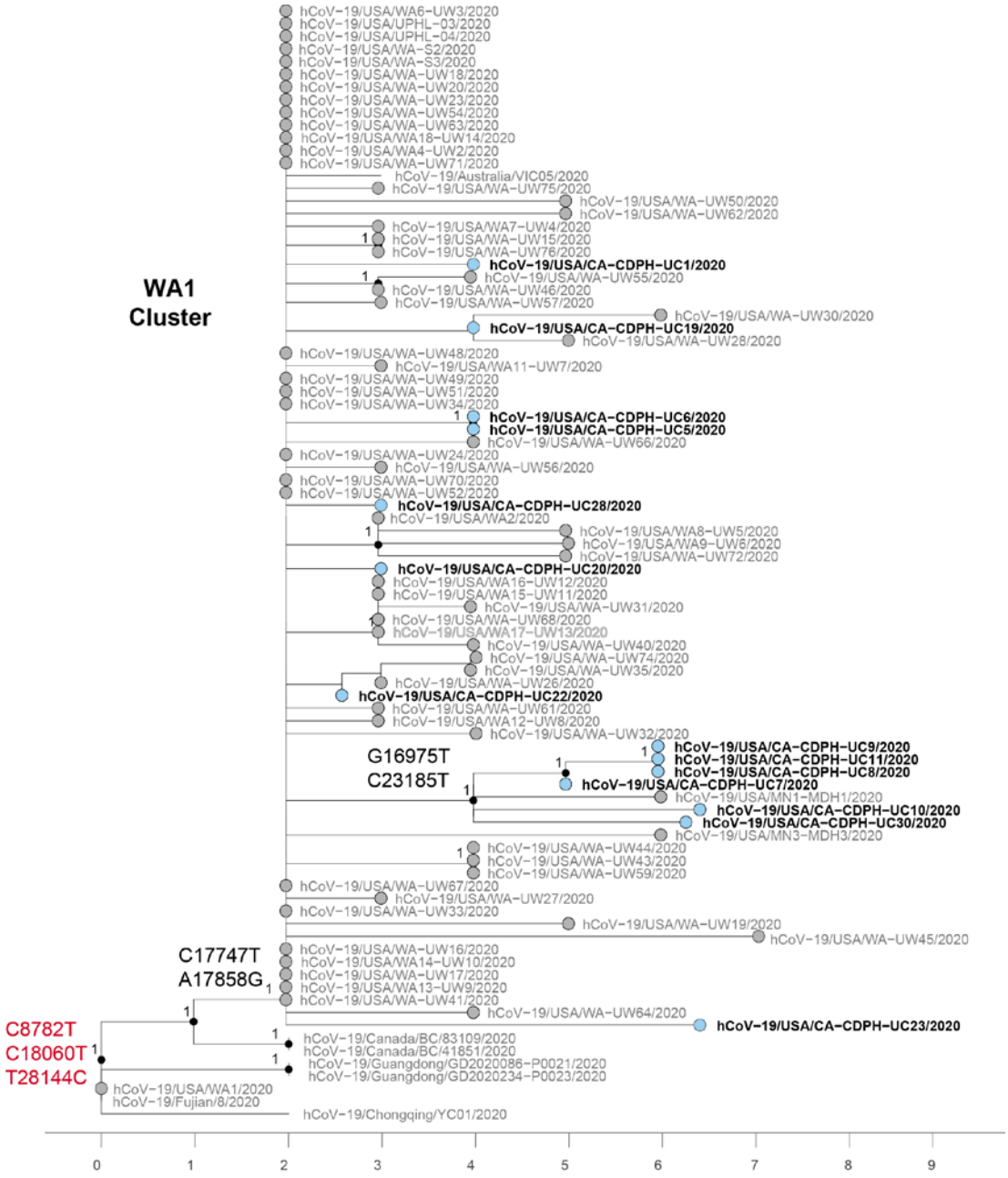

Fig. S2. Initial phylogeny of the WA1 lineage pruned from the global phylogenetic tree of 789 SARS-CoV-2 genomes. Light blue circles denote the genome sequences obtained in this study, while gray circles denote sequences from Washington State (WA). The 3 key SNVs defining the WA1 lineage are highlighted in red, while an additional 2 SNVs found in viral genomes from the Grand Princess passengers and crew as well as the majority of the WA1 lineage viruses are highlighted in black. The WA1 virus from Washington State (first reported COVID-19 case in the US) is positioned at the root of the tree and is closely related to Fujian and Chongqing viruses from China. Bootstrap values (converted from the approximate likelihood ratio test, or aLRT score) are displayed at each node, with a value of 1 indicating 100% support.

Table S1. (separate file in Excel format) List of all 62 SARS-CoV-2 samples analyzed in the study along with sequencing controls.

Table S2. (separate file in Excel format) Patient epidemiologic metadata and sequencing metrics for 36 SARS-CoV-2 genomes recovered from COVID-19 infected patients.

Table S3. (separate file in Excel format) Custom primer sequences for the MSSPE method.

Table S4. (separate file in Excel format) Tiling multiplex PCR primers.

Table S5. (separate file in Excel format) PCR primers and Sanger sequencing reads obtained for confirmation of SNV G29711T from SARS-CoV-2 respiratory nasal swab samples.

Data S1. (separate file in Nexus tree format) Tree file for the global phylogeny of 789 SARS-CoV-2 genomes.

References

1. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020). doi:10.1016/S1473-3099(20)30120-1 Medline

2. J. F. Chan, S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C.-Y. Yip, R. W.-S. Poon, H.-W. Tsoi, S. K.-F. Lo, K.-H. Chan, V. K.-M. Poon, W.-M. Chan, J. D. Ip, J.-P. Cai, V. C.-C. Cheng, H. Chen, C. K.-M. Hui, K.-Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **395**, 514–523 (2020). doi:10.1016/S0140-6736(20)30154-9 Medline

3. R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020). doi:10.1126/science.abb3221 Medline

4. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* **395**, 565–574 (2020). doi:10.1016/S0140-6736(20)30251-8 Medline

5. A. Otani, P. Santilli, The week that wiped $3.6 trillion off the stock market. *Wall Street Journal*, 29 February 2020.

6. T. Bedford *et al*., Cryptic transmission of SARS-CoV-2 in Washington State. medRxiv 20051417 [preprint]. 16 April 2020.

7. M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural, G. Diaz, A. Cohn, L. Fox, A. Patel, S. I. Gerber, L. Kim, S. Tong, X. Lu, S. Lindstrom, M. A. Pallansch, W. C. Weldon, H. M. Biggs, T. M. Uyeki, S. K. Pillai, First Case of 2019 Novel Coronavirus in the United States. *N. Engl. J. Med.* **382**, 929–936 (2020). doi:10.1056/NEJMoa2001191 Medline

8. C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons, T. Jombart, W. R. Hinsley, N. C. Grassly, F. Balloux, A. C. Ghani, N. M. Ferguson, A. Rambaut, O. G. Pybus, H. Lopez-Gatell, C. M. Alpuche-Aranda, I. B. Chapela, E. P. Zavala, D. M. E. Guevara, F. Checchi, E. Garcia, S. Hugonnet, C. Roth, WHO Rapid Pandemic Assessment Collaboration, Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* **324**, 1557–1561 (2009). doi:10.1126/science.1176062 Medline

9. J. L. Gardy, N. J. Loman, Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* **19**, 9–20 (2018). doi:10.1038/nrg.2017.88 Medline

10. R. A. Urbanowicz, C. P. McClure, A. Sakuntabhai, A. A. Sall, G. Kobinger, M. A. Müller, E. C. Holmes, F. A. Rey, E. Simon-Loriere, J. K. Ball, Human Adaptation of Ebola Virus during the West African Outbreak. *Cell* **167**, 1079–1087.e5 (2016). doi:10.1016/j.cell.2016.10.013 Medline

11. S. E. Mate, J. R. Kugelman, T. G. Nyenswah, J. T. Ladner, M. R. Wiley, T. Cordier-Lassalle, A. Christie, G. P. Schroth, S. M. Gross, G. J. Davies-Wayne, S. A. Shinde, R. Murugan, S. B. Sieh, M. Badio, L. Fakoli, F. Taweh, E. de Wit, N. van Doremalen, V. J. Munster, J. Pettitt, K. Prieto, B. W. Humrighouse, U. Ströher, J. W. DiClaro, L. E. Hensley, R. J. Schoepp, D. Safronetz, J. Fair, J. H. Kuhn, D. J. Blackley, A. S. Laney, D. E. Williams, T. Lo, A. Gasasira, S. T. Nichol, P. Formenty, F. N. Kateh, K. M. De Cock, F. Bolay, M. Sanchez-Lockhart, G. Palacios, Molecular Evidence of Sexual Transmission of Ebola Virus. *N. Engl. J. Med.* **373**, 2448–2454 (2015). doi:10.1056/NEJMoa1509773 Medline

12. N. R. Faria, J. Quick, I. M. Claro, J. Thézé, J. G. de Jesus, M. Giovanetti, M. U. G. Kraemer, S. C. Hill, A. Black, A. C. da Costa, L. C. Franco, S. P. Silva, C.-H. Wu, J. Raghwani, S. Cauchemez, L. du Plessis, M. P. Verotti, W. K. de Oliveira, E. H. Carmo, G. E. Coelho, A. C. F. S. Santelli, L. C. Vinhal, C. M. Henriques, J. T. Simpson, M. Loose, K. G. Andersen, N. D. Grubaugh, S. Somasekar, C. Y. Chiu, J. E. Muñoz-Medina, C. R. Gonzalez-Bonilla, C. F. Arias, L. L. Lewis-Ximenez, S. A. Baylis, A. O. Chieppe, S. F. Aguiar, C. A. Fernandes, P. S. Lemos, B. L. S. Nascimento, H. A. O. Monteiro, I. C. Siqueira, M. G. de Queiroz, T. R. de Souza, J. F. Bezerra, M. R. Lemos, G. F. Pereira, D. Loudal, L. C. Moura, R. Dhalia, R. F. França, T. Magalhães, E. T. Marques Jr., T. Jaenisch, G. L. Wallau, M. C. de Lima, V. Nascimento, E. M. de Cerqueira, M. M. de Lima, D. L. Mascarenhas, J. P. M. Neto, A. S. Levin, T. R. Tozetto-Mendoza, S. N. Fonseca, M. C. Mendes-Correa, F. P. Milagres, A. Segurado, E. C. Holmes, A. Rambaut, T. Bedford, M. R. T. Nunes, E. C. Sabino, L. C. J. Alcantara, N. J. Loman, O. G. Pybus, Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* **546**, 406–410 (2017). doi:10.1038/nature22401 Medline

13. J. Thézé, T. Li, L. du Plessis, J. Bouquet, M. U. G. Kraemer, S. Somasekar, G. Yu, M. de Cesare, A. Balmaseda, G. Kuan, E. Harris, C. H. Wu, M. A. Ansari, R. Bowden, N. R. Faria, S. Yagi, S. Messenger, T. Brooks, M. Stone, E. M. Bloch, M. Busch, J. E. Muñoz-Medina, C. R. González-Bonilla, S. Wolinsky, S. López, C. F. Arias, D. Bonsall, C. Y. Chiu, O. G. Pybus, Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico. *Cell Host Microbe* **23**, 855–864.e7 (2018). doi:10.1016/j.chom.2018.04.017 Medline

14. N. D. Grubaugh, J. T. Ladner, M. U. G. Kraemer, G. Dudas, A. L. Tan, K. Gangavarapu, M. R. Wiley, S. White, J. Thézé, D. M. Magnani, K. Prieto, D. Reyes, A. M. Bingham, L. M. Paul, R. Robles-Sikisaka, G. Oliveira, D. Pronty, C. M. Barcellona, H. C. Metsky, M. L. Baniecki, K. G. Barnes, B. Chak, C. A. Freije, A. Gladden-Young, A. Gnirke, C. Luo, B. MacInnis, C. B. Matranga, D. J. Park, J. Qu, S. F. Schaffner, C. Tomkins-Tinch, K. L.

West, S. M. Winnicki, S. Wohl, N. L. Yozwiak, J. Quick, J. R. Fauver, K. Khan, S. E. Brent, R. C. Reiner Jr., P. N. Lichtenberger, M. J. Ricciardi, V. K. Bailey, D. I. Watkins, M. R. Cone, E. W. Kopp 4th, K. N. Hogan, A. C. Cannons, R. Jean, A. J. Monaghan, R. F. Garry, N. J. Loman, N. R. Faria, M. C. Porcelli, C. Vasquez, E. R. Nagle, D. A. T. Cummings, D. Stanek, A. Rambaut, M. Sanchez-Lockhart, P. C. Sabeti, L. D. Gillis, S. F. Michael, T. Bedford, O. G. Pybus, S. Isern, G. Palacios, K. G. Andersen, Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* **546**, 401–405 (2017). doi:10.1038/nature22400 Medline

15. X. Deng, A. Achari, S. Federman, G. Yu, S. Somasekar, I. Bártolo, S. Yagi, P. Mbala-Kingebeni, J. Kapetshi, S. Ahuka-Mundeke, J. J. Muyembe-Tamfum, A. A. Ahmed, V. Ganesh, M. Tamhankar, J. L. Patterson, N. Ndembi, D. Mbanya, L. Kaptue, C. McArthur, J. E. Muñoz-Medina, C. R. Gonzalez-Bonilla, S. López, C. F. Arias, S. Arevalo, S. Miller, M. Stone, M. Busch, K. Hsieh, S. Messenger, D. A. Wadford, M. Rodgers, G. Cloherty, N. R. Faria, J. Thézé, O. G. Pybus, Z. Neto, J. Morais, N. Taveira, J. R. Hackett Jr., C. Y. Chiu, Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat. Microbiol.* **5**, 443–454 (2020). doi:10.1038/s41564-019-0637-9 Medline

16. Centers for Disease Control and Prevention, CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel, Revision 03 (30 March 2020); fda.gov/media/134922/download.

17. J. Quick, N. D. Grubaugh, S. T. Pullan, I. M. Claro, A. D. Smith, K. Gangavarapu, G. Oliveira, R. Robles-Sikisaka, T. F. Rogers, N. A. Beutler, D. R. Burton, L. L. Lewis-Ximenez, J. G. de Jesus, M. Giovanetti, S. C. Hill, A. Black, T. Bedford, M. W. Carroll, M. Nunes, L. C. Alcantara Jr., E. C. Sabino, S. A. Baylis, N. R. Faria, M. Loose, J. T. Simpson, O. G. Pybus, K. G. Andersen, N. J. Loman, Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017). doi:10.1038/nprot.2017.066 Medline

18. C. Y. Chiu, S. A. Miller, Clinical metagenomics. *Nat. Rev. Genet.* **20**, 341–355 (2019). doi:10.1038/s41576-019-0113-7 Medline

19. W. Gu, S. Miller, C. Y. Chiu, Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu. Rev. Pathol.* **14**, 319–338 (2019). doi:10.1146/annurev-pathmechdis-012418-012751 Medline

20. A. S. Gonzalez-Reiche, M. M. Hernandez, M. J. Sullivan, B. Ciferri, H. Alshammary, A. Obla, S. Fabre, G. Kleiner, J. Polanco, Z. Khan, B. Alburquerque, A. van de Guchte, J. Dutta, N. Francoeur, B. S. Melo, I. Oussenko, G. Deikus, J. Soto, S. H. Sridhar, Y.-C. Wang, K. Twyman, A. Kasarskis, D. R. Altman, M. Smith, R. Sebra, J. Aberg, F. Krammer, A. García-Sastre, M. Luksza, G. Patel, A. Paniz-Mondolfi, M. Gitman, E. M. Sordillo, V. Simon, H. van Bakel, Introductions and early spread of SARS-CoV-2 in the

New York City area. *Science* **368**, eabc1917 (2020). doi:10.1126/science.abc1917 Medline

21. M. T. Maurano *et al*., Sequencing identifies multiple, early introductions of SARS-CoV2 to the New York City Region. medRxiv 20064931 [preprint]. 23 April 2020.

22. A. Brufsky, Distinct Viral Clades of SARS-CoV-2: Implications for Modeling of Viral Spread. *J. Med. Virol.* jmv.25902 (2020). doi:10.1002/jmv.25902 Medline

23. L. F. Moriarty, M. M. Plucinski, B. J. Marston, E. V. Kurbatova, B. Knust, E. L. Murray, N. Pesik, D. Rose, D. Fitter, M. Kobayashi, M. Toda, P. T. Cantey, T. Scheuer, E. S. Halsey, N. J. Cohen, L. Stockman, D. A. Wadford, A. M. Medley, G. Green, J. J. Regan, K. Tardivel, S. White, C. Brown, C. Morales, C. Yen, B. Wittry, A. Freeland, S. Naramore, R. T. Novak, D. Daigle, M. Weinberg, A. Acosta, C. Herzig, B. K. Kapella, K. R. Jacobson, K. Lamba, A. Ishizumi, J. Sarisky, E. Svendsen, T. Blocher, C. Wu, J. Charles, R. Wagner, A. Stewart, P. S. Mead, E. Kurylo, S. Campbell, R. Murray, P. Weidle, M. Cetron, C. R. Friedman, CDC Cruise Ship Response Team, California Department of Public Health COVID-19 Team, Solano County COVID-19 Team, Public Health Responses to COVID-19 Outbreaks on Cruise Ships - Worldwide, February-March 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 347–352 (2020). doi:10.15585/mmwr.mm6912e3 Medline

24. M. Klompas, Coronavirus Disease 2019 (COVID-19): Protecting Hospitals From the Invisible. *Ann. Intern. Med.* **172**, 619–620 (2020). doi:10.7326/M20-0751 Medline

25. J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R. A. Neher, Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018). doi:10.1093/bioinformatics/bty407 Medline

26. J. R. Fauver, M. E. Petrone, E. B. Hodcroft, K. Shioda, H. Y. Ehrlich, A. G. Watts, C. B. F. Vogels, A. F. Brito, T. Alpert, A. Muyombwe, J. Razeq, R. Downing, N. R. Cheemarla, A. L. Wyllie, C. C. Kalinich, I. M. Ott, J. Quick, N. J. Loman, K. M. Neugebauer, A. L. Greninger, K. R. Jerome, P. Roychoudhury, H. Xie, L. Shrestha, M. L. Huang, V. E. Pitzer, A. Iwasaki, S. B. Omer, K. Khan, I. I. Bogoch, R. A. Martinello, E. F. Foxman, M. L. Landry, R. A. Neher, A. I. Ko, N. D. Grubaugh, Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell* **181**, 990–996.e5 (2020). doi:10.1016/j.cell.2020.04.021 Medline

27. M. Bouvet, I. Imbert, L. Subissi, L. Gluais, B. Canard, E. Decroly, RNA 3′-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9372–9377 (2012). doi:10.1073/pnas.1201130109 Medline

28. V. Hill, A. Rambaut, Phylodynamic analysis of SARS-CoV-2. Virological.org (6 March 2020); https://virological.org/t/phylodynamic-analysis-of-sars-cov-2-update-2020-03-06/420.

29. W. Gu, X. Deng, K. Reyes, E. Hsu, C. Wang, A. Sotomayor-Gonzalez, S. Federman, B. Bushnell, S. Miller, C. Chiu, Associations of Early COVID-19 Cases in San Francisco with Domestic and International Travel. *Clin. Infect. Dis.* ciaa599 (2020). doi:10.1093/cid/ciaa599

30. L. du Plessis *et al*., Zenodo DOI:10.5281/zenodo.3779312 (2020).

31. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, 30494 (2017). doi:10.2807/1560-7917.ES.2017.22.13.30494 Medline

32. S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **1**, 33–46 (2017). doi:10.1002/gch2.1018 Medline

33. S. Miller, S. N. Naccache, E. Samayoa, K. Messacar, S. Arevalo, S. Federman, D. Stryke, E. Pham, B. Fung, W. J. Bolosky, D. Ingebrigtsen, W. Lorizio, S. M. Paff, J. A. Leake, R. Pesano, R. DeBiasi, S. Dominguez, C. Y. Chiu, Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res.* **29**, 831–842 (2019). doi:10.1101/gr.238170.118 Medline

34. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009). doi:10.1186/1471-2105-10-421 Medline

35. K. Katoh, D. M. Standley, MAFFT: Iterative refinement and additional methods. *Methods Mol. Biol.* **1079**, 131–146 (2014). doi:10.1007/978-1-62703-646-7_8 Medline

36. S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010). doi:10.1093/sysbio/syq010 Medline

37. Z. Yang, Estimating the pattern of nucleotide substitution. *J. Mol. Evol.* **39**, 105–111 (1994). doi:10.1007/BF00178256 Medline

38. M. Hasegawa, H. Kishino, T. Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985). doi:10.1007/BF02101694 Medline

39. M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006). doi:10.1080/10635150600755453 Medline