Cite as: X. Deng *et al.*, *Science* 10.1126/science.abb9263 (2020).

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

Xianding Deng[1,2]*, Wei Gu[1,2]*, Scot Federman[1,2]*, Louis du Plessis[3]*, Oliver G. Pybus[3], Nuno Faria[3], Candace Wang[1,2], Guixia Yu[1,2], Brian Bushnell[4], Chao-Yang Pan[5], Hugo Guevara[5], Alicia Sotomayor-Gonzalez[1,2], Kelsey Zorn[6], Allan Gopez[1], Venice Servellita[1], Elaine Hsu[1], Steve Miller[1], Trevor Bedford[7,8], Alexander L. Greninger[7,9], Pavitra Roychoudhury[7,9], Lea M. Starita[8,10], Michael Famulare[11], Helen Y. Chu[8,12], Jay Shendure[8,9,13], Keith R. Jerome[7,9], Catie Anderson[14], Karthik Gangavarapu[14], Mark Zeller[14], Emily Spencer[14], Kristian G. Andersen[14], Duncan MacCannell[15], Clinton R. Paden[15], Yan Li[15], Jing Zhang[15], Suxiang Tong[15], Gregory Armstrong[15], Scott Morrow[16], Matthew Willis[17], Bela T. Matyas[18], Sundari Mase[19], Olivia Kasirye[20], Maggie Park[21], Godfred Masinde[22], Curtis Chan[22], Alexander T. Yu[5], Shua J. Chai[5,15], Elsa Villarino[23], Brandon Bonin[23], Debra A. Wadford[5], Charles Y. Chiu[1,2,24]†

[1]Department of Laboratory Medicine, University of California, San Francisco, CA, USA. [2]UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA, USA. [3]Department of Zoology, University of Oxford, Oxford, UK. [4]Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]California Department of Public Health, Richmond, CA, USA. [6]Department of Biochemistry and Biophysics, University of California, San Francisco, CA, USA. [7]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [8]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [9]Department of Laboratory Medicine, University of Washington, Seattle, WA, USA. [10]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [11]Institute for Disease Modeling, Bellevue, WA, USA. [12]Department of Medicine, University of Washington, Seattle, WA, USA. [13]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [14]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA. [15]U.S. Centers for Disease Control and Prevention, Atlanta, GA, USA. [16]San Mateo County Department of Public Health, San Mateo, CA, USA. [17]Marin County Division of Public Health, San Rafael, CA, USA. [18]Solano County Department of Public Health, Fairfield, CA, USA. [19]Sonoma County Department of Public Health, Santa Rosa, CA, USA. [20]Sacramento County Division of Public Health, Sacramento, CA, USA. [21]San Joaquin County Department of Public Health, Stockton, CA, USA. [22]San Francisco County Department of Public Health, San Francisco, CA, USA. [23]County of Santa Clara, Public Health Department, Santa Clara, CA, USA. [24]Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, CA, USA.

*These authors contributed equally to this work.

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

The COVID-19 pandemic caused by the novel coronavirus SARS-CoV-2 has spread globally, with >52,000 cases in California as of May 4, 2020. Here we investigate the genomic epidemiology of SARS-CoV-2 in Northern California from late January to mid-March 2020, using samples from 36 patients spanning 9 counties and the Grand Princess cruise ship. Phylogenetic analyses revealed the cryptic introduction of at least 7 different SARS-CoV-2 lineages into California, including epidemic WA1 strains associated with Washington State, with lack of a predominant lineage and limited transmission between communities. Lineages associated with outbreak clusters in 2 counties were defined by a single base substitution in the viral genome. These findings support contact tracing, social distancing, and travel restrictions to contain SARS-CoV-2 spread in California and other states.

The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19), is a pandemic that has infected more than 3.2 million people around the world and caused more than 250,000 deaths as of May 4, 2020 (*1*), including >1.2 million cases in the United States (US) and >52,000 in California. An exponential growth in the number of cases has overburdened clinical care facilities and threatens to overwhelm the medical workforce. The reported case numbers also underestimate the true number of infections due to the limited volume of diagnostic testing to date and the presence of asymptomatic or mild cases (*2–4*). As a result, California, along with many other states and countries, has issued a "shelter-in-place" policy for all residents, effective March 20, 2020, and ongoing at the time of this report. These unprecedented measures have disrupted daily life significantly for ~40 million inhabitants for an indefinite period, with the potential for incurring profound economic losses (*5*).

Until late Feb 2020, the majority of infections identified in the US were related to travelers returning from high-risk

countries, repatriated citizens under quarantine, or close contacts of infected patients. Community spread, in which the source of the infection is unknown, has since been documented in multiple states. In particular, Washington State has reported a series of COVID-19 cases from Jan 21 to Mar 18, following the identification of the earliest case reported in the US, WA1, on Jan 19, suggesting the presence of a persistent transmission chain in the community (*6*, *7*).

Genomic epidemiology of emerging viruses has proven to be a useful tool for outbreak investigation and for tracking virus evolution and spread (*7–9*). During the Ebola virus disease epidemic of 2013-2016 in West Africa, genomic analyses established that the outbreak had a single zoonotic origin (*9*), that two major viral lineages were circulating (*10*), and that sexual transmission played a role in maintaining some transmission chains (*11*). Viral genome sequencing also uncovered the route that Zika virus traveled from northern Brazil to other regions (*12*), including Central America and Mexico (*13*) and the Caribbean and US (*14*). However, real-time genomic epidemiology data of COVID-19 to inform public health interventions in California have been lacking to date.

We recently developed a method called MSSPE (Metagenomic Sequencing with Spiked Primer Enrichment) to rapidly enrich and assemble viral genomes directly from clinical samples (*15*). Here we used this method and/or tiling multiplex PCR to recover viral genomes from COVID-19 patients in Northern California and perform phylogenetic analyses to better understand the genetic diversity of SARS-CoV-2 in the US and the nature of transmission of virus lineages in the community.

We screened a total of 62 respiratory swab samples from 54 COVID-19 patients available from hospitals and clinics at University of California, San Francisco (UCSF), the California Department of Public Health (CDPH), and 8 county public health departments in Northern California (table S1). Presumptive positive cases were confirmed to be SARS-CoV-2 infected by testing using a CDC assay approved by a Food and Drug Administration (FDA) Emergency Use Authorization (EUA) on February 4, 2020 (*16*). SARS-CoV-2 genomes (>65% coverage) were recovered from 36 patients (Fig. 1A and table S2). The 36 infected patients for whom viral genomes were obtained were collected from January 29 to March 20, 2020 and spanned 9 counties in Northern California (Fig. 1B and table S2). The patient samples included (i) 11 samples collected from the Grand Princess cruise ship, during its two voyages from San Francisco to Mexico and Hawaii in February and March 2020, (ii) 3 samples from a Solano County cluster that included the first reported case of community transmission in the US with subsequent spread to two health care workers, (iii) 7 samples from Santa Clara County from a local outbreak cluster associated with workspace transmission, (iv) 3 samples from patients who contracted the infection from a sick contacts,(v) 5 samples related to domestic or international travel, and (vi) 7 samples from additional cases of community transmission.

We performed MSSPE (*15*) and/or tiled multiplex PCR (*17*) on each sample to enrich for the SARS-CoV-2 RNA genome, followed by metagenomic next-generation sequencing (mNGS) of pooled and indexed samples on Illumina NextSeq, HiSeq or MiSeq instruments (*18*, *19*). The PCR cycle thresholds ranged from 15.3 to 33.4, corresponding to virus loads of $1.1 \times 10^4 - 2.7 \times 10^8$ copies/mL (fig. S1 and table S2). An average of 31 million (interquartile ratio, IQR, 23-57 million) and 2.2 ± 0.2 million reads were generated per sample for using MSSPE and tiling multiplex PCR respectively, and virus genomes were assembled by mapping to reference genome NC_045512 (Wuhan-Hu-1). The assembly yielded 34 SARS-CoV-2 genomes with genome coverage >65% and these were included in the study. An additional two genomes sequenced from samples of a returning traveler from Wuhan, China and a household contact collected on January 29th by the CDC (CA3 and CA4) were also included in the analysis. The median coverage achieved across all samples was 97.7% (IQR 90.4.0%-99.7%).

Phylogenetic analysis revealed that the 36 SARS-CoV-2 genomes from California generated in this study were dispersed across the evolutionary tree of SARS-CoV-2 that was built from 789 worldwide genomes deposited into GISAID as of March 20, 2020 (Fig. 2A). The 36 genomes included 14 in the Washington State (WA1) lineage, 10 in a lineage associated with the Santa Clara County outbreak cluster (henceforth referred to as the SCC1 lineage), 3 from a Solano County cluster of 3 individuals, 5 related to lineages circulating in Europe and New York, and 4 related to early lineages from Wuhan or other regions of China (including 2 patients from San Benito County with identical genomes) (Figs. 1, 2A, and 3 and table S2).

A large outbreak was associated with travel on the US Grand Princess cruise ship (with at least 78 confirmed positive cases out of 469 tested) as of March 26 (*20*). The Grand Princess undertook two consecutive voyages from San Francisco (voyage A to Mexico on February 11 – 21 and voyage B to Hawaii on February 22 – March 4), with much of the same crew and a shared subset of passengers. Samples from 11 infected patients were sequenced, 3 of whom had been on voyage A and became sick after returning to their home county, and 8 from crew members and passengers aboard the cruise ship on voyage B. Importantly, all 11 available sequenced genomes from the Grand Princess were part of the WA1 lineage (Fig. 2, A and B, and Fig. 3). In addition to sharing 3 single nucleotide variants (SNVs) that define WA1 (C8782T, C18060T, and T28144C), the sequences from cruise

ship passengers and crew also shared two additional SNVs, C17747T and A17858G common to nearly all WA1 sequences sampled from Washington and California but not the basal WA1 case (Figs. 2B and 3).

The WA1 case was reported on January 19 (*6*), and thus substantially predated the voyages of the Grand Princess cruise ship (*7*, *20*). In addition, 6 of 8 passengers on voyage A (UC 7 –11, 30) carried at least 2 new mutations (G16975T and C23185T) not observed in UC1, UC19, and UC20, who were all on the first cruise (Fig. 3). This suggested that the virus from UC19 could be basally positioned relative to the cruise ship strains from voyage B, and that COVID-19 infections associated with voyage A may have been passed onto passengers and crew on voyage B. However, the initial WA1 subtree extracted from the global maximum-likelihood phylogenetic tree did not place UC19 basal to sequences from voyage B passengers due to artifacts from shared areas of low coverage (fig. S2). To establish a more accurate tree topology, we therefore reconstructed a new phylogenetic subtree of the WA1 lineage after excluding all ambiguous sites. In this new subtree (Fig. 2B), UC19 is basal to all other California genomes within the WA1 lineage. In addition, among the sequences from patients on voyage B, UC5 and UC6 group together, while UC7-11 and UC30 group together with a sequence sampled in Minnesota.

The chronology and phylogeny of the cruise ship outbreak, along with the predominance of the WA1 lineage in Washington State (*7*), suggest that the virus on the Grand Princess likely came from Washington State, although the cases may also have originated from a different region in which the WA1 strain is circulating. In addition to passengers and crew members aboard the Grand Princess, virus genomes sampled from three cases of community transmission in different counties of the Bay area (UC22, UC23 and UC28) were also of the WA1 lineage. UC22 was the son of an infected Grand Princess passenger (UC20) on voyage A and most likely contracted the virus from household contact. The UC23 and UC28 cases may also reflect transmission from disembarking Grand Princess passengers on voyage A, or pre-existing circulation of the WA1 strain in the community.

Three patients examined in this study (CA3, CA4, and UC12) had COVID-19 infections associated with international travel or exposure to international travelers. CA3 corresponds to a resident of San Benito County who became sick shortly after returning from Wuhan, China in late January. The sequence of his SARS-CoV-2 genome is identical to that of CA4, a household contact who was also infected with the virus. Their viral genomes were found to be closely related to early lineages from China (Fig. 2A and data S1). UC12 had a prolonged exposure to a known positive traveler from Switzerland while attending a conference. The genome from UC12 fell within a lineage containing many sequences from European residents or travelers from Europe (Fig. 2A). Interestingly, four additional genomes (UC24, UC26, UC27 and UC36) were also grouped within the European lineage. UC27 and UC36 were both diagnosed shortly after returning to California from New York, consistent with reports that the New York outbreak that began in March 2020 originated with travelers coming from Europe (*21*, *22*). UC26 also reported domestic travel from Los Angeles, while UC24 had no known travel history.

In Santa Clara County, we sequenced 7 genomes from individuals who were part of a local outbreak of COVID-19 at a large workplace facility with multiple employers, large areas of shared space, and heavy pedestrian traffic. The genomes all shared the G29711T SNV that defines the SCC1 lineage (Figs. 2C and 3). Four employees (UC13, UC14, UC15, and UC34) had dates of symptom onset within two weeks of each other, although they did not know each other. UC16 and UC17 were family members of UC13 and lived in the same residence, while UC35 transported UC14 to the hospital via emergency medical services. Notably, the genomes from a Solano county resident (UC21) and a San Mateo couple (UC18 and UC25) were also placed in the SCC1 lineage, suggesting possible spread to different counties. Further epidemiological investigation found that UC21 had visited a merchant in Santa Clara, during which he likely became infected.

In Solano County, a small cluster of 3 cases included the first reported instance of community transmission in the US on February 26 (UC4) (Figs. 2D and 3). The two other cases (UC2 and UC3) were healthcare workers who were taking care of patient UC4 and likely contracted the disease in the hospital, consistent with transmission of the disease from patient to health care providers (*23*). The genomic epidemiology of the COVID-19 cases associated with community spread studied here do not show any predominant SARS-CoV-2 lineage circulating in Northern California. In California, multiple recent and unrelated introductions of SARS-CoV-2 into the state via different routes appear to give rise to the diversity of virus lineages reported in this study, with no single predominant lineage observed. We note that this does not exclude the possibility of cryptic transmission of multiple lineages in California simultaneously, as the current level of sampling is not dense enough to confidently estimate the dates of the seeding events, nor the subsequent periods of cryptic transmission before a lineage was identified.

There is growing evidence that the WA1 is now an established lineage of SARS-CoV-2 in the US. Here we found that viruses in the WA1 lineage from Grand Princess cruise ship passengers as well as from residents of several Northern California counties. In addition, WA1 lineage viruses

have been identified in COVID-19 cases from many states including Minnesota, Connecticut, Utah, Virginia, and New York (*24*, *25*). The early date and basal phylogenetic position of the WA1 virus make it likely that the direction of dissemination was from Washington State to California and other states; however, this conclusion could change if further genomic sampling in the US revealed additional virus genetic diversity. Notably, SARS-CoV-2 sequences from Connecticut (*25*) and British Columbia, Canada (Fig. 2B) are positioned close to the root of the subtree containing the WA1 sequences, raising the possibility that the virus may not have been first introduced into the US via Washington State.

SARS-CoV-2, like other coronaviruses, contains a nonstructural gene with proofreading activity (*26*). Consequently, the virus evolves more slowly than many other human RNA viruses, on the order of 1 to 2 DNA base substitutions a month across its ~29 kB genome (*27*). Thus, only 1-3 SNVs in general are needed to define a distinct lineage. The WA1 lineage consists of 3 key SNVs, C8782T, C18060T, and T28144C, while the SCC1 lineage associated with the Santa Clara County cluster and the Solano County cluster are each defined by only one SNV, G29711T and C9924T, respectively (Figs. 2 and 3).

Our epidemiological and genomic survey of SARS-CoV-2 has several limitations. First, this initial analysis represents a relatively sparse sampling of cases. Undersampling of virus genomes is due in part to the high proportion of cases (80%) with asymptomatic or mild disease (*2*–*4*) and limited diagnostic testing for COVID-19 infection to date in California and throughout the US. Second, the majority of samples analyzed were obtained from public health laboratories and thus may not be representative of the general population. Finally, phylogenetic grouping of viruses from different locations, such as Washington State and California in the same WA1 lineage, does not prove the directionality of spread. Despite this, our study shows that robust insights into COVID-19 transmission are achievable if virus genomic diversity is combined and jointly interpreted with detailed epidemiological case data. In particular, we found that a returning traveler from New York was infected with a lineage circulating widely in Europe, thus suggesting an association between the New York outbreak and intercontinental travel to and from Europe before this was widely recognized (*21*, *22*).

Public health containment measures such as isolation and contact tracing, as performed in the Solano County and Santa Clara County outbreak clusters, become more difficult to maintain once a lineage becomes established in the community. Our data suggest concerning trends in this direction, such as the association between the WA1 lineage and community-acquired COVID-19 cases in several counties of Northern California, and a virus from the SCC1 lineage detected in residents of neighboring San Mateo and Solano County. Social distancing interventions, such as the "shelter-in-place" directive that was issued by the governor of California on March 20, 2020, may assist in stemming spread from community to community. Interstate dissemination of SARS-CoV-2 lineages has also been demonstrated coast-to-coast between Washington State and Connecticut (*25*), and from domestic and international travel into the San Francisco Bay Area in the current study. Suspension of non-essential travel may thus be necessary to prevent ongoing importation of new cases in California and other states.

REFERENCES AND NOTES

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

ACKNOWLEDGMENTS

We acknowledge the help and advice of James T. Lee at the US CDC. We thank all of the authors who have contributed genome data on GISAID. Author credits for specific GISAID contributions can be found on www.gisaid.org/. Clinical samples from UCSF were processed according to protocols approved by the UCSF Institutional Review Board (protocol number 10-01116, 11-05519). Samples collected by the CDPH were de-identified and deemed not research or exempt by the Committee for the Protection of Human Subjects (project number 2020-30) issued under the California Health and Human Services Agency's Federal Wide Assurance #00000681 with the Office of Human Research Protections. A non-research determination for this project was also granted by Sonoma County as SARS-CoV-2 genome sequencing was designated an epidemic disease control activity, with collected data directly related to disease control. **Funding:** This work was funded by NIH grants R33-AI129455 (CYC) from the National Institute of Allergy and Infectious Diseases and K08-CA230156 (WG) from the National Cancer Institute, the California Initiative to Advance Precision Medicine (CYC), the Charles and Helen Schwab Foundation (CYC), and the Burroughs-Wellcome CAMS Award (WG). OP and LdP acknowledge support from the Oxford Martin School and the European Research Council under the Seventh Framework Program of the European Commission (Pathogen Phylodynamics Grant 614725). **Author contributions:** CYC conceived, designed, and supervised the study. AG, AS-G, CW, CYP, GY, HG, VS, and XD performed experiments. CYC, SF, and BB assembled and curated the viral genomes. CYC, WG and XD analyzed data. CYC, EH, KZ, SM, and WG collected patient samples at UCSF. DM and GA analyzed genomic and epidemiologic data. TB, AG, PR, LMS, MF, HYC, JS, and KRJ collected, assembled, and provided viral genome data from Washington and contributed to the phylogenetic analysis. CA, KG, MZ, ES, and KGA provided viral genome data from Southern California. CP, JTL, JZ, ST, and YL sequenced and analyzed viral genomes at the CDC. LDP, NF and OP performed phylogenetic analysis of genomes. AY, BTM, BB, CC, DAW, EV, GM, MP, MW, OK, SC, SM collected samples, extracted the viral RNA and/or provided epidemiology data from counties in California. CYC, WG, and XD wrote the manuscript. CYC, XD, SF, CW, DAW, LDP, OP, and WG edited the manuscript. All authors read the manuscript and agree to its contents. **Competing interests:** CYC is the director of the UCSF-Abbott Viral Diagnostics and Discovery Center (VDDC) and receives research support funding from Abbott Laboratories. CYC and XD are inventors on a patent application on the MSSPE method titled "Spiked Primer Design for Targeted Enrichment of Metagenomic Libraries" (US Application No. 62/667,344, filed 05/04/2018 by University of California, San Francisco). HYC is a consultant for Merck and GlaxoSmithKline, and receives research funding from Sanofi-Pasteur, Ellume and Cepheid, unrelated to this work. All other authors have no conflicts to declare. The opinions expressed by the authors contributing to this journal do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated. **Data and materials availability:** Assembled SARS-CoV-2 genomes in this study were uploaded to GISAID (*28*, *29*) as FASTA files (accession numbers in table S2) and can be visualized on a continually updated phylogenetic tree on NextStrain (*24*). Viral genomes were submitted to the National Center for Biotechnology Information (NCBI) GenBank database (accession numbers MT419827 – MT419860). Raw sequence data were submitted to the NCBI Sequence Read Archive (SRA) database (BioProject accession number PRJNA 629889 and umbrella BioProject accession number PRJNA171119). Locations of SNVs aligned to the reference sequence (NC_045512), was done by custom scripts (*30*). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/cgi/content/full/science.abb9263/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S5
Data S1
References (*31*–*39*)

27 March 2020; accepted 3 June 2020
Published online 8 June 2020
10.1126/science.abb9263

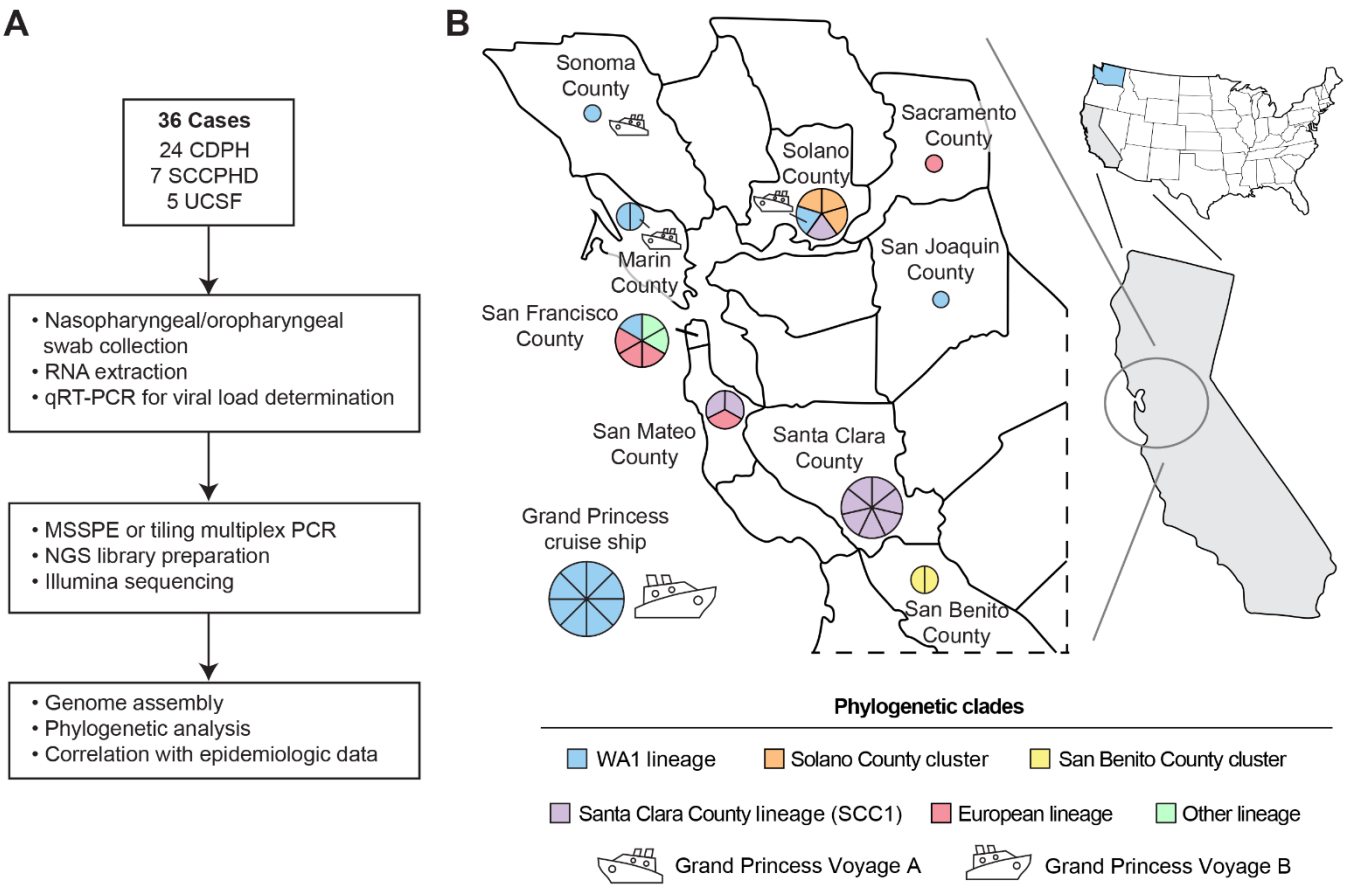


Fig. 1. Genomic survey of SARS-CoV-2 genomes in Northern California. (**A**) Analysis workflow. (**B**) Map of the Northern California survey region divided by county. The pie charts for each county are subdivided according to the number of patients whose viral genome was sequenced, and the color corresponds to the viral lineage as determined by phylogenetic analysis. Passengers (n = 3) who were on the Grand Princess cruise ship during voyage A to Mexico and disembarked to return to their home communities are denoted by a ship icon facing left, while passengers (n = 8) aboard the Grand Princess cruise ship during voyage B to Hawaii are denoted by a ship icon facing right. Abbreviations: SCCPHD, Santa Clara County Public Health Department; CDPH, California Department of Public Health; UCSF, University of California San Francisco; MSSPE, Metagenomic Sequencing with Spiked Primer Enrichment; NGS, next-generation sequencing.

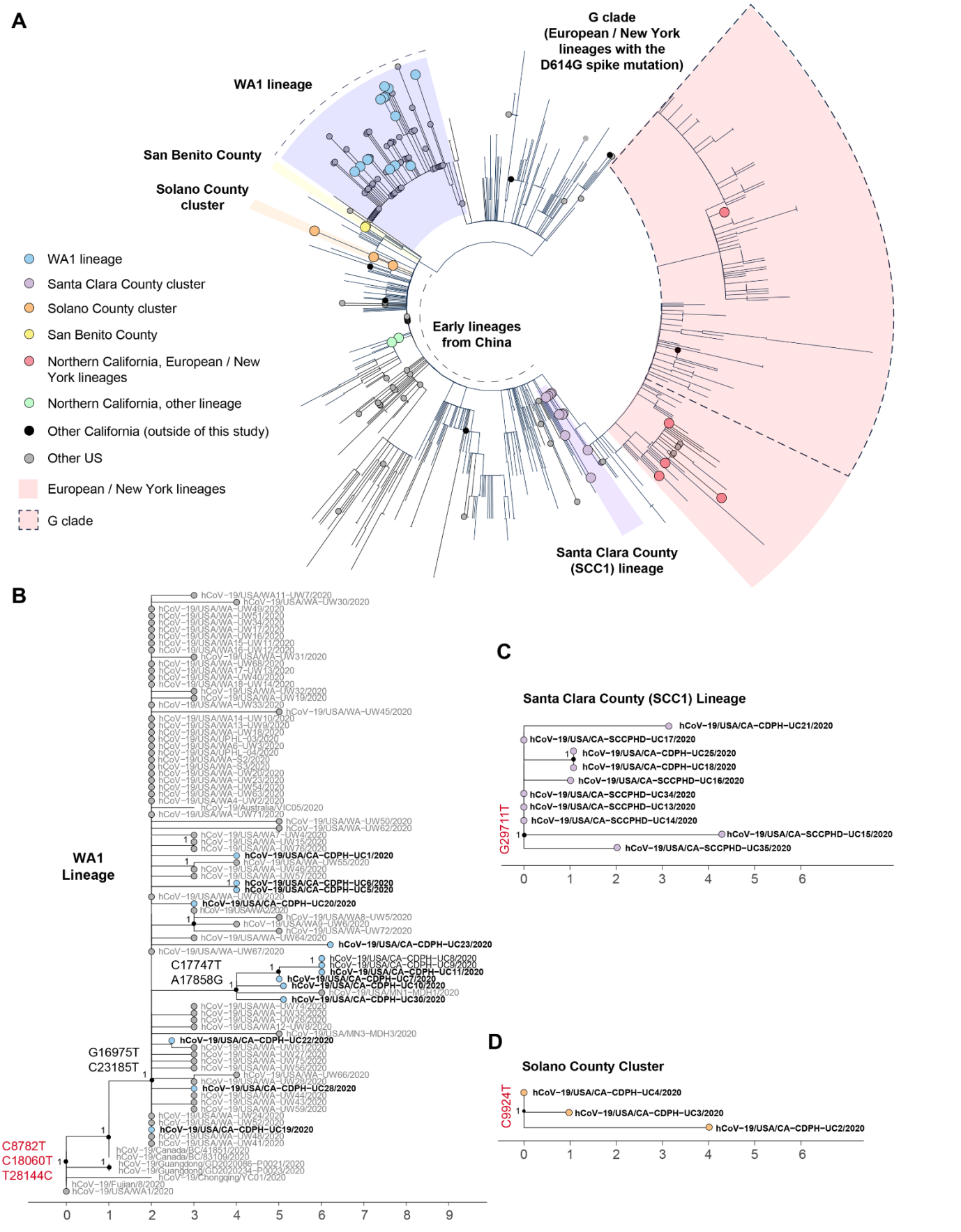
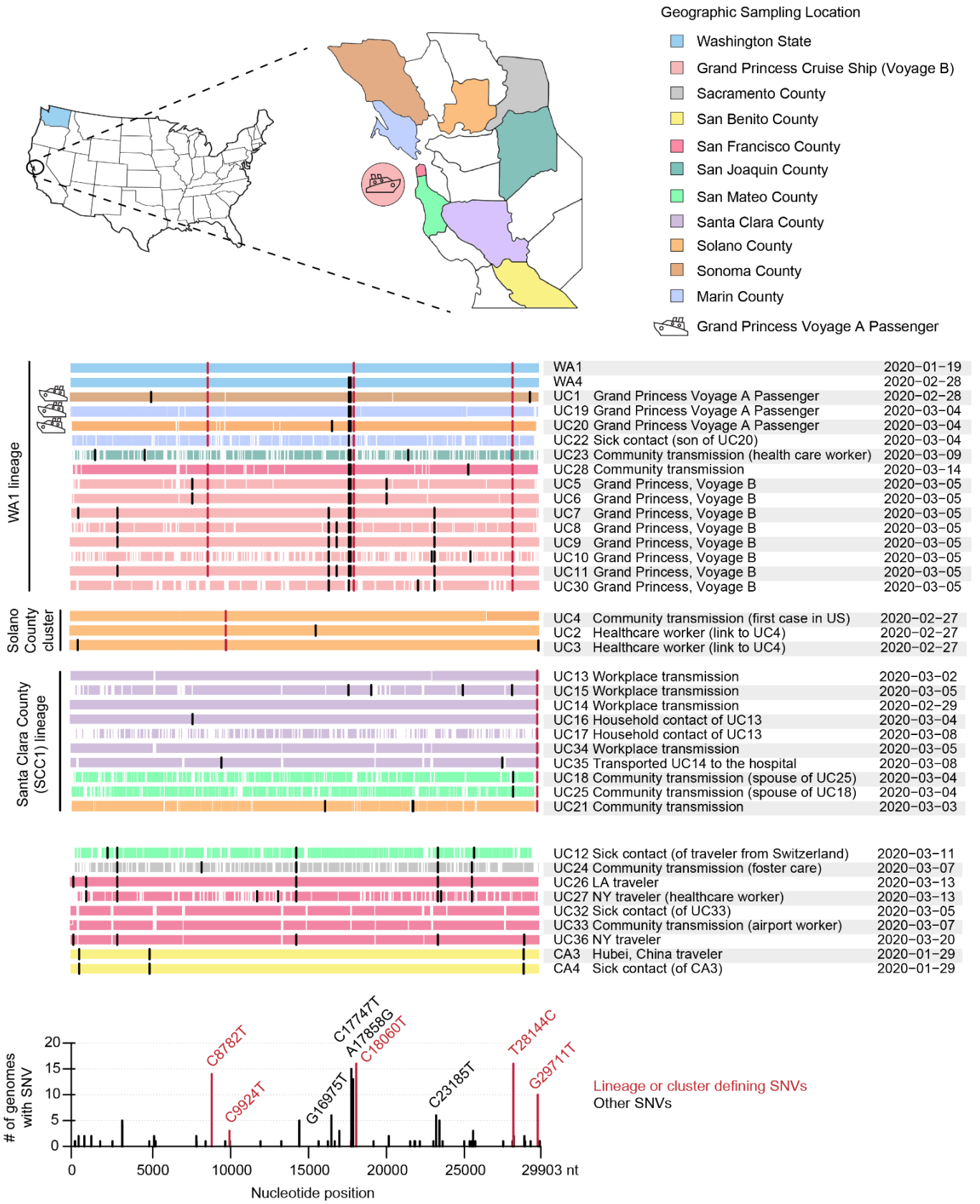

A
G clade
(European / New York
lineages with the
D614G spike mutation)
WA1 lineage
San Benito County
Solano County
cluster
Early lineages
from China
Santa Clara County
(SCC1) lineage
WA1 lineage
Santa Clara County cluster
Solano County cluster
San Benito County
Northern California, European / New York lineages
Northern California, other lineage
Other California (outside of this study)
Other US
European / New York lineages
G clade
B
WA1
Lineage
C17747T
A17858G
G16975T
C23185T
C8782T
C18060T
T28144C
hCoV-19/USA/CA-CDPH-UC1/2020
hCoV-19/USA/CA-CDPH-UC5/2020
hCoV-19/USA/CA-CDPH-UC6/2020
hCoV-19/USA/CA-CDPH-UC7/2020
hCoV-19/USA/CA-CDPH-UC8/2020
hCoV-19/USA/CA-CDPH-UC9/2020
hCoV-19/USA/CA-CDPH-UC10/2020
hCoV-19/USA/CA-CDPH-UC11/2020
hCoV-19/USA/CA-CDPH-UC19/2020
hCoV-19/USA/CA-CDPH-UC20/2020
hCoV-19/USA/CA-CDPH-UC22/2020
hCoV-19/USA/CA-CDPH-UC23/2020
hCoV-19/USA/CA-CDPH-UC28/2020
hCoV-19/USA/CA-CDPH-UC30/2020
0 1 2 3 4 5 6 7 8 9
C
Santa Clara County (SCC1) Lineage
G29711T
hCoV-19/USA/CA-CDPH-UC21/2020
hCoV-19/USA/CA-SCCPHD-UC17/2020
hCoV-19/USA/CA-CDPH-UC25/2020
hCoV-19/USA/CA-CDPH-UC18/2020
hCoV-19/USA/CA-SCCPHD-UC16/2020
hCoV-19/USA/CA-SCCPHD-UC34/2020
hCoV-19/USA/CA-SCCPHD-UC13/2020
hCoV-19/USA/CA-SCCPHD-UC14/2020
hCoV-19/USA/CA-SCCPHD-UC15/2020
hCoV-19/USA/CA-SCCPHD-UC35/2020
0 1 2 3 4 5 6
D
Solano County Cluster
C9924T
hCoV-19/USA/CA-CDPH-UC4/2020
hCoV-19/USA/CA-CDPH-UC3/2020
hCoV-19/USA/CA-CDPH-UC2/2020
0 1 2 3 4 5 6

Fig. 2 (preceding page). Phylogeny of SARS-CoV-2 lineages in California. (**A**) Phylogenetic tree of 753 SARS-CoV-2 genomes from GISAID (till Mar 20, 2020) along with the 36 genomes in this survey (tree file attached in supplementary material). Genomes from Northern California sequenced in the current study are denoted with colored circles, while other genomes sequenced from California and from other states in the US are denoted with black and gray circles, respectively. The name of each lineage or outbreak cluster is shown next to the arc line. (**B**) Phylogenetic subtree corresponding to the WA1 lineage. This subtree was reconstructed from 88 SARS-CoV-2 genomes after removal of ambiguous nucleotide sites that had generated low-coverage artifacts (see text). The WA1 virus from Washington state (first case in the US) is at the root of the subtree along with a virus sequenced in China. The UC19 virus (from a Grand Princess voyage A passenger) is basal to the viruses sequenced from crew members and passengers on voyage B. (**C**) Zoomed view of the SCC1 lineage associated with the Santa Clara County outbreak cluster. (**D**) Zoomed view of the Solano County cluster. The x-axis shows the number of base substitutions relative to the root of the phylogenetic tree. The key SNVs defining a lineage or cluster are shown in red text. 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.

Fig. 3 (next page). Multiple sequence alignment of all SARS-CoV-2 genomes reported across 9 Northern California counties and the Grand Princess cruise ship. Single nucleotide variants (SNVs) with respect to the reference genome (NC_045512) are shown as vertical red and black lines for lineage defining SNVs and other SNVs, respectively. Cases that are part of the WA1 lineage include the first case of COVID-19 infection (WA1) in the US, 8 passengers and crew members aboard the Grand Princess cruise ship during its second trip (voyage B), and 3 individuals surveyed from 3 Northern California counties as passengers on the ship's first trip (voyage A). The three SNVs C8782T, C18060T, and T28144C define the WA1 lineage, and the two SNVs C17747T, A17858G are common to Grand Princess passengers and crew. Viruses from voyage B passengers and crew share SNVs G16975T and C23185T that are lacking in viruses from voyage A passengers. Single SNV variants C9924T and G29711T define the lineages from Solano County and Santa Clara County, respectively. The putative epidemiological link and sample collection date are shown beside the sequence alignment.


Geographic Sampling Location
Washington State
Grand Princess Cruise Ship (Voyage B)
Sacramento County
San Benito County
San Francisco County
San Joaquin County
San Mateo County
Santa Clara County
Solano County
Sonoma County
Marin County
Grand Princess Voyage A Passenger
WA1 lineage
Solano County cluster
Santa Clara County (SCC1) lineage
WA1 2020-01-19
WA4 2020-02-28
UC1 Grand Princess Voyage A Passenger 2020-02-28
UC19 Grand Princess Voyage A Passenger 2020-03-04
UC20 Grand Princess Voyage A Passenger 2020-03-04
UC22 Sick contact (son of UC20) 2020-03-04
UC23 Community transmission (health care worker) 2020-03-09
UC28 Community transmission 2020-03-14
UC5 Grand Princess, Voyage B 2020-03-05
UC6 Grand Princess, Voyage B 2020-03-05
UC7 Grand Princess, Voyage B 2020-03-05
UC8 Grand Princess, Voyage B 2020-03-05
UC9 Grand Princess, Voyage B 2020-03-05
UC10 Grand Princess, Voyage B 2020-03-05
UC11 Grand Princess, Voyage B 2020-03-05
UC30 Grand Princess, Voyage B 2020-03-05
UC4 Community transmission (first case in US) 2020-02-27
UC2 Healthcare worker (link to UC4) 2020-02-27
UC3 Healthcare worker (link to UC4) 2020-02-27
UC13 Workplace transmission 2020-03-02
UC15 Workplace transmission 2020-03-05
UC14 Workplace transmission 2020-02-29
UC16 Household contact of UC13 2020-03-04
UC17 Household contact of UC13 2020-03-08
UC34 Workplace transmission 2020-03-05
UC35 Transported UC14 to the hospital 2020-03-08
UC18 Community transmission (spouse of UC25) 2020-03-04
UC25 Community transmission (spouse of UC18) 2020-03-04
UC21 Community transmission 2020-03-03
UC12 Sick contact (of traveler from Switzerland) 2020-03-11
UC24 Community transmission (foster care) 2020-03-07
UC26 LA traveler 2020-03-13
UC27 NY traveler (healthcare worker) 2020-03-13
UC32 Sick contact (of UC33) 2020-03-05
UC33 Community transmission (airport worker) 2020-03-07
UC36 NY traveler 2020-03-20
CA3 Hubei, China traveler 2020-01-29
CA4 Sick contact (of CA3) 2020-01-29
# of genomes with SNV
20
15
10
5
0
C8782T
C9924T
C17747T
A17858G
C18060T
G16975T
C23185T
T28144C
G29711T
0
5000
10000
15000
20000
25000
29903 nt
Nucleotide position
Lineage or cluster defining SNVs
Other SNVs