CORONAVIRUS

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[3,4], Nuno R. Faria[3,5], Candace Wang[1,2], Guixia Yu[1,2], Brian Bushnell[6], Chao-Yang Pan[7], Hugo Guevara[7], Alicia Sotomayor-Gonzalez[1,2], Kelsey Zorn[8], Allan Gopez[1], Venice Servellita[1], Elaine Hsu[1], Steve Miller[1], Trevor Bedford[9,10], Alexander L. Greninger[9,11], Pavitra Roychoudhury[9,11], Lea M. Starita[10,12], Michael Famulare[13], Helen Y. Chu[10,14], Jay Shendure[10,11,15], Keith R. Jerome[9,11], Catie Anderson[16], Karthik Gangavarapu[16], Mark Zeller[16], Emily Spencer[16], Kristian G. Andersen[16], Duncan MacCannell[17], Clinton R. Paden[17], Yan Li[17], Jing Zhang[17], Suxiang Tong[17], Gregory Armstrong[17], Scott Morrow[18], Matthew Willis[19], Bela T. Matyas[20], Sundari Mase[21], Olivia Kasirye[22], Maggie Park[23], Godfred Masinde[24], Curtis Chan[24], Alexander T. Yu[7], Shua J. Chai[7,17], Elsa Villarino[25], Brandon Bonin[25], Debra A. Wadford[7], Charles Y. Chiu[1,2,26]†

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread globally, with >365,000 cases in California as of 17 July 2020. We investigated the genomic epidemiology of SARS-CoV-2 in Northern California from late January to mid-March 2020, using samples from 36 patients spanning nine counties and the Grand Princess cruise ship. Phylogenetic analyses revealed the cryptic introduction of at least seven different SARS-CoV-2 lineages into California, including epidemic WA1 strains associated with Washington state, with lack of a predominant lineage and limited transmission among communities. Lineages associated with outbreak clusters in two counties were defined by a single base substitution in the viral genome. These findings support contact tracing, social distancing, and travel restrictions to contain the spread of SARS-CoV-2 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 13.8 million people around the world and caused more than 591,000 deaths as of 17 July 2020 (*1*), including >3.5 million cases in the United States and >365,000 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 because of the presence of asymptomatic or mild cases who do not get tested (*2*–*4*). As a result, California, along with many other states and countries, has issued a "shelter-in-place" policy for all residents, effective 20 March 2020 and ongoing at the time of this report. These unprecedented measures have disrupted daily life for ~40 million inhabitants of the state and have incurred profound economic losses (*5*).

Until late February 2020, the majority of infections identified in the United States 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 reported a series of COVID-19 cases from 21 January to 18 March, following the identification of the earliest case reported in the United States, WA1, on 19 January; this suggests that a persistent WA1 lineage transmission chain was present in the community during that time period (*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 United States (*14*). However, real-time genomic epidemiology data for COVID-19 to inform public health interventions in California have been limited 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 polymerase chain reaction (PCR) to recover viral genomes from COVID-19 patients in Northern California and to perform phylogenetic analyses, with the goal of better understanding the genetic diversity of SARS-CoV-2 in the United States 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 the University of California, San Francisco (UCSF), the California Department of Public Health (CDPH), and eight county public health departments in Northern California (table S1). Presumptive positive cases were confirmed to be SARS-CoV-2–infected by a U.S. Centers for Disease Control and Prevention (CDC) assay approved by a Food and Drug Administration (FDA) Emergency Use Authorization (EUA) on 4 February 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 29 January to 20 March 2020 and spanned nine 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) three samples from a Solano County cluster that included the first reported case of community transmission in the United States with subsequent spread to two health care workers, (iii) seven samples from a local outbreak cluster in Santa Clara County associated with workspace transmission, (iv) three samples from patients who contracted the infection from a sick contact, (v) five samples related to domestic or international travel, and (vi) seven samples from additional cases of community transmission.

We performed MSSPE (*15*) and/or tiling multiplex PCR (*17*) on each sample to enrich for the

[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]Department of Pathobiology and Population Sciences, Royal Veterinary College, London, UK. [5]MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, UK. [6]Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [7]California Department of Public Health, Richmond, CA, USA. [8]Department of Biochemistry and Biophysics, University of California, San Francisco, CA, USA. [9]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [10]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [11]Department of Laboratory Medicine, University of Washington, Seattle, WA, USA. [12]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [13]Institute for Disease Modeling, Bellevue, WA, USA. [14]Department of Medicine, University of Washington, Seattle, WA, USA. [15]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [16]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA. [17]U.S. Centers for Disease Control and Prevention, Atlanta, GA, USA. [18]San Mateo County Department of Public Health, San Mateo, CA, USA. [19]Marin County Division of Public Health, San Rafael, CA, USA. [20]Solano County Department of Public Health, Fairfield, CA, USA. [21]Sonoma County Department of Public Health, Santa Rosa, CA, USA. [22]Sacramento County Division of Public Health, Sacramento, CA, USA. [23]San Joaquin County Department of Public Health, Stockton, CA, USA. [24]San Francisco County Department of Public Health, San Francisco, CA, USA. [25]County of Santa Clara, Public Health Department, Santa Clara, CA, USA. [26]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

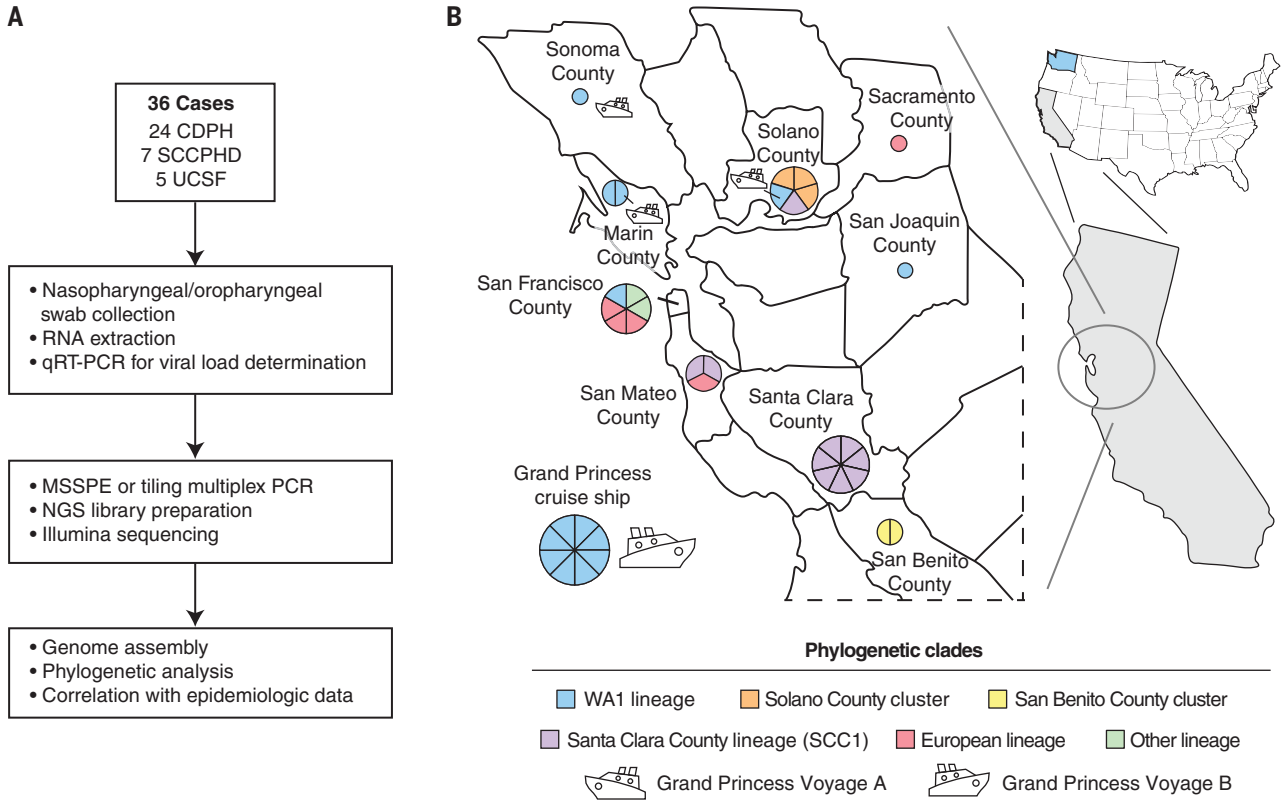

Fig. 1. Genomic survey of SARS-CoV-2 genomes in Northern California. (**A**) Analysis workflow. (**B**) Map of the Northern California survey region divided by county. The pie charts for each county are subdivided according to the number of patients whose viral genome was sequenced; colors correspond to viral lineages 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; passengers (n = 8) aboard the Grand Princess 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.

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 1.1×10^4 to 2.7×10^8 copies/ml (fig. S1 and table S2). An average of 31 million [interquartile ratio (IQR), 23 million to 57 million] and 2.2 ± 0.2 million reads were generated per sample for use in 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 exceeding 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 29 January 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% to 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 constructed from 789 worldwide genomes deposited into GISAID as of 20 March 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 the SCC1 lineage), three from a Solano County cluster of three individuals, five related to lineages circulating in Europe and New York, and four related to early lineages from Wuhan or other regions of China (including two 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 Grand Princess cruise ship (with at least 78 confirmed positive cases out of 469 tested) as of 26 March (*20*). The Grand Princess undertook two consecutive voyages from San Francisco (voyage A to Mexico, 11 to 21 February; voyage B to Hawaii, 22 February to 4 March), with much of the same crew and a shared subset of passengers. Samples from 11 infected patients were sequenced, three of whom had been on voyage A and became sick after returning to their home county, and eight from crew members and passengers aboard the cruise ship on voyage B. Note that 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 three 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 19 January (*6*) and thus substantially predated the voyages of the Grand Princess cruise ship (*7*, *20*). In addition, six of eight passengers on voyage B (UC7 to UC11 and UC30) each carried at least two new mutations (G16975T and C23185T) not observed in UC1, UC19, and UC20, who were all on voyage A (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 on to passengers and crew on voyage B. However, because of sequencing artifacts from areas of low coverage, the initial WA1 subtree extracted from the global maximum likelihood phylogenetic tree did not place UC19 basal to

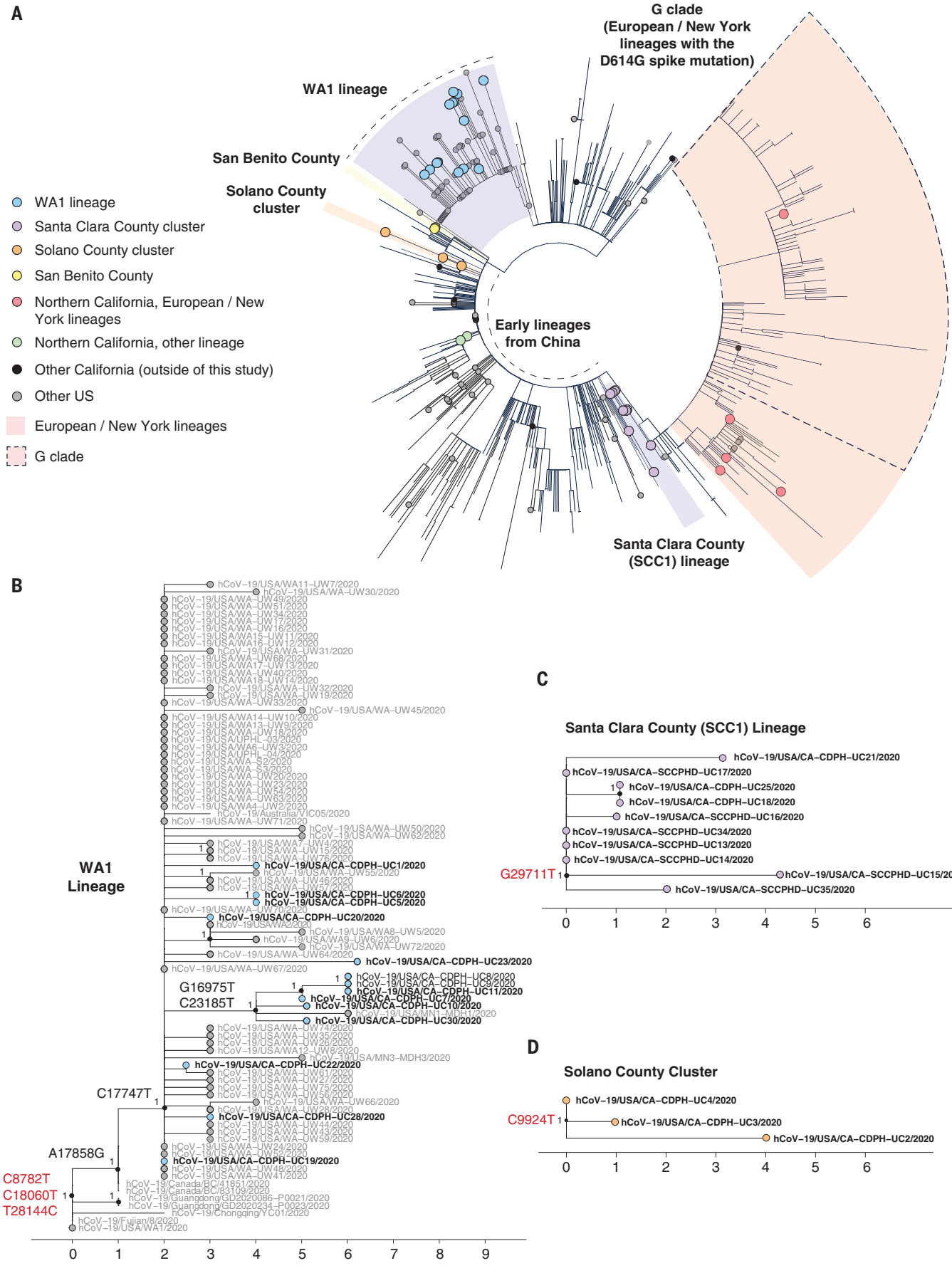

Fig. 2. Phylogeny of SARS-CoV-2 lineages in California. (**A**) Phylogenetic tree of 753 SARS-CoV-2 genomes from GISAID (until 20 March 2020) along with the 36 genomes in this survey (see supplementary materials for tree file). Genomes from Northern California sequenced in the current study are denoted by colored circles; other genomes sequenced from California and from other U.S. states are denoted by 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 United States) 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. In (B) to (D), the key single-nucleotide variants (SNVs) defining a lineage or cluster are shown in red or black 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.

sequences from voyage B passengers (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, whereas UC7 to UC11 and UC30 group together with a sequence sampled in Minnesota.

The chronology and phylogeny of the cruise ship outbreak, and the predominance of the WA1 lineage in Washington state (*7*), together suggest that the virus on the Grand Princess likely came from Washington, 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 San Francisco 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 preexisting 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, whereas UC24 had no known travel history.

In Santa Clara County, we sequenced seven 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 2 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 County 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 County, during which he likely became infected.

In Solano County, a small cluster of three cases included the first reported instance of community transmission in the United States on 26 February (UC4) (Figs. 2D and 3). The two other cases (UC2 and UC3) were health care workers who were taking care of patient UC4 and likely contracted the disease in the hospital, consistent with transmission of the disease from patients 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 have given 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 WA1 is now an established lineage of SARS-CoV-2 in the United States. Here, we found WA1-lineage viruses 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. 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 United States 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 one to two DNA base substitutions per month across its ~29-kb genome (*27*). Thus, only one to three SNVs in general are needed to define a distinct lineage. The WA1 lineage consists of three key SNVs (C8782T, C18060T, and T28144C), whereas 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 who do not get tested (*2*–*4*). 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

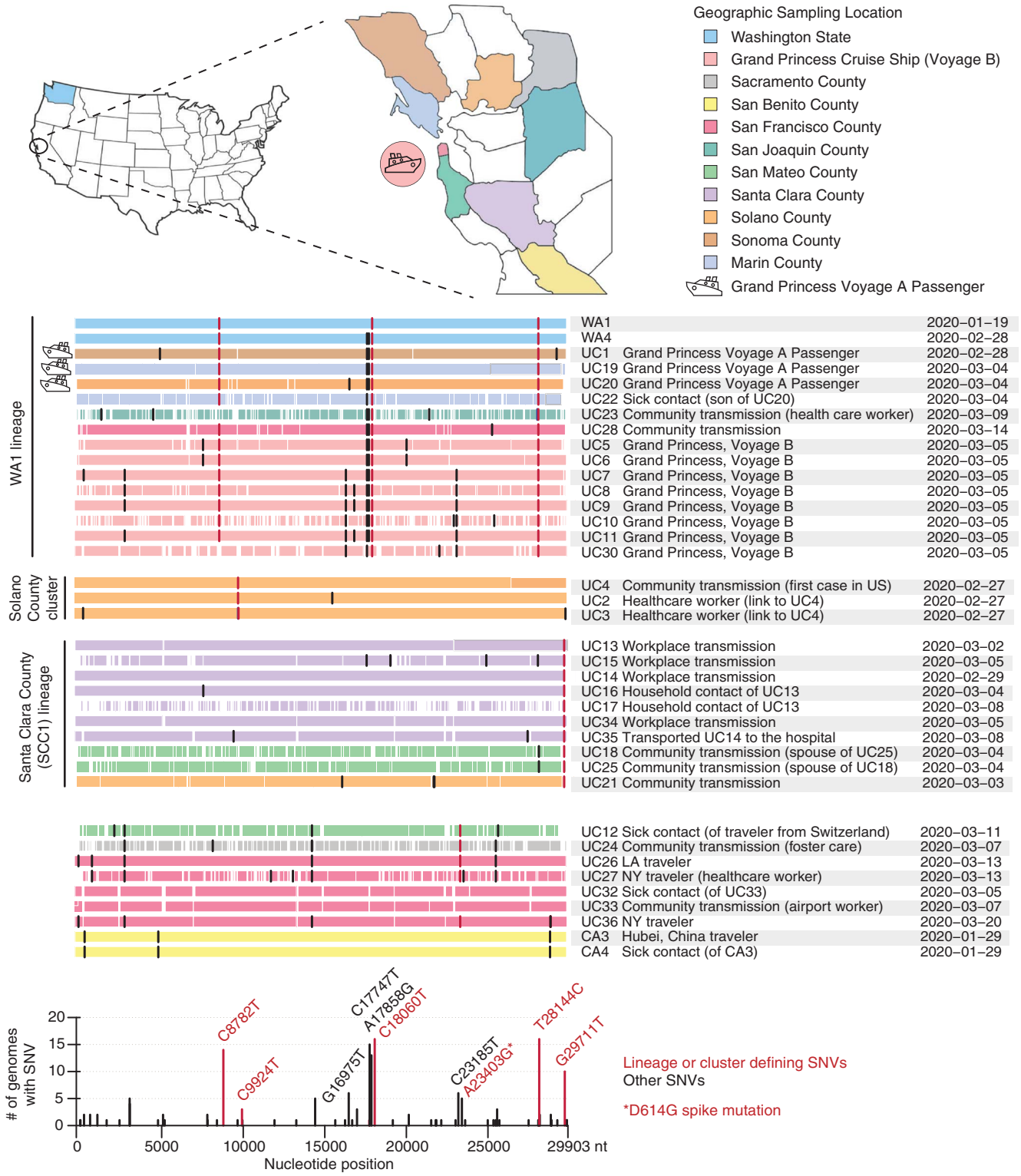

Fig. 3. Multiple sequence alignment of all SARS-CoV-2 genomes reported across nine Northern California counties and the Grand Princess cruise ship. SNVs with respect to the reference genome (NC_045512) are shown as vertical red and black lines for lineage-defining SNVs and other SNVs, respectively. Cases that are part of the WA1 lineage include the first case of COVID-19 infection (WA1) in the United States, eight passengers and crew members aboard the Grand Princess cruise ship during its second trip (voyage B), and three individuals surveyed from three 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 and 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. European lineages share SNV A23403G. The putative epidemiological link and sample collection date are shown beside each sequence alignment.

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 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 the detection of a virus from the SCC1 lineage in residents of neighboring San Mateo and Solano counties. Social distancing interventions, such as the "shelter-in-place" directive that was issued by the governor of California on 20 March 2020, may have assisted 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 Bay Area in the current study. Suspension of non-essential travel may help to prevent importation of new cases in California and other states.

REFERENCES AND NOTES

1. E. Dong, H. Du, L. Gardner, *Lancet Infect. Dis.* **20**, 533–534 (2020).
2. J. F. Chan *et al.*, *Lancet* **395**, 514–523 (2020).
3. R. Li *et al.*, *Science* **368**, 489–493 (2020).
4. R. Lu *et al.*, *Lancet* **395**, 565–574 (2020).
5. A. Otani, P. Santilli, *Wall Street Journal*, 29 February 2020.
6. T. Bedford *et al.*, medRxiv 20051417 [preprint]. 16 April 2020.
7. M. L. Holshue *et al.*, *N. Engl. J. Med.* **382**, 929–936 (2020).
8. C. Fraser *et al.*, *Science* **324**, 1557–1561 (2009).
9. J. L. Gardy, N. J. Loman, *Nat. Rev. Genet.* **19**, 9–20 (2018).
10. R. A. Urbanowicz *et al.*, *Cell* **167**, 1079–1087.e5 (2016).
11. S. E. Mate *et al.*, *N. Engl. J. Med.* **373**, 2448–2454 (2015).
12. N. R. Faria *et al.*, *Nature* **546**, 406–410 (2017).
13. J. Thézé *et al.*, *Cell Host Microbe* **23**, 855–864.e7 (2018).
14. N. D. Grubaugh *et al.*, *Nature* **546**, 401–405 (2017).
15. X. Deng *et al.*, *Nat. Microbiol.* **5**, 443–454 (2020).
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 *et al.*, *Nat. Protoc.* **12**, 1261–1276 (2017).
18. C. Y. Chiu, S. A. Miller, *Nat. Rev. Genet.* **20**, 341–355 (2019).
19. W. Gu, S. Miller, C. Y. Chiu, *Annu. Rev. Pathol.* **14**, 319–338 (2019).
20. A. S. Gonzalez-Reiche *et al.*, *Science* **368**, eabc1917 (2020).
21. M. T. Maurano *et al.*, medRxiv 20064931 [preprint]. 23 April 2020.
22. A. Brufsky, *J. Med. Virol.* jmv.25902 (2020).
23. L. F. Moriarty *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **69**, 347–352 (2020).
24. M. Klompas, *Ann. Intern. Med.* **172**, 619–620 (2020).
25. J. Hadfield *et al.*, *Bioinformatics* **34**, 4121–4123 (2018).
26. J. R. Fauver *et al.*, *Cell* **181**, 990–996.e5 (2020).
27. M. Bouvet *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9372–9377 (2012).
28. V. Hill, A. Rambaut, Virological.org (6 March 2020); https://virological.org/t/phylodynamic-analysis-of-sars-cov-2-update-2020-03-06/420.
29. W. Gu *et al.*, *Clin. Infect. Dis.* ciaa599 (2020).
30. L. du Plessis *et al.*, Zenodo DOI:10.5281/zenodo.3922369 (2020).

ACKNOWLEDGMENTS

We acknowledge the help and advice of J. T. Lee at the U.S. CDC. We thank all of the authors who have contributed genome data on GISAID. Author credits for specific GISAID contributions can be found at www.gisaid.org/. Clinical samples from UCSF were processed according to protocols approved by the UCSF Institutional Review Board (protocol numbers 10-01116 and 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 nonresearch 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:** Supported by NIH grants R33-AI129455 (C.Y.C.) from the National Institute of Allergy and Infectious Diseases and K08-CA230156 (W.G.) from the National Cancer Institute, the California Initiative to Advance Precision Medicine (C.Y.C.), the Charles and Helen Schwab Foundation (C.Y.C.), the Burroughs-Wellcome CAMS Award (W.G.), the Oxford Martin School and the European Research Council under the Seventh Framework Program of the European Commission (Pathogen Phylodynamics grant 614725) (O.G.P. and L.d.P.), Wellcome Trust and Royal Society (204311/Z/16/Z), and Medical Research Council–São Paulo Research Foundation (MR/S0195/1, FAPESP 18/14389-0). **Author contributions:** C.Y.C. conceived, designed, and supervised the study; A.G., A.S.-G., C.W., C.-Y.P., G.Y., H.G., V.S., and X.D. performed experiments; C.Y.C., S.F., and B.B.. assembled and curated the viral genomes; C.Y.C., W.G., and X.D. analyzed data; C.Y.C., E.H., K.Z., S.M., and W.G. collected patient samples at UCSF; D.M. and G.A. analyzed genomic and epidemiologic data; T.B., A.L.G., P.R., L.M.S., M.F., H.Y.C., J.S., and K.R.J. collected, assembled, and provided viral genome data from Washington and contributed to the phylogenetic analysis; C.A., K.G., M.Z., E.S., and K.G.A. provided viral genome data from Southern California; C.R.P., J.Z., S.T., and Y.L. sequenced and analyzed viral genomes at the CDC; L.d.P., N.F., and O.G.P. performed phylogenetic analysis of genomes; A.T.Y., B.T.M., B.B., C.C., D.A.W., E.V., G.M., M.P., M.W., O.K., S.J.C., and S.M. collected samples, extracted the viral RNA, and/or provided epidemiology data from counties in California; C.Y.C., W.G., and X.D. wrote the manuscript; and C.Y.C., X.D., S.F., C.W., D.A.W., L.d.P., O.G.P., and W.G. edited the manuscript. All authors read the manuscript and agree to its contents. **Competing interests:** C.Y.C. is the director of the UCSF-Abbott Viral Diagnostics and Discovery Center and receives research support funding from Abbott Laboratories. C.Y.C. and X.D. are inventors on a patent application on the MSSPE method titled "Spiked Primer Design for Targeted Enrichment of Metagenomic Libraries" (U.S. Application No. 62/667,344, filed 05/04/2018 by University of California, San Francisco). H.Y.C. 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 to 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*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/369/6503/582/suppl/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S5
Data S1
References (*31*–*39*)

View/request a protocol for this paper from *Bio-protocol*.

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