

Associations of Early COVID-19 Cases in San Francisco with Domestic and International Travel

Wei Gu^{1,2&}, Xianding Deng^{1,2&}, Kevin Reyes^{1,2}, Elaine Hsu^{1,2}, Candace Wang^{1,2}, Alicia Sotomayor-Gonzalez^{1,2}, Scot Federman^{1,2}, Brian Bushnell³, Steve Miller^{1,2}, Charles Chiu^{1,2,4*}

¹Department of Laboratory Medicine, University of California, San Francisco, California, USA

²UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, California, USA

³DOE Joint Genome Institute, Walnut Creek, CA, United States of America

⁴Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, California, USA

*Corresponding Author: Charles Chiu, Charles.Chiu@ucsf.edu

&These authors have equal contribution

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Abstract

In early-to-mid March 2020, 20 of 46 (43%) COVID-19 cases at a tertiary care hospital in San Francisco, California were travel-related. Cases were significantly associated with travel to Europe or New York (odds ratio 32.9). Viral genomes recovered from 9 of 12 (75%) cases co-clustered with lineages circulating in Europe.

Accepted Manuscript

Main Text

As of April 4th, 2020, the COVID-19 pandemic, caused by the novel SARS-CoV-2 coronavirus¹, has infected more than 1.2 million people worldwide and the rise in cases has been exponential. In particular, New York cases in the United States quickly surged from 22 to >10,000 between March 10 and 22². By April 4th there are >150,000 cases in New York and nearby New Jersey, threatening to overwhelm hospitals and other regional health care systems in the city.

In San Francisco, we validated a qRT-PCR test to detect SARS-CoV-2 infection from nasopharyngeal swab samples based on the EUA (Emergency Use Authorization)-approved US CDC assay³. During the first 10 days since launch, we performed SARS-CoV-2 testing on 947 samples collected from March 10 through March 20 from patients with suspected SARS-CoV-2 infection at University of California, San Francisco. We reviewed the electronic medical records from the first 46 consecutive SARS-CoV-2 positive cases admitted to University of California, San Francisco hospitals or seen in outpatient clinics from March 10 through March 20. Data from these COVID-19 patients were matched with 102 randomly selected negative controls who were patients who tested negative for SARS-CoV-2 over the same time period. Documented history was recorded by a physician or nurse practitioner and included sick contacts, health care worker status, and travel history. Among the 46 COVID-19 positive patients, the median age was 44 years, 46% were female, and 65% were outpatients (**Table S1**).

We noted that a travel history within 2 weeks of symptom onset (median date Mar 11, 2020) was significantly associated with COVID-19 infection (OR 3.8 [1.8-8.4]),

comprising 43% (20/46) of newly diagnosed cases (**Figure 1a**). Out of the 20 travelers with COVID-19 infection, there were significant associations for prior travel to Europe (5 travelers, OR 6.1 [1.1-32.7]), USA (14 travelers, OR 4.0 [1.6-10.0]), and specifically New York (6 travelers, OR 32.9 [1.8-598]) as compared to 17 travelers without infection (**Figure 1b and Table S2, S3**). The association with travel may be due to direct exposure to SARS-CoV-2 while in high prevalence regions (e.g. NY) or exposure while traveling (close contact with fellow travelers or airport personnel). One cluster of 3 positive cases associated with COVID-19 infection in an airport worker was categorized as a case of community rather than travel-associated transmission. No significant associations were found with regards to close contacts with known COVID-19 infected persons or frontline healthcare workers. Those who did not have a recent travel history, a close contact who was COVID-19 positive, or were not a frontline healthcare worker were categorized as community transmission with an unknown source of infection and comprised 39% of cases.

We conducted viral genomic sequencing and phylogenetic analysis of SARS-CoV-2 viruses from 12 of 20 travelers for whom the breadth of coverage of the viral genome was >90%⁴⁻⁶. These viral genomes were aligned using MAFFT v7.427⁷ with 762 high-coverage viral genomes deposited in the GISAID database^{8,9} as of March 20, 2020, in addition to the most recent viral genomes sequenced in California as of May 3, 2020⁴, for a total of 983 sequences. A maximum likelihood phylogenetic tree was constructed using IQTREE (version 2) using an HKY substitution model¹⁰ (**Figure 2**).

We defined genomic clades through the GISAID nomenclature found at that point in time on March 20, 2020^{8,9}. The majority (9 of 12) of all travel cases clustered in the G

clade as defined by the spike protein D614G variant marker (**Figure 2, S1, S2**), including 3 cases from Europe (UC40, UC45, UC46), 4 cases from New York (UC27, UC36, UC44, UC47), 1 case from Los Angeles (UC26), and 1 case from Chicago (UC48). Viruses in the G clade comprise most of the genomes sequenced from patients in Europe^{8,9}, but notably have also been identified in the vast majority of cases associated with the New York SARS-CoV-2 outbreak in March to April of 2020, which occurred after the timeline of this study^{11,12}. The detection of G clade viral genomes in travelers to Los Angeles and Chicago suggests the possibility of dissemination of this clade to other states, either indirectly via New York or directly from Europe. Another case involving travel to Denver (UC42) was part of the WA1 lineage, which is associated with the first reported case of SARS-CoV-2 infection in the United States and is currently circulating in local communities in Washington State^{4,13} and California⁴. Viruses from two additional travel-associated cases from Europe (UC43) and New York (UC41) were mapped to other clades circulating in Europe (**Figure 2**). The additional case from Europe was found to be part of the V clade, defined by a G251V mutation in the NS3 protein^{8,9}.

Limitations of our study include the use of epidemiological data from only the first 10 days of testing at a single institution. Nevertheless, in the setting of an emergent pandemic with shifting epidemiology, the results of our study reached statistical significance over 4 categories of travel (all travel, New York, USA, and Europe), and yielded data that may have presaged the exponential rise of New York cases and subsequent large-scale outbreak in the New York metropolitan area^{11,12}.

Real-time dissemination of epidemiological survey data from positive COVID-19 cases is critical to support efforts to contain or reduce spread of viral infection in the community. Our evaluation of diagnosed COVID-19 cases in San Francisco in early March 2020 associates with travelers from New York prior to the recognized spike in New York cases in late March (**Figure 1c**). Travel from New York was underrecognized as a risk factor for COVID-19 infection in the United States in early March. Guidelines for COVID-19 testing have not included screening for domestic travel. Our findings in San Francisco here can be extrapolated across America as there are over 100 direct domestic destinations and more than 6 million domestic flights a month from JFK, Newark, and LaGuardia airports in the New York metropolitan area¹⁴. Similarly, travel by motor vehicle or train is also a plausible means of spread, especially if there are disproportionate numbers of cases between closely situated major population centers. Cryptogenic transmission of COVID-19 by individuals with mild illness or asymptomatic infection is a tremendous challenge to the containment of COVID-19^{15,16}. As demonstrated here, stratifying the general population by their exposure risks such as travel to specific hotspot regions is one containment strategy that can be informed by real-time epidemiological and phylogenetic surveillance.

NOTES

Data Availability

Assembled SARS-CoV-2 genomes in this study were uploaded to GISAID^{8,9} as FASTA files and were also submitted to the National Center for Biotechnology Information (NCBI) GenBank database (accession numbers pending). Raw sequence data were submitted to the NCBI Sequence Read Archive (SRA) database (BioProject accession number PRJNA629889 and umbrella BioProject accession number PRJNA171119). De-identified data from UCSF patients is available upon request.

The institutional review board at University of California, San Francisco approved the clinical and epidemiological association study (IRB #20-30538) and the phylogenetic study (IRB #10-01116, 11-05519). Non-identifying clinical, demographic, and laboratory data were extracted from clinical testing results and the electronic medical board by retrospective chart review. Informed consent was waived for this minimal risk study.

Funding

This work was funded by the National Institutes of Health NIH grants R33-AI129455 (CYC) from the National Institute of Allergy and Infectious Diseases and K08-CA230156 (WG) from the National Cancer Institute, the Charles and Helen Schwab Foundation (CYC), and the Burroughs-Wellcome Career Awards for Medical Scientists Award (WG).

Competing Interests

CYC is the director of the UCSF-Abbott Viral Diagnostics and Discovery Center (VDDC) and receives research support funding from Abbott Laboratories. The other authors have declared no competing interest.

Accepted Manuscript

References

1. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 2020;395(10224):565–74.
2. NYS-COVID19-Tracker [Internet]. [cited 2020 Mar 30];Available from: <https://covid19tracker.health.ny.gov>
3. US Centers for Disease Control and Prevention. Real-Time RT-PCR Panel for Detection 2019-nCoV [Internet]. 2020;Available from: <https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-detection-instructions.html>
4. Deng X, Gu W, Federman S, et al. A Genomic Survey of SARS-CoV-2 Reveals Multiple Introductions into Northern California without a Predominant Lineage. *medRxiv* 2020;2020.03.27.20044925.
5. Deng X, Achari A, Federman S, et al. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nature Microbiology* 2020;5(3):443–54.
6. Quick J, Grubaugh ND, Pullan ST, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nature Protocols* 2017;12(6):1261–76.
7. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30(14):3059–66.
8. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* 2017;1(1):33–46.
9. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;34(23):4121–3.
10. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol* 2015; 32(1):268–274.
11. Gonzalez-Reiche AS, Hernandez MM, Sullivan M, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *medRxiv* 2020;2020.04.08.20056929.
12. Maurano MT, Ramaswami S, Westby G, et al. Sequencing identifies multiple, early introductions of SARS-CoV2 to New York City Region. *medRxiv* 2020;2020.04.15.20064931.

13. Bedford T, Greninger AL, Roychoudhury P, et al. Cryptic transmission of SARS-CoV-2 in Washington State. medRxiv 2020;2020.04.02.20051417.
14. Port Authority of New York and New Jersey Airport Traffic Statistics [Internet]. [cited 2020 Mar 27];Available from: <https://www.panynj.gov/airports/en/statistics-general-info.html>
15. Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science [Internet] 2020 [cited 2020 Mar 19];Available from: <https://science.sciencemag.org/content/early/2020/03/13/science.abb3221>
16. McMichael TM, Currie DW, Clark S, et al. Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. New England Journal of Medicine 2020,doi:10.1056/NEJMoa2005412.

Accepted Manuscript

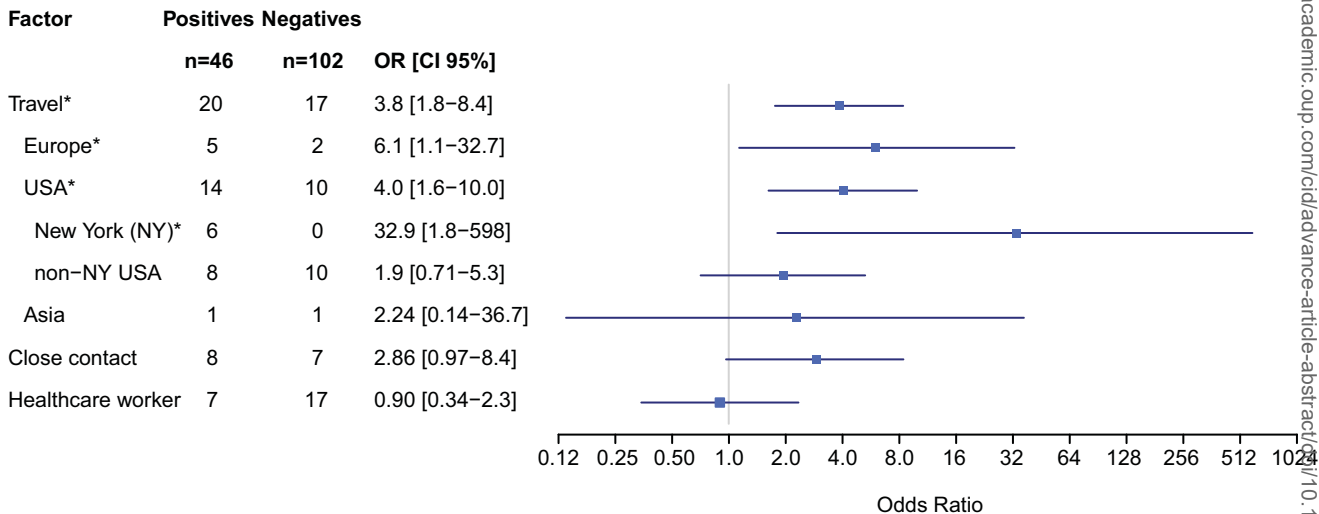
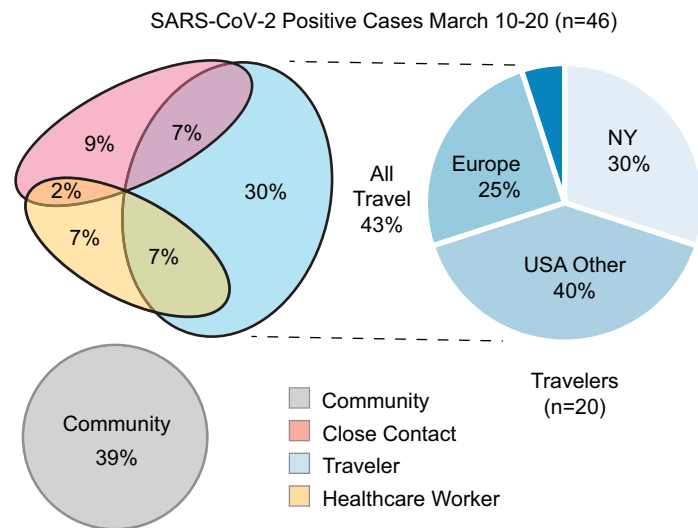
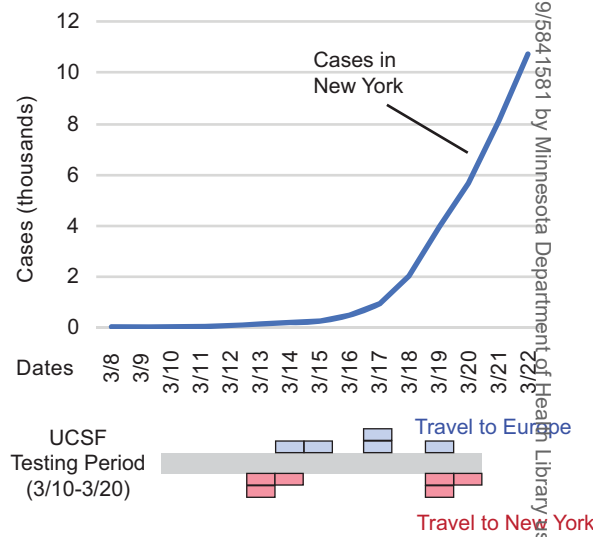
FIGURES

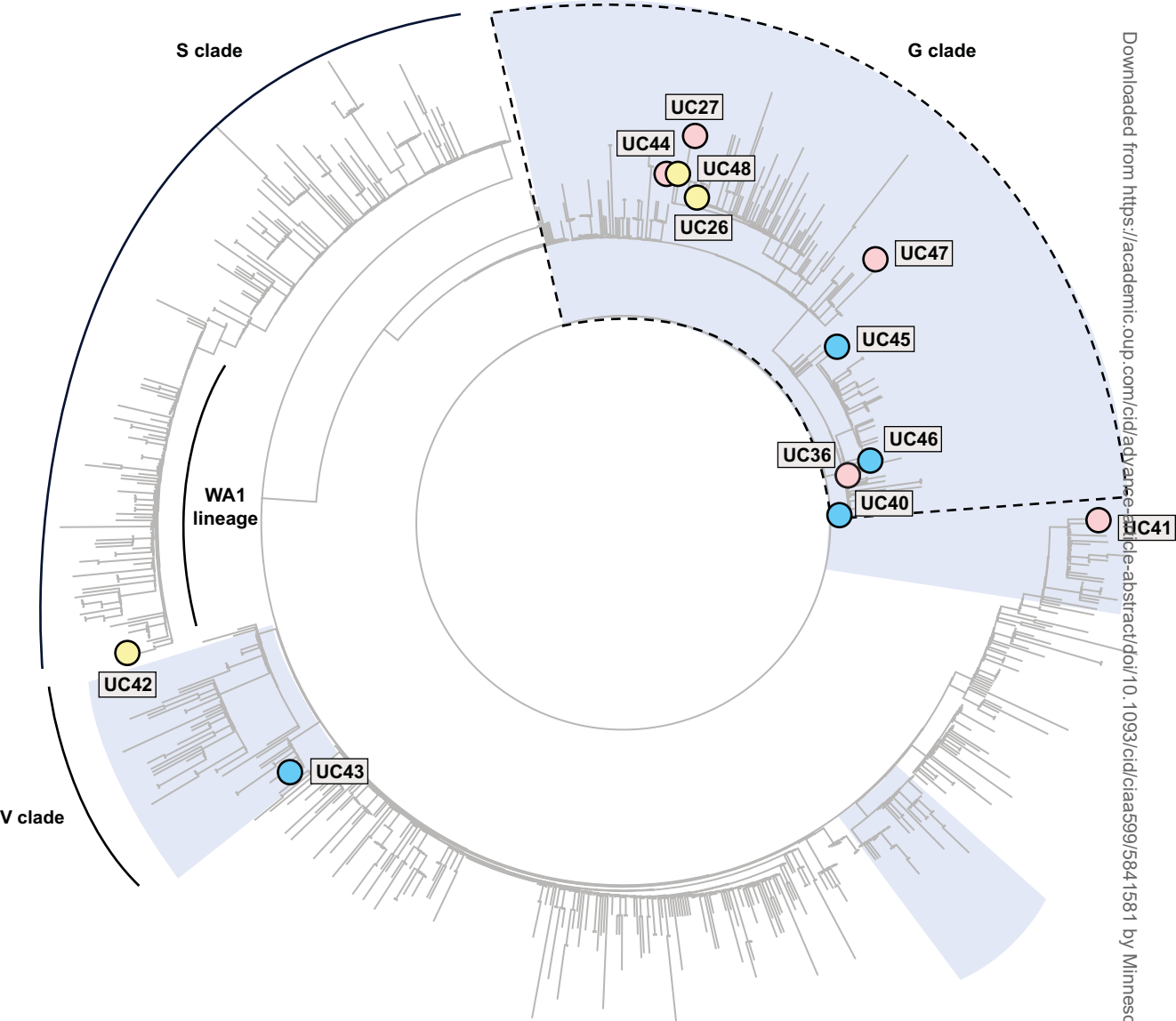
Figure 1. (A) Associations with Positive COVID-19 RT-PCR Testing. The odds ratios (OR) with 95% confidence intervals (CI) are displayed. Positives (n=46) were consecutive cases from March 10 to 20, 2020. Negatives (n=102) were randomized from the same time period. Significant risk factors ($p < 0.05$) are designated with '*' and were recent travel, including Europe, USA (domestic), and/or New York (p-values are found in **Table S2**). Abbreviations: Pos, positive; Neg, negative; OR, odds ratio; CI, confidence interval.

(B) Venn diagram of risk factors in positive SARS-CoV-2 cases. All positive cases and their associations are shown here categorized as those with a recent travel history, had a close contact who was COVID-19 positive, a frontline healthcare worker, or a combination of the previous categories (left). Those who did not match one of those categories were uniformly categorized as a community case. The most common association with a positive case was a travel history immediately prior to symptoms. Travelers (n=20) are subdivided by travel region: New York (NY), non-NY USA, Europe, or Asia (right).

(C) Timeline of cumulative COVID-19 cases diagnosed in New York (top), and UCSF positive cases found in San Francisco who recently traveled to New York or Europe over time (bottom). Each colored block represents a single patient.

Figure 2: Phylogenetic Analysis of SARS-CoV-2 Viral Genomes from Domestic and International Travelers. The 12 cases with sufficient viral genome coverage for phylogenetic analysis ($\geq 90\%$) are highlighted by colored circle overlaying a global phylogenetic tree of 983 viruses, including 762 viruses in GISAID as of March 20, 2020 and most recent viral genomes sequenced from California patients. The G, S, and V clades and the lineages dominated by genomes in Europe are highlighted.

A**B****C**



UC26	● Los Angeles, CA USA	2020-03-13
UC27	● New York City, NY USA	2020-03-13
UC36	● New York City, NY USA	2020-03-13
UC40	● Spain	2020-03-14
UC41	● New York City, USA	2020-03-14
UC42	● Denver, CO USA	2020-03-14
UC43	● London, United Kingdom	2020-03-14
UC44	● New York City, NY, USA	2020-03-19
UC45	● Italy	2020-03-19
UC46	● France	2020-03-17
UC47	● New York City, NY, USA	2020-03-20
UC48	● Chicago, IL, USA	2020-03-20

● travel to Europe

● travel to New York

● travel to other states in the US

lineages dominated by genomes in Europe

clade G