REPORT

CORONAVIRUS

Cryptic transmission of SARS-CoV-2 in Washington state

Trevor Bedford[1,2,3]*†, Alexander L. Greninger[1,4]†, Pavitra Roychoudhury[1,4]†, Lea M. Starita[2,3]†, Michael Famulare[5]†, Meei-Li Huang[1,4], Arun Nalla[4], Gregory Pepper[4], Adam Reinhardt[4], Hong Xie[4], Lasata Shrestha[4], Truong N. Nguyen[4], Amanda Adler[6], Elisabeth Brandstetter[7], Shari Cho[2,3], Danielle Giroux[3], Peter D. Han[2,3], Kairsten Fay[1], Chris D. Frazar[3], Misja Ilcisin[1], Kirsten Lacombe[6], Jover Lee[1], Anahita Kiavand[2,3], Matthew Richardson[3], Thomas R. Sibley[1], Melissa Truong[2,3], Caitlin R. Wolf[7], Deborah A. Nickerson[2,3], Mark J. Rieder[2,3], Janet A. Englund[2,6,8], The Seattle Flu Study Investigators‡, James Hadfield[1], Emma B. Hodcroft[9,10], John Huddleston[1,11], Louise H. Moncla[1], Nicola F. Müller[1], Richard A. Neher[9,10], Xianding Deng[12], Wei Gu[12], Scot Federman[12], Charles Chiu[12], Jeffrey S. Duchin[7,13], Romesh Gautom[14], Geoff Melly[14], Brian Hiatt[14], Philip Dykema[14], Scott Lindquist[14], Krista Queen[15], Ying Tao[15], Anna Uehara[15], Suxiang Tong[15], Duncan MacCannell[16], Gregory L. Armstrong[16], Geoffrey S. Baird[4], Helen Y. Chu[2,7]§, Jay Shendure[2,3,17]§, Keith R. Jerome[1,4]§

After its emergence in Wuhan, China, in late November or early December 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus rapidly spread globally. Genome sequencing of SARS-CoV-2 allows the reconstruction of its transmission history, although this is contingent on sampling. We analyzed 453 SARS-CoV-2 genomes collected between 20 February and 15 March 2020 from infected patients in Washington state in the United States. We find that most SARS-CoV-2 infections sampled during this time derive from a single introduction in late January or early February 2020, which subsequently spread locally before active community surveillance was implemented.

The novel coronavirus, referred to alternately as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (1) or human coronavirus 2019 (hCoV-19) (2), emerged in Wuhan, Hubei, China, in late November or early December 2019 (3). As of 18 May 2020, there have been >4 million confirmed cases of coronavirus disease 2019 (COVID-19)—the disease caused by SARS-CoV-2—that have resulted in >300,000 deaths (4). After its initial emergence in China, travel-associated cases with travel histories related to Wuhan appeared in other parts of the world (5). The first confirmed case in the United States was travel associated and was detected in Snohomish County, Washington state, on 19 January 2020. Until 27 February 2020, the U.S. Centers for Disease Control and Prevention (CDC) guidance recommended prioritizing testing for COVID-19 on persons with direct travel history from an affected area or with exposure to a known case. Cases of respiratory disease with no known risk factors were not routinely tested. In the 6 weeks between 19 January and 27 February, 59 confirmed cases were reported in the United States (6), all outside of Washington state and with either direct travel history or exposure to a known, confirmed case. On 28 February 2020, a community case was identified in Snohomish County (7). One month later, on 25 March, as a result of increased testing and ongoing transmission, Washington state reported 2580 confirmed cases and 132 deaths (8). Here, we report on the putative history of early community transmission in Washington state as revealed by genomic epidemiology. We conclude that SARS-CoV-2 was circulating for several weeks undetected by the surveillance apparatus in Washington state from late January to early February 2020.

Although publicly available SARS-CoV-2 genomes (9, 10) are not sampled in strict proportion to the burden of infections through time and across geography, their genetic relationships can still shed light on underlying patterns of spread. SARS-CoV-2 genomes sampled between December 2019 and 15 March 2020 appear to be closely related, with between 0 and 12 mutations relative to a common ancestor estimated to exist in Wuhan between late November and early December 2019 (Fig. 1). This pattern is consistent with a reported rate of molecular evolution of ~0.8 × 10^{-3} substitutions per site per year or approximately two substitutions per genome per month (3). After its initial zoonotic emergence in Wuhan (11), SARS-CoV-2 viral genomes began to accumulate substitutions and spread from Wuhan to other regions in the world (3). During December 2019, the Wuhan outbreak was too small to seed many introductions outside of China, but by January 2020, it had grown large enough to begin seeding cases elsewhere (12).

Sequencing of viruses from the Washington state outbreak began on 28 February 2020 and has continued since then. We analyzed the sequences of 455 SARS-CoV-2 viruses from this outbreak collected between 19 January and 15 March 2020 (Fig. 1). Virus sequences from Washington state are closely related to those from viruses collected elsewhere. Clusters of closely related viruses indicate separate introduction events followed by local spread. The majority (n = 384; 84%) of these viruses fall into a closely related clade (marked by the larger arrow in Fig. 1), and these viruses have single-nucleotide polymorphisms (SNPs) C8782T, C17747T, A17858G, C18060T, and T28144C relative to the basal virus at the root of the phylogeny, which is equivalent to the reference virus Wuhan/Hu-1/2019. This clade derives from viruses circulating in China (Fig. 1, in blue), is closely related to viruses sampled in British Columbia (Fig. 1, in orange), and is labeled as Pangolin lineage A.1 (13). Going forward, we refer to this clade as the Washington state outbreak clade. Other viruses (n = 39; 9%) fall into a separate, smaller clade (marked by the smaller arrow in Fig. 1) and derive from viruses circulating in Europe. The remaining 33 viruses (7%) from Washington state are distributed across the phylogeny. Thus, we conclude that most early cases descend from a single introduction event followed by local amplification.

The Washington state outbreak clade has a highly comb-like structure (Fig. 2A), which is indicative of rapid exponential growth (14). This clade has a C17747T change relative to viruses sampled in British Columbia and a A17858G change relative to viruses sampled in

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [2]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. [3]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [4]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. [5]Institute for Disease Modeling, Bellevue, WA, USA. [6]Division of Infectious Disease, Seattle Children's Hospital, Seattle, WA, USA. [7]Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, WA, USA. [8]Department of Pediatrics, University of Washington, Seattle, WA, USA. [9]Biozentrum, University of Basel, Basel, Switzerland. [10]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [11]Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA. [12]Department of Laboratory Medicine, University of California San Francisco, San Francisco, CA, USA. [13]Public Health — Seattle & King County, Seattle, WA, USA. [14]Washington State Department of Health, Shoreline, WA, USA. [15]Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA. [16]Office of Advanced Molecular Detection, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA. [17]Howard Hughes Medical Institute, Seattle, WA, USA.
*Corresponding author. Email: tbedford@fredhutch.org
†These authors contributed equally to this work.
‡The Seattle Flu Study Investigators and affiliations are listed in the supplementary materials.
§These authors contributed equally to this work.

Fig. 1. Maximum-likelihood phylogeny of 455 SARS-CoV-2 viruses collected from Washington state on a background of 493 globally collected viruses. Viruses collected from Washington state are shown as red circles. Tips and branches are colored on the basis of location, branch lengths are proportional to the number of mutations along a branch, and the *x* axis is labeled with the number of substitutions relative to the root of the phylogeny—here equivalent to basal Wuhan outbreak viruses. The clustering of related viruses indicates community transmission after an introduction event. Branch locations are estimated on the basis of a discrete traits model. We observe a single introduction leading to a large outbreak clade of 384 sampled viruses from Washington state (marked by the larger arrow), and we observe a second introduction leading to a smaller outbreak clade of 39 viruses (marked by the smaller arrow). An interactive version of this figure is available at https://nextstrain.org/community/blab/ncov-cryptic-transmission/introductions.
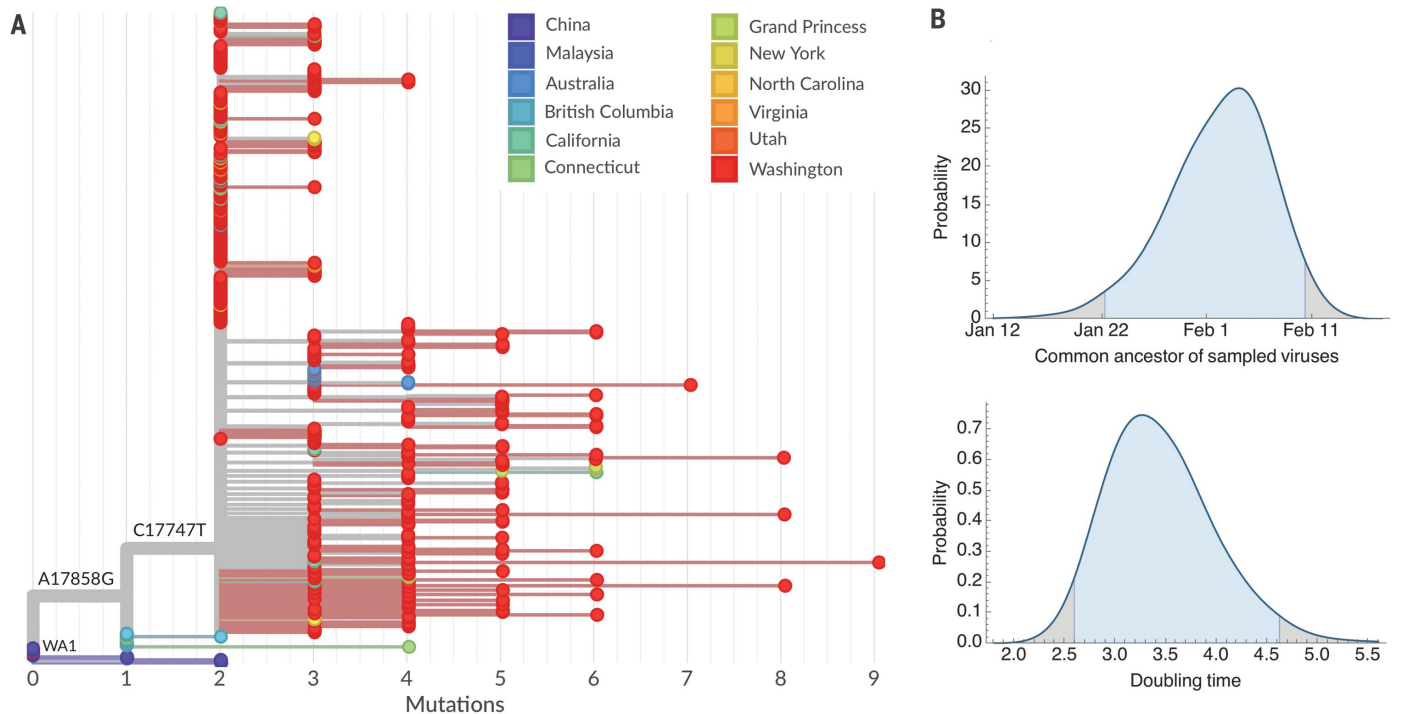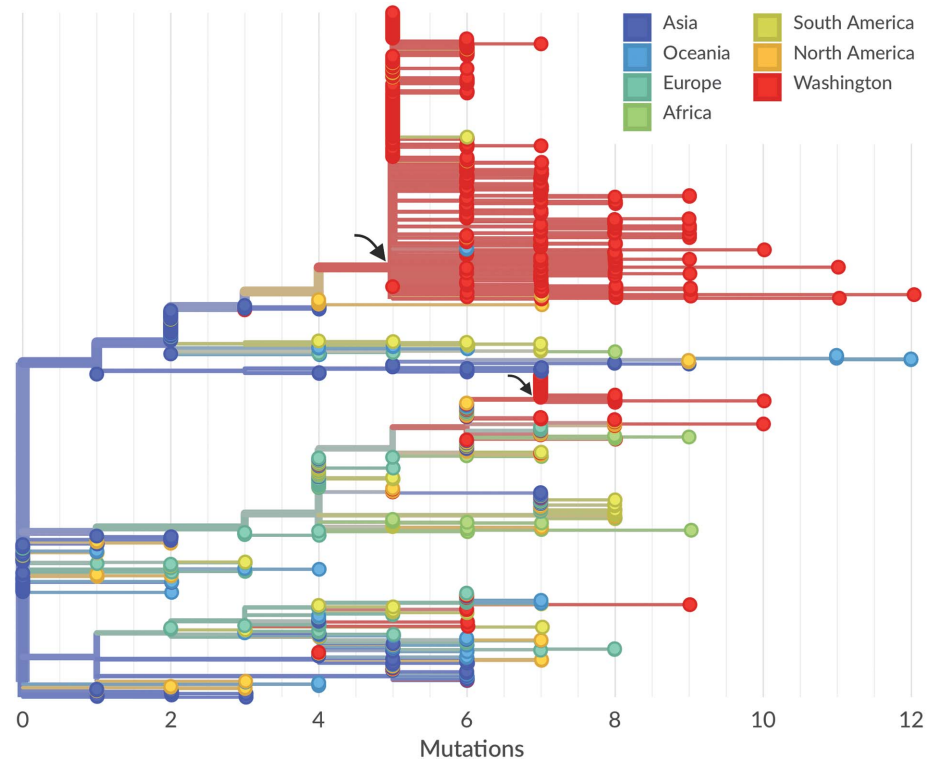
Fig. 2. Maximum-likelihood phylogeny of the Washington state outbreak clade and immediately ancestral variants containing 448 SARS-CoV-2 viruses and Bayesian estimates of the date of the outbreak common ancestor and outbreak doubling time. (**A**) Maximum-likelihood phylogeny. Tips are colored on the basis of location, branch lengths are proportional to the number of mutations between viruses, and the *x* axis is labeled with the number of substitutions relative to the root of the phylogeny—here equivalent to the WA1 haplotype. This comb-like phylogenetic structure of the Washington state outbreak clade is consistent with rapid exponential growth of the virus population. An interactive version of this figure is available at https://nextstrain.org/community/blab/ncov-cryptic-transmission/wa-clade. (**B**) Highest posterior density estimates for the date of the common ancestor of viruses from the Washington state outbreak clade (top) as well as the doubling time in days of the growth of this clade (bottom).

Fujian, Chongqing, Hangzhou, and Guangdong. Given the limited and nonrepresentative sampling of viruses for sequencing, along with the rate of molecular evolution, it is difficult to make detailed assessments of geographic origins. However, we can be confident that this clade represents an introduction from China followed by local spread within the United States and Canada. British Columbia may have been the entry point or the location at which the first virus was sampled.

We analyzed the Washington state outbreak clade in a coalescent analysis to estimate evolutionary dynamics. Here, we assume a prior on evolutionary rate based on analysis of viruses sampled globally between December 2019 and July 2020 (see materials and methods). This analysis uses the degree and pattern of genetic diversity of sampled genomes to estimate the date of a common ancestor and the exponential growth rate of the virus population. We obtained a median estimate for the date of the clade's common ancestor of 2 February 2020,

with a 95% Bayesian credible interval of 22 January to 10 February 2020 (Fig. 2B). We note that the initiation of a transmission chain may slightly predate the common ancestor belonging to this chain in sampled viruses, as initial transmission events after introduction may not result in branching of the transmission tree. We calculated a rate of exponential growth from the coalescent analysis for this clade and found a median doubling time of 3.4 days, with a 95% Bayesian credible interval of 2.6 to 4.6 days (Fig. 2B).

In addition to the 384 viruses from Washington state identified in the Washington state outbreak clade, we observed 12 viruses from elsewhere, including from California, Connecticut, Minnesota, New York, North Carolina, Virginia, Utah, Australia, and the Grand Princess cruise ship (Fig. 2A). Viruses from outside Washington state nest within the diversity found in Washington state. In the case of the Grand Princess, the genetic relationship among these viruses is consistent with a single introduction onto the

cruise ship of the basal outbreak variant—having C17747T and A17858G changes—and subsequent transmission and evolution on the ship.

The first confirmed case recorded in the United States was a travel-associated case from an individual returning from Wuhan on 15 January 2020, who presented for care at an outpatient clinic in Snohomish County on 19 January 2020 and tested positive (*15*). This infection is recorded as strain USA/WA1/2020 (referred to here as WA1 and annotated in Fig. 2A), and it appears to be closely related to viruses from infections in China (Fujian, Hangzhou, and Guangdong provinces). Viruses from the Washington state outbreak clade group together as direct descendants of WA1 and its identical relatives (Fig. 2A). This tree structure is consistent with the WA1 strain transmitting locally after arrival into the United States. The rarity of the C8782T, T28144C, and C18060T mutations—characteristic of WA1—in viruses sampled from China (found in 6 of 224 or 3% of sequenced viruses) indicates that this is a parsimonious explanation for the origin of the Washington state outbreak clade. However, because the evolution rate for SARS-CoV-2 (one mutation per ~15 days) is slower than the transmission rate (one transmission event every 4 to 8 days) (*16*, *17*), it is possible that WA1 sits on a side branch of the underlying transmission tree even if it appears as a direct ancestor in the maximum-likelihood tree. The fact that viruses sampled from British Columbia interdigitate between WA1 and the Washington state outbreak clade indicates that this clade may have been introduced into North America by a closely related infection to—but one distinct from—WA1 (Fig. 2A). Additionally, it remains possible that multiple viruses with the basal Washington state outbreak clade genotype were introduced, which resulted in the local amplification of this clade; however, this is markedly less likely than a single introduction of the virus.

Given that community transmission was first detected on 28 February 2020 from a transmission chain originating between 22 January and 10 February 2020, we sought to address community prevalence during this period. Here, we analyzed 10,382 acute respiratory specimens collected as part of the Seattle Flu Study between 1 January and 15 March 2020 (Fig. 3A). These specimens represented a mix of residual samples collected as part of routine clinical testing and samples collected as part of prospective community enrollment of individuals with acute respiratory illness. In total, 5270 samples collected between 1 January and 20 February tested negative. The first positive sample was collected on 21 February (Fig. 3B). From 21 February to 15 March, of 5112 samples collected, 65 samples tested positive. On
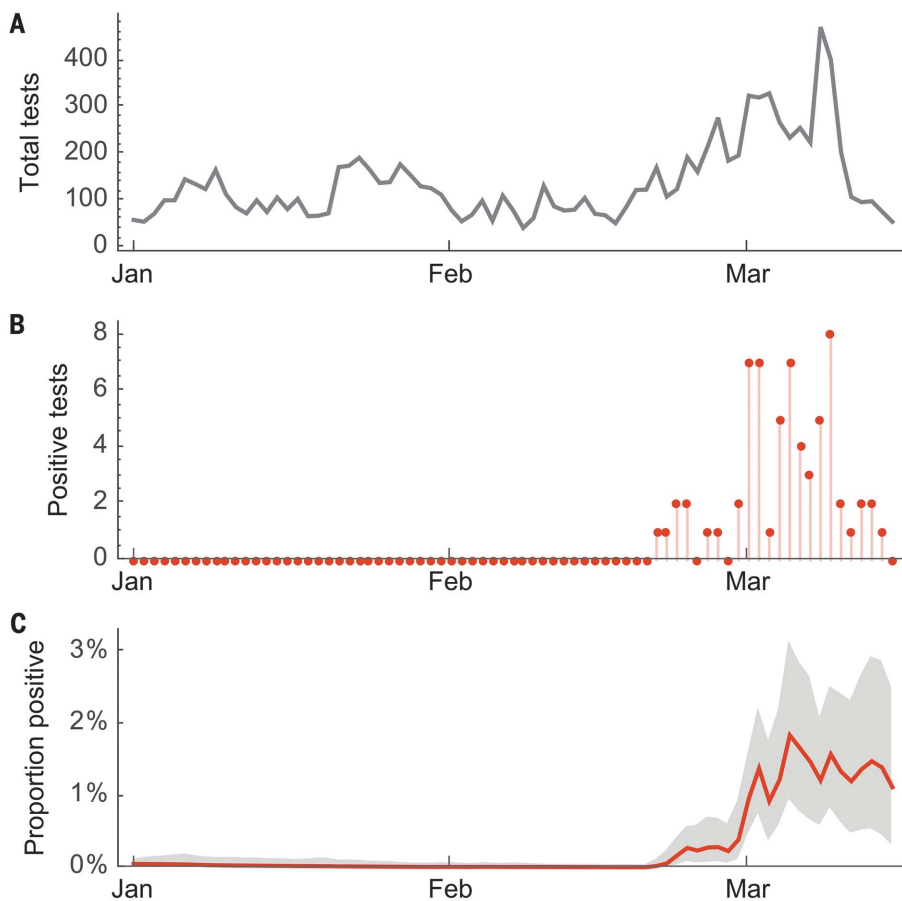

Fig. 3. Acute respiratory samples tested for SARS-CoV-2 collected as part of the Seattle Flu Study between 1 January and 15 March 2020. (**A**) Total samples tested per day. In total, 10,382 samples collected between 1 January and 15 March were tested. (**B**) Number of samples testing positive per day. (**C**) Estimated proportion positive using a sequential Monte Carlo model to provide day-to-day smoothing. The solid red line is the mean estimate of proportion positive, and the gray shaded region is the 95% credible interval. All dates are those of sample collection, not dates of testing.

1 March, a sequential Monte Carlo procedure estimated the proportion of acute respiratory specimens positive for SARS-CoV-2 as 1.1% with a 95% credible interval of 0.5 to 2.0% (Fig. 3C). It is challenging to directly convert this value into population prevalence of SARS-CoV-2; however, U.S. Health Weather data show a 4.5% prevalence of influenza-like illness on 1 March (18), from which we estimated a 0.05% population prevalence of SARS-CoV-2.

In January and February 2020, screening for SARS-CoV-2 in the United States was directed at travelers with fever, cough, and shortness of breath, with the point of origin broadening as new outbreaks were identified but continuing to solely specify travel to China up until 24 February 2020 (19, 20). Our analysis indicates that at least one clade of SARS-CoV-2 had been circulating in the Seattle area for 3 to 6 weeks by the time the virus was first detected in a non-traveler on 28 Feb 2020. By then, variants within this clade constituted the majority of confirmed infections in the region (384 of 455; 84%). Several factors could have contributed to the delayed detection of presumptive community spread, including limited testing among nontravelers or the presence of asymptomatic or mild illnesses.

Both the WA1 strain sampled in Snohomish County, Washington, on 19 January as well as viruses sampled from British Columbia in early March appear to be phylogenetically ancestral to viruses from the Washington state outbreak clade (Fig. 2A), which suggests a possible route of introduction. However, in both of these cases, a lack of comprehensive geographic sampling makes it difficult to rely on phylogenetic structure for transmission inference. Viruses sampled from British Columbia may derive from local spread after a direct introduction event, or they may be offshoots of an introduction elsewhere that subsequently spread to British Columbia. Refining the time and geographic origin of the introduction into Washington state will require a combination of earlier samples and samples from other geographic locations. Other states in the United States have shown different genetic histories from that seen in Washington state, with most SARS-CoV-2 sequences from New York (21) and Connecticut (22) clustering with European lineages, which indicates repeated introductions from Europe. We also observed a second cluster of Washington state viruses related to a later introduction from Europe.

Our results highlight the critical need for widespread surveillance for community transmission of SARS-CoV-2 throughout the United States and the rest of the world, even after the current pandemic is brought under control. The broad spectrum of disease severity (23) makes such surveillance challenging (24).

The combination of traditional public health surveillance and genomic epidemiology can provide actionable insights, as happened in this instance: Upon sequencing the initial community case on 29 February 2020, results were immediately shared with national, state, and local public health agencies, which resulted in the rapid rollout of social distancing policies as Seattle and Washington state came to grips with the extent of existing COVID-19 spread. The confirmation of local transmission in Seattle prompted a change in testing criteria to emphasize individuals with no travel history. From 29 February onward, genomic data were immediately posted to the GISAID EpiCoV sequence database (9, 10) and analyzed alongside other public SARS-CoV-2 genomes by means of the Nextstrain online platform (25) to provide immediate and public situational awareness. We see the combination of community surveillance, genomic analysis, and public real-time sharing of results as a pathway to empower infectious disease surveillance systems.

REFERENCES AND NOTES

1. A. E. Gorbalenya, S. C. Baker, R. S. Baric, R. J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman, D. Penzar, S. Perlman, L. L. M. Poon, D. Samborskiy, I. A. Sidorov, I. Sola, J. Ziebuhr, Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. bioRxiv 2020.02.07.937862 [Preprint]. 11 February 2020. https://doi.org/10.1101/2020.02.07.937862.
2. S. Jiang et al., Lancet 395, 949 (2020).
3. A. Rambaut, Phylogenetic analysis of nCoV-2019 genomes (Virological, 2020); https://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356.
4. World Health Organization (WHO), "Coronavirus disease 2019 (COVID-19): Situation report – 55" (WHO, 2020); www.who.int/docs/default-source/coronaviruse/situation-reports/20200315-sitrep-55-covid-19.pdf?sfvrsn=33daa5cb_8.
5. F. Schlosser, B. F. Maier, O. Baranov, D. Brockmann, C. Jongen, A. Zachariae, A. Rose, Coronavirus COVID-19 Global Risk Assessment, Event Horizon - COVID-19 (2020); http://rocs.hu-berlin.de/corona/.
6. World Health Organization (WHO), "Coronavirus disease 2019 (COVID-19): Situation report – 38" (WHO, 2020); www.who.int/docs/default-source/coronaviruse/situation-reports/20200227-sitrep-38-covid-19.pdf?sfvrsn=2db7a09b_4.
7. K. Bray, "Coronavirus update: Addressing questions about a presumptive positive case in an adolescent" (Snohomish Health District Public Health Essentials, 2020); www.snohd.org/Blog.aspx?IID=13.
8. Washington State Department of Health, "2019 Novel Coronavirus Outbreak (COVID-19)" (2020); www.doh.wa.gov/Emergencies/Coronavirus.
9. Y. Shu, J. McCauley, Euro Surveill. 22, 30494 (2017).
10. S. Elbe, G. Buckland-Merrett, Glob Chall 1, 33–46 (2017).
11. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, Nat. Med. 26, 450–452 (2020).
12. N. Imai, I. Dorigatti, A. Cori, C. Donnelly, S. Riley, N. M. Ferguson, "Report 2: Estimating the potential total number of novel Coronavirus cases in Wuhan City, China" (Imperial College London, 2020); http://hdl.handle.net/10044/1/77150.
13. A. Rambaut et al., Nat. Microbiol. 10.1038/s41564-020-0770-5 (2020).
14. E. M. Volz, K. Koelle, T. Bedford, PLOS Comput. Biol. 9, e1002947 (2013).
15. M. L. Holshue et al., N. Engl. J. Med. 382, 929–936 (2020).
16. Q. Li et al., N. Engl. J. Med. 382, 1199–1207 (2020).
17. H. Nishiura, N. M. Linton, A. R. Akhmetzhanov, Int. J. Infect. Dis. 93, 284–286 (2020).
18. A. C. Miller, I. Singh, E. Koehler, P. M. Polgreen, Clin. Infect. Dis. 67, 388–397 (2018).
19. U.S. Centers for Disease Control and Prevention (CDC), "Update and Interim Guidance on Outbreak of 2019 Novel Coronavirus (2019-nCoV)" (CDC Health Alert Network, 2020); https://emergency.cdc.gov/han/han00427.asp.
20. U.S. Centers for Disease Control and Prevention (CDC), "Criteria to Guide Evaluation of Persons Under Investigation (PUI) for 2019-nCoV" (CDC, 2020); https://web.archive.org/web/20200222215422/https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-criteria.html.
21. A. S. Gonzalez-Reiche et al., Science 369, 297–301 (2020).
22. J. R. Fauver et al., Cell 181, 990–996.e5 (2020).
23. W.-J. Guan et al., N. Engl. J. Med. 382, 1708–1720 (2020).
24. R. Li et al., Science 368, 489–493 (2020).
25. J. Hadfield et al., Bioinformatics 34, 4121–4123 (2018).
26. T. Bedford, blab/ncov-cryptic-transmission: Release 2020-05-31, version 2020-05-31, Zenodo (2020); https://doi.org/10.5281/zenodo.3871089.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6516/571/suppl/DC1
Materials and Methods
Fig. S1
Seattle Flu Study Investigators List
References (27–40)
MDAR Reproducibility Checklist
Data S1