Before the coronavirus disease 2019 (COVID-19) outbreak on July 16, 2020, no new COVID-19 cases were reported in Urumqi City of Xinjiang for 150 consecutive days. During this outbreak, 826 confirmed COVID-19 cases have been reported in Urumqi. The phylogenetic characteristics of COVID-19 virus from clinical specimens of the initial 4 cases in the Urumqi outbreak were sequenced.

The full-length genome sequencing of COVID-19 virus was performed on the Illumina MiSeq platform, and 4 full-length COVID-19 virus genome sequences were obtained from clinical specimens of the initial 4 cases in this outbreak. Compared with the reference sequence EPI_ISL_402119 (1), which was isolated from Wuhan City in Hubei Province on January 7, 2020. A total of 13 nucleotide variations were determined, among which 7 nucleotide variations (C241T, C3037T, C14408T, A23403G, G28881A, G28882A, and G28883C) were identified in each full-length genome of the 4 Xinjiang strains, which were consistent with the characteristics of the L-lineage European branch 1/B.1.1 according to the latest classification principle (2) (Figure 1). The Xinjiang strains were different from the previous hCoV-19/Wuhan/IVDC-HB-01/2019 hCoV-19/Wuhan/IVDC-HB-03/2019 hCoV-19/Wuhan/IVDC-HB-05/2019 (belonging to S-lineage/A) strains in Wuhan in December 2019, indicating that they were not continuous transmission of indigenous COVID-19 virus strains. Other nucleotide substitutions were also identified in these Xinjiang strains, including C12789T and G14118T substitutions in all 4 genomes of the viruses, C2197T in specimen 1, C12809T in specimen 2, and C19718T and C23481T in specimen 3.

Since no Xinjiang strains contained the characteristic substitutions of Beijing Xinfadi strains (C6024T) (3) and Dalian strains (C2091T, A5128G, A8360G, C13860T, T19839C, G19999T, and C28905T) (4) and because the Beijing Xinfadi strains and Dalian strains also did not contain the characteristic substitutions of Xinjiang strains (C12789T and G14118T), the correlation between Xinjiang strains, Beijing Xinfadi strains, and Dalian strains were preliminarily excluded. The Xinjiang strains had high nucleotide similarity with several genomes of the COVID-19 virus (Shulan strains, Harbin strains) from the recently imported COVID-19 cases in northeastern China from April to May that also belonged to L-lineage European branch 1 (5–6). However, because of the interval (more than 2 months) and epidemiological evidence (no epidemiological link of the cases and contacts from Shulan and Harbin with the Xinjiang outbreak), the possibility of Harbin strains and Shulan strains spreading in Xinjiang was also ruled out. In addition, the full-length genome sequence analysis of the Xinjiang strains further confirmed that the virus source of the outbreak was not transmitted from natural animal hosts or intermediate animal hosts to human beings.

Although the source of COVID-19 virus of the Urumqi outbreak has not been determined, the possibility of transmission of the COVID-19 virus through importing contaminated products could not be excluded according to the characteristics of the recent COVID-19 outbreaks. Therefore, the COVID-19 virus detection in imported products should be further strengthened. Meanwhile, entry control at ports in Xinjiang and among key populations from Central Asia should be strengthened. A comprehensive analysis of epidemiology, serology, and etiology is needed to determine the source of the virus.

**Fundings:** This work was supported by the National Key Research and Development Program of China (Program No. 2018YFC1200305), National Science and Technology Major Project of China (Project No. 2017ZX10104001, 2018ZX10102001, 2018ZX10711 001, 2018ZX10713002).

Shulan Strains
Heilongjiang Strains

FIGURE 1. Phylogenetic tree based on the full-length genome sequences of the COVID-19 virus. The genomes of COVID-19 virus from Xinjiang were highlighted in blue. The genomes of reference COVID-19 virus from Wuhan in December 2019 were highlighted in grey. The recent reemergence of COVID-19 virus in Beijing Xinfadi were highlighted in green, and the recent reemergence of COVID-19 virus in northeastern China (Shulan and Heilongjiang) that related to imported cases were highlighted in pink and brown, respectively. S(A)- or L(B)-lineage of the COVID-19 virus were marked and colored on the right.
Corresponding authors: Xucheng Fan, 27980310@qq.com; Yong Zhang, zhangyong8@ivdc.chinacdc.cn.

1 National Health Commission Key Laboratory for Medical Virology and Viral Diseases, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China; 2 Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, China; 3 Xinjiang Uygur Autonomous Region Center for Disease Control and Prevention, Xinjiang, China; 4 Urumqi Center For Disease Control And Prevention, Xinjiang, China.

Joint first authors.

Submitted: September 15, 2020; Accepted: September 22, 2020

REFERENCES


