



LETTER

SARS-CoV-2 Genomic Sequencing Revealed N501Y and L452R Mutants of S/A Lineage in Tianjin Municipality, China

Xiaoyan Li¹ · Xin Gao¹ · Ming Zou¹ · Zhichao Zhuang¹ · Zhaolin Tan¹ · Baolu Zheng¹ · Aiping Yu¹ · Xu Su¹

Received: 31 March 2021 / Accepted: 26 May 2021 / Published online: 11 August 2021
© Wuhan Institute of Virology, CAS 2021

Dear Editor,

Coronavirus Disease 2019 (COVID-19) outbreak caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) presents a global pandemic which has resulted in more than 4 million people death in the world. In China, the initial transmission of COVID-19 has already been blocked by strict control strategies and effective treatment for patients. All local COVID-19 outbreaks after April 2020 were related to overseas importing. Up to February 2021, 207 imported COVID-19 cases have been reported in Tianjin, China (Tianjin Health Commission, 2021).

SARS-CoV-2, a kind of positive-sense single stranded RNA virus and belongs to the *Betacoronavirus* genus of the *Coronaviridae* family. Fourteen open reading frames (ORFs) constitute the majority of the ~ 2.9-kb SARS-CoV-2 genome, which encodes four structural proteins including nucleocapsid protein (N), membrane protein (M), spike protein (S) and RNA independent RNA polymerase (RdRp) and other non-structural proteins (Lu *et al.* 2020). The intrinsically high error rates of RdRp result in the stochastic introduction of mutations during viral genome replication (Liu *et al.* 2021). SARS-CoV-2 has evolved into two main lineages according to genome-based typing method established by China CDC (China CDC Lineage) and PANGO Lineages method (PANGO Lineage) (Rambaut *et al.* 2020; Wu *et al.* 2020; Yang and Xu 2021). The S/A-lineage defined by two nucleotides (8782T,

28144C) mainly circulates in Asia and the L/B-lineage defined by 8782C and 28144T mainly circulates in Europe and America. B.1.1.7-lineage/501Y.V1 mutation exhibited a rapid increase in its range and incidence, and other variants emerged in succession namely B.1.351-lineage/501Y.V2 from South Africa, P.1-lineage/501Y.V3 from Brazil, and B.1.207-lineage from Nigeria (Hodcroft *et al.* 2021; Oude Munnink *et al.* 2021). In this study, high through-put whole genome sequencing (WGS) was used to sequence all the imported SARS-CoV-2 samples for showing the genome characteristics and helping to design control strategies.

The study involved 94 throat swabs and one swab collected on the outer packaging of cold-chain food products (sample ID: TJ_20TF_FH158) from March 2020 to February 2021 (Supplementary Table S1). The nucleic acid was extracted using viral RNA extraction kits and instruments (Xi'an Tianlong Science and Technology Co., China). WGS was carried out using Nextera XT DNA Library Preparation Kit and MiniSeq System (Illumina, USA). Fifty-six SARS-CoV-2 genomes with the coverage higher than 98% and the sequencing depth higher than 100× were obtained by *de novo* assembly.

For nucleotide variation analysis, 309 single nucleotide polymorphisms (SNPs) and 14 deletion/insertion sites were found in all 56 SARS-CoV-2 genomes by using Bowtie2 software (Supplementary Figure S1). Less than ten SNPs were identified from each of genomes collected from March 2020 to June, and more than 17 SNPs were found from genomes collected from October 2020 to February 2021 (Supplementary Table S1). According to China CDC Lineages and PANGO Lineages, T8782C and C28144T were two specific SNPs of S/A-lineage. In this study, four sequences fell into the S/A-lineage, and 34 SNPs and 3 deletion/insertion sites were found in these strains. The other 52 sequences all fell into the L/B-lineage. Four SNPs (C241T, C3037T, C14408T and A23403G) were detected as the characteristics of L-lineage European branch (China

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12250-021-00432-5>.

✉ Xu Su
13323360318@163.com

✉ Xiaoyan Li
xiaoyanli1291@163.com

¹ Department of Pathogenic Microbiology, Tianjin Centers for Disease Control and Prevention, Tianjin 300011, China

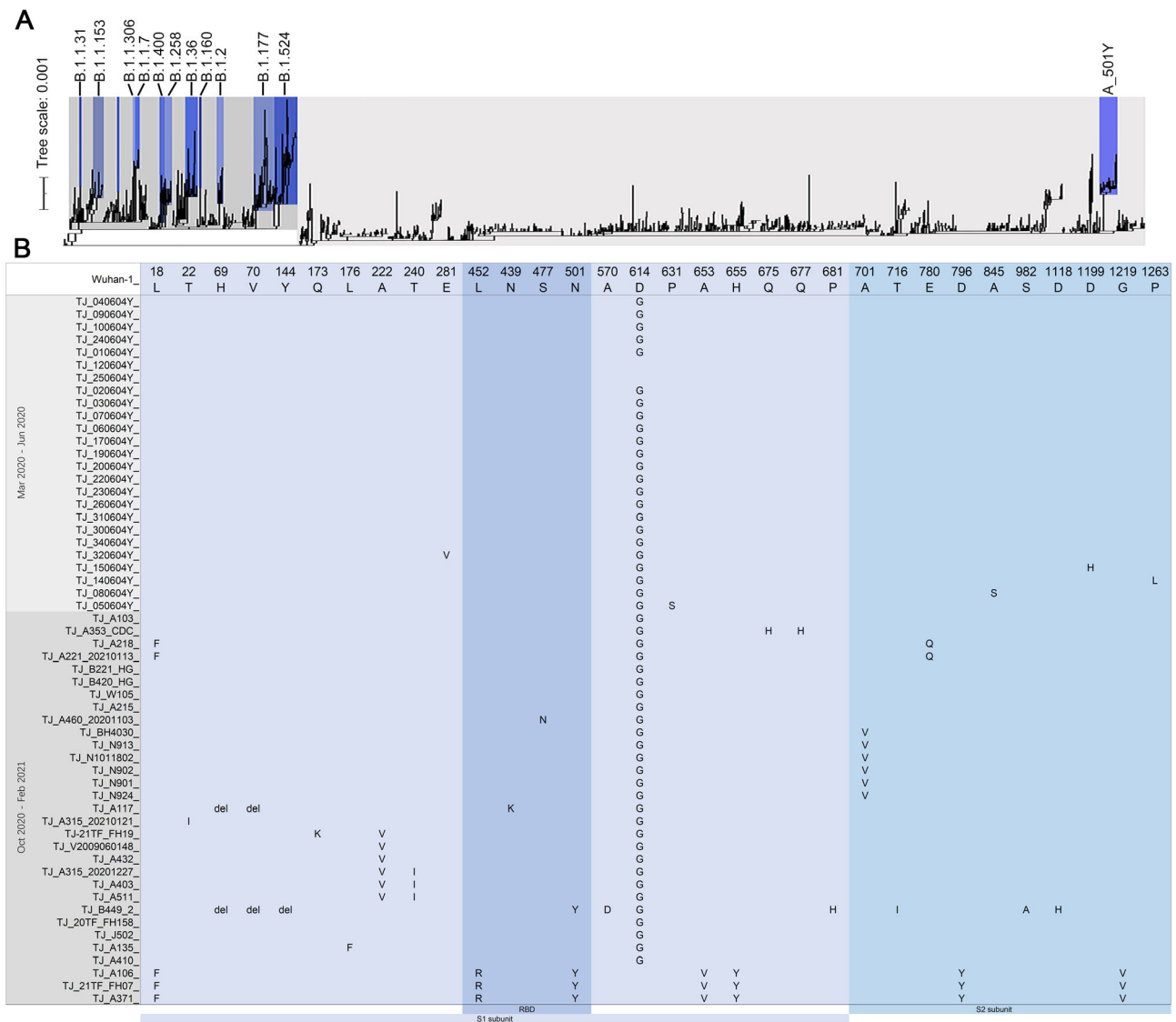


Fig. 1 **A** Phylogenetic analysis of the SARS-CoV-2 isolated in Tianjin, China. Maximum likelihood phylogeny inferred using 56 genome sequences of SARS-CoV-2 generated in this study and 1494 sequences already deposited on the GISAID database. The tree is

rooted between A-lineage and B-lineage. **B** Variant sites of S protein amino acid identified in 56 SARS-CoV-2 genomes obtained in this study.

CDC lineage method)/B.1-lineage (PANGO Lineage) in 50 genomes. Nine sequences had three SNPs (G28881A, G28882A and G28883C) which were detected as the characteristics of L-lineage European branch I/B.1.1-lineage. The other nine sequences had two SNPs (G25563T and C1059T) which were characteristics of L-lineage European branch II.1. Four sequences had the three SNPs (G25563T and C2416T) which were detected as the characteristics of L-lineage European branch II.2. Five sequences had the two SNPs (G25563T and C18877T) which were detected as the characteristics of L-lineage European branch II.3.

For phylogenetic analysis, 1494 high-quality and high-coverage SARS-CoV-2 genomic sequences were downloaded from Global Initiative on Sharing Avian Influenza Data (GISAID) and aligned by MAFFT v7.42 and IQ-tree v2.1.2 (Shu and McCauley 2017). 895 sequences of S/A-lineage included 18 strains which all had the N501Y and L452R mutation of S protein and can constitute an individual sub-lineage, named as the A_501Y lineage including TJ_A371, TJ_21TF_FH07 and TJ_A106 (Fig. 1A). Two genomes (EPI_ISL_801441, EPI_ISL_801442) were closely related to TJ_A371, TJ_21TF_FH07 and TJ_A106. The original region of these strains included Belgium, Turkey, Mayotte, Spain, *et al.*, and the date of specimen

collection ranged from December 2020 to January 2021. The other 599 sequences of L/B-lineage included the B.1.1.7/501Y.V1 variant (sample ID: TJ_B449_2) which was the first detection of a major international variant in Tianjin Municipality in January 2021 and was closely related to two genomes, EPI_ISL_1060597, EPI_ISL_1018092, collected from France and Ghana, respectively. Four genomes (EPI_ISL_806290, EPI_ISL_428910, EPI_ISL_1383183, EPI_ISL_1111200) were closely related to TJ_20TF_FH158 according to phylogenetic analysis, all fell into the B.1.1.1-lineage and shared four specific SNPs, C4002T, G10097A, C13536T and C23731T, and other five unique SNPs (C1612T, G3606T, C7772T, C25665T and C26600T) were found in genome of TJ_20TF_FH158.

S protein comprises S1 subunit locating at the N terminal and S2 subunit locating at the C terminal. The receptor-binding domain (RBD, S protein aa319-541) of S1 subunit binds to the human angiotensin I converting enzyme 2 (ACE2) receptor which mainly depends on the receptor binding motif (RBM, S protein aa438-506) region, and S2 subunit helps virus fusion into cells (Rathnasinghe *et al.* 2021). RBD is also a target of the SARS-CoV-2 neutralizing antibody, and mutations in this region may affect the neutralizing titer of antibody (Massacci *et al.* 2020). 29 amino acid (aa) variants and three aa deletion sites of spike protein were found in all 56 genomes, and 22 variants of them located in the S1 subunit including four variants in the RBD region of S1 subunit (Fig. 1B). Seven aa variants including L18F, L452R, N501Y, A653V, H655Y, D796Y and G1219V were identified in TJ_A371, TJ_21TF_FH07 and TJ_A106 of A_501Y lineage (Fig. 1B). For B.1.1.7/501Y.V1 variant, 28 nucleotide variants as the characteristics were identified in the genome of TJ_B449_2 including ten aa variants/deletion of S gene (Fig. 1B), and other six unique SNPs including C2110T, G2914T, T7984C, G10887A, C14120T, C19390T and nucleotide deletion of 27792–27794 and 28271 were found in this genome. Nucleotide deletions of 26160–26167, 27386 and 28248–28253 were found in genomes of TJ_21TF_FH07 and TJ_A371 (Supplementary Table S2).

Initial low percentage of sequences in S/A-lineage exhibited a slight increase from December 2020 to the beginning of 2021 and were discovered in many countries worldwide with several aa mutations of S protein. In this study, N501Y and L452R mutants locating in RBD region of S protein were found in three sequences belonging to A_501Y lineage. L452R may decrease the sensitivity of the virus to neutralizing antibody. N501Y mutation may affect the immunogenicity of the virus (Liu *et al.* 2021; Gu *et al.* 2020). N501Y variant of A lineage have been identified in Turkey, the United Kingdom, France, Denmark, Niger,

et al (Li *et al.* 2020). The pathogenicity and transmissibility of A_501Y mutation were still unknown.

Two waves of COVID-19 outbreaks emerged in Tianjin Municipality. At the beginning of 2020, the outbreak was mainly resulted by small-scale local area transmission. In November 2020, the outbreak was associated with imported cold chain products emerged in local communities. Numbers of accumulated COVID-19 positive travelers from overseas in Tianjin continually rise, and the risk of imported COVID-19 outbreaks remains. Sustained SARS-CoV-2 genome sequencing and monitoring of variants can be used to enrich genome data of SARS-CoV-2 which is of great significance for implementing prevention and control strategy of COVID-19 outbreak and tracing the infection sources (Chen *et al.* 2021; Eden *et al.* 2020).

Acknowledgements This work was supported by Tianjin Municipal Commission of Health COVID-19 Prevention and Control Technology General Project (No. 2020xkm01 and No. 2020xkz01), and Science and Technology Program of Tianjin, China (No. 20JCZDJC00130).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Animal and Human Rights Statement Additional informed consent was obtained from all patients for whom identifying information is included in this article.

References

- Chen AT, Altschuler K, Zhan SH, Chan YA, Deverman BE (2021) COVID-19 CG: tracking SARS-CoV-2 mutations by locations and dates of interest. *Elife* 10:e63409
- Eden JS, Rockett R, Carter I, Rahman H, de Ligt J, Hadfield J, Storey M, Ren X, Tulloch R, Basile K, Wells J, Byun R, Gilroy N, O'Sullivan MV, Sintchenko V, Chen SC, Maddocks S, Sorrell TC, Holmes EC, Dwyer DE, Kok J (2020) An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. *Virus Evol* 6:veaa027
- Gu H, Chen Q, Yang G, He L, Fan H, Deng YQ, Wang Y, Teng Y, Zhao Z, Cui Y, Li Y, Li XF, Li J, Zhang NN, Yang X, Chen S, Guo Y, Zhao G, Wang X, Luo DY, Wang H, Yang X, Li Y, Han G, He Y, Zhou X, Geng S, Sheng X, Jiang S, Sun S, Qin CF, Zhou Y (2020) Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* 369:1603–1607
- Hodcroft EB, Domman DB, Snyder DJ, Oguntuyo K, Van Diest M, Densmore KH, Schwalm KC, Femling J, Carroll JL, Scott RS, Whyte MM, Edwards MD, Hull NC, Kevill CG, Vanchiere JA, Lee B, Dinwiddie DL, Cooper VS, Kamil JP (2021) Emergence in late 2020 of multiple lineages of SARS-CoV-2 Spike protein variants affecting amino acid position 677. medRxiv. <https://doi.org/10.1101/2021.02.12.21251658>
- Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H, Wang M, Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y (2020) The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 182:1284–1294

- Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS, Whelan SPJ (2021) Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* 29:477–488
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395:565–574
- Massacci A, Sperandio E, D’Ambrosio L, Maffei M, Palombo F, Aurisicchio L, Ciliberto G, Pallocca M (2020) Design of a companion bioinformatic tool to detect the emergence and geographical distribution of SARS-CoV-2 Spike protein genetic variants. *J Transl Med* 18:494
- Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, van der Spek A, Tolsma P, Rietveld A, Brouwer M, Bouwmeester-Vincken N, Harders F, Hakze-van der Honing R, Wegdam-Blans MCA, Bouwstra RJ, GeurtsvanKessel C, van der Eijk AA, Velkers FC, Smit LAM, Stegeman A, van der Poel WHM, Koopmans MPG (2021) Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* 371:172–177
- Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 5:1403–1407
- Rathnasinghe R, Jangra S, Cupic A, Martínez-Romero C, Mulder LCF, Kehrer T, Yildiz S, Choi A, Mena I, De Vrieze J, Aslam S, Stadlbauer D, Meekins DA, McDowell CD, Balaraman V, Richt JA, De Geest BG, Miorin L, Krammer F, Simon V, García-Sastre A, Schotsaert M (2021) The N501Y mutation in SARS-CoV-2 spike leads to morbidity in obese and aged mice and is neutralized by convalescent and post-vaccination human sera. medRxiv. <https://doi.org/10.1101/2021.01.19.21249592>
- Shu Y, McCauley J (2017) GISAID: global initiative on sharing all influenza data—from vision to reality. *Eur Surveill* 22:30494
- Tianjin Health Commission (2021) Covid-19 epidemic situation in Tianjin, 26th February 2021. http://wsjk.tj.gov.cn/XWZX6600/YQFKDT1752/202102/t20210226_5367775.html
- Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T (2020) Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 27:325–328
- Yang S, Xu W (2021) Genomic analysis platforms and typing methods for SARS-CoV-2 genome sequences. *Chin J Virol* 37:181–190 (in Chinese)