



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Original Article

**Genomic Epidemiology of Imported Cases of COVID-19 in Guangdong Province, China, October 2020 – May 2021***

LIANG Dan^{1,2,3,&}, WANG Tao^{4,&}, LI Jiao Jiao^{5,&}, GUAN Da Wei³, ZHANG Guan Ting⁶, LIANG Yu Feng⁷,
LI An An⁷, HONG Wen Shan⁴, WANG Li⁶, CHEN Meng Lin⁸, DENG Xiao Ling³, CHEN Feng Juan⁵,
PAN Xing Fei⁹, JIA Hong Ling^{10,#}, LEI Chun Liang^{5,#}, and KE Chang Wen^{2,3,4,6,7,#}

1. Guangdong Provincial Key Laboratory of Virology, Institute of Medical Microbiology, Jinan University, Guangzhou 510632, Guangdong, China; 2. Guangzhou National Laboratory, Guangzhou 510700, Guangdong, China; 3. Guangdong Provincial Center for disease control and prevention, Guangdong Workstation for Emerging Infectious Disease Control and Prevention, Guangzhou 511430, Guangdong, China; 4. Shantou University Medical College, Shantou 515041, Guangdong, China; 5. Guangzhou Eighth People's Hospital, Guangzhou 510030, Guangdong, China; 6. School of public health, Sun Yat-sen University, Guangzhou 510080, Guangdong, China; 7. School of Public Health, Southern Medical University, Guangzhou 510515, Guangdong, China; 8. Guangzhou Mendel Genomics and Medical Technology Co., Ltd., Guangzhou 510535, Guangdong, China; 9. Department of Infectious Diseases, the Third Affiliated Hospital of Guangzhou Medical University, Guangzhou 510150, Guangdong, China; 10. Department of Medical Biochemistry and Molecular Biology, School of Medicine, Jinan University, Guangzhou 510632, Guangdong, China

Abstract

Objective The pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been engendering enormous hazards to the world. We obtained the complete genome sequences of SARS-CoV-2 from imported cases admitted to the Guangzhou Eighth People's Hospital, which was appointed by the Guangdong provincial government to treat coronavirus disease 2019 (COVID-19). The SARS-CoV-2 diversity was analyzed, and the mutation characteristics, time, and regional trend of variant emergence were evaluated.

Methods In total, 177 throat swab samples were obtained from COVID-19 patients (from October 2020 to May 2021). High-throughput sequencing technology was used to detect the viral sequences of patients infected with SARS-CoV-2. Phylogenetic and molecular evolutionary analyses were used to evaluate the mutation characteristics and the time and regional trends of variants.

Results We observed that the imported cases mainly occurred after January 2021, peaking in May 2021, with the highest proportion observed from cases originating from the United States. The main lineages were found in Europe, Africa, and North America, and B.1.1.7 and B.1.351 were the two major sublineages. Sublineage B.1.618 was the Asian lineage (Indian) found in this study, and B.1.1.228 was not included in the lineage list of the Pangolin web. A reasonably high homology was observed among

*This work was supported by the grant National Key Research and Development Program of China [2018YFC1603804 and 2021YFC0863300]; Emergency Key Program of Guangzhou Laboratory [EKPG21-27]; and Sino-German Center for Research Promotion (SGC)'s Rapid Response Funding Call for Bilateral Collaborative Proposals Between China and Germany in COVID-19 Related Research [C-0032].

&These authors contributed equally to this work.

#Correspondence should be addressed to JIA Hong Ling, E-mail: jiahongling@aliyun.com; Tel: 86-20-85228359; LEI Chun Liang, E-mail: gz8hlc1@126.com; Tel: 86-13802917518; KE Chang Wen, E-mail: kecw1965@aliyun.com; Tel: 86-20-31051236.

Biographical notes of the first authors: LIANG Dan, female, born in 1992, PhD Candidate, majoring in microbiology; WANG Tao, male, born in 1998, Master Candidate, majoring in laboratory diagnosis of infectious diseases; LI Jiao Jiao, female, born in 1991, MD, majoring in infectious disease.

all samples. The total frequency of mutations showed that the open reading frame 1a (ORF1a) protein had the highest mutation density at the nucleotide level, and the D614G mutation in the spike protein was the commonest at the amino acid level. Most importantly, we identified some amino acid mutations in positions S, ORF7b, and ORF9b, and they have neither been reported on the Global Initiative of Sharing All Influenza Data nor published in PubMed among all missense mutations.

Conclusion These results suggested the diversity of lineages and sublineages and the high homology at the amino acid level among imported cases infected with SARS-CoV-2 in Guangdong Province, China.

Key words: Severe acute respiratory syndrome coronavirus 2; Next-Generation sequencing; Phylogenetic tree; Variants

Biomed Environ Sci, 2022; 35(5): 393-401 doi: 10.3967/bes2022.055

ISSN: 0895-3988

www.besjournal.com (full text)

CN: 11-2816/Q

Copyright ©2022 by China CDC

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the pathogenic agent of coronavirus disease 2019 (COVID-19), has caused a global pandemic and conferred a profound effect on all aspects of society. As of June 15, 2021, 175 million infections have been confirmed, and 3.8 million deaths have been reported all over the world^[1]. The Chinese government managed to contain the virus through a series of draconian control measures. However, the number of newly confirmed cases outside of China has been rising rapidly. Thus, China prioritized protecting itself from imported cases from other countries or territories.

Guangdong Province is known as the “South Gate of China,” with a large number of foreign passengers; hence, the most number of imported cases was detected here compared with other regions of Mainland China^[2]. The World Health Organization announced several variants of concern (VOCs) and variants of interest (VOIs) of SARS-CoV-2 on May 31, 2021^[3]. With the increase in the number of genome sequences available in the public database, analysis using more sequences by sampling from a wider area over a longer period was imperative. In the present study, we employed second-generation sequencing to obtain virus sequence information and performed advanced comprehensive methods to analyze the genetic characteristics of SARS-CoV-2 in the imported cases based on the demographic and baseline characteristics of these patients. Our studies enriched the knowledge of the existing epidemic strains of SARS-CoV-2 in the world and provided valuable clues for defense against this infectious

disease imported to China.

MATERIALS AND METHODS

Ethics Approval Statement

The protocol of the study was reviewed and approved by Guangdong Provincial Center for Disease Control and Prevention (approval number: W96-027E-202121).

Sample Sources

This study included human throat swab samples that were collected and processed in the Guangdong Provincial Center for Disease Control and Prevention, China, and throat swab samples, which were de-identified prior to analysis, collected from the Guangzhou Eighth People’s Hospital from October 2020 to May 2021. All the samples were inactivated at 56 °C for 30 min. Demographic characteristics and clinical information were collected from electronic medical records.

Nucleic acid Test (NAT) and Next-generation Sequencing (NGS)

All the throat swab samples were screened by quantitative real-time polymerase chain reaction method using the SARS-CoV-2 NAT kits (BioGerm, Shanghai, China). The positive specimens were processed for NGS using the MGISEQ-2000 platform of the Beijing Genomics Institute (BGI, Shenzhen, China). DNA fragment assembly was performed using the iGenome® VirusDetector system (BGI, Shenzhen, China), with the Wuhan-hu-01 (GenBank access no. MN908947.3) genome as the reference sequence. All ambiguous nucleotides in the alignment analysis were set to N. The ratio of non-N

greater than 85% indicated a good fragment assembly in accordance with the manufacturer's instruction.

Phylogenetic Analysis

The complete reference genomes were retrieved from the National Center for Biotechnology Information database (<https://www.ncbi.nlm.nih.gov/>). Pangolin web tool (<https://cov-lineages.org/>) was used for the origin of clades. Mattf v7.037 (Kazutaka Katoh, Japan, 2002) and BioEdit v7.0.9.0 (Borland, USA, 2009) were applied to sequence alignment and trimming, respectively. The maximum-likelihood (ML) phylogenetic tree was constructed using IQ-Tree v1.6.12 (Bui Quang Minh, Nguyen Lam Tung, Olga Chernomor, et al., USA, 2011). A total of 1,000 bootstrap replicates were used to deduce branch support. The phylogenetic tree was visualized using FigTree v1.4.2 (Andrew Rambaut, UK, 2006) and Adobe Illustrator CS6 (Adobe, USA, 1987). In addition, the Nextclade web tool (<https://clades.nextstrain.org>) was used for putting up information about nucleotide mutations and amino acid mutations. R v3.6.1 (Ross Ihaka, Robert Gentleman, New Zealand, 1995) was used to calculate the genetic distance of sequences, obtain the homology, and cleanse the mutation data at the nucleotide and amino acid levels.

Statistical Analyses

The data were managed using EpiData v3.1 (Jens M. Lauritsen, Mark Myatt, Michael Bruus, DK, 1999). R v3.6.1 (Ross Ihaka, Robert Gentleman, New Zealand, 1995) was used for descriptive analysis and plotting. All the bar graphs were drawn using the GraphPad Prism software v8.0.2 (San Diego, CA, USA, 1992).

Age variables were expressed using "median {Interquartile range [(IQR): 25%–75%]," and categorical variables including age group, gender, and NAT results were presented as percentages. A two-tailed *P*-value of < 0.05 was considered statistically significant throughout the study.

RESULTS

Demographic Information of the Patients

A total of 177 patients were included in this study, and their demographic characteristics are shown in Table 1. More than half of the respondents were male (136, 76.84%). A total of 78 individuals (44.07%) were aged 31–40 years, and 6 people were

over 61. The minimum age was 16, and the maximum was 85 (median age, 34 years). A total of 123 out of 177 (69.49%) individuals showed positive results for SARS-CoV-2 RNA by the NATs, and among them, 117 cases (95.12%) were sequenced successfully. Exactly 43 patients in our studies (24.29%) were asymptomatic cases (Table 1).

Sequence Distribution in the Imported Cases

A total of 117 complete genomes sequences were obtained from this study, and the time and regional distribution through sequence clades are shown in Figure 1. With the sampling dated between June 2020 and May 2021, the time distributions of 177 patients were mainly in May 2021 (65/177, 36.72%), February 2021 (43/177, 24.29%), January 2021 (33/177, 18.64%), and March 2021 (24/177, 13.56%) (Figure 1A), and the time distributions of sequences were primarily in May 2021 (40/117, 34.19%), January 2020 (25/117, 21.37%), February 2020 (24/117, 20.51%), and March 2021 (18/117, 15.38%). The sequence distribution relevant to the timelines was

Table 1. Demographic information of patients with SARS-CoV-2 infection

Variables	Patients (N = 177)
Age, years	
Median (P25, P75)	34 (31, 45)
Age group, n (%)	
≤ 30	39 (22.03)
31–40	78 (44.07)
41–50	34 (19.21)
51–60	20 (11.30)
≥ 61	6 (3.39)
Gender, n (%)	
Female	41 (23.16)
Male	136 (76.84)
NATs results, n (%)	
Positive	123 (69.49)
Sequenced successfully	117 (95.12)
Failed to sequence	6 (4.87)
Negative	54 (30.51)
Clinical classification, n (%)	
Asymptomatic case	43 (24.29)
Confirmed case	134 (75.71)

Note. NATs, nucleic acid detection

asymmetrical due to the successful epidemic control in China after April 2020^[4]. The sequences from this study were classified into 39 sublineages: A, A.23.1, B.1, B.1.1, B.1.1.157, B.1.1.214, B.1.1.222, B.1.1.228, B.1.1.368, B.1.1.420, B.1.1.7, B.1.170, B.1.177.81, B.1.2, B.1.214.2, B.1.351, B.1.36, B.1.36.16, B.1.36.17, B.1.36.38, B.1.36.22, B.1.402, B.1.413, B.1.427, B.1.429, B.1.438.1, B.1.466.2, B.1.468, B.1.524, B.1.525, B.1.533, B.1.580, B.1.595, B.1.596, B.1.618, B.1.84, C.36, L.3, and P.2, where the clades located were chiefly concentrated on B.1.1.7 (28/117, 23.93%) and B.1.351 (20/117, 17.09%). The lineage distribution included Europe ($n = 50$), Africa ($n = 28$), North America ($n = 22$), Asia (excluding China) ($n = 12$), China ($n = 3$), South Africa ($n = 1$), South America ($n = 1$), and B.1.1.228 ($n = 1$) (Figure 1B). The largest diversity of lineages was observed in January 2021, with six lineages, followed by February 2021, with five lineages. European and African lineages mainly appeared in May 2021 (25/117, 21.37% vs. 10/117, 8.55%), the North American lineage emerged in January 2021 (11/117, 9.40%), and that from South America and variant B.1.1.228 materialized in January 2021 (not classified on the Pangolin web tool). China lineages were interspersed in October 2020 and January and February 2021 (Figure 1C).

The regional distribution was chiefly imported from the United States (USA; 13/117, 11.11%), Zambia (8/117, 6.84%), Nigeria (8/117, 6.84%), and the United Arab Emirates (UAE; 6/117, 5.13%). Additionally, the regional distribution included Indonesia ($n = 5$), Cambodia ($n = 5$), Tanzania ($n = 4$), Bangladesh ($n = 4$). Notably, the UK strains of the European lineage (B.1.1.7 and B.1.36.2) mostly came from Cambodia and Nigeria. The USA strains of the North American lineage (B.1.1.222, B.1.1.368, B.1.2, B.1.402, B.1.427, B.1.429, B.1.580, B.1.595, and B.1.596) were predominantly from the USA. The two patients infected with sublineage B.1.618 (Indian lineage) were asymptomatic and imported from UAE and Mozambique. Sublineage B.1.170, belonging to the European lineage (Egypt and UAE), first appeared in the world on March 19, 2020, but occurred in June 2020 in our study. Sublineage B.1.468 (Asian lineage (Indonesia), a member of the early epidemiological transmission chains, first emerged on February 13, 2020, and was included in our study in July 2020.

Phylogenetic Tree and Variation Analyses of SARS-CoV-2 at the Nucleotide and Amino Acids Levels

For this study, a GTR + F + I + G4 nucleotide

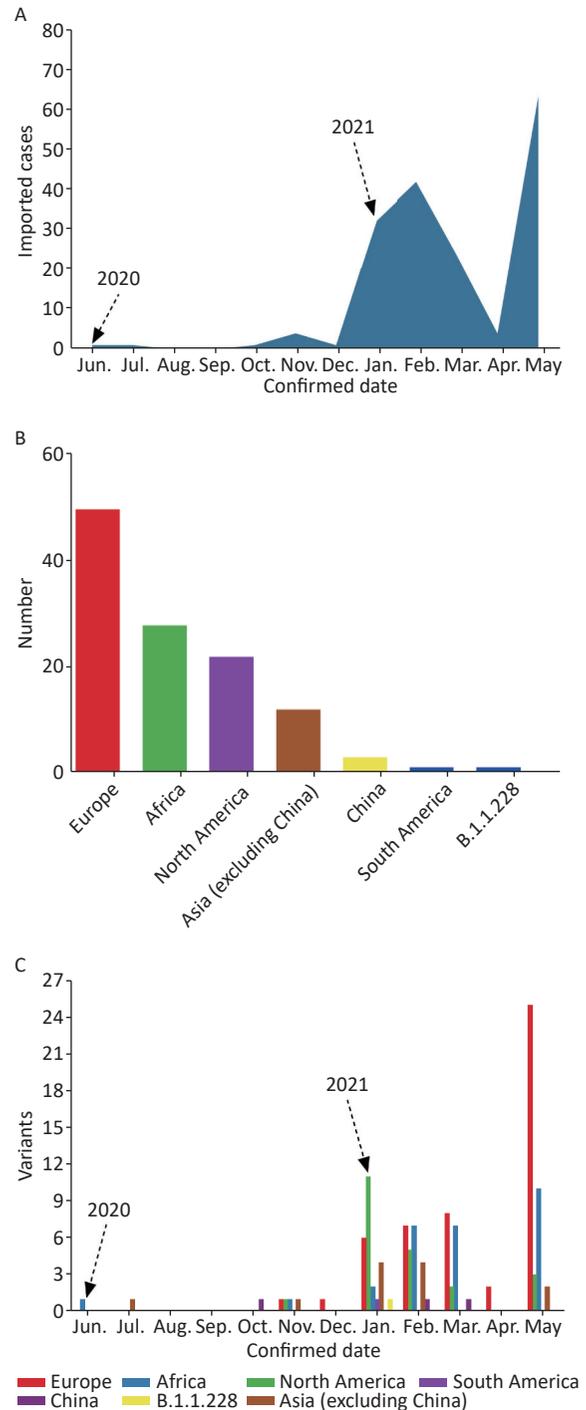


Figure 1. Distribution of imported cases. (A) Time distribution of 177 imported cases during the whole study period (from June 2020 to May 2021); (B) total number of lineages (including those from Europe, Africa, North America, Asia (excluding China), China, and South America and sublineage B.1.1.228) in 117 imported cases; (C) temporal distribution of each lineage in 117 imported cases.

substitution model, the best-fit model based on the Akaike Information Criterion according to the Bayesian information criterion, was used for the phylogenetic analyses. The virus species in this study clustered in two large groups. Most viral sequences from patients with SARS-CoV-2 infection clustered in the African, North American, and European lineages. The rest of the sublineages were scattered in diverse branches of the phylogenetic tree. The diversity of molecular evolution of SARS-CoV-2 cases imported to Guangdong Province was revealed, which was consistent with the case distribution (left part of Figure 3). In the homology analyses, a comparison of the reference sequences with the sequences of respective samples indicated that the homology among genomic sequences was 99.68%–100%. In addition, the viral genomes from infected patients were consistent with the animal reference sequences, including the *Canis lupus* family, feline, mink, ferret, *Panthera leo*, tiger, and *Neovison vison* (99.76%–99.98%). Moreover, we analyzed the correlations between age, gender, clinical classification, and VOCs, with a statistically significant difference between males and females ($P < 0.05$, not shown in the table).

The mutations in nucleotide that occurred in 117 sequences were found in these loci: S (nt23403), 1a (nt241, 3037), 1b (nt14408), 3a (nt25563), N (nt28881–28883), 8 (nt27972), E (nt26456), and M (nt26735). Position 1a (nt241 and nt3037) exhibited the highest mutation rate, with T substituting for C, followed by nt14408 (position 1b) with C displacing T. At the amino acid level, mutations were discovered in the following sites: S (aa614), 1b (aa314), 1a (aa3676–aa3677), N (aa203–aa205), 3a (aa57), and 8 (aa52). D614G had the highest mutation rate, followed by P314L (Figure 2 and Table 2). Supplementary Table S1 (available in www.

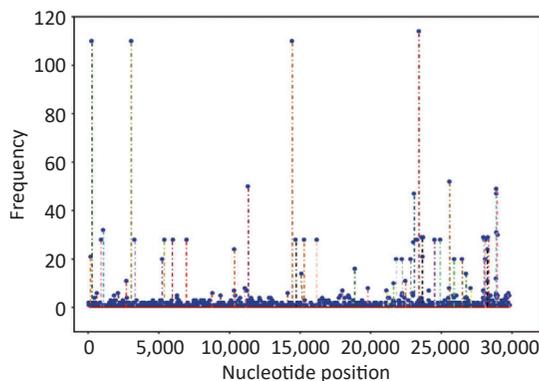


Figure 2. Frequencies of all mutations on the genome at the nucleotide level.

besjournal.com) lists all substitutions in the nucleotides and amino acids.

Supplementary Tables S2 and S3 (available in www.besjournal.com) show the nonsense and missense mutation distributions in the non-synonymous mutation. Additionally, Supplementary Tables S4 and S5 (available in www.besjournal.com) list the reported amino acid mutation loci on the Global Initiative of Sharing All Influenza Data (GISAID) website (<https://www.gisaid.org/>) and publications in PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), respectively. In total, 1,076 nonsense mutations were observed on 476 nucleotide sites in all the analyzed sequences. A total of 2,010 missense mutations were present in 441 loci, and they were largely disseminated in positions ORF 1a and S. The missense mutations were demonstrated in all 12 gene fragments (right part of Figure 3). We also identified some missense mutations that had neither been reported on the GISAID database nor published in PubMed, including position non-S (replacements at five amino acid residue lineages, including those from the North American (sublineage B.1.2): Ile5Val of ORF9b and European lineages (sublineage B.1.214.2: Ala75Val of ORF9b; sublineage B.1.1.7: Met1- and Leu6Met of ORF7b; sublineage B.1.1.7: Met1- and Leu6Met of ORF7b and Arg25Ile of ORF9b), and position S (eight substitution sites, including those in lineages from North America (sublineage B.1.595): Leu226Phe; Europe (sublineage B.1): Tyr248Asn; North America (sublineage B.1.402): Cys1247Arg; Asia (not China) (B.1.618): Tyr145- and His146-; Asia (not China) (B.1.1.25): His66Asp; Africa (sublineage L.3): Gly75Arg, Tyr1155Phe; Africa (sublineage L.3): Tyr1155Phe, whose corresponding coding nucleotide sites were demonstrated as the points marked in red in the right panel of Figure 4.

DISCUSSION

We narrated the complete sequences of 117 patients with SARS-CoV-2 infection from the Guangzhou Eighth People's Hospital, Guangdong, China, during the whole study period (June 2020 to May 2021). Based on our basic description and phylogenetic analysis of imported cases, the median age of patients in this study was younger than that of patients in earlier studies^[5-7], and we noticed a close identity between the genome of the respondents and animal reference sequences [including *Canis lupus* family, feline, mink, ferret, *Panthera leo*, tiger, and *Neovison vison*

(99.76%–99.98%]). Similar findings were described previously in other studies (homology between rhinolophine, bat, pangolin, and human)^[8-11]. New variants with several critical mutations, such as

B.1.1.7, B.1.351, B.1.617, and P.1, to which the available vaccines or potential drugs may be ineffective, emerged in 2021^[12-14]. The B.1.617 variant rapidly spread to India and several countries

Table 2. The main location of nucleotide or amino acid variant in SARS-CoV-2 patients

Regions	Nucleotide mutations					Amino acid mutations				
	Site	No.	Type	Total types	Total frequency	Site	No.	Type	Total types	Total frequency
Gene E	26456	20	C→T	4	24	71	20	P→L	2	22
Gene M	26735	14	C→T	22	59	82	2	I→T	7	10
Gene N	28881	49	G→A			203	49	R→K		
	28882	47	G→A	91	462	204	46	G→R	45	278
	28883	47	G→C			205	31	T→I		
Gene ORF										
1a	241	110	C→T			3677	50	F→L		
	3037	110	C→T	345	1104	3676	50	G→S	184	540
1b	14408	110	C→T	148	465	314	108	P→L	74	226
3a	25563	52	G→T	46	153	57	52	Q→H	38	140
6	27213	2	C→T	6	9	2	2	F→M	2	4
7a	27577	2	C→T	10	11	51	2	D→A	7	8
7b	27881	2	C→T	6	9	6	2	L→M	5	8
8	27972	29	C→T	24	185	52	28	R→I	18	151
9b						32	3	R→P	9	12
Gene S	23403	114	A→G	123	617	614	114	D→G	77	544

Note. Abbreviations: E, envelope protein; M, membrane; N, nucleocapsid protein; ORF, open reading frame; S, spike protein; 1a, 1b, 3a, 6, 7a, 7b, 8, and 9b represent ORFs 1a, 1b, 3a, 6, 7a, 7b, 8, and 9b, respectively.

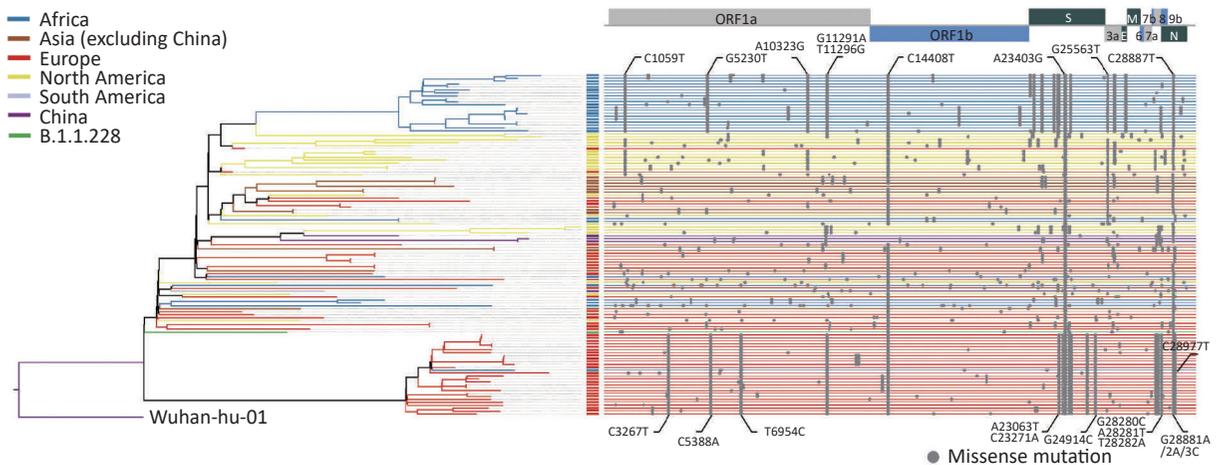


Figure 3. Maximum-likelihood phylogenetic dendrogram constructed based on Pangolin’s classifications from the 117 sequences and Wuhan-hu-01 in GenBank as a reference sequence. The positions of all missense mutation sites on genes and their distribution, regardless of whether they were published or unreported and unpublished and unreported, are revealed in the right part of the graph.

in the world. In this study, the Indian lineage we found was sublineage B.1.618 and not the B.1.617.2 variant that was first detected and caused a pandemic in India^[15]. This variant was first isolated and sequenced in Guangzhou in May 2021^[16]. Sublineages B.1.17 and B.1.351 were considered to be the main variants. They had been affiliated with the increased transmissibility of COVID-19 epidemiology and virulence, changes in clinical disease presentation, and reduced effectiveness of available vaccines and therapeutics^[17,18]. In addition, we discovered the B.1.427/B.1.429, P.2, and B.1.525 variants in our studied cases. They caused community transmission or multiple clusters of COVID-19, which should arouse our attention^[19-22]. Previous studies showed that D614G endowed a replication advantage to SARS-CoV-2 by increasing the feasibility of human-to-human transmission^[23-25]. Given the limitations of epidemiological and clinical data, with a conservative interpretation method, we can only discover the D614G in the spike protein that presented the highest mutation density, presenting another limitation of our research.

According to current studies, the ORF7b gene of SARS-CoV-2 with a procaspase-3 cleavage activity has a prerequisite role in viral replication^[17,26]. It potentially impedes certain cellular processes that dominate several common symptoms of SARS-CoV-2 infection, such as leucine zipper formation and epithelial cell-cell adhesion, which may contribute to heart-rate disorders^[26-28]. The ORF9b of SARS-CoV-2, encoding a 97-amino acid protein on the mitochondrial membrane, was located within the nucleocapsid gene (N gene)^[27]. Previous studies revealed that SARS-CoV-2 ORF9b, as an accessory

protein, inhibits innate immunity by targeting mitochondria. It can potentially suppress the antiviral response of infected cells and regulate the host immune response by suppressing type I interferon synthesis^[29-31]. Among the various missense mutations in the 117 sequences, we disclosed some missense mutations in several sublineages (such as B.1, B.1.1.25, B.1.1.7, B.1.2, B.1.214.2, B.1.402, B.1.595, B.1.618, and L3) of partial lineages, including those from North America, Europe, Asia (excluding China), and Africa. They had neither been reported in the GISAID database nor published on PubMed and included five amino acid substitutions of the non-S positions and eight substitution sites in the S position. Given our lack of access to the information on clinical symptoms and dissemination chain in patients infected with the corresponding sublineages, we cannot prove whether these novel mutations contribute to its pathological functions (like serious symptoms: loss of speech or mobility, or confusion), spread, or adaptation to particular surroundings, including the age, pre-existing medical conditions, and living environment of the host. Our findings nevertheless provide important resources for tracking SARS-CoV-2 single-nucleotide variations, lineages, and clades in the imported cases of Southern China.

CONCLUSION

These imported cases in Guangdong, China, were identified, sequenced, and characterized epidemiologically and genetically, portraying the current studies on the molecular evolution of SARS-CoV-2. The descriptive epidemiology approaches

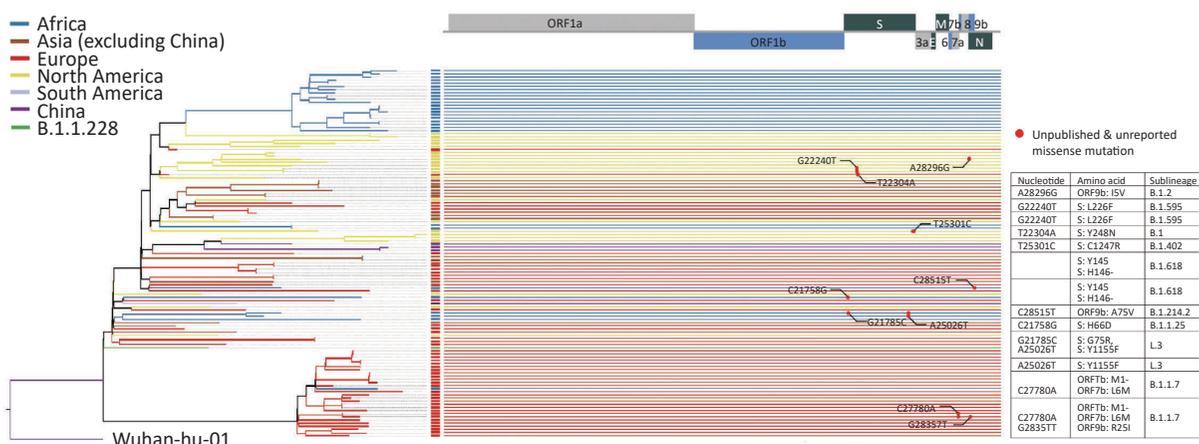


Figure 4. Maximum-likelihood phylogenetic dendrogram. The red dots in the right panel indicate that the missense mutations were not reported and not published.

demonstrated the distribution of this virus. Our phylogenetic and homology analyses increased the awareness of the detailed description of this virus. Based on the results of our viral sequence analysis, sublineage diversity and high homology at the amino acid level were present in the imported cases in Guangdong. This finding confirmed the result of a previous study reporting that SARS-CoV-2 exhibited genetic diversity over time^[32]. All our results provide an important step in the search for viral variants with reduced pathogenicity or without potential pathogenicity.

AUTHOR CONTRIBUTIONS

No potential conflicts of interest were disclosed. LIANG Dan performed the experiment, drew figures, wrote the manuscript, analyzed phylogenetic relationships, completed statistical tables, and interpreted the results. WANG Tao experimented and undertook fragment assembly. ZHANG Guan Ting and LIANG Yu Feng were in charge of data analysis and statistics and performed the literature search and the experiment. HONG Wen Shan and LI An An performed the experiment and recorded the results. LI Jia Jiao, GUAN Da Wei, DENG Xiao Ling, CHEN Feng Juan, PAN Xing Fei, and LEI Chun Liang were responsible for collecting samples and collating the basic information about patients. WANG Li and CHEN Meng Lin gave advice on the diagrams in this manuscript. JIA Hong Ling was in charge of editing the English text of a draft of this manuscript. KE Chang Wen designed the whole study framework and took responsibility for the integrity and accuracy of the data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors wish to thank Dr. LI Peng (Chinese PLA Center for Disease Control and Prevention, Beijing 510030, China) for his guidance on the diagrams in our manuscript.

Received: November 29, 2021;

Accepted: March 29, 2022

REFERENCES

- World Health Organization. Coronavirus disease (COVID-19) pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. [2021-06-17].
- Baidu. Real-time notification of new pneumonia (last updated: 2021/6/17). https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_aladin_top1. [2021-06-17]. (In Chinese)
- World Health Organization. Tracking SARS-CoV-2 variants. https://www.who.int/health-topics/nipah-virus-infection/tracking-SARS-CoV-2-variants#tab=tab_1. [2021-06-17].
- Song N, Cui GL, Zeng QL. Genomic epidemiology of SARS-CoV-2 from mainland China with newly obtained genomes from Henan province. *Front Microbiol*, 2021; 12, 673855.
- Xiong YB, Tian YX, Ma Y, et al. Factors defining the development of severe illness in patients with COVID-19: a retrospective study. *Biomed Environ Sci*, 2021; 34, 984–91.
- Ma Y, Zhu DS, Chen RB, et al. Association of overlapped and un-overlapped comorbidities with COVID-19 severity and treatment outcomes: a retrospective cohort study from nine provinces in China. *Biomed Environ Sci*, 2020; 33, 893–905.
- Li LH, Tu HW, Liang D, et al. Kinetic characteristics of neutralizing antibody responses vary among patients with COVID-19. *Biomed Environ Sci*, 2021; 34, 976–83.
- Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*, 2020; 579, 265–9.
- Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*, 2019; 17, 181–92.
- Zhu N, Zhang DY, Wang WL, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*, 2020; 382, 727–33.
- Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020; 579, 270–3.
- Gupta RK. Will SARS-CoV-2 variants of concern affect the promise of vaccines? *Nat Rev Immunol*, 2021; 21, 340–1.
- Tegally H, Wilkinson E, Giovanetti M, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*. 2020.
- Volz E, Mishra S, Chand M, et al. Transmission of SARS-CoV-2 Lineage B. 1.1. 7 in England: Insights from linking epidemiological and genetic data. *medRxiv*. 2021.
- Iacobucci G. Covid-19: local councils initiate surge vaccination to tackle B. 1. 617. 2 variant. *BMJ*, 2021; 373, n1361.
- Guangzhou Municipal Health Commission. Coronavirus disease (COVID-19) pandemic. http://wjw.gz.gov.cn/ztzl/xfyyqfk/yqtb/content/post_7284283.html. [2021-5]
- Galloway SE, Paul P, MacCannell DR, et al. Emergence of SARS-CoV-2 B. 1.1. 7 lineage - United States, December 29, 2020-January 12, 2021. *MMWR Morb Mortal Wkly Rep*, 2021; 70, 95-9.
- Horspool AM, Ye CJ, Wong TY, et al. SARS-CoV-2 B. 1.1. 7 and B. 1.351 variants of concern induce lethal disease in K18-hACE2 transgenic mice despite convalescent plasma therapy. *bioRxiv*, 2021.
- Aleem A, Akbar SAB, Slenker AK. Emerging variants of SARS-CoV-2 and novel therapeutics against coronavirus (COVID-19). StatPearls Publishing. 2021.
- Maison DP, Nerurkar VR. Research methodology to define the introduction of the SARS-CoV-2 B. 1.429 variant in Hawaii. *Res Sq*, 2021.
- Campoy PJS, Buenestado-Serrano S, Pérez-Lago L, et al. First importations of SARS-CoV-2 P. 1 and P. 2 variants from Brazil to Spain and early community transmission. *Enferm Infecc Microbiol Clin*, 2021; 40, 262-5.
- Ozer EA, Simons LM, Adewumi OM, et al. High prevalence of SARS-CoV-2 B. 1.1. 7 (UK variant) and the novel B. 1.525 lineage in Oyo State, Nigeria. *medRxiv*, 2021.

23. Yurkovetskiy L, Wang X, Pascal KE, et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell*, 2020; 183, 739–51.e8.
24. Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, 2020; 182, 812–27.e19.
25. Volz E, Hill V, McCrone JT, et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*, 2021; 184, 64–75.e11.
26. Nyayanit DA, Sarkale P, Baradkar S, et al. Transcriptome & viral growth analysis of SARS-CoV-2-infected Vero CCL-81 cells. *Indian J Med Res*, 2020; 152, 70–6.
27. Fogeron ML, Montserret R, Zehnder J, et al. SARS-CoV-2 ORF7b: is a bat virus protein homologue a major cause of COVID-19 symptoms? *bioRxiv*, 2021.
28. Kaushal N, Gupta Y, Goyal M, et al. Mutational frequencies of SARS-CoV-2 genome during the beginning months of the outbreak in USA. *Pathogens*, 2020; 9, 565.
29. Shi CS, Qi HY, Boularan C, et al. SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. *J Immunol*, 2014; 193, 3080–9.
30. Redondo N, Zaldívar-López S, Garrido JJ, et al. SARS-CoV-2 accessory proteins in viral pathogenesis: knowns and unknowns. *Front Immunol*, 2021; 12, 708264.
31. Kreimendahl S, Rassow J. The mitochondrial outer membrane protein tom70-mediator in protein traffic, membrane contact sites and innate immunity. *Int J Mol Sci*, 2020; 21, 7262.
32. Colson P, Fournier PE, Chaudet H, et al. Analysis of SARS-CoV-2 variants from 24, 181 patients exemplifies the role of globalization and zoonosis in pandemics. *Front Microbiol*, 2022; 12, 786233.