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.

Virus Research 305 (2021) 198553

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

The dominant strain of SARS-CoV-2 is a mosaicism

Wei Wang [1], Cheng-Peng Li [1], Mei He, Sheng-Wen Li, Lin Cao, Nai-Zheng Ding, Cheng-Qiang He [*]

Shandong Provincial Key Laboratory of Animal Resistance Biology, College of Life Science, Shandong Normal University, Jinan, Shandong 250014, China

ARTICLE INFO

Keywords:
COVID-19
SARS-CoV-2
Homologous recombination
Dominant strain

ABSTRACT

COVID-19 is seriously threatening human health all over the world. A comprehensive understanding of the genetic mechanisms driving the rapid evolution of its pathogen (SARS-CoV-2) is the key to controlling this pandemic. In this study, by comparing the entire genome sequences of SARS-CoV-2 isolates from Asia, Europe and America, and analyzing their phylogenetic histories, we found a lineage derived from a recombination event that likely occurred before March 2020. More importantly, the recombinant offspring has become the dominant strain responsible for more than one-third of the global cases in the pandemic. These results indicated that the recombination might have played a key role in the pandemic of the virus.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, also known as 2019-nCoV) is the pathogen of the coronavirus disease 2019 (COVID-19). First noticed in Wuhan, China in December 2019, the virus soon hit the entire globe badly. The pandemic crisis is even worsening now. As of August 2021, the virus has affected more than 220 countries, areas or territories, caused more than 209 million infections, and killed more than 4.4 million people (https://www.who.int/emergencies/diseases/novel-coronavirus-2019). To control SARS-CoV-2, scientists are concentrating on characterizing the virus and its replication dynamics (Abuin et al., 2020), tracking its movement through the human population (Lee et al., 2020; Worobey et al., 2020), exploring virus origin (Zhou et al., 2020), and developing vaccines (Sharma et al., 2020). The results of these works are based on the genetics and evolution of the virus. Continuous evolution of SARS-CoV-2 leads to the emergence of new variants, which has brought challenges to vaccine development and other control measures against the virus (Khurshid et al., 2020). Therefore, a comprehensive understanding of the genetic mechanisms underlying its evolution is the basic issue of controlling the virus. It is unknown yet whether homologous recombination, an important genetic mechanism, influences SARS-CoV-2 evolution. Here, we reported the discovery of a dominant SARS-CoV-2 lineage with a mosaic genome, revealing that homologous recombination is a notable evolutionary power of the virus.

Results and discussion In order to explore the origin of SARS-CoV-2 strains circulating in China and its neighboring countries in the first half of this year, we selected and analyzed their genome sequences deposited in the Genbank database. When comparing their genomic sequences with those of isolates from Germany and the United States (US), it was found that the genomes of a group from Bangladesh exhibited mosaic characteristic (Fig. 1A). Taking the nucleotide position 7434 of the viral genome as the boundary, the Bangladeshi group (e.g., NIB-1) shared higher genome sequence similarity with one US group (e.g., UNC_200428) forward, but with the German group (e.g., NRW-04) afterward. We also compared all variable genomic sites of three representatives of the Bangladeshi group with those of their putative parents (Fig. 1B). The inflexion of sequence similarity change was clearer (Fig. 1C). According to Fisher's exact statistics, the putative recombination breakpoint with the maximum chi-square value was located at the region around the position 7000. Delimited by the putative breakpoint, there was significant difference in the similarity of the recombinant virus to the two parents ($p < 0.01$).

Phylogenetic reconstruction based on the representatives of these virus groups (listed in Table 1) showed that they constituted three parallel lineages (Fig. 1D). Lineage I covers isolates from China, Germany, and the US, with Wuhan-Hu-1 collected in December 2019 being in an ancestral status, lineage II is mainly endemic to the US, and lineage III is composed of the mosaic isolates. Thus, one parent is a member of lineage I type and the other from lineage II type. Further, to figure out when and where this recombination event took place, a recombinant isolate was used as the query to perform BLAST in GenBank to seek its sisters. It was discovered that many isolates collected in March 2020 were clustered with the Bangladeshi group into lineage III. Notably,

* Corresponding author.
E-mail address: hchqiang@sdnu.edu.cn (C.-Q. He).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.virusres.2021.198553
Received 3 April 2021; Received in revised form 25 August 2021; Accepted 26 August 2021
Available online 4 September 2021

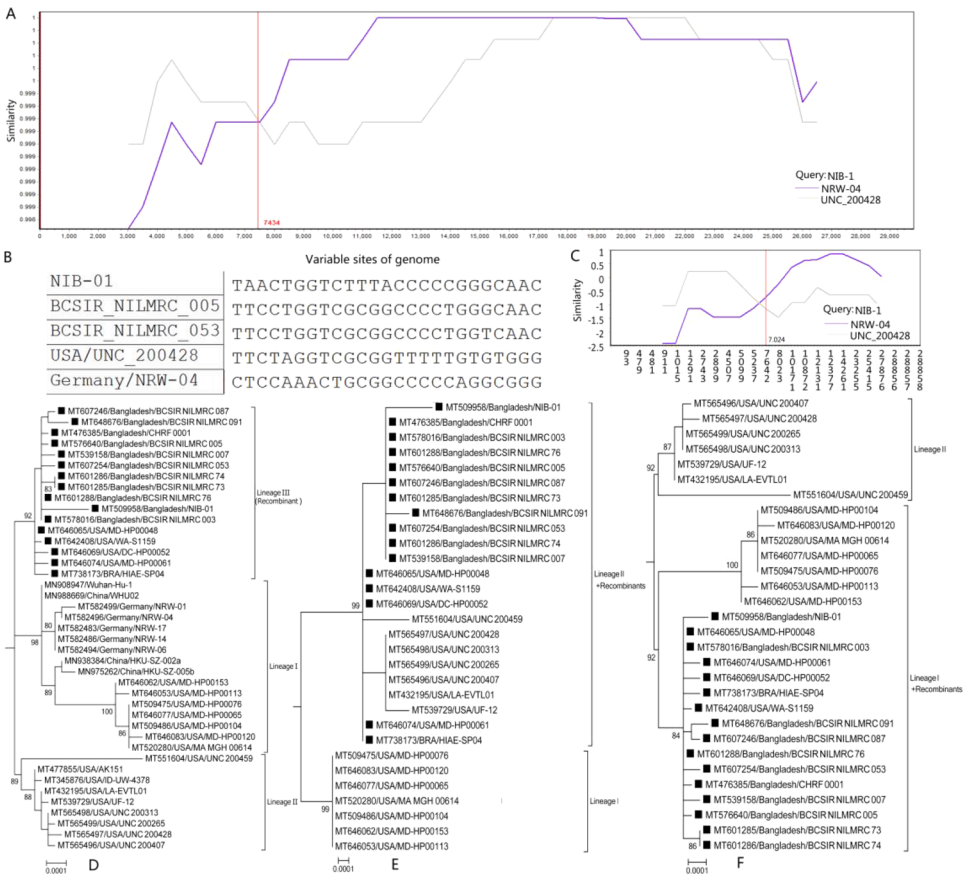

Fig. 1. Recombination evidence of SARS-CoV-2 isolates from Bangladesh

A. Comparison of the complete genome sequence of a Bangladeshi isolate (NIB-1) with those of the putative parents (NRW-04 and UNC_200428). The vertical axis is the sequence similarity; the horizontal axis is the genomic positions. B. Alignment of variant sites between the Bangladeshi group and its putative parents. C. Comparison of the variant sites between NIB-1 and its putative parents. The horizontal axis is the genomic positions of the variable sites. The vertical line indicates the location with the maximum chi-square value that is shown beside. D. Phylogenetic history of the recombinant and parent lineages inferred from complete genome sequence by Maximum Likelihood method. The recombinants were marked with "square". E. Phylogenetic history of the recombinant and parent lineages inferred from genomic region before position 7000. F. Phylogenetic history inferred from genomic region after position 7434. The bootstrap values (> 70%) were showed on each branch of the trees.

some US isolates were in the ancestral status of this mono-phylogenetic lineage (Fig. 1D). Given that both lineages I and II were prevalent in the US early in the virus outbreak, it is more likely that the recombination event occurred in the US before March 2020.

The discordant phylogenetic pattern of a virus genome is a golden indicator for judging the origins of the recombination. To further confirm this recombination event, different genomic parts of the recombinant and parent lineages were used to reconstruct their phylogenetic histories through Maximum Likelihood and Neighbor-Joining methods. For each tree, the two methods obtained the same topology except for the different bootstrap values (Figs. 1 and S1). As expected, the recombinant viruses fell into lineage II before the putative breakpoint (Fig. 1E), but were clustered with lineage I as a mono-phylogenetic group after the breakpoint (Fig. 1F). Each mono-phylogenetic group with the recombinant lineage was supported by the robust bootstrap value. Thus, the mosaic genome indeed has double origins involving the I- and II-type viruses.

To understand the impact of the recombination event on the pandemic of SARS-CoV-2, we also analyzed the virus information deposited in the GSIAID database (https://www.gisaid.org/). Based on the phylogenetic history, the SARS-CoV-2 isolates are divided into 8 clades in the database: S, O, L, V, G, GH, GR, and GV (Fig. S2). Among them, GR is the largest clade accounting for more than 30% of the available isolates in total, 65% in South America, 56% in Asia, 54% in Oceania, 36% in Africa, 30% in Europe, and 12% in North America (Fig. 2). Therefore, the GR group should be the dominant strain in the pandemic. By December 2020, GR had differentiated into a new sub-lineage GRY (Fig. S4). Interestingly, we found that these GR isolates are the offspring of the recombinant virus (Figs. S3 and S4), indicating that the recombination might have played a key role in the SARS-CoV-2 pandemic.

High mutation rates are generally considered deleterious for RNA virus with asexual reproduction (Chao, 1990). A decrease in the mean fitness of its population is continually driven by the evolutionary mechanism known as Muller's ratchet. This is where the load of deleterious mutations increases in a ratchet-like manner with successive loss of the fittest viruses (Donis, 1991; Muller, 1964). As a form of sexual reproduction, genetic recombination is thought the power to stop Muller's ratchet in asexual reproduction biology (Naito and Pawlowska, 2016). This process enables some viruses to acquire key adaptive mutations to fill a major fitness gap in a single step. Usually, a dominant strain represents the highly adaptable variant. The recombinant SARS-CoV-2 has become the largest dominant strain, suggesting that the recombination might have conferred high adaptability on the virus.

For coronavirus, its replication is mediated by the polymerase Nsp12 without proofreading function (Robson et al., 2020). Although its non-structure protein Nsp14 has a limited proofreading function (Eckerle et al., 2010), rapid mutation is still an important feature of the virus. This suggests the need for recombination to tune balance between its quasispecies diversity and replicative fitness. In addition, its transcription of the 3′proximal structural and accessory protein genes is a discontinuous process and result in a nested set of subgenomic mRNAs (sgmRNAs) (Sethna et al., 1989). These sgmRNA might play an important role in template-switching during minus-strand synthesis (Robson et al., 2020). This may provide a key genetic basis for the frequent recombination of coronavirus.

Genomic epidemiology allows for effective reconstruction of the geographical spread as well as estimation of the key epidemiological quantities of SARS-CoV-2 (Rambaut et al., 2020). Along with the phylogeny, viral genome sequencing has become a powerful tool in understanding and tracking the dynamics of SARS-CoV-2. However, the accuracy of phylogenetic history may be severely impaired by genetic recombination (Martin et al., 2005), resulting in an unreasonable conclusion about the origin of the virus. Therefore, it might be necessary

Table 1
SARS-CoV-2 representative used in this study.

Access_number_	Isolate	Isolation_source	Country	Collection_date
MT646074	MD-HP0006	Human	USA	2020-03
MT551604	UNC_200459	Human	USA	2020-04
MT539729	UF-12	Environment	USA	2020-04
MT345876	ID-UW-4378	Human	USA	2020-03
MT477855	AK151	Human	USA	2020-04
MT432195	LA-EVTL01	Human	USA	2020-04
MT565496	UNC_200407	Human	USA	2020-04
MT565496	UNC_200407	Human	USA	2020-04
MT565497	UNC_200428	Human	USA	2020-04
MT565499	UNC_200265	Human	USA	2020-04
MT565498	UNC_200313	Human	USA	2020-04
MT539729	UF-12	Environment	USA	2020-04
MT432195	LA-EVTL01	Human	USA	2020-04
MT477855	AK151	Human	USA	2020-04
MT345876	ID-UW-4378	Human	USA	2020-03
MT551604	UNC_200459	Human	USA	2020-4
MT509486	MD-HP00104	Human	USA	2020-03
MT646083	MD-HP00120	Human	USA	2020-03
MT520280	MA_MGH_00614	Human	USA	2020-03
MT646077	MD-HP00065	Human	USA	2020-03
MT509475	MD-HP00076	Human	USA	2020-03
MT646053	MD-HP00113	Human	USA	2020-3
MT646062	MD-HP00153	Human	USA	2020-3
MN975262	HKU-SZ-005b	Human	China	2020-01
MN938384	HKU-SZ-002a	Human	China	2020-01
MN988669	WHU02	Human	China	2020-01
MN908947	Wuhan-Hu-1	Human	China	2019-12
MT582499	NRW-01	Human	Germany	2020-02
MT582494	NRW-06	Human	Germany	2020-02
MT582496	NRW-04	Human	Germany	2020-02
MT582483	NRW-17	Human	Germany	2020-03
MT582486	NRW-14	Human	Germany	2020-03
MT509958	NIB-1	Human	Bangladesh	2020-05
MT646065	MD-HP00048	Human	USA	2020-03
MT578016	BCSIR_NILMRC_003	Human	Bangladesh	2020-05
MT648676	BCSIR_NILMRC_091	Human	Bangladesh	2020-06
MT607246	BCSIR_NILMRC_087	Human	Bangladesh	2020-05
MT601288	BCSIR_NILMRC_76	Human	Bangladesh	2020-06
MT601285	BCSIR_NILMRC_73	Human	Bangladesh	2020-06
MT601286	BCSIR_NILMRC_74	Human	Bangladesh	2020-06
MT646069	DC-HP00052	Human	USA	2020-03
MT576640	BCSIR_NILMRC_005	Human	Bangladesh	2020-05
MT738173	HIAE-SP04	Human	Brazil	2020-03

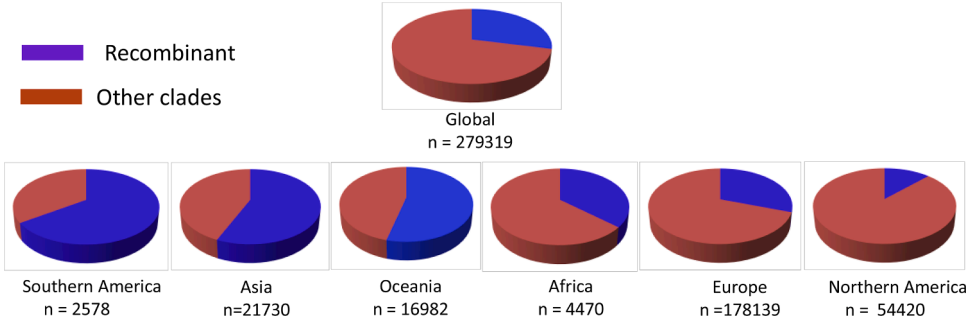

Fig. 2. The distribution of recombinant offspring in each continent n: available SARS-CoV-2 genome number in GISAID.

to carefully exclude the recombination event when molecular epidemiology of COVID-19 is surveyed using SARS-CoV-2 genome sequences.

Recombination is an important evolution power of CoVs. Genetic recombination has been previously documented for different CoVs, including SARS-CoV (Lau et al., 2010) and MERS-CoV (Sabir et al., 2016). For the occurrence of recombination, co-infection of SARS-CoV-2 in a host cell is necessary. Long-term existence in COVID-19 patients (Wang et al., 2020) and high prevalence of asymptomatic infection of SARS-CoV-2 might increase the chance of co-infection. To data, virus co-infection has not been reported in the COVID-19 patients; however, re-infection has been found repeatedly. The emergence of the recombinant lineage suggested that a COVID-19 patient might be subjected to co-infection of two distinct SARS-CoV-2 strains.

Currently, there is an urgent need to develop SARS-CoV-2 vaccines to control the ongoing pandemic. However, recombination between viruses might result in their antigen shift and fitness change in the host (Lowen, 2017; Ludwig-Begall et al., 2020). Thus, the recombination between SARS-CoV-2 strains might bring a challenge to the development of effective vaccines against the virus.

In all, our analyses demonstrated that SARS-CoV-2 can undergo

recombination during its natural infection and circulation. The dominant strain of the virus is mosaic, suggesting that homologous recombination might have played a key role in the pandemic. And also, SARS-CoV-2 may adopt recombination for rapid evolution, leading to the emergence of novel variants with higher adaption, which we should be aware of and pay more attention to.

2. Materials and methods

SARS-CoV-2 genomes were collected from GenBank and GISAID, and aligned with CLUSTAL W. The representatives of the recombinant and parent lineages were listed in Table 1. Phylogenetic histories were reconstructed employing Maximum Likelihood and Neighbor-Joining methods implemented in MEGA X with the optimal substitution models and rates among sites (Kumar et al., 2018). The robustness of lineage was tested by bootstrap method with 1000 replicates. A lineage with more than 70% bootstrap value was considered robust. Sequence similarity comparison and similarity graphical representation were carried out with the help of SimPlot program (Lole et al., 1999). Fisher's exact statistics were used to determine whether there was a significant difference in the similarity between the recombinant and parent sequences in different regions. Using one Bangladeshi isolate as the query, basic local alignment search tool (BLAST) was employed to search for its sisters in GenBank.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by grants from the Key R&D project of Shandong Province (2017GNC10125, 2019GSF107020). The funding bodies did not play a role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2021.198553.

References

Abuin, P., Anderson, A., Ferramosca, A., Hernandez-Vargas, E.A., Gonzalez, A.H., 2020. Characterization of SARS-CoV-2 dynamics in the host. Annu. Rev. Control 50, 457–468. https://doi.org/10.1016/j.arcontrol.2020.09.008.

Chao, L., 1990. Fitness of RNA virus decreased by Muller's ratchet. Nature 348, 454–455.

Donis, R.O., 1991. Muller's ratchet and flu virus. Nature 353, 308–309.

Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R.L., Baric, R.S., Stockwell, T.B., Spiro, D.J., Denison, M.R., 2010. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. PLoS Pathog. 6, e1000896.

Khurshid, A., Ammar Ahmed, M., Aziz, A., Amin, R., 2020. Living with coronavirus (COVID-19): a brief report. Eur. Rev. Med. Pharmacol. Sci. 24, 10902–10912.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.

Lau, S.K., Li, K.S., Huang, Y., Shek, C.T., Tse, H., Wang, M., Choi, G.K., Xu, H., Lam, C.S., Guo, R., Chan, K.H., Zheng, B.J., Woo, P.C., Yuen, K.Y., 2010. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. J. Virol. 84, 2808–2819.

Lee, E.C., Wada, N.I., Grabowski, M.K., Gurley, E.S., Lessler, J., 2020. The engines of SARS-CoV-2 spread. Science 370, 406–407.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype *C*-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–160.

Lowen, A.C., 2017. Constraints, drivers, and implications of influenza A virus reassortment. Annu. Rev. Virol. 4, 105–121.

Ludwig-Begall, L.F., Lu, J., Hosmillo, M., de Oliveira-Filho, E.F., Mathijs, E., Goodfellow, I., Mauroy, A., Thiry, E., 2020. Replicative fitness recuperation of a recombinant murine norovirus - *in vitro* reciprocity of genetic shift and drift. J Gen. Virol. 101, 510–522.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.

Muller, H.J., 1964. The relation of recombination to mutational advance. Mutat. Res. 106, 2–9.

Naito, M., Pawlowska, T.E., 2016. Defying Muller's ratchet: ancient heritable endobacteria escape extinction through retention of recombination and genome plasticity. mBio 7 e02057–e02015.

Rambaut, A., Holmes, E.C., O'Toole, A., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403–1407.

Robson, F., Khan, K.S., Le, T.K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., Ng, W.L., 2020. Coronavirus RNA proofreading: molecular basis and therapeutic targeting. Mol. Cell 79, 710–727.

Sabir, J.S., Lam, T.T., Ahmed, M.M., Li, L., Shen, Y., Abo-Aba, S.E., Qureshi, M.I., Abu-Zeid, M., Zhang, Y., Khiyami, M.A., Alharbi, N.S., Hajrah, N.H., Sabir, M.J., Mutwakil, M.H., Kabli, S.A., Alsulaimany, F.A., Obaid, A.Y., Zhou, B., Smith, D.K., Holmes, E.C., Zhu, H., Guan, Y., 2016. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science 351, 81–84.

Sethna, P.B., Hung, S.L., Brian, D.A., 1989. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. Proc. Natl. Acad. Sci. U S A. 86, 5626–5630.

Sharma, O., Sultan, A.A., Ding, H., Triggle, C.R., 2020. A review of the progress and challenges of developing a vaccine for COVID-19. Front. Immunol. 11, 585354.

Wang, X., Huang, K., Jiang, H., Hua, L., Yu, W., Ding, D., Wang, K., Li, X., Zou, Z., Jin, M., Xu, S., 2020. Long-term existence of SARS-CoV-2 in COVID-19 patients: host immunity, viral virulence, and transmissibility. Virol. Sin.

Worobey, M., Pekar, J., Larsen, B.B., Nelson, M.I., Hill, V., Joy, J.B., Rambaut, A., Suchard, M.A., Wertheim, J.O., Lemey, P., 2020. The emergence of SARS-CoV-2 in Europe and North America. Science 370, 564–570.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature.