Omicron: a chimera of two early SARS-CoV-2 lineages

Chenglong Xiong (xiongchenglong@fudan.edu.cn)
School of Public Health, Fudan University

Xiuliang Liu
Department of Epidemiology, School of Public Health, Fudan University

Jiasheng Xiong
Department of Social Medicine, School of Public Health, Fudan University

Zhong Sun
Universiti Putra Malaysia

Wei Hu
Fudan university, Shanghai

Jingjing Hu
Shanghai Pinnacles Medical Technology Co., Ltd

Yuqian Wang
Department of Epidemiology, School of Public Health, Fudan University

Kunyu Li
Department of Integrative Medicine and Neurobiology, School of Basic Medical Sciences; Institutes of Brain Science, Brain Science Collaborative Innovation Center, State Key Laboratory of Medical Neu

Karuppiah Thilakavathy
Universiti Putra Malaysia

Mingquan Chen
Department of Emergency, and Department of Infectious Diseases, Huashan Hospital, Fudan University, 12 Middle Urumqi Road, Shanghai, 200040, China

Qi Zhao
Department of Social Medicine, School of Public Health, Fudan University

Yi Feng
Fudan University https://orcid.org/0000-0002-7156-6023

Qingwu Jiang
Department of Epidemiology, School of Public Health, Fudan University

---

Letter

Keywords:

Posted Date: January 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1302993/v1

Abstract

The current global epidemiology of COVID-19 is now characterized by the emergence and rapid spread of the SARS-CoV-2 Omicron variant on a global scale[1,2]. Despite the variant's prompt predominance, there remain knowledge gaps in its origin and evolution history[3–6]. Here, we show that Omicron lineage SARS-CoV-2 is characterized by the feature of chimera. It was generated by genomic recombination of two early PANGO lineages of SARS-CoV-2. In the recombination event, strains with medium or high circulating intensity like SARS-CoV-2/human/USA/COR-21-434196/2021 belonging to PANGO lineage BA.1 provided the fundamental genome and served as the major parents, while the rare lineage strains like SARS-CoV-2/human/IRN/Ir-3/2019 belonging to B.35, as the minor parents, hybridized their genomic fractions into the major genomes at position 21593-23118nt. This recombination event results in 22 amino acid residue substitutions for the variant of Omicron, including 16 in the pivotal RBD of the spike protein. These substitutions have led to some subtle variations in the spatial structure and the affinity to hACE2 receptor of the spike protein[7,8], thereby raising concerns about the effectiveness of available vaccines and antibody therapeutics[9–12]. The global spread and explosive growth of the SARS-CoV-2 in human population increase opportunities for future recombination[13–15].

Main Text

The outbreak of the current coronavirus disease (COVID-19) occurred in late 2019 and quickly spread globally. It has posed a remarkable threat to public health around the world[16,17]. This pandemic is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which belongs to the *Betacoronavirus* genus and is characterized by an unsegmented, positive-sense, single-stranded RNA genome[18,19]. It is known that SARS-CoV-2 is a genetically diverse group that mutates continuously leading to the emergence of multiple variants[9,10,15].

As new variants become more widespread, additional genetic markers emerge[12,14]. By determining the phyletic lineages of the SARS-CoV-2 genomic variants and those of the conserved regions in the accessory and spike proteins of all the SARS-related coronaviruses, research in the evolution of molecular pathways involved in emergence of pandemic is critical for the development of therapeutics and vaccines as well as the prevention of future zoonosis[9,10,13]. One tool for this purpose is the PANGO (Pangolin lineage) nomenclature by Rambaut et al which takes a granular approach to classify and describe viral evolution with detailed lineages; as evidence becomes available, classifications of variants will be revised to reflect the continuous evolution of circulating variants and their changing epidemiology[20,21]. Potential variants of concern (VOCs), variants of interest (VOIs) or variants under monitoring (VUMs) are regularly assessed based on the risk posed to global public health[22].

Following the identification of a novel variant in South Africa on 24 November 2021, WHO designated Omicron (clade GRA, PANGO lineage B.1.1.529 and descendants BA.1 and BA.2) as the fifth SARS-CoV-2 VOC two days later due to its large number of substitutions[2,4]. The variant has since spread to most

countries. The current global epidemiology of SARS-CoV-2 is now characterized by the emergence and rapid spread of the Omicron variant on a global scale, continued decline in the prevalence of the previous Delta and other variants[1]. Despite its prompt predominance, there remain knowledge gaps in its origin and evolution, which has attracted people's interests and speculations[3–6]. Here, we propose that Omicron variant may be derived from recombination of two early PANGO lineages of SARS-CoV-2.

We retrieved a total of 4,192 whole-length genomes of SARS-CoV-2 from EpiCoV^{TM} database of Global Initiative on Sharing All Infuenza Data (GISAID) and SARS-CoV-2 data (NCBI). These genome sequences belong to 1,263 PANGO lineages, including 29 lineages of VOCs, VOIs, VUMs and formerly monitored variants (FMVs) according WHO's Tracking SARS-CoV-2 variants (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/, accessed December 18, 2021), and are those with the earliest collection times within each PANGO lineage (**Extended Data 1 and 2**: **Tab. S1** and **S2**). After quality control which is mainly assessed by the extent of sequencing completion, 2609 whole-length genomes of SARS-CoV-2 were used for the first round rapid screen (**Extended Data 3**, **Genome sequence matrix 1**), while the genome sequences involved in all putative recombination events identified by the first round screen were picked out for further verification (**Extended Data 4**, **Genome sequence matrix 2**). Taking SARS-CoV-2/human/USA/UT-UPHL-211211887190/2021 (Accession, OL920485) as the query genome sequence, recombination events were detected and verified by Recombination Detection Program (RDP) v4.101[23,24] and the SimPlot Program package[25].

We confirmed that at least one recombination event occurred in the origin and evolution history of Omicron variant of SARS-CoV-2. In this event, strains like SARS-CoV-2/human/USA/COR-21-434196/2021 (Accession, OL849989) belonging to PANGO lineage BA.1 provided the fundamental genome for VOC Omicron and served as its major parents, while strains like SARS-CoV-2/human/IRN/Ir-3/2019 (Accession, MW737421) belonging to PANGO lineage B.35, as the minor parents, hybridized the genomic fractions into the major genome at the position of 21593-23118 nt (Fig. 1, **Extended Data 5**: **Fig. S1**). From the perspective of encoding protein by the recombination fraction, it may have a profound impact on the pathogenicity and transmission potential of the novel variant, i.e., Omicron lineage of SARS-CoV-2. This fraction encodes 144-505 amino acid residues of SARS-CoV-2's spike protein (S). As a result of the recombination, VOC Omicron did derived the substitutions of N211I, L212V, V213R, R214E, deletion215P, deletion216E, R346K, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, from the minor parent of SARS-CoV-2/human/IRN/Ir-3/2019-like strains, while the substitution of G/D339D may come from a back mutation after recombination. All these substitutions locate in the NTD (N-terminal domain, residues 18~330) and RBD (receptor-binding domain, residues 331~528) of the S1 subunit of spike protein[7,26,27], and even up to 16 in RBD (Tab. 1). The consistency of amino acid residues encoded by VOC Omicron and its minor parent in the corresponding fraction, as well as the difference between it and the major parent, proved at the level of amino acids that the recombination event may have actually happened. It is known that SARS-CoV-2 Omicron variant encodes 37 amino acid substitutions (including insertions and deletions) in the spike protein[7,8], and then the recombination event alone leads to 22 of them.

By checking the isolation frequency of BA.1 in the databases, it is a lineage with medium or high circulating intensity. There are 292,755 and 82,610 isolates in GISAID's EpiCoV^{TM} database and SARS-CoV-2 data of NCBI (both accessed January 13, 2022), accounting for 4.17% (292755/7022141) and 2.57% (82610/3218586) of the total isolates in the two databases, respectively, while B.35 is obviously a rare lineage, accounting for 0.0019% (130/7022141) and 0.0018% (59/3218586) respectively in the two databases, and its striking area is more limited to several countries such as United Kingdom, Iceland, Australia, USA, Jordan, Timor-Leste and New Zealand. Fig. 2 shows the temporal order and isolation frequency of the PANGO lineages involved in the recombination event and the VOCs, VOIs, VUMs and FMVs then.

Interestingly, both major parent BA.1 and minor parent B.35 are lineages that emerged in the early outbreak of COVID-19. Considering the rapid transmission potential of lineage BA.1, which accounted for 71.9% of the isolating proportion in less than two months after the emergence, BA.1 cannot be the descendant but the parent of VOC Omicron. In fact, when investigating 5,100 genomes of BA.1 PANGO lineage SARS-CoV-2, we found these viruses were different from the major parent of recombination only at 346 amino acid residue (R *vs.* K) but fundamentally different from the index Omicron variants (Accessions, OL920485, OL901845, and OL902308), which indicated that BA.1 lineage should be the recombination parent of VOC Omicron rather than the descendant of it (Tab. 1). However, the recombination event for VOC Omicron did not occurred until recently. The reason may be that the circulating frequency of the lineage B.35 is too low and its striking area is too limited, which reduces the chance of recombination between it and other lineages of SARS-CoV-2. After all, the prerequisite for recombination is that no less than two lineages of viruses co-infect an individual simultaneously[28,29], but the isolation frequency of the minor parent lineage B.35 is so rare that it hardly has the opportunity to infect an individual with a prior SARS-CoV-2 infection, and *vice versa*.

Our study suggested that Omicron PANGO lineage SARS-CoV-2 is characterized by the feature of chimera. It was generated by genomic recombination of two early SARS-CoV-2 lineages in the coding sequence (CDS) of spike protein, and the recombination event results in 22 amino acid residue substitutions for the variant of Omicron, including 16 in the pivotal RBD. Spike protein is the most critical structural protein of SARS-CoVs, which is responsible for recognizing and binding to the surface receptors of host cells[26,27,30]. The recently emerged SARS-CoV-2 Omicron variant encodes 37 amino acid substitutions in the spike protein, thereby raising concerns about the effectiveness of available vaccines and antibody therapeutics[9–12]. Unfortunately, these concerns do not seem unreasonable. It has been reported that these substitutions have led to some subtle variations in the spatial structure and the affinity to hACE2 receptor of the spike protein[3,8,31]. More importantly, it has caused the immune escape of Omicron variant to the available vaccines and antibody therapeutics[11,32–34].

Recombination is proposed to be critical for coronavirus diversity and emergence of SARS-CoV-2, MERS-CoV, SARS-CoV (2002), and other zoonotic CoVs. It allows viruses to overcome selective pressure and

adapt to new hosts and environments[35–37]. Viral recombination between different CoVs within animal populations may lead to the emergence of novel zoonotic CoVs that are lethal to humans[38]. Spike proteins are type I membrane glycoproteins with signal peptides used for receptor binding and play a crucial role in viral attachment, fusion and entry, being a target for development of antibodies, entry inhibitors and vaccines[39,40]. Recombination events in the evolutionary history of the spike protein have particular significance for the current pandemic. Studies comparing coronavirus strains that are closely related to SARS-CoV-2 have proposed that SARS-CoV-2 acquired the ability to infect human cells through recombination within the spike protein sequence[37,41,42]. The enrichment for recombination found at spike protein is in agreement with many other recently published works as well[37,41,43,44].

Undoubtedly, recombination may be occurring during infections in humans[36,45]. The global spread and explosive growth of the SARS-CoV-2 in human population has contributed additional mutational variability into this genome, increasing opportunities for future recombination. It has been reported that recombination among SARS-CoV-2 is associated with increased spread and severe disease, and has resulted in vaccine failure[41,46]. Thus, targeting the ability of the virus to recombine is a critical consideration for vaccine development in the ongoing SARS-CoV-2 pandemic as well as future animal and zoonotic CoVs.

Declarations

Acknowledgements

This research was funded by the National Natural Science Foundation of China (Grant No. 81872673), the Three-Year Action Plan of Shanghai Public Health System Construction-Key Discipline Construction (2020-2022, No. GWV-10.1-XK03), and the Three-Year Initiative Plan for Strengthening Public Health System Construction in Shanghai (2020-2022, GWV-10.1-XK23). The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report. We acknowledge the contributions of scientists and researchers from all over the world for depositing the genomes of SARS-CoV-2 in SARS-CoV-2 data (NCBI) and EpiCoV^{TM} database of Global Initiative on Sharing All Infuenza Data (GISAID).

Author contributions

Conceived research and designed study: C.X., Q.J., Q.Z., Y.F., and M.C. Bioinformatic analyses: X.L., J.X., Z.S., J.H., and Y.W. Epidemiology analyses: J.X., W.H., and K.L. Interpreted Data: C.X., K.T., Q.J., Q.Z., and Y.F. First draft: C.X., X.L, and J.X. Revision: C.X., X.L, J.X., Z.S., W.H., J.H., Y.W., K.L., K.T., M.C., Q.Z., Y.F., and Q.J. Supervision: M.C., and Q.J.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Chenglong Xiong (xiongchenglong@fudan.edu.cn), Yi Feng (fengyi17@fudan.edu.cn), or Qi Zhao (zhaoqi@shmu.edu.cn). The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication, but GISAID data access, if needed, requires registration.

Peer review information

We thanks the reviewers for their contribution to the peer review of this work.

References

1 WHO, COVID-19 Weekly Epidemiological Update. **Edition 75** 18, 2022 (2022).

2 CDC, SARS-CoV-2 B.1.1.529 (Omicron) Variant - United States, December 1-8, 2021. *MMWR Morb Mortal Wkly Rep* **70** 1731 (2021).

3 Ma, W. *et al.*, Genomic perspectives on the emerging SARS-CoV-2 omicron variant. *Genomics Proteomics Bioinformatics* (2022).

4 Ferre, V. M. *et al.*, Omicron SARS-CoV-2 variant: What we know and what we don't. *Anaesth Crit Care Pain Med* **41** 100998 (2021).

5 Tong, C., Shi, W., Zhang, A. & Shi, Z., Tracking and controlling the spatiotemporal spread of SARS-CoV-2 Omicron variant in South Africa. *Travel Med Infect Dis* **46** 102252 (2021).

6 Kupferschmidt, K., Where did 'weird' Omicron come from? *SCIENCE* **374** 1179 (2021).

7 Cameroni, E. *et al.*, Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *NATURE* (2021).

8 A. J. Venkatakrishnan, P. A. P. J., Omicron variant of SARS-CoV-2 harbors a unique insertion mutation of putative viral or human genomic origin. *Preprint* **December** 10 (2021).

9 Bano, I., Sharif, M. & Alam, S., Genetic drift in the genome of SARS COV-2 and its global health concern. *J MED VIROL* **94** 88 (2022).

10 Akram, F. *et al.*, Insights into the evolutionary and prophylactic analysis of SARS-CoV-2: A review. *J VIROL METHODS* **300** 114375 (2022).

11 Gharpure, R. *et al.*, Multistate Outbreak of SARS-CoV-2 Infections, Including Vaccine Breakthrough Infections, Associated with Large Public Gatherings, United States. *EMERG INFECT DIS* **28** 35 (2022).

12 Araf, Y. *et al.*, Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. *J MED VIROL* (2022).

13 Singh, D. & Yi, S. V., On the origin and evolution of SARS-CoV-2. *EXP MOL MED* **53** 537 (2021).

14 Angeli, F., Spanevello, A., Reboldi, G., Visca, D. & Verdecchia, P., SARS-CoV-2 vaccines: Lights and shadows. *EUR J INTERN MED* **88** 1 (2021).

15 Awadasseid, A., Wu, Y., Tanaka, Y. & Zhang, W., SARS-CoV-2 variants evolved during the early stage of the pandemic and effects of mutations on adaptation in Wuhan populations. *INT J BIOL SCI* **17** 97 (2021).

16 Zhou, P. *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *NATURE* **579** 270 (2020).

17 Lu, R. *et al.*, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *LANCET* **395** 565 (2020).

18 Shang, J. *et al.*, Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* **117** 11727 (2020).

19 Jackson, C. B., Farzan, M., Chen, B. & Choe, H., Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol* **23** 3 (2022).

20 Rambaut, A. *et al.*, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *NAT MICROBIOL* **5** 1403 (2020).

21 O'Toole, A. *et al.*, Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* **7** b64 (2021).

22 WHO, Tracking SARS-CoV-2 variants. (2021).

23 Martin, D. & Rybicki, E., RDP: detection of recombination amongst aligned sequences. *BIOINFORMATICS* **16** 562 (2000).

24 Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B., RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* **1** v3 (2015).

25 Lole, K. S. *et al.*, Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *J VIROL* **73** 152 (1999).

26 Wang, Q. *et al.*, Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *CELL* **181** 894 (2020).

[27] Lan, J. *et al.*, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *NATURE* **581** 215 (2020).

[28] Soli, R., Kaabi, B., Barhoumi, M., Maktouf, C. & Ahmed, S. B., Bayesian phylogenetic analysis of the influenza-A virus genomes isolated in Tunisia, and determination of potential recombination events. *MOL PHYLOGENET EVOL* **134** 253 (2019).

[29] Savolainen-Kopra, C. & Blomqvist, S., Mechanisms of genetic variation in polioviruses. *REV MED VIROL* **20** 358 (2010).

[30] Walls, A. C. *et al.*, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *CELL* **181** 281 (2020).

[31] Wang, R., Chen, J. & Wei, G. W., Mechanisms of SARS-CoV-2 Evolution Revealing Vaccine-Resistant Mutations in Europe and America. *J PHYS CHEM LETT* **12** 11850 (2021).

[32] Lu, L. *et al.*, Neutralization of SARS-CoV-2 Omicron variant by sera from BNT162b2 or Coronavac vaccine recipients. *CLIN INFECT DIS* (2021).

[33] Wang, Y. *et al.*, The significant immune escape of pseudotyped SARS-CoV-2 variant Omicron. *Emerg Microbes Infect* **11** 1 (2022).

[34] Nemet, I. *et al.*, Third BNT162b2 Vaccination Neutralization of SARS-CoV-2 Omicron Infection. *N Engl J Med* (2021).

[35] Perez-Losada, M., Arenas, M., Galan, J. C., Palero, F. & Gonzalez-Candelas, F., Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *INFECT GENET EVOL* **30** 296 (2015).

[36] Gribble, J. *et al.*, The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLOS PATHOG* **17** e1009226 (2021).

[37] Li, X. *et al.*, Emergence of SARS-CoV-2 through recombination and strong purifying selection. *SCI ADV* **6** (2020).

[38] Li, C. X. *et al.*, A critical analysis of SARS-CoV-2 (COVID-19) complexities, emerging variants, and therapeutic interventions and vaccination strategies. *BIOMED PHARMACOTHER* **146** 112550 (2022).

[39] Woo, P. C., Huang, Y., Lau, S. K. & Yuen, K. Y., Coronavirus genomics and bioinformatics analysis. *Viruses* **2** 1804 (2010).

[40] Tai, W. *et al.*, Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *CELL MOL*

IMMUNOL **17** 613 (2020).

41 Patino-Galindo, J. A. *et al.*, Recombination and lineage-specific mutations linked to the emergence of SARS-CoV-2. *GENOME MED* **13** 124 (2021).

42 Haddad, D. *et al.*, SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. *PLOS ONE* **16** e251368 (2021).

43 Wu, F. *et al.*, A new coronavirus associated with human respiratory disease in China. *NATURE* **579** 265 (2020).

44 Bobay, L. M., O'Donnell, A. C. & Ochman, H., Recombination events are concentrated in the spike protein region of Betacoronaviruses. *PLOS GENET* **16** e1009272 (2020).

45 Yi, H., 2019 Novel Coronavirus Is Undergoing Active Recombination. *CLIN INFECT DIS* **71** 884 (2020).

46 Kow, C. S., Merchant, H. A. & Hasan, S. S., Mortality risk in patients infected with SARS-CoV-2 of the lineage B.1.1.7 in the UK. *J Infect* **83** e14 (2021).

Table

Table 1 is available in the supplementary files section.

Figures

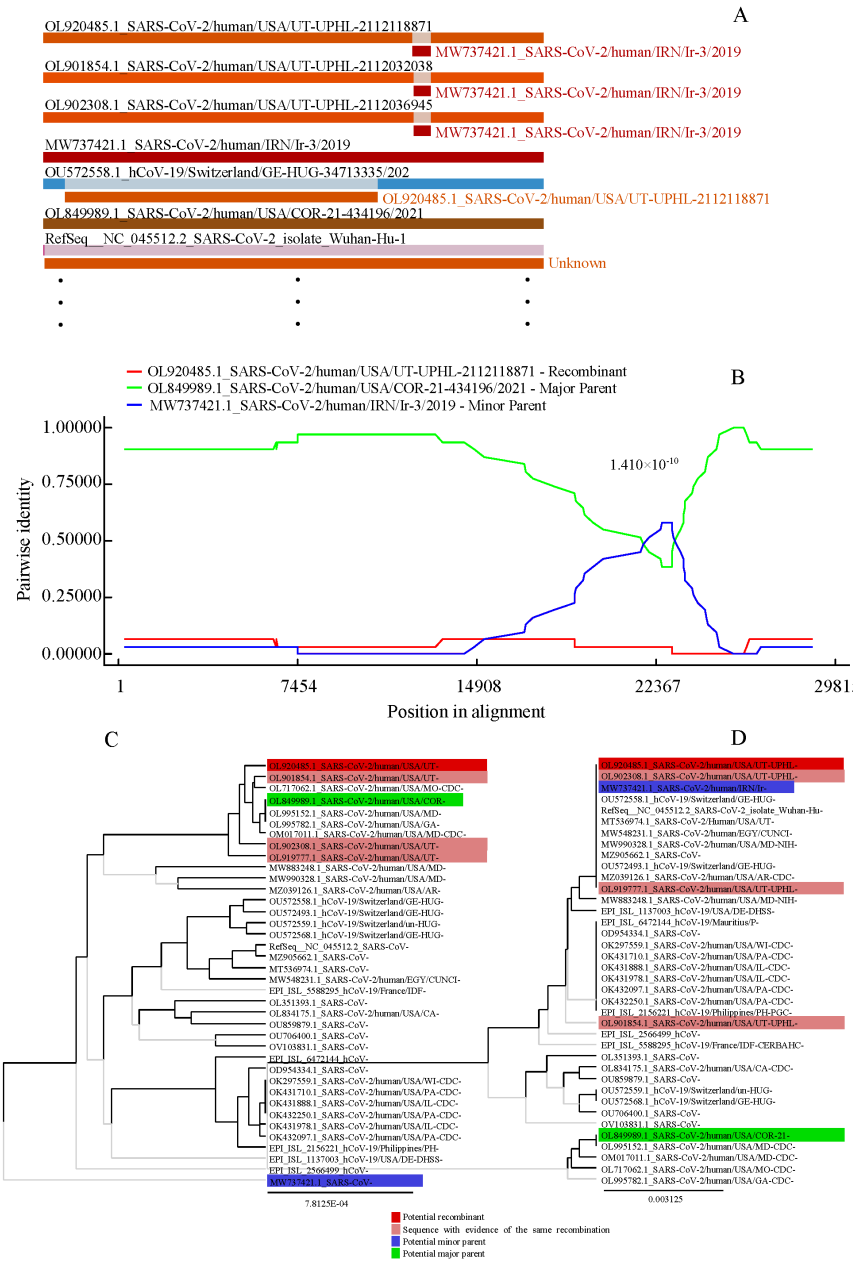

Figure 1

Panel of information related to the recombination event

A, schematic sequence display for overview the recombination event(s) about the Omicron SARS-CoV-2 variants. B, the plot diagram checked by RDP method, shows the *p* value of it. C, and D, split UPGMA trees

of the fractions derived from major and minor parents. In parts B, C, and D, curve or sequence(s) in red, green, and blue are potential recombinants, the major and minor parents, respectively.

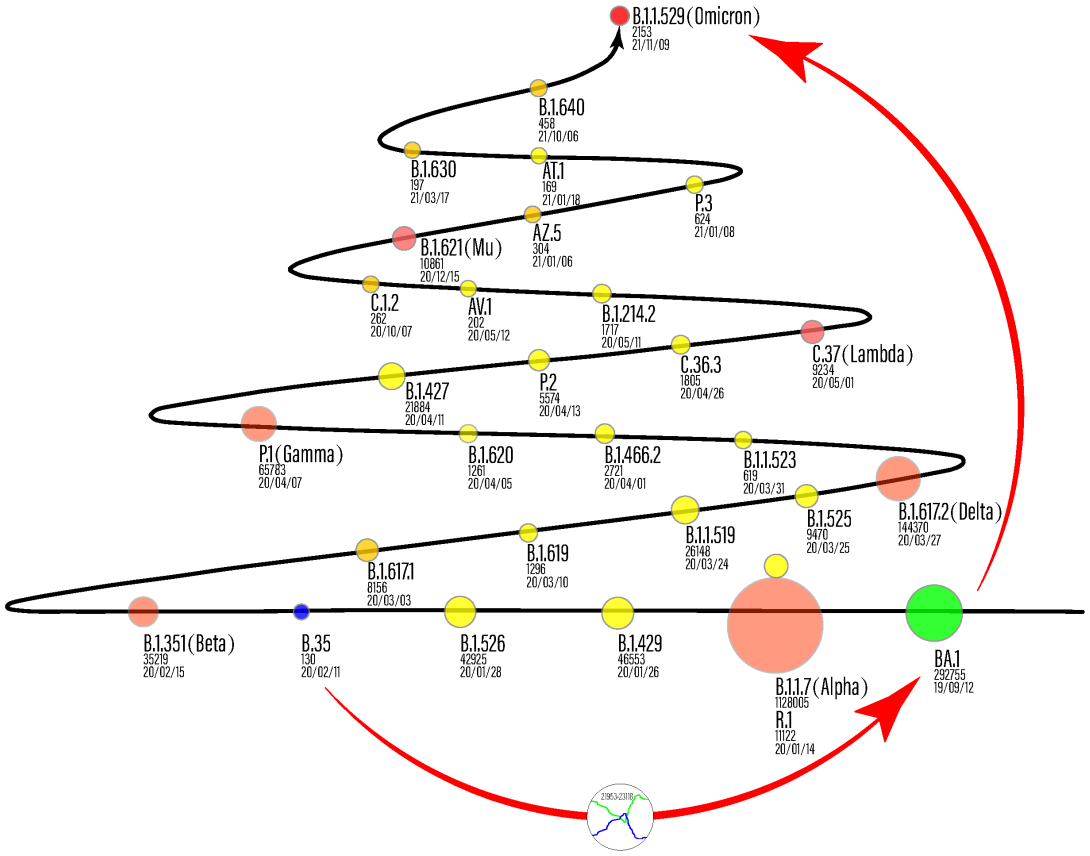

Figure 2

The temporal order and isolation frequency of the PANGO lineages involved in the recombination event and the VOCs, VOIs, VUMs and FMVs

The dot area is calculated according to the natural logarithm of the isolate numbers within each PANGO lineage, and that of B.35 is used as a reference to show the proportion of others. The data derived from EpiCoV^{TM} database of Global Initiative on Sharing All Influenza Data (GISAID, accessed December 18, 2021). Dots in red, green, and blue are potential recombinants, the major and minor parents, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ExtendedData1Tab.S1.xlsx
- ExtendedData2Tab.S2.xlsx
- ExtendedData4Genomesequencematrix238SQ.txt
- ExtendedData5Fig.S1.tif
- ExtendedData3Genomesequencematrix12609SQ.rar
- Tab.1Aminoacidsubstitutionscorrespondingtotherecombinationfraction.docx