Whole-genome comparison of representatives of all variants of SARS-CoV-2, including subvariant BA.2 and the GKA clade

Ida Bagus Kade Suardana
Udayana University
Bayu K. Mahardika
Udayana University
Made Pharmawati
Udayana University
Putu H. Sudipa
Udayana University
Tri Komala Sari
Udayana University
Nyoman B. Mahendra
Udayana University https://orcid.org/0000-0001-9336-8186
Gusti N. Mahardika (gnmahardika@unud.ac.id)
Udayana University https://orcid.org/0000-0001-5525-0793

Research Article

Keywords: SARS-CoV-2, Variants, BA.2, GKA clade, Deltacron

Posted Date: April 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1526043/v2

Abstract

Since its discovery at the end of 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has rapidly evolved into many variants, including subvariant BA.2 and the GKA clade. Genomic clarification is needed for better management of the current pandemic as well as the possible re-emergence of novel variants. The sequence of the original strain Wuhan-Hu-1 and approximately 20 representatives of each variant were downloaded from GenBank and GISAID. Two representatives with no track of un-definitive nucleotides were selected. The sequences were aligned using Muscle. The location of insertion/deletion (indel) in the genome was mapped following the open reading frame (ORF) of Wuhan-Hu-1. Amino acid substitutions in all ORFs were analysed separately. Evolutionary analysis was inferred using maximum likelihood. There are two indel sites in ORF1AB, eight in spike, and one each in ORF3A, Matrix (MA), Nucleoprotein (NP), and the 3'-untranslated regions (3'UTR). Some indel sites are not unique, and some are variant specific. There are 10 residues shared by the Omicron, BA2, and GKA lineages. Three deletions in NP are unique to Omicron and BA.2; two substitutions in ORF1AB, four in spike, three in NP, and two in ORF7A and ORF8 were exclusive to the Delta and GPA clades. Delta, Omicron, and BA.2 share the same single residue spike. Three insertions in spike are unique for Omicron and GKA. Phylogenetic analysis shows that the Omicron, BA.2, and GKA clades share a common cluster that emerged from Delta. In conclusion, whole-genome comparison reveals indels and polymorphic amino acids specific to variants or sub-variants. We propose that the GKA clade is an Omicron subvariant. Finally, the higher transmissibility of BA.2 might be attributed to a 48-nucleotide deletion in the 3'UTR.

Introduction

The rapid evolution of severe acute respiratory syndrome 2 virus (SARS-CoV-2), the causative agent of the coronavirus 2019 (COVID-19) pandemic, requires immediate scientific clarification to better manage the current pandemic and to serve as a reference for the possible emergence of novel variants. Since it was discovered at the end of 2019 [1], the original virus has evolved into many variants. Established variants of concern (VOCs) are Alpha, Beta, Gamma, Delta, and Omicron; Lambda and Mu are variants of interest (VOIs), and GH/490R is a variant under monitoring (VUM). Some other clusters of viruses that are closely related to Omicron, the so-called BA.2 and GKA clades, which carry molecular markers of Delta and Omicron variants, require special attention. Originally, detection of the GKA clade, popularly known as Deltacron, led to arguments that it may be the result of sequencing error [2]. However, there are an increasing number of SARS-CoV-2 whole-genome sequences labelled as clade GKA with the notification "this submission requires investigation! It appears to contain markers of multiple lineages from both Delta and Omicron variants" in the database.

The capacity to spread globally as well other biological properties of each variant must be encoded, at least in part, by its genome. Whole-genome comparison should also be able to confirm the establishment of the BA.2 subvariant and GKA clade. Based on previous publications [3, 4], the whole genome organization of SARS-CoV-2 after the ORF annotation of Wuhan-Hu-1, and adding 5'- and 3'UTR and intergenic sequence (IGS), is 5'UTR-ORF1AB- IGS-Spike-IGS-ORF3A-ORF3B-IGS-Protein E-IGS-membrane (MA)-IGS-ORF6-ORF7A-ORF7B-ORF8-IGS-Nucleoprotein NP-ORF-10-3'UTR. Intergenic sequences (IGSs) have been identified previously [5–7].

Large submissions of whole genome sequences pose a major computational challenge, and some portions of submitted sequences contain a long track of non-definitive nucleotides. Here, we identify insertion/deletion (indel) and amino acid substitution patterns in the whole genome of representative variants, including the BA.2 subvariant and the GKA clade.

Materials And Methods

The sequence of the original SARS-CoV-2 strain Wuhan-Hu-1 (Accession Number NC_045512) was downloaded from GenBank. Ten to 20 complete sequences of each definitive variant as well as the subvariant BA.2 and GKA clade were selected randomly from GISAID and downloaded. Two representatives of each variant with no undetermined nucleotide of "N" or other IUPAC nucleotide codes were selected. Sequences with a single N or un-definitive nucleotide were accepted. The sequences were aligned using Muscle in MEGA-X software [8]. The locations of deletions/insertions in the genome of SARS-CoV-2 were mapped following the open reading frame of Wuhan-Hu-1, as available in the GenBank file. The fasta file of the final dataset is available in Supplementary Material 1. ORFs were analysed separately to determine the effects of mutations and deletions/insertions.

Using the corresponding ORF of the coding region of Wuhan-Hu-1, the first 15 nucleotides of the 5'-terminus were searched, and the sequence prior to the marked sequence was deleted. The last 15 nucleotides of Wuhan-Hu-1 were used. The selected sequences were translated into amino acid sequences and aligned using MEGA-X software. Using the same software, the data were exported in Mega format and analysed further for polymorphic amino acids. We identified amino acids that were consistently substituted from Wuhan-Hu-1 across all variants and amino acids shared by the Omicron, BA2, and GKA lineages, Omicron and BA.2, Delta and GKA, Delta, Omicron, BA.2, and GKA, as well as Omicron and GKA. Evolutionary history was inferred by using the maximum likelihood method and the Kimura 2-parameter model [9]. Initial tree(s) for the heuristic searching were obtained automatically by applying the neighbour-joining and BioNJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach and then selecting the topology with a superior log likelihood value. Evolutionary analyses were conducted in MEGA X [8]. The final fasta file of the data set is available in Supplementary Material 1.

Results

The indel pattern and its location in the whole genome of representatives of various variants of SARS-CoV-2 are presented in Table 1. There are two indel sites in ORF1AB, eight in spike, and one each in ORF3A, MA, NP, and the 3'UTR. No indel occurs in IGS. Some indel sites are not unique, as they

occur in more than one variant. D21605-21613 is unique to Omicron and DA.2 variants, and D21965-21967 is unique to the Alpha variant; I21968-21971 and D26143-26146 are unique to the Mu variant, D28351-28359 is unique to the Omicron and DA.2 variants; and D29723-29748 is unique to the DA.2 variant. Some indels occur in one representative of the variant.

All polymorphic amino acids of all proteins of two representatives of each variant of SARS-CoV-2 are listed in Supplementary Material 2. A summary of unique amino acids across entire genes in at least one of the representative strains of SARS-CoV-2 variants is presented in Table 2. Amino acids consistently substituted from Wuhan-Hu-1 across all variants are ORF1AB P4715L/F and spike D618G. Regarding amino acids shared by Omicron, BA2, and GKA, there are 10 in ORF1AB, 21 in spike, and one each in ORF3A, ORF6, and NP. Three deletions in NP are unique to Omicron and BA.2. Exclusive to Delta and GKA are two deletions in ORF1AB, four in spike, three in NP, and two in ORF7A and ORF8; spike amino acids shared by Delta, Omicron, BA.2, and GKA occur only once. Three insertions, Ins216E, Ins217P, and Ins218E, are unique to Omicron and GKA. GH/490 harbors 16 variants specific to Wuhan-Hu-1, namely, five in ORF1AB, eight in spike, two in NP, and one in ORF3A.

To assess the rapid evolution of all variants, evolutionary analysis of the whole genome of SARS-CoV-2 variants based on the maximum likelihood method is presented in Fig. 1. The results show that Omicron, BA.2, and GKA share a common cluster emerging from the Delta variant sequence with boostrap support of 98%.

Discussion

The genetic diversity of coronaviruses occurs through mutation and recombination, as has been described for SARS-CoV-2 [10]. Although the RNA-dependent RNA polymerases of coronaviruses possess proof-reading capacity [11], the virus still undergo mutation, which might lead to amino acid replacement. Such changes impact the biology of the virus as well as the clinical manifestation of its infection. Recombination involves viral RNA merging with other RNAs, either its own RNA, the RNA of other viruses, or cellular RNA; thus, template switching occurs during transcription [12]. This process leads to RNA indels. Mutations in SARS-CoV-2 prior to the emergence of variants have been reported [5]. In HIV, the deletions occurred by at least three different mechanisms: (i) misalignment of the growing point; (ii) incorrect synthesis and termination in the primer-binding sequence during synthesis of the plus-strand strong-stop DNA; and (iii) incorrect synthesis and termination before the primer-binding sequence during synthesis of the plus-strand strong-stop DNA [13].

Previous whole-genome comparisons have been conducted, including for Omicron [14]. However, that work focused on phylogeny and did not cover the recently identified BA.2 and GPA lineages, which are colloquially known as Deltacron. Indels and amino acid substitutions unique to specific variants were not described.

Through random selection of variant representatives with definitive sequences across the genome, we managed to identify unique patterns of indels and amino acid substitutions. Even with only two representatives for each variant, we identified quasispecies or, in the case of a variant, quasivariants. Viral quasispecies refers to a population structure that consists of extremely large numbers of variant genomes, termed mutant spectra, mutant swarms or mutant clouds [15]. For SARS-CoV-2, this phenomenon has been discovered even in single infected individuals [16–20]. We propose the term quasivariant, as many indels and amino acid substitutions occur in one of only two representatives. We believe that we will find more variation if we analyse more variant representatives. There must have been no variant or clade annotation error in the GISAID database, as phylogenetic analysis (Fig. 1) shows each sequence cluster with its own variant partner.

Amino acids consistently substituted from Wuhan-Hu-1 across all variants are ORF1AB P4715L/F and spike D618G. The D618G has been covered in previous works [5, 21–28]. ORF1AB P4715L/F has also been described [29, 30]. A database-wide survey is needed to understand the frequency of those substitutions.

The variant that harbours the most variant-specific substitution to Wuhan-Hu-1 is VUM GH/490. Both representatives show five, eight, two, and one amino acid substitution(s) in ORF1AB, spike, NP, and ORF3A, respectively. This VUM is being tracked in Europe, Africa, Asia, and America; however, the genome frequency for access of GISAID dated March 30, 2022, is lower than 0.3%.

The GKA clade does not comprise a unique variant. It harbours no unique indel or substitution compared with Wuhan-Hu-1, but it does share 34 amino acid replacements with Wuhan-Hu-1 with Omicron and BA2, 13 with Delta, one in spike with Delta, Omicron, and BA.2. Three insertions in spike in one representative of Ins216E, Ins217P, and Ins218E of the GKA clade are shared with Omicron. The molecular signatures of Delta and Omicron are obvious in the GKA clade. It is plausible that the GKA clade is an Omicron subvariant. The Delta signatures are understandable, as phylogenetically, the Omicron, BA.2, and GKA clades emerged from Delta (Fig. 1) with a high boostrap value. We suggest that the clade is not the result of sequencing error, as previously thought [2].

Interestingly, we identified truncated ORF3A in the Mu variant. Deletion of four nucleotides generates a stop codon; thus, ORF3A in this variant is 257 amino acids in length, whereas the others are 275 residues long. This accessory protein contributes to the pathogenesis of SARS-CoV-2 by inducing pathological apoptosis [31]. The effect of the Mu variant at the cellular level has not yet been described. One paper on this variant covered the neutralization effect of antibodies [32]. According to the GISAID database accessed on March 30, 2022, this variant has been identified in many countries, with a maximum global genome frequency of less than 1%, which has declined recently.

BA.2 differs from Omicron in the deletion of 48 nucleotides from the 3'UTR. The 3'UTR of coronaviruses contains all of the *cis*-acting sequences necessary for viral replication and binds to cellular as well as the viral components nsp1 and N proteins [33], which are required for minus-strand RNA synthesis [34]. This has also been described in SARS-CoV-2, whereby the 3'UTR is involved in genomic dimerization and interacts with cellular micro-RNA [35]. BA.2 has recently increased in frequency in multiple regions of the world, suggesting that it has a selective advantage over Omicron [36–39]. The genome frequency of BA.2 has increased exponentially to 90% of total Omicron submission, as based on GISAID accessed on the above date. As the original strain of Wuhan-Hu-1 has a basic reproduction number (R0) of 2.4-3 [40], Delta has an R0 of 5 [41], and Omicron has an R0 of estimated to be higher than 10, or three times greater than Delta [42], BA.2 subvariant might has R0 of 15 or higher. The higher transmissibility of BA.2 might be attributed, at least in part, to the shorter 3'UTR, which results in higher speed of viral replication, which needs to be investigated further. However, because the coding region across the whole genome, particularly for the spike protein, of BA.2 is very close to that of Omicron, people who survived Omicron infection should be naturally protected against BA.2.

In conclusion, whole-genome comparison of representatives of all variants revealed indel patterns that are specific to SARS-CoV-2 variants or sub-variants. Polymorphic amino acid comparison across all coding regions also showed amino acid residues shared by specific groups of variants. Additionally, based on the findings, it is plausible that the GKA clade is an Omicron subvariant. Finally, the higher transmissibility of BA.2 might be due at least in part to the 48 nucleotides deletion in the 3'UTR, which results in higher speed of viral replication.

Declarations

Acknowledgements

This study was supported by Research, Technology and Higher Education (RISTEKDIKTI) of Indonesia through World Class Research. The English language of the manuscript has been edited by a global professional editing service.

Conflict of Interest

The authors have declared no conflict of interest.

References

1. Wang H, Li X, Li T, Zhang S, Wang L, Wu X, Liu J: The genetic sequence, origin, and diagnosis of SARS-CoV-2. Eur J Clin Microbiol Infect Dis 2020, 39(9):1629–1635; DOI 10.1007/s10096-020-03899-4.
2. Kreier F: Deltacron: the story of the variant that wasn't. Nature 2022, 602(7895):19; DOI 10.1038/d41586-022-00149-9.
3. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY *et al*: A new coronavirus associated with human respiratory disease in China. Nature 2020, 579(7798):265–269; DOI 10.1038/s41586-020-2008-3.
4. Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, Duan G: Virology, Epidemiology, Pathogenesis, and Control of COVID-19. Viruses 2020, 12(4); DOI 10.3390/v12040372.
5. Mercatelli D, Giorgi FM: Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front Microbiol 2020, 11:1800; DOI 10.3389/fmicb.2020.01800.
6. Rahimi A, Mirzazadeh A, Tavakolpour S: Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. Genomics 2021, 113(1 Pt 2):1221–1232; DOI 10.1016/j.ygeno.2020.09.059.
7. Jungreis I, Sealfon R, Kellis M: SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. Nat Commun 2021, 12(1):2642; DOI 10.1038/s41467-021-22905-7.
8. Kumar S, Stecher G, Li M, Knyaz C, Tamura K: MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 2018, 35(6):1547–1549; DOI 10.1093/molbev/msy096.
9. Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980, 16(2):111–120; DOI 10.1007/BF01731581.
10. Pollett S, Conte MA, Sanborn M, Jarman RG, Lidl GM, Modjarrad K, Maljkovic Berry I: A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic. Sci Rep 2021, 11(1):17365; DOI 10.1038/s41598-021-96626-8.
11. Robson F, Khan KS, Le TK, Paris C, Demirbag S, Barfuss P, Rocchi P, Ng WL: Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. Mol Cell 2020, 80(6):1136–1138; DOI 10.1016/j.molcel.2020.11.048.
12. Kim MJ, Kao C: Factors regulating template switch in vitro by viral RNA-dependent RNA polymerases: implications for RNA-RNA recombination. Proc Natl Acad Sci U S A 2001, 98(9):4972–4977; DOI 10.1073/pnas.081077198.
13. Pulsinelli GA, Temin HM: Characterization of large deletions occurring during a single round of retrovirus vector replication: novel deletion mechanism involving errors in strand transfer. J Virol 1991, 65(9):4786–4797; DOI 10.1128/JVI.65.9.4786-4797.1991.
14. Kandeel M, Mohamed MEM, Abd El-Lateef HM, Venugopala KN, El-Beltagi HS: Omicron variant genome evolution and phylogenetics. J Med Virol 2022, 94(4):1627–1632; DOI 10.1002/jmv.27515.
15. Domingo E, Perales C: Viral quasispecies. PLoS Genet 2019, 15(10):e1008271; DOI 10.1371/journal.pgen.1008271.

16. Sun F, Wang X, Tan S, Dan Y, Lu Y, Zhang J, Xu J, Tan Z, Xiang X, Zhou Y *et al*: SARS-CoV-2 Quasispecies Provides an Advantage Mutation Pool for the Epidemic Variants. Microbiol Spectr 2021, 9(1):e0026121; DOI 10.1128/Spectrum.00261 – 21.
17. Ghorbani A, Samarfard S, Ramezani A, Izadpanah K, Afsharifar A, Eskandari MH, Karbanowicz TP, Peters JR: Quasi-species nature and differential gene expression of severe acute respiratory syndrome coronavirus 2 and phylogenetic analysis of a novel Iranian strain. Infect Genet Evol 2020, 85:104556; DOI 10.1016/j.meegid.2020.104556.
18. Lau BT, Pavlichin D, Hooker AC, Almeda A, Shin G, Chen J, Sahoo MK, Huang C, Pinsky BA, Lee H *et al*: Profiling SARS-CoV-2 mutation fingerprints that range from the viral pangenome to individual infection quasispecies. *medRxiv* 2020; DOI 10.1101/2020.11.02.20224816.
19. Armero A, Berthet N, Avarre JC: Intra-Host Diversity of SARS-Cov-2 Should Not Be Neglected: Case of the State of Victoria, Australia. Viruses 2021, 13(1); DOI 10.3390/v13010133.
20. Jary A, Leducq V, Malet I, Marot S, Klement-Frutos E, Teyssou E, Soulie C, Abdi B, Wirden M, Pourcher V *et al*: Evolution of viral quasispecies during SARS-CoV-2 infection. *Clin Microbiol Infect* 2020, 26(11):1560 e1561-1560 e1564; DOI 10.1016/j.cmi.2020.07.032.
21. Andres C, Pinana M, Borras-Bermejo B, Gonzalez-Sanchez A, Garcia-Cehic D, Esperalba J, Rando A, Zules-Ona RG, Campos C, Codina MG *et al*: A year living with SARS-CoV-2: an epidemiological overview of viral lineage circulation by whole-genome sequencing in Barcelona city (Catalonia, Spain). Emerg Microbes Infect 2022, 11(1):172–181; DOI 10.1080/22221751.2021.2011617.
22. Leung K, Pei Y, Leung GM, Lam TT, Wu JT: Estimating the transmission advantage of the D614G mutant strain of SARS-CoV-2, December 2019 to June 2020. Euro Surveill 2021, 26(49); DOI 10.2807/1560–7917.ES.2021.26.49.2002005.
23. Cheng YW, Chao TL, Li CL, Wang SH, Kao HC, Tsai YM, Wang HY, Hsieh CL, Lin YY, Chen PJ *et al*: D614G Substitution of SARS-CoV-2 Spike Protein Increases Syncytium Formation and Virus Titer via Enhanced Furin-Mediated Spike Cleavage. *mBio* 2021, 12(4):e0058721; DOI 10.1128/mBio.00587 – 21.
24. Yang TJ, Yu PY, Chang YC, Hsu SD: D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation. J Biol Chem 2021, 297(4):101238; DOI 10.1016/j.jbc.2021.101238.
25. Bhattacharya M, Chatterjee S, Sharma AR, Agoramoorthy G, Chakraborty C: D614G mutation and SARS-CoV-2: impact on S-protein structure, function, infectivity, and immunity. Appl Microbiol Biotechnol 2021, 105(24):9035–9045; DOI 10.1007/s00253-021-11676-2.
26. Mat Yassim AS, Asras MFF, Gazali AM, Marcial-Coba MS, Zainulabid UA, Ahmad HF: COVID-19 outbreak in Malaysia: Decoding D614G mutation of SARS-CoV-2 virus isolated from an asymptomatic case in Pahang. Mater Today Proc 2022, 48:828–836; DOI 10.1016/j.matpr.2021.02.387.
27. Eaaswarkhanth M, Al Madhoun A, Al-Mulla F: Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? Int J Infect Dis 2020, 96:459–460; DOI 10.1016/j.ijid.2020.05.071.
28. Muttineni R, Kammili N, Bingi TC, Rao MR, Putty K, Dholaniya PS, Puli RK, Pakalapati S, S S, K S *et al*: Clinical and whole genome characterization of SARS-CoV-2 in India. PLoS One 2021, 16(2):e0246173; DOI 10.1371/journal.pone.0246173.
29. De Marco C, Marascio N, Veneziano C, Biamonte F, Trecarichi EM, Santamaria G, Leviyang S, Liberto MC, Mazzitelli M, Quirino A *et al*: Whole-genome analysis of SARS-CoV-2 in a 2020 infection cluster in a nursing home of Southern Italy. Infect Genet Evol 2022, 99:105253; DOI 10.1016/j.meegid.2022.105253.
30. Banerjee S, Seal S, Dey R, Mondal KK, Bhattacharjee P: Mutational spectra of SARS-CoV-2 orf1ab polyprotein and signature mutations in the United States of America. J Med Virol 2021, 93(3):1428–1435; DOI 10.1002/jmv.26417.
31. Ren Y, Shu T, Wu D, Mu J, Wang C, Huang M, Han Y, Zhang XY, Zhou W, Qiu Y *et al*: The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. Cell Mol Immunol 2020, 17(8):881–883; DOI 10.1038/s41423-020-0485-9.
32. Uriu K, Kimura I, Shirakawa K, Takaori-Kondo A, Nakada TA, Kaneda A, Nakagawa S, Sato K, Genotype to Phenotype Japan C: Neutralization of the SARS-CoV-2 Mu Variant by Convalescent and Vaccine Serum. N Engl J Med 2021, 385(25):2397–2399; DOI 10.1056/NEJMc2114706.
33. Yang D, Leibowitz JL: The structure and functions of coronavirus genomic 3' and 5' ends. Virus Res 2015, 206:120–133; DOI 10.1016/j.virusres.2015.02.025.
34. Lin YJ, Zhang X, Wu RC, Lai MM: The 3' untranslated region of coronavirus RNA is required for subgenomic mRNA transcription from a defective interfering RNA. J Virol 1996, 70(10):7236–7240; DOI 10.1128/JVI.70.10.7236-7240.1996.
35. Imperatore JA, Cunningham CL, Frye CJ, Pellegrene KA, Evanseck JD, Mihailescu M-R: Conserved Elements in the 3'-UTR of SARS-CoV-2: Involvement in Genomic Dimerization and Interactions with Cellular Micrornas. Biophys J 2021, 120(3):313a-313a; DOI 10.1016/j.bpj.2020.11.1982.
36. Yu J, Collier AY, Rowe M, Mardas F, Ventura JD, Wan H, Miller J, Powers O, Chung B, Siamatu M *et al*: Comparable Neutralization of the SARS-CoV-2 Omicron BA.1 and BA.2 Variants. *medRxiv* 2022; DOI 10.1101/2022.02.06.22270533.
37. Cheng VC, Ip JD, Chu AW, Tam AR, Chan WM, Abdullah SMU, Chan BP, Wong SC, Kwan MY, Chua GT *et al*: Rapid spread of SARS-CoV-2 Omicron subvariant BA.2 in a single-source community outbreak. Clin Infect Dis 2022; DOI 10.1093/cid/ciac203.
38. Mahase E: Omicron sub-lineage BA.2 may have "substantial growth advantage," UKHSA reports. BMJ 2022, 376:o263; DOI 10.1136/bmj.o263.
39. Desingu PA, Nagarajan K: Omicron BA.2 lineage spreads in clusters and is concentrated in Denmark. J Med Virol 2022; DOI 10.1002/jmv.27659.
40. D'Arienzo M, Coniglio A: Assessment of the SARS-CoV-2 basic reproduction number, R 0, based on the early phase of COVID-19 outbreak in Italy. Biosaf Health 2020, 2(2):57–59; DOI 10.1016/j.bsheal.2020.03.004.

41. Liu Y, Rocklov J: The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. J Travel Med 2021, 28(7); DOI 10.1093/jtm/taab124.

42. Ito K, Piantham C, Nishiura H: Relative instantaneous reproduction number of Omicron SARS-CoV-2 variant with respect to the Delta variant in Denmark. J Med Virol 2022, 94(5):2265–2268; DOI 10.1002/jmv.27560.

Tables

Table 1

Insertion/deletion pattern and its location at the whole genome of representatives of various variants of SARS-CoV-2

No	Position*	D6485-6487	D11260-11268	D21605-21613	D21959-21967	D21965-21967	I21968-21971	D22005-22010	D22170-22172	I22181-22189	D22266-22274	D26143-26146	D27055-27067	D28351-28359	D29723-29748
	Location at the genome**	ORF1AB	ORF1AB	Spike	Spike	Spike	Spike	Spike	Spike	Spike	Spike	ORF3A	MA	NP	3'UTR
	Virus Strain														
1	WH-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	A-1	-	+	-	-	+	-	-	-	-	-	-	-	-	-
3	A-2	-	+	-	-	+	-	-	-	-	-	-	-	-	-
4	B-1	-	+	-	-	-	-	-	-	-	+	-	-	-	-
5	B-2	-	+	-	-	-	-	-	-	-	-	-	-	-	-
6	G-1	-	+	-	-	-	-	-	-	-	-	-	-	-	-
7	G2	-	+	-	-	-	-	-	-	-	-	-	-	-	-
8	D-1	-	-	-	-	-	-	+	-	-	-	-	-	-	-
9	D2	-	-	-	-	-	-	-	-	-	-	-	+	-	-
10	L-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	L-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	M-1	-	-	-	-	-	+	-	-	-	-	+	-	-	-
13	M-2	-	-	-	-	-	+	-	-	-	-	+	-	-	-
14	O-1	-	+	+	-	-	-	-	-	-	-	-	-	+	-
15	O-2	+	+	-	+	-	-	-	+	+	-	-	-	+	-
16	BA2-1	-	+	+	-	-	-	-	+	+	-	-	-	+	+
17	BA2-2	-	+	+	-	-	-	-	-	-	-	-	-	+	+
18	GKA-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19	GKA-2	-	-	-	-	-	-	+	-	-	-	-	-	-	-
20	GH-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	GH-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-

N: No insertion/deletion; P: positive insertion/deletion; *Position based on aligned sequence start from nucleotide no. 1 of Wuhan-Hu-1; **Location in the genome v based on GenBank data of Wuhan-Hu-1; Virus strains and GISAID codes WH-1 is Wuhan Hu-1, A-1 is Alpha EPI ISL 1489174, A-2 is Alpha EPI ISL 1489173, B-1 is B EPI ISL 660190, B-2 is Beta EPI ISL 660613, G-1 is Gamma EPI ISL 811149, G-2 is Gamma EPI ISL 87219, D-1 is Delta EPI ISL 1544014, D-2 is Delta EPI ISL 18386 1 is Lambda EPI ISL 1111127, L-1 is Lambda EPI ISL 1111341, M-1 is Mu EPI ISL 2339799, M-2 is Mu EPI ISL 2339801, O-1 is Omicron EPI ISL 9010924, O-2 is Omicron EPI ISL 9042199, BA2-1 is BA2 LSPA-3B08C6A, BA2-2 is BA2 HSLL-3B08857, GKA-1 is EPI ISL 9021856, GKA-2 is HDF-IPP08027, GH-1 is GH/490 EPI ISL 6316493, GH-2 is GH/490 ISL 7702915.

Table 2

Unique amino acids across whole genes in at least one of the representing strains of various variants of SARS-CoV-2

Consistent substitution to Wuhan-Hu-1	Specific for Omicron, BA2, and GKA	Specific for Omicron and BA2	Specific for Delta and GKA	Specific to Delta, Omicron, BA.2, and GKA	Specific to Omicron and GKA	Unique to GH
ORF1AB (1) P4715L/F	ORF1AB (10) S135R, T842I, G1307S, L3027F, T3090I, L3201F, P3395H, R5716C, I5967V, T6594I		ORF1AB (2) G5063S, P5401L			ORF1AB (5) E452G, L2119F, A2355S, S3149F, R3164H
Spike (1) D618G	Spike (21) T19I, A27S, G142D, V214G, G343D, S375F/L, S377P, S379F, T380A, D409N, R412S, N444K, E488A, Q497R, Q502R, Y509H, N683K, N768K, D800Y, Q958H, N973K		Spike (4) T19R, E157G, F158del, R159del	Spike (1) T482K	Spike (3) Ins216E, Ins217P, Ins218E	Spike (8) P9L, E96Q, I211T, R350S, N398S, Y453N, F494R, D940H
	NP (1) S413R	NP (3) E31del, R32del, S33del	NP (3) D63G, R203M, D377Y			NP (2) D22Y, E378Q
	ORF3A (1) T223I		ORF7A (2) V82A, T120I			ORF3A (1) T32I
	ORF6 (1) D61L		ORF8 (2) D119del, F120del			

Numbers in parentheses were the number of unique amino acid(s) in corresponding gene;

Figures

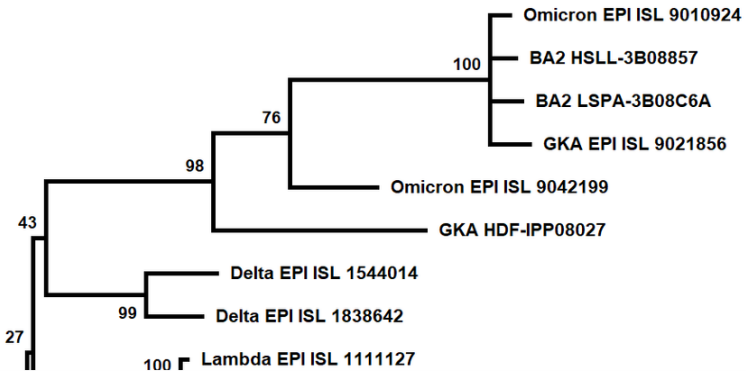

Figure 1

Evolutionary analysis of all established variants of SARS-CoV-2 including BA.2 and GKA subvariants by Maximum Likelihood method and Kimura 2-parameter model [9]. The tree with the highest log likelihood is shown. The phylogeny was tested using bootstrap method with 100 replicates. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. Evolutionary analyses were conducted in MEGA X [8].

Supplementary Files

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

- supplementarymaterial1.fas
- supplementarymaterial2.pdf