Heliyon 7 (2021) e08170

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Evaluation of global evolutionary variations in the early stage of SARS-CoV-2 pandemic

Sanghyun Lee [a,b,d,1], Chi-Hwan Choi [a,1], Mi-Ran Yun [a,1], Dae-Won Kim [a], Sung Soon Kim [c], Young Ki Choi [d,**], Young Sill Choi [a,*]

[a] Division of Pathogen Resource Management, Center for Public Vaccine Development and Support, National Institute of Infectious Diseases, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju-si, Republic of Korea
[b] Division of Bio Bigdata, Department of Precision Medicine, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju-si, Republic of Korea
[c] Center for Public Vaccine Development and Support, National Institute of Infectious Diseases, Korea National Institute of Health, Korea Disease Control and Prevention Agency, Cheongju-si, Republic of Korea
[d] Department of Microbiology, College of Medicine and Medical Research Institute, Chungbuk National University, Cheongju, Republic of Korea

ARTICLE INFO

Keywords:
COVID-19
SARS-CoV-2
Evolution
Genetic variation
Phylogeny

ABSTRACT

To understand the origin of variants and their evolutionary history in the early stage of the COVID-19 pandemic, time-scaled phylogenetic and gene variation analyses were performed. The mutation patterns and evolution characteristics were examined using the Bayesian Evolutionary Analysis Sampling Trees (BEAST) with 349 whole-genome sequences available by March 2020. The results revealed five phylogenetic clusters (Groups A–E), with 408 nucleotide variants. The mutations including the deletion of three nucleotides underwent various and complicated changes in the whole genome over time, while some frequency or transient mutations were also observed. Phylogenetic analysis demonstrated that SARS-CoV-2 originated from China and was transmitted to other Asian countries, followed by North America and Europe. This study could help to comprehensively understand the evolutionary characteristics of SARS-CoV-2 with a special emphasis on its global variation patterns.

1. Introduction

The *coronavirus* disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome *coronavirus* 2 (SARS-CoV-2), has rapidly spread worldwide after being first detected in December 2019 in Wuhan, Hubei, China, and has become a major public health concern globally (Lu et al., 2020; Bogoch et al., 2020). Previous studies have confirmed that this virus can spread from person to person, after identifying clusters of cases among families, including transmission from patients to healthcare workers (Chan et al., 2020). COVID-19 mainly causes respiratory illnesses, including cough, sputum production, dyspnoea, and haemoptysis, along with other symptoms such as fatigue, headache, diarrhoea, muscle pain, and lymphopenia (Rothan and Byrareddy, 2020). As of August 17, 2020, there have been 21,549,706 cases of SARS-CoV-2 confirmed worldwide, including 767,158 deaths (WHO, 2020).

SARS-CoV-2 has been showing dynamic transmission patterns during its spread by creating random mutations over time. The mutations retained after the virus's error correction machinery may help understand the origin and evolution of SARS-CoV-2 (Kupferschmidt and Cohen, 2020). With the rapidly increasing number of infections, the Global Initiative on Sharing All Influenza Data (GISAID) provided a platform for sharing SARS-CoV-2 sequences and their metadata (Shu and McCauley, 2017). Six major types, including S (C8782T, T28144C), L (C241, C3037, A23403, C8782, G11083, G25563, G26144, T28144, G28882), V (G11083T, G26144T), G (C241T, C3037T, A23403G), GH (C241T, C3037T, A23403G, G25563T), and GR (C241T, C3037T, A23403G, G28882A) have been designated by the GISAID for this virus. Currently, the latest update from GISAID on July 7, 2020, reported that the manifestation of these six types of SARS-CoV-2 has been recorded in five continents (GISAID, 2020a, b).

** Corresponding author.
* Corresponding author.
E-mail addresses: choiki55@chungbuk.ac.kr (Y.K. Choi), kcdc.yschoi83@gmail.com (Y.S. Choi).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2021.e08170
Received 24 September 2020; Received in revised form 16 January 2021; Accepted 8 October 2021

Bioinformatics analysis studies on COVID-19 have been carried out to gain better insight into its evolution and transmission, including cross-species transmission (Li et al., 2020a,b; Benvenuto et al., 2020; Liu et al., 2020). However, the studies were unable to clarify the evolutionary history of sequences using neighbour-joining trees and simple comparison of gene variants (Forster et al., 2020; Phan, 2020; Sánchez-Pacheco et al., 2020). Therefore, we analysed the evolutionary characteristics and variations in SARS-CoV-2 during the early days of the COVID-19 pandemic using the Bayesian Evolutionary Analysis by Sampling Trees (BEAST) method. These analyses should extend our understanding of the origins and evolutionary dynamics as well as subsequent transmission of the SARS-CoV-2 outbreak.

2. Materials and methods

2.1. Collation of SARS-CoV-2 whole-genome data sets

In early March 2020, 99 and 668 sequences were obtained from NCBI GenBank, a representative database of sequences, and GISAID, which stores information on multiple SARS-CoV-2 sequences, respectively (Supplementary Table S1; GISAID, 2020a, b). Among these 767 sequences, 418 sequences, containing non-standard nucleotides, derived from animals, consisting of partial (non-complete) sequences, or showing the same metadata were excluded. Therefore, we obtained a total of 349 unique target sequences for our further study.

2.2. Reconstruction of time-scaled phylogenies

BEAST (ver. 2.5.0) was used for phylogenetic analysis (Bouckaert et al., 2019). The Transition Model 2, Empirical base frequency type (TIM2+F + I) was chosen for the best fitting nucleotide substitution model by ModelFinder (ver. 2.0) based on its minimum Bayesian Information Criterion value. The best model was selected among six clock-tree model combinations with an appropriate effective sample size (ESS >100) by Path sampling and Stepping Stone of model-selection package of BEAST. For the model selection, we used log Bayes factors (BF) of 0 as a cutoff for binary classification of model (Baele et al., 2012). Then, Relaxed Clock Log Normal, Coalescent Exponential Tree, and a chain length of 2×10^8 with every 2×10^4 iterations with the highest value of log-scale marginal likelihood estimate were used for further analysis (Supplementary Table S2). To avoid that the single MCMC run may be stuck in a local optimum, two independent replicate runs using BEAST were performed, and then the runs were combined with the LogCombiner program. All resulted statistic values were similar (Supplementary Table S3).

The trees were summarized in a target tree by using the TreeAnnotator included in the BEAST package by choosing the tree with the maximum sum of posterior probabilities (maximum clade credibility) after a 10% burn-in. A sufficiency of 10% burn-in was inspected by ESS values using Tracer, and ESS for all statistic except TreeHeight (ESS = 128) showed >200 (Supplementary Table S4). Finally, the tree was visualized with FigTree (ver. 1.4). The mean time to the most recent common ancestor (tMRCA) and the 95% highest posterior density interval (95% HPDs) were calculated.

Groups were determined based on the topology of phylogenetic trees generated by BEAST. We named the branches at the time-scaled phylogenetic tree's root as Group A - E, and defined the descended branches as a numerical value (for example, Group A1). Subsequently, the groups were compared to the types classified by GISAID and PANGOLIN, which had previously shown the SARS-CoV-2 classification system (GISAID, 2020a, b; Rambaut et al., 2020). The GISAID classified the types depending on specific genetic mutations, while the PANGOLIN's nomenclature was made based on the generated maximum likelihood tree.

2.3. Gene variation analysis

For sequence analysis, multiple sequence alignment was performed using Multiple Alignment with Fast Fourier Transform (MAFFT; ver. 7), and variations were extracted using a self-developed Python programme in accordance with the reference genome (Wuhan-Hu-01; NC_045512, isolated on December 30, 2019). The scripts developed in Python are available through GitHub, at https://github.com/ivareve125/Variation (file name; VariatGenome_sars-cov-2.py).

3. Results

3.1. SARS-CoV-2 whole-genome data collection

The data set included 349 genomes from five continents, viz., Asia: China (n = 95), Republic of Korea (n = 13), Japan (n = 11), Singapore (n = 11), Hong Kong (n = 8), Taiwan (n = 6), India (n = 3), Thailand (n = 2), Nepal (n = 1), Cambodia (n = 1), and Vietnam (n = 1), Europe: the Netherlands (n = 46), Italy (n = 4), Portugal (n = 2), France (n = 23), Switzerland (n = 10), Germany (n = 6), Finland (n = 7), England (n = 6), Luxembourg (n = 1), Sweden (n = 1), Belgium (n = 1), and Ireland (n = 1), North America: United States of America (n = 67) and Canada (n = 2), South America: Chile (n = 2) and Brazil (n = 1); and Oceania: Australia (n = 17), with sampling dates between December 24, 2019 and March 11, 2020.

3.2. Genetic characteristics of SARS-CoV-2 variations

Among the genome sequences analysed herein, 408 nucleotide variations and 238 amino acid variations were observed. Regarding the individual genes, 262 variations were observed in ORF1ab, followed by 57 in S, 23 in ORF3a, 5 in E, 9 in M, 3 in ORF7a, 1 in ORF7b, 7 in ORF8, 35 in N, and 5 in ORF10 (Figure 1a). The **S gene** contains a binding region for host receptors, i.e. a receptor-binding domain (RBD), associated with infectivity (Shang et al., 2020; Lan et al., 2020), and 8 nucleotide variations and 7 amino acid variations were observed in this region (Figure 1b), with the A23403G (D614G; GISAID-G type, PANGOLIN-B.1 type) variant being observed in nucleotide sequences of 71 isolates. The **N gene** was primarily a target diagnosis region (Li et al., 2020a, b), and 3 nucleotide variations (G28881A; R203K, G28882A; R203K, G28883C; G204R) were observed in one Chilean isolate and 30 European isolates (Figure 1c). No variation was observed in **the ORF6 gene, and the C29543T** variation was observed in the non-coding region between genes N and ORF10.

3.3. Characteristics of variations by continent

Isolates from Asia (152 strains) harboured 192 nucleotide variations and 125 amino acid variations. Regarding the individual genes, 119 variations in ORF1ab, 25 in S, 15 in ORF3a, 2 in E, 7 in M, 3 in ORF7a, 6 in ORF8, 15 in N, and 2 in ORF10 were observed, whereas no variations were observed in ORF6 and ORF7b genes. **Isolates from Europe** (108 strains) harboured 119 nucleotide variations and 71 amino acid variations. Regarding the individual genes, 69 variations in ORF1ab, 19 in S, 7 in ORF3a, 1 in E, 2 in M, 1 in ORF7b, 1 in ORF8, 15 in N, and 2 in ORF10 were observed, whereas no variations were observed in ORF6 and ORF7a genes. **Isolates from the United States** (72 isolates) harboured 89 nucleotide mutations and 52 amino acid mutations. Regarding the individual genes, 57 variations in ORF1ab, 10 in S, 3 in ORF3a, 2 in E, 1 in M, 3 in ORF8, 11 in N, and 1 in ORF10 were observed, whereas no variations were observed in ORF6, ORF7a, and ORF7b. No unique nucleotide or amino acid variations were found in the isolates from Brazil, Chile, and Canada. **Isolates from Oceania** (17 isolates) harboured 25 nucleotide variations and 15 amino acid variations. Regarding the individual genes,

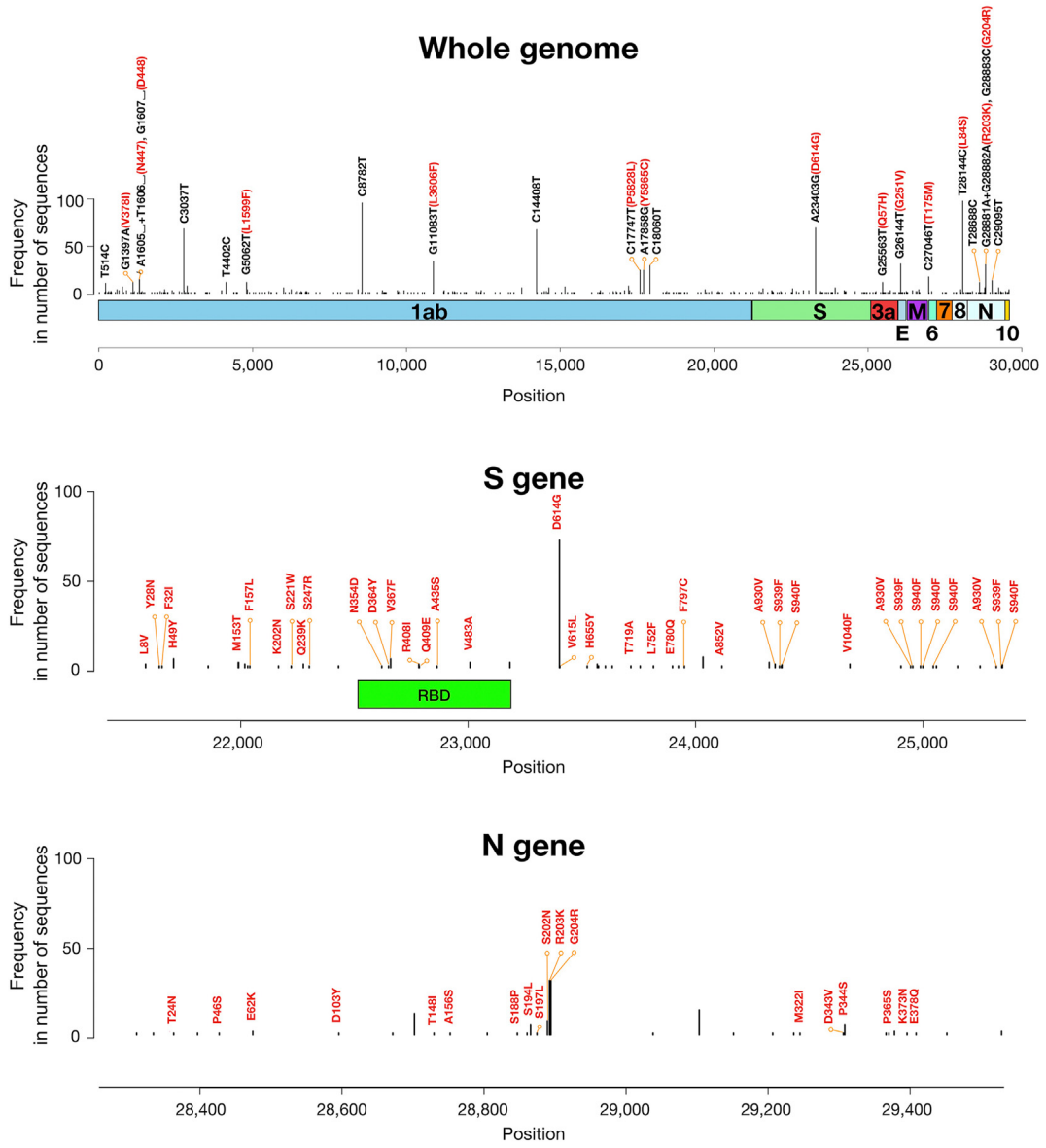

Figure 1. Distribution of variations in the entire genome, S gene, and N gene. a) Distribution of genome-wide variations: 408 nucleotide variations and 238 amino acid variations were observed herein. A relatively large number of variations were observed in ORF1ab, S, and N genes. b) Distribution of S gene variations: eight nucleotide variations and seven amino acid variations were observed in the receptor-binding domain (RBD), which is associated with infectivity. A23403G (D614G), the GISAID-G variant, was observed in 71 strains. c) Distribution of N gene variations: three consecutive nucleotide variations (G28881A; R203K, G28882A; R203K, and G28883C; G204R) were observed in 31 strains.

11 variations in ORF1ab, 5 in S, 2 in ORF3a, 1 in ORF8, 5 in N, and 1 in ORF10 were observed, whereas no variations were observed in five genes (E, M, ORF6, ORF7a, and ORF7b).

3.4. Global evolutionary pattern via a phylogenetic tree and gene variation analysis

On analysing evolutionary patterns using a phylogenetic tree (Figure 2, Supplementary Figure S1) and a variation table (Supplementary Table S5), **Group A** was the ancestor group of all 349 isolates, and it likely evolved from two Chinese isolates (A1), isolated on December 24 (EPI_ISL_402123) and 26 (EPI_ISL_406798). Since it was observed in China, Group A seems to have evolved separately in Canada and the United States (A1→A2) and then in Japan (A1→A3). In this study, we found the Group A samples by early February.

Group B was classified into several groups, which were evolved from the original Chinese isolate (B1) to Europe and Asia (B2 and B3) and Europe, Asia, and Oceania (B4 and B5). Group B2 contains an isolate as the reference genome as well as 10 isolates having the same sequence as that of the reference genome (shown as ⊖ in Supplementary Figure S1). In group B4, 7 isolates from Europe and America containing both variants, b-① and b-③, were clustered to the GISAID-V type, which was later found by July 2020. In group B5, after the b-① variant, the b-⑦ and b-⑧ variants were detected in Australia, Europe, and the Americas, and subsequently, the b-⑨ and b-⑩ variants were detected in Australia.

Group C evolved from China (A) and branched into groups that evolved generally into Asia, America, and Europe (C1) and those that evolved exclusively within Europe (C2 and C3). Group C1 contains 24 isolates having the same sequence as that of the reference genome (expressed as ⊖ in Supplementary Figure S1). In addition, in the c-⑤ variant, deletions of three contiguous bases (A1605-,T1606-,G1607-; ORF1ab gene) were found in isolates from the Netherlands in early March (indicated by dotted box and Del. in Figure 1). Groups C2 and C3, clustered to the GISAID-G (c-①, c-②, and c-③ variants)/PANGOLIN B.1,

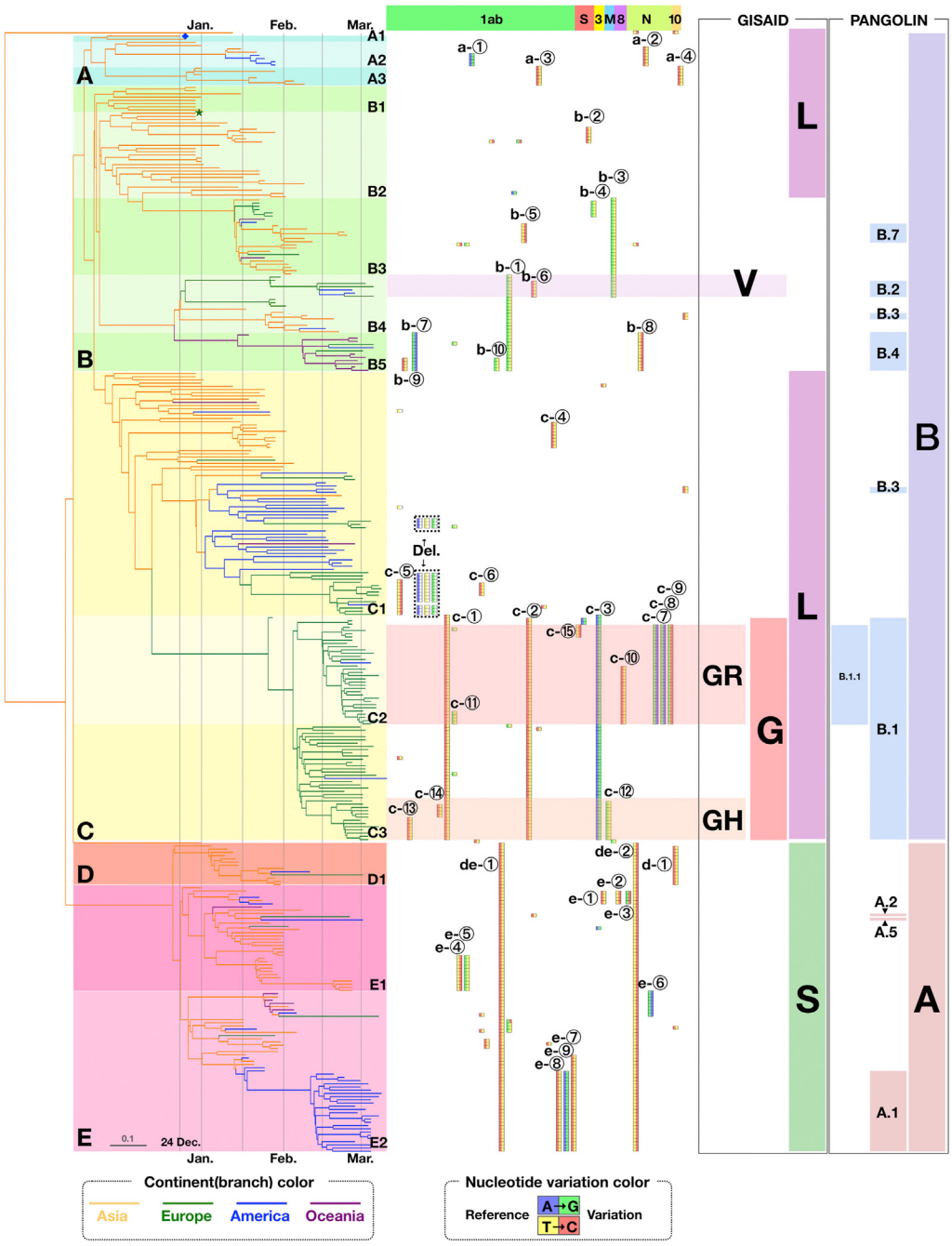

Figure 2. Phylogenetic tree (prepared using Bayesian Evolutionary Analysis Sampling Trees) and variations in the SARS-CoV-2 genome during the early stages of the COVID-19 pandemic. The groups are classified on the basis of the branch points of the phylogenetic tree (A–E), and variant groups are indicated by circled numbers. Each variation due to nucleotide substitutions is indicated with different colours. We applied the PANGOLIN classification and GISAID classification criteria herein. The GISAID classified the types depending on the existence of specific genetic mutations, while the PANGOLIN nomenclature in PANGOLIN was made based on the maximum likelihood tree-generated. The results of variation analysis are only expressed when the average number of variations per nucleotide exceeds 3.6. The colour of branches indicates each continent. A root sequence (occurred at December 24, 2019) is indicated by blue diamond.

started from a strain isolated in Germany (EPI_ISL_406862) at the end of January and later evolved to the GISAID-GR (c-⑦, c-⑧, and c-⑨)/PANGOLIN B.1.1 and GISAID-GH (c-⑫) types.

Group D and E were similar to GISAID-S/PANGOLIN A, containing de-① and de-②, compared with Group A, B, and C, and seemed to have evolved from the Chinese strain isolated on January 5, 2020 (EPI_ISL_406801).

Group D, a Group E ancestor, evolved from China to the United States, Germany, and Japan and subsequently evolved independently by branching to E1 and E2.

Group E branched from China to other Asian countries, including the isolate from Korea (E1), and subsequently independently branched to Oceania and America (E2). In particular, starting with one isolate (EPI_ISL_404895) obtained from the United States on January 19, 2020, isolates from United States generated a subgroup with common variants, e-⑧ and e-⑨.

The evolution rate of the 349 isolates, determined using BEAST, was 1.062×10^{-3} substitutions/site/year (95% HPD interval, 8.207×10^{-4}–1.334×10^{-3}), and tMRCA was October 19, 2019 (95% HPD, August 15, 2019–December 6, 2019).

3.5. Characteristics of variations by time

Concerning time, the evolution and variation patterns can be classified into three phases depending on when the major variations occurred. In the first phase from the end of December 2019 to early January 2020, internal transmission was observed in China, and L and S type variations were observed. In the second phase from the end of January to early February 2020, the V type mutation was observed in Europe and the Americas. In the third phase from the end of February to early March

2020, beginning with Germany, the G, GR, and GH type variations were observed in Europe and the Americas.

4. Discussion

In this study, we analysed the mutation patterns and evolution characteristics using time-scaled phylogenies with 349 SARS-CoV-2 whole-genome sequences obtained from GISAID and GenBank until March 11, 2020. The phylogenetic tree constructed in the present study showed five main clusters (Groups A–E) and 14 sub-clusters. The A, B, and C groups generated by our analysis were involved in GISAID-L, V, and G, and PANGOLIN-B types, while D and E groups were associated with GISAID-S and PANGOLIN-A types. Group C2 and C3 were related to GISAID-G, in which GR and GH types were located in Group 2 and 3, respectively. This proves that our method was able to reliably classify the isolates based on the genetic mutations.

The estimated evolutionary rate in the current study (1.062×10^{-3}) is similar when compared to other previously reported rates of SARS-CoV-2 (1.16×10^{-3}) as well as other coronaviruses, MERS-CoV (1.12×10^{-3}) and SARS-CoV ($0.80–2.38 \times 10^{-3}$) (Taiaroa et al., 2020; Cotten et al., 2014; Zhao et al., 2020). This value could be varied if the analysis is performed with longer time windows, since the estimated evolutionary rate showed time-dependent pattern in a previous study (Ghafari et al., 2020). Moreover, it seems that all SARS-CoV-2 mutation types in the most recent report of GISAID-hCoV-19 Analysis (reference, July 7, 2020, updated version) were already present in the early stage of the COVID-19 pandemic (until the beginning of March 2020), indicating that those mutations used as marker variants in GISAID nomenclature had already originated from the beginning of the COVID-19 pandemic. In particular, common mutations were detected in GISAID-G and GISAID-GR types. In GISAID-G type, c-② mutation was found beyond the known common mutations, c-① and c-③ mutations. Furthermore, c-⑦ and c-⑧ mutations were simultaneously observed with well-known c-⑨ mutation in GISAID-GR.

Interestingly, 34 whole-genome sequences identical to the reference sequence were found in two different groups, viz., Group B2 (10 isolates; located within close distance from the reference sequence) and Group C1 (24 isolates; later evolved to GISAID-G). Group B2 and C1 evolved to GISAID-V and GISAID-GH and GR, respectively, indicating that genetically identical SARS-CoV-2 could evolve to viral strains having different genetic characteristics and located in different groups in the future.

Moreover, specific deletions or mutations were detected in certain countries as shown in Figures 1 and 2. The deletion of three continuous nucleotides (A1605, T1606, and G1607) was found only in the Netherlands (15 isolates) (Phan, 2020). To determine whether the mutations and deletions affect the characteristics of SARS-CoV-2 in terms of morbidity and mortality, further studies with recent isolates and clinical information are necessary. Furthermore, the sequencing data in early pandemic of COVID-19 were deposited from countries that were able to prepare samples to be sequenced and had available sequencing machines during the COVID-19 pandemic's onset. Therefore, the geographic distribution of sequence samples in this study was not proportionally identical to the COVIDS-19's known number of cases. In addition, after our sample collection date, a number of sequence samples were added in the database, which also could affect the results of genetic variations in the early stage of COVID-19. To reach more precise insight regarding COVID-19's global spread, transmission and genetic variation characteristics, we need further studies with a geographically proportional number of sequenced samples and recently collected samples from the database. This larger amount of samples will allow us to evaluate the virus's incidence better. For the future study, we also need to consider filtering and masking alignments of SARS-CoV-2 sequence to avoid the oddities in genome sequences caused by contamination, recurrent sequencing errors, or hyper mutability for the reliable analysis (De Maio et al., 2020).

In summary, this study emphasizes the importance of time-scaled phylogenetic analysis to provide insights into the time of origin, genetic diversity, and transmission dynamics of COVID-19. Such phylogenetic research could directly influence public health in terms of adoption of preventive measures to reduce virus transmission in real-time. This study provides basic information for tracking future variations and transmission rates, and insights on COVID-19's genome variation and will improve the disease's diagnosis development, vaccines, and effective therapeutic pharmaceutics for its treatment.

5. Conclusion

This study analysed global variations and evolutionary patterns through the analysis of SARS-CoV-2 whole genomes at the onset of the COVID-19 pandemic. Notwithstanding the limitation regarding non-uniformity of the distribution of the sampling region of the analysed full-length nucleotide sequence, this study potentially provides evidence regarding the variations and evolutionary patterns of SARS-CoV-2 during the early days of the COVID-19 pandemic.

Declarations

Author contribution statement

Sanghyun Lee: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Chi-Hwan Choi, Mi-Ran Yun: Performed the experiments; Analyzed and interpreted the data.

Dae-Won Kim: Conceived and designed the experiments.

Sung Soon Kim, Young Ki Choi: Contributed reagents, materials, analysis tools or data.

Young Sill Choi: Analyzed and interpreted the data.

Funding statement

This work was supported by the Korea National Institutes of Health, the Korea Disease Control and Prevention Agency, Republic of Korea under Grant 4837-301.

Data availability statement

Data included in article/supp. material/referenced in article.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e08170.

Acknowledgements

We gratefully acknowledge the Authors, the Originating and Submitting Laboratories for their sequence and metadata shared through GISAID and NCBI, on which this research is based. All submitters of data may be contacted directly via www.gisaid.org. The Acknowledgments Table for GISAID and NCBI is reported as Supplementary material (Supplementary Table S1).

References

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A., 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29, 2157–2167.

Benvenuto, D., Giovanetti, M., Salemi, M., Prosperi, M., De Flora, C., Junior Alcantara, L.C., Angeletti, S., Ciccozzi, M., 2020. The global spread of 2019-nCoV: a molecular evolutionary analysis. Pathog. Glob. Health 114, 64–67.

Bogoch, I.I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M.U.G., Khan, K., 2020. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. J. Trav. Med. 27.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15, e1006650.

Chan, J.F.W., Yuan, S., Kok, K.H., To, K.K.W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C.-Y., Poon, R.W.S., Tsoi, H.-W., Lo, S.K.-F., Chan, K.H., Poon, V.K.M., Chan, W.M., Ip, J.D., Cai, J.P., Cheng, V.C.C., Chen, H., Hui, C.K.M., Yuen, K.Y., 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395, 514–523.

Cotten, M., Watson, S.J., Zumla, A.I., Makhdoom, H.Q., Palser, A.L., Ong, S.H., Al Rabeeah, A.A., Alhakeem, R.F., Assiri, A., Al-Tawfiq, J.A., Albarrak, A., Barry, M., Shibl, A., Alrabiah, F.A., Hajjar, S., Balkhy, H.H., Flemban, H., Rambaut, A., Kellam, P., Memish, Z.A., 2014. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. mBio 5 (1), e01062, 13.

De Maio, N., Walker, C., Borges, R., Weilguny, L., Slodkowicz, G., Goldman, N., 2020. Issues with SARS-CoV-2 Sequencing Data. https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473.

Forster, P., Forster, L., Renfrew, C., Forster, M., 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc. Natl. Acad. Sci. Unit. States Am. 117, 9241–9243.

Ghafari, M., du Plessis, L., Pybus, O., Katzourakis, A., 2020. Time Dependence of SARS-CoV-2 Substitution Rates. https://virological.org/t/time-dependence-of-sars-cov-2-substitution-rates/542.

Global Initiative for Sharing All Influenza Data (GISAID), 2020. https://www.gisaid.org/. (Accessed 17 March 2020).

Global Initiative for Sharing All Influenza Data (GISAID), 2020b. Full Genome Tree Derived from All Outbreak Sequences 2020-07-07. https://www.gisaid.org/. (Accessed 1 July 2020).

Kupferschmidt, K., Cohen, J., 2020. Race to find COVID-19 treatments accelerates. Science 367, 1412–1413.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., Wang, X., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220.

Li, X., Geng, M., Peng, Y., Meng, L., Lu, S., 2020a. Molecular immune pathogenesis and diagnosis of COVID-19. J. Pharmaceut. Anal. 10, 102–108.

Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B.T., Chaillon, A., 2020b. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J. Med. Virol. 92, 602–611.

Liu, S., Shen, J., Yang, L., Hu, C., Wan, J., 2020. Distinct Genetic Spectrums and Evolution Patterns of SARS-CoV-2. medRxiv.

Lu, H., Stratton, C.W., Tang, Y., 2020. Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. J. Med. Virol. 92, 401–402.

Phan, T., 2020. Genetic diversity and evolution of SARS-CoV-2. Infect. Genet. Evol. 81, 104260.

Rambaut, A., Holmes, E., Hill, V., O'Toole, Á., McCrone, J., Ruis, C., Plessis, L., Pybus, O., 2020. A Dynamic Nomenclature Proposal for SARS-CoV-2 to Assist Genomic Epidemiology. bioRxiv.

Rothan, H.A., Byrareddy, S.N., 2020. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J. Autoimmun. 109, 102433.

Sánchez-Pacheco, S.J., Kong, S., Pulido-Santacruz, P., Murphy, R.W., Kubatko, L., 2020. Median-joining network analysis of SARS-CoV-2 genomes is neither phylogenetic nor evolutionary. Proc. Natl. Acad. Sci. Unit. States Am. 117, 12518–12519.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., Li, F., 2020. Structural basis of receptor recognition by SARS-CoV-2. Nature 581, 221–224.

Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data – from vision to reality. Euro Surveill. 22.

Taiaroa, G., Rawlinson, D., Featherstone, L., Pitt, M., Caly, L., Druce, J., Purcell, D., Harty, L., Tran, T., Roberts, J., Scott, N., Catton, M., Williamson, D., Coin, L., Duchene, S., 2020. Direct RNA Sequencing and Early Evolution of SARS-CoV-2. bioRxiv.

World Health Organization (WHO), 2020. WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/. (Accessed 17 August 2020).

Zhao, J., zhai, X., Zhou, J., 2020. Snapshot of the Evolution and Mutation Patterns of SARS-CoV-2. bioRxiv.