Heliyon 10 (2024) e30222

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Molecular evolutionary characteristics of severe acute respiratory syndrome coronavirus 2 and the relatedness of epidemiological and socio-environmental factors

Kyuyoung Shim [a,b], Eun-Ha Hwang [a], Green Kim [a], Young Min Woo [a,b], You Jung An [a], Seung Ho Baek [a], Taehwan Oh [a], Yujin Kim [a], Kiwon Jang [c], Jung Joo Hong [a,b], Bon-Sang Koo [a,b,*]

[a] National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju, Republic of Korea
[b] Department of Biomolecular Science, KRIBB School of Bioscience, Korea University of Science and Technology, Daejeon, Republic of Korea
[c] Korea Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea

ARTICLE INFO

Keywords:
SARS-CoV-2
Bayesian evolutionary analysis
Sequence analysis
Socio-environmental factors
Epidemiological triad

ABSTRACT

After the first outbreak, SARS-CoV-2 infection continues to occur due to the emergence of new variants. There is limited information available on the comparative evaluation of evolutionary characteristics of SARS-CoV-2 among different countries over time, and its relatedness to epidemiological and socio-environmental factors within those countries. We assessed comparative Bayesian evolutionary characteristics for SARS-CoV-2 in eight countries from 2020 to 2022 using BEAST version 2.6.7. Additionally, the relatedness between virus evolution factors and both epidemiological and socio-environmental factors was analyzed using Pearson's correlation coefficient. The estimated substitution rates in the gene encoding S protein of SARS-CoV-2 exhibited a continuous increase from 2020 to 2022 and were divided into two distinct groups in 2022 (p value < 0.05). Effective population size (*Ne*) generally showed decreased patterns by time. Notably, the change rates of the substitution rates were negatively correlated with the cumulative vaccination rates in 2021. A strict and rapid vaccination policy in the United Arab Emirates dramatically reduced the evolution of the virus, compared to other countries. Also, the average yearly temperature in countries were negatively correlated with the substitution rates. The changes of six epitopes in SARS-CoV-2 were related to various socio-environmental factors. We figured out comparative virus evolutionary traits and the association of epidemiological and socio-environmental factors especially cumulative vaccination rates and average temperature.

1. Introduction

After the first outbreak in China in 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been one of the most prevalent and devastating respiratory pathogens. Despite broad vaccination and strict biosecurity measures, these diseases have continued to emerge by new variants. These variants of SARS-CoV-2 have been continuously reported worldwide [1]. Some variants with biological significance have been designated as Variants of Concerns (VOCs) by the World Health Organization, namely alpha,

* Corresponding author. 30, Yeongudanji-ro, Ochang-eup, Cheongwon-gu, Cheongju-si, Chungcheongbuk-do, 28116, Republic of Korea
E-mail address: porco9@kribb.re.kr (B.-S. Koo).

https://doi.org/10.1016/j.heliyon.2024.e30222
Received 18 August 2023; Received in revised form 19 April 2024; Accepted 22 April 2024
Available online 28 April 2024

beta, gamma, delta, and omicron. VOCs exhibit distinct phenotypes in terms of antigenicity, virulence, transmissibility and tissue tropisms [2]. The disease severity of SARS-CoV-2 infection varied from asymptomatic to mild and severe illness, and factors affect the disease severity such as age, immune status and other diseases [3]. Various non-respiratory symptoms were reported including gastrointestinal, neurological, reproductive and cardiovascular diseases [4].

Among virus families, single-stranded, positive sense RNA viruses, including SARS-CoV-2, generally exhibit higher substitution rates in glycoproteins [5]. The estimated substitution rates of the gene encoding S protein in SARS-CoV-2 were approximately 10^{-3} and 10^{-4} substitutions per site per year, which were in the range of other RNA viruses [6]. Mutations in the gene encoding the S protein can alter the virus phenotypes and influence antigenicity, virulence, transmissibility and tissue tropisms [7]. Mutations within the epitopes of each VOC have been continuously identified in S and other viral proteins [8]. The receptor binding domain (RBD) in S protein directly interacts with human cells via angiotensin-converting enzyme 2 (ACE2) receptors [9]. Therefore, S protein and its RBD were the target regions of commercial SARS-CoV-2 vaccines and therapeutics [10]. SARS-CoV-2 VOCs achieved vaccine escape against commercial SARS-CoV-2 vaccine, causing large outbreaks in vaccinated populations [11]. Also, therapeutic efficacy of FDA-approved monoclonal antibodies diminished against Omicron variants [12]. ACE2 receptors are expressed in a variety of human tissues including the respiratory tract, gastrointestinal tract, kidney, heart, testis, lymphoid organs, brain and urinary tract [13]. Altered affinity of SARS-CoV-2 variants to ACE2 receptors in organs can alter tissue tropism and related clinical signs [14].

Virus mutation is affected by several factors including polymerase fidelity, host immune pressure, viral transmission mode, replication frequencies and duration of infection [15–17]. Viruses originated in one species can cross species barriers and infect another species, which is called a host jump. Several bat origin coronaviruses have been linked to pandemics in humans including human coronavirus NL63 and 229E, SARS-CoV, Middle East Respiratory Syndrome Coronavirus and SARS-CoV-2 [17]. After the host jump, the virus faces different environments including host receptors, immune responses and intracellular factors [18]. In general, the rapid mutation especially in S protein of SARS-CoV-2 occurred for adaptation to new host for increased fitness optimization expressed as the basic reproduction number (R0) [19]. After the host jump, a higher proportion of beneficial mutations were identified under lower adapted to the host niche, which could alter viral phenotypes. After adaptation to new host niche, mutation types are neutral and deleterious mutations, not beneficial mutations [20].

The interactions between epidemiological triads including host, pathogen and environments affect the dynamics of viral outbreaks and spread [21]. To the best of our knowledge, there has been limited information on a comparative evaluation of evolutionary characteristics of SARS-CoV-2 (pathogen) among countries by time and the relatedness with epidemiological factors (host) and socio-environmental factors of countries (environment). In this report, eight countries with different vaccination rates and socio-environmental conditions were evaluated for comparative virus evolutionary traits of SARS-CoV-2 by country and time and the relatedness with various factors (Fig. 1).

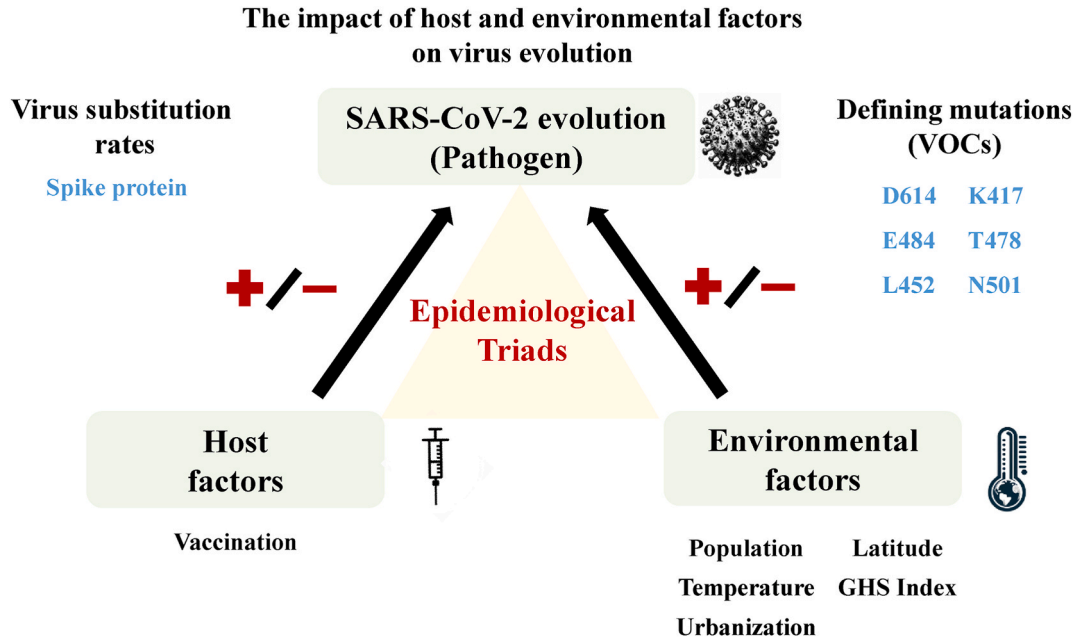

Fig. 1. The description of the impact of host and environmental factors on the evolutionary properties of viruses. The epidemiological triad, consisting of the host, pathogen, and environment, interacts dynamically to influence the outbreak and spread of viruses. This report highlights the significance of host and environmental factors in understanding the evolution of viruses.

2. Materials and methods

2.1. Virus sequences

Eight countries were selected based on the criteria of vaccination rates in 2021 provided by Our World in Data and availability of SARS-CoV-2 sequence data from 2020 to 2022 for Bayesian evolutionary analysis [22]. Five countries had high vaccination rates above 59 %: United Arab Emirates (UAE), Portugal, United Kingdom (UK), India and Italy. The three countries with low vaccination rates below 30 % were Kenya, Egypt and South Africa. Virus sequences data with known collection dates and country were retrieved by year and country from the Global Initiative for Sharing All Influenza Data platform (GISAID) from 2020 to 2022 (Table 1) [23]. Multiple alignments were performed based on the gene encoding the S protein of SARS-CoV-2 and reference strain (Wuhan-Hu-1) in CLC Genomics Workbench 20.0.4. (CLC Bio, Qiagen, Hilden, Germany). After then, sequences with homologous sequences (100 %) and ambiguous symbols (Ns or IUPAC codes) were excluded for further evaluation (Table 1). If more than 100 sequences remained after filtration, the 100 sequences were finally selected based on dates and a similarity criterion of more than 99 %.

2.2. Bayesian evolutionary analysis

All sequence alignment sets were imported in BEAUti program and analyzed using the Bayesian Markov Chain Monte Carlo (MCMC) method in BEAST version 2.6.7 [24]. The best fit substitution models were determined based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) using modeling test software in CLC genomics workbench. A relaxed uncorrelated lognormal molecular clock was performed for each alignment data set with Bayesian skyline coalescent prior. The procedures were run in duplicate for 60,000,000 generations with sampling per each 6000 steps. The duplicated log and tree files were combined using LogCombiner except for 10 % burn-ins per each run. The log results with over 200 values of the effective sample size (ESS) were calculated and visualized using Tracer (version 1.7.2.), and the effective population size (*Ne*) were calculated using Bayesian skygrid reconstruction in Tracer (version 1.7.2.).

2.3. Virus evolutionary, epidemiological and socio-environmental factors

Virus evolutionary factors include substitution rates per site per year in the S gene, annual change rates of the substitution rates and relative composition rates of amino acid in each epitope. Epidemiological factors were cases per tests, cases per million and cumulative vaccination rates. Socio-environmental factors were population size, population density, urbanization rates, latitude, average yearly temperature and global health index. These data were sourced from Our World in Data (https://ourworldindata.org/) [25], Global Health Security Index (GHS) Index (https://www.ghsindex.org/), World Population Prospects 2022 (https://population.un.org/wpp/) and the world factbook (https://www.cia.gov/the-world-factbook/) and Climate Change Knowledge Portal (https://climateknowledge-portal.worldbank.org/). Data sources are detailed in Supplement Table 1.

2.4. Mutation in VOCs

Six epitopes were selected for mutation analysis at position D614, N501, K417, E484, L452, and T478 of S protein of SARS-CoV-2 that were closely related to alpha, beta and delta variants [26]. These variants were designated as VOCs by the World Health Organization [27]. Nucleotide and amino acid sequence alignments for each country were conducted annually from 2020 to 2022. The relative composition rates of prior amino acid sequences were calculated for each position, and their correlation with other factors was assessed (Fig. 1). The numbering of amino acid sequences was performed based on the reference sequence (GenBank accession no. MN908902).

Table 1
Numbers of SARS-CoV-2 sequences categorized by country and year.

Countries	2020		2021		2022	
	Retrieved sequence numbers[a]	Final sequence numbers[b]	Retrieved sequence numbers	Final sequence numbers	Retrieved sequence numbers	Final sequence numbers
United Arab Emirates	152	99	135	100[c]	295	74
Portugal	136	99	182	100	377	100
United Kingdom	179	47	267	90	233	57
India	155	90	239	88	243	63
Italy	209	99	201	83	221	73
Kenya	147	93	328	99	328	62
Egypt	166	75	255	99	199	100
South Africa	143	86	401	58	315	75

[a] Sequences without ambiguous letters (Ns and IAPUC codes) were retrieved from GISAID database.
[b] Homologous sequences were excluded for the Bayesian evolutionary analysis.
[c] The final selection of 100 sequences was determined based on their collection dates and a criterion that required a similarity exceeding 99 %.

2.5. Statistical analysis

Significant statistical differences between the means of two groups were determined using *t*-test with independent samples. A p value less than 0.05 was considered statistically significant. Correlation analysis between virus evolution and other factors was calculated using Pearson correlation coefficient. The final values are expressed as r values, and those below 0.05 considered statistically significant. All statistical analysis was performed using GraphPad Prism software, version 8.4.3 (Graphpad Software, Boston, MA, USA).

3. Results

3.1. Bayesian evolutionary analysis

The substitution rates per site per year in the gene encoding S protein of SARS-CoV-2 were successfully determined and represented greater than 200 values of effective sample size (ESS). The ESS of a parameter sampled via MCMC represents the number of effectively independent draws from the posterior distribution equivalent to those produced by the Markov chain. This estimate is utilized to evaluate the statistical reliability of parameter estimates derived from MCMC sampling. Generally, the substitution rates per site per year continuously increased in countries from 2020 to 2022 (Fig. 2A, Supplement Table 2). In 2020, the substitution rates per site per year were clustered among countries. In 2021, the overall substitutions rates significantly increased compared with those of 2020, except for the UAE. After 2021, the substitution rates showed higher values and exhibited a broader range among different countries in 2022. Virus substitution rates were $9.9 \times 10^{-4} \pm 0.00055$ per site per year in 2020, $2.0 \times 10^{-3} \pm 0.00094$ per site per year in 2021 and $3.7 \times 10^{-3} \pm 0.00186$ substitution per site per year in 2022 (expressed as average $\pm$ standard deviation). In 2022, the substitution rates in India, Italy and Kenya were significantly higher compared with other countries (p value < 0.05) (Fig. 2B). In the UAE, the substitution rates were significantly lower compared with other countries in 2021 and 2022. The effective population size (*Ne*) in each country was determined by year using Bayesian skyline plot in BEAST 2 program. The overall *Ne* values showed a tendency to decrease by year, although significant differences and fluctuation patterns were identified according to country and year (Fig. 3A–H).

3.2. The relatedness between virus evolutionary values and the values of epidemiological and socio-environmental factors

Correlation analysis was performed between virus evolution values (virus substitution rates per site per year, increased rates of virus substitution rates compared with the previous year) and the values of epidemiological and socio-environmental factors (cumulative vaccination rates, SARS-CoV-2 cases per million people, population size, population density, urbanization rates, global health index, latitude and average temperature) using Pearson correlation coefficients (Supplement Table 3). A correlation between substitution rates and factors in 2020 was not identified. In 2021, increased substitution rates per site per year were negatively correlated

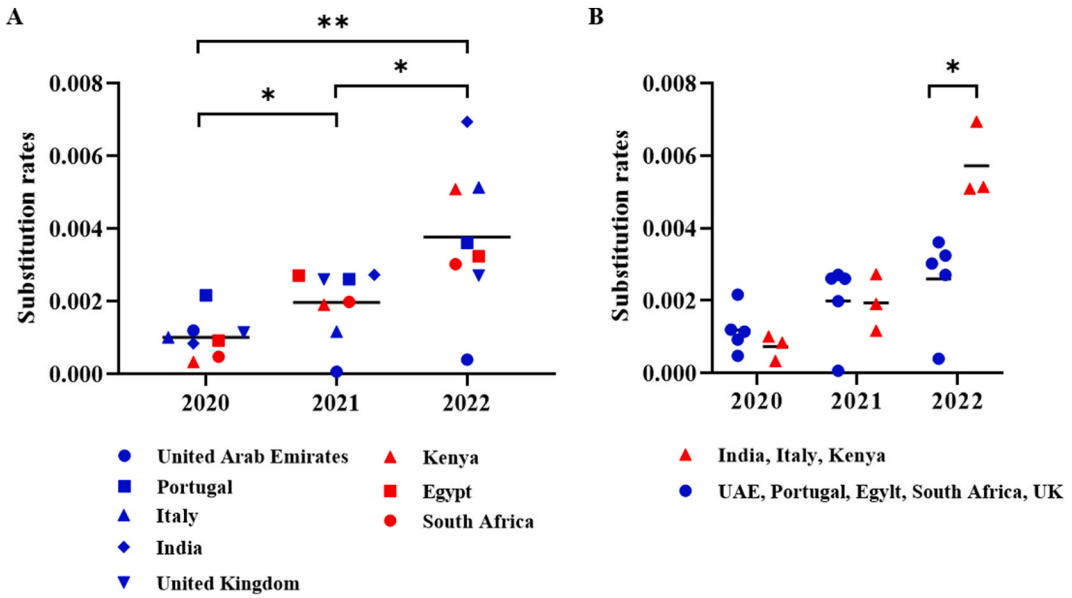

Fig. 2. The substitution rates per site per year in the gene encoding spike protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The substitution rates by year from 2020 to 2022 (A), Differences in the substitution rates between two groups of countries by year (B). (A) Blue letter indicate five countries with high vaccination rates above 59 %, while red letters highlight three countries with low vaccination rates below 30 %. (B) Blue letter represent five countries with low substitution rates in 2022, and red letters indicate three countries with high substitution rates. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

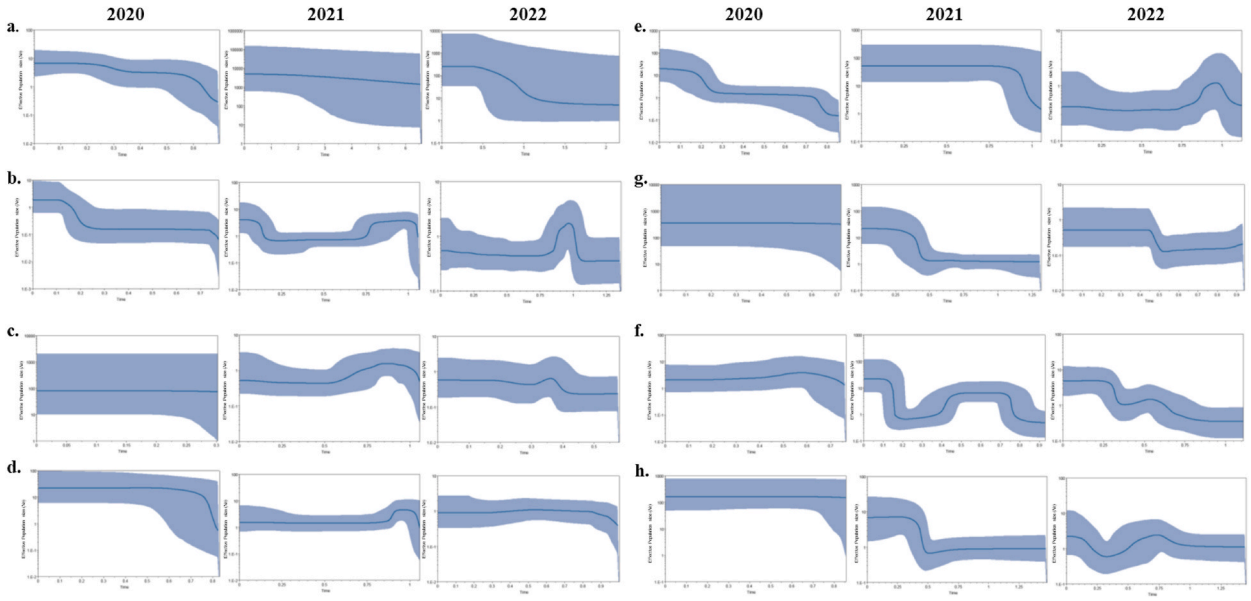

Fig. 3. The effective population sizes (*Ne*) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in countries by year from 2020 to 2022. United Arab Emirates (A), Portugal (B), United Kingdom (C), India (D), Italy (E), Kenya (F), Egypt (G), South Africa (H). The *x*-axis indicates the time in year units, and the *y*-axis represents *Ne* values. The parenthesis indicate 95 % high probability density interval.

with cumulative vaccination rates (p value = 0.01) (Fig. 4A). Additionally, the average yearly temperature in countries was negatively correlated with the substitution rates per site per year (p value = 0.01) (Fig. 4B). Other factors including cases per millions were not correlated to virus evolutionary values.

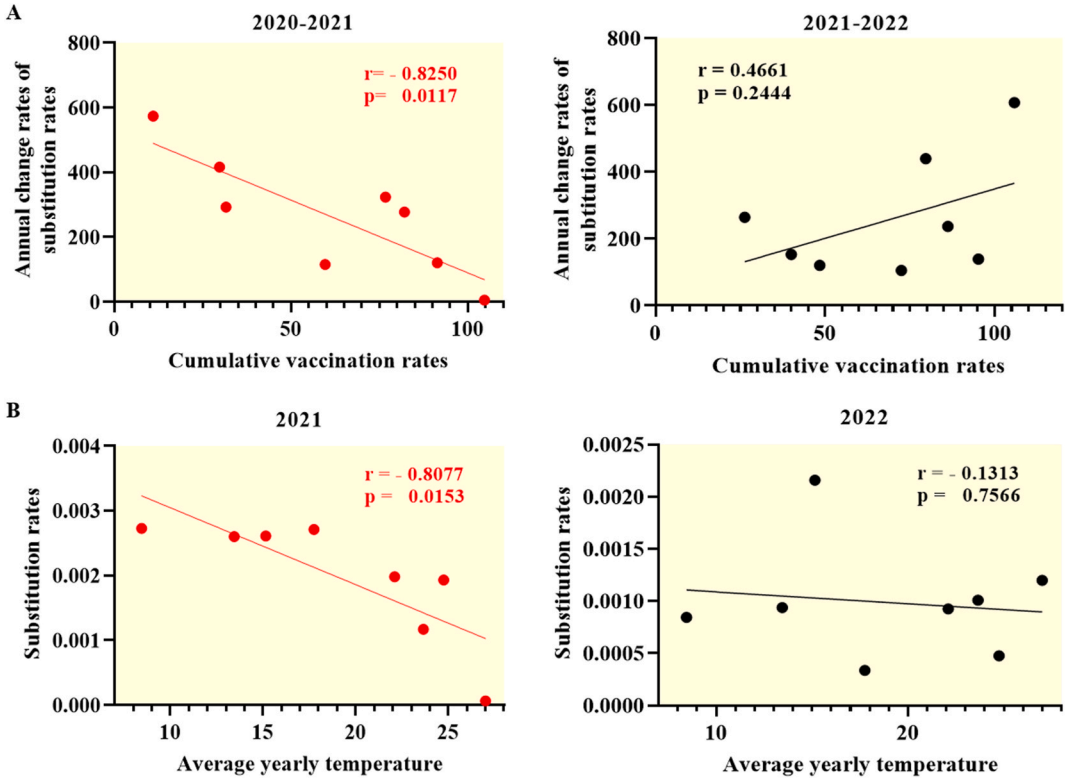

Fig. 4. The Pearson correlation coefficient (r) and p-value (p) between the virus evolutionary factors and epidemiological and socio-environmental factors. The correlation between virus substitution rates and cumulative vaccination rates (A), and average yearly temperature (B).

3.3. The correlation between mutation in VOCs and virus evolutionary, epidemiological and socio-environmental factors

The relative composition rates of amino acid sequences in six amino acid epitopes are shown in Supplement Table 4**.** The correlation between the relative composition rates in six epitopes in VOCs, vaccination and socio-environmental factors were statistically determined by year using Pearson correlation coefficient. Amino acid changes in various epitopes were found to be associated with one or more socio-environmental factors during the years 2020, 2021 and 2022 (Fig. 5A–C).

4. Discussion

New variants of SARS-CoV-2 viruses continue to emerge due to rapid virus evolution, leading to differences in antigenicity, virulence and tissue tropism. For this reason, SARS-CoV-2 has not been effectively controlled, despite strict biosecurity measures and vaccination policies. In this report, we performed Bayesian evolutionary analysis for the gene encoding the S protein in SARS-CoV-2 from eight countries and evaluated the relatedness among evolutionary traits, epidemiological and socio-environmental factors.

The substitution rates per site per year in the full length S protein of SARS-CoV-2 were successfully determined and were over 200 ESS; we compared the rates among eight countries by year. The overall tendency of virus substitution rates tended to continuously increase in counties from 2020 to 2022 except for UAE in 2021 (Fig. 2). After a host jump, viruses need rapidly adapt to new environments in the new host species for increased fitness of the host niche [19]. Therefore, SARS-CoV-2 from bats would experience rapid evolution rates in human beings for increased virus fitness. Over time, the viral evolutionary rates of DNA and RNA viruses generally decrease by time and approach those of their hosts [19]. Therefore, considering the continuous increased tendency of virus evolutionary rates observed in this study, SARS-CoV-2 viruses are in the process of adapting human beings. In 2022, the variation in virus substitution rates across at least eight countries broadened compared to the previous years, with some countries exhibiting more rapid evolutionary characteristics (Fig. 2). The differentiation between two groups could not be statistically substantiated in terms of their correlation with the values of epidemiological and socio-environmental factors. The interactions among numerous factors including immune pressure, transmission dynamics and environmental factors could affect the virus evolutionary differentiation between countries over time [16,28]. This phenomenon warrants further evaluation over time and across a broader range of countries. The *Ne* values showed a tendency to decrease by year in all eight countries (Fig. 3A–H). In general, a smaller *Ne* promotes a higher rate of genetic drift and lesser potential for maintaining genetic diversity [29]. This finding coincides with the continuous increased evolutionary rates by year.

The correlation between virus evolution properties and both epidemiological and socio-environmental factors was successfully determined in this report. Evolutionary rates that increased in 2021 compared to 2020 were found to be negatively correlated with the cumulative vaccination rates in 2021 (p value = 0.01) (Fig. 4A). After the initiation of the first SARS-CoV-2 vaccination in UK in December 2020, various vaccination policies were implemented in different countries in 2021. Vaccination enhances the host's immune response against SARS-CoV-2, which can reduce virus replication and consequently influence virus evolution within the host [5]. However, the relationship between the absolute values of the substitution rates and cumulative vaccination rates were not statistically established in 2021 and 2022. Increased rates of the substitution rates in 2021 may be primarily influenced by cumulative vaccination rates without other factors. However, the absolute substitution rates themselves could be simultaneously affected by both epidemiological and socio-environmental factors. Therefore, only the increase in substitution rates could be statistically validated. The virus substitution rates per site per year for SARS-CoV-2 in the UAE were significantly lower than those in other countries in 2021 and 2022, as shown in Fig. 2**.** This difference was not observed in 2020 before the vaccination policy was adopted worldwide. In the UAE, the cumulative vaccination rates exceeded 100 % in 2021, reflecting the rapid and strict vaccination policy that was applied to nearly the entire population, including multiple doses for some individuals. Therefore, a stringent vaccination policy, with coverage exceeding the total population count, could effectively decrease virus evolution, regardless of the number of confirmed cases per million.

This report determined the correlation between epidemiological and socio-environmental factors and virus evolutionary properties, including the substitution rates and main epitopes of VOCs. The average yearly temperature in eight countries was found to be negatively correlated with the absolute substitution rates in S genes of SARS-CoV-2. Higher incidence rates of SARS-CoV-2 have been identified during winter seasons compared to other seasons, as reflected on the WHO COVID-19 dashboard. The increased incidence of SARS-CoV-2 in winter heightened virus transmission among people, which could have contributed to an increase in the substitution rates per site per year [30]. Six epitopes in VOCs were selected, which are antibody epitopes and have the potential to influence the ACE2 binding capacity [26]. Amino acid changes in various epitopes were found to be associated with one or more socio-environmental factors during the years 2020, 2021 and 2022 (Fig. 5A–C). Generally, the association between changes in epitopes and related factors has increased over time. Mutation in three epitopes (D614G, K417 and E484A) were correlated with disease incidence-related values such as cases per test and cases per million. Changes in some epitopes, including E484A, K417 N and T478, were association with social factors (urbanization and population size and density). Additionally, virus substitution related values were statistically correlated with changes of epitopes, such as K417 N, L452 and N501Y. Therefore, various factors influence changes in epitopes, with their effects varying according to the specific factors involved.

In this report, we determined the comparative virus evolutionary properties among countries and their correlation with epidemiological and socio-environmental factors. The virus evolutionary rates increased continuously from 2020 to 2022, displaying different patterns among countries in 2022. A significant finding is that vaccination policies can decrease virus evolution rates. Additionally, data showing the correlation between average temperature and both substitution rates and specific epitopes suggest that new variants with altered epitopes affecting phenotypes are more likely to emerge in cold countries and seasons. These insights into SARS-CoV-2 evolutionary traits and their interplay with various factors may assist in shaping global quarantine policies. Our analysis

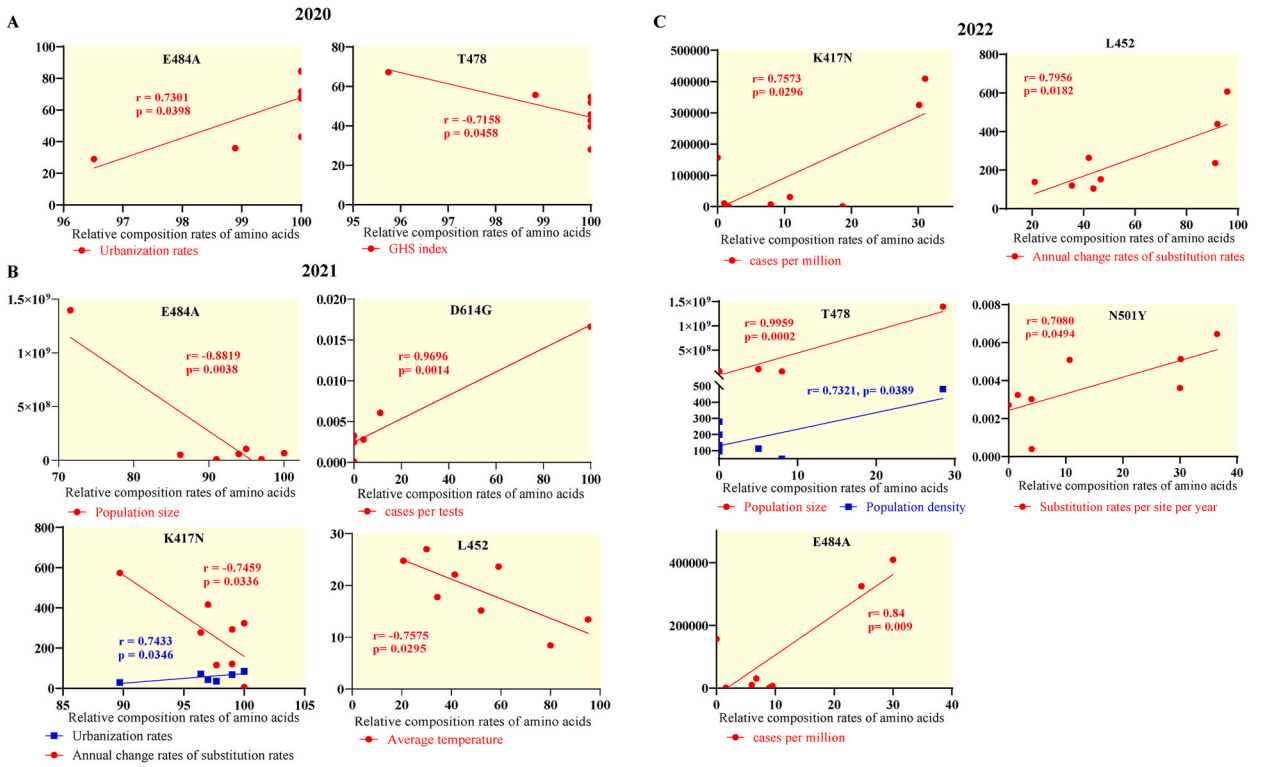

Fig. 5. The Pearson correlation coefficient (r) and p-value (p) between the relative composition rates of amino acids in epitopes of variants of concern (VOCs) and evolutionary, epidemiological and socio-environmental factors for years 2020 (A), 2021 (B), and 2022 (C).

of evolutionary properties was confined to eight countries, limiting our capacity to elucidate global phenomena. To overcome these constraints and gain a more comprehensive understanding, further research encompassing a broader range of countries is essential.

Funding

This work was supported by the National Research Foundation of Korea grant funded by the Korean government (NRF-2021M3H9A2098600).

Data availability statement

Most of the data are available in the main text and the supplementary information. The data from this study will be available upon reasonable request.

CRediT authorship contribution statement

Kyu Young Shim: Writing – original draft, Formal analysis. **Eun-Ha Hwang:** Methodology. **Green Kim:** Formal analysis, Data curation. **YoungMin Woo:** Validation. **You Jung An:** Project administration. **Seung Ho Baek:** Validation. **Taehwan Oh:** Validation. **Yujin Kim:** Validation. **Kiwon Jang:** Software. **Jung Joo Hong:** Supervision. **Bon-Sang Koo:** Writing – review & editing, Funding acquisition, Conceptualization.

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.

Acknowledgements

We extent our thanks to all the researchers at the National Primate Research Center who contributed to infrastructure related to infectious diseases.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e30222.

References

[1] B. Salzberger, F. Buder, B. Lampl, B. Ehrenstein, F. Hitzenbichler, T. Holzmann, B. Schmidt, F. Hanses, Epidemiology of SARS-CoV-2, Infection 49 (2) (2021 Apr) 233–239, https://doi.org/10.1007/s15010-020-01531-3.

[2] A.M. Carabelli, T.P. Peacock, L.G. Thorne, W.T. Harvey, J. Hughes, T.I. de Silva, et al., SARS-CoV-2 variant biology: immune escape, transmission and fitness, Nat. Rev. Microbiol. 21 (3) (2023) 162–177, https://doi.org/10.1038/s41579-022-00841-7.

[3] S. Samadizadeh, M. Masoudi, M. Rastegar, V. Salimi, M.B. Shahbaz, A. Tahamtan, COVID-19: why does disease severity vary among individuals? Respir. Med. 180 (2021) 106356 https://doi.org/10.1016/j.rmed.2021.106356.

[4] P. Guadarrama-Ortiz, J.A. Choreño-Parra, C.M. Sánchez-Martínez, F.J. Pacheco-Sanchez, A.I. Rodríguez-Nava, G. García-Quintero, Neurological aspects of SARS-CoV-2 infection: mechanisms and manifestations, Front. Neurol. 11 (2020) 1039, https://doi.org/10.3389/fneur.2020.01039.

[5] S. Duffy, L.A. Shackelton, E.C. Holmes, Rates of evolutionary change in viruses: patterns and determinants, Nat. Rev. Genet. 9 (4) (2008) 267–276, https://doi.org/10.1038/nrg2323.

[6] Q. Nie, X. Li, W. Chen, et al., Phylogenetic and phylodynamic analyses of SARS-CoV-2, Virus Res. 287 (2020) 198098, https://doi.org/10.1016/j.virusres.2020.198098.

[7] A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2) (2020) 281–292. e6, https://doi.org/10.1016/j.cell.2020.02.058.

[8] D. Tian, Y. Sun, J. Zhou, Q. Ye, The global epidemic of SARS-CoV-2 variants and their mutational immune escape, J. Med. Virol. 94 (3) (2022) 847–857, https://doi.org/10.1002/jmv.27376.

[9] J. Lan, J. Ge, J. Yu, et al., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, Nature 581 (7807) (2020) 215–220, https://doi.org/10.1038/s41586-020-2180-5.

[10] W. Tai, L. He, X. Zhang, 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 (6) (2020) 613–620, https://doi.org/10.1038/s41423-020-0400-4.

[11] S.W.X. Ong, T. Chia, B.E. Young, SARS-CoV-2 variants of concern and vaccine escape, from Alpha to Omicron and beyond, Expet Rev. Respir. Med. 16 (5) (2022) 499–502, https://doi.org/10.1080/17476348.2022.2057299.

[12] R. Uraki, M. Kiso, M. Imai, S. Yamayoshi, M. Ito, S. Fujisaki, et al., Therapeutic efficacy of monoclonal antibodies and antivirals against SARS-CoV-2 Omicron BA. 1 in Syrian hamsters, Nature Microbiology 7 (8) (2022) 1252–1258, https://doi.org/10.1038/s41564-022-01170-4.

[13] I. Hamming, W. Timens, M. Bulthuis, A. Lely, Gv Navis, H. van Goor, Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis, J. Pathol.: A Journal of the Pathological Society of Great Britain and Ireland 203 (2) (2004) 631–637, https://doi.org/10.1002/path.1570.

[14] J. Liu, Y. Li, Q. Liu, Q. Yao, X. Wang, H. Zhang, et al., SARS-CoV-2 cell tropism and multiorgan infection, Cell discovery 7 (1) (2021) 17, https://doi.org/10.1038/s41421-021-00249-2.

[15] V.N. Petrova, C.A. Russell, The evolution of seasonal influenza viruses, Nat. Rev. Microbiol. 16 (1) (2018) 47–60, https://doi.org/10.1038/nrmicro.2017.118.

[16] K. Hanada, Y. Suzuki, T. Gojobori, A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes, Mol. Biol. Evol. 21 (6) (2004) 1074–1080, https://doi.org/10.1093/molbev/msh109.

[17] K. Moelling, Within-host and between-host evolution in SARS-CoV-2—new variant's source, Viruses 13 (5) (2021) 751, https://doi.org/10.3390/v13050751.

[18] B. Longdon, M.A. Brockhurst, C.A. Russell, J.J. Welch, F.M. Jiggins, The evolution and genetics of virus host shifts, PLoS Pathog. 10 (11) (2014) e1004395, https://doi.org/10.1371/journal.ppat.1004395.

[19] P. Simmonds, P. Aiewsakun, A. Katzourakis, Prisoners of war—host adaptation and its constraints on virus evolution, Nat. Rev. Microbiol. 17 (5) (2019) 321–328, https://doi.org/10.1038/s41579-018-0120-2.

[20] S. Duffy, Why are RNA virus mutation rates so damn high? PLoS Biol. 16 (8) (2018) e3000003 https://doi.org/10.1371/journal.pbio.3000003.

[21] M. Cevik, S.D. Baral, Networks of SARS-CoV-2 transmission, Science 373 (6551) (2021) 162–163, https://doi.org/10.1126/science.abg0842.

[22] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, M. Roser, J. Hasell, C. Appel, et al., A global database of COVID-19 vaccinations, Nat. Human Behav. 5 (7) (2021) 947–953, https://doi.org/10.1038/s41562-021-01122-8.

[23] Y. Shu, J. McCauley, GISAID: global initiative on sharing all influenza data–from vision to reality, Euro Surveill. 22 (13) (2017) 30494, https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

[24] R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, et al., Beast 2: a software platform for Bayesian evolutionary analysis, PLoS Comput. Biol. 10 (4) (2014) e1003537, https://doi.org/10.1371/journal.pcbi.1003537.

[25] M.H. Cheng, J.M. Krieger, A. Banerjee, Y. Xiang, B. Kaynak, Y. Shi, et al., Impact of new variants on SARS-CoV-2 infectivity and neutralization: a molecular assessment of the alterations in the spike-host protein interactions, iScience 25 (3) (2022), https://doi.org/10.1016/j.isci.2022.103939.

[26] J.Y. Choi, D.M. Smith, SARS-CoV-2 variants of concern, Yonsei Med. J. 62 (11) (2021) 961, https://doi.org/10.3349/ymj.2021.62.11.961.

[27] S.D. Smith-Tsurkan, R.A. Herr, S. Khuder, C.O. Wilke, I.S. Novella, The role of environmental factors on the evolution of phenotypic diversity in vesicular stomatitis virus populations, J. Gen. Virol. 94 (Pt 4) (2013) 860, https://doi.org/10.1099/vir.0.048082-0.

[28] R. Sanjuán, P. Domingo-Calap, Genetic diversity and evolution of viral populations, Encyclopedia of virology 53 (2021), https://doi.org/10.1016/B978-0-12-809633-8.20958-8.

[29] S. Inaida, R.E. Paul, S. Matsuno, Viral transmissibility of SARS-CoV-2 accelerates in the winter, similarly to influenza epidemics, Am. J. Infect. Control 50 (9) (2022) 1070–1076, https://doi.org/10.1016/j.ajic.2022.05.009.