Journal of Umm Al-Qura University for Applied Sciences
https://doi.org/10.1007/s43994-025-00270-1

ORIGINAL ARTICLE

The influence of gene flow on the genetic diversity of SARS-CoV-2 and its variants (British, South African, and Brazilian)

Nabil Benazi[1] · Sabrina Bounab[2] · Mohammed Khodja[2] · Azzedine Melouki[2] · Hadia Benhalima[3] · Zahra Sadok[4]

Received: 5 February 2025 / Accepted: 29 August 2025

Abstract
The inference of the history of the genetic variability of the novel coronavirus and its variants is essentially based on the presence of gene flow between the viral populations of the different existing and ancestral variants. Our study allowed the detection of numerous gene flows between the endemic variant and a clade composed of two South African and Brazilian variant, using the AIM model part of the StarBeast2 package with MCMC algorithm sampling, we were able to co-infer the SARS_CoV2 tree and its variants alongside the evolutionary parameters of interest. Our study demonstrated a number of gene flows between geographically separated gene pools of SRAS_CoV2 and its mutations, resulting in competition between introduced viral lineages (British; South African; Brazilian) and their endemic ancestors for the sensitive human host. This process will have lasting effects on the structure of the viral population, including the extinction of the endemic lineage, followed by the British variant, and will give way mainly to the clade (South African; Brazilian) Hence the value of increased full genome monitoring of new variants in other isolated geographic areas in order to gain an understanding of the gene flow that occurs between these variants.

Keywords SARS-CoV-2 · Variants · Gene flow · Phylogenetic · AIM model

1 Introduction

RNA viruses such as riboviruses and retroviruses are characterized by a genetic information carrier of great plasticity and great adaptability to environmental variations because they use RNA and DNA in an alternative way of the host cell during the replication of their genomes, which generally gives them the advantage of having a large viral population, which can reach 10^{12} viral particles in an infected organism,, possessing both a fast rate of replication and a short generation period, which enables the production of approximately one hundred thousand copies on average of viral RNA per infectious particle in about 10 h [1]. These properties, with a particularly high rate of mutations are the basis of the extreme genetic variability of these RNA viruses [2]. Among the processes of this genetic variability are distinguished:

- Variability by mutations, which will result in substitutions, deletions or insertions [3], and in the absence of mechanisms for maintaining and/or restoring the integrity of genetic information carried by RNA, the frequency of RNA virus mutations, estimated after a single replication cycle, is of the order of 10^{-3} to 10^{-5} per nucleotide site [4, 5], whereas the frequency of mutations is a thousand to a million times lower for DNA genomes. As a result, some sites or regions of the RNA virus genome may correspond to unstable points of mutations while others will appear more stable. Such a higher mutation frequency results in an irreversible loss of information and leads to catastrophic error for the genome.
- Genetic variability by recombination, which involves the exchange of genetic information within genomic segments, as well as genetic variability by reassortment,

✉ Nabil Benazi
benmsila@hotmail.fr

1 Institut PASTEUR Algérie, Annexe M'sila 28,000 M'Sila, Algeria

2 Faculty of Sciences, University of M'sila, 28,000 M'sila, Algeria

3 Laboratory of Cellular Toxicology, Department of Biology, Faculty of Sciences, Badji Mokhtar University, Annaba, Algeria

4 Faculty of Science, University of Medea, Medea, Algeria

Published online: 25 September 2025

which involves the exchange of whole genomic segments in the case of viruses with segmented genome [6].

Although RNA recombination does not appear or rarely occurs for negative RNA viruses, on the other hand in the case of positive RNA viruses such as coronavirus, the frequency of recombinant events is particularly high and can be estimated at 1% for 1500 nucleotides [7]. Thus variants from recombinant events are frequently isolated in the case of coronavirus avian infectious bronchitis virus [8, 9]. Genetic variability by reassortment concerns much more the RNA-segmented viruses, namely Orthomyxoviridue, Bunyuviridue, Arenuviridue and Reoviridue, during this process, which occurs in the favor of a co-infection of the same host by different viruses, the genomic segments of the co-infecting viruses are randomly redistributed within the virions produced. It can occur between human viruses and viruses of animal origin such as influenza viruses including reassortment events between viruses of avian and human origin which are at the origin of the antigenic breaks which have resulted in the emergence of viruses of a new subtype responsible for major influenza pandemics [10]. This scenario cannot be excluded for the emergence of the SARS_CoV2 virus. This emergence process depends largely on gene flow between these virions produced. In order for there to be gene flow between two produced virions, it is necessary that these two produced virions share a common host. This situation is likely to be produced in the case for SARS_COV2 whose phylogenetic groups are congruent with geographic groups and not host species [4]. It is therefore possible to find several variants of SARS_COV2 infecting the same host. In order to study the way in which evolution occurs, by which new SARS_COV2 variants are formed from common ancestors (the endemic variant), we use the AIM (Approximate Isolation with Migration) model which allows joint inference of the tree of these variants with gene flow rates and effective population sizes.

2 Materials and method

2.1 Statement of ethics

In light of the fact that the sequences utilized in this investigation were obtained from the Los Alamos database NCBI-GenBank, and from the GISAID website, which do not contain a patient identifier. The use of a consent form is not necessary.

3 Collection of SARS_COV2 sequences and phylogenetic analysis

The presented work uses the whole genome sequences with high coverage, its length of over 29,000 base pairs. Metadata and annotations for SARS_CoV2 and its variants were obtained from the GISAID [11] and NCBI-GenBank [12] from date: 21jeullet 2020. For four variants of SARS_COV2, from 3 countries: 30 sequences of SARS_CoV2 B.1.1.7 lineage (and its variant 501Y.V1, or “British”); 30 sequences of SARS_CoV2 B.1.351 lineage (and its variant 501Y.V2, or “South African”); and 30 sequences of SARS_CoV2 P.1 lineage or SARS_CoV2 B.1.1.248 lineage (and its variant 501Y.V3, or “Brazilian”). And finally 30 sequences of the endemic SARS_CoV2 of British origin. All of these SARS-CoV2 viral sequences were split and picked individually by nation, sometimes manually, in order to optimize the length of the segment that was being studied. MEGA X v10.1 [13] was used to visualize the multiple alignments. MEGA X v10.1 was used also to perform a phylogenetic analysis using the maximum likelihood approach (1000 bootstrap replicates). The most suitable nucleotide substitution model was chosen according to the Akaike information criterion (AIC) and the GTR+4 model was carried out. Another phylogenetic inference was also performed; the program jModeltest [14] was utilized in order to determine the nucleotide substitution model that was the most appropriate, which resulted in the model HKY+4 with a strict clock of the dataset. The results were similar, but the tree generated with the HKY+4 model had better bootstrap values. The tree topologies were validated based on the results of the Shimodaira-Hasegawa (HS) tests, which were carried out using a total of one thousand bootstraps in FastTree [15]. Having a score that is higher than 0.98 indicates that you have a high level of confidence in the particular division of the tree or sub-tree.

4 Inference of history of SARS_CoV2 variants in the presence of gene flow

We will infer the history of three variants (British, Brazilian, South African) and endemic variant of SARS_CoV2, using the AIM software. The AIM is part of the StarBeast2 package [16]. Because we want to jointly infer the stories of variants and gene flows between these existing and ancestral variants (endemic variant). We ran the MCMC algorithm over 200 million generations for the 4 variants sampled every 5000 steps with a burn-in of 10%. The effective sample size (ESS) values for the estimates were mostly greater than 200, and the uncertainty in the estimates was represented by the probability density (HPD) values of 95%.

The HKY+4 nucleotide substitution model was used for all sequences to match the simulation conditions."To control for the potential confounding effects of temporal variation in the sampling times of the sequences, all analyses were conducted using a tip-dated approach. The collection date for each sequence was used to calibrate the molecular clock, allowing the model to distinguish between signals of shared ancestry and recent gene flow."Because of this, none of our estimations will be expressed in terms of time units; rather, they will be expressed in terms of the number of substitutions. The priors that are most crucial to specify are those that pertain to the number of active gene flow routes, the rates of gene flow, and the effective population size. The term"active gene flow route"refers to a route of gene flow between two variations that is not equal to zero. By default, a Poisson Prior with lambda equal to 0.693 is used to determine the prior on the number of active routes (migIndicatorSum.species) of gene flow. Due to this, almost fifty percent of the probability mass is assigned to 0 active gene flow paths. The conclusion that can be drawn from this is that the prior likelihood of having a gene flow is quite low when there is no information available on gene flow. We placed our tree at a height of approximately 0.02 substitutions per location in order to have a better understanding of migration rates. If we had a migration rate of one to two, which would equal fifty, then it would mean that one lineage of a gene from the present to the root would have to migrate an average of one time. The previous on the migration rates is established in the species block that is dedicated to migration rates. Based on the assumption that we will set the mean of the log normal distribution to 25, we may anticipate that around one out of every two lineages will experience a migration event over the whole species tree. This is a reasonable estimate of the number of migration events we can predict under this prior [16]."The significance of inferred gene flow events between variants was quantified using Bayes Factors (BF). The marginal likelihoods of models including and excluding each potential migration route were compared. Following Kass and Raftery (1995) [17], a Bayes Factor threshold of>100 was chosen as decisive evidence for the presence of gene flow, providing a conservative criterion to minimize false positives."

5 Results

5.1 Phylogenetic history between SARS_CoV2 variants from genetic sequences

We have built a phylogenetic tree, while using the maximum likelihood approach launches on all 130 complete sequences of SRAS-CoV-2 and its British South African and Brazilian variants available on GISAID as of January 4, 2021. The first observation about this phylogenetic tree of SRAS-CoV-2 and its variants, is the presence of a clade made only of the two variants South African and Brazilian, this monophyletic group is characterized by a rapid divergence, because this clade has the shortest lengths of branches (see Fig. 1), and if we compare it with the sister group (British variant) we see that this lineage is moderately

Fig. 1 Phylogenetic tree of the SARS-CoV-2 and its variants sequences. The Phylogenetic tree was built by an approximately maximum likelihood method on the full genomes of SARS-CoV-2 and its novel variants. The width of the edges indicates how confident we can be of the position estimation at each point on the tree to 98%. The four lineages have been classified into three groups: clade (South African-Brazil) sister group (England2) ancestral group (England1)

divergent, on the other hand the endemic lineage (SARS-CoV-2) represents the slowest divergence compared to the other variants.

6 Inference of histories of SARS_CoV2 variants in the presence of gene flow

To assess potential divergence events and predict the number of active gene flow pathways, gene flow rates and effective population size of SARS_CoV2 and its variants selected for this study, we reconstructed evolutionary trees according to the Bayesian model AIM [16], using full-length genomes of SRAS-COV-2 isolates from"england"with variants 501Y.V1, or"British", variant 501Y.V2, or"South African", variant 501Y.V3, or"Brazilian", from consecutive rows B.1.1.7; B.1.351; P.1 or B.1.1.248. Figure 1 shows the tree of inferred SARS_CoV2 species and its variants. We deduce that the order of the speciation event is consistent with the chronology of the appearance of its new variant according to the scientific platforms in particular GISAID and GenBank. Together with the speciation history of its variants, In addition, we have determined that there is a flow of genes between all of the species that are coexisting. At the intersection of two species that coexist, arrows represent the flow of genes. In order to illustrate the flow of genes between species, arrows are shown, and the posterior support must be at least 0.5. We found support for the unidirectional gene flow between the endemic variation SARS_COV2 symbolises"ENGLAND1"and the common ancestor of all other variants in all sub-variants random sets, as indicated by the chronology of speciation of its new variant with the endemic variant (see Fig. 2), which depicts the sequence of events that led to the emergence of the new variant (see Table 1 and Fig. 2). As we deduce a fairly robust unidirectional gene flow also penetrating the clade (South African; Brazil), and positioned between the British variant group (501Y.V1) symbolizes «ENGLAND2» and the South African variant(501Y.V2), as well as a third gene flow which is also unidirectional and consists but this time in the same clade between the Brazilian variant(501Y.V3) and the South African variant(501Y.V2) (see Fig. 2). In accordance with the AIM model, and as we mentioned in the section on the methodology and materials, in order to determine which

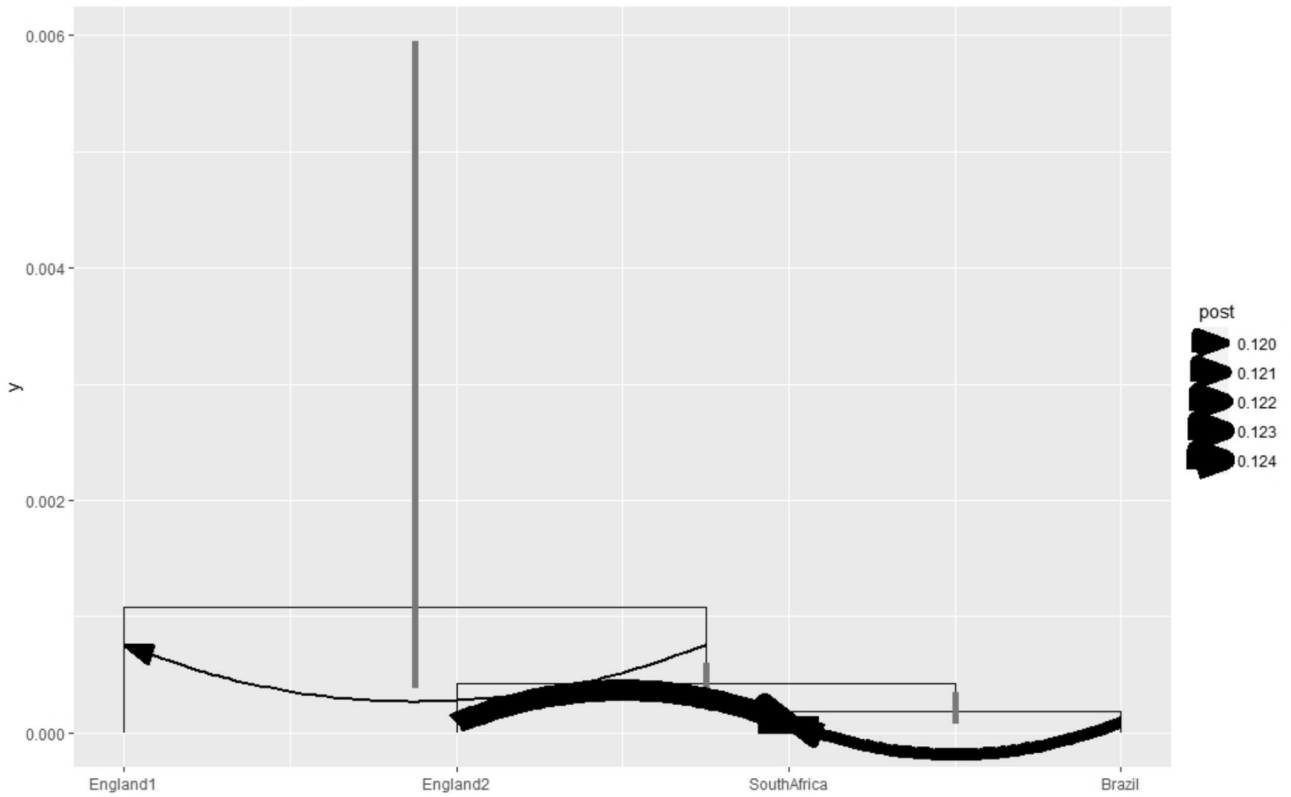

Fig. 2 Best supported ranked tree with gene flow and node height bars. The inferred variants history of novel coronaviruses is presented in units of substitutions per site averaged over all 4 random subsets of SARS-CoV-2 and its variants. The node heights are the median inferred speciation times. The grey bars show 95% highest posterior density intervals for speciation times. The heights are given in substitutions per site. The cutoff for an arrow to be plotted is support for gene flow with posterior support of at least 0.5. The second number is for a Bayes Factor values threshold for Bayes values calculated using the third number as a prior probability for gene flow

Table 1 Epidemiological parameters estimated by trees obtained from the AIM model

Parameter	Units	Mean	95% HPD Lower	95% HPD Upper
migrationIndicators.Variants				
migIndicators.Brazil_to_SouthAfrica	(10^{-2} sub/site)	0.051	0	1
migIndicators.England2_to_SouthAfrica	(10^{-2} sub/site)	0.0516	0	1
migIndicators.England1_to_SouthAfrica:Brazil:England1	(10^{-2} sub/site)	0.0541	0	1
migrationRates.Variants				
bmig.Brazil_to_SouthAfrica	(dimensionless rate*)	0.9272	0	399.269
bmig.England2_to_SouthAfrica	(dimensionless rate*)	0.8176	0	186.684
bmig.England1_to_SouthAfrica:Brazil:England1	(dimensionless rate*)	1.0655	0	203.154
populationSizes.Variants				
Ne_SouthAfrica	(viral cycle/lineage)	4.955 × 10^{-4}	3.135× 10^{-4}	6.92× 10^{-4}
Ne_Brazil	(viral cycle/lineage)	1.72 × 10^{-4}	1.027 × 10^{-4}	2.362 × 10^{-4}
Ne_England2	(viral cycle/lineage)	2.054 × 10^{-4}	1.30 × 10^{-4}	2.815 × 10^{-4}
Ne_SouthAfrica:Brazil:England2	(viral cycle/lineage)	2.4738 × 10^{-4}	1.268 × 10^{-4}	4.059 × 10^{-4}
Ne_SouthAfrica:Brazil:England2:England1	(viral cycle/lineage)	2.598 × 10^{-4}	1.414 × 10^{-4}	4.391 × 10^{-4}

HPD, highest posterior density. Posterior parameter estimates of SRAS-COV-2 and its variants analysis. Mean posterior estimates and 95% HPD intervals. migrationIndicators.Variants: the number of active gene flow routes(in 10^{-2}sub/site). migrationRates.Variants: migration rates of active gene flow. populationSizes.Variants: effective population size (viral cycle/lineage)

*BEAST analyses

gene flows are significant, we have sampled the probability that each migration rate is zero. Within the MCMC, the probability of each rate being zero was assumed to be equal to 0.5 a priori. Following this, the Bayes Factor (BF) for gene flow between different clades was calculated as the Bayes Factor (BF) of any migration rate from clade a to clade b that was nonzero [18]. The arrows are displayed between the clades for which the Bayes Factor (BF) is equal to or greater than 2, as demonstrated in this study (see Fig. 2).

7 Discussion

In spite of the fact that gene flow is an essential component of population genomics research and an essential evolutionary process that plays a vital part in the creation of genetic variation both within and between populations and species, gene flow is recognised as an integral feature of the field [19]. Highlighting his inference remains a notoriously difficult statistical problem. Our study demonstrated a number of gene flows between the geographically separated gene pools of SARS_CoV2 and its variants. Although the appearance of this new virus on a pandemic scale is very recent and is in its early stages, this virus tends to lead to competition between the introduced viral lineages (British; South African; Brazilian) and their endemic ancestors for limited resources such as the sensitive human host. This process will have lasting effects on the structure of the viral population, including the extinction of the endemic lineage, followed by the British variant, and will give way mainly to the clade (South African; Brazilian) because the analysis of the phylogenetic tree of SARS_COV2 and its variants, showed the presence of a clade made only of the two South African and Brazilian variants, this monophyletic group is characterized by a rapid divergence because this clade proceeds the shortest lengths of branches (see Fig. 1), if we compare it with the sister group (British variant) where this lineage is moderately divergent in comparison with the endemic lineage (SARS_COV2) which represents the slowest divergence from the other variants. This is the case for viruses that are derived from close evolutionary lineages and are sufficiently antigenically similar (that is, those that belong to the same subtype) to elicit a cross-protective immune response. In contrast, viruses that are antigenically divergent (that is, viruses of different subtypes) and that originate from lineages that are phylogenetically too far apart would be able to infect the same susceptible individual without being subjected to immune selection. For the purpose of explaining the evolutionary behaviour of these new RNA viruses, we will adopt the postulate of competitive exclusion, which states that when two species compete for limited resources, one species will eventually outcompete the other and become dominant in the population [20], Additionally, we will adopt the principle that in RNA viruses, a combination of high replication numbers and high nucleotide substitution rates may make it unlikely that two or more genetically distinct viral populations will coexist for an extended period of time [21, 22]. This is consistent with endemic SARS_CoV2 isolates and its new variants, including the most divergent variants of SARS_CoV2 such as the south-African that has active evolutionary behaviour with corresponding gene flow exchanges in the evolutionary dynamics of the gene pool of the SRAS_CoV2 variants. Because according to this study,

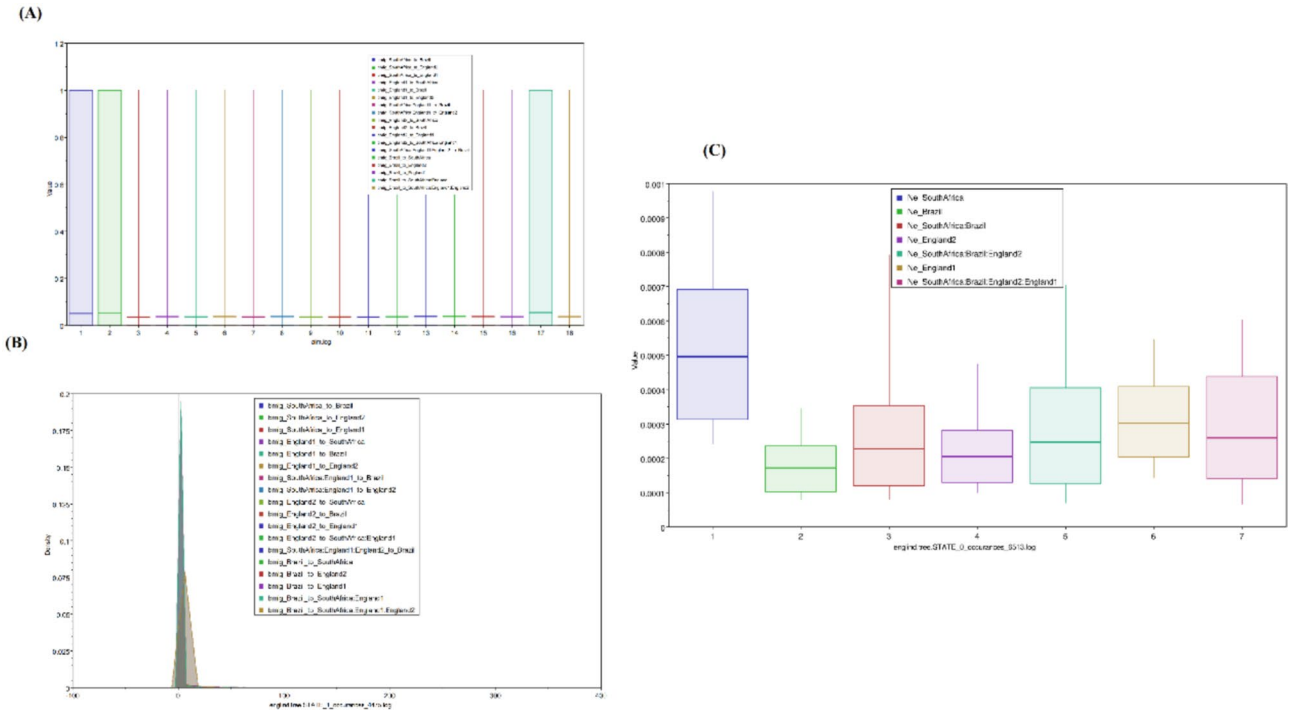

Fig. 3 Inferred variants history of SRAS-CoV-2. **A** Box and whisker plots of the numbers of active gene flow routes of each extant and ancestral variants. The colors of the Box and whisker plots correspond to the colors of the variants tree. **B** Compare the inferred migration Rates variants of active gene flow between SARS-CoV-2 and its variants. **C** Box and whisker plots of the effective population sizes of each extant and ancestral variant. The colors of the Box and whisker plots correspond to the colors of the variants tree

as we have seen from the results obtained (see Table 1 and Fig. 3), this South African variant received two large gene flows, one from the British variant and the other from the Brazilian variant, which makes it too divergent from the endemic strain, and even exceeds the Brazilian variant that forms with it, the active monophyletic group. When analyzing parameter values such as the number of active gene flow pathways, gene flow rates and effective population size obtained by the Approximate Isolation with Migration (AIM) model that is part of the BEAST StarBeast2 package [23] (see Table 1 and Fig. 3), these values are observed to be almost homogeneous and there is no big difference between the numbers of these parameters for each variant or between its variants and their endemic variants (see Table 1), which is simply to say that there is an early process of genetic diversity, and for there to be differences in the values of these evolutionary parameters, the process of genetic diversity must take the necessary time to evolve for this pandemic coronavirus species and its variants. But the peculiarity of this study showed that the appearance of new variants in the endemic gene pool will radically change the evolutionary dynamics of the SARS-COV-2 virus in the human reservoir. In another way it seems to us that we will probably see the appearance of new variants of SARS-CoV-2 with too much genetic power to diversify because these new variants receive very consistent gene flow to build a very competitive genome, and the disappearance (extinction) of endemic variant since they will compete unsuccessfully with the newly introduced variants.

8 Conclusion

Based on the findings of this study, it can be concluded that the physical form of each viral lineage belonging to its geographical pool has undergone profound genetic diversification as a result of extended geographic isolation and harsh measures of border closures. The high frequency of the formation of novel variations is most likely owing to antigenic differences between lineages within the same geographical pool. These differences are caused by the exchange of gene flow between lineages and lineages, as well as between lineages and hosts. It is important to note that the primary host, which is the human, was frequently responsible for the transmission of the variations across continents. Every one of these evolutionary scenarios of this pandemic calls for decision-making authorities to place the epidemiology of gene migration at the forefront of surveillance of emerging viral populations. This includes increased monitoring of the complete genome of new variants in other geographically isolated areas in order to gain a better understanding of the effects of gene flow between these variants. It is imperative

that these impacts be taken into consideration when formulating and assessing vaccine programs since they have the potential to have an impact.

Author contributions In terms of the conception and design of the study, each author made a contribution. Nabile Benazi and Azzedine Melouki were the individuals responsible for the preparation of the material. Both Nabile Benazi and Mohammed Abdallah Khodja were responsible for the collecting of data and the analysis of that data. Sabrina Bounab was the one who penned the initial draft of the manuscript. The manuscript was written and edited by Hadia Benhalima and Zahra Sadok, both contributed to the editing process. It was read and agreed by all of the authors before it was finalised.

Funding This research did not receive any kind of funding.

Data availability The authors confirm that all data fully available without restriction in NCBI and GSAID website mentioned in [11] and [12].

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. Domingo E, Holland JJ (1997) RNA virus mutations and fitness for survival. Annu Rev Microbiol 51:151–178
2. Drake J, W Holland JJ (1999) Mutation rates among RNA viruses. Proc Nat Acad SCI USA 96:1391C-13913
3. Holland JJ, de la Torre JC, Steinhauer DA (1992) RNA virus populations as quasispecies. Curr Topics Microbial Immunol 176: l-20
4. Benazi N, Bounab S, Bounab A (2020) Transmission route and introduction of pandemic SARS-CoV-2 between China, Italy and Spain. J Med Virol 93:564–568. https://doi.org/10.1002/jmv.26333
5. Drake JW (1993) Rates of spontaneous mutation among RNA viruses. Proc Natl Acad Sci U S A 90(9):4171–4175. https://doi.org/10.1073/pnas.90.9.4171
6. Worobey M, Holmes EC (1999) Evolutionary aspects of recombination in RNA viruses. J Gen Virol 80(Pt 10):2535–2543. https://doi.org/10.1099/0022-1317-80-10-2535
7. Lai MM (1992) RNA recombination in animal and plant viruses. Microbiol Rev. 56(1):61–79. PMID: 1579113; PMCID: PMC372854
8. Jia W, Karaca K, Parrish CR, Naqi SA (1995) A novel variant of avian infectious bronchitis virus resulting from recombination among 3 different strains. Arch Virol 140:259–271
9. Wang L, Junker D, Hock L, Ebiary E, Collisson EW (1994) Evolutionary implications of genetic variations in the S1 gene of infectious bronchitis virus. Virus Res 34(3):327–338. https://doi.org/10.1016/0168-1702(94)90132-5.PMID:7856318;PMCID:PMC7134089
10. Webster RG, Wright SM, Castrucci MR, Bean WJ, Kawaoka Y (1993) Influenza–a model of an emerging virus disease. Intervirology 35(1–4):16–25. https://doi.org/10.1159/000150292
11. Elbe S, Buckland-Merrett G (2017) Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall 1:33–46. https://doi.org/10.1002/gch2.1018
12. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I (2019) GenBank. Nucleic Acids Res 47(D1):D94–D99. https://doi.org/10.1093/nar/gky989
13. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547–1549. https://doi.org/10.1093/molbev/msy096
14. Posada D (2008) Jmodeltest: phylogenetic model averaging. Mol Biol Evol 25(7):1253–1256. https://doi.org/10.1093/molbev/msn083
15. Price MN, Dehal PS, Arkin AP (2010) Fasttree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE 5(3):e9490. https://doi.org/10.1371/journal.pone.0009490
16. Joëlle BS, Veronika B, Louis d P, Denise K, Carsten M, Venelin M, Nicola FM, Jūlija P, David AR, Chi Z, Alexei JD, Tracy AH, Oliver GP, Timothy GV, Tanja S (2018) Taming the BEAST – A community teaching material resource for BEAST 2. System Bio 67(1):170-174. https://doi.org/10.1093/sysbio/syx060
17. Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430):773–795
18. Nicola FM, HuwA O, Chi Z, Alexei D, Tanja S (2018) Inference of species histories in the presence of gene flow. 348391. https://doi.org/10.1101/348391
19. Adams RH, Schield DR, Castoe TA (2019) Recent advances in the inference of gene flow from population genomic data. Curr Mol Bio Rep 5:107–115. https://doi.org/10.1007/s40610-019-00120-0
20. Hardin G (1960) The competitive exclusion principle. Science 131(3409):1292–1297. https://doi.org/10.1126/science.131.3409.1292
21. Clarke DK, Duarte EA, Elena SF, Moya A, Domingo E, Holland J (1994) The red queen reigns in the kingdom of RNA viruses. Proc Natl Acad Sci U S A 91(11):4821–4824. https://doi.org/10.1073/pnas.91.11.4821
22. Andrés M, Santiago FE, Alma B, Rosario M, Eladio B (2000) The evolution of RNA viruses: A population genetics view. Proceedings of the National Academy of Sciences 97(13):6967–6973. https://doi.org/10.1073/pnas.97.13.6967.
23. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS computational biology. 8;15(4):e1006650