ELSEVIER

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Evolution and spread of SARS-CoV-2 likely to be affected by climate

Priyanka Bajaj [a,1,*], Prakash Chandra Arya [b,1,*]

[a] *Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India*
[b] *Center for Earth Sciences, Indian Institute of Science, Bangalore 560012, India*

ARTICLE INFO

ABSTRACT

COVID-19 pandemic has been a subject of extensive study. However, it is still unclear why it was restricted to higher latitudes during the initial days and later cascaded in the tropics. Here, we analyzed 176 SARS-CoV-2 genomes across different climate zones and Köppen's climate that provided insights about within-species virus evolution and its relation to abiotic factors. Two genetically variant groups, named G1 and G2, were identified, well defined by four mutations. The G1 group (ancestor) is mainly restricted to warm and moist, temperate climate (Köppen's *C* climate) while its descendent G2 group surpasses the climatic restrictions of G1, initially cascading into neighboring cold climate *(D)* of higher latitudes and later into the hot climate of the tropics *(A)*. It appears that the gradation of temperate climate *(Cfa-Cfb)* to cold climate *(Dfa-Dfb)* drives the evolution of G1 into the G2 variant group, which later adapted to tropical climate *(A)* as well. It seems this virus followed an inverse latitudinal gradient in the beginning due to its preference towards temperate *(C)* and cold climate *(D)*. Our work elucidates virus evolutionary studies combined with climatic studies can provide crucial information about the pathogenesis and natural spreading pathways in such outbreaks, which is hard to achieve through individual studies. Mutational insights gained may help design an efficacious vaccine.

Introduction

The first case of Corona Virus Disease-19 (COVID-19) pandemic caused by (Severe Acute Respiratory Syndrome Coronavirus-2) SARS-CoV-2 pathogen was first reported from Wuhan, China [1]. Despite various precautions such as lockdown, social distancing, wearing a mask, and sanitization, the disease reached almost every part of the world [2]. This zoonotic virus is like SARS-CoV-1 (Severe Acute Respiratory Syndrome Coronavirus-1) (79% similarity), and MERS-CoV (Middle East Respiratory Syndrome Coronavirus) (50% similarity) and is closely related to bat-derived coronaviruses [1]. The SARS-CoV-2 can survive up to 3, 4, and 24 h on aerosols, copper, and cardboard, respectively. It can survive up to 3 days on stainless-steel or plastic [1]. SARS-CoV-2 spreads faster than its ancestors SARS-CoV-1 and MERS-CoV [1,3]. This led it to concur a larger geographical region by infecting a larger population. Since the social behavior and traveling of humans have not changed much, what makes few respiratory viruses confined locally and others spread globally is still unclear. COVID-19 outbreak led to an extensive discussion, does climate have a role in the spread of the disease? The ancestor SARS-CoV-1 losses its viability at higher temperature (38°C) and relatively higher humidity (>95%) [4]. Experiments support that SARS-CoV-2 is highly stable at 4°C but is sensitive to heat [5]. Studies both

in favor and against have been published, but the topic is still debatable [6,7]. A recent review compiles 61 studies relating COVID-19 with climate [1]. Several climatic factors such as humidity, precipitation, radiation, temperature, and wind speed affecting this virus spread have been incorporated. Both positive and negative associations of COVID-19 with temperature and humidity have been published [1]. Recently, Carlson et al. mentioned that COVID-19 transmission could be affected by climate but discouraged the use of SDMs (Species Distribution Models) for COVID-19 transmission due to their limitations, which generally takes only climatic parameters as input, as they may not be appropriate tools [7]. This study was challenged by Araujo et al., mentioning the strengths and limitations of the tools and reasoned that R_0 of COVID-19 depends on several factors; it may also be affected by climate [6]. Since only climatic parameters are insufficient to capture climatic signatures of COVID-19 spread, such patterns can be recognized by combining phylogenetic and climatic studies [8]. Such approach enables to probe the similarities or differences in virus genome across similar or different climate types present all over the world. To understand such a behavior, we have attempted to study the genomic sequence of the SARS-CoV-2 across different climates.

Till now, pathogens are poorly mapped, and little is known about their underlying ecological and evolutionary causes [9]. Nucleotide substitution is considered to be one of the most important mechanisms of

viral evolution in nature [10]. However, factors responsible for the generation of these mutations are not well understood. One of the possible factors is adaptation to new environments, dictated by natural selection that discriminates among genetic variations and favors survival of the fittest [11]. Virus evolution as a consequence of climate change is poorly understood. SARS-CoV-2 consists of large single-stranded ~30 kb long positive-sense RNA. These viruses majorly have a conserved genomic organization, consisting of a unique 265 bp long leader sequence, ORF1ab polyprotein, and structural proteins like S (spike glycoprotein), *E* (Envelope), *M* (Membrane), and *N* (Nucleocapsid). ORF1ab encodes replicase, transcriptase, and helicase, essential enzymes required for replication, along with non-structural and accessory proteins. Expression of non-structural proteins is facilitated by ribosomal frameshifting [12,46]. All coronaviruses express structural proteins *S, E, M, and N*; spike glycoprotein *(S)* being the most immunogenic to T-cell response [13]. Spike glycoprotein of coronaviruses binds to human angiotensin-converting enzyme 2 (hACE2) receptor for viral fusion and entry and is the main target for neutralizing antibodies and development of vaccines [14]. The membrane protein is also antigenic as it stimulates a humoral immune response [15]. The Envelope protein (*E*) protein is responsible for virus assembly and release of virion particles [16]. Nucleocapsid protein *(N)* packages RNA genome into a helical ribonucleocapsid protein (RNP) complex during virion assembly and can elicit an immune response [17]. Since it is still unclear whether SARS-CoV-2 evolution and spread have a relation with climate, our study will act as a missing link between genomic sequence, climate, and COVID-19 severity. If SARS-CoV-2 responds to external climate, it can be delineated by superimposing its genomic variants across different climate zones and Köppen's climate [8]. The earliest and the most simple classification of Earth's climate is based on latitudes which divide the Earth's climate into seven climate zones, North Frigid Zone (NFZ), North Temperate Zone (NTZ), North Subtropical Zone (NSTZ), Tropical Zone (TZ), South Subtropical Zone (SSTZ), South Temperate Zone (STZ) and South Frigid Zone (SFZ) lying between 90°N to 66.5°N, 66.5°N to 30°N, 30°N to 23.5°N, 23.5°N to 23.5°S, 23.5°S to 30°S, 30°S to 66.5°S and 66.5°S to 90°S, respectively [18]. Based on temperature and precipitation, Wladimir Köppen divided Earth's climate into five major climates, *A* (Tropical), *B* (Arid), *C* (Temperate), *D* (Cold or Continental), and *E* (Polar), which are further subdivided into 30 climate types [19]. To understand the effect of climate on SARS-CoV-2 evolution, the present study comprises two parts: (1) Sequence analysis of SARS-CoV-2 strains, (2) Mapping SARS-CoV-2 strains across different climates. These combined studies can provide insights on within-species evolution and preferential distribution of SARS-CoV-2 across different climates which might be difficult to probe through individual studies.

Methodology

Molecular phylogenetic analysis

Approximately 11,000 full-length genome sequences of SARS-CoV-2 were available in Global Initiative on Sharing Avian Influenza Data (GISAID) database, accessed till 2nd May 2020. 185 full-length SARS-CoV-2 genomic sequences from countries worldwide, with genome length more than 29 kb and high coverage, were obtained from the GISAID database, and the reference genome was retrieved from GenBank24 (Supplementary Table S1). This sample size was selected by taking a 95% confidence interval, 0.5 standard deviation, and 7% margin of error. To avoid bias related to the geographical area covered by a country, the genomic sequence of strains isolated from different locations from each country/ climate type was retrieved, depending on the availability of data. The corresponding location, latitude, Köppen's climate, Köppen's climate type, SARS-CoV-2 variant group, environment/ region, climate zone, temperature, and precipitation of each strain are provided in Supplementary Table S2. These sequences were aligned to the full reference genome [20] using Biomanager and Seqinr packages of *R*

(version 3.6.3). Among 185 genomes, some partial genomes were discarded. Thus, finally, 176 genomes were analyzed. NC_045512 genome sequence was used as a reference, and the genomic coordinate in this study is based on this reference genome. Based on protein annotations, nucleotide level variants were converted into amino acid codon variants for alignments when their location within a gene was identified. The amino acid position numbering is according to its position within the specified gene (Coding Sequence) as annotated in the reference sequence (NC_045512, NCBI) [20]. To ensure comparability, we trimmed the flanks of all sequences. The aligned sequences were used to construct a phylogenetic tree using MEGA X [21]. The evolutionary history was inferred using the Neighbor-Joining method (500 bootstrap tests) [22]. The optimal tree with the sum of branch length = 0.01116462 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [23] and are in the units of the number of base substitutions per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option). A total of 29,408 positions were present in the final dataset. The results are presented in the form of DNA sequencing, i.e., *U* (uracil) is read as *T* (thymine). We have labeled each virus strain by the GISAID Accession ID and the location from which it was isolated in the format "Location|EPI ISL Accession ID", in the constructed phylogenetic tree. For ease of visualization, we have marked a new Strain ID (1 to 176) against each SARS-CoV-2 isolate in the phylogenetic tree (Fig. 1). The same Strain ID is used for the climatic studies in this article. High-frequency SNPs (Single Nucleotide Polymorphisms) distinguishing one virus cluster from the others is referred to as "virus cluster SNPs" throughout this paper.

Mapping virus strain on the Köppen's climate map

The location of each SARS-CoV-2 strain is obtained from the META-DATA file provided in the GISAID database for each viral isolate (Supplementary Table S1). The coordinates of the locations were taken from the official website of USGS Earth Explorer [24]. Köppen-Geiger map is used for climatic studies [19]. The Köppen's climate type, temperature, precipitation of each strain was assessed from weatherbase [25] and climate.org [26]. The above information for each strain is tabulated in Supplementary Table S2. The map is georeferenced by using 'Arc-GIS 10.1′ [27]. The locations of all strains (*n*=176) were transferred to the georeferenced map [27]. On the map, the G1 strains were symbolized as 'Yellow-circle' and G2 as 'Square' (Fig. 3). Each strain in the map is labeled as per their Strain ID (1 to 176) (Fig. 3). The map combines information of the phylogeny, climate, and global distribution of SARS-CoV-2. These locations were classified into the coastal and continental regions. We define the coastal region as land region < 500 km from the ocean/sea and the continental region as land lying >500 km from the coastline measured through Google maps.

Statistical analyses

Chi-square tests were performed in Microsoft Excel (2016) to test our null hypothesis. The null hypotheses for these tests are mentioned in the text, and its corresponding contingency table is mentioned in Supplementary Table S6. Histograms depicting the distribution of coronavirus in the coastal region, continental region, Köppen's climate, and climate type were plotted using *R* (version 3.6.3). SigmaPlot10 was used to generate box plots, regression plots, and mesh plots to statistically compare the frequency distribution of latitude, temperature, and precipitation of G1 and G2 strains. We performed one-way ANOVA to estimate statistical differences in the latitude, temperature, and precipitation between G1 and G2 virus populations. Various scatterplots between latitude, temperature, and precipitation of G1 and G2 strains were plotted in *R* (version 3.6.3). *P* values below 0.05 were considered statistically significant. Exact *P* values are given in respective figures.
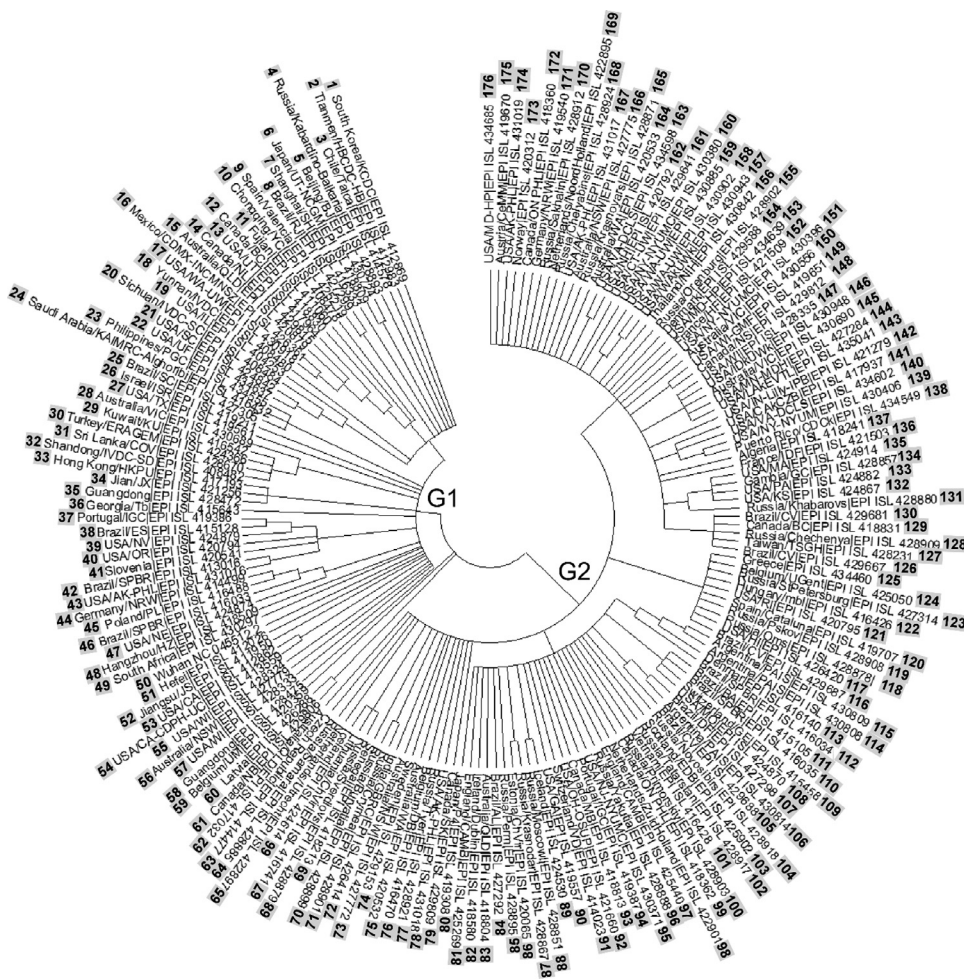
Data accessibility

The full-length genomic sequences were downloaded from the GISAID website (https://www.gisaid.org/), an open-source database for influenza viruses. The data was downloaded as a FASTA file along with the acknowledgment. The location of each strain is accessed from its METADATA file. Köppen's climate map was taken from Peel et al. (2007) [19]. The Köppen's climate type, temperature, and precipitation for each strain were taken from weatherbase (https://www.weatherbase.com/) and climatedata.org (https://en.climate-data.org/). Refer Supplementary Tables S1–S5. The code is available from the corresponding authors on request.

Results

Molecular phylogeny analysis to infer genomic similarities and their distribution in different climates

To probe genomic similarities between SARS-CoV-2 virus isolates, a phylogenetic tree was constructed by aligning 176 virus genomes to the reference genome [20] retrieved from GISAID. Interestingly, our Multiple Sequence Alignment (MSA) results reveal sixty virus cluster SNPs (see methodology). Table 1 comprises SNPs of virus clusters across different climatic zones, Köppen's climate, and climate0 type. Climatic parameters (temperature and precipitation) for each virus strain are mentioned in Supplementary Table S2. Based on phylogenetic clustering, 176 SARS-CoV-2 strains are majorly divided into two groups, and we named them as G1 (1-58) and G2 (59-176) (Fig. 1). Predominantly

four mutations distinguish G2 from G1 group, i.e., (1) a synonymous mutation (C241T) appeared in the unique leader sequence, (2) F924 (C3037T) appeared in nsp3, encoding for papain-like proteinase [28], (3) a non-synonymous mutation, P214L (C14408T) arose in ORF1b, that codes for four putative non-structural proteins (nsp13, nsp14, nsp15, and nsp16), functionally involved in replication-transcription complex [29], and 4) D614G (A23403G) arose in *S* gene, encoding spike glycoprotein [13] (Fig. 2a). Among four mutations in G2, the D614G mutation, lying in spike glycoprotein, was widely studied due to its higher infectivity and involvement in entering the host cell through hACE2 receptors [30–33]. The other three mutations in G2 have coevolved with D614G making it distinguishable from G1. We explored the extent of genome-wide divergence of the G1 and G2 groups across different climate zones and Köppen's climate (Fig. 2b). 59% of G1 viruses fall in NTZ, 14% in NSTZ, 12% in TZ, 10% in SSTZ and 5% in STZ. 76% of the virus isolates in the G2 group are present in the NTZ, 13.5% in TZ, 7.6% in STZ, and the remaining 2% is equally distributed in NSTZ and SSTZ, showing G2 strain variants evolved to adapt to temperate zones as their population decreased drastically in the subtropical zones. These results show both G1 and G2 strains have a strong preference towards higher latitudes, i.e., NTZ (Fig. 2c). Mapping viral strains on Köppen's map (thoroughly discussed in the next section) reveal their prevalence majorly in the *C* and *D* climate (Fig. 2d). 71% of G1 lie in the *C* climate, 17% in *D*, and the remaining are equally distributed in the *A* and *B* climate. 54% of G2 lie in *C* climate, 36% in *D*, 9% in *A*, and 1% in *B* climate, pointing towards a preferential shift of the novel coronavirus towards *D* climate (Fig. 2b), alluding G2 is climatically and genomically more diverse than G1. The analysis suggests that the G1 group is mainly

Table 1
SNPs representing virus cluster and their distribution across varied climates.

Virus cluster	Nucleotidemutation	Amino acid mutation	Gene	Climate Zone	KCT	KC
1–22	C8782T	S2839	ORF1a	NTZ	Cfa	C
	T28144C	L84S	ORF8			
5–6	C29095T	F274	N	NTZ	Cfa	C
8–9	T9477A	F3071Y	ORF1a	NTZ, TZ	Mix	C-A
	G25979T	G196V	ORF3a			
	C28657T	D128	N			
	C28863T	S197L	N			
10–17	C18060T	L1431	ORF1b	NTZ	Cfa-Cfb	C
12–17	A17858G	Y1364C	ORF1b	NTZ	Cfa-Cfb	C
13–17	C17747T	P1327L	ORF1b	NTZ	Cfa	C
20–22	C24034T	N824	S	NTZ	Cfa	C
	T26729C	A69	M			
	G28077C	V62L	ORF8			
21–22	T490A	D75	ORF1a	NTZ, NTSZ	Cfa	C
	C3177T	P971L	ORF1a			
	T18736C	F1657L	ORF1b			
23–25	C6312A	T2016K	ORF1a	NTZ, TZ, SSTZ	Mix	Mix
	C13730T	L4489	ORF1a			
	C23929T	Y789	S			
	C28311T	P13L	N			
28–32	G1397A	D392G	ORF1a	NTZ	Mix	Mix
	T28688C	L139	N			
33–43	G26144T	G251V	ORF3a	NTZ	Cfa-Cfb	C
37–39	A2480G	I739V	ORF1a	NTZ	Mix	Mix
	C2558T	P765S	ORF1a			
37–43	C14805T	Y346	ORF1b	NTZ	Cfa	C
42–43	T17247C	R1160	ORF1b	NTZ	Cfb	B
44–47	G1440A	G392D	ORF1a	NTZ	Cfb	C
	G2891A	A876T	ORF1a			
58–61	C15324T	N519	ORF1b	NTZ	Cfa-Dfb	C-D
59–176	C3037T	F924	ORF1a	NTZ	Cfa-Cfb-Dfb-Aw	C-D -A
	A23403G	D614G	S			
	C14408T	P214L	ORF1b			
59–125, 127–176	C241T		Leader seq.	NTZ	Cfa-Cfb-Dfa-Dfb	C-D-A
	C241T					C-D
66–68	A26530G	D3G	M	NTZ	Cfc-Dfb	C-D
70–71	G4201T	M1312I	ORF1a	NTZ	Cfa-Dwc	C-D
	C26527T	A2V	M			
80–115	G28881A	R203K	N	NTZ	Cfa-Cfb-Dfb-Aw	C-D-A
	G28882A	R203K	N			
	G28883C	G204R	N			
86–87	C27046T	T175M	M	NTZ	Cfa-Dfb	C-D
88–89	C3373A	D1036E	ORF1a	NTZ	Dfb-Cfb	C-D
105–107	T29148C	I292T	N	TZ, STZ	Cfa-Aw	C-A
106–107	A27299C	I33T	ORF6	NTZ, TZ	Cfa-Aw	C-A
108–111	C313T	L16	ORF1a	NTZ, TZ	Cfa-Cfb-Aw	C
113–115	C4002T	T1246I	ORF1a	STZ	Cfa-Cfb-Am	C-A
	G10097A	G3278S	ORF1a			
	C13536T	T4424I	ORF1a			
	C23731T	T723	S			
116–125	A20268G	L2167	ORF1b	NTZ	Cfa-Cfb-Dfa-Dfb	C-D
126–176	G25563T	Q57H	ORF3a	NTZ	Cfa-Cfb-Dfa-Dfb	C-D
126–130	C18877T	L1704	ORF1b	NTZ	Cfa-Dfa-Dcb	C
131–135	C2416T	Y717	ORF1a	NTZ	Cfa-Dfa-Aw	D
136–176	C1059T	T265I	ORF1a	NTZ	Cfa-Cfb-Dfa-Dfb	C-D
138–139	C18998T	A1744V	ORF1b	NTZ, TZ	Cfa-Am	C-A
	G29540A					
138–141	C11916T	S3884L	ORF1a	NTZ	Cfa-Csb-Am	C
143–147	C27964T	S24L	ORF8	NTZ	Cfa-Cfb-Dfa-Dfb	C-D
148–149	C11224T	V3653	ORF1a	NTZ	Dfa-Dfb	D
157–159	G29553A			NTZ	Cfa	C

NOTE: Virus clusters are named by Strain ID as depicted on the tree (Supplementary Tables S1 and S2). Genomic coordinates in this study are based on reference genome [20]. The SNP positions are based on the reference genome. Nucleotide *T* represents nucleotide *U* in the SARS-CoV-2 RNA genome. Mutation at the protein level is not mentioned for the SNPs arising in the non-coding region. The amino acid position numbering is according to its position within the specified gene (CDS). In the Climate zone column, we have mentioned the major climate zone for the corresponding virus cluster [18]. KCT is Köppen's Climate Type, and KC is Köppen's Climate columns display the main Köppen's climate in which most of the virus isolates of the corresponding virus cluster lie. 'Mix' implies no particular climate type is favored [19].
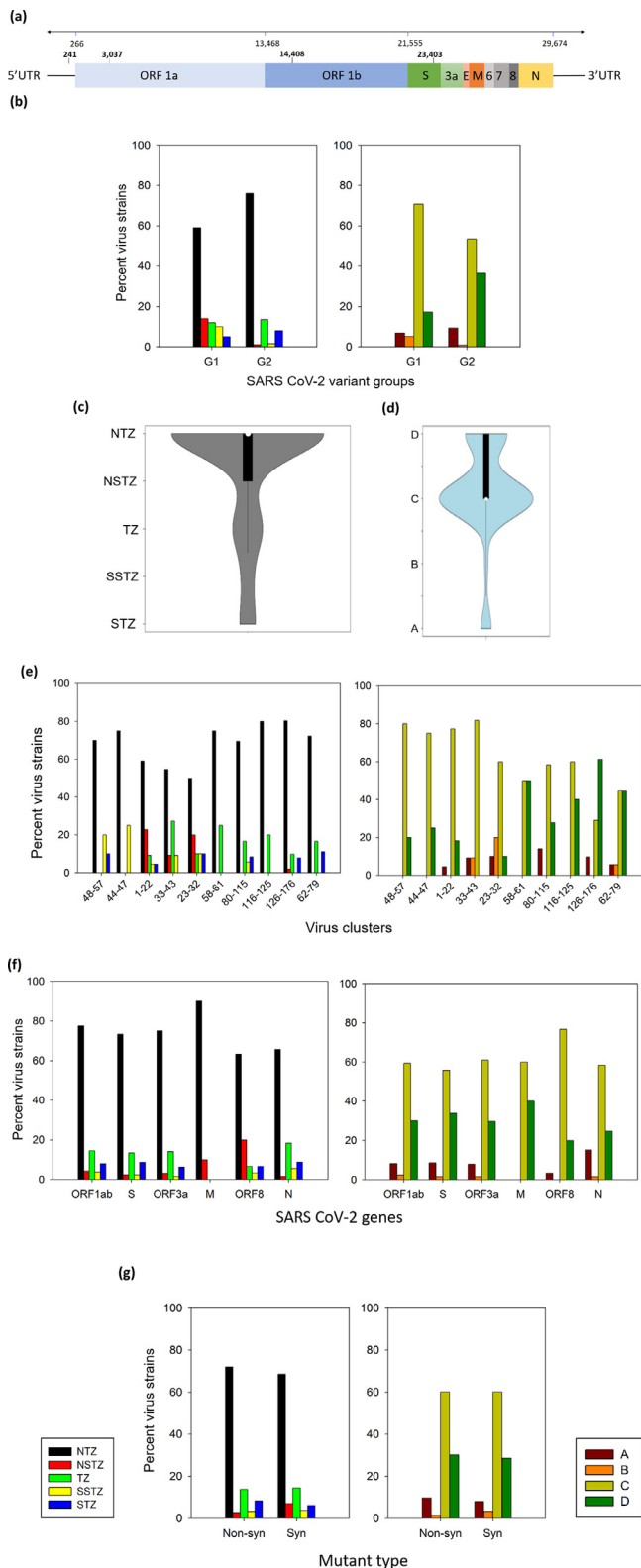
restricted to temperate climate *(C)* and G2 is climatically and geographically widely distributed. G1 might have acquired these mutations to sustain in different climates, hence allowing it to spread globally. Similar climatic concordance with the temperate climate *(C)* was also observed for SARS-CoV-1 that was responsible for the 2002–2004 epidemic as it prevailed in regions of Australia, Europe, Canada, and China [34], having Köppen's *C* climate. Such similar occurrence of SARS-CoV-1 and G1 group of SARS-CoV-2 hints towards why initially G1 variant group (consisting of the reference genome NC_045512[20]) that has 79% similarity to SARS-CoV-1[1] was majorly located in the temperate climate *(C)*. Later it evolved into G2 variant group that allowed it to extend its climatic boundaries into temperate, cold, and tropical climate. These results suggest these four SNPs could be the key factors in increasing the virulence, transmission, and sustainability of the virus in humans.

We further analyzed the order in which the phylogenetic clusters evolved from the ancestor 45–57 cluster (containing the reference genome, Strain ID: 50) based on nodes, mutational branches, and branch length. The order in which the virus evolved is 44–47 (G1440A, G392D; G2891A, A876T), 1–22 (C8782T, S2839; T28144C, L84S), 33–43 (G26144T, G251V), 23–32, 58–61 (C15324T, N519), 80–115 (G28881A, G28882A, R203K; G28883C, G204R), 116–125 (A20268G, L2167), 126–176 (G25563T, Q57H) and 62–79 (cluster, acquired genomic mutation and its corresponding amino acid mutation). In Fig. 2e, looking at the distribution of the viruses in different climate zones, no such preference was observed as the virus evolved. Virus cluster 58–61, linking G1 and G2, has an equal distribution of virus strains in *C* and *D* climate. The virus cluster 80–115 of G2, more closely related to G1, is widely distributed in *A, C,* and *D* climates. Within the 80–115 virus cluster, 106–115 subcluster shows the distribution in *C* and *A* climate. A trend was observed that virus clusters in the G2 group gradually evolved to sustain in Köppen's *D* climate (80–115 to 116–125 to 126–176). Within these major virus clusters, small clusters also exist, as shown in Table 1, with their mutations along with their climatic distribution.

We have examined whether climatic conditions exhibit any selective pressure on each gene (Fig. 2f). Since the present picture of the data shows that SARS-CoV-2 spreads widely in NTZ, as expected, all genes are having mutations in NTZ, suggesting the virus is probably using varied mechanisms to adapt to the two main climates of NTZ, i.e., temperate *(C)* and cold or continental *(D)*. Mutations in the *M* gene are only pertaining to NTZ and NSTZ and are present in *C* and *D* climate. In particular, there is a surge in the virus strains carrying SNPs in ORF8 in the NSTZ (20%). 77% of the SNPs in ORF8 lie in the *C* and 20% in the *D* climate. Overall, the distribution of virus cluster SNPs of ORF1ab, *S*, ORF3a, and *N* gene follows a similar pattern across all the climatic zones and Köppen's climate, implying no difference in selective pressure of the climate in generating mutations in these genes. *S, M,* and *N* proteins are immunogenic [13,15,17], implicating virus evades immune response by introducing these substitutions.

Apart from non-synonymous mutations, synonymous mutations within the gene can also significantly affect protein function due to codon usage bias [35] and mechanisms such as ribosome stalling [36] and mRNA secondary structure formation [37]. We probed the frequency of derived synonymous versus non-synonymous mutations and

Fig. 2. *Molecular phylogeny analysis to infer genomic similarities of SARS-CoV-2 and their distribution across different climate zones [18] and Köppen's climate types [19].* (a) Genomic architecture of SARS-CoV-2 genome highlighting four positions, substitutions on these positions probably enabled evolution of G1 into G2 variant group. (b, e–g) Strains found within a virus cluster (as shown in the phylogenetic tree and mentioned in Table 1) were analyzed for significant mutations that may have arisen due to climatic pressure. Hence, the percentage of such virus strains is plotted according to the geographical location of the climate zone from where they were isolated. The height of the bar is proportional to percent virus strain occurring in the specified condition i.e., labelled on the *x*-axis. Box in the left panel consist of color code for each climate zone and box in the right panel consist of color code for Köppen's climate. Left panel shows distribution of percent virus strains in different climate zones and right panel shows distribution of percent virus strain in Köppen's climate (b) Percent virus strains prevailing in different climate zones, stratified by SARS-CoV-2 variant groups. Width of curves of violin plot is proportional to the number of SARS-CoV-2 strains (*n* = 176) in varied (c) climate zones and (d) Köppen's climate. (e) Abiotic factors influencing evolutionary dynamics of phylogenetic virus clusters. (f) Percent of virus strains with high frequency SNPs in each gene. (g) Type of mutation, i.e., non-synonymous (Non-syn) or synonymous (Syn) exhibited by viruses.
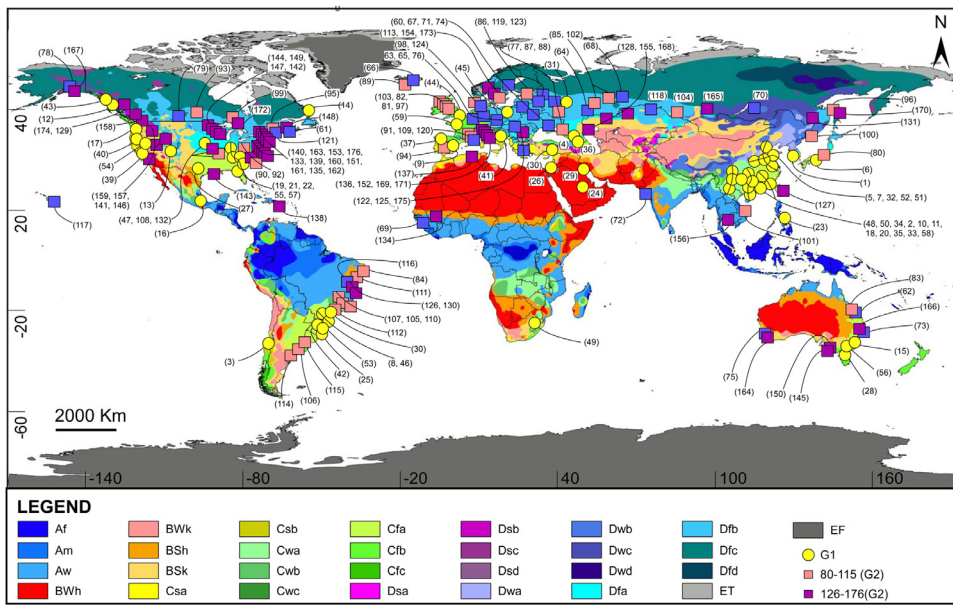
Fig. 3. *Global distribution of SARS-CoV-2 strains on the Köppen-Geiger map displaying different climate types* [19]. *Each strain is labeled as per the strain ID (1 to 176) within the parenthesis. The G1 strains were symbolized as 'Yellow-circle,' and G2 as 'Square', pink square denotes strain clusters (80–115) stable across C, D, and A climate, the purple square represents strain cluster (126–176) stable majorly in D climate, the remaining G2 strains (blue squares) are stable across C and D climate. Standard Köppen's climate-type symbols are mentioned in the legend. The criteria for distinguishing these climate types are mentioned in Table S3. Table S4 contains the full form of these symbols. All symbols with initials 'A' (Af, Am, Aw) are of tropical climate, initials with 'B' belong to desert climate, 'C' to temperate and 'D' to cold, and 'E' to polar climate. The shades of blue on the map, in North America and Russia, belong to the 'D' climate. The shades of blue in South America, Africa, and South Asia belong to tropical climate. Shades of yellow and green belong to the 'C' climate. Shades of red, orange, and pink belong to the Desert climate (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).*
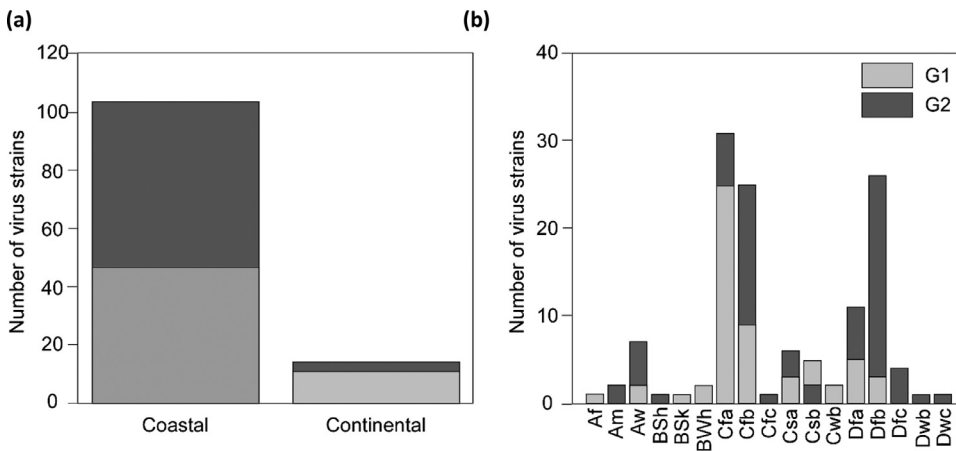

Fig. 4. *Global distribution of SARS-CoV-2 strains (n = 176)* (a) *in the coastal and continental region* (b) *and in different Köppen's climate types* [19]. *The number of virus strains in the G1 population is represented by light gray color, and virus strains in the G2 population are represented by dark gray color.*

observed a very similar distribution pattern of the derived synonymous versus missense mutations across all climate zones and Köppen's climate (Fig. 2g). These analyses suggest novel coronavirus is using varied mechanisms both at the transcriptional as well as the translational level to adapt, survive, and increase infectivity in all types of climates. These findings unequivocally bolster a requirement for further prompt, comprehensive studies that join genomic information, epidemiological information, and climatic distribution with COVID-19 severity.

Distribution of strains across Köppen's climate

To probe the relation between climate and SARS-CoV-2 strains, we superimposed genomic information along with their geolocations on the climate map of Wladimir Köppen (Fig. 3). We carefully examined the distribution of strains on Köppen-Geiger map, and an overview of the map shows, the distribution of 176 strains is mainly concentrated in the western coasts of Europe and North America, and eastern coasts of China, North America, Australia, and South America (Fig. 3). Throughout the text, Köppen's climate type is marked within quotations, and its standard symbol is written within brackets, e.g., "humid-subtropical" (Cfa). A list of Köppen's symbols of each climate type is given in Supplementary Table S3, and its criteria for classification is provided in Supplementary Table S4. Primarily the SARS-CoV-2 strains are distributed in

the "humid-subtropical" (Cfa) and "marine-temperate" (Cfb) and "humid-continental" (Dfa-Dfb) climate and, two strains from virus clusters (80–115 and 126–176) belonging to South America, are found in "tropical-savanna" (Aw) of A climate (Supplementary Table S5). The map displays ~86% (151/176) of virus isolates are distributed in the coastal regions and the remaining in the continental region (Chi-square test, $P < .001$ considering the null hypothesis that an equal number of virus isolates are found in coastal and continental region, Fig. 4a, Supplementary Table S6). Around ~74% (130/176) of the total strains are distributed in "humid-subtropical" (Cfa) and "marine-temperate" (Cfb) climate type of C climate and "humid-continental" (Dfa-Dfb) climate type of D climate. The remaining ~26% (46/176) strains are distributed in other climate types of Köppen's climate, including non 'Cfa-Cfb' of C climate and non 'Dfa-Dfb' of D climate (Fig. 4b). It seems that the spread of COVID-19 is maximally in areas with 'Cfa' and 'Cfb' climate types. The climatic parameters (temperature and precipitation) in which these strains lie were analyzed. Statistically, a significant difference was found in the latitudes of G1 ~24.14±3.5 (mean±s.e.) and G2 ~34.03 ± 2.7 (one-way ANOVA, $P = .03251$, Fig. 5a). Statistically, a significant difference was observed in the temperatures of G1 (15.82 ± 0.75 °C (mean±s.e.) and G2 (11.67 ± 0.68 °C) strains (one-way ANOVA, $P < .001$, Fig. 5b). However, the difference in precipitation for G1 (1046.95 ± 80 mm) and G2 (896.64 ± 35.48 mm) strains is statistically not significant (one-way
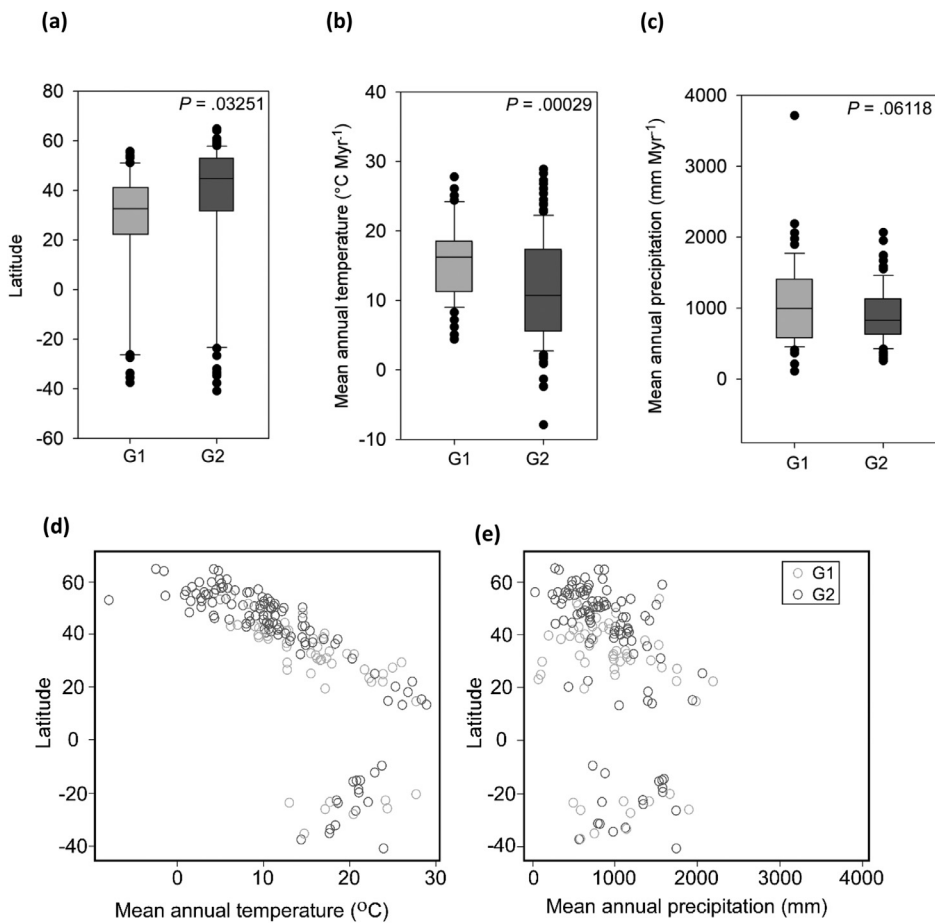
Fig. 5. *Comparative analysis of different climatic parameters such as latitude, temperature, and precipitation between G1 and G2 variant groups.* (a) Positive values represent the latitude range falling in Northern Hemisphere, and negative values represent the latitude range falling in Southern Hemisphere. The G2 strains preferentially occur towards the higher latitudes than G1 (P = .032; 95% CI 17.12–31.12 for G1; 95% CI 28.67–68.06 for G2). (b) The mean annual temperature of G2 is significantly lower than the G1 strains (P < .001; 95% CI 17.32–14.32 for G1; 95% CI 13.02–10.33 for G2) (c) Mean annual precipitation of G1 and G2 strains is nearly the same (P = .061; 95% CI 1207.16–886.75 for G1; 95% CI 966.91–826.37 for G2). (a–c) The black horizontal line in the middle of the box is median, upper and lower limits of the box indicate the first and third quartile. Black dots represent outliers. P values are based on one-way ANOVA. Scatter plot for (d) latitude and annual temperature and (e) latitude and precipitation for each SARS-CoV-2 strain (n = 176) belonging to the G1 group (n = 58, shown in light gray) and G2 group (n = 118, shown in dark gray).

ANOVA, $P = .06118$, Fig. 5c). The latitudes and temperature are inversely related to each other ($r = -0.6649$, Supplementary Fig. S1a), which explains the occurrence of G1 strains in lower and G2 strains in higher latitudes (Fig. 5d). Such relation between latitude and precipitation has not been observed ($r = -0.3064$, Supplementary Fig. S1b and Fig. 5e). A mesh plot simultaneously evaluates all climatic parameters for both G1 and G2 strains, the results agree to the limited temperature and wider precipitation range of the G1 group, and interestingly, the G2 group appears in a wider temperature range and shows a preferential shift towards lower temperature, which is evident from the fact that it initially appeared more in higher latitudes (Supplementary Fig. S2). A complete description of the distribution of G1 and G2 strains lying in different countries and/or continents of the world is provided in Supplementary Fig. S3 and Supplementary Material.

Discussion

A pattern observed and discussed by several authors is that SARS-CoV-2 predominantly affected higher latitude countries (e.g., China, Europe, USA and Australia, Japan, Turkey, etc.) in its preliminary stage (November 2019 to March 2020). This led many researchers to predict a reduced spread in regions with higher temperatures, such as tropical countries [38,39]. Our phylogenetic analysis provides one more dimension to the above observations in which we have identified two groups G1 and G2, similar observations were also reflected in other studies [40,41]. We have further integrated the genomic information with climatic data and analyzed the spread of these two variant groups across the globe and their climate.

Many reasons can unravel the observed pattern of spread. First, it could be because of less testing in tropical countries. However, if this would be true, local media would have reported over-crowding of patients in hospitals. For example, Italy escalated its testing when the media reported over-crowding of hospitals with patients having COVID symptoms. But such incidences were not reported from tropical countries. Second, it could be higher human mobility between China-US and China-Europe. But mobility between China and South-East Asia is also very high. Thus, no outbreak in tropical countries initially is perplexing (till March 2020). Third, most countries implemented lock-down measures from mid-March. Until March, people were working in normal mode, hence, it is not possible that stringent measures in containing the virus evaded it from cascading into tropical countries. Fourth, the population density of the tropics could be lower than in temperate countries. But this is also not true; most of the world's population resides in tropical countries and are mainly developing with limited proper health and hygiene facilities compared to temperate countries. In such a situation, the disease should have equally spread in tropics like that of temperate countries as positive cases of COVID-19 were identified from many tropical countries in January itself. Fifth, our analyses show that the G1 group spread largely in the temperate climate. A possible reason behind this could be that the host immune response of people residing in tropical countries resisted the infection. However, no supporting evidence has been found till now and needs detailed investigation.

Hence, these reasons negate the fact that the observed patterns could solely be due to community structure, social dynamics, government policies, global connectivity, population density, and the number of reported cases. Similar observations, i.e., more COVID cases in higher latitudes, were reflected in COVID-19 growth curves of Yusuf and Bukhari until March 2020 for different countries, which led them to interpret

that natural factors (temperature and humidity) are responsible for less number of COVID cases in tropical countries [42]. Thus, we are inclined to say that environmental factors directly or indirectly play a role in explaining the observed pattern of spread in cold countries during the initial stages. Since in winters mostly people stay indoors, thus climate can indirectly be involved in the increased growth rate of the virus in cold countries. However, if the climate is indirectly responsible for the spread of the disease, it would have been limited in summers in tropical countries. On the contrary, the COVID-19 outbreak in the tropical countries started during summers, soon after the emergence of G2 strains.

Our results depict that in the initial stages of the pandemic (until March 2020), the G1 group was restricted to temperate climate (Köppen's *C* type) whereas the G2 group spread worldwide. Our analyses suggest that the same regions were affected by the G1 variant group and SARS-CoV-1. It is well known that SARS-CoV-1 could not survive in regions with warm climate as its lipid bilayer was prone to degradation at higher temperature [43], and it loses its viability at higher temperature [4]. A recent study shows the structural stability of SARS-CoV-2 Virus-Like Particles (VLPs) degrades on increasing temperature [44], supporting our observations.

Our analyses favor that evolution of G1 into G2 helps to sustain this virus from temperate to cold and tropical climate successively, mainly due to four mutations, i.e., in leader sequence, ORF1ab, and *S* gene. The leader sequence and ORF1ab are involved in replication and transcription, and the *S* gene is involved in binding to the host cell through hACE2 receptors. Substitutions in the ORF1ab gene may increase the synthesis of the replicase-transcriptase complex, thus, increasing the replication rate of the virus and blocking the host's innate immune response. 614 position in spike glycoprotein lies near the S1/S2 subunit junction where the furin-cleavage site is present (R667) that enhances virion cell-cell fusion [45]. This suggests aspartate to glycine substitution in the vicinity of the furin-recognition site may result in a conformational change of the spike glycoprotein that favors a higher affinity of the Receptor Binding Domain (RBD) to hACE2. A recent article showed retroviruses pseudotyped with Glycine at 614 position infected ACE2-expressing cells markedly more efficiently than those with Aspartic acid due to less S1 shedding and greater incorporation of the S protein into the pseudovirion [30]. Several studies reported D614G mutation is increasing at an alarming rate [31,32]. Few observed that this alteration correlated with increased viral loads in COVID-19 patients [31]. This is consistent with the epidemiological data showing the proportion of viruses bearing G614 is correlated to increased case-fatality rate on a country-by-country basis [33]. This substitution coevolved with substitution in the leader sequence, nsp3, and RdRp (RNA dependent RNA polymerase) proteins, suggesting these mutations allow the virus to transmit more efficiently. This explains these mutations have not emerged merely because of the founder's effect, but this virus under selection pressure has made itself more stable and infective. Also, Forster et al. (2020) observed in their phylogenetic analysis the preferential geographical spread of SARS-CoV-2 and provided a plausible cause which could be founders effect or immunological or environmental effect [*41*]. Although there is a possibility that the stable variant might have appeared because of the host's innate immune response or some unknown reason, in such a case, it would not show any close association with climate. Through our analyses, we are inclined to say that climate affects SARS-CoV-2 evolution. However, in particular, the selective pressure of climate on each gene of SARS-CoV-2 is not visible. Our genomic analysis of virus strains shows that the novel coronavirus undergoes both synonymous as well as non-synonymous mutations throughout its genome in various climates, suggesting the novel coronavirus uses multiple mechanisms both at the transcriptional and translational level for evading the immune response, developing drug resistance, and increasing pathogenesis. However, the actual role of these mutations is not yet determined, and these studies need to be further enlightened by biophysical and biochemical studies. Such mutational insights will aid in designing efficacious vaccines that

can be stored and transported in a wide range of temperatures and conditions, thereby minimizing cold storage costs.

To delineate the signatures of underlying abiotic factors (temperature, precipitation, and latitude) responsible for the evolution of SARS-CoV-2 (*n* = 176), spreading patterns of G1 and G2 strains were carefully examined on Köppen-Geiger map. Fig. 3 shows an elevated spread of COVID-19 in the western and eastern coasts of the continents and a diminished spread in the hot and cold deserts. The G1 strains are majorly present in the eastern and western coasts of the continents, and G2 strains lie in both the coastal regions and the continent's interior. On a closer inspection, the eastern coasts of continents consist of "*humid-subtropical*" *(Cfa)* climate while the western coasts of continents consist of "*marine-temperate*" *(Cfb)*, commonly known as east and west coast climate, respectively. These two climates are very similar to each other and belong to the temperate climate, also known as *C* type climate of Köppen-Geiger classification scheme. A substantial portion (~94%) of habitable China consists of temperate climate *(C)*, i.e., "*humid-subtropical*" *(Cfa)* climate, which explains the presence of only G1 strains in China and one strain of G1 is present in cold climate *(D)* present near the transition of temperate *(C)* to cold climate *(D)*, thus probably temperate climate was suitable for G1. A similar association of G1 with a temperate climate *(C)* was found in the eastern and western coast of North America, the eastern coast of South America, the western coast of Europe, and the eastern and western coast of Australia. Statistically, the distribution of G1 strains all over the globe is in concordance with the temperate climate and strongly favors *C* climate (Chi-square test, *P* < .01 for the null hypothesis that G1 strains are equally distributed in all climates, Supplementary Table S6) as compared to any other climate. If the climate does not have any role in the evolution and preferential spread of coronavirus, in such a case, G1 would have been evenly distributed in all climate types, which is not the case. Few exceptions of G1 seen in other climate types are most probably because of travel as they remained subsided in that climate, implying their inability to sustain in different climate types. It appears that the G1 strains existed in temperate climates all over the world but could not extend their geographical territories beyond temperate climates. Contrastingly, the evolved G2 strains can sustain in temperate *(C)*, cold *(D)*, and tropical *(A)* climate surpassing the climatic restrictions of G1. Map interpretation suggests that G2 strains enter the continent's interior through *D* climate (e.g., North America and Russia). Temperate climate *(C)* generally grades into cold climate *(D)* and deserts *(B)* in the northern hemisphere (e.g., *C* to *D*: Europe to Russia, and the USA to Canada; *C* to *B*: China, and the USA). In the southern hemisphere, gradation of temperate climate *(C)* into tropical climate *(A)* and deserts *(B)* exists (e.g., *C* to *A*, Brazil; *C* to *B*, Australia), *C* to *A* transition is identified by virus cluster 105–115 in the phylogenetic tree. In Russia, 91.3% (21/23) of the strains belong to G2 (Fig. 3), are mainly present in the ~8500 km long and 600–1700 km wide *D* climate belt *(Dfa-Dfb-Dw)*, suggesting the G2 strains might have adapted to the *D* climate (Chi-square test, *P* < .001 for the null hypothesis that both G1 and G2 strains are equally found in *D* climate of Russia, Supplementary Table S6). Similar observations are seen for North America, South America, and Australia. The eastern and western coasts of North America have temperate climates and are connected by cold climate along the USA-Canada boundary (i.e., having "*humid-subtropical*" *(Cfa)* on the eastern coast and "*marine-temperate*" *(Cfb)* climate on the western coast) (Fig. 3). The G2 strains follow this cold climate *(Dfa-Dfb)* belt, which is ~3800 km long and ~600 to 1000 km wide. The dominance of G2 and nearly absence of the G1 population in the cold climate of North America is similar to the observations of Russia. Our analyses suggest that a fall of temperature from temperate to cold climate might have dictated the evolution of G1 into the G2 variant group.

Similarly, a change in climate from *C* to *A* probably made the strains stable in tropical regions. Overall, our analyses suggest that SARS-CoV-2 has likely evolved to sustain in different climates, thereby increasing its spread. Studies combining genetic information with climate can provide

useful information about virus evolution and possible climatic pathways during an outbreak.

Conclusion

It is reasonable to assume COVID-19 transmission pathway and evolution are influenced by climate. The phylogenetic network classified 176 SARS-CoV-2 strains into two variant groups G1 and G2. The G1 strains were habituated to *C* climate that evolved into G2 by undergoing significant mutations (C241T in leader sequence, F924 in ORF1a, P214L in ORF1b, and D614G in *S* gene), plausibly extended its climatic boundaries from *C* to *D* climate, displaying the role of natural selection on virus evolution. In our analysis, SARS-CoV-2 was found resistive to desert climate *(B)*. Gradually, strains are adapting to *A* climate in South America. The strains adapted to the *"tropical-savannah" (Aw)* climate are a threat to all the tropical countries, which were initially less affected by COVID-19. Nevertheless, due to the uncertainty of COVID-19 data, the results should be carefully interpreted and should not be extrapolated to climate types and climatic conditions other than those analyzed here for the early evolution period. The study agrees that viruses are sensitive to their environment and respond towards naturally occurring abiotic factors such as temperature, latitude, and humidity to sustain in different climates of the Earth, which also provides insights about seasonal variations possibly being a strong reason for the spread of other viral diseases as well. Here we showed a more refined description of genes based on phylogenetics and their distribution across different climates. These finer-grained analyses led to highly relevant insights on the evolutionary dynamics of the poorly understood SARS-CoV-2 genome that provides vital information about the direction of the spread and highlight vulnerable regions of the Earth. Such interdisciplinary studies will play an imperative role in designing antiviral strategies and taking pre-emptive precautionary measures to combat such a pandemic.

Potential caveats

We acknowledge that there are few caveats due to uncertainty in the COVID-19 data. The data from the tropical regions is limited because, at the time of data collection (SARS-CoV-2 strains) from all over the world, the strains from the tropical countries were very limited, from few tropical regions strains were available (e.g., Ghana, India, Mexico, Nepal, Pakistan) but the data has been discarded due to the travel history of the strains, a large fraction of strains without travel history have large gaps in genomic sequences which were not suitable for the present study. Also, the case history of each patient is not reported in the METADATA file as collecting all information from each patient is time-consuming. Hence, there are chances patients from whom these strains were isolated may have a migratory history. Data from different individual locations without travel history and large gaps in genomic sequences have been incorporated in this study. Due to these reasons, our analyses should be carefully interpreted and should not be extrapolated to climate types and climatic conditions other than those analyzed here for the early evolution period.

Declaration of Competing Interest

The authors declare no conflict of interest.

CRediT authorship contribution statement

Priyanka Bajaj: Conceptualization, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **Prakash Chandra Arya:** Conceptualization, Data curation, Formal analysis, Data Analysis, Writing - original draft, Writing - review & editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecochg.2021.100005.

References

[1] Á. Briz-Redón, Á. Serrano-Aroca, The effect of climate on the spread of the COVID-19 pandemic: a review of findings, and statistical and modelling techniques, Prog. Phys. Geogr. (2020), doi:10.1177/0309133320946302.

[2] WHOCoronavirus Disease (COVID-19), WHO, 2020.

[3] B. Vellingiri, et al., COVID-19: a promising cure for the global panic, Sci. Total Environ. (2020), doi:10.1016/j.scitotenv.2020.138277.

[4] K.H. Chan, et al., The effects of temperature and relative humidity on the viability of the SARS coronavirus, Adv. Virol. (2011), doi:10.1155/2011/734690.

[5] A.W.H Chin, et al., Stability of SARS-CoV-2 in different environmental conditions, Lancet Microbe (2020), doi:10.1016/s2666-5247(20)30003-3.

[6] M.B. Araújo, F. Mestre, B. Naimi, Ecological and epidemiological models are both useful for SARS-CoV-2, Nat. Ecol. Evol. (2020), doi:10.1038/s41559-020-1246-y.

[7] C.J. Carlson, J.D. Chipperfield, B.M. Benito, R.J. Telford, R.B O'Hara, Species distribution models are inappropriate for COVID-19, Nat. Ecol. Evol. (2020), doi:10.1038/s41559-020-1212-8.

[8] P. Chandra Arya, P. Bajaj, Mini-review evolutionary studies combined with species distribution models are useful to understand COVID-19 spread and aid vaccine design, J. Vaccines Vaccin. 11 (2020) 1000438.

[9] V. Guernier, M.E. Hochberg, J.F. Guégan, Ecology drives the worldwide distribution of human diseases, PLoS Biol. 2 (2004) 740–746.

[10] A.S. Lauring, R. Andino, Quasispecies theory and the behavior of RNA viruses, PLoS Pathog. (2010), doi:10.1371/journal.ppat.1001005.

[11] E. Racevska, Natural selection, in: J. Vonk, T Shackelford (Eds.), Encyclopedia of Animal Cognition and Behavior, Springer International Publishing, 2018, pp. 1–14, doi:10.1007/978-3-319-47829-6_542-1.

[12] A.R. Fehr, S. Perlman, Coronaviruses: an overview of their replication and pathogenesis, Coronaviruses: Methods and Protocols, 1–23, Springer, New York, 2015.

[13] C.K. Li, et al., T cell responses to whole SARS coronavirus in humans, J. Immunol. 181 (2008) 5490–5500.

[14] F. Li, W. Li, M. Farzan, S.C. Harrison, Structural biology: structure of SARS coronavirus spike receptor-binding domain complexed with receptor, Science 80 (309) (2005) 1864–1868.

[15] J. Liu, et al., The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes, J. Infect. Dis. 202 (2010) 1171–1180.

[16] T.R. Ruch, C.E. Machamer, The coronavirus E protein: assembly and beyond, Viruses (2012), doi:10.3390/v4030363.

[17] C.K. Chang, M.H. Hou, C.F. Chang, C.D. Hsiao, T.H. Huang, The SARS coronavirus nucleocapsid protein - forms and functions, Antivir. Res. (2014), doi:10.1016/j.antiviral.2013.12.009.

[18] M. Allaby, Atmosphere. A scientific history of air, weather, and climate, Adv. Space Res. (2009).

[19] M.C. Peel, B.L. Finlayson, T.A. McMahon, Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci. 11 (2007) 1633–1644.

[20] F. Wu, et al., A new coronavirus associated with human respiratory disease in China, Nature (2020), doi:10.1038/s41586-020-2008-3.

[21] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, X. MEGA, Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.

[22] M. Nei, N. Saitou, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. (1987) 406–425, doi:10.1093/oxfordjournals.molbev.a040454.

[23] K. Tamura, M. Nei, S. Kumar, Prospects for inferring very large phylogenies by using the neighbor-joining method, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 11030–11035.

[24] EarthExplorer.

[25] Travel Weather Averages (Weatherbase).

[26] Climate data for cities worldwide - Climate-Data.org.

[27] M. Herbei, V. Ciolac, A. Smuleac, L. Ciolac, Georeferencing of topographical maps using the software ArcGIS, Res. J. Agric. Sci. 42 (2010) 595–606.

[28] B.H. Harcourt, et al., Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity, J. Virol. (2004), doi:10.1128/jvi.78.24.13600-13612.2004.

[29] E.J. Snijder, E. Decroly, J. Ziebuhr, The nonstructural proteins directing coronavirus RNA synthesis and processing, Adv. Virus Res. (2016), doi:10.1016/bs.aivir.2016.08.008.

[30] L. Zhang, C.B. Jackson, H. Mou, et al., SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity, Nat Commun 11 (2020) 6013, doi:10.1038/s41467-020-19808-4.

[31] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI; Sheffield COVID-19 Genomics Group, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020 Aug 20;182(4):812–827.e19. doi: 10.1016/j.cell.2020.06.043. Epub 2020 Jul 3. PMID: 32697968; PMCID: PMC7332439.

[32] I.J. Morais, R.C. Polveiro, G.M. Souza, et al., The global population of SARS-CoV-2 is composed of six major subtypes, Sci Rep 10 (2020) 18289, doi:10.1038/s41598-020-74050-8.

[33] M. Becerra-Flores, T. Cardozo, SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate, Int. J. Clin. Pract. (2020), doi:10.1111/ijcp.13525.

[34] WHOSARS (Severe Acute Respiratory Syndrome), WHO, 2020 Available at https://www.who.int/ith/diseases/sars/en/ Accessed: 10th September.

[35] G. Boël, et al., Codon influence on protein expression in E. coli correlates with mRNA levels, Nature 529 (2016) 358–363.

[36] C.J. Tsai, et al., Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima, J. Mol. Biol. 383 (2008) 281–291.

[37] S.A. Shabalina, A.Y. Ogurtsov, N.A. Spiridonov, A periodic pattern of mRNA secondary structure created by the genetic code, Nucleic Acids Res. (2006), doi:10.1093/nar/gkl287.

[38] J. Jia, et al., Modeling the control of COVID-19: impact of policy interventions and meteorological factors, Electron. J. Differ. Equ. (2020).

[39] A. Meyer, R. Sadler, C. Faverjon, A.R. Cameron, M. Bannister-Tyrrell, Evidence That Higher Temperatures Are Associated With a Marginally Lower Incidence of COVID-19 Cases, Frontiers in public health 8 (2020) 367.

[40] Q. Nie, et al., Phylogenetic and phylodynamic analyses of SARS-CoV-2, Virus Res. (2020), doi:10.1016/j.virusres.2020.198098.

[41] P. Forster, L. Forster, C. Renfrew, M. Forster, Phylogenetic network analysis of SARS-CoV-2 genomes, Proc. Natl. Acad. Sci. U. S. A. (2020), doi:10.1073/pnas.2004999117.

[42] Q. Bukhari, Y. Jameel, Will coronavirus pandemic diminish by summer? SSRN Electron. J. (2020), doi:10.2139/ssrn.3556998.

[43] M.J.B Raamsman, et al., Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E, J. Virol. (2000), doi:10.1128/jvi.74.5.2333-2342.2000.

[44] A. Sharma, B. Preece, H. Swann, X. Fan, R.J. McKenney, K.M. Ori-McKenney, S. Saffarian, M.D. Vershinin, Structural stability of SARS-CoV-2 virus like particles degrades with temperature, Biochemical and biophysical research communications 534 (2021) 343–346.

[45] K.E. Follis, J. York, J.H. Nunberg, Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry, Virology 350 (2006) 358–369.

[46] Fehr, Coronaviruses: An Overview of Their Replication and Pathogenesis, Methods in Molecular Biology 1282 (2015) https://link.springer.com/content/pdf/10.1007%2F978-1-4939-2438-7.pdf, doi:10.1007/978-1-4939-2438-7_1.