



OPEN ACCESS

EDITED BY

Xin Yin,
Chinese Academy of Agricultural
Sciences (CAAS), China

REVIEWED BY

Dharmendra Kumar Yadav,
Gachon University, South Korea
Jinzhao Song,
University of Chinese Academy of
Sciences, China

*CORRESPONDENCE

Naveed Iqbal
Naveed.iqbal@buitms.edu.pk

SPECIALTY SECTION

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

RECEIVED 17 April 2022

ACCEPTED 27 June 2022

PUBLISHED 04 August 2022

CITATION

Iqbal N, Rafiq M, Masooma, Tareen S,
Ahmad M, Nawaz F, Khan S, Riaz R,
Yang T, Fatima A, Jamal M, Mansoor S,
Liu X and Ahmed N (2022) The
SARS-CoV-2 differential genomic
adaptation in response to varying
UVindex reveals potential genomic
resources for better COVID-19
diagnosis and prevention.
Front. Microbiol. 13:922393.
doi: 10.3389/fmicb.2022.922393

COPYRIGHT

© 2022 Iqbal, Rafiq, Masooma, Tareen,
Ahmad, Nawaz, Khan, Riaz, Yang,
Fatima, Jamal, Mansoor, Liu and
Ahmed. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

The SARS-CoV-2 differential genomic adaptation in response to varying UVindex reveals potential genomic resources for better COVID-19 diagnosis and prevention

Naveed Iqbal^{1*}, Muhammad Rafiq¹, Masooma¹,
Sanaullah Tareen¹, Maqsood Ahmad¹, Faheem Nawaz¹,
Sumair Khan¹, Rida Riaz², Ting Yang³, Ambrin Fatima⁴,
Muhsin Jamal⁵, Shahid Mansoor⁶, Xin Liu³ and Nazeer Ahmed¹

¹Faculty of Life Sciences and Informatics, Baluchistan University of Information Technology, Engineering and Management Sciences (BUIITEMS), Quetta, Pakistan, ²Department of Microbiology, Quaid i Azam University, Islamabad, Pakistan, ³Beijing Genomic Institute (BGI), Shenzhen, China, ⁴Department of Biological and Biomedical Sciences, Aga Khan University, Karachi, Pakistan, ⁵Department of Microbiology, Abdul Wali Khan University Mardan, Mardan, Pakistan, ⁶Agriculture Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan

Coronavirus disease 2019 (COVID-19) has been a pandemic disease reported in almost every country and causes life-threatening, severe respiratory symptoms. Recent studies showed that various environmental selection pressures challenge the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infectivity and, in response, the virus engenders new mutations, leading to the emergence of more virulent strains of WHO concern. Advance prediction of the forthcoming virulent SARS-CoV-2 strains in response to the principal environmental selection pressures like temperature and solar UV radiation is indispensable to overcome COVID-19. To discover the UV-solar radiation-driven genomic adaptation of SARS-CoV-2, a curated dataset of 2,500 full-grade genomes from five different UVindex regions (25 countries) was subjected to in-depth downstream genome-wide analysis. The recurrent variants that best respond to UV-solar radiations were extracted and extensively annotated to determine their possible effects and impacts on gene functions. This study revealed 515 recurrent single nucleotide variants (rcntSNVs) as SARS-CoV-2 genomic responses to UV-solar radiation, of which 380 were found to be distinct. For all discovered rcntSNVs, 596 functional effects (rcntEfts) were detected, containing 290 missense, 194 synonymous, 81 regulatory, and 31 in the intergenic region. The highest counts of missense rcntSNVs in spike (27) and nucleocapsid (26) genes explain the SARS-CoV-2 genomic adjustment to escape immunity and prevent UV-induced DNA damage, respectively. Among all, the most commonly observed rcntEfts were four missenses (RdRp-Pro327Leu, N-Arg203Lys, N-Gly204Arg, and Spike-Asp614Gly) and one synonymous (ORF1ab-Phe924Phe) functional effects. The

highest number of rcntSNVs found distinct and were uniquely attributed to the specific UVindex regions, proposing solar-UV radiation as one of the driving forces for SARS-CoV-2 differential genomic adaptation. The phylogenetic relationship indicated the high UVindex region populating SARS-CoV-2 as the recent progenitor of all included samples. Altogether, these results provide baseline genomic data that may need to be included for preparing UVindex region-specific future diagnostic and vaccine formulations.

KEYWORDS

SARS COVID-19, genomic adaptation, UV-solar radiation, COVID diagnosis, comparative genomics

Introduction

In December 2019, clusters of pneumonia cases were reported from the Wuhan city, Hubei province, China. Some of the early disease cases were reported working in the live animal market. On 11 March 2020, the WHO announced the disease outbreak, now named coronavirus diseases 2019 (COVID-19), as a public health emergency of international concern and declared it a pandemic (Koyama et al., 2020). As of June 2022, $\sim >528.82$ million positive cases were reported to WHO across the world [WHO Coronavirus (COVID-19) Dashboard, 2022], with more than 6.29 million deaths. The COVID-19 symptoms range from mild fever, cough and fatigue to severe shortness of breath, and loss of taste and smell (Guan, 2020; Wang D. et al., 2020), with the 5% average fatality rate of all confirmed positive cases, which is of lower than SARS-CoV (9.6%) and MERS (34.3%) (World Health Organization., 2003, 2019; Wang C. et al., 2020).

After the preliminary etiological investigations based on the exclusion of all common respiratory pathogens, the deep meta-transcriptomic sequencing of the patient's bronchoalveolar lavage fluid revealed the abundance of a viral strain from β -coronavirus (CoV) genus (Shi et al., 2016, 2018; McMullan et al., 2019; Yadav et al., 2019; Abdelrahman et al., 2020; Wu et al., 2020). The COVID-19-causing virus showed 89.1%, 79.5%, and 50% sequence homology to previously reported SARS-like coronavirus strains, namely, bat SL-CoVZC45, SARS-CoV, and MERS, respectively (Wang et al., 2015; Wu et al., 2020). Based on the sequence homology to SARS-like viruses, the crown-like viral structure, and the consequent manifestation of severe respiratory disease symptoms, the COVID-19-causing virus is designated as SARS-CoV-2 (severe acute respiratory syndrome coronavirus-2) (Lu et al., 2020; Wu et al., 2020). Furthermore, most SARS-like coronaviruses have been identified in bats (Hamre and Procknow, 1966; McIntosh et al., 1967; Li et al., 2005), and the SARS-CoV-2 shares 100% amino-acid sequence similarity with NSP7 and E protein of the bat SARS-like coronavirus strain (bat SL-CoVZC45) (Wu et al., 2020). These

findings suggest that bats are the possible natural reservoirs for most SL-CoVs, including SARS-CoV-2.

The SARS-CoV-2 genomic characterization revealed 29,903 nucleotide long single-stranded positive-sense RNA (ribonucleic acids) comprising a multi-domain nonstructural protein (NSP) encoding ORF1ab, four structural protein genes (spike "S," envelope "E," membrane "M," and nucleocapsid "N"), and six accessory protein-encoding genes (ORF2a, ORF6, ORF7a, ORF7b, ORF8, and ORF10) (Koyama et al., 2020). The SARS-CoV-2 was found capitalizing its spike structural protein for host cell (respiratory epithelial) attachment and subsequent entries *via* the angiotensin-converting enzyme 2 (ACE2) receptor (Hoffmann et al., 2020).

Since December 2019, whole-genome sequence analysis revealed hundreds of viable genetic variants of SARS-CoV-2 from different parts of the globe. Within SARS-CoV-2, the observed predominating drivers of genetic variation are the single-nucleotide variants (SNVs) caused by the error-prone viral polymerases (Smertina et al., 2019; Lu et al., 2022) and endogenous mutagenesis *via* the host RNA-editing enzymes (nucleotide deaminases APOBEC: C>U and ADAR: A>G) (Placido et al., 2007; Moris et al., 2014; Mourier et al., 2021; Tong et al., 2022). The genome-wide studies of large sets of SARS-CoV-2 revealed SNV-based nucleotide substitution rate of $\sim 1 \times 10^{-3}$ per year (Duchene et al., 2022), closer to the 1.42×10^{-3} Ebola virus substitution rate reported from West Africa during 2013–2016. However, SNVs are not the only genetic variations discovered in coronaviruses, but the small insertions/deletions of viral or non-viral sequences were also reported in various genetic variants of coronavirus genomes possibly caused by the discontinuous nature of viral transcriptase for sub-genomic mRNA synthesis (Licitra et al., 2013; V'kovski et al., 2021). In total, a large proportion of the mutations represent neutral "genetic drift" or have died out quickly, and a small subset is affecting viable viral traits, such as host range, transmissibility, antigenicity, pathogenicity, and adaptability of the virus to various selection pressures.

Various biotic and abiotic selection pressures challenge the SARS-CoV-2 persistence, transmission, infectivity, host cell entry efficacy, and pathogenesis (Pica and Bouvier, 2012). Since RNA viruses, *via* high mutation rate, have demonstrated a great potential for rapid evolution and adaptation to new environmental conditions in the absence of a proper proofreading RNA polymerases activity (Holland et al., 1982; Rubio et al., 2013). Therefore, to escape stress conditions, the coronaviruses continuously engender new genomic variations, potentially resulting in the emergence of more virulent SARS-CoV-2 strains of WHO concern with higher transmission and mortality rates (Sanjuán and Domingo-Calap, 2016; Chin et al., 2020; Koyama et al., 2020; Seyer and Sanlidag, 2020; Kumar et al., 2021; Soh et al., 2021). The commonly experienced biotic selection pressures in human hosts may include natural immunity (Clapham et al., 2020), host genetic makeup (COVID-19 Host Genetics Initiative, 2021), monoclonal antibodies produced in response to vaccines (Rella et al., 2021; Shah et al., 2021), antiviral drugs, and convalescent sera, whereas solar radiation (Chiyomaru and Takemoto, 2020) (ultraviolet radiations) (Seyer and Sanlidag, 2020), temperature (Chin et al., 2020; Wang J. et al., 2020), relative humidity (Ahlawat et al., 2020; Ghoushchi et al., 2020), and air pollutants (Coccia, 2020) are the widely studied abiotic selection pressures on viral populations (Tan et al., 2005; Shaman et al., 2010; Otter et al., 2016; Chattopadhyay et al., 2018; Dalziel et al., 2018; Gardner et al., 2019). Studies revealed a negative correlation between the environmental conditions (temperature and humidity) and the H3N2 strain of the influenza flu virus (Lowen et al., 2007; Reich et al., 2019). Additionally, ultraviolet radiation imposed negative selection pressure on strains of influenza and related coronaviruses (Darnell et al., 2004), and more recently, Ratnesar-Shumate et al. showed that the UV-solar radiation induced SARS-CoV-2 nucleic-acid damage and subsequent viral inactivation (Ratnesar-Shumate et al., 2020).

Predicting genomic level adaptation of SARS-CoV-2 in response to various selection pressures is indispensable in understanding the viral spread, mutation, pathogenicity, control, and future treatment options to effectively tackle COVID-19 (O'Reilly et al., 2020). Solar ultraviolet radiation is thought to have a great impact on the formation of viral populations by selecting variants that can withstand UV-solar radiations (Ratnesar-Shumate et al., 2020). In this study, to investigate the SARS-CoV-2 genomic adaptation in response to UV solar radiation, we analyzed 2,500 high-quality, full-length genomes from five different WHO's defined UVindex regions. The comparative genome-wide analysis of SARS-CoV-2 populations revealed differential genomic adjustments in response to different ultraviolet solar radiations. All identified differential genomic signatures in response to various UVindex ranges provide baseline data for future more effective molecular COVID-19 diagnosis and region-specific vaccine production against COVID-19.

Methods

Sampling

In this study, to reveal the genomic adaptation of SARS-CoV-2 in response to UV-radiation, all COVID-19 experienced countries, which have uploaded at least 100 full-length, high-quality SARS-CoV-2 genomes, are included. Based on the WHO and US-EPA ultraviolet (UV) radiation exposure categories (Table 1), all included countries are divided into the following five groups according to their respective ultraviolet index (UVindex) records (World Health Organization, 2002; Fioletov et al., 2004). Low UVindex countries (UVindex range: <2), Moderate UVindex countries (UVindex range: 3–5), High UVindex countries (UVindex range: 6 to 7), Very_High UVindex countries (UVindex range: 8–10), and Extreme UVindex countries (UVindex range: >11).

UVindex mean data for 12 months (from 7 December 2020 to 8 December 2021) for all included countries were obtained from the monthly weather forecast and climate by WeatherAtlas (retrieved on 08 December 2021, at 15:30 GMT/UTC + 5h; <https://www.weather-atlas.com/>). The UVindex value for each country was presented as a single value rounded to the nearest whole number. For each category, irrespective of the country's geographical location, the most relevant (top of the category's list) five countries were selected provided that the country experiencing UVindex falls in the specified category range and must have at least 100 full-length, high-quality genome sequences reported in publicly accessible databases (Supplementary Table 1). Initially, for all UVindex categories, the all available (total of 8,631) full-length SARS-CoV-2 genomes were downloaded from GISAID on 11 December 2021, GenBank on 15 December 2021, the Chinese National Genomics Data Center Genome Warehouse on 23 December 2021, and the Chinese National Microbiology Data Center on 23 December 2021 (Benson et al., 2012; Shu and McCauley, 2017; CNCB-NGDC Members and Partners, 2021). To process high-quality, full-length genomes in each of the UVindex category, downloaded sequences shorter than 29,700

TABLE 1 Ultraviolet radiation exposure categories by WHO UVindex guide.

Exposure categories	UVindex range
Low	≤2
Moderate	3–5
High	6–7
Very high	8–10
Extreme	≥11

Each row is filled with different shade using the WHO assigned color for each UVindex range.

bps and containing seven consecutive ambiguous nucleotides (NNNs) were excluded from the downstream analysis. The China National Center for Bioinformatics annotations was used to remove redundancy (Gong et al., 2020). Downloaded sequences containing 50 ambiguous bases were removed from the downstream analysis to reduce the number of false-positive variants using Trimmomatic version 0.39 (Bolger et al., 2014). Finally, using the accustomed Perl script, a 100 high-quality genome sequences from each of the five included countries in a UVindex category were randomly selected, so in a nutshell, for all five UVindex categories, 2,500 full-length SARS-CoV-2 reported genomes were retained for analysis.

Reference genome

The SARS-CoV-2 (NC_045512.2) sequence was used as a reference genome in this study. The NC_045512.2 was sequenced in December 2019 from a sample recovered from Wuhan, China (Wu et al., 2020). According to the standard procedure for variant detection (DePristo et al., 2011), to retrieve high-quality variants, first, each sample was converted to short FastQ reads using emboss-splitter (Rice et al., 2000) and an accustomed fasta-to-fastq.pl script available in GitHub (Dabbish et al., 2012).

Read mapping

High-quality reads from each sample were mapped to the latest available reference SARS-CoV-2 genome NC_045512.2 using the BWA-MEM algorithm with the default minimum seed length of 20, gap open penalty 6, gap extension penalty 1, and matching score 1 (Li, 2013). For variant identification and downstream processing, open-source software packages were used. The “RealignerTargetCreator” and “InDelRealigner” command-line tools of the Genome Analysis Toolkit (GATK version 3.3.0) were used to fix all mapping issues through locally realigning improperly mapped reads, possessing variant artifacts at their terminals (McKenna et al., 2010). Before calling variants, Picard, Samtools, and BWA were used to generate the reference and bam file indexes (Li and Durbin, 2009; Li et al., 2009; McKenna et al., 2010; DePristo et al., 2011).

Variant calling and quality filtration

Any deviation of the properly mapped read sequence to the reference genome NC_045512.2 was called as a variation. For variant discovery, initially, the “mpileup” utility of bcftools, with default parameters, was used to call genotypes for each of the samples included in this study. From the derived genotypes, high-quality variants were identified as any deviation of the

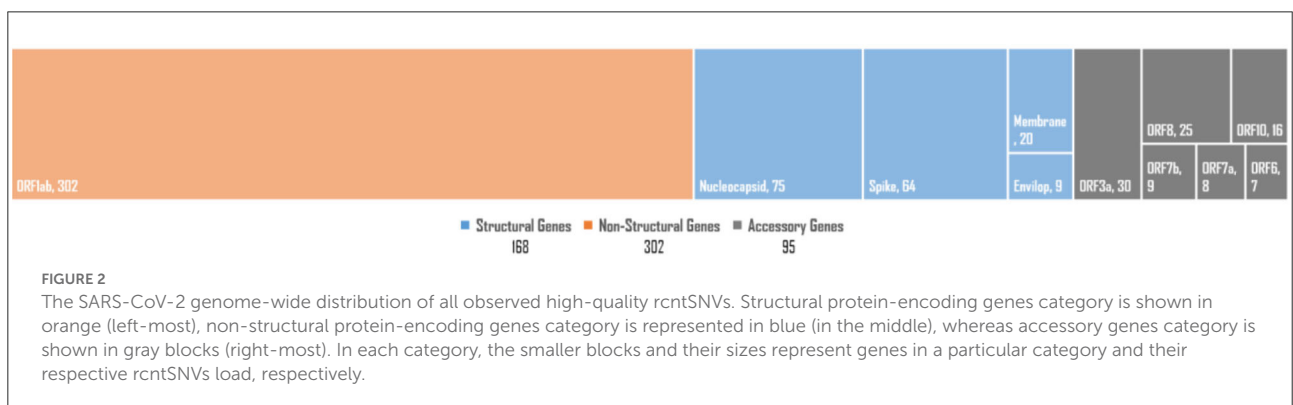
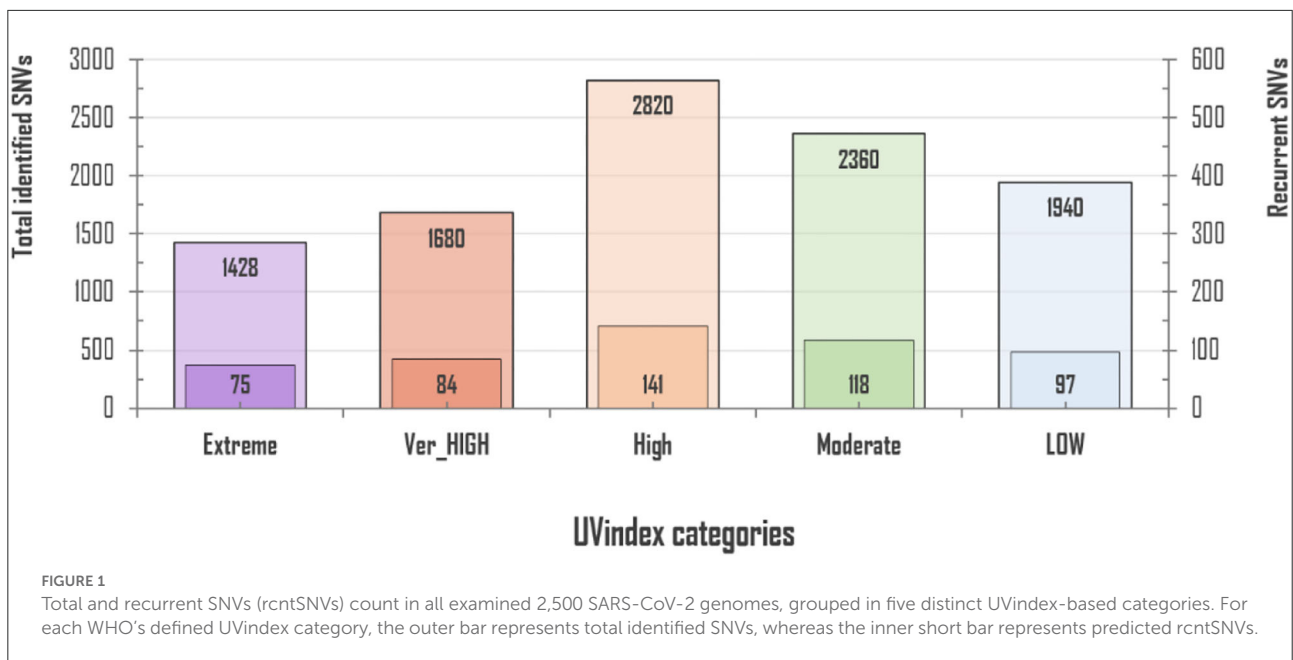
mapped read sequences from the reference genome using the bcftools “call” command (Li, 2011). To differentiate between real hereditary variants from the false-positive data-processing artifacts (caused ambiguous bases), a calibrated statistical likelihood was generated for each of the identified variant loci using the GATK “Variant Recalibrator” and “ApplyRecalibrator” functions (McKenna et al., 2010). Finally, false-positive data-processing artifacts were removed using the following options of bcftools filter and GATK variant filtration; (a) variants were removed with a Phred quality score ≤ 20 ; (b) since Fisher’s exact test-based Phred-scaled *P*-value (FS) represents strand bias for the reference and alternative allele, a sign for the false-positive variant. Therefore, variants with FS values > 60 were filtered out from the downstream analysis (Kim et al., 2017; Iqbal et al., 2019).

Variant functional annotation and prioritization

After filtration, high-quality variants were retained for each of the UVindex categories. Furthermore, high-quality variants to predict possible variant functional effects, impact, and their respective distribution across the reference NC_045512.2 genome were comprehensively investigated. The SnpEff_4.3 was used to attribute each variant by a functional class and offered various annotation levels to identify potential coding variants. For functional annotation, the SnpEff database was developed according to the SnpEff database building protocol (Cingolani et al., 2012) using the NCBI SARS-CoV-2 sequence annotation resources (NC_045512.2; Bio-Project, PRJNA485481; <https://www.ncbi.nlm.nih.gov/sars-cov-2/>). For all potential coding variants, the assigned SnpEff functional class vocabularies were UTR 3 prime, UTR 5 prime, splice site donor, splice site acceptor, splice site region, downstream, upstream, disruptive in-frame deletion and insertion, and conserved in-frame insertion and deletion. The results are provided in the list of functionally annotated variants (Supplementary Material: rcntSNV_UVindex.snpEff.vcf). A customized script was developed in Python to extract all identified variants for each of the genes in all UVindex categories (Supplementary Material: rcntSNVs_genes_functional_effects_UV.Case.genes). Following variant functional annotation, all coding region variants were compared to find UVindex category-specific and overlapping variants using vcftools (Danecek et al., 2011), the bioinformatics, and evolutionary genomics resources (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

Phylogeny

For phylogeny, sequences were precisely chosen with < 30 variations, and the lengths were adjusted by 5′ UTR and 3′

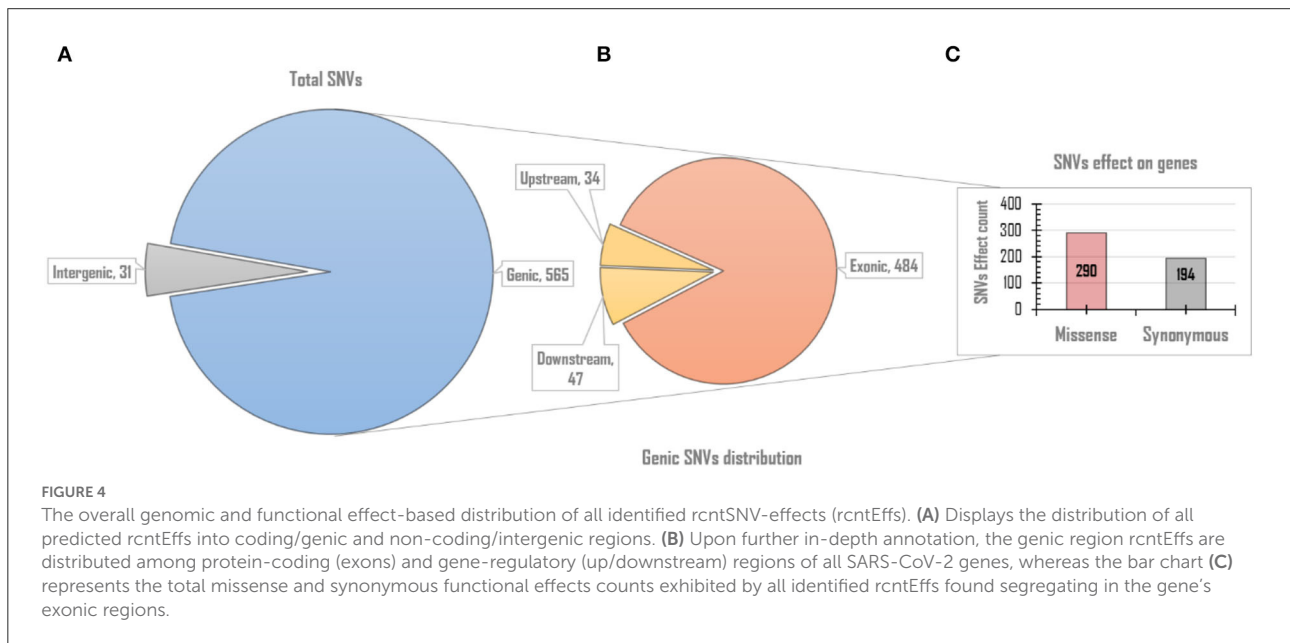
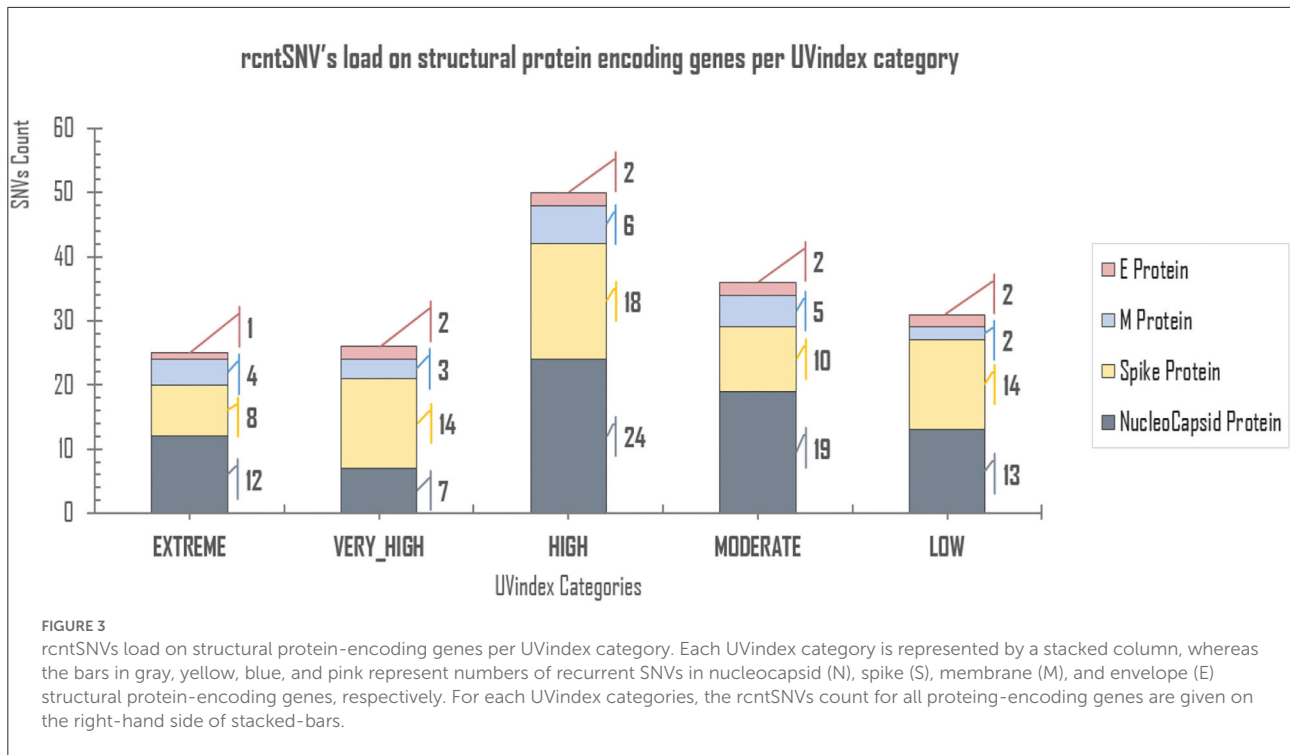


UTR truncation, without losing the key sequence sites. From this sequence pool, for an optimal phylogenetic relationship, a subset of 125 high-quality SARS-CoV-2 whole-genome samples (25 from each of the UVindex category) randomly selected in Perl by using a random number generator. All selected genomes were first aligned using the progressive multiple sequence alignment method of ClustalW (Thompson et al., 1994). The MEGA X (version 11.0.10) was used to produce and visualize the phylogenetic tree (Kumar et al., 2018). The maximum likelihood approach with Tamura-Nei substitution model, uniform rates among sites, all sites' data treatment, 1,000 bootstrap value, and nearest neighbor interchange (NNI) heuristic method was used for the best interfacing of a tree.

Results and discussions

To determine the differential genomic adaptation of SARS-CoV-2 in response to different UVindex ranges, 2,500 full-length, high-quality reported genomes were investigated from

25 countries, classified into five distinct categories based on the country's UVindex exposures. UVindex-based categories are described in the "Methods" section (Table 1). A total of 500 full-grade genomes were included from each of the defined UVindex-based categories; for the Low UVindex category, genomes were obtained from Estonia, Faroe Islands, Iceland, Norway, and Sweden; for the Moderate UVindex category, genomes were retained from Kazakhstan, North Macedonia, South Korea, Spain, and Georgia, the United States; for the High UVindex category, genomes were maintained from Cyprus, Iran, Japan, New Zealand, and Florida, the United States; for the Very_High UVindex category, genomes were acquired from Bahrain, Bangladesh, Egypt, Kuwait, and Saudi Arabia; and for the Extreme UVindex category, genomes were included from Brazil, Ecuador, Singapore, Suriname, and Uganda (Supplementary_info_file.docx, Supplementary Table 1, and for geographical location, please see the map from Supplementary_map1, Supplementary Material). Accustomed



Perl script was used to randomly select 100 high-quality SARS-CoV-2 genomes from each of the included countries.

Variant discovery (Total/RcntSNVs)

For 2,500 SARS-CoV-2 complete genome samples, we discovered a total of 10,228 single nucleotide variants (SNVs)

with an average variation load of one SNV after every 15.49 nucleotides per UVindex category (averaging ~3.92 SNVs/sample). In each UVindex category, countries are included based on their commonly shared UVindex ranges, irrespective of their relative humidity, temperature, altitude, geographical location, and many other selection pressures. Considering our sampling strategy, all identified SNVs in

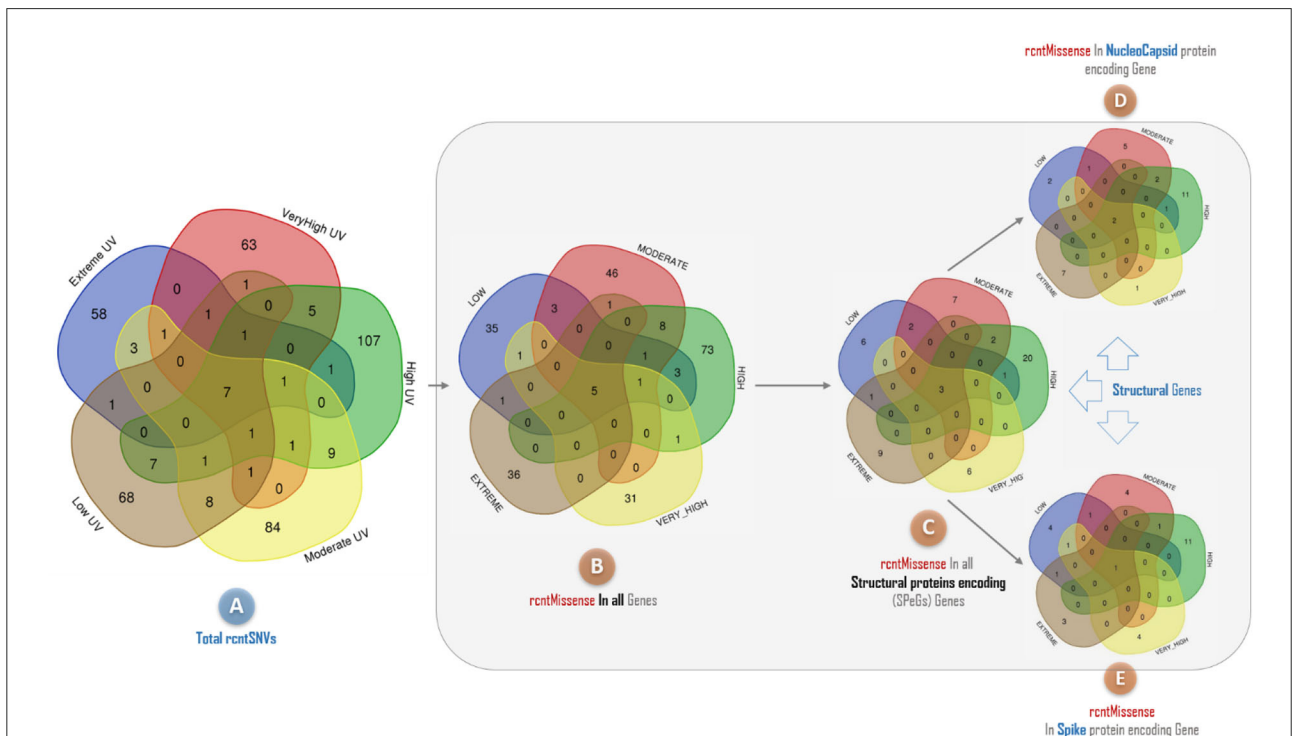
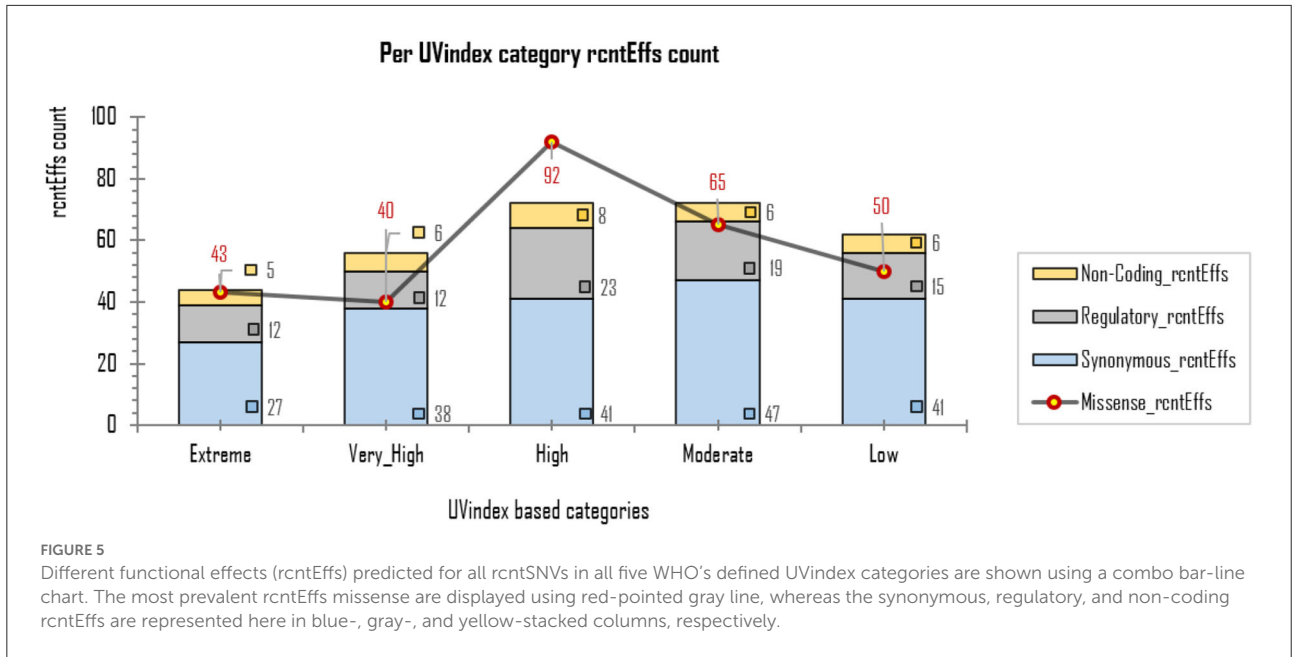


TABLE 2 Functional effects of all identified category-specific recurrent SNVs (CaSp-rcntEfts) counts identified in all 2,500 SARS-CoV-2 genomes and their respective per WHO's defined UVindex category distribution.

rcntEfts load on	Structural Genes				non-Structural Protein gene	Accessory Protein Genes						Total
	N gene	S gene	M gene	E gene	ORF1ab	ORF3a	ORF10	ORF8	ORF7b	ORF7a	ORF6	
Extreme	07	06	03	01	38	01	02	03	00	01	02	64
Very_High	02	11	02	02	37	05	01	01	02	03	02	68
High	16	15	05	02	60	02	04	07	03	04	02	120
Moderate	11	07	03	02	54	06	03	05	02	00	01	94
Low	05	10	01	02	45	07	02	02	00	00	00	74
TOTAL rcntEfts	41	49	14	09	234	21	12	18	07	08	7	420

TABLE 3 Functional effects of all identified category-specific recurrent SNVs (CaSp-rcntEfts) count across all SARS-CoV-2 genes.

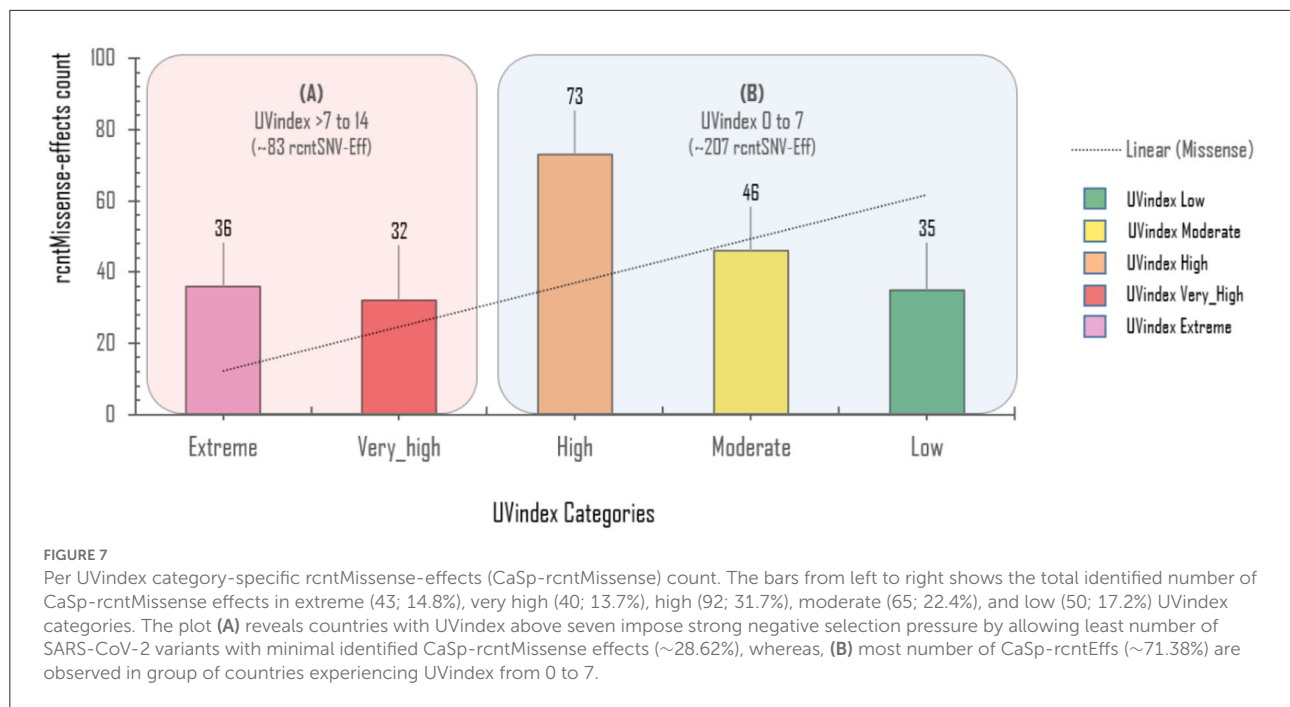
Genes groups	Genes	Missense *[Ex:Vh:Hi:Mo:Lo]	Synonymous [Silent]	Regulatory	Total	Effects per gene group
Non-structural genes	ORF1ab	141 [22:21:42:32:24]	85	08	234	234
Structural genes	E Protein	03 [0:1:1:0:1]	01	05	09	113
	M Protein	02 [1:1:0:0:0]	08	04	14	
	N Protein	26 [7:1:11:5:2]	08	07	41	
	S Protein	27 [3:5:11:4:4]	20	02	49	
Accessory genes	ORF3a	09 [1:1:1:2:4]	07	05	21	73
	ORF6	03 [1:0:1:1:0]	00	04	07	
	ORF7a	03 [0:2:1:0:0]	02	03	08	
	ORF7b	00 [0:0:0:0:0]	01	06	07	
	ORF8	07 [1:0:4:2:0]	04	07	18	
	ORF10	01 [0:0:1:0:0]	00	11	12	
Grand total		222	136	62	420	

The horizontal-rows and vertical columns represent the CaSp-rcntEfts count by functional effect classes and by gene groups, respectively.

*[Ex:Vh:Hi:Mo:Lo] denotes the UVindex-specific rcntSNV-missense effect counts in Extreme, Very_High, High, Moderate, and Low categories, respectively.

each UVindex category are the probable genomic adjustments against all experienced biotic and abiotic selection pressures, whereas only the most common SNVs in a UVindex category are the potential genomic adaptation of SARS-CoV-2 in response to UVindex. Therefore, based on a 25% reoccurrence rate in a UVindex category, a sum of 515 (5.03% of a total of 10,228) recurring SNVs were carefully prioritized to discover the SARS-CoV-2 genomic responses to a commonly experienced environmental selection pressure, the UV solar radiation. These SNVs with at least 25% reoccurrences in each UVindex category are termed recurrent-SNVs (rcntSNVs). For all UVindex categories, lists of all discovered rcntSNVs are given in Supplementary_info_file.docx [Supplementary Tables 2–6](#). Of the total, the least number of rcntSNVs (75) were observed in SARS-CoV-2 genomes included from countries exposed to Extreme UVindex solar radiation, revealing that the Extreme UVindex solar radiation employs negative selection pressure by damaging viral DNA and thus limits the diversity of

SARS-CoV-2 strains. Our finding is consistent with the hypothesis that Extreme UVindex radiations induces viral DNA damage to disinfect the SARS-CoV-2 without altering its morphology (Lo et al., 2021). Furthermore, the solar UV radiation of extreme intensity inactivates SARS-CoV-2 and other related strains of corona and influenza viruses on surfaces (Pi et al., 2003; Darnell et al., 2004; Ianevski et al., 2019; Ratnesar-Shumate et al., 2020). On the contrary, the highest number of rcntSNVs (141) was discovered in the High UVindex region, suggesting that the large majority of SARS-CoV-2 variants/strains are adapted to High UVindex solar radiation. A. Ianevski et al. also showed the highest counts for the active influenza virus strains populating High UVindex experiencing parts of Northern Europe from 2010 to 2018 (Ianevski et al., 2019). Based on these findings, we propose that COVID-19-causing viruses have had sufficient evolutionary time to acquire genomic-level adaptation in High UVindex regions, probably in their primary natural reservoir (bat). Our findings



are scientifically in line with the Li et al.'s work that found bats families, being the zoonotic origin of several SARS-like coronaviruses, greatly enriched in tropical regions experiencing High UVindex solar radiations (e.g., Guangdong, Guangxi, Hubei, and Tianjin) (Hamre and Procknow, 1966; McIntosh et al., 1967; Li et al., 2005; Wu et al., 2020). Figure 1 shows the total number of identified and rcntSNVs in each of the UVindex category.

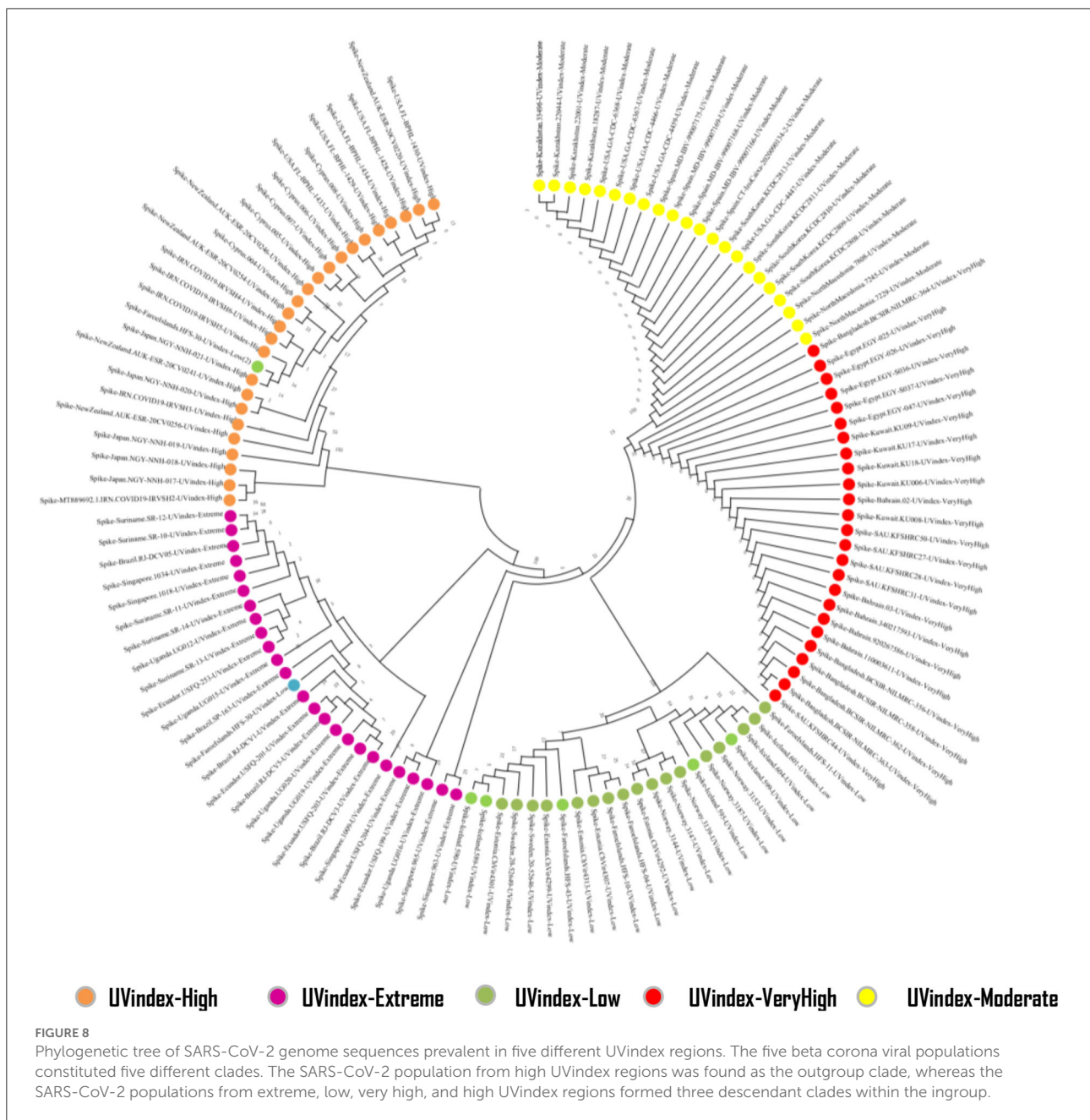
rcntSNVs genomic distribution

The SARS-CoV-2 genome exhibits two non-structural multi-domain protein-encoding genes (ORF1a and ORF1b), four structural protein-encoding genes (SPeGs; S, E, M, and N), and up to six genes that encode accessory proteins 3a, 6, 7a, 7b, 8, and 10a (Brant et al., 2021). Our in-depth analysis for gene-set-based distribution of all potentially UVindex responding variants revealed the large majority of the total rcntSNVs (302: 53.45%) in the non-structural protein-encoding genes (ORF1ab), followed by 168 (29.73%) in four SPeGs (N = 75, S = 64, M = 20, and E = 9), whereas only 95 (16.81%) were found in six accessory genes (Figure 2). These inferences are in agreement with the genomic architecture of the SARS-CoV-2 (Wu et al., 2020) and illustrate that SARS-CoV-2 has done most (approximately >53%) of the genomic-level adaptation in non-structural multi-domain protein-encoding genes (ORF1ab) to adapt various UVindex regions, where the accessory protein-encoding genes were the most conserved gene-set of SARS-CoV-2.

Of all the virion proteins, the structural gene products were directly exposed to environmental selection pressures, like solar UV radiation. Therefore, the downstream analysis was focused to identify rcntSNVs in E, M, S, and N SPeGs for each of the UVindex category (Figure 3). Of the total identified 168 structural rcntSNVs, we discovered 75, 64, 20, and 9 in nucleocapsid, spike, membrane, and envelope SPeGs, respectively. Of all four SPeGs, the nucleocapsid gene has gone through most of the genomic rearrangements, possibly to shield the nucleic acid damaging effects of UV radiation *via* adaptation in response to differential UVindex exposures. These findings support recent studies on SARS-CoV-2, revealing the adverse effects of UV radiation (UVC) on nucleic acid without affecting viral proteins (Chang et al., 2014), and the nucleocapsid protein's key role in packaging and protecting COVID-19 viral genome in a viable virion (Tahara et al., 1994, 1998; Lai and Cavanagh, 1997).

rcntSNVs functional effects

Since rcntSNVs in each of the five UVindex categories best represent differentially adapted SARS-CoV-2 populations. Therefore, all rcntSNVs were functionally annotated to predict their direct effects and impacts on the genes' functions. One SNV may have more than one effects, possibly due to the gene overlapping (Cingolani et al., 2012; Iqbal et al., 2019). As a result, slightly more rcntSNV-effects (rcntEfts) were observed compared to the total rcntSNV count. In this study, a total of 596 functional rcntEfts were discovered for all rcntSNVs.



Functional annotation revealed only 31 (5.2%) rcntEffs in the non-coding intergenic regions, and the remaining 565 (94.8%) were located in the genic regions of the SARS-CoV-2 genome. Of the total genic region rcntEffs, 81 (14.3%) were detected in the gene's regulatory regions, positioned 200 bp upstream (34 count) and downstream (47 count) of all genes, and the remaining 484 (85.7%) were found in the coding regions (exonic). These results are scientifically in line with the genomic architecture of the SARS-CoV-2, and similar results were also shown by [Koyama et al. \(2020\)](#). The overall functional rcntEffs count for all rcntSNVs and their corresponding distribution across the SARS-CoV-2 genome are shown in [Figure 4](#).

The exonic rcntEffs set comprises 290 missenses and 194 synonymous genes' functional effects. Interestingly, of the total identified rcntSNVs in all UVindex categories, the highest number of the variants are with missense functional effects (290; 48.7%), suggesting that in response to immense selection pressure imposed by varying degrees of UV radiation, the SARS-CoV-2 has capitalized on the high impact missense variation enrichment to qualify for UV radiation stress. More than 71.38% (~207) of the total missense rcntEffs are found segregating in High (92; 31.7%), Moderate (65; 22.4%), and Low (50; 17.2%) UVindex categories. Suggesting that the UVindex range \leq seven allows more SARS-CoV-2 strains to survive. On the contrary, the

UVindex \geq eight imposes strong negative selection pressure on SARS-CoV-2 as only \sim 28.62% (83) of the total missense rcntEffs are identified segregating in the Extreme (43; 14.8%) and the Very_High (40; 13.7%) UVindex categories. Furthermore, the ORF1ab, which occupies two-thirds of the SARS-CoV-2 genome and expresses into 16 non-structural proteins (NSPs), harbors the highest number (163) of missense rcntEffs. We also observed that the nucleocapsid protein (N) and spike glycoprotein (S) encoding genes carry the second and third highest number of missense rcntEffs, 43 and 40, respectively. The rcntEffs counts observed in all UVindex categories are presented in [Figure 5](#).

Comparative genomic analysis

The rcntSNVs-based comparative analysis of all studied full-length SARS-CoV-2 genomes revealed a total of 380 (\sim 73.8% of the all rcntSNVs) UVindex category-specific rcntSNVs (*CaSp*-rcntSNVs), not being shared among any two or more categories (Extreme 58, Very_High 63, High 107, Moderate 84, and Low 68). The comprehensive annotation of each category-specific rcntSNV is given in [Supplementary_info_file.docx](#), [Supplementary Tables 2–6](#). A total of seven rcntSNVs, five missense and two synonymous, observed commonly shared among all UVindex categories, with at least 3,217 overall recurrences, suggesting that all these common rcntSNVs are conserved and near to fixation (rcntSNVs-based comparison is shown in [Figure 6A](#)). Of seven shared rcntSNVs, the ORF1ab 14159C>T (missense; Pro4720Leu) is the most common rcntSNV found in RNA-dependent RNA polymerase (missense; RdRp Pro327Leu; 4,683/8,631 samples), followed by the N gene 608G>A (missense; N Arg203Lys; samples 35,98/8,631), 610G>C (missense; N Gly204Arg; 3,384/8,631 samples), S gene 1841A>G (missense; S Asp614Gly; samples 3,259/8,631), ORF1ab gene 2772C>T (synonymous; ORF1ab Phe924Phe; samples 3,238/8,631), and N gene 610G>C (synonymous; N Gly204Arg; samples 3,217/8,631). All commonly shared rcntSNVs and their respective annotations are given in [Supplementary_info_file.docx](#), [Supplementary Table 8](#). To effectively combat COVID-19, all seven commonly shared rcntSNVs may play a key role in universal vaccine preparation against SARS-CoV-2.

Functional annotation of all 380 *CaSp*-rcntSNVs revealed a sum of 420 category-specific rcntSNV effects (*CaSp*-rcntEffs) on genes products (Extreme 64, Very_High 68, High 120, Moderate 94, and Low 74). Of the total genes, the ORF1ab harbors the highest number of *CaSp*-rcntEffs (234), followed by all four structural genes (103) and six accessory genes (73). The detailed number of *CaSp*-rcntEffs loads per gene for each of the UVindex categories is given in [Table 2](#).

Of the total UVindex *CaSp*-rcntEffs, 222 are found changing codons to specify biochemically different amino acids (*CaSp*-rcntMissense-effects), 136 are observed without consequent

changes in the amino-acid compositions (*CaSp*-rcntSilent-effects), and 62 are detected in the genes' regulatory region (*CaSp*-rcntRegulatory-effects). Most *CaSp*-rcntMissense effects observed in ORF1ab (141), S (27), and surprisingly, the N (26) protein-encoding structural genes. These results showed that SARS-CoV-2 capitalized *CaSp*-rcntMissense, likely the gain of function variant, in ORF1ab and structural protein-encoding genes to adapt to varying UVindex ranges ([Table 3](#)).

Approximately 69.4% (154/222) of the overall *CaSp*-rcntMissense effects are detected in the UVindex range \leq 7 (UVindex categories: Low 35, Moderate 46, and High 73), whereas the remaining 30.6% (68/222) are observed in the Extreme UVindex (36) and Very_High UVindex (31) categories (for details, see [Figures 6B–E](#)). The negatively related linear-trending line with the UVindex implies that the UVindex is inversely proportional to the *CaSp*-rcntMissense effects count. Suggesting that a higher UVindex (mostly \geq 8) allows significantly fewer SARS-CoV-2 viral strains to survive hence imposing strong negative selection pressure ([Figure 7](#)). A set of all category-specific rcntMissense effects causing rcntSNVs may serve as potential resource for considerably more effective region-specific vaccine production.

Phylogeny

To find the evolutionary relationship between SARS-CoV-2 populations prevailing in different UVindex regions, we constructed a phylogenetic tree based on high-quality whole-genome sequences of 125 randomly selected SARS-CoV-2 samples, 25 from each of the UVindex categories ([Figure 8](#)).

Our phylogenetic analysis revealed five different branches for all randomly selected 125 high-quality SARS-CoV-2 genome samples (25 from each of the UVindex region). The tree displays separate branches for SARS-CoV-2 retrieved from UVindex regions, namely High (orange), Extreme (purple), Low (green), Very_High (red), and Moderate (yellow). The phylogenetic analysis has shown High UVindex inhabiting SARS-CoV-2 population as an outgroup and the SARS-CoV-2 prevailing Extreme, Low, Very_High, and Moderate UVindex regions as ingroup populations. To accommodate four SARS-CoV-2 populations, three main lineages were found within the ingroup, revealing the extent of relationships between different populations. The Extreme and Low UVindex populations are placed in two separate ingroup lineages and SARS-CoV-2 populations from the Very_High and Moderate UVindex regions are found sharing the third lineage. This relationship reflects that all SARS-CoV-2 samples, which are included in this study, are descended from the High UVindex region's inhabiting populations, whereas the SARS-CoV-2 populations from Very_High and Moderate UVindex regions are closely related to others.

Conclusion

SARS-CoV-2 is the pandemic COVID-19-causing coronavirus, which has raised a great threat to human health in almost all regions of the world. The genome-wide analysis of the rapidly evolving SARS-CoV-2 genomes discovered a large majority of the rcntSNVs as distinctive (found uniquely in a specific UVindex region), revealing the SARS-CoV-2 differential genomic responses to WHO's defined five different UVindex regions. Based on the total number of rcntSNVs predicted in all included SARS-CoV-2 genomes, our analysis showed that the Extreme UVindex applies negative selection pressure, whereas UVindex range of 6–7 provides the most suitable conditions for SARS-CoV-2 endurance. The phylogenetic relationship indicated the high UVindex region inhabiting the SARS-CoV-2 population as the recent progenitor of all included samples. To help in immune evasion and tolerate the DNA damaging effects of varying UV-solar radiation, the SARS-CoV-2 has acquired the highest number of missense rcntSNVs in their spike glycoprotein and nucleocapsid-encoding genes. Since COVID-19 diagnostic tests and vaccines are based on the spike or the nucleocapsid viral proteins, all missense rcntSNVs may need to be included in future diagnostic and vaccine formulations.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Author contributions

Writing—review and editing: MR, MA, AF, SM, NA, XL, FN, and NI. Writing—original draft preparation: NI, M, AF, and SK. Validation: NI, TY, RR, and XL. Supervision and project administration: SM. Software: NI, TY, and SK. Resources: NI, SM, MJ, and XL. Methodology:

References

- Abdelrahman, Z., Li, M., and Wang, X. (2020). Comparative review of SARS-CoV-2, SARS-CoV, MERS-CoV, and influenza a respiratory viruses. *Front. Immunol.* 11, 552909. doi: 10.3389/fimmu.2020.552909
- Ahlawat, A., Wiedensohler, A., and Mishra, S. (2020). An overview on the role of relative humidity in airborne transmission of SARS-CoV-2 in indoor environments. *Aerosol and Air Quality Research* 20, 1856–1861. doi: 10.4209/aaqr.2020.06.0302
- Benson, D., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D., Ostell, J., et al. (2012). GenBank. *Nucleic Acids Res.* 41, D36–D42. doi: 10.1093/nar/gkr1202
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Brant, A. C., Tian, W., Majerciak, V., Yang, W., and Zheng, Z. (2021). SARS-CoV-2: from its discovery to genome structure, transcription, and replication. *Cell and Bioscience* 11, 1–17. doi: 10.1186/s13578-021-00643-z
- Chang, C., Hou, M., Chang, C., Hsiao, C., and Huang, T. (2014). The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral Res.* 103, 39–50. doi: 10.1016/j.antiviral.2013.12.009
- Chattopadhyay, I., Kiciman, E., Elliott, J., Shaman, J. L., and Rzhetsky, A. (2018). Conjunction of factors triggering waves of seasonal influenza. *Elife* 7, e30756. doi: 10.7554/eLife.30756
- Chin, A. W. H., Chu, J. T. S., Perera, M. R. A., Hui, K. P. Y., Yen, H., Chan, M. C. W., et al. (2020). Stability of SARS-CoV-2 in different environmental conditions. *Lancet Microbe.* 1, e10. doi: 10.1016/S2666-5247(20)30003-3

NI, TY, XL, and MR. Investigation: NI, MR, MA, and FN. Funding acquisition, conceptualization, and formal analysis: NI. Data curation: NI, M, MJ, SK, and ST. All authors contributed to the article and approved the submitted version.

Acknowledgments

We gratefully acknowledge the authors for generating and submitting the laboratories of the sequences to publically accessible GISAID's EpiFlu Database, GenBank, NGDC Genome Warehouse, and the National Microbiology Data Center, on which this research is based. The list of the genomic variations detected from all included genomes is provided in the [Supplementary File](#).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.922393/full#supplementary-material>

- Chiyomaru, K., and Takemoto, K. (2020). Global COVID-19 transmission rate is influenced by precipitation seasonality and the speed of climate temperature warming. *medRxiv* (2020). doi: 10.1101/2020.04.10.20060459
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 6, 80–92. doi: 10.4161/fly.19695
- Clapham, H., Hay, J., Routledge, I., Takahashi, S., Choisy, M., Cummings, D., et al. (2020). Seroepidemiologic study designs for determining SARS-CoV-2 transmission and immunity. *Emerging Infect. Dis.* 26, 1978. doi: 10.3201/eid2609.201840
- CNCB-NGDC Members and Partners (2021). Database Resources of the National Genomics Data Center, China national center for bioinformatics in 2021. *Nucleic Acids Res.* 49, D18–D28. doi: 10.1093/nar/gkaa1022
- Coccia, M. (2020). Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Sci. Total Environ.* 729, 138474. doi: 10.1016/j.scitotenv.2020.138474
- COVID-19 Host Genetics Initiative (2021). Mapping the human genetic architecture of COVID-19. *Nature*. 600, 472–477. doi: 10.1038/s41586-021-03767-x
- Dabbish, L., Stuart, C., Tsay, J., and Herbsleb, J. (2012). in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. p. 1277–1286. doi: 10.1145/2145204.2145396
- Dalziel, B. D., Kissler, S., Gog, J. R., Viboud, C., Bjornstad, O. N., Metcalf, C., et al. (2018). Urbanization and humidity shape the intensity of influenza epidemics in US cities. *Science*. 362, 75–79. doi: 10.1126/science.aat6030
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics*. 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Darnell, M. E. R., Subbarao, K., Feinstone, S. M., and Taylor, D. (2004). Inactivation of the coronavirus that induces severe acute respiratory syndrome, SARS-CoV. *J. Virol. Methods* 121, 85–91. doi: 10.1016/j.jviromet.2004.06.006
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K., Maguire, R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., Baele, G., et al. (2022). Temporal signal and the phylogenetic threshold of SARS-CoV-2. *J. Virus Evolut.* 6, veaa061. doi: 10.1093/ve/veaa061
- Fioletov, V. E., Kimlin, M., Krotkov, N., McArthur, L., Kerr, J. B., Wardle, D. I., et al. (2004). UV index climatology over the United States and Canada from ground-based and satellite estimates. *J. Geophysical Res.* 109, 2004. doi: 10.1029/2004JD004820
- Gardner, E. G., Kelton, D., Poljak, Z., Van Kerkhove, M., Von Dobschuetz, S., and Greer, A. L. A. (2019). case-crossover analysis of the impact of weather on primary cases of Middle East respiratory syndrome. *BMC Infect. Dis.* 19, 1–10. doi: 10.1186/s12879-019-3729-5
- Ghoushchi, S., Ahmadi, M., Sharifi, A., Dorosti, S., Jafarzadeh Ghoushchi, S., Ghanbari, N., et al. (2020). Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Sci Total Environ.* 729, 138705. doi: 10.1016/j.scitotenv.2020.138705
- Gong, Z., Zhu, J., Li, C., Jiang, S., Ma, L., Tang, B., et al. (2020). An online coronavirus analysis platform from the National Genomics Data Center. *Zoological Res.* 41, 705. doi: 10.24272/zj.issn.2095-8137.2020.065
- Guan, W. (2020). Ni, Zheng-yi H, Yu L, Wen-hua O, Chun-quan H, Jian-xing L, et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. doi: 10.1056/NEJMoa2002032
- Hamre, D., and Procknow, J. J. A. (1966). new virus isolated from the human respiratory tract. *Proc. Soc. Exp. Biol. Med.* 121, 190–193. doi: 10.3181/00379727-121-30734
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 181, 271–280, e278. doi: 10.1016/j.cell.2020.02.052
- Holland, J., Spindler, K., Horodyski, F., Grabau, E., and Nichol, S. (1982). Rapid evolution of RNA genomes. *J. Science*. 215, 1577–1585. doi: 10.1126/science.7041255
- Ianevski, A., Zusinaite, E., Shtaida, N., Kallio-Kokko, H., Valkonen, M., Kantele, A., et al. (2019). Low temperature and low UV indexes correlated with peaks of influenza virus activity in Northern Europe during 2010–2018. *Viruses* 11, 207. doi: 10.3390/v11030207
- Iqbal, N., Liu, X., Yang, T., Huang, Z., Hanif, Q., Asif, M., et al. (2019). Genomic variants identified from whole-genome resequencing of indicine cattle breeds from Pakistan. *PLoS ONE*. 14, e0215065. doi: 10.1371/journal.pone.0215065
- Kim, J., Hanotte, O., Mwai, O. A., and Dessie, T. (2017). BashirS, Diallo B, et al. The genome landscape of indigenous African cattle. *Genome Biol.* 18, 1–14. doi: 10.1186/s13059-017-1153-y
- Koyama, T., Platt, D., and Parida, L. (2020). Variant analysis of SARS-CoV-2 genomes. *Bull. World Health Organ.* 98, 495. doi: 10.2471/BLT.20.253591
- Kumar, S., Singh, R., Kumari, N., Karmakar, S., Behera, M., Siddiqui, A. R., et al. (2021). Current understanding of the influence of environmental factors on SARS-CoV-2 transmission, persistence, and infectivity. *Environ. Sci. Pollut. Res.* 1–22. doi: 10.1007/s11356-020-12165-1
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547. doi: 10.1093/molbev/msy096
- Lai, M. M. C., and Cavanagh, D. (1997). The molecular biology of coronaviruses. *Adv. Virus Res.* 48, 1–100. doi: 10.1016/S0065-3527(08)60286-9
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 1303, 3997.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*. 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, H. A. (2011). statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., et al. (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science*. 310, 676–679. doi: 10.1126/science.1118391
- Licitra, B. N., Millet, J. K., Regan, A. D., Hamilton, B. S., Rinaldi, V. D., Duhamel, G. E., et al. (2013). Mutation in spike protein cleavage site and pathogenesis of feline coronavirus. *Emerging Infect. Dis.* 19, 1066. doi: 10.3201/eid1907.121094
- Lo, C. W., Matsuura, R., Iimura, K., Wada, S., Shinjo, A., Benno, Y., et al. (2021). UVC disinfects SARS-CoV-2 by induction of viral genome damage without apparent effects on viral morphology and proteins. *Sci. Rep.* 11, 1–11. doi: 10.1038/s41598-021-93231-7
- Lowen, A. C., Mubareka, S., Steel, J., and Palese, P. (2007). Influenza virus transmission is dependent on relative humidity and temperature. *JPLoS Pathogens*. 3, e151. doi: 10.1371/journal.ppat.0030151
- Lu, H., Li, J., Yang, P., Jiang, F., Liu, H., Cui, F., et al. (2022). Mutation in the RNA-Dependent RNA Polymerase of a Symbiotic Virus Is Associated With the Adaptability of the Viral Host. *Hosts*. 13. doi: 10.3389/fmicb.2022.883436
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 395, 565–574. doi: 10.1016/S0140-6736(20)30251-8
- McIntosh, K., Becker, W. B., and Chanock, R. M. (1967). Growth in suckling-mouse brain of “IBV-like” viruses from patients with upper respiratory tract disease. *Proc. Natl. Acad. Sci. USA*. 58, 2268. doi: 10.1073/pnas.58.6.2268
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- McMullan, L. K., Flint, M., Chakrabarti, A., Guerrero, L., Lo, M. K., Porter, D., et al. (2019). Characterisation of infectious Ebola virus from the ongoing outbreak to guide response activities in the Democratic Republic of the Congo: a phylogenetic and in vitro analysis. *Lancet Infect. Dis.* 19, 1023–1032. doi: 10.1016/S1473-3099(19)30291-9
- Moris, A., Murray, S., and Cardinaud, S. (2014). AID and APOBECs span the gap between innate and adaptive immunity. *Front. Microbiol.* 5, 534. doi: 10.3389/fmicb.2014.00534
- Mourier, T., Sadykov, M., Carr, M. J., Gonzalez, G., Hall, W. W., Pain, A. J., et al. (2021). Host-directed editing of the SARS-CoV-2 genome. *Biochem. Biophys. Res. Commun.* 538, 35–39. doi: 10.1016/j.bbrc.2020.10.092
- O’Reilly, K. M., Auzenbergs, M., Jafari, Y., Liu, Y., Flasche, S., Lowe, R., et al. (2020). Effective transmission across the globe: the role of

- climate in COVID-19 mitigation strategies. *Lancet Planetary Health*. 4, e172. doi: 10.1016/S2542-5196(20)30106-6
- Otter, J. A., Donskey, C., Yezli, S., Douthwaite, S. G., Dea, S., Weber, D. J., et al. (2016). Transmission of SARS and MERS coronaviruses and influenza virus in healthcare settings: the possible role of dry surface contamination. *J Hospital Infect.* 92, 235–250. doi: 10.1016/j.jhin.2015.08.027
- Pi, S., Seng-Libi, L., and Xiao-Ping, D. (2003). Stability of SARS coronavirus in human specimens and environment and its sensitivity to heating and UV irradiation. *Biomed. Environ. Sci.* 16, 246–255. https://www.researchgate.net/profile/Xiao-Ping-Dong-2/publication/8995908_Stability_of_SARS_Coronavirus_in_Human_Specimens_and_Environment_and_Its_Sensitivity_to_Heating_and_UV_Irradiation/links/5e2fa67f299bf10a6598fa29/Stability-of-SARS-Coronavirus-in-Human-Specimens-and-Environment-and-Its-Sensitivity-to-Heating-and-UV-Irradiation.pdf
- Pica, N., and Bouvier, N. M. (2012). Environmental factors affecting the transmission of respiratory viruses. *Curr. Opin. Virol.* 2, 90–95. doi: 10.1016/j.coviro.2011.12.003
- Placido, D., Brown, I. I. B. A., Lowenhaupt, K., Rich, A., and Athanasiadis, A. A. (2007). left-handed RNA double helix bound by the α domain of the RNA-editing enzyme ADAR1. *J Structure.* 15, 395–404. doi: 10.1016/j.str.2007.03.001
- Ratnesar-Shumate, S., Williams, G., Green, B., Krause, M., Holland, B., Wood, S., et al. (2020). Simulated sunlight rapidly inactivates SARS-CoV-2 on surfaces. *J. Infect. Dis.* 222, 214–222. doi: 10.1093/infdis/jiaa274
- Reich, N. G., Brooks, L. C., Fox, S. J., Kandula, S., McGowan, C. J., Moore, E., et al. (2019). A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proc. Natl. Acad. Sci. U.S.A.* 116, 3146–3154. doi: 10.1073/pnas.1812594116
- Rella, S. A., Kulikova, Y. A., Dermitzakis, E. T., and Kondrashov, F. A. (2021). Rates of SARS-CoV-2 transmission and vaccination impact the fate of vaccine-resistant strains. *Sci. Rep.* 11, 1–10. doi: 10.1038/s41598-021-95025-3
- Rice, P., Longden, I., and Bleasby, A. E. M. B. O. S. S. (2000). the European molecular biology open software suite. *Trends in genetics* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Rubio, L., Guerri, J., and Moreno, P. (2013). Genetic variability and evolutionary dynamics of viruses of the family Closteroviridae. *Front. Microbiol.* 4, 151. doi: 10.3389/fmicb.2013.00151
- Sanjuán, R., and Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cell. Mol. Life Sci.* 73, 4433–4448. doi: 10.1007/s00018-016-2299-6
- Seyer, A., and Sanlidag, T. (2020). Solar ultraviolet radiation sensitivity of SARS-CoV-2. *The Lancet Microbe.* 1, e8–e9. doi: 10.1016/S2666-5247(20)30013-6
- Shah, A. S. V., Gribben, C., Bishop, J., Hanlon, P., Caldwell, D., Wood, R., et al. (2021). Effect of vaccination on transmission of SARS-CoV-2. *N. Engl. J. Med.* 385, 1718–1720. doi: 10.1056/NEJMc2106757
- Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T., and Lipsitch, M. (2010). Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* 8, e1000316. doi: 10.1371/journal.pbio.1000316
- Shi, M., Lin, X., Chen, X., Tian, J., Chen, L., Li, K., et al. (2018). The evolutionary history of vertebrate RNA viruses. *Nature.* 556, 197–202. doi: 10.1038/s41586-018-0012-7
- Shi, M., Lin, X., Tian, J., Chen, L., Chen, X., Li, C., et al. (2016). Redefining the invertebrate RNA virosphere. *Nature.* 540, 539–543. doi: 10.1038/nature20167
- Shu, Y., and McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance.* 22, 30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494
- Smertina, E., Uraikova, N., Strive, T., and Frese, M. (2019). Calicivirus RNA-dependent RNA polymerases: evolution, structure, protein dynamics, and function. *Front. Microbiol.* 10. doi: 10.3389/fmicb.2019.01280
- Soh, S. M., Kim, Y., Kim, C., Jang, U., and Lee, H. (2021). The rapid adaptation of SARS-CoV-2—rise of the variants: transmission and resistance. *J. Microbiol.* 59, 807–818. doi: 10.1007/s12275-021-1348-5
- Tahara, S. M., Dietlin, T. A., Bergmann, C. C., Nelson, G. W., Kyuwa, S., Anthony, R. P., et al. (1994). Coronavirus translational regulation: leader affects mRNA efficiency. *Virology* 202, 621–630. doi: 10.1006/viro.1994.1383
- Tahara, S. M., Dietlin, T. A., Nelson, G. W., Stohlman, S. A., and Manno, D. J. (1998). “Mouse hepatitis virus nucleocapsid protein as a translational effector of viral mRNAs,” in *Coronaviruses and Arteriviruses* (Boston, MA: Springer), 313–318. doi: 10.1007/978-1-4615-5331-1_41
- Tan, J., Mu, L., Huang, J., Yu, S., Chen, B., Yin, J., et al. (2005). An initial investigation of the association between the SARS outbreak and weather: with the view of the environmental temperature and its variation. *J. Epidemiol. Community Health.* 59, 186–192. doi: 10.1136/jech.2004.020180
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Tong, J., Zhang, W., Chen, Y., Yuan, Q., Qin, N., Qu, G., et al. (2022). The emerging role of RNA modifications in the regulation of antiviral innate immunity. *Front. Microbiol.* 13, 845675. doi: 10.3389/fmicb.2022.845675
- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *J Nat. Rev. Microbiol.* 19, 155–170. doi: 10.1038/s41579-020-00468-6
- Wang, C., Horby, P. W., Hayden, F. G., and Gao, G. F. A. (2020). novel coronavirus outbreak of global health concern. *Lancet.* 395, 470–473. doi: 10.1016/S0140-6736(20)30185-9
- Wang, D., Bo, H., Chang, Z., Fangfang, L., Xing, Z., Jing, W., et al. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 323, 1061–1069. doi: 10.1001/jama.2020.1585
- Wang, J., Tang, K., Feng, K., Li, X., Lv, W., Chen, K., et al. (2020). High temperature and high humidity reduce the transmission of COVID-19. *arXiv.* 2003, 05003. doi: 10.2139/ssrn.3551767
- Wang, W., Lin, X., Guo, W., Zhou, R., Wang, M., Wang, C., et al. (2015). Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology.* 474, 19–27. doi: 10.1016/j.virol.2014.10.017
- World Health Organization, World Health Protection, and International Commission on Non-Ionizing Radiation. (2002). *Global solar UV index: a practical guide. Report No. 9241590076.* World Health Organization 2002. Available online at: <https://apps.who.int/iris/bitstream/handle/10665/42459/9241590076.pdf>
- World Health Organization. (2003). Summary table of SARS cases by country, 1 November 2002–7 August 2003. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire* 78, 310–311.
- World Health Organization. (2019). *Middle East respiratory syndrome coronavirus (MERS-CoV).*
- Wu, F., Zhao, S., Yu, B., Chen, Y., Wang, W., Song, Z., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature.* 579, 265–269. doi: 10.1038/s41586-020-2008-3
- Yadav, P. D., Shete, A. M., Kumar, G. A., Sarkale, P., Sahay, R. R., Radhakrishnan, C., et al. (2019). Nipah virus sequences from humans and bats during Nipah outbreak, Kerala, India, 2018. *Emerging Infect. Dis.* 25, 1003. doi: 10.3201/eid2505.181076