



Genomic Characterization of SARS-CoV2 from Peshawar Pakistan Using Next-Generation Sequencing

Ome Kalsoom Afridi^{1,2} · Nousheen Bibi³ · Syed Adnan Haider⁴ · Bibi Sabiha⁴ · Hanifullah Jan⁴ · Abid Ali Khan⁵ · Shireen Akhter^{6,7} · Valeed Khan⁸ · Johar Ali^{4,9}

Received: 15 March 2021 / Accepted: 6 December 2021 / Published online: 4 January 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

This study aimed to characterize the whole genome of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2) isolated from an oropharyngeal swab specimen of a Pashtun Pakistani patient using next-generation sequencing. Upon comparing the SARS-CoV2 genome to the reference genome, a total of 10 genetic variants were identified. Among the 10 genetic variants, 1 missense mutation (c.1139A > G, p.Lys292Glu) in the Open Reading Frame 1ab (ORF1ab) positioned at 112 in the non-structural protein 2 (NSP2) was found to be unique. Phylogenetic analysis ($n = 84$) revealed that the current SARS-CoV2 genome was closely clustered with 8 Pakistani strains belonging to Punjab, Federal Capital, Azad Jammu and Kashmir (AJK), and Khyber Pakhtunkhwa (KP). In addition, the current SARS-CoV2 genome was very similar to the genome of SARS-CoV2 reported from Guam, Taiwan, India, the USA, and France. Overall, this study reports a slight mismatch in the SARS-CoV2 genome, indicating the presence of a single unique missense mutation. However, phylogenetic analysis revealed that the current SARS-CoV2 genome was closely clustered with 8 other Pakistani strains.

Introduction

Coronavirus infects a variety of hosts, including bats, snakes, birds, mice, wild animals, and humans [1–3]. Severe Acute Respiratory Syndrome (SARS) coronavirus (SARS-CoV) occurred in late 2002 in Guangdong, China. SARS-CoV has infected 8098 people worldwide, killing a total of 774 individuals. After SARS-CoV infection, another SARS virus called Middle East Respiratory Syndrome (MERS) emerged in Saudi Arabia in 2012. MERS-CoV caused 2494 infections and killed 858 people [3]. At the end of 2019, another coronavirus with the usual symptoms of pneumonia appeared in Wuhan, China. This virus is called SARS-CoV2 and the COVID19 term is used to refer to infected cases. To date, 199 million people worldwide have been infected with SARS-CoV2 and 4.24 million have expired. SARS-CoV2 is the 7th member of coronavirus (CoV) family *Coronaviridae*, which is known to cause serious infections in various hosts, such as birds, mammals, and humans [4, 5]. In humans, the six coronaviruses are known to cause different infections with different symptoms. Symptoms range from mild cold-like symptoms often caused by alpha (229E, NL63) and beta coronaviruses (OC43, HKU1) to severe respiratory illnesses caused by SARS-CoV and the MERS-CoV [6]. These viruses are usually known for having larger genetic

✉ Ome Kalsoom Afridi
ummeafridi@gmail.com

¹ Department of Biochemistry, Women University Swabi, Swabi, Khyber Pakhtunkhwa (KP), Pakistan

² Department of Biology Education, Kyungpook National University, Daegu 41566, Republic of Korea

³ Department of Bioinformatics, Shaheed Benazir Bhutto Women University, Peshawar, KP, Pakistan

⁴ Center for Genome Sciences, Rehman Medical College (RMC), Peshawar, KP, Pakistan

⁵ Department of Chemical Sciences, University of Lakki Marwat, Lakki Marwat, KP, Pakistan

⁶ Executive Development Centre, Sukkur IBA University, Sukkur, Sindh, Pakistan

⁷ Biotech, Sukkur IBA University, Sukkur, Sindh, Pakistan

⁸ Pathology Laboratory, Department of Molecular Diagnostics, Rehman Medical Institute (RMI), Peshawar, Khyber Pakhtunkhwa (KP), Pakistan

⁹ AlviArmani International, Mississauga, ON, Canada

material (RNA; 26–32 kilobases in length), composed of an enveloped positive-sense single-stranded RNA. Electron microscope observations of these RNA viruses show a crown-shaped spherical sequence arrangement commonly referred to as coronaviruses.

Whole-genome sequencing (WGS) played an important role in the understanding of various emerging viruses outbreaks, such as Zika, Ebola, Usutu, and Yellow fever viral infections [6]. Owing to the powerful role of WGS during various outbreaks, large-scale next-generation sequencing (NGS)-based efforts were initiated globally to explore the genomic insights of SARS-CoV2 following the first outbreak of COVID19. On January 5, 2020, the first whole-genome sequence of SARS-CoV2 was published. Following the first complete genome sequencing of SARS-CoV2, several countries submitted the complete genome sequence of SARS-CoV2 to the Global Initiative on Sharing All Influenza Data (GISAID) database [7, 8]. Pakistan submitted its first coronavirus whole-genome sequence (SARS-CoV2/Gilgit1/human/2020/PAK; accession number: MT240479) to the GenBank on March 25, 2020. Following the first WGS of SARS-CoV2 from Pakistan, several WGS-based attempts were made and the sequencing data were submitted to the GenBank (SARS-CoV2/Manga1/human/2020/PAK; accession number: MT262993) [9]. The initial two sequenced strains of Pakistani coronavirus (accession numbers: MT240479 and MT262993) exhibited 99.98% and 100% sequence similarity to the Chinese SARS-CoV2 isolate [9]. During the COVID19 pandemic, the mortality pattern of SARS-CoV2 in Pakistan was found to be considerably different from the rest of the world. For instance, the neighboring country of Pakistan, such as India, has so far been ranked the second largest in the world with the highest number of COVID-19-positive cases after the USA [10]. The variation in COVID-19-positive cases in Pakistan from its neighboring countries warrants extensive WGS-based investigation in order to rule out the key variants associated with Pakistani SARS-CoV2 strains and determine if such mutations exist in other countries with similar prevalence patterns. Therefore, we sequenced and characterized the complete genome of SARS-CoV2 isolated from a patient of Pashtun ethnicity in the Khyber Pakhtunkhwa (KP) region of Pakistan. In addition to WGS, phylogenetic analysis also performed to compare the current SARS-CoV2 strain genome with the publicly available genomes.

Materials and Methods

Extraction of RNA and Quality Control

An oropharyngeal swab specimen was collected from a symptomatic patient registered at Rehman Medical Institute

(RMI) in Peshawar in June 2020 by following World Health Organization (WHO) guidelines [11]. Viral RNA was extracted from the swab specimen using the QIAasympphony SP/AS instruments as per the manufacturer's instructions (QIAGEN). SARS-CoV2 was quantified using Rotor-Gene Q real-time PCR and a Novel Coronavirus Nucleic Acid Diagnostic Kit (Sansure Biotech, Inc. China). Quality control (QC) of the extracted RNA was performed by quantification using the Qubit™ RNA Broad Range (BR) assay kit (Cat. No. Q10211).

Primer Designing and PCR Amplification

Extracted RNA was subjected to cDNA synthesis using RevertAid First Strand cDNA Synthesis Kit (Cat. No. K1622) following the manufacturer's guidelines (Thermo Fisher Scientific). Using Primer3 software (v 0.4.0) [12] seven sets of primers were designed to cover the genome of SARS-CoV2 (GenBank accession number: NC_045512). In addition, two sets of overlapping primers were also designed for PCR troubleshooting (Supplementary Table S1). Then, according to the recommended guidelines, PCR (Bio-Rad) was performed using the Phusion Flash High-Fidelity PCR Master Mix (Cat. No. F548S, Thermo Fisher Scientific). The PCR conditions were optimized as follows: (i) initial denaturation (95 °C for 10 s), (ii) 35 cycles of denaturation (95 °C for 3 s), and (iii) a final extension (72 °C for 4 min) were kept similar for all sets of primers while the annealing temperature and extension time were different for each set of primers (Supplementary Table S1). After amplification, the PCR products were run on an agarose gel and then quantified using the Qubit dsDNA High Sensitivity (HS) Assay Kit (Cat. No. Q32851; Invitrogen).

NGS Library Preparation and Whole-Genome Sequencing

The quantified PCR product was normalized to 0.2 ng/μL. The paired-end sequencing library was prepared using the Illumina Nextera XT DNA Library Preparation Kit (Cat. No. FC131-1096) by incorporating the following main steps: (i) tagmentation (enzymatic fragmentation and tagging of the dsDNA), (ii) PCR amplification, (iii) PCR clean up, (iv) bead-based normalization of NGS library, and (v) library pooling. The pooled library was then loaded onto the MiSeq (Illumina) for paired-end sequencing using sequencing reagent cartridge, MiSeq Reagent Micro Kit v2, 300-cycles (Cat. No. MS-103-1002).

NGS Bioinformatics Analysis

Paired-end NGS data were analyzed using publicly available bioinformatics softwares. Firstly, the read quality of

FASTQ files were checked using the FastQC tool (v0.11.8) [13]. The Trimmomatic tool (v0.39) was used to remove low-quality base calls ($Q < 30$) and index adapter sequences from both ends of sequenced reads [14]. The filtered reads were aligned with the Wuhan reference genome (GenBank accession number: NC_045512) using the default settings for the Burrows–Wheeler Aligner (BWA, v0.7.17) [15]. The reference sequence of SARS-CoV2 (accession number: NC_045512) was used as a control for the viral genome alignment [16]. Genome annotation and variant calling were performed using China’s National Genomics Data Center (NGDC) [17]. The identified variants were cross-validated and analyzed using Genome Detective Coronavirus Typing tool [18] and BioEdit [19].

Phylogenetic Analysis

FASTQ files of various publicly available SARS-CoV2 genomes were retrieved from the GISAID and NCBI databases. A total of 85 SARS-CoV2 genomes including the current sequenced genome (accession number: MW242667) were clustered using Augur Nextstrain’s phylodynamic pipeline (Supplementary Material) [20]. The sequences were then aligned against Wuhan reference genome (accession number: NC_045512.2) using MAFFT [21]. Phylogenetic tree was constructed using IQ-TREE [22]. The phylogenetic tree was then visualized using FigTree v1.4.4 [23].

Results and Discussion

The whole genome of SARS-CoV2 isolated from a patient belonging to KP region of Pakistan was sequenced using NGS. A total of 166,007 reads (22,292,523 bases) were aligned against the reference genome (accession number: NC_045512). Sequence Statistics of the sequenced reads are listed in Supplementary Table S2. The SARS-CoV2-sequenced genome revealed the presence of 10 genomic variants encoding various genes, such as open reading frame 1ab (ORF1ab), spike glycoprotein (S), ORF8, nucleocapsid (N), and ORF10 genes. The 10 identified variants consist of 8 missense and 2 synonymous mutations. Out of total 10 variants, the following 5 mutations were identified in the ORF1ab region: 1139A > G (codons AAG to GAG, missense mutation), 2144G > T (codons GTC to TTC, missense mutation), 11083G > T (codons TTG to TTT, missense mutation), 13730C > T (codons: GCT to GTT, missense mutation), and 6312 C > A (codons ACA to AAA, missense mutation). A single mutation was identified in S gene 23929 C > T (codons TAC to TAT, synonymous mutation). Similarly, a single genomic variant was detected in the ORF8 gene at 28253 genomic location (C > T, codons TTC > TTT, synonymous mutation). Similarly, out of 10

mutations, 1139A > G, p.Lys292Glu (position 112 of NSP2 protein) was identified as a unique genetic variant. The NSP2 protein of SARS-CoV2 contains 61 amino acids, which are different from SARS-CoV [24]. In coronavirus infection, double-membrane vesicles filled with replication–transcription complexes (RTCs) are formed in the infected cells and the NSP2 protein is essential for the RTC formation [25]. Other variants identified in this study are also detected in the SARS-CoV2 genome isolated from different countries around the world (Table 1). The higher number of variants identified in SARS-CoV2 replicase polyproteins at different positions are supported by a previous study indicating that replicase polyprotein of 13 SARS-CoV2 isolates from different countries harbored mutations at different locations of different amino acids [26]. The various missense mutations in the SARS-CoV2 replicase polyprotein detected in this study is consistent with previous studies. For instance, a missense mutation in the replicase polyprotein at position 3606 (L to F) was reported in a previous study [27]. Mutations in replicase polyprotein (ORF 1 ab) are associated with different mechanisms of SARS-CoV2 [28]. The low mortality rate of Pakistan can be attributed to its diverse climatic conditions. A recent study compared the death rates of SARS-CoV2 in various countries ($n = 45$) of the world according to the country-specific climatic condition. They found that temperate countries such as Italy, Spain, the Netherlands, France, and England exhibited higher mortality rates compared to countries with a diverse climate, such as Brazil, Australia, and Pakistan [29].

Based on phylogenetic analysis, the SARS-CoV2 genome sequenced in the current study formed clusters with isolates from Guam (accession numbers: MT459985.1, MT459986.1, and MT459987.1), Taiwan (accession numbers: MT517436.1 and MT517437.1), India (accession numbers: MT477885.1, MT457403.1, and MT415322.1), the USA (accession numbers: MT499206.1, and MT344946.1), France (accession number: MT470111.1), and 8 other Pakistani strains including, Punjab (accession numbers: MW422100.1), Federal Capital (accession numbers: MW422012.1, MW422082.1, MW422089.1, MW422099.1, and MW421988.1), Azad Jammu and Kashmir (AJK; accession number: MW422086.1), and KP (accession numbers: MW422088.1) (Fig. 1). Details of all annotated SARS-CoV2 genomic sequences from Pakistan and other countries used for phylogenetic analysis are listed in Supplementary Material. The close association between the current SARS-CoV2 genome and local Pakistani strains is inconsistent with a recent study suggesting differences between the local Pakistani strains [30]. The close association of local SARS-CoV2 strains can be attributed to the “Smart lockdown” in Pakistan, during which no restrictions were imposed on traveling within the country. In

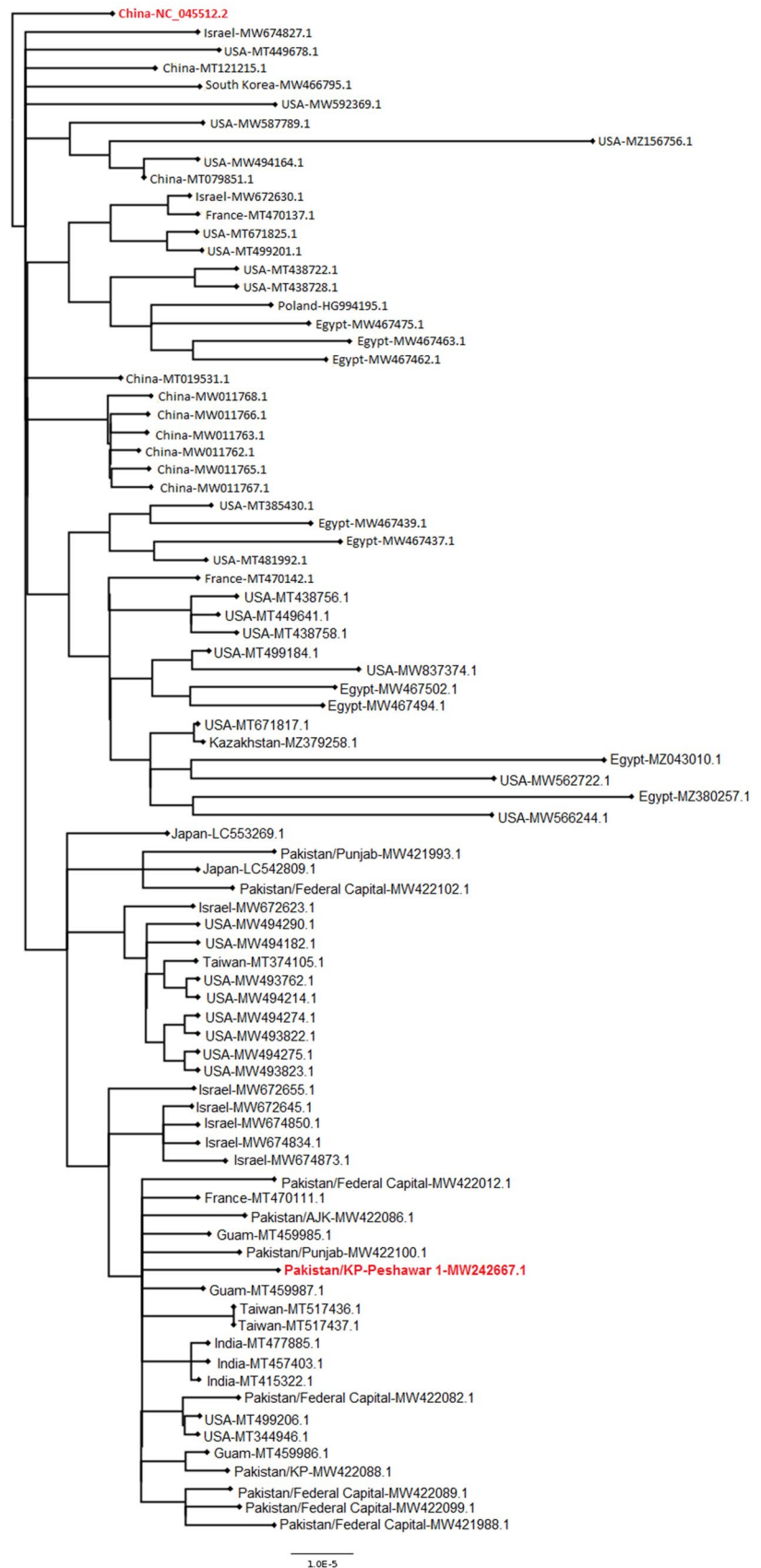
Table 1 List of mutations detected in the genome of SARS-CoV2 from other geographic regions besides Pakistan

Genomic position	Mutation	Region	Simple ID		
orf1ab	2144 (G>T)	England	EPI_ISL_425449		
		Australia	EPI_ISL_427753		
	11083 (G>T)	Australia	EPI_ISL_419793		
		13730 (C>T)	Saudi Arabia	EPI_ISL_416432	
	Australia		EPI_ISL_419761		
	6312 (C>A)	USA	EPI_ISL_434297		
			England	EPI_ISL_433944	
			India	EPI_ISL_437438	
			Brunei	EPI_ISL_435674	
			Saudi Arabia	EPI_ISL_416432	
			Australia	EPI_ISL_419761	
			India	EPI_ISL_437438	
			Brunei	EPI_ISL_435674	
			S gene	23929(C>T)	Saudi Arabia
Iceland					EPI_ISL_417752
Australia	EPI_ISL_419761				
USA	EPI_ISL_434297				
India	EPI_ISL_437438				
Brunei	EPI_ISL_435674				
ORF8	28253 (C>T)	India			EPI_ISL_436456
		Denmark	EPI_ISL_437041		
		England	EPI_ISL_453633		
		Switzerland	EPI_ISL_476097		
		Netherlands	EPI_ISL_461144		
		England	EPI_ISL_461916		
		USA	EPI_ISL_414483		
		Wales	EPI_ISL_418137		
		France	EPI_ISL_420064		
		N	28311 (C>T)	Australia	EPI_ISL_419728
Saudi Arabia	EPI_ISL_416432				
USA	EPI_ISL_434297				
India	EPI_ISL_437438				
28887 (C>T)	Beijing			EPI_ISL_430722	
	Singapore		EPI_ISL_435687		
	Wales		EPI_ISL_432250		
	USA		EPI_ISL_454682		
	England		EPI_ISL_452867		
ORF10 protein	29645 (G>T)		Australia	EPI_ISL_419961	
		DRC*	EPI_ISL_437343		

addition, the current SARS-CoV2 genome was very similar to the genome of SARS-CoV2 reported from Guam, Taiwan, the USA, India, and France. The close clustering of current-sequenced SARS-CoV2 genome with the genome of SARS-CoV2 reported from Guam, Taiwan, the USA, India, and France is in agreement with the recent study [31]. Sequenced in the first wave, the current SARS-CoV2 genome close clustering with the genomes of SARS-CoV2 of various countries can be attributed to flexible international travel. International travel has been considered as one of the potential risk factors for the transmission and circulation of different variants of SARS-CoV2 in Pakistan [31–34]. The initial-sequenced local SARS-CoV2 genomes (accession numbers: MT240479 and MT262993) are closely related to the Chinese SARS-CoV2 isolates [9], while the currently sequenced SARS-CoV2 isolate differs from that of the Chinese isolates (Fig. 1). This implies that the current-sequenced SARS-CoV2 genome evolved and acquired genetic variants from the Wuhan reference genome (accession number: NC_045512).

In conclusion, using NGS, this study has revealed new information about the SARS-CoV2 genome and comprehensively characterized the full sequence of SARS-CoV2. A missense mutation (1139A>G) detected in the NSP2 protein of SARS-CoV2 in this study is a novel genetic variant. Furthermore, the identification of a novel variant in the SARS-CoV2 genome collected in Peshawar, Pakistan confirmed previous findings that SARS-CoV2 is not similar around the world. Therefore, there is a need for SARS-CoV2 genome sequencing in all regions of the world. The region-specific sequencing and reporting of the identified variant will greatly contribute to the development of vaccines and diagnostic kits. Furthermore, it has also been concluded that the SARS-CoV2 diagnostic kit developed in one part of the world may misdiagnose patients in another part. In conclusion, genome sequencing of local SARS-CoV2 strains could provide crucial information for improving diagnostic, prognostic, and therapeutic interventions. Overall, this study reports a minor deviation in the SARS-CoV2 genome showing the presence of 1 unique missense mutation. However, phylogenetic analysis revealed that the current SARS-CoV2 genome was closely clustered with 8 other Pakistani strains.

Fig. 1 A phylogenetic tree of 85 genomes obtained from GISAID and sequenced in the current study. The selected genomes were clustered using Augur, the Nextstrain phylogenetic pipeline. The SARS-CoV2 genome (accession number: MW242667) sequenced in the current study is highlighted in red



Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00284-021-02743-y>.

Acknowledgements This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Author Contribution OKA wrote the manuscript and assisted the team in experimental and bioinformatics analysis, NB helped in data analysis and manuscript preparation, SAH performed bioinformatics analysis and reviewed the manuscript, BS and HJ performed the experiments, AAK supported the lab work, reviewed, and helped in the manuscript preparation, SA reviewed the manuscript and assisted in editing the first draft, VK helped with the sampling, and JA developed the idea and led the execution of the experiment, supervised, and advised on the data analysis and manuscript.

Data Availability SARS-CoV2 full genome sequence generated in this study has been deposited in GenBank under the accession number: MW242667 (available at URL: <https://www.ncbi.nlm.nih.gov/nuccore/MW242667.1/>).

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical Approval The present study reporting research involving Human Participants was carried out following ethics principles of the Declaration of Helsinki.

Informed Consent This study was approved by our institutional review committee (Rehman Medical Institute-Research Ethics Committee, Peshawar, Pakistan; RMI/RMI-REC/Article Approval/48). Informed consent was obtained from the enrolled patient.

References

- Hirschey M (2020) Lab life-rebuild it better after coronavirus lockdowns ease. *Nature* 582(7811):184
- Corman VM, Muth D, Niemeyer D, Drosten C (2018) Hosts and sources of endemic human coronaviruses. *Adv Virus Res* 100:163–188
- Guo D (2020) Old weapon for new enemy: drug repurposing for treatment of newly emerging viral diseases. *Viol Sin.* 35:253–255
- Cui J, Li F, Shi Z-L (2019) Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 17(3):181–192
- Monchatre-Leroy E, Boué F, Boucher J-M, Renault C, Moutou F, Ar Guilh M et al (2017) Identification of alpha and beta coronavirus in wildlife species in France: bats, rodents, rabbits, and hedgehogs. *Viruses* 9(12):364
- Munnink BBO, Nieuwenhuijse DF, Stein M, O’Toole Á, Haverkate M, Mollers M et al (2020) Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med* 26(9):1405–1410
- Elbe S, Buckland-Merrett G (2017) Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Global Chall* 1(1):33–46
- Shu Y, McCauley J (2017) GISAID: global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* 22(13):30494
- Saif R, Mahmood T, Ejaz A (2020) Whole genome comparison of Pakistani corona virus with Chinese and US strains along with its predictive severity of COVID-19. *bioRxiv*. <https://doi.org/10.1016/j.genrep.2021.101139>
- Worldometers (2020) COVID-19 coronavirus pandemic
- World Health Organization (2020) Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 March 2020. World Health Organization, Geneva.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M et al (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res* 40(15):e115
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760
- Wu F, Zhao S, Yu B, Chen Y, Wang W, Song Z et al (2020) Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. *Nature* 579(7798):265–269
- National Genomics Data Center Members and Partners (2020) Database resources of the national genomics data center in 2020. *Nucleic Acids Res* 48:D24–D33
- Cleemput S, Dumon W, Fonseca V, Abdool Karim W, Giovanetti M, Alcantara LC et al (2020) Genome detective coronavirus typing tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics* 36(11):3552–3555
- Hall T, Biosciences I, Carlsbad C (2011) BioEdit: an important software for molecular biology. *GERF Bull Biosci* 2(1):60–61
- Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C et al (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34(23):4121–4123
- Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. *Bioinformatics for DNA sequence analysis*. Springer, New York, pp 39–64
- Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1):268–274
- Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C et al (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 5(11):1403–1407
- Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P et al (2020) Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 27:325–328
- Hagemeijer MC, Verheije MH, Ulasli M, Shaltiél IA, De Vries LA, Reggiori F et al (2010) Dynamics of coronavirus replication-transcription complexes. *J Virol* 84(4):2134–2149
- Khan MI, Khan ZA, Baig MH, Ahmad I, Farouk A-E, Song YG et al (2020) Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: an in silico insight. *PLoS ONE* 15(9):e0238344
- Wang C, Liu Z, Chen Z, Huang X, Xu M, He T et al (2020) The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol* 92(6):667–674
- Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M (2020) COVID-2019: the role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol* 92(6):584–588
- Islam MR, Hoque MN, Rahman MS, Alam ARU, Akther M, Puspo JA et al (2020) Genome-wide analysis of SARS-CoV-2

- virus strains circulating worldwide implicates heterogeneity. *Sci Rep* 10(1):1–9
30. Khan MT, Ali S, Khan AS, Muhammad N, Khalil F, Ishfaq M et al (2021) SARS-CoV-2 Genome from the Khyber Pakhtunkhwa Province of Pakistan. *ACS Omega* 6(10):6588–6599
 31. Umair M, Salman M, Rehman Z, Badar N, Ali Q, Ahad A et al (2021) Proliferation of SARS-CoV-2 B. 1.1.7 variant in Pakistan—a short surveillance account. *Frontiers Public Health*. 9:702
 32. Atif M, Malik I (2020) Why is Pakistan vulnerable to COVID-19 associated morbidity and mortality? A scoping review. *Int J Health Plann Manage* 35(5):1041–1054
 33. Saqlain M, Munir MM, Ahmed A, Tahir AH, Kamran S (2020) Is Pakistan prepared to tackle the coronavirus epidemic? *Drugs Therapy Perspect* 36:213–214
 34. Nafees M, Khan F (2020) Pakistan's response to COVID-19 pandemic and efficacy of quarantine and partial lockdown: a review. *Electron J Gen Med*. 17(2):em240

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.