# Comparative genomics of spike, envelope, and nucleocapsid protein of severe acute respiratory syndrome coronavirus 2

Sufyan Sohail Khan, Anwar Ullah

Department of Biosciences, COMSATS University Islamabad, Islamabad, Pakistan.

### Abstract

**Background:** Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) upsurge sprang up in Wuhan, China, in late December 2019.

**Objectives:** Due to the exceptionally high mutation frequency, comparative genomics of viruses isolated throughout time and in various geographical locations are crucial. To better understand how SARS-CoV-2 heterogeneity has changed around the globe, this research was conducted.

**Methods:** Nucleotide and protein sequences of *SARS-CoV-2*, *SARS-CoV*, and bat *SARS-like CoV* were extracted from the NCBI Virus database. The Wuhan *SARS-CoV-2* variant was used as a reference. Molecular Evolutionary Genetics Study performed the phylogenetic analysis, while the Genome Detective Coronavirus Typing Tool performed the mutational analysis.

**Results:** The evolutionary research has revealed that bats are the primary host for coronavirus evolution and the origin of the formation of *SARS-CoV* and *SARS-CoV-2*. Numerous mutations have been discovered in the spike, envelope, and nucleocapsid protein.

**Conclusions:** The current research findings may have an implication that facilitates the development of prospective immunization candidates/small pharmacological compounds targeting COVID-19.

Keywords: *SARS-CoV-2*; COVID-19; pandemic; comparative genomics; spike protein; envelope protein; nucleocapsid protein. DOI: https://dx.doi.org/10.4314/ahs.v23i3.45

**Cite as:** Khan SS, Ullah A. Comparative genomics of spike, envelope, and nucleocapsid protein of severe acute respiratory syndrome coronavirus 2. Afri Health Sci. 2023;23(3): 384-399. https://dx.doi.org/10.4314/ahs.v23i3.45

### Introduction

Today, the entire world is facing a new global public health crisis. In December 2019, an upsurge sprang up in Wuhan, China, where a multitudinous of patients were suffering from a bizarre agglomeration of respiratory afflictions <sup>1</sup>. The International Committee on Taxonomy of Viruses (ICTV) cleped the virus as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) on February 11<sup>th</sup>, 2020, and the ailment effectuated by SARS-CoV-2 was cleped as COVID-19 by the World Health Organization (WHO) <sup>2,3</sup>. The SARS-CoV-2 virus, which first appeared in Wuhan, progressively propagated throughout China and to other developed and developing

### Corresponding author:

Anwar Ullah, Faculty of Science, Department of Biosciences, COMSATS University Islamabad, Park Road, Chak Shahzad, Islamabad, Pakistan Tel: (+92) 334-0111362 Email: anwar.ms90@yahoo.com countries throughout the world, ultimately erupting into a global pandemic, as proclaimed by WHO on March 11<sup>th</sup>, 2020 <sup>2,4</sup>. The pandemic has inflicted a myriad of people. As stated by Worldometer, an evaluation of 201, 064, 392 coronavirus cases, 4, 271, 371 demises, and 181, 054, 513 recovered have been reported (accessed on August 5<sup>th</sup>, 2021) <sup>5</sup>.

Among all other nations, Pakistan has also been inflicted by the Wuhan virus. On February 26th, 2020, the Ministry of Health, Government of Pakistan, promulgated the first case of COVID-19 in Karachi. Another case was announced by the Federal Ministry of Health in Islamabad on the same day <sup>6</sup>. As of 5th August 2021, there were 105, 3660 confirmed cases of which 23, 635 have died, 952, 616 have recovered, and 4050 are critical. The SARS-CoV-2 has greatly affected the Sindh province and is positioned at number 01 in proportion to the quota of COVID cases (with 392, 433 cases as of 5th August 2021) <sup>7</sup>. Presently, Pakistan is enduring a fourth wave of COVID-19 endangered by the delta variant (B.1.617.2)

African Health Science © 2023 Khan SS et al. Licensee African Health Sciences. This is an Open Access article distributed under the terms of the Creative commons Attribution License (https://creativecommons.org/licenses/BY/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

also avowed as the Indian variant <sup>8</sup>. The newly discovered variants of COVID-19, that is, South Africa and Brazilian variants, have been identified in Karachi during the third wave of COVID-19 <sup>9</sup>. In another report, researchers from Karachi disclosed that 70% of COVID-19 infirmities throughout Pakistan were attributed to the UK variant <sup>10</sup>. However, based on new findings and data, researchers from the National Institute of Health, Islamabad, have claimed that 50% of COVID-19 manifestations in Pakistan are now attributed to the Delta (Indian) variant <sup>11</sup>.

Genomics and Proteomics of COVID-19 have been delineated in relevant studies. Genomic studies have imparted that the structure of SARS-CoV-2 is made up of single-strand ribonucleic acid (ssRNA), comprising of four structural proteins, namely, Spike protein (S), an Envelope protein (E), Membrane protein (M), and Nucleocapsid (N) protein. In line with evolutionary studies, it portrays that it exhibits 96.2% homogenous to bat coronaviruses and 79.5% (<50% sequence identity) with SARS-CoV and MERS-CoV<sup>12,13</sup>. SARS-CoV-2 utilizes a distinct protein known as S-protein to bind specifically to the Angiotensin-Converting Enzyme 2 (ACE 2) of the host cell 14. Asides from spike protein, SARS-CoV-2 also embodies additional structural proteins such as E-protein and N-protein. E-protein is crucial for viral genome packaging and the synthesis of ion channels (IC), which are vital for virus-host contact and are frequently associated with pathogenicity 15,16. In contrast, the N-protein serves a variety of roles in the CoV virus replication. For example, in SARS-CoV, the N-protein has been shown to adhere to viral RNA and assemble it into ribonucleoprotein (RNP) complexes. 17,18

Due to the extremely high mutation frequencies, SARS-CoV-2 is naturally prone to mutations, ensuing in genomic diversity <sup>19</sup>. Because of the virus's fast evolution, generating vaccines and therapeutics may be problematic; consequently, comparative genomics of viruses isolated throughout time and in different regional areas is essential. A comparative study of the genomes of distinct SARS-CoV-2, SARS-CoV, and bat SARS-like CoV strains isolated would enable the identification and evaluation of the variable and preserved genomic regions moreover, this scientific understanding can help create efficacious vaccines and molecular epidemiological tracking. Thus, this research was implemented to investigate the evolution of

*SARS-CoV-2* heterogeneity in other countries worldwide afflicted with COVID-19.

# Methods

# Ethical Approval

Ethical approval was attained by the Dean, Faculty of Science of COMSATS University Islamabad, Main Campus, Islamabad, Pakistan. This comparative genomic research was effectuated in August 2021 at the Department of Biosciences. This research entails no patient or animal models and was implemented by computational tools/ software.

### Sequence Retrieval

Nucleotide sequences of SARS-CoV-2 from Pakistan (Accession # MT240479.1), India (Accession # MZ558086.1), UK (Accession # MZ376737.1), South Africa (Accession # MZ202314.1), and Brazil (Accession # MZ397163.1) were retrieved from the NCBI Virus database. Likewise, the nucleotide sequences of SARS-CoV (Accession # NC\_004718.3), and bat SARS-like CoV (Accession # MG772934.1) were also retrieved from the NCBI Virus database (https://www.ncbi.nlm.nih.gov/ labs/virus/vssi/#/virus?SeqType\_s=Nucleotide&VirusLineage\_ss=SARS-CoV-2,%20taxid:2697049).Correspondingly, in synchrony to nucleotide sequences, protein sequences of S, E, and N-proteins of SARS-CoV-2 variants, SARS-CoV, and bat SARS-like CoV were also retrieved from the NCBI Virus database. SARS-CoV-2 from Wuhan, China, was used as a reference (Accession # NC\_045512.2).

# **Phylogenetic Analysis**

Molecular evolutionary genetics analysis (MEGA) 11 version 11.0.10 software was exploited to generate and visualize the phylogenetic tree. A maximum likelihood approach with a bootstrap of 1000 replicates was employed for determining the best interfacing tree. The substitution model was computed to find the best DNA substitution model for the phylogenetic analysis.

# Mutational Analysis and Impact on Protein Stability

Genome detective Coronavirus typing tool (Version 1.17), which enables rapid identification and depiction of novel COVID genomes, executed mutational analysis. This tool allows for the input of up to 2000 sequences and completes the probe within seconds. This typing tool

has been approved for classifying novel *SARS-CoV-2* among COVID species. We employed this typing method to detect mutation in the genome of *SARS-CoV-2* variants S, E, and N in comparison to Wuhan (China) *SARS-CoV-2*. Finally, we utilized an in-house software built in Perl and Python, MUpro server to forecast the impact of mutation on protein stability on *SARS-CoV-2* S, E, and N-protein.

#### Results

#### **Evolutionary analysis**

A phylogenetic probe was performed to determine variation among the genomes of *SARS-CoV-2*, *SARS-CoV*, and bat *SARS-like CoV*. MEGA 11 version 11.0.10 software was exploited for generating a phylogenetic tree. A maximum likelihood statistical method with a bootstrap of 1000 replicates was utilized to discern the best interfacing tree. The substitution model was computed, and the General Time Reversible + proportion of Invariant sites (GTR+I) model was found to be the best substitution model for the evolutionary analysis.

Our evolutionary research is evident that all the *SARS*-*CoV-2* variants form a clade that is all closely related to each other and to bat *SARS-like CoV*, which successively is related to *SARS-CoV*. This analysis reveals that only bat *SARS-like CoV* has a very close evolutionary relationship with *SARS-CoV-2* encountering an independent bifurcation from bat *SARS-like CoV*. The branch length of *SARS-CoV* portrays that it has diverged very early from bat SARS-like CoV. This evolutionary analysis strongly concurs with the fact that bats are the primary host for coronavirus evolution (Figure 1). The bootstrap values (100%) as depicted in Figure 1 robustly support this analysis.



**Figure 1:** Phylogenetic Tree displaying an evolutionary variation of *SARS-CoV-2, SARS-CoV*, and bat *SARS-like CoV* complete (nucleotide) genome. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model <sup>57</sup>. The tree with the highest log likelihood (-72607.67) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The rate variation model allowed for some sites to be evolutionarily invariable ([+1], 46.45% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). Evolutionary analyses were conducted in MEGA11 <sup>58</sup>.

#### **Comparative Genomics**

Multalin online tool <sup>20</sup> and Nucleotide BLAST (BLASTN) were used for comparative genomics, employing the *SARS-CoV-2* variant from Wuhan as a reference (Figure 2-4). The results of BLASTN have disclosed that all the SARS-CoV-2 S, E, and N-genomes exhibit 99% homogenous with Wuhan (China) *SARS-CoV-2* excluding the E-genomes of *SARS-CoV-2* (Pakistan, UK, Brazil, and

0.02

India) and N-genome of *SARS-CoV-2* (Pakistan) which displays 100% similarity. In addition, *SARS-CoV* S, E, and N-genome display's a 78%, 94%, and 89% analogous with Wuhan (China) *SARS-CoV-2* whereas bat *SARS-like CoV* S, E, and N-genome displays 83%, 99%, and 91% homogenous (Supplementary Table 1-3). These results also authenticate the evolutionary variation among the genomes of *SARS-CoV-2*, *SARS-CoV*, and bat *SARS-like CoV* as delineated in Figure 1.

					50	60
CHINA PAKISTAN	MFVFLVLLPLVS: MFVFLVLLPLVS	SQCVNLTTRTQ SQCVNLTTRTQ	LPPAYTNSFTF LPPAYTNSFTF	RGVYYPDKVFI RGVYYPDKVFI	RSSVLHSTQDI RSSVLHSTQDI	LFLPFFS LFLPFFS
SOUTHAFRICA	MFVFLVLLPLVS	SQCVNFTTRTQ	LPPAYTNSFTF	GVYYPDKVFI	RSSVLHSTQD	LFLPFFS
DKATT	MFVFLVLLPLVS.	SOCVNE INKIQ.	LP <mark>SATINSFIF</mark> LPPAYTNSFTI	RGVYYPDKVFI	RSSVLHSTQD	LFLPFFS LFLPFFS
INDIA	MFVFLVLLPLVS	SQCVNL <mark>X</mark> TRTQ	LPPAYTNSFT	RGVYYPDKVFI	RSSVLHSTQD	LFLPFFS
	70	80	90	100	110	120
CHINA	NVTWFHAIHVSG	TNGTKRFDNPV	LPFNDGVYFAS	STEKSNIIRG	WIFGTTLDSK	TOSLLIV
PAKISTAN	NVTWFHAIHVSG	INGTKRFDNPV	LPFNDGVYFAS	STEKSNIIRG	WIFGTTLDSK	TQSLLIV
SOUTHAFRICA	NVTWFHAIHVSG	ING TKRFANPV	LPFNDGVYFAS	STEKSNIIRG	WIFGTTLDSK	FQSLLIV
UK	NVTWFHAIHVSG	TNGTKRFDNPV TNGTKRFDNPV	LPFNDGVIFA: LPFNDGVYFA:	STEKSNIIRG	WIFGTTLDSK'	TOSLLIV
INDIA	NVTWFHAIHVSG	INGTKRFDNPV	LPFNDGVYFAS	STEKSN <mark>XXXX</mark>	XXXXXXXXXXK	rõslliv
	130	140	150	160	170	190
CHINA	NNATNVVIKVCE	FOFCNDPFLGV	YYHKNNKSWMI	ESEFRVYSSA	NNCTFEYVSQ	PELMDLE
PAKISTAN	NNATNVVIKVCE	FQFCNDPFLGV	YYHKNNKSWME	SEFRVYSSA	NNCTFEYVSQ	PFLMDLE
SOUTHAFRICA	NNATNVVIKVCE	FQFCNDPFLGV	YYHKNNKSWME	SEFRVYSSA	NNCTFEYVSQ1	PFLMDLE
UK	NNATNVVIKVCE NNATNVVIKVCE	FOFCNDPFLGV	Y <mark>HK-NNKSWM</mark>	SEFRVISSA	NNCTFEYVSQ NNCTFEYVSQ	PFLMDLE
INDIA	NNATNVVIKVCE	FQFCNDPFLDF	YYHKNNKSWME	SEFRVYSSA	NNCTFEYVSQ	PFLMDLE
		200	210			
PAKISTAN	GKQGNFKNLREF GKOGNFKNLREF	/FKNIDGYFKI VFKNIDGYFKI	YSKHTPINLVR YSKHTPINLVR	RDLPQGFSALE	:PLVDLPIGI :PLVDLPIGI	NITREQT NITREOT
SOUTHAFRICA	GKQGNFKNLREF	VFKNIDGYFKI	YSKHTPINLVF	GLPQGFSALF	SPLVDLPIGIN	NITRFQ-
BRAZIL	GKQGNFKNL <mark>S</mark> EFV	/FKNIDGYFKI	YSKHTPINLVR	NDLPQGFSALE	SPLVDLPIGIN	NITRFQT
UK TNDIA	GKQGNEKNLREE GKOGNEKNLREE	/FKNIDGYFKI VFKNIDGYFKI	YSKHTPINLVR YSKHTPINLVR	RDLPQGF SALE	PLVDLPIGI	NITRFQT NITRFOT
СПТИХ				280		
PAKISTAN	LLALHRSYLTPGI	DSSSGWTAGAA	AYYVGYLQPRI	FLLKYNENG	fitdavdcali	DPLSETK DPLSETK
SOUTHAFRICA	TLHRSYLTPGI	DSSSGWTAGAA	AYYVGYLQPRI	FLLKYNENGT	TITDAVDCALI	DPLSETK
BRAZIL	LLALHRSYLTPGI	DSSSGWTAGAAA	AYYVGYLQPRI	FLLKYNENG	LITDAVDCALI	DPLSETK
INDIA	LLALHRSYLTPGI	DSSSGWTAGAA	AYYVGYLQPRI	FLLKINENG	<u>TIT</u> DAVDCALI	DPLSEIK DPLSETK
СИТИЛ						
PAKISTAN	CTLKSFTVEKGI	YOT SNFRVQFTI	ESIVRFPNITN	ILCPFGEVFN/	ATRFASVIAWI	NRKRISN NRKRISN
SOUTHAFRICA	CTLKSFTVEKGI	YQTSNFRVQPTI	ESIVRFPNITN	ILCPFGEVFNA	ATRFASVYAWI	NRKRISN
BRAZIL	CTLKSFTVEKGI	IQT SNFRVQPTI	ESIVRFPNITN	ILCPFGEVFNA	ATRFASVYAW	NRKRISN
UK INDIA	CTLKSFIVERGI	YOT SNFRVQPTI YOT SNFRVQPTI	ESIVREPNIIK ESIVRFPNITN	ILCPFGEVFN/	ATRFASVIAWI ATRFASVYAWI	NRKRISN

Figure 2: Multiple Sequence Alignment of SARS-CoV-2, SARS-CoV, and bat SARS-like CoV Spike Protein.

CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	CVADYSVLYNS CVADYSVLYNS CVADYSVLYNS CVADYSVLYNS CVADYSVLYNS CVADYSVLYNS CVADYSVLYNS	380 SASFSTFKCY SASFSTFKCY SASFSTFKCY SASFSTFKCY SASFSTFKCY	) YGVSPTKL YGVSPTKL YGVSPTKL YGVSPTKL YGVSPTKL	390 NDLCFTNVY NDLCFTNVY NDLCFTNVY NDLCFTNVY NDLCFTNVY	400 ADSFVIRG ADSFVIRG ADSFVIRG ADSFVIRG ADSFVIRG	410 DEVRQIAPG DEVRQIAPG DEVRQIAPG DEVRQIAPG DEVRQIAPG DEVRQIAPG	420 QTGKIAD QTGKIAD QTG <mark>NIAD</mark> QTG <mark>T</mark> IAD QTGKIAD
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	YNYKLPDDFTC YNYKLPDDFTC YNYKLPDDFTC YNYKLPDDFTC YNYKLPDDFTC YNYKLPDDFTC	GCVIAWNSNN GCVIAWNSNN GCVIAWNSNN GCVIAWNSNN GCVIAWNSNN GCVIAWNSNN	ILDSKVGGI ILDSKVGGI ILDSKVGGI ILDSKVGGI ILDSKVGGI ILDSKVGGI	450 NYNYLYRLF NYNYLYRLF NYNYLYRLF NYNYLYRLF NYNYLYRLF NYNY <mark>R</mark> YRLF	.460 RKSNLKPFI RKSNLKPFI RKSNLKPFI RKSNLKPFI RKSNLKPFI	ERDISTEIY SRDISTEIY SRDISTEIY SRDISTEIY SRDISTEIY SRDISTEIY SRDISTEIY	2AGSTPC 2AGSTPC 2AGSTPC 2AGSTPC 2AGSTPC 2AGSTPC 2AGS <mark>K</mark> PC
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA		>LQSYGFQP7 PLQSYGFQP7 PLQSYGFQP7 PLQSYGFQP7 PLQSYGFQP7 PLQSYGFQP7 PLQSYGFQP7	NGVGYQP NGVGYQP YGVGYQP YGVGYQP YGVGYQP NGVGYQP	510 YRVVVLSFE YRVVVLSFE YRVVVLSFE YRVVVLSFE YRVVVLSFE YRVVVLSFE	.520 LLHAPATVO LLHAPATVO LLHAPATVO LLHAPATVO LLHAPATVO		540 VKNKCVN VKNKCVN VKNKCVN VKNKCVN VKNKCVN
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	FNFNGLTGTGV FNFNGLTGTGV FNFNGLTGTGV FNFNGLTGTGV FNFNGLTGTGV FNFNGLTGTGV	/LTESNKKFI /LTESNKKFI /LTESNKKFI /LTESNKKFI /LTESNKKFI /LTESNKKFI	) .PFQQFGRI .PFQQFGRI .PFQQFGRI .PFQQFGRI .PFQQFGRI .PFQQFGRI	570 DIADTTDAV DIADTTDAV DIADTTDAV DIADTTDAV DIDDTTDAV DIADTTDAV	.580 RDPQTLEII RDPQTLEII RDPQTLEII RDPQTLEII RDPQTLEII RDPQTLEII	LDITPCSFG LDITPCSFG LDITPCSFG LDITPCSFG LDITPCSFG LDITPCSFG LDITPCSFG	600 GVSVITP GVSVITP GVSVITP GVSVITP GVSVITP GVSVITP
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	GTNTSNQVAVL GTNTSNQVAVL GTNTSNQVAVL GTNTSNQVAVL GTNTSNQVAVL GTNTSNQVAVL GTNTSNQVAVL	YQDVNCTEV YQDVNCTEV YQGVNCTEV YQGVNCTEV YQGVNCTEV YQGVNCTEV YQGVNCTEV	PVAIHADÇ PVAIHADÇ PVAIHADÇ PVAIHADÇ PVAIHADÇ PVAIHADÇ	30 2LTPTWRVY 2LTPTWRVY 2LTPTWRVY 2LTPTWRVY 2LTPTWRVY 2LTPTWRVY	.640 STGSNVFQT STGSNVFQT STGSNVFQT STGSNVFQT STGSNVFQT STGSNVFQT	650 'RAGCLIGAE 'RAGCLIGAE 'RAGCLIGAE 'RAGCLIGAE 'RAGCLIGAE 'RAGCLIGAE	EHVNNSY EHVNNSY EHVNNSY EHVNNSY EHVNNSY EHVNNSY
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	ECDIPIGAGIC ECDIPIGAGIC ECDIPIGAGIC ECDIPIGAGIC ECDIPIGAGIC ECDIPIGAGIC	ASYQTQTNS ASYQTQTNS ASYQTQTNS ASYQTHTNS ASYQTQTNS ASYQTQTNS ASYQTQTNS	PRRARSVA PRRARSVA PWRARSVA PRRARSVA HRRARSVA RRRARSVA	590 ASQSIIAYT ASQSIIAYT ASQSIIAYT ASQSIIAYT ASQSIIAYT	.700 MSLGAENSV MSLGAENSV MSLG <mark>V</mark> ENSV MSLGAENSV MSLGAENSV	710 YAYSNNSIAI YAYSNNSIAI YAYSNNSIAI YAYSNNSIAI YAYSNNSIAI YAYSNNSIAI	720 PTNFTI PTNFTI PTNFTI PTNFTI PTNFTI PTNFTI

Figure 2: continued

		0	.740	750	760.	770.	
CHINA	SVTTEILPV	SMTKTSV	DCTMYICGI	OSTECSNLL	LQYGSFCT	QLNRALTGIA	VEQDKNTQE
PAKISTAN	SVTTEILPV	SMTKTSV	DCTMYICGI	OSTECSNLL	LOYGSFCT	JLNRALTGIA	VEODKNTOE
SOUTHAFRICA	SVTTEILPV	SMTKTSV	DCTMYICGI	OSTECSNLL	LOYGSFCT	~ DLNRALTGIA	VEODKNTÕE
BRAZIL	SVTTEILPV	SMTKTSV	DCTMYICGI	OSTECSNLL	LÔYGSFCT	~ DINRALTGIA'	VEÔDKNTÔE
IIK	SVTTETLPV	SMTKTSV	DCTMYICGI	OSTECSNLL	LOYGSECT	OLNRALTGTA	VEODKNTOE
TNDTA	SVTTEILPV	SMTKTSV		OSTECSNI.I.	LOYGSECT	DINRALTGIA	VEODKNTOE
110111	SVIIIIIV	5111(15)	0011111001	DELICENTE	nőiger (†		A P Z P I (14 1 Z P
	79	0	800	810	820	830	840
CHINA	VEAOVKOTY	VTPPTKD	FGGENESO	TTEDESKES	KRSETEDLI	LENKVTLADA	GETKOYGDC
PAKISTAN	VFAOVKOIY	KTPPIKD	FGGENESO	LTDDDSKDS	KRSEIEDI.	LENKVTLADA(	SEIKONGDC
COUTHAEDICA	VEVOIKOIA	KTI I IKD.	FCCENESO	TT DDDGVDG	KDGEIEDI I		CEIKOVGDC
DDATT	VEAQVIQUU	KILLIKD.	FCCENESQ	LT DDDGKDG	NRSFIEDLI		JF INQIODC
DKAZIL	VEAQVKQII	KIFFIKD.	FCCENESO	LLEDESKES	VDGETEDI		JF IKQIGDC
UN	VEAQVKQII	KIFFIKD.	rggrnrsv.	LLEDESKES	KDGPIPDI		JE IKQIGDC
INDIA	VFAQVKQII	KTPPIKDI	EGGENESQ1	LEPDPSKPS	KRSFIEDLI	JENKVTLADA	JF IKQIGDC
		) <b></b>	860	870			
CHINA	LGDIAARDL	ICAQKFN	GLTVLPPLL	TDEMIAQY	TSALLAGTI	TSGWTFGAGA	ALQIPFAM
PAKISTAN	LGDIAARDLI	ICAQKFNC	GLTVLPPLL	TDEMIAQY	<b>TSALLAGTI</b>	TSGWTFGAGA	ALQIPFAM
SOUTHAFRICA	LGDIAARDLI	ICAQKFN	GLTVLPPLL	TDEMIAQY	<b>TSALLAGTI</b>	TSGWTFGAGA	ALQIPFAM
BRAZIL	LGDIAARDL	ICAQKFNO	GLTVLPPLL	TDEMIAQY	<b>TSALLAGTI</b>	TSGWTFGAGA	ALQIPFAM
UK	LGDIAARDL	ICAQKFN	GLTVLPPLL	TDEMIAQY	<b>TSALLAGTI</b>	TSGWTFGAGA	ALQIPFAM
INDIA	LGDIAARDLI	ICAQKENC	GLTVLPPLL	TDEMIAQY	TSALLAGTI	TSGWTFGAGA	ALQIPFAM
			920	930			
CHINA		) GVTQNVLY	920 ENQKLIAN	930 IQFNSAIGK	940 IQDSLSSTA	950 SALGKLQDVV	960 VNQNAQALN
CHINA PAKISTAN	QMAYRFNGIO	) GVTQNVLY GVTQNVLY	920 ENQKLIAN ENQKLIAN	930 IQFNSAIGK IQFNSAIGK	940 IQDSLSSTA IQDSLSSTA	950. SALGKLQDVV SALGKLQDVV	960 VNQNAQALN VNQNAQALN
CHINA PAKISTAN SOUTHAFRICA	QMAYRFNGIC QMAYRFNGIC QMAYRFNGIC OMAYRFNGIC	) GVTQNVLY GVTQNVLY GVTQNVLY	920 ENQKLIAN ENQKLIAN ENQKLIAN	930 IQFNSAIGK IQFNSAIGK IQFNSAIGK	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV	960 /NQNAQALN /NQNAQALN /NONAQALN
CHINA PAKISTAN SOUTHAFRICA BRAZIL	QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO	) GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY	920 ENQKLIAN ENQKLIAN ENQKLIAN ENQKLIAN	930 IQFNSAIGK IQFNSAIGK IQFNSAIGK IQFNSAIGK	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA	SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV	960 WQNAQALN WQNAQALN WQNAQALN WQNAQALN
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK	2000 2000 2000 2000 2000 2000 2000 200	O GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY	920 ENQKLIAN ENQKLIAN ENQKLIAN ENQKLIAN	930 IQFNSAIGK IQFNSAIGK IQFNSAIGK IQFNSAIGK IOFNSAIGK	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA	950 SALGKLQDV SALGKLQDV SALGKLQDV SALGKLQDV	960 /NQNAQALN /NQNAQALN /NQNAQALN /NQNAQALN /NONAOALN
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDTA	QMAYRFNGIC QMAYRFNGIC QMAYRFNGIC QMAYRFNGIC QMAYRFNGIC OMAYRFNGIC	O GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY	920 (ENQKLIAN (ENQKLIAN (ENQKLIAN (ENQKLIAN (ENQKLIAN (ENQKLIAN	930 IQFNSAIGK IQFNSAIGK IQFNSAIGK IQFNSAIGK IQFNSAIGK	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IODSLSSTA	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV	960 /NQNAQALN /NQNAQALN /NQNAQALN /NQNAQALN /NQNAQALN /NONAQALN
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	QMAYRFNGIC QMAYRFNGIC QMAYRFNGIC QMAYRFNGIC QMAYRFNGIC QMAYRFNGIC	O GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY	920 YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN	.930 IQFNSAIGK IQFNSAIGK IQFNSAIGK IQFNSAIGK IQFNSAIGK	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNV	960 /NQNAQALN /NQNAQALN /NQNAQALN /NQNAQALN /NQNAQALN
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	QMAYRFNGI QMAYRFNGI QMAYRFNGI QMAYRFNGI QMAYRFNGI QMAYRFNGI 2000	O GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY	920 YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN 980	930 IQFNSAIGK IQFNSAIGK IQFNSAIGK IQFNSAIGK IQFNSAIGK	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV	960 /NQNAQALN /NQNAQALN /NQNAQALN /NQNAQALN /NQNAQALN
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA		) SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY O SVTQNVLY O SVTQNVLY	920. YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN	930 JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK 990	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA 1000	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV 1010 SSLOTYVTOOI	960 NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN		D SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY D FGAISSVI FGAISSVI	920. ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN 980. .NDILSRLD	930 JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK 990 DKVEAEVQI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA 1000 DRLITGRLQ	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV 1010 SSLQTYVTQQI	960 MQNAQALN NQNAQALN NQNAQALN MQNAQALN MQNAQALN MQNAQALN 1020 .IRAAEIRA .IRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA		D SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY FGAISSVI FGAISSVI FGAISSVI	920 YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN 980 NDILSRLD NDILSRLD	930 JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK 990 DKVEAEVQI DKVEAEVQI DKVEAEVQI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA 1000 DRLITGRLQ DRLITGRLQ DRLITGRLQ	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV SALGKLQNVV SALGKLQNVV SSLGTVVTQQI SSLQTVVTQQI SSLQTVVTQQI	960 MQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL		D SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY FGAISSVI FGAISSVI FGAISSVI	920 ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN MDILSRLD NDILSRLD NDILSRLD	930 JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK 990 DKVEAEVQI DKVEAEVQI DKVEAEVQI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA 1000 DRLITGRLQ DRLITGRLQ DRLITGRLQ	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV SALGKLQNVV SLGTYVTQQI SSLQTYVTQQI SSLQTYVTQQI SSLQTYVTQQI	960 WQNAQALN WQNAQALN WQNAQALN WQNAQALN WQNAQALN WQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL		GUTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY FGAISSVI FGAISSVI FGAISSVI	920 ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN SOBO ZENQKLIAN SOBO ZENDILSRLD JOILSRLD JOILSRLD	930 JQFNSAIGK JQFNSAI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA 1000 DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV SALGKLQNVV SLGTYVTQQI SLQTYVTQQI SSLQTYVTQQI SSLQTYVTQQI	960 WQNAQALN WQNAQALN WQNAQALN WQNAQALN WQNAQALN WQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK		D GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY C GAISSVI FGAISSVI FGAISSVI FGAISSVI FGAISSVI FGAISSVI	920 YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENDILSRLD NDILSRLD NDILSRLD NDILSRLD	930 JQFNSAIGK JQFNSAI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV SALGKLQNVV SALGTYVTQQI SLQTYVTQQI SLQTYVTQQI SSLQTYVTQQI	960 MQNAQALN MQNAQALN MQNAQALN MQNAQALN MQNAQALN MQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO DMAYRFNGIO TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE	D GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY C GAISSVI FGAISSVI FGAISSVI FGAISSVI FGAISSVI	920 YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN YENQKLIAN	930 JQFNSAIGK JQFNSAI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA ORLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV SALGKLQNVV SALGYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI	960 NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA	910 QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO CMAYRFNGIO TLVKQLSSNI TLVKQLSSNI TLVKQLSSNI TLVKQLSSNI TLVKQLSSNI TLVKQLSSNI 1030	DD GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY D FGAISSVI FGAISSVI FGAISSVI FGAISSVI FGAISSVI	920 ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN 980 NDILSRLD NDILSRLD NDILSRLD NDILSRLD NDILSRLD NDILSRLD NDILSRLD NDILSRLD	930 JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK 990 PKVEAEVQI PKVEAEVQI PKVEAEVQI PKVEAEVQI PKVEAEVQI PKVEAEVQI PKVEAEVQI PKVEAEVQI PKVEAEVQI PKVEAEVQI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA ORLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQ NVV 1010 SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI 1070	960 NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA	910 QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO CMAYRFNGIO TLVKQLSSNI TLVKQLSSNI TLVKQLSSNI TLVKQLSSNI TLVKQLSSNI TLVKQLSSNI SANLAATKMS	D GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY C GAISSVI GGAISSVI FGAISSVI FGAISSVI FGAISSVI FGAISSVI GGAISSVI D D C C C C C C C C C C C C C	920 (ENQKLIAN (ENQKLIAN (ENQKLIAN (ENQKLIAN (ENQKLIAN (ENQKLIAN (ENQKLIAN 0980 NDILSRLD NDILSRLD NDILSRLD NDILSRLD NDILSRLD 040 KRVDFCGK	930 JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK JQFNSAIGK SVEAEVQI SKVEAEVQI SKVEAEVQI SVEAEVQI SVEAEVQI SVEAEVQI SVEAEVQI SVEAEVQI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA ORLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV SALGKLQNVV SALGKLQNVV SALGTVVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI 1070	960 NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN	910 QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE SANLAATKMS	D GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY GVTQNVLY D GAISSVI FGAISSVI FGAISSVI FGAISSVI FGAISSVI GAISSVI GAISSVI D D SECVLGQS SECVLGQS	920 (ENQKLIAN (E	930 QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK VEAEVQI KVEAEVQI KVEAEVQI KVEAEVQI KVEAEVQI QKVEAEVQI QKVEAEVQI QKVEAEVQI QYHLMSFPC CGYHLMSFPC	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ SAPHGVVF QSAPHGVVF	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV SALGKLQNVV SALGTVVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI 1070 CHVTYVPAQF	960 NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA .IRAAEIRA 1080 .KNFTTAPA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA	910 QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE SANLAATKMS SANLAATKMS	D SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SECVLGQS SECVLGQS	920 (ENQKLIAN (ENQKL	930 QFNSAIGK QFNSAI	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA ORLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ SAPHGVVF QSAPHGVVF	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQNVV SALGKLQNVV SALGKLQNVV SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI 1070 THVTYVPAQE LHVTYVPAQE	960 NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL	910 QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE TLVKQLSSNE SANLAATKMS SANLAATKMS SANLAATKMS	D SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SECVLGQS SECVLGQS SECVLGQS	920 ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN MOILSRLD NDILSRLD NDILSRLD NDILSRLD SKRVDFCGK SKRVDFCGK SKRVDFCGK	930 QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAEVQI XVEAEVQI XVEAEVQI XVEAEVQI XVEAEVQI QKVEAEVQI CGYHLMSFPC GYHLMSFPC GYHLMSFPC	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA ORLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ SAPHGVVF QSAPHGVVF DSAPHGVVF	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI CLHVTYVPAQE LHVTYVPAQE LHVTYVPAQE	960 MQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA SKNFTTAPA KNFTTAPA KNFTTAPA
CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK INDIA CHINA PAKISTAN SOUTHAFRICA BRAZIL UK	91( QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO QMAYRFNGIO TLVKQLSSNF TLVKQLSSNF TLVKQLSSNF TLVKQLSSNF TLVKQLSSNF TLVKQLSSNF TLVKQLSSNF SANLAATKMS SANLAATKMS SANLAATKMS SANLAATKMS	D SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SVTQNVLY SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SGAISSVI SECVLGQS SECVLGQS SECVLGQS SECVLGQS	920 ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN ZENQKLIAN DILSRLD NDILSRLD NDILSRLD NDILSRLD NDILSRLD SKRVDFCGK SKRVDFCGK SKRVDFCGK SKRVDFCGK	930 QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAIGK QFNSAEVQI OKVEAEVQI O	940 IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA IQDSLSSTA 1000 DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ DRLITGRLQ SAPHGVVF QSAPHGVVF QSAPHGVVF SAPHGVVF	950 SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGKLQDVV SALGYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI SLQTYVTQQI CLVTYVPAQE LHVTYVPAQE LHVTYVPAQE	960 MQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN NQNAQALN 1020 JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA JIRAAEIRA SKNFTTAPA SKNFTTAPA SKNFTTAPA SKNFTTAPA SKNFTTAPA

Figure 2: continued

	1090		1110	1120	1130	1140
CHINA	ICHDGKAHFPRE	GVFVSNGTHWF	VTQRNFYEPQ]	IITTDNTFVS	GNCDVVIGIV	NTVYDP
PAKISTAN	ICHDGKAHFPRE	GVFVSNGTHWF	VTQRNFYEPQ	IITTDNTFVS	GNCDVVIGIV	NTVYDP
SOUTHAFRICA	ICHDGKAHFPRE	GVFVSNGTHWF	VTQRNFYEPQ	IITTDNTFVS	GNCDVVIGIV	NTVYDP
BRAZIL	ICHDGKAHFPRE	GVFVSNGTHWF	VTQRNFYEPQ	IITTDNTFVS	GNCDVVIGIV	NNTVYDP
UK	ICHDGKAHFPRE	GVFVSNGTHWF	VTQRNFYEPQ	IITT <mark>H</mark> NTFVS(	GNCDVVIGIV	NNTVYDP
INDIA	ICHDGKAHFPRE	GVFVSNGTHWF	VTQRNFYEPQ]	IITTONTFVS	GNCDVVIGIV	NNTVYDP
		1160				1200
CHINA	LQPELDSFKEEL	DKYFKNHTSPD	VDLGDISGINA	ASVVNIQKEII	DRLNEVAKNLI	NESLIDL
PAKISTAN	LQPELDSFKEEL	DKYFKNHTSPD	VDLGDISGINA	ASVVNIQKEII	DRLNEVAKNLI	NESLIDL
SOUTHAFRICA	LQPELDSFKEEL	DKYFKNHTSPD	VDLGDISGINA	ASVVNIQKEII	DRLNEVAKNLI	NESLIDL
BRAZIL	LQPELDSFKEEL	DKYFKNHTSPD	VDLGDISGINA	AS <mark>F</mark> VNIQKEII	DRLNEVAKNLI	NESLIDL
UK	LQPELDSFKEEL	DKYFKNHTSPD	VDLGDISGINA	ASVVNIQKEII	DRLNEVAKNLI	NESLIDL
INDIA	LQPELDSFKEEL	DKYFKNHTSPD	VDLGDISGINA	ASVVNIQKEII	DRLNEVAKNLI	NESLIDL
		1220	1230	1240	1250	1260
CHINA	QELGKYEQYIKW	PWYIWLGFIAG	LIAIVMVTIMI	LCCMTSCCSCI	LKGCCSCGSC	CKFDEDD
PAKISTAN	QELGKYEQYIKW	PWYIWLGFIAG	LIAIVMVTIMI	LCCMTSCCSCI	LKGCCSCGSC	CKFDEDD
SOUTHAFRICA	QELGKYEQYIKW	PWYIWLGFIAG	LIAIVMVTIMI	LCCMTSCCSCI	LKGCCSCGSC	CKFDEDD
BRAZIL	OELGKYEOYIKW	PWYIWLGFIAG	LIAIVMVTIM		LKGCCSCGSC	CKFDEDD
UK	QELGKIEQIIKW	PWIIWLGFIAG				CKEDEDD
INDIA	QELGKIEQIIKW	PWIIWLGFIAG	LIAIVMVIIM	LUCMISUUSU.	LKGCCSCGSCI	SKEDEDD
	1270					
CHINA	SEPVLKGVKLHY	· T				
PAKISTAN	SEPVLKGVKLHY	т Т				
SOUTHAFRICA	SEPVLKGVKLHY	т Т				
BRAZIT.	SEPVLKGVKLHY	т Т				
IIK	SEPVLKGVKLHY	т Т				
INDIA	SEPVLKGVKLHY	T				

### Figure 2: continued

		20		40		60
CHINA	MYSFVSEETGTLIVN	SVLLFLAFV	VFLLVTLAILT.	ALRLCAYCCN	IVNVSLVKP	SFYVYS
PAKISTAN	MYSFVSEETGTLIVN	SVLLFLAFV	VFLLVTLAILT.	ALRLCAYCCN	IVNVSLVKP	SFYVYS
INDIA	MYSFVSEETGTLIVN	SVLLFLAFV	VFLLVTLAILT.	ALRLCAYCCN	IVNVSLVKP	SFYVYS
UK	MYSFVSEETGTLIVN	SVLLFLAFV	VFLLVTLAILT.	ALRLCAYCCN	IVNVSLVKP	SFYVYS
BRAZIL	MYSFVSEETGTLIVN	SVLLFLAFV	VFLLVTLAILT.	ALRLCAYCCN	IVNVSLVKP	SFYVYS
SOUTHAFRICA	MYSFVSEETGTLIVN	SVLLFLAFV	VFLLVTLAILT.	ALRLCAYCCN	IVNVSLVKP	SFYVYS
	<u></u> 70					
CHINA	RVKNLNSSRVPDLLV					
PAKISTAN	RVKNLNSSRVPDLLV					
INDIA	RVKNLNSSRVPDLLV					
UK	RVKNLNSSRVPDLLV					
BRAZIL	RVKNLNSSRVPDLLV					
SOUTHAFRICA	RVKNLNSSRV <mark>L</mark> DLLV					

Figure 3: Multiple Sequence Alignment of SARS-CoV-2, SARS-CoV, and bat SARS-like CoV Envelope Protein.



Figure 4: Multiple Sequence Alignment of SARS-CoV-2, SARS-CoV, and bat SARS-like CoV Nucleocapsid Protein.

#### **Mutational Analysis**

Mutational analysis was implemented by Genome Detective Coronavirus Typing Tool (Version 1.17) and validated by Multalin online tool<sup>20</sup>. This tool has unearthed mutations eventuating in S, E, and Nprotein of *SARS*- CoV-2 variants. We found out that SARS-CoV-2 (India) has undergone 7 mutations, SARS-CoV-2 (UK) has undergone 9 mutations, SARS-CoV-2 (South Africa) has undergone 11 mutations, and SARS-CoV-2 (Brazil) has undergone 12 mutations, respectively. We have not witnessed any mutations in the S-protein of SARS-CoV-2 (Pakistan) (Table 1, Figure 2).

Protein	Origin	No. of Mutation	Mutation
	Pakistan	0	-
			G142D (21987G>A)
			V143F (21989G>T)
			L452R (22917T>G)
	India	7	T478K (22995C>A)
			D614G (23403A>G)
			P681R (23604C>G)
0			D950N (24410G>A)
5			H69_V70del
			(21766_21771delACATGT)
			Y144del (21992_21994delTAT)
			N501Y (23063A>T)
	UK	9	A570D (23271C>A)
			D614G (23403A>G)
			P681H (23604C>A)
			T716I (23709C>T)
			\$982A (24506T>G)
			D1118H (24914G>C)
	South Africa	11	L18F (21614C>T)
			D80A (21801A>C)
			D215G (22206A>G)
			L242_L244del
			(22286_22294delCT*TGCT*TA)
			K417N (22813G>T)
			E484K (23012G>A)
			N501Y (23063A>T)
			D614G (23403A>G)
			Q677H (23593G>T)
			R682W (23606C>T)
			A701V (23664C>T)
			L18F (21614C>T)
			T20N (21621C>A)
			P26S (21638C>T)
			D138Y (21974G>T)
	D "I	10	R190S (22132G>T)
	Brazil	12	K417T (22812A>C)
			E484K (23012G>A)
			N501Y (23063A>T)
			D614G (23403A>G)
			H655Y (23525C>T)
			T1027I (24642C>T)
			V1176F (25088G>T)

**Table 1:** Mutation in Spike protein of SARS-CoV-2 strains from diverse geographical regions.

Further, we have also observed one amino acid mutation of Proline (P) to Leucine (L) at the  $71^{st}$  locus in the E-protein of *SARS-CoV-2* (South Africa). However, we have not witnessed any mutation in the E-protein of *SARS-CoV-2* (Pakistan, India, UK, and Brazil) as well in the N-protein of *SARS-CoV-2* (Pakistan) but have noted several mutations in the N-protein of *SARS-CoV-2* (India, UK, South Africa, and Brazil) (Table 2).

Protein	Origin	No. of Mutation	Mutation
	South Africa	1	P71L (26456C>T)
Ε	Pakistan, India, UK, and Brazil	0	-
	Pakistan	0	-
			D63G (288461A>G)
			L139F (28690G>T)
	India	5	R203M (28881G>T)
			D377Y (29402G>T)
N			R385K (29427G>A)
N			D3L (28280G>C
	UK		28281A>T 28282T>A)
		4	R203K (28881G>A
			28882G>A)
			G204R (28883G>C)
			S235F (28977C>T)
	South Africa	1	T205I (28887C>T)
			P80R (28512C>G)
	Brazil	4	R203K (28881G>A
			28882G>A)
			G204R (28883G>C)
			D288N (29135G>A)

**Table 2:** Mutation in Envelope and Nucleocapsid protein of SARS-CoV-2strains from various geographical locations.

### Impact on Protein Stability

We have distinguished the repercussion of these mutations on the protein stability by the MUpro server. It seems that the consequence of protein mutation diminishes the stability of the protein structure as demonstrated as negative (-)  $\Delta\Delta G$  (Table 3).

Protein Name	Mutation	Origin	Stability Effect (MUpro)
1 loteni i vanie	G142D	Oligin	Decrease Stability
	0112D		$(\Delta \Lambda G = -1.6291959)$
S	V143E	India	Decrease Stability
0	V 1 151	incia	$(\Delta \Lambda G = -1.0725209)$
	I 452R		Decrease Stability
	1.1521		$(\Delta \Delta G = -0.4547574)$
	T478K		Decrease Stability
	1 1701		$(\Delta \Lambda G = -0.3452853)$
	D614G		Decrease Stability
	Doirio		$(\Delta\Delta G = -0.93148242)$
	P681R		Decrease Stability
			$(\Delta \Delta G = -1.390031)$
	D950N		Decrease Stability
			$(\Delta\Delta G = -0.61717494)$
	N501Y		Decrease Stability
			$(\Delta \Delta G = -1.7152495)$
	A570D		Decrease Stability
			$(\Delta \Delta G = -1.8979773)$
	D614G		Decrease Stability
		UK	$(\Delta \Delta G = -0.93148242)$
	P681H		Decrease Stability
			$(\Delta\Delta G = -1.158464)$
	T716I		Decrease Stability
			$(\Delta \Delta G = -1.7281241)$
	S982A		Decrease Stability
			$(\Delta \Delta G = -1.2467257)$
	D1118H		Decrease Stability
			$(\Delta\Delta G = -0.90371524)$
	L18F		Decrease Stability
		South Africa	$(\Delta \Delta G = -0.61079093)$
	D80A		Decrease Stability
	DALEO		$(\Delta\Delta G = -0.85116553)$
	D215G		Decrease Stability $(AAC = 0.00(40755))$
	IZ 44 7N I		$(\Delta\Delta G = -0.88640/55)$
	K41/IN		Decrease Stability $(AAC = 1.3481028)$
	E494V		$(\Delta\Delta G = -1.3481028)$
	L404K		(AAC = 0.0000422203)
	N501V		$(\Delta\Delta G = -0.0000+22203)$
	193011		$(\Lambda\Lambda G = -1.7152495)$
	D614G		Decrease Stability
	Dorre		$(\Delta \Lambda G = -0.93148242)$
	O677H		Decrease Stability
	2011		$(\Delta\Delta G = -0.80644929)$
	R682W		Decrease Stability
			$(\Delta \Delta G = -0.62478789)$
	A701V		Decrease Stability
			$(\Delta \Delta G = -1.5740659)$
	L18F		Decrease Stability
			$(\Delta \Delta G = -0.61079093)$
	T20N	Brazil	Decrease Stability
			$(\Delta \Delta G = -1.1205078)$

Table 3: Repercussion of Mutation on Protein Stability.

	<b>D2</b> ( 0		D 0.111
	P268		Decrease Stability $(AAC = 0.30125180)$
	DAGONZ	-	$(\Delta\Delta G = -0.39123189)$
	D138Y		Decrease Stability $(AAC = 1.01235)$
	D1000	-	$\frac{(\Delta\Delta G = -1.01233)}{(\Delta\Delta G = -1.01233)}$
	R1905		Decrease Stability $(AAG = -0.76374131)$
	K417T	-	Decrease Stability
	1271/1		$(\Delta\Delta G = -1.2705224)$
	E484K	-	Decrease Stability
			$(\Delta\Delta G = -0.0090422293)$
	N501Y		Decrease Stability
			$(\Delta\Delta G = -1.7152495)$
	D614G		Decrease Stability
			$(\Delta\Delta G = -0.93148242)$
	H655Y		Decrease Stability
		_	$(\Delta\Delta G = -0.8723021)$
	T1027I		Decrease Stability
		4	$(\Delta\Delta G = -2.663/486)$
	V1176F		Decrease Stability
	D741		$(\Delta\Delta G = -1.4982363)$
Е	P/1L	South Africa	Decrease Stability
	D(2C		$(\Delta\Delta G = -1.9631091)$
N	D03G	India	Decrease Stability $(AAC = 0.44071754)$
1	L 120E	India	$\frac{(\Delta\Delta G = -0.449/1/34)}{Decrease Stability}$
	L139F		(AAC = 0.61518665)
	D202M	-	$(\Delta\Delta G = -0.01518005)$
	<b>K203</b> M		(AAC = 1.3051696)
	D377V	-	$(\Delta\Delta\Theta = -1.5051090)$
	DJ//I		$(\Delta \Lambda G = -1.277213)$
	R 385K	-	Decrease Stability
	RSOOT		$(\Delta\Delta G = -0.44097048)$
	D3L		Decrease Stability
			$(\Delta \Delta G = -1.53012)$
	R203K	1	Decrease Stability
		UK	$(\Delta\Delta G = -0.73977168)$
	G204R	_	Decrease Stability
			$(\Delta \Delta G = -1.7379677)$
	S235F		Decrease Stability
			$(\Delta \Delta G = -1.8826398)$
	T205I	South Africa	Decrease Stability
			$(\Delta\Delta G = -1.9524886)$
	P80R		Decrease Stability $(AAC = 1.7719400)$
	<b>D202</b> <i>V</i>	4	$(\Delta \Delta G = -1.0/18409)$
	K203K	Brazil	Decrease Stability $(\Lambda\Lambda G = -0.73077168)$
	G204B		Decrease Stability
			$(\Delta\Delta G = -1.7379677)$
	D288N	1	Decrease Stability
			$(\Delta \Delta G = -0.39236791)$

#### Discussion

In this study, we have examined SARS-CoV-2 genomes from Pakistan, India, the UK, South Africa, and Brazil, SARS-CoV, and bat SARS-like CoV to Wuhan (China) SARS-CoV-2. Evolutionary research has revealed that all the SARS-CoV-2 variants form a clade that is all closely related to each other and to bat SARS-like CoV, which successively is related to SARS-CoV. This analysis reveals that only bat SARS-like CoV has a very close evolutionary relationship with SARS-CoV-2 encountering an independent bifurcation from bat SARS-like CoV. The branch length of SARS-CoV portrays that it has diverged very early from bat SARS-like CoV. This evolutionary analysis strongly concurs with the fact that bats are the primary host for coronavirus evolution and the genesis of SARS-CoV and SARS-CoV-2, prompting scientists worldwide to ponder bats as a natural reservoir. (Figure 1). Our evolutionary analysis also concurs with other prior studies <sup>21–25</sup> moreover, is authenticated by BLASTN.

Coronaviruses (CoVs) have the longest genomes (26.4 to 31.7 kb) of any well-known RNA virus  $^{4,26-28}$ . The enormous genome size makes it flexible in acclimatizing and manipulating genes  $^{26}$ . The frequency of recombination in RNA viruses is rather substantial, henceforth enhancing virulence and thus is responsible for the development of speciation  $^{29}$ . The high frequency of recombination within the viral genome at various locations is perhaps one of the causes whereby *SARS-CoV-2* is accountable for both the variation in deaths and medical manifestations  $^{30}$ . The viral genome of *SARS-CoV-2* encodes four prime structural proteins: the S-protein, the N-protein, the M-protein, and an E-protein, which are all critical for the production of a functionally mature virion  $^{31-37}$ .

The *S*-protein *RBD* is the domain that precisely combines with ACE 2 to induce viral ingress into the host cell 38– 42. An assessment of the polypeptides of the S-protein of five *SARS-CoV-2* variants discovered polymorphisms in India, the UK, South Africa, and Brazil except for Pakistan at numerous nucleotide and amino acid positions (Table 1). The prognostication of protein stability employing theoretical or experimental techniques has been a significant topic of research for some years <sup>43</sup>. Previous research suggests that a single point mutation at *RBD* is responsible for altering the epitope organization and, hence impairing *RBD* binding to *ACE* 2 <sup>44,45</sup>. A modification in this area of the S-protein may impair *RBD* adherence towards its receptors, consequently impacting viral penetration into the host genome.

E-protein is crucial for viral genome packaging and the synthesis of ion channels (IC), which are vital for virus-host contact and are frequently associated with pathogenicity 15,31,46. We have observed one amino acid mutation of Proline (P) to Leucine (L) at the 71st locus in the E-protein of SARS-CoV-2 (South Africa). We have not witnessed any mutation in the E-protein of SARS-CoV-2 (Pakistan, India, UK, and Brazil) (Table 2). E-proteins are polypeptides with approximately 100 residues that are miniature components of virions but are extensively synthesized within infected cells <sup>47,48</sup>. They exhibit a small hydrophilic N-terminus, one or more putative terminal transmembrane (TM) domains, and a less hydrophobic C-terminal tail (15). Previously, it was demonstrated that SNPs within the TMD domain of the E-protein impaired IC function and resulted in reduced viral virulence 49. Henceforth, E is a viable antiviral therapeutic target and immunization candidate against SARS-CoV-2.

The N-protein serves a variety of roles in the  $C_{\theta}V$  virus replication <sup>17,18</sup>. For example, in *SARS-C\_{\theta}V*, the N-protein has been shown to adhere to viral *RNA* and assemble it into *RNP* complexes. The packed *RNPs* particles are found on the viral membrane's internal face, generating a distinct layer from the envelope proteins M, E, and S. Moreover, the association between N and the C-terminus of the M-protein may facilitate *RNP* localization <sup>50,51</sup>. Numerous mutations at the nucleotide and amino acid positions in the N-protein of *SARS-C\_{\theta}V-2* have been discerned (Table 3). Prior studies have shown that Carboxyl-Terminal Domain (CTD) is essential for oligomerization <sup>52</sup>.

It was also discovered that the *S-protein* was the most mutated of all the structural proteins investigated in this study. Among these variants, the most prominent are D614G, N501Y, E484K, K417N, K417T, and L452R. L452R mutation was discovered in the Indian *SARS-CoV-2 variant*, also referred to as *delta variant* (B.1.617.2), whereas conjunction of D614G and N501Y mutations was discovered in the UK *SARS-CoV-2* variant, also recognized as alpha variant (B.1.1.7). In South African *SARS-CoV-2 variants*, a combination of E484K, K417N, N501Y, and D614G was seen, widely known as beta variant (B.1.351), while a blend of K417T, E484K, N501Y, and D614G was discerned in the Brazilian *SARS-CoV-2* variant, commonly known as the gamma variant (P.1) <sup>53–55</sup>. Our results concur with CDC <sup>56</sup>. It is envisaged that the significant number of mutations found in structural proteins, particularly S-protein, will have an impact on the development of a vaccine/inhibitor against COVID-19.

To sum up, nucleotide and protein sequences of *SARS*-*CoV-2* from Pakistan, India, the UK, South Africa, and Brazil, *SARS-CoV*, and bat *SARS-like CoV* were evaluated and compared with Wuhan (China) *SARS-CoV-2*. Investigators uncovered variants in structural proteins that were unique to each nation (S-protein, E-protein, and N-protein). Furthermore, the MUpro server investigation indicated that mutations impair protein stability and impede inhibitor adhesion. The current research findings might facilitate the development of prospective immunization candidates/small pharmacological compounds targeting COVID-19.

### Acknowledgments

I would like to thank COMSATS University Islamabad for providing me with an excellent academic environment, teachers, and resources to carry out my research. Also, I want to extend my sincere gratitude to my supervisor Dr. Anwar Ullah for his kind support and guidance.

# Conflict of interest

There was no funding associated with this study. The authors declare that the research was performed in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# References

1. Huang Y, Yang C, Xu X feng, Xu W, Liu S wen. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin.* 2020;41(9):1141–9.

2. WHO. Coronavirus Disease [Internet]. Available from: https://www.who.int/emergencies/diseases/novel-coro-navirus-2019

3. Nurlela N, Awaluddin F, Batubara I, Wahyudi ST. Computational study of kaurene diterpenoids for antivirals against SARS-CoV-2. *J Appl Pharm Sci.* 2022;12(8):112–29.

4. Jain M, Anand A, Shah A. Exploring the Potential Role of Theaflavin-3,3'-Digallate in Inhibiting Various Stages

of SARS-CoV-2 Life Cycle: An In-Silico Approach. *Chem Africa*. 2022;5(4):883–98.

5. Worldometer. COVID LIVE UPDATE [Internet]. Worldometer. 2021. Available from: https://www.worldometers.info/coronavirus/

6. Waris A, Atta UK, Ali M, Asmat A, Baset A. COVID-19 outbreak: current scenario of Pakistan. *New Microbes New Infect.* 2020;35:1–6 PubMed .

7. Government of Pakistan. National Command and Operation Center [Internet]. Gop. 2021. p. 1. Available from: https://ncoc.gov.pk/

8. DAWN NEWS. Clear signs 4th Covid wave is starting, warns Asad Umar - Pakistan - DAWN [Internet]. Available from: https://www.dawn.com/news/1634076

9. GEO NEWS. COVID-19\_ Pakistan confirms presence of South African, Brazilian variants [Internet]. Available from: https://www.geo.tv/latest/348183-covid-19-pakistan-confirms-presence-of-south-african-brazilian-variants

10. THE EXPRESS TRIBUNE. "UK variant accounts for 70% of Covid cases in Pakistan" [Internet]. Available from: https://tribune.com.pk/story/2298965/uk-variant-accounts-for-70-of-covid-cases-in-pakistan-researcher

11. ARAB NEWS. Delta variant accounts for 50% new Pakistani COVID-19 cases — National Institute of Health [Internet]. Available from: https://www.arabnews.pk/node/1891751/pakistan

12. Umar HI, Josiah SS, Saliu TP, Jimoh TO, Ajayi A, Danjuma JB. In-silico analysis of the inhibition of the SARS-CoV-2 main protease by some active compounds from selected African plants. *J Taibah Univ Med Sci.* 2021;16(2):162–76.

13. Ahmed-Abakur EH, Ullah MF, Elssaig EH, Alnour TMS. In-silico genomic landscape characterization and evolution of SARS-CoV-2 variants isolated in India shows significant drift with high frequency of mutations. *Saudi J Biol Sci.* 2022;29(5):3494–501.

14. Turoňová B, Sikora M, Schürmann C, Hagen WJH, Welsch S, Blanc FEC, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* (80-). 2020;370(6513):203–8.

15. Jalily PH, Jalily Hasani H, Fedida D. In Silico Evaluation of Hexamethylene Amiloride Derivatives as Potential Luminal Inhibitors of SARS-CoV-2 E Protein. *Int J Mol Sci.* 2022;23(18).

16. Schoeman D, Fielding BC. Coronavirus envelope protein: Current knowledge. *Virol J.* 2019;16(1):1–22. 17. McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses*. 2014;6(8):2991–3018.

18. URAS ME. In silico comparative analysis of SARS-CoV-2 Nucleocapsid (N) protein using bioinformatics tools. *Front Life Sci Relat Technol.* 2021;2(1):1–9.

19. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. Coronaviruses: An RNA proofreading machine regulates replication fidelity and diversity. Vol. 8, RNA Biology. 2011.

20. Florence Corpet. Multalin Interface Page [Internet]. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.*; 1988. Available from: http://multalin. toulouse.inra.fr/multalin/

21. Baruah C, Devi P, Sharma DK. Sequence Analysis and Structure Prediction of SARS-CoV-2 Accessory Proteins 9b and ORF14: Evolutionary Analysis Indicates Close Relatedness to Bat Coronavirus. *Biomed Res Int.* 2020;2020.

22. Rehman HA, Ramzan F, Basharat Z, Shakeel M, Khan MUG, Khan IA. Comprehensive comparative genomic and microsatellite analysis of SARS, MERS, BAT-SARS, and COVID-19 coronaviruses. *J Med Virol.* 2021;93(7):4382–91.

23. Cui J, Han N, Streicker D, Li G, Tang X, Shi Z, et al. Evolutionary relationships between bat coronaviruses and their hosts. *Emerg Infect Dis.* 2007;13(10):1526–32.

24. D. Vijaykrishna, G.J.D. Smith, J.X.Zhang, J.S.M. Peiris, H. Chen and YG. Evolutionary Insights into the Ecology of Coronaviruses. *Am Soc Microbiol Virol.* 2007;81(8):4012–20.

25. P Z, XL Y, XG W, B H, L Z, W Z. A pneumonia outbreak associated with anew coronavirus of probable bat origin. *Nature*. 2020;579(7798):270–3.

26. Khan MI, Khan ZA, Baig MH, Ahmad I, Farouk AEA, Song YG, et al. 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.* 2020;15(9 September):1–15.

27. Woo PCY, Lau SKP, Lam CSF, Lai KKY, Huang Y, Lee P, et al. Comparative Analysis of Complete Genome Sequences of Three Avian Coronaviruses Reveals a Novel Group 3c Coronavirus. *J Virol.* 2009;83(2):908–17.

28. Mihindukulasuriya KA, Wu G, St. Leger J, Nordhausen RW, Wang D. Identification of a Novel Coronavirus

from a Beluga Whale by Using a Panviral Microarray. J Virol. 2008;82(10):5084–8.

29. Duffy S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* 2018;16(8):1–6.

30. Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol.* 2020;92(6):584–8.

31. Schoeman D, Fielding BC. Coronavirus envelope protein: Current knowledge. Vol. 16, *Virology Journal.* 2019. 1–122 p.

32. Masters PS. The Molecular Biology of Coronaviruses. *Adv Virus Res.* 2006;65(06):193–292.

33. Mortola E, Roy P. Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system. *FEBS Lett.* 2004;576(1–2):174–8.

34. Wang C, Zheng X, Gai W, Zhao Y, Wang H, Wang H, et al. MERS-CoV virus-like particles produced in insect cells induce specific humoural and cellular imminity in rhesus macaques. *Oncotarget.* 2017;8(8):12686–94.

35. Mithilesh S, Raghunandan D, Suresh PK. In-Silico Identification of Natural Compounds from Traditional Medicine as Potential Drug Leads against SARS-CoV-2 Through Virtual Screening. *Proc Natl Acad Sci India Sect B - Biol Sci.* 2022;92(1):81–7.

36. Low ZY, Zabidi NZ, Yip AJW, Puniyamurti A, Chow VTK, Lal SK. SARS-CoV-2 Non-Structural Proteins and Their Roles in Host Immune Evasion. *Viruses*. 2022;14(9):1–27.

37. Junior NN, Santos IA, Meireles BA, Nicolau MSP, Lapa IR, Aguiar RS, et al. In silico evaluation of lapachol derivatives binding to the Nsp9 of SARS-CoV-2. *J Biomol Struct Dyn.* 2022;40(13):5917–31.

38. Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S. The spike protein of SARS-CoV - A target for vaccine and therapeutic development. *Nat Rev Microbiol.* 2009;7(3):226–36. 39. Kuhn JH, Li W, Choe H, Farzan M. Angiotensin-converting enzyme 2: A functional receptor for SARS coronavirus. *Cell Mol Life Sci.* 2004;61(21):2738–43.

40. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol.* 2020;17(6):613–20.

41. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. J Med Virol. 2022;94(4):1641–9.

42. Math RK, Mudennavar N, Javaregowda PK, Savanur A. In Silico Comparative Analysis of the Functional, Structural, and Evolutionary Properties of SARS-CoV-2 Variant Spike Proteins. *JMIR Bioinforma Biotechnol.* 2022;3(1): e37391.

43. Wang L, Veenstra DL, Radmer RJ, Kollman PA. Can one predict protein stability? An attempt to do so for residue 133 of T4 lysozyme using a combination of free energy derivatives, PROFEC, and free energy perturbation methods [Internet]. Vol. 32, Proteins: Structure, Function and Genetics. 1998. p. 438–58. Available from: https:// pubmed.ncbi.nlm.nih.gov/9726415/

44. Babcock GJ, Esshaki DJ, Thomas WD, Ambrosino DM. Amino Acids 270 to 510 of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Are Required for Interaction with Receptor. *J Virol.* 2004;78(9):4552–60.

45. Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, Xiao X, et al. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J Biol Chem.* 2006;281(23):15829–36.

46. Ruch TR, Machamer CE. The coronavirus E protein: Assembly and beyond. *Viruses*. 2012;4(3):363–82.

47. Nieto-Torres JL, DeDiego ML, Álvarez E, Jiménez-Guardeño JM, Regla-Nava JA, Llorente M, et al. Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology.* 2011;415(2):69–82.

48. Mandala VS, McKay MJ, Shcherbakov AA, Dregni AJ, Kolocouris A, Hong M. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat Struct Mol Biol.* 2020;27(12):1202–8.

49. Nieto-Torres JL, DeDiego ML, Verdiá-Báguena C, Jimenez-Guardeño JM, Regla-Nava JA, Fernandez-Delgado R, et al. Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Ion Channel Activity

Promotes Virus Fitness and Pathogenesis. *PLoS Pathog.* 2014;10(5):1–16.

50. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein - Forms and functions. *Antiviral Res.* 2014;103(1):39–50.

51. Cortés-Sarabia K, Luna-Pineda VM, Rodríguez-Ruiz HA, Leyva-Vázquez MA, Hernández-Sotelo D, Beltrán-Anaya FO, et al. Utility of in silico-identified-peptides in spike-S1 domain and nucleocapsid of SARS-CoV-2 for antibody detection in COVID-19 patients and antibody production. *Sci Rep.* 2022;12(1):15057.

52. Lo YS, Lin SY, Wang SM, Wang CT, Chiu YL, Huang TH, et al. Oligomerization of the carboxyl terminal domain of the human coronavirus 229E nucleocapsid protein. *FEBS Lett.* 2013;587(2):120–7.

53. Idowu KA, Onyenaka C, Olaleye OA. A computational evaluation of structural stability of omicron and delta mutations of SARS-CoV-2 spike proteins and human ACE-2 interactions. *Informatics Med unlocked* [Internet]. 2022; 33:101074. Available from: http://www.ncbi.nlm. nih.gov/pubmed/36092780%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC9450468

54. Sun C, Xie C, Bu GL, Zhong LY, Zeng MS. Molecular characteristics, immune evasion, and impact of SARS-CoV-2 variants. *Signal Transduct Target Ther.* 2022;7(1):1–72.

55. Fidan O, Mujwar S, Kciuk M. Discovery of adapalene and dihydrotachysterol as antiviral agents for the Omicron variant of SARS-CoV-2 through computational drug repurposing. *Mol Divers.* 2022;

56. Centers for Disease Control and Prevention. SARS-CoV-2 Variant Classifications and Definitions [Internet]. Cdc. 2021. Available from: https://www.cdc.gov/coro-navirus/2019-ncov/variants/variant-info.html

57. Nei M and Kumar S. Molecular Evolution and Phylogenetics. Oxford Univ Press New York. 2000;

58. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol.* 2021;38(7):3022–7.