­­­­­­

**Deciphering Whole Genomic Sequence of SARS-CoV-2 Isolated from COVID 19 Patients in Nepal**

Submitted by:

Nepal Health Research Council

Ram Shah Path, Kathmandu, Nepal

Submitted to

Government of Nepal

Ministry of Health and Population

Ram Shah Path, Kathmandu

November 2020

**Acknowledgements**

I am grateful to all the members of the study team for their individual contribution at different aspects of the study in the process of completing it and producing this report. I am thankful to all other co-investigators of this study Prof. Dr. Sanjib Kumar Sharma, Head of Department of Internal Medicine, B.P. Koirala Institute of Health Sciences (BPKIHS), Dr. Basudev Pandey, Director at Epidemiology and Diseases Control Division (EDCD), Dr. Narayan Raj Bhattarai, Additional Professor, B.P. Koirala Institute of Health Sciences (BPKIHS), Dr. Runa Jha, Director of National Public Health Laboratory (NPHL), Dr. Meghnath Dhimal,Chief, Research Section of Nepal Health Research Council (NHRC), Dr. Guna Nidhi Sharma, Senior Health Administrator, Ministry of Health and Population (MoHP), Mr. Puspa Raj Khanal, Registrar, Nepal Health Professional Council, Dr. Basudha Khanal, Professor and Head, Department of Microbiology, B.P. Koirala Institute of Health Sciences (BPKIHS), Dr. Megha Raj Banjara, Associate Professor, Central Department of Microbiology, Tribhuvan University, Dr. Sameer M Dixit, Founder Director of Public Health Research, Center for Molecular Dynamics Nepal (CMDN), Dr. Rajeev Shrestha, Associate Professor, Kathmandu University,Dr.Anup Bastola, Chief Consultant at Sukraraj Tropical and infectious Disease Hospital, Dr. Binod Kumar Yadav, Associate Professor at Institute of Medicine (IOM), Tribhuwan University, Dr. Nilam Thakur, Clinical Geneticist, Incharge, Medical Genetics Unit, NAMS, Bir Hospital, Dr. Krishna Kumar Aryal , Deputy Team Leader, UK aid, Nepal Helath Sector Programme 3, Monitoring Evaluation and Operational research, Abt Associates, Dr. Nirajan Bhusal, Assistant Professor, Kathmandu University, Dr. Shyam Prakash Dumre, Assistant Professor, Institute of Tropical Medicine, Nagasaki University, Dr. Suman Pant, Research officer NHRC, Mr. Kamal Ranabhat, Medical Laboratory Inspector, Ministry of Health and Population (MoHP), Mr. Balram Gautam, Genetic Scientist, NAMS, Bir Hospital and Mr. Saroj Bhattrai, Research Officer of NHRC who provided technical support for conceptualizing, designing and implementation of this study.

Similarly, I would like to acknowledge the contribution of Ms. Neelam Dhakal, Research Officer of NHRC, Mr. Nishan Katuwal, Biotechnologist from Kathmandu Univeristy,inhelping to prepare the report. I am also grateful to Intrepid Nepal and CMDN for their contribution in carrying out gene sequencing.I am also thankful to the staffof National Public Health Laboratory,Teku, Nepal Korea Hospital, Bhaktapur and Dhulikhel Hospital for their support in providing sample for this study. Last but not least, I would like to acknowledge the support provided by the NHRC team and all other colleagues of NHRC.

**………………………………**

**Dr. Pradip Gyanwali**

**Executive Chief**

**Nepal Health Research Council**

**and Principal Investigator**

**Acronyms**

|  |  |
| --- | --- |
| aa  A  ACE2  BPKIHS  C  CMDN  COVID-19  D  E gene protein  EDCD  GISAID  G  HMM  IOM  M gene  MERS  MoHP  MUSCLE  N gene  NCBI  NGS  NHRC  NPHL  nsp  nt  ORF  rRT-PCR  S gene  SARS  SARS-CoV-2  SNPs  T  TMPRSS  WGS | Amino acids  Adenine  Angiotensin-converting enzyme  B.P. Koirala Institute of Health Sciences  Cytosine  Center for Molecular Dynamics Nepal  Coronavirus disease 2019  Aspartic acid  Envelope protein  Epidemiology and Diseases Control Division  Global Initiative on Sharing All Influenza Data  Glycine  Hidden Markov Models  Institute of Medicine  Membrane Glycoprotein,  Middle East Respiratory Syndrome  Ministry of Health and Population  Multiple Sequence Comparison by Log- Expectation  Nucleocapsid Glycoprotein  National Center for Biotechnology Information  Next generation sequencing  Nepal Health Research Council  National Public Health Laboratory  Non-structural proteins  Nucleotide Deletion  Open Reading Frame  Real-time Reverse Transcription Polymerase Chain Reaction  Spike Glycoprotein,  Severe Acute Respiratory Syndrome  Severe Acute respiratory Syndrome Coronavirus 2  Single Nucleotide Polymorphisms  Thymine  TM protease serine 2  Whole-Genome Sequencing |

**Executive summary**

**Introduction**

COVID-19 is spreading exponentially globally and Nepal is experiencing exponential increase in COVID 19 cases with significant clinical severity. The major contributor of such scenario could be due to the mutations in SARS-CoV-2. It is important to explore the transmission of the virus in Nepalese communities focusing on virus genome isolated from Nepalese patients. Therefore, the objective of this study was to characterize the whole genome sequence isolated from patients diagnosed in the country, and look at the genetic variation of SARS-CoV2 .

**Methods:**

Fifteen (15) retrospective/archived samples (naropharyngeal/throat swabs) within 26 June to 10 August, 2020 andobtained from 10 districts were collected in the month of August from confirmed COVID-19 patients from National Public Health Laboratory,Teku, Kathmandu University Hospital, Dhulikhel and Nepal Korea Hospital Friendship Municipality Hospital, Bhaktapur using convenient sampling technique. The RNA genome of the virus was extracted and sequenced using next generation sequencing (NGS), MiSeq platform using Nextera XT DNA Library Preparation Kit and analysed to identify diversity and mutations among viruses. Genome Sequence Data was analyzed using CLC Genomics Workbench 5 software. Phylogenetic tree was created and compared among sequence data of virus isolates from Nepal and available Genome sequence from other countries from NCBI/GSAID. Strains of virus circulating in Nepal and any mutation that had occurred in gene were identified using GSAID CoV Surver mutation app and Pangolin COVID 19 Lineage Assigner.

**Findings:**

Description of the sample

* Male: Female ratio was 2.75.
* Median age of the participants was 36 years.

Results from Bioinformatic analysis

* Samples were organized based on GISAID, Nextstrain and Pangolin nomenclature system
* GISAID Clade: GH & GR; 8 sequences fell in GH clade while, 7 sequences fell in GR cladeNextstrain Clade: 20 A & 20 B;8 sequences fell into 20A and 7 sequences fell into 20B
* Pangolin Lineage: B.1.36.1, B.1.1, B.1.1
* Genomes of SARS-CoV-2 isolated among Nepalese population fall into the existing clades assigned by GISAID/ Nextstarin/Pangolin lineage that were collected from February-August 2020, from from India, Saudi Arabia, Singapore, USA, Europe, and Australia.

Phylogenetic analysis

* Phylogenetic analysis of 15 Nepalese genome showed that genome sequences were located in a cluster with genomes mainly from India, USA, Bangladesh and Saudi Arabia.
* Identical strains have been introduced from different geographic locations of Nepal.
* The first known SARS-CoV-2 sequence, in Nepal, from January 2020 was found to be clade 19A, while the fifteen sequences, used in this analysis are in recent clade i.e.,GH, GR (GISAID) 20A and 20B (Nextstrain) which indicate that the SARS-CoV-2 was introduced from late February.

Mutation pattern

* With respect to the *hCoV-19/Wuhan/WIV04/2019 GISAID reference sequence/strain* compared to 15 sequences, maximum of 9 existing and 3 unique non-synonymous mutations were observed in non-structural protein (nsp2, nsp 3, nsp 12 (RdRp), spike protein (S) and nucleocapsid protein (N)).
* D614G in the S protein, were present on all of the sequences that have emerged in Europe starting from February 2020.
* Mutational profile and SNPs of the fifteen SARS-CoV-2 sequences based on NCBI reference sequence/strain showed maximum 17 existing and 3 unique non-synonymous mutations and 25 single nucleotide polymorphisms (SNPs).

**Conclusion:**

We conclude that SARS-CoV-2 from Nepal, mostly, belongs to GH and GR (that mutated from strain G at the end of February). Sequences of SARS-CoV-2 from Nepal resembles with the sequence worldwide, however; we may assume that, the probable source of SARS-CoV-2 in Nepal would be from India, Saudi Arabia, Europe and USA.

**INTRODUCTION**

**Background**

Coronavirus disease 2019 (COVID-19) is a respiratory disease caused by a novel strain of coronavirus, following the Severe Acute Respiratory Syndrome (SARS) in 2003 and Middle East Respiratory Syndrome (MERS) in 2012 [1]. The initial cases can be traced back to December 2019, in Wuhan, Hubei province, China, when clusters of patients began to present with pneumonia-like symptoms, and were later reported to be infected with a novel coronavirus named SARS-CoV-2 [2, 3]. Although initial cases were suspected to be acquired from a zoonotic source (Wuhan’s seafood market), more cases of human-to-human transmission were soon reported, leading to explosive outbreaks [3]. The COVID-19 has rapidly spread to different countries/territories, gradually affected the entire world with infecting 47 million people, and killing 1207 thousands as of November 03, 2020 [4]. After the first imported case of COVID-19 (a Nepalese student from Wuhan) on 13 January 2020, the number of cases has been increasing gradually and reached 179614 and 1004 deaths as of 03 November 2020[5].The virus is transmitted primarily via droplets from infected personsand contaminated objects through close – contacts, within families or gathering events. The estimated COVID-19 basic reproductive ratio (R0) of 1.25 to 3.0 is similar to, or slightly higher than, that of seasonal (1.3) or pandemic influenza [6,7]. The appearance of multiple reports regarding super-spreading events, once again, demonstrated a largely unpredictable characteristics of this viral infection.

As COVID-19 is caused by a novel virus, there are neither approved vaccines nor therapy available. Lack of standard treatment protocol to certain extend the treatment approach of each hospital might have been different depending upon the local guidelines and expertise. The complications of the disase might also have influence the treatment modality depending upon the available resources [8, 9]. Up to now, clinical management includes infection control and prevention and supportive cares by oxygen supplementation and mechanical ventilation when indicated [10]. Dexamethasone and heparin also became the standard mode of therapy in hypoxic patients after Recovery trial [10].

SARS-CoV-2 is an enveloped, single-stranded RNA virus with a genome length of 29.9 kilobases, which belongs to the β coronavirus genera. Gene fragments express structural and non-structural proteins, encoding 980 amino acids[11]. The (S gene) glycoprotein, envelope (E gene) protein, membrane (M gene) glycoprotein, nucleocapsid (N gene) protein encode structural proteins whereas non-structural proteins (nsp) such as 3-chymotrypsin-like protease, papain-like protease, and RNA-dependent RNA polymerase, are encoded by Open Reading Frame (ORF)[11-13]. Originating from an intermediate animal host the SARS-CoV-2 intermediate host remains unknown; however, SARS-CoV-2 are closely related and originated in bats, who most likely serve as reservoir host for these two viruses. Whereas palm civets and racoon dogs have been recognized as intermediate hosts for zoonotic transmission of SARS-CoV between bats and humans[14-16]. Transmission of SARS-CoV-2 occurs when a healthy individual inhales or comes into contact with respiratory droplets from an infected person [17]. The glycosylated spike (S) protein of SARS-CoV-2 binds to host cell receptor via human angiotensin-converting enzyme (ACE2), TM protease serine 2 (TMPRSS2), a type 2 TM serine protease located on the host cell membrane, promotes virus entry into the cell by activating the S protein[18,19]. The average incubation period before patient exhibit disease symptoms ranges from 2 to 14 days with clinical presentations ranging from no symptoms to mild fever, cough, and dyspnea to cytokine storm, respiratory failure, and death. SARS-CoV-2 is the seventh coronavirus known to infect humans after SARS-CoV [11, 20].

RNA viruses' mutation rate is dramatically high, up to million times higher than that of their hosts and this high rate is correlated with virulence modulation and evolvability [15]. RNA virus mutation rate contributes to viral adaptation creating a balance between the integrity of genetic information and genome variability [22]. Viral genome mutagenic process depends on the viral enzymes that replicate the nucleic acids, influenced by few or no proofreading capability and/or post-replicative nucleic acid repair. Other mutation-generating processes include: host enzymes, spontaneous nucleic acid damages due to physical and chemical mutagens, recombination events and also particular genetic elements responsible for production of new variants[21, 22]. Mutation rates are modulated by other factors such as determinants of the template sequence and structure involved in virus replication. Like other RNA virus e.g. Influenza, SARS-CoV-2, accumulate mutations over time resulting in sequence diversity. The current mutation rate of SARS-CoV-2 is estimated to be approximately 2•5 nucleotides per month[17]. SARS-CoV-2 phylogenetic analysis of three super-clades (S, V, and G) revealed 1,516 nucleotide-level variations at different positions throughout the entire genome. Frequency of aa (amino acids) mutations were relatively higher in Europe (43.07%) followed by Asia (38.09%) and North America (29.64%). Nucleotide deletion (nt) analysis found twelve sites at coding sequence of the ORF8, spike, and ORF7 protein specifically in polyprotein ORF1ab(n=9), ORF10(n=1) and 3-UTR(n=2) suggesting viral evolution is mainly driven by genttic drift and founder events besides local/regional adaptations [24, 25].

Although its clinical significance remains to be known, genomic characteristics of the virus can influence clinical outcome. Moreover, most of the patients (>80%) presented/detected with mild disease. In the background of overall case fatality rate of 2.3%, none of the patients with non-critical disease had died [18]. Therefore, it is pertinent to study if specific genetic makeupis associated with critical/non-critical clinical outcome of the disease or if the outcome is influenced along with other co-factors, if any.

**Rationale**

The strain of this virus is novel and understanding the Genomic epidemiology of SARS-CoV-2 allow recognition of transmission clusters, its biological evolutionary rate as well as can provide precious insights for assessing viral drug resistance, immune escape and virulence/ pathogenesis related mechanisms [23]. In various countries of the world, obtaining new SARS-CoV-2 genomes is ongoing, and this would allocate monitoring multiple aspects of this pandemic. These include genetic diversity, association with clinical and epidemiological patterns and profiles, the usefulness of diagnostic methods, and the rational design of therapeutics and vaccine candidates [23, 26]. Sharing SARS‐CoV‐2 strains data with global scientific community is very important for collaborative scientific efforts to combat this pandemic.

Currently, despite a high burden of COVID-19 in Nepal, very little information is available to date from full-length high-quality sequences. New SARS-CoV-2 virus needs to be understood further at the genetic level in order to assess if new viral variants were spreading across the countries. Therefore, to gain further understanding on the molecular epidemiology of the COVID-19 in Nepal, we aimed to characterize the whole genome sequence SARS-CoV-2 isolated from patients diagnosed in the country, and look at the genetic variation,on top of it, genome analysis for mutations (synonymous/ non-synonymous mutations) which are significant enough to impact functional properties or not. Further, whole genome analysis is the best way to understand the genetic changes and the potential selection pressures in the viral genome of locally circulating strains.

**General Objective:**

The main objective of this study was to decipher the whole genomic sequence of SARS-CoV-2 from COVID-19 cases.

**Specific Objectives:**

To explore the variants of SARS-CoV-2based on epidemiological information.

To identify the possible mutations in different genes of SARS-CoV-2.

To analyze the viral variants circulating in Nepal by using molecular phylogeny tools.

**MATERIALS AND METHODS**

**Study population**

Fifteen (15) retrospective/archived samples from confirmed COVID-19 patients stored at (– 800 C) were collected in the month of August from National Public Health Laboratory,Teku, Kathmandu University Hospital, Dhulikhel and Nepal Korea Hospital Friendship Municipality Hospital, Bhaktapur using convenient sampling technique. The samples used for this study were both orophyarangeal/ nasopharyngeal swabs.

**Inclusion criteria:**

Specimens from patients (any age and sex) with confirmed diagnosis of SARS-CoV-2 infection by real-time reverse transcription polymerase chain reaction (rRT-PCR) (both symptomatic and asymptomatic) was included for further analysis in this study.

**Whole-Genome Sequencing (WGS) of SARS-CoV-2 Strains**

The viral RNA extraction was performed using Omega Biotek MagBind Viral DNA/RNA kit: M6246. The extracted RNA was converted to cDNA using SuperScript IV Vilo Master Mix: 11756050.Amplification of cDNA was performed using COVID-19 primers Q5® High-Fidelity 2X Master Mix: M0492S. Viral extraction and amplification was followed by DNA library preparation that was conducted with MiSeq platform using Nextera XT DNA Library Preparation Kit, Nextera XT Index Kit v2 Set A-D (96 indexes, 384 samples): FC-131-2001 - FC-131-2004. The quantification of the extracted RNA was evaluated with Qubit High Sensitivity DNA kit: Q3285. The average size of the libraries was 350-400 bp. The detail procedures and materials used is depicted in **Annex I.**

**WGS assembling and reference mapping**

The consensus sequencewas generated using Linux platform. Bowtie 2 was used for aligning the sequencing reads to the reference genome for SARS-CoV-2 (GenBank number, MN908947.3 for GSAID referencesequence EPI\_ISL\_402124). SAMtools were used for manipulating the alignments following a double-assembly approachin which multiple rounds of mapping combined with de novo assembling.

**Phylogenetic analysis**

To analyse the obtained SARS-CoV-2 genomes, a dataset of 66 SARS-CoV-2 complete genomes from differentcountries/continent between December 2019-October 2020 was retrieved from NCBI (https://www.ncbi.nlm.nih.gov/genome/viruses/, last access 4 November 2020). Sequence alignment was performed using Multiple Sequence Comparison by Log- Expectation (MUSCLE) using CLC Genomics Workbench 5 software. The phylogenetic tree were conducted by a maximum likelihood approach using CLC Genomics Workbench 5 software.

**Lineage and mutations prediction**

Complete genome sequences of the Nepal’s SARS-CoV-2 strains were subjected to nucleotide/amino acids alignment against the reference strain both from NCBI/ GSAID(Accession NC\_045512 for NCBI Reference Sequence, GenBank number, MN908947.3 for GSAID reference sequence EPI\_ISL\_402124). The genome-wide single nucleotide polymorphisms (SNPs) were also described.The following online tools were used for these purposes: Pangolin COVID-19 Lineage Assigner—https://pangolin.cog-uk.io, CoV-GLUE—http://cov-glue.cvr.gla.ac.uk/#/home for NCBI Reference sequence and https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/ for GISAID Reference Sequence.Current definition of characterizing mutations of SARS-CoV-2 phylogenetic categorization systems (GISAID clades and PANGOLIN lineages) is depicted in **Annex II.**

**Verification and comparision**

Three samples were sent to WHO Collaborating School of Public Health Lab, University of Hong Kong for verification and comparison with the results obtained from WGS in Nepal.

**Ethics**

The study protocol of this study was approved by the Ethical Review Committee of the NHRC. This an emergency situation and the Ministry of Health and Population is seriously needing the information on the viral genetics for control strategy updating based on scientific evidence as necessary. It was expected to get exempt from human participants’ research/ethical regulations as it involved identification, control or prevention of disease in response to an immediate public health threat. Therefore, obtaining the informed consent was not applicable for previously collected and frozen-stored samples. However, the patient’s confidentiality was strictly maintained during samples collection, handling, investigation, analysis and reporting without using their identifier.

**FINDINGS**

Extraction date of samples range from 26th June to 10th of August, 2020. Male: female ratio was 2.75 and the median age of the participants was 36 years. **Table 1** below demonstrates the description of sample participants. Samples were received from National Public Health Laboratory, Kathmandu University Hospital and Nepal Korea Hospital where samples from 10 districts were collected including Kapilvastu, Rautahat, Baglung, Rolpa, Sarlahi, Dolakha,Banepa, Dhulikhel, Bhaktapur and Kathmandu.

Table 1 Description of the patient characteristics of the samples

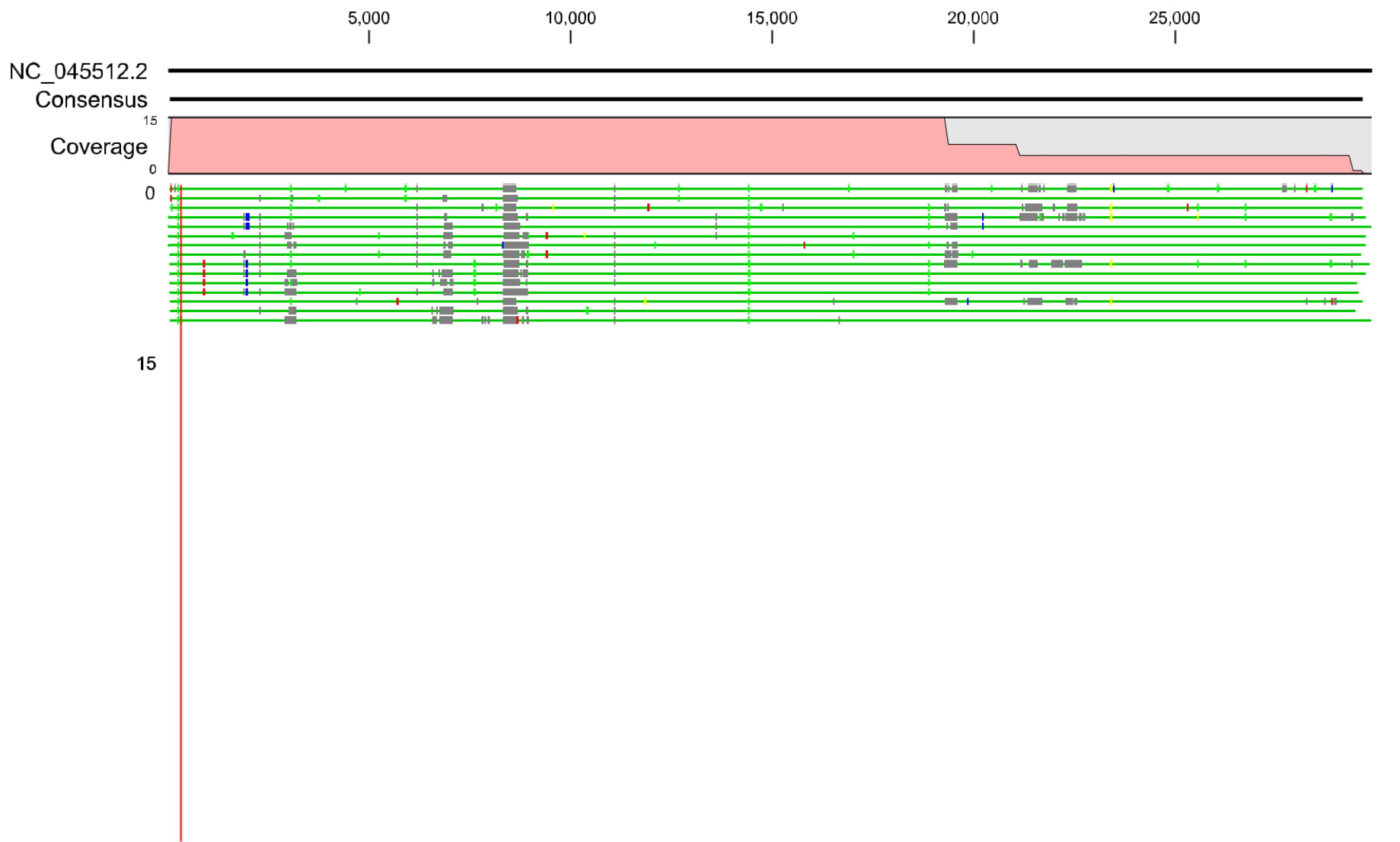
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample Id | Sample collection laboratory/hospital | District | Age | Sex | Collection date |
| 962I | NPHL | Kapilvastu | 19 | M | 3-Aug-20 |
| H353 | NPHL | Sarlahi | 28 | M | 3-Aug-20 |
| 614 | KU | Kavre | 28 | M | 7-Aug-20 |
| 616 | KU | Dolakha | 26 | M | 10-Aug-20 |
| 617 | KU | Kavre | 42 | F | 10-Aug-20 |
| 618 | KU | Kavre | 37 | M | 10-Aug-20 |
| 619 | KU | Kavre | 37 | M | 10-Aug-20 |
| 620 | KU | Kavre | 98 | M | 9-Aug-20 |
| STM-118 | Nepal Korea | Bhaktapur | 30 | M | 26-June-20 |
| CoV-2097 | Nepal Korea | Kathmandu | 66 | F | 30-July-20 |
| CoV-2178 | Nepal Korea | Bhaktapur | 42 | M | 31- July -20 |
| CoV-2673 | Nepal Korea | Rolpa | 13 | F | 7-Aug-20 |
| CoV-3031 | Nepal Korea | Kathmandu | 73 | M | 10-Aug-20 |
| 459A | NPHL | Baglung | 18 | F | 3-Aug-20 |
| 70N | NPHL | Rautahat | 36 | M | 3-Aug-20 |

**Sequence validation:**

As mentioned before, three extracted samples were sent to School of Public Health Lab, University of Hong Kong for whole genome sequencing. These three samples were also sequenced here in Nepal at Center for Molecular Dynamics Nepal (CMDN) among other 12 samples. These three sequences generated from each site were compared using pairwise alignment from EMBOSS Water. This tool uses Smith-Waterman algorithm and evaluates on the basis of %identity and %gaps using EDNAFULL matrix with gap penalty (Pairwise alignment score for the first residue in a gap) of 10 and gap extend (Pairwise alignment score for each additional residue in a gap) of 0.5.

**Table 2:** Comparison of sequences, with sequencing performed at Intrepid Nepal and HK, for each of the three samples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sample/Sequence from CMDN** | **Sample/Sequence** | **Nucleotide** | **%identity** | **%gap** |
|  | **from University of** | **length (nt)** |  |  |
|  | **HongKong** |  |  |  |
| Nepal\_Rautahat\_NPHL-70N | Nepal\_70N\_WHV460 | 30344 | 28646/30344 | 1670/30344 |
|  |  |  | (94.4%) | (5.5%) |
| Nepal\_Baglung\_NPHL-459A | Nepal\_459A\_WHV461 | 24270 | 22475/24270 | 1748/24270 |
|  |  |  | (92.6%) | (7.2%) |
| Nepal\_Kapilvastu\_NPHL-962l | Nepal\_962I\_WHV459 | 31704 | 26896/31704 | 4772/31704 |
|  |  |  | (84.8%) | (15.1%) |



**Figure 1:** Coverage of sequence and conflict region in relation to refertence sequence

When the individual sequence sets were investigated, the sequences performed in Nepal had relatively more undetermined regions, in the sequence, compared to that performed at University of Hong Kong and Reference sequence NCBI\_NC\_0455122. This could be because of the difference in protocol and sensitivity of the instruments used at each site.

Nevertheless, the comparison above is done on the basis of %identity, which is above 90% for samples set containing Nepal\_Rautahat\_NPHL-70N and Nepal\_Baglung\_NPHL-459A, andaround 85% for Nepal\_Kapilvastu\_NPHL-962l. Percent identity is a good metrics for anlaysis,as it is a quantitative measure of similarity between any two sequences, where closely related sequences are expected to have higher identity percent, thus indicating relatedness. *(BlastGlossary, Jan Fassler, Ph.D. and Peter Cooper, Ph.D, 2011)* The third sets of sequences Nepal\_Kapilvastu\_NPHL-962l have relatively lower identity percent(84.4%). However, when that particular sequence from Hong Kong (Nepal\_962I\_WHV459) was analysed in GISAID, it had similar metrics and mutation pattern.

Thus, evaluating the %identity and results in GISAID, it can be concluded that the samples performed in Nepal had similar sequences compared to the samples sent for validation to University of Hong Kong for whole genome sequences.

**Table 3** below demonstrates the statistical distribution of sequence data. Mean length of consensus sequence nucleotide obtained was29653.53 base pairs. Likewise, mean length of amino acid was 8885.533. Mean percentage of non assigned nucleotide out of total nucleotide length was 7.04. Further, mean sequence of Adenine (A), Cytosine(C), Guanine(G) and Thymine (T) were 8220,5084.2, 5424.53 and 8827.8 bases respectively.

**Table 3:**Sequence data statistics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample ID** | **Length(nt)** | **Length(aa)** | **%N** | Adenine (A) | Cytosine (C) | Guanine (G) | Thymine (T) |
| Nepal\_NA\_KU-614 | 29686 | 8436 | 8.82% | 8,066 | 5,005 | 5,325 | 8,658 |
| Nepal\_NA\_KU-617 | 29797 | 9007 | 6.33% | 8,308 | 5,144 | 5,490 | 8,963 |
| Nepal\_NA\_KU-618 | 29470 | 9018 | 7.60% | 8,124 | 5,013 | 5,366 | 8,718 |
| Nepal\_NA\_KU-619 | 29520 | 8714 | 9.31% | 7,986 | 4,942 | 5,264 | 8,573 |
| Nepal\_Rautahat\_NPHL-70N | 29611 | 9229 | 3.23% | 8,527 | 5,271 | 5,635 | 9,213 |
| Nepal\_Bhaktapur\_NK-2178 | 29724 | 9139 | 4.56% | 8,455 | 5,234 | 5,572 | 9,096 |
| Nepal\_Kathmandu\_NK-3031 | 29868 | 9060 | 7.31% | 8,272 | 5,111 | 5,464 | 8,826 |
| Nepal\_Bhaktapur\_NK-STM118 | 29739 | 8790 | 9.02% | 8,097 | 4,980 | 5,322 | 8,652 |
| Nepal\_Baglung\_NPHL-459A | 29611 | 9463 | 4.06% | 8,466 | 5,226 | 5,577 | 9,124 |
| Nepal\_NA\_KU-616 | 29731 | 8425 | 8.69% | 8,108 | 5,015 | 5,346 | 8,671 |
| Nepal\_NA\_KU-620 | 29571 | 9005 | 7.59% | 8,144 | 5,038 | 5,389 | 8,745 |
| Nepal\_Kapilvastu\_NPHL-962l | 29417 | 8352 | 8.53% | 8,022 | 4,985 | 5,300 | 8,589 |
| Nepal\_Kathmandu\_NK-2097 | 29626 | 9221 | 3.46% | 8,524 | 5,262 | 5,607 | 9,200 |
| Nepal\_Rolpa\_NK-2673 | 29613 | 8989 | 6.71% | 8,230 | 5,089 | 5,441 | 8,861 |
| Nepal\_Sarlahi\_NPHL-H353 | 29819 | 8435 | 10.38% | 7,971 | 4,948 | 5,270 | 8,528 |
| Mean | 29653.53 | 8885.533 | 7.04 | 8220 | 5084.2 | 5424.53 | 8827.8 |

**Bioinformatics results of WGS of SARS-CoV-2**

Table4illustrates clade assignment of samples on the basis of GISAID, Nextstrain and Pangolin nomenclature system. According to GISAID nomenclature system, sequences were assigned into two clades i.e GH and GR. Among 15 sequences, 8 sequencesfell in GH clade while, 7 sequencesfell in GR clade.

Sequences from sample Nepal\_Baglung\_NPHL-459A, Nepal\_Kapilvastu\_NPHL-962l, Nepal\_Kathmandu\_NK-2097, Nepal\_NA\_KU-616, Nepal\_NA\_KU-620, Nepal\_Rolpa\_NK-2673 and Nepal\_Sarlahi\_NPHL-H353 fell in GR Clade. Similarly, samples Nepal\_Bhaktapur\_NK-2178, Nepal\_Bhaktapur\_NK-STM118, Nepal\_Kathmandu\_NK-3031, Nepal\_NA\_KU-614, Nepal\_NA\_KU-617, Nepal\_NA\_KU-618, Nepal\_NA\_KU-619, and Nepal\_Rautahat\_NPHL-70N fell in clad GH.

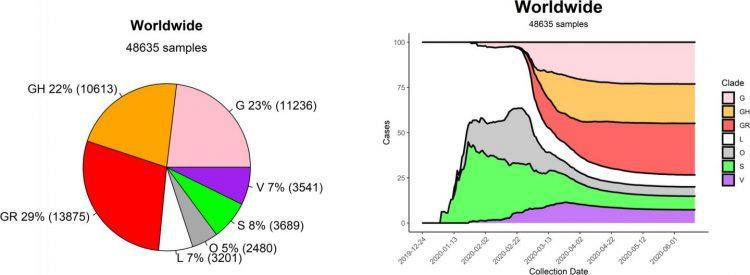
Likewise, samples were also distributed into two clades 20 A and 20 B based on Nextstrain nomenclature system where 8 sequences fell into 20A and 7 sequences fell into 20B. Whereas, according to Pangolin nomenclature system, most of the sequences fell into B.1.36.1, B.1.1, B.1.1. Most of the circulating strains detected in Nepal were also reported from India, UK, Australia, USA and Portugal. Interestingly, the subject Nepal\_Bhaktapur\_NK-STM118 from whom the sample was collected had travel history from Kuwait matched with the strain detected in Saudi Arabia.

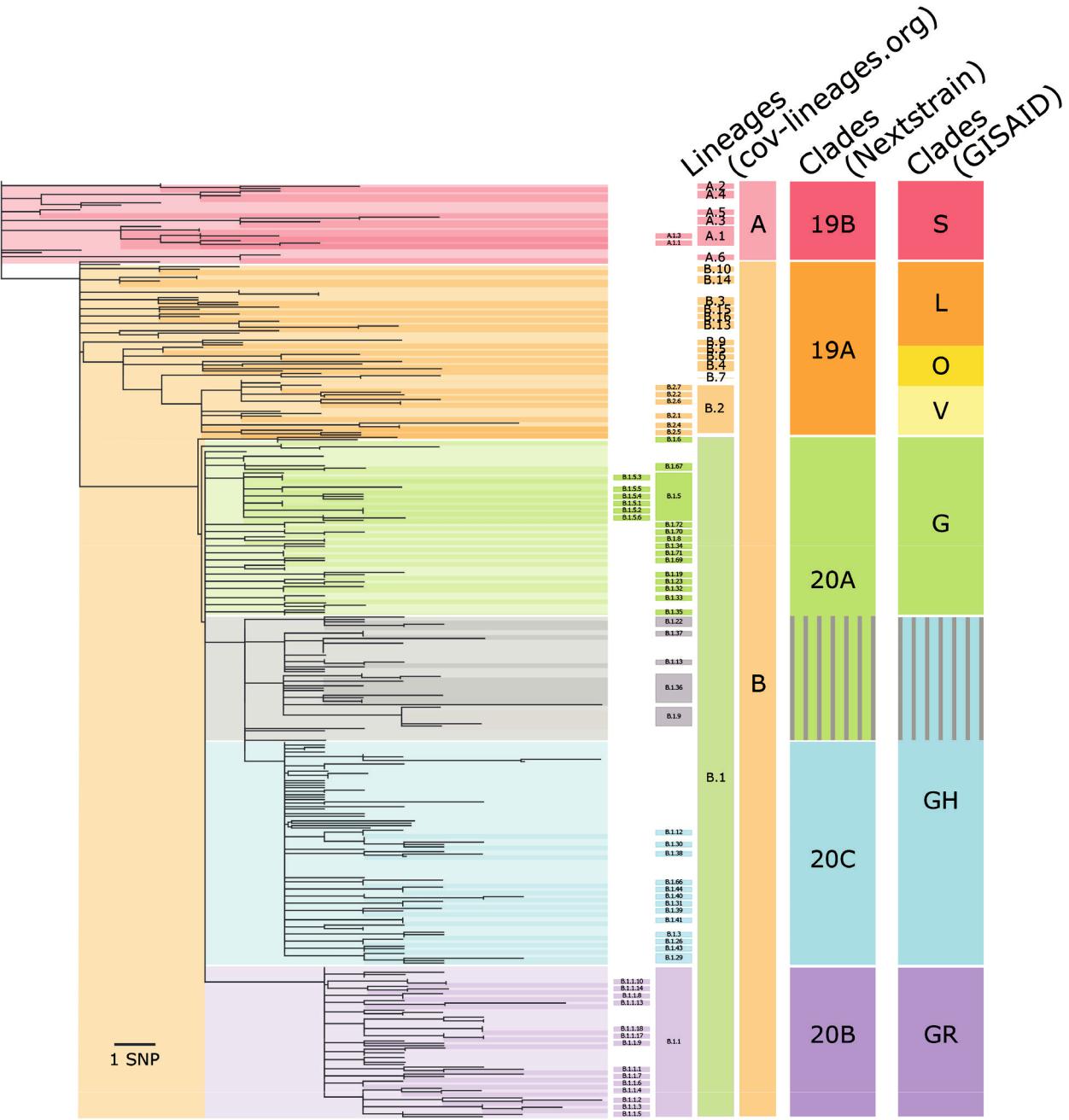
Genomes of SARS-CoV-2 isolated among Nepalese population fall into the existing clades assigned by GISAID/ Nextstarin/Pangolinlineage that were collected from February-August 2020.

**Table 4:** Clade assignment based on nomenclature system of GISAID, Nextstrain and Pangolin

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample ID** | **GISAID** | **Nextstrain** | **Pangolineage** | **Most common countries** | **Date range of Sample cxollection** |
| Nepal\_Banepa\_KU-614 | GH | 20A | B.1.36.1 | India,UK, Australia | May-12, August-05 |
| Nepal\_Banepa\_KU-617 | GH | 20A | B.1.36.1 | India, UK, Australia | May-12, August-05 |
| Nepal\_Banepa\_KU-618 | GH | 20A | B.1.36.1 | India, UK, Australia | May-12, August-05 |
| Nepal\_Banepa\_KU-619 | GH | 20A | B.1.36.1 | India, UK, Australia | May-12, August-05 |
| Nepal\_Rautahat\_NPHL-70N | GH | 20A | B.1.160 | Saudi\_Arabia, India, UK | February-04, August-05 |
| Nepal\_Bhaktapur\_NK-2178 | GH | 20A | B.1.36.1 | India, UK, Australia | May-12, August-05 |
| Nepal\_Kathmandu\_NK-3031 | GH | 20A | B.1.36.1 | India, UK, Australia | May-12, August-05 |
| Nepal\_Bhaktapur\_NK-STM118 | GH | 20A | B.1.36 | India, Saudi\_Arabia, Bangladesh | February-16, August-22 |
| Nepal\_Baglung\_NPHL-459A | GR | 20B | B.1.1.5 | UK, Belgium, Iceland | March-16, May-22 |
| Nepal\_Dolakha\_KU-616 | GR | 20B | B.1.1.72 | India, UK, Singapore | March-28, August-04 |
| Nepal\_Dulikhel\_KU-620 | GR | 20B | B.1.1.72 | India, UK, Singapore | March-28, August-04 |
| Nepal\_Kapilvastu\_NPHL-962l | GR | 20B | B.1.1 | UK, USA, Portugal | February-16, August-17 |
| Nepal\_Kathmandu\_NK-2097 | GR | 20B | B.1.1 | UK, USA, Portugal | February-16, August-17 |
| Nepal\_Rolpa\_NK-2673 | GR | 20B | B.1.1 | UK, USA, Portugal | February-16, August-17 |
| Nepal\_Sarlahi\_NPHL-H353 | GR | 20B | B.1.1 | UK, USA, Portugal | February-16, August-17 |

The presence of GH and GR strain, in Nepal, was expected as GR and GH strains are by far the most widespread, representing 74% of all sequences analysed, worlwide. These strains have four mutations, two of which are able to change the sequence of RNA polymerase and spike protein, thus facilitating the spread of the virus. This was also supported by that fact that, Wuhan, China which initially had L strain, now reported the increasing spread of G, GH and GR strains in Asia.*(Geographic and Genomic Distribution of SARS-CoV-2 Mutations, Frontiers in Microbiology)*





**Figure 2:**The prevalence of different strains of SARS-CoV-2 and their emergence timeline, worldwide [27, 28].

**Mutation patterns**

The whole genome sequences of viruses from the fifteen different COVID-19 subjects (sequenced at CMDN) were, first, analysed for presence of non-synonymous mutations. The sequences were subjected to analysis through CoVsurver from GISAID (*https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/*). GISAID has previously been recognized as a database for outbreak data related of H1N1 pandemic in 2009, H7N9 pandemic in 2013 and COVID-19 pandemic in 2020.

**Table 5:Mutation profile of the fifteen SARS-CoV-2 sequences based on*hCoV-19/Wuhan/WIV04/2019* as the reference sequence/strain(GenBank number, MN908947.3 for GSAID reference sequence EPI\_ISL\_402124)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample ID** | **Mutations** | **Uniquemutations** | **Existing**  **mutations** | **Uniquemutation list** | **Existingmutation list** |
| Nepal\_Banepa\_KU-614 | 7 | 1 | 6 | (NSP3\_D1283B) | (NSP2\_V381A, NSP2\_A26T, NSP12\_P323L, Spike\_D614G, NS3\_Q57H, N\_S194L) |
| Nepal\_Banepa\_KU-617 | 6 | 0 | 6 | - | (NSP2\_V381A, NSP2\_A26T, NSP12\_P323L, Spike\_D614G, NS3\_Q57H, N\_S194L) |
| Nepal\_Banepa\_KU-618 | 6 | 0 | 6 | - | (NSP2\_V381A, NSP2\_A26T, NSP12\_P323L, Spike\_D614G, NS3\_Q57H, N\_S194L) |
| Nepal\_Banepa\_KU-619 | 7 | 0 | 7 | - | (NSP2\_V381A, NSP2\_A26T, NSP3\_P679S, NSP12\_P323L, Spike\_D614G, NS3\_Q57H, N\_S194L) |
| Nepal\_Rautahat\_NPHL-70N | 7 | 2 | 5 | (NSP7\_S25T, Spike\_C1248Y) | (NSP12\_P323L, NSP12\_V424L, Spike\_D614G, NS3\_Q57H, N\_Q418H) |
| Nepal\_Bhaktapur\_NK-2178 | 9 | 2 | 7 | (NSP15\_L200S, Spike\_D389B) | (NSP2\_I393T, NSP2\_V381A, NSP12\_P323L, Spike\_D614G, NS3\_Q57H, NS3\_V55G, N\_S194L) |
| Nepal\_Kathmandu\_NK-3031 | 8 | 1 | 7 | (NSP15\_L200S) | (NSP2\_I393T, NSP2\_V381A, NSP12\_P323L, Spike\_D614G, NS3\_Q57H, NS3\_V55G, N\_S194L) |
| Nepal\_Bhaktapur\_NK-STM118 | 5 | 0 | 5 |  | (NSP12\_P323L, Spike\_D614G, NS3\_Q57H, N\_S194L, N\_L139F) |
| Nepal\_Baglung\_NPHL-459A | 7 | 3 | 4 | (NSP3\_A991T, NSP13\_N95B, NSP16\_N233B) | (NSP12\_P323L, Spike\_D614G, N\_G204R, N\_R203K) |
| Nepal\_Dolakha\_KU-616 | 10 | 1 | 9 | (NSP2\_E261D) | (NSP3\_K837N, NSP4\_D279N, NSP5\_K90R, NSP12\_P323L, NSP13\_L256F, Spike\_D614G, N\_G204R, N\_T296I, N\_R203K) |
| Nepal\_Dulikhel\_KU-620 | 8 | 0 | 8 |  | (NSP3\_K837N, NSP4\_D279N, NSP12\_P323L, NSP13\_L256F, Spike\_D614G, N\_G204R, N\_T296I, N\_R203K) |
| Nepal\_Kapilvastu\_NPHL-962l | 4 | 0 | 4 |  | (NSP12\_P323L, Spike\_D614G, N\_G204R, N\_R203K) |
| Nepal\_Kathmandu\_NK-2097 | 9 | 0 | 9 |  | (NSP3\_M560I, NSP8\_A194V, NSP12\_P323L, Spike\_D614G, Spike\_A1087S, NS3\_T223I, N\_D63Y, N\_G204R, N\_R203K) |
| Nepal\_Rolpa\_NK-2673 | 7 | 0 | 7 |  | (NSP8\_A194V, NSP12\_P323L, Spike\_D614G, NS3\_T223I, N\_G204R, N\_T135I, N\_R203K) |
| Nepal\_Sarlahi\_NPHL-H353 | 4 | 0 | 4 |  | (NSP12\_P323L, Spike\_D614G, N\_G204R, N\_R203K) |

**Table 6: Mutation profile and SNPs of the fifteen SARS-CoV-2 sequences based on NCBI reference sequence/strain (NC\_045512)**

|  |  |  |
| --- | --- | --- |
| **Sample ID** | SNPs | Number of mutation |
| Nepal\_Baglung\_NPHL-459A | 24 | 17 |
| Nepal\_Banepa\_KU-614 | 23 | 16 |
| Nepal\_Dolakha\_KU-616 | 23 | 14 |
| Nepal\_Banepa\_KU-617 | 18 | 10 |
| Nepal\_Banepa\_KU-618 | 20 | 15 |
| Nepal\_Banepa\_KU-619 | 19 | 14 |
| Nepal\_Dulikhel\_KU-620 | 25 | 17 |
| Nepal\_Rautahat\_NPHL-70N | 21 | 13 |
| Nepal\_Kapilvastu\_NPHL-962l | 19 | 13 |
| Nepal\_Kathmandu\_NK-2097 | 23 | 15 |
| Nepal\_Bhaktapur\_NK-2178 | 25 | 19 |
| Nepal\_Rolpa\_NK-2673 | 17 | 11 |
| Nepal\_Kathmandu\_NK-3031 | 22 | 17 |
| Nepal\_Sarlahi\_NPHL-H353 | 16 | 8 |
| Nepal\_Bhaktapur\_NK-STM118 | 15 | 8 |

With respect to the ***hCoV-19/Wuhan/WIV04/2019* GISAID reference sequence/strain**compared to 15 sequences maximum of 9 existing and 3 unique non-synonymous mutations were observed in (non-structural protein (nsp2, nsp 3, nsp 12 (RdRp), spike protein (S) and nucleocapsid protein (N)). D614G in the S protein, were present on all of the sequences that have emerged in Europe starting from February 2020. Similarly, mutation profile and SNPs of the fifteen SARS-CoV-2 sequences based on NCBI reference sequence/strain showed maximum 17 existing and 3 unique non-synonymous mutationsand 25 SNPs.Complete mutation list is depicted in **Annex III**. In comparision to GISAID reference sequence, NCBI refrence sequences showed more number of mutations.

**Unique mutation**

When these unique mutations were investigated, they were found to be present in various non-structural proteins (nsp) of the SARS-CoV-2 virus. The severe acute respiratory syndrome coronavirus (SARS-CoV) is known to generate 16 nsp through proteolytic cleavage of a large precursor protein. (Ref: Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling, ASM). The functions of these nsp vary from catalytic activities that are important forviral replication to transcription. And, few of them have less defined roles. *(Ref: Severe AcuteRespiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling, ASM).*

For instance, the coronavirus nsp's 3, 4, and 6 are integral membrane proteins. These proteins are responsible for membrane rearrangements and to provide scaffolds for assembly for replication complexes [29].

In our analysis, we found that there were several mutations at aminoacid level of the respective proteins with few unique mutations. For instance, in genome sequence of Nepal\_Baglung\_NPHL-459A, a unique mutation **(NSP3\_A991T)** was observed in non-structural protein 3 at position 991. Strain (hCoV-19/Spain/VC-IBV-98003455/2020) that had first reported mutation of this kind was collected in April 2020. The particular sample (Nepal\_Baglung\_NPHL-459A) that had this mutation was collected in Nepal on August 2020.

Similarly, another unique mutation **(Spike\_D389B)** in spike protein was seen in position 389 of sequence from sample Nepal\_Bhaktapur\_NK-2178. The position equivalent to this mutation is reported to be involved in antibody recognition sites and viral oligomerisation interfaces.

Interestingly, mutation in **Spike\_C1248Y**, a spike protein mutation, was seen in sequence from sample Nepal\_Rautahat\_NPHL-70N which was collected in Nepal on August 2020 and was first reported only in August 2020 in strain hCoV-19/England/QEUH-963600/2020. This would mean that the virus strain in sample Nepal\_Rautahat\_NPHL-70N, with mutation similar to that isolated in England, was already circulating in Nepal before August 2020.

**Known mutation**

**Spike\_D614G**

* SARS-CoV-2 spike D614G (aspartic acid (SD614) and glycine (SG614) at residue 614) variant exhibits enhanced replication and transmissibility.
* Mutation Spike D614G already occurred 141412 times (87.11% of all samples with Spike sequence) in 116 countries. The first strain with this mutation, collected inJanuary 2020, was hCoV-19/Zhejiang/HZ103/2020. The mutation most recentlyoccurred in strain hCoV-19/England/NOTT-116DE1/2020, collected in October 2020.
* D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity.
* No effect on neutralizing antibody and varied pathogenicity reported.

**P314L**

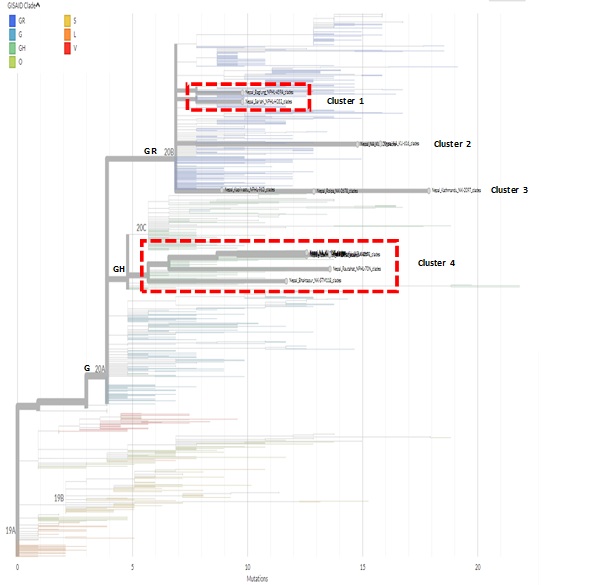
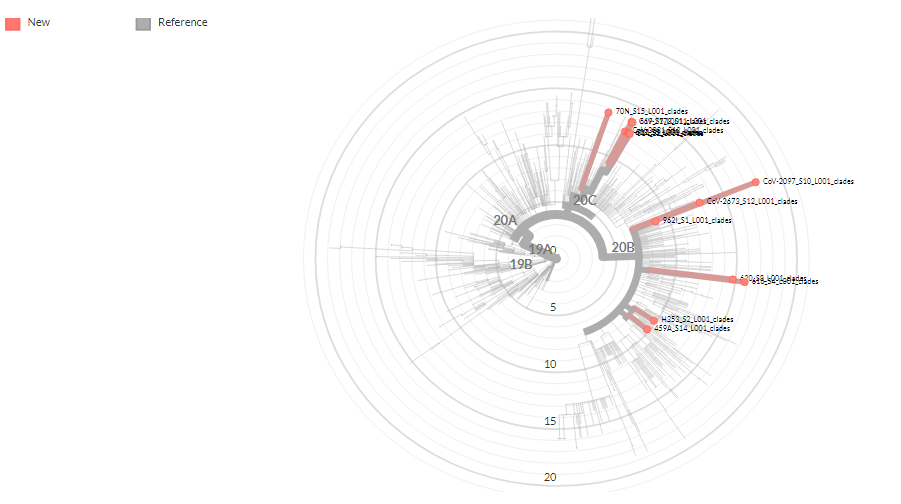
* Affect the non-structural protein 12 (NSP12), the viral RNA-dependent RNA polymerase.

**N\_S194L, N\_G204R, N\_R203K**

* Mutation is correlated with enhanced virulence, evolvability, and traits considered beneficial for the virus, however need to be elucidated further if such kind of mutationcould influence protein function and even virus infectivity.

**Phylogeny placement Nextstrain:**

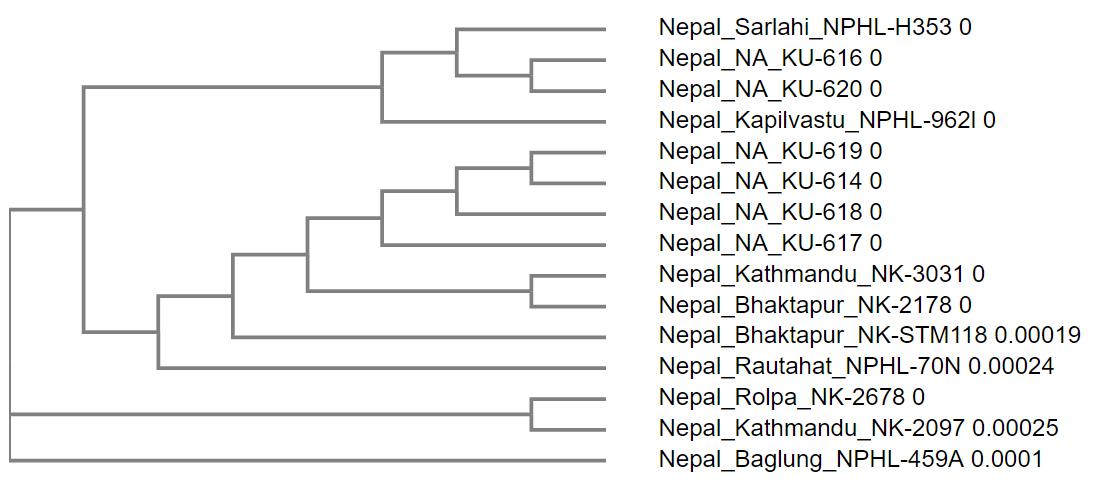
Phylogeny placement of the 15 sequences by nextstrain which uses data from GISAID/ NCBI which are curated and analysed through maximum likelihood ancestral state reconstruction of discrete traits such as country or region of isolation showed that our sequences resembles to the sequences isolated from Europe, US, India, Asia, Saudi Arabia and India. Phylogeny placement of the 15 samples based on GISAID nomenclature system found that samples that fell into GR clade were categorized into three clusters. Where, cluster 1 includes Europe, US, India, cluster 2 includes India and cluster 3 includes Europe, Asia. While, samples that fell into GR clades were categorized into cluster 4 which includes Saudi Arabia, India, and US.

****

**Figure 3**: Phylogenetic assignment of SARS-CoV-2 sequences from Nepalese patients by nextstrain platform (https://nextstrain.org/).

**Phylogeny placement of sequence of SARS-CoV-2 from Nepalese patients**

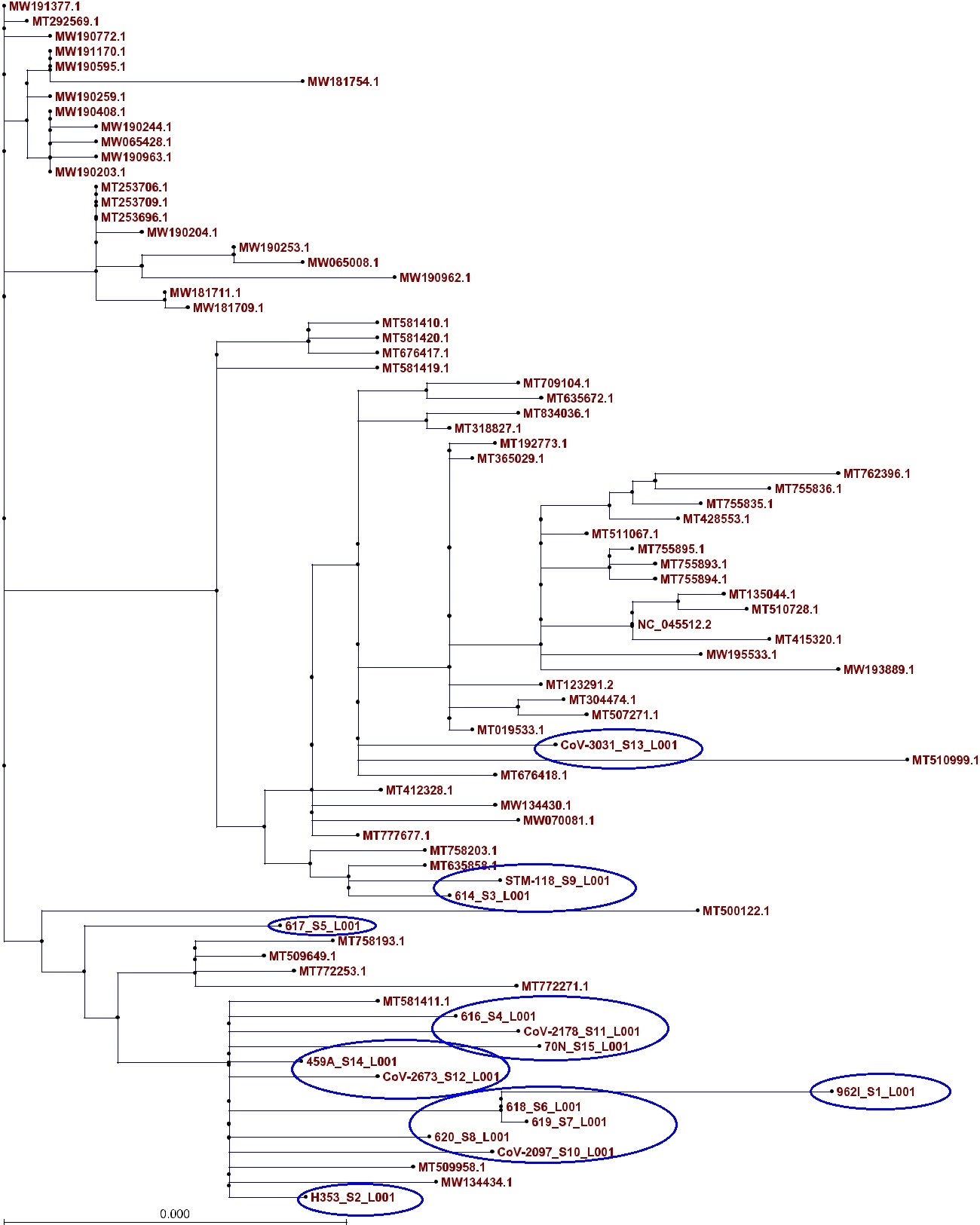
All 15 sequences were subjected to multiple sequence alignment followed by phylogenetic analysis using CLUSTAL OMEGA [30]*.* Clustal Omega utilizes seeded guide trees and HMM (Hidden Markov Models)profile-profile techniques to generate alignments.



**Figure 4:** The phylogenetic tree generated from the fifteen query sequences, is a Neighbour-joining tree without distance corrections.

When compared to the clade generated by GISAID (from the current hCoV-19 database), the clade generated by Clustal Omega have some similarities. As mentioned before, samples Nepal\_Baglung\_NPHL-459A, Nepal\_Kapilvastu\_NPHL-962l, Nepal\_Kathmandu\_NK-2097, Nepal\_Dolakha\_KU-616, Nepal\_Dulikhel\_KU-620, Nepal\_Rolpa\_NK-2678 and Nepal\_Sarlahi\_NPHL-H353 are all in a single clade, GR. In the phylogenetic tree above, Nepal\_Sarlahi\_NPHL-H353, Nepal\_NA\_KU-616, Nepal\_NA\_KU-620 and Nepal\_Kapilvastu\_NPHL-962l generate from a single node, diverging at different periods from Nepal\_NA\_KU-616 and Nepal\_NA\_KU-620, that appear in a single cluster.

Also, the samples that were in same clade in GISAID database are also in same clade when analysed through Nextstrain. In Nextstrain, all of the 15 sequences are tagged unknown when filtered for the geographical region/countries they could be related to. Additionally, the first known SARS-CoV-2 sequence, in Nepal, from January 2020 was found to be clade 19A, while the fifteen sequences, used in this analysis are in recent clade i.e., 20A and 20B.

****

**Figure 5:**Phylogenetic trees showing the relationship of the 15 virus geneome sequences from Nepalese patients to other SARS-CoV-2 genomes available from **NCBI (ANNEX-IV).**

Phylogenetic analysis of 15 genome sequences (marked oval in fig.5) revealed that all of the genome sequences were located in a cluster with genomes mainly from India, USA, Bangladesh and Saudi Arabia. Several of our sequences have formed clusters that the identical strains have been introduced from different geographic locations. Additionally, the first known SARS-CoV-2 sequence, in Nepal, from January 2020 was found to be clade 19A, while the fifteen sequences, used in this analysis are in recent clade i.e., 20A and 20B which indicate that the SARS-CoV-2 was introduced from late February.

**Conclusion:**

We conclude that SARS-CoV-2 from Nepalese patients, mostly, belongs to GH and GR (that mutated from strain G at the end of February). These strains, too, have several mutations, compared to the mutations listed to calde placement in non-structural proteins and spike proteins, possibly affecting their properties in compared to the first strain hCoV-19/Wuhan/WIV04/2019.

SARS-CoV-2 genome sequences from Nepalese patients resemble with the sequence worldwide; however, we may assume that, the probable source of Nepalese SARS-CoV-2came from India, Saudi Arabia, Europe and USA. In the absence of the information related to travel/contact history of patients, more inference and definite conclusions on the possible origin and source could not be made..

**Recommendations:**

**Immedaite**

* Validation of the PCR primer and probes (kits) need to be done in order to ensure the accuracy of the results
* The spike protein of SARS-CoV-2, one of the target region used for the development of Vaccine against SARS-CoV-2, found to have undergone several mutation. Thus there is dire need for further investigation to identify if the vaccine candidate developed targeting spike protein be worthwhile against SARS-CoV-2 reported from Nepal Needs to be elucidated further.

**Future**

* For more prolific and comprehensive analysis, it is recommended to obtain the clinical and epidemiological data that could assist in study of importation and circulation of the virus in the country.
* It is recommended to obtain more diverse samples in terms of geographical area, demography and clinical variables. In addition, protein structure determination, analysis and comparison need to be done on the consequences of mutation observed.
* Our study highlights the further need of sequencing SARS-CoV-2 genomes in conjunction with the clinical history in terms of recovery, hospitalization and co-morbidity might allow identification of variants that should be actionable and would also have relevance for prognosis and epidemiological understanding of the virus and the disease.

**References:**

1. Morens DM, Daszak P, Taubenberger JK. Escaping Pandora's Box - Another Novel Coronavirus. *The New England journal of medicine*. 2020;382(14):1293-5.
2. WHO. Clinical management of severe acute respiratory infection when Novel coronavirus (nCoV) infection is suspected: interim guidance. <https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected>. (Accessed 11 Jan 2020)
3. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *Journal of medical virology*. 2020;92(4):401-2.
4. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
5. *Ministry of Health and Population*. 2020 [cited 2020 11/02/2020]; Available from: <https://covid19.mohp.gov.np/>.
6. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). <https://coronavirus.jhu.edu/map.html>. (Accessed 3May 2020)
7. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). <https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)>. (Accessed 3may 2020)
8. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. Int J Infect Dis. 2020.
9. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging infectious diseases*. 2020;26(7).
10. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama*. 2020.
11. Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19.* Acta Pharmacologica Sinica, 2020: p. 1-9.
12. Yin, C., *Genotyping coronavirus SARS-CoV-2: methods and implications.* Genomics, 2020.
13. Wu, A., et al., *Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China.* Cell host & microbe, 2020.
14. Ye, Z.-W., et al., *Zoonotic origins of human coronaviruses.* International journal of biological sciences, 2020. **16**(10): p. 1686.
15. Walls, A.C., et al., *Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein.* Cell, 2020.
16. 1Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* nature, 2020. **579**(7798): p. 270-273.
17. Dhand, R. and J. Li, *Coughs and sneezes: their role in transmission of respiratory viral infections, including SARS-CoV-2.* American journal of respiratory and critical care medicine, 2020. **202**(5): p. 651-659.
18. McKee, D.L., et al., *Candidate drugs against SARS-CoV-2 and COVID-19.* Pharmacological Research, 2020: p. 104859.
19. Maggi, E., G.W. Canonica, and L. Moretta, *COVID-19: unanswered questions on immune response and pathogenesis.* Journal of Allergy and Clinical Immunology, 2020.
20. Shah, V.K., et al., *Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past.* Frontiers in immunology, 2020. **11**: p. 1949-1949.
21. Pachetti, M., et al., *Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant.* Journal of Translational Medicine, 2020. **18**: p. 1-9.
22. Duffy, S., *Why are RNA virus mutation rates so damn high?* PLoS biology, 2018. **16**(8): p. e3000003.
23. Meredith, L.W., et al., *Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study.* The Lancet Infectious Diseases, 2020.
24. Islam, M.R., et al., *Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity.* Sci Rep, 2020. **10**(1): p. 020-70812.
25. Chiara, M., et al., *Comparative genomics suggests limited variability and similar evolutionary patterns between major clades of SARS-CoV-2.* bioRxiv, 2020: p. 2020.03.30.016790.
26. Uddin, M., et al., *SARS-CoV-2/COVID-19: Viral genomics, epidemiology, vaccines, and therapeutic interventions.* Viruses, 2020. **12**(5): p. 526
27. Mercatelli, D. and F. M. Giorgi (2020). "Geographic and Genomic Distribution of SARS-CoV-2 Mutations." Frontiers in Microbiology 11: 1800.
28. Alm, E., Broberg, E. K., Connor, T., Hodcroft, E. B., Komissarov, A. B., Maurer-Stroh, S., ... & Pereyaslov, D. (2020). Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Eurosurveillance, 25(32), 2001410.
29. Hagemeijer, M. C., Ulasli, M., Vonk, A. M., Reggiori, F., Rottier, P. J., & de Haan, C. A. (2011). Mobility and interactions of coronavirus nonstructural protein 4. Journal of virology, 85(9), 4572-4577.
30. Chitranshi, N., Gupta, V. K., Rajput, R., Godinez, A., Pushpitha, K., Sheng, T., ... & Graham, S. (2020). Evolving geographic diversity in SARS-CoV2 and in silico analysis of replicating enzyme 3CLPro targeting repurposed drug candidates.

**Annex-I**

**SARS-CoV-2 Whole genome sequencing using ARTIC primer panel**

**Materials:**

Omega Biotek MagBind Viral DNA/RNA kit: M6246

SuperScript IV Vilo Master Mix: 11756050

COVID-19 primers

Q5® High-Fidelity 2X Master Mix: M0492S

AMPure XP beads: A63881

Qubit High Sensitivity DNA kit: Q32851

Nextera XT DNA Library Preparation Kit: FC-131-1096

Nextera XT Index Kit v2 Set A-D (96 indexes, 384 samples): FC-131-2001 - FC-131-2004

PhiX control v3: FC-110-3001

2N NaOH: R007

Preparation of cDNA

1. Isolate viral RNA using Viral DNA/RNA kit, Trizol, or equivalent.

*Many different cDNA synthesis kits can be used. The current protocol uses Superscript III*

a) Mix the following components in an 0.2mL 8-strip tube;

**Component Volume**

50µM random hexamers **1 µl**

10mM dNTPs mix (10mM each) **1 µl**

Template RNA **11 µl**

**Total** **13 µl**

Viral RNA input from a clinical sample should be between Ct 18-35. If Ct is between 12-15, then dilute the sample 100-fold in water, if between 15-18 then dilute 10-fold in water. This will reduce the likelihood of PCR-inhibition.

A mastermix should be made up in the **mastermix cabinet** and aliquoted into PCR strip tubes. Tubes should be wiped down when entering and leaving the mastermix cabinet.

1. Gently mix by pipetting and pulse spin the tube to collect liquid at the bottom of the tube.
2. Incubate the reaction as follows:

**65 °C** for **00:05:00**

Place on ice for **00:01:00**

d) Add the following to the annealed template RNA:

**Component Volume**

SSIII Buffer **4 µl**

100mM DTT **1 µl**

RNaseOUT RNase Inhibitor **1 µl**

SSIII Reverse Transcriptase **1 µl**

**Total** **20 µl**

A master mix should be made up in the **master mix cabinet** and added to the denatured RNA in the **extraction and sample addition cabinet**. Tubes should be wiped down when entering and leaving the master mix cabinet.

1. Gently mix by pipetting and pulse spin the tube to collect liquid at the bottom of the tube.
2. Incubate the reaction as follows:

|  |  |
| --- | --- |
| 42 °C | 00:50:00 |
| 70 °C | 00:10:00 |
| Hold at | 5 °C |

PCR generation of tiled amplicons

1. Prepare two primer pools by mixing equal volumes of each 10 μM primer.
2. Prepare two PCR Reactions for each sample (one for each primer pool). Add 22 μl of reaction master mix to 3 μl of cDNA for each sample.

**Component** **1x**

|  |  |
| --- | --- |
| Ampligold 360 2X MM | 12.5 μl |
|  |  |
| Primer pool (#1 or #2) | 3.6 μl |
|  |  |
| Nuclease-free water | 6.9 μl |
|  |  |
| Total | 23 |

4. Run the following cycles on a thermocycler:

|  |  |
| --- | --- |
| 98°C | 30 seconds |
|  |  |
| 95°C | 15 seconds |
|  |  |
| 65°C | 5 minutes |
|  |  |

**Repeat steps 2 & 3 for a total of 25-35 cycles**

4°C ∞

Post PCR cleanup and quantification

1. Allow AMPure XP beads to equilibrate to room temperature, vortex until homogenous.
2. Combine PCR products from the primer pool reactions for each sample to obtain a final volume of 50 μl.
3. Add 50 μl of beads to 50 μl of combined PCR product, mix well, and incubate at room temperature for 10 minutes.
4. Place tubes on a magnetic stand and incubate until solution appears clear.
5. Discard supernatant without disturbing the beads.
6. While tubes are on the magnet, add 200 μl of 80% EtOH, incubate for 30 seconds, and discard the EtOH wash.
7. Repeat previous 80% EtOH wash and remove as much EtOH as possible.
8. Leave tubes on magnet and air dry for 5 min.
9. Remove tubes from magnet and add 20 μl of nuclease-free water. Mix well by pipetting and incubate at room temperature for 10 minutes.
10. Place tubes on magnet stand. When solution appears clear, remove supernatant without disturbing the beads and place into new tubes.
11. Quantify the DNA concentration using the Qubit High Sensitivity DNA kit (or equivalent) from 1 μl of each product. Expected range = 10-100 ng/μl DNA.

Nextera XT DNA Library preparation:

1. Dilute each PCR product to 1ng/μl. Transfer 2 μl of the diluted sample to a 0.2 PCR-tube or 96 well plate.
2. Add the following (a master-mix may be prepared, if desired):

**Component** **1x**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Amplicon Tagment Mix (ATM) | 0.5 μl | | |  |  |
|  |  |  |  |  |  |  |
|  | Tagment DNA Buffer (TD) | 2.5 μl | | |  |  |
|  |  | |  |  | |  |
| 18. | Spin down, then incubate at 55°C for 5 min, hold at 10°C. | | | | | |
| 19. | Once at 10°C, immediately add 1.25 μl of Neutralize Tagment Buffer (NT) to end | | | | | |
|  | reaction. Mix by pipetting up and down. | | | |  |  |
| 20. | Incubate at room temp for 5 min. Centrifuge at 280xg (at room temp) for 1 min. | | | | | |
| 21. | Add the following to the 6.25 μl reaction: | | | |  |  |
|  | **Component** |  | **1x** | |  |  |
|  |  |  |  | |  |  |
|  | Nextera PCR Master Mix (NPM) |  | 3.75 | | l |  |
|  |  |  |  | |  |  |
|  | Index 1 primer (i7) |  | 1.25 μl | |  |  |
|  |  |  |  | |  |  |
|  | Index 2 primer (i5) |  | 1.25 μl | |  |  |
|  |  |  |  |  |  |  |

22. Pipet up and down to mix and spin down.

23.

24. Perform PCR on thermal cycler using the following conditions:

|  |  |
| --- | --- |
| 72°C | 3 min |
|  |  |
| 95°C | 30s |
|  |  |
| 95°C | 10s |
|  |  |
| 55°C | 30s |
|  |  |
| 72°C | 30s |
|  |  |
|  |  |

**Repeat steps 3-5 for a total of 18 cycles**

|  |  |
| --- | --- |
| 72°C | 5 min |
|  |  |
| 4°C | ∞ |
|  |  |

**Clean-up & Quality Control**

1. Warm AMPure XP SPRI beads to room temperature for 30 minutes.
2. Bring sample up to 50μl with H2O (Add 37.5 μl of H2O).
3. Gently shake AMPure XP bottle to resuspend any magnetic particles that may have settled.
4. Add 0.7x volume (35 μl) of AMPure XP beads to each sample; vortex and spin down.

Incubate for 10 min in a shaker (low speed) at room temperature.

1. Place mixture on the magnetic station. Wait for the solution to clear.
2. Remove cleared solution from the station and discard
3. Wash beads with 200 μl of 80% ethanol and incubate for 30 sec. Remove ethanol and discard. Repeat for a total of two washes.
4. Keep plate on the magnetic station and air-dry for ~3 min to allow the beads to dry completely. Avoid overdrying.
5. Add 20 μl of H2O to each well. Remove plate from the station to mix (vortex and spin quickly) the beads and buffer thoroughly. Incubate for 10 min.
6. Analyze fragment size using a Tapestation (D5000 chip). Determine concentration using Qubit High Sensitivity DNA kit.
7. Dilute sample down to 2nM. Molecular weight [nM] = Library concentration [ng/μl] / ((average library size x 660)/1,000,000)

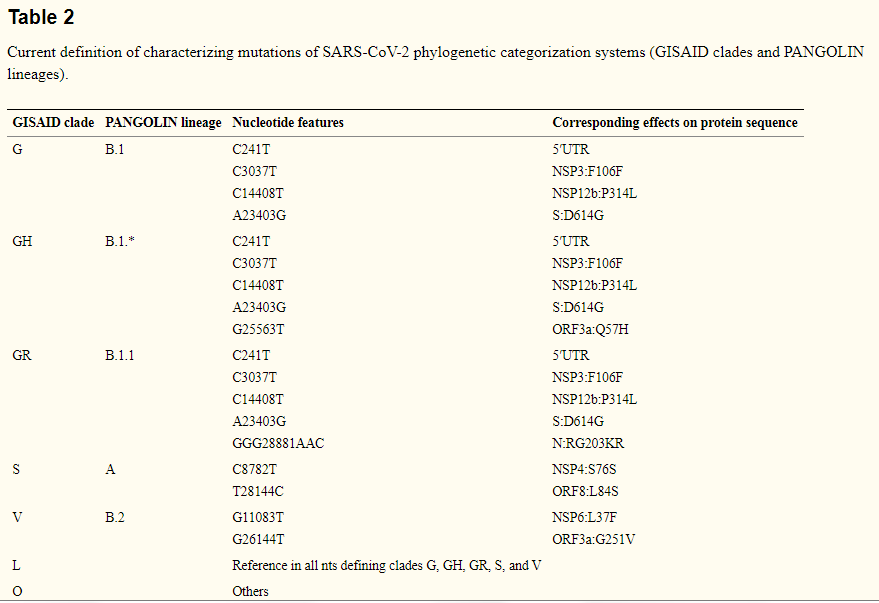
Loading the MiSeq

1. Thaw the MiSeq sequence cassette and the Hyb Buffer.
2. Pool samples in equimolar amounts at 2nM. The total volume should be more than

20 μl.

1. Prepare a 2nM solution of PhiX control by combining 1 μl of 10 nM PhiX control with 4 μl of H2O.
2. Combine 1 μl of 2nM PhiX control with 19 μl of 2nM sample pool to obtain a 2nM solution of sample pool and PhiX control.
3. Prepare a 0.2 N NaOH solution by combining 2 μl of 2N NaOH with 18 μl of H2O.
4. Combine 10 μl of 0.2 N NaOH solution with 10 μl of 2nM sample pool and PhiX control solution. Incubate at room temperature for 5 min.
5. Add 980 μl of Hyb Buffer to the 20 μl. Mix by vortexing. (This is your 20 pM solution).
6. Pipet 500 μl of the sample pool/Hyb Buffer mix from the previous step into a 1.5 ml tube. Add another 500 μl of Hyb Buffer to obtain a 10pM solution. Mix by vortexing.
7. Load 600 μl of the 10 pM solution into the designated well of the thawed MiSeq sequence cassette and follow the instructions on the MiSeq to start sequencing.

**Annex-II**

****

**Annex-III**

|  |  |  |  |
| --- | --- | --- | --- |
|  | SNPs | Number of Mutation | Mutation |
| 459A\_S14\_L001 | 24 | 17 | |  |  |  |  | | --- | --- | --- | --- | | nsp3 | V | 653 | VL | | nsp3 | A | 991 | T | | nsp3 | A | 1652 | TA | | nsp12 | P | 323 | L | | nsp13 | N | 95 | ND | | nsp14 | S | 450 | NS | | nsp16 | V | 194 | VF | | nsp16 | F | 230 | SF | | nsp16 | N | 233 | ND | | S | A | 292 | TA | | S | R | 319 | R\* | | S | D | 614 | G | | N | H | 145 | HY | | N | R | 203 | K | | N | G | 204 | R | | N | D | 225 | DV | | N | G | 238 | SG | |
| 614\_S3\_L001 | 23 | 16 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | A | 26 | T | | nsp2 | A | 357 | AV | | nsp2 | V | 381 | A | | nsp2 | C | 487 | YC | | nsp3 | E | 81 | EA | | nsp3 | D | 1283 | ND | | nsp3 | T | 1335 | TS | | nsp12 | P | 323 | L | | S | N | 30 | NH | | S | G | 35 | SG | | S | D | 53 | DG | | S | D | 614 | G | | ORF 3a | Q | 57 | H | | E | V | 47 | VF | | N | S | 194 | L | | N | N | 196 | NS | |
| 616\_S4\_L001 | 23 | 14 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | E | 261 | D | | nsp2 | C | 487 | YC | | nsp3 | K | 837 | N | | nsp3 | V | 1150 | MV | | nsp4 | D | 279 | N | | nsp5 | K | 90 | R | | nsp12 | P | 323 | L | | nsp13 | L | 256 | F | | S | D | 614 | G | | ORF 7b | F | 9 | LF | | N | R | 203 | K | | N | G | 204 | R | | N | T | 296 | I | | N | F | 403 | LF | |
| 617\_S5\_L001 | 18 | 10 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | A | 26 | T | | nsp2 | A | 357 | AV | | nsp2 | V | 381 | A | | nsp2 | C | 487 | YC | | nsp3 | V | 1150 | MV | | nsp12 | P | 323 | L | | S | D | 614 | G | | ORF 3a | Q | 57 | H | | N | S | 194 | L | | N | K | 369 | KT | |
| 618\_S6\_L001 | 20 | 15 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | A | 26 | T | | nsp2 | V | 381 | A | | nsp3 | F | 123 | LF | | nsp3 | A | 149 | AG | | nsp3 | S | 1285 | PS | | nsp4 | R | 55 | PR | | nsp4 | T | 114 | TA | | nsp12 | P | 323 | L | | nsp16 | M | 224 | MV | | S | S | 12 | PS | | S | N | 30 | NH | | S | F | 43 | LF | | S | D | 614 | G | | ORF 3a | Q | 57 | H | | N | S | 194 | L | |
| 619\_S7\_L001 | 19 | 14 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | A | 26 | T | | nsp2 | A | 357 | AV | | nsp2 | V | 381 | A | | nsp2 | C | 487 | YC | | nsp3 | P | 679 | S | | nsp3 | V | 1150 | MV | | nsp12 | P | 323 | L | | S | S | 112 | SL | | S | D | 614 | G | | ORF 3a | Q | 57 | H | | ORF 7a | A | 50 | TA | | ORF 7a | D | 51 | DG | | ORF 7b | M | 1 | MV | | N | S | 194 | L | |
| 620\_S8\_L001 | 25 | 17 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | A | 360 | AV | | nsp2 | C | 487 | YC | | nsp3 | K | 837 | N | | nsp3 | V | 1150 | MV | | nsp4 | I | 49 | IV | | nsp4 | Q | 77 | QR | | nsp4 | I | 95 | IV | | nsp4 | D | 279 | N | | nsp12 | P | 323 | L | | nsp13 | L | 256 | F | | nsp14 | F | 444 | SF | | nsp14 | S | 454 | PS | | S | S | 45 | PS | | S | D | 614 | G | | N | R | 203 | K | | N | G | 204 | R | | N | T | 296 | I | |
| 70N\_S15\_L001 | 21 | 13 | |  |  |  |  | | --- | --- | --- | --- | | nsp3 | V | 1150 | MV | | nsp3 | K | 1693 | KN | | nsp7 | S | 25 | T | | nsp12 | P | 323 | L | | nsp12 | V | 424 | L | | nsp12 | V | 605 | AV | | nsp16 | K | 182 | KE | | nsp16 | D | 220 | DG | | S | L | 141 | LF | | S | D | 614 | G | | S | C | 1248 | Y | | ORF 3a | Q | 57 | H | | N | Q | 418 | H | |
| 962I\_S1\_L001 | 19 | 13 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | C | 487 | YC | | nsp3 | P | 1316 | PS | | nsp3 | S | 1443 | TS | | nsp12 | P | 323 | L | | nsp15 | A | 231 | AV | | nsp15 | F | 279 | LF | | S | N | 30 | NH | | S | S | 71 | SF | | S | R | 78 | RG | | S | D | 614 | G | | ORF 7b | A | 15 | AV | | N | R | 203 | K | | N | G | 204 | R | |
| CoV-2097\_S10\_L001 | 23 | 15 | |  |  |  |  | | --- | --- | --- | --- | | nsp3 | M | 560 | I | | nsp3 | V | 1150 | MV | | nsp8 | A | 194 | V | | nsp12 | P | 323 | L | | nsp14 | Y | 446 | YC | | nsp16 | N | 177 | NS | | S | F | 59 | SF | | S | D | 614 | G | | S | A | 1087 | S | | ORF 3a | T | 223 | I | | ORF 7a | E | 95 | EG | | ORF 8 | Q | 23 | KQ | | N | D | 63 | Y | | N | R | 203 | K | | N | G | 204 | R | |
| CoV-2178\_S11\_L001 | 25 | 19 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | A | 357 | AV | | nsp2 | V | 381 | A | | nsp2 | I | 393 | T | | nsp2 | C | 487 | YC | | nsp3 | V | 1150 | MV | | nsp12 | P | 323 | L | | nsp15 | L | 200 | S | | S | A | 27 | S | | S | T | 51 | TA | | S | G | 184 | DG | | S | Q | 218 | QR | | S | R | 246 | RG | | S | D | 389 | ND | | S | D | 614 | G | | ORF 3a | V | 55 | G | | ORF 3a | Q | 57 | H | | N | S | 194 | L | | N | K | 369 | KR | | N | D | 377 | DG | |
| CoV-2673\_S12\_L001 | 17 | 11 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | C | 487 | YC | | nsp3 | E | 120 | EA | | nsp8 | A | 194 | V | | nsp12 | P | 323 | L | | S | F | 58 | SF | | S | N | 148 | ND | | S | D | 614 | G | | ORF 3a | T | 223 | I | | N | T | 135 | I | | N | R | 203 | K | | N | G | 204 | R | |
| CoV-3031\_S13\_L001 | 22 | 17 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | A | 357 | AV | | nsp2 | V | 381 | A | | nsp2 | I | 393 | T | | nsp2 | C | 487 | YC | | nsp3 | E | 81 | EV | | nsp3 | S | 126 | PS | | nsp3 | V | 1150 | MV | | nsp12 | P | 323 | L | | nsp15 | L | 200 | S | | S | V | 42 | AV | | S | G | 142 | GC | | S | D | 614 | G | | ORF 3a | V | 55 | G | | ORF 3a | Q | 57 | H | | N | S | 194 | L | | N | A | 381 | AG | | N | \* | 420 | Q\* | |
| H353\_S2\_L001 | 16 | 8 | |  |  |  |  | | --- | --- | --- | --- | | nsp3 | E | 124 | EG | | nsp3 | K | 1715 | KR | | nsp12 | P | 323 | L | | nsp13 | K | 139 | KR | | S | D | 614 | G | | S | D | 745 | DG | | N | R | 203 | K | | N | G | 204 | R | |
| STM-118\_S9\_L001 | 15 | 8 | |  |  |  |  | | --- | --- | --- | --- | | nsp2 | C | 487 | YC | | nsp3 | E | 134 | EG | | nsp3 | V | 1150 | MV | | nsp12 | P | 323 | L | | S | D | 614 | G | | ORF 3a | Q | 57 | H | | N | L | 139 | F | | N | S | 194 | L | | Non-codon-aligned | ORF 7a | 27567 | 27618 | |

**Annex-IV: Whole Genome Sequence of SARS-CoV-2 retived from NCBI**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accession | Release\_Date | Length | Geo\_Location | Collection\_Date |
| NC\_045512 | 2020-01-13. | 29903 | China | 2019-12 |
| MW193889 | 2020-11-02. | 29895 | Israel | 7/20/2020 |
| MW195533 | 2020-11-02. | 29899 | USA: Virginia | 2020-09 |
| MW181709 | 2020-11-01. | 29782 | Hong Kong | 3/27/2020 |
| MW181711 | 2020-11-01. | 29782 | Hong Kong | 3/20/2020 |
| MW181754 | 2020-11-01. | 29782 | Hong Kong | 10/2/2020 |
| MT581410 | 2020-10-31. | 29822 | Bangladesh | 4/29/2020 |
| MT581411 | 2020-10-31. | 29822 | Bangladesh | 4/29/2020 |
| MT581419 | 2020-10-31. | 29822 | Bangladesh | 4/29/2020 |
| MT581420 | 2020-10-31. | 29822 | Bangladesh | 4/29/2020 |
| MT676417 | 2020-10-31. | 29824 | Bangladesh | 5/18/2020 |
| MT676418 | 2020-10-31. | 29851 | Bangladesh | 5/18/2020 |
| MW190203 | 2020-10-30. | 29782 | USA: CA | 3/17/2020 |
| MW190204 | 2020-10-30. | 29782 | USA: CA | 3/17/2020 |
| MW190244 | 2020-10-30. | 29782 | USA: CT | 3/16/2020 |
| MW190253 | 2020-10-30. | 29782 | USA: MA | 3/16/2020 |
| MW190259 | 2020-10-30. | 29782 | USA: MA | 3/12/2020 |
| MW190408 | 2020-10-30. | 29782 | USA: MI | 3/17/2020 |
| MW190595 | 2020-10-30. | 29773 | USA: UT | 3/15/2020 |
| MW190772 | 2020-10-30. | 29782 | USA: LA | 3/17/2020 |
| MW190962 | 2020-10-30. | 29782 | USA: GA | 3/13/2020 |
| MW190963 | 2020-10-30. | 29782 | USA: GA | 3/15/2020 |
| MW191170 | 2020-10-30. | 29782 | USA: WA | 3/16/2020 |
| MW191377 | 2020-10-30. | 29782 | USA: MO | 3/15/2020 |
| MW134430 | 2020-10-20. | 29854 | USA: CA, Orange County | 8/5/2020 |
| MW134434 | 2020-10-20. | 29733 | USA: CA, Orange County | 8/13/2020 |
| MW065008 | 2020-10-05. | 29782 | USA: CA | 3/17/2020 |
| MW065428 | 2020-10-05. | 29782 | USA: MN | 3/16/2020 |
| MW070081 | 2020-10-05. | 29846 | USA: New Mexico | 8/31/2020 |
| MT834036 | 2020-08-04. | 29870 | USA: Washington,Yakima County | 6/1/2020 |
| MT777677 | 2020-07-20. | 29809 | France | 5/1/2020 |
| MT772253 | 2020-07-17. | 29791 | India: Surat | 6/13/2020 |
| MT772271 | 2020-07-17. | 29800 | India: Choryasi | 6/13/2020 |
| MT758193 | 2020-07-15. | 29800 | India: Vadodara | 6/5/2020 |
| MT758203 | 2020-07-15. | 29800 | India: Modasa | 6/14/2020 |
| MT762396 | 2020-07-15. | 29903 | Bangladesh: Jashore | 7/7/2020 |
| MT755835 | 2020-07-14. | 29903 | Bangladesh: Barishal | 7/6/2020 |
| MT755836 | 2020-07-14. | 29903 | Bangladesh: Barishal | 7/6/2020 |
| MT755893 | 2020-07-14. | 29903 | Saudi Arabia | 2/18/2020 |
| MT755894 | 2020-07-14. | 29903 | Saudi Arabia | 2/20/2020 |
| MT755895 | 2020-07-14. | 29903 | Saudi Arabia | 2/23/2020 |
| MT709104 | 2020-07-06. | 29838 | France | 5/14/2020 |
| MT635672 | 2020-06-18. | 29832 | Bangladesh | 6/7/2020 |
| MT635858 | 2020-06-18. | 29800 | India: Gujarat, Nadiad | 6/8/2020 |
| MT510728 | 2020-05-26. | 29903 | China | 2/13/2020 |
| MT510999 | 2020-05-26. | 29890 | Netherlands: Leiden | 3/9/2020 |
| MT511067 | 2020-05-26. | 29903 | Poland | 4/2/2020 |
| MT509958 | 2020-05-25. | 29724 | Bangladesh | 5/11/2020 |
| MT509649 | 2020-05-24. | 29793 | India: Vadodara | 5/2/2020 |
| MT507271 | 2020-05-22. | 29882 | Jamaica | 3/9/2020 |
| MT500122 | 2020-05-21. | 29819 | Pakistan: Karachi | 3/16/2020 |
| MT365029 | 2020-05-05. | 29891 | Hong Kong | 1/30/2020 |
| MT428553 | 2020-05-05. | 29903 | Kazakhstan | 3/20/2020 |
| MT318827 | 2020-05-04. | 29870 | Germany | 3/19/2020 |
| MT412328 | 2020-04-30. | 29834 | USA: CT | 4/2/2020 |
| MT415320 | 2020-04-30. | 29901 | India | 3/1/2020 |
| MT304474 | 2020-04-07. | 29882 | South Korea | 2/27/2020 |
| MT292569 | 2020-04-06. | 29782 | Spain | 3/9/2020 |
| MT253696 | 2020-03-27. | 29781 | China: Zhejiang, Hangzhou | 1/23/2020 |
| MT253706 | 2020-03-27. | 29781 | China: Zhejiang, Hangzhou | 1/22/2020 |
| MT253709 | 2020-03-27. | 29781 | China: Zhejiang, Hangzhou | 1/21/2020 |
| MT192773 | 2020-03-16. | 29890 | Viet Nam: Ho Chi Minh city | 1/22/2020 |
| MT135044 | 2020-03-04. | 29903 | China: Beijing | 1/28/2020 |
| MT123291 | 2020-02-28. | 29882 | China: Guangzhou | 1/29/2020 |
| MT019533 | 2020-02-05. | 29883 | China: Hubei, Wuhan | 1/1/2020 |