Whole Genome Sequence Analysis to Identify SARS-CoV-2 Variant in Nepal

Shrestha R,[1,2,3†] Katuwal N,[2,3†] Adhikari N,[2,3†] Vanaerschot M,[4] Tamrakar D,[2,5] Dhimal M,[6] Gyanwali P,[6] Bhattarai S,[6] Madhup SK,[7] Devkota B[8]

[1]Department of Pharmacology,
Kathmandu University School of Medical Sciences,
Dhulikhel, Kavre, Nepal.

[2]Research and Development Division,
Dhulikhel Hospital, Kathmandu University Hospital,
Dhulikhel, Kavre, Nepal.

[3]Genomic Research Lab,
Dhulikhel Hospital, Kathmandu University Hospital,
Dhulikhel, Kavre, Nepal.

[4]Chan Zuckerberg Biohub,
San Francisco, USA.

[5]Department of Community Medicine,
Kathmandu University School of Medical Sciences,
Dhulikhel, Kavre, Nepal.

[6]Nepal Health Research Council, Ramshah Path,
Kathmandu, Nepal.

[7]Department of Microbiology,
Dhulikhel Hospital, Kathmandu University Hospital,
Dhulikhel, Kavre, Nepal.

[8]Ministry of Health and Population,
Kathmandu, Nepal.

([†]Equal Contribution)

Corresponding Author
Rajeev Shrestha
Department of Pharmacology,
Kathmandu University School of Medical Sciences,
Research and Development Division,
Dhulikhel Hospital, Kathmandu University Hospital,
Dhulikhel, Kavre, Nepal.
E-mail: rajeev.shrestha@kusms.edu.np

Citation

Shrestha R, Katuwal N, Adhikari N, Vanaerschot M, Tamrakar D, Dhimal M, et al. Whole Genome Sequence Analysis to Identify SARS-CoV-2 Variant in Nepal. *Kathmandu Univ Med J.* 2021;74(2):237-42.

ABSTRACT

Background

The spread of SARS-CoV-2 has become a global public health crisis. Nepal is facing the second wave of COVID-19 pandemic but, there is still a limited data on the genomic sequence of SARS-CoV-2 variants circulating in Nepal.

Objective

The objective of this study is to sequence the whole genome of SARS-CoV-2 in Nepal to detect possible mutation profiles and phylogenetic lineages of circulating SARS-CoV-2 variants.

Method

In this study, swab samples tested positive for SARS-CoV-2 were investigated. After RNA extraction, the investigation was performed through real-time PCR followed by whole genome sequencing. The consensus genome sequences were, then, analyzed with appropriate bioinformatics tools.

Result

Sequence analysis of two SARS-CoV-2 genomes from patient without travel history (Patient A1 and A2) were found to be of lineage B.1.1. Similarly, among other four samples from subjects returning from the United Kingdom, genomes of two samples were of lineage B.1.36, and the other two were of lineage B.1.1.7 (Alpha Variant). The mutations in the consensus genomes contained the defining mutations of the respective lineages of SARS-CoV-2.

Conclusion

We confirmed two genomic sequences of variant of concern VOC-202012/01 in Nepal. Our study provides the concise genomic evidence for spread of different lineages of SARS-CoV-2 – B.1.1, B.1.36 and B.1.1.7 of SARS-CoV-2 in Nepal.

KEY WORDS

COVID-19, Nepal, SARS-CoV-2, Variant of concern, Whole genome sequencing

INTRODUCTION

Coronavirus diseases (COVID-19) is caused by severe acute respiratory syndrome - coronavirus 2 (SARS-CoV-2) that became pandemic since detected on December 2019 in Wuhan, China.[1] Since then, COVID-19 has spread rapidly across the world and has caused a wide range of clinical manifestation in humans, from asymptomatic conditions to critical and fatal cases.[2]

The high mutation rate of SARS-CoV-2 is an alarming situation, as such mutations help the virus to adapt, even in unfavorable conditions.[3] Despite its lower evolutionary rate in compare to other RNA viruses, the proofreading function of the RNA polymerase might causing mutations resulting to its continuous evolution.[4] A related study in mutation pattern, found that 80% of the mutations occurring in SARS-CoV-2 genome, are in the spike protein.[5] Similarly, Orf1ab gene is harbour to a large number of mutations causing changes in important structural proteins and non-structural proteins such as Nsp6, Nsp11 and Nsp13.[5] Although most viral mutations are benign, mutations on spike protein (like D614G) has found to strengthen viral survival capability.[6]

In course of genomic investigation, it is important to decipher the effects of mutations on emergence and spread of genetic variants.[7] In December 2020, COVID-19 Genomics UK (COG-UK) consortium identified a new variant of SARS-CoV-2, with enhanced transmission, named VOC-202012/01 (also known as lineage B.1.1.7, 20I/501Y.V1).[8] Recently, a new variant B.1.617, first identified in India, has been circulating with higher transmissibility and pathogenicity.[9,10] The mutations such as substitution N501Y (substitution of Asparagine by Tyrosine), in the receptor binding domain of the viral genome, reportedly increases the affinity of viral proteins to human ACE2 receptor (Angiotensin-converting enzyme 2).[11] Similarly, 69-70 deletion is known to increase infectivity and leading to greater spike cleavage. It is evident from studies that, B.1.17 requires this deletion for infectivity and optimal cleavage.[12] Usually, the mutations in Spike gene have been associated with escape from immune response, increased transmissibility, ACE2 receptor binding, and severity.[7]

Consequently, it is of great importance to comprehend the existing mutations on the viral genome in reference with the infectivity and subsequently speculate immune protection, transmission dynamics and future infection tendency.[13,14] Since the start of the pandemic, researches all around the world have successfully generated and shared viral genome-sequence data (GISAID, Nextstrain, NCBI). The investigations with genomic data focusing on emerging mutations has helped new researches to understand the outbreak in their respective countries and manage them. This has also aided to gain concurring insights into the pandemic, understand the disease and curate medical countermeasures.[15]

Therefore, through whole genome data of SARS-CoV-2 strains circulating in Nepal, we aimed to understand molecular epidemiology of COVID-19 in Nepal with in-depth information on virulence/pathogenesis, transmission clusters, mutation patterns, genetic variations and understanding of the biological relevance those genetic changes. Furthermore, studying and disseminating genomic data of SARS-CoV-2 with the global scientific society is vital for collaborative scientific efforts to collectively combat this global pandemic.

Thus, in this study, we intend to perform whole genome sequencing of SARS-CoV-2 from samples in Nepal, and analyze the data to identify the genetic mutations and circulating variants.

METHODS

Sample

The samples tested positive for SARS-CoV-2 were obtained for the first sequencing. The sample had tested positive on August 24, 2020 (First episode –A1) and had been tested negative on September 3, 2020 prior to tested positive again on October 6, 2020 (Second episode –A2). Additionally, for second run, aliquot of six positive samples were obtained from National Public Health Laboratory (NPHL), through Ministry of Health and Population (MOHP)-Nepal on 10 January 2021. These samples (R1, R2, R3, R4, R5 and R6) were of people who had recently travelled the United Kingdom (UK) and those samples were selected as a purposive sampling.

RNA extraction and Real Time PCR

The RNA was extracted using the extraction kit QIAamp Viral RNA Mini Kit (Cat. No. 52904) from Qiagen and RT-PCR was done in BioRad CFX96 Detection System (1855195) using SARS-CoV-2 Fluorescent PCR Kit (Cat. No. BUSGN7103109, Maccura).

Whole Genome Sequencing

The sequencing was performed in two batches (A1 and A2 in first batch and R1, R2, R3, R4, R5 and R6 in second batch). The library preparation was performed using protocol from Illumina for Ampliseq Amplicon Sequencing, with few modifications.[16] Firstly, extracted samples were treated with DNase I (Zymo Research, Cat. No. E1010) to eliminated any DNA present in the sample. After that, cleanup was performed using 2X SPRI (Solid Phase Reversible Immobilization) beads. The cleaned RNA was quantified using Qubit™ RNA HS Assay kit (Invitrogen Cat. No. Q32852). cDNA was synthesized for all the samples using AmpliSeq™ cDNA synthesis for Illumina® (Cat. No. 20022654). Subsequently, the library preparation was achieved by AmpliSeq™ Library PLUS for Illumina® (Cat. No. 20019101), in which cDNA was amplified using two primer pools (AmpliSeq Custom RNA Panel Pool 1 and Pool

2). Post-amplification, pool 1 and pool 2 amplicons were pooled, partially digested with FuPa reagent, ligated with AmpliSeq™ UD Indexes for Illumina® (Cat. No. 20019104) and thus prepared library was cleaned up with 1X SPRI beads. The individual libraries were amplified with Library Amplification Mix, followed by two step clean-ups (0.5X and 0.8X SPRI Beads). The quality control (QC) of the individual libraries was achieved through Kapa Library Quantification Primer Premix for Illumina (KAPA Biosystems Cat. No. KP0005), followed by agarose gel electrophoresis to envisage the amplicons. Finally, all the individual libraries were quantified using Qubit™ 1X dsDNA HS Assay kit (Invitrogen Cat. No. Q33231). The libraries were pooled together and re-quantified. Finally, 100pM of the pooled library, along with 5% PhiX control, was loaded on the Illumina iSeq100.

Sequence data study

The genomic sequences from the Illumina iSeq100 were base called and subsequently, demultiplexed using Illumina BaseSpace. The generated fastq files were imported to IDSeq that uses Consensus Genome Piepline v3.1.1 and aligned to the reference genome (MN908947.3). IDseq creates consensus fasta sequence from the aligned reads. Thus obtained consensus genome were further analysed in Nextclade and GISAID Covsurver platforms. Additionally, for lineage assignment, the genomes were subjected to analysis in Phylogenetic Assignment of Named Global Outbreak LINeages (PANGOLIN).[17]

RESULTS

A Real Time PCR for diagnosis

As mentioned before, the samples selected in this study were previously confirmed RT-PCR positives for SARS-CoV-2. However, after re-extraction at DHKUH, RT-PCR was again performed to re-verify the presence of SARS-CoV-2 in the samples. A summary of RT-PCR data is provided in Table 1. For instance, the Sample A1 shows the higher Ct values compared to the Sample A2.

Table 1. RT-PCR data of the analyzed samples

Nomenclature	RT-PCR Date	Ct Values	
		N gene	E gene
Sample A1	24 August 2020	33.6	31.0
Sample A2	6 October 2020	20.69	19.23
Sample R1	January 2021	35.41	34.74
Sample R2	January 2021	25.63	25.36
Sample R3	January 2021	35.82	36.06
Sample R4	January 2021	17.75	20.52
Sample R5	January 2021	26.94	30.08
Sample R6	January 2021	25.82	27.21

Genome analysis reveals different SARS-CoV-2 lineages in Nepal

The consensus genomes generated form our analysis had average genome coverage of ~99.6% (Table 2). In GISAID, clade-based analysis of the Sample A1 and A2 showed that both the sequences belonged to clade GR (20B in Nextclade nomenclature). Among the UK returned subjects, Samples R1 and R3 failed the quality check analysis that followed the library preparation, therefore, were not sequenced. After sequencing of quality-control passed samples (R2, R4, R5 and R6), it was observed that Samples R4 and R5 belonged to clade GH (20A in Nextclade nomenclature), while R2 and R6 belonged to clade GRY (20I/501Y.V1 in Nextclade nomenclature). Further, analysis of lineage by PANGOLIN revealed the presence of three variants: B.1.1 for sample Sample A1 and Sample A2, B.1.36 of Sample R4 and R5, and B.1.1.7 of Samples R2 and R6 (shown in table 2 and figure 1).

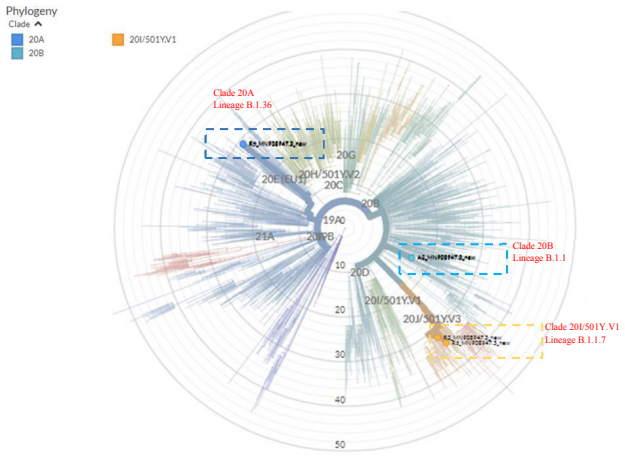

Figure 1. Phylogenetic analysis of SARS-CoV-2 sequences from Nepal (https://nextstrain.org/).

In both samples A1 and A2, mutation analysis in the GISAID CoVsurver mutations app showed six existing mutations. The mutations were identical in the both of Sample A genomes (A1 and A2), including the characteristic mutations of clade GR: Spike_D614G and NSP12_P323L. Similarly, the Sample R4 and R5 had characteristic mutations of clade GH: Spike_D614G and NS3_Q57H. Interestingly, both of these samples had two unique mutations: Spike_ins674QT, which is observed in only 5 of GISAID submissions until May 2021, and NSP3_M1788L, which is observed in 30 GISAID submissions until May 2021. Genomes from Sample R2 and R6 belonged to clade 20I/501Y.V1. This clade of Alpha Variant has already been labeled as a variant of concern (VOC). Additionally, these two genomes (R2 and R6) had unique characteristic mutations of clade GRY: Spike-H69del, Spike-V70del, Spike-Y144del, Spike-N501Y, Spike-D614G and N-G204R (Table 3).

Table 2. Clade, lineage of patients

Nomenclature	GISAID ID	Genome coverage	Sequencing depth	Lineage (PANGOLIN)	Most common countries (PANGOLIN)	Nextclade	GISAID
Sample A1	EPI_ISL_2171040	99.4	237.9	B.1.1	Global distribution	20B	GR
Sample A2	EPI_ISL_2171041	99.8	35644.0	B.1.1			
Sample R2	EPI_ISL_2171042	99.7	10548.2	B.1.1.7	UK	20I/501Y.V1	GRY
Sample R4	EPI_ISL_2171144	99.7	11397.3	B.1.36	India, Canada, UK	20A	GH
Sample R5	EPI_ISL_2171240	99.5	11531.0	B.1.36			
Sample R6	EPI_ISL_2171241	99.6	6999.5	B.1.1.7	UK	20I/501Y.V1	GRY

Table 3. Mutation analysis with reference to hCoV19/Wuhan/WIV04/2019 (GISAID)

Nomenclature	Total mutation	#Unique mutation	Unique mutations list	#Existing mutation	Existing mutations list
Sample A1	6	0		6	NSP3_K1693N, NSP12_P323L, Spike_D614G, Spike_E654Q, N_G204R, N_R203K
Sample A2	6	0		6	NSP3_K1693N, NSP12_P323L, Spike_D614G, Spike_E654Q, N_G204R, N_R203K
Sample R2	26	0		26	NSP3_T183I, NSP3_A890D, NSP3_I1412T, NSP5_L75F, NSP6_F108del, NSP6_G107del, NSP6_S106del, NSP12_P323L, NSP13_K460R, Spike_A570D, Spike_S982A, Spike_P681H, Spike_D614G, Spike_T716I, Spike_Y144del, Spike_N501Y, Spike_D1118H, Spike_V70del, Spike_H69del, NS8_R52I, NS8_Q27stop, NS8_Y73C, N_S235F, N_D3L, N_G204R, N_R203K
Sample R4	16	2	NSP3_M1788L, Spike_ins674QT	14	NSP2_V381A, NSP3_S403L, NSP6_I168V, NSP12_P323L, NSP13_A598V, Spike_P681H, Spike_D936H, Spike_D614G, Spike_A570S, Spike_K1073N, Spike_S477N, NS3_Q57H, NS3_V55G, N_S194L
Sample R5	16	2	NSP3_M1788L, Spike_ins674QT	14	NSP2_V381A, NSP3_S403L, NSP6_I168V, NSP12_P323L, NSP13_A598V, Spike_P681H, Spike_D936H, Spike_D614G, Spike_A570S, Spike_K1073N, Spike_S477N, NS3_Q57H, NS3_V55G, N_S194L
Sample R6	26	0		26	NSP3_T183I, NSP3_A890D, NSP3_I1412T, NSP5_C160F, NSP6_F108del, NSP6_G107del, NSP6_S106del, NSP12_P323L, NSP13_P47S, Spike_A570D, Spike_S982A, Spike_P681H, Spike_D614G, Spike_T716I, Spike_Y144del, Spike_N501Y, Spike_D1118H, Spike_V70del, Spike_H69del, NS8_R52I, NS8_Q27stop, NS8_Y73C, N_S235F, N_D3L, N_G204R, N_R203K

Sequence Submission

All the six SARS-CoV-2 genomes generated in this study were submitted to GISAID on May 20, 2021. The accession number, for the genomes, provided by GISAID are: EPI_ISL_2171040 for genome Sample A1, EPI_ISL_2171041 for genome Sample A2, EPI_ISL_2171042 for Sample R2, EPI_ISL_2171144 for Sample R4, EPI_ISL_2171240 for Sample R5 and EPI_ISL_2171241 for Sample R6.

DISCUSSION

This study investigated the possible mutation profiles of SARS-CoV-2 circulating in Nepal. All the samples included in this study were tested positive for SARS-CoV-2 using RT-PCR. Based on the Ct values, the viral loads were different in the samples and can be correlated to the adverse outcomes independent of other demographic (age) and clinical parameters (comorbidities and severity of illness).[18] However, due to absence of demographic and clinical data, this proposition was not pursued. Various factors like sample collection method, its source, sample transfer including incubation period can affect the Ct value.

As far as the library preparation is concerned, samples R1 and R3 failed the quality check as they only had adapter dimers without any libraries, after library preparation. This could be due to their high Ct values which correlates to low viral load in the sample. For this and several such reasons, CDC (Centre for Disease Control and Prevention) also suggests the use of samples with RT-PCR Ct value<28 for sequencing.[19]

The coverage width amongst the consensus genome were highly similar, with average genome coverage of 99.6%, indicating excellent coverage. The sequencing depth, on the other hand, was highly variable among the consensus genome. Sequencing depth represents the number of

nucleotides contributing to a portion of an assembly. For a specific nucleotide, it represents the number of sequences that added information about that nucleotide. However, a steady pattern can be seen, where samples with Ct value < 25 have high sequencing depth. This is expected as samples with low Ct value they have high viral load, corresponding to high amount of viral genome in the sample. This abundance leads to optimum utilization of sequencing reagents leading to high sequencing depth.

When compared to the reference genome hCoV19/Wuhan/WIV04/2019 in GISAID, genomes A1 and A2 fell in clade 20B and pangolin lineage B.1.1 and genomes R4 and R5 belonged to clade 20A and pangolin lineage B.1.36. Additionally, R4 and R5 belonged to clade 20I/501Y.V1 and pangolin lineage B.1.1.7. The presence of B.1.1 lineage of SARS-CoV-2 was expected in Nepal, in sample A1 and A2 which were collected respectively in August and October 2020. This particular variant was the most widespread variant (74% worldwide) till June 2020 and was also spreading in Nepal during August 2020.[20,21] The samples R4 and R5 were obtained from those patients who had travelled UK suggesting that B.1.36 strain was the imported strain from UK as this strain was circulating since February, 2020 including UK.[21] Therefore, this strain was expected to appear in the patients with travel history from UK.

The genomes A1 and A2 of B.1.1 lineage have two mutations, which could have facilitated the spread of the virus. One of the mutations was in the spike protein (Spike_D614G) and the other in RNA polymerase (NSP12_P323L). The substitution mutation D614G (Aspartic acid substituted by Glycine) resulting in enhanced replication and transmissibility, is known to strengthen viral survival.[6,13] Similarly, through structural study and homology modeling, substitution mutation in orf3a, of B.1.36 lineage, is known to have important role to form hydrophilic constrictions in the transmembrane helix 1 (TM-1) and help in viral oligomerization.[22] Interestingly, both of B.1.36 genomes have a unique mutation ins674QT (insertion mutation of Glutamine and Threonine), which has been documented only five times worldwide.[23] This mutation is present in the region between Receptor Binding Domain (RBD) and Fusion Protein (FP) of the Spike Protein, and is close to the furin cleavage site (FRRG motif) as well.[24] Consequently, this mutation could influence the protein folding mechanism, thus possibly altering the protein's function. Therefore, with in-vitro cell study, the role of this mutation should be investigated further. Another unique mutation NSP3_M1788L was seen in genomes of R4 and R5. This mutation, though present in non-structural protein, has occurred only 35 times in 12 countries until May 2021. The first strain with this substitution, collected in June 2020, was hCoV-19/Japan/IC-0065/2020 and has occurred most recently in strain hCoV-19/India/SK-1122500044182/2021, collected in March 2021.

Interestingly, the genomes of R2 and R6 belong to clade 20I/501Y.V1 and are part of lineage B.1.1.7., currently known as Alpha Variant. This can be confirmed with the fact that samples R2 and R6 were from the patients who had returned from UK on January 2021. The SARS-CoV-2 variant of B.1.1.7 lineage was first identified in early December 2020 and was declared as a variant of concern VOC 202012/01 by Public Health England (PHE) and World Health Organization (WHO) on 21 December 2020.[25,26] This lineage is reported to have 23 mutations, including mutations, of biological significance, in the spike protein, The defining mutations of B.1.1.7 were found in the genomes of both R2 and R6. The most notable mutations are: substitution N501Y (Asparagine substituted by Tyrosine) in the receptor-binding domain (RBD) which increases the binding affinity to human ACE2 receptor; deletion 69-70del and substitution P681H present adjacently to the furin cleavage site, a known location of biological significance. Therefore, these mutation in S gene has, therefore, been associated with increased transmissibility, ACE-2 receptor binding, severity and escape from immune response.[7]

However, due to unavailability of clinical data, this study was not able to co-relate the Ct-value and types of variants with the clinical status. In addition, the clinical outcomes also could not be known due to lack of metadata.

CONCLUSION

In our study, we identified two samples each of lineage B.1.1 (lineage B.1.36 and lineage B.1.1.7) of SARS-CoV-2. In addition, this study provides a genomic evidence for spread of different lineages of SARS-CoV-2 at different time intervals during the first wave of COVID-19 in Nepal.

It is evident that viral genomic sequences allow the identification of pathogens, their origin, transmission, genetic diversity and outbreak dynamics. Furthermore, genomic analyses, in general, are capable of depicting the attributes of the epidemiological dynamics of viral outbreaks, which are irredeemable with just epidemiological data. Therefore, we advocate for investigation of burden of COVID-19 pandemic through whole genome sequencing.

ACKNOWLEDGEMENT

We want to acknowledge Bill & Melinda Gates Foundation for supporting iSeq100 platform which helped us to perform whole genome sequencing of SARS-CoV-2. We also would like to thank the Ministry of Health and Population of Nepal for facilitating in acquiring samples.

REFERENCES

1. Morens DM, Daszak P, Taubenberger JK. Escaping Pandora's Box — Another Novel Coronavirus. *N Engl J Med*. 2020 Apr 2;382(14):1293-5.
2. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N Engl J Med*. 2020 Mar 5;382(10):970-1.
3. Martinez IL, Llinás DT, Romero MB, Salazar LM. High Mutation Rate in SARS-CoV-2: Will It Hit Us the Same Way Forever? *Journal of Infectious Diseases and Epidemiology*. 2020; 6:176.
4. Surleac M, Banica L, Casangiu C, Cotic M, Florea D, Sandulescu O, et al. Molecular Epidemiology Analysis of SARS-CoV-2 Strains Circulating in Romania during the First Months of the Pandemic. *Life* (Basel) [Internet]. 2020 Aug 14 [cited 2021 May 9];10(8). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7460100/
5. Dorp LV, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*. 2020 Sep 1;83:104351.
6. Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *Nature Communications*. 2020; 11:6013.
7. Maio ND, Walker CR, Turakhia Y, Lanfear R, Corbett-Detig R, Goldman N. Mutation Rates and Selection on Synonymous Mutations in SARS-CoV-2. *Genome Biology and Evolution*. 2021;13(5).
8. Tang JW, Tambyah PA, Hui DS. Emergence of a new SARS-CoV-2 variant in the UK. *Journal of Infection*. 2021 Apr 1;82(4):e27–8.
9. World Health Organisation. Weekly epidemiological update on COVID (Emergency Situational Updates). Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---11-may-2021 [Accessed on 14 May, 2021]
10. Yadav PD, Mohandas S, Shete AM, Nyayanit DA, Gupta N, Patil DY, et al. SARS CoV-2 variant B.1.617.1 is highly pathogenic in hamsters than B.1 variant. *bioRxiv.* 2021 May 5;2021.05.05.442760.
11. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, et al. The N501Y spike substitution enhances SARS-CoV-2 transmission. *bioRxiv*. 2021; 2021.03.08.434499.
12. Meng, Bo, Kemp SA, Ferriera IATM, et al. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70. *Cell Reports*. 2021; 45(13); 109292.
13. Korber B. Fischer WM, Gnanakran S, Yoon H, Theiler J, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 2020; 182(4): 812-27.
14. Abdullahi IN, Emeribe AU, Ajayi OA, Oderinde BS, Amadu DO, Osuji AI. Implications of SARS-CoV-2 genetic diversity and mutations on pathogenicity of the COVID-19 and biomedical interventions. *Journal of Taibah University Medical Sciences*. 2020; 15(4): 258-64.
15. Nie Q, Li X, Chen W, Liu D, Chen Y, Li H, et al. Phylogenetic and phylodynamic analyses of SARS-CoV-2. *Virus Research*. 2020 Oct 2;287:198098.
16. Illumina Inc. AmpliSeq for Illumina On-Demand, Custom, and Community Panels: Reference Guide. November 2019. Available from: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/ampliseq-for-illumina/ampliseq-for-illumina-custom-and-community-panels-reference-guide-1000000036408-09.pdf
17. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*. 2020 Nov;5(11):1403-7.
18. Zacharioudakis IM, Prasad PJ, Zervou FN, Basu A, Inglima K, Weisenberg SA, et al. Association of SARS-CoV-2 Genomic Load with Outcomes in Patients with COVID-19. Annals of the American Thoracic Society. 2021; 18(5): 900-3.
19. Center for Disease Control and Prevention. COVID-19 vaccine breakthrough case investigation. Available From: https://www.cdc.gov/vaccines/covid-19/downloads/Information-for-laboratories-COVID-vaccine-breakthrough-case-investigation.pdf [Accessed on 2 June, 2021]
20. Nepal Health Research Council. Deciphering Whole Genomic Sequence of SARS-CoV-2 Isolated from COVID-19 Patients in Nepal. November 2020. Available From: http://nhrc.gov.np/wp-content/uploads/2021/01/WGS_Report.docx
21. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Frontiers in Microbiology.* 2020;11: 1800.
22. Alosaimi B, Naeem A, Alghoribi MF, Okdah L, Hamed ME, AlYami AS, et al. Structural Mapping of Mutations in Spike, RdRp and Orf3a Genes of SARS-CoV-2 in Influenza Like Illness (ILI) Patients. *Viruses.* 2021:13(1).
23. Spike_ins674QT. Available from: https://mendel.bii.a-star.edu.sg/METHODS/corona/current/MUTATIONS/hCoV-19_Human_2019_WuhanWIV04/geo/Spike_ins674QT_geo.html [Accessed on May 24, 2021]
24. Örd M, Faustova I, Loog M. The sequence at Spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. *Scientific Reports.* 2020 Oct 9;10(1):16944.
25. Public Health England. Investigation of novel SARS-COV-2 variant : Variant of Concern 202012/01; 2020 Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf [Accessed on 17 May, 2021].
26. World Health Organisation. WHO | SARS-CoV-2 Variant – United Kingdom of Great Britain and Northern Ireland. Available from: http://www.who.int/csr/don/21-december-2020-sars-cov2-variant-united-kingdom/en/ [Accessed on 28 May, 2021]