

# Emergence of European and North American mutant variants of SARS-CoV-2 in South-East Asia

Ovinu Kibria Islam<sup>1</sup>  | Hassan M. Al-Emran<sup>1</sup>  | Md. Shazid Hasan<sup>1</sup>  |  
Azraf Anwar<sup>2</sup> | Md. Iqbal Kabir Jahid<sup>1</sup>  | Md. Anwar Hossain<sup>1,3</sup> 

<sup>1</sup>Jashore University of Science and Technology, Jashore, Bangladesh

<sup>2</sup>New York, USA

<sup>3</sup>University of Dhaka, Dhaka, Bangladesh

## Correspondence

M. Anwar Hossain, Jashore University of Science and Technology, Jashore-7408, Bangladesh.

Email: hossaina@du.ac.bd

## Abstract

The SARS-CoV-2 coronavirus is responsible for the current COVID-19 pandemic, with an ongoing toll of over 5 million infections and 333 thousand deaths worldwide within the first 5 months. Insight into the phylodynamics and mutation variants of this virus is vital to understanding the nature of its spread in different climate conditions. The incidence rate of COVID-19 is increasing at an alarming pace within subtropical South-East Asian nations with high temperatures and humidity. To understand this spread, we analysed 444 genome sequences of SARS-CoV-2 available on the GISAID platform from six South-East Asian countries. Multiple sequence alignments and maximum-likelihood phylogenetic analyses were performed to analyse and characterize the non-synonymous (NS) mutant variants circulating in this region. Global mutation distribution analysis showed that the majority of the mutations found in this region are also prevalent in Europe and North America, and the concurrent presence of these mutations at a high frequency in other countries indicates possible transmission routes. Unique spike protein and non-structural protein mutations were observed circulating within confined area of a given country. We divided the circulating viral strains into four major groups and three subgroups on the basis of the most frequent NS mutations. Strains with a unique set of four co-evolving mutations were found to be circulating at a high frequency within India, specifically. Group 2 strains characterized by two co-evolving NS mutants which alter in RdRp (P323L) and spike (S) protein (D614G) were found to be common in Europe and North America. These European and North American variants have rapidly emerged as dominant strains within South-East Asia, increasing from a 0% prevalence in January to an 81% by May 2020. These variants may have an evolutionary advantage over their ancestral types and could present a large threat to South-East Asia for the coming winter.

## KEYWORDS

COVID-19, genome sequence, non-synonymous mutation, phylogeny, SARS-CoV-2

## 1 | INTRODUCTION

The SARS-CoV-2 ssRNA virus was initially detected in December 2019 within China's Hubei province and its associated COVID-19 pandemic is

currently ongoing, with a global toll of over 5 million infections and 333 thousand deaths so far (WHO coronavirus disease [COVID-19] situation report-124, published on 23 May 2020). While China encountered the pandemic first, North American and European nations have also

suffered crippling blows from this pandemic. Moreover, the infection numbers within South-East Asian, South American and African nations are growing. The number of infections in resource-poor, low-income countries may be underestimated due to their limited testing capacity. A number of earlier studies indicated that high temperatures and high humidity could decrease the spread of the virus (Demongeot, Flet-Berliac, & Seligmann, 2020; Sajadi et al., 2020; Wang, Tang, Feng, & Lv, 2020). Despite having such climate conditions, COVID-19 cases are increasing at an alarming rate in relatively hot and humid subtropical South-East Asian countries. As of 22 May, a total of 182,278 cases of SARS-CoV-2-infected cases were identified in South-East Asia, with a death toll of over 5,500 (WHO coronavirus disease [COVID-19] situation report-124, published on 23 May 2020).

Scientists from all over the world are making an unprecedented effort to expose the genomic profiles and characterize the mutation variants of the circulating virus to get insight into its evolutionary patterns and driving force. Tang et al. analysed 103 genomic sequences and indicated that the circulating SARS-CoV-2 strains have two major lineages, one with a synonymous mutation (NSP4\_S75S) and the other with a non-synonymous (NS) mutation (NS8\_L84S; Tang et al., 2020). In another study, Forster, Peter et al. found 3 central variants by analysing 160 sequences and claimed that B-type viruses (with amino acid [aa] substitution, NS8\_L84S) were common in East Asia, whereas A type (ancestral lineage) and C type (NS3\_G251V variant) were prevalent in Europe and North America (Forster, Forster, Renfrew, & Forster, 2020). Changchuan Yin analysed 558 genome sequences and found 15 high-frequency single SNP genotypes (Yin, 2020). He suggested four major groups with either single or co-evolving mutations: group 1 with NSP6\_L37F, group 2 with NS3\_G251V, group 3 with co-evolving mutation at two sites (NSP4\_S75S and NS8\_L84S) and group 4 with co-evolving mutations at four sites (241C>T-leader sequence, NSP3\_F105F, NSP12\_P323L and S\_D614G). In addition, GISAID differentiated COVID-19 into three major clades: Clade S (prevalent in North America), Clade V (prevalent in Asia and Europe) and Clade G (prevalent in Europe), having NS mutations with aa substitutions at NS8\_L84S, NS3\_G251V and S\_D614G, respectively (Fuentes et al., 2020). Later on, they divided the G clade into GR clade and GH clade that contain N\_203-204: RG>KR and NS3\_Q57H aa substitutions, respectively. Finally, based on genomic data available from the GISAID, Nextstrain's SARS-CoV-2 global genomic epidemiology analysis shows 10 major clades for this virus. In this study, we investigated 444 sequenced genomes of SARS-CoV-2 available from South-East Asia to provide an insight into the genomic divergence, phylogeny and recurrent mutations in this region.

## 2 | METHODS AND MATERIALS

### 2.1 | Data collection

In South-East Asia, a total of 444 high-coverage SARS-CoV-2 genomes from India ( $n = 327$ ), Bangladesh ( $n = 20$ ), Indonesia ( $n = 8$ ), Thailand ( $n = 83$ ), Sri Lanka ( $n = 4$ ), Nepal ( $n = 1$ ) and Myanmar ( $n = 1$ )

have been submitted to GISAID ([www.gisaid.org](http://www.gisaid.org)) as of 23 May 2020. This study analysed all 444 genome sequences using hCoV19/Wuhan/WIV04/2019 (Accession: EPI\_ISL\_402124) as a reference genome for phylogenetic and mutation analysis.

### 2.2 | Mutation analyses

Genome sequences of all 444 SARS-CoV-2 were aligned using the MAFFT (multiple alignment using fast Fourier transform) algorithm (Katoh, Misawa, Kuma, & Miyata, 2002), and the alignment was visualized in Jalview 2.11.0. A subset of 60 genome sequences were selected from India ( $n = 30$ ), Bangladesh ( $n = 10$ ), Indonesia ( $n = 8$ ), Thailand ( $n = 7$ ), Sri Lanka ( $n = 4$ ) and Nepal ( $n = 1$ ) for intensive analysis of unique and co-evolving mutations and their phylogeny. The number and ratio of sequence by country were determined based on the available number of sequences and the total number of COVID-19 cases up to 23 May 2020. The selection of sequences analysed was based on genome quality, discreteness of their location and random sample collection dates. Accession ID, collection dates and locations of this subset are provided in Supporting Information S1. Non-synonymous mutations with their specific mutation sites in the 60 genomes were determined using 'CoVsurver enabled by GISAID' in GISAID's EpiCoV database. Additionally, the frequencies of those recurrent NS mutations in all 444 genomes of South-East Asia were determined using Spyder 4.0.1 application under Anaconda Navigator, a free and open-source distribution of the python programming language (Van Rossum & Drake, 2009).

### 2.3 | Spike protein analysis

The S protein sequences were compared with the reference S protein to identify NS mutation sites. The 3D structure of the S protein was constructed using Phyre2 (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015), and the mutation sites were visualized with PyMOL(TM) 2.4.0 software (Schrodinger, 2017).

### 2.4 | Phylogenetic and evolutionary analysis

The aligned sequences were used for the construction of a phylogenetic tree using the maximum-likelihood method in Molecular Evolutionary Genetics Analysis (MEGA X) software (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). Interactive Tree Of Life v5 (Letunic & Bork, 2019) was used to adjust the branch and label colour of the phylogenetic tree.

### 2.5 | Mutation cluster-time plot

A mutation-time plot for South-East Asia was prepared by analysing all 444 sequences available in GISAID. A proportion of mutation-based clusters were plotted monthly from 1 January until 23 May 2020.

Additionally, the numbers of COVID-19 infections and deaths were added in the same time plot collected from WHO situation reports.

## 2.6 | Transmission route analysis

Transmission patterns (performed on 23 May 2020) were observed in South-East Asia using country filters on Nextstrain, an online-based real-time pathogen evolution tracking tool (Hadfield et al., 2018). Nextstrain server filtered 329 sequences from 444 available sequences to create the transmission map.

The frequencies of NS mutations found after in depth analysis of 60 genomes were analysed globally using GISAID-based SARS-CoV-2 mutation statistics server. Countries with a high frequency of those specific mutations were marked on a geographic heat map using the online tool Maptive.

## 3 | RESULTS

Preliminary mutation analysis with 60 of 444 SARS-CoV-2 sequences from South-East Asia revealed 78 NS mutations. Most of them ( $n = 52$ ) were found in non-structural proteins. Other mutations with aa substitutions were present in S protein ( $n = 13$ ), N

protein ( $n = 9$ ), M protein ( $n = 3$ ) and E protein ( $n = 1$ ). The Nepal SARS-CoV-2 genome, which was sequenced early, had no NS mutations compared with our reference sequence (Table 1).

This subset of data identified 21 NS mutations including four aa alterations in S proteins that solely observed in South-East Asia (Table 2). The majority of these unique mutations ( $n = 15$ ) have arisen only once to date. The remaining six were present more than once, but each of these variants circulated in a specific country or region. Moreover, we found 13 mutations with aa replacement in S protein. We also identified the 10 most frequent mutations with aa substitution (N\_R203K, N\_G204R, N\_P13L, NS3\_Q57H, NS3\_Q57H, NS8\_L84S, NSP12\_A97V, NSP12\_P323L, NSP3\_T1198K, NSP6\_L37F and S\_D614G) and separated the variants into four major groups and three subgroups accordingly (Figure 1). Group 1 consists of 11 sequences, including the reference sequence hCoV19/Wuhan/WIV04/2019 (Accession: EPI\_ISL\_402124). All sequences within this group were observed earlier in this year (13 January–1 April) and that do not possess any of those predominant mutations. Group 2 consists of the co-evolving mutations NSP12\_P323L and S\_D614G. Most of the sequences ( $n = 24$ ) belong to this group and their isolation dates range from 10 March to 4 May 2020. We further separated this group into two subgroups: those in 2a have additional N\_R203K and N\_G204R (28881–28883: GGG>AAC) trinucleotide mutations and those in 2b have mutations with aa substitution at NS3 (Q57H).

**TABLE 1** The frequency of SARS-CoV-2 cases identified in South-East Asia region by country and their possessed mutations among selected 60 strains. Case reports were until 23 May 2020

Country	India	Bangladesh	Indonesia	Thailand	Sri Lanka	Nepal	Overall
Selected sequences for detailed study	30	10	8	7	4	1	60
Collection dates	31 Jan–6 May	18 Apr–13 May	17 Mar–14 Apr	22 Jan–3 Apr	4–31 Mar	13 Jan	13 January–13 May
Non-synonymous (N-S) mutations	39	22	12	8	13	0	78
Unique N-S mutations	11	4	2	3	1	0	21
Spike protein N-S mutations	8	2	4	2	1	0	13
E protein N-S mutation	1	0	0	0	0	0	1
M protein N-S mutations	1	0	0	1	1	0	3
N protein N-S mutations	4	6	1	2	3	0	9
NS3 N-S mutations	1	3	1	1	2	0	6
NS7b N-S mutations	0	0	0	1	0	0	1
NS8 N-S mutations	1	2	0	1	0	0	2
NSP2 N-S mutations	8	1	0	1	1	0	10
NSP3 N-S mutations	5	2	1	2	1	0	11
NSP4 N-S mutations	2	0	0	0	0	0	2
NSP5 N-S mutations	0	2	0	0	0	0	2
NSP6 N-S mutations	1	0	1	0	1	0	1
NSP8 N-S mutations	0	1	0	0	0	0	1
NSP12 N-S mutations	3	1	4	1	1	0	6
NSP13 N-S mutations	2	1	0	0	0	0	3
NSP14 N-S mutations	2	1	0	1	1	0	5
NSP15 N-S mutations	0	0	0	1	1	0	2

**TABLE 2** Unique mutations with amino acid substitutions observed solely in South-East Asia until 23 May 2020

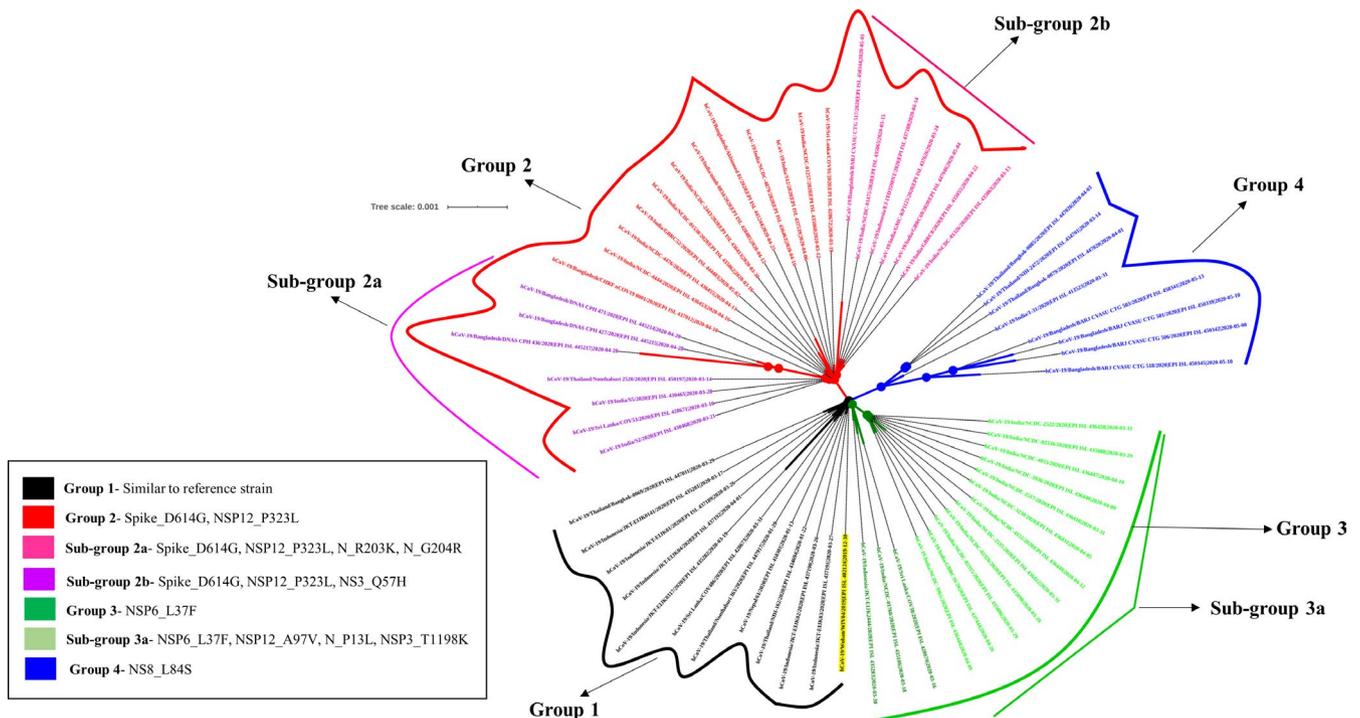
Mutation sites	Country	No. of virus (study)	No. of virus (total)
E_L65M	India	1	1
M_L29J	India	1	1
N_K347N	Indonesia	2	2
NS3_I263M	Sri Lanka	1	1
NSP2_G212D	Thailand	1	1
NSP2_I120F	Bangladesh	4	9
NSP2_L266I	India	1	1
NSP3_L1553J	India	1	1
NSP3_N1337S	Bangladesh	1	1
NSP3_S1485Y	India	1	1
NSP5_D92G	Bangladesh	3	4
NSP12_V880I	India	1	4
NSP13_K40R	India	1	2
NSP13_T214I	India	1	1
NSP14_D415G	Thailand	1	1
NSP14_S434N	India	1	1
NSP14_V459I	Bangladesh	1	1
Spike_A829T	Thailand	3	31
Spike_A930V	India	1	1
Spike_E471Q	India	1	1
Spike_S116C	Indonesia	1	1

Group 3 is characterized by NSP6\_L37F and three other mutations (NSP12\_A97V, N\_P13L, NSP3\_T1198K) co-evolved with this mutation in 12 Indian sequences and form a unique subgroup 3a. Group 4 consists of eight variants from Bangladesh, India and Thailand with a common aa substitution at NS8 (L84S).

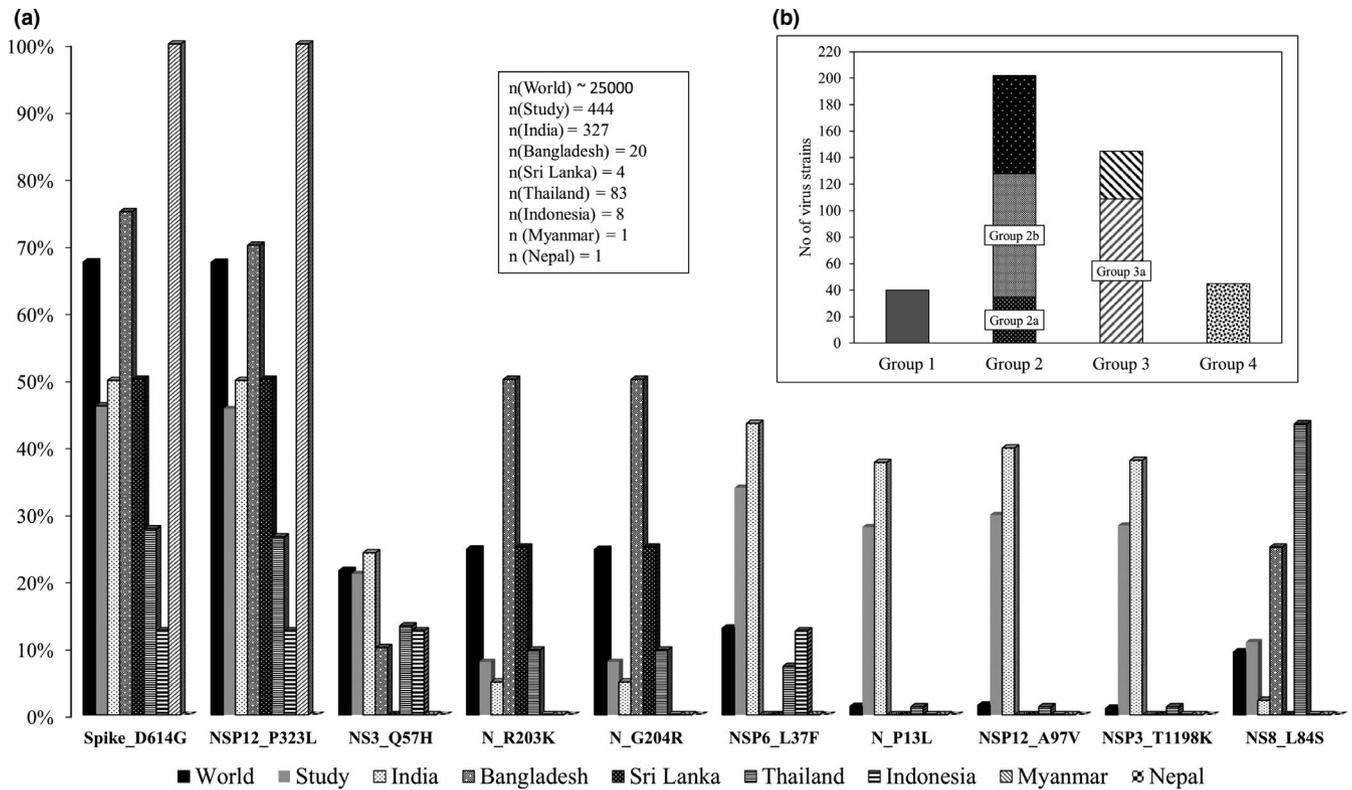
The predominant mutations found from the primary analysis were also observed in 444 genomes and the frequency percentage of those mutations are shown country wise in Figure 2a. Analysis with 444 sequences also shows that 40 sequences belong to group 1, 203 to group 2, 145 to group 3 and 45 to group 4 (Figure 2b). Moreover, subgroups 2a, 2b and 3b contain 35, 93 and 109 genomes, respectively.

S protein analysis of 444 sequences revealed 49 sites for aa substitution (Supporting Information S2). Only four of the aa substitution sites were found in receptor-binding domain (R408I, I434K, E471Q and S494P), two of them resided in critical receptor-binding motif. All the aa substitution sites were marked in 3D structure of S protein (Figure 3). In the cluster-based time plot, 444 available sequences were studied from January until May 2020. The group 2 cluster was not observed until February (0%, n = 0), but was 36% (n = 58) in March, 48% (n = 84) in April and 81% (n = 61) in May. The group 3 cluster was 0% to 7% in those months, with a sudden increase of more than 40% in March and April. The infections increased from 17 cases in January up to 128,257 in May 2020 (Figure 4). During those months, the number of deaths increased up to 3,468 cases by May.

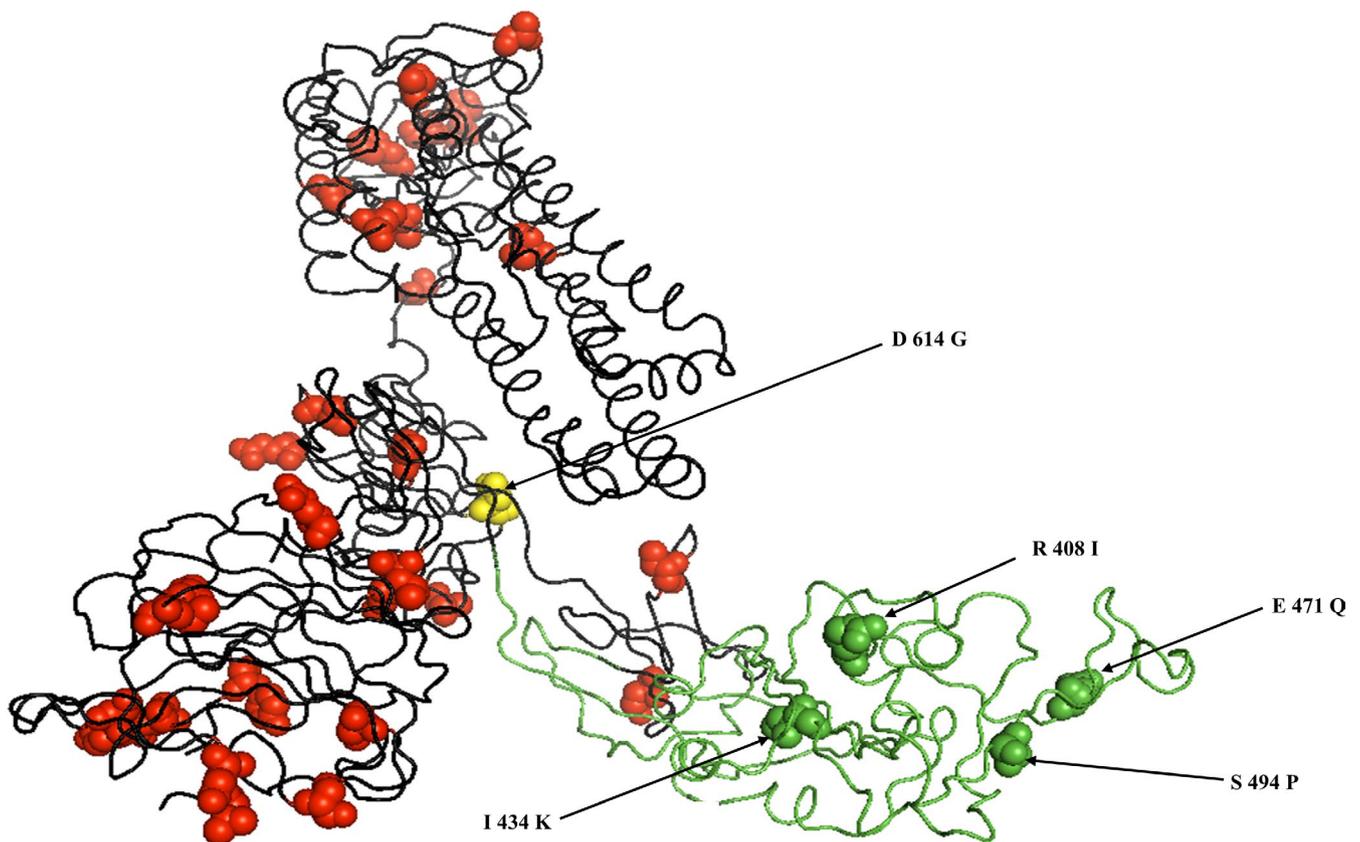
A geographic heat map (Figure 5) revealed that most of the 78 NS mutations found from 60 sequences were also common in Europe and North America (according to global mutation statistics database; Supporting Information S3). The transmission map also revealed that



**FIGURE 1** Phylogenetic relations of selected 60 SARS-CoV-2 strain sequences separated into four clusters



**FIGURE 2** (a) The frequencies of 10 most recurrent mutations in countries of South-East Asia compared with worldwide frequencies in percentage. (b) Number of sequences separated in different group and subgroup



**FIGURE 3** Mutation sites with amino acid substitution in spike proteins found in South-East Asia. Receptor-binding domain (RBD) chain is marked with green. Most abundant mutation D614G is marked with yellow and mutations reside in RBD are marked with green balls

group 2 sequences (A2 clade of Nextstrain) from this study were found to be dominant among viruses circulating in South-East Asia (Figure 6). This clade contains characteristic S D614G mutation also found to be prevalent in Europe and North America (Cao et al., 2020; Gonzalez-Reiche et al., 2020).

## 4 | DISCUSSION

The SARS-CoV-2 pandemic has cost more than 333 thousand lives already, with many more deaths being reported each day. During this period, the global community has yet to predict its virulence, seasonal variation, transmission properties and immunity. However, it is apparent that the fatality rate varies by region and that the degree of virulence varies from person to person (Khafaie & Rahim, 2020; Yao et al., 2020). Some regions in Europe and North America were affected the more than others, while most of Asia, Africa and Australia remain less affected. Analysis of the ssRNA genome is now crucial to understand the pathogenesis and transmission of the virus.

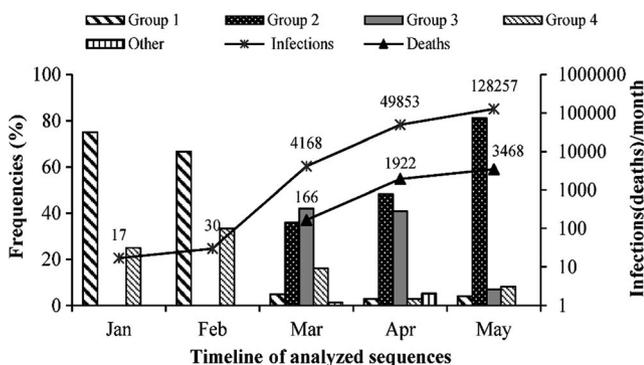
This study has characterized the SARS-CoV-2 virus circulating in South-East Asia into four major groups and three subgroups by studying common NS mutations. Group 1 consists of 40 strains that do not have the recurrent mutations. Those strains are mostly found in January and February in this subcontinent (Figure 2a,b). Group 2 involves 46% ( $n = 203$ ) of the variants in this study. Strains belonging to this group co-evolved with characteristic NS mutation, NSP12\_P323L and S\_D614G. These variants were initially prevalent in Europe and North America, and now constitute 68% of the virus all over the world. A recent study analysed 95 sequences and also found NSP12\_P323L variants to be at a higher frequency, and reported that this variant was mostly found outside of Wuhan, China (Khailany, Safdar, & Ozaslan, 2020). Another study suggests that RNA-dependent RNA-polymerase (RdRp) aa substitution at the 323rd position (NSP12\_P323L) causes RdRp fidelity, which, in turn, increases the number of mutations within the virus and causes co-evolved mutations (Pachetti et al., 2020). NSP12\_P323L

was co-evolved with S\_D614G; this particular non-silent S protein mutation generates an additional elastase cleavage site near the S1-S2 junction and thus facilitates fusion and cell entry (Koyama, Weeraratne, Snowdon, & Parida, 2020). This variant (S\_D614G) was first observed in 28 January 2020 and was initially prevalent in Europe. Within 4 months, this variant has now rapidly outcompeted its ancestral subtype all over the world (Bhattacharyya et al., 2020). These studies explain the frequency of group 2 variants in South-East Asia and why these variants have subdivided into additional subgroups involving co-evolving mutations.

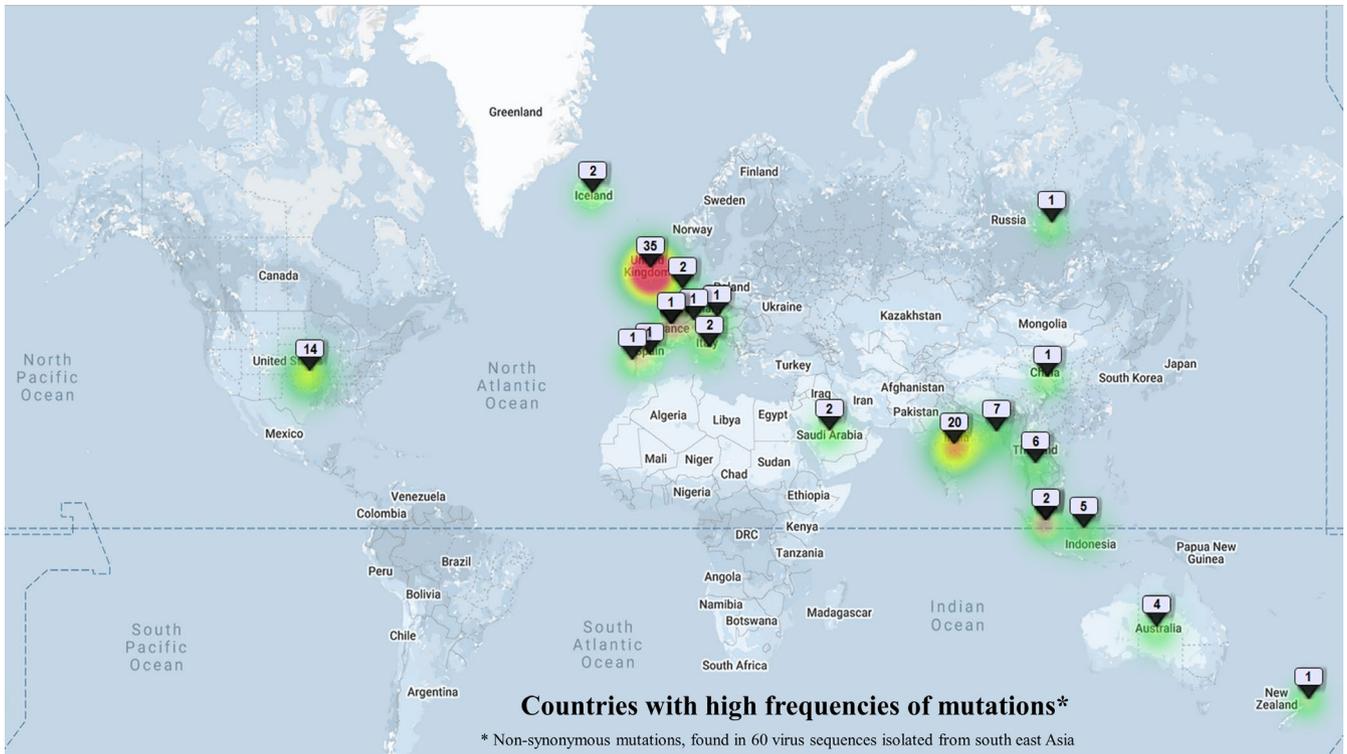
We differentiated group 2 into two subgroups, 2a and 2b, which involve N\_203-204: RG>R and NS3\_Q57H aa substitutions, respectively, along with NSP12\_P323L and S\_D614G. Several studies (Ayub, 2020; Lorusso et al., 2020; Yin, 2020) mention trinucleotide block mutations in nucleotides (28881-28883: GGG>AAC) which resulted in 2 aa changes (N\_203-204: RG>KR) that affect the serine-arginine-rich motif of N protein. This trinucleotide block mutations were found in 35 sequences, 10 of them were from Dhaka, Bangladesh. NS3\_Q57H mutation variants have been commonly found in the USA (Mercatelli & Giorgi, 2020) and Europe and are predicted to be deleterious (Issa, Merhi, Panossian, Salloum, & Tokajian, 2020). Subgroup 2a strains are predominant in Bangladesh (50%) while 2b strains are common in India (25%) (Figure 2a).

Unlike the others, group 3a was unique, with four co-evolving mutations. Of these, the NSP6\_L37F mutation variant was common (Mercatelli & Giorgi, 2020); this mutation variant has also been frequently found in the UK, USA, Australia and India. The other three mutations are relatively less common and found mostly in India and Australia. Group 4, on the other hand, consists of a characteristic NS8\_L84S mutation variant, which was declared as S type by Tang et al. (Tang et al., 2020). This mutation was later reported as C type by another group (Forster, Forster, Renfrew, & Forster, 2020) and was clustered as S clade by GISAID (Fuentes et al., 2020). Group 4 included 45 variants prevalent mostly in Thailand (43%). The analysis of S proteins revealed 49 aa substitution sites. Among them, four sites reside in receptor-binding domain. The most abundant aa substitution D614G was found 203 times in all countries except Nepal. The effect of D614G mutation is described earlier, and the effect of other S protein mutations is yet to be evaluated.

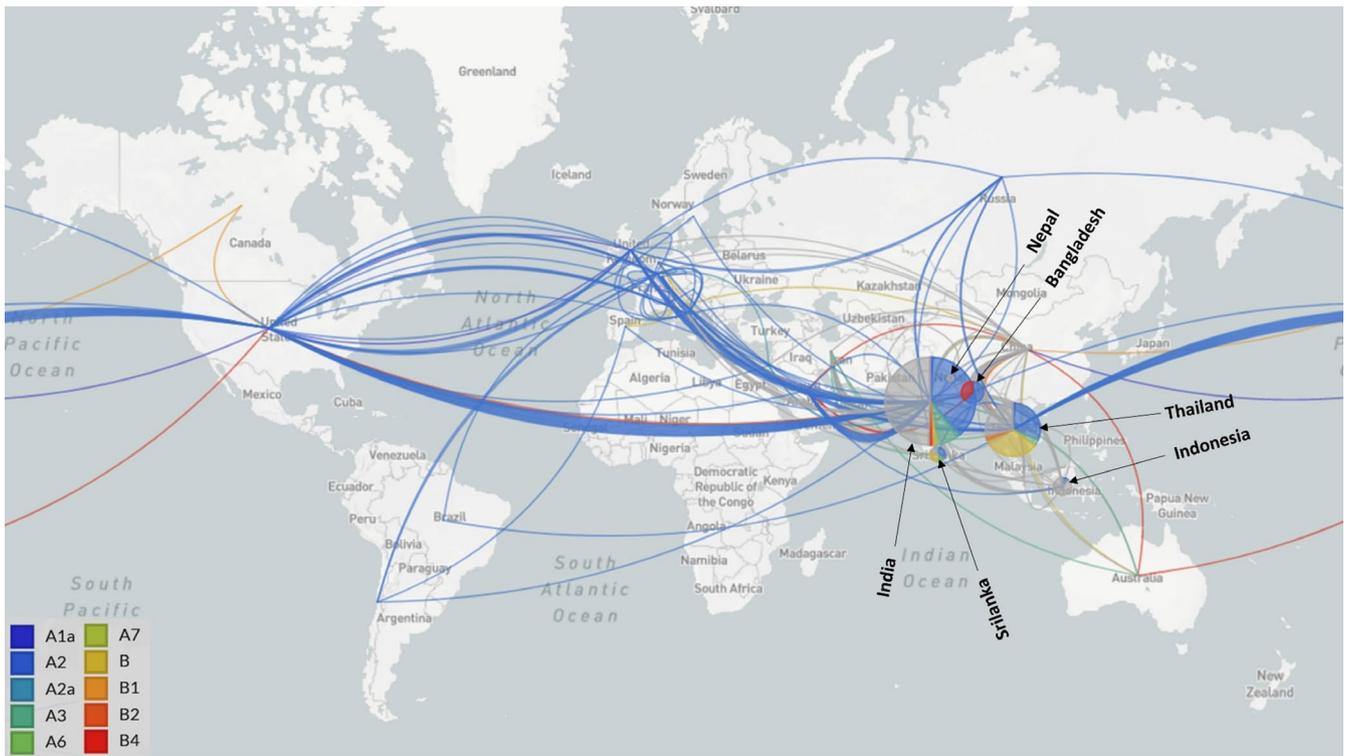
A recent study conducted with 10,014 sequences identified 13 frequent NS mutations (Mercatelli & Giorgi, 2020), while we found only seven of them, along with three less common mutation, at high frequency in this region (Figure 2a). Most of the S protein mutations identified in this region (Supporting Information S2) were also observed in Europe and North America according to the GISAID global mutation database. S protein mutations with aa substitution at 614 position, found in 46% of the studied strains in this region, were also prevalent in Europe and North America. On the contrary, the aa substitutions found at the 1109th position of the S protein found in one Bangladeshi strain was found in another strain from Switzerland. We observed another aa substitution in the S protein at the 76th (S\_T76I) position in an Indonesian strain, which was also found in two strains from West Bengal, India (data



**FIGURE 4** The cluster-based time plot of 444 complete genome sequences available in GISAID from January to May 2020. Bar charts indicate the frequency of mutants identified in the respective months. Line graph indicates the number of COVID-19 infection and death cases



**FIGURE 5** A total of 78 non-synonymous mutations found in 60 countries of South-East Asia. Markers showing the number of these mutations present in relatively high frequency in respective countries (Supporting Information S3)



**FIGURE 6** Transmission map constructed by Nextstrain with 329 sequences showing possible transmission routes into six South-East Asian countries. This map is also separated the sequences of different clades

not shown). This specific aa substitution was identified on 55 occasions according to the global database. Among them, 49 were

from Australia, suggesting that this variant might have been transmitted from Australia.

Additionally, global mutation distribution statistics showed that S\_A829T mutation was observed in 31 sequences, all of them from Thailand (Table 2; Supporting Information S3). NSP2\_I120F mutation was found in nine of the 12 cases from Dhaka, Bangladesh, and NSP2\_D92G mutation was present in four out of the five sequences from Chittagong, Bangladesh (data not shown). These cities are separated by a distance of 250 km, suggesting that those viruses carrying novel mutations were circulating in an area-specific manner.

Non-synonymous mutation and phylogenetic analysis conducted through the Nextstrain database was particularly useful in getting a closer look at mutation variants and their possible routes of transmission. We found a common N\_203-204: RG>KR aa substitution (nine out of 12 strains) in Dhaka, Bangladesh. However, instead of the common N\_203-204: RG>KR aa substitution, a less common aa substitution was observed at the 202nd position of N protein (N\_S202N) among the most (five out of seven) strains of Chittagong. The mutation distribution database showed that strains having trinucleotide block mutation in N protein were prevalent in Europe and that the N\_S202N mutant was found more commonly in recent strains of Saudi Arabia. Phylogenetic analysis by Nextstrain also revealed that the Chittagong strains (belongs to Nextstrain B4 clade and group 4 of our study) have close relationship with the Saudi Arabian strains, while Dhaka strains (A2 clade in Nextstrain, group 2 in this study) are similar to the European ones.

The geographical heat map (Figure 6) of these NS mutations indicate that most of these mutations were also frequently found in the UK, USA, Australia, Saudi Arabia and other European countries, revealing possible transmission routes to South-East Asia. Phylogenetic analysis by Nextstrain produced a similar transmission route map (Figure 5). This study also confirmed, through phylogenetic and mutation analysis, that a high percentage of group 2 strains are linked to European and North American strains (A2 clade in Nextstrain analysis) in India and Bangladesh.

We could not analyse the strains from Maldives, Bhutan and Timor-Leste because they do not have whole genome sequence data of the virus at the time of our analysis. Among the seven countries with available genome sequences, only India, Bangladesh and Indonesia have reported a higher number of SARS-CoV-2 infections. The frequencies of infection have increased exponentially from mid-April, 2020. In our study, it was also shown that group 2 variants containing NSP12\_P323L and S\_D614G mutations were not found earlier than March. The time plot data (Figure 4) delineate that this group 2 cluster is emerging rapidly from 0% in January and February, to 81% in May 2020, suggesting that the European and North American strains are the most recent predominant strains in this subcontinent. A study conducted in early March reported that NSP12\_P323L (14408C>T) and S\_D614G (23403A>G) mutations were recurrent in Europe and had not been detected in Asia until then, supporting our statement (Pachetti et al., 2020). Along with other co-evolving mutations, NSP12\_P323L and S\_D614G probably provide variants with an evolutionary advantage over their ancestral types, allowing them to survive and circulate in this densely

populated region. Time plot is also indicating that the infection and death cases increased along with the increased presence of group 2 strains in this region. Manuel Becerra-Flores and Timothy Cardozo recently reported that prevalence of D614G mutant correlates with higher case fatality rate in different states of the USA (Becerra-Flores & Cardozo, 2020).

In our study, we focused on the NS mutations found in SARS-CoV-2 variants in South-East Asia. More experimental work is required to determine the biological effect of these variants on transmission and pathogenesis of this virus. A number of earlier studies hypothesized that high temperatures and high humidity could result in reduced SARS-CoV-2 transmission. A recent study dealt with combined genomic and climatic approach to provide crucial information about the pathogenesis and spread of the virus (Bajaj & Arya, 2020). However, the seasonality of SARS-CoV-2 has not been established yet and despite of having comparatively hot and humid climate, the infection rate of SARS-CoV-2 is increasing in South-East Asia.

Our study shows that the European and North American mutant variant containing aa substitutions in S protein (D614G) and Rdrp (P323L) have recently become the dominant variant in this region. With the arrival of this variant in this region, the infection and death cases are also increasing. Given that this variant is emerging rapidly and that winter is approaching, the next wave of SARS-CoV-2 may take place in South-East Asia.

## ACKNOWLEDGEMENT

We would like to acknowledge Dr. Mohammad Mahfuzur Rahman, Chairman, Dept. of Climate and Disaster Management, Jashore University of Science and Technology for his valuable suggestion on graphical presentation.

## CONFLICT OF INTEREST

All authors declared that they do not have any conflicts of interests or relationship with this study financially or otherwise.

## ETHICS STATEMENT

The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to. No ethical approval was required as *this study did not collect any samples or questionnaires from animals or humans*.

## DATA AVAILABILITY STATEMENT

Complete and near-complete genome sequences of SARS-CoV-2 are available in the GISAID database. The accession numbers of the genome sequences are available in Supporting Information S1.

## ORCID

Ovinu Kibria Islam  <https://orcid.org/0000-0002-1114-3125>

Hassan M. Al-Emran  <https://orcid.org/0000-0003-1185-6720>

Md. Shazid Hasan  <https://orcid.org/0000-0002-2021-4693>

Md. Iqbal Kabir Jahid  <https://orcid.org/0000-0003-0717-0806>

Md. Anwar Hossain  <https://orcid.org/0000-0001-9777-0332>

## REFERENCES

- Ayub, M. I. (2020). Reporting two SARS-CoV-2 strains based on a unique trinucleotide-bloc mutation and their potential pathogenic difference. Preprints, 2020040337. Published online April, 19. <https://doi.org/10.20944/preprints202004.0337.v1>
- Bajaj, P., & Arya, P. C. (2020). Climatic-niche evolution of SARS CoV-2. *bioRxiv*. <https://doi.org/10.1101/2020.06.18.147074>
- Becerra-Flores, M., & Cardozo, T. (2020). SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *International Journal of Clinical Practice*, 74(8). <https://doi.org/10.1111/ijcp.13525>
- Bhattacharyya, C., Das, C., Ghosh, A., Singh, A. K., Mukherjee, S., Majumder, P. P., ... Biswas, N. K. (2020). Global spread of SARS-CoV-2 subtype with spike protein mutation D614G is shaped by human genomic variations that regulate expression of TMPRSS2 and MX1 genes. *bioRxiv*. <https://doi.org/10.1101/2020.05.04.075911>
- Cao, C., Huang, L., Liu, K., Ma, K., Tian, Y., Qin, Y., ... Wu, P. (2020). Amino acid variation analysis of surface spike glycoprotein at 614 in SARS-CoV-2 strains. *Genes & Diseases*. <https://doi.org/10.1016/j.gendis.2020.05.006>
- Demongeot, J., Flet-Berliac, Y., & Seligmann, H. (2020). Temperature decreases spread parameters of the new Covid-19 case dynamics. *Biology*, 9(5), 94. <https://doi.org/10.3390/biology9050094>
- Forster, P., Forster, L., Renfrew, C., & Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9241–9243. <https://doi.org/10.1073/pnas.2004999117>
- Fuertes, F. D., Caballero, M. I., Monzón, S., Jiménez, P., Varona, S., Cuesta, I., ... Casas, I. (2020). Phylodynamics of SARS-CoV-2 transmission in Spain. *bioRxiv*. <https://doi.org/10.1101/2020.04.20.050039>
- Gonzalez-Reiche, A. S., Hernandez, M. M., Sullivan, M. J., Ciferri, B., Alshammari, H., Obla, A., ... van Bakel, H. (2020). Introductions and early spread of SARS-CoV-2 in the New York City area. *Science*, eabc1917. <https://doi.org/10.1126/science.abc1917>
- Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., ... Neher, R. A. (2018). Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics*, 34(23), 4121–4123. <https://doi.org/10.1093/bioinformatics/bty407>
- Issa, E., Merhi, G., Panossian, B., Salloum, T., & Tokajian, S. T. (2020). SARS-CoV-2 and ORF3a: Non-synonymous mutations and polyproline regions. *bioRxiv*. <https://doi.org/10.1101/2020.03.27.012013>
- Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858. <https://doi.org/10.1038/nprot.2015.053>
- Khafaie, M. A., & Rahim, F. (2020). Cross-country comparison of case fatality rates of COVID-19/SARS-COV-2. *Osong Public Health and Research Perspectives*, 11(2), 74–80. <https://doi.org/10.24171/j.phrp.2020.11.2.03>
- Khailany, R. A., Safdar, M., & Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene Reports*, 19, 100682. <https://doi.org/10.1016/j.genrep.2020.100682>
- Koyama, T., Weeraratne, D., Snowdon, J. L., & Parida, L. (2020). Emergence of drift variants that may affect COVID-19 vaccine development and antibody treatment. *Pathogens*, 9(5), 324. <https://doi.org/10.3390/pathogens9050324>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256–W259. <https://doi.org/10.1093/nar/gkz239>
- Lorusso, A., Calistri, P., Mercante, M. T., Monaco, F., Portanti, O., Marcacci, M., ... D'Alterio, N. (2020). A “One-Health” approach for diagnosis and molecular characterization of SARS-CoV-2 in Italy. *One Health*, 10, 100135. <https://doi.org/10.1016/j.onehlt.2020.100135>
- Mercatelli, D., & Giorgi, F. M. (2020). Geographic and genomic distribution of SARS-CoV-2 mutations. *Frontiers in Microbiology*, 11. <https://doi.org/10.3389/fmicb.2020.01800>
- Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., ... Ippodrino, R. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*, 18, 1–9. <https://doi.org/10.1186/s12967-020-02344-6>
- Sajadi, M. M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F., & Amoroso, A. (2020). Temperature and latitude analysis to predict potential spread and seasonality for COVID-19. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3550308>
- Schrodinger, L. (2017). *The PyMol molecular graphics system, version 2.0*. New York, NY: Schrödinger, LLC.
- Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., ... Lu, J. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, 7(6), 1012–1023. <https://doi.org/10.1093/nsr/nwaa036>
- Van Rossum, G., & Drake, F. L. (2009). PYTHON 2.6 reference manual. Retrieved from <https://docs.python.org/2.0/ref/ref.html>
- Wang, J., Tang, K., Feng, K., & Lv, W. (2020). High temperature and high humidity reduce the transmission of COVID-19. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3551767>
- WHO. (2020). *Corona virus disease (COVID-19) situation report-124*. Retrieved from [https://www.who.int/docs/default-source/coronavirus/situation-reports/20200523-covid-19-sitrep-124.pdf?sfvrsn=9626d639\\_2](https://www.who.int/docs/default-source/coronavirus/situation-reports/20200523-covid-19-sitrep-124.pdf?sfvrsn=9626d639_2)
- Yao, H.-P., Lu, X., Chen, Q., Xu, K., Chen, Y., Cheng, L., ... Jin, C. (2020). Patient-derived mutations impact pathogenicity of SARS-CoV-2. *medRxiv*. <https://doi.org/10.1101/2020.04.14.20060160>
- Yin, C. (2020). Genotyping coronavirus SARS-CoV-2: Methods and implications. *Genomics*. <https://doi.org/10.1016/j.ygeno.2020.04.016>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Islam OK, Al-Emran HM, Hasan MS, Anwar A, Jahid MIK, Hossain MA. Emergence of European and North American mutant variants of SARS-CoV-2 in South-East Asia. *Transbound Emerg Dis*. 2021;68:824–832. <https://doi.org/10.1111/tbed.13748>