Journal of Science and Technology
ISSN: 2456-5660 Volume 7, Issue 02 (MAR-APR 2022)

www.jst.org.in *DOI: https://doi.org/10.46243/jst.2022.v7.i02.pp112-120*

STUDY OF COVID-19 AND PREDICTION OF FUTURE MODEL USING MACHINE LEARNING

Ravindra Nath[1], Nidhi Katiyar[2], Ashish Kumar Saini[3], Astha Agarwal[4]

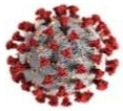
[1]*Department of Computer Science and Engineering Technology CSJM University, Kanpur UP (India)*
[2]*DR. APJ Abdul Kalam Technical University (AKTU), Lucknow*
[3]*Department of Computer Science and Engineering*
[4]*University Institute of Engineering and Technology CSJM University, Kanpur UP (India)*
Corresponding Author: rnkatiyar@gmail.com

To Cite this Article
Ravindra Nath, Nidhi Katiyar, Ashish Kumar Saini, Astha Agarwal, *"***STUDY OF COVID-19 AND PREDICTION OF FUTURE MODEL USING MACHINE LEARNING***", Journal of Science and Technology, Vol. 07, Issue 02, March-April 2022.*

Article Info
Received: 15-02-2022 *Revised: 03-03-2022* *Accepted:16-03-2022* *Published: 21-03-2022*

Abstract: *SARS-CoV-2, a novel coronavirus mostly known as COVID-19 has created a global pandemic. The world is now immobilized by this infectious RNA virus. This RNA virus has the ability to do the mutation in the human body. This study explores the mutation rate of the whole genomic sequence gathered from the patient's dataset of different countries. The collected dataset is processed to determine the nucleotide mutation and codon mutation separately. It has been found that a huge amount of Thymine (T) and Adenine (A) are mutated to other nucleotides for all regions, but codons are not frequently mutating like nucleotides. Using this training and testing process, the nucleotide mutation rate of 400th patient in future time has been predicted. About 0.1% increment in mutation rate is found for mutating of nucleotides from T to C and G, C to G and G to T. While a decrement of 0.1% is seen for mutating of T to A, and A to C. It is found that this model can be used to predict day basis mutation rates if more patient data is available in updated time.*

Keywords: *SARS-Cov-2, Gene sequence, Mutation rate, Neural Network, LSTM mode.*



I. Introduction

The novel human coronavirus disease 2019 (COVID-19) was first reported in Wuhan, China, in 2019, and subsequently spread globally to become the fifth documented pandemic since the 1918 flu pandemic. By September 2021, almost two years after COVID-19 was first identified, there had been more than 200 million confirmed cases and over 4.6 million lives lost to the disease. Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. Most people who fall sick with COVID-19 will experience mild to moderate symptoms and recover without special treatment. However, some will become seriously ill and require medical attention.

The whole world is suffering by an ongoing pandemic due to Coronavirus disease brought by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was an outbreak from Wuhan, the capital of Hubei province in China during

December 2019. The virus was identified on 7th January and observed that it is spread by human-to-human transmission via droplets or direct contact. Its infection has been estimated to be a mean incubation period of 6.4 days and a basic reproduction number of 2.24-3.58. Since its identification, it has already been spread speedily over the whole globe, therefore the world health organization (WHO) had declared COVID-19 a global pandemic on 11th March 2020. The SARS-CoV-2 is a pathogenic human coronavirus under the Betacoronavirus genus[3,4].

In the recent decade, the other two pathogenic species SARS-CoV and MERS-CoV were outbreaks in 2002 and 2012 in China and the Middle East, respectively. The complete genomic sequence (Wuhan-HU1) of this large RNA virus (SARS-CoV-2) was first discovered in the laboratory of China on 10th January and placed in the NCBI GenBank. The SARS-CoV-2 is a single positive-stranded RNA virus having non-segmented in nucleic acid sequence. Although it is an RNA virus but for simplicity of understanding the gene sequence has been given as DNA type which means nucleobase Uracil (U) has been replaced by Thymine (T). The genomic sequence of SARS-CoV-2 virus shows about 79% and 50% similarity with the SAR-CoV and MARS-CoV, respectively. SARS-CoV-2 performs mutation during replication of genomic information. The mutation occurs due to some errors when copying RNA to a new cell. Mutations are mainly three kinds: Base substitution, Insertion, and Deletion. Further, in base substitutions, there are some more divisions: silent, nonsense, missense, and frameshift. Micro-level alteration of mutation rate is also detectable for virus infection in host immune systems and drastically change the virus characteristic and virulence. To understand viral evolution, the mutation rate is one of the crucial parameters[6,7].

Furthermore, it is one of the most important factors for the assessment of the risk of emergent infectious disease, like due to SARS-CoV-2[8,9]. Therefore, an accurate estimation of this parameter finds a great significance. In connection to this and following the current pandemic, many researchers and scientists are working relentlessly to understand the evolution of SARS-CoV-2. Asim et.al has performed Phylogenetic analysis of SARS-CoV-2 virus based on the spike gene of the genomic sequence[17].

II. Material and Methods

In this study, they described a detailed genomic sequence of the SARS-CoV-2 virus. They identified the factor of endemicity of SARS-CoV-2 and then focused on to find out the next reservoir of the SARS-CoV-2 virus. Based on the case study, the authors reported that all sequence of this virus is constituted in a single cluster without making any branching on it but not validated the findings with detailed statistical analysis. An analysis on Gene signature of SARS-CoV-2 virus has been performed by Ranajit and Sudeep[18]. They estimated the ancestry rate of the European genome from the reference population by applying a statistical tool qpAdm. Then they applied Pearson's correlation coefficient between various ancestry rate of European genome and performed statistical analysis on death/recovery ratio by using GraphPad Prism v8.4.0, GraphPad Software. In this study, they developed different linear regression models.

Finally, they performed Genome-wide association analyses (GWAS) among European and East Asian genomes to examine single-nucleotide polymorphism (SNP) which is correlated to the infection of the SARS-CoV-2 virus. From the SNP association, they observed a huge difference in allele frequencies between European and Eastern Asian countries. Debaleena et al. analyzed the statistical changes of signature from different variant of SARS-CoV-2 virus[19]. They calculated diversity, non-synonymous, synonymous, and substitution rates for each gene of the nucleotide sequence by using DnaSP. They also employed time zone software for phylogenetic analysis and mutation detection for each gene. After that, they compared the sequence alignment of a protein of Wuhan and India by using multiple sequence alignment. Note that, in their study, the mutation rate was not calculated based on the patient's genomic sequence. However, the contemporary literature shows adequate studies on the genomic sequence but very few studies on the mutation rate. Therefore, the present study is designed to perform the mutation rate prediction for SARS-CoV-2 on the basis of the time series. Unfortunately, the current data shows that the SARS-CoV-2 virus is highly infectious than the other harmful species of pathogenic human coronaviruses[20].

World populations are now suffering and are in great anxiety by observing the deadliest effect of this virus. But what can be done to stay healthy or avoid getting infected with the virus is still undiscovered. To stop SARS-CoV-2 virus, there is a critical need to invent proper vaccine and antibody based therapy against this virus[21]. Scientists and Researchers are trying their best to discover suitable drugs or vaccines to neutralize the effect of this virus on the human body, or at least in helping to create an effective resistance against the spreading out of this virus. For inventing proper drugs and vaccines against COVID-19 RNA viruses, genomic sequence and mutation analysis are crucially required[22]. In fact, accurate information on the viral mutation rate may play a vital role in the assessment of possible vaccination strategies.

In this regard, we performed a detailed study on the mutation rate of this virus using the available dataset in the NCBI GenBank. From this dataset, we have analyzed the Genomic sequence of 3408 patients from different countries for a period of 12th January to 11th May 2020. We focus specifically on the mutations that have developed freely on different dates (homoplasies) as these are likely possibilities for progressing adjustment of SARS-CoV2 to its novel human host. Specifically, we have calculated the base substitution mutation rates. Due to the lack of necessary information for insertion and deletion, we have considered those as substitution mutations to ensure that no nucleotide goes out of count. It is

expected that the present analysis would help to understand the changing behaviour of this virus in the human body and set up strategies to combat the epidemiological and evolutionary levels[23].

How it Spread?

The virus can spread from an infected person's mouth or nose in small liquid particles when they cough, sneeze, speak, sing or breathe. These particles range from larger respiratory droplets to smaller aerosols.

You can be infected by breathing in the virus if you are near someone who has COVID-19, or by touching a contaminated surface and then your eyes, nose or mouth. The virus spreads more easily indoors and in crowded settings

What is a COVID-19 Variants?

Viruses are always changing, and that can cause a new variant, or strain, of a virus to form. A variant usually doesn't affect how the virus works. But sometimes they make it act in different ways.

Scientists around the world are tracking changes in the virus that causes COVID-19. Their research is helping experts understand whether certain COVID-19 variants spread faster than others, how they might affect your health, and how effective different vaccines might be against them.

How do Variants happen?

Coronaviruses have all their genetic material in something called RNA (ribonucleic acid). RNA has some similarities to DNA, but they aren't the same.

When viruses infect you, they attach to your cells, get inside them, and make copies of their RNA, which helps them spread. If there's a copying mistake, the RNA gets changed. Scientists call those changes mutations.

These changes happen randomly and by accident. It's a normal part of what happens to viruses as they multiply and spread. Because the changes are random, they may make little to no difference in a person's health. Other times, they may cause disease.

Major Variants of COVID-19

As of July 2021, there are four dominant variants of SARS-CoV-2 spreading among global populations:

1. The Alpha Variant (formerly called the UK Variant and officially referred to as B.1.1.7), first found in London and Kent
2. The Beta Variant (formerly called the South Africa Variant and officially referred to as B.1.351)
3. The Gamma Variant (formerly called the Brazil Variant and officially referred to as P.1)
4. The Delta Variant (formerly called the India Variant and officially referred to as B.1.617.2).
5. The currently designated variants are-
6. The Lambda Variant (formerly called the Peru Variant and officially referred to as C.37)
7. The Mu Variant (formerly called the Colombia Variant and officially referred to as B.1.621)
8. Omicron Variant

Variants of Concern (VOC)

A SARS-CoV-2 variant that meets the definition of a VOI (see below) and, through a comparative assessment, has been demonstrated to be associated with one or more of the following changes at a degree of global public health significance:

1. Increase in transmissibility or detrimental change in COVID-19 epidemiology; OR
2. Increase in virulence or change in clinical disease presentation; OR
3. Decrease in effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics.

Dataset Analysis and Pre-processing

An adequate amount of gene dataset is currently available in the NCBI GenBank which contains the complete genome sequence of SARS-CoV-2. Among the many entities, we have filtered the gene sequence, date of collection, and country of the sample. All genes are taken from the human body that are affected by COVID-19. There are genes from almost 33 countries but China, Australia and the United States has a considerable number of patients' data. Though some countries like England, Italy, France, Spain, and Brazil has a very high mortality rate but for the lack of available data in the NCBI GenBank till 15th May, we were unable to calculate the mutation rates for these countries separately. Therefore, we have considered these countries along with others those have low gene data sequence available in the GenBank as the rest of the World category to cover as much region as possible.

III. Result and Discussion

In this dataset, there are also some partial genes. So we filtered them and take only with the level of the complete genome. There is a reference gene sequence of length 29903. Finally, we have filtered the dataset by taking a maximum gene length of 29903 and a minimum of 29161, and avoided the copy sequences. With this filtering process, the total number of patients come down to 3068 from 3408, patients from China come down to 40 from 86, The United States come down to 1903 from 2103 and Australia come down to 918 from 925. Following the size of the available dataset, the mutation rate calculations have been set for four categories: China, the United States of America (USA), Australia and the rest of the World. Furthermore, the dataset is arranged in a suitable way to separately calculate the nucleotide mutation and codon mutation. The first filtered dataset is to find the nucleotide mutation rate. Then we have converted the four raw nucleotides (A = Adenine, T = Thymine, C = Cytosine and G = Guanine) into codon set. A codon consists of three nucleotides and forms a unit of genetic code in DNA or RNA.

Gene Mutation

Gene mutates for many reasons. When RNA tries to copy genetic codes form DNA it may cause some error which causes mutation. Also, errors in DNA replication, recombination, and chemical damage in DNA or RNA cause mutation. There are basically three types of mutations: base substitutions, deletions, and insertions. From this dataset, we can find out the three kinds of substitution mutation which are silent, missense, and nonsense. A silent mutation is the change of codon by which the resulting amino acid remains unchanged. If the resulting amino acid changes then it is called a missense mutation. On the other hand, when changing codon produces the stop signal for gene translation which causes a nonfunctional protein then it is called a nonsense mutation. These three types of substitution mutation of the observed dataset where the missense rate is 34.3%, the nonsense mutation rate is 6.7% and the silent mutation rate is 0.8%.

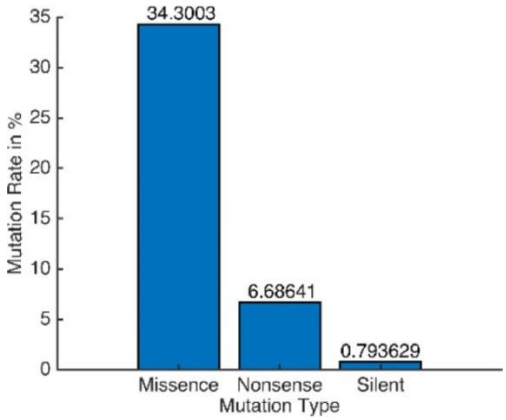


Figure 1 Types of substitution mutation rate.

Nucleotide Mutation

If the mutation type is missense then it can be said that the change of nucleotide has affected the protein generation, which may change the behaviour of the virus. Also, it is hard to identify the cure's gene sequence. The missense nucleotide mutation rate has been calculated by the given algorithm. In this process, we have calculated the nucleotide mutation rate for the prepared dataset. The mutation rate for China has been shown fig. 2. It shows that a huge percent of Thymine (T) are being mutated to other nucleotides but not producing the same amount of T again. Also, a huge amount of Adenine (A) is mutated to other nucleotides. Comparing to T and A, Cytosine (C) and Guanine (G) were not changed much.

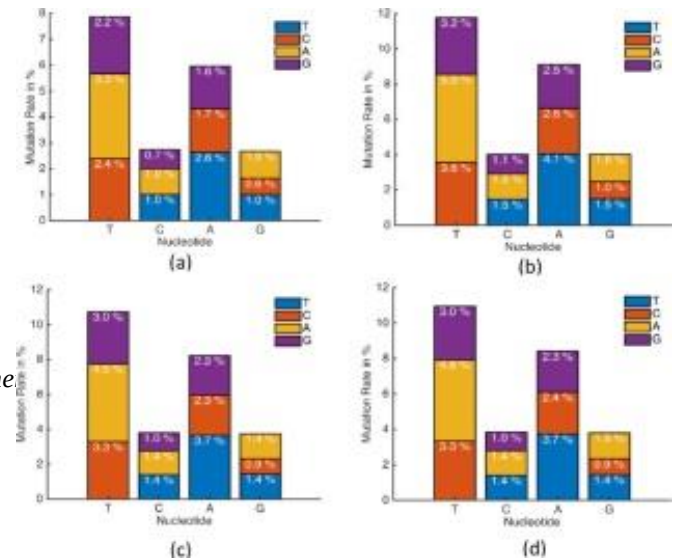


Figure 2 Mutation rate for China of Adenine (A) Thymine (T) Cytosine (C) and Guanine (G).

The main objective of this studies the data of various variants and finds patterns between them. For finding patterns we will use three algorithms:

1. Longest Common Subsequence (LCS)
2. Binary String
3. Rabin-Karp

After finding patterns between different datasets and doing analysis we can predict the future variants and their infectivity.

Longest Common Subsequence (LCS)

The longest common subsequence (LCS) problem is the problem of finding the longest subsequence common to all sequences in a set of sequences (often just two sequences). It differs from the longest common substring problem: unlike substrings, subsequence is not required to occupy consecutive positions within the original sequences. The longest common subsequence problem is a classic computer science problem, the basis of data comparison programs such as the diff utility (data comparison tool that calculates and displays the difference between the text), and has applications in computational linguistics and bioinformatics. It is also widely used by revision control systems such as Git for reconciling multiple changes made to a revision-controlled collection of files.

For example, consider the sequences (ABCD) and (ACBAD). They have 5 length-2 common subsequence: (AB), (AC), (AD), (BD), and (CD); 2 length-3 common subsequence: (ABD) and (ACD); and no longer common subsequence. So (ABD) and (ACD) are their longest common subsequence.

Algorithm

X and Y be two given sequences
Initialize a table LCS of dimension X.length * Y.length
X.label = X
Y.label = Y
LCS [0][] = 0
LCS [][0] = 0
Start from LCS [1][1]
Compare X[i] and Y[j]
If X[i] = Y[j]
LCS[i][j] = 1 + LCS[i-1, j-1]
Point an arrow to LCS[i][j]
Else
LCS[i][j] = max(LCS[i-1][j], LCS[i][j-1])
Point an arrow to max (LCS[i-1][j], LCS[i][j-1])

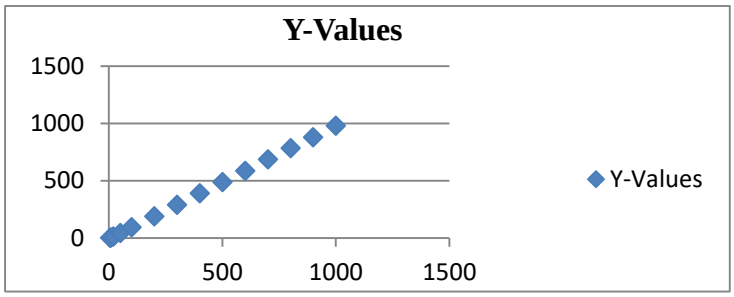


Figure 3 LCS Graph of subsequence data set.

Table 1 Length of sequence A and sequence B with LCS length.

Serial No	Length of sequence A	Length of Sequence B	LCS Length
1	5	5	2
2	10	10	6
3	20	20	15
4	50	50	45
5	100	100	94
6	200	200	190
7	300	300	290
8	400	400	390
9	500	500	489
10	600	600	587
11	700	700	687
12	800	800	784
13	900	900	881
14	1000	1000	981

Binary String Algorithm

This algorithm is used for exact matching of a pattern in a given text the most costly part of any string matching algorithm is to check whether the character of the pattern match the character of the window and it is also easily observed that the probability of Pattern match is much higher than the probability of pattern match over a given text so if you find the occurrence of character mismatch in the Pattern quickly we will get efficiency here we do match in of the character in a window in a match similar way as better as in a binary search. But in this does not necessarily that characters of the text are in sorted order similarly it is also not required for the pattern and to be in sorted order.

In this algorithm firstly the $w_{n)}$ of the window are matched character(p_1 & p_n) of the pattern compared their middle characters the characters the window is the binary searching The First w_2 to w_{mid-1} and second substring further match second &(mid -1)th window with their corresponding occurs, we match middle the window and pattern. In this characters and breaking into the second half of the window at which is very likely, we shift our forward direction (i.e, now first and the last character(w_1 & with the first and the last respectively if they match we after the successful matching of divided into two substring as in substring contains character from contains w $_{mid+1}$ to W_{n-1} then we character of the first substring in characters in pattern if matching characters of the first substring of way we keep on comparing the substring. Similarly we do with any time if any mismatch occurs window by one character in window becomes w_2 to w_{n+1}).

Algorithm

This algorithm can be explain in very simple way Binary string match (t, p, s, s+pl, s+pm)

1. input the text t and length of text n
2. input the pattern p and length of text=m
3. string match up to the length s=0 to n-m
4. if (first char of pattern =first char of text after shifts s)
5.{
a. if (last character of patter character after (s+(n-m+1)) of text.
b.{
i Find middle of pattern and mid2 of text after shifts s equal to length of the pattern
ii. (midi=mid2)
iii.{
1.RecursiveBinar_string match(t,p,s,s+pl+i,s+mid1l-1)
2.RecursiveBinar_string_match(t,p,s,s+mid1+1,s+m-1)
iv. }
c. Else
i. Print (string does not match).
d.}
e. Print (string does not match).
6. }
7. else {Print (string does not match)}

Rabin-Karp Algorithm

Rabin Karp algorithm matches the hash value of the pattern with the hash value of current substring of text, and if the hash values match then only it starts matching individual characters. So Rabin Karp algorithm needs to calculate hash values for following strings.

1) Pattern itself.
2) All the substrings of the text of length m.

Since we need to efficiently calculate hash values for all the substrings of size m of text, we must have a hash function which has the following property. Hash at the next shift must be efficiently computable from the current hash value and next character in text or we can say hash(txt[s+1 .. s+m]) must be efficiently computable from hash(txt[s .. s+m-1]) and txt[s+m] i.e., hash(txt[s+1 .. s+m])= rehash(txt[s+m], hash(txt[s .. s+m-1])) and rehash must be O(1) operation.
The hash function suggested by Rabin and Karp calculates an integer value. The integer value for a string is the numeric value of a string.

For example, if all possible characters are from 1 to 10, the numeric value of "122" will be 122. The number of possible characters is higher than 10 (256 in general) and pattern length can be large. So the numeric values cannot be practically stored as an integer. Therefore, the numeric value is calculated using modular arithmetic to make sure that the hash values can be stored in an integer variable (can fit in memory words). To do rehashing, we need to take off the most significant digit and add the new least significant digit for in hash value. Rehashing is done using the following formula.
hash(txt[s+1 .. s+m]) = (d (hash(txt[s .. s+m-1]) – txt[s]*h) + txt[s + m]) mod q
hash (txt[s .. s+m-1]) : Hash value at shift s.
hash (txt[s+1 .. s+m]) : Hash value at next shift (or shift s+1)
d: Number of characters in the alphabet
q: A prime number
h: d^(m-1)

Algorithm

n = t.length
m = p.length
h = dm-1 mod q
p = 0
t_0 = 0
for i = 1 to m
p = (dp + p[i]) mod q
t_0 = (dt_0 + t[i]) mod q

for s = 0 to n - m
if p = t_s
if p[1.....m] = t[s + 1..... s + m]
print "pattern found at position" s
If s < n-m
$t_s + 1 = (d (ts - t[s + 1]h) + t[s + m + 1]) m_0$

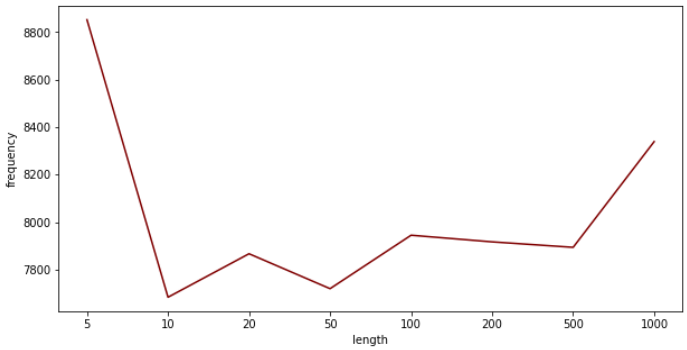


Figure 4 Graph between frequency vs length of strings using Rabin-Karp Algorithm.

IV. Conclusion

The COVID-19 pandemic has almost immobilized the world in this twenty-first century. The great spreading power mixing with mutation turns this virus very difficult and deadly, and the cumulative incidence of COVID-19 is rapidly increasing day-by-day. Lockdown has limited the spreading power of this virus temporarily but the mutation power cannot be controlled till now as no reliable vaccine has invented yet. In this research, we have explained the nucleotide mutation rate and pattern in the codon mutation set. A RNN-based LSTM model has been created to predict the future rate of mutation in person's body if affected with COVID-19. Also, we have explained this LSTM-RNN model for time series prediction based on patients' nucleotide mutation rate, and predicted 400th patient's mutation rate in future time. By analyzing more patient data in updated time, this model can be used to predict day basis mutation rates. The situation may change if a reliable way of cure would be invented. Also in this paper, the mutation rate is limited to base substitution only, insertion and deletion rate can be determined in further research.

References

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7293453
[2] kaggle.com/paultimothymooney/coronavirus-genome-sequence.
[3] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology. 2020;5(4): 536-544
[4] Laboratory testing of 2019 novel coronavirus (2019-nCoV) in suspected human cases: interim guidance, 17 January 2020.
[5] Wang C., Horby P.W., Hayden F.G., Gao G.F. A novel coronavirus outbreak of global health concern. Lancet North Am Ed. 2020;395(10223): 470–473.
[6] Cucinotta D., Vanelli M. WHO declares COVID-19 a pandemic. Acta Bio-medica. 2020;91(1): 157–160.
[7] Huang C., Wang Y., Li X., Ren L., Zhao J., Hu Y., Cheng Z. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet North Am Ed. 2020;395(10223): 497–506
[8] Lu R., Zhao X., Li J., Niu P., Yang B., Wu H., Bi Y. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet North Am Ed. 2020;395 (10224): 565–574
[9] Wu A., Peng Y., Huang B., Ding X., Wang X., Niu P., Sheng J. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe. 2020
[10] Graham R.L., Baric R.S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J Virol. 2010;84(7): 3134–3146
[11] Cui J., Li F., Shi Z.L. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17(3): 181–192
[12] Yang J. Inhibition of SARS-CoV-2 replication by acidizing and RNA lyase-modified carbon nanotubes combined with photodynamic thermal effect. J Explor Res Pharmacol. 2020: 1–6
[13] Fehr A.R., Perlman S. Coronaviruses. Humana Press; New York, NY. Coronaviruses: an overview of their replication and pathogenesis. 2015: 1–23
[14] Pachetti M., Marini B., Benedetti F., Giudici F., Mauro E., Storici P., Zella D. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med. 2020;18(1): 1–9
[15] Lodish H., Berk A., Zipursky S.L., Matsudaira P., Baltimore D., Darnell J. Mutations: types and causes. Mol Cell Biol. 2000; 4
[16] Sanjuán R., Nebot M.R., Chirico N., Mansky L.M., Belshaw R. Viral mutation rates. J Virol. 2010;84(19): 9733–9748

[17] Pfeiffer J.K., Kirkegaard K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog. 2005;1(2)
[18] Vignuzzi M., Stone J.K., Arnold J.J., Cameron C.E., Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature. 2006;439(7074): 344–348
[19] Biswas A., Bhattacharjee U., Chakrabarti A.K., Tewari D.N., Banu H., Dutta S. Emergence of Novel Coronavirus and COVID-19: whether to stay or die out? Crit Rev Microbiol. 2020; 1–12
[20] Das R., Ghate S.D. Investigating the likely association between genetic ancestry and COVID-19 manifestation. medRxiv. 2020
[21] Bhowmik D., Pal S., Lahiri A., Talukdar A., Paul S. Emergence of multiple variants of SARS-CoV-2 with signature structural changes. bioRxiv. 2020
[22] Wrapp D., Wang N., Corbett K.S., Goldsmith J.A., Hsieh C.L., Abiona O., McLellan J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020;367(6483): 1260–1263
[23] Kumar G.V., Jeyanthi V., Ramakrishnan S. A short review on antibody therapy for COVID-19. New Microbes New Infect. 2020
[24] Ojosnegros S., Beerenwinkel N. Models of RNA virus evolution and their roles in vaccine design. Immunome Res. 2010;6(S2): S5