Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computational Biology and Chemistry 93 (2021) 107532

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac

A computational analysis of molecular evolution for virulence genes of zoonotic novel coronavirus (COVID-19)

Priya Kumari, Raju Poddar *

Department of Bioengineering, Birla Institute of Technology-Mesra, Ranchi, JH 835 215, India

ARTICLE INFO

Keywords:
Zoonotic
COVID-19
Adaptive molecular evolution
Hemagglutinin
RdRP
Positive selection
Markov model
Likelihood ratio test

ABSTRACT

Zoonotic Novel coronavirus disease 2019 (COVID-19) is highly pathogenic and transmissible considered as emerging pandemic disease. The virus belongs from a large virus *Coronaviridae* family affect respiratory tract of animal and human likely originated from bat and homology to SARA-CoV and MERS-CoV. The virus consists of single-stranded positive genomic RNA coated by nucleocapsid protein. The rate of mutation in any virulence gene may influence the phenomenon of host radiation. We have studied the molecular evolution of selected virulence genes (HA, N, RdRP and S) of novel COVID-19. We used a site-specific comparison of synonymous (silent) and non-synonymous (amino acid altering) nucleotide substitutions. Maximum Likelihood genealogies based on differential gamma distribution rates were used for the analysis of null and alternate hypothesis. The null hypothesis was found more suitable for the analysis using Likelihood Ratio Test (LRT) method, confirming higher rate of substitution. The analysis revealed that RdRP gene had the fastest rate evolution followed by HA gene. We have also reported the new motifs for different virulence genes, which are further useful to design new detection and diagnosis kit for COVID -19.

1. Introduction

The Novel coronavirus disease 2019 (COVID-19) is a communicable and fatal disease caused by the infection of SARS CoV2 (Severe Acute Respiratory Syndrome Coronavirus-2). The COVID-19 a member of SARS coronavirus lineage, belongs from betacoronavirus genera of huge coronaviridae family (Cui et al., 2019). The virus consists of single-stranded positive genomic RNA coated by nucleocapsid protein. Four virus families i.e. *Arteriviridae, Coronaviridae, Mesoniviridae* and *Roniviridae* are the part of *Nidovirales* order known for largest viral genome (Gorbalenya et al., 2006). It has been ranked among the most complexed

viral genome infecting broad range of host. The subfamily Coronavirinae of *Coronaviridae* encompasses the most significant number of virus including human pathogens (Cong et al., 2017). The species of this subfamily is categorized into four main subgroups namely alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus (Fig. 1). The sequence of novel coronavirus strain belonging from beta-coronavirus revealed similarity with SARS-like BAT coronaviruses bat-SL-CoVZC45 and bat-SL-CoVZXC21 and homology with MERS (Wu and McGoogan, 2020; Rothan and Byrareddy, 2020; Koyama et al., 2020). Illinois COVID-19 investigation team framed different types of exposure risk in consultation with CDC (Ghinai et al., 2020). The comparison of COVID-19 with SARS and MERS showed similarity as well as dissimilarity initiated by zoonotic transmission.

The virus consists of many structural as well as non-structural protein which help as accessory protein playing role in pathogenicity indirectly or directly (refer Fig. 2). The coronaviruses initially infect ciliated bronchial epithelial cells using angiotensin-converting enzyme 2 (ACE2) receptor and type II pneumocytes (Origin and evolution of pathogenic). Transmembrane spike glycoprotein (S) of the virus infect and bind to the host receptor leading to the viral fusion and initiation of viral mechanism for its replication (Tortorici et al., 2019).

The outbreak of novel coronavirus disease 2019 (COVID-19) was first reported in Wuhan city of China in December 2019. It has become potent life-threatening disease to mankind in across the world. The rapidly emerging effect of COVID-19 resulting in declaration of pandemic by WHO (World Health Organization, 2020) with maximum death cases in Italy though it has been initiated from China. According to the WHO report (20th April 2020), there were total 2319066 confirmed cases due to COVID-19 infection and 157970 confirmed death due to the locally transmitted or imported (World Health Organization, 2020).

* Corresponding author.
E-mail address: rpoddar@bitmesra.ac.in (R. Poddar).

https://doi.org/10.1016/j.compbiolchem.2021.107532
Received 5 June 2020; Received in revised form 4 June 2021; Accepted 16 June 2021
Available online 19 June 2021

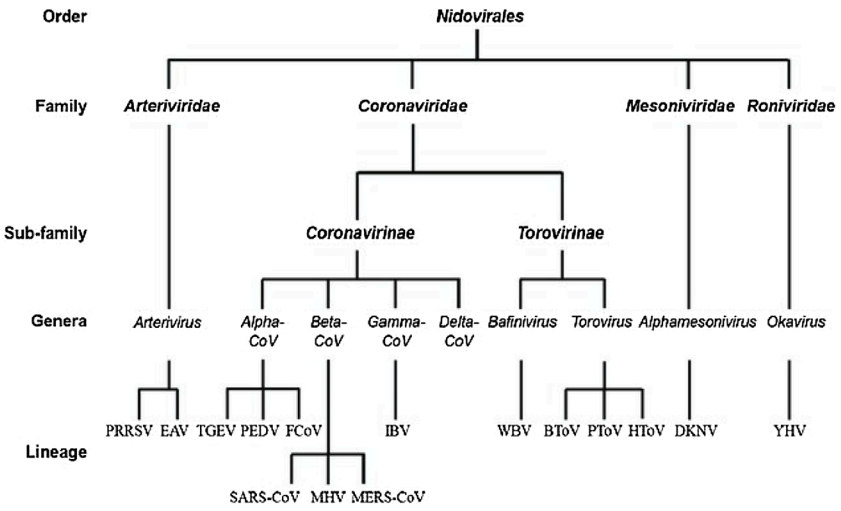


Fig. 1. Classification of order Nidovirales (Cong et al., 2017).

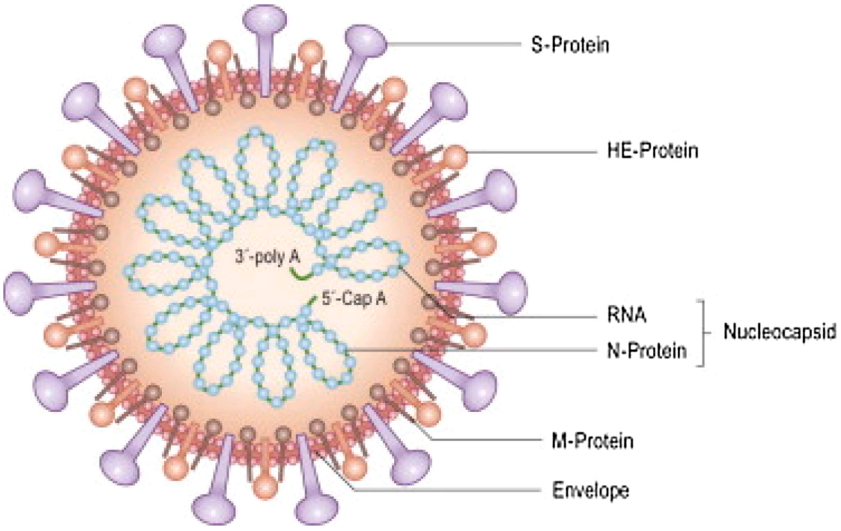
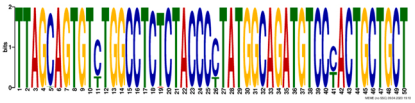
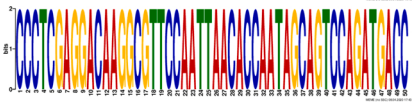
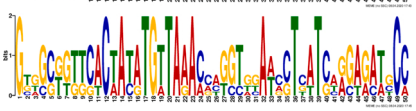


Fig. 2. The schematic representation of cellular organization of SARS CoV2 (https://www.sciencedirect.com/science/article/pii/B9780443073670000409).

Table 1
Motif analysis of virulent genes extracted from SARS-CoV 19.

Gene name	Logo	Consensus sequence (position specific maximum occurrence)	E-value
HA		TTAGCAGTGTCTGGCCTCTCTACCCCTATGGCAGATGTCCYACTGCTGCT	9.5e-2015
N		CCCTCGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTCCAGATGACC	8.6e-484
RdRP		GTGGCGGTTCACTATATGTTAAACCAGGTGGAACCTCATCAGGAGATGCC	9.3e-291
S		ATGTCTATGCAGATTCATTTG	9.1e-053

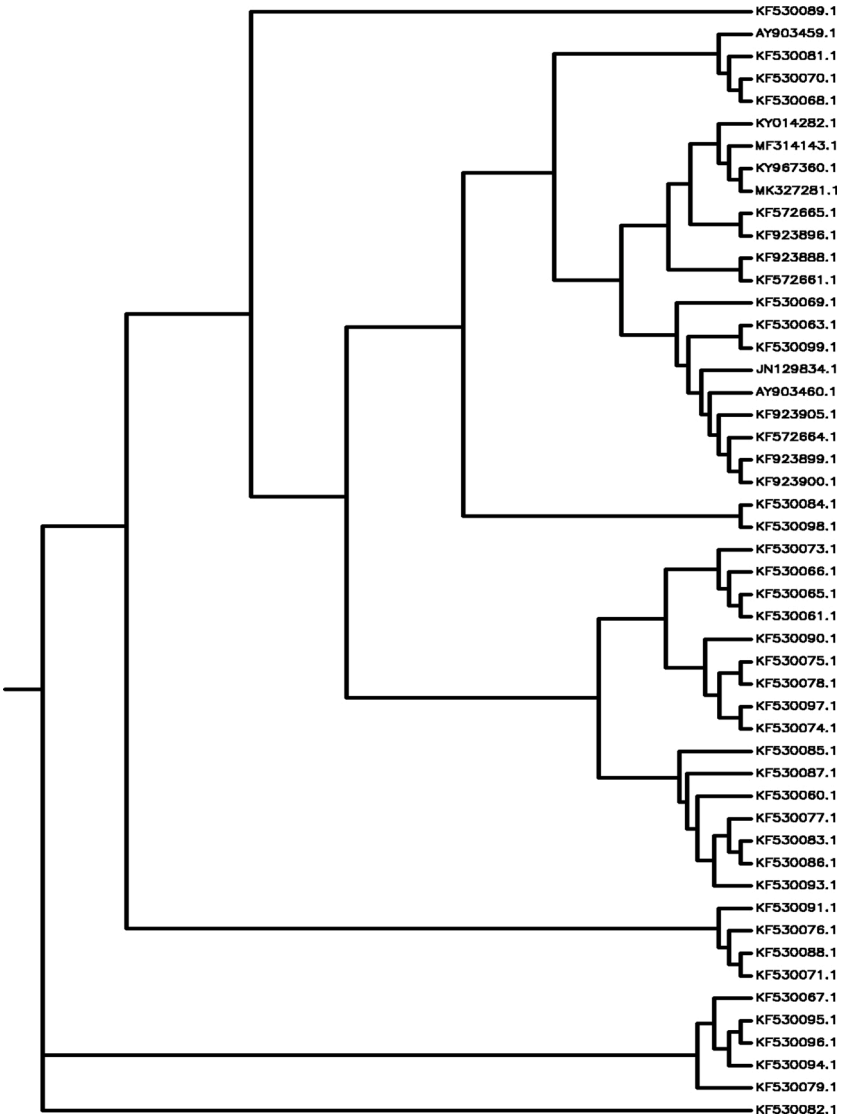


Fig. 3. The genealogy of HA gene sequences used in the study. Each node representing a sequence with its corresponding GenBank accession number.

Coronavirus disease 2019 (COVID-19): situation report, 90). Based on the Centres for Disease Control and Prevention (CDC), the incubation period of this virus is 14 days along with there might be carriers, mild symptom or asymptomatic patients who play role in transmitting (Rothe et al., 2020).

Human coronavirus was first characterized in the year 1960s which are responsible for upper respiratory tract infection. At least 5 new strains of human coronaviruses have been identified after 2003 (Kahn and McIntosh, 2005). Betacoronavirus including SARS-CoV and MERS-CoV are reported as highly pathogenic to humans causing respiratory symptoms. Some of the strain of human betacoronavirus has been identified like human coronavirus 229E and OC43 very earlier. Later, SARS-Cov, HCoV NL63, HKU1, MERS-CoV and SARS-CoV2 were identified in the year 2003, 2004, 2005, 2012 and 2019 respectively (Corman et al., 2018; Kahn and McIntosh, 2005). The novel coronavirus is phylogenetically distinct from all previously known coronavirus indicating evolution towards human host. The recent isolated strain showed 29 nucleotide deletion in ORF8. The viral surface glycoprotein contains higher rates of nonsynonymous mutation reflecting current adaptation to the new host. However, the biological significance of this deletion is not clear so far (Peiris et al., 2003).

COVID-19 has affected approximately all the countries in the whole world either in the form of financially or threatening to mankind. Due to the unavailability of the cure, medicine or vaccine it is extremely important to study about the virus genome to explore and invent new directions for the human welfare. Though research has predicted that chloroquine phosphate, an old drug for the treatment of malaria, might be useful but not clinically proved (Rothe et al., 2020; Gao et al., 2020; Cortegiani et al., 2020). In addition, China International Exchange and Promotive Association for Medical and Health Care (CPAM) recommends ritonavir, lopinavir in combination with nebulized alfa-interferon (dose undefined). Nobody knows the exact mechanism of COVID-19 to get potential treatment so a range of treatments are in trail and regularly reviewed so that hospitals will get a any effective standard treatment quickly for all infected patients. The recovery trail suggested treatments are (https://www.recoverytrial.net/):

- Lopinavir-Ritonavir (commonly used to treat HIV)
- Low dose Dexamethasone (a type of steroid used for mainly inflammation reduction)
- Hydroxychloroquine (anti-malarial drug)
- Azithromycin (antibiotic)
- Tocilizumab (anti-inflammatory treatment given by injection)

The major structural proteins of CoV-2 namely matrix protein (M protein), Nucleocapsid protein (N protein), Envelop protein (E protein)

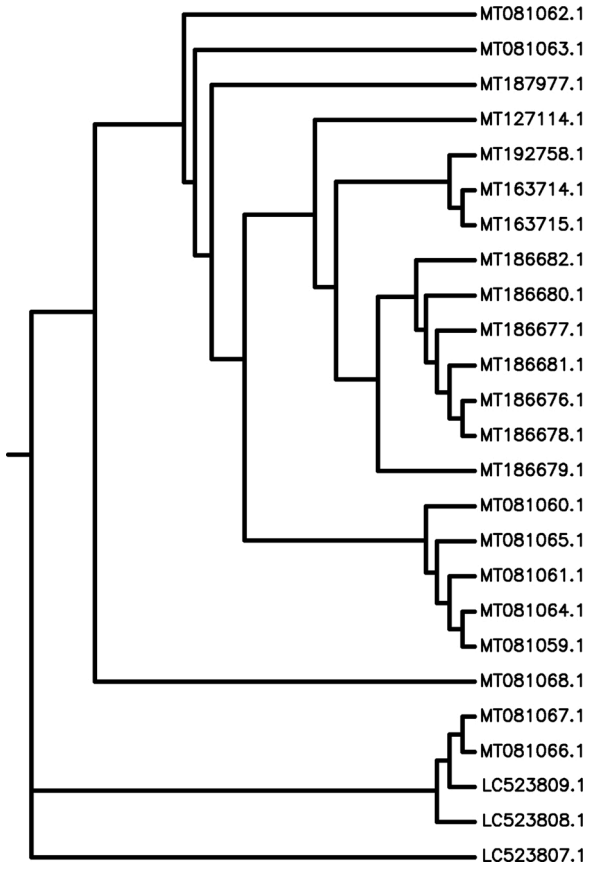


Fig. 4. The genealogy of N gene sequences used in the study. Each node representing a sequence with its corresponding GenBank accession number.

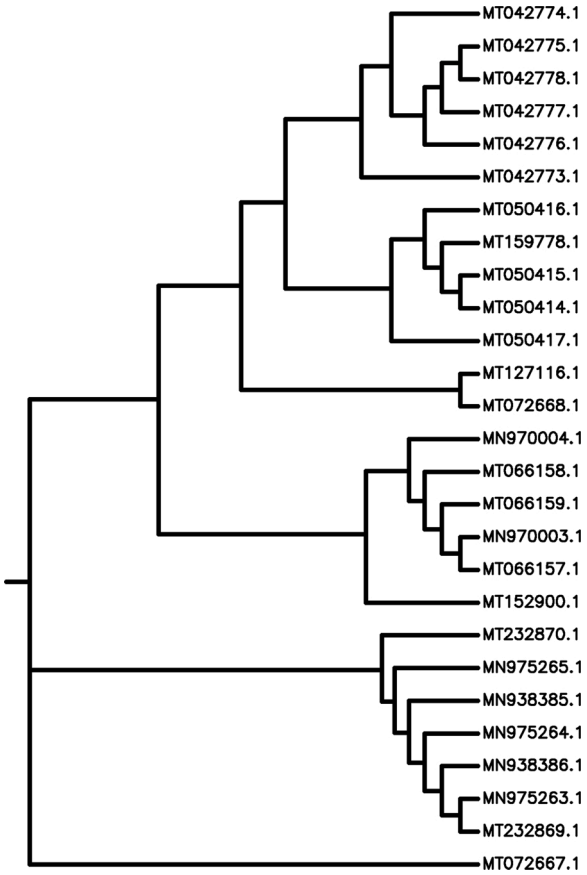


Fig. 5. The genealogy of RdRP gene sequences used in the study. Each node representing a sequence with its corresponding GenBank accession number.

and Spike surface glycoprotein (S protein) involve in the pathogenesis. The matrix protein and nucleoprotein are responsible in the replication and virion formation of the virus in the host (Brian and Baric, 2005) via attaching and fusion with the help of ACE2 (Human cell) to get the entry which is mediated by S glycoprotein (Hasan et al., 2020). The N proteins are highly conserved structural protein and found abundant in C0V which is also involved in the host-pathogen interaction and bind RNA genome to produce virion by forming ribonucleoprotein complex (Lin et al., 2014; Cong et al., 2020; Du et al., 2008). The nucleotide sequence alignment revealed 93 mutations in the SARS-CoV-2 genomes among them, 42 mutations were recognized in all the major proteins except envelop protein. Four, one and twenty-nine missense mutation were found in the nucleocapsid protein, spike glycoprotein and ORF1an polyprotein respectively (Phan, 2020). The identification of alteration in sequences would be significant in terms of potential drug target. All structural proteins of SARS-CoV-2 could be a potential drug targets such as spike protein (Hasan et al., 2020), RNA-dependent RNA polymerase (Elfiky, 2020).

The pathway and functional enrichment analysis of gene/protein set of SARS-CoV-2 was performed using many computational approaches. The protein interactions in enrichment analysis revealed that these proteins are involved in many biological processes (Gollapalli et al., 2020). In some literatures, it has been found that integrins are used as receptors by SARS-CoV-2 which bind with very conserved motif i.e., Arg-Gly-Asp (Sigrist et al., 2020). Alam et al. suggested that a variant cluster of envelope protein with two completely conserved feature i.e., ion-channel and PDZ binding motif shared by SARS and SARS-CoV-2 (Alam et al., 2020). A comparative approach was attempted using spike glycoprotein of SARS-Cov-2, SARS-CoV and MERS-CoV to develop therapeutic and vaccination strategies. Structural quality verification, gene ontology, epitope prediction and motif prediction were performed

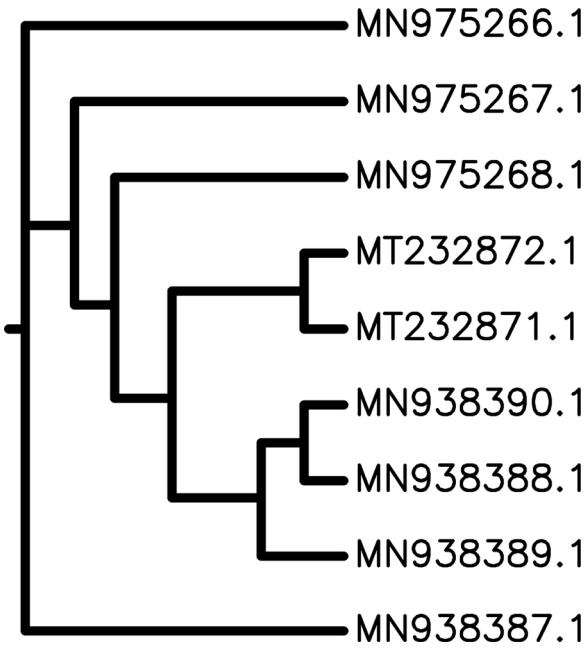


Fig. 6. The genealogy of S gene sequences used in the study. Each node representing a sequence with its corresponding GenBank accession number.

(Banerjee et al., 2020).

The molecular evolution of SARS-Cov-2 and other viruses were studied and revealed that pangolin SARS-CoV are caused due to the

Table 2
Analysis of virulent genes of corona virus using ML method.

Virulent genes	Likelihood value of Hypothesis H_0	Likelihood value of Hypothesis H_1	U (log likelihood ratio)	P value
HA	−2457.66405	−2557.30893	5.3834	0.8761
N	−2624.31513	−2823.77067	4.1465	0.5532
RdRP	−3321.25451	−4056.70751	7.2237	0.8853
S	−211.38457	−211.38458	3.5612	0.5781

*relationship of log likelihood values follows $L_1 \leq L_0$ thus the value of U will always be positive.

Table 3
Estimation of Rates of changes in virulent genes with their probability for both hypotheses.

States	Rates of change H_1 H_0	Probability H_1 H_0
1	0.323 0.264	0.483 0.522
2	1.746 1.413	0.286 0.286
3	4.537 3.596	0.031 0.076
4	9.395 7.086	0.00043 0.0036
5	0.000 12.641	0.200 0.000023

natural selection rather recombination. The analysis predicted that variations in the spike protein (especially receptor binding domain) cause development of new variation. Population genetics predicted that these viruses have two main lineage which can be described by two different SNP that show linkage with nearly all the sequenced virus strain (Tang et al., 2020).

In-silico study will be significantly useful in predicting the hypothesis and ultimately helping the researcher to develop a cure as fast as possible. The fundamental objective of the work is to study how selection acts on variants of genes. The gene were Hemagglutinin (HA), Nucleocapsid phosphoprotein (N), Surface Glycoprotein (S), and RNA dependent RNA polymerase (RdRP) chosen for the analysis. To do so, we use a site-specific comparison of synonymous (silent) and non-synonymous (amino acid altering) nucleotide substitution in humans. The main purpose of the work is to increase understanding of the evolutionary process in corona virus.

2. Methodology

The nucleotide sequences of virulent genes of coronavirus isolated from humans were extracted from the list of SARS-CoV-2 sequences available in GenBank and sequence read archive (SRA) provided by NCBI database (https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/). A total of 528 genome sequences were considered. This short study is based on the Goldman and Yang model in which Markov process predict the substitution between the codons excepting the biasness of transition/transversion rate and codon usage. Further, Investigation of changes in evolution was done by developing a likelihood ratio test based on Markov model of codon substitution for detecting significant rate shifts. The assumption behind this approach is based on functional constraint i.e. functionally important residues and sequences are under stronger selective constraints that lower their evolutionary rates.

The collected sequences were utilized for the discovery of motifs using MEME (Multiple EM for Motif Elicidation) (Bailey et al., 2009) in respective virulent genes which use EM (Expectation Maximization) algorithm. Multiple sequence alignment was performed using ClustalX software v.1.2 (Thompson et al., 1997). Further, Maximum Likelihood (ML) algorithm of PHYLIP package v. 3.698 (Felsenstein, 1993) was used to infer the genealogy of selected genes and also significant branches with their branch length was predicted with the help of drawgram program provided by PHYLIP package.

The Hypothesis was tested by using the Log Likelihood Ratio Test method (Knudsen and Miyamoto, 2001) which can be mathematically expressed as:

$$X = -2log\, L1/L0$$

Where, X is the log likelihood ratio of given model and L_1 and L_0 log likelihood values for the hypothesis H_1 (alternate hypothesis) and H_0 (null hypothesis) respectively. P value for accepting the null hypothesis (H_0) can be calculated by the equation:

$$X \leq (1-\alpha) * 100\%$$

Where, α is the probability of not accepting H_0.

The posterior genealogy distribution was achieved with the help of MrBayes program v.3.2.7 (Ronquist et al., 2012). The estimation was performed with a general time reversible (GTR) model of substitution and a gamma distribution on rate heterogeneity.

3. Results and discussion

Motifs are specific sequence that signifies its specific function mainly use for the characterization and classification of the protein. The analysis of motifs obtained from the MEME suite 5.0.2 using Expectation Maximization algorithm was done. The e-value was low and 'window' ranges ~ 6−50. The predicted motifs and respective details are shown below in tabular form (Table 1). The conserveness of the motifs obtained from the RdRP (RNA dependent RNA Polymerase) is not so strong resulting in fastest rate of evolution and adapting the host significantly. Single Nucleotide Polymorphism may occur in the HA (Haemagglutinin) sequences which can be observed in motifs obtained by HA sequences.

The Maximum Likelihood genealogies estimation were carried out using those virulent sequences of four selected genes. There were two hypotheses assumed in which H_0 was calculated by Gamma distribution plus a class of invariant sites where as H_1 was calculated by Gamma distribution of rates. Results of both hypothesis for the four selected genes are shown in Table 1.

The analysis predicted that the hypothesis H_0 (null hypothesis) could be accepted against the H_1 hypothesis (alternate hypothesis). Gamma distribution plus a class of invariant sites algorithm for H_0 calculation is mainly responsible for the faster rate of evolution in the selected virulent genes. The probability value is higher in the case of RdRP gene followed by HA gene inferring the fastest rate of evolution and to be most significant for molecular adaptation.

The result obtained by ML method was confirmed by generation of rooted tree which is shown in Figs. 3–6 for each selected virulent gene of coronavirus.

The phylogenetic tree in Figs. 3–6 shows distribution of different virulent genes (N, RdRP and S) from COVID 19. It is observed that, the RdRP genes are having distinct branches and most evolving. Diverging branches of RdRP genes (Fig. 6) and HA genes (Fig. 3) represent the divergent evolution of RdRP genes. Further, the estimation of rates of changes was performed using PAML 4: Phylogenetic Analysis by Maximum Likelihood (http://abacus.gene.ucl.ac.uk/software/paml.html) to verify the hypothesis. Table 2 predict the rates of changes in virulent gene with its approximate probability. The wide range of rate of changes for all the selected genes in H_0 hypothesis predict the chance of positive selection is more as compared to that of H_1 model. The results were also confirmed with the simulation studies using Markov Chain Monte Carlo (MCMC) algorithm provided by Mr Bayes (data not shown). Assuming the rates of synonymous and nonsynonymous substitution were identical where all amino acid substitution are either neutral or strongly deleterious (Table 3).

4. Conclusion

The present study reveals about the molecular evolution of the SARS

virus which was earlier isolated from bat. The virus is adapting from bat to human resulting in the process of evolution in fastest rate. We have put our effort towards the study of evolution of virus adapting unique features with due course of time leading to the novelty. The gamma distribution of rates among the sites along with a class of invariant sites parameter used in the study for the crude indication of the rate of the evolution which was further confirmed by rooted tree generation using ML method, motif analysis and estimation of the rate of change confirmed by Markov Chain Monte Carlo (MCMC) algorithm. The most important virulence gene responsible for causing pandemic is RdRP followed by HA gene. Diverging branches of RdRP genes and HA genes represented the divergent evolution of each RdRP genes and HA genes. We believe that our study may prove to be useful to identify candidate genes and codons for the molecular biological investigation of species-specific adaptation in viruses.

Declaration of Competing Interest

The authors have no conflict of interest.

Acknowledgment

Authors are thankful to DBT, Sub-Distributed Information Centre (BTISnet subDIC) government of India and TEQIP (Phase III) for financial support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.compbiolchem.2021.107532.

References

Alam, I., Kamau, A.A., Kulmanov, M., Jaremko, Ł., Arold, S.T., Pain, A., et al., 2020. Functional pangenome analysis shows key features of e protein are preserved in sars and sars-cov-2. Front. Cell. Infect. Microbiol. 10, 405.
Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37 (suppl_2), W202–W208.
Banerjee, A., Santra, D., Maiti, S., 2020. Energetics based epitope screening in SARS CoV-2 (COVID 19) spike glycoprotein by Immuno-informatic analysis aiming to a suitable vaccine development. BioRxiv.
Brian, D.A., Baric, R.S., 2005. Coronavirus genome structure and replication. Coronavirus Replication and Reverse Genetics, pp. 1–30.
Cong, Y., Verlhac, P., Reggiori, F., 2017. The interaction between nidovirales and autophagy components. Viruses 9 (7), 182.
Cong, Y., Ulasli, M., Schepers, H., Mauthe, M., V'kovski, P., Kriegenburg, F., et al., 2020. Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle. J. Virol. 94 (4).
Corman, V.M., Muth, D., Niemeyer, D., Drosten, C., 2018. Hosts and sources of endemic human coronaviruses. In: Advances in Virus Research, Vol. 100. Academic Press, pp. 163–188.
Cortegiani, A., Ingoglia, G., Ippolito, M., Giarratano, A., Einav, S., 2020. A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. J. Crit. Care.
Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17 (3), 181–192.
Du, L., Zhao, G., Lin, Y., Chan, C., He, Y., Jiang, S., et al., 2008. Priming with rAAV encoding RBD of SARS-CoV S protein and boosting with RBD-specific peptides for T cell epitopes elevated humoral and cellular immune responses against SARS-CoV infection. Vaccine 26 (13), 1644–1651.
Elfiky, A.A., 2020. SARS-CoV-2 RNA dependent RNA polymerase (RdRP) targeting: an in silico perspective. J. Biomol. Struct. Dyn. 1–9.
Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package), Version 3.5 c. Joseph Felsenstein.
Gao, J., Tian, Z., Yang, X., 2020. Breakthrough: Chloroquine Phosphate Has Shown Apparent Efficacy in Treatment of COVID-19 Associated Pneumonia in Clinical Studies. Bioscience trends.
Ghinai, I., McPherson, T.D., Hunter, J.C., Kirking, H.L., Christiansen, D., Joshi, K., Fricchione, M.J., 2020. First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. Lancet.
Gollapalli, P., Rimac, H., Patil, P., Nalilu, S.K., Kandagalla, S., Shetty, P., 2020. Pathway enrichment analysis of virus-host interactome and prioritization of novel compounds targeting the spike glycoprotein receptor binding domain–human angiotensin-converting enzyme 2 interface to combat SARS-CoV-2. J. Biomol. Struct. Dyn. 1–14.
Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Nidovirales: evolving the largest RNA virus genome. Virus Res. 117 (1), 17–37.
Hasan, A., Paray, B.A., Hussain, A., Qadir, F.A., Attar, F., Aziz, F.M., et al., 2020. A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. J. Biomol. Struct. Dyn. 1–9.
Kahn, J.S., McIntosh, K., 2005. History and recent advances in coronavirus discovery. Pediatr. Infect. Dis. J. 24 (11), S223–S227.
Knudsen, B., Miyamoto, M.M., 2001. A likelihood ratio test for evolutionary rate shifts and functional divergence among proteins. Proc. Natl. Acad. Sci. 98 (25), 14512–14517.
Koyama, T., Platt, D., Parida, L., 2020. Variant analysis of COVID-19 genomes. Bull. World Health Organ. http://dx. doi. org/10.2471/BLT, 20.
Lin, S.Y., Liu, C.L., Chang, Y.M., Zhao, J., Perlman, S., Hou, M.H., 2014. Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target. J. Med. Chem. 57 (6), 2247–2257.
Peiris, J.S., Yuen, K.Y., Osterhaus, A.D., Stöhr, K., 2003. The severe acute respiratory syndrome. N. Engl. J. Med. 349 (25), 2431–2441.
Phan, T., 2020. Genetic diversity and evolution of SARS-CoV-2. Infect. Genet. Evol. 81, 104260.
Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542.
Rothan, H.A., Byrareddy, S.N., 2020. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J. Autoimmun., 102433
Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Seilmaier, M., 2020. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N. Engl. J. Med. 382 (10), 970–971.
Sigrist, C.J., Bridge, A., Le Mercier, P., 2020. A potential role for integrins in host cell entry by SARS-CoV-2. Antiviral Res. 177, 104759.
Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., et al., 2020. On the origin and continuing evolution of SARS-CoV-2. Sci. Rev. 7 (6), 1012–1023.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25 (24), 4876–4882.
Tortorici, M.A., Walls, A.C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Veesler, D., 2019. Structural basis for human coronavirus attachment to sialic acid receptors. Nat. Struct. Mol. Biol. 26 (6), 481–489.
World Health Organization, 2020. Coronavirus Disease 2019 (COVID-19): Situation Report, 90.
Wu, Z., McGoogan, J.M., 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. Jama.