

Emerging SARS-CoV-2 Genetic Variations and Mutations in the COVID-19 genomic sequence: A Systematic and Meta-analysis review

Reza Valadan

Department of Immunology, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

Reza Alizadeh-Navaei

3- Gastrointestinal Cancer Research Center, Mazandaran University of Medical Sciences, Sari, Iran

Monireh Golpour

4- Molecular and Cell Biology Research Center, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

Mehrdad Gholami

5- Department of Microbiology and Virology, Faculty of Medicine, Mazandaran University of Medical Sciences, Sari, Iran.

Mehryar Zargari

6- Department of biochemistry and genetic, Molecular and cell biology research center, Faculty of Medicine, Mazandaran University of medical sciences, Sari, Iran

Mohammadreza Haghshenas

7- Department of Microbiology, Molecular and Cell-Biology Research Center, Faculty of Medicine, Mazandaran, University of Medical Sciences, Sari, Iran

Tahoora Mousavi (✉ stm.jmums@gmail.com)

8- Molecular and Cell Biology Research Center, Hemoglobinopathy Research Institute, Mazandaran University of Medical Sciences, Sari, Iran

Systematic Review

Keywords: SARS-CoV-2, Genetic Variation, Mutation, COVID-19 sequences

Posted Date: March 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-307547/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Objectives

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of coronavirus disease 2019 (COVID-19). The high mutation rate of RNA viruses causes genetic variation, virus evolution and it is a strategy to escape the immune system. In the present study, all researches and evidence were extracted from the available online national databases. Two researchers randomly evaluated the assessment of the research sensitivity. Finally, after quality assessment and specific inclusion and exclusion criteria, the eligible articles were entered for meta-analysis. The heterogeneity between the results of studies was measured using test statistic (Cochran's Q) and I^2 index. The forest plots illustrated the point and pooled estimates with 95% confidence intervals (crossed lines). All statistical analyses were performed using Comprehensive meta-Analysis V.2 software. This meta-analysis included 13 primary studies investigating the SARS-CoV-2 genetic variations and mutations in the COVID-19 genomic sequence. According to the pooled prevalence (95% confidence interval) of mutations, the spike gene variations showed the highest non-synonymous mutation frequency (16.4%, CI: 13.6, 16.6) and the Non-structural protein (NSP) genes possess the highest mutation frequency among total mutations (31.6%, CI: 21, 44.6). Genomic mutation analysis of SARS-CoV-2 strains may provide knowledge about different biological infrequent mutations and their relationships of viral transmission, pathogenicity, infectivity, and fatality rates between SARS-CoV-2 and human cells.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) the causative agent of coronavirus disease 2019 (COVID-19), poses a foremost challenge to public health. Since the primary appearance of SARS-CoV-2 in late December 2019 in Wuhan, Hubei province, central China, a high dissemination rate has been observed worldwide [1]. Based on information released from the World Health Organization (WHO) on 29 December 2020, the present pandemic COVID-19 has nearly 79 million confirmed cases worldwide and over 1.7 million [2]. The SARS-CoV-2 is classified in the family of Coronaviridae, the order of Nidovirales, and the genus Betacoronavirus [3]. Similar to other coronaviruses, the genome of SARS-CoV-2 consists of specific genes encoding some structural/non-structural proteins [4]. Mutation level among RNA viruses is notably high, which this phenomenon is essential for viral adaptation [5]. Though, coronaviruses have been introduced to have proofreading systems and so, nucleotide sequence variety in SARS-CoV-2 has been observed at a very low level [6]. In a study, Wang et al. reported the presence of 13 variations site in open-reading frames (ORF) of SARS-CoV-2 [in 1a, 1b, S, 3a, M, 8] and N regions, which among them positions nt28144 and nt8782 in ORF 8 and ORF 1a indicated mutation rate of 30.53% and 29.47%, respectively [7]. In addition, based on the evidence obtained from a study on 48,635 SARS-CoV-2 sequences, 353,341 mutations have been detected throughout the world. Among them, D614G mutation in C-terminal of the spike protein (Aspartate to Glycine substitution at position 614) is one such evolutionary alteration detected in the SARS-CoV-2 and has become the most common type reported in many regions of the world such as Europe, Oceania, South America and Africa [8]- [9]. The present study was aims to assess the prevalence of SARS-CoV-2 genetic variation and mutation in COVID-19 sequences.

Genetic diversity and mutations of the COVID-19

There are several reports of unusual public health due to variants of SARS-CoV-2, which changes in transmissibility, clinical features, and severity. Table 1 shows the list of significant mutations in the world.

Table 1: The list of significant mutations in the world.

References	Out comes	Location	Area	Time of mutation	Name of Variant or mutation
[10], [11]	<ul style="list-style-type: none"> - D614G have indicates greater transmissibility in humans rather than greater pathogenicity. -D614G produce higher viral loads. - D614G variant more susceptible to neutralizing antibodies and does not causes serious disease or alter the efficacy of vaccines. 	spike protein	China	early February 2020	D614G
[12], [13]	<ul style="list-style-type: none"> -Increased transmissibility -No change in disease severity. -No evidence that this variant has any impact on the vaccine. - Mutation of N501Y is detected in B.1.1.7. 	RBD	UK and 31 other countries	December 2020	SARS-CoV-2 VOC 202012/01 or B.1.1.7 or 20B/501Y.V1
[14]	<ul style="list-style-type: none"> -Higher viral load -Increased transmissibility - No associated with more severe disease or worse outcomes. - K417N mutation effect on monoclonal and poly clonal antibody. - Mutation of N501Y, E484K and K417N are detected in B.1.351. -E484K makes the vaccine less effective against it. 	RBD	South Africa other countries	18 December, 2020 30 December, 2020	B.1.351 or 501Y.V2
[15], [16]	<ul style="list-style-type: none"> - This variant contain N501Y (More transmission) and E484K (Escape of antibody) and K417N. - Effective on the production of antibody, vaccination or virus neutralization. 	RBD	Tokyo and 3 other countries (Brazilian)	January 2021	B.1.1.248 (P.1) or 501Y.V3
[17]	<ul style="list-style-type: none"> - Escape from immune system. - SARS-CoV-2 Bind to Human ACE2 more strongly than original strain. - N439K escape from polyclonal and neutralizing antibody responses. 	RBD	Wuhan, Europe 12 countries	March 2020 October 2020	N439K
[18]	<ul style="list-style-type: none"> -It is not clear for increasing of the transmissibility of the virus. -Mutation of A222V and A220V are detected in A.EU1. - Spike mutations A222V had a functional effect on spike's ability to mediate cell entry. - Less effective against vaccine. 	spike protein	Spain, UK and 12 countries	June 2020	A.EU1
[19]	<ul style="list-style-type: none"> -Decrease the duration of immune protection following natural infection or vaccination - Cluster 5 variant identified only in 12 human casesand this variant dose not spread widely -Might effect on vaccine development. 	spike protein	Denmark	August and September 2020	Cluster 5

RBD: (Receptor-Binding Domain)

Methods

Search strategy

In the present study, the search strategy was done using available online national databases, including ISI, Science direct, Scopus, Pubmed, Wiley and Google scholar between December 2019 and January 2020. The search was performed based on appropriate keywords of SARS-CoV-2, Variation, Mutation and COVID-19 sequences, which were combined with and/or/not to determine and screen articles in the search strategy. Besides, it is investigated the references of the published studies to improve the sensitivity of the search. The assessment of the research was randomly evaluated by two researchers and confirmed that all suitable studies had been detected.

Study selection

At first, articles of all researches, evidence or reports were extracted from the electronic database. After examinations of studies, duplicate articles were identified and removed from the study. Then, after analyzing the articles, the irrelevant articles were excluded by reviewing of title, abstract, and full text. Also, articles screened for eligibility and review articles and articles published in other languages were extracted from this study.

Quality assessment

The PRISMA checklist was used for evaluation of the quality of the related studies and determination of the selected studies based on title and contents. The PRISMA checklist consists of 27 items covering different aspects of research methodology such as determining Protocol and registration, eligibility criteria, search, study selection, defining variables, method of data collection, risk of bias in individual studies, presentation of results and statistical tests. Each question was required one score.

Inclusion/exclusion criteria

All articles approved by the above assessment phases were considered eligible for final meta-analysis: 1) All English studies. 2) Studies based on the prevalence of SARS-CoV-2 genetic variation among total mutation. 3) Reported prevalence of SARS-CoV-2 genetic variation among non-synonymous mutation. The following studies were ruled out: 1) Duplicated studies. 2) Non-relevant articles. 3) Article with non-full length sequence. 4) Abstracts, letters or review studies. 5) Studies published in languages other than English. 6) Articles with no access to the full-text.

Data extraction

After selection of appropriate articles, the following data for each research were extracted based on first author's name, geographical regions, publication year, language, the number of total mutations, non-synonym mutations, Mutation in S protein, Mutation in N protein, Mutation in M protein, Mutation in E protein, ORF 1a/1b, ORF 3a, ORF 7a, ORF 7b, ORF8a, ORF 10a, ORF6, ORF 1a and NSP. The data were extracted and entered into a Microsoft Excel spreadsheet.

Statistical analysis

The primary outcome was the SARS-CoV-2 genetic variation and mutation in COVID-19 sequences. In our research, the heterogeneity between the results of studies was measured using the test statistic (Cochran's Q) and the I^2 index. P-value less than 0.1 was used to consider significant heterogeneity.

The forest plots illustrated the point and pooled estimates with 95% confidence intervals (crossed lines). Each box in a forest plot indicated the study's weight. The heterogeneity and homogeneity of the suspected factors were performed using random and fixed effects models, respectively and more than 50% were considered as high degrees of heterogeneity. All statistical analyses were performed using Comprehensive Meta Analysis V.2 software.

Results

In the present study, 1370 articles were identified in the starting process. The number of studies was reduced to 1209 following the removal of duplicate articles. In the next step, 890 irrelevant documents were removed after reviewing the full texts. Then, 319 articles were considered for further screening. After the exclusion of 291 articles, 28 articles were assessed for eligibility. 6 articles with non-full length sequence, 8 review articles and one article with other languages were excluded. Finally, 13 relevant articles were included in the Meta-analysis review (Fig.1). In addition, the geographic distribution and frequently mutated residues among COVID-19 sequences are shown in table 2 and fig.2 respectively.

Analysis of mutations among non-synonymous mutation

In the current study, the prevalence of S, N, M, E, ORF 1a/1b, ORF 3a, ORF 7a, ORF 7b, ORF 8a, ORF 10a and ORF 6 mutations among non-synonymous mutation is varied from 0.06% (ORF7b) to 70.44% (ORF 1a/1b) (Table 2, A). Also, it is shown that the highest and lowest frequency of S, N, M, ORF 3a, ORF 7a, ORF 7b and NSP mutations among total mutation belongs to N (0.07%) and NSP (44.14%) respectively (Table 2,B).

In this review 8 sectional studies, S, N, M, E, ORF 1a/1b, ORF 3a, ORF 7a, ORF 7b, ORF 8a, ORF 10a and ORF 6 mutations were assessed among non-synonymous mutation.

1 - Analysis of S mutation

Our analysis revealed that the D614G spike mutation has the highest frequency. This mutation improved spike protein fitness with cell surface receptors and increased the virus's transduction compared to the wild type [20]. Other S mutations, **P1263L, V483A, and L54F**, have a low frequency. The forest plot shows that the overall frequency of S mutation is 16.4% (13.6, 16.6) and with the compounding of the results, the overall prevalence of S mutation with the confidence interval of 95 % and based on random effect model is (I^2 : 85.98%, $Q=49.947$, $P < 0.001$). Also, the results of the heterogeneity studies show that there is heterogeneity among the primary results of the studies (Fig 3, A).

2 - Analysis of N mutation

Other frequent mutations are **R203K and G204R** located in the N-area. N genes encode the nucleocapsid protein that contributes to the formation of helical ribonucleoproteins in the virus [21]. These mutations modify miRNAs' binding mechanism and changed the pathogenesis and development of COVID-19 infection in subjects [22]. Other mutations in region N include **S197L, P13L, L37F, P323L, and P1103L**, which are less frequent, respectively. As can be seen the total prevalence of N, mutations are estimated as 11.7% (7, 19.1). Generally, with the compounding of the results, the overall prevalence of N mutation with the confidence interval of 95 % based on the random effect model is (I^2 : 98.23%, $Q=396.15$, $P < 0.001$). Besides, the results of the heterogeneity studies show that there is a heterogeneity among the initial results of the studies (Fig 3, B).

3 - Analysis of M mutation

The M protein plays a part in the viral envelope packaging by interacting with the S protein [23]. Our analysis revealed two low-frequency **T175M and D3G** mutations in the M gene. Accordingly, analysis of M mutation is calculated 1.9% (0.9, 4.1). The overall prevalence of M mutation with the confidence interval of 95 % based on the random effect model is (I^2 : 84.70%, $Q=45.76$, $P < 0.001$). The results of the heterogeneity studies describe that there is a heterogeneity among the result of these studies (Fig 3, C).

4 - Analysis of ORF1a/1b mutation

ORF1ab is a large gene that coded polyprotein (16 proteins) involved in virus genome synthesis and replication [24]. **P4715L, L3606F, C8517T, A876T and F3071Y** mutations are more frequent in ORF1ab. Due to the overall distribution of ORF 1a/1b mutation 12.8% (5.7, 26.4) with the confidence interval of 95 % based on random effect model is (I^2 : 97.09%, $Q=240.66$, $P < 0.001$) and it is shown that there is a heterogeneity among the results of the studies (Fig 3, D).

5 - Analysis of ORF3a mutation

Q57H, G251V, S193I, and G196V are more frequent mutations in ORF3a. ORF3a proteins are located in host cells and found in the endoplasmic reticulum or Golgi intermediate space, acting as ion channels and controlling the virus's release [25]. moreover, ORF3a

triggers pro-inflammatory pathways and assists in severing modes of infection [26]. It is noteworthy that the ORF3a gene shows a high level of non-synonymous and neutral mutations with a potential effect on B-cells like epitope generation that is a significant point [27]

The incidence of non-synonymous mutation according to ORF 3a group by 95% confidence interval in different studies is shown in the forest plot 5.7% (4.3, 7.6). The results of the analysis demonstrated that the heterogeneity among reported studies is ($P < 0.001$; $I^2 = 78.67\%$, $Q = 32.81$) (Fig 3.E).

6 - Analysis of ORF7a mutation

Test results of forest plot shows that the average rate of ORF 7a is reported to be 2.1% (1.3,3.3) and the overall prevalence of ORF 7a mutation with the confidence interval of 95 % is ($I^2: 60.03\%$, $Q=17.51$, $P = <0.014$) So, there is a heterogeneity among these studies (Fig 3, F).

7 - Analysis of ORF7b mutation

The forest plot shows the prevalence of the non-synonymous mutation based on ORF 7b mutation and confidence intervals (95 % CI). The average frequency of ORF 7b mutation is estimated to be 0.4% (0.1, 1.4). We observed heterogeneity ($I^2: 72\%$, $Q=25$, $P < 0.001$) among these studies (Fig 3, G).

8 - Analysis of ORF8a mutation

In all non-synonymous mutation groups, the average rate of ORF 8a mutation is 1.8% (1.5, 2.1). Based on analysis by 95% confidence interval on fixed-effect model, there is no heterogeneity across these studies ($I^2: 29.82\%$, $Q=9.97$, $P < 0.190$) (Fig 3, H).

9 - Analysis of ORF10a mutation

According to the heterogeneity between the results of the studies, the overall prevalence of ORF 10a mutation 0.5% (0.2, 1) with the confidence interval of 95 % based on random effect model is ($I^2: 50.84\%$, $Q=14.24$, $P < 0.047$) (Fig 3, I).

10 - Analysis of ORF6 mutation

Based on the heterogeneity for ORF 6 mutation ($I^2: 74.18\%$, $Q = 27.11$, $P < 0.001$) using the random effects model, the prevalence of mutation is estimated as 0.7% (95% CI: 0.2, 1.7) (Fig 3, J).

11 - Analysis of E mutation

The heterogeneity indices show the heterogeneity between the primary results of E mutation. Therefore, the random effect model is applied for combining the results ($I^2 = 56.68\%$ $Q = 16.16$, $P < 0.024$). The pooled event rates for mutations of ORF6 is estimated as 0.4% (0.2, 1.1) (Fig 3, K).

Analysis of mutations among total mutation

In the current Meta-analysis, review 5 primary studies. S, N, M, ORF 3a, ORF 7a, ORF 7b and NSP mutations were examined among total mutations.

1 - Analysis of S mutation

Based on the significant heterogeneity observed among the results ($Q=45.6$, $P = 0.000$ and $I^2=91.12\%$), the pooled event rate (95% CI) of developing S mutation using random model was estimated as 18.4% (13.7, 24.4) (Fig 3, L).

2 - Analysis of ORF3a mutation

The forest plot indicated that the overall frequency of ORF3a mutation is 3.9% (2.5, 6) and with the compounding of the results, the overall prevalence of ORF3a mutation with the confidence interval of 95 % and based on random effect model is ($I^2: 60.33\%$,

Q=10.08, P <0.039). Also, the results of the heterogeneity studies show that there is a heterogeneity among the primary results of the studies (Fig 3, M).

3 - Analysis of M mutation

The prevalence of total mutation according to the NSP group by 95% confidence interval in different studies is shown in the forest plot 31. 6% (21, 44.6). The results of the analysis manifest heterogeneity among reported studies (P=0.00; I²=90.47%, Q=42) (Fig 3, N).

4 - Analysis of N mutation

According to the severe heterogeneity, the random effect meta-analysis is performed (P=0.00; I²=96.35%, Q=109.83). The overall mutation of N using the random effect model meta-analysis is 10.5% (95% CI: 5.1, 20.4) (Fig 3, O).

5 - Analysis of M mutation

Heterogeneity indices for primary results for M were not statistically significant (I²: 16.55%, Q=4.79, P <0.309). Therefore, using fixed-effect model, the event rate for M mutation was estimated as 2.1% (95% CI: 1.7, 2.5) (Fig 3, P).

6 - Analysis of ORF7a mutation

More ever there was no significant heterogeneity between the results of primary studies regarding the effect of ORF 7a (I²: 46.89%, Q=7.53, P <0.11). The pooled event rate for ORF7a was estimated at 1.8% (95% CI: 1.5, 2.2) (Fig 3, Q).

7 - Analysis of ORF7b mutation

In this study, it is observed a great heterogeneity between the results of studies regarding the effect of ORF 7b (I²: 58.07%, Q=9.54, P <0.049). Therefore, the random effects model was applied that estimated the pooled event rate for this mutation as 0.4% (95% CI: 0.1, 1.2). (Fig 2, R).

Discussion

Research on the variation in the SARS-CoV-2 genome sequence is necessary for the examination of disease course of COVID-19, disease progression, monitoring, controlling and treatment of SARS- CoV-2 infection [28]. In this present study, the genome sequences of MERS-CoV-2 isolates were examined. The impact of epitope deletion among non-synonymous mutations was the aim of this study which is related to immune escape and pathogenesis. Our study showed that according to the pooled prevalence (95% confidence interval) of mutations, the S variation was shown high frequency 16.4% (13.6, 16.6) among non-synonymous and NSP was the most common mutant among total mutation 31. 6% (21,44.6).

The high mutation of RNA viruses causes genetic variation, virus evolution and it is a strategy to escape the immune system and drug resistance. The SARS-CoV-2 complete genomes with different geographical locations are essential for detecting the genetic variations in the virus that causes viral shedding [29]. Several genome variations in the SARS- CoV- 2, such as nucleocapsid (N) protein, ORF4a and the surface protein S associated with the host immune system [30]. Research indicates that genetic variations of SARS-COV-2 can transmit during the early stage of the epidemic; however, genomes are remarkably stable, and they are not able to evolve rapidly [31].

It is demonstrated that the fatality rate of COVID-19 can vary in different populations, and the level of virulence varies among humans [32]. A larger number of specific mutations with a rapid transmission is detected in Italy, Spain and US and it is related to critical conditions [33]. However, it is demonstrated that genome sequences of SARS-CoV-2 are similar with only a few mutations, but some countries such as North America and Europe are shown the heavily affected regions and Australia, Asia and Africa less affected with sequence variation [34, 35]. Research shows that the variation of RNA viruses is pivotal during an outbreak and it depends on nucleotide substitutions. Based on the viral transmission, the viral mutation rates vary in different viruses and help the virus in host adaptation [36].

Finding non-synonymous mutations through the database is useful for identifying mutations and their modes of transmission [37]. There are some new variants such as (deletion 69-70, deletion 144, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H) are defined in the spike protein of SARS-CoV-2. The novel mutation, (N501Y) which is found in the UK virus variant is located in the receptor-binding domain (RBD). The severity and infectious diseases of the UK variant remain unknown [13]. In SARS-CoV-2 viruses, D614G, is a common mutation spike protein around the world. Also, it is proposed that the highest frequency of spike D614G mutation (S) may be associated with higher viral loads, cellular infectivity, infection severity and lethal outcome in COVID-19 [38]. The relation between high viral loads in the upper respiratory tracts and G clade is measured by RT PCR. It is suggested that the sensitivity of the G variant of SARS-CoV-2 spike to neutralizing antibody is more sensitive than D variant [8]. It is reported that D to G mutation at position 614 (D614G) in the spike glycoprotein which is originated from Europe or China is a significant variation in changes of the secondary structure of protein [39] [8]. D614G mutation started in all affected regions such as Bangladesh (with 95.6% D614G mutation), Italy, Spain, North America and European countries [40] [41]. , amino-acid substitution 1109 (F→L) and 76th (S_T76I) position at spike protein found in Bangladeshi and Indonesian strain respectively [37]. It is also suggested that mutation in RNA-dependent RNA-polymerase (RdRp) and D614G increase SARS-CoV-2 transmission and promote the infectivity of SARS-CoV-2 [42]. The study of 12,300 SARS-CoV-2 genome sequences from different countries reported that D614G and P4715L variation was associated with higher COVID-19 mortality [43].

It is evident that ORF1ab P4715L (nsp 12) plays a pivotal role in viral replication and it is reported that ORF1ab-V378I mutation is associated with COVID-19 infection in Taiwan, Australia and Germany [4] [44]. Also, three mutations, including (M5865V, S5932F) and (R203K) described in ORF1ab and N respectively [45]. It is noticed that mutation in Nucleocapsid (N protein) (R203K and G204R) observed in Italy, Spain, India and France and also N_S202N mutant was detected in Saudi Arabia [46] [47].

Our study shows that substitution in S protein (D614G) is the dominant variant in Asia, Oceania, Europe, North America mutant, Italy, Morocco and Saudi Arabia and led to severe respiratory infections and death in these regions. Genomic mutation analysis of SARS-CoV-2 strains may provide knowledge about different biological infrequent mutations and their relationships of viral transmission, pathogenicity, infectivity, and fatality rates between SARS-CoV-2 and human cells.

Declarations

Acknowledgments

This present study was funded by, Mazandaran University of Medical Sciences, Grant/Award Number: IR.MAZUMS.REC.1399.8702

Conflict of interest

The authors declare there is no conflict of interest.

Authors contributions

Tahoor Mousavi and Reza Valadan conceived and designed the study and wrote the manuscript. Mehrdad Gholami, and Monireh Golpour performed the literature search and collected the data. Mehryar Zargari and Mohammadreza Haghshenas performed the literature review. Reza Alizadeh-Navaei performed the statistical analysis, and all authors read and approved the final version of the manuscript.

ORCID

Tahoor Mousavi [0000-0003-2505-370X](https://orcid.org/0000-0003-2505-370X)

References

1. Pimentel, R.M.M. et al. (2020) The dissemination of COVID-19: an expectant and preventive role in global health. *Journal of Human Growth and Development* 30 (1), 135-140.
2. Organization, W.H. (2020) COVID-19 weekly epidemiological update, 1 December 2020.

3. Adedokun, K.A. et al. (2020) A close look at the biology of SARS-CoV-2, and the potential influence of weather conditions and seasons on COVID-19 case spread. *Infectious Diseases of Poverty* 9 (1), 1-5.
4. Toyoshima, Y. et al. (2020) SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of human genetics* 65 (12), 1075-1082.
5. Pachetti, M. et al. (2020) Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine* 18, 1-9.
6. Kumar, B.K. et al. (2020) Mutational analysis unveils the temporal and spatial distribution of G614 genotype of SARS-CoV-2 in different Indian states and its association with case fatality rate of COVID-19. *bioRxiv*.
7. Wang, C. et al. (2020) The establishment of reference sequence for SARS-CoV-2 and variation analysis. *Journal of medical virology* 92 (6), 667-674.
8. Korber, B. et al. (2020) Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182 (4), 812-827. e19.
9. Biswas, N.K. and Majumder, P.P. (2020) Analysis of RNA sequences of 3636 SARS-CoV-2 collected from 55 countries reveals selective sweep of one virus type. *Indian J. Med. Res.*
10. Fernández, A. (2020) Structural impact of mutation D614G in SARS-CoV-2 spike protein: enhanced infectivity and therapeutic opportunity. *ACS medicinal chemistry letters* 11 (9), 1667-1670.
11. Plante, J.A. et al. (2020) Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 1-6.
12. Bal, A. et al. (2021) Two-step strategy for the identification of SARS-CoV-2 variant of concern 202012/01 and other variants with spike deletion H69–V70, France, August to December 2020. *Eurosurveillance* 26 (3), 2100008.
13. Leung, K. et al. (2021) Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance* 26 (1), 2002106.
14. Cheng, M.H. et al. (2021) Impact of South African 501. V2 Variant on SARS-CoV-2 Spike Infectivity and Neutralization: A Structure-based Computational Assessment. *bioRxiv*.
15. Franceschi, V.B. et al. (2021) Genomic Epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil. *medRxiv*.
16. Voloch, C.M. et al. (2020) Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *medRxiv*.
17. Thomson, E.C. et al. (2020) The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. *bioRxiv*.
18. Hodcroft, E.B. et al. (2020) Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *MedRxiv*.
19. Assessment, R.R. (2020) Detection of new SARS-CoV-2 variants related to mink. *Eur. Cent. Dis. Prev. Control*.
20. Daniloski, Z. et al. (2020) The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *bioRxiv*.
21. Dutta, N.K. et al. (2020) The nucleocapsid protein of SARS–CoV-2: a target for vaccine development. *Journal of virology* 94 (13).
22. Maitra, A. et al. (2020) Mutations in SARS-CoV-2 viral RNA identified in Eastern India: Possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. *Journal of Biosciences* 45 (1).
23. Jamwal, S. et al. (2020) An updated insight into the molecular pathogenesis, secondary complications and potential therapeutics of COVID-19 pandemic. *Life Sciences* 257, 118105.
24. Rehman, S. et al. (2020) Identification of novel mutations in SARS-COV-2 isolates from Turkey. *Archives of Virology* 165 (12), 2937-2944.
25. Hassan, S.S. et al. (2020) Pathogenetic perspective of missense mutations of orf3a protein of sars-cov2. *bioRxiv*.
26. Tozzi, A. SUSD1: A CANDIDATE GENE FOR COVID-19 SEVERITY.
27. Issa, E. et al. (2020) SARS-CoV-2 and ORF3a: Non-Synonymous Mutations and Polyproline Regions. *BioRxiv*.
28. Wang, F. et al. (2020) Initial Whole Genome Sequencing and Analysis of the Host Genetic Contribution to COVID-19 Severity and Susceptibility. *medRxiv*.
29. Gong, Y.N. et al. (2020) SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. *Emerging Microbes and Infections* 9 (1), 1457-1466.

30. AlBalwi, M.A. et al. (2020) Evolving sequence mutations in the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). *J Infect Public Health* 13 (10), 1544-1550.
31. Zhao, J. et al. (2020) Snapshot of the evolution and mutation patterns of SARS-CoV-2. *bioRxiv*.
32. Malki, Z. et al. (2020) Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. *Chaos, Solitons & Fractals* 138, 110137.
33. Laamarti, M. et al. (2020) Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations. *PLoS One* 15 (11), e0240345.
34. Yamamoto, N. and Bauer, G. (2020) Apparent difference in fatalities between Central Europe and East Asia due to SARS-COV-2 and COVID-19: Four hypotheses for possible explanation. *Medical hypotheses* 144, 110160.
35. Lu, R. et al. (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395 (10224), 565-574.
36. Hanada, K. et al. (2004) A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. *Molecular biology and evolution* 21 (6), 1074-1080.
37. Islam, O.K. et al. (2020) Emergence of European and North American mutant variants of SARS-CoV-2 in South-East Asia. *Transbound Emerg Dis*.
38. Kim, J.S. et al. (2020) Genome-Wide Identification and Characterization of Point Mutations in the SARS-CoV-2 Genome. *Osong Public Health Res Perspect* 11 (3), 101-111.
39. Gupta, A.M. et al. (2020) Non-synonymous mutations of SARS-CoV-2 leads epitope loss and segregates its variants. *Microbes Infect*.
40. Hodcroft, E.B. et al. (2020) Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv*.
41. Hasan, M.M. et al. (2020) Global and Local Mutations in Bangladeshi SARS-CoV-2 Genomes. *bioRxiv*.
42. Biswas, S.K. and Mudi, S.R. (2020) RNA-dependent RNA polymerase and spike protein mutant variants of SARS-CoV-2 predominate in severely affected COVID-19 patients.
43. Pandey, U. et al., High Prevalence of SARS-CoV-2 Genetic Variation and D614G Mutation in Pediatric Patients with COVID-19, *Open Forum Infectious Diseases*, 2020.
44. Gong, Y.-N. et al. (2020) SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. *Emerging microbes & infections* 9 (1), 1457-1466.
45. Ippodrino, R. et al. (2020) Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant.
46. Koyama, T. et al. (2020) Variant analysis of SARS-CoV-2 genomes. *Bulletin of the World Health Organization* 98 (7), 495.
47. Islam, O. et al. (2020) Emergence of European and North American mutant variants of SARS-CoV-2 in Southeast Asia.
48. Korber, B. et al. (2020) Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv*.
49. Su, Y.C. et al. (2020) Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. *MBio* 11 (4).
50. Lai, A. et al. (2020) Molecular tracing of SARS-CoV-2 in Italy in the first three months of the epidemic. *Viruses* 12 (8), 798.
51. Zhao, J. et al. (2020) Snapshot of the evolution and mutation patterns of SARS-CoV-2. *bioRxiv*.
52. Poterico, J.A. and Mestanza, O. (2020) Genetic variants and source of introduction of SARS-CoV-2 in South America. *Journal of Medical Virology*.
53. Nguyen, T.T. et al. (2020) Genetic diversity of SARS-CoV-2 and clinical, epidemiological characteristics of COVID-19 patients in Hanoi, Vietnam. *PloS one* 15 (11), e0242537.
54. Laamarti, M. et al. (2020) Genome sequences of six SARS-CoV-2 strains isolated in morocco, obtained using oxford nanopore minion technology. *Microbiology Resource Announcements* 9 (32).
55. AlBalwi, M.A. et al. (2020) Evolving sequence mutations in the Middle East Respiratory Syndrome Coronavirus (MERS-CoV). *Journal of infection and public health* 13 (10), 1544-1550.
56. Gupta, A.M. et al. (2020) Non-synonymous mutations of SARS-CoV-2 leads epitope loss and segregates its variants. *Microbes and infection* 22 (10), 598-607.

57. Kim, J.-S. et al. (2020) Genome-Wide Identification and Characterization of Point Mutations in the SARS-CoV-2 Genome. *Osong Public Health and Research Perspectives* 11 (3), 101.
58. Laha, S. et al. (2020) Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infection, Genetics and Evolution* 85, 104445.
59. Laamarti, M. et al. (2020) Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geodistribution and a rich genetic variations of hotspots mutations. *bioRxiv*.
60. Lai, A. et al. (2020) Molecular tracing of SARS-CoV-2 in Italy in the first three months of the epidemic. *Viruses* 12 (8).
61. Jin, X. et al. (2020) Virus strain from a mild COVID-19 patient in Hangzhou represents a new trend in SARS-CoV-2 evolution potentially related to Furin cleavage site. *Emerging Microbes and Infections* 9 (1), 1474-1488.
62. Wang, R. et al. (2020) Mutations on COVID-19 diagnostic targets. *arXiv preprint arXiv:2005.02188*.

Table 2

Table 2: Geographic distribution of mutant variants of SARS-CoV-2

Name	Location	Mutation Position	Reference	
apore	ORF1ab	C8517T, T17459C (V5820A)	[44]	
	S	T2449C (F817L)		
	ORF3a	C176A (A59D)		
	N	C595T (P199S)		
CGU- (an)	ORF1ab	C8517T, A16577G (K5526R)	[44]	
	S	C145T (H49Y), C2651T (S884F)		
)/Los	ORF1ab	F924F, P4715L	[43]	
	S	D614G		
)/ Asia, i, North t	ORF1ab	(1397 nsp2, 2891 nsp3, 14408 RdRp, 17746 and 17857 nsp143, 18060 nsp14),	[45]	
	S	(23403, spike protein)		
	ORF9a	(28881, nucleocapsid phosphoprotein)		
	ORF3a	(nt 26143)		
	ORF8a	(nt 28144)		
	ORF1ab	P4715L, L3606F	[4]	
	S	D614G		
	N	R203K/G204R, P13L, 203K/204R		
AID	S	D614G, L5F, L8V/W, H49Y, Y145H/del, Q239K, V367F, G476S, V483A, V615I/F, A831V, D839Y/N/E, P1263L,	[48]	
apore	ORF7b		382-nt deletion	[49]
	ORF8			
ralia	ORF7b		138-nt deletion	[49]
	ORF8			
gladesh	ORF8		345-nt deletion	[49]
9/Spain	ORF8		62-nt deletion	[49]
)/Italy	ORF 1ab	(S443F, H3076Y, L3606F, P4715L, E5689D, R5919K)	[50]	
	M	(D3G)		
	ORF 7a	(G70C)		
	S	(A570D, D614G, G1046V)		
	ORF 3a	(G251V)		
N	(R203K-G204R, V246I)			
GISAID atabase	ORF 1ab, ORF 1a, ORF8 S, ORF 3a, N	(nsp12, nsp13, RdRp) (nsp2, nsp6)	[51]	
gladesh	S	I300F (nsp2), P4715L (nsp12), D614G	[41]	
	N	R203K, G204R (N protein)		
n states	2'-O-ribose methyltransferase	N298L	[6]	
	RNA-dependent RNA polymerase	V871I, A88V, P314L		
	Predicted phosphoesterase, papain-like proteinase	P1103L, S1285F, S1197R, A994D, T1198K		

	Transmembrane protein	D279N, L37F, A380V,	
	NSP	G339S, Q496P, S202N	
	3'-to-5' exonuclease	T372I, L177F	
	ORF3a	L46F, Q57H	
	ORF8	L84S	
	S	L54F, D614G	
	NP	P13L, S194L, RG203KR	
South American	ORF 1b	D614G, E1207V	[52]
	ORF 1a	G392D, T708I, I739V, P765S, A876T, A1043V, N2894D, F3071Y, G3334S, L3606F Q57H, G196V, G251V	
	ORF 3a	T175M	
	M	L84S	
	ORF 8	D103Y, R191C, S197L, R203K, G204R, G238C	
e	3' UTR	G204R	[42]
	5' UTR	-	
	N	S194L, R203K, S202N	
	ORF 8	L84S	
	M	-	
	ORF 3	Q57H	
	S	D614G, A879S	
ORF1ab	A1812D (nsp3), L3606F (nsp6), P4715L (RdRp)		
Northern Vietnam	S	L54F, S254F, C1250F, D614G	[53]
	ORF 3a	Q57H, G251V	
	M	D3G, V70F	
	ORF 7a	S81L, L96F, L102_I103del	
	ORF 7b	R203K; G204R, S180I, A211V, Q283H	
N	Gly82_Val86del, Met85del (nsp1), T85I, G212D, 559V, P585S (nsp2) A58T, T428I, R646W, L672F, G730D, P1103L, K1186R, M1901I (nsp3), D477N (nsp4), G15S (nsp5), L37F (nsp6), D161V, P323L, V338F (nsp12) R595S(nsp13) V320L(nsp15) P134S, T140I (nsp16)		
South-East Asia	S	D614G	[47]
	N	R203K, G204R, P13L, Q57H, NS8_L84S, L37F, P323L	
	NSP	A97V, T1198K	
Morocco	S	D614G	[54]
Saudi Arabia	S	P97 L, T424I, C1313S, W553R, S950T, R700L, S191 P, S459T, V26 L, Q1009L	[55]

	ORF	S733R/E736 K, F1609 L, P1883S, M4574I, V51I, T649I, P777S, A1045 V, V1202I, E1835A, M2119I, P2742S, G3117E/V3120D, F5011I, R5027S/D5028Y/R5029 K, G5061R, H243C, H73N, 167 G, L3785F, A5362 L, V117D, D2204A, V5551A, E102C, T5560S, V6030 F, T636I, G981S, V1375I, A1462 T, F3098 L/T3101N, F5072 L, I5559 M/T5560 F/G5561W/L5562 V/Y5563 V, I6668 M, V6431F, G85D/P86 F/T87N, K1255R, H1714Y, N4385 K, I4611 V, A5764S, M6272I, L6373R, L6958 P, V62F, G85D/P86F, A218S, T586S, P1099R, E1428 G, A1301 V, P1971S, S2111R, K5019 N, L6412 F, P86L, I147L	
	M	T127I	
	N	W293C, A300V, V178A, S11 F, L283 F, G28 V, D242E, V263A, P7L, G198S, R292 P, S391I,	
GISAID atabase	ORF 1ab	P4715L, Y232C, F1657L, A1906V, V1973L, G2374R	[56]
	ORF 1a	D58E, L952P, E955K, S1498F, N1559T, A3203V, G4227R, A4297G, F4304L	
	S	Y145del, N354D, D364Y, R416I, S438F, Y508H, D614G ^C , Q675H, T791I, F797C, A930V, I1216T, P1263L	
	ORF 3a	A31T, Q57H ^{GH} , V88L, H93Y, G196V, G251V ^V	
	M	D3G, T175M	
	ORF7a	V74F, S81L	
	ORF8	V62L, L84S ^S	
	N	L121H, T148I, S193I, S197L, R203K ^{GR} , G204R ^{GR} , I292T	
GISAID atabase	S	D614G	[57]
	ORF 1b	P214L	
	ORF 3a	G251V, Q57H	
	ORF8	L84S	
	N	R203K, G204R	
	ORF 1a, M, ORF6, ORF7a, ORF7b, ORF10		
NCBI atabase	ORF 1ab	D75E, T265I, P971L, L3606F, P4715L, V5550L, P5828L, Y5865C, F6158L	[58]
	S	D614G	
	ORF 3a	Q57H, G251V	
	ORF 8	S24L, V62L, L84S	
	N	R203K, G204R	
GISAID atabase	ORF 1ab	M4555T, T4847I, T5020I, V5661A, P5703L, M5865V, G3278S, K3353R, I6525T, Ter6668W, A876T, T1246I, S5932F, F3071Y,	[59]
	S	V483A	
	M	D3G, T175M	
	ORF 3	S197L, S202N, R203K, G204R	
	N	S193I, S194L, S197L, S202N, R203K, G204R	
	ORF 8		

Table 2(A): Frequency of mutations among non-synonymous mutation included in Meta-analysis

Reference	First author	Language	Area of study	non-synonymous sample size	S%	N%	M%	ORF 1a/1b %	ORF 3a%	ORF 7a%	ORF 7b%	ORF 8a%	ORF 10a%	ORF6%	E%
	Gupta (2020)	English	GISAID	47	27.65	14.89	4.25	12.76	12.76	4.25	NA	4.25	NA	NA	NA
	Alessia (2020)	English	Italy	159	11.94	3.77	2.51	70.44	7.54	1.88	0.62	0.62	0.62	NA	NA
	Kumar (2020)	English	Indian	4648	19.75	19.16	NA	NA	6.92	NA	NA	1.48	NA	NA	NA
	Kim (2020)	English	GISAID	1352	13.46	8.8	1.55	NA	5.76	2.58	0.59	2.36	0.81	1.40	0.88
	Hasan (2020)	English	Bangladesh	1602	15.60	36.14	0.811	39.51	3.30	2.18	0.06	1.87	0.18	0.18	0.12
	Jin (2020)	English	Zhejiang	37	27.02	NA	13.51	NA	NA	NA	NA	NA	NA	NA	NA
	Laha (2020)	English	NCBI	351	12.53	7.40	1.42	66.95	5.69	1.7	0	2.27	0.56	0.854	0.56
	Islam (2020)	English	South-East Asia	78	16.66	11.53	3.84	NA	NA	NA	NA	NA	NA	NA	1.2

Table 2(B): Frequency of mutations among total mutation included in Meta-analysis

Reference	First author	Language	Area of study	Total mutation sample size	S%	N%	M%	ORF 3a%	ORF 7a%	ORF 7b%	NSP %
	Biswas (2020)	English	GISAID	504	16.26	7.14	1.78	3.96	N/A	N/A	N/A
	Wang (2020)	English	GISAID	4796	1.2	0.07	2.18	4.81	1.83	0.2	44.14
	Nguyen (2020)	English	GISAID	167	26.34	23.95	1.19	N/A	2.39	1.19	40.11
	Utsav (2020)	English	GISAID	273	15.38	N/A	N/A	N/A	N/A	N/A	N/A
	Nguyen (2020)	English	GISAID	171	25.73	23.39	1.16	4.67	2.33	1.16	41.52

Figures

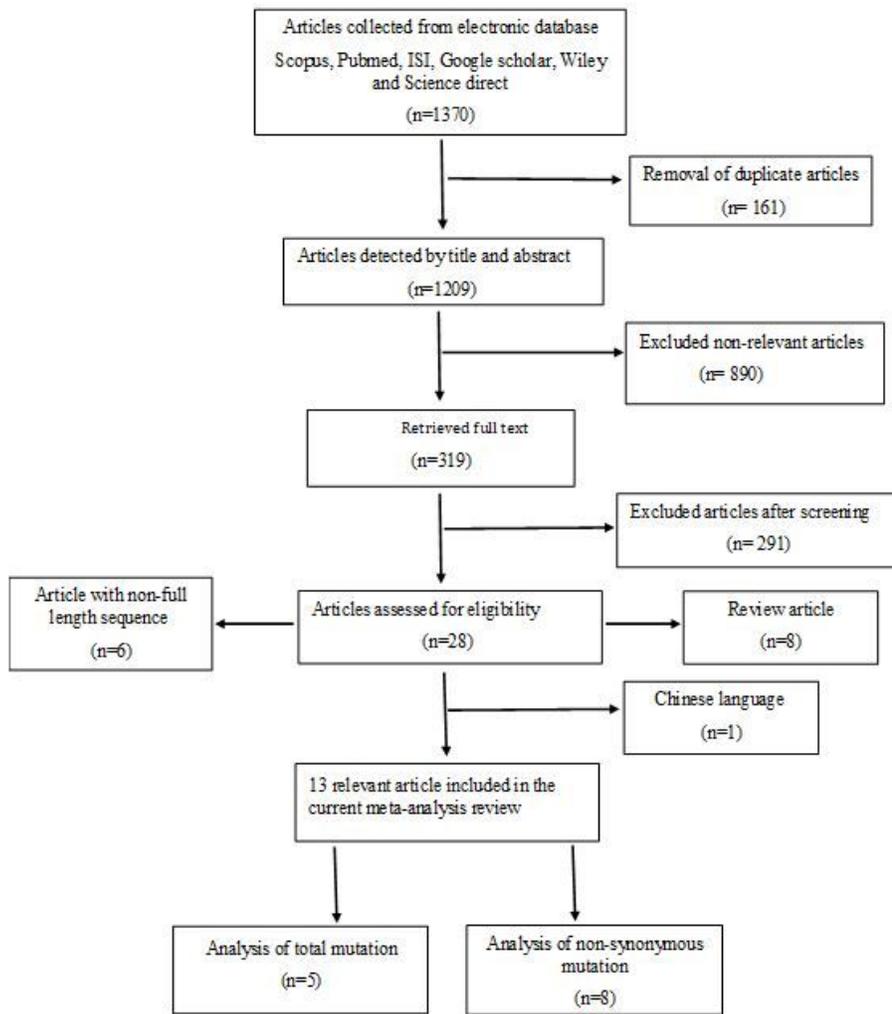


Figure 1

Flow chart of the literature search strategy for selection and including primary articles

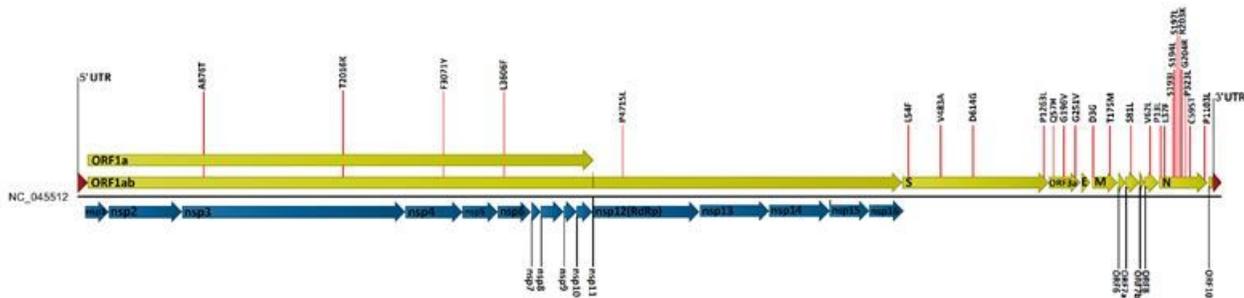


Figure 2

Frequently mutated residues among COVID-19 sequences from different locations

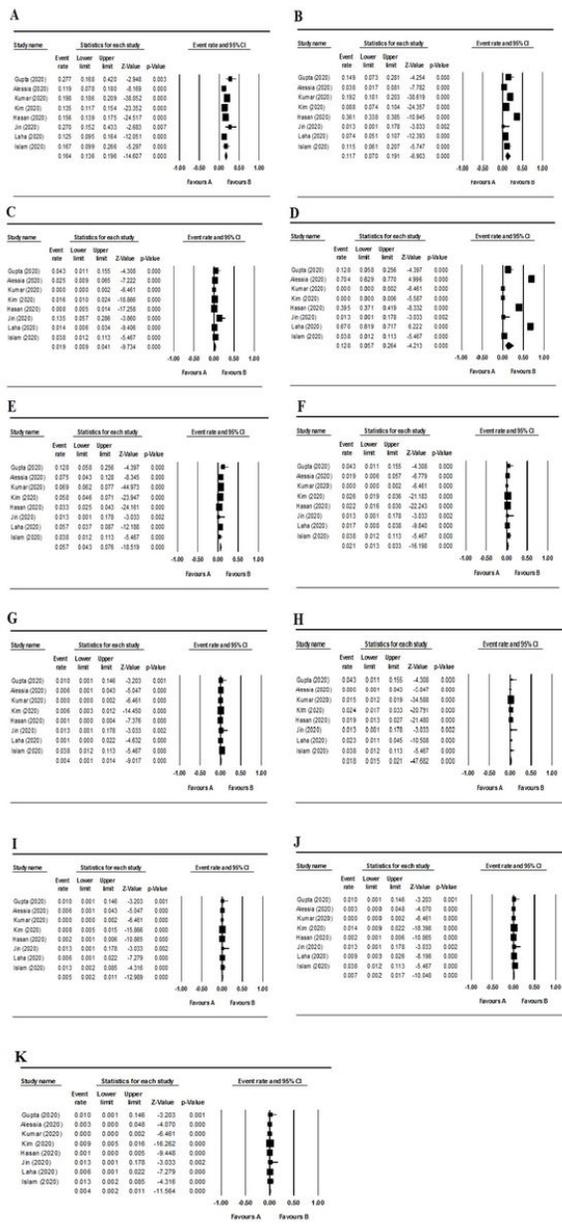


Figure 3

A: Estimation of mutations in S protein among non-synonymous mutations. B: Estimation of mutations in N protein among non-synonymous mutations. C: Estimation of mutations in M protein among non-synonymous mutations. D: Estimation of mutations in ORF1/ab among non-synonymous mutations. E: Estimation of mutations in ORF3a among non-synonymous mutations. F: Estimation of mutations in ORF7a among non-synonymous mutations. G: Estimation of mutations in ORF7b among non-synonymous mutations. H: Estimation of mutations in ORF8a among non-synonymous mutations. I: Estimation of mutations in ORF10a among non-synonymous mutations. J: Estimation of mutations in ORF6 among non-synonymous mutations. K: Estimation of mutations in E protein among non-synonymous mutations. L: Estimation of mutations in S protein among total mutations. M: Estimation of mutations in ORF3a among total mutations. N: Estimation of mutations in NSP among total mutations. O: Estimation of mutations in N protein among total mutations. P: Estimation of mutations in M protein among total mutations. Q: Estimation of mutations in ORF7a among total mutations. R: Estimation of mutations in ORF7b among total mutations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [prismachecklist.docx](#)