Severity of COVID-19 infection and different SARS-CoV-2 variants: current evidence

Thi Loi Dao[1,2,3], Van Thuan Hoang[1,2,3], Philippe Colson[2,4], Jean Christophe Lagier[2,4], Matthieu Million[2,4], Didier Raoult[2,4], Anthony Levasseur[2,4], Philippe Gautret[1,2]*

[1]Aix Marseille Univ, IRD, AP-HM, SSA, VITROME, Marseille, France

[2]IHU-Méditerranée Infection, Marseille, France

[3]Thai Binh University of Medicine and Pharmacy, Thai Binh, Vietnam

[4]Aix Marseille Univ, IRD, AP-HM, MEPHI, Marseille, France

*Corresponding author:

Philippe Gautret

VITROME, Institut Hospitalo-Universitaire Méditerranée Infection, 19-21 Boulevard Jean Moulin 13385 Marseille Cedex 05, France. Phone: + 33 (0) 4 13 73 24 01. Fax: + 33 (0) 4 13 73 24 02. E-mail address: philippe.gautret@club-internet.fr

Abstract (145/150 words)

Since the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), genetic variants have been identified. The virus mutation is also thought to affect the infectivity of the virus or severity of the disease. To date, most studies showed that the viral mutations, especially the D614G variant, correlated with a higher infectivity than the wild-type virus. However, the evidence of the association between viral mutation and severity of the disease is scant. A SARS-CoV-2 variant with a 382-nucleotide deletion (Δ382) was associated with less severe infection in patients. The 11083G>U mutation was significantly associated with asymptomatic patients. By contrast, ORF1ab 4715L and S protein 614G variants were significantly more frequent in patients from countries where high fatality rates were also reported. The COVID-19 pandemic continues to spread worldwide. It is necessary to anticipate large clinical cohorts to evaluate the virulence and transmissibility of SARS-CoV-2 mutants.

Key words: SARS-CoV-2; COVID-19; mutants; variants; infectivity; severity

Introduction

At the end of 2019, an epidemic of severe respiratory infections and pneumonia (named COVID-19) has begun in Wuhan, China. The cause of this outbreak is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. The disease is highly contagious, and the spread of COVID-19 has been taking place at varying rates globally. The World Health Organization (WHO) declared it a Public Health Emergency of International Concern on 30 January 2020, and then a global pandemic on 11 March 2020, less than three months after its appearance (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen). This pandemic is the cause of an unprecedented health care crisis worldwide with more than 55 million confirmed cases and more than 1,300,000 caused deaths to date (https://www.worldometers.info/coronavirus/).

Differences in severity have been observed with respiratory viruses including influenza viruses, rhinovirus, and coronaviruses [1-3]. SARS-CoV-2 affects primarily the respiratory system and the severity of the disease ranges from asymptomatic infection to severe acute respiratory distress [4]. Also, a broad spectrum of neurological symptoms including notably anosmia and ageusia was frequent, [4] and some patients may also present with cutaneous manifestations [5] and gastrointestinal symptoms [5]. Finally, thrombotic, and thromboembolic diseases appeared to be frequent complications in COVID-19 patients [5]. As a consequence, the severity of the disease may greatly vary depending on the clinical presentation and the organs affected by the disease. In addition, the severity of the disease and the mortality rate are related to many host factors, including age, gender, chronic condition, comorbidities, race, and ethnicity [4].

On other hand, the virus mutation is also thought to affect the severity of the disease [6-9]. Since the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), genetic

variants have been identified. In a study conducted on more than 10,000 SARS-CoV-2 genomes from four databases from patients in 68 countries, 5775 distinct genomes were identified including 2969 missense mutations and 36 stop-gained variants [10]. Investigation of a possible selective advantage or of an association with clinical severity of these variants is of paramount importance. Mutations in the gene encoding Spike protein of SARS-CoV-2 have been showed to affect both the virus infectivity and antigenicity, *in vitro* [11]. In Marseille, France, a two-act pattern of incidence of COVID-19 cases occurred and significant differences in clinical outcomes were observed between patients seen in March-April 2020 and those seen in June-August [12,13].

We aim to conduct this review to summarize the relation between genetic variation in SARS-CoV-2 virus and severity of COVID-19 infection, *in vivo* and *in vitro*.

Material and Methods

Search strategy and selection criteria

This review was conducted according to the Preferred Reporting Items for Systematic Reviews and MetaAnalyses (PRISMA) guidelines (http://www.prisma-statement.org). The following databases were investigated in an attempt to identify all relevant studies published on: PubMed (http://www.ncbi.nlm.nih.gov/pubmed), Web of Science (https://www.webofknowledge.com/) and Google Scholar (http://scholar.google.fr/). The most recent search was conducted on December 15, 2020. The topic search terms used for searching through the databases were the following:

#1: “variant” OR “variation” OR “clade” OR “mutation”.

#2: “SARS-CoV-2” OR “COVID-19”

#3: “severity” OR infectivity”

#4: #1 AND #2 AND #3

Only articles published in English were included. For inclusion, articles had to fulfill two criteria: (1) be related to variants of SARS-CoV-2 virus and (2) describe the relation between viral mutation and severity or infectivity of COVID-19 infection or infectivity of SARS-CoV-2. Reference lists of selected articles were screened to identify studies that might have been missing from the research.

After manually removing duplicates, three researchers (TLD, HVT and GP) independently performed the screening of the abstracts, applying the inclusion and exclusion criteria. In addition, articles without an abstract were included for full-text screening and assessed at this stage. Any discordant results were discussed in a consensus meeting. After screening the abstracts, the full texts of the articles were assessed for eligibility by the same three researchers and selected or rejected for inclusion in the systematic review.

Data collection process

The following data (if available) were extracted from each article: country where patients were sampled, time period of the study, number of patients, type of clinical sample, genomic methods, characteristics of variants and outcome.

Data synthesis and analysis

As a result of the nature of the studies and the heterogeneity in patient populations, a formal meta-analysis was not possible. Therefore, the study results were summarized to describe the

relation between genetic variation in SARS-CoV-2 and severity of COVID-19 infection. When possible, percentages not presented in the articles were calculated from the available data.

Results

Study selection and types of studies

The study selection is presented in the flow-diagram (Figure 1). The search algorithm produced 758 articles from the PubMed, Web of Science and Google Scholar databases. After removing duplicates, 237 articles were scanned, based on their title and abstract. A total of 86 articles were processed for full text screening. Twenty-nine articles met the inclusion criteria and were included in the qualitative synthesis of the systematic review (Figure 1) [6-9,11-35].

Of the 29 publications included, ten were preprints [12,13,15,22,26-29,33,35]. Four articles reported *in vitro* studies [11,14-16], eleven articles reported clinical studies [12,13,17-25]. The remaining 14 articles analyzed over 330,000 genomes of SARS-CoV-2 downloaded from the GISAD database with patient status [6-9,26-35].

Most clinical studies were conducted before May 2020, corresponding to the first wave of the COVID-19 pandemic. Only one study was done during the two waves (March to May and May to July, 2020) [19]. Finally, a study was conducted from June to September, 2020, during the second wave of the pandemic [13]. Most of the studies were conducted in the USA (3), followed by France (2), China (2), UK (1) Singapore (1), Vietnam (1) and Uruguay (1).

Relation of viral mutation and infectivity of SARS-CoV-2

Thirteen studies reported the relation between viral mutations and virus load or infectivity of SARS-CoV-2. Of which, four articles were *in vitro* studies [11,14-16], five analyzed sequences downloaded from GISAID database [7,27,30-32] and four were clinical studies [12,17,18,23]. Six studies addressed specially the effect of SARS-CoV-2 D614G mutation in the Spike glycoprotein [11,14-16,18,32]. Eleven studies showed that the viral mutations correlated with a higher infectivity than the wild-type virus [7,11,14-18,23,27,30,32]. Only two studies showed no correlation between viral load and diverse mutational events [12,31].

Relation of viral mutation and severity of COVID-19 infection

A total of 21 articles addressed the effect of mutations on severity of COVID-19 in patients. Of which, ten were clinical studies [12,13,18-25] and eleven were analyzes of sequences from GISAID with patient status [6-9,26,28,29,32-35].

A study by Long et al showed that infection with SARS-CoV-2 variants harboring the D614G substitution was not associated with disease severity, overall mortality, transfer to ICU, mechanical ventilation and length of stay at hospital [19]. This result was supported by other researches [25,32]. In a study conducted among 44 Vietnamese patients, 85 mutations covering 67 variant types were reported. Of which, P323L and D614G variants were the most frequent (present in 40/44 patients), followed by C241U (39/44) and GGG to AAC at 28881-3 variants (33/44). But these mutations were not significantly associated with phenotype of illness [24]. Genomic investigation of 309 SARS-CoV-2 isolates obtained from patients seen in Marseille, in March-April revealed specific mutations clustering in five main groups with no marked correlation with clinical severity of the disease [12]. A study by Zhang et al conducted in China

showed no significant differences between two variants (clade I (ORF3a: p.251G>V, or S: p.614D>G (subclade G)); clade II (ORF8: p.84L>S (28144U>C) and ORF1ab: p.2839S (8782C>U)) regarding disease severity and blood parameters indicative of severity [20]. A study conducted on isolates from patients in Washington, US, allowed identifying two major clades distinguished by twelve polymorphisms in five genes. No significant difference concerning mortality was observed among patients infected with these two clades. [22]. In a study conducted in 88 patients in the USA, most of the sequences (93%) clustered in three main clades (Clade 1, 2 and 3), defining mutations at the US level. The authors showed that the viral mutations have had no effect on time to symptom onset or disease severity [23]. Pawan et al identified seven different variants from 3068 SARS-CoV-2 genomes obtained from GISAID. Of which, three clades (V, S and GH) were of no effect on the outcome of patients [29].

A negative relation between severity of COVID-19 infection and virus mutations was described in recent studies. In Singapore, a SARS-CoV-2 variant with a 382-nucleotide deletion (Δ382) that eliminates open reading frame (ORF) 8 transcription was detected in a cluster of cases in January and February, 2020 and was associated with less severe infection in patients, in terms of hypoxia requiring supplemental oxygen [21]. A French study conducted during the second wave of the epidemic in the country showed that SARS-CoV-2 mutation rate was negatively associated with mortality rate [13]. In addition, 11083G>U mutation was significantly associated with asymptomatic patients [34].

By contrast, ORF1ab 4715L and S protein 614G variants were significantly more frequent in patients from countries where high fatality rates were also reported [6,8,9]. Patients infected with virus clades L, G and O were also exposed to higher risk of severe infection than the base level [29]. Mutation at NSP6 and S protein has had a tendency to increase the death rate [35].

Majumda et al analyzed 218 viral strains obtained from 15 countries. Their result showed that mutation in ORF3 protein increased the mortality of COVID-19 infection [7]. Among 1096 SARS-CoV-2 complete sequences downloaded from UK Biobank, 216 different verified super-variants were identified with 8 predominant generic variants (chr6_148, chr7_23, chr2_197, chr2_221, chr8_99, chr10_57, chr16_4 and chr17_26). These variants were significantly associated with increase of COVID-19 mortality [33].

Discussion

SARS-CoV-2 virus, due to the lack of proofreading activity of the RNA-dependent RNA polymerase, has high mutation rates that may have important effects on the pathogenicity and transmissibility of the virus [12]. The identification of genome variations of SARS-CoV-2 and their relationships with severity of COVID-19 disease is therefore important for controlling and surveying the evolution of the pandemic [10,36]. In addition, mutation rate of SARS-CoV-2 determines the evolution of this virus and the risk of emergent infectious disease [36]. In a study by Koyama et al, median mutation rate of SARS-CoV-2 was estimated at $1.12.10^{-3}$ mutations per site-year 95%CI = [$9.86.10^{-4}$ – $1.85.10^{-4}$] [36]. A high mutation rate around 30% was observed among 95 full-length genomic sequences [37]. An analysis of 48,635 samples showed an average of 7.23 mutations per sample [38]. To date, 32435 SARS-CoV-2 mutations were documented in the public databases (http://cov-glue.cvr.gla.ac.uk/#/replacement). Numbers of variations are the highest in NSP3 protein, followed by S protein, NSP12 protein, NSP2 protein, NSP 13 protein, NSP14, and NSP4 protein. By contrast, very little divergence was documented in NSP11, ORF10, ORF7b and E protein [36]. The figure 2 shows the positions of the mutations and deletions in the genome and of amino acid substitutions in the virion.

A key element of coronavirus host range is determined by the binding affinity between the spike S protein and the cellular receptor. All mutations in protein S could influence host range and transmissibility of the virus [12]. The SARS-CoV-2 Spike protein is a class I fusion protein that forms trimers on the viral surface: it is heavily glycosylated, which enables entry into host cells [39]. Angiotensin-converting enzyme 2 (ACE2) is the target receptor of SARS-CoV-2 virus for entry into the host cell [39]. The main effect of the D614G mutation is to increase the availability of spike trimer components in the conformation and that permits enhancing the binding of the virus spike to the ACE2 receptor. *In vitro* and *in vivo* studies to date showed that the mutation D614G in Spike protein was associated with higher viral loads and probably with enhanced transmissibility of the virus [14,15,18]. Therefore, this mutation emerged and has become the dominant form in the global pandemic worldwide within a matter of months. It suggests that G614 may have a fitness advantage. [40]. The frequency of S protein 614G was significantly associated with high fatality rates, in several countries as reported in studies which analyzed SARS-CoV-2 sequences from GISAID database [6,8,9]. However, clinical studies showed that this mutation did not correlate with severity of COVID-19 disease, including mortality, transfer to ICU, mechanic ventilation, or length of stay at hospital [18,19,20,23-25].

In addition, clinical studies have shown that other viral mutations were not related to severity of COVID-19 infection or were associated with less severe infection in patients. Young et al showed that the patients infected with Δ382 have had lower concentrations of inflammatory biomarkers. Furthermore, these patients have had a higher concentration of SDF-1 α which is low in patients with hypoxemia [21]. Interestingly, the replication capacity of Δ382 variant *in-vitro* is similar to that of wild-type SARS-CoV-2. It suggests that this mutation does not reduce replicative fitness [21]. In a study by Colson et al, the authors demonstrated seven new mutations

of SARS-CoV-2, named “Marseille 1” to “Marseille 7”. Moreover, heterogeneity of the sequences produced from June to August 2020 (second outbreak) was higher than in sequences produced from February to May 2020 (first outbreak) ($7.6.10^{-4} \pm 3.8.10^{-4}$ versus $2.3.10^{-4} \pm 1.1.10^{-4}$). This result indicates that the rate of virus mutation has increased rapidly. By contrast, the mortality of COVID-19 patients during the second outbreak was lower than that of those in the first one [13].

Our study has some limitations. We screened papers published only in English and reported in PubMed, Web of Science and Google scholar. Ongoing research projects have not been captured. For example, in our University Hospital Institute, a large cohort study aiming at comparing the demographic and clinical characteristics of patients infected with several new variants of SARS-CoV-2 virus during July to September 2020 is ongoing. Also, a variant with multiple spike protein mutations named SARS-CoV-2 VUI 202012/01 has been recently identified in the United Kingdom [https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf]. Preliminary results showed that this variant was significantly associated with increase of transmissibility compared to the previously circulating variants but no relation with disease severity was observed to date. Investigations to understand the spread of this new variant across the UK and Europe and evaluation of clinical severity and antigenic change are ongoing. Nevertheless, our review gives an overview on the relation between SARS-CoV-2 genetic variations and infectivity or severity of COVID-19 infection. In conclusion, most studies showed that some genetic variants of the virus were associated with high virus load. But to date, the evidence of the association between viral mutation and severity of the disease is scant. On the other hand, severity and outcome of

COVID-19 infection depend also on the host's genetic factors, on treatment and clinical management which have been improved, and on increased hospital capacity and response speed. The COVID-19 pandemic continues to spread worldwide. It is necessary to anticipate large clinical cohorts to evaluate the virulence and transmissibility of SARS-CoV-2 mutants.

This manuscript has been edited by a native English speaker.

Ethical Approval

NA

Consent to participate

NA

Consent to Publish

NA

Authors Contributions

Conceptualization: Philippe Gautret, Didier Raoult

Methodology : Thi Loi Dao, Van Thuan Hoang, Philippe Colson, Jean-Christophe Lagier, Matthieu Million, Didier Raoult, Antony Levasseur

Data collection: Thi Loi Dao, Van Thuan Hoang, Philippe Gautret

Formal analysis and investigation: Thi Loi Dao, Van Thuan Hoang, Philippe Colson, Antony Levasseur, Philippe Gautret

Writing original draft and preparation: Thi Loi Dao, Van Thuan Hoang, Philippe Gautret

Writing – review and editing: Thi Loi Dao, Van Thuan Hoang, Philippe Colson, Jean-Christophe Lagier, Matthieu Million, Didier Raoult, Antony Levasseur, Philippe Gautret

Supervision: Philippe Gautret

Funding

No funding.

Conflict of Interest

The author declare that they have no conflict of interest.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from corresponding author [P.G] on reasonable request.

References

1. Daoud A, Laktineh A, Macrander C, Mushtaq A, Soubani AO. Pulmonary complications of influenza infection: a targeted narrative review. Postgrad Med **2019**;131:299-308.
2. Hayden FG. Rhinovirus and the lower respiratory tract. Rev Med Virol **2004**;14:17-31.
3. Su S, Wong G, Shi W, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends Microbiol **2016**;24:490-502.
4. Chidambaram V, Tun NL, Haque WZ, et al. Factors associated with disease severity and mortality among patients with COVID-19: A systematic review and meta-analysis. PLoS One **2020**; 15: e0241541
5. Lai CC, Ko WC, Lee PI, Jean SS, Hsueh PR. Extra-respiratory manifestations of COVID-19. Int J Antimicrob Agents **2020**;56:106024.
6. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. J Hum Genet **2020**; 65:1075–1082.
7. Majumdar P, Niyogi S. ORF3a mutation associated with higher mortality rate in SARS-CoV-2 infection. Epidemiology and Infection **2020**;148:1–6.
8. Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. Int J Clin Pract **2020**;74:e13525.
9. Eaaswarkhanth M, Al Madhoun A, Al-Mulla F. Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? Int J Infect Dis **2020**;96:459-460.
10. Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. Bull World Health Organ **2020**; 98:495–504.

11. Li Q, Wu J, Nie J, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell **2020**; 182:1284-1294.e9.
12. Levasseur A, Delerce J, Caputo A, Brechard L, et al. Genomic diversity and evolution of coronavirus (SARS-CoV-2) in France from 309 COVID-19-infected patients. bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.282616
13. Colson P, Levasseur A, Delerce J, et al. Dramatic increase in the SARS-CoV-2 mutation rate and low mortality rate during the second epidemic in summer in Marseille. Preprints. https://doi.org/10.35088/68c3-ew82
14. Hou YJ, Chiba S, Halfmann P, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science **2020**: eabe8499. doi: 10.1126/science.abe8499
15. Zhang L, Jackson CB, Mou, et al. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv preprint doi: https://doi.org/10.1101/2020.06.12.148726
16. Plante JA, Liu Y, Liu J, et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature **2020**. doi: 10.1038/s41586-020-2895-3
17. Yao H, Lu X, Chen Q, et al. Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity in vitro and have clinical implications in vivo. Cell Discov. **2020**;6:76.
18. Korber B, Fischer WM, Gnanakaran S, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell **2020**; 182:812-827.e19.

19. Long SW, Olsen RJ, Christensen PA, et al. Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area. mBio **2020**; 11.
20. Zhang X, Tan Y, Ling Y, et al. Viral and host factors related to the clinical outcome of COVID-19. Nature **2020**; 583:437-440.
21. Young BE, Fong S-W, Chan Y-H, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet **2020**; 396:603–611.
22. Nakamichi K, Shen JZ, Lee CS, et al. Outcomes associated with SARS-CoV-2 viral clades in COVID-19. medRxiv preprint doi: https://doi.org/10.1101/2020.09.24.20201228
23. Lorenzo-Redondo R, Nam HH, Roberts SC, et al. A clade of SARS-CoV-2 viruses associated with lower viral loads in patient's upper airways. EBioMedicine **2020**; 62:103112.
24. Nguyen TT, Pham TN, Van TD, et al. Genetic diversity of SARS-CoV-2 and clinical, epidemiological characteristics of COVID-19 patients in Hanoi, Vietnam. PLoS One **2020**; 15:e0242537.
25. Elizondo V, Harkins GW, Mabvakure B, et al. SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay. Emerg Microbes Infect **2020**;1-52.
26. Aiewsakun P, Wongtrakoongate P, Thawornwattana Y, Hongeng S, Thitithanyanont A. SARS-CoV-2 genetic variations associated with COVID-19 severity. medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20114546

27. Wang R, Chen J, Gao K, Hozumi Y, Yin C, Wei G. Characterizing SARS-CoV-2 mutations in the United States. Preprint. Res Sq **2020**;rs.3.rs-49671.
28. Nagy A, Pongor S, Gyorffy B. Different mutations in SARS-CoV-2 associated with severe and mild outcome. medRxiv preprint doi: https://doi.org/10.1101/2020.10.16.20213710
29. Pawan V, Rasha E, Oluwasefunmi S, et al. COVID-19 Mortality Risk Assessment among Various Age Groups Using Phylogenetic Analysis. Preprints. doi:10.20944/preprints202009.0487.v1
30. Chen J, Wang R, Wang M, Wei G-W. Mutations Strengthened SARS-CoV-2 Infectivity. J Mol Biol **2020**; 432:5212–5226.
31. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F. No evidence for increased transmissibility of recurrent mutations in SARS-CoV-2. Nat Commun. **2020**;11:5986.
32. Volz E, Hill V, McCrone JT, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell **2020**;S0092-8674(20)31537-3.
33. Hu J, Li C, Wang S, Li T, Zhang H. Genetic variants are identified to increase risk of COVID-19-related mortality from UK Biobank data. medRxiv preprint doi: https://doi.org/10.1101/2020.11.05.20226761
34. Wang R, Chen J, Hozumi Y, Yin C, Wei GW. Decoding Asymptomatic COVID-19 Infection and Transmission. J Phys Chem Lett **2020**;11:10007-10015.
35. Jalali S, Bhadra B, Dasgupta S. Mutation analysis of structural and non-structural proteins indicated that the death rate of COVID-19 pandemic may significantly reduce by the end of 2020. Preprints. doi: 10.31219/osf.io/gubys

355 36. Sardar R, Satish D, Birla S, Gupta D. Integrative analyses of SARS-CoV-2 genomes
356 from different geographical locations reveal unique features potentially consequential to
357 host-virus interaction, pathogenesis and clues for novel therapies. Heliyon.
358 **2020**;6:e04658.
359 37. Wang C, Liu Z, Chen Z, et al. The establishment of reference sequence for SARS-CoV-2
360 and variation analysis. J Med Virol **2020**; 92:667–674.
361 38. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2
362 Mutations. Front Microbiol **2020**; 11 :1800
363 39. Yurkovetskiy L, Wang X, Pascal KE, et al. Structural and Functional Analysis of the
364 D614G SARS-CoV-2 Spike Protein Variant. Cell **2020**; 183:739-751.e8.
365 40. Groves DC, Rowland-Jones SL, Angyal A. The D614G mutations in the SARS-CoV-2
366 spike protein: Implications for viral infectivity, disease severity and vaccine design.
367 Biochem Biophys Res Commun **2020**;S0006-291X(20)32038-6.

368

Table 1: Relation of viral mutation and infectivity of SARS-CoV-2 and/or severity of COVID-19 infection

Ref	Country where patients were sampled	Period of time	Number of patients	Type of samples	Sequencing methods / data availability	Variants	Outcome
6	50 countries from six geographic areas		12,343	ND	12,343 SARS-CoV-2 sequences isolated in 50 different countries from six geographic areas obtained from GISAID database	1234 mutations, including 57Q>H, 251G>V (ORF3 protein), 265U>I, 378V>I, 5865Y>C, 5828P>L, 4489A>V, 2016U>K, 3606L>F, 4715P>L (ORF1ab protein), 614D>G (S protein), , 204G>R, 203R>K, 13P>L (N protein), 175U>M (M protein), 84L>S (ORF8 protein)	ORF1ab 4715L and S protein 614G variants were significantly more frequent in patients from countries where high fatality rates were reported.
7	23 countries		Approximately 20,000 case reports	ND	SARS-CoV-2 strains for each country were extracted from NextStrain open-source project. Amino acid sequences of ORF3a protein were downloaded from NCBI protein database	218 viral strains from 15 countries were further analyzed for amino acid mutations from NextStrain database	Mutation in ORF3a protein was associated with increased infection and mortality rate of SARS-CoV-2.

8	Various		ND	ND	SARS-CoV-2 viral spike sequences were accessed from the GISAID database	D614G variant	Both the average and median case fatality rates correlate strongly ($p < 0.02$) with the proportion of G614 variant
9	Various		4246	ND	4,246 SARS-CoV-2 genomes downloaded from GISAID	D614G variant	D614G variant was associated with high mortality related to COVID-19 in European populations
11	NA	NA	*In vitro* study	NA	NA	S mutants reported in the public domain or mutants at putative N-linked glycosylation sites. We analyzed their infectivity and reactivity to neutralizing antibodies using the high-throughput pseudotyped virus system	Pseudotyped viruses expressing either the D614G single mutation or a combination of mutations that included D614G are more infectious than the reference strain, whereas no difference was found between single D614G and D614G combination variants
12	France	February 29th to April 4th, 2020	309	Nasopharyngeal swabs	Sequencing by Illumina protocols on MiSeq platform (Illumina).	A total of 321 mutational events were reported in the SARS-CoV-2 genomes divided in 5 clusters. Cluster 1 (44 patients, 14.2%, positions [28881-28882-28883])	Poor clinical outcome (PClinO, defined by either death or transfer to intensive care unit or hospitalization for 10 days or more) and poor virological outcome (PVirO, defined by viral shedding persistence at

						with two nonsynonymous mutations in protein N (R203K; G204R). Cluster 2 (39 patients, 12.6%, position 15324) contains a synonymous mutation (C15324U). Cluster 3 (126, 100 and 211 patients, at positions 2416, 8371, 25563, respectively) includes one synonymous mutation (C2416U), and two nonsynonymous mutations (nsp3: Q1884H; ORF3a: Q57H). Cluster 4 (68 patients, 22%, position 1059) contains one nonsynonymous mutation (nsp2: T85I). Finally, cluster 5 (from 297 to 303 patients, 96-98%,	day 10). Coronavirus genome isolates from 38 patients' isolates with PVirO were widely distributed across the groups, including diverse mutational events meaning that there is no correlation between higher viral loads. For the 10 patients with PClinO, a majority of isolates were also distributed into all groups. An exception concerned two patients (IHUCOVID-0318, IHUCOVID-0333) with one PClinO and one death that were clusterized together in a group of seven different isolates.

						positions 241, 3037, 14408, 23403) displays one mutation in 5'UTR (C241U), one synonymous mutation (C3037U) and two nonsynonymous mutations (nsp12b: P314L, S protein 133 D614G).	
13	France	June - September, 2020	691	Nasopharyngeal swabs	Next-generation sequencing using Illumina technology with the Illumina Nextera XT Paired end strategy on a MiSeq instrument	Marseille-1 to Marseille-7, located in most SARS-CoV-2 genes including structural and non-structural genes among which nsp2, nsp3 (predicted phosphoesterase), nsp5 (membrane glycoprotein), nsp12 (RNA-dependent RNA polymerase), S (Spike glycoprotein), ORF3a, E (membrane glycoprotein), M (membrane glycoprotein),	SARS-CoV-2 mutation rate was negatively associated with mortality rate.

						ORF8 and N (Nucleocapsid phosphoprotein)	
14	NA	NA	*in vitro* and animal model study with a mutant virus and a wild-type virus	NA	NA	SARS-CoV2 D614G mutation in the Spike glycoprotein	SARS-CoV-2 variants harboring the D614G substitution replicated more efficiently in some immortalized epithelial cell lines and exhibited significantly faster droplet transmission between infected hamsters than the wild-type virus.
15	NA	NA	*In vitro* study	NA	NA	SARS-CoV2 D614G mutation in the Spike glycoprotein	Pseudovirus G614 infected hACE2-293T cells with approximately 9-fold higher efficiency than did Pseudovirus D614
16	NA	NA	*In vitro* and animal model study	NA	NA	SARS-CoV2 D614G mutation in the Spike glycoprotein	D614G mutation increases the infectivity of SARS-CoV-2 produced from a human lung cell line. Hamsters infected with the G614 variant produced higher infectious titers in the nasal washes and trachea, but not lungs, confirming clinical evidence that the D614G mutation enhances viral loads in

							the upper respiratory tract of COVID-19 patients
17	China	January 22nd to February 4th, 2020	*In vitro* and clinical study among 11 patients	Sputum, stool and nasopharyngeal swabs	Deep sequencing by the Novaseq 6000 platform (Illumina)	33 mutations were identified in 11 isolates	Different viral isolates, exhibit a significant variation of viral load when infecting Vero-E6 cells. ZJU-1, which clusters with the S-D614G clade, has a viral load 19 times higher than ZJU-2 and ZJU-8. A near 270-fold difference in viral load was observed between ZJU-10 and ZJU-2 at 24 hours post infection. In addition, a higher viral load leads to a higher cell death ratio
18	UK	Between March and May 2020	999	throat or combined nose/throat swabs	Long-read whole genome sequencing (Oxford Nanopore Technologies (ONT), Oxford, UK) using the ARTIC network protocol	SARS-CoV2 D614G mutation in the Spike glycoprotein	SARS-CoV-2 variants harboring the D614G substitution were associated with potentially higher viral loads in COVID-19 patients but not with disease severity
19	USA	March 5th to May 11th, 2020 (first wave) and May 12th to July 7th, 2020	1026 (first wave) and 4059 (second wave)	Nasopharyngeal swabs	Long reads were generated with the LSK-109 sequencing kit, 24 native barcodes (NBD104 and NBD114 kits), and a GridION instrument (Oxford Nanopore). Short reads were generated with a NexteraXT kit and a NextSeq	SARS-CoV2 D614G mutation in the Spike glycoprotein	No relationship between virus clades and disease severity (overall mortality, transfer to ICU, mechanical ventilation and length of stay).

		(second wave)			550 instrument (Illumina)		
20	China	January 20th - February 25th, 2020	112	Sputum or nasopharyngeal swabs	Sequencing by Illumina protocols on MiSeq platform (Illumina)	Clade I (ORF3a: p.251G>V (subclade V), or S: p.614D>G (subclade G)). Clade II (ORF8: p.84L>S (28144U>C) and ORF1ab: p.2839S (8782C>U))	Patients with Clade II viruses were younger than those with Clade I viruses (median age = 46.5 vs 57.5, p = 0.02). There were no significant differences between variants regarding disease severity, leukocytes, lymphocyte and platelet count, CD3 T cell count, Haemoglobin , C-reactive protein, Lactose dehydrogenase, complement C3,D-dimer or IL-6 and IL-8 level, or the duration of virus shedding after onset
21	Singapore	January 22nd to March 21st, 2020	131	Respiratory sample	2 specific PCRs were used to detect the 382-nucleotide deletion in the SARS-CoV-2	92 (70%) were infected with the wild-type virus, ten (8%) had a mix of wild-type and Δ382-variant viruses, and 29 (22%) had only the Δ382 variant	Infection with the Δ382 variant was only associated with lower odds of developing hypoxia requiring supplemental oxygen (adjusted odds ratio 0·07 [95% CI 0·00–0·48]) compared with infection with wild-type virus only
22	USA	March 1st	190	Nasopharyngeal	Samples were sequenced on	97 samples	A trend toward higher

		to April 15th, 2020		samples	MiSeq, NextSeq or NovaSeq instruments (Illumina) using 1x185, 1x75, or 1x100 runs respectively.	corresponded to what we refer to as 'Clade 1' and 91 corresponded to 'Clade 2'. Two of 190 samples did not fall into either of the two major clades. When mapped onto GISAID and NextStrain clades: in clade 1, 89 corresponded to clades GH/20C, 6 mapped to G/20A, and 2 mapped to G/20B. In clade 2, 86 corresponded to S/19B, and 5 mapped onto L/19A. The 2 of the 190 samples that did not fall into either of the major clades corresponded to GH/20C and S/19B.	rates of hospitalization of patients with Clade 2 virus was observed (p=0.06). Mortality was not significantlly different in patients infected with Clade 1 and 2 viruses
23	USA	Mid-March 2020	88	Nasopharyngeal swabs	Library sequencing performed on the Nanopore MinION device using FLO-MIN106D Type R9.4.1 flow cells	Most of the sequences (93%) clustered in three main clades (Clade 1, 2 and 3), defining mutations at the US level	Patients infected with Clade 1 viruses had significantly higher average viral loads in their upper airways relative to patients infected with Clade 2

							viruses, independently of time to symptom onset and disease severity
24	Vietnam	March 6th to April 15th, 2020	44	Nasopharyngeal and oropharyngeal swabs	Sequencing was performed on an Illumina Miseq platform (Nextera XT Library preparation kit)	85 mutations covering 67 variant types among the 44 SARS-CoV-2 genomes. The most ubiquitous modifications were C3037U, C14408U (P323L) and A23403G (D614G) occurring in 40/44 samples. Two other variants C241U and GGG to AAC at 28881–3 were detected in 39 and 33 sequences, respectively	These mutations were not associated with differences in phenotype of illness
25	Uruguay	March to May 2020	44	Naso-oropharyngeal swabs	Whole SARS-CoV-2 genomes were sequenced using Illumina NovaSeq 6000.	D614G mutation	The spike D614G mutation and clade G-related viruses, were not associated with any clinical parameters, severity, or lethality of COVID-19 infection
26	Various		152	Not documented	Genomes of SARS-CoV-2 with patient status. Criteria for selection were full-length sequences and high sequencing	Two genetic variations were observed at the nucleotide	Asymptomatic SARS-CoV-2 tended to have 11083U (N(11083U)/N(11083

					coverage (downloaded from the GISAID database)	position 11,083, namely thymine (11083U, 75/152 = 49.34%) and guanine (11083G, 72/152 = 47.37%)	G) = 60/7), while viruses causing symptomatic cases tended to have 11083G (N(11083U)/N(11083 G) = 15/65). The relative risk ratio of developing symptoms given 11083G to 11083U was (65/72)/(15/75) = 4.51 times (95% confidence interval = 2.85–7.14), and the odd ratio was estimated to be 37.14 by the Wald method (95% confidence interval = 14.17–97.33).
27	USA		7823		28726 complete SARS-CoV-2 genome sequences downloaded from GISAID	4968 single mutations are detected with top eight missense mutations (i.e., 14408C>U-(P323L), 23403A>G-(D614G), 25563G>U-(Q57H), 1059C>U-(T85I), 28144U>C-(L84S), 17858A>G-(Y541C), 17747C>U-	Based on co-mutation and time evolution analysis, three concurrent mutations 17747C>U-(P504L), 17858A>G- (Y541C), and 28144U>C tend to fade out, while the other five concurrent mutations can enhance the infectivity of SARS-CoV-2

						(P504L), and 27964C>U-(S24L)) are identified	
28	Various		73020	ND	72,331 viral sequences downloaded from GISAID database. Clinical data was available for 5,094 patients, and 3,184 of them had also follow-up data	2,121 different mutations affecting the protein structure were identified	Mutations correlated with mild outcome were located in the ORF8, NSP6, ORF3a, NSP4, and in the nucleocapsid phosphoprotein N. Mutations associated with inferior outcome were located in the surface (S) glycoprotein, in the RNA dependent RNA polymerase, in the 3'-to5' exonuclease, in ORF3a, NSP2 and N. Mutations leading to severe outcome with low prevalence were found in the surface (S) glycoprotein and in NSP7
29	Various		3608	ND	3068 SARS-CoV-2 genomes downloaded from GISAID	7 different variants (Clade G, GH, GR, L, O, S and V)	Patients infected with virus clades L, G and O are exposed to higher risk than the base level. Patients infected with clade GR were associated with the low risk. Clade V, S and GH were of no effect on the outcome

							of patients.
30	17 countries		24175	ND	24175 complete SARS-CoV-2 genomes downloaded from GISAID	11904 single mutations found in 6 distinct clusters	Mutations on the RBD strengthen the binding of S protein and ACE2, leading to more infectious SARS-CoV-2
31	Various		46,723	ND	46,723 complete SARS-CoV-2 genomes downloaded from GISAID	12,706 variable positions	None of the recurrent SARS-CoV-2 mutations were associated with increased viral transmission
32	UK	January 29th to June 16th, 2020	ND	ND	21,231 614G and 5,755 614D de-duplicated whole genome sequences were downloaded from The COVID-19 Genomics UK consortium dataset	245 and 62 clusters of 614G and 614D variants containing UK virus genomes from 10 or more different patients were identified, respectively,	614G variant was not associated with mortality or severity of COVID-19. But this mutation was associated with higher viral load and younger age of patient.
33	UK	-	1096	ND	1096 SARS-CoV-2 complete sequences were downloaded from UK Biobank	216 different verified super-variants across 10 repetitions of the discovery-validation procedure were found. Two super-variants chr6_148 and chr7_23, identified in 4 out of 10 repetitions. Six other super-variants,	Eight genetic variants are identified to significantly increase risk of COVID-19 mortality

						chr2_197, chr2_221, chr8_99, chr10_57, chr16_4 and chr17_26 identified in 3 out of 10 repetitions.	
34	Various	-	75775	ND	75775 SARS-CoV-2 complete genome sequences were downloaded from GISAID database. 9912 samples have patient status information recorded as asymptomatic, symptomatic, hospitalized, intensive care unit, deceased. Of which, 537 samples are labeled with asymptomatic (76) and symptomatic (461) cases	11083G>U mutation changes leucine to phenylalanine residue at position 37 of NSP6 protein	11083G>U mutation was significantly associated with asymptomatic patients (OR = 33.4, p = $8.45.10^{-35}$)
35	Various	April-July, 2020	41304	ND	41304 SARS-CoV-2 protein sequences from 49 different countries were downloaded from NCBI GenBank.	Mutation at NSP6, ORF8, S, M, E and N protein	A relationship of positive tendency between the death rate and the mutation rate was noted in case of NSP6 and S proteins

Figure 1: Study flow chart

Figure 2: Positions of mutations and deletions in the genome and of amino acid substitutions in the virion