OPEN ACCESS Freely available online

Journal of Vaccines & Vaccination

Opinion Article

Preliminary Estimates of the Mutational Frequencies of Individual Genes in SARS-CoV-2 Genome

Wenfa Ng*

Department of Biomedical Engineering, National University of Singapore, Novena, Singapore

ABSTRACT

Evolutionary selection pressure or presence of error-prone polymerase in virus would result in emergence of variants that are more infectious or more deadly. But, different genes encode proteins of different functions in the pathogenic process of the virus, and thus, are subjected to different evolutionary selection pressure. This necessarily leads to different mutational frequencies which have implications for the evolution of the virus, and its associated pathogenic potential. This work sought to determine a preliminary estimate of the mutational frequency of each gene in SARS-CoV-2 genome using 10 assembled genomes deposited in European Nucleotide Archive. Results reveal that five genes are commonly mutated. Specifically, these genes are RNA-dependent RNA polymerase, spike protein, *ORF3a*, ORF8 and nucleocapsid protein. In particular, RNA-dependent RNA polymerase and spike protein have the highest mutational frequency at 90% followed by *ORF3a* (30%), nucleocapsid protein (20%) and ORF8 protein (10%). Overall, estimating the mutational frequencies of each gene in a virus genome provides critical knowledge for correlating severity of disease with genotype as well as understanding the epidemiological profile of the virus. Such efforts in SARS-CoV-2 conducted in this work has illuminated that spike protein and RNA-dependent RNA polymerase have the highest mutational rate which correlates with publicized frequent emergence of variants of concern in different parts of the world.

Keywords: Mutational frequency; SARS-CoV-2; Evolutionary pressure; RNA-dependent; RNA polymerase; Spike protein

DESCRIPTION

Mutations are common in viral genes and genomes as they are the bedrock on which the virus evolves to evade immune surveillance and attack, as well as adapt to new host environment [1-4].

Hence, it is important to gain an understanding of the mutational frequencies of each gene in the genome of a virus in order to appreciate its evolutionary trajectory, and predict how it will evolve in future [5-7].

This work sought to obtain preliminary estimates of the mutational frequencies of each gene in the genome of SARS-CoV-2 to help decipher the underlying reasons why there have been frequent emergence of variants of concern.

A total of 10 different genomes of the virus belonging to individual isolates were downloaded from European NucleotideArchive.

The genome sequence of each isolate is used as template on which different genes in the reference genome of SARS-CoV-2 downloaded from Genbank was aligned.

Mutation in gene is called when there is no exact match in the isolate's genome sequence for a particular gene in the reference genome. Mutational data was thus aggregated over ten isolates and help produce the mutational frequency table. All computations and

Correspondence to: Wenfa Ng, Department of Biomedical Engineering, National University of Singapore, Novena, Singapore, E-mail: ngwenfa@nus.edu.sg

Received date: November 5, 2021; **Accepted date:** November 19, 2021; **Published date:** November 26, 2021

Citation: Wenfa Ng (2021) Preliminary Estimates of the Mutational Frequencies of Individual Genes in SARS-CoV-2 Genome. J Vaccines Vaccin. 12:467.

Table 1: Mutational frequency of each gene in the genome of SARS-CoV-2.

Gene	Mutational frequency (%)
ORF1ab polyprotein	90
Spike protein	90
ORF3a protein	30
Envelope protein, E protein	0
Membrane glycoprotein, M protein	0
ORF6 protein	0
ORF7a protein	0
ORF7b protein	0
ORF8 protein	10
Nucleocapsid phosphoprotein	20
ORF10 protein	0

Results presented in Table 1 reveal that five genes suffer high mutational burden compared to other genes in the genome of SARS-CoV-2. These genes are *ORF1ab* encoding RNA-dependent RNA polymerase, spike protein, *ORF3a* protein, nucleocapsid protein and ORF8 protein. In particular, RNA-dependent RNA polymerase and spike protein suffer the highest mutational burden at 90% mutational frequency, which holds significant implications for vaccine development and containing the spread of variants of concern. Specifically, mutations in RNA-dependent RNA polymerase changes the shape and possibly the functional efficiency of the enzyme, which suggests that errors will creep in during transcription of large genes that may increase the pathogenic potential of the virus. In the case of spike protein, mutations in this protein will alter its shape, and thwart efforts to develop neutralizing antibodies for this antigen or in vaccine development.

CONCLUSION

Overall, mutations drive the evolution of virus in unpredictable ways. Although the course of evolution can be shaped by natural selection pressure in the lab, the same cannot be said of virus in the environment exposed to changing environmental conditions and host factors. This work probes the mutational frequency of each gene in the genome of SARS-CoV-2 and reveals that five genes are of highest mutational burden. These genes are *ORF1ab*, spike protein, *ORF3a* protein, ORF8 protein, and nucleocapsid protein. Of particular concern is the exceedingly high mutational frequency (90%) of *ORF1ab* (RNA-dependent RNA polymerase) and spike protein, which explains the frequent emergence of new variants of concern with higher infectivity, as well portends a future in which new variants will emerge if we fail to contain the spread of the virus, as mutations potentially occur with each replication of the virus in cells.

REFERENCES

1. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol. 2021;19(7):409-424.
2. Van Egeren D, Novokhodko A, Stoddard M, Tran U, Zetter B, Rogers M, et al. Risk of rapid evolutionary escape from biomedical interventions targeting SARS-CoV-2 spike protein. PloS One. 2021;16(4):e0250780.
3. Lazarevic I, Pravica V, Miljanovic D, Cupic M. Immune Evasion of SARS-CoV-2 Emerging Variants: What Have We Learnt So Far?. Viruses. 2021;13(7):1192.
4. Rashid F, Suleman M, Shah A, Dzakah EE, Wang H, Chen S, et al. Mutations in SARS-CoV-2 ORF8 Altered the Bonding Network With Interferon Regulatory Factor 3 to Evade Host Immune System. Front Microbiol. 2021:1811.
5. Jaroszewski L, Iyer M, Alisoltani A, Sedova M, Godzik A. The interplay of SARS-CoV-2 evolution and constraints imposed by the structure and functionality of its proteins. PLoS Comput Biol. 2021 ; 17(7):e1009147.
6. Yi K, Kim SY, Bleazard T, Kim T, Youk J, Ju YS. Mutational spectrum of SARS-CoV-2 during the global pandemic. Exper Molecular Med. 2021;53(8):1229-1237.
7. De Maio N, Walker CR, Turakhia Y, Lanfear R, Corbett-Detig R, Goldman N. Mutation rates and selection on synonymous mutations in SARS-CoV-2. Genome Biol Evol. 2021;13(5):087.