Print ISSN : 0972-8813
e-ISSN : 2582-2780

[Vol. 20(1), January-April, 2022]

Pantnagar Journal of Research

(Formerly International Journal of Basic and Applied Agricultural Research ISSN : 2349-8765)

G.B. Pant University of Agriculture & Technology, Pantnagar

PANTNAGAR JOURNAL OF RESEARCH

Vol. 20(1) January-April, 2022

CONTENTS

An analysis of selective pressure on Delta variants of SARS CoV2 circulating in India

BEENU JAIN[1], ANUJ TEWARI[2], ALOK KUMAR[3], PREMA S GODI[4] and SHRADDHA DWIVEDI[5]

[1]Department of Animal Husbandry (Uttar Pradesh),[2]Department of Veterinary Microbiology, College of Veterinary and Animal Sciences, G. B. Pant University of Agriculture and Technology, Pantnagar-263145 (U.S. Nagar, Uttarakhand), [3]Department of Public Health (Uttar Pradesh), [4]Maharishi Vishwamitra Medical College, Ghazipur (Uttar Pradesh), [5]College of Veterinary and Animal Sciences, G. B. Pant University of Agriculture and Technology, Pantnagar-263145 (U.S. Nagar, Uttarakhand)

ABSTRACT: In this study 423 SARS-CoV-2 genome encoding for spike protein of Delta lineage circulating upto December 2022 in India were downloaded from NCBI and GISAID and aligned against the reference sequence MN908947.3. Total 192 mutations were found to be prevalent in the spike protein of Delta variants circulating in India out of which 125 were non-synonymous mutations and 67 were synonymous mutations. The non-synonymous mutations were further subjected for analysis to analyse the effect of mutation on the functionality of the protein. Out of the total 125 non-synonymous mutations, nine were found to have deleterious effect on the functionality of spike protein. Remaining 116 mutations were found to have neutral effect on functionality of spike protein. The selection pressure imparted by the host was analysed using Data monkey sever. The selective pressure analysis revealed that site numbers 142, 152 and 222 of SARS-CoV-2 Delta variants were evolving under positive selection pressure.

Key words: Delta variants, FEL, MEME, mutations, positive selection, SARS-CoV-2, SLAC, selection pressure

Viruses mutate with due course of time but not all mutations are beneficial to virus. The mutational rate differs in viruses (Gralinski and Menachery, 2020; Tang *et al*., 2020). RNA viruses exhibit high mutational rate (Gralinski and Menachery, 2020; Tang *et al*., 2020) due to which they exhibit high adaptability to the environmental conditions. In RNA viruses the mutations arise due to the lack of proof reading mechanism of RNA polymerase and due to recombination between two viral lineages. Although most common mutations found are neutral, however, some mutations may affect the replication of virus and it's infectivity (Loewe and Hill, 2010; Alexander *et al.*, 2017; Gralinski and Menachery, 2020; Tang *et al*., 2020). Therefore, new emerging variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) poses threat to the global health.

Mutations in the variants of interest (VOIs), Variants of concern (VOCs) are associated with increased transmissibility (Plante *et al*., 2021) and reduced neutralization by antibody (Planas *et al*., 2021). The genome analysis of the SARS CoV-2 strains has revealed that the mutations are distributed heterogeneously in the genome of SARS-CoV-2 (Dearlove *et al*., 2020; van Dorp *et al*., 2020).

The spike protein of SARS CoV-2 has accumulated several mutations which have increased affinity to human ACE2 receptor (Luan *et al*., 2021). Moreover, the genome of SARS CoV-2 is subjected to selection pressure affecting the evolution of the virus. Therefore, the main aim of this study was to study the selection pressure analysis on the spike protein of SARS-CoV-2.

MATERIALS AND METHODS

Genome sequences retrieval and alignment

A total of 423, SARS-CoV-2 complete, and high coverage viral genome sequences of Delta variants were downloaded from Global Initiative on Sharing All Influenza Database (GISAID) platform and National Center for Biotechnology Information (NCBI). Only viruses affecting human hosts were selected and low-quality sequences having sequences more than 5% NNNs and other ambiguous characters were removed using BioEdit 7.2 software. Only full-length sequences including reference sequence MN908947.3 having more than 29000 nucleotides were included in the analysis.

Multiple sequence alignment and lineage prediction

Sequences of SARS-CoV-2 Delta variants from NCBI & GISAID were downloaded and consensus sequences were aligned with the reference sequence of the SARS-CoV-2 Wuhan-Hu-1 isolate (GenBank accession number MN908947.3) using Multiple Alignment using Fast Fourier Transform (MAFFT) (Katoh *et al*., 2002).

Mapping nucleotide substitutions in the genome and expressed proteins of SARS Co-V2

The mutations across the genome of aligned sequences were mapped with the reference sequence of the SARS-CoV-2 Wuhan-Hu-1 isolate (GenBank accession number MN908947.3) using an online software described previously (Mercatelli *et al*., 2021).

Selection pressure amongst the lineages of SARS-CoV2

The mutations across the genomes and the expressed proteins were analyzed and annotated separately. The genes coding for all proteins were analyzed separately after the removal of all internal stop codons. The selection pressure amongst the lineages of SARS CoV2 from June 2020 to December 2021 was assessed using Datamonkey Adaptive Evolution Server by Single-Likelihood Ancestor Counting (SLAC) model, Fixed Effect Likelihood (FEL) and Mixed Effects Model of Evolution (MEME). FEL and SLAC report both positively and negatively selected sites but MEME reports sites under positive selection. The results were expressed as dN-dS at a 0.1 level of significance. The mutated sites exhibiting selective pressure were annotated using Molecular Evolutionary Genetics Analysis (MEGA 11) software.

Prediction of effect of non-synonymous mutations under positive selective pressure on the functionalities of SARS-CoV-2

The effect of the non-synonymous mutations on the functionalities of SARS Co-V2 was predicted using the software Protein Variation Effect Analyser (PROVEAN) on the assumption that protein sequences evolutionarily conserved among living organisms have survived natural selection (Choi *et al*., 2012). The threshold value was subjected to a -2.5 default value below which the mutation was predicted to have a deleterious effect.

RESULTS AND DISCUSSION

Synonymous and non-synonymous mutations

Total 192 mutations were found to be present in the spike protein of 423 Delta variants out of which 67 were synonymous and 125 were non-synonymous mutations. The complete list of synonymous and non-synonymous mutations is given in Table 1 and Table 2, respectively.

Prediction of effect of non-synonymous mutations on protein functionality

The effect of non-synonymous substitution on the protein functionality was estimated using PROVEAN. Out of total 125 non-synonymous substitutions, nine mutations were found to have a deleterious effect on the functionality of Spike protein. Remaining 116 non-synonymous mutations were found to have neutral effect on the functionality of spike protein (Table 1). The SARS Co-V2 has got well adapted in the human population so it was anticipated that most of the mutations to be neutral (Duffy, 2018).

Selection pressure analysis on the spike protein

The spike protein sequences were subjected to selection pressure analysis using data monkey.org. FEL and SLAC address the question that which site in a gene is subjected to pervasive mutation that is consistently present across the entire phylogeny. MEME addresses which site in a gene is subjected to pervasive or episodic mutation. Therefore, all three procedures were adopted for selection pressure analysis. The details of the sites under positive and negative selection have been depicted in Table 3. In the SLAC model one site was found under positive selection and three sites under negative selection. In the FEL model two sites were found under positive selection and twenty three sites under negative selection. In the MEME model, two sites were found under positive selection. The positive selection sites

Table 1: List of non-synonymous mutations in the genome of Delta variants of SARS-CoV-2 circulating in India, their PROVEAN score and their effect on functionality of protein. The threshold value is -2.5 default value below which the mutation was predicted to have a deleterious effect.

Mutation	PROVEAN SCORE	Prediction Cut-off -2.5	Mutation	PROVEAN SCORE	Prediction Cut-off -2.5
A1020S	-0.811	Neutral	**L1141W**	-1.571	Neutral
A1078S	-0.359	Neutral	**L176F**	-0.175	Neutral
A222S	0.204	Neutral	**L249F**	0.526	Neutral
A222V	-0.096	Neutral	**L335F**	-0.29	Neutral
A243S	0.592	Neutral	**L452R**	0.559	Neutral
A263E	-1.048	Neutral	**L517H**	-0.482	Neutral
A27S	0.905	Neutral	**L54S**	-0.861	Neutral
A288S	-1.506	Neutral	**L5F**	-1.126	Neutral
A623S	0.3	Neutral	**L849I**	-1.471	Neutral
A623T	-0.635	Neutral	**L922F**	-2.286	Neutral
A701S	-0.12	Neutral	**M1237I**	-1.382	Neutral
A701V	0.597	Neutral	**M153K**	-0.122	Neutral
A831V	1.231	Neutral	**N1173K**	**-4.298**	**Deleterious**
A845S	0.924	Neutral	**N148D**	-0.458	Neutral
A879S	-0.361	Neutral	**N440S**	-0.49	Neutral
A892S	0.698	Neutral	**N460Y**	-0.299	Neutral
C1236F	**-4.061**	**Deleterious**	**N501Y**	-0.09	Neutral
C1247F	-1.723	Neutral	**N532S**	-1.081	Neutral
D1163Y	-2.465	Neutral	**P1162S**	**-2.722**	**Deleterious**
D1199Y	**-6.528**	**Deleterious**	**P26S**	1.34	Neutral
D1260N	-0.87	Neutral	**P621S**	1.593	Neutral
D215Y	-0.215	Neutral	**P681R**	0.741	Neutral
D228Y	0.226	Neutral	**P812R**	-1.019	Neutral
D574N	0.606	Neutral	**Q1071H**	-1.72	Neutral
D614G	0.598	Neutral	**Q1208H**	-1.327	Neutral
D796H	-0.263	Neutral	**Q14R**	0.258	Neutral
D80Y	-0.102	Neutral	**Q23R**	-0.239	Neutral
D950N	-1.631	Neutral	**Q613H**	-0.917	Neutral
E1262D	-0.59	Neutral	**Q677H**	0.002	Neutral
E156G	-1.631	Neutral	**R214L**	-0.393	Neutral
E471Q	0.445	Neutral	**R21T**	1.484	Neutral
E484K	0.128	Neutral	**R346K**	0.344	Neutral
E484Q	0.353	Neutral	**S221A**	-0.374	Neutral
E516Q	0.237	Neutral	**S477G**	0.294	Neutral
E868D	-0.226	Neutral	**S477I**	-1.31	Neutral
F157I	-0.174	Neutral	**S494L**	-0.308	Neutral
F515L	-0.203	Neutral	**S698L**	-0.345	Neutral
G1124V	0.478	Neutral	**S939C**	**-2.853**	**Deleterious**
G1167V	0.293	Neutral	**S939F**	**-3.094**	**Deleterious**
G1219C	-0.855	Neutral	**S943T**	-0.711	Neutral
G1219V	-0.54	Neutral	**S98F**	-1.636	Neutral
G142D	-0.277	Neutral	**T109A**	-2.024	Neutral
G75V	-0.282	Neutral	**T1100I**	**-3.154**	**Deleterious**
H1101N	-0.055	Neutral	**T19R**	-0.839	Neutral
H1101Y	1.26	Neutral	**T20I**	-0.373	Neutral
H1159R	-2.163	Neutral	**T274I**	-2.006	Neutral
I1179V	-0.795	Neutral	**T284I**	-0.717	Neutral
I410S	-1.527	Neutral	**T299I**	3.189	Neutral
I934V	-0.683	Neutral	**T478K**	-0.524	Neutral
K1191N	-1.707	Neutral	**T478R**	-0.839	Neutral
K417N	0.27	Neutral	**T547I**	**-3.72**	**Deleterious**
K77T	0.643	Neutral	**T732I**	-2.143	Neutral

Mutation	PROVEAN SCORE	Prediction Cut-off -2.5
T76I	-0.115	Neutral
T791I	-0.896	Neutral
T791S	-0.848	Neutral
T95I	-1.214	Neutral
V1033A	**-3.741**	**Deleterious**
V1104L	-0.604	Neutral
V1129L	-0.598	Neutral
V1133F	-1.666	Neutral
V1176F	-0.161	Neutral
V1228L	-0.457	Neutral
V143F	-0.622	Neutral
V16F	-0.797	Neutral
V483F	-0.372	Neutral
V615I	0.146	Neutral
V622F	-0.593	Neutral
V622I	-0.349	Neutral
V635A	0.209	Neutral
V6F	-0.739	Neutral
V70F	0.189	Neutral
W152K	-0.801	Neutral
W258L	-1.084	Neutral

Table 2: List of synonymous mutations in the genome of Delta variants of SARS-CoV-2 circulating in India

A1020A	F1148F	N122N	S514S
A263A	F275F	N606N	S591S
A263A	F306F	N679N	S691S
A411A	F456F	N777N	T385T
A475A	F55F	N824N	T716T
A694A	F888F	N925N	T778T
A701A	G1246G	P491P	T912T
C1240C	G413G	P728P	V1061V
C432C	G757G	L517L	V11V
D111D	H66H	Q134Q	V382V
D1146D	I1114I	Q14Q	V620V
D1168D	I410I	R682R	V83V
D1199D	I624I	R905R	W1217W
D294D	I692I	S1147S	Y505Y
D578D	I834I	S1175S	Y707Y
D737D	L335L	S221S	Y789Y
D848D	L7L	S477S	

were found in low number. Site numbers 142, 152 and 222 were found to be evolving under positive selection. Twenty three sites were found evolving under negative selection pressure.

In general, mutations imparting fitness to virus have been found within few weeks during the initial stage of the pandemic as exhibited by the spike protein amino acid change in D614G with high dN/dS ratio (Korber *et al.*, 2020; Zhan *et al.*, 2020).

Interestingly, the sites under RBD were not found under positive selection. In this study the sites 142, 152 and 222 which are other than RBD were found to be evolving under positive selection. It becomes paramount to determine the role of these sites in the virus transmission and pathogenesis.

CONCLUSION

The present study signifies that the delta variants of SARS CoV-2 has got well adapted in the human population as most of the non-synonymous mutations have neutral effect on protein functionality. Few sites, other than the RBD have been found to evolving under positive selection pressure, hence, it is paramount to conduct regular genomic surveillance.

REFERENCES

Alexander, H. K., Mayer, S. I., and Bonhoeffer, S. (2017). Population Heterogeneity in Mutation Rate Increases the Frequency of Higher-Order Mutants and Reduces Long-Term Mutational Load. *Mol. Biol. Evol.,* 34:419–436.

Choi, Y., Sims , G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One.*, 7. doi:10.1371/

Table 3: Sites under positive and negative selection pressure in the spike protein of Delta variants circulating in India. FEL and SLAC report both positively and negatively selected sites but MEME reports sites under positive selection. FEL and SLAC address the question that which site in a gene is subjected to pervasive mutation that is consistently present across the entire phylogeny. MEME addresses which site in a gene is subjected to pervasive or episodic mutation

Model	Sites under positive selection	Sites under negative selection
SLAC	142	294, 905, 1061
FEL	142, 222	66, 111, 134, 151, 294, 411, 456, 491, 514, 578, 591, 624, 679, 682, 716, 728,737, 888, 905, 912, 1061, 1147, 1175
MEME	142 and 152	Nil

journal.pone.0046688.

Dearlove, B., Lewitus, E., Bai, H., Li, Y., Reeves, D. B., Joyce, M. G., Scott, P. T., Amare, M., Vasan, F. S., Michael, N. L., Modjarrad, K., and Rolland, M. (2020). A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc. Natl. Acad. Sci., U. S. A.* 117:23652–23662. doi:10.1073/pnas. 2008281117.

van Dorp, L., Acman, M., Richard, D., Shaw, L. P., Ford, C. E., Ormond, L., Owen,C. J., Pang , J., Tan, C. C. S., Boshier, F. A. T., Ortiz, A. T., and Balloux, F. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.,* 83. doi:10.1016/j.meegid.2020.104351.

Duffy, S. (2018). Why are RNA virus mutation rates so damn high? *PLoS Biol.*, 16. doi:10.1371/journal.pbio.3000003.

Gralinski, L. E., and Menachery,V. D. (2020). Return of the coronavirus: 2019-nCoV. *Viruses,* 12. doi:10.3390/v12020135.

Katoh, K., K. Misawa, K. I. Kuma, and T. Miyata. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.,* 30:3059–3066. doi:10.1093/nar/gkf436.

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva,T. I., Angyal, A., Brown, R. L., Carrilero,L., Green, L. R., Groves, D. C., Johnson, K. J., Keeley, A. J., Lindsey, B. B., Parsons, P. J.,Raza, M., Rowland-Jones, S., Smith, N., Tucker, R. M., Wang, D.,Wyles, M. D., McDanal,C. , Perez, L. G., Tang, H., Moon-Walker, A. , Whelan, S. P., LaBranche, C. C., Saphire, E. O., and Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182:812-827.

Loewe, L., and Hill, W. G. (2010). The population genetics of mutations: Good, bad and indifferent. Philos. *Trans. R. Soc. B Biol. Sci.,* 365:1153–1167.

Luan, B., Wang, H., and Huynh, T. (2021). Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. *FEBS Lett.,* 595:1454–1461.

Mercatelli, D., Triboli, L., Fornasari, E., Ray, F., and Giorgi, F. M. (2021). Coronapp: A web application to annotate and monitor SARS-CoV-2 mutations. *J. Med. Virol.,* 93:3238–3245.

Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine F., Rajah, M. M., Planchais, C., Porrot, F., Robillard, N., Puech, J., Prot, M., Gallais, F., Gantner, P., Velay, A., Guen, J. Le, Kassis-Chikhani, Edriss, N., D., Belec, L.,Seve, A., Courtellemont, L., Pere, H., Hocqueloux, L., Fafi-Kremer, S., Prazuck, T., Mouquet, H., Bruel, T., Simon-Lorière, E., Rey, F. A., and Schwartz, O. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*, 596:276–280. doi:10.1038/s41586-021-03777-9.

Plante, J. A., Liu, Y., Liu, J.,Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfia, C. R., Mirchandani, D., Scharton, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Friberg, A. N., Menachery, V. D., Xie, X., Plante, K. S., Weaver, S. C., and Shi, P. Y. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 592:116–121.

Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., and Lu, J. (2020). On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.,* 7:1012–1023.

Zhan, X. Y., Zhang, Y., Zhou, X., Huang, K., Qian,Y., Leng, Y., Yan, L., Huang, B., and He.,Y. (2020). Molecular evolution of SARS-CoV-2 structural genes: Evidence of positive selection in spike glycoprotein. bioRxiv. doi:10.1101/2020.06.25.170688.

Received: March 22, 2022
Accepted: April 23, 2022