

## Coronavirus Pandemic

# Emergence of Novel SARS-CoV-2 variants in India: second wave

Rama Adiga<sup>1</sup>, Varun Nayak<sup>1</sup>

<sup>1</sup> Nitte (Deemed to be University), Nitte University Center for Science Education and Research (NUCSE), Mangalore, India

### Abstract

**Introduction:** Globally South-East Asia reported 40% of SARS-CoV-2 infected cases in the fourth week of April 2021. It continued to show an increase with India accounting for 50% of cases worldwide and 30% of global deaths. Genomic surveillance should continue at a rapid pace because of the continuously evolving nature of the virus. The time period of sample collection from the Global Initiative on Sharing All Influenza Data database was concurrent with the surge in new cases seen in the Indian subcontinent.

**Methodology:** 7,415 sequences were downloaded from Global Initiative on Sharing All Influenza Data between January and April 2021; out of which 4,411 were high coverage genome sequences and were considered for analysis. Phylogenetic analysis were carried out using Nextstrain.

**Results:** 21A or B.1.617 or delta was the most prevalent lineage in India accounting for 67.7% of the genomes. Next important clades were 20A, 20B and 20I accounting for 23.6%, 11.8% and 12.1% respectively collected between January 2021 and April 2021. The remaining sequences were assigned to clade 20H, 20J, 20D, 20C, 20G, 20E, 19A and 19B. The spike mutation frequencies of L452R, E484Q and P681R in Indian state of Maharashtra were 62.4%, 66.5% and 61.5% respectively. Two unique N-terminal domain deletion of spike protein were found at position 67 and 68.

**Conclusions:** The phylogenomics of the delta variant or 21A emerged in neighboring Asian countries of Thailand, Bangladesh, Indonesia and Japan. We analyzed the SARS-CoV-2 genomes from India for mutation characterization of the spike glycoprotein and the nucleocapsid protein.

**Key words:** SARS-CoV-2; molecular mutation; delta; India.

*J Infect Dev Ctries* 2021; 15(11):1578-1583. doi:10.3855/jidc.15484

(Received 16 June 2021 – Accepted 19 September 2021)

Copyright © 2021 Adiga *et al.* This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Introduction

The extremely contagious severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread all over the globe since its emergence in China in 2019. Globally the pandemic is showing no signs of ebbing with increasing death rate since February 2021 [1,2]. A sharp rise in cases were observed in South-East Asia, West Pacific and Eastern Mediterranean regions with accelerated death rate. Globally the mortality which took nine months to reach 1 million last year, reached 3 million deaths in just three months in 2021. The number of weekly new cases globally showed an increasing trend with 5.2 million reported during the third week of April 2021 [2]. The largest increase was reported by the South-East Asian region mainly India.

The number of new cases reported from India, however, which was 1.5 million cases in the third week of April rose to 2.7 million new cases in the first week of May [2]. The mortality was also at an all-time high of 90,000 deaths in the week.

Comparative genomic studies have made possible the study of SARS-Cov-2 genomic sequences through

the open access repositories such as Global Initiative on Sharing All Influenza Data (GISAID) [3].

The nomenclature system of GISAID classified the SARS-CoV-2 genome into seven major clades. The clade belonging to SARS-CoV-2 strains include L, S, V, G, GH, GR and GV. The GR was most prevalent worldwide followed by GV and GH [4].

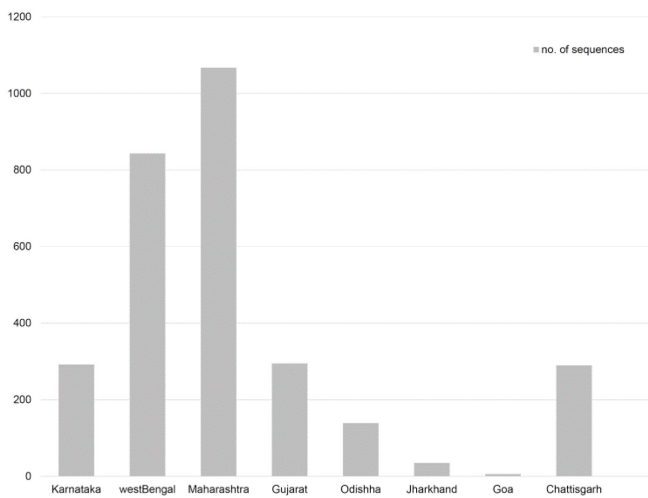
The pattern of SARS-CoV2 global spread can be understood by the Phylogenetic Assignment of Named Global Outbreak LINEages (PANGOLIN) tool [5] explaining detailed lineages. Nextclade assignment supports the PANGOLIN lineage system characterizing the mutations into nine clades [6]. New major clades are named once the frequency of a clade exceeds 20% in a representative global sample and that clade differs in at least two positions from its parent clade. The Nextstrain clades 19A, 19B, 20A, 20B, 20C and 20I are used for annotation of strains. Variants belonging to clade 20I, 20H and 20J originated in U.K., South Africa and Brazil respectively were synonymous with SARS-CoV-2 virus lineage B1.1.7 [5], B.1.351 [7] and P1.

Due to low sequencing carried out and the slow rate of genomic data availability in databases, tracking of SARS-CoV-2 variants circulating in Indian states has been limited. The current study highlights the major lineages and the important mutations and their distribution in the Indian subcontinent.

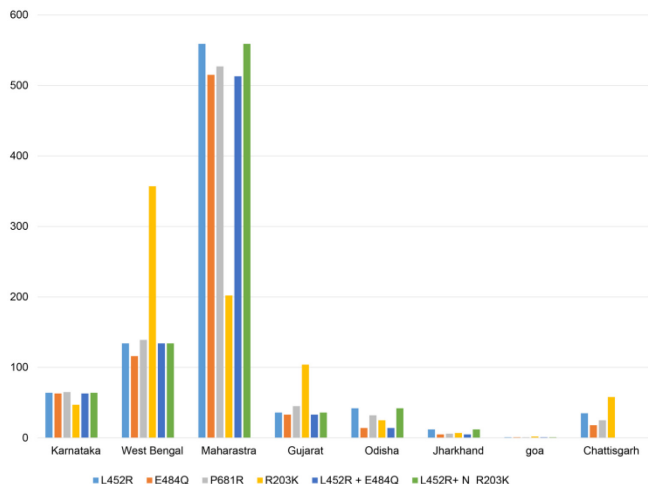
**Methodology**

The Indian SARS-CoV-2 genomes deposited in GISAID [3] during the period 1<sup>st</sup> January - 25<sup>th</sup> May 2021 were selected and their metadata were retrieved for the analysis. A total of 7415 sequences were available from which 4,411 genome sequences with

**Figure 1.** Sequences in GISAID. The number of sequences collected from GISAID and analyzed from eight states of India from January 2021 to 26th April 2021.



**Figure 2.** The western state of Maharashtra showing high prevalence of L452R, E484Q and P681R substitution are shown in the graph depicting the frequency with respect to the total sequences available in the database for the states during the period.



high coverage were considered for the analysis. We considered the sequences with high coverage for amino acid substitution analysis and phylogenetic clustering. Furthermore, for molecular analysis of sequences from non-Indian origin, high quality genome sequences with full coverage were retrieved from GISAID.

Multiple Alignment using Fast Fourier Transform (MAFFT) was used to perform the multiple sequence alignment [8] implemented via an Augur phylodynamic analysis pipeline. Afterwards, the open-source interactive tool Auspice for visualizing genomic data implemented via Nextstrain was used.

Nextclade beta [6] was used to check the quality of sequence and to generate the nextclade designated tree. The joint global effort by Nextstrain [6,9] which is an open-source project was initiated in 2018 and integrates the genomic data of SARS-CoV-2. Amino acid mutation analysis was done using CoVserver mutation analysis with reference to hcov19/Wuhan strain.

**Results**

The GISAID sequences of Indian SARS-CoV-2 from collection date and submission date January 2021 to 25<sup>th</sup> May 2021 were downloaded. The available 4,411 sequences having full coverage were analysed using Nextstrain analysis which identified four main clades: 67.7% (n = 2589) of the genomes belonged to Nextstrain clade 21A, 23.6% (n = 902) belonged to 20A, 11.8% (n = 424) belonged to 20B while 12.1% (n = 434) belonged to the variant 20I. The 0.01% (n = 63) of remaining sequences were assigned to clade 20H. Sequences comprising less than 0.01% were assigned to 20J (n = 1), 20D (n = 5), 20C (n = 1), 20G (n = 2), 20E (EU1) (n = 1), 19A (n = 5), and 19B (n = 4). There were twelve Nextstrain clades from sequences of India namely 21A, 20C, 20D, 20E, 20G, 20I, 20H, 20J, 19A, 19B, 20A and 20B as accessed on 25<sup>th</sup> May 2021.

The sequences available in GISAID from eight states in India had variable number of genome sequences deposited. The proportion of sequences available in the database are depicted (Figure 1) which were analysed for the geographical distribution of signature mutations in spike protein L452R, E484Q, P681R from West to East of India (Figure 1, Figure 2). The nucleoprotein substitution R203K showed a prevalence of only 0.18% (n = 202) in the western state of Maharashtra as compared to the eastern state of West-Bengal which has a prevalence of 42.3% (n = 357) (Figure 2). The 203K substitution were predominantly found in 20I and 20B.

The mutations of spike protein H69del located in the N-terminal domain (NTD) were found with a

**Table 1.** Mutations/deletions Y144del, E780Q, H69del in amino acid in positions indicated from various continents as compared to India.

Mutation	India	USA	Europe	Oceania	South America
Y144del	129	92	162	2	3
E780Q	6	1,623	38	4	16
H69del	216	54,635	369,419	420	891
total genomes	3,084	214,952	458,053	1,023	5,543

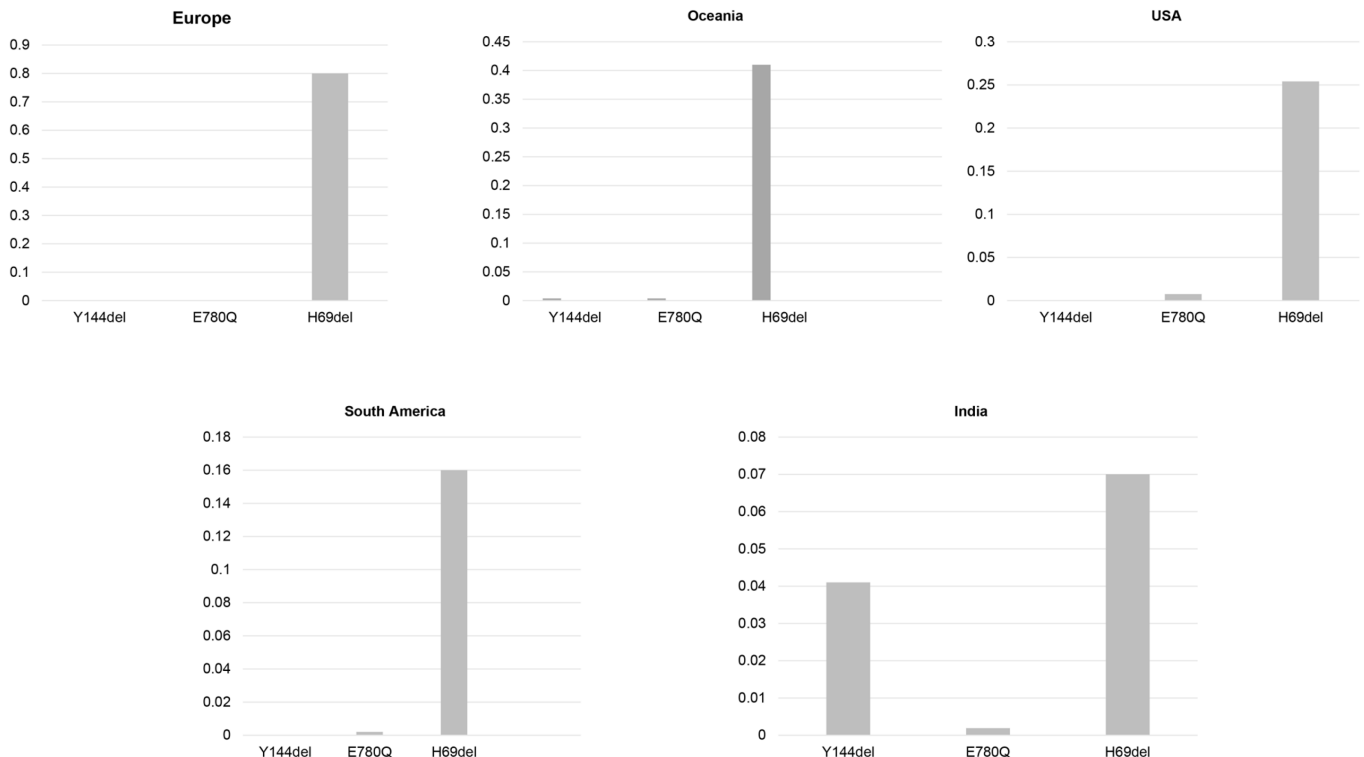
frequency of 80%, 40%, 25% and 16% from Europe, Oceania, U.S.A. and South America respectively. In India it was present at a frequency of 7% (n = 216) of genomic sequences between January and April 2021 (Figure 3, Table 1). The H69del was found in higher abundance in 20I within the twelve clades and found to be emerging in 20A, 20B, 20C,20D, 19B and 21A. The other significant mutations observed in N-terminal domain (NTD) of spike protein were Y144del/Y145del (n = 129), V70del (n = 215), A67del (n = 5) and I68del (n = 3). The last two substitutions were not reported from any other country. The NTD deletion region are depicted in sequence alignment (Table 2).

The Nextstrain clade 21A or B.1.617.2 or delta was the most prevalent lineage among the sequences reported from India. The other countries where the 21A lineage has been identified were Bangladesh, Japan, Malaysia, Thailand, Singapore.

**Discussion**

Although mortality is variable in geographic regions, the sudden surge in cases in South-east Asia largely contributed by India is alarming. Scientist around the globe have realized the necessity of rapid large scale genome analysis of the virus SARS-CoV-2. The classification the genomes of the second wave is essential for surveillance. With the recent nomenclature of 21A or B.1.617.2 as “delta” by the WHO risk assessment team of SARS-CoV2 variant on 11<sup>th</sup> May 2021, there is an urgent need to characterize mutations which are crucial for high transmissibility. B.1.617 sublineages have high transmission rates and the dip in proportion of B.1.617.1 as reported in April in comparison to B.1.617.2 was not observed in our study. The B.1.617.1 or kappa and B.1.617.2 or delta accounted for 19.6% (n = 2,105) and 29.6% (n = 3,180) of sequenced samples from India respectively until June 8<sup>th</sup> 2021.

**Figure 3.** Percentage of mutation observed for Y144del, E780Q, H69del from continents analyzed for the time period (Jan 2021 to April 2021). The graph depicts the frequency with respect to the total genomic sequences for Oceania, Europe, South America, USA and India.



**Table 2.** The alignment shows four A67X sequences and five I68X sequences from India in the deletion region of NTD. The deletions are at 69 and 70 which are labelled H69X and V70X where ‘X’ denotes deletion of amino acid.

West Bengal	W	F	H	V	I	X	X	S	G
Karnataka	W	F	H	X	X	X	X	S	G
Karnataka	W	F	H	A	I	H	V	S	G
Orissa	W	F	H	V	X	X	X	S	G
Telangana	W	F	H	X	X	X	X	S	G
Telangana	W	F	H	X	X	X	X	S	G
Telangana	W	F	H	X	X	X	X	S	G
Gujarat	W	F	H	A	I	X	X	S	G
Gujarat	W	F	H	A	I	X	X	S	G

The sequences have GISAID accession no from Karnataka (EPI ISL 1595876,EPI ISL 1595879), Orissa (EPI ISL 1703836),Telangana (EPI ISL 1838789,EPI ISL 1838837,EPI ISL 1838857) and Gujarat (EPI ISL 1841277,EPI ISL 1841278).

The signature mutations in spike protein receptor binding domain involved in increased ACE2 binding have been studied [10]. Substitutions at position 452, 484 and 681 of receptor binding domain have been reported in other globally circulating lineages recently. The distribution of the substitutions at the above-mentioned positions from India were studied (Figure 3). The western state of Maharashtra showed high prevalence of L452R, E484Q and P681R substitution at 62.4% (n = 559), 66.5% (n = 774), 61.5% (n = 527) as compared to eastern states of West Bengal which had frequency of 14.9% (n = 134),14.9% (n = 116) or 16.2% (n = 139) respectively (Table 3).

Of particular mention is the nucleocapsid protein (N) substitution R203K which showed lower frequency of 23% (n = 202) in Maharashtra as compared to West Bengal which had a frequency of 40.7% (n = 357). The N protein is involved in virus assembly and replication. It plays pivotal role in targeting innate host immunity and IFN inhibition [11]. The viral strategy is to shield the RNA from being detected by host immune RNA sensors [12]. Zhao *et al*, showed that mutation at 203 had a greater effect on suppression of IFN bringing about immune evasion. It may be hypothesized that the substitution of arginine with lysine may also be involved in immune escape. Overall, the mutation warrants further investigation to understand the effect of alteration. The other mutations found in our analysis were D63G, D377Y and Q384H, R191H and S180I in

the N protein. The N protein has also been a target for drug discovery [12] and the mutation studies will help in bringing many antivirals to the market.

Antibodies have been developed with neutralizing activity towards NTD which may be of therapeutic potential [13-15]. Deletions in NTD of spike have been reported to bring about reduced sensitivity to neutralizing antibodies [16]. NTD deletions abolish the binding of antibody and have been termed as ‘escape variants’ and resist neutralization by antibody [17-19].

The deletion at 69–70 and 144/145 found in B.1.1.7 lineage have been studied (Figure 3). The deletion region identified by us (Table 2) at position A67X and I68X in a total of eight sequences comprised of the N2 loop of NTD epitope recognition region [10]. Any deletion in the region may bring about loss of conformational stability. Surveillance of deletion/insertion region is important in epidemiology for monitoring vaccine efficacy and primer-based detection methods. The analysis of variant of concern 20H or B.1.351 first identified in South Africa and 20I or B.1.1.7 lineage which first emerged in UK have been studied with respect to important mutations. The appearance of clade and naming of 21A by WHO as delta and kappa variant has highlighted the need for further studies.

The frequency of the variant subsequently labelled 21A [10] or delta variant or B.1.617.2 in India increased from 0% at the beginning of March 2021 to > 33% of

**Table 3.** Percent mutations in amino acid in positions indicated from various states of India.

State	L452R (%)	L452R + E484Q (%)	E484Q (%)	L452R+ N_R203K (%)	N_R203K (%)	P681R (%)
	(N) = 896	(N) = 896	(N) = 774	(N) =896	(N) = 876	(N) = 857
Karnataka	21.9	21	0.21	21.9	16	22
West Bengal	15.8	15	0.13	15	42.3	16.4
Maharashtra	52	48	0.48	52.3	18	49.3
Gujarat	12.2	0.11	11	12.2	35	15.2
Odisha	30	0.1	10	30	17.9	23
Jharkhand	34	0.14	14	34.2	20	18
Goa	16.6	0.166	16.6	16	33	16
Chhattisgarh	12	0	6	0	20	8.6



sequenced cases by the end of May 2021, with a similar trajectory to the surge of COVID-19 cases observed during the same time.

The changing nomenclature of clade by Nextstrain was fruitful in tracking variants as compared to previous analysis conducted a month ago as more sequences were deposited in GISAID.

## Conclusions

Focusing on the phylogenetic classification and the clade assignment it was clear that sequences from India classified as 21A or B.1.617.1/B.1.617.2 may probably be the cause of high transmission. Other variants identified were twelve in number, out of which the variant of concern were 20H, 20I and 20J. B.1.617 was named as a global Variant of Concern on 11 May 2021 and the sublineage 1 and 2 as kappa and delta respectively. Both sub lineages are associated with increased transmissibility. The tracking of mutation is also of significance in development of antigen detection kit and drug discovery.

## Data Availability Statement

The data used in the study were downloaded and can be found on the official GISAID repository. The Accession ID of the sequences used in this study can be found in Supplementary Table 1.

## Acknowledgements

The author wishes to thank Dr. Anirban Chakraborty, Director, Nitte University Centre for Science Education and Research (NUCSER), Prof. Dr. Indrani Karunasagar Director (Projects) and the Management of Nitte (Deemed to be University), Deralakatte, Mangalore, Karnataka, India for the support in establishing the center, providing facilities and continuous encouragement in research, including the present work.

## References

1. World Health Organization (WHO) (2020) Coronavirus disease 2019 (COVID-19) Situation Report - 38. NUCSER (2021) Emergence of Novel SARS-CoV-2 variants in India. Available: [https://www.who.int/docs/default-source/coronavirus/situation-reports/20200227-sitrep-38-covid-19.pdf?sfvrsn=9f98940c\\_2](https://www.who.int/docs/default-source/coronavirus/situation-reports/20200227-sitrep-38-covid-19.pdf?sfvrsn=9f98940c_2). Accessed: 28 February 2020.
2. World Health Organization (WHO) (2021) Tracking SARS-CoV-2 variants. NUCSER (2021) Available: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>. Accessed: 20 January 2021
3. Shu Y, McCauley J (2017) GISAID: Global initiative on sharing all influenza data from vision to reality. *Eurosurveillance* 22: 30494 NUCSER (2021) Emergence of Novel SARS-CoV-2 variants in India Available: <https://www.gisaid.org/>. Accessed: 26 April 2021 and 26 May 2021
4. Hamed SM, Elkhatib WF, Khairalla AS, Noreddin AM (2021) Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. *Sci Rep* 11: 8435.
5. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG (2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5: 1403–1407.
6. Updated Nextstrain SARS-CoV-2 Clade Naming Strategy-SARS-CoV-2 Coronavirus/nCoV-2019 Genomic Epidemiology (NUCSER) (2021) Emergence of Novel SARS-CoV-2 variants in India. Available: <https://virological.org/t/updated-nextstrain-sars-cov-2-clade-naming-strategy/581>. Accessed: 15 February 2021 and 31 May 2021.
7. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao NY, Korsman S, Davies MA, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T (2021) Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592: 438-443.
8. Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software version 7: improvements in performance and usability. *Mol Biol Evol* 30 : 772–780.
9. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34: 4121-4123.
10. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A; COVID-19 Genomics UK (COG-UK) Consortium, Peacock SJ, Robertson DL (2021) SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 19: 409–424.
11. Mu J, Fang Y, Yang Q, Shu T, Wang A, Huang M, Jin L, Deng F, Qiu Y, Zhou X (2020) SARS-CoV-2 N protein antagonizes type I interferon signaling by suppressing phosphorylation and nuclear translocation of STAT1 and STAT2. *Cell Discov* 6: 65.
12. Zhao M, Yu Y, Sun LM, Xing JQ, Li T, Zhu Y, Wang M, Yu Y, Xue W, Xia T, Cai H, Han QY, Yin X, Li WH, Li AL, Cui

- J, Yuan Z, Zhang R, Zhou T, Zhang XM, Li T (2021) GCG inhibits SARS-CoV-2 replication by disrupting the liquid phase condensation of its nucleocapsid protein. *Nat Commun* 12: 2114.
13. McCallum M, Marco A, Lempp F, Tortorici MA, Pinto D, Walls AC, Beltramello M, Chen A, Liu Z, Zatta F, Zepeda S, di Iulio J, Bowen JE, Montiel-Ruiz M, Zhou J, Rosen LE, Bianchi S, Guarino B, Fregni CS, Abdelnabi R, Caroline Foo SY, Rothlauf PW, Bloyet LM, Benigni F, Camerini E, Neyts J, Riva A, Snell G, Telenti A, Whelan SPJ, Virgin HW, Corti D, Pizzuto MS, Veesler D (2021) N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* 184: 2332-2347.e16.
  14. Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, Zhang Z, Fan P, Dong Y, Yang Y, Chen Z, Guo Y, Zhang J, Li Y, Song X, Chen Y, Xia L, Fu L, Hou L, Xu J, Yu C, Li J, Zhou Q, Chen WA (2020) A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* 369: 650-655.
  15. Suryadevara N, Shrihari S, Gilchuk P, VanBlargan LA, Binshtein E, Zost SJ, Nargi RS, Sutton RE, Winkler ES, Chen EC, Fouch ME, Davidson E, Doranz BJ, Carnahan RH, Thackray LB, Diamond MS, Crowe JE (2021) Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell* 184: 2316-2331.e15.
  16. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, Duprex WP (2021) Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* 371: 1139-1142.
  17. Ribes M, Chaccour C, Moncunill G (2021) Adapt or perish: SARS-CoV-2 antibody escape variants defined by deletions in the Spike N-terminal Domain. *Signal Transduct Target Ther* 6: 164.
  18. Wang R, Zhang Q, Ge J, Ren W, Zhang R, Lan J, Ju B, Su B, Yu F, Chen P, Liao H, Feng Y, Li X, Shi X, Zhang Z, Zhang F, Ding Q, Zhang T, Wang X, Zhang L (2021) Analysis of SARS-CoV-2 variant mutations reveals neutralization escape mechanisms and the ability to use ACE2 receptors from additional species. *Immunity* 54: 1611-1621.e5.
  19. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, Lambson BE, Oliveira TD, Vermeulen M, Berg KV, Rossouw T, Boswell M, Ueckermann V, Meiring S, von Gottberg A, Cohen C, Morris L, Bhiman JN, Moore PL (2021) Sars-CoV-2 501y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med* 27: 622.

### Corresponding author

Rama Adiga PhD

Nitte (Deemed to be University)

Nitte University Center for Science Education and Research (NUCSER)

Division of Bioinformatics and Computational Genomics, Mangalore, India 575018

Phone: 09880438741

Fax: +91-824-2204162

Email: rama\_adiga@nitte.edu.in

**Conflict of interests:** No conflict of interests is declared.

**Annex – Supplementary Items****Supplementary Table 1.** Submission acknowledgement of sequences to GISAID.

<b>Accession ID</b>	<b>Originating Laboratory</b>	<b>Submitting Laboratory</b>	<b>Authors</b>
see above	INSACOG-WB	National Institute of Biomedical Genomics	Arindam Maitra, Bhaswati Bandyopadhyay, Nidhan Kumar Biswas, Tamal Ghosh, Sreedhar Chinnaswamy, Ajay Chakraborti, Saumitra Das
EPI_ISL_1360324	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360325, EPI_ISL_1360326	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Sofia Banu, Lamuk Zaveri, Ara Sreenivas, Shreekant Verma, Onkar Kulkarni, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360327, EPI_ISL_1360328	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Onkar Kulkarni, Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360329	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Onkar Kulkarni, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360330	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Onkar Kulkarni, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360331, EPI_ISL_1360332	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati

EPI_ISL_1360333	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Sofia Banu, Lamuk Zaveri, Ara Sreenivas, Shreekant Verma, Onkar Kulkarni, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360335, EPI_ISL_1360336	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Onkar Kulkarni, Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360338	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Onkar Kulkarni, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360339, EPI_ISL_1360340	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360342	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Sofia Banu, Lamuk Zaveri, Ara Sreenivas, Shreekant Verma, Onkar Kulkarni, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360343, EPI_ISL_1360344	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Onkar Kulkarni, Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360345	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Onkar Kulkarni, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati



EPI_ISL_1360346	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Onkar Kulkarni, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360348	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360349, EPI_ISL_1360350	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Sofia Banu, Lamuk Zaveri, Ara Sreenivas, Shreekant Verma, Onkar Kulkarni, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360351, EPI_ISL_1360352	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Onkar Kulkarni, Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360353	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Onkar Kulkarni, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360354, EPI_ISL_1360355	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360356	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Sofia Banu, Lamuk Zaveri, Ara Sreenivas, Shreekant Verma, Onkar Kulkarni, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati

EPI_ISL_1360359	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Onkar Kulkarni, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360360	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Onkar Kulkarni, Sofia Banu, Shreekant Verma,Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360361, EPI_ISL_1360362	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360364	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Sofia Banu, Lamuk Zaveri, Ara Sreenivas, Shreekant Verma, Onkar Kulkarni, Amareshwar Vodapalli ,Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360365, EPI_ISL_1360366	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Onkar Kulkarni, Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360367	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Onkar Kulkarni, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360368	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Onkar Kulkarni, Sofia Banu, Shreekant Verma,Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati

EPI_ISL_1360369, EPI_ISL_1360370	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360371, EPI_ISL_1360372	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Sofia Banu, Lamuk Zaveri, Ara Sreenivas, Shreekant Verma, Onkar Kulkarni, Amareshwar Vodapalli ,Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360373	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Onkar Kulkarni, Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360375	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Onkar Kulkarni, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360377, EPI_ISL_1360378	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360381, EPI_ISL_1360382	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Onkar Kulkarni, Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati
EPI_ISL_1360383	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Sofia Banu, Onkar Kulkarni, Shreekant Verma, Amareshwar Vodapalli , Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra,Divya Tej Sowpati

EPI_ISL_1360384	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Lamuk Zaveri, Ara Sreenivas, Onkar Kulkarni, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360385	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Ara Sreenivas, Onkar Kulkarni, Lamuk Zaveri, Sofia Banu, Shreekant Verma, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1360387, EPI_ISL_1360388	CSIR-National Environmental Engineering Research Institute	CSIR-Centre for Cellular and Molecular Biology - INSACOG	Sofia Banu, Lamuk Zaveri, Ara Sreenivas, Shreekant Verma, Onkar Kulkarni, Amareshwar Vodapalli, Viswagithe S L, B Himasri, Sharath Chandra Thota, Karthik Bharadwaj Tallapaka, Krishna Khairnar, Rakesh K Mishra, Divya Tej Sowpati
EPI_ISL_1372009	Patna Medical College, Patna	Institute of Life Sciences - INSACOG	Sunil K. Raghav, Safal Walia, Arup Ghosh, Atimukta Jha, Amol M. Kanampalliwar, Shifu Aggarwal, Rupesh Dash, Rajeeb Swain, Punit Prasad, INSACOG Consortium, Ajay Parida
EPI_ISL_1372010	Immunogenomics lab, Institute of Life Sciences, Bhubaneswar	Institute of Life Sciences - INSACOG	Sunil K. Raghav, Safal Walia, Arup Ghosh, Atimukta Jha, Amol M. Kanampalliwar, Shifu Aggarwal, Rupesh Dash, Rajeeb Swain, Punit Prasad, INSACOG Consortium, Ajay Parida
EPI_ISL_1372011	Capital hospital, Bhubaneswar	Institute of Life Sciences - INSACOG	Sunil K. Raghav, Safal Walia, Arup Ghosh, Atimukta Jha, Amol M. Kanampalliwar, Shifu Aggarwal, Rupesh Dash, Rajeeb Swain, Punit Prasad, INSACOG Consortium, Ajay Parida

We gratefully acknowledge the following authors from the originating laboratories responsible for obtaining the specimens, as well as the submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based. All Submitters of data may be contacted directly via [www.gisaid.org](http://www.gisaid.org). Authors are sorted alphabetically.