

Short Communication

A comprehensive report on severe acute respiratory syndrome coronavirus-2 lineages in Puducherry from 2020 to 2023

Ferdinamarie Sharmila Philomenadin, Reshma Rajendran, Vimal R. Ratchagadasse, Rahul Dhodapkar*

Department of Microbiology, JIPMER, Puducherry, India

Received: 17 June 2025

Accepted: 13 October 2025

*Correspondence:

Dr. Rahul Dhodapkar,

E-mail: rahuldhodapkar@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

This study presents the first comprehensive molecular epidemiology of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) in Puducherry, India, from 2020 to 2023, analyzing 1,840 sequences from S gene and whole-genome sequencing. Initial waves saw clade G dominance, followed by delta (clade GK) and Omicron (clade GRA) variants. High mutation density was observed in the spike protein's N-terminal domain, including key mutations like D614G, E484 variants, N501Y, and P681H. The beta-associated A701V mutation was also detected. Findings align with global trends, showing variant transitions and spike mutations influencing infectivity and immune evasion. This work enhances understanding of regional SARS-CoV-2 evolution and variant dynamics.

Keywords: SARS-CoV-2, India, Evolution, Molecular epidemiology, Spike gene, Sequencing

INTRODUCTION

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) which emerged in late 2019, in Wuhan, China has caused a pandemic of acute respiratory disease named coronavirus disease 2019 (COVID-19).¹ Over 572 million confirmed cases are there worldwide with 6 million deaths.² The first case in India was reported in Kerala, Thrissur on 30 January 2020. Currently, in India, there are more than 45 million cases reported, with deaths exceeding 5 lakhs. About 65.4% of the world population is fully vaccinated with primary series worldwide and in India it is about 68.9%.³ So far, the SARS-CoV-2 variants observed worldwide include alpha, beta, gamma, delta, epsilon, kappa, iota, lambda, and mu. Currently, there are no variants of concern circulating worldwide but 1 variant of interest (VOIs) in circulation is JN.1, a descendant of BA.2.86.⁴

Puducherry, an earlier French colonial town lies on the South East coast of India and is a major recreational hub attracting more tourists for its pristine beaches. Not many reports are available exclusively on the status of COVID-

19 from this Union territory. A recent study reported the mean case fatality rate around 2.6 in phase I, and 1.7 in both resting and phase II stages.⁵

Another study reported that the seropositivity rate increased dramatically and was 20-fold higher than the number of diagnosed cases.⁶ However, there are no studies on the molecular characterization of the virus circulating during these periods and hence the current study was designed to address this issue.

SARS-CoV-2 shares 79% genome sequence identity with SARS-CoV and 50% with MERS CoV. The 6 functional ORFs are arranged as follows, in the order from 5' to 3': replicase (ORF1a/ ORF1b), spike(S), envelope (E), membrane (M) and nucleocapsid (N).

The spike protein of SARS COV2 involved in virus attachment and entry into cells is one of the most variable genes to be encoded. Mutations in the spike surface glycoprotein may induce some conformational changes, which probably led to the changing antigenicity.⁷ Hence

variations occurring in the spike glycoprotein region have received more clinical and scientific attention.

METHODS

This study was designed to record the S gene mutations accumulated over the pandemic from patients visiting Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), a tertiary-care hospital located in Puducherry. The study was approved by the Institute Ethics Committee for Human Studies (IEC No JIP/ IEC/ 2021/069). Data taken for this study has been derived from two different methods and timelines.

Method 1 sequencing of S gene

Study participants comprised of OP or IP patients positive for SARS-CoV-2 Realtime PCR from May 2020 to May 2021. Briefly, SARS-CoV-2 RNA was extracted using magnetic bead extraction from lab-confirmed positive cases of SARS-CoV-2.

Reverse transcription PCR was done using overlapping primers to amplify the S gene (base pair: 21563-25384) and sequenced by the Sanger sequencing (ABI3500) technique. The phylogenetic tree was constructed to identify the clades of the 65 samples along with reference sequences using Mega v11.⁸

Method 2 whole genome sequencing

Around December 2020, the Government of India launched an initiative to track the SARS-CoV2 variants by whole genome sequencing. A consortium was formed for this purpose and samples from our region were also sequenced and submitted in GISAID. The earliest submission starts in February 2021. As of January 2024, there are about 1909 sequences available from this region.

We have included these two data to get a comprehensive idea of the SARS-CoV-2 lineages in Pondicherry from 2020 to 2023. The final analysis included 1840 sequences, 64 from method one and 1776 from method 2.

RESULTS

SARS-CoV-2 clade distribution, based on GISAID classification, varied across these waves: clade G dominated the first wave (May 2020–March 2021), clade GK (delta) was prevalent during the second wave (April–November 2021), and clade GRA (Omicron) during the third wave (January 2022–May 2023) (Figure 1). Table 1 describes the distribution of clades from May 2020 to May 2023.

Mutation analysis revealed a total of 432 variations distributed across the S gene (Table 2). Frequency of occurrence of variations was higher in S1 subunit (n=294) with more variations in the NTD region (n=173) followed

by RBD (n=88). In the S2 subunit more variations were observed in the heptapeptide repeat sequence 1 (n=17).

Table 1: Distribution of SARS COV-2 clades from May 2020 to May 2023).

Month	G	GH	GK	GR	GRA	O	S
May-20	6	1	0	3	0	0	0
Jun-20	8	2	0	0	0	0	0
Jul-20	9	0	0	1	0	0	0
Aug-20	6	1	0	1	0	0	0
Sep-20	7	0	0	2	0	0	0
Oct-20	8	0	0	2	0	0	0
Nov-20	11	0	0	3	0	0	0
Dec-20	8	1	1	0	0	0	0
Jan-21	9	0	1	0	0	0	0
Feb-21	3	5	0	3	0	0	0
Mar-21	5	4	0	18	0	0	1
Apr-21	6	2	144	28	0	0	0
May-21	2	2	382	6	0	0	0
Jul-21	0	0	40	0	0	0	0
Aug-21	0	0	72	0	0	1	0
Oct-21	0	0	40	0	0	0	0
Nov-21	0	0	28	0	0	0	0
Dec-21	0	0	22	0	8	1	0
Jan-22	0	0	0	1	130	0	0
Feb-22	0	0	0	0	274	4	0
Jul-22	0	0	0	0	130	0	0
Aug-22	0	0	0	0	86	0	0
Sep-22	0	0	0	0	192	0	0
Oct-22	0	0	0	7	95	0	0
Nov-22	0	0	0	0	9	0	0
Feb-23	0	0	0	0	2	0	0
Mar-23	0	0	0	0	6	0	0
Apr-23	0	0	0	0	34	0	0
May-23	0	0	0	0	1	0	0
Total	88	18	730	75	967	6	1

Table 2: Distribution of variations across s gene.

Region and sub-region	Amino-acid positions	No of variations	Mutation density
Signal peptide	1-13	6	0.46
S1 subunit	14-685	294	0.44
N terminal	14-305	173	0.59
RBD	319-541	88	0.39
Furin cleavage sequence	680-685	2	0.33
S2 subunit	686-1273	112	0.19
Fusion peptide	788-806	3	0.16
Heptapeptide repeat sequence 1	912-984	17	0.23
Heptapeptide repeat sequence 2	1163-1213	7	0.14
Trans membrane domain	1213-1237	7	0.28
Cytoplasmic domain	1237-1273	7	0.19

The frequency of the variations during the study period ranged from as low as 0.001 to as high as 0.995, with NTD region having the maximum mutation density (0.59). Variations with frequency ≥ 0.4 includes S1- T19I, L24del, P25del, P26del, A27S, G142D, V213G, S371F, S373P,

S375F, T376A, D405N, R408S, K417N, N440K, L452R, S477N, T478K, E484A, Q498R, N501Y, Y505H, D614G, H655Y, N679K, Furin cleavage sequence- P681H, and S2-N764K, D796Y, Q954H, N969K (Table 3).

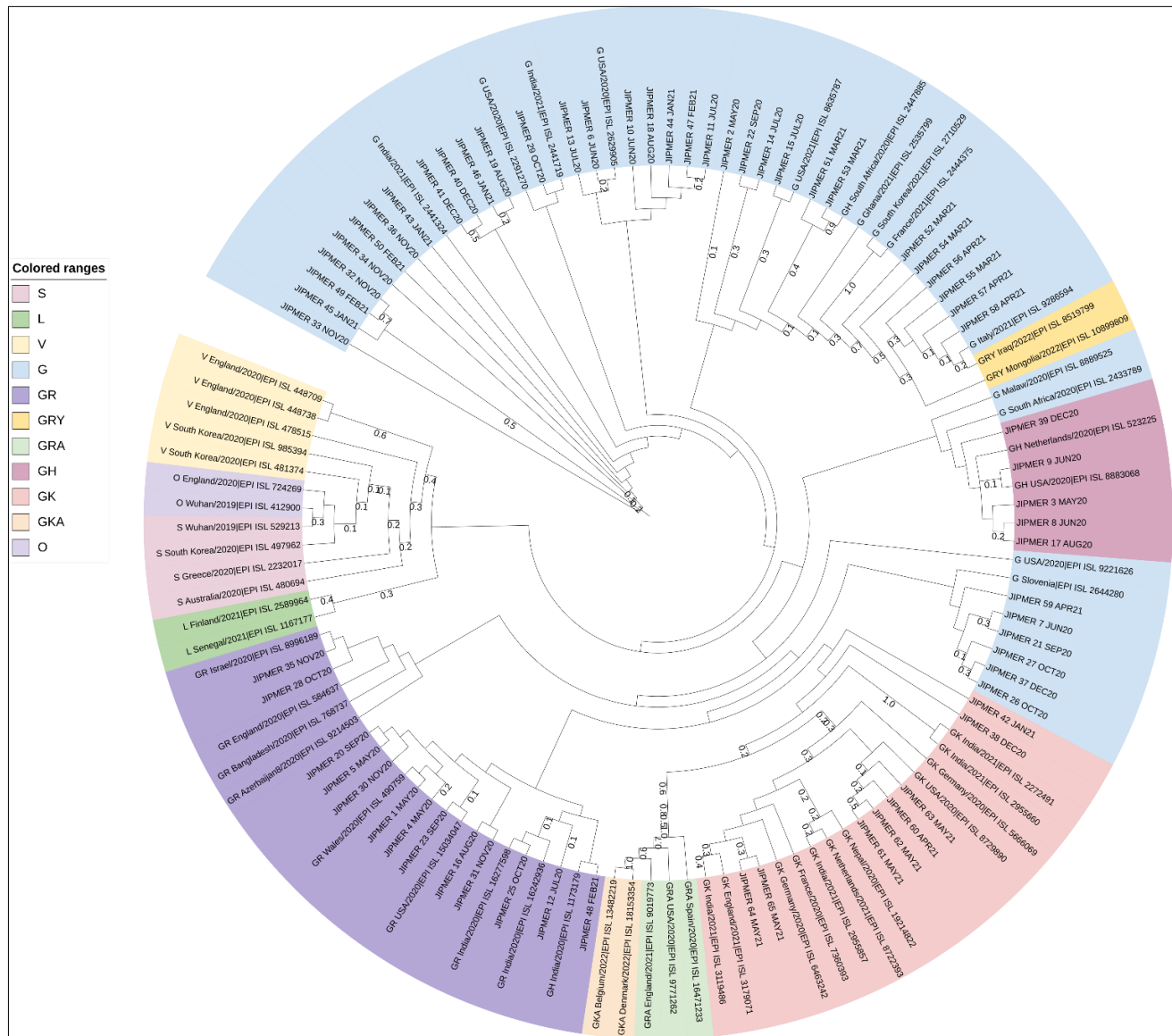


Figure 1: Circular dendrogram visualization of the SARS-CoV-2 spike gene of 65 samples obtained in the study and the closest matching sequences along with reference sequence for various SARS-CoV-2 variants. This dendrogram was constructed by neighbor joining method using Mega 11.0, bootstrap of 1000.

Table 3: Variations across S gene with frequencies >0.4 .

Region in S protein	Position of amino acid	Frequency of the variation
S1 subunit		
N terminal domain	T19I	0.499
	L24del	0.499
	P25del	0.499
	P26del	0.499
	A27S	0.499
	G142D	0.643
	V213G	0.424

Continued.

Region in S protein	Position of amino acid	Frequency of the variation
RBD	S371F	0.486
	S373P	0.502
	S375F	0.502
	T376A	0.487
	D405N	0.486
	R408S	0.406
	K417N	0.441
	N440K	0.436
	L452R	0.480
	S477N	0.498
	T478K	0.866
	E484A	0.502
	Q498R	0.500
	N501Y	0.533
	Y505H	0.503
Carboxy(C)-terminal region of the S1 domain	D614G	0.995
	H655Y	0.536
	N679K	0.536
Furin cleavage	P681H	0.564
S2 subunit		
	N764K	0.520
Fusion peptide	D796Y	0.535
Heptapeptide repeat sequence 1	Q954H	0.536
	N969K	0.536

DISCUSSION

India experienced three major COVID-19 waves, peaking in September 2020, May 2021, and December 2021.⁹ The second wave was particularly severe, with daily cases exceeding 200,000 by mid-April 2021. SARS-CoV-2 clade distribution, based on GISAID classification, varied across these waves: clade G dominated the first wave (May 2020–March 2021), clade GK (Delta) was prevalent during the second wave (April–November 2021), and clade GRA (Omicron) during the third wave (January 2022–May 2023). Notably, the epsilon variant, prevalent in Iraq during mid-2020 to early 2021, was not observed in our study area.¹⁰ Another study report also showed the predominance of the epsilon variant (32.48%), followed by B.1 (20.9%), alpha (8.7%), and beta (6.06%) variants.¹¹ Similarly, the gamma variant was absent between May 2020 and May 2021, aligning with INSACOG data. The D614G mutation, associated with increased infectivity and higher upper respiratory tract viral loads, was predominant.¹²

Variations at spike position 484 (E484A/G/K/Q) which were reported as potent escape mutation were observed.^{13–15} E484K and E484Q, known for reducing neutralization by convalescent sera and monoclonal antibodies, were found in Beta, Gamma, Eta, Kappa, and other variants.¹⁶ Another critical mutation, N501Y, seen in Alpha, Beta, and Gamma variants, is linked to enhanced ACE2 binding and immune evasion when combined with K417N.^{17,18}

Meta-analyses have indicated mutation hotspots in the spike protein, particularly in the N-terminal domain (NTD) and protease cleavage sites. A Hyderabad-based study (January–August 2020) identified L5, L54, and D614G as key mutation sites.¹⁹ Our study of 1,840 isolates confirmed a high mutation density in the NTD region, supporting prior findings. Mutations such as T19R, G142D, P681H/R, and D950N appeared from March 2021, matching earlier reports.²⁰ We observed their increasing frequency, indicating the start of spike region mutation accumulation by January 2021. Among 91 Iraqi isolates (June 2020–March 2021), mutations like S982A, A570D, P681H, and D1118H co-occurred with D614G, patterns also reflected in our dataset. Additionally, the A701V mutation, linked to the Beta variant and affecting antibody binding at the furin cleavage site, was detected in a sample from March 2021 alongside D80A, K417N, E484K, and N501Y mutations.²¹ Overall, our genomic analysis aligns with global trends, highlighting key mutations and clade transitions over time in India's COVID-19 trajectory.

CONCLUSION

In conclusion, we report here for the first time the molecular epidemiology of SARS-CoV-2 in Puducherry from 2020 to 2023. While the clades distribution varied across continents during the early phase of the pandemic, it was more or less uniform during the subsequent waves when the Delta variant dominated followed by its replacement with Omicron. Several sub-lineages of Omicron have continued to emerge around the time of

writing, but none have been recognized as variants of concern.

ACKNOWLEDGEMENTS

Authors would like to thank RVRDL staff.

Funding: The study was funded by Department of Health Research, India. (Project No: VIR/14/2012/ECD-I) and JIPMER - Institutional Intramural grant (JIPMER/01/109/2021/01469: screening for mutations in spike protein gene of SARS-COV-2 genome)

Conflict of interest: None declared

Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021;19(3):141-54.
2. Cascella M, Michael R, Abdul A, Scott C D, Raffaella DN. Features, Evaluation, and Treatment of Coronavirus (COVID-19). *StatPearls*. 2024.
3. World Health Organization. Coronavirus (COVID-19) Dashboard. Available at: <https://covid19.who.int/>. Accessed on 18 August 2025.
4. World Health Organization. COVID-19 Epidemiological Update. 2024. Available at: https://www.who.int/health-topics/coronavirus#tab=tab_1. Accessed on 18 August 2025.
5. George N, Prasad JB, Verma P. Statistical Model for COVID-19 in Different Waves of South Indian States. *Dialogues Health*. 2022;1:100016.
6. Kar SS, Sarkar S, Murali S, Dhodapkar R, Joseph NM, Aggarwal R. Prevalence and Time Trend of SARS-CoV-2 Infection in Puducherry, India, August-October 2020. *Emerg Infect Dis*. 2021;27(2):666-9.
7. Phan T. Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol*. 2020;81:104260.
8. Yuan Y, He J, Gong L, Li W, Jiang L, Liu J, et al. Molecular epidemiology of SARS-CoV-2 clusters caused by asymptomatic cases in Anhui Province, China. *BMC Infect Dis*. 2020;20(1):930.
9. Sarkar A, Chakrabarti AK, Dutta S. Covid-19 Infection in India: A Comparative Analysis of the Second Wave with the First Wave. *Pathogens*. 2021;10(9):1222.
10. Sabir DK. Analysis of SARS-COV2 spike protein variants among Iraqi isolates. *Gene Rep*. 2022;26:101420.
11. Benslimane FM, Al Khatib HA, Al-Jamal O, Albatesh D, Boughattas S, Ahmed AA, et al. One Year of SARS-CoV-2: Genomic Characterization of COVID-19 Outbreak in Qatar. *Front Cell Infect Microbiol*. 2021;11:768883.
12. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell*. 2021;184(1):64-75.
13. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe*. 2021;29(3):463-76.
14. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JCC, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *bioRxiv*. 2020;2020:214759.
15. Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science*. 2020;369(6506):1014-8.
16. Andreano E, Piccini G, Licastro D, Casalino L, Johnson NV, Paciello I, et al. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv*. 2020;2020:424451.
17. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature*. 2022;602(7896):294-9.
18. Tian F, Tong B, Sun L, Shi S, Zheng B, Wang Z, et al. N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. *Elife*. 2021;10:e69091.
19. Guruprasad L. Human SARS CoV-2 spike protein mutations. *Proteins*. 2021;89(5):569-76.
20. Limaye S, Kasibhatla SM, Ramtirthkar M, Kinikar M, Kale MM, Kulkarni-Kale U. Circulation and Evolution of SARS-CoV-2 in India: Let the Data Speak. *Viruses*. 2021;13(11):2238.
21. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*. 2021;593(7857):130-5.

Cite this article as: Philomenadin FS, Rajendran R, Ratchagadasse VR, Dhodapkar R. A comprehensive report on severe acute respiratory syndrome coronavirus-2 lineages in Puducherry from 2020 to 2023. *Int J Community Med Public Health* 2025;12:5225-9.