Archives of Virology (2022) 167:393–403
https://doi.org/10.1007/s00705-021-05320-7

ORIGINAL ARTICLE

Emergence of two distinct variants of SARS-CoV-2 and an explosive second wave of COVID-19: the experience of a tertiary care hospital in Pune, India

Shubham Shrivastava[1] · Suhas T. Mhaske[1] · Meera S. Modak[2] · Rashmi G. Virkar[1] · Shamburaje S. Pisal[1] · Akhilesh Chandra Mishra[1] · Vidya A. Arankalle[1]

Received: 17 August 2021 / Accepted: 19 October 2021 / Published online: 9 January 2022

Abstract

The emergence of novel variants of SARS-CoV-2 in several countries has been associated with increased transmissibility or reduced neutralization potential of antibodies against the Wuhan virus (wild type). From August 2021 onwards, India experienced a progressive decline in the number of active SARS-CoV-2 infections, indicative of a downward trend in the explosive second wave. This prospective study was conducted quarterly for one year (May 2020 to June 2021) at a tertiary care hospital in the city of Pune in western India. Receptor-binding domain (RBD, n = 319) and full genome (n = 20) sequences from viral-RNA-positive nasopharyngeal swabs of COVID-19 patients representing the first and second waves were used for analysis. No Brazilian, South African, or California variants were detected in this study. Until December 2020, only the wild-type strain was prevalent. Concurrent with the upsurge of the second wave in March 2021, 73% (33/45) of RBD sequences harboured L452R/E484Q mutations characteristic of the Kappa variant. In April 2021, co-circulation of Kappa (37%) and Delta (L452R/T478K, 59%) variants was recorded. During May and June 2021, the Delta variant became the predominant circulating variant, and this coincided with a significant decline in the number of COVID-19 cases. Of the 20 full genome sequences, six isolates each exhibited signature mutations of the Kappa and Delta variant. With several states witnessing a reduction in the number of COVID-19 cases, continuous monitoring of newer mutations and assessment of their effect on virus transmissibility and their impact on vaccinated or previously exposed individuals is necessary.

Introduction

Coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, was declared a pandemic on March 11, 2020, and continued to be a global public health concern in 2021. Several countries have experienced a resurgence leading to second or third waves of the disease [1]. Several vaccines have been approved or approved for emergency use and are being used in different countries, depending on their availability and national policies [2]. However, the proportion of the global population that is vaccinated remains low.

The original SARS-CoV-2 strain from Wuhan (wild-type) was rapidly transmitted in a large number of countries through infected travellers, followed by establishment of community transmission and further rapid spread. In India, the first COVID-19 case was reported on January 29, 2020, in a student returning from China [3]. At the peak in September 2020, 97,860 cases were recorded [4]. Subsequently, the number of cases decreased considerably, with the lowest number of cases in January and February 2021 [4]. The second wave started from the middle of March 2021, with the highest number of cases (414,188) recorded on May 6, 2021. Currently, India is experiencing a significant drop in the number of active infections, with 41,831 as of July 31, 2021 [4]. The state of Maharashtra was the first to report an

Handling Editor: Tim Skern.

Shubham Shrivastava and Suhas T. Mhaske contributed equally to this work.

✉ Vidya A. Arankalle
varankalle@yahoo.com

[1] Department of Communicable Diseases, Interactive Research School for Health Affairs (IRSHA), Bharati Vidyapeeth (Deemed to be University), Katraj-Dhankawadi, Pune 411043, India

[2] Department of Microbiology, Bharati Vidyapeeth (Deemed to be University) Medical College, Katraj-Dhankawadi, Pune 411043, India

increase in the number of cases during the second wave and remains the worst-affected state, so far.

Although mutation is an ongoing process for RNA viruses, the identification of a variant harbouring a set of mutations in the spike protein accompanying enhanced transmissibility of SARS-CoV-2 in the UK (UK variant, B.1.1.7, Alpha) was alarming [5, 6]. Subsequently, additional variants of concern with higher transmissibility, virulence, or resistance to the vaccines currently in use were identified in different continents. These include the South African variant B.1.351 (Beta) [7], the Brazilian variants P.1 and P.2, (Gamma) [8, 9], the California variants B.1.429, (Epsilon) and B.1.427 [10, 11], and the most recent Indian variant, B.1.617.2, (Delta) [12].

The first case in the state of Maharashtra was reported on March 19, 2020, in Pune, a city with a population of 5,057,709 that has remained a major hotspot. To monitor SARS-CoV-2 strains over time on a quarterly basis, a single-centre study was undertaken at Pune. Changes in the variants that occurred during the two waves of disease are reported here.

Materials and methods

Clinical specimens

In May 2020, eight nasopharyngeal swab (NPS) specimens from patients with confirmed COVID-19 were collected from a designated COVID treatment facility. Subsequent NPS samples were collected from Bharati Vidyapeeth (Deemed to be University) Medical College and Hospital (BVDUMCH), Pune. NPS specimens collected from suspected COVID-19 patients for diagnosis were tested at BVDUMCH for the presence of SARS-CoV-2 RNA. Depending on their availability, three different kits were used for RT-PCR: Mylab, India; Covipath, Thermo Fisher, USA; and Aargene, France. RNA was extracted from NPS specimens (QIAGEN, USA), and RT-PCR was performed according to the instructions provided by the manufacturers. SARS-CoV-2-positive NPS specimens were stored in aliquots at -80°C until testing. Samples yielding Ct values of <20 in RT-PCR were selected for amplification of the RBD region (669 nt) using PCR. Of the total 319 samples selected for sequencing, 30 and 289 were collected during the first and second wave of the pandemic, respectively. The samples collected in September 2020 represented the peak of the first wave, while those obtained in April 2021 corresponded to the second wave. Five samples were collected in May 2020, 10 in September 2020, 15 in December 2020, 45 in March 2021 (including 20 samples collected in the first week and 25 collected in the fourth week), 91 in April 2021, 132 in May 2021, and 21 in June 2021.

Viral RNA extraction, PCR, and sequencing

Viral RNA was extracted using a QIAamp viral RNA Mini Kit (QIAGEN, CA, USA), and cDNA was synthesized using a high-capacity reverse transcription kit using the reverse primer RBD-R. The PCR product was amplified using AmpliTaq polymerase (Thermo Fisher Scientific, MA, USA) with specific primers. The forward primer RBD-F (5'-GCGGTCTAGAATGAATATTACAAACTT GTGCCCT-3') and the reverse primer RBD-R (5'-TAC ATGCACCAGCAACTGTTTAACTCGAGCCAT-3') were designed to amplify the RBD region, encoding amino acid residues 331-524 of the SARS-CoV-2 spike protein, using a reference sequence from NCBI (accession number NC045512). PCR products (~600 bp) were gel purified and sequenced from both the ends using a Big Dye Terminator Cycle Sequencing Kit (Thermo Fisher Scientific, MA, USA). RBD sequences were confirmed using BLAST (www.ncbi.nlm.nih.gov/BLAST). The forward and reverse sequences were aligned and edited manually using CodonCode Aligner v.9.0.1 software to obtain a consensus sequence. Partial RBD sequences were submitted to the GenBank database at www.ncbi.nlm.nih.gov (Supplementary Table S1).

Sequencing of full SARS-CoV-2 genomes

Full genomes of SARS-CoV-2 were sequenced from NPS specimens using an Ion Torrent targeted next-generation sequencing platform (Thermo Fisher Scientific, MA, USA). The analysis was performed on Torrent Suite Server v5.14 using plugins specifically for the Ion Ampliseq SARS-CoV-2 panel, including IRMA report, to build a consensus sequence. The consensus sequences were assembled using Trinity Assembler v. 1.3.0.2, and contigs were annotated using the NCBI database. Complete genome sequences were submitted to the GenBank database at www.ncbi.nlm.nih.gov (accession numbers MT416725, MT416726, MW969752 to MW969758, MZ021503 to MZ021506, and MZ574051 to MZ574056). One sequence was deposited in the GISAID database (EPI_ISL_1710598).

Sequence analysis

The sequences obtained during the present study and other sequences retrieved from GenBank and GISAID were aligned against that of the Wuhan-Hu-1 reference strain (NC045512) using the MAFFT online alignment tool (v7.475, 2020), and Pangolin v2.1.6 (www.github.com/cov-lineages/pangolin) was used to determine their lineages. The Nextclade web tool (v0.14.2) (https://clades.nextstrain.org)

was used to compare the sequences to those of reference strains for clade assignment.

Results

Comparison of COVID-19 cases in Pune and those diagnosed at Bharati Hospital (March 2020-June 2021)

Figure 1A shows the number of COVID-19 cases each month in the city of Pune as reported by the Pune Municipal Corporation [13]. The first case was detected on March 19, 2020, and the number of cases increased steadily until June, with a sharp rise in July 2020. At the peak in September 2020, 49,918 cases were reported, and this peak was followed by a sharp decline until February 2021. In March 2021, there was a rapid increase in the number of COVID-19 cases, which reached a peak of 150,175 cases in April 2021. Starting in the third week of April, the number declined sharply again, continuing through May and June 2021. During June 2021, 8,590 cases were reported. Figure 1B presents the number of patients advised and the number with confirmed COVID-19 diagnosis at BVDUMCH after permission was given by the government to provide diagnostic services. At Bharati Hospital, during the first wave of disease, the number of COVID-19 cases increased steadily from June to September 2020. The highest number (n = 800) was recorded in September 2020, followed by a sharp decline (n = 285) in October 2020. An increasing trend in the number of cases was observed during March 2021 that reached a peak (n = 3203) in April 2021 during the second wave of the disease. Clearly, the patterns of the number of COVID-19 cases detected at different times were identical when the city of Pune and Bharati Hospital were compared.

RBD sequence analysis

In view of the role of RBD in the attachment of the virus to the ACE2 receptor, in neutralization of the virus by antibodies, and as a major target for vaccine development, RBD sequences were monitored over time. Table 1 shows the frequency of selected mutations in the RBD region over a period of one year (May 2020 to June 2021). Until December 2020, none of the samples exhibited the characteristic mutations (K417N/T, E484K, N501Y, T478K) found in the variants of concern that have been identified so far. Two of the 15 samples from December 2020 contained the N440K mutation. At the beginning of March 2021, when the number of cases started showing an upward trend (Fig. 1B), we did not find the N501Y mutation, which is a characteristic of the UK variant (and shared by the Brazil and SA variants), or the K417N mutation that is found in the Brazil/SA variants.

Strikingly, 70% of the NPS specimens sequenced during the beginning of March 2021 and 83% of the NPS samples sequenced at the end of March contained the L452R mutation observed in the California variant. In addition, instead of the E484K mutation seen in strains from Brazil and South Africa, Indian strains exhibited an E484Q mutation, defined later as the Kappa variant. Overall, in the month of March, 33 out of 45 samples (73%) harboured the India-specific L452R/E484Q mutations. Week-wise analysis revealed the same trend, with 70% and 76% of Kappa-related cases in the first and last week of March 2021. A simultaneous exponential rise in the number of COVID-19 cases (Fig. 1B) revealed a clear association of the emergence of this mutant with the second wave of the disease. Unfortunately, we did

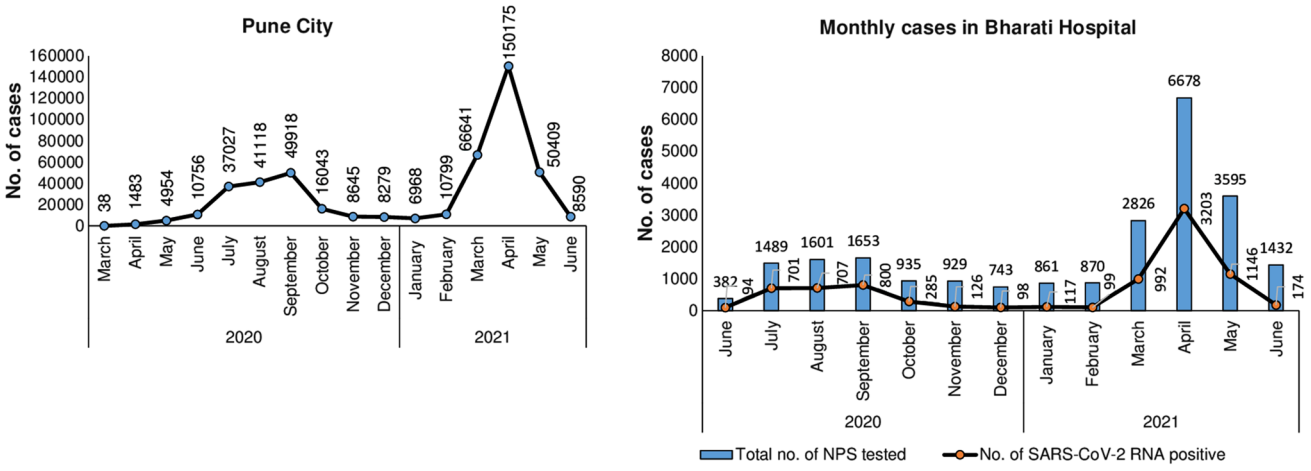

Fig. 1 (A) The number of confirmed COVID-19 cases in the city of Pune from March 2020 to June 2021 and (B) the number of NPS specimens from suspected COVID-19 cases tested at the tertiary care hospital and the number that scored positive for SARS-CoV-2 RNA by RT-PCR (June 2020 to June 2021)

Table 1 SARS-CoV-2 RBD mutant profiles during the first (2020) and second (2021) waves of COVID-19 in Pune, Western India*

Collection month (no. of samples sequenced)	Original (Wuhan)	UK Alpha (N501Y)	Indian**		Others		
			Kappa (L452R, E484Q)	Delta (L452R, T478K)	N440K	E484K	E484Q
May 2020 (5)	5 (100%)	0	0	0	0	0	0
Sep 2020 (10)	10 (100%)	0	0	0	0	0	0
Dec 2020 (15)	13 (87%)	0	0	0	2 (13%)	0	0
Mar 2021 (45)	4 (8.9%)	0	**33 (73.3%)**	**2 (4.4%)**	2 (4.4%)	3 (6.7%)	1 (2.2%)
Apr 2021 (91)	1 (1.1%)	1 (1.1%)	**34 (37.3%)**	**54 (59.3%)**	1 (1.1%)	0	0
May 2021 (132)	0	0	**10 (7.6%)**	**122 (92.4%)**	0	0	0
June 2021(21)	0	0	**0**	**21 (100%)**	0	0	0

*Mutants specific to Brazil/South Africa were not found. **Additionally, in 2021, the V382L mutation was found in 14 out of 289 COVID-19 cases (4.8%).

not collect samples during January or February, and hence, the possibility of earlier detection was missed. As shown in Table 1, we identified the V382L mutation in 4.8% of NPS specimens in 2021.

Analysis of 91 RBD sequences from the samples collected during the month of April led to striking observations. At that time, 34 variants contained L452R/E484Q (37%), while 54 sequences harboured L452R/T478K mutations (59%), characteristic of the Delta variant. One sequence each exhibited the N501Y mutation characteristic of the UK strain, an N440K mutation, and no mutation. Overall, the L452R mutation increased from 0/30 in December to 123/136 (90%) in April ($p < 0.001$). Importantly, as compared to March (35/45), a significant rise in L452R mutants was seen in April (88/91, $p = 0.001$). Interestingly, at the same time, the frequency of the T478K mutation increased dramatically from 4% in March to 59% in April ($p < 0.001$). The increasing frequency of the dominant mutation L452R/T478K continued in May and June 2021, but we experienced a significant decline in the number of patients seeking COVID-19 diagnosis at the hospital (Table 1). Our results indicated that the dominant clade G virus strains in Pune seem to have been replaced by the variant of concern Delta (L452R/T478K) during the second wave of the disease.

Analysis of SARS-CoV-2 full genome sequences from Pune, India

During 2020-21, 20 full SARS-CoV-2 genomes from Pune were sequenced (two in May and two in September of 2020; 10 in March, two in April, and four in May of 2021). The four mutations, C241T, C3037T, C14408T, and A23403G, were observed in all 20 genome sequences from the clade "G" isolates (named after the D614G mutation). Of the 10 genomes from March 2021 selected for sequencing, four were wild type, while six were the Kappa variant (B.1.617.1) according to RBD analysis.

As shown in Table 2, two sequences each from May and September 2020 belonged to lineage B.1.1.306 and clade 20B. Of the 10 sequences from March 2021, two (CD211295 [MW969753] and CD210761 [MZ021503] belonged to the original prevalent lineage B.1.1.306, clade 20B, while the other two sequences, CD210922 (MW969752) and CD210896 (EPI_ISL_1710598) belonged to separate lineages, B.1.1 (clade 20B) and B.1.36.29 (clade 20A), respectively, suggesting simultaneous low-level circulation of the wild-type virus. The remaining six sequences from March 2021 (CD210871 [MZ021506], CD210927 [MZ021505], CD210929 [MZ021504], CD211290 [MW969754], CD211294 [MW969755], and CD211406 [MW969756] formed a distinct lineage, B.1.617.1, Kappa variant (clade 21B). Two sequences from April 2021 (CD212095, CD212098) and four sequences from May 2021 (CD213366, CD213522, CD213523, and CD213570) formed a distinct clade, 21A, and belonged to the B.1.617.2 lineage (Delta variant, VoC).

Complete genome sequence analysis of sample CD210896 (EPI_ISL_1710598) revealed an in-frame stop codon in Orf3a at amino acid position 261 (nucleotide position 26,173). The first strain with this amino acid change was reported in Germany (hCoV-19/Germany/NW-KRO-2355/2020) in March 2020 and more recently in England in June 2021 (hCoV-19/England/HSLL-17FBAD3/2021). It has been reported in 35 countries so far (https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/).

Table 2 Comparison of spike protein mutations in the Indian variant and other variant viruses (reference: Wuhan strain, accession number, NC045512)

Pangolin lineage	Sequence ID*	N-terminal domain in S1 subunit														RBD									C-terminal domain in S1 subunit							S2 subunit										
		aa position														aa position									aa position							aa position										
		18	**19**	20	26	27	70	95	138	142	143	154	171	190	**222**	382	**417**	433	**440**	**452**	**478**	**484**	501	570	614	647	655	675	677	**681**	716	950	982	1027	1071	1101	1118	1133	1153	1176	1251
B	NC-045512, Wuhan	L	**T**	T	P	A	V	T	D	G	V	E	V	R	**A**	V	**K**	V	**N**	**L**	**T**	**E**	N	A	D	A	H	Q	Q	**P**	T	D	S	T	**Q**	H	D	V	**D**	V	G
B.1.1.306	**8003-IRSHA (MT416725)**																								G				H												
	8004-IRSHA (MT416726)																								G				H												
B.1.1.306	**CD208550 (MW969758)**																								G																
	CD208560 (MW969757)																	I							G																V
B.1.1	**CD210922 (MW969752)**										F														G				H								F			
B.1.36.29	**CD210896 (EPI_ISL_1710598)**																		**K**						G																
B.1.1.306	**CD211295 (MW969753)**	F				S							I									**K**			G			H												
B.1.1.306	**CD210761 (MZ021503)**	F				S																**K**			G			H												
B.1.617.1	**CD210871 (MZ021506)**							I		D											**R**		**Q**		G					**R**				**H**						
B.1.617.1	**CD210927 (MZ021505)**									D		K				L					**R**		**Q**		G					**R**				**H**				**Y**		
B.1.617.1	**CD210929 (MZ021504)**									D		K				L					**R**		**Q**		G					**R**				**H**				**Y**		
B.1.617.1	**CD211290 (MW969754)**									D		K				L					**R**		**Q**		G	S				**R**				**H**				**Y**		
B.1.617.1	**CD211406 (MW969756)**									D		K				L					**R**		**Q**		G					**R**				**H**				**Y**		
B.1.617.1	**CD211294 (MW969755)**									D		K									**R**		**Q**		G					**R**				**H**	D					
B.1.617.2	**CD212098 (MZ574054)**		**R**							D					V						**R**	**K**			G					**R**		N								
B.1.617.2	**CD213366 (MZ574051)**		**R**							D											**R**	**K**			G					**R**		N								
B.1.617.2	**CD212095 (MZ574052)**		**R**							D											**R**	**K**			G					**R**		N								
B.1.617.2	**CD213570 (MZ574053)**		**R**							D											**R**	**K**			G					**R**		N								
B.1.617.2	**CD213522 (MZ574055)**		**R**							D											**R**	**K**			G					**R**		N								
B.1.617.2	**CD213523 (MZ574056)**		**R**							D											**R**	**K**			G					**R**		N								
B.1.617.2	EPI_ISL_2880930 (India/HR)		**R**												V						**R**	**K**			G					**R**										
B.1.617.2	EPI_ISL_2272481 (India/DL)		**R**							D					V						**R**	**K**			G					**R**		N								
B.1.617.2	EPI_ISL_2036277 (India/WB)		**R**							D											**R**	**K**			G					**R**		N								
B.1.617.2	EPI_ISL_2897823 (India/MH)		**R**																		**R**	**K**			G					**R**										
B.1.617.2	EPI_ISL_2956019 (Sweden)		**R**							D					V						**R**	**K**			G					**R**		N								
B.1.617.2	EPI_ISL_2958767 (France)		**R**																		**R**	**K**			G					**R**		N								

Table 2 (continued)

Pangolin lineage	Sequence ID*	N-terminal domain in S1 subunit														RBD								C-terminal domain in S1 subunit							S2 subunit									
		aa position														aa position								aa position							aa position									
		18	**19**	20	26	27	70	95	138	142	143	154	171	190	**222**	382	**417**	433	**440**	**452**	**478**	**484**	501	570	614	647	655	675	677	**681**	716	950	982	1027	1071	1101	1118	1133	1153	1176	1251
AY.2	EPI_ISL_2923711 (USA/CA-CDPH)		**R**				F								V		N			**R**	**K**				G					**R**		N								
AY.2	EPI_ISL_2545667 (USA/HI-TAMC)		**R**				F			D					V		N			**R**	**K**				G					**R**		N								
AY.2	EPI_ISL_2929274 (USA/CA-CDC)		**R**				F			D					V		N			**R**	**K**				G					**R**		N								
AY.2	EPI_ISL_2928214 (USA/CA-OC)		**R**				F								V		N			**R**	**K**				G					**R**		N								
B.1.1.7 (UK)	EPI_ISL_601443																						Y	D	G					H	I		A			H				
B.1.1.7 (L18F)	EPI_ISL_720875	F																					Y	D	G					H	I		A			H				
B.1.1.7 (F490S)	EPI_ISL_736026																						Y	D	G					H	I		A			H				
B.1.1.7 (S494P)	EPI_ISL_741039																						Y	D	G					H	I		A			H				
B.1.1.7 (E484K)	EPI_ISL_782148																					K	Y	D	G					H	I		A			H				
B.1.351 (SA)	EPI_ISL_1012924																N					K	Y		G																
P.1 (Brazil)	EPI_ISL_792681	F		N	S				Y					S			T					K	Y		G		Y						I						F	
P.1 (Brazil)	EPI_ISL_875566	F		N	S				Y					S			T					K	Y		G		Y						I						F	
P.1 (Brazil)	EPI_ISL_875567	F		N	S				Y					S			T					K	Y		G		Y						I						F	
P.1 (Brazil)	EPI_ISL_875568	F		N	S				Y					S			T					K	Y		G		Y						I						F	
B.1.427 (California, L452R)	EPI_ISL_1620465																			R					G																
B.1.429	EPI_ISL_824555																			R					G																
B.1.525 (E484K)	EPI_ISL_1615794																					K			G																
B.1.2	EPI_ISL_824741																								G																
B.1.1.432	EPI_ISL_913915																								G																

*Sequence IDs in bold denote sequences from India. Black, wild type; red, variant viruses

Spike protein

Next, spike protein sequences from the Indian variants were compared with those of other known variants (Table 2). The Indian variants formed two distinct clusters, B.1.617.1 (Kappa) and B.1.617.2 (Delta). The presence of the D614G mutation in all of the Indian sequences revealed that clade G continued to be the only clade circulating so far in Pune. Mutations specific to the recently emergent variants (UK, South Africa, Brazil, and California) were not shared by the Indian variants; however, signature mutations of the Delta variant were observed in all six sequences obtained in April and May of 2021.

Two sequences from 2021 (CD211295 [MW969753] and CD210761 [MZ021503]) belonging to the original prevalent lineage B.1.1.306 had four characteristic mutations, L18F, A27S, E484K, and Q675H, in the spike protein. None of these mutations were present in the four sequences from 2020, suggesting that SARS-CoV-2 is continuously evolving. Two isolates, CD210922 (MW969752) and CD210896 (EPI_ISL_1710598), had the unique mutations V143F, Q677H, and N440K (Table 2). All six Indian variants of the B.1.617.1 lineage exhibited five unique mutations: G142D, L452R, E484Q, P681R, and Q1071H. Additionally, the E154K mutation was found in five and the V382L and D1153Y mutations were found in four of the six Kappa variants. For the Delta variant, all six sequences belonged to B.1.617.2 lineage and had six unique mutations (T19R, G142D, L452R, T478K, P681R, and D950N). Together with the substitution E156G, two deletions at F157 and R158 were identified. Notably, the position of the P681R mutation in the Kappa and Delta variants is immediately adjacent to the furin cleavage site (682-685). *In vitro* site-directed mutagenesis experiments have shown that this mutation leads to an increase in the fusion activity of the SARS-CoV-2 spike protein [14]. Importantly, P681H was present in the UK variant B.1.1.7 (Table 2). The acquisition of A222V and K417N by the Delta variant (AY.2 lineage) was associated with an upsurge in COVID-19 cases in Europe and the USA, and importantly, these amino acid changes were not found in the Delta variants from this study.

Nucleoprotein

The Indian variants were distinct in this protein as well (Table 3) and did not share the D3L and S236F mutations and the P80R mutation seen in the UK and Brazilian variants, respectively. The two Indian sequences (CD211295, CD210761) had tyrosine (Y) instead of leucine (L) at the third amino acid position, as was observed in the UK strain. The R203M and D377Y mutations were observed in all twelve Kappa and Delta variants. At position 203, the change is from a hydrophilic arginine to a hydrophobic methionine, while at position 377, aspartic acid is changed to tyrosine. The biological significance of these mutations needs to be evaluated. The remaining Indian sequences continued to harbour the R203K and G204R mutations, which were present since May 2020. One Indian sequence (CD210896_EPI_ISL_1710598) was unique, without amino acid substitutions at positions 203 and 204. Like the Delta variant, the six Indian variants contained the unique mutation D63G.

Other genomic regions

The other genome regions were also compared with the Wuhan strain as well. The Kappa-variant-specific mutations were Nsp3-T749I, Nsp6-T77A (Orf1a), Nsp12-P323L, Nsp13-M429I, Nsp15-K259R (Orf1b), S26L (Orf3a), I33T (Orf6), and V82A (Orf7a). The other proteins (E, M, Orf7b, Orf8, and Orf10) remained unchanged in the Kappa variant. The Indian Delta variant-specific mutations were G210T (5'UTR) Nsp2-P129L, Nsp3-P822L, Nsp4-A446V, Nsp6-V149A (Orf1a), Nsp12-P323L, Nsp12-G671S, Nsp13-P77L (Orf1b), S26L (Orf3a), I82T (M), V82A, T120I (Orf7a), and deletions of D119 and F120 in Orf8. The other proteins (E, Orf6, Orf7b, and Orf10) remained unchanged in the Delta variant.

Discussion

Here, we report the association of the emergence of two SARS-CoV-2 variants with the progression of the second wave of COVID-19, as evidenced by the data from a large tertiary care hospital in Pune, India. At the beginning of the pandemic, diagnostic facilities were available only at a few national laboratories, and hence, the first five samples analysed in May 2020 came from a central treatment facility. From the first week of June onward, BVDUMCH, a tertiary care hospital at Pune serving a population of 5 million (2011 census), was permitted to undertake RT-PCR testing for COVID-19 diagnosis. Therefore, the subsequent 314 samples were collected at BVDUMCH. Although our study is restricted to a single hospital, it does reflect situation in the city of Pune.

The variants of concern (VoC) identified as UK, Brazil, and SA were not detected during the first wave of the disease in 2020. It seems that the spread of SARS-CoV-2 during the first wave of disease in India could be attributed to the replacement of clade L (original Wuhan strain) with clade G, characterized by the spike mutation D614G. All of the Pune strains exhibited this mutation and belonged to clade G. Notably, the D614G mutation was shown to enhance the rate of SARS-CoV-2 infection in the upper airway [15, 16].

Table 3 Mutations in the nucleocapsid protein with reference to the Wuhan strain (accession number, NC045512).

Domain of the nucleocapsid protein				N-terminal domain				SR-rich			Linker region	
Amino acid position		2	3	**63**	80	119	142	194	**203**	204	236	**377**
Pangolin lineage	Sequence ID*											
B	NC-045512 (Wuhan)	S	D	**D**	P	A	P	S	**R**	G	S	**D**
B.1.1.306	**8003-IRSHA (MT416725)**								**K**	R		
	8004-IRSHA (MT416726)								**K**	R		
B.1.1.306	**208550 (MW969758)**								**K**	R		
	208560 (MW969757)								**K**	R		
B.1.1	**CD210922 (MW969752)**								**K**	R		
B.1.36.29	**CD210896 (EPI_ISL_1710598)**	P						L				
B.1.1.306	**CD211295 (MW969753)**		Y				S		**K**	R		
B.1.1.306	**CD210761 (MZ021503)**		Y			T			**K**	R		
B.1.617.1	**CD210871 (MZ021506)**								**M**			**Y**
B.1.617.1	**CD210927 (MZ021505)**								**M**			**Y**
B.1.617.1	**CD210929 (MZ021504)**								**M**			**Y**
B.1.617.1	**CD211290 (MW969754)**								**M**			**Y**
B.1.617.1	**CD211406 (MW969756)**								**M**			**Y**
B.1.617.1	**CD211294 (MW969755)**								**M**			**Y**
B.1.617.2	**CD212098 (MZ574054)**			**G**					**M**			**Y**
B.1.617.2	**CD213366 (MZ574051)**			**G**					**M**			**Y**
B.1.617.2	**CD212095 (MZ574052)**			**G**					**M**			**Y**
B.1.617.2	**CD213570 (MZ574053)**			**G**					**M**			**Y**
B.1.617.2	**CD213522 (MZ574055)**			**G**					**M**			**Y**
B.1.617.2	**CD213523 (MZ574056)**			**G**					**M**			**Y**
B.1.617.2	EPI_ISL_2880930 (India/HR)			G					M			Y
B.1.617.2	EPI_ISL_2272481 (India/DL)			G					M			Y
B.1.617.2	EPI_ISL_2036277 (India/WB)			G					M			Y
B.1.617.2	EPI_ISL_2897823 (India/MH)			G					M			Y
B.1.617.2	EPI_ISL_2956019 (Sweden)			G					M			Y
B.1.617.2	EPI_ISL_2958767 (France)			G					M			Y
AY.2	EPI_ISL_2923711 (USA/CA-CDPH)			G					M			Y
AY.2	EPI_ISL_2545667 (USA/HI-TAMC)			G					M			Y
AY.2	EPI_ISL_2929274 (USA/CA-CDC)			G					M			Y
AY.2	EPI_ISL_2928214 (USA/CA-OC)			G					M			Y
B.1.1.7 (UK)	EPI_ISL_601443		L						**K**	R	F	
B.1.1.7 (L18F)	EPI_ISL_720875		L						**K**	R	F	
B.1.1.7 (F490S)	EPI_ISL_736026		L						**K**	R	F	
B.1.1.7 (S494P)	EPI_ISL_741039		L						**K**	R	F	
B.1.1.7 (E484K)	EPI_ISL_782148		L						**K**	R	F	
B.1.351 (SA)	EPI_ISL_1012924											
P.1 (Brazil)	EPI_ISL_792681				R				**K**	R		
P.1 (Brazil)	EPI_ISL_875566				R				**K**	R		
P.1 (Brazil)	EPI_ISL_875567				R				**K**	R		
P.1 (Brazil)	EPI_ISL_875568				R				**K**	R		
B.1.427 (California, L452R)	EPI_ISL_1620465											
B.1.429	EPI_ISL_824555											
B.1.525 (E484K)	EPI_ISL_1615794											
B.1.2	EPI_ISL_824741											
B.1.1.432	EPI_ISL_913915								**K**	R		

*Sequence IDs in bold denote sequences from India. Black, wild type; red, variant viruses

The second wave in India was first noticed in the state of Maharashtra, to which Pune belongs. The emergence of the second wave was characterized by the appearance of the Kappa variant in the first week of March 2021 (Table 1). Within weeks, the virus evolved further, replacing the E484Q mutation with T478K, and this variant was later named the Delta variant. Interestingly, the L452R mutation was maintained. Later, both variants circulated for a time until the Delta variant became dominant. The reasons for the simultaneous mutations at positions 484 and 478 in the RBD and the replacement of the highly transmissible Kappa variant with the Delta variant are not yet clear. The role of factors other than the error-prone replication of RNA viruses remains to be determined. Another puzzling fact is the association of the emergence of the Delta variant in Pune with the sharp decline in the number of cases. The role of other mutations in the viral genome and host immunity needs to be examined.

Because India is a large country with distinct geographic and socioeconomic differences, the second wave of COVID-19 was recorded in different states at different time points, starting with the state of Maharashtra. As a result, although there was an early peak in Pune and the state of Maharashtra, the peak of the second wave for the country as a whole did not occur until 2021 [13, 17]. It was observed that the UK variant was predominant in the North Indian states of Punjab and Delhi [18]. At the time of submission of this manuscript, similar observations of co-circulation of novel Kappa and Delta variants had been reported in Maharashtra [12]. According to WHO, the Kappa and Delta variants accounted for 21% and 7% of the sequenced samples from India in late April 2021 [19]. In a recent study in Houston, USA, all of the known mutants, in variable proportions, were detected at a single centre [20].

Outside India, the United Kingdom reported a large upsurge in the number of COVID-19 cases caused by the Delta variant, prompting WHO to declare this variant a VoC, on May 6, 2021. In the third week of June, a 79% rise in the number of COVID-19 cases in the UK was due to the Delta variant. Around the same time, the Delta variant accounted for more than 20% of the cases in USA. This variant was reported to have a faster growth rate than the UK variant [19, 21]. According to the technical briefings by Public Health England, despite the dominance of the Delta variant in UK, the rise in hospital admissions and deaths was lower than the rate of new infections [22, 23].

Regarding the L452R and E484Q mutations in the Kappa variant, the substitution of a hydrophobic leucine (L) with a hydrophilic arginine (R) is similar to what occurred in the California variant, which was shown to correlate with increased transmissibility and decreased antibody neutralization [24]. Instead of lysine at position 484, as is found in the SA and Brazil variants, the Kappa variant has glutamine. In two different sequences, E484K was present without any change at position 452 (Table 2). The E484K variant is of serious concern because of its resistance to some therapeutic monoclonal antibodies and decreased *in vitro* neutralization by the serum samples from patients infected with the original SARS-CoV-2. In this regard, an elegant study by Greaney et al. is noteworthy [25]. By employing a deep mutational scanning method, they were able to assess the effect of even a single mutation in the RBD region on the binding/neutralization of well-characterized plasma samples from COVID-19 patients. Mutations at position E484 were most striking in terms of reduction in neutralization titers. More importantly, mutations to K, Q, or P led to a 10- to 60-fold decrease in the neutralizing antibody titer. Thus, the position 484 is crucial for the interaction of neutralizing antibodies with the RBD region, and mutations at this position result in a reduction in neutralizing antibody titers.

Several countries have witnessed a striking increase in the number of COVID-19 cases attributed to the emergence of the Delta variant. This variant is characterized by a mutation at T478 in the RBD, which is located near the interface of the RBD binding site with the ACE2 receptor. With the T478I mutation, moderate resistance to neutralization by two monoclonal antibodies and convalescent sera from two COVID-19 patients has been documented [26]. In in-vitro experiments, an increased binding affinity of the T478K mutant for the ACE2 receptor has been shown [27]. It is pertinent to note here that a recent increase in COVID-19 cases in New York, USA, was observed to be due to the introduction of B.1.243 lineage strain with P681H and T478K mutations [28]. The immunoreactivity of the variants currently in circulation in Pune with antibodies induced by wild-type viral infection and vaccines would decide the degree of reinfections/infections among previously infected COVID-19 patients and vaccine recipients. In vaccine breakthrough infections in healthcare workers at three centres in Delhi, India, the Delta variant B.1.617.2 had 3.2-fold higher transmissibility than the UK variant, B.1.1.7 [29]. In line with these observations, the Delta variant caused a higher rate of re-infections when compared to the Alpha strain in the UK [23].

Interestingly, two Indian strains (CD211295-MW969753, CD210761-MZ021503) had a combination of two substitutions, L18F and E484K. The L18F substitution was dominant in England before the emergence of the UK strain. It was shown that the proportion of genomes with the L18F mutation increased from 35% in September to 43% in October and to 52% in November 2020, suggesting that this mutation facilitates virus transmission [30]. The role of simultaneous L18F and E484K mutations in virus transmission needs to be evaluated.

Apart from the mutations of concern described above, V382L (9/55, 16.4%) and N440K (4/55, 7.3%)

substitutions were found in the 2021 sequences (Table 1). The prevalence of an RBD variant with N440K was estimated to be 2.1% in India, with ~34% in Andhra Pradesh, a southern Indian state [31]. The N440K mutation was also reported in COVID-19 reinfection cases from northern India [32]. The N440K mutant showed enhanced binding affinity to the human ACE2 receptor and resulted in immune escape from a panel of neutralizing monoclonal antibodies [33, 34].

The protease cleavage site (aa 675-692) in the spike protein is the most prone to mutation. The Kappa and Delta variants both contained a mutation at position 681, located adjacent to the furin cleavage site (aa 682-685), that was shown to facilitate efficient SARS-CoV-2 transmission and infection [14, 35]. The rapid global spread of the UK variant with P681H and P681R mutations is well known. Although we also found substitutions in other regions of the genome, their significance remains unclear.

The Delta variant is continuously evolving, and a new variant, AY.1 (delta plus, B.1.617.2.1), carrying an additional K417N mutation, is emerging in India and the UK. This particular mutation was first identified in B.1.351, the Beta variant, in Brazil. Acquisition of the K417N and A222V mutations by Delta resulted in the emergence of AY.2, which is being associated with an upsurge in COVID-19 infections in the United States.

In summary, the beginning of the second wave of COVID-19 in Pune in March 2021 was associated with the first emergence of the Kappa variant, followed immediately by the Delta variant, which then almost completely replaced the Kappa variant. Although the cases have declined significantly since May 2021, molecular surveillance needs to be continued. Monitoring of the emergence of newer mutations and assessment of their association with virus transmission and infection of vaccinated or previously exposed individuals continues to be of utmost importance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00705-021-05320-7.

Acknowledgements We thank Ms. Abha Thite, Ms. Meghana Walke, and Ms. Shraddha Hatangadi for excellent technical assistance, Mr. Rahul Patil for statistical analysis, and Mr. Rahul Kadu and Dinesh Nande for BSL-3 laboratory support. Timely help from the staff of the Microbiology Department, Bharati Vidyapeeth (Deemed to be University) Medical College and Hospital, is gratefully acknowledged.

Author contributions VAA: concept, study design, and supervision. SS, STM, RGV, and SSP: performance of experiments, data analysis, and interpretation. MM: recruitment of patients and clinical management. SS and VAA: writing of the manuscript. ACM: review of the results and the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Department of Biotechnology (DBT) – BIRAC (grant number BT/NBM0095/02/18). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability All data generated or analyzed during this study are included in this article. A full list of sequence identifiers is given in the text.

Declarations

Conflict of interest None of the authors have any conflict of interest.

Institutional review board statement This study was approved by the Human Ethics Committee of Bharati Vidyapeeth Deemed to be University Medical College, Pune. The study was conducted according to the guidelines of the Declaration of Helsinki.

References

1. Geneva: World Health Organization (2020) WHO coronavirus disease (COVID-19) dashboard. https://covid19.who.int/. Accessed 23 Apr 2021
2. World Health Organization (2020) Status of COVID-19 vaccines within WHO EUL/PQ evaluation process. https://extranet.who.int/pqweb/sites/default/files/documents/Status_COVID_VAX_16Feb2021.pdf. Accessed 23 Apr 2021
3. Andrews MA, Areekal B, Rajesh KR et al (2020) First confirmed case of COVID-19 infection in India: a case report. Indian J Med Res 151:490–492. https://doi.org/10.4103/ijmr.IJMR_2131_20
4. (2021) https://www.covid19india.org/state/TT. Accessed 23 Apr 2021
5. Leung K, Hh SM, Leung GM, Lam TT, Wu JT (2021) Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill 26:2002106. https://doi.org/10.2807/1560-7917.ES.2020.26.1.2002106
6. Volz E, Mishra S, Chand M et al (2020) Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: insights from linking epidemiological and genetic data. medRxiv. https://doi.org/10.1101/2020.12.30.20249034
7. Tegally H, Wilkinson E, Giovanetti M et al (2020) Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv. https://doi.org/10.1101/2020.12.21.20248640
8. Sabino EC, Buss LF, Carvalho MPS et al (2021) Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. Lancet 397:452–455. https://doi.org/10.1016/S0140-6736(21)00183-5
9. Claro IM, da Silva Sales FC, Ramundo MS et al (2021) Local transmission of SARS-CoV-2 lineage B.1.1.7, Brazil, December 2020. Emerg Infect Dis 27:970–972. https://doi.org/10.3201/eid2703.210038
10. Zhang W, Davis BD, Chen SS, Martinez JMC, Plummer JT, Vail E (2021) Emergence of a novel SARS-CoV-2 variant in Southern California. JAMA 325:1324–1326. https://doi.org/10.1001/jama.2021.1612
11. Tchesnokova V, Kulakesara H, Larson L et al (2021) Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. bioRxiv. https://doi.org/10.1101/2021.02.22.432189
12. Cherian S, Potdar V, Jadhav S et al (2021) Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R in

the second wave of COVID-19 in Maharashtra, India. bioRxiv. https://doi.org/10.1101/2021.04.22.440932
13. COVID-19 progression in Pune and Maharashtra (2021) http://www.cessi.in/coronavirus/pune. Accessed 9 Aug 2021
14. Hoffmann M, Kleine-Weber H, Pohlmann S (2020) A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol Cell 78:779-784.e5. https://doi.org/10.1016/j.molcel.2020.04.022
15. Plante JA, Liu Y, Liu J et al (2021) Spike mutation D614G alters SARS-CoV-2 fitness. Nature 592:116–121. https://doi.org/10.1038/s41586-020-2895-3
16. Hou YJ, Chiba S, Halfmann P et al (2020) SARS-CoV-2 D614G variant exhibits enhanced replication ex vivo and earlier transmission in vivo. Science 370:1464–1468. https://doi.org/10.1126/science.abe8499
17. Boehm E, Kronig I, Neher RA et al (2021) Novel SARS-CoV-2 variants: the pandemics within the pandemic. Clin Microbiol Infect 27:1109–1117. https://doi.org/10.1016/j.cmi.2021.05.022
18. Vaidyanathan G (2021) Coronavirus variants are spreading in India—what scientists know so far. Nature 593:321–322
19. World Health Organization (2021) Weekly epidemiological update on COVID-19-11 May 2021 (2021) https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---11-may-2021
20. Long SW, Olsen RJ, Christensen PA et al (2021) Sequence analysis of 20,453 severe acute respiratory syndrome coronavirus 2 genomes from the houston metropolitan area identifies the emergence and widespread distribution of multiple isolates of all major variants of concern. Am J Pathol 191:983–992. https://doi.org/10.1016/j.ajpath.2021.03.004
21. Bolze A, Cirulli ET, Luo S et al (2021) Rapid displacement of SARS-CoV-2 variant B.1.1.7 by B.1.617.2 and P.1 in the United States. medRxiv. https://doi.org/10.1101/2021.06.20.21259195
22. O'Dowd A (2021) Covid-19: cases of delta variant rise by 79%, but rate of growth slows. BMJ 373:n1596. https://doi.org/10.1136/bmj.n1596
23. Public Health England (2021) SARS-CoV-2 variants of concern and variants under investigation (publishing.service.gov.uk). Technical briefing 16. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1001359/Variants_of_Concern_VOC_Technical_Briefing_16.pdf
24. Deng X, Garcia-Knight MA, Khalid MM et al (2021) Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. medRxiv. https://doi.org/10.1101/2021.03.07.21252647
25. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD (2021) Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe 29:463-476.e6. https://doi.org/10.1016/j.chom.2021.02.003
26. Liu Z, VanBlargan LA, Bloyet LM et al (2021) Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe 29:477-488.e4. https://doi.org/10.1016/j.chom.2021.01.014
27. Zahradnik J, Marciano S, Shemesh M et al (2021) SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. bioRxiv. https://doi.org/10.1101/2021.01.06.425392
28. Lasek-Nesselquist E, Pata J, Schneider E, St. George K (2021) A tale of three SARS-CoV-2 variants with independently acquired P681H mutations in New York State. medRxiv. https://doi.org/10.1101/2021.03.10.21253285
29. Mlcochova P, Kemp SA, Dhar MS et al (2021) SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. Nature. https://doi.org/10.1038/s41586-021-03944-y
30. Grabowski F, Preibisch G, Giziński S, Kochańczyk M, Lipniacki T (2021) SARS-CoV-2 variant of concern 202012/01 has about twofold replicative advantage and acquires concerning mutations. Viruses 13:392. https://doi.org/10.3390/v13030392
31. Jolly B, Rophina M, Shamnath A et al (2020) Genetic epidemiology of variants associated with immune escape from global SARS-CoV-2 genomes. bioRxiv. https://doi.org/10.1101/2020.12.24.424332
32. Gupta V, Bhoyar RC, Jain A et al (2020) Asymptomatic reinfection in two healthcare workers from India with genetically distinct SARS-CoV-2. Clin Infect Dis 23:ciaa1451. https://doi.org/10.1093/cid/ciaa1451
33. Thomson EC, Rosen LE, Shepherd JG et al (2020) Circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. Cell 184:1171-1187.e20. https://doi.org/10.1016/j.cell.2021.01.037
34. Weisblum Y, Schmidt F, Zhang F et al (2020) Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife 9:e61312. https://doi.org/10.7554/eLife.61312
35. Peacock TP, Goldhill DH, Zhou J et al (2021) The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. Nat Microbiol 6:899–909. https://doi.org/10.1038/s41564-021-00908-w