Replication of SARS-CoV-2 in human population is defined by distributions of mutants that are present at different frequencies within the infected host, and can be detected by ultra-deep sequencing techniques. In this study, we have examined the SARS-CoV-2 mutant spectra of amplicons from the spike (S)-coding region of five nasopharyngeal isolates derived from vaccine-breakthrough patients. Interestingly, all patients became infected with the Alpha variant but amino acid substitutions that correspond to the Delta Plus, Iota and Omicron variants were present in the mutant spectra of the resident virus. Deep sequencing analysis of SARS-CoV-2 from vaccine-breakthrough patients revealed a rich reservoir of mutant types, and may also inform of tolerated substitutions that can be represented in epidemiological dominant variants.
Vaccine-breakthrough infections with SARS-CoV-2 Alpha mirror
mutations in Delta Plus, Iota and Omicron

Brenda Martínez-González¹, Lucía Vázquez-Sirvent¹, María Eugenia Soria¹,², Pablo Mínguez³,⁴,⁵, Llanos Salar-Vidal¹, Carlos García-Crespo², Isabel Gallego²,⁶, Ana Isabel de Ávila², Carlos Llorens⁷, Beatriz Soriano⁷, Ricardo Ramos-Ruiz⁸, Jaime Esteban¹, Ricardo Fernandez-Roblas¹, Ignacio Gadea¹, Carmen Ayuso³,⁴, Javier Ruiz-Hornillos⁹,¹⁰,¹¹, Concepción Pérez-Jorge¹, Esteban Domingo²,⁶ and Celia Perales¹,⁶,¹²*

¹Department of Clinical Microbiology, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM) Av. Reyes Católicos 2, 28040 Madrid, Spain, ²Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Consejo Superior de Investigaciones Científicas (CSIC), Campus de Cantoblanco, 28049 Madrid, Spain, ³Department of Genetics & Genomics, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Av. Reyes Católicos 2, 28040 Madrid, Spain, ⁴Centro for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, 28029 Madrid, Spain, ⁵Bioinformatics Unit, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid 28040, Spain, ⁶Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, 28029 Madrid, Spain, ⁷Biotechvana, “Scientific Park”, Universidad de Valencia, 46980 Valencia, Spain, ⁸Unidad de Genómica, “Scientific Park of Madrid”, Campus de Cantoblanco, 28049 Madrid, Spain, ⁹Allergy Unit, Hospital Infanta Elena, Valdemoro, Madrid, Spain; ¹⁰Instituto de Investigación Sanitaria-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Av. Reyes Católicos 2, 28040 Madrid, Spain; ¹¹Faculty of Medicine, Universidad Francisco de Vitoria, Madrid, Spain, ¹²Department of Molecular and Cell Biology, Centro Nacional de Biotecnología (CNB-CSIC), Consejo Superior de Investigaciones Científicas (CSIC), Campus de Cantoblanco, 28049 Madrid, Spain

*Address correspondence to: Celia Perales, Centro Nacional de Biotecnología (CNB-CSIC), Consejo Superior de Investigaciones Científicas (CSIC), C/ Darwin, 3, Campus
Abstract

Replication of SARS-CoV-2 in human population is defined by distributions of mutants that are present at different frequencies within the infected host, and can be detected by ultra-deep sequencing techniques. In this study, we have examined the SARS-CoV-2 mutant spectra of amplicons from the spike (S)-coding region of five nasopharyngeal isolates derived from vaccine-breakthrough patients. Interestingly, all patients became infected with the Alpha variant but amino acid substitutions that correspond to the Delta Plus, Iota and Omicron variants were present in the mutant spectra of the resident virus. Deep sequencing analysis of SARS-CoV-2 from vaccine-breakthrough patients revealed a rich reservoir of mutant types, and may also inform of tolerated substitutions that can be represented in epidemiological dominant variants.

Introduction

SARS-CoV-2 continues its diversification worldwide and a new variant termed Omicron (B.1.1.529) carrying a large number of mutations has been recently described in South Africa and classified as a potential variant of concern (VOC) by the World Health Organization (WHO) [https://www.who.int/news/item/26-11-2021-classification-of-omicron-(B.1.1.529)-sars-cov-2-variant-of-concern]. As compared to other VOCs, current evidence suggests an increased risk of reinfection with this variant.

It has been reported that SARS-CoV-2 replicates as distribution of mutants in infected hosts (1-3), as was also previously described for other coronaviruses (4, 5), and in general for RNA viruses. This implies that a consensus sequence of an isolate determined for diagnostic purposes in reality hides a mixture of different variants present in different proportions within the same population (6).

Despite vaccination being highly effective to prevent severe COVID-19, vaccine breakthrough infections are observed (7, 8). Little is known of the mutant spectrum composition of the SARS-CoV-2 that infect fully vaccinated individuals. This raises the question of whether a vaccine failure could be associated with an ensemble of variant genome that can facilitate replication in the face of an effective anti-SARS-CoV-2 immune response (9, 10). Here we show that the virus replicating in vaccinated
individuals who developed COVID-19 as a consequence of infection with the Alpha variant includes signature mutations of Delta Plus, Iota and Omicron SARS-CoV-2.

**Results and Discussion**

We have studied five patients who had been fully vaccinated (two-doses) with BNT162b2 (Pfizer-BioNTech), and who mounted an effective antiviral response [>2,000 arbitrary units (AU)/ml]. They were subsequently infected with SARS-CoV-2 in April 2021, and developed COVID-19 clinical symptoms. Nasopharyngeal swabs were collected between the period of 6 to 14 of April 2021 that corresponds to the fourth pandemic wave in Madrid (Spain) associated with the Alpha variant. RNA extracted from the diagnostic samples of these vaccinated and infected patients was used to amplify six overlapped amplicons of the genomic region of the spike (S) protein (covering nucleotides 21,488 to 23,666; residue numbering is according to the genomic nucleotide sequence of the Wuhan-Hu-1 isolate, NCBI reference NC_045512.2) that were analyzed by ultra-deep sequencing with a cut-off value of 0.1%. Two deletions (Δ69-70 and Δ144) and four amino acid substitutions (N501Y, A570D, D614G, P681H), characteristic of the Alpha variant, were dominant variations (termed “Divergence” mutations) relative to the reference sequence (Wuhan isolate) (Figure 1A). Interestingly, in addition to these “Divergence” mutations, amino acid substitutions representative of the Delta Plus, Iota and Omicron variants were also found in the mutant spectra of the five vaccine-breakthrough patients. In particular, substitution L5F in patient Pt449 present in the Iota variant, A222V in patients Pt450 and Pt453 present in the Delta Plus variant, N679K in patient Pt451 present in the Omicron variant, and P681R in patient Pt452 present in the Delta Plus variant were found at frequencies of 2.2%, 0.6%, 0.2%, 12.6%, 0.2% respectively, within their corresponding mutant spectra (Figure 1). Additionally, previously undescribed amino acid replacements at positions that were also substituted to other amino acids in the Iota (amino acids 157, 452), Delta Plus (amino acids 417, 452) and Omicron (amino acids 417, 547) variants were also present (Figure 1).

As a comparison, the mutant spectrum of diagnostic samples of five unvaccinated patients who were infected by the Alpha variant in Madrid at about the same time (January 16- February 13, 2021) was analyzed. These viral samples did not include substitution N679K, which was present at a frequency of 12.64% in one of the vaccine breakthrough samples. The remaining substitutions were shared by the two groups at similar frequencies, with the exception of L5F which was present at a frequency of 0.24% in
virus from one of the unvaccinated patients, and 2.2% in virus from a vaccine breakthrough case. These data do not support that the majority of substitutions scored in mutant spectra of virus from vaccine-breakthrough patients was influenced by the immune pressure exerted by the vaccine. Establishing a possible role of N679K in immune escape will require further studies. Thus, despite the Omicron variant being first reported to WHO from South Africa on 24 November 2021, SARS-CoV-2 mutant spectra from a vaccine-breakthrough infected patient from the fourth wave in Madrid already included Omicron-associated mutations.

The presence in the mutant spectra of isolates assigned to the Alpha variant of minority mutations that were dominant in Delta Plus, Iota and Omicron variants reflects a complex intra-host SARS-CoV-2 dynamics, with variants that incorporate tolerated mutations. The variants are present at different frequencies, now amenable to scrutiny by deep-sequencing that can attain cut-off detection levels of 0.1% with the number of clean reads produced (see Methods). Replacements of some minority mutant subpopulations by others are continuously produced and frequency variations depend on selective pressures applied to the viral population. Mutant spectra may be predictors of the mutation repertoires with potential to become dominant at the epidemiological level.

The total number of mutations identified in the S-coding region of the SARS-CoV-2 of the five patients analyzed amounted to a maximum mutation frequency of $4.09 \times 10^{-5}$ mutations per nucleotide, in line with typical values for RNA viruses in general. This mutational level in nasopharyngeal diagnostic samples suggest the presence of abundant mutant reservoirs. Mutations need not be directly beneficial but may become so in another environment, in a different individual genome sequence, or under another viral population context (intra-mutant spectrum interaction set) (11). Yet another implication is that successive COVID-19 waves that are associated with variants with a defined name cannot be regarded as compartmentalized entities. The mutant spectrum of epidemiologically relevant SARS-CoV-2 isolates can be permeated by genomes with minority mutations with past or future prominence.

**Methods**

**Patient cohort and stratification.** The viral samples were collected during the fourth COVID-19 outbreak in Spain between the period of 6 to 14 of April 2021. The cohort of the study included five patients diagnosed to be positive for SARS-CoV-2 at Hospital
Universitario Rey Juan Carlos (Móstoles, Madrid, Spain) in April, 2021. All patients had been fully vaccinated (two-doses) with BNT162b2 (Pfizer-BioNTech). Patients were considered fully vaccinated since the second dose of BNT162b2 was administered at least 14 days before the patients were positive in the standard PCR test for SARS-CoV-2. All patients were confirmed to be positive for SARS-CoV-2, with a Ct between 19 and 30. Data collected included patient demographics, risk factors for SARS-CoV-2 disease, and clinical information related to the time of SARS-CoV-2 diagnosis (Table S1). Patients were not immunocompromised. This study was approved by the Ethics Committee and the Institutional Review Board of the Fundación Jiménez Díaz (FJD) hospital (no. PIC-087-20-FJD).

Oligonucleotide design. The oligonucleotide primers used for viral RNA amplifications and nucleotide sequencing were designed on the basis of a total of 663 SARS-CoV-2 sequences from NCBI database ([https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/](https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/)). They were retrieved and aligned to the Wuhan-Hu-1 NCBI reference sequence NC_045512.2 (12) (sequences to design oligonucleotides are described in Table S2). Six pairs of oligonucleotides were used for the amplification and sequencing of the overlapping amplicons corresponding to the end of nsp12 (polymerase) genomic coding region and the genomic region of the spike (S) protein (nucleotides 21,424 to 23,666; residue numbering is according to reference sequence NC_045512.2). The nucleotides analyzed encode amino acids 7062 to 7096 of the nsp12 (polymerase), and amino acids 1 to 694 of S (Table S3).

RNA extraction and viral RNA amplification of SARS-CoV-2 from infected patients. SARS-CoV-2 RNA of vaccinated patients was extracted and amplified from 140 μl of medium from nasopharyngeal swabs using the QIAamp Viral RNA Mini Kit (250) (Qiagen), following the indications of the manufacturer. Amplifications were performed from 5 μl of the purified RNA solution mixed with 10 μl of 5x buffer, and 2 μl of forward and 2 μl of reverse PCR primers (50 ng/μl) and 1 μl for polymerase for each amplicon using Transcriptor One Step RT-PCR kit (Roche Applied Science). The RT-PCR parameters were 50 °C for 30 min for the reverse transcription, an initial denaturing step at 94 °C for 7 min, followed by 45 cycles of a denaturing step at 94 °C for 10 s, an annealing step at 46-48 °C for 30 s, an extension step at 68 °C for 40 s, and then a final extension at 68 °C for 7 min. Amplifications in the absence of RNA were performed in parallel, as negative controls. No amplification was observed in any of the negative
control runs, and no Iota, Delta Plus or Omicron SARS-CoV-2 variants were epidemiologically relevant in Spain prior to or during April, 2021. The amplification products were analyzed by 2% agarose gel electrophoresis, including the Gene Ruler 1 Kb Plus DNA ladder (Thermo Scientific) as molar mass standard. PCR products were purified using the QIAquick Gel Extraction Kit (Qiagen), and then quantified employing the Qubit dsDNA Assay kit (Thermofisher Scientific), and, finally, tested for quality (TapeStation System, Agilent Technologies), prior to nucleotide sequencing using the Illumina MiSeq platform.

**Ultra-Deep Sequencing of SARS-CoV-2 from infected patients.** To obtain DNA pools, PCR products were adjusted to 4 x 10⁹ molecules/μl, and they were purified using Kapa Pure Beads (Kapabiosystems, Roche). Pools quantifications were performed using Qubit as previously described, and then adjusted to 1.5 ng/μl. DNA pools were processed using the DNA library preparation kit Kapa Hyper Prep kit (Roche), during which each pool was indexed using SeqCap Adapter Kit A/B (Nimblegen) (24 Index). Final DNA pools was quantified by LightCycler 480, and sequenced using MiSeq sequencing platform with MiSeq Reagent kit v3 (2 x 300 bp mode with the 600 cycle kit) (Illumina).

**Bioinformatics analyses.** Basal error, recombination frequency and reproducibility of results were previously performed (13). Given the sequence coverage with 37,311 to 197,230 clean reads per amplicon and patient (Table 1), mutations considered for the analysis were those with a frequency above a 0.1% cut-off value. For the characterization of the SARS-CoV-2 mutant spectra, the Fastq data was analyzed using the SeekDeep pipeline (14) with the following options: --extraExtractorCmds=--checkRevComplementForPrimers --primerNumOfMismatches 3" "--extraProcessClusterCmds=--fracCutOff 0.001 --rescueExcludedOneOffLowFreqHaplotypes”.

**Data availability.** Fastq files of SARS-CoV-2 samples included in the patient cohort are available in ENA under project id “PRJEB49400”.

**Study approval**

This study was approved by the Ethics Committee and the Institutional Review Board of the FJD hospital (no. PIC-087-20-FJD).

**Author contributions**
CP, CPJ conceived and designed the study. BMG, MES, LVS performed the experiments. PM, CL, BS and RR implemented the computational methods. CPJ, LSV, JE, RFR and IG (I. Gadea) provided viral samples and Ct values. LSV, CA and JRH provided clinical data. CGC, IG (I. Gallego), AIA analyzed the data. All authors contributed to writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by Instituto de Salud Carlos III, Spanish Ministry of Science and Innovation (COVID-19 Research Call COV20/00181), and co-financed by European Development Regional Fund ‘A way to achieve Europe’. The work was also supported by grants CSIC-COV19-014 from Consejo Superior de Investigaciones Científicas (CSIC), project 525/C/2021 from Fundació La Marató de TV3, PID2020-113888RB-I00 from Ministerio de Ciencia e Innovación, BFU2017-91384-EXP from Ministerio de Ciencia, Innovación y Universidades (MCIU), PI18/00210 and PI21/00139 from Instituto de Salud Carlos III and S2018/BAA-4370 (PLATESA2 from Comunidad de Madrid/FEDER). This research work was also funded by the European Commission – NextGenerationEU (Regulation EU 2020/2094), through CSIC’s Global Health Platform (PTI Salud Global). C.P. and P.M. are supported by the Miguel Servet programme of the Instituto de Salud Carlos III (CPII19/00001 and CP16/00116, respectively) cofinanced by the European Regional Development Fund (ERDF). CIBERehd (Centro de Investigación en Red de Enfermedades Hepáticas y Digestivas) is funded by Instituto de Salud Carlos III. Institutional grants from the Fundación Ramón Areces and Banco Santander to the CBMSO are also acknowledged. The team at CBMSO belongs to the Global Virus Network (GVN). B.M.-G. is supported by predoctoral contract PFIS FI19/00119 from Instituto de Salud Carlos III (Ministerio de Sanidad y Consumo) cofinanced by Fondo Social Europeo (FSE). C.G.-C. is supported by predoctoral contract PRE2018-083422 from MCIU. BS was supported by a predoctoral research fellowship (Doctorados Industriales, DI-17-09134) from Spanish MINECO.

Competing interests

The authors declare no competing financial interests.

References


Figure 1. (A) Heat maps of mutations that define SARS-CoV-2 variants Alpha, Delta Plus, Iota as described by WHO and Omicron as described in (15) (top heat map) and the corresponding mutations detected in our cohort (n=5) (bottom heat map). Amino acid substitutions are displayed in the top row, and their presence in variants Alpha, Delta Plus, Iota and Omicron is indicated by a filled square. The frequency of the substitutions within the mutant spectrum of each sample (with patient identification code on the left of each row) is color coded, as specified in the bottom box. Substituted amino acids that map at the same position but that are not identical with the substitutions reported for that position in the Delta Plus, Iota, and Omicron variants are indicated in red at the bottom heat map (from left to right: F157L, K417R, L452P, and T547A). Mutations and deletions have been identified relative to Wuhan-Hu-1 NCBI reference sequence NC_045512.2. (B) Representation of the six spike (S) amplicons used to perform UDS analysis, with indication of the relevant protein domains: signal peptide (SP), N-terminal domain (NTD), the receptor binding domain (RBD) and the S1/S2 cleavage site (S1/S2). Flanking black boxes indicate the amino acids (aa) of the S covered by the amplicons. Mutations and deletions that coincide with the Alpha, Delta Plus, Iota and Omicron variants found in our patient cohort are indicated.
Table 1. Number of clean reads obtained in SARS-CoV-2 from post-vaccination infected patients (identification code given in the column at the left) distributed according to amplicons. Reads refer to the number of sequences analyzed by ultra-deep sequencing (UDS) in MiSeq platform (procedure detailed in Methods).

<table>
<thead>
<tr>
<th>Patient</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt449</td>
<td>62,684</td>
<td>56,031</td>
<td>69,219</td>
<td>68,754</td>
<td>177,739</td>
<td>152,919</td>
<td>587,346</td>
</tr>
<tr>
<td>Pt450</td>
<td>68,515</td>
<td>50,946</td>
<td>68,990</td>
<td>52,383</td>
<td>151,637</td>
<td>143,877</td>
<td>536,348</td>
</tr>
<tr>
<td>Pt451</td>
<td>79,752</td>
<td>77,568</td>
<td>89,701</td>
<td>97,455</td>
<td>197,230</td>
<td>102,124</td>
<td>643,830</td>
</tr>
<tr>
<td>Pt452</td>
<td>83,893</td>
<td>65,296</td>
<td>90,668</td>
<td>80,073</td>
<td>111,557</td>
<td>124,134</td>
<td>555,621</td>
</tr>
<tr>
<td>Pt453</td>
<td>139,720</td>
<td>67,217</td>
<td>69,649</td>
<td>37,311</td>
<td>114,977</td>
<td>140,903</td>
<td>569,777</td>
</tr>
<tr>
<td>Average</td>
<td>86,912</td>
<td>63,411</td>
<td>77,645</td>
<td>67,195</td>
<td>150,628</td>
<td>132,791</td>
<td></td>
</tr>
</tbody>
</table>