Received: 20 December 2021 | Accepted: 16 March 2022

DOI: 10.1002/jmv.27726

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

Low-frequency variants in mildly symptomatic vaccine breakthrough infections presents a doubled-edged sword

Brittany R. Magalis[1,2] | Carla Mavian[1,2] | Massimiliano Tagliamonte[1,2] | Shannan N. Rich[1,3,4] | Melanie Cash[1,2] | Alberto Riva[5] | Julia C. Loeb[1,6] | Michael Norris[1,6] | David M. Amador[5] | Yanping Zhang[5] | Jerne Shapiro[3,4] | Petr Starostik[2] | Simone Marini[1,3] | Paul Myers[4] | David A. Ostrov[2] | John A. Lednicky[1,6] | J. Glenn Morris Jr.[1,7] | Michael Lauzardo[1,4,7] | Marco Salemi[1,2]

[1]Emerging Pathogens Institute, University of Florida, Gainesville, Florida, USA

[2]Department of Pathology, University of Florida, Gainesville, Florida, USA

[3]Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, Florida, USA

[4]Florida Department of Health, Gainesville, Florida, USA

[5]Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, Florida, USA

[6]Department of Environmental and Global Health, University of Florida, Gainesville, Florida, USA

[7]Division of Infectious Diseases and Global Medicine, College of Medicine, University of Florida, Gainesville, Florida, USA

Correspondence

Marco Salemi, Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610, USA.

Email: salemi@pathology.ufl.edu

Funding information

National Science Foundation

Abstract

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern (VOC) has raised questions regarding vaccine protection against SARS-CoV-2 infection, transmission, and ongoing virus evolution. Twenty-three mildly symptomatic "vaccination breakthrough" infections were identified as early as January 2021 in Alachua County, Florida, among individuals fully vaccinated with either the BNT162b2 (Pfizer) or the Ad26 (Janssen/J&J) vaccines. SARS-CoV-2 genomes were successfully generated for 11 of the vaccine breakthroughs, and 878 individuals in the surrounding area and were included for reference-based phylogenetic investigation. These 11 individuals were characterized by infection with VOCs, but also low-frequency variants present within the surrounding population. Low-frequency mutations were observed, which have been more recently identified as mutations of interest owing to their location within targeted immune epitopes (P812L) and association with increased replicative capacity (L18F). We present these results to posit the nature of the efficacy of vaccines in reducing symptoms as both a blessing and a curse—as vaccination becomes more widespread and self-motivated testing reduced owing to the absence of severe symptoms, we face the challenge of early recognition of novel mutations of potential concern. This case study highlights the critical need for continued testing and monitoring of infection and transmission among individuals regardless of vaccination status.

KEYWORDS

epidemiology, genetic variation, genetics, SARS coronavirus, virus classification

Brittany R. Magalis and Carla Mavian are co-first authors.

1 | INTRODUCTION

Several rapidly spreading variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) rose to the status of "variants of concern" (VOC), according to the Centers for Disease Control (CDC, https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html), accumulating several unique genomic mutations with respect to the original viral strain originating from the Wuhan province of China.[1] These variants have been demoted to Variants Being Monitored (VBM) in the wake of the Omicron variant (also referred to as B.1.1.529/BA[2]),[3] largely responsible for 15 million new cases as of January 2022 according to the World Health Organization, though collectively they have been responsible for over 900 000 deaths in the United States alone. The VBM referred to as Alpha (or B.1.1.7/Q), was one of the first variants to be identified as of potential concern, carrying the notable mutations N501Y and P681H in the spike protein. Alpha emerged in September of 2020 in Kent, UK, but quickly made its way across the globe,[4] including at least two separate introductions into the United States.[4] One introduction was estimated to occur early November 2020, into California, and was characterized by a virus resembling more closely the traditional UK variant, whereas the other, differing by a single mutation (C15720T) was estimated to have been introduced in late November. The expanding virus populations originating from these two events were referred to as "Clade 1" and "Clade 2." Beta (B.1.351) was first detected around the same time (summer of 2020) in South Africa[5] and was subsequently observed in seven other countries by the end of that year, though its presence was limited in the United States (data available from https://covid.cdc.gov/covid-data-tracker/#variant-proportions). Beta carried the N501Y mutation observed in Alpha, along with two additional mutations—K417N and E484K—deemed mutations of "high concern" owing to their ability to compromise antibody neutralization.[6] The Gamma variant (P.1), emerging in Brazil in November of that year carried similar (K417T, E484K, and N501Y) in the Spike protein, resulting also in rapid spread across the globe.[7] The two Epsilon (B.1.427/B.1.429) lineages, collectively dubbed the "California variant", are often left unmentioned in the context of VOCs, as the variant was ultimately unable to compete with the more transmissible Alpha variant following its emergence in California in July 2020.[8] However, its notable L452R mutation in spike was shared by the subsequent Delta (B.1.617/AY) variant that became the predominant lineage worldwide in 2021.[3,9]

The emergence of these variants and their level of evolutionary divergence from the original strain has raised questions regarding the extent of protection of currently implemented vaccines against infection with future arising variants.[10–13] Approximately 61.6% of the world population has received at least one dose of a coronavirus disease 2019 (COVID-19) vaccine. However, vaccination coverage is uneven across countries, as only 10.6% of people in low-income countries have received at least one dose. For example, Haiti vaccinated approximately 0.9% of the country's population, while the Dominican Republic (Haiti's neighboring country on the other half of the Island of Hispaniola) vaccinated 53.5%.[14] Vaccination efforts have largely focused on the spike protein, as it is structurally important for coronaviruses, rendering it susceptible to recognition by the host immune response. Existing vaccines specifically target the receptor-binding domain (RBD),[15–17] though there is evidence that the host immune system can also target the N-terminal domain (NTD).[6,18] The natural accumulation of mutations in these regions owing to the error-prone viral replication machinery can lead to the ability of the virus to evade the host immune response, including that of vaccinated individuals, and may be responsible for enhanced transmissibility of the variant.[19] With the growing evidence of SARS-CoV-2's genome plasticity since the beginning of the pandemic,[20] tracking of mutations in these regions of the genome and investigation into their relevance for current and upcoming vaccines is critical. Although vaccination protects against hospitalization with the associated disease (COVID-19), since the campaign began, vaccination breakthrough cases have been reported globally. As of April 30, 2021, 10 262 SARS-CoV-2 vaccine breakthrough infections had been reported in the US, where approximately 101 million individuals had been fully vaccinated, and approximately 355 000 COVID-19 cases were reported nationally during the week of April 24–30, 2021.[21] Though demonstrated to be both highly effective against infection during initial clinical trials (>90% for Pfizer's BNT162b2[15] and 66.3% for the J&J/Janssen[22]), the question of the role of emerging mutations in current and future breakthrough infections remains an important one.

2 | METHODS

2.1 | Participant involvement

UF Health Screen, Test & Protect (STP) assists the Florida Department of Health in Alachua County with COVID-19 case and contact tracing efforts in its UF students, faculty, staff, and other UF-affiliated people including the UF Health Academic Medical Center (~123 000 total UF Affiliates). Full epidemiological investigations were conducted on positive cases to collect exposure information, trace contacts, and provide disease transmission education. Fully vaccinated individuals who became a contact (defined as ≥15 min and closer than 6 feet) were called and provided public health education but not placed into quarantine. For 14 days after their last exposure, they received a daily text or email to record symptom development. Immediate polymerase chain reaction (PCR) testing was recommended for anyone newly reporting symptoms and Day 7 PCR testing was recommended for all fully vaccinated but exposed individuals, regardless of symptom development. Individuals working in healthcare settings may have been sampled more frequently due to internal hospital policies.

2.2 | Criteria of inclusion

Fully vaccinated UF affiliates deemed PCR-positive for SARS-CoV-2 were eligible for molecular epidemiology investigation as part of the STP program if they met the following criteria:

1) The case must be infectious at the time of saliva sample donation. Infectiousness was defined ≤10 days after the onset of symptoms or for an asymptomatic individual ≤10 days after the positive lab collection date.
2) The case must meet the definition of a vaccine-breakthrough case. Defined as PCR-positive for SARS-CoV-2 and ≥14 days after the second dose of Pfizer or Moderna or first dose of Janssen/J&J.

If both criteria were met and individuals volunteered to provide a sample for the purpose of public health molecular surveillance, they were scheduled to arrive on-site for sample collection as soon as possible to increase the probability of detectable virus at the time of collection.

2.3 | Sample collection and processing

Each participant was asked to give a saliva sample of at least 2 ml in total volume and was instructed not to drink anything for 10 min before giving the sample. The saliva was collected in a 15 ml conical tube, filling it to the 2 ml marking on the tube, not including froth. Patient samples were deidentified following Institutional Review Board approval before viral processing.

2.4 | RNA extraction and library preparation

Viral RNA was extracted from 180 µl of each saliva sample using the QIAamp 96 Viral RNA Kit with the QIAcube HT (Qiagen) using the following settings with a filter plate: the lysed sample was premixed eight times before subjecting to vacuum for 5 min at 25 kP and vacuum for 3 min at 70 kPa. Following three washes using the same vacuum conditions above, the samples were eluted in 100 µl AVE buffer followed by a final vacuum for 6 min at 60 kPa. Nine microliters of RNA was used for cDNA synthesis and library preparation using the COVIDSeq Test kit (Illumina) and Mosquito HV Genomics Liquid Handler (SPT Labtech Inc.). The size and purity of the library were determined using the 4200 TapeStation System (Agilent) and the Qubit dsDNA HS Assay Kit (Life Technologies) according to the manufacturer's instructions. Constructed libraries were pooled and sequenced using the NovaSeq. 6000 Sequencing System SP Reagent Kit and the NovaSeq Xp 2-Lane Kit. Illumina's DRAGEN pipeline was used to derive sample consensus sequences, which were filtered based on a minimum of 70% coverage of the genome.

2.5 | Database sequence retrieval and phylogenetic analysis

Each Floridian sequence was used in a local alignment (BLAST)[23] search for the most (genetically) similar non-Floridian sequence in the GISAID database as of June 12, 2021, and linked to two reference sequences including the best match (highest E-value) with a date occurring within one month following, as well as 1 month before the sampling date of the Floridian sequence.[24] This method was termed "FLACO-BLAST,"[24] and the script is available from https://github.com/salemilab/flaco. After removing duplicate sequences (sequences with same GISAID ID), sequences were aligned in viralMSA[25] using the MN908947 reference sequence, and mutations potentially associated with contamination, recurrent sequencing errors, or hypermutability were masked using a vcf filter (https://virological.org/t/masking-strategies-for-sars-cov-2-alignments/480). The maximum likelihood phylogenetic relationship with IQ-TREE software[26] based on the best-fit model according to the Bayesian Information Criterion (BIC) with ultrafast bootstrap approximation.[27]

2.6 | Modeling mutations in the SARS-CoV-2 spike glycoprotein

The cryoEM structure of the SARS-CoV-2 (Wuhan-Hu-1) spike protein complexed to a neutralizing human antibody (4A8)[28] bound to the N-terminal domain of the Spike protein (PDB 7C2L) was used as the basis for structural analyses and modeling L18F. Sidechains were mutated in COOT[29] using rotamers that represent a local energy minimum of torsional angles.

2.7 | Cell culture for viral infectivity

Vero E6 cells were used for SARS-CoV-2 isolation attempts. The cells had been obtained from the American Type Culture Collection (catalog no. ATCC CRL-1586) and have been used for our SARS-CoV-2 projects,[30–33] including the isolation of >30 SARS-CoV-2 isolates from human and environmental samples. The cells were propagated in a cell culture medium comprised of advanced Dulbecco's modified essential medium (aDMEM; Invitrogen) supplemented with 10% low antibody, heat-inactivated, gamma-irradiated fetal bovine serum (FBS; Hyclone, GE Healthcare Life Sciences), L-alanine, L-glutamine dipeptide supplement (GlutaMAX,), and 50 µg/ml penicillin, 50 µg/ml streptomycin, 100 µg/ml neomycin (PSN antibiotics, Invitrogen) with incubation at 37°C in 5% CO_2.

2.8 | Isolation of virus in cultured cells

Virus isolation attempts were performed in a BSL3 laboratory at the University of Florida Emerging Pathogens Institute (EPI) by analysts who wore powered air-purifying respirators and engaged in BSL3 work practices. Vero E6 cells grown as monolayers in a T-25 flask (growing surface 25 cm^2) were inoculated when they were at 80% of confluency as follows: for each flask, the spent cell culture medium was removed and replaced with 1 ml of supplemented aDMEM medium ("complete medium") with 10% FBS, and the cells inoculated with 50 µl of unfiltered saliva. Before inoculation, samples were frozen (−80°C) within 15 min following collection. Samples were only

thawed once to produce an aliquot for processing. The inoculated cell cultures were incubated at 37°C in 5% CO_2, and rocked every 15 min for 1 h, after which 4 ml of complete medium with 10% FBS was added. The following day, the cell culture media was completely removed and replenished with 5 ml of maintenance medium (complete medium with 3% FBS). Mock-infected cell cultures were maintained in parallel with the other cultures. The cell cultures were refed every 3 days by the replacement of 2 ml of spent media with a maintenance medium. The cells were observed daily for 1 month before being judged negative for virus isolation, with a blind passage performed 15 days postinoculation of the cells. When virus-induced cytopathic effects (CPE) were evident, the presence of SARS-CoV-2 in the cell culture medium was examined by real-time reverse transcription PCR (rRT-PCR). In the event that SARS-CoV-2 strains that were not cytolytic or did not produce CPE had been isolated, the culture media were blindly tested at weekly intervals.

2.9 | Detection of SARS-CoV-2 genomic RNA (vRNA) in cell culture medium by rRT-PCR

vRNA was extracted from virions in collection media in a Class II biosafety cabinet in a BSL3 laboratory at the EPI by analysts wearing appropriate personal protective equipment (chemically impervious Tyvek lab coats and gloves) and using powered-air purifying respirators. The vRNA was extracted from 140 µl aliquots of the collection media using a QIAamp Viral RNA Mini Kit (Qiagen), and purified RNA eluted from the RNA-binding silicon column in a volume of 80 µl. Twenty-five microliter (final volume) rRT-PCR tests were performed in a BioRad CFX96 Touch Real-Time PCR Detection System using 5 µl of purified vRNA and the N1 and N2 primers and their corresponding probes of the CDC 2019-Novel Coronavirus (2019-nCoV) rtRT-PCR test.[34] The primers and probes were purchased from Integrated DNA Technologies (IDT). A plasmid that encodes the SARS-CoV-2 N-gene sequence was purchased from IDT and used in positive control reactions for the CDC N1 and N2.

The rRT-PCR tests were performed using the following parameters: 400 nM final concentration of forward and reverse primers and 100 nM final concentration of probe using a SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (Thermo Fisher Scientific). Cycling conditions were 20 min at 50°C for reverse transcription, followed by 2 min at 95°C for Taq polymerase activation, then 45 cycles of denaturation for 15 s at 95°C, annealing for 30 s at 55°C, and extension at 68°C for 20 s.

3 | RESULTS

3.1 | Vaccine breakthrough infection rate and epidemiology

From January to June 2021, Alachua County, Florida, experienced a reduced number of infections (~11% drop in positive cases), with a general increase in a number of vaccinations, particularly since the last week of February (Figure S1). Regardless of the reduced COVID-19 burden in the county, continued testing of vaccinated individuals has provided insight into susceptibility of vaccines to circulating viral variants within the region. Between February and June 2021, 23 individuals identified by the UF Health STP program tested positive using the standard PCR-based assay, as well as via antigenic testing for one sample (STP-Saliva-1678), for SARS-CoV-2 infection after completing the BNT162b2 mRNA two-dose vaccine series.[15] Individuals were considered fully vaccinated two weeks after their second vaccination in a two-dose series of Moderna or Pfizer or after one dose of J&J. By June 3, 2021, more than 130 000 individuals were vaccinated, of which more than 113 000 received complete series of Pfizer, Moderna, or J&J/Janssen vaccines in Alachua County, Florida.[35] The number of infections among vaccinated individuals suggested a relatively low breakthrough rate consistent with the reported efficacy of Pfizer's BNT162b2[15] and J&J/Janssen's[22] vaccines demonstrated in earlier clinical trials. The twenty-three "vaccine-breakthrough" individuals all experienced symptom onset between 0 and 97 (mean = 35.4) days following the final dose, sixteen (70%) of whom presented with symptoms characteristic of COVID-19 (Tables 1 and S1). The majority of these individuals identified as female (78%), White (74%), under the age of 35 (mean = 33.1), and reported their occupation as healthcare-related (56.5%), including students working in patient care. Case investigations revealed that 14 (61%) of these individuals were exposed to individuals previously identified as COVID-19 cases in the 2 weeks before their disease onset, with 10 reporting the nature of the relationship as household and four as community/social. Hence, despite an increased risk of exposure in the healthcare setting, infection in these individuals was likely the result of inadvertent exposure during social contact. Contact tracing performed for 21 of these 23 breakthrough cases reported no secondary cases, indicating at least limited onward transmission at the time of the study.

3.2 | Genomic sequencing of vaccine breakthrough samples

Secondary saliva samples were collected from the 23 vaccine breakthrough individuals within 3–7 days of testing positive for infection (Tables 1 and S1). Saliva collection was used in this study for the isolation of viral RNA, as this bodily fluid has demonstrated prolonged presence of viral RNA (up to 25 days postsymptom onset[36]), irrespective of disease severity.[37] The research protocol was approved by the University of Florida institutional review board. SARS-CoV-2 full-genome (>70%) sequences were successfully generated for 11 of the vaccinated individuals. A panel of additional respiratory viruses was also targeted during sequencing (Table S2), confirming the absence in all patients of additional infection with common viruses, such as influenza. Thus, clinical symptoms, when present, were likely the result of productive coronavirus infection. The reason for insufficient SARS-CoV-2 genome quality for the

TABLE 1 Reported cases of vaccination breakthroughs in Alachua County, Florida

Characteristic	
Age (years)	33.1 (13.2)
Sex	
Female	18 (78.3%)
Race	
White	17 (73.9%)
African American/Black	2 (8.7%)
Asian/Pacific Islander	3 (13.0%)
Ethnicity	
Hispanic	4 (17.4%)
Symptoms (Y)	16 (69.6%)
Fever	2 (12.5%)
Cough	4 (25.0%)
Dyspnea	0 (0.0%)
Anosmia	5 (31.3%)
Ageusia/dysgeusia	3 (18.8%)
Sore throat	8 (50.0%)
Headache	7 (30.4%)
Runny nose	11 (47.8%)
Fatigue	7 (43.8%)
Comorbidities	
Asthma	1 (4.3%)
Immunocompromised	1 (4.3%)
Former smoker	1 (4.3%)
BMI	26.7 (6.2)
Known exposure (Y)	14 (60.9%)
Household	10 (71.4%)
Community	4 (28.6%)
Occupation	
Healthcare worker	10 (43.5%)
Student engaged in patient care	3 (13.0%)
None or nonsensitive occupation	10 (43.5%)
Time between second vaccination dose and disease onset (days)	35.4 (24.8)
Time between disease onset (for symptomatic cases) and sample collection date (days)	4.4 (2.1)

Note: Results are presented as frequency (%) for categorical variables and mean (standard deviation) for continuous variables.

remainder of the vaccine breakthrough cases is not fully clear. Results from COVID-19 testing for the majority of these individuals were limited to qualitative data (positive or negative), though the number of PCR cycles (Cq) required for reliable viral RNA amplification for two of the individuals was provided and already >25 at the time of diagnosis (Table S3), which has been considered relatively high for genomic sequencing.[38] Given the time between diagnosis and saliva sampling (Table S1), saliva viral load may have been too low for the genomic amplification required for amplicon-based sequencing for some individuals. While mutations in primer-binding regions can reduce genome coverage for amplicon-based sequencing,[38] coverage mapping did not indicate this phenomenon – alternating regions of high- and low-coverage were not observed (Figure S2). Two of these individuals were, however, asymptomatic whereas asymptomatic individuals were not found among the successful full-genome sequences (Table S2), supporting the link between severity of symptoms and rapidity of viral clearance.[39]

3.3 | Molecular epidemiology of Florida SARS-CoV-2 infections

To understand the context of the vaccine breakthrough infections, we assembled a full-genome sequence data set ($n = 8619$) including sequences generated from hospital samples in Alachua County between January and June 2021 ($n = 889$), Miami area ($n = 485$), and central Florida ($n = 62$), and a total of epidemiologically relevant sequences from Florida ($n = 5693$) and the global population ($n = 1490$) deposited into the GISAID database (https://gisaid.org). Epidemiological relevance was defined on an individual sequence basis, restricting the global GISAID search to the two sequences most similar genetically to each Floridian sequence and sampled within a high-confidence transmission time window (30 days) based on sample collection date, as described in Giovanetti et al.[40] GISAID IDs and corresponding submission information are provided in Table S4. Owing to low genetic variability, as well as potentially shared epidemiological linkages, sequences retrieved from GISAID were often shared by more than one query (Florida) sequence, resulting in a total of 1490 non-Floridian sequences, of which 83% were located within the Unites States, 9% from Europe; the rest of the Americas, Africa, and Asia contributed with less than 5%. The resulting data set spanned October 10, 2020–June 2, 2021.

3.4 | SARS-CoV-2 lineage distribution in Florida

Lineages for all sequences were determined using the PangoLEARN model (Pangolin v 3.1), which was trained using approximately 60 000 GISAID SARS-CoV-2 sequences to classify incoming sequences based on molecular and epidemiological criteria.[3] The full data set in this study was characterized by a total of 146 lineages, for which 12 lineages (11 B.1.X, and Gamma) represented >99% of samples (Figure1). The distribution of lineages for Floridian sequences outside of Alachua County largely resembled that of the non-Floridian reference sequences, as expected given the filtering approach for genetic similarity described above; the exception to this similarity was the presence of the B.1.375 lineage within the Florida

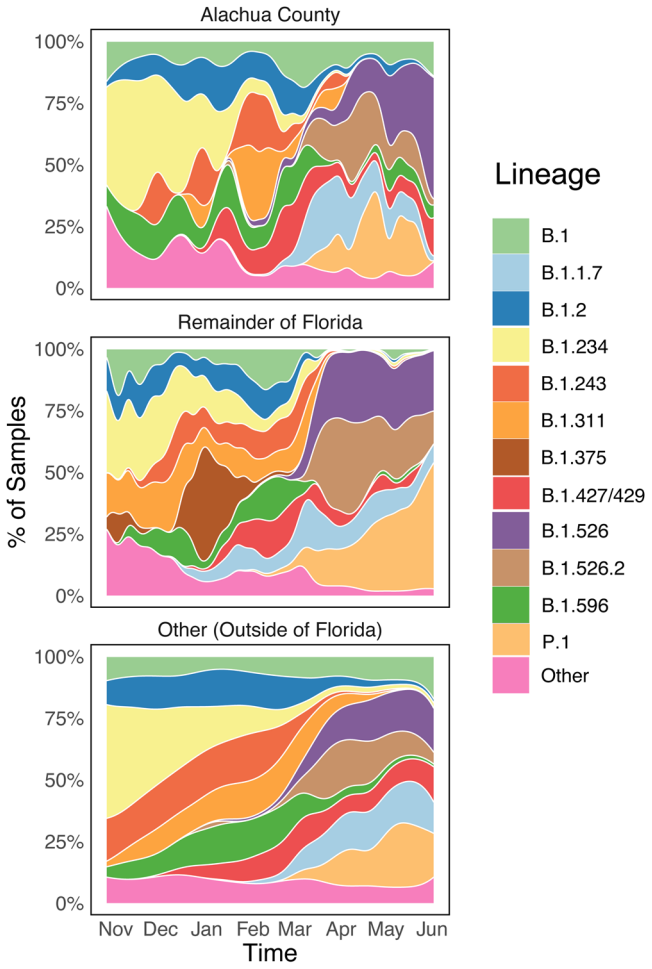

FIGURE 1 Distribution of identified lineages. Lineage as identified using Pangolin[2] distribution over time for Alachua County, the surrounding Florida areas, and locations outside of Florida linked to Florida sequences via genetic similarity

data set, which was not present among reference sequences (Figure 1A). As this lineage was not observed among the Alachua County sequences, further investigation into the potential misclassification of these sequences as the B.1.375 lineage, more notably associated with the northern states,[41] was considered outside the scope of this study. Whereas both the Alachua County and remaining Floridian samples were dominated by the B.1.234 lineage in November 2020, both regions quickly expanded to include at least seven other lineages in January 2021 (Figure 1). This expansion included a growing presence of the B.1.1.7 (or Alpha VOC) and B.1.427/429 (Epslion VOC) lineages (Figure 1). Mirroring the pandemic outside of Florida, within Florida and in Alachua County, six lineages were dominating the epidemic between March and June 2021: B.1, Alpha, Epslion, Iota (B.1.526, B.1.526.2), and Gamma; the latter decreasing in Alachua County by the end of June, while accounting for approximately 50% of the samples in the remainder of Florida (Figure 1). Given the growth in the Alpha and Epsilon variants within the Florida population, it is not surprising that of the 11 vaccinated individuals reported in this study. Approximately half of the breakthrough infections belonged to Alpha ($n = 3$) and Epsilon ($n = 2$). The remaining vaccinated individuals, however, presented with the high-frequency B.1 and B.1.2 lineages covering approximately 10% and 25% of the breakthrough infections, respectively. Low-frequency B.1.377 and B.1.596 lineages, not considered VOCs at the time of the study, were responsible for less than 1% of breakthrough infections.

3.5 | Evolutionary relationships of vaccine breakthrough sequences in the context of Florida infection

Following lineage classification, a maximum likelihood phylogenetic tree was reconstructed from the sequences to verify lineage classification and to determine relationships among vaccinated individuals in the context of geographical and temporal information (Figure 2A). Bootstrap replicates for the sequence data were used to provide support for branching patterns within the tree,[27,42,43] and the smallest (number of sequences), well-supported (>90% of replicates) clade involving each vaccination-breakthrough case was examined individually. Whereas 6 of the 11 successfully sequenced, vaccinated individuals belonged to relatively small, definable clades (Figure 2B–F), reliable placement of five individuals (two B.1, two B.1.2, and one B.1.596 lineage) within the tree could not be obtained, despite >90% coverage of the genome (Figure S2). However, these sequences (STP-VTM-513: B.1 lineage, STP-Saliva-1337: B.1.2 lineage, STP-Saliva-1582: B.1 lineage, STP-Saliva-1678: B.1.596 lineage, and STP-Saliva-1680: B.1.2 lineage) did share common ancestry with sequences of similar lineage, supporting proper lineage assignment (Figure S3).

3.6 | Detection of low-frequency variant transmission among vaccine breakthrough cases

The two individuals harboring the Epsilon variant belonged to a clade of 28 individuals comprised of additional Epsilon variants, confirming lineage classification using Pangolin (Figure 2B). The remaining 26 sequences within this clade consisted primarily of Floridian sequences (nine from Alachua County), with the exception of one individual from Minnesota, suggesting largely local transmission of this variant, though directionality of transmission could be inferred. The two vaccine-breakthrough cases with the Epsilon VOC in this study (STP-Saliva-412 and STP-Saliva-413) reported separate exposures; the two individuals were also not more closely related to each other phylogenetically than the remainder of the sequences (with significant support), so we could neither confirm nor exclude the possibility of a relationship via direct transmission between these two individuals specifically.

Approximately 36.8% of Alpha individuals within the total Florida sample, and 61.9% of Alpha in Alachua County, harbored the "Clade 2" mutation, including three Alpha-lineage vaccine-breakthrough cases.

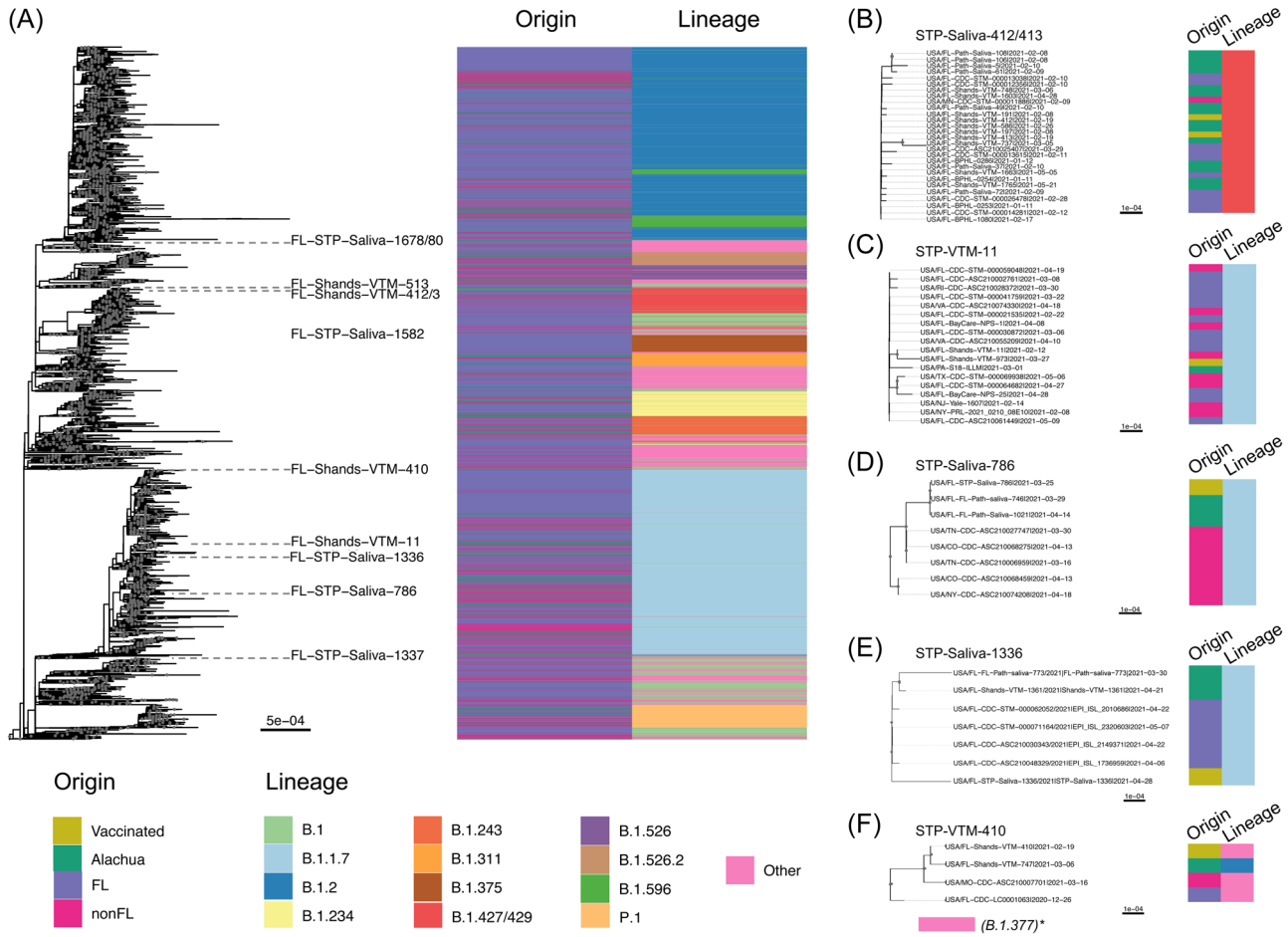

FIGURE 2 Geographical origin of sampling over time and within the phylogenetic tree of SARS-CoV-2 data collected from Florida and relevant non-Florida locations. (A) Distribution of both assigned lineages and geographic origin (as in Figure 1) across the maximum likelihood phylogenetic tree. Branches are scaled in genetic substitutions/site, and nodes with ≥90% support using bootstrap sampling are indicated by gray dots. Vaccinated individuals within well-supported clades have been emphasized, and corresponding clades are represented as insets in panels (B–F). *Other lineages, defined as present within <1% of the total sample population. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

These cases are clustered in three distinct, well-supported clades. The STP-VTM-11 sample belonged to a well-supported clade in this study comprised of 22 total individuals (Figure 2C), all sharing the Alpha lineage designation (again confirming pangolin classification) and "Clade 2" mutation, though originating from other parts of Florida (10), Virginia (2), Pennsylvania (1), New York (1), New Jersey (1), Rhode Island (1), Tennessee (1), and Texas (1), consistent with the widespread presence of Alpha in the United States at the time.[4] The individual was reportedly exposed to a recent COVID-19 case outside of the household. The other Alpha-lineage vaccine-breakthrough cases—STP-Saliva-786 (Figure 2D) and STP-Saliva-1336 (Figure 2E)—similarly clustered together with other Alpha-classified genomes, though sharing more recent ancestry with Alachua County, indicating local spread. Contact tracing of these three Alpha-lineage vaccine-breakthrough cases' exposed contacts did not identify any secondary cases associated with vaccine failure.

The final well-supported clade containing the low-frequency B.1.377 vaccinated individual (STP-VTM-410) shared significant common ancestry with a single Floridian B.1.377 sequence, confirming lineage classification and suggesting transmission within Florida, but not confined to the county (Figure 2F).

3.7 | Mutational profiles of vaccine breakthrough variants

Even though infection in fully vaccinated individuals was not limited to VOC, mutational analysis was necessary to determine if (1) recently acquired mutations (i.e., not conserved among the corresponding lineage) could potentially be responsible for limited protection of the vaccine against infection or (2) evidence existed for the emergence of vaccine-resistant variants. The B.1, B.1.377, and Epsilon variants found within breakthrough-vaccinated individuals did not appear to have acquired any additional Spike mutations outside of those associated with the parental lineage, suggesting newly acquired mutations were not responsible for breakthrough infections and that these individuals did not harbor a

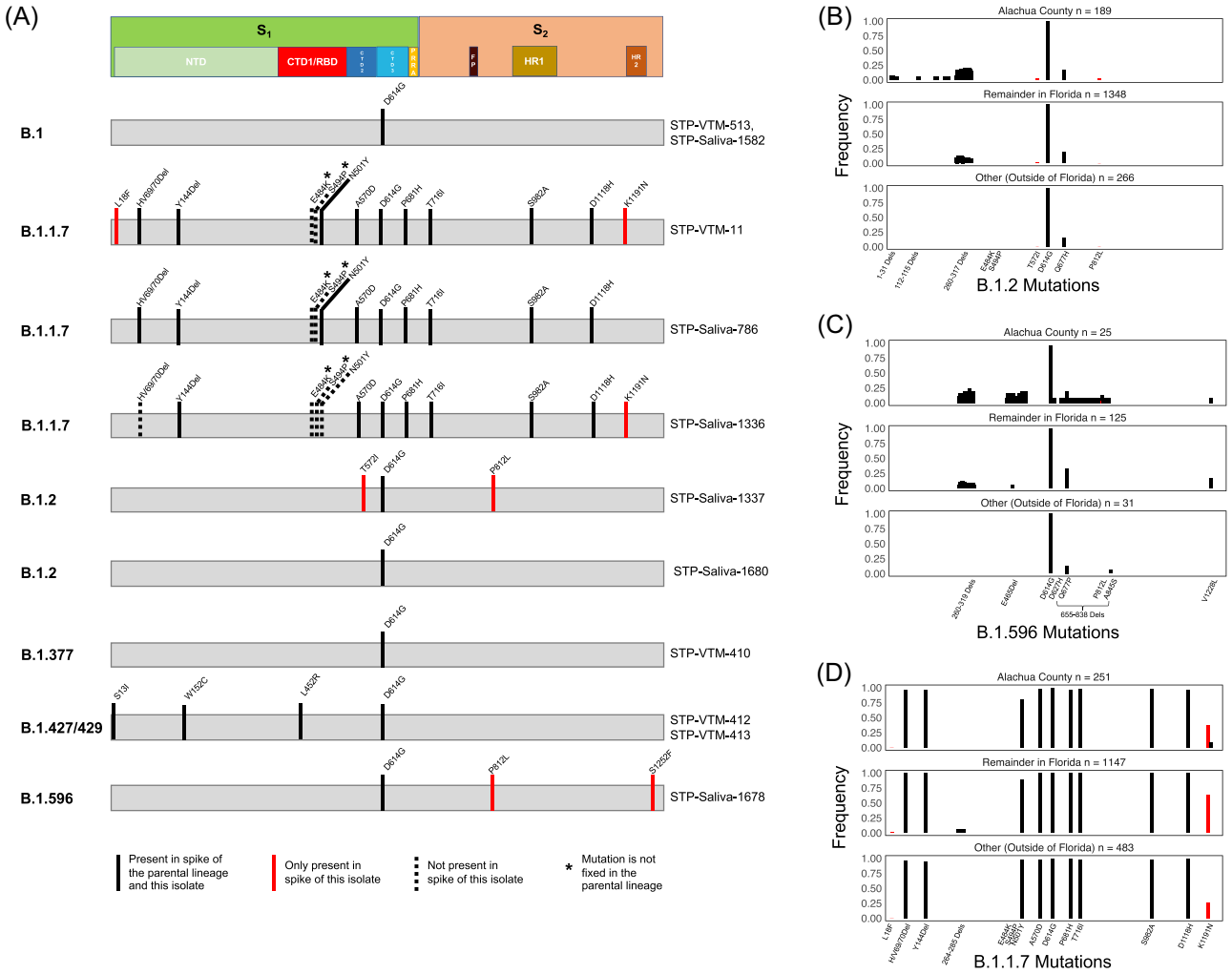

FIGURE 3 Mutations of interest and of concern in Spike found in the 11 vaccine-breakthrough individuals and their frequencies across Alachua, remainder, and outside of Florida. Mutational profiles for vaccine-breakthrough individuals (A) and comparison of (B) B.1.2 and (C) Alpha individuals with remaining sequence data. Mutations in red are present in vaccinated individual (VAC), but not fixed in the parental lineage. Spike protein architecture is displayed at the top of both panels, wherein CTD (1–3), C-terminal domain; FP, fusion peptide; HR (1–2), heptad repeat; NTD, N-terminal domain; PRRA, SARS-CoV-2 characteristic PRRA insertion at the S1/S2 cleavage site; RBD, receptor-binding domain; S, spike subunit (1–2). Positions numbers are relative to the spike protein in the MN908947 reference sequence

vaccine-resistant variant (Figure 3A). However, three vaccine breakthrough cases presented unique mutations relative to the parental lineage. The first mutation of interest was observed in two breakthrough-vaccinated cases—STP-Saliva-1337 (B.1.2) and STP-Saliva-1678 (B.1.569) (Figures 3A and S4). The mutation P812L (amino acid coordinates relative to spike starting position, Figure 3A) was found at very low frequencies (less than 4% within Alachua County, and <1% in the remainder of Florida and United States (Figure 3B,C). This mutation was first detected in India in 2020[44] and is located within a predicted CD4 + T cell epitope[45] within the Spike S2' cleavage site, which plays an essential role in the interaction of the virus with the host cell receptor[46] (Figure S4A,B). Two distinct lineages harboring P812L within this Florida cohort (Figure 3) suggest convergent evolution of a mutation that may play a crucial role in modulating tropism and pathogenicity of the virus. Although this mutation remains at low frequencies in the current population, its location and distribution warrant further studies and surveillance.

A B.1.2-lineage vaccination breakthrough case—STP-Saliva-1337—presented with an additional unique mutation: T572I within the fraction of the spike protein referred to as CTD2. This mutation was similarly found at low frequency (<3%, six individuals in total) among Alachua samples, as well as the remainder of the Florida and US sequences (<1%) (Figure 3C). In vitro assays revealed that mutation T572I reduced the exposure of the RBD and binding efficacy to the ACE2 receptor[47]; our structural analysis of T572I also shows that the residue is located in a position that may cause interference with intermolecular contact (Figure S4C). Changes in these regions involved in the interface between SARS-CoV-2 spike proteins are expected to influence protein cleavage, protein structural rearrangement, and host cell membrane fusion and are one example of the types of mutations involved in recent SARS-CoV-2 adaptation.[48]

The vaccinated individual STP-VTM-11 presented two unique, and previously not described, mutations within the spike protein relative to

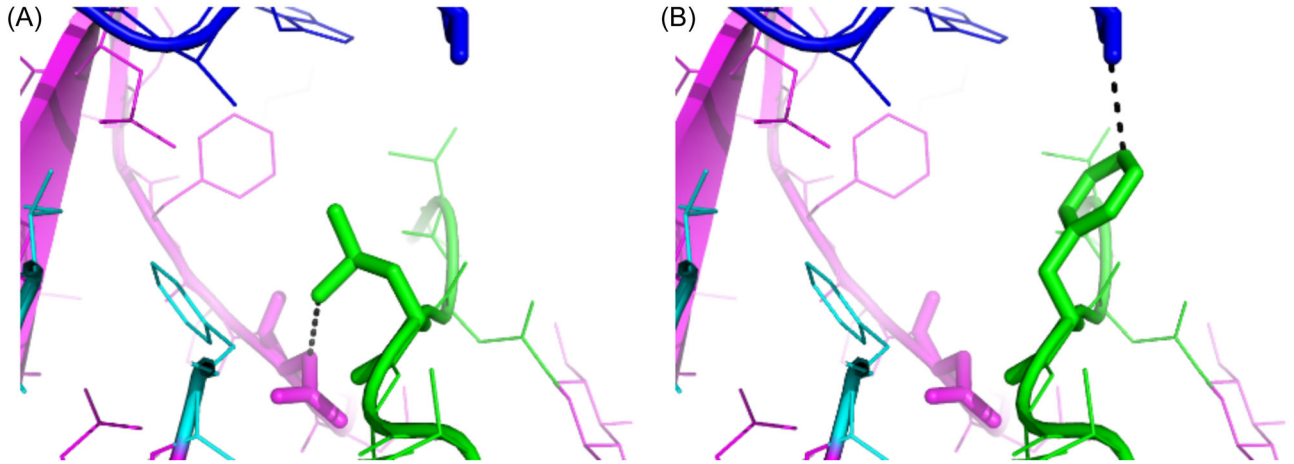

FIGURE 4 Potential effect on trimer stabilization of Floridian Alpha variant mutation of interest from lysine (A) to phenylalanine (B) at position 18 of the Spike protein N-terminal domain (NTD). Loops N1, N2, and N5 of the NTD are represented in green, pink, and blue, respectively. Modified interactions as a result of the mutation are represented as dotted lines, with original L18 oriented toward position F79 (cyan) of the N1 loop (A), and the variant F18 toward S252 of the N5 loop (B), potentially acting to stabilize. Dotted lines represent distances 3.8 (A) and 3.6 (B) Angstrom. Positions numbers are relative to the Spike protein in the MN908947 reference sequence

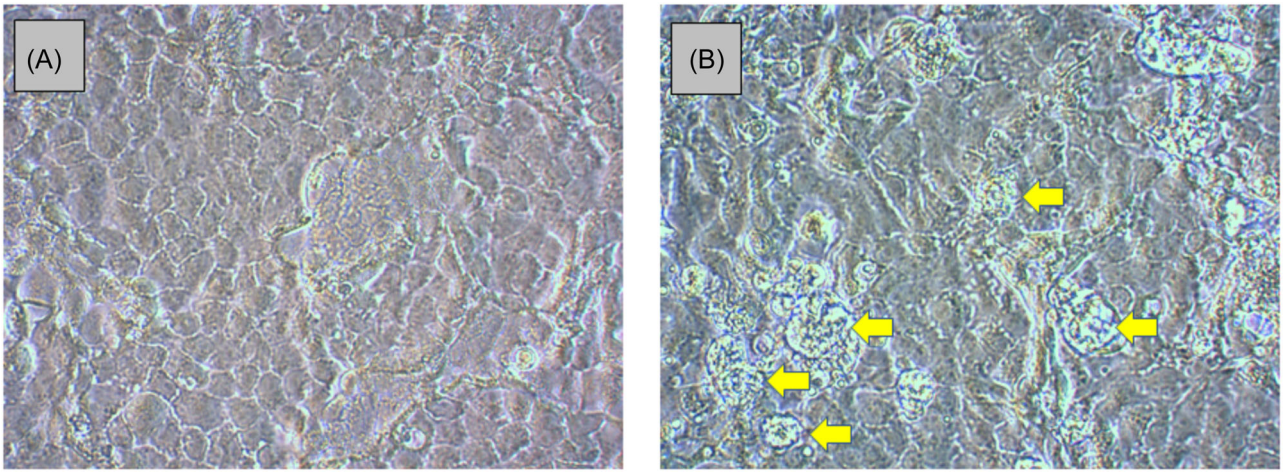

FIGURE 5 Cytopathic effects in Vero E6 cells inoculated with saliva sample FL-STP-VTM-11. (A) Mock-infected Vero E6 cells, 12 dpi. (B) Early SARS-CoV-2-specific CPE, 12 dpi. Rounded cells, some in the process of detaching from the growth surface, are pointed out by yellow arrows. Original magnification at 400X. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

the parental Alpha lineage[1] - L18F and K1191N within the NTD and C-terminal domain (CTD), respectively (Figure 3A). HR2 K1191N mutation was also observed in a second vaccinated individual (STP-Saliva-1336); yet, the frequency of the mutation among the rest of the genomes (~37% of samples for Alachua County, ~63% of the remainder of Florida samples, and ~27% of US and international sequences) suggested a less recent acquisition of this change in the population and no relation to vaccine-mediated adaptation or an enhanced ability to infect vaccinated individuals (Figure 3D). Alternatively, the NTD L18F mutation (structural diagram in Figure 4) was far less prevalent, only observed in 16 individuals (1.4%) from Florida (dating back to January 15, 2021) (Figure 3A). Among the rest of United States and international genomes, only one genome collected in Minnesota (EPI_ISL_2375747) on May 15, 2021, presented with this mutation at the time. Though recent studies have posited involvement of this mutation in an ascribed nearly twofold replicative advantage for the spread of the Alpha in the UK,[49] experimental evidence confirming infectivity of this particular variant has not been presented. Infectivity of the virus isolated from individual STP-VTM-11 was measured in vitro, revealing viability of the L18F mutation, defined by the presence of virus-specific cytopathic effects (CPE) on Vero E6 cells (African green monkey kidney cells), first noted 12 days postinoculation (dpi) (Figure 5). Quantification of the virus at 12 dpi for the STP-VTM-11 sample-infected cells and control (no viral inoculation) confirmed viral replication (C_q 8.14 and $C_q > 39$, respectively),

4 | DISCUSSION

A previous study reported that individuals vaccinated between January and March 2021 in Israel were disproportionally infected with VOCs relative to the unvaccinated population[50]; subsequently, during the month of June 2021, Israel faced an outbreak of the Delta variant (B.1.617.2 lineage) among those fully inoculated with the Pfizer's vaccine which prompted the government to reimpose indoor mask requirement and other measures to contain the new variant.[51] Whereas we cannot exclude the impact of prevalence of VOCs within the Florida population on their rate of breakthrough, the findings of our study, collectively with,[50] indicate limited protection of the BNT162b2 mRNA vaccine (and potentially others) against not only VOC but also emerging, low-frequency variants of SARS-CoV-2. It is important to note that none of the breakthrough individuals in this study was hospitalized, corroborating the vaccine's 100% efficacy (at the time) against severe disease caused by currently known variants.[15] While these individuals presented with only mild symptoms (or no symptoms at all), the number of vaccine breakthrough cases might be expected to be under-reported. Hence, if we assume at least a minority of test-positive vaccinated individuals harbor infectious virus, the potential for hidden reservoirs within the global populations is increased. Hidden reservoirs in asymptomatic or mild symptomatic individuals, as has been proposed,[24] pose a particular threat to early recognition of novel mutations of potential concern, such as the L18F and P812L described herein, particularly once vaccination is more widespread. Moreover, given the effectiveness of the vaccine in limiting symptom presentation, vaccinated individuals may engage more frequently in social activities, increasing the risk of exposure. Continued testing and case management, assessing contacts and exposure, for vaccinated individuals is thus encouraged and will be forthcoming in determining whether the vaccine is protective against the ongoing spread of SARS-CoV-2 variants that may emerge in these individuals. This strategy is particularly important in the face of relaxed guidelines regarding masked protection and social distancing for vaccinated individuals.

AUTHOR CONTRIBUTIONS

Brittany Rife Magalis and Carla Mavian wrote the manuscript and carried out data analyses with help from Massimiliano Tagliamonte, Shannan N. Rich, Alberto Riva, and David A. Ostrov. Melanie Cash, Julia C. Loeb, Michael Norris, David Moraga Amador, and Yanping Zhang performed the experiments to generate the data. Jerne Shapiro, Petr Starostik, Simone Marini, Paul Myers, John A. Lednicky, J. Glenn Morris Jr., and Michael Lauzardo helped supervise the project. Marco Salemi conceived of the original idea and supervised the project.

ACKNOWLEDGMENTS

This study was supported by the Stephany W. Holloway University of Florida Chair and by funds of the University of Florida Office of Research and Health Science Center with resources from the Interdisciplinary Center for Biotechnology Research Gene Expression Core (RRID:SCR_019145), NextGen Sequencing Core (RRID:SCR_019152) and Bioinformatics Core (RRID:SCR_019120). Funding for this study was also provided by National Science Foundation (NSF)—Division of Environmental Biology (DEB) award no. 2028221.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from GISAID (https://www.gisaid.org). GISAID identifiers are supplied in Table S4.

ORCID

Brittany R. Magalis https://orcid.org/0000-0001-6088-4651
Carla Mavian https://orcid.org/0000-0003-3819-5520

REFERENCES

1. Rambaut A, Loman N, Pybus O, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virologicalorg*; 2020. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563
2. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020;5(11):1403-1407.
3. CDC. *SARS-CoV-2 Variant Classifications and Definitions*. Accessed on February 9, 2022. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html
4. Larsen B, Worobey M. Phylogenetic evidence that B.1.1.7 has been circulating in the United States since early- to mid-November. *Virologicalorg*. 2021. (https://virological.org/t/phylogenetic-evidence-that-b-1-1-7-has-been-circulating-in-the-united-states-since-early-to-mid-november/598)
5. Tegally H, Wilkinson E, Giovanetti M, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021;592(7854):438-443.
6. Greaney AJ, Loes AN, Crawford KHD, 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):463-476e6.
7. Faria NR, Mellan TA, Whittaker C, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 2021;372(6544):815-821.
8. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E. Emergence of a novel SARS-CoV-2 variant in Southern California. *JAMA*. 2021;325(13):1324-1326.
9. Adam D. What scientists know about new, fast-spreading coronavirus variants. *Nature*. 2021;594(7861):19-20.
10. Hitchings MDT, Ranzani OT, Lind ML, et al. Effectiveness of CoronaVac in the setting of high SARS-CoV-2 P.1 variant transmission in Brazil: a test-negative case-control study. *medRxiv*. 2021:21255081.
11. Madhi SA, Baillie V, Cutland CL, et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 vaccine against the B.1.351 variant. *N Engl J Med*. 2021;384:1885-1898.
12. Buss LF, Prete CA Jr, Abrahim CMM, et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science*. 2021;371(6526):288-292.
13. Sabino BD, Alonso FOM, Oliveira MSC, Venceslau MT, Guimarães MAAM, Varella RB. Long-term intermittent detection

of SARS CoV 2 in the upper respiratory tract: what is the meaning of it? *Infect Dis (Lond)*. 2021;53(2):151-153.
14. University CaJH. *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*; 2019. Accessed February 9, 2022. https://github.com/CSSEGISandData/COVID-19.
15. Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med*. 2020;383(27):2603-2615.
16. Baden LR, El Sahly HM, Essink B, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med*. 2021;384(5):403-416.
17. Sadoff J, Le Gars M, Shukarev G, et al. Interim results of a phase 1-2a trial of Ad26.COV2.S Covid-19 vaccine. *N Engl J Med*. 2021;384:1824-1835.
18. Suryadevara N, Shrihari S, Gilchuk P, et al. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell*. 2021;184:2316-2331.
19. Ostrov D. Structural consequences of variation in SARS-CoV-2 B.1.1. *Scientific Archives*. 2021;3(2):103-108.
20. Velazquez-Salinas L, Zarate S, Eberl S, Gladue DP, Novella I, Borca MV. Positive selection of ORF1ab, ORF3a, and ORF8 genes drives the early evolutionary trends of SARS-CoV-2 during the 2020 COVID-19 pandemic. *Front Microbiol*. 2020;11:11.
21. COVID-19. Vaccine breakthrough infections reported to CDC - United States, January 1-April 30, 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(21):792-793.
22. CDC. Johnson & Johnson's Janssen COVID-19 Vaccine Overview and Safety. Accessed June 27, 2021. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/janssen.html
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410.
24. Giovanetti M, Cella E, Benedetti F, et al. SARS-CoV-2 shifting transmission dynamics and hidden reservoirs limited the efficacy of public health interventions in Italy. *medRxiv*. 2020:20248355.
25. Moshiri N. ViralMSA: massively scalable reference-guided multiple sequence alignment of viral genomes. *Bioinformatics*. 2020
26. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268-274.
27. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35(2):518-522.
28. Chi X, Yan R, Zhang J, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science*. 2020;369(6504):650-655.
29. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 4):486-501.
30. Lednicky JA, Lauzard M, Fan ZH, et al. Viable SARS-CoV-2 in the air of a hospital room with COVID-19 patients. *Int J Infect Dis*. 2020;100:476-482.
31. Lednicky JA, Cherabuddi K, Tagliamonte MS, et al. In-frame 12-nucleotide deletion within open reading frame 3a in a SARS-CoV-2 strain isolated from a patient hospitalized with COVID-19. *Microbiol Resour Announc*. 2021;10(8).
32. Carbone M, Lednicky J, Xiao SY, Venditti M, Bucci E. Coronavirus 2019 infectious disease epidemic: where we are, what can be done and hope for. *J Thorac Oncol*. 2021;16(4):546-571.
33. Lednicky JA, Lauzardo M, Alam MM, et al. Isolation of SARS-CoV-2 from the air in a car driven by a COVID patient with mild illness. *Int J Infect Dis*. 2021;108:212-216.
34. CDC. CDC real-time reverse transcriptase PCR (rRT-PCR) diagnostic test for COVID-19, 2021.
35. CDC. COVID-19: vaccine summary. Accessed June 27, 2021. http://ww11.doh.state.fl.us/comm/_partners/covid19_report_archive/vaccine/vaccine_report_latest.pdf
36. To KK, Tsang OT, Leung WS. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis*. 2020;20(5):565-574.
37. Zhu J, Guo J, Xu Y, Chen X. Viral dynamics of SARS-CoV-2 in saliva from infected patients. *J Infect*. 2020;81(3):e48-e50.
38. Charre C, Ginevra C, Sabatier M, et al. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol*. 2020;6(2):veaa075.
39. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *Lancet Microbe*. 2021;2(1):e13-e22.
40. Giovanetti M, Cella E, Benedetti F, et al. SARS-CoV-2 shifting transmission dynamics and hidden reservoirs potentially limit efficacy of public health interventions in Italy. *Commun Biol*. 2021;4(1):489.
41. Moreno G, Braun K, Larsen B, et al. Detection of non-B.1.1.7 Spike Δ69/70 sequences (B.1.375) in the United States. *Virologicalorg*. 2021. https://virological.org/t/detection-of-non-b-1-1-7-spike-69-70-sequences-b-1-375-in-the-united-states/587
42. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530-1534.
43. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 2013;30(5):1188-1195.
44. Begum FB, Ray AK. U. Mutation hot spots in spike protein of SARS-CoV-2 virus. *Preprints*. 2020. 2020040281.
45. Peng Y, Mentzer AJ, Liu G, et al. Broad and strong memory. *Nat Immunol*. 2020;21(11):1336-CD45.
46. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*. 2020;181(2):281-292.
47. Juraszek J, Rutten L, Blokland S, et al. Stabilizing the closed SARS-CoV-2 spike trimer. *Nat Commun*. 2021;12(1):244.
48. Ostrov DA, Knox GW. Emerging mutation patterns in SARS-CoV-2 variants. *Biochem Biophys Res Commun*. 2021;586:87-92.
49. Grabowski F, Preibisch G, Giziński S, Kochańczyk M, Lipniacki T. SARS-CoV-2 variant of concern 202012/01 has about twofold replicative advantage and acquires concerning mutations. *Viruses*. 2021;13(3).
50. Kustin T, Harel N, Finkel U, et al. Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2 mRNA vaccinated individuals. *medRxiv*. 2021:21254882.
51. Lieber D. *Delta Variant Outbreak in Israel Infects Some Vaccinated Adults*. The Wall Street Journal; 2021.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Magalis BR, Mavian C, Tagliamonte M, et al. Low-frequency variants in mildly symptomatic vaccine breakthrough infections presents a doubled-edged sword. *J Med Virol*. 2022;94:3192-3202. doi:10.1002/jmv.27726