Selective adaptation of SARS-CoV-2 Omicron under booster vaccine pressure: a multicentre observational study

Ralf Duerr,[a,b,c,*] Dacia Dimartino,[d] Christian Marier,[d] Paul Zappile,[d] Guiqing Wang,[e] Fritz François,[b] Mila B. Ortigoza,[a,b] Eduardo Iturrate,[b] Marie I. Samanovic,[b,c] Mark J. Mulligan,[a,b,c] and Adriana Heguy[d,e,**]

[a]Department of Microbiology, NYU Grossman School of Medicine, USA
[b]Department of Medicine, NYU Grossman School of Medicine, USA
[c]Vaccine Center, NYU Grossman School of Medicine, New York, NY, USA
[d]Genome Technology Center, Office of Science and Research, NYU Langone Health, USA
[e]Department of Pathology, NYU Grossman School of Medicine, USA

eBioMedicine
2023;97: 104843
Published Online 20 October 2023
https://doi.org/10.1016/j.ebiom.2023.104843

Summary

Background High rates of vaccination and natural infection drive immunity and redirect selective viral adaptation. Updated boosters are installed to cope with drifted viruses, yet data on adaptive evolution under increasing immune pressure in a real-world situation are lacking.

Methods Cross-sectional study to characterise SARS-CoV-2 mutational dynamics and selective adaptation over >1 year in relation to vaccine status, viral phylogenetics, and associated clinical and demographic variables.

Findings. The study of >5400 SARS-CoV-2 infections between July 2021 and August 2022 in metropolitan New York portrayed the evolutionary transition from Delta to Omicron BA.1-BA.5 variants. Booster vaccinations were implemented during the Delta wave, yet booster breakthrough infections and SARS-CoV-2 re-infections were almost exclusive to Omicron. In adjusted logistic regression analyses, BA.1, BA.2, and BA.5 had a significant growth advantage over co-occurring lineages in the boosted population, unlike BA.2.12.1 or BA.4. Selection pressure by booster shots translated into diffuse adaptive evolution in Delta spike, contrasting with strong, receptor-binding motif-focused adaptive evolution in BA.2-BA.5 spike (Fisher Exact tests; non-synonymous/synonymous mutation rates per site). Convergent evolution has become common in Omicron, engaging spike positions crucial for immune escape, receptor binding, or cleavage.

Interpretation. Booster shots are required to cope with gaps in immunity. Their discriminative immune pressure contributes to their effectiveness but also requires monitoring of selective viral adaptation processes. Omicron BA.2 and BA.5 had a selective advantage under booster vaccination pressure, contributing to the evolution of BA.2 and BA.5 sublineages and recombinant forms that predominate in 2023.

Funding The study was supported by NYU institutional funds and partly by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center.

Copyright © 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: SARS-CoV-2; Omicron; Variant of concern (VOC); Booster vaccinations; Vaccine pressure; Breakthrough; Re-infections; Selective adaptation; Convergent evolution; Spike RBD; Receptor-binding motif; Imprinted immunity; Clinical variables

Introduction

The SARS-CoV-2 pandemic has been a showcase for viral evolution in real-time. Variants of concern (VOC) Alpha, Beta, Gamma, Delta, and, more recently, Omicron evolved and increased rapidly despite worldwide vaccination efforts.[1–4] In the United States, vaccinations started in late December 2020, with four vaccines currently employed: two mRNA-based vaccines, BNT162b2 (Pfizer/BioNTech) and mRNA-1273 (Moderna), the adenovirus-based Janssen COVID-19 vaccine,

*Corresponding author. Department of Medicine, NYU Grossman School of Medicine, Alexandria Center for Life Science (ACLS), West Tower, 430 East, 29th Street, Room 314, New York, NY, 10016, USA.
**Corresponding author. Department of Pathology, NYU Grossman School of Medicine, Genome Technology Center, NYU Langone Health, 550 First Avenue, MSB 294A, New York, NY, 10018, USA.
E-mail addresses: ralf.duerr@nyulangone.org (R. Duerr), adriana.heguy@nyulangone.org (A. Heguy).

Research in context

Evidence before this study

References for this Original Research article were searched on PubMed, medRxiv, and bioRxiv for SARS-CoV-2 literature published from December 2019 to the time of submission, and for general articles in virology and vaccinology without date restriction, by use of the terms "SARS-CoV-2", "COVID-19", "vaccination", "booster", "immunization", "viral evolution", "Omicron", "Delta", "breakthrough", "mutations", "vaccine escape", "selective adaptation", "convergent evolution", "epistasis", and "epidemiology". Articles published in English, Spanish, French, and German were included in our search, with a focus on English articles. In 2022, Omicron became the predominant SARS-CoV-2 variant causing a massive wave globally. Omicron carries an unusually high amount of mutations compared to the original virus or other variants of concern (VOC), which has been linked to Omicron's fast transmission and efficient vaccine escape. Since Omicron replaced Delta in early 2022, Omicron keeps diversifying and evolving into subvariants and descendants thereof, while at the same time, recurrent mutations at key functional sites, particularly in spike, have been observed in numerous lineages. This process, called convergent evolution, has been linked to growing but incomplete immunity in the human population. It is sustained by the antigenic distance between the initial immune imprinting by vaccines/primary infections and the contemporary SARS-CoV-2 challenges, according to the concept of original antigenic sin. A detailed view of Omicron's mutational forces and selective adaptation processes in humans under increasing vaccine pressure is lacking.

Added value of this study

Here, we characterized the mutational dynamics and selective adaptation processes of SARS-CoV-2 Omicron in a multicentre cohort in metropolitan New York. The study comprises more than 5400 sequences collected over more than a year (July 2021–August 2022) that cover the replacement of Delta by Omicron and its evolution into BA.1 to BA.5 sublineages. During this period, primary series-vaccinated individuals increased from ~55% to 80% and boosted individuals from zero to ~40%. Fine-scale variant and mutation analysis included the statistical comparison of full genome base pair and amino acid mutations in unvaccinated, primary series-vaccinated, and boosted individuals, and logistic regression analyses of Omicron subvariants by vaccine status. These data revealed that boosters but not primary vaccination were the main source of vaccine pressure leading to mutational imprints, adaptive evolution, and variant selection during the transition from Delta to Omicron and evolution into Omicron subvariants. BA.1, BA.2, and BA.5 had selective advantages under booster vaccine pressure, but not BA.2.12.1, a variant first detected in New York, or BA.4. Booster vaccinations exerted strong mutational pressure on Delta that resulted in diffuse, low-level adaptive evolution. In contrast, Omicron BA.1, and even more BA.2-BA.5 variants had undergone extensive diversification and acquired a critical set of immune escape, including convergent mutations, that precluded further significant mutations after their introduction into metropolitan New York.

Implications of all the available evidence

Booster shots are a key component in public health to increase immunity against infectious agents. They fill in immunity gaps that typically result from adaptations of a virus in response to increasing immune pressure by vaccination and/or natural infection. SARS-CoV-2 and Influenza are prime examples of these adaptive processes, and as we enter a new respiratory season with updated boosters, knowledge of immunity gaps and their underlying processes of adaptive viral evolution are crucial. Since its introduction in early 2022, SARS-CoV-2 BA.2 has constituted the framework for an extensive diversification of Omicron into BA.2-BA.5 lineages. BA.2 and BA.5's epidemiological advantages in the boosted population were a strong foundation that drove the evolution of BA.2 and BA.5 sublineages and recombinants predominating in 2023. Vaccine pressure and the genomic background of a virus are critical co-factors that predict the path of adaptive evolution under epistatic rules. High mutational divergence under booster vaccine pressure in Delta did not translate into high rates of adaptive evolution, whereas Omicron variants, equipped with strong marks of adaptive evolution, exhibited modest mutational divergence in boosted versus unvaccinated individuals. Recurrent mutations across variants, which helped to shape spike immune escape, receptor binding, and cleavage, are a hallmark of SARS-CoV-2's convergent evolution in the ongoing, diversifying BA.2-BA.5 wave. The critical extent of viral diversification and immune escape justifies the use of booster shots, with a need for monitoring of selective adaptation processes. Booster vaccinations for vulnerable people and healthcare workers remain a priority.

JNJ-78436735, and most recently, the protein subunit vaccine, NVX-CoV2373 (Novovax).[5–8] The vaccines are highly effective at preventing symptomatic disease, hospitalisation, and death,[9] though with Omicron, protection from symptomatic disease dropped to <70% efficacy after receipt of the primary vaccination series, and large numbers of vaccine breakthrough infections have been recorded.[10] In late September 2021, booster vaccinations became available in the US in high-risk individuals and occupational sectors; in mid-November 2021, the CDC expanded the eligibility for booster shots to all adults.[11] High immunity due to booster

vaccination, previous infection, or both (hybrid immunity) is a bottleneck that selects for transmissions with the most evasive variants. The bottleneck furthermore depends on host factors, co-morbidities, and the extent and duration of the viral challenge.[12]

Omicron was first detected in November 2021 in Botswana and South Africa and spread rapidly around the globe replacing Delta as the dominant variant.[4] The origin of Omicron remains a subject of ongoing investigation, yet it is theorized that its evolution took place within an immunocompromised host, subjected to persistent but imperfect immune pressure. Omicron acquired >60 amino acid mutations compared to the ancestral Wuhan-Hu-1 virus, more than double the number of Delta's defining set of amino acid mutations (<30), resulting in functional, evolutionary, and immunological changes. Omicron spike is less efficiently cleaved at S1/S2 compared to previous variants, thus shifting the cellular tropism away from TMPRSS2-expressing cells and membrane fusion towards an endosomal route of entry, more dependent on cathepsins.[13,14] In consequence, Omicron replicates faster than all previous variants in the upper airways but less in the lung. In line with the latter finding, epidemiological studies described a reduced clinical severity of Omicron, uncoupling clinical outcomes from skyrocketing infections.[15,16] The highly mutated Omicron genome caused epistatic shifts that reframed the evolutionary track of the virus.[17,18] There are principally two Omicron evolutionary paths, based on either BA.1 or BA.2 genomic background. BA.1 caused a massive but short wave, whereas BA.2 set the ground for an extended, highly diversifying BA.2-BA.5 evolution. By late 2022/early 2023, Omicron evolved into BA.2 and BA.5 variants that are antigenically as different to the ancestral SARS-CoV-2 as SARS-CoV-1.1[19] Consequently, Omicron subvariants broadly escape vaccine-mediated neutralization and most therapeutic neutralizing antibodies, facilitated by the antigenic distance between vaccine/primary infection-imprinted immunity and chronologically phased SARS-CoV-2 challenges.[19,20]

Rising breakthrough infections were recorded with Delta in mid-2021, however, they were not directly linked with the Delta variant *per se*, but rather with waning immune responses and minor adaptations within Delta subvariants.[21] Omicron is equipped with a significantly higher potential of immune escape compared to Delta with a >2x and >12x increase in 50% neutralization titers for vaccinated individuals who received the primary series with or without booster, respectively.[22] However, while Omicron's high capacity of immune escape is known, there is a lack of data on the evolutionary pressure and continuous adaptive evolution of Omicron under boosted vaccine pressure in a real-world environment. Here we characterized more than 5400 high-quality SARS-CoV-2 full genomes and associated demographic and clinical variables in a highly vaccinated, large multicentre cohort in metropolitan New York. The study covers the transition from Delta to Omicron BA.1—BA.5 subvariant evolution in a time of increasing immune pressure by booster vaccinations up to the point when bivalent booster vaccinations became available in the US in early September 2022.[23] Our study reveals that booster vaccinations provided discriminative immune pressure that was effectively evaded by Omicron variants, involving receptor-binding motif-focused adaptive evolution in BA.2 to BA.5 lineages.

Methods

Study design and sample collection

Clinical SARS-CoV-2 infections were studied in vaccinated, unvaccinated, and re-infected individuals collected between July 16th, 2021 and August 9th, 2022 in the multicentre NYU Langone Health (NYULH) system. The three NYU Langone Health centres included the Langone hospital in Manhattan, Midtown East, the Winthrop hospital on Long Island, Nassau County, and the Lutheran hospital in the highly diverse Sunset Park neighbourhood of Brooklyn. The observational, cross-sectional study was designed to cover the transition from Delta to Omicron including BA.1 to BA.5 variants, and the introduction of booster vaccinations (September 24th, 2021) until the plateauing phase and before bivalent boosters became available (September 2, 2022). Cases were identified using DataCore, NYULH's clinical data storage, management, and extraction resource. SARS-CoV-2 infection was defined by positive real-time (RT)-qPCR test for SARS-CoV-2 RNA. Participants were stratified by vaccine and infection status as follows, 1) unvaccinated: no COVID-19 vaccine shot by the time of SARS-CoV-2 infection; 2) partial vaccination: SARS-CoV-2 infection between first vaccine shot and <14 days after completion of the primary vaccination series; 3) primary series: at least 14 days after the second dose of BNT162b2 (Pfizer/BioNTech) or mRNA-1273 (Moderna) vaccines, or the single-dose COVID-19 Janssen vaccine; 4) boosted: at least one additional COVID-19 vaccine shot after completion of the primary series; 5) re-infected: SARS-CoV-2 infection in people with a previous SARS-CoV-2 infection, irrespective of vaccine status. Participants from groups 1 to 4 reported no prior SARS-CoV-2 infection. We sequenced all SARS-CoV-2-positive cases with Ct < 30 up to our maximum sequencing capacity of 94 cases per week. Nasopharyngeal swabs were sampled from individuals with exposure to SARS-CoV-2, suspected to have an infection with SARS-CoV-2, or as part of clinical diagnostics or hospital admission. Follow-up visits were not scheduled. Potential sampling and sequencing biases included an underrepresentation of asymptomatic cases or re-infections, particularly individuals with hybrid immunity, because of their stronger immunity, associated with lower viral load

upon re-infection, milder symptoms, reduced testing, and lower chances of obtaining high-quality sequences. Unequal viral introductions, founder effects, and demographic and clinical variables are possible confounders. Adjusted analyses were used to control for selected confounders.

This study was approved by the NYULH Institutional Review Board, protocol numbers i21-00493 and i21-00561. Overall outcome measures were SARS-CoV-2 full genome sequences, their mutations, lineage/variant designations, demographic, and clinical variables (binary, count, and categorical variables). All sexes were included in the study and sex was self-reported by study participants.

Blood samples collection for immunological assays

Twenty adults who received their primary vaccination series and one booster dose (a total of three doses of BNT162b2 vaccine) in accordance to CDC guidelines provided written consent for enrollment with approval from the NYU Institutional Review Board (protocols 18-02035 and 18-02037). Eight participants had a breakthrough infection, laboratory confirmed by RT-qPCR, and their blood was collected in average 21 days post-onset, 69 days post-booster dose. Blood from 12 control individuals, with no reported history of SARS-CoV-2 infection ("Naïve" group), was collected one month post-booster dose. Blood was collected by standard phlebotomy, in SST tubes (BD Biosciences). Serum was isolated and frozen immediately at –80 °C.

ELISA

Direct ELISA was used to quantify antibody endpoint titers (continuous variable) in participant serum as previously described.[24] S1 protein antigen concentration was 1 μg/mL for the ancestral and BA.1 variants (Sino Biological 40591-V08H and 40591-V08H41), and 0.5 μg/mL for the Delta variant (40591-V08H23). The outcome measure was a comparison of antibody titers among SARS-CoV-2 naïve, Delta-, and Omicron-infected individuals who previously received a booster vaccination.

RNA extraction, cDNA synthesis, library preparation and sequencing

Clinical testing was performed using various FDA emergency use authorization (EUA) assays for detection of SARS-CoV-2 RNA, i.e., the Roche Cobas 6800 SARS-CoV-2 (90% of the samples in this study), Cepheid Xpert SARS-CoV-2, or SARS-CoV-2/Flu/RSV assays.

RNA was extracted from 400 μl of each nasopharyngeal swab specimens using the MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit on the KingFisher flex system (Thermo Fisher Scientific). Eleven μl of RNA were used for first strand cDNA synthesis using the Superscript IV first-strand synthesis kit (Invitrogen, ref# 180901050). Libraries were prepared using Swift Normalase Amplicon SARS-CoV-2 Panel (SNAP) and SARS-CoV-2 additional Genome Coverage Panel (Cat# SN-5X296 core kit, 96rxn), using 10 μl of first strand cDNA, with 24 PCR cycles.[25] Libraries were pooled and run on the Illumina NovaSeq 6000 system on SP300 flow cells, as paired-end 150 cycles with dual indexing reads. Typically, two pools representing two full 96 well plates (192 samples) were sequenced on each SP300 NovaSeq flow cell. Reads were demultiplexed using the Illumina bcl2fastq2 Conversion Software v2.20, and adapters and low-quality bases were trimmed with Trimmomatic v0.36.[26] BWA v0.7.17[27] was utilized for mapping reads to the SARS-CoV-2 reference genome (NC_045512.2, wuhCor1). SNAP tiled primer sequences were removed with Primerclip v0.3.8.[28] BCFtools v1.9[29] was used to call mutations and assemble consensus sequences, which were then assigned phylogenetic lineage designations according to PANGO nomenclature, version 2021-09-28.[30] SARS-CoV-2 sequences that were <23,000 bp or <4000x genome coverage were considered inadequate and not included in the analyses. These quality controls reduced PCR and sequencing artifacts that however could not be fully excluded.

For the Illumina COVIDSeq test (catalog number #20049393), 8.5 μl of total RNA extracted from the swab were used as input. For optimal coverage of Omicron variants, the ARTIC V4.1 nCOV-2019 amplicon panel was spiked into the Illumina primer pool, following the Illumina Technical Note (M/GL/00691 v1.0). A pool of 96 samples was sequenced on an SP300 flow cell on the NovaSeq 6000. After demultiplexing with bcl2fastq2 v2.20, the FASTQ files were processed with the Illumina DRAGEN COVID Lineage v3.5.11 (https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/dragen-covid-lineage.html). In order to perform Kmer-based detection of SARS-CoV-2, read alignment to the SARS-CoV-2 reference genome (NC_045512.2), variant calling, consensus genome sequence generation, and lineage/clade analysis, Pangolin (https://cov-lineages.org/resources/pangolin.html) and NextClade (https://docs.nextstrain.org/projects/nextclade/en/stable/index.html) were used.

Phylogenetic analyses

Time-calibrated maximum-likelihood (IQ) trees were generated using the Nextstrain CLI package WSL on Windows.[31] As input, we used 5432 high-quality NYULH sequences and 1714 North-America-focused global reference sequences from GISAID[32] until the end of the study period (August 9th, 2022). The trees were constructed using a focal-contextual subsampling, focusing on sequences from NYU Langone Health as originating lab and prioritized by proximity to the focal data set. Except for the minimum length of 23,000 bp per sequence, default parameters were used, including the masking of the first 100 and last 50 bp of the SARS-CoV-2 alignment, restricting the sequences to a minimum date of "2019-12-01", and creating a

maximum-likelihood time-calibrated tree under a general time-reversible (GTR) substitution model (1 thread) and a fixed molecular clock with an evolutionary rate of 0.0008 substitutions per site per year (0.0004 standard deviations), rooted to Wuhan-Hu-1. The base maximum-likelihood tree was generated with IQ-tree and then scaled over time using TreeTime.[31]

Mutation and adaptive evolution analyses

Mutations compared to Wuhan-Hu-1 as reference were determined on MAFFT-aligned SARS-CoV-2 sequences in program R v.4.1.0[33] and R Studio v.1.4.1106[34] using scripts based on the seqinr and tidyverse packages. The statistical analyses considered residues that were covered by sequencing, i.e., non-ambiguous characters and gaps, whereas ambiguous and undefined characters were excluded. Full genome and spike base pair/amino acid counts and site-specific mutation frequencies were analysed in program R using Fisher exact tests with multiplicity corrections (Benjamini-Hochberg).[35] P values < 0.05 were considered significant. Adaptive evolution in a coding gene was tested using a fast, unconstrained Bayesian approximation for inferring selection (FUBAR, Datamonkey).[36,37] Outcome measures were the comparison of vaccinated (primary series/booster) and unvaccinated individuals regarding the frequency of mutations per full-genome basepair site and the non-synonymous (dN) and synonymous (dS) substitution rates per spike site for the given coding alignment and corresponding phylogeny.

Mutational effects on antibody immune escape and ACE2-binding affinity

The predicted effect of mutations on neutralizing antibody immune escape were determined with the SARS-CoV-2 receptor-binding domain (RBD) escape calculator. It calculates the difference in polyclonal antibody binding by all mutations in the SARS-CoV-2 RBD in a given virus compared to BA.2.[38] The calculations are based on deep mutational scanning of RBD targeting antibodies.[19] An escape value of 0 means all binding is retained, and a value of 1 means $e-1 = 0.4$ of binding is retained. The score is the sum of the impact of each mutation independently and irrespective of the amino acids that are subsituted. All RBD mutations are considered (spike residues 330–531).

The predicted impact of RBD mutations of study sequences on ACE2 binding relative to BA.2 was determined using Bloom lab's ACE2 binding calculator.[18] A logarithmic scale on the y-axis is used where a score of 1 means 10-fold higher affinity compared to background BA.2. The score is the sum of the impact of each mutation independently assuming that mutations do not interact.

Visualization

Mutation data tables and multi-layered donut charts were created in Microsoft Excel 2016, bar, line, and scatter plots in GraphPad Prism v.9.4.1, and mirror bar graphs, stacked bar graphs, bubble plots, and lollipop plots in program R with the ggplot2 package. Logistic regression analyses were done with the stats package, correlation heatmaps with the corrplot package, and geomapping with the tigris and ggplot2 packages in program R. Circular edge bundling plots were generated in undirected mode in R using ggraph, igraph, tidyverse, and RColorBrewer packages. Edges are only shown if $P < 0.05$, and nodes are sized according to the connecting edges' r values. Nodes are color-coded according to groups of variables. Stacked bar graphs and bubble plots were created using the ggplot2 package in R; box plots were made in Prism.

External data sources

New York City-wide, daily COVID-19 case numbers and hospitalisations were obtained from NYC Open Data with data provided by the Department of Health and Mental Hygiene (DOHMH)[39] and NYC Health with data provided by the Citywide Immunization Registry (CIR).[40]

Statistics

The statistical comparison of genomic and clinical variables in vaccinated versus unvaccinated individuals were primary goals of the study. In a post hoc power analysis, the comparison of a total of 1582 individuals vaccinated with the primary series or 1875 boosted individuals with 1876 unvaccinated individuals achieved 99.3% and 99.6% power, respectively, in detecting a 5% difference (10% versus 15%) for a specific variable in a two-tailed Fisher Exact test with a type I error of 5% (G*Power v.3.1.9.4).[41] To compare the probability of a variant/mutation for the vaccinated/boosted and unvaccinated groups over time, we performed logistic regression analyses on the full data adjusting for sex, age (centred and standardized), and month of test. We did not adjust for other confounding factors, e.g., co-morbidities, because of potential bias in reporting, documentation, or incompleteness. We compared clinical and demographic variables between groups using non-parametric Mann–Whitney, Kruskal–Wallis with Dunn's multiplicity correction, or chi-square tests in Prism. Correlation analyses were done using two-sided Spearman rank tests in Prism using continuous and binary variables. For all analyses unless otherwise stated, a two-sided Type I error rate of 0.05 was applied. Missing data were not imputed.

Ethics/study approval

This study was approved by the NYULH Institutional Review Board, protocol numbers i21-00493, i21-00561, 18-02035, and 18-02037. Ethics approval and informed patient consent were not required on unidentifiable viral sequences.

Role of funders

The funders played no role in the design, conduct, interpretation, or reporting of this study.

Results

A New York metropolitan cohort of vaccinated, unvaccinated, and re-infected SARS-CoV-2 positive cases during increasing booster vaccinations

Between July 2021 and August 2022, New York city faced four SARS-CoV-2 waves (Fig. 1a and b, Figure S1): the remainder of the Delta wave, a massive Omicron BA.1 wave, and the merging BA.2/BA.3 and BA.4/BA.5 waves. During this period, the percentage of vaccinated individuals (one shot or more) rose from ~60% to almost 90%. After the introduction of booster vaccinations in September 2021, the percentage of boosted individuals steeply increased until March 2022 and successively tapered towards 40% coverage (Fig. 1b). The high COVID-19 case numbers during the Omicron waves, particularly BA.1, were associated with a low rate of hospitalisations and deaths compared with Delta (Fig. 1b). During this time frame, i.e., between July 16th, 2021 and August 9th, 2022, we obtained a total of n = 5432 high-quality near-full genome SARS-CoV-2

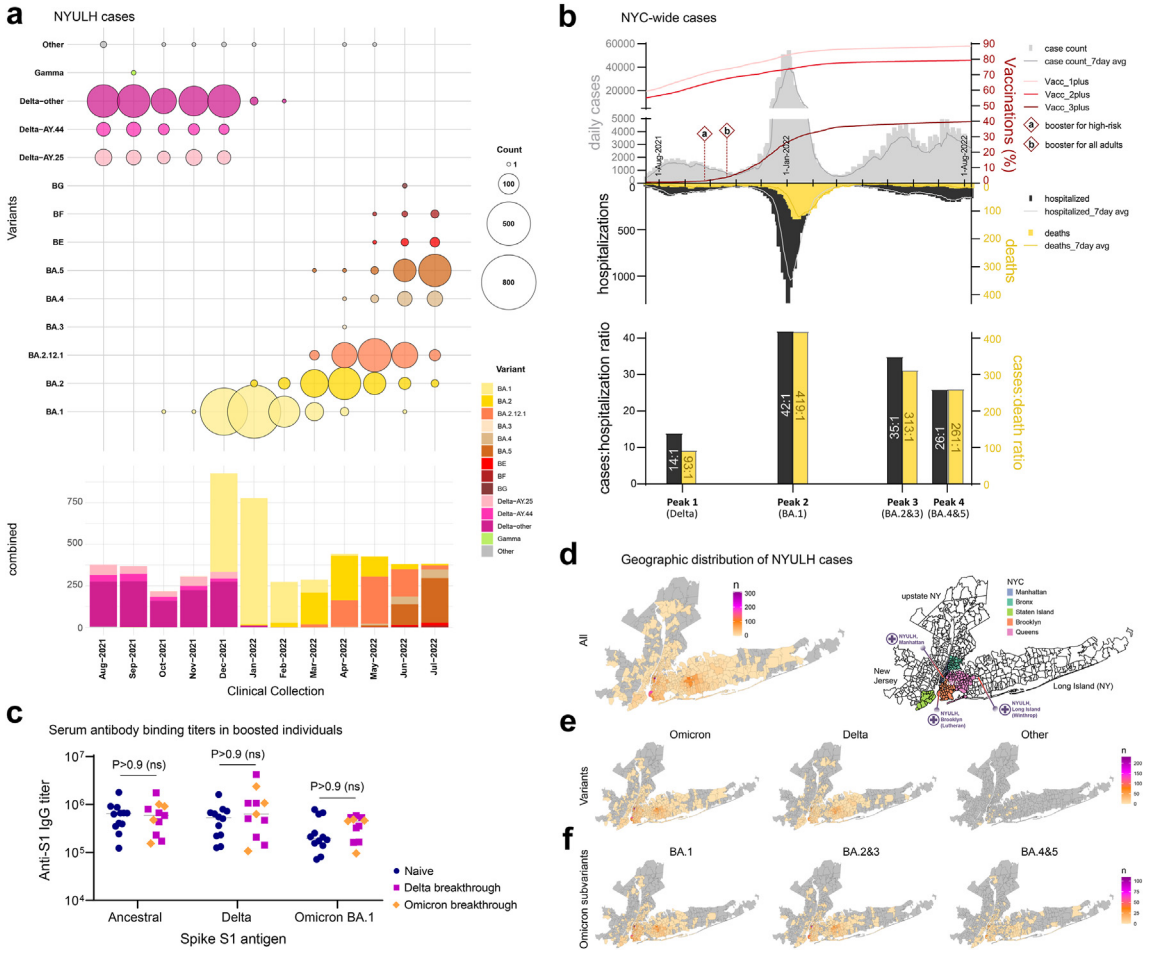

Fig. 1: **COVID-19 cases and clinical outcome in relation to vaccination rates, antibody titers, and geographic distribution in New York City, July 2021–August 2022. a.** Variant distribution of SARS-CoV-2 infections at NYULH, shown separated into the main variants observed (top) and aggregated (bottom). All fully-covered months are shown (July 2021 and August 2022 were omitted). **b.** Data facing up: Daily NYC-wide COVID-19 cases (grey bars; ~1.5 Mio in total), 7-day averages (grey line), and cumulative vaccination data (red) by number of shots (Vacc_1plus: ≥1shot, Vacc_2plus: ≥2 shots, Vacc_3plus: ≥3 shots; a: implementation of booster shots for high-risk patients, September 24, 2021; b: free access to booster shots for all adults, November 19, 2021). Data facing down: daily hospitalisations due to COVID-19 (black bars) and 7-day averages (grey line), and daily deaths due to COVID-19 (yellow bars) and 7-day averages (ochre line). The plot below depicts case:hospitalisation (black bars) and case:death ratios (yellow bars), shown for the four main peaks. Source of data: NYC Open Data and NYC Health, Citywide Immunization Registry (CIR); data from between July 16th, 2021 and August 9th, 2022. **c.** Anti-SARS-CoV-2 spike S1 antibody titers in boosted individuals, measured in COVID-19-naïve (never infected), Delta- and Omicron (BA.1)-infected breakthrough individuals by ELISA using three different S1 antigens. Statistical comparisons were done using Kruskal–Wallis tests (ns: not significant). **d–f.** Geographic mapping of SARS-CoV-2 infections in the greater NYC region, shown for all variants combined (**d**), Omicron compared to Delta and other variants (**e**), and major Omicron subvariants (**f**).

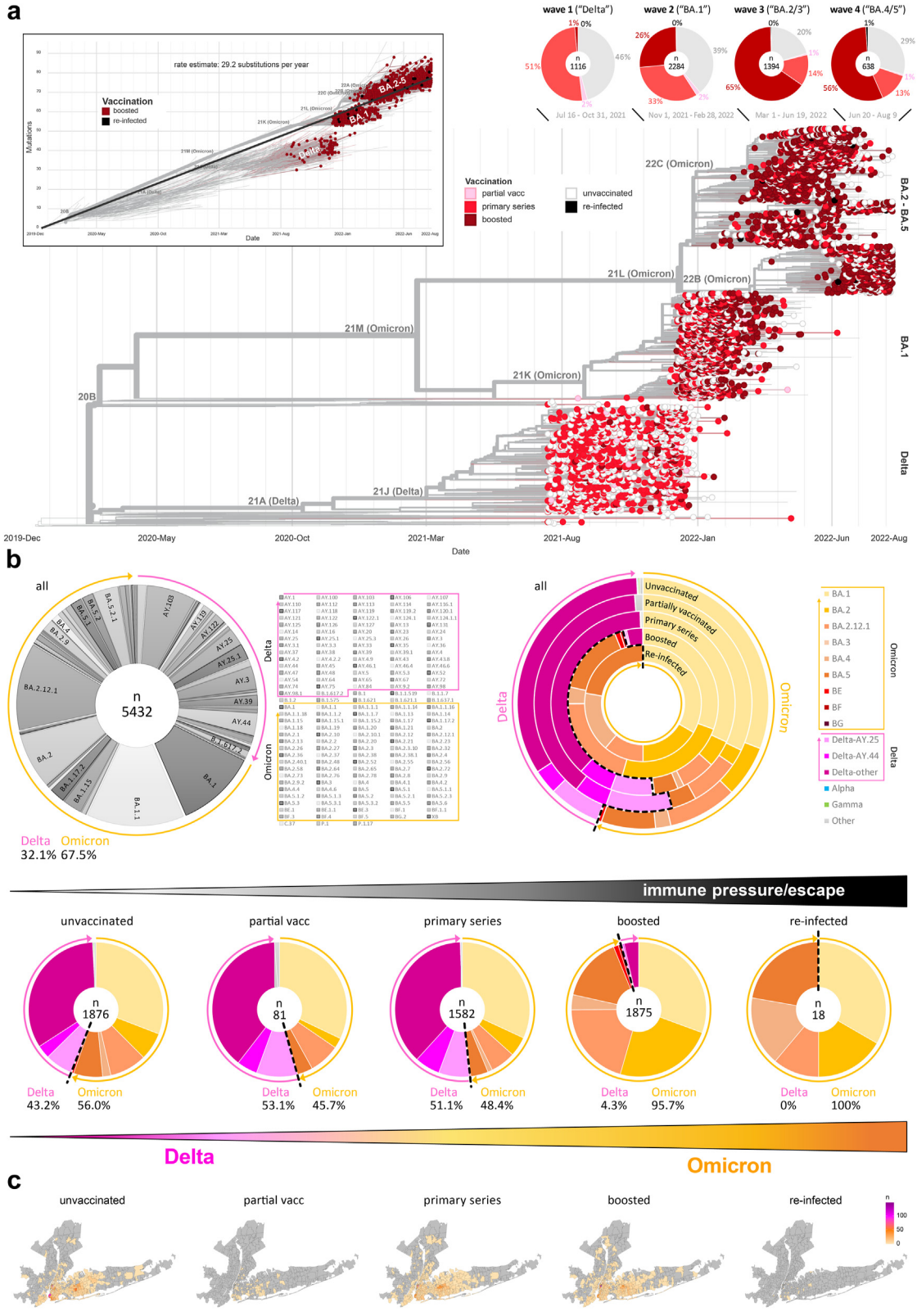

***Fig. 2:* Variant distribution by vaccination and re-infection status. a.** Time-calibrated maximum-likelihood tree of 7146 North-America-focused SARS-CoV-2 sequences (as in Figure S1), filtered for the 5432 NYULH sequences and annotated by vaccination/re-infection status.

sequences from the NYULH multicentre healthcare system in metropolitan New York. The majority of SARS-CoV-2 sequences were obtained from unvaccinated individuals (35%), individuals vaccinated with the primary series and not boosted (29%), and boosted individuals (35%), whereas partially vaccinated (1%) and re-infected individuals (0.3%) were a minority of cases (Table S1). Notably, booster vaccinations yielded saturating anti-spike IgG titers that did not further increase after breakthrough infections, neither by Delta nor Omicron (Fig. 1c, Table S2). It facilitated the comparison of Delta and Omicron mutation trajectories under comparable, saturating antibody binding immune pressure, i.e., by booster vaccinations.

The age distribution was comparable across variants (Figure S2); removing paediatric patients (age <16), who might not have been eligible for vaccinations or had a lower vaccine uptake, we found that boosted individuals were younger than unvaccinated ($P < 0.0001$) or primary series-vaccinated individuals ($P = 0.0032$). The sex distribution was slightly shifted towards females in Omicron versus Delta infections (60% versus 56% females, respectively) (Figure S3, Table S3). Hospitalisation and death rates significantly decreased in a stepwise manner from pre-Delta to Delta and Omicron infections ($P < 0.006$) and also with stronger immunization ($P < 0.0001$), i.e., from unvaccinated to partially vaccinated, primary series, boosted, and re-infected individuals (Figure S4, Table S3). In our cohort, we found marginally higher hospitalisation and death rates in individuals vaccinated and boosted with Moderna than Pfizer-BioNTech ($P < 0.008$) (Figure S5). A detailed set of demographic and clinical specifications can be found in Tables S4–S8.

Variant evolution and their geographic distribution in the greater NYC area

Until November 2021, Delta was the dominant variant in metropolitan New York. Omicron BA.1 started to increase rapidly in December 2021 with a high co-circulation of both variants in that month (Fig. 1a and b, Figure S1, Table S9). BA.2 and BA.3 succeeded BA.1, albeit BA.3 occurred at very low numbers. BA.2 became dominant in March 2022, including a prominent subpeak of BA.2.12.1, a variant first detected in New York.[42] Starting in June 2022, BA.5 variants became predominant, accompanied by a smaller number of BA.4 infections (Fig. 1a and b, Figure S1).

SARS-CoV-2 samples were collected from all over NYC's five boroughs, wide areas of Long Island (East), and adjacent counties of upstate New York (North) and New Jersey (West) (Fig. 1d). Sampling foci were around the three main NYU healthcare locations in Manhattan, Brooklyn, and Long Island. We recorded a comparable distribution of Delta and Omicron infections (Fig. 1e) and likewise for the three Omicron subpeaks composed of BA.1, BA.2 & BA.3, and BA4 & BA.5 (Fig. 1f). These findings indicate no significant differences in the catchment area or founder effects of the major variants but highlight a sampling gradient towards the collection centers.

Variant distribution by vaccination/infection status

Breakthrough infections in individuals vaccinated with the primary series became frequent since the introduction of Delta (Fig. 2a).[21] Booster shots were approved in the Delta wave, and booster breakthrough infections became abundant in the Omicron BA.1 wave (Fig. 2a), coincident with the strong surge in booster shot administrations in the study population (Fig. 1b). We performed stratified analyses by vaccination/infection status (Fig. 2b and c) and logistic regression analyses correcting for potential confounders, including study month, sex, and age of participants (Fig. 3a and b). Overall, the 5432 study sequences were composed of 32.1% Delta, 67.5% Omicron, and a minor 0.4% of other SARS-CoV-2 variants. The Delta and Omicron waves were highly diverse; Delta subvariants AY.25 and AY.44, which increased significantly during the first half of the Delta wave in New York and particularly in vaccinated individuals,[21] still made up approximately a quarter of Delta infections in the second half of the Delta wave, but they declined in the later months of this wave. Omicron infections included a broad variety of subvariants, specifically within the abundant BA.1, BA.2, and BA.5 clades (Fig. 2b). Whereas Delta infections constituted between 43 and 53% of infections in unvaccinated, partially vaccinated, and primary series vaccinated individuals, Delta infections were rare in boosted individuals (4%) and were absent among re-infected individuals (Fig. 2b). Geographically, sampling was comparably spread over metropolitan New York for unvaccinated, primary series-vaccinated, and boosted individuals (focussing on well-powered groups of >1500 each) (Fig. 2c).

Selective advantage of BA.1, BA.2, and BA.5 lineages in boosted individuals

To determine whether a SARS-CoV-2 variant had a selective advantage over co-circulating lineages in vaccinated compared to unvaccinated individuals, we performed

Pie charts indicate distribution by vaccination/re-infection status for each wave (according to Fig. 1b). Box: molecular clock highlighting boosted and re-infected samples. **b.** Pango lineage distribution of all 5432 NYULH sequences and frequency of Delta and Omicron lineages (upper left). Distribution of main variants by vaccination/re-infection status are shown in overlay in a multi-layer donut plot (upper right) or separately, sorted from left to right by increasing immune pressure/escape (down below). Frequencies of Delta and Omicron infections are indicated. **c.** Geographic mapping of SARS-CoV-2 infection events in the greater NYC region by vaccination/re-infection status.

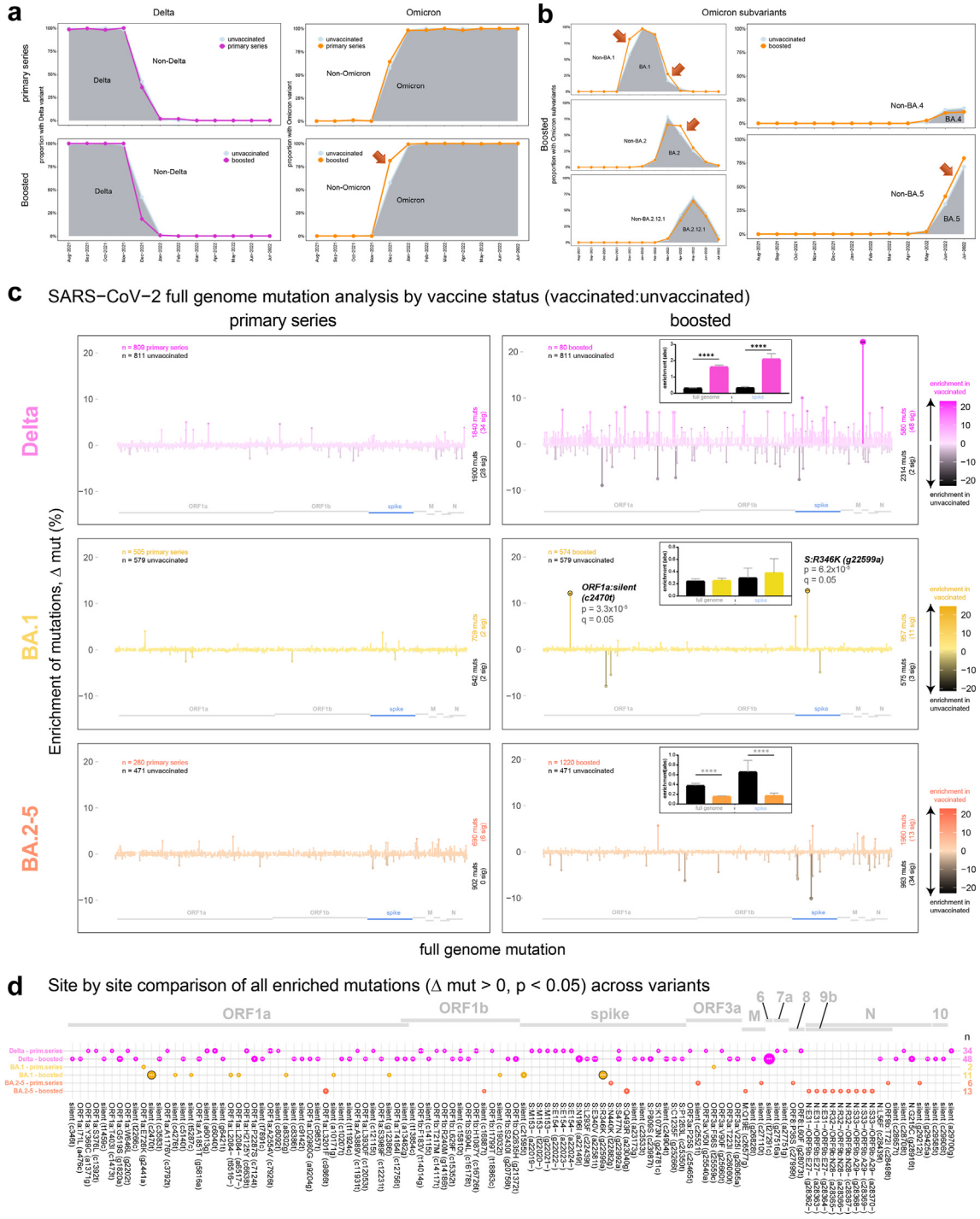

Fig. 3: **Relative growth advantage of Omicron subvariants in boosted individuals and selection of Omicron full genome mutations under vaccine pressure. a.** Probability of detection of Delta (left) or Omicron (right) by month in fully vaccinated (top) and boosted (bottom) individuals compared to unvaccinated individuals, adjusted for month of test, sex, and age of participants. **b.** Relative growth analyses as in (**a**) but for Omicron subvariants and focusing on boosted individuals versus unvaccinated. Arrows indicate relative growth advantages of the studied variant for the given month in the boosted study population against all other circulating variants compared with unvaccinated individuals. **c.** Site-specific spike mutation analysis in SARS-CoV-2 vaccine breakthrough sequences compared to unvaccinated controls. The analysis was done for both primary series-vaccinated and boosted individuals, and separated into Delta, BA.1 and BA.2-BA.5 (BA.2-5) infections. The Wuhan-Hu-1 sequence served as reference for mutation calling. The mirror plot displays differences of mutation frequencies per full

relative growth and logistic regression analyses adjusted for study month, sex, and age of participants (Fig. 3a and b, Table S10). Both Omicron and Delta were effective in infecting individuals vaccinated with the primary series, and Omicron was only marginally superior to Delta in this study group. However, Omicron's advantage over Delta became strongly significant in the boosted population, which became most evident in December 2021 when both variants co-circulated in high numbers in metropolitan New York (Fig. 3a, Table S10a). Within the boosted population, we further compared the main Omicron lineages sampled >200x. BA.1, BA.2, and BA.5 lineages had distinct advantages over co-occurring lineages under booster vaccine pressure, but not BA.2.12.1, the variant first detected in New York, or the BA.4 lineage (Fig. 3b, Table S10b). Of note, BA.2 and BA.5 remained the main source for the development of new sublineages and recombinants to predominate during the first months of 2023.[2] A comparison of BA.2-BA.5 sequences from unvaccinated, primary series, and boosted individuals revealed no significant differences in predicted immune escape or ACE2 binding affinity between groups (Figure S6).

High mutational booster vaccine pressure on Delta that gradually settled with the introduction of BA.1—BA.5

To screen for mutational marks of vaccine pressure, we assessed all base pair mutations over the SARS-CoV-2 full genome, and compared the mutation levels and their enrichment in vaccinated versus unvaccinated individuals (Fig. 3c and d, Figure S7a and b, Table S11). We performed subanalyses to determine the differential effect of vaccination with the primary series (left) and booster (right), and further stratified into Delta (top), BA.1 (middle), and BA.2-BA.5 infections (bottom) (Fig. 3c), according to their staggered occurrence in time and genomic backgrounds (Figs. 1a and b and 2a). In viruses from individuals vaccinated with the primary series, we recorded flat differential mutation profiles when compared against viruses from unvaccinated controls (Fig. 3c, left). In contrast, mutational vaccine pressure became evident in viruses from boosted individuals (Fig. 3c, right), who exhibit saturating but comparable antibody binding immune pressure against Delta and Omicron (Fig. 1c). Overall, Delta exhibited the strongest signs of mutational vaccine pressure, compared with the flatter mutational profiles of BA.1 and BA.2-BA.5. For example, Delta had the highest number of differential mutation sites (2894x), the highest number of significantly enriched mutations (50x; Fisher Exact test), and the highest enrichment scores in average (e.g., ~2% Δmut in spike, boosted versus unvaccinated). Notably, in Delta, the mutation scores were significantly higher in viruses from boosted compared to unvaccinated individuals (boxed graph; colored versus black bars), a pattern that gradually switched along the evolutionary timeline, i.e., the mutation scores were comparable between boosted and unvaccinated in BA.1, but were higher in unvaccinated compared to boosted individuals in BA.2-BA.5. The latter pattern might be partly explained by the longer time intervals after completion of vaccination in BA.1 and BA.2-5 compared to Delta (Figure S7c, upper row). Particularly BA.2-5 infections were highly associated with long intervals between vaccination/boosting and infection (Figure S7d), which raises the possibility that waned adaptive immune responses played a role. Time intervals between the different vaccinations were less associated with a particular variant; absolute $|r|$ values (Spearman rank) did not exceed 0.3 (Figure S7d). Screening the subsets site-by-site for significantly enriched mutations in boosted individuals, corrected for multiple comparisons, revealed two hits in BA.1, i.e., spike RBD immune escape mutation R346K and silent ORF1a c2470t mutation, characteristic for subvariant BA.1.1 (Fig. 3c and d, Table S11).

Diffuse adaptive evolution in Delta spike contrasts strong, receptor-binding motif-focused adaptive evolution in BA.2-BA.5 spike

To profile adaptive evolution in Delta, BA.1, and BA.2-BA.5, we compared synonymous (dS) and non-synonymous (dN) substitution rates in spike using a Bayesian analysis for adaptive selection (Fig. 4, Figure S8).[37] We differentiated positive, diversifying selection (shown in colors) and negative, purifying selection (black), and stratified the participants by vaccination status. The strength of positive and negative selection (selection scores) increased from Delta to BA.2-BA.5 in the unvaccinated and primary series vaccinated groups, but not in the boosted group (Fig. 4a, Figure S8). In the boosted group, adaptive evolution was enhanced in Delta already and remained at similar levels in BA.1 and

genome residue between vaccinated and unvaccinated groups, shown along the x-axis. Bars facing up and down refer to elevated mutation frequencies in vaccinated and unvaccinated individuals, respectively. Numbers of observed mutations (muts) and significant mutations (sig) are indicated to the right. Boxes: Statistical comparison of positive (Δmut >0) and negative (Δmut <0) enrichment scores of mutations (absolute values) over the full genome and for spike. Means and 95% confidence intervals are shown. Positive and negative scores are statistically compared using a Kruskal–Wallis test with Dunn's multiplicity correction (****P < 0.0001; grey asterisks indicate greater negative than positive enrichment/higher mutation rate in unvaccinated). **d.** Summary of all enriched mutations for the different variants tested. Significant results are displayed with asterisks (Fisher Exact test, *P < 0.05; **P < 0.01; ***P < 0.005). Significant results after Bonferroni-Hochberg multiplicity correction (q < 0.05) are encircled in black. 6: ORF6; 7a: ORF7a; 8: ORF8; 9 b: ORF9b; 10: ORF10; bp: base pairs; E: envelope; M: matrix; mut: mutation; nsp: non-structural protein; N: nucleocapsid; ORF: open reading frame.

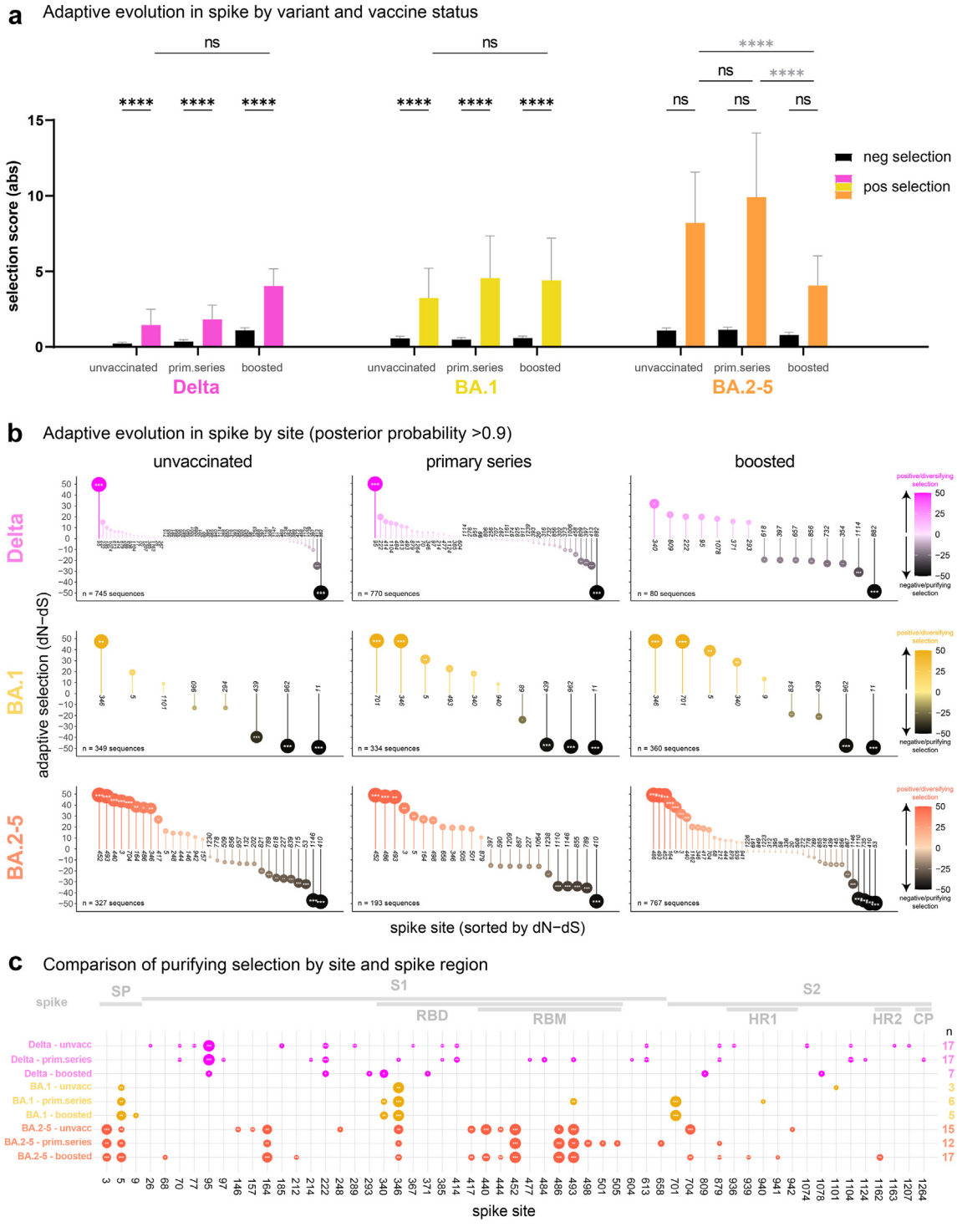

Fig. 4: **Adaptive evolution in spike by variant and vaccine status. a.** Adaptive evolution analysis of all spike residues using a fast, unconstrained Bayesian approximation for inferring selection (FUBAR, Datamonkey) based on the differences of non-synonymous (dN) and synonymous (dS) mutation rates per site. The analyses were done for unvaccinated, vaccinated, and boosted individuals infected with Delta, BA.1, or BA.2-BA.5 (BA.2-5). The barplot displays the means and 95% confidence intervals (absolute values) of all positive, diversifying (dN-dS >0) and all negative, purifying (dN-dS <0) selection scores in each vaccination group. The statistical comparison includes the comparison of all positive and negative scores per vaccination group and all positive scores among the different vaccination groups per variant (Kruskal–Wallis test

BA.2-BA.5, suggesting comparable evolutionary pressures by booster shots or comparable bottlenecks of breakthrough in terms of amino acid acquisition irrespective of variant. Furthermore, positive selection was significantly higher than negative selection in Delta and in BA.1, but not in BA.2-BA.5 (Fig. 4a). The latter is based on relatively high negative selection scores in BA.2-BA.5, indicative of reversions of mutations to wildtype in the process of fine-tuning immune evasion and viral fitness in variants such as BA.2-BA.5 that have undergone a multitude of evolutionary changes.

We further looked into individual sites with significantly increased adaptive selection scores (posterior probability >0.9) and studied their location across spike (Fig. 4b and c, Figure S8). Delta and BA.2-BA.5 had the highest number of positive (n = 17) and negative selection sites (n = 26/24, for Delta and BA.2-BA.5, respectively), whereas BA.1 had no more than 6 positive and 5 negative selection sites per vaccination group. Notably, in RBD, positive selection was observed at multiple sites but at moderate levels in Delta, whereas BA.1 had a distinct, strong positive selection at immune evasion site 346 (mutation found in BA.1.1). In BA.2-BA.5, multipe sites underwent strong positive selection with a particular focus on the receptor-binding motif (RBM), including sites 417, 440, 444, 452, 486, and 493 that represent the most immune evasive mutation sites.[38] Of note, positive selection at 486, a key immune escape site in Omicron,[38] was significantly higher in viruses from primary series and boosted compared to unvaccinated individuals (Fig. 4c).

Convergent evolution and correlations with vaccination status and clinical factors

SARS-CoV-2 evolution occurred independently in Delta, BA.1, and BA.2-BA.5 variants, however, distinct mutations were recurrently selected despite differing genomic backgrounds/epistasis, either as imprinted clade-defining mutations (T19X, G142D, S371X, L452X, T478K, P681X), or as recurrent minority mutations (E340X, R346X, P1162) (Fig. 5a and b, Figure S9). Many of these recurrent mutations engage key spike positions critical for immune escape, receptor binding, or spike cleavage, and are found among the hits of our selective adaptation screen (Figs. 3 and 4). Some other mutations we identified adaptively selected under immune pressure did not show a recurrent pattern at the given sampling depth (Figure S10).

To determine whether adaptively selected/enriched mutations under vaccine pressure are associated with vaccination status, viral, demographic, or clinical variables, we performed a multivariable correlation analysis (Fig. 5c–e, Figures S11 and S12). The identified mutations were similarly distributed across clinical outcomes, with a few exceptions, such as the infrequent E340X mutations, preferentially found in BA.1 lineages, being associated with higher hospitalisation and death rates. The recurrent mutations did not follow a consistent pattern of being associated with vaccination status, suggesting that vaccine pressure is not the only determinant for convergent evolution. The recurrent mutations became more prevalent/evident in the later stages of the pandemic, and were found co-associated with different BA.2-BA.5 subvariants (Fig. 5c–e, Figures S11 and S12).

Discussion

Booster shots are crucial to boost immunity against evolving viruses. The transition from Delta to Omicron and its onward evolution into subvariants marks a cornerstone in SARS-CoV-2 evolution, curbed and shaped by infection and booster-mediated immunity. We studied more than 5400 sequences during this time, collected from a multicentre cohort in an urban metropolis with high viral throughput, diversity, and high vaccination rates, with the goal to determine SARS-CoV-2 genomic fingerprints of increasing booster vaccine pressure. Omicron, equipped with an elite set of immune escape mutations, adapted strikingly different under vaccine pressure than Delta with ongoing plasticity through the evolution of BA.1 to BA.5 variants.

The transition from Delta to Omicron coincided with the introduction and increased uptake of booster vaccinations (Fig. 1b and 2a) that induced saturating anti-spike antibody titers that did not further increase after breakthrough infections, neither by Delta nor Omicron (Fig. 1c). Booster shots significantly reduced severe COVID-19 and symptomatic SARS-CoV-2 infection in Delta and previous variants but less in Omicron, where strong immune escape was accompanied by substantial waning of the booster immune response.[43] Strikingly, all re-infections detected in our cohort were Omicron (Fig. 2a and b), which is in line with a meta-analysis of 65 studies from 19 countries showing strong protection

with Dunn's multiplicity correction; ****P < 0.0001; black and grey asterisks indicate increasing and decreasing values from left to right, respectively). **b.** All sites with significant positive (facing up and colored) and negative (facing down and dark gradient) selection are shown along the x-axis. The y-axis displays the difference of non-synonymous and synonymous mutation rates per site. Posterior probabilities (PP) > 0.9 are considered significant (* PP > 0.9; **P > 0.95; ***P > 0.99) and are indicated by an asterisk inside the circles. **c.** Comparative summary of all sites where significant positive, diversifying selection was detected in at least one data set. CP: cytoplasmic domain; dN: mean posterior non-synonymous mutation rate at a site; dS: mean posterior synonymous mutation rate at a site; HR: heptad repeat; RBD: receptor-binding domain; RBM: receptor-binding motif; SP: signal peptide.

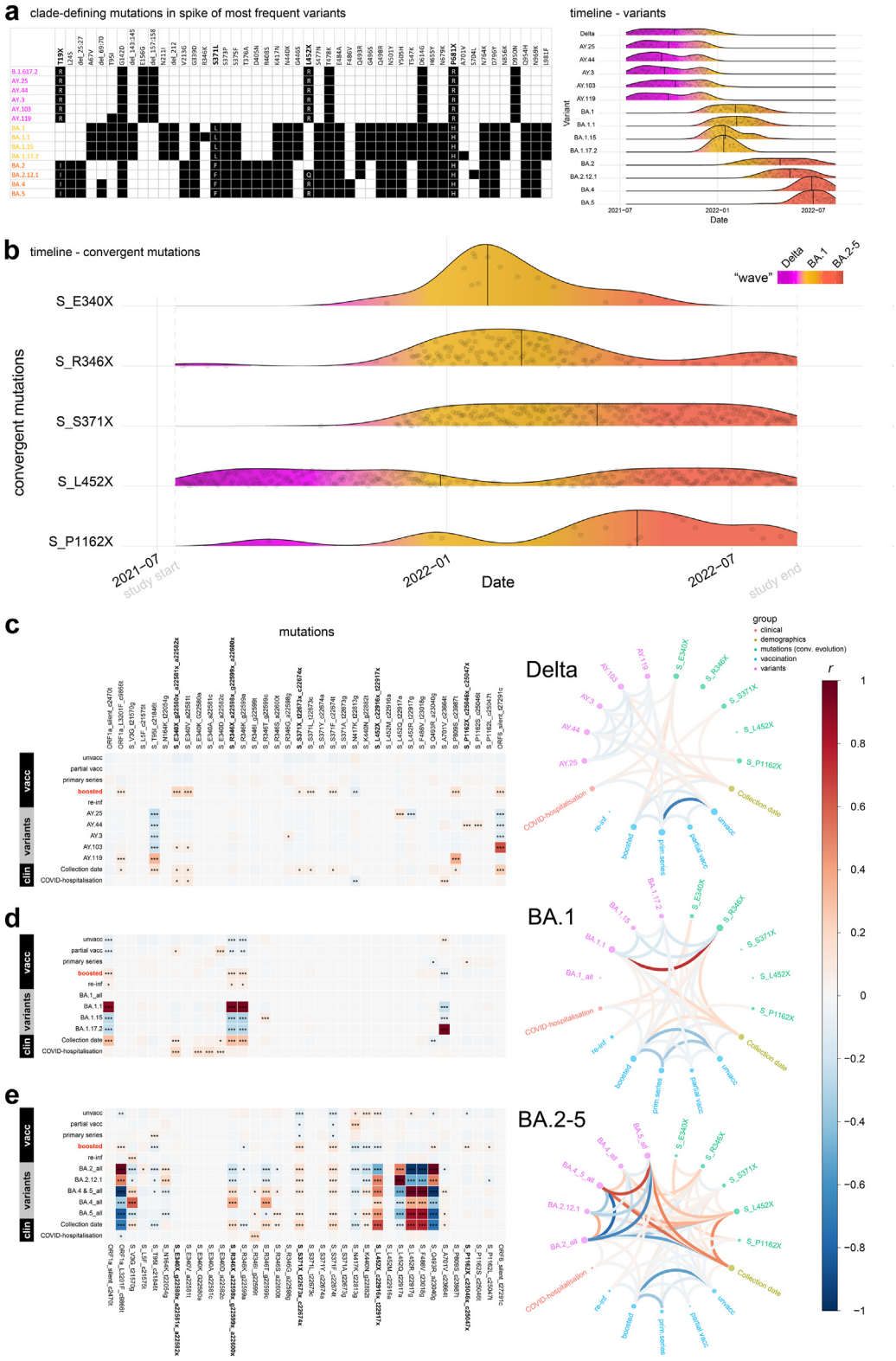

Fig. 5: **Convergent evolution and correlations with vaccination status and clinical factors. a.** Spike mutations in major circulating variants compared to Wuhan-Hu-1. Sites of convergent evolution are highlighted in bold and the different amino acids are listed in the table. A timeline

from re-infection with previous variants, including the ancestral strain, Alpha, or Delta, but not with Omicron.[44]

Omicron's diversifying evolution into a multitude of BA.2-BA.5 variants has been attributed to increasing immune pressure in the population, as immunological and global sequence analyses suggest,[19,45] supported by our differential genomic analyses by vaccine status in a model cohort (Figs. 3–5). In particular the increasing booster immune pressure may have shaped further Omicron adaptations including convergent evolution (Figs. 3c, 4c and 5b). Compared to previous variants, adaptive evolution in Omicron occurred against a different antigenic background and associated with different biological and clinical features, i.e., preferential replication in the upper airways,[13,14] reduced clinical severity (Fig. 1b), and growth advantages over Delta, which essentially drove the lineage replacement.[15,16,46] From an immunological point of view, it was proposed that immune imprinting by vaccinations/infections with previous variants curbed the diversification of the immune repertoire that could not cope with the rapid evolution of Omicron variants. Instead, Omicron immunity remained incomplete, and the increasing but imprinted immunity supposedly drove the selection of comparable mutations across variants, resulting in convergent evolution.[19,45] Since the neutralizing antibody response primarily targets spike RBD, convergent evolution has focussed on this region, as our BA.2-BA.5 adaptive evolution analysis confirmed (Fig. 4c).

Dependent of the underlying virus and epistatic landscape, genomic marks of immune escape and selective adaptations differ along the evolutionary track. Before the advent of Omicron, convergent evolution focused on spike residues K417, L452, E484, N501, and P681, as found in Alpha to Delta variants.[19,45] Immune escape mutations such as K417N that decreased spike's ACE2 affinity were compensated by affinity-increasing mutation N501Y.[17] Furthermore, in Delta, our previous study identified selective adaptation processes in individuals vaccinated with the primary series that eventually led to the accumulation of spike NTD (S112L) and nsp12 (F192V) mutations.[21] The current study showed that the evolutionary pressure in primary series-vaccinated individuals waned in the second half of the Delta wave (Fig. 3c). Booster vaccine pressure again lifted the bar of immune pressure driving the selection of further mutations in Delta. We observed a high number of enriched mutations under booster vaccine pressure (Fig. 3c and d) that however translated into low rates of adaptive evolution (non-synonymous/synonymous mutation rates) (Fig. 4). In BA.1, the mutational divergence along the full SARS-CoV-2 genome levelled off, however a booster-associated selection of immune escape mutation R346K in spike RBD was recorded (Fig. 3c and d), accompanied by additional sites of adaptive evolution (Fig. 4). The high rates of asymptomatic infections that were particularly abundant with BA.1 and presumably underrepresented in this study, might have biased these selection processes.[47,48] BA.2-BA.5 had the highest rates of adaptive evolution in spike, which contrasted with the low mutational divergence over the full genome under booster vaccine pressure and the stronger enrichment of mutations in viruses from unvaccinated than boosted individuals (Figs. 3c, d and 4).

Omicron spike amino acid residues that we found selected under boosted vaccine pressure (346, 417, 440, 444, 452, 486, and 493) are crucial sites of immune escape.[19,38] Notably, enriched/fixed mutations under vaccine pressure were not restricted to immune escape mutations (as preferentially found in spike RBD) but also mutations affecting viral fitness and biological functions (e.g., in spike signal peptide, NTD, or near the S1/S2 cleavage site), which demonstrates that immune escape is not a uni-directional process but intertwined with adaptations to maintain/improve viral transmission, cellular entry, and replication (Figs. 3d and 4c). Furthermore, evolution requires a fine balance of diversifying mutations (positive selection) and compensatory/purifying mutations (negative selection), as mutational enrichment and adaptive evolution analyses suggested, particularly for BA.2-BA.5 (Figs. 3c and 4a).

Notably, selective advantage under booster vaccine pressure differed among Omicron subvariants. While BA.2 and BA.5 lineages had advantages in boosted individuals, BA.2.12.1 and BA.4 were inferior in relative epidemiologic growth in boosted individuals (Fig. 3b). It remains elusive why BA.4 lacked behind BA.5, given that both variants have the same clade-defining spike

of detection of these variants is shown to the right. The density plots are colored by wave, the median is shown as vertical black line, and each detected sequence is shown as a grey dot within the shapes. **b.** Density plots indicating the detection of convergent mutations, formatted as in (**a**). **c–e.** Multivariable correlation analyses of mutations with adaptive evolution under vaccine pressure (as identified in Figs. 3 and 4) in Delta (**c**), BA.1 (**d**), and BA.2-5 (**e**) infections against vaccination status, variants, and clinical variables. The heatmaps (left) are color-coded according to linear regression coefficients (*r*) between the respective pairwise variables. Asterisks indicate statistically significant correlations (Spearman rank, *P < 0.05; **P < 0.01; ***P < 0.005). The circular network plots (right) highlight the network of correlations of the convergent mutations. Red and blue edges represent positive and negative correlations between connected variables, respectively, according to the scale of *r* values to the right. Only significant correlations (P < 0.05, Spearman rank test) are displayed. Nodes are color-coded based on the grouping of variables. Node size corresponds to the strength of correlations. Clin: clinical variables, re-inf: re-infected, S: spike, unvacc: unvaccinated, vacc: vaccination, X: any other residue than wildtype.

mutations. Reasons might be the more rapid/extensive diversification into subvariants in BA.5 compared to BA.4 (Fig. 2b) with the accumulation of clade-defining as well as non-clade-defining mutations in spike and beyond, which may have affected immune escape, viral fitness, transmission, and eventually positive selection. Of note, despite large differences in selective advantages across variants and mutational profiles under vaccine pressure, convergent evolution has become a constant factor particularly since BA.2 (Fig. 5).[19,45] It highlights common evolutionary tracks, including in spike, despite epistatic differences between variants.[17,49] High rates of convergent evolution in Omicron spike were particularly striking for R346X, an immune escape mutation identified in >100 lineages, S371X, a stabilizing mutation to remedy the thermodynamically less stable Omicron RBD, and L452X, Delta's main immune escape mutation that recurred in numerous Omicron lineages.[18,19,45]

Booster shots are a pivotal tool to lift SARS-CoV-2 immunity and are particularly recommended in vulnerable populations and healthcare workers. They provide immune pressure to curb infections and their severity, while at the same time requiring genomic monitoring of selective viral adaptation processes.

Limitations of our observational study include potentially confounding demographic, behavioural, socio-economic, and temporal factors (e.g., time post boost and waning immunity), differing testing practices, unequal sampling, and underrepresentation of asymptomatic/re-infections with low viral loads that could not be fully compensated by our unbiased testing and sampling procedures, stratified, and adjusted analyses.

Contributors

Ralf Duerr, MD, PhD designed the study, analysed data, created figures, and wrote the first version of the manuscript; Dacia Dimartino, PhD and Paul Zappile made libraries and sequenced the samples; Christian Marier analysed genomic data; Guiqing Wang, MD, PhD collected samples and performed clinical SARS-CoV-2 detection; Fritz Francois, MD and Eduardo Iturrate, MD generated total numbers of vaccinations for the entire healthcare system; Eduardo Iturrate, MD monitored epidemiological data and generated lists of breakthrough infections; Mila B. Ortigoza, MD supervised the study and reviewed clinical information; Marie I. Samanovic, PhD and Mark J. Mulligan, MD generated immunological data and reviewed clinical information; and Adriana Heguy, PhD designed the study, generated genomic data and wrote the manuscript. All authors directly accessed and verified the underlying data, reviewed, and edited the manuscript, and read and approved the final version of the manuscript.

Data sharing statement

All deidentified viral sequences and metadata are publicly available in GISAID (Tables S4–S9).

Declaration of interests

Mark J. Mulligan reports potential competing interests: clinical trials and laboratory research contract funding for vaccines and monoclonal antibodies for SARS-CoV-2 with Lilly, Pfizer, and Sanofi; personal fees for Scientific Advisory Board service from Merck, Meissa Vaccines, Inc., and Pfizer. All other authors declare no conflicts of interest.

Acknowledgements

We thank NYU Langone Health DataCore for support extracting the data for this study from clinical databases and the clinical laboratory technicians for assistance in testing, saving and retrieving specimens, especially Joanna Fung and Dawn Ip. We also thank Dr. Joan Cangiarella and Dr. Dafna Bar-Sagi for their continuous support of genomic surveillance for SARS-CoV-2 at NYULH, including providing institutional funding for this study. We are grateful to the submitting laboratories who deposited data in GISAID, in particular to those whose sequences we used to create the phylogenetic tree (Table S9). The NYULH Genome Technology Center is partially supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center and the National Institutes of Health Clinical and Translational Science award 3UL1TR001445-06A1S1 (M.B.O.). This research was also funded in part by the National Institutes of Health, grant number AI148574, and NIAID contract funding via 75N93021C00014 to M.J.M.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2023.104843.

References

1 World Health Organization (WHO). *WHO coronavirus (COVID-19) dashboard*; 2023. https://covid19.who.int/. Accessed February 3, 2023.
2 GISAID. *GISAID database*; 2023. https://www.gisaid.org/. Accessed March 7, 2023.
3 Cele S, Jackson L, Khoury DS, et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature*. 2022;602(7898):654–656.
4 Viana R, Moyo S, Amoako DG, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*. 2022;603(7902):679–686.
5 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.
6 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.
7 Sadoff J, Gray G, Vandebosch A, et al. Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *N Engl J Med*. 2021;384:2187.
8 Dunkle LM, Kotloff KL, Gay CL, et al. Efficacy and safety of NVX-CoV2373 in adults in the United States and Mexico. *N Engl J Med*. 2021;386(6):531–543.
9 Haas EJ, Angulo FJ, McLaughlin JM, et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet*. 2021;397(10287):1819–1829.
10 Andrews N, Stowe J, Kirsebom F, et al. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *N Engl J Med*. 2022;386(16):1532–1546.
11 Centers for Disease Control and Prevention (CDC). *COVID-19 vaccination clinical & professional resources*; 2023. https://www.cdc.gov/vaccines/covid-19/index.html. Accessed February 23, 2023.
12 Marc A, Kerioui M, Blanquart F, et al. Quantifying the relationship between SARS-CoV-2 viral load and infectiousness. *Elife*. 2021;10:e69302.
13 Hui KPY, Ho JCW, Cheung M-C, et al. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature*. 2022;603(7902):715–720.
14 Meng B, Abdullahi A, Ferreira IATM, et al. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature*. 2022;603(7902):706–714.
15 Madhi SA, Kwatra G, Myers JE, et al. Population immunity and Covid-19 severity with Omicron variant in South Africa. *N Engl J Med*. 2022;386(14):1314–1326.

16 Nyberg T, Ferguson NM, Nash SG, et al. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. *Lancet*. 2022;399(10332):1303–1312.
17 Starr TN, Greaney AJ, Hannon WW, et al. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science*. 2022;377(6604):420–424.
18 Starr TN, Greaney AJ, Stewart CM, et al. Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. *PLoS Pathog*. 2022;18(11):e1010951.
19 Cao Y, Jian F, Wang J, et al. Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. *Nature*. 2022;614:521.
20 Dejnirattisai W, Huo J, Zhou D, et al. SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. *Cell*. 2022;185(3):467–484.e15.
21 Duerr R, Dimartino D, Marier C, et al. Clinical and genomic signatures of SARS-CoV-2 Delta breakthrough infections in New York. *eBioMedicine*. 2022;82:104141.
22 Hoffmann M, Krüger N, Schulz S, et al. The Omicron variant is highly resistant against antibody-mediated neutralization: implications for control of the COVID-19 pandemic. *Cell*. 2022;185(3):447–456.e11.
23 Centers for Disease Control and Prevention (CDC). *Stay up to date with COVID-19 vaccines including boosters*; 2023. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/stay-up-to-date.html. Accessed February 23, 2023.
24 Samanovic MI, Cornelius AR, Gray-Gaillard SL, et al. Robust immune responses are observed after one dose of BNT162b2 mRNA vaccine dose in SARS-CoV-2-experienced individuals. *Sci Transl Med*. 2022;14(631):eabi8961.
25 SWIFT Protocol. *Swift Normalase™ amplicon panels (SNAP)*; 2021. https://swiftbiosci.com/wp-content/uploads/2021/06/PRT-028-Swift-Normalase-Amplicon-Panels-SNAP-SARS-CoV-2-Panels-Rev-9.pdf.
26 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
27 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
28 SWIFT Biosciences. *Primerclip*; 2021. https://github.com/swiftbiosciences/primerclip.
29 Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.
30 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.
31 Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121–4123.
32 Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data–from vision to reality. *Euro Surveill*. 2017;22(13).
33 R Core Team. *R: a language and environment for statistical computing*; 2013. http://www.R-project.org/. Accessed October 3, 2020.
34 RStudio Team. *RStudio: integrated development for R.*; 2015. http://www.rstudio.com/. Accessed October 3, 2020.
35 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B*. 1995;57(1):289–300.
36 Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 2005;22(5):1208–1222.
37 Murrell B, Moola S, Mabona A, et al. FUBAR: a fast, unconstrained Bayesian AppRoximation for inferring selection. *Mol Biol Evol*. 2013;30(5):1196–1205.
38 Greaney AJ, Starr TN, Bloom JD. An antibody-escape estimator for mutations to the SARS-CoV-2 receptor-binding domain. *Virus Evol*. 2022;8(1):veac021.
39 NYC OpenData. *COVID-19 daily counts of cases, hospitalizations, and deaths. October 22, 2021*; 2021. https://data.cityofnewyork.us/Health/COVID-19-Daily-Counts-of-Cases-Hospitalizations-an/rc75-m7u3. Accessed October 23, 2021.
40 NYC Health. *COVID-19: data–vaccines*; 2021. https://www1.nyc.gov/site/doh/covid/covid-19-data-vaccines.page. Accessed October 23, 2021.
41 Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175–191.
42 Centers for Disease Control and Prevention (CDC). *COVID data tracker*; 2022. https://covid.cdc.gov/covid-data-tracker/#datatracker-home. Accessed August 25, 2022.
43 Chemaitelly H, Ayoub HH, Tang P, et al. Long-term COVID-19 booster effectiveness by infection history and clinical vulnerability and immune imprinting: a retrospective population-based cohort study. *Lancet Infect Dis*. 2023;23(7):816–827.
44 COVID-19 Forecasting Team. Past SARS-CoV-2 infection protection against re-infection: a systematic review and meta-analysis. *Lancet*. 2023;401(10379):833–842.
45 Focosi D, Quiroga R, McConnell S, Johnson MC, Casadevall A. Convergent evolution in SARS-CoV-2 spike creates a variant soup from which new COVID-19 waves emerge. *Int J Mol Sci*. 2023;24(3):2264.
46 Liu Y, Rocklöv J. The effective reproductive number of the Omicron variant of SARS-CoV-2 is several times relative to Delta. *J Trav Med*. 2022;29(3):taac037.
47 Joung SY, Ebinger JE, Sun N, et al. Awareness of SARS-CoV-2 omicron variant infection among adults with recent COVID-19 seropositivity. *JAMA Netw Open*. 2022;5(8):e2227241–e.
48 Garrett N, Tapley A, Andriesen J, et al. High asymptomatic carriage with the omicron variant in South Africa. *Clin Infect Dis*. 2022;75(1):e289–e292.
49 Rochman ND, Faure G, Wolf YI, Freddolino PL, Zhang F, Koonin EV. Epistasis at the SARS-CoV-2 receptor-binding domain interface and the propitiously boring implications for vaccine escape. *mBio*. 2022;13(2):e00135522.