

Title: SARS-CoV-2 Omicron subvariant genomic variation associations with immune evasion in Northern California: A retrospective cohort study

Authors: Joshua R. Nugent, PhD¹; Mariah S. Wood, MPH¹; Liyan Liu, MSc¹; Teal Bullick²; Jeffrey M. Schapiro, MD³; Phacharee Arunleung²; Gautham Gautham²; Shiffen Getabeche, MPH²; Christina Morales, PhD²; Laura B. Amsden, MSW, MPH¹; Crystal A. Hsiao, MPH¹; Debra A. Wadford, PhD, MS²; Stacia K. Wyman, PhD, MS⁴; Jacek Skarbinski, MD^{1,3,5,6}

Affiliations:

1. Division of Research, Kaiser Permanente Northern California, Oakland, CA, 94612
2. Viral and Rickettsial Disease Laboratory, Center for Laboratory Sciences, California Department of Public Health, Richmond, California, 94804
3. The Permanente Medical Group, Kaiser Permanente Northern California, Oakland, CA, 94612
4. Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, 94720
5. Department of Infectious Diseases, Oakland Medical Center, Kaiser Permanente Northern California, Oakland, CA, 94611
6. Physician Researcher Program, Kaiser Permanente Northern California, Oakland, CA, 94612

Running title: Omicron genomic variation and immune evasion

Word count: Abstract (249); Manuscript text (2,859)

Corresponding author:

Joshua R. Nugent
Kaiser Permanente Division of Research
4480 Hacienda Drive, Building B
Pleasanton, CA 94588
Email: joshua.r.nugent@kp.org
Mobile: 628-282-1934

Author Contributions:

- Joshua R. Nugent; Mariah S. Wood; Liyan Liu; Jeffrey M. Schapiro; Laura B. Amsden; Crystal A. Hsiao; Stacia K. Wyman; Jacek Skarbinski contributed to conceptual ideas and methodology;
- Joshua R. Nugent; Mariah S. Wood; Liyan Liu; Teal Bullick; Stacia K. Wyman; Jacek Skarbinski contributed to formal analysis;
- Joshua R. Nugent; Liyan Liu; Stacia K. Wyman; Jacek Skarbinski contributed to writing and original draft preparation;
- Joshua R. Nugent; Mariah S. Wood; Liyan Liu; Teal Bullick; Jeffrey M. Schapiro; Phacharee Arunleung; Gautham Gautham; Christina Morales; Laura B. Amsden; Crystal A. Hsiao; Debra A. Wadford; Stacia K. Wyman; Jacek Skarbinski contributed to writing, reviewing, and editing the manuscript;
- Jacek Skarbinski; Phacharee Arunleung; Christina Morales; Debra A. Wadford supervised the work.

Declaration of interests: All authors: No conflicts of interest identified. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Abstract

Background: The possibility of association between SARS-CoV-2 genomic variation and immune evasion is not known among persons with Omicron variant SARS-CoV-2 infection.

Methods: In a retrospective cohort, using Poisson regression adjusting for sociodemographic variables and month of infection, we examined associations between individual non-lineage defining mutations and SARS-CoV-2 immunity status, defined as a) no prior recorded infection, b) not vaccinated but with at least one prior recorded infection, c) complete primary series vaccination, and/or d) primary series vaccination and ≥ 1 booster. We identified all non-synonymous single nucleotide polymorphisms (SNPs), insertions and deletions in SARS-CoV-2 genomes with $\geq 5\%$ allelic frequency and population frequency of $\geq 5\%$ and $\leq 95\%$. We also examined correlations between the presence of SNPs with each other, with subvariants, and over time.

Results: Seventy-nine mutations met inclusion criteria. Among 15,566 persons infected with Omicron SARS-CoV-2, 1,825 (12%) were unvaccinated with no prior recorded infection, 360 (2%) were unvaccinated with a recorded prior infection, 13,381 (86%) had a complete primary series vaccination, and 9,172 (58%) had at least one booster. After examining correlation between SNPs, 79 individual non-lineage defining mutations were organized into 38 groups. After correction for multiple testing, no individual SNPs or SNP groups were significantly associated with immunity status levels.

Conclusions: Genomic variation identified within SARS-CoV-2 Omicron specimens was not significantly associated with immunity status, suggesting that contribution of non-lineage defining SNPs to immune evasion is minimal. Larger-scale surveillance of SARS-CoV-2 genomes linked with clinical data can help provide information to inform future vaccine development.

Key words: SARS-CoV-2; genomic variation; immune evasion

Background

Since its emergence, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been evolving, with evidence for increased transmissibility and ability to evade existing immune responses elicited by prior infection or COVID-19 vaccination.^{1,2} Multiple mutations that increase transmission and immune evasion were identified when the Delta and subsequent Omicron variants (and sub-variants) emerged. However, few in-depth analyses characterizing the genomic variation within a circulating variant and identifying potential mutations associated with immune evasion have been conducted. In addition, during the Omicron wave, individual immunity could be due to prior infection or multiple levels of vaccination, due to the availability of vaccine boosters in this timeframe, and any immune evasion characteristics found among mutations could be tied to a specific type of immunity, derived from, for example, prior infection, immunization dose(s), or both.

To address these data gaps, we conducted a retrospective cohort study of persons with nucleic-acid amplification test (NAAT)-confirmed SARS-CoV-2 infection with the Omicron variant confirmed by whole genome sequencing. Patients were members of a large, integrated health system that had detailed epidemiologic and clinical data on all participants. The study assessed overall SARS-CoV-2 genetic variation over time as well as the relationship between mutations and the individual's source(s) of potential COVID-19 immunity. The key assumption in this study is that after adjusting for sociodemographic characteristics and timeframe, an association between mutation presence and vaccine- or infection-derived immunity could imply greater immune evasion for that mutation.

Methods

Setting

Kaiser Permanente Northern California (KPNC) is an integrated health system that serves over 4.5 million members in Northern and Central California and provides comprehensive preventive and

curative care in inpatient and outpatient settings across 266 medical offices and 21 hospitals. Members receive most clinical services, including laboratory testing, outpatient, and inpatient care in KPNC facilities. Members have similar sociodemographic characteristics to the population of Northern and Central California.³ SARS-CoV-2 NAAT for outpatients and hospitalized patients is conducted at a regional laboratory or local hospital laboratories using various platforms and specimens are sent to the California Department of Public Health (CDPH) for whole genome sequencing.

Study design

We conducted a retrospective cohort study of persons with NAAT-confirmed SARS-CoV-2 infection and whole genome sequencing-confirmed Omicron variant infection and assessed patterns among non-lineage-defining mutations as well as the association between SARS-CoV-2 genetic variation and immunity source (if any). This immunity was defined by prior reported SARS-CoV-2 infection or vaccination. Epidemiologic and clinical data were obtained from the KPNC Virtual Data Warehouse, a common data model into which standardized data are extracted from clinical and administrative databases including an integrated electronic health record database (Epic, Verona, WI, USA), and from other sources.^{4,5} The study was approved by the KPNC and University of California, Berkeley Institutional Review Boards with waivers of the requirement for informed consent. This activity was reviewed by the Centers for Disease Control and Prevention (CDC) and conducted consistent with applicable federal law and CDC policy (see e.g., 45 C.F.R. part 46.102(l)(2), 21 C.F.R. part 56; 42 U.S.C. §241(d); 5 U.S.C. §552a; 44 U.S.C. §3501 et seq.).

Inclusion criteria

We included all persons with an incident NAAT-confirmed SARS-CoV-2 infection and whole genome sequencing-confirmed Omicron variant infection from January 1, 2022 to October 31, 2022 who

met the following criteria: 1) KPNC membership for at least one year prior to SARS-CoV-2 diagnosis; 2) aged ≥ 5 years and < 90 years; 3) no prior receipt of COVID-19 vaccine or the completion of a primary COVID-19 vaccine series of two doses of BNT162b2 [Pfizer/BioNTech] or m-RNA-1973 [Moderna/National Institutes of Health] or one dose of Ad.26.COV2.s [Janssen] more than 14 days prior to index date; 4) incident infection defined as the first NAAT-confirmed SARS-CoV-2 infection during the study period. The index date was defined as the date of NAAT-confirmation of SARS-CoV-2 infection.

SARS-CoV-2 whole genome sequencing and processing

KPNC is a partner in COVIDNet with CDPH and submits a majority of NAAT-confirmed SARS-CoV-2 specimens to COVIDNet for genomic surveillance; a subset of specimens undergo whole genome sequencing at CDPH or affiliated laboratories based on CDPH priorities, including prioritizing hospitalized cases and a convenience sample of the first specimens received per week (number varies by burden).⁶ COVIDNet consists of a set of (primarily University of California-affiliated) laboratories tapped to assist CDPH because they had surplus sequencing capacity. Illumina sequencing was used by most of the sites, with about 10% of genomes being sequenced by Oxford Nanopore. For this study, we only included Illumina-sequenced genomes because the error rate of Nanopore is too high to accurately identify low-frequency mutations. Varying sets of primers were used by different institutions, but the most common for the Omicron genomes were the ARTICv3 and ARTICv4 primers. The University of California, San Francisco lab used their own custom set of primers with more primers and shorter amplicons. After sequencing, the CDPH Viral and Rickettsial Disease Laboratory links lineage information with patient identifiers and shares the results with KPNC. The raw sequence files are shared with the University of California, Berkeley (UCB); KPNC and UCB collaborate further to link demographic, clinical, and epidemiologic data from KPNC with subsequent sequence analysis data from UCB.

The raw sequencing data were processed through a SARS-CoV-2 analysis pipeline that was modified for this work as follows. Adapter removal and trimming were performed using bbduk (<http://sourceforge.net/projects/bbmap/>). The reads were then aligned to the Wuhan reference genome (GISAID ID: EPI_ISL_402125, GenBank accession number MN908947) using minimap2⁷ followed by primer trimming using iVar.⁸ Samtools was used to create a pileup file from which a consensus sequence was generated. This consensus genome was created with iVar using a minimum depth of 10 reads and majority rule for base calling. We next used iVar to call variants from the pileup file where we set the sensitivity threshold for calling a mutation to 2%. This process calls mutations for any loci where at least 2% of the reads are non-reference. Using this very low threshold allowed us to capture even very low frequency mutations in the sequencing data. This threshold resulted in many spurious mutations from sequencing and alignment errors and therefore we later set a threshold of 5% when processing mutations. This original list of variants was then annotated with the gene and amino acid change (if there was one) or insertion or deletion.

After calling with iVar using a low-frequency threshold, mutations were compiled into a master frequency matrix listing all mutations that appeared in any genome in rows, and all the genomes that each mutation appeared in columns (https://github.com/staciawyman/SC2_GenVar). The cell values were frequency of the alternate reads at the location. A mutation had to have an allele frequency of at least 5% to be included to rule out sequencing errors. This matrix was partitioned into variant-specific parts to facilitate analysis. We constructed an Omicron matrix including all individuals that had the Omicron lineage⁹ between January 1, 2022 and October 31, 2022 including non-synonymous SNPs, insertions and deletions (indels). Pipeline code is available at https://github.com/staciawyman/SC2_GenVar.

Outcome measure

The primary outcome measure was, among persons with an incident NAAT-confirmed SARS-CoV-2 infection, the presence of individual mutations (nonsynonymous SNPs, insertions and deletions) with $\geq 5\%$ allelic frequency and $\geq 5\%$ and $\leq 95\%$ population frequency, as well as mutation groups created by grouping SNPs with a pairwise Pearson correlation coefficient > 0.7 . Additionally, patterns in the prevalence of these mutations were assessed as follows: as noted earlier, pairwise Pearson correlations with other mutations in each infection; presence within Omicron subvariant (BA.1, BA.2, BA.4, BA.5, BCD, BE, BF, Bdot), and prevalence over time.

Primary exposure: SARS-CoV-2 immunity status

The primary exposure of interest in this analysis was immunity level, defined in mutually exclusive categories: i) no recorded vaccination or recorded prior infection; ii) recorded prior infection but no recorded vaccination; and iii) recorded vaccination (primary series complete and/or more than primary series), regardless of prior infection status.

Statistical analysis

Adjusted relative prevalence, shown as an adjusted rate ratio (aRR), and 95% confidence intervals (CIs) for the estimates for our exposure and adjustment variables were obtained from modified Poisson regression models with robust standard errors using PROC GENMOD with a log link function in SAS (Version 9.4, Cary, North Carolina); statistical significance was defined as a p-value < 0.05 . All p-values were adjusted for multiple comparisons to control the false discovery rate at 5%, using the Benjamini-Hochberg procedure.¹⁰ To adjust for confounding, all models included age category, sex, race/ethnicity, Charlson comorbidity index score¹¹, and index month. Adjusting for these demographic and clinical factors accounted for differential vaccination status of these subgroups, increasing

vaccination rates over time, and the inherent uncertainty as to where in a population a new mutation might arise.

Data Availability

De-identified data that underlie the results reported in this article will be made available upon reasonable request.

Role of funding source

The funders had no role in study design, data collection, data analysis, interpretation, or writing of the report.

Results

From January 1, 2022 to October 31, 2022, 312,249 SARS-CoV-2-positive specimens were submitted for sequencing, 18,603 (6.0%) underwent sequencing (Supplemental Figure 1), and 18,476 (5.9%) had reported lineages consistent with the Omicron variant. After applying selection criteria, 15,556 persons with SARS-CoV-2 Omicron variant infection were included in the analysis: 9,172 (58%) vaccinated with a booster; 4,209 (26%) vaccinated with a complete primary series [two doses of Pfizer or Moderna vaccine]; and 2,185 (14%) with no vaccination (Supplemental Figure 1). The prevalence of Omicron subvariants in our specimens, aggregated to the week level, showed increasing diversity over time (Figure 1).

After applying selection criteria to the sequences, we identified 78 non-lineage defining mutations (all non-synonymous SNPs) in 1,456 SARS-CoV-2 SNPs included in this analysis, shown in Supplemental Figure 2. Identified mutations were in the spike gene (n=44), nucleocapsid gene (n=8) open reading frame 1a (ORF1a) (n=19), ORF1b (n=4), ORF3a (n=1), ORF6c (n=1), ORF10 (n=1). The

population frequency within our genomes of individual identified mutations ranged from 5.0% to 95.0% and 46 out of 79 (58.2%) mutations were present in <50% of the overall population. Although we called low-frequency mutations with our pipeline, none of them met the criteria for inclusion in our analysis.

Of the 78 identified SNPs that met our inclusion criteria, many co-occurred on the same specimens. Supplemental Figure 2 shows the pairwise correlations of SNP presence in our dataset. Given the strong pairwise correlations observed, we formed groups of SNPs that had a pairwise Pearson correlation $> .7$, reducing the dimension to 33 SNPs/SNP groups, also shown in Supplemental Figure 2. These mutation groups are largely monophyletic and segregate together across the phylogeny of Omicron genomes (Supplemental Figure 3).

The demographics of persons with SARS-CoV-2 infection included in this analysis, stratified by vaccination and prior infection status, are shown in Table 1, and the number of total reported infections and those sequenced over time is shown in Supplemental Figure 4. Significant differences were found between groups, thus supporting their inclusion in our covariate adjustment set. Figure 2 details some of the relationships between SNP groups, time, and subvariant among the specimens in our data set. While certain mutations/groups appear essentially only in a single subvariant (for example, 99.1% of S:H69L mutations appear in specimens classified as BA.1 in Figure 2c), the majority of mutations/groups appear in many subvariants, suggesting that they are not simply proxy measures of subvariant. Further, the presence of mutations/groups do not follow clear temporal patterns (Figure 2b). While some mutations such as Group 12 arise, become plentiful, then wane, others disappear and re-appear over time without a clear pattern (for example, S:D1259H). Smoothed over time, grouped SNP prevalence by month is shown in Supplemental Figure 5.

Using modified Poisson regression models with robust standard errors, we found several of our 33 SNP groupings (Table 2) and/or 78 individual mutations (Supplemental Table 1) that were significantly associated with immunity status; however, after correcting p-values to adjust for multiple

comparisons, none of the associations were statistically significant. These findings were not meaningfully changed when people with a booster vaccine were/were not aggregated with the complete primary series group (Supplemental Table 2, Supplemental Table 3). The SNPs/groups with the largest absolute associations with vaccination (Table 2) included some that were strongly associated with a single variant (for example, ORF10:L37F, which occurred almost exclusively in BA.5 subvariants) as well as some that frequently appeared in multiple variants (for example, ORF1b:N498I and group 45).

When looking at the evolutionary relationships with the genomes colored by vaccination status (Supplemental Figure 6), vaccination status did not show a correlation with a clade/subvariant or distance from the root (number of mutations with respect to the reference SARS-CoV-2 genome).

Discussion

Characterizing the continuing evolution of SARS-CoV-2, along with active surveillance of infections, is critical to targeting public health measures and vaccine development. Large-scale global whole genome sequencing efforts have contributed substantially to our understanding of the genetic variation of SARS-CoV-2 and its associations with transmissibility and immune evasion. Our current study adds to this evidence base by assessing within-variant (Omicron) variation in the genome beyond the already-identified subvariants such as BA.1 and BA.2 using a unique whole genome surveillance system linking genomic data with detailed data on demographic, clinical and epidemiologic characteristics in a well-defined population.

Understanding the potential for immune evasion based on mutations beyond the spike protein is critical for vaccine development and understanding the mechanisms of the human immune response. In this detailed analysis of genomic variation within the Omicron variant over a 10-month period in California, we identified only 78 mutations that met our inclusion criteria among 15,566 genomes. These mutations occurred across the entire genome, not just in the spike protein. Spike protein mutations

previously implicated with immune evasion were excluded based on population frequency. We encourage future research to examine the entire viral genome for potential immune evasion or increased replication or infectivity.

Our results showed that there appear to be some mutations that are tightly grouped with certain subvariants, while a number of other mutations appear across multiple subvariants. While we did not find evidence for immune evasion among our identified mutations/groups after correction for multiple testing, it is possible that we are underpowered to detect small effects over a short time period.¹² Many of the SNPs with the highest adjusted risk ratios appeared in a small number of specimens (Supplemental Table 1). While one would expect more variability from the smallest table cells if we were simply observing statistical noise, it is also possible that a larger sample could confirm these stronger adjusted risk ratios.

Our analysis has several limitations. First, not all possible demographic and clinical variables were included in our models, and if those variables are associated with immunity status and SNP presence, unmeasured confounding could bias our analysis. Second, there is a risk of misclassification of immunity status; though the KPNC surveillance system for vaccination is quite robust, prior infections (say, diagnosed by a home test) may have gone unreported. Third, we focused on a set of specimens from Northern and Central California, limiting the generalizability of our findings to this context. Finally, there may be hidden bias as to which specimens were successfully sequenced based on the nucleic acid viral load in the specimen (e.g., specimens with lower viral loads might not be successfully sequenced and thus would have been excluded from this analysis). Fourth, although our surveillance system has a very robust system for assessing individual-level vaccination status, there is a small risk of misclassification of persons as having incorrect immunity status.

As part of the global effort to conduct near real-time genomic surveillance for SARS-CoV-2, there has been tremendous progress in developing national genomic surveillance systems with nearly real-

time public reporting of circulating SARS-CoV-2 variants and uploading of consensus sequence data to publicly available genomic databases.^{13 14 15 16} However, these efforts can be improved. First, enhancing methods to define consensus sequences to also assess for minor variation with <50% allelic frequency would better define the true genomic variation of circulating SARS-CoV-2, including capturing within-host genetic diversity.^{17,18} Second, current SARS-CoV-2 genomic surveillance systems in the United States have limited metadata on human demographics, clinical factors, epidemiologic factors, and immunity status (e.g. COVID-19 vaccination and documented prior infection), particularly when compared to the United Kingdom.¹⁹ This limits the ability to correlate noted patterns in SARS-CoV-2 genetic variation and real-world clinical and epidemiologic outcomes in the United States. We present one potential analysis that uses this data linkage to better understand how SARS-CoV-2 genetic variation might be associated with human disease, in this case evidence for immune evasion associated with specific mutation clusters. Improving the clinical metadata associated with SARS-CoV-2 genomic data should be a priority as we continue to improve and expand genomic surveillance. We have demonstrated this capability through partnership and collaboration between a large healthcare system and a state public health department, namely KPNC and CDPH/COVIDNet.

In summary, we demonstrate that linking SARS-CoV-2 genomic sequence data with detailed clinical data revealed no significant associations between genetic variation among Omicron SARS-CoV-2 mutations and immunity status. Further research using larger sample sets that link SARS-CoV-2 genetic data with clinical data is warranted for the potential to reveal whether SARS-CoV-2 non-lineage defining genetic variation might contribute to the fitness of SARS-CoV-2 by either immune evasion or other mechanism.

Funding Statement: This study was supported by the Rockefeller Foundation (SKW, JS), National Cancer Institute of the National Institutes of Health under Award Number U01CA260584 as part of the Serological Sciences Network (SeroNet) (JS), the Physician Researcher Program of The Permanente Medical Group Delivery Science and Applied Research Program (JS), The Packard Foundation (SKW), and The Sergey Brin Family Foundation (SKW). CDPH/COVIDNet genomic surveillance work was funded by Centers for Disease Control and Prevention, Epidemiology and Laboratory Capacity for Infectious Diseases, Cooperative Agreement Number 5 NU50CK000539.

Declaration of interests: All authors: No conflicts of interest identified. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Acknowledgments: The authors are grateful to all Kaiser Permanente members without whom this study would not have been possible. In addition, we wish to acknowledge our partnership with the California Department of Public Health and participation in the California SARS-CoV-2 Whole Genome Sequencing Initiative (COVIDNet). We thank the VRDL and COVIDNet Teams: Summer Adams, Allison Bailey, Matt Bacinskas, Nikki Baumrind, Elizabeth Baylis, John Bell, Ricardo Berumen III, Yocelyn Cruz, Mojgan Deldari, Alex Espinosa, Sabrina Gilliam, Madeline Glenn, Bianca Gonzaga, Melanie Greengard, Jill Hacker, Kim Hansard, Monica Haw, Thalia Huynh, Chantha Kath, Ruth Lopez, Sharon Messenger, Alexa Quintana, Chris Preas, Clarence Reyes, Maria Salas, Hilary Tamnanchit, Serena Ting, Kathy Jacobson, MD, and Carol Glaser, MD.

Data sharing statement: De-identified data that underlie the results reported in this article will be made available upon reasonable request.

332

333 **Disclaimer:** The content is solely the responsibility of the authors and does not represent the official
334 views or opinions of the National Institutes of Health, Kaiser Permanente, California Department of
335 Public Health or the California Health and Human Services Agency.

336

337 ***Additional Disclaimer:** Use of trade names and commercial sources is for identification only and does
338 not imply endorsement by the California Department of Public Health or the California Health and
339 Human Services Agency.

340

References

1. Lou F, Li M, Pang Z, et al. Understanding the Secret of SARS-CoV-2 Variants of Concern/Interest and Immune Escape. *Front Immunol.* 2021;12. Accessed November 13, 2023. <https://www.frontiersin.org/articles/10.3389/fimmu.2021.744242>
2. Harvey WT, Carabelli AM, Jackson B, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol.* 2021;19(7):409-424. doi:10.1038/s41579-021-00573-0
3. Davis AC, Voelkel JL, Remmers CL, Adams JL, McGlynn EA. Comparing Kaiser Permanente Members to the General Population: Implications for Generalizability of Research. *Perm J.* 2023;27(2):87-98. doi:10.7812/TPP/22.172
4. Ross TR, Ng D, Brown JS, et al. The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. *EGEMS Wash DC.* 2014;2(1):1049. doi:10.13063/2327-9214.1049
5. Weeks J, Pardee R. Learning to Share Health Care Data: A Brief Timeline of Influential Common Data Models and Distributed Health Data Networks in U.S. Health Care Research. *EGEMS Wash DC.* 2019;7(1):4. doi:10.5334/egems.279
6. Wadford DA, Baumrind N, Baylis EF, et al. Implementation of California COVIDNet – a multi-sector collaboration for statewide SARS-CoV-2 genomic surveillance. *Front Public Health.* 2023;11. Accessed November 26, 2023. <https://www.frontiersin.org/articles/10.3389/fpubh.2023.1249614>
7. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma Oxf Engl.* 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191
8. Grubaugh ND, Gangavarapu K, Quick J, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* 2019;20(1):8. doi:10.1186/s13059-018-1618-7
9. O’Toole Á, Scher E, Underwood A, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* 2021;7(2):veab064. doi:10.1093/ve/veab064

10. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol.* 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x
11. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8
12. Volz E. Fitness, growth and transmissibility of SARS-CoV-2 genetic variants. *Nat Rev Genet.* 2023;24(10):724-734. doi:10.1038/s41576-023-00610-z
13. Chen Z, Azman AS, Chen X, et al. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. *Nat Genet.* 2022;54(4):499-507. doi:10.1038/s41588-022-01033-y
14. World Health Organization. *Guidance for Surveillance of SARS-CoV-2 Variants: Interim Guidance, 9 August 2021.* World Health Organization; 2021. <https://apps.who.int/iris/handle/10665/343775>
15. Lambrou AS. Genomic Surveillance for SARS-CoV-2 Variants: Predominance of the Delta (B.1.617.2) and Omicron (B.1.1.529) Variants — United States, June 2021–January 2022. *MMWR Morb Mortal Wkly Rep.* 2022;71. doi:10.15585/mmwr.mm7106a4
16. Robishaw JD, Alter SM, Solano JJ, et al. Genomic surveillance to combat COVID-19: challenges and opportunities. *Lancet Microbe.* 2021;2(9):e481-e484. doi:10.1016/S2666-5247(21)00121-X
17. Tonkin-Hill G, Martincorena I, Amato R, et al. Patterns of within-host genetic diversity in SARS-CoV-2. Neher RA, Sawyer SL, Neher RA, Luring AS, eds. *eLife.* 2021;10:e66857. doi:10.7554/eLife.66857
18. Gao R, Zu W, Liu Y, et al. Quasispecies of SARS-CoV-2 revealed by single nucleotide polymorphisms (SNPs) analysis. *Virulence.* 2021;12(1):1209-1226. doi:10.1080/21505594.2021.1911477
19. Nicholls SM, Poplawski R, Bull MJ, et al. CLIMB-COVID: continuous integration supporting decentralised sequencing for SARS-CoV-2 genomic surveillance. *Genome Biol.* 2021;22(1):196. doi:10.1186/s13059-021-02395-y

Figure 1. SARS-CoV-2-positive specimens tested by Kaiser Permanente Northern California, proportion by subvariant, by week from January 1, 2022 to October 31, 2022

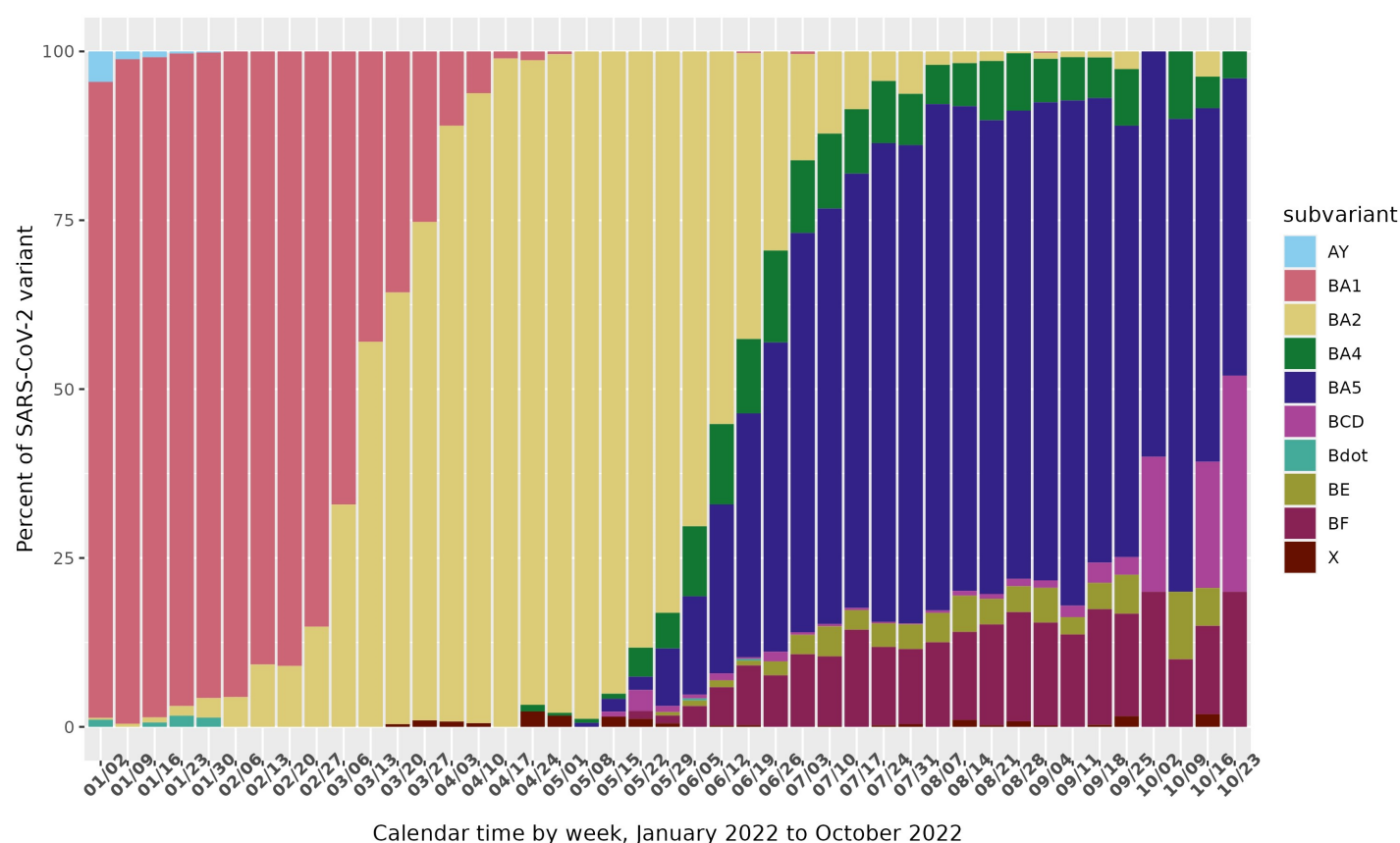
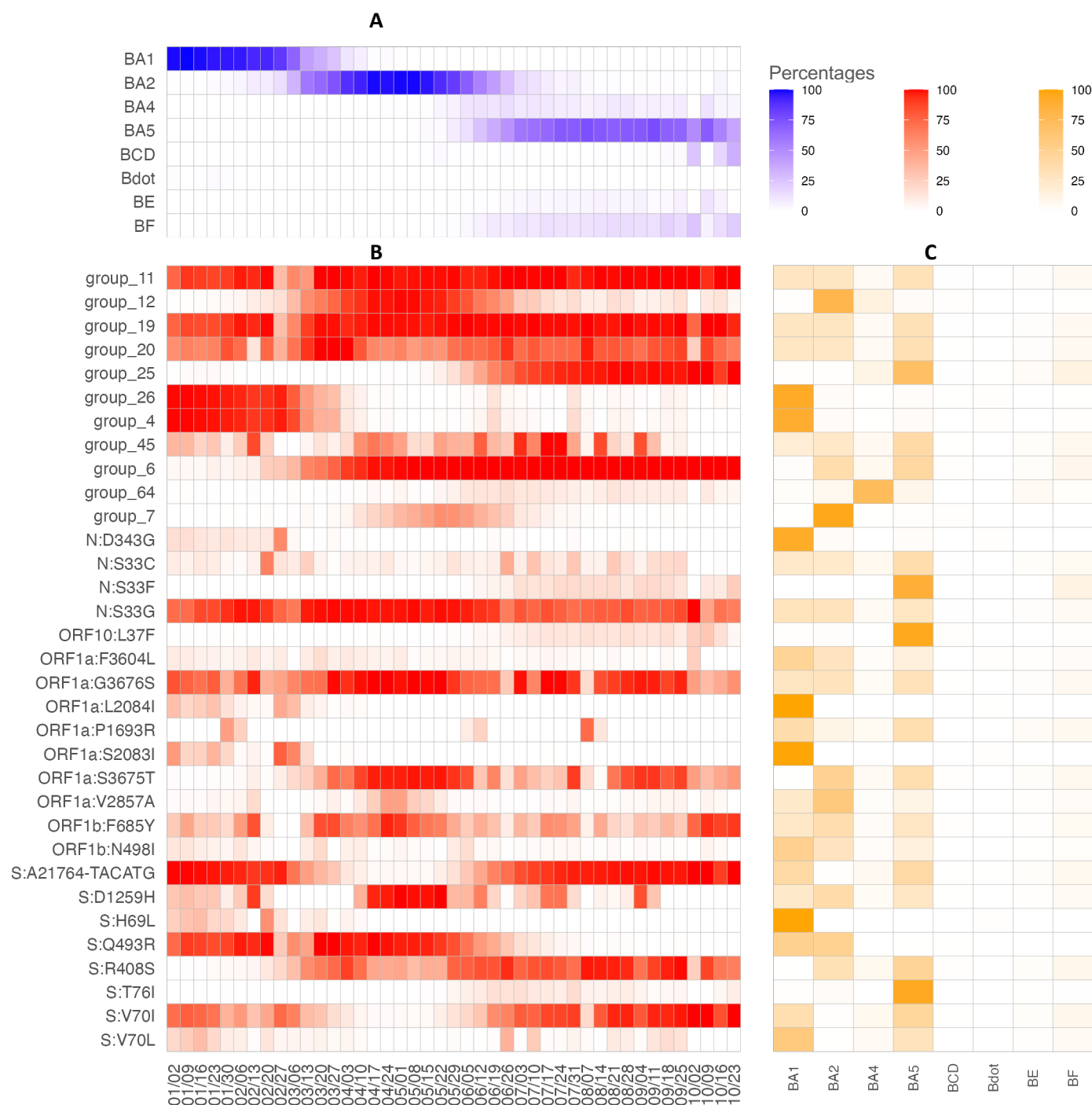


Figure 2: SNP/group prevalence by subvariant and week in 2022



- (A) Proportion of specimens in each sublineage by week
- (B) Proportion of specimens with a given SNP/group by week
- (C) Proportion of subvariant by SNP/group (rows sum to 100%)

Table 1: Demographics of persons included in this analysis with SARS-CoV-2 infection with Omicron variant in Northern California stratified by vaccination/infection status, January 1, 2022 - October 31, 2022

Characteristic	All n (%)	Not vaccinated, no prior reported infection n (%)	Not vaccinated, prior reported infection n (%)	Vaccinated n (%)	p-value ¹
Total	15,566 (100%)	1,825 (11.7%)	360 (2.3%)	13,381 (86.0%)	
Median age in years (Interquartile range)	43.0 (30.0, 57.0)	32.0 (13.0, 49.0)	32.0 (19.0, 43.0)	44.0 (32.0, 58.0)	<0.0001
Age in years					<0.0001
5-19	2,019 (13.0%)	594 (32.5%)	95 (26.4%)	1,330 (9.9%)	
20-29	1,776 (11.4%)	219 (12.0%)	57 (15.8%)	1,500 (11.2%)	
30-39	3,054 (19.6%)	323 (17.7%)	97 (26.9%)	2,634 (19.7%)	
40-49	2,859 (18.4%)	242 (13.3%)	53 (14.7%)	2,564 (19.2%)	
50-59	2,572 (16.5%)	189 (10.4%)	35 (9.7%)	2,348 (17.5%)	
60-69	1,879 (12.1%)	135 (7.4%)	13 (3.6%)	1,731 (12.9%)	
70-79	1,023 (6.6%)	88 (4.8%)	8 (2.2%)	927 (6.9%)	
80-89	384 (2.5%)	35 (1.9%)	2 (0.6%)	347 (2.6%)	
Sex					0.0003
Female	8,665 (55.7%)	929 (50.9%)	204 (56.7%)	7,532 (56.3%)	
Male	6,901 (44.3%)	896 (49.1%)	156 (43.3%)	5,849 (43.7%)	
Race/ethnicity					<0.0001
LatinX/Hispanic ethnicity	4,303 (27.6%)	591 (32.4%)	158 (43.9%)	3,554 (26.6%)	
Black/African descent	1,368 (8.8%)	267 (14.6%)	45 (12.5%)	1,056 (7.9%)	
Asian descent	4,090 (26.3%)	211 (11.6%)	38 (10.6%)	3,841 (28.7%)	
White/European or middle eastern descent	4,788 (30.8%)	603 (33.0%)	100 (27.8%)	4,085 (30.5%)	
Other/unknown	1,017 (6.5%)	153 (8.4%)	19 (5.3%)	845 (6.3%)	
Charlson comorbidity index score					0.44
Score 0	10,082 (64.8%)	1,218 (66.7%)	271 (75.3%)	8,593 (64.2%)	
Score 1	2,278 (14.6%)	225 (12.3%)	48 (13.3%)	2,005 (15.0%)	
Score 2	881 (5.7%)	60 (3.3%)	4 (1.1%)	817 (6.1%)	
Score 3+	1,141 (7.3%)	118 (6.5%)	11 (3.1%)	1,012 (7.6%)	
Missing (no visits in prior year)	1,184 (7.6%)	204 (11.2%)	26 (7.2%)	954 (7.1%)	
COVID-19 vaccination					
None	2,185 (14.0%)	1,825 (100.0%)	360 (100.0%)		
Complete primary series					
Johnson & Johnson	803 (5.2%)			803 (6.0%)	
Moderna	4,359 (28.0%)			4,359 (32.6%)	
Pfizer	8,219 (52.8%)			8,219 (61.4%)	
Month of SARS-CoV-2 infection					0.007
January 2022	3,120 (20.0%)	663 (36.3%)	65 (18.1%)	2,392 (17.9%)	
February 2022	1,280 (8.2%)	250 (13.7%)	26 (7.2%)	1,004 (7.5%)	
March 2022	791 (5.1%)	92 (5.0%)	13 (3.6%)	686 (5.1%)	
April 2022	1,090 (7.0%)	88 (4.8%)	10 (2.8%)	992 (7.4%)	
May 2022	1,210 (7.8%)	95 (5.2%)	21 (5.8%)	1,094 (8.2%)	
June 2022	1,837 (11.8%)	152 (8.3%)	53 (14.7%)	1,632 (12.2%)	
July 2022	2,770 (17.8%)	212 (11.6%)	75 (20.8%)	2,483 (18.6%)	
August 2022	2,263 (14.5%)	182 (10.0%)	62 (17.2%)	2,019 (15.1%)	
September 2022	1,070 (6.9%)	83 (4.5%)	32 (8.9%)	955 (7.1%)	
October 2022	135 (0.9%)	8 (0.4%)	3 (0.8%)	124 (0.9%)	
Months from last vaccine dose to SARS-CoV-2 infection					
Mean (standard deviation)	6.5 (3.7)			6.5 (3.7)	

Characteristic	All n (%)	Not vaccinated, no prior reported infection n (%)	Not vaccinated, prior reported infection n (%)	Vaccinated n (%)	p-value ¹
Median (25 th percentile, 75 th percentile)	6.3 (3.7, 8.6)			6.3 (3.7, 8.6)	
Immune compromising conditions²					
Any immune compromising condition	1,092 (7.0%)	92 (5.0%)	16 (4.4%)	984 (7.4%)	0.0009
Severe clinical outcomes					
Any hospitalization	483 (3.1%)	164 (9.0%)	5 (1.4%)	314 (2.3%)	<0.0001

¹Wilcoxon rank sum test; Pearson's Chi-squared test

²Immune compromising conditions: HIV, organ transplant, cancer, rheumatologic/inflammatory conditions, and other immune deficiencies.

Table 2: Association between mutation groups¹ and vaccine status (pooling over complete and boosted) among persons included in this analysis with SARS-CoV-2 infection with Omicron variant in Northern California, January 1, 2022 - October 31, 2022 (N=15,566)²

SNP/group	Not vaccinated, no recorded prior infection	Not vaccinated, recorded prior infection			Vaccinated, regardless of prior infection status		
		aRR (95% CI)	p (raw)	p (FDR)	aRR (95% CI)	p (raw)	p (FDR)
group64	Reference	1.26 (0.82- 1.93)	0.2845	0.97743	1.22 (0.96- 1.56)	0.1097	0.72385
ORF10:L37F	Reference	1.00 (0.61- 1.62)	0.9849	0.98491	1.18 (0.91- 1.53)	0.2210	0.86611
ORF1b:N498I	Reference	0.96 (0.57- 1.63)	0.8889	0.97743	1.13 (0.92- 1.40)	0.2404	0.86611
group45	Reference	1.13 (0.96- 1.34)	0.1504	0.97743	1.10 (1.01- 1.19)	0.0282	0.46532
ORF1a:V2857A	Reference	1.38 (0.84- 2.29)	0.2047	0.97743	1.09 (0.86- 1.39)	0.4703	0.90946
S:D1259H	Reference	1.06 (0.87- 1.29)	0.5509	0.97743	1.08 (0.99- 1.19)	0.0834	0.68786
S:T76I	Reference	1.25 (0.80- 1.95)	0.3219	0.97743	1.07 (0.83- 1.40)	0.5906	0.91512
N:S33C	Reference	1.07 (0.78- 1.45)	0.6751	0.97743	1.06 (0.92- 1.24)	0.4194	0.89674
group25	Reference	1.04 (0.88- 1.23)	0.6343	0.97743	1.04 (0.95- 1.14)	0.4105	0.89674
S:V70L	Reference	0.89 (0.62- 1.29)	0.5376	0.97743	1.04 (0.90- 1.20)	0.6297	0.91512
group7	Reference	0.95 (0.66- 1.36)	0.7830	0.97743	1.04 (0.87- 1.24)	0.6892	0.91512
ORF1a:F3604L	Reference	0.98 (0.59- 1.64)	0.9420	0.97743	1.04 (0.84- 1.28)	0.7322	0.91512
group19	Reference	1.02 (0.91- 1.15)	0.6952	0.97743	1.03 (0.98- 1.09)	0.2357	0.86611
N:S33G	Reference	0.97 (0.85- 1.10)	0.6454	0.97743	1.03 (0.98- 1.09)	0.2335	0.86611
ORF1a:G3676S	Reference	1.02 (0.90- 1.16)	0.7179	0.97743	1.03 (0.97- 1.09)	0.4073	0.89674
ORF1b:F685Y	Reference	1.01 (0.84- 1.21)	0.9202	0.97743	1.03 (0.95- 1.12)	0.4348	0.89674
S:H69L	Reference	0.61 (0.36- 1.03)	0.0638	0.97743	1.03 (0.88- 1.19)	0.7379	0.91512
group11	Reference	0.99 (0.88- 1.11)	0.8721	0.97743	1.02 (0.97- 1.08)	0.4105	0.89674
S:Q493R	Reference	0.92 (0.78- 1.10)	0.3603	0.97743	1.02 (0.95- 1.09)	0.6063	0.91512
group12	Reference	0.92 (0.74- 1.15)	0.4770	0.97743	1.02 (0.92- 1.13)	0.6863	0.91512
S:V70I	Reference	1.03 (0.89- 1.19)	0.6590	0.97743	1.01 (0.95- 1.08)	0.6802	0.91512
S:A21764-TACATG	Reference	1.04 (0.92- 1.18)	0.5241	0.97743	1.01 (0.95- 1.07)	0.8278	0.93114
group6	Reference	0.99 (0.86- 1.13)	0.8384	0.97743	1.01 (0.94- 1.08)	0.7811	0.92055
S:R408S	Reference	1.01 (0.87- 1.18)	0.8820	0.97743	1.01 (0.93- 1.09)	0.8725	0.93114
ORF1a:S3675T	Reference	0.96 (0.80- 1.16)	0.7049	0.97743	1.01 (0.92- 1.10)	0.9048	0.93114
group20	Reference	1.00 (0.88- 1.15)	0.9478	0.97743	1.00 (0.94- 1.07)	0.9311	0.93114
group26	Reference	1.06 (0.88- 1.29)	0.5288	0.97743	0.99 (0.92- 1.06)	0.7487	0.91512
ORF1a:P1693R	Reference	1.06 (0.68- 1.63)	0.8043	0.97743	0.99 (0.81- 1.21)	0.9139	0.93114
group4	Reference	1.06 (0.88- 1.28)	0.5443	0.97743	0.98 (0.91- 1.05)	0.4961	0.90946
N:D343G	Reference	0.59 (0.31- 1.11)	0.1019	0.97743	0.90 (0.75- 1.08)	0.2628	0.86611
N:S33F	Reference	1.34 (0.94- 1.92)	0.1096	0.97743	0.89 (0.71- 1.11)	0.2887	0.86611
ORF1a:S2083I	Reference	1.06 (0.74- 1.53)	0.7369	0.97743	0.87 (0.76- 1.00)	0.0431	0.47401
ORF1a:L2084I	Reference	0.90 (0.58- 1.39)	0.6213	0.97743	0.84 (0.72- 0.97)	0.0177	0.46532

¹ Mutations with Pearson pairwise correlation >0.70 combined into groups for analysis:

group 4: M:D3G, ORF1a:I3758V, ORF1a:K856R, S:A67V, S:G496S, S:N211I, S:N856K, S:R346K, S:S371P, S:T22204+GAGCCAGAA, S:T547K, S:T95I

group 6: N:S413R, ORF1a:F3677L, ORF1a:G1307S, ORF1a:L3027F, ORF1a:S135R, ORF1a:T3090I, ORF1a:T842I, ORF1b:R1315C, ORF1b:T2163I, ORF3a:T223I, S:D405N, S:G142D, S:L24S, S:T19I, S:T376A, S:V213G

group 7: S:L452Q, S:S704L

group 11: S:E484A, S:N501Y, S:Q498R, S:S477N, S:T478K, S:Y505H

group 12: ORF1a:L3201F, ORF6:D61L

group 19: S:S371F, S:S373P, S:S375F

group 20: S:K417N, S:N440K

group 25: M:D3N, S:F486V, S:L452R

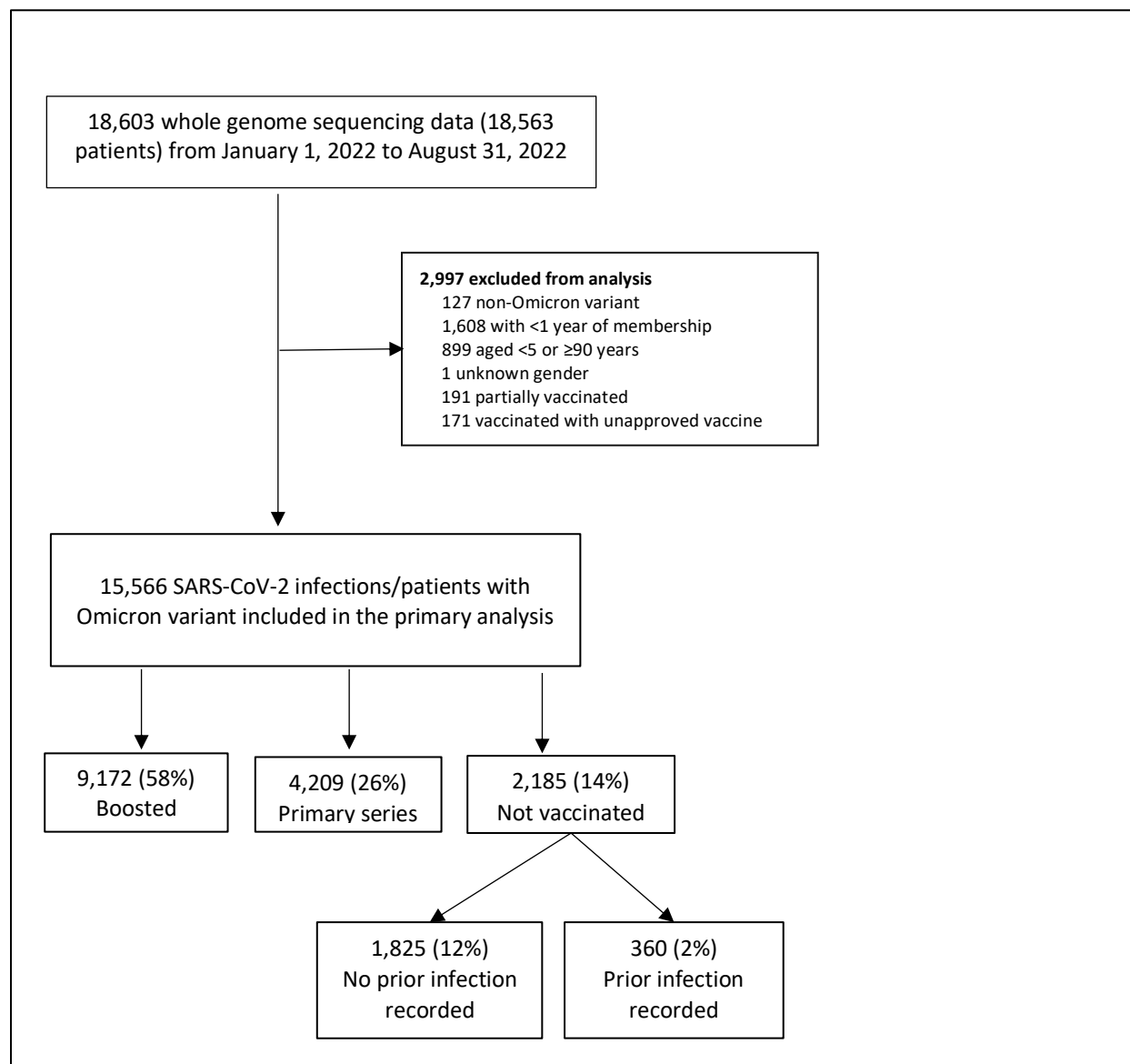
group 26: ORF1a:A2710T, ORF1a:L3674F, S:G21986-GTGTATT, S:G22193-ATT, S:G446S, S:L981F

group 45: S:A27S, S:C1254*

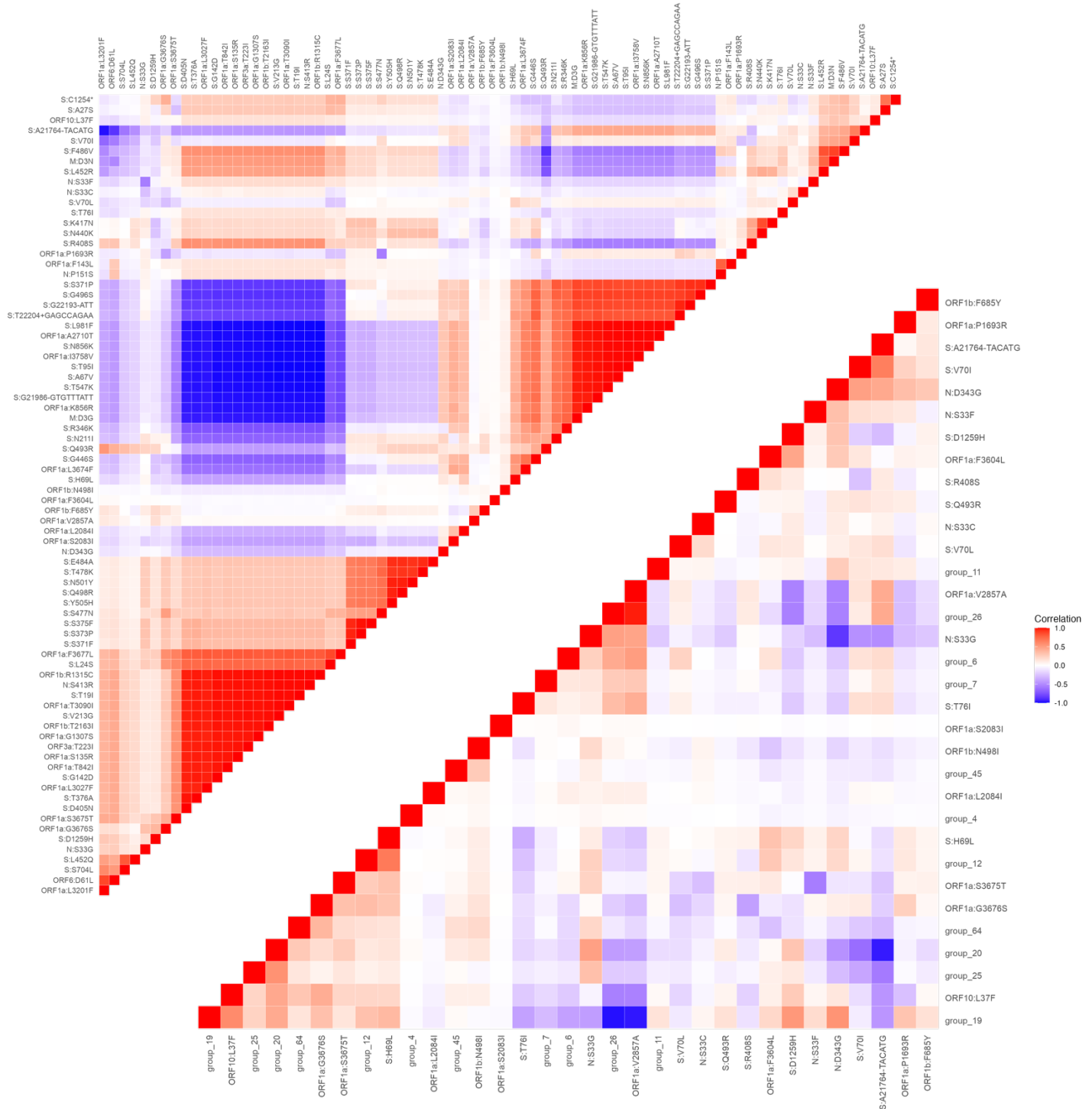
group 64: N:P151S, ORF1a:F143L

² Adjusted risk ratios derived from Poisson regression models to assess the association of immunity status with each mutation group adjusting for age, sex, race/ethnicity, Charlson comorbidity index score, and month of SARS-CoV-2 infection.

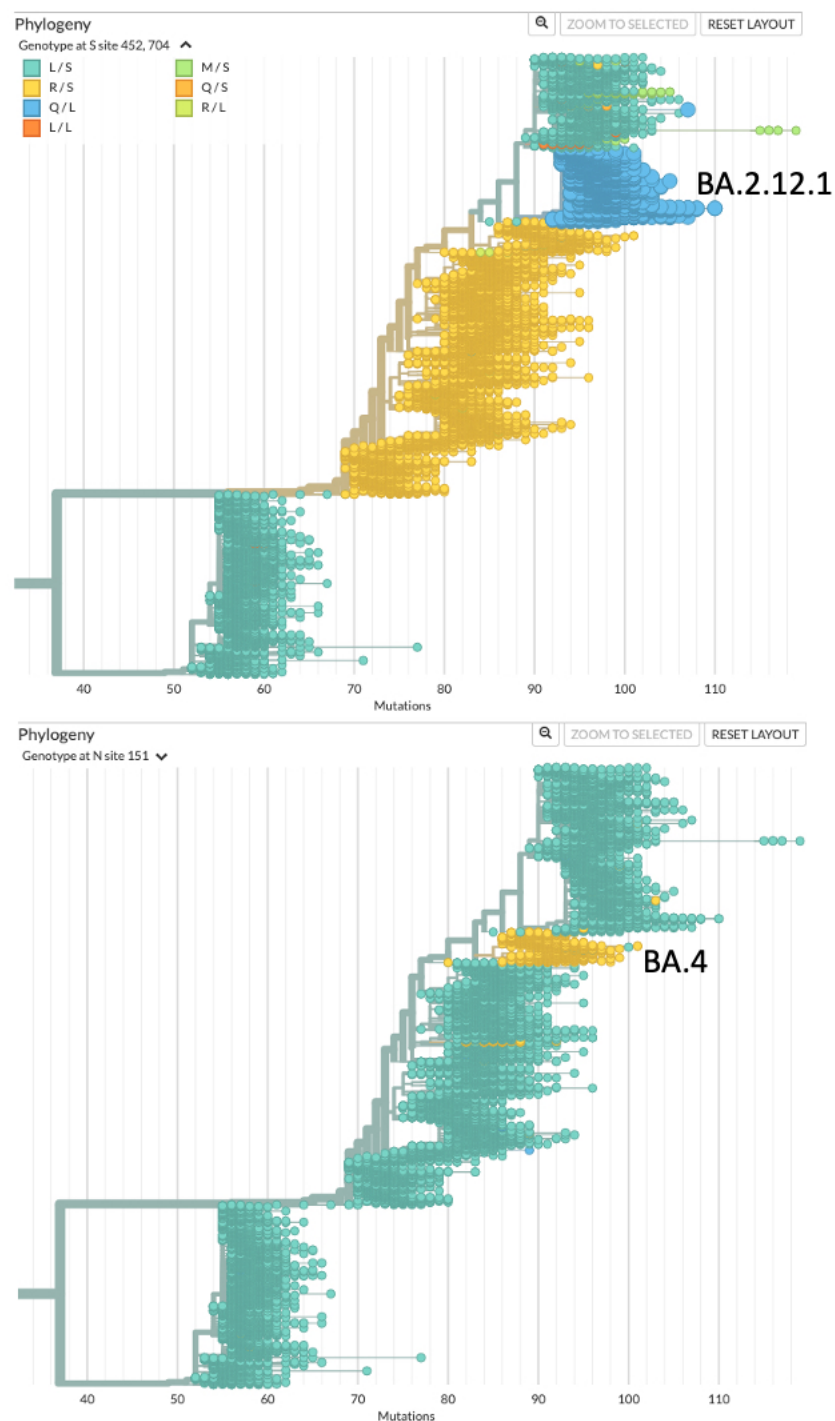
Supplemental Figure 1. Cohort flowchart



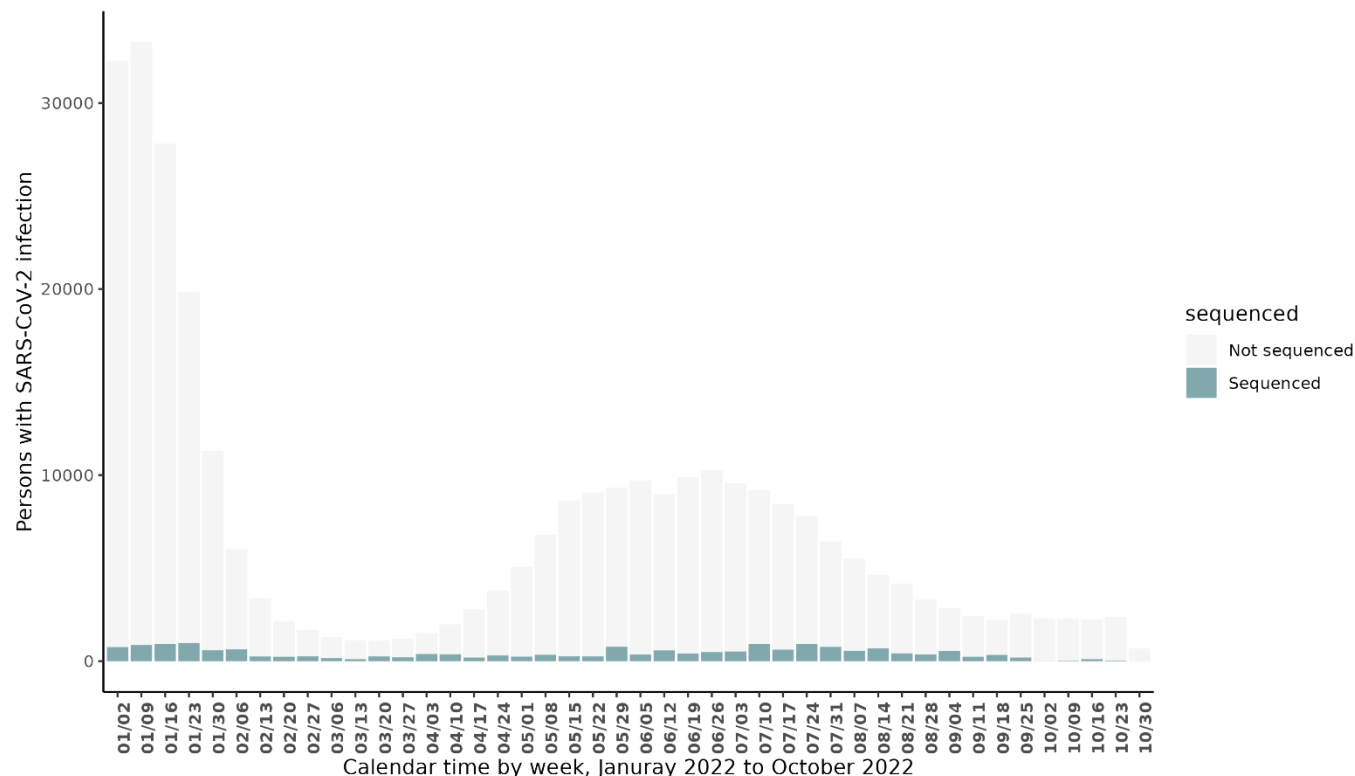
Supplemental Figure 2: Pearson correlations of SNP presence in specimens, before and after clustering.



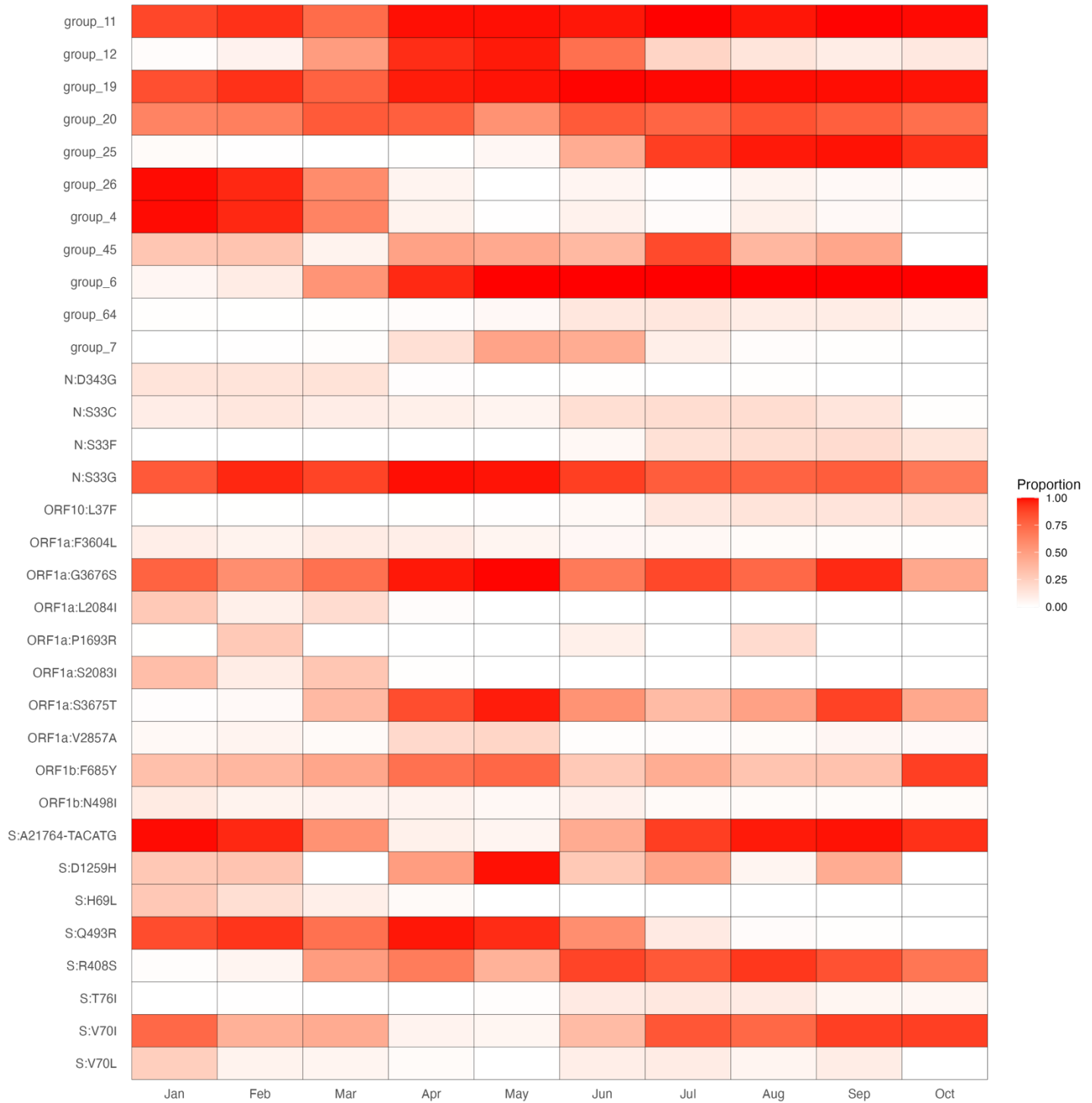
Supplemental Figure 3. Phylogenies of 15,045 Omicron genomes colored by group. Mutations in group 52 (top) represent a clade defined by the BA.2.12.1 subvariant (in blue). Mutations in group 4 (bottom) fall into a subclade of BA.4 (in yellow).



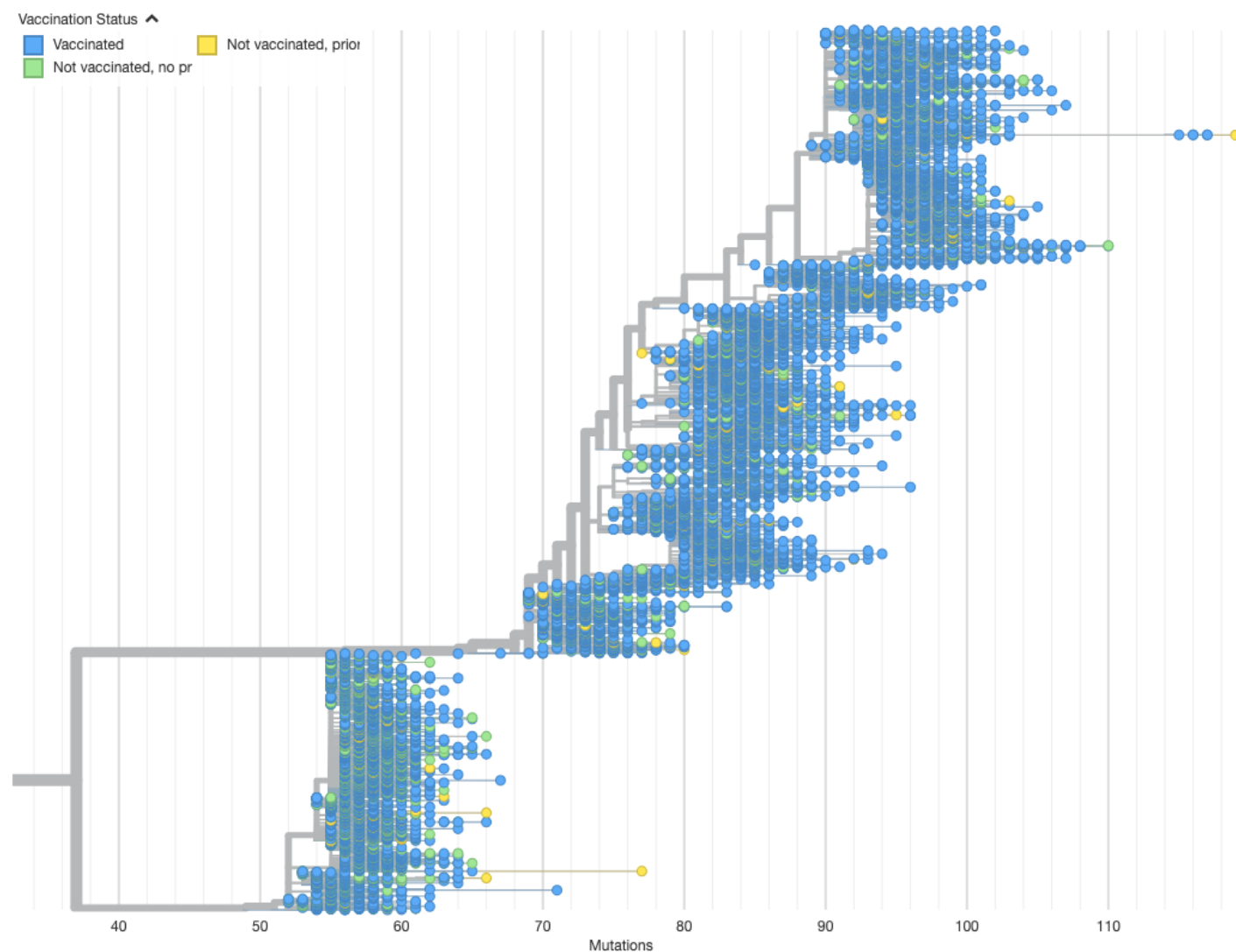
Supplemental Figure 4. Persons with SARS-CoV-2 infection tested by Kaiser Permanente Northern California stratified by subvariant by week from January 1, 2022 to October 31, 2022



Supplemental Figure 5. Proportion of specimens with a given SNP or group, by month of 2022



Supplemental Figure 6. Phylogeny of 15,045 Omicron genomes sampled between 1/1/2022 and 9/25/22. Branch lengths represent divergence from Wuhan reference genome and nodes are colored by vaccination status. Vaccinated category includes primary series and primary plus booster.



Supplemental Table 1: Association between SNP presence and vaccine status (pooling over complete and boosted) among persons included in this analysis with SARS-CoV-2 infection with Omicron variant in Northern California, January 1, 2022 - October 31, 2022 (N=15,566).

	Not vaccinated, no recorded prior infection		Not vaccinated, recorded prior infection		Vaccinated, regardless of prior infection status	
SNP	N (%)	Adjusted RR (95% CI)	N (%)	Adjusted RR (95% CI)	N (%)	Adjusted RR (95% CI)
ORF1A:F143L	53 (2.9%)	Reference	24 (6.7%)	1.38 (0.85- 2.25)	705 (5.3%)	1.29 (0.97- 1.72)
N:P151S	65 (3.6%)	Reference	26 (7.2%)	1.19 (0.75- 1.87)	773 (5.8%)	1.18 (0.91- 1.54)
ORF10:L37F	63 (3.5%)	Reference	22 (6.1%)	1.00 (0.61- 1.62)	820 (6.1%)	1.18 (0.91- 1.53)
ORF1B:N498I	113 (6.2%)	Reference	16 (4.4%)	0.96 (0.57- 1.63)	729 (5.4%)	1.13 (0.92- 1.40)
S:N211I	499 (27.3%)	Reference	57 (15.8%)	1.11 (0.84- 1.46)	2,132 (15.9%)	1.12 (1.01- 1.24)
S:C1254*	615 (33.7%)	Reference	164 (45.6%)	1.15 (0.97- 1.37)	5,506 (41.1%)	1.11 (1.02- 1.21)
ORF1A:V2857A	82 (4.5%)	Reference	19 (5.3%)	1.38 (0.84- 2.29)	744 (5.6%)	1.09 (0.86- 1.39)
S:D1259H	561 (30.7%)	Reference	121 (33.6%)	1.06 (0.87- 1.29)	4,756 (35.5%)	1.08 (0.99- 1.19)
S:T76I	64 (3.5%)	Reference	28 (7.8%)	1.25 (0.80- 1.95)	755 (5.6%)	1.07 (0.83- 1.40)
N:S33C	209 (11.5%)	Reference	51 (14.2%)	1.07 (0.78- 1.45)	1,770 (13.2%)	1.06 (0.92- 1.24)
S:T22204+GAGCCAGAA	742 (40.7%)	Reference	84 (23.3%)	1.07 (0.85- 1.34)	3,088 (23.1%)	1.05 (0.97- 1.15)
S:G22193-ATT	780 (42.7%)	Reference	88 (24.4%)	1.07 (0.86- 1.34)	3,237 (24.2%)	1.05 (0.97- 1.14)
S:S371P	785 (43.0%)	Reference	91 (25.3%)	1.10 (0.88- 1.36)	3,254 (24.3%)	1.05 (0.96- 1.13)
S:F486V	505 (27.7%)	Reference	186 (51.7%)	1.04 (0.88- 1.23)	5,949 (44.5%)	1.04 (0.95- 1.14)
S:V70L	235 (12.9%)	Reference	33 (9.2%)	0.89 (0.62- 1.29)	1,325 (9.9%)	1.04 (0.90- 1.20)
ORF1A:F3604L	112 (6.1%)	Reference	17 (4.7%)	0.98 (0.59- 1.64)	709 (5.3%)	1.04 (0.84- 1.28)
N:S33G	1,513 (82.9%)	Reference	286 (79.4%)	0.97 (0.85- 1.10)	11,437 (85.5%)	1.03 (0.98- 1.09)
S:S371F	1,626 (89.1%)	Reference	341 (94.7%)	1.02 (0.91- 1.15)	12,659 (94.6%)	1.03 (0.98- 1.09)
S:S373P	1,628 (89.2%)	Reference	342 (95.0%)	1.02 (0.91- 1.15)	12,717 (95.0%)	1.03 (0.98- 1.09)
S:S375F	1,629 (89.3%)	Reference	343 (95.3%)	1.03 (0.91- 1.15)	12,727 (95.1%)	1.03 (0.98- 1.09)
S:K417N	1,180 (64.7%)	Reference	254 (70.6%)	1.03 (0.90- 1.18)	9,298 (69.5%)	1.03 (0.97- 1.10)
ORF1A:G3676S	1,376 (75.4%)	Reference	285 (79.2%)	1.02 (0.90- 1.16)	10,715 (80.1%)	1.03 (0.97- 1.09)
S:E484A	1,640 (89.9%)	Reference	333 (92.5%)	0.99 (0.88- 1.12)	12,662 (94.6%)	1.03 (0.97- 1.08)
S:Y505H	1,643 (90.0%)	Reference	334 (92.8%)	0.99 (0.88- 1.12)	12,699 (94.9%)	1.03 (0.97- 1.08)
ORF1A:L3027F	844 (46.2%)	Reference	253 (70.3%)	1.00 (0.87- 1.16)	9,562 (71.5%)	1.03 (0.96- 1.11)
ORF1B:F685Y	693 (38.0%)	Reference	137 (38.1%)	1.01 (0.84- 1.21)	5,570 (41.6%)	1.03 (0.95- 1.12)
ORF1A:F3677L	767 (42.0%)	Reference	223 (61.9%)	0.98 (0.84- 1.14)	8,689 (64.9%)	1.03 (0.95- 1.11)
S:G496S	802 (43.9%)	Reference	82 (22.8%)	0.99 (0.79- 1.25)	3,188 (23.8%)	1.03 (0.95- 1.11)
S:H69L	238 (13.0%)	Reference	15 (4.2%)	0.61 (0.36- 1.03)	952 (7.1%)	1.03 (0.88- 1.19)
S:S704L	141 (7.7%)	Reference	38 (10.6%)	0.96 (0.67- 1.38)	1,630 (12.2%)	1.03 (0.87- 1.24)
S:N501Y	1,652 (90.5%)	Reference	334 (92.8%)	0.99 (0.88- 1.11)	12,716 (95.0%)	1.02 (0.97- 1.08)

	Not vaccinated, no recorded prior infection		Not vaccinated, recorded prior infection		Vaccinated, regardless of prior infection status	
SNP	N (%)	Adjusted RR (95% CI)	N (%)	Adjusted RR (95% CI)	N (%)	Adjusted RR (95% CI)
S:Q498R	1,652 (90.5%)	Reference	332 (92.2%)	0.98 (0.87- 1.11)	12,695 (94.9%)	1.02 (0.97- 1.08)
S:S477N	1,612 (88.3%)	Reference	326 (90.6%)	0.98 (0.87- 1.10)	12,528 (93.6%)	1.02 (0.97- 1.08)
S:T478K	1,643 (90.0%)	Reference	331 (91.9%)	0.98 (0.87- 1.11)	12,651 (94.5%)	1.02 (0.97- 1.08)
ORF1A:S135R	862 (47.2%)	Reference	254 (70.6%)	0.98 (0.86- 1.13)	9,651 (72.1%)	1.02 (0.95- 1.10)
ORF1A:T842I	862 (47.2%)	Reference	255 (70.8%)	0.99 (0.86- 1.14)	9,632 (72.0%)	1.02 (0.95- 1.10)
ORF1B:R1315C	868 (47.6%)	Reference	259 (71.9%)	1.00 (0.87- 1.15)	9,701 (72.5%)	1.02 (0.95- 1.10)
S:T376A	860 (47.1%)	Reference	257 (71.4%)	1.00 (0.87- 1.15)	9,602 (71.8%)	1.02 (0.95- 1.10)
S:V213G	863 (47.3%)	Reference	257 (71.4%)	0.99 (0.86- 1.14)	9,670 (72.3%)	1.02 (0.95- 1.10)
N:S413R	866 (47.5%)	Reference	257 (71.4%)	0.99 (0.86- 1.14)	9,672 (72.3%)	1.02 (0.95- 1.09)
ORF1A:T3090I	866 (47.5%)	Reference	258 (71.7%)	1.00 (0.87- 1.14)	9,686 (72.4%)	1.02 (0.95- 1.09)
ORF1B:T2163I	865 (47.4%)	Reference	258 (71.7%)	1.00 (0.87- 1.14)	9,654 (72.1%)	1.02 (0.95- 1.09)
ORF3A:T223I	865 (47.4%)	Reference	257 (71.4%)	1.00 (0.87- 1.15)	9,668 (72.3%)	1.02 (0.95- 1.09)
S:D405N	858 (47.0%)	Reference	256 (71.1%)	1.00 (0.87- 1.15)	9,556 (71.4%)	1.02 (0.95- 1.09)
S:Q493R	1,141 (62.5%)	Reference	145 (40.3%)	0.92 (0.78- 1.10)	6,743 (50.4%)	1.02 (0.95- 1.09)
S:T19I	864 (47.3%)	Reference	257 (71.4%)	0.99 (0.86- 1.14)	9,667 (72.2%)	1.02 (0.95- 1.09)
ORF6:D61L	424 (23.2%)	Reference	91 (25.3%)	0.91 (0.72- 1.14)	4,440 (33.2%)	1.02 (0.92- 1.13)
S:V70I	1,078 (59.1%)	Reference	229 (63.6%)	1.03 (0.89- 1.19)	7,707 (57.6%)	1.01 (0.95- 1.08)
S:A21764-TACATG	1,477 (80.9%)	Reference	295 (81.9%)	1.04 (0.92- 1.18)	9,809 (73.3%)	1.01 (0.95- 1.07)
ORF1A:G1307S	866 (47.5%)	Reference	258 (71.7%)	1.00 (0.87- 1.14)	9,633 (72.0%)	1.01 (0.94- 1.09)
S:G142D	889 (48.7%)	Reference	257 (71.4%)	0.98 (0.85- 1.13)	9,732 (72.7%)	1.01 (0.94- 1.08)
S:L24S	710 (38.9%)	Reference	213 (59.2%)	1.01 (0.87- 1.18)	7,962 (59.5%)	1.01 (0.93- 1.09)
S:R408S	705 (38.6%)	Reference	217 (60.3%)	1.01 (0.87- 1.18)	7,727 (57.7%)	1.01 (0.93- 1.09)
M:D3N	453 (24.8%)	Reference	163 (45.3%)	1.01 (0.85- 1.21)	5,285 (39.5%)	1.01 (0.92- 1.12)
S:L452R	433 (23.7%)	Reference	153 (42.5%)	1.02 (0.85- 1.23)	4,878 (36.5%)	1.01 (0.92- 1.12)
ORF1A:S3675T	527 (28.9%)	Reference	143 (39.7%)	0.96 (0.80- 1.16)	5,891 (44.0%)	1.01 (0.92- 1.10)
S:A27S	207 (11.3%)	Reference	73 (20.3%)	1.07 (0.82- 1.39)	2,309 (17.3%)	1.01 (0.87- 1.17)
S:L452Q	99 (5.4%)	Reference	28 (7.8%)	0.96 (0.63- 1.47)	1,116 (8.3%)	1.01 (0.82- 1.25)
ORF1A:L3201F	365 (20.0%)	Reference	72 (20.0%)	0.88 (0.69- 1.14)	3,801 (28.4%)	1.00 (0.89- 1.11)
S:N440K	1,221 (66.9%)	Reference	251 (69.7%)	0.97 (0.85- 1.11)	9,418 (70.4%)	0.99 (0.93- 1.05)
ORF1A:K856R	954 (52.3%)	Reference	103 (28.6%)	1.05 (0.86- 1.29)	3,747 (28.0%)	0.99 (0.92- 1.07)
S:A67V	974 (53.4%)	Reference	109 (30.3%)	1.08 (0.88- 1.31)	3,848 (28.8%)	0.99 (0.92- 1.06)
S:L981F	972 (53.3%)	Reference	107 (29.7%)	1.06 (0.87- 1.29)	3,830 (28.6%)	0.99 (0.92- 1.06)
S:T547K	966 (52.9%)	Reference	105 (29.2%)	1.06 (0.86- 1.29)	3,784 (28.3%)	0.99 (0.92- 1.06)
S:G446S	567 (31.1%)	Reference	53 (14.7%)	0.91 (0.68- 1.20)	2,220 (16.6%)	0.99 (0.90- 1.09)

	Not vaccinated, no recorded prior infection		Not vaccinated, recorded prior infection		Vaccinated, regardless of prior infection status	
SNP	N (%)	Adjusted RR (95% CI)	N (%)	Adjusted RR (95% CI)	N (%)	Adjusted RR (95% CI)
ORF1A:P1693R	119 (6.5%)	Reference	25 (6.9%)	1.06 (0.68- 1.63)	790 (5.9%)	0.99 (0.81- 1.21)
ORF1A:A2710T	978 (53.6%)	Reference	110 (30.6%)	1.08 (0.88- 1.31)	3,878 (29.0%)	0.98 (0.91- 1.06)
ORF1A:I3758V	974 (53.4%)	Reference	109 (30.3%)	1.08 (0.88- 1.31)	3,834 (28.7%)	0.98 (0.91- 1.06)
S:G21986-GTGTTTATT	967 (53.0%)	Reference	106 (29.4%)	1.06 (0.87- 1.30)	3,767 (28.2%)	0.98 (0.91- 1.06)
S:T95I	974 (53.4%)	Reference	108 (30.0%)	1.07 (0.88- 1.30)	3,839 (28.7%)	0.98 (0.91- 1.06)
M:D3G	965 (52.9%)	Reference	106 (29.4%)	1.05 (0.86- 1.29)	3,779 (28.2%)	0.98 (0.91- 1.05)
S:R346K	660 (36.2%)	Reference	89 (24.7%)	1.28 (1.03- 1.60)	2,600 (19.4%)	0.98 (0.90- 1.07)
S:N856K	986 (54.0%)	Reference	109 (30.3%)	1.06 (0.87- 1.29)	3,860 (28.8%)	0.97 (0.90- 1.05)
ORF1A:L3674F	521 (28.5%)	Reference	56 (15.6%)	1.05 (0.80- 1.38)	2,013 (15.0%)	0.95 (0.86- 1.05)
N:D343G	160 (8.8%)	Reference	10 (2.8%)	0.59 (0.31- 1.11)	624 (4.7%)	0.90 (0.75- 1.08)
N:S33F	93 (5.1%)	Reference	44 (12.2%)	1.34 (0.94- 1.92)	974 (7.3%)	0.89 (0.71- 1.11)
ORF1A:S2083I	308 (16.9%)	Reference	33 (9.2%)	1.06 (0.74- 1.53)	1,082 (8.1%)	0.87 (0.76- 1.00)
ORF1A:L2084I	255 (14.0%)	Reference	22 (6.1%)	0.90 (0.58- 1.39)	865 (6.5%)	0.84 (0.72- 0.97)

¹ Poisson regression models to assess the association of breakthrough infection with each individual mutation adjusting for age, sex, race/ethnicity, Charlson comorbidity index score, and month of SARS-CoV-2 infection. None of the p-values were significant after adjustment for multiple testing.

Supplemental Table 2: Association between SNP presence and vaccine status (separating out complete and boosted) among persons included in this analysis with SARS-CoV-2 infection with Omicron variant in Northern California, January 1, 2022 – October 31, 2022 (N=15,566)

SNP	Not vaccinated, no prior recorded infection	Not vaccinated, prior recorded infection (95% CI)	Primary vaccine series complete (95% CI)	Boosted (95% CI)
ORF1a:F143L	Reference	1.38 (0.85- 2.24)	1.43 (1.05- 1.95)	1.22 (0.91- 1.63)
ORF10:L37F	Reference	1.00 (0.61- 1.62)	1.17 (0.88- 1.56)	1.18 (0.90- 1.55)
S:N211I	Reference	1.11 (0.84- 1.46)	1.10 (0.98- 1.22)	1.16 (1.04- 1.31)
ORF1b:N498I	Reference	0.96 (0.57- 1.63)	1.12 (0.90- 1.40)	1.15 (0.91- 1.44)
N:P151S	Reference	1.18 (0.75- 1.87)	1.24 (0.94- 1.65)	1.15 (0.88- 1.50)
ORF1a:F3604L	Reference	0.99 (0.59- 1.65)	0.97 (0.77- 1.22)	1.12 (0.89- 1.41)
S:C1254*	Reference	1.15 (0.97- 1.37)	1.10 (1.01- 1.21)	1.11 (1.01- 1.22)
S:D1259H	Reference	1.06 (0.87- 1.30)	1.06 (0.96- 1.17)	1.11 (1.00- 1.22)
S:T76I	Reference	1.25 (0.80- 1.96)	1.05 (0.79- 1.41)	1.09 (0.83- 1.42)
N:S33C	Reference	1.07 (0.79- 1.45)	1.06 (0.90- 1.25)	1.07 (0.91- 1.25)
S:S371P	Reference	1.10 (0.88- 1.36)	1.04 (0.95- 1.13)	1.06 (0.96- 1.16)
ORF1b:F685Y	Reference	1.01 (0.84- 1.22)	1.02 (0.93- 1.11)	1.05 (0.96- 1.15)
S:T22204+GAGCCAGAA	Reference	1.07 (0.85- 1.34)	1.06 (0.97- 1.16)	1.05 (0.95- 1.15)
N:S33G	Reference	0.97 (0.86- 1.10)	1.03 (0.97- 1.09)	1.04 (0.98- 1.11)
S:S371F	Reference	1.02 (0.91- 1.15)	1.03 (0.97- 1.09)	1.04 (0.98- 1.10)
S:S373P	Reference	1.02 (0.91- 1.15)	1.03 (0.97- 1.09)	1.04 (0.98- 1.10)
S:S375F	Reference	1.03 (0.91- 1.15)	1.03 (0.97- 1.09)	1.04 (0.98- 1.10)
ORF1a:L3027F	Reference	1.00 (0.87- 1.16)	1.02 (0.95- 1.11)	1.04 (0.96- 1.12)
S:F486V	Reference	1.04 (0.88- 1.23)	1.03 (0.93- 1.14)	1.04 (0.95- 1.15)
S:G22193-ATT	Reference	1.07 (0.86- 1.34)	1.06 (0.97- 1.16)	1.04 (0.95- 1.15)
S:E484A	Reference	0.99 (0.88- 1.12)	1.02 (0.97- 1.08)	1.03 (0.97- 1.09)
S:N501Y	Reference	0.99 (0.88- 1.11)	1.02 (0.97- 1.08)	1.03 (0.97- 1.09)
S:S477N	Reference	0.98 (0.87- 1.10)	1.02 (0.97- 1.09)	1.03 (0.97- 1.09)
S:T478K	Reference	0.98 (0.87- 1.11)	1.02 (0.96- 1.08)	1.03 (0.97- 1.09)
S:Y505H	Reference	0.99 (0.88- 1.12)	1.03 (0.97- 1.09)	1.03 (0.97- 1.09)
S:K417N	Reference	1.03 (0.90- 1.18)	1.03 (0.96- 1.10)	1.03 (0.96- 1.10)
ORF1a:G3676S	Reference	1.02 (0.90- 1.16)	1.02 (0.96- 1.09)	1.03 (0.96- 1.09)
ORF1a:F3677L	Reference	0.98 (0.84- 1.14)	1.03 (0.95- 1.12)	1.03 (0.95- 1.11)
ORF1a:V2857A	Reference	1.37 (0.83- 2.27)	1.16 (0.90- 1.50)	1.03 (0.80- 1.33)
S:Q498R	Reference	0.98 (0.87- 1.11)	1.02 (0.96- 1.08)	1.02 (0.97- 1.08)
N:S413R	Reference	0.99 (0.86- 1.14)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
ORF1a:G1307S	Reference	1.00 (0.87- 1.14)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
ORF1a:S135R	Reference	0.98 (0.86- 1.13)	1.01 (0.94- 1.10)	1.02 (0.95- 1.10)

SNP	Not vaccinated, no prior recorded infection	Not vaccinated, prior recorded infection (95% CI)	Primary vaccine series complete (95% CI)	Boosted (95% CI)
ORF1a:T3090I	Reference	1.00 (0.87- 1.14)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
ORF1a:T842I	Reference	0.99 (0.86- 1.14)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
ORF1b:R1315C	Reference	1.00 (0.87- 1.15)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
ORF1b:T2163I	Reference	1.00 (0.87- 1.15)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
ORF3a:T223I	Reference	1.00 (0.87- 1.15)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
S:D405N	Reference	1.00 (0.87- 1.15)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
S:T19I	Reference	0.99 (0.86- 1.14)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
S:T376A	Reference	1.00 (0.87- 1.15)	1.01 (0.93- 1.09)	1.02 (0.95- 1.10)
S:V213G	Reference	0.99 (0.86- 1.14)	1.01 (0.94- 1.10)	1.02 (0.95- 1.10)
S:Q493R	Reference	0.92 (0.78- 1.10)	1.02 (0.95- 1.09)	1.02 (0.95- 1.09)
M:D3N	Reference	1.01 (0.85- 1.21)	0.99 (0.89- 1.11)	1.02 (0.93- 1.13)
S:G496S	Reference	0.99 (0.79- 1.25)	1.03 (0.95- 1.12)	1.02 (0.93- 1.12)
orf6:d61I	Reference	0.91 (0.72- 1.14)	1.01 (0.90- 1.13)	1.02 (0.92- 1.14)
S:G142D	Reference	0.98 (0.85- 1.13)	1.00 (0.92- 1.08)	1.01 (0.94- 1.09)
S:V70I	Reference	1.03 (0.89- 1.19)	1.02 (0.95- 1.10)	1.01 (0.94- 1.08)
S:L24S	Reference	1.01 (0.87- 1.18)	1.01 (0.93- 1.10)	1.01 (0.93- 1.10)
S:R408S	Reference	1.01 (0.87- 1.18)	1.00 (0.91- 1.09)	1.01 (0.93- 1.10)
ORF1a:S3675T	Reference	0.97 (0.80- 1.16)	1.00 (0.90- 1.11)	1.01 (0.92- 1.11)
S:L452R	Reference	1.02 (0.85- 1.23)	1.01 (0.91- 1.13)	1.01 (0.91- 1.13)
S:V70L	Reference	0.89 (0.62- 1.28)	1.06 (0.91- 1.23)	1.01 (0.86- 1.19)
S:H69L	Reference	0.61 (0.36- 1.03)	1.03 (0.88- 1.21)	1.01 (0.85- 1.20)
S:S704L	Reference	0.96 (0.67- 1.37)	1.09 (0.90- 1.32)	1.01 (0.84- 1.21)
S:A21764-TACATG	Reference	1.04 (0.92- 1.18)	1.01 (0.95- 1.07)	1.00 (0.94- 1.07)
ORF1a:L3201F	Reference	0.88 (0.69- 1.14)	0.99 (0.88- 1.12)	1.00 (0.89- 1.12)
S:A27S	Reference	1.07 (0.82- 1.39)	1.03 (0.88- 1.21)	1.00 (0.86- 1.16)
S:N440K	Reference	0.97 (0.84- 1.11)	0.99 (0.93- 1.06)	0.99 (0.92- 1.05)
S:L452Q	Reference	0.96 (0.63- 1.46)	1.05 (0.84- 1.33)	0.99 (0.79- 1.23)
ORF1a:K856R	Reference	1.05 (0.86- 1.29)	1.00 (0.92- 1.08)	0.98 (0.90- 1.06)
S:L981F	Reference	1.06 (0.87- 1.29)	0.99 (0.92- 1.07)	0.98 (0.90- 1.06)
S:G446S	Reference	0.91 (0.68- 1.20)	1.00 (0.90- 1.11)	0.98 (0.88- 1.10)
S:A67V	Reference	1.08 (0.88- 1.31)	1.00 (0.92- 1.08)	0.97 (0.89- 1.06)
S:T547K	Reference	1.06 (0.86- 1.29)	0.99 (0.92- 1.07)	0.97 (0.89- 1.06)
ORF1a:A2710T	Reference	1.08 (0.88- 1.31)	0.99 (0.91- 1.07)	0.97 (0.89- 1.05)
ORF1a:I3758V	Reference	1.08 (0.88- 1.31)	0.99 (0.92- 1.07)	0.97 (0.89- 1.05)
S:G21986-GTGTATT	Reference	1.06 (0.87- 1.30)	0.99 (0.92- 1.07)	0.97 (0.89- 1.05)

SNP	Not vaccinated, no prior recorded infection	Not vaccinated, prior recorded infection (95% CI)	Primary vaccine series complete (95% CI)	Boosted (95% CI)
S:T95I	Reference	1.07 (0.88- 1.31)	0.99 (0.92- 1.07)	0.97 (0.89- 1.05)
S:R346K	Reference	1.28 (1.03- 1.60)	0.99 (0.90- 1.09)	0.97 (0.88- 1.08)
ORF1a:P1693R	Reference	1.05 (0.68- 1.63)	1.01 (0.81- 1.26)	0.97 (0.78- 1.21)
M:D3G	Reference	1.05 (0.86- 1.29)	0.99 (0.91- 1.07)	0.96 (0.88- 1.05)
S:N856K	Reference	1.06 (0.87- 1.29)	0.98 (0.91- 1.06)	0.96 (0.88- 1.04)
N:D343G	Reference	0.59 (0.31- 1.11)	0.87 (0.72- 1.06)	0.95 (0.77- 1.16)
ORF1a:L3674F	Reference	1.05 (0.80- 1.39)	0.96 (0.87- 1.07)	0.94 (0.83- 1.05)
n:s33f	Reference	1.34 (0.94- 1.92)	0.94 (0.73- 1.20)	0.86 (0.69- 1.08)
ORF1a:S2083I	Reference	1.07 (0.74- 1.53)	0.92 (0.80- 1.06)	0.80 (0.68- 0.93)
ORF1a:L2084I	Reference	0.90 (0.58- 1.40)	0.94 (0.81- 1.10)	0.69 (0.58- 0.82)

¹ Poisson regression models to assess the association of breakthrough infection with each individual mutation adjusting for age, sex, race/ethnicity, Charlson comorbidity index score, and month of SARS-CoV-2 infection. None of the p-values were significant after adjustment for multiple testing.

Supplemental Table 3: Association between mutation groups¹ and vaccine status (separating complete and boosted) among persons included in this analysis with SARS-CoV-2 infection with Omicron variant in Northern California, January 2022-October 2022 (N=15,566)²

	Not vaccinated, No prior infection	Not vaccinated, Recorded prior infection			Vaccinated with complete series, regardless of prior infection status			Vaccinated with more than complete series, regardless of prior infection status		
SNP or group		aRR (95% CI)	p (raw)	p (FDR)	aRR (95% CI)	p (raw)	p (FDR)	aRR (95% CI)	p (raw)	p (FDR)
ORF10:L37F	Reference	1.00 (0.61- 1.62)	0.9854	0.99996	1.17 (0.88- 1.56)	0.2897	0.98013	1.18 (0.90- 1.55)	0.2188	0.75760
group64	Reference	1.26 (0.82- 1.93)	0.2893	0.99996	1.31 (1.00- 1.71)	0.0460	0.63809	1.17 (0.91- 1.51)	0.2127	0.75760
ORF1b:N498I	Reference	0.96 (0.57- 1.63)	0.8930	0.99996	1.12 (0.90- 1.40)	0.3044	0.98013	1.15 (0.91- 1.44)	0.2418	0.75760
ORF1a:F3604L	Reference	0.99 (0.59- 1.65)	0.9670	0.99996	0.97 (0.77- 1.22)	0.8029	0.99285	1.12 (0.89- 1.41)	0.3409	0.84737
S:D1259H	Reference	1.06 (0.87- 1.30)	0.5325	0.99996	1.06 (0.96- 1.17)	0.2355	0.98013	1.11 (1.00- 1.22)	0.0437	0.38280
group45	Reference	1.13 (0.96- 1.34)	0.1487	0.99996	1.09 (1.00- 1.20)	0.0532	0.63809	1.10 (1.01- 1.20)	0.0299	0.34911
S:T76I	Reference	1.25 (0.80- 1.96)	0.3213	0.99996	1.05 (0.79- 1.41)	0.7269	0.99285	1.09 (0.83- 1.42)	0.5496	0.93632
N:S33C	Reference	1.07 (0.79- 1.45)	0.6744	0.99996	1.06 (0.90- 1.25)	0.4647	0.98013	1.07 (0.91- 1.25)	0.4358	0.84737
ORF1b:F685Y	Reference	1.01 (0.84- 1.22)	0.9039	0.99996	1.02 (0.93- 1.11)	0.7211	0.99285	1.05 (0.96- 1.15)	0.2874	0.77368
N:S33G	Reference	0.97 (0.86- 1.10)	0.6580	0.99996	1.03 (0.97- 1.09)	0.4080	0.98013	1.04 (0.98- 1.11)	0.1683	0.75760
group19	Reference	1.02 (0.91- 1.15)	0.6887	0.99996	1.03 (0.97- 1.09)	0.3374	0.98013	1.04 (0.98- 1.10)	0.2138	0.75760
group25	Reference	1.04 (0.88- 1.23)	0.6335	0.99996	1.03 (0.93- 1.14)	0.5399	0.98077	1.04 (0.95- 1.15)	0.3822	0.84737
ORF1a:G3676S	Reference	1.02 (0.90- 1.16)	0.7164	0.99996	1.02 (0.96- 1.09)	0.4624	0.98013	1.03 (0.96- 1.09)	0.4181	0.84737
group12	Reference	0.92 (0.74- 1.15)	0.4800	0.99996	1.01 (0.91- 1.13)	0.8274	0.99285	1.03 (0.92- 1.14)	0.6258	0.93632
ORF1a:V2857A	Reference	1.37 (0.83- 2.27)	0.2144	0.99996	1.16 (0.90- 1.50)	0.2416	0.98013	1.03 (0.80- 1.33)	0.8122	0.93632
group11	Reference	0.99 (0.88- 1.11)	0.8758	0.99996	1.02 (0.96- 1.08)	0.4901	0.98013	1.02 (0.97- 1.09)	0.4009	0.84737
S:Q493R	Reference	0.92 (0.78- 1.10)	0.3606	0.99996	1.02 (0.95- 1.09)	0.6369	0.99285	1.02 (0.95- 1.09)	0.6295	0.93632
group6	Reference	0.99 (0.86- 1.13)	0.8414	0.99996	1.00 (0.93- 1.08)	0.9600	0.99860	1.01 (0.94- 1.09)	0.6974	0.93632
S:V70I	Reference	1.03 (0.89- 1.19)	0.6699	0.99996	1.02 (0.95- 1.10)	0.5570	0.98077	1.01 (0.94- 1.08)	0.8514	0.93632
group20	Reference	1.00 (0.88- 1.15)	0.9439	0.99996	1.00 (0.94- 1.07)	0.9986	0.99860	1.01 (0.94- 1.07)	0.8779	0.93632
S:R408S	Reference	1.01 (0.87- 1.18)	0.8801	0.99996	1.00 (0.91- 1.09)	0.9742	0.99860	1.01 (0.93- 1.10)	0.7979	0.93632
ORF1a:S3675T	Reference	0.97 (0.80- 1.16)	0.7068	0.99996	1.00 (0.90- 1.11)	0.9848	0.99860	1.01 (0.92- 1.11)	0.8491	0.93632
S:V70L	Reference	0.89 (0.62- 1.28)	0.5264	0.99996	1.06 (0.91- 1.23)	0.4816	0.98013	1.01 (0.86- 1.19)	0.9096	0.93632
S:H69L	Reference	0.61 (0.36- 1.03)	0.0640	0.98819	1.03 (0.88- 1.21)	0.6712	0.99285	1.01 (0.85- 1.20)	0.9075	0.93632
group7	Reference	0.95 (0.66- 1.36)	0.7696	0.99996	1.09 (0.90- 1.32)	0.3774	0.98013	1.01 (0.84- 1.21)	0.9374	0.93741
S:A21764-TACATG	Reference	1.04 (0.92- 1.18)	0.5275	0.99996	1.01 (0.95- 1.07)	0.7878	0.99285	1.00 (0.94- 1.07)	0.8937	0.93632
group26	Reference	1.06 (0.88- 1.29)	0.5289	0.99996	0.99 (0.92- 1.07)	0.7873	0.99285	0.99 (0.91- 1.07)	0.7439	0.93632
group4	Reference	1.06 (0.88- 1.28)	0.5447	0.99996	0.98 (0.91- 1.06)	0.5721	0.98077	0.97 (0.89- 1.05)	0.4805	0.88512
ORF1a:P1693R	Reference	1.05 (0.68- 1.63)	0.8134	0.99996	1.01 (0.81- 1.26)	0.9443	0.99860	0.97 (0.78- 1.21)	0.7922	0.93632
N:D343G	Reference	0.59 (0.31- 1.11)	0.1022	0.98819	0.87 (0.72- 1.06)	0.1683	0.98013	0.95 (0.77- 1.16)	0.5944	0.93632
N:S33F	Reference	1.34 (0.94- 1.92)	0.1098	0.98819	0.94 (0.73- 1.20)	0.6012	0.98381	0.86 (0.69- 1.08)	0.2069	0.75760
ORF1a:S2083I	Reference	1.07 (0.74- 1.53)	0.7251	0.99996	0.92 (0.80- 1.06)	0.2454	0.98013	0.80 (0.68- 0.93)	0.0046	0.07981
ORF1a:L2084I	Reference	0.90 (0.58- 1.40)	0.6480	0.99996	0.94 (0.81- 1.10)	0.4431	0.98013	0.69 (0.58- 0.82)	<.0001	0.00095

¹ Mutations with Pearson pairwise correlation >0.70 combined into groups for analysis:

group 4: M:D3G, ORF1a:I3758V, ORF1a:K856R, S:A67V, S:G496S, S:N211I, S:N856K, S:R346K, S:S371P, S:T22204+GAGCCAGAA, S:T547K, S:T95I

group 6: N:S413R, ORF1a:F3677L, ORF1a:G1307S, ORF1a:L3027F, ORF1a:S135R, ORF1a:T3090I, ORF1a:T842I, ORF1b:R1315C, ORF1b:T2163I, ORF3a:T223I, S:D405N, S:G142D, S:L24S, S:T19I, S:T376A, S:V213G

group 7: S:L452Q, S:S704L

group 11: S:E484A, S:N501Y, S:Q498R, S:S477N, S:T478K, S:Y505H

group 12: ORF1a:L3201F, ORF6:D61L

group 19: S:S371F, S:S373P, S:S375F

group 20: S:K417N, S:N440K

group 25: M:D3N, S:F486V, S:L452R

group 26: ORF1a:A2710T, ORF1a:L3674F, S:G21986-GTGTATT, S:G22193-ATT, S:G446S, S:L981F

group 45: S:A27S, S:C1254*

group 64: N:P151S, ORF1a:F143L

² Adjusted risk ratios derived from Poisson regression models to assess the association of immunity status with each mutation group adjusting for age, sex, race/ethnicity, Charlson comorbidity index score, and month of SARS-CoV-2 infection.