Mutations in Persistent SARS-CoV-2 Cases

Lacy M. Simons[1], Ramon Lorenzo-Redondo[1], Estefany Guzman[1], Egon A. Ozer[1], Michael G. Ison[1], Chad J. Achenbach[1], Judd F. Hultquist[1]

[1] Department of Infectious Diseases, Northwestern University Feinberg School of Medicine, [2]Center for Pathogen Genomics and Microbial Evolution, Institute for Global Health, Northwestern University Feinberg School of Medicine

Northwestern Medicine® Feinberg School of Medicine

00179

Background

The evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants with decreased susceptibility to neutralizing antibodies are of clinical importance. While several demographic and clinical correlates of Coronavirus Disease 2019 (COVID-19) outcome have been identified, their relationship to virological and immunological parameters remain poorly defined. Here, we evaluate viral diversity and the accumulation of intra-host mutations over time in a population of hospitalized adults positive for SARS-CoV-2.

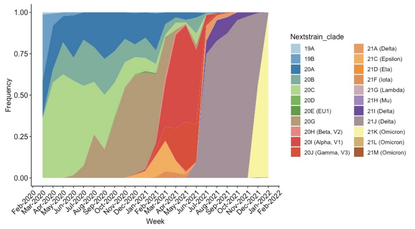


Figure 1 | The distribution of clades from February 2020 to February 2022 in Chicago. Monthly cumulative frequency of each clade using NextStrain nomenclature. Sequences obtained from patient samples at Northwestern Memorial Hospital; encompassing both inpatients and outpatients.

Methods

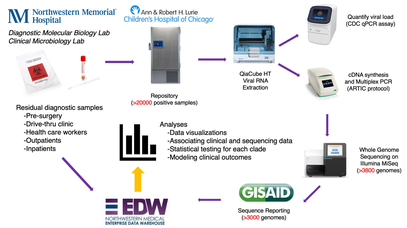


Figure 2 | Schematic of laboratory methods including sample collection and processing.

We developed a biobank of residual nasal swabs collected at Northwestern Memorial Hospital, encompassing inpatients and outpatients. Inpatients that longitudinal specimen in the biobank were assessed for viral load, genotype and intra-host mutations. To expand our study on intra-host variation, we performed longitudinal collection of nasopharyngeal swabs and blood samples from a small cohort of hospitalized adults with COVID-19. Clinical information regarding study subject's immunocompromised status was collected. Samples were assessed for SARS-CoV-2 viral load, viral genotype, viral diversity, and antibody titer. We then identified cases where mutations emerged that were consistent with variants of concerns and performed functional studies.

Inpatients with Longitudinal Samples

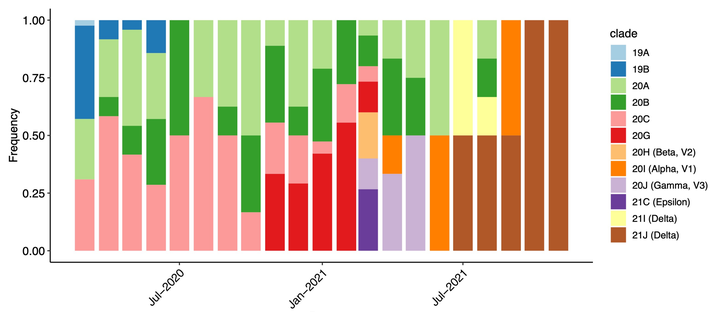


Figure 3 | Frequency of clades limited to inpatients samples at Northwestern Memorial Hospital. Clade distribution closely resembles trends in Chicago at the time.

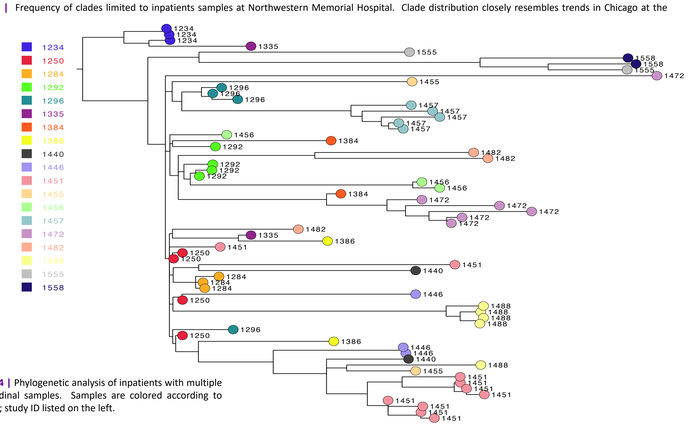


Figure 4 | Phylogenetic analysis of inpatients with multiple longitudinal samples. Samples are colored according to patient; study ID listed on the left.

Viral Evolution in Serial Samples

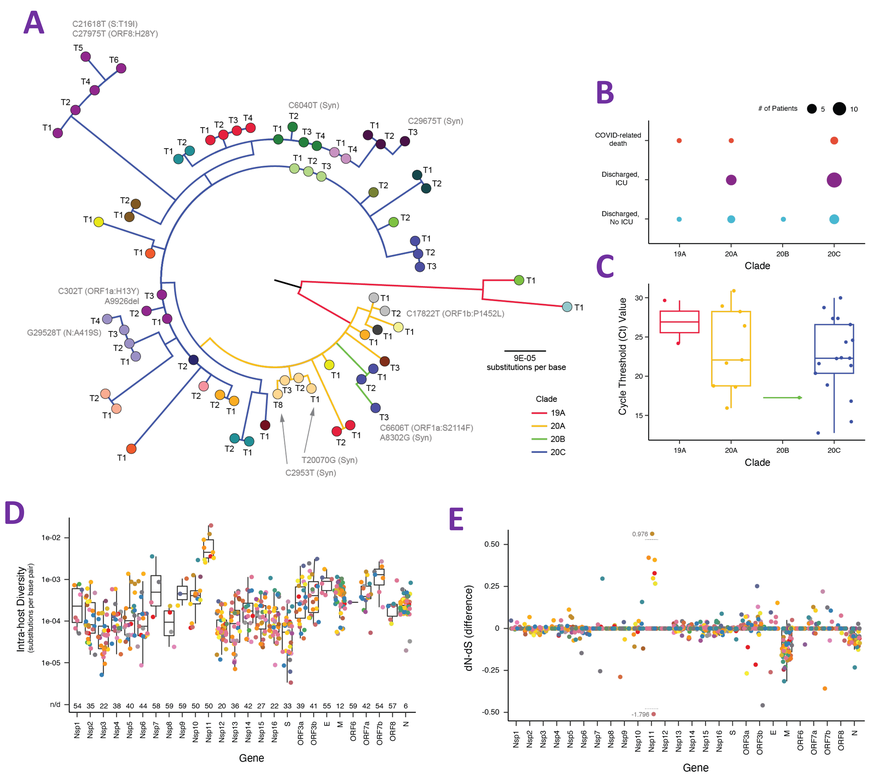


Figure 5 | Analysis of correlations between viral genotype, intra-host diversity, and participant outcome. A) Phylogenetic tree of SARS-CoV-2 whole genome sequences from each study participant and timepoint. Each branch is colored by corresponding viral clade and each branch tip is colored by study participant; tips are labeled chronologically from timepoint 1 (T1) onward ordered clockwise. Changes in consensus viral sequence observed are labeled by nucleotide and corresponding amino acid change, when applicable. B) Dot plot depicting the clade membership of the virus isolated from each study participant by outcome group. C) Box plot comparing the SARS-CoV-2 N1 cycle threshold (Ct) value at timepoint 1 for each participant separated by clade. Tests for significance were performed with the Wilcoxon rank sum test with Benjamini-Hochberg correction for multiple comparisons (all comparisons not significant, p-value >0.05). D) Box plot of the intra-host diversity observed within each indicated open reading frame in each participant specimen. Dots are colored by study participant. The number of specimens in which no intra-host diversity was observed is indicated on the bottom (n/d) E) Box plot of the dN-dS observed within each indicated open reading frame in each participant specimen. Dots are colored by study participant.

Viral Load

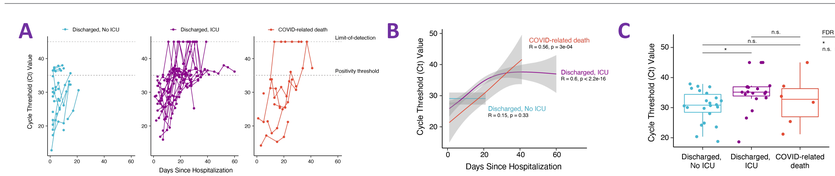


Figure 6 | Analysis of correlations between viral load and participant outcome. A) Plot of the quantitative PCR cycle threshold (Ct) values for SARS-CoV-2 N1 in nasopharyngeal swabs from each study participant relative to hospitalization time, separated by participant outcome. Specimens that did not amplify were assigned a value at the limit-of-detection. B) Generalized additive model fit to the average Ct value of hospitalized study participants over time separated by outcome. The correlation between average DI score and time during hospitalization is provided (R) alongside the p-value for each outcome group. C) Box plot of SARS-CoV-2 N1 Ct values from the final nasopharyngeal swab collected from each participant within 10 days of hospital discharge or death grouped by outcome (Wilcoxon rank sum test with Benjamini-Hochberg procedure to control FDR for multiple comparisons; n.s. = FDR >0.05, * = FDR <0.05).

Serology

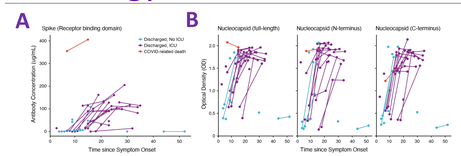


Figure 7 | Analysis of correlations between serology and participant outcome A) Line graph of the concentration of IgG antibody (micrograms per milliliter) recognizing the receptor binding domain (RBD) of Spike as determined by ELISA per study participant relative to the time since symptom onset. Lines are colored by outcome group. B) Line graph of the Optical Density (OD) value reflective of the concentration of antibody recognizing the full-length (left), N-terminus (center), or C-terminus (right) of the Nucleocapsid (N) protein as determined by ELISA per study participant. Values are presented relative to time since symptom onset and lines are colored by outcome group.

Immunosuppressed Patients

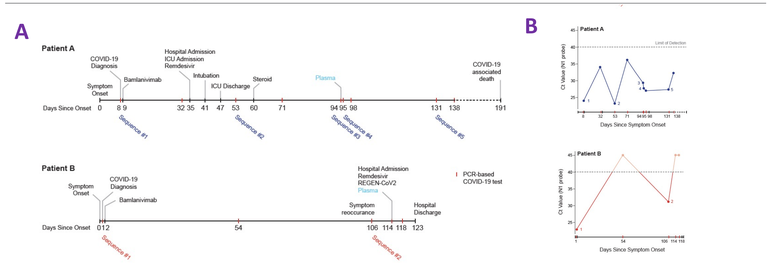


Figure 8 | Histories of Patients A and B with Persistent SARS-CoV-2 Infection. Panel A summarizes the clinical progression of two immunosuppressed patients starting at the time of COVID-19 symptom onset. Red tick marks represent dates of PCR-based COVID-19 diagnostic testing. The date of plasma collection from each patient is indicated in light blue. Panel B shows the cycle threshold (Ct) value of the diagnostic COVID-19 PCR-based tests administered during each patient's clinical course. Dates of collection for SARS-CoV-2 isolates that successfully underwent whole genome sequencing are annotated with numbers. Specimens with Ct values beyond the limit of detection of 40 were artificially represented by points at 45.

Emergence of Spike Mutations

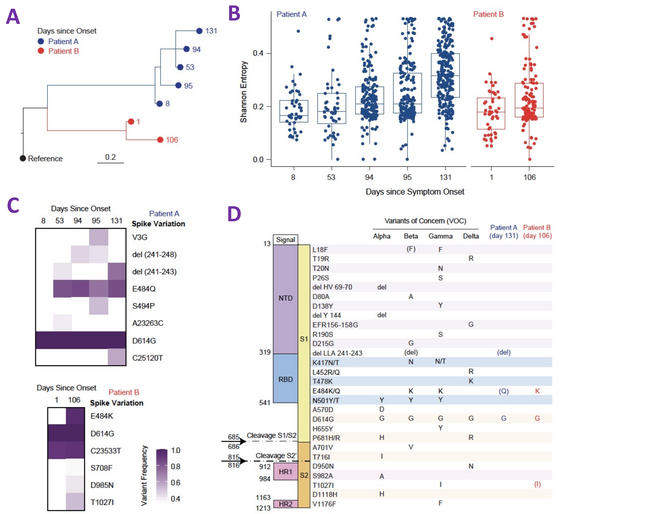


Figure 9 | Intra-host Emergence of Spike mutations in SARS-CoV-2 Isolates from Immunosuppressed Patients. Panel A shows a temporal phylogenetic tree comparing consensus SARS-CoV-2 whole genome sequences from Patient A (blue, 5 sequences) and Patient B (red, 2 sequences). The time since symptom onset of isolate collection is indicated to the right of each node. Panel B depicts one measure of viral diversity (Shannon entropy) calculated for each sequenced isolate labeled by the collection time since symptom onset. Each dot represents the Shannon entropy at a given nucleotide; the box-and-whisker plot reflects the median and interquartile range. Panel C shows a heat map of Spike variant frequency found in each viral isolate. Only variants that achieved at least 30% frequency in at least one sample are shown. Panel D summarizes the Spike mutations found in four variants of concern (Alpha, Beta, Delta, and Gamma) and in the SARS-CoV-2 isolates from each patient at the time of last sampling.

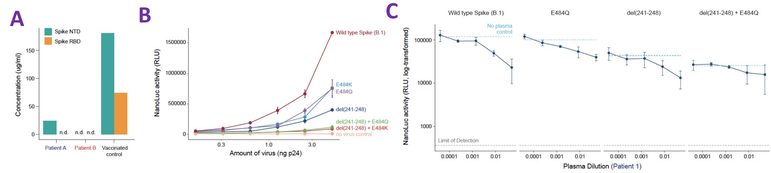


Figure 10 | Emergent Spike Mutations confer Decreased Viral Infectivity, but Enhanced Escape from Antibody Neutralization. Panel A depicts the levels of anti-Spike antibodies targeting the N-terminal domain (NTD) or receptor-binding domain (RBD) in plasma collected from each patient or a fully vaccinated control subject (n.d. = not detectable). Panel B depicts luciferase activity in cells after *in vitro* challenge with reporter viruses pseudotyped with SARS-CoV-2 Spike proteins containing the indicated mutations. Each point reflects the average +/- standard deviation of two independent technical replicates. Panel C depicts log-transformed nanoluc luminescence signal in HeLa-ACE2 cells after challenge with reporter viruses pseudotyped with SARS-CoV-2 Spike proteins containing the indicated mutations that were first incubated with dilutions of Patient A plasma. Each point reflects the average +/- standard deviation of two independent technical replicates. The limit of detection in the assay is depicted by a dashed gray line; the no plasma control in the assay is depicted by a dashed blue line..

Conclusions

Intra-host viral genetic diversity remained constant through disease course in study subjects that were non-immunocompromised and resulted in changes in viral genotype in some participants. We report the *de novo* emergence of Spike mutations that have been previously associated with circulating variants of concern in two immunosuppressed patients with persistent SARS-CoV-2 infection. Constant rates of viral evolution suggest the emergence of variants as a function of time, emphasizing the need for effective antivirals to control viral load over long disease courses.