

SUPPLEMENTARY INFORMATION

for

The Impact of Remdesivir on SARS-CoV-2 Evolution *in vivo*

Ted Ling-Hu^{1,2}, Lacy M. Simons^{1,2}, Estefany Rios-Guzman^{1,2}, Alexandre Machado de Sant'Anna
Carvalho^{1,2}, Maria Francesca Agnes^{1,2}, Arghavan Alisoltanidehkordi^{1,2}, Egon A. Ozer^{1,2}, Ramon
Lorenzo-Redondo^{1,2,+}, Judd F. Hultquist^{1,2,+}

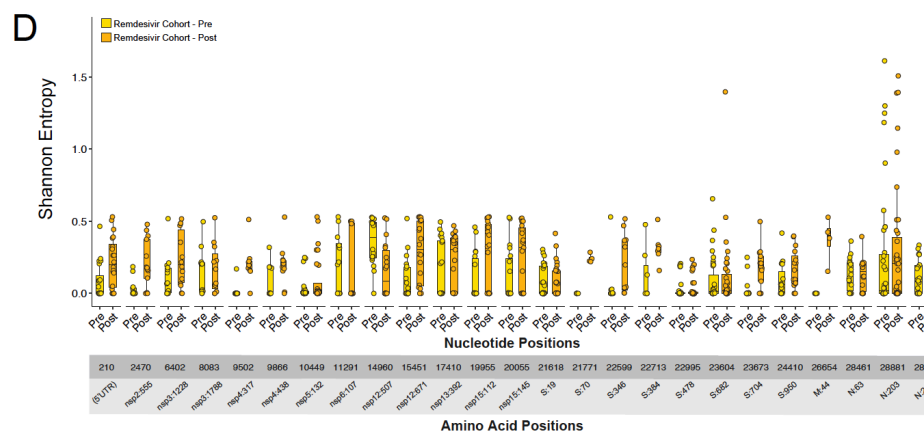
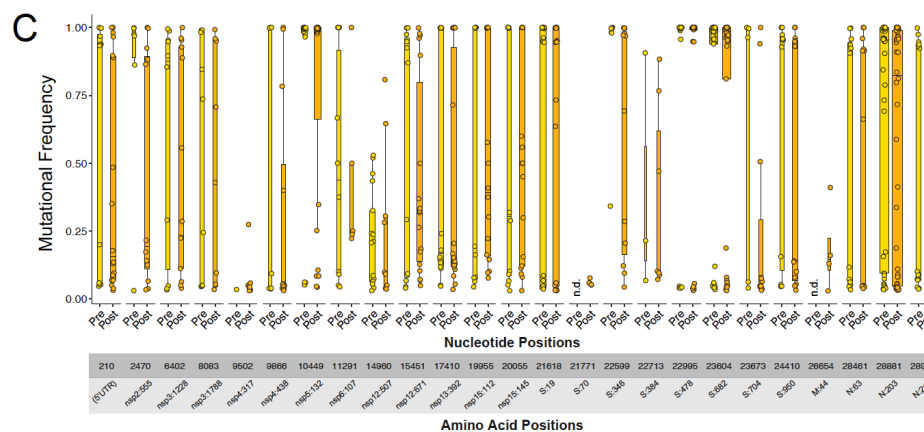
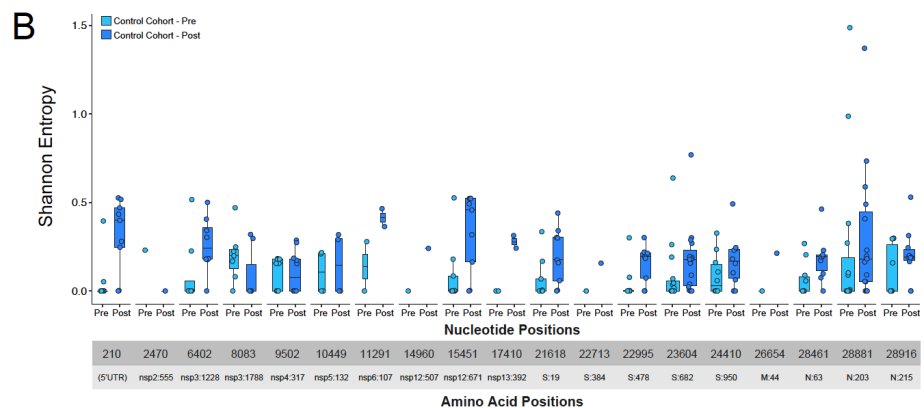
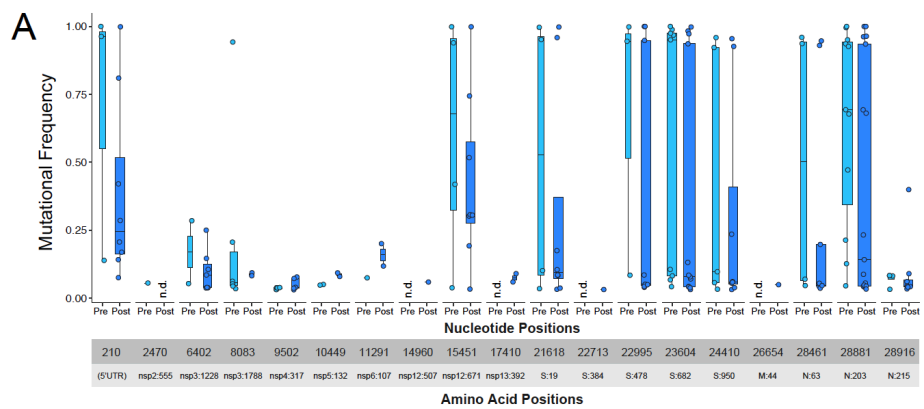
¹ Division of Infectious Diseases, Northwestern University Feinberg School of Medicine, Chicago,
IL 60611, USA.

² Center for Pathogen Genomics and Microbial Evolution, Northwestern University Havey Institute
for Global Health, Chicago, IL 60611, USA.

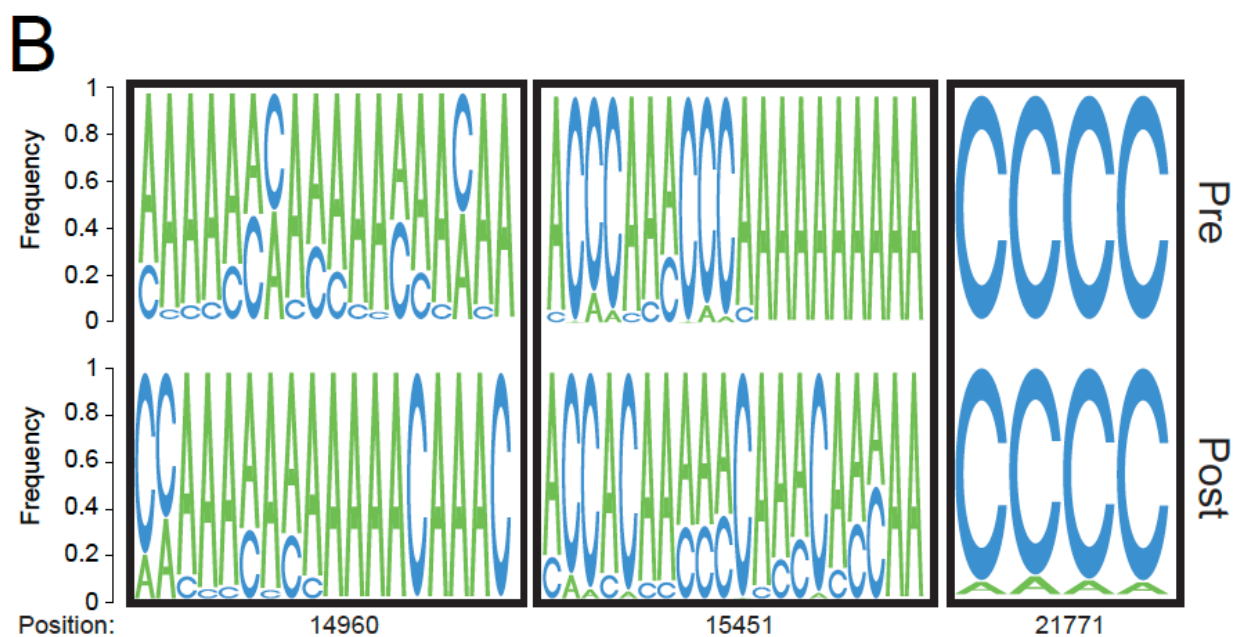
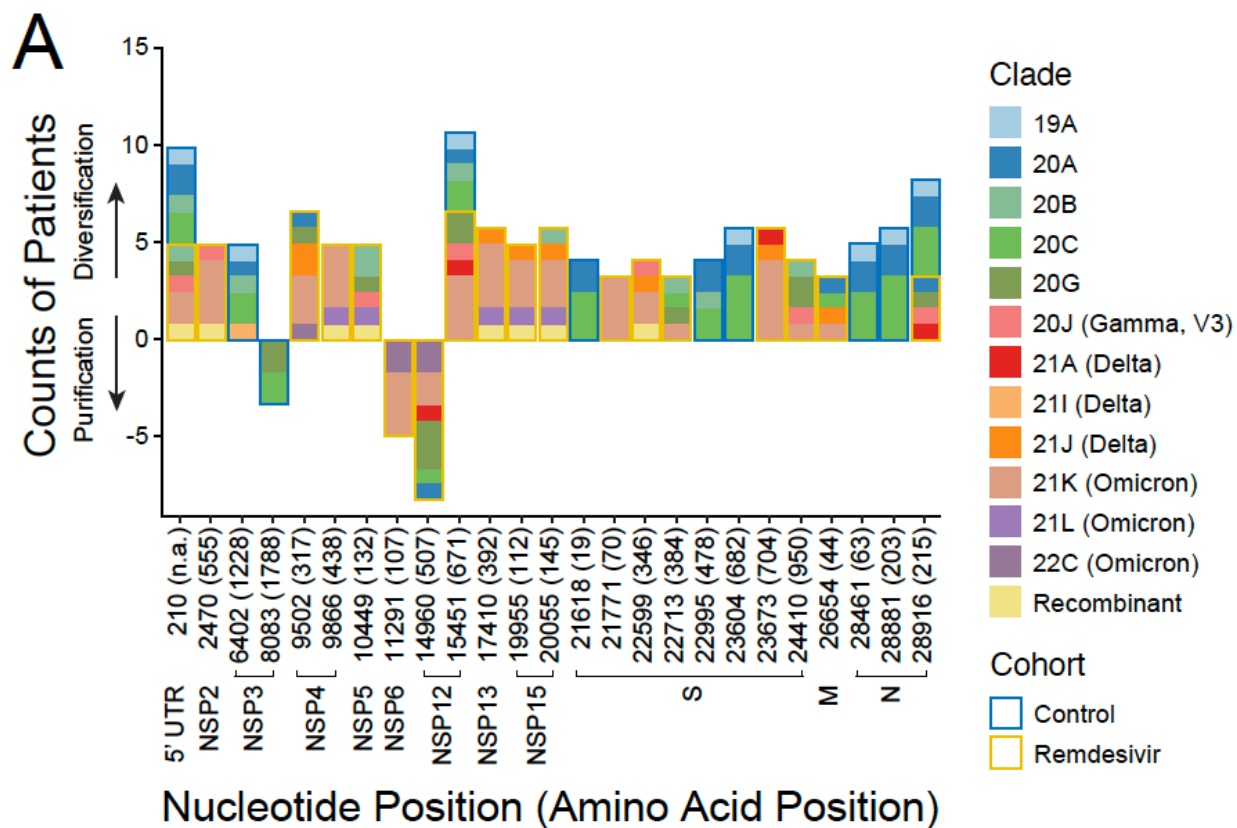
⁺ Corresponding authors: Ramon Lorenzo-Redondo, Ph.D. (ramon.lorenzo@northwestern.edu);
Judd F. Hultquist, Ph.D. (judd.hultquist@northwestern.edu)

TABLE OF CONTENTS

Supplemental Figure 1.....	pages 2-3
Supplemental Figure 2.....	pages 4-5
Supplemental Table Legends.....	page 6



Supplemental Figure 1 | Mutational frequency and Shannon entropy at sites of diversification or purification over time. A) Mutational frequency at the indicated positions from **Figure 4B** in the SARS-CoV-2 genome relative to the Wuhan-Hu-1 reference genome (accession number: MN908947.3) in “pre” and “post” specimens from the control cohort. Nucleotide and amino acid positions are provided below. Only specimens with >1% mutational frequency at a given position are shown (positions that were not detected listed as “n.d.”). Boxplots represent the median (center line) and first/third quartiles (box), with tails extending 1.5 times the interquartile range (IQR). **B)** Shannon entropy [$Sh = SUM[-(p_i) * \log_2(p_i)]$, where p_i is nucleotide frequency] in “pre” and “post” specimens from the control cohort at the positions indicated in A. Boxplots represent the median (center line) and first/third quartiles (box), with tails extending 1.5 times the interquartile range (IQR). **C)** Mutational frequency and **D)** Shannon entropy at the indicated positions from **Figure 4B** in “pre” and “post” specimens from the remdesivir cohort.



Supplemental Figure 2 | Clade distribution and nucleotide frequencies of selected positions following remdesivir treatment. A) Stacked bar plot illustrating the number of patients in each cohort that exhibit change in Shannon entropy (diversity) at a given position in paired specimens. Shannon entropy was calculated as $Sh = SUM[-(pi) * \log_2(pi)]$, where pi is nucleotide frequency. Positions where at minimum four patients in either cohort showed increase in Shannon entropy from zero (*i.e.*, diversification) or decrease to zero (*i.e.*, purification) are shown. Cohort membership is indicated by the blue (control) and yellow (remdesivir) outlines. The colors inside the bar chart refer to the clade of the virus in which the diversification or purification event was observed. **B)** Sequence logo plots depicting nucleotide frequency at the three nucleotide positions (14960, 15451, and 21771) that had significant changes following remdesivir administration in each set of paired patient isolates for which some mutational frequency was observed either in the "pre" or "post" specimen.

SUPPLEMENTAL TABLES

Supplemental Table 1 | Summary of previously identified remdesivir resistance mutations and their respective references.

Supplemental Table 2 | Summary of isolates analyzed in this study including clade designations, lineage designations, GISAID accession numbers, and NCBI accession numbers.

Supplemental Table 3 | Summary of consensus level mutations found between matched “pre” and “post” specimens in the control and remdesivir cohorts.