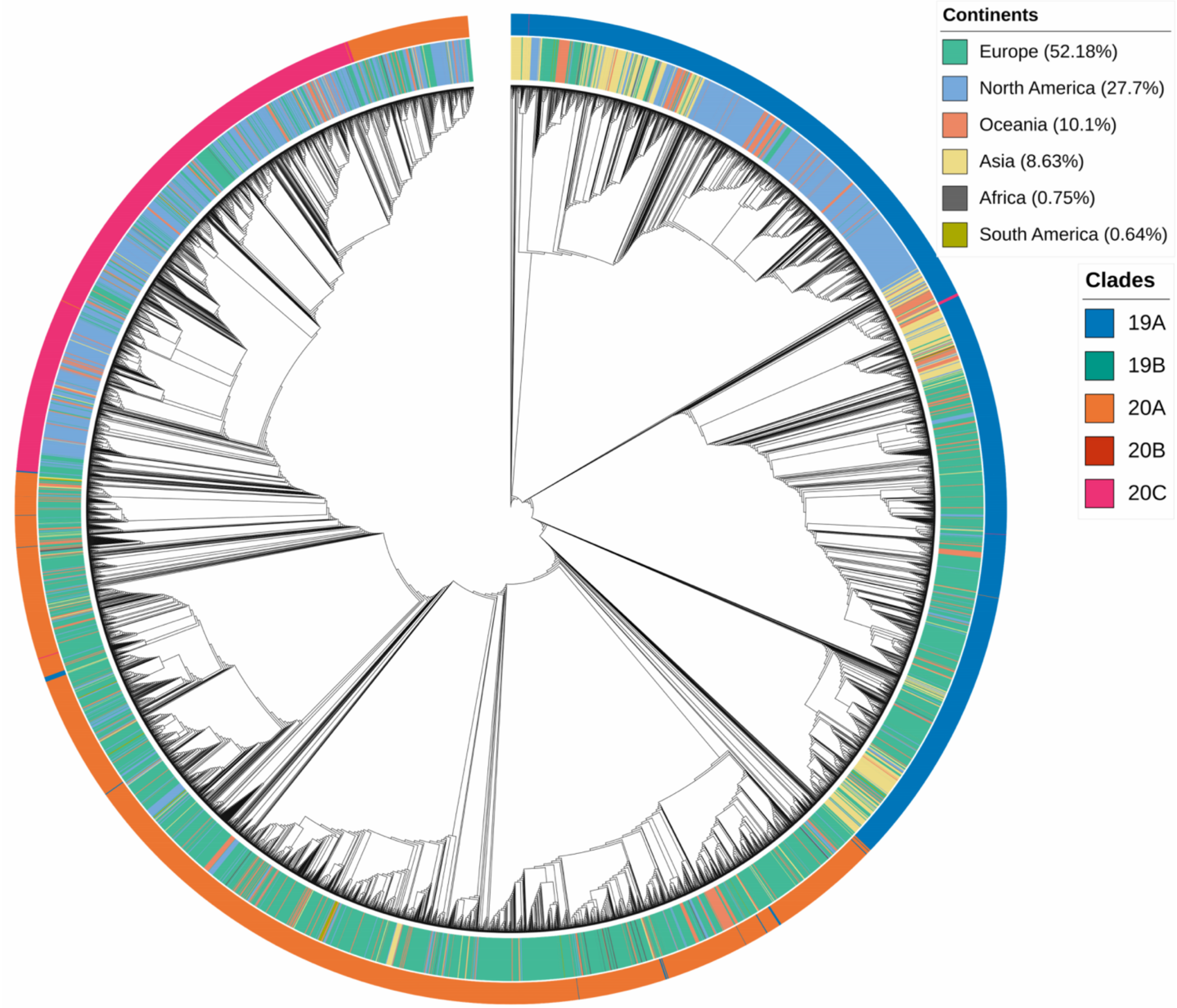
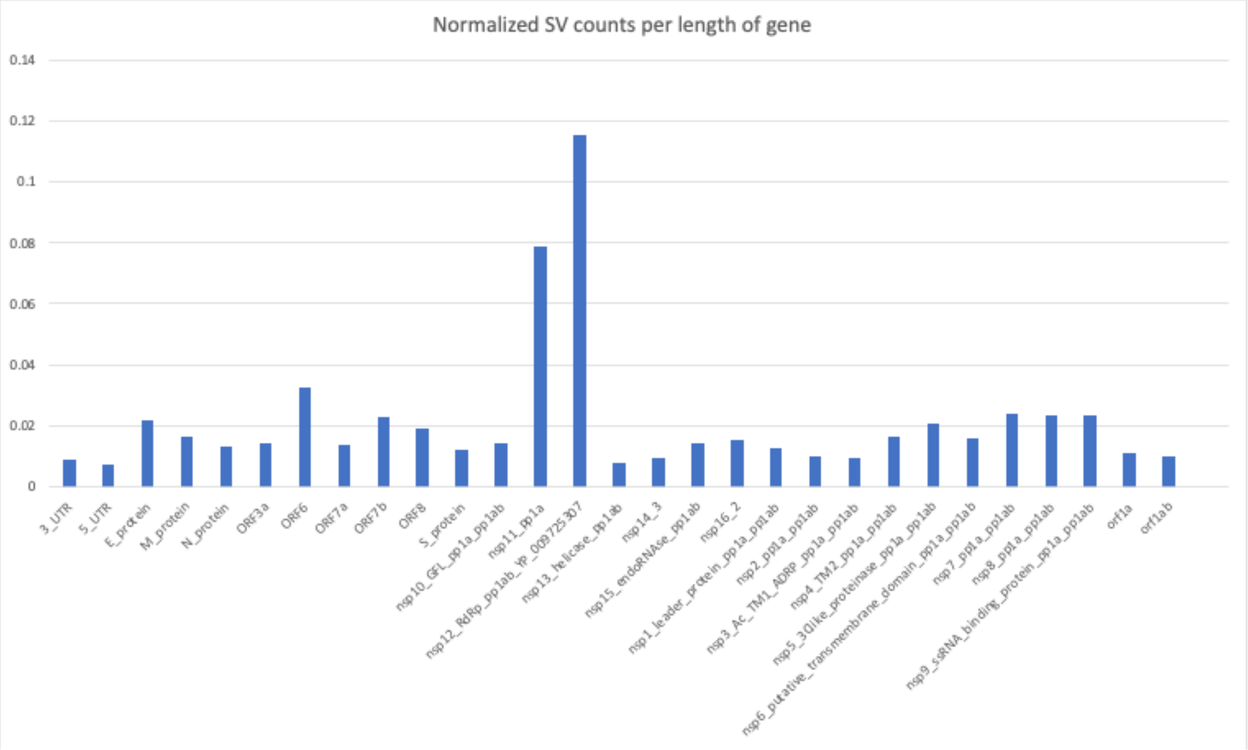
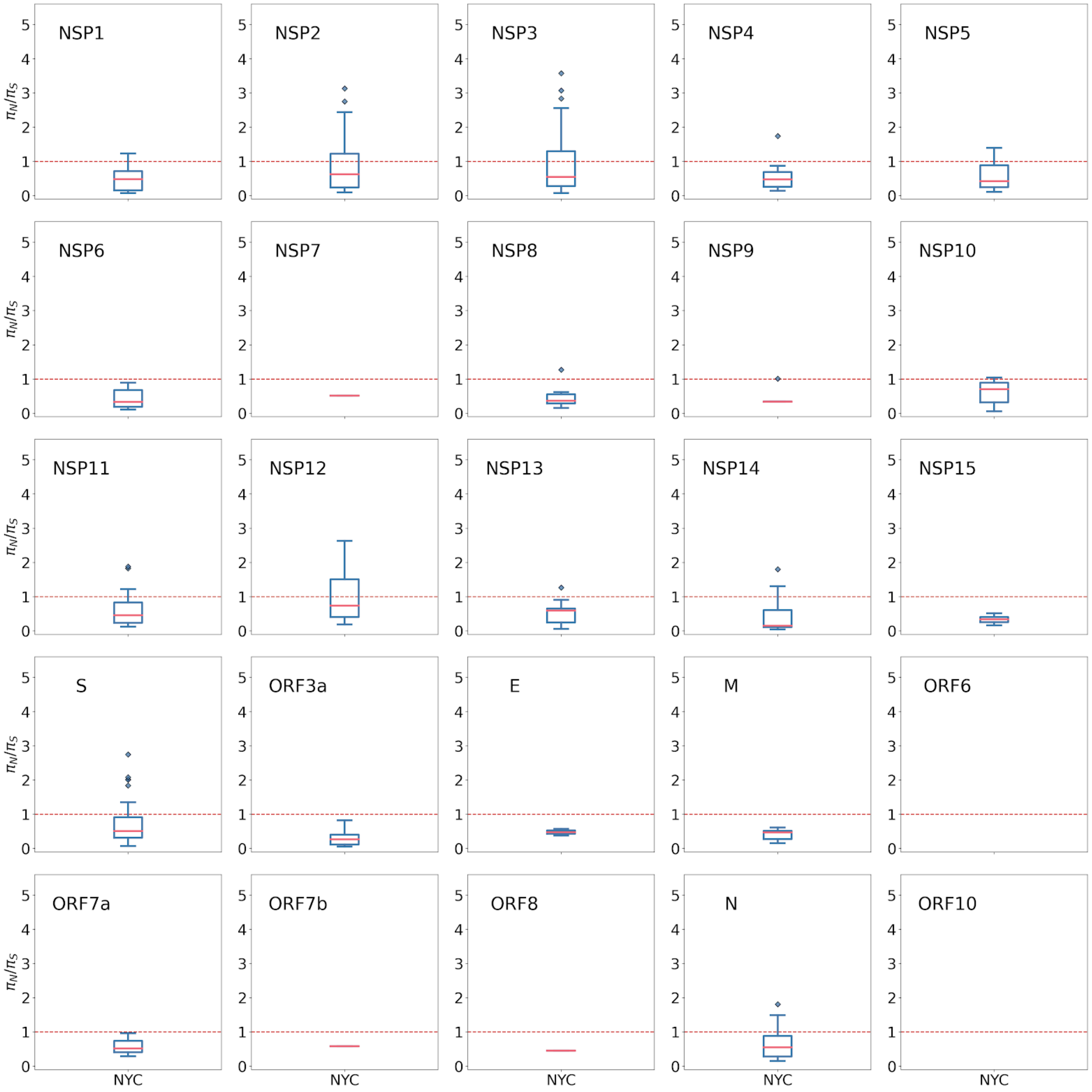
Supplementary Material



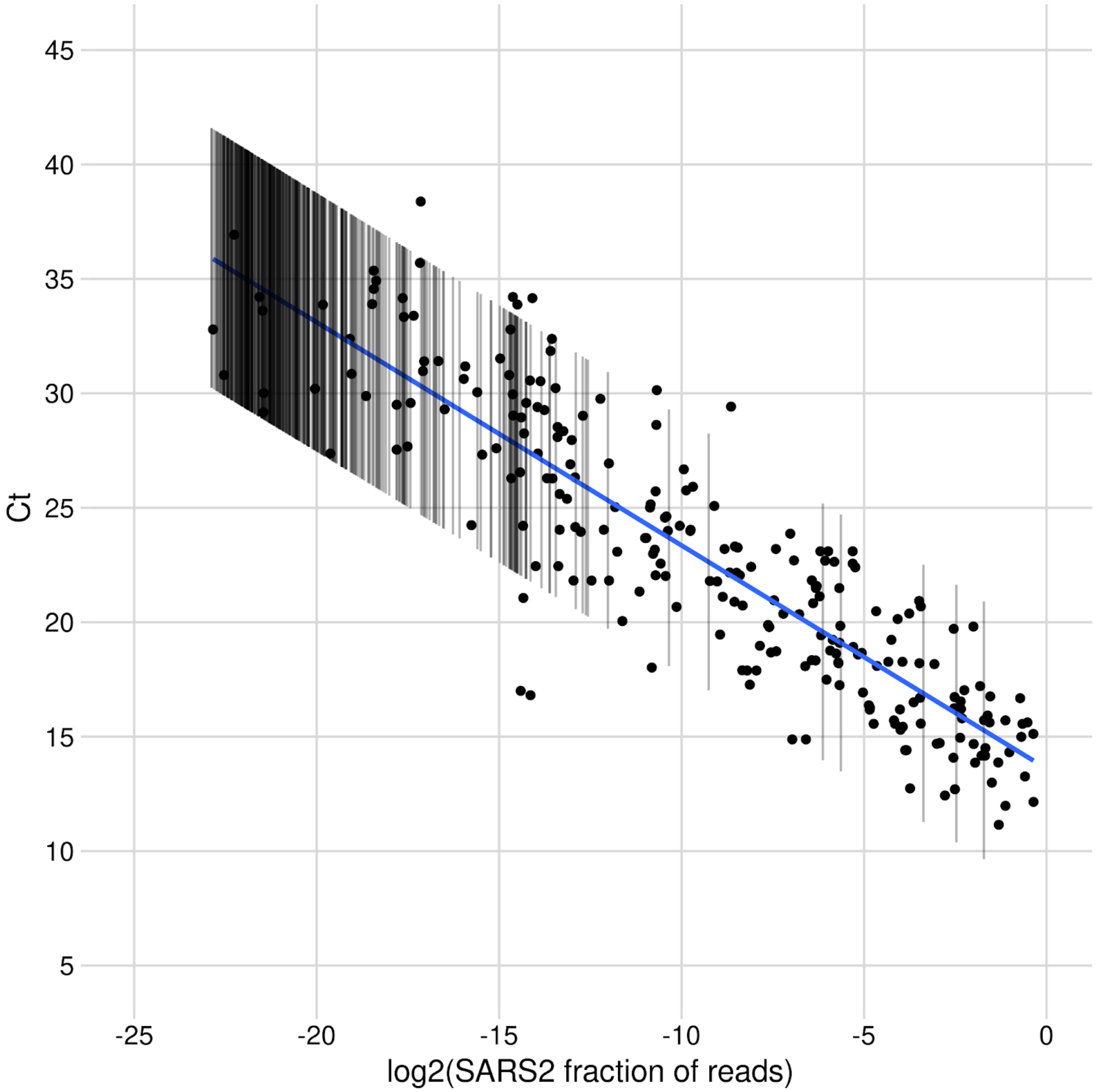
Supplemental Figure S1: Inferred phylogenetic tree for the 6,928 GISAID genomes. Directly outside of the tree branches is the continuous annotation ring for the continents corresponding to each sample. The second ring indicates clade membership (using Nextstrain clade nomenclature) of the respective sequences.



Supplemental Figure S2: Structural Variation concentration across annotated regions for SARS-CoV-2. Number of insertion and deletions below 1000bp are counted per region and are normalized by the size of the region showing a clear enrichment in certain parts.



Supplemental Figure S3: Non-synonymous to synonymous diversity ratios for iSNVs and SNPs in NYC Weill Cornell data mapped onto NSP/ORFs of SARS-CoV-2 genome.



Supplemental Figure S4: SARS-CoV-2 mapped read fractions vs Ct. PCR positive samples are shown as points. Lines represent the 95% prediction interval for PCR negative samples with a non-zero number of SARS-CoV-2 mapped reads.

Supplemental File S1: A table of recurring SV including position, sample count and the type of the SV.

Supplemental File S2: WHO detection probes and primers, Artic and Tomer Altman sequencing primers with SNP and iSNV occurrences indicated are summarized in this Excel file.

Supplemental File S3: A table describing how many samples from each dataset were used for each of the preformed analyses.

Supplemental File S4: A VCF file containing all SV calls for the 129 RNA-seq samples used in our study.

Supplemental File S5: A set of VCF files containing all iSNV and SNP calls for the Houston RNA-seq samples used in our study.

Supplemental File S6: A set of VCF files containing all iSNV and SNP calls for the NYC Weill Cornell RNA-seq samples used in our study.

Supplemental File S7: A VCF file containing all SNP calls for the GISAID genomes used in our study.