

Dear Editor,

We are writing to submit our manuscript entitled “Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2,” which we hope you will consider for publication in Cell.

During the recent months of the SARS-CoV-2 pandemic, basal lineages of the virus have been almost completely replaced by derived, variant lineages. Many of the lineages possess mutations that have been shown to enhance receptor-binding or alter antigenicity. Despite these clear indications of adaptive evolution, the traditional methods used to identify adaptive evolution from sequence data alone (McDonald-Kreitman,  $d_N/d_S$ , Tajima’s J, etc) fail to detect adaptive evolution in SARS-CoV-2. This is because these methods rely on fixation (or near fixation) of substitutions and, thus, are sensitive to adaptive evolution on time scales of years to decades.

In this manuscript, we present a new method for identifying adaptive evolution on shorter time scales by correlating protein-coding changes with clade-level evolutionary success. This method relies only on genome sequences and the phylogenetic relationships between them and could be applied to the genomes of any organism with dense-temporal sampling. Here, we apply this method to SARS-CoV-2 to perform a comprehensive analysis of adaptive evolution across the SARS-CoV-2 genome. Our analysis provides strong evidence that SARS-CoV-2 is evolving adaptively, and that spike S1 is the primary locus for this evolution.

Spike S1 contains the receptor-binding domain and is a main target of neutralizing antibodies. As such, evolution in this domain may affect the durability of infection- or vaccine-induced immunity. In influenza H3N2, adaptive evolution in HA1 (the equivalent to S1) necessitates nearly-annual updates to the vaccine. We compare the surface proteins of SARS-CoV-2 and H3N2 and find that adaptive evolution at nonsynonymous sites in S1 is occurring at roughly 4.5 times the pace of HA1.

Though the majority of adaptive evolution is concentrated in S1, we also identify mutations elsewhere in the genome that show strong signatures of positive selection. Many of these mutations have been under-appreciated in the literature thus far and we direct attention to a particular deletion in the *nsp6* gene. We believe this deletion has played an important role in the adaptive evolution of SARS-CoV-2 and is deserving of experimental characterization.

Thank you for considering our manuscript for publication in Cell. We look forward to your response.

Sincerely,

Kathryn Kistler, John Huddleston, and Trevor Bedford