

RAPID ASSESSMENT OF SARS-COV-2 CLADES THROUGH MOLECULAR SEQUENCE ANALYSIS

Alexander G Lucaci, Jordan D Zehr, Stephen D Shank, Steven Weaver, Sergei L Kosakovsky Pond

Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania, USA



INTRODUCTION

In order to monitor the continued evolution of the SARS-CoV-2 genome, we developed a rapid assessment tool designed to investigate the nature and extent of selective forces acting on viral genes in clade sequences.

OBJECTIVES

We investigated the nature and extent of selective forces acting on the viral genes in B.1.617.2 (Delta) clade sequences.

METHODS

Our application relies on performing comparative phylogenetic analyses to identify clade-specific molecular evolution by evaluating selective pressures for query clade sequences against automatically selected background sequences. Full details, including software implementation, are provided in the QR code link.

A key observation is that selection appears to be operating on many sites that are not a part of the clade-defining signature mutation set, implying that there may be ongoing diversification and adaptation in these clades.



INSTITUTE FOR GENOMICS AND EVOLUTIONARY MEDICINE



RESULTS

Within B.1.617.2 (Delta) sequences we find statistical support for selection acting on several sites outside of the clade-defining set for the Spike gene. Below, we list two examples in the NTD and describe interesting features for each site within clade sequences.

- Spike/77
 - K77T is noted in clade sequences
 - Pervasive and Episodic selection
 - 77T occurs at a frequency of 0.19% of Delta sequences
- Spike/222
 - Directional selection for A222V
 - Pervasive and Episodic selection
 - 222V occurs at a frequency of 1.83% of Delta sequences



Figure 1. Demonstrates an example of the Spike gene trimer with sites (in Magenta) where we find statistical support for ongoing evolution. The NTD of Spike is colored in Green. The RBD of Spike is colored in Blue.