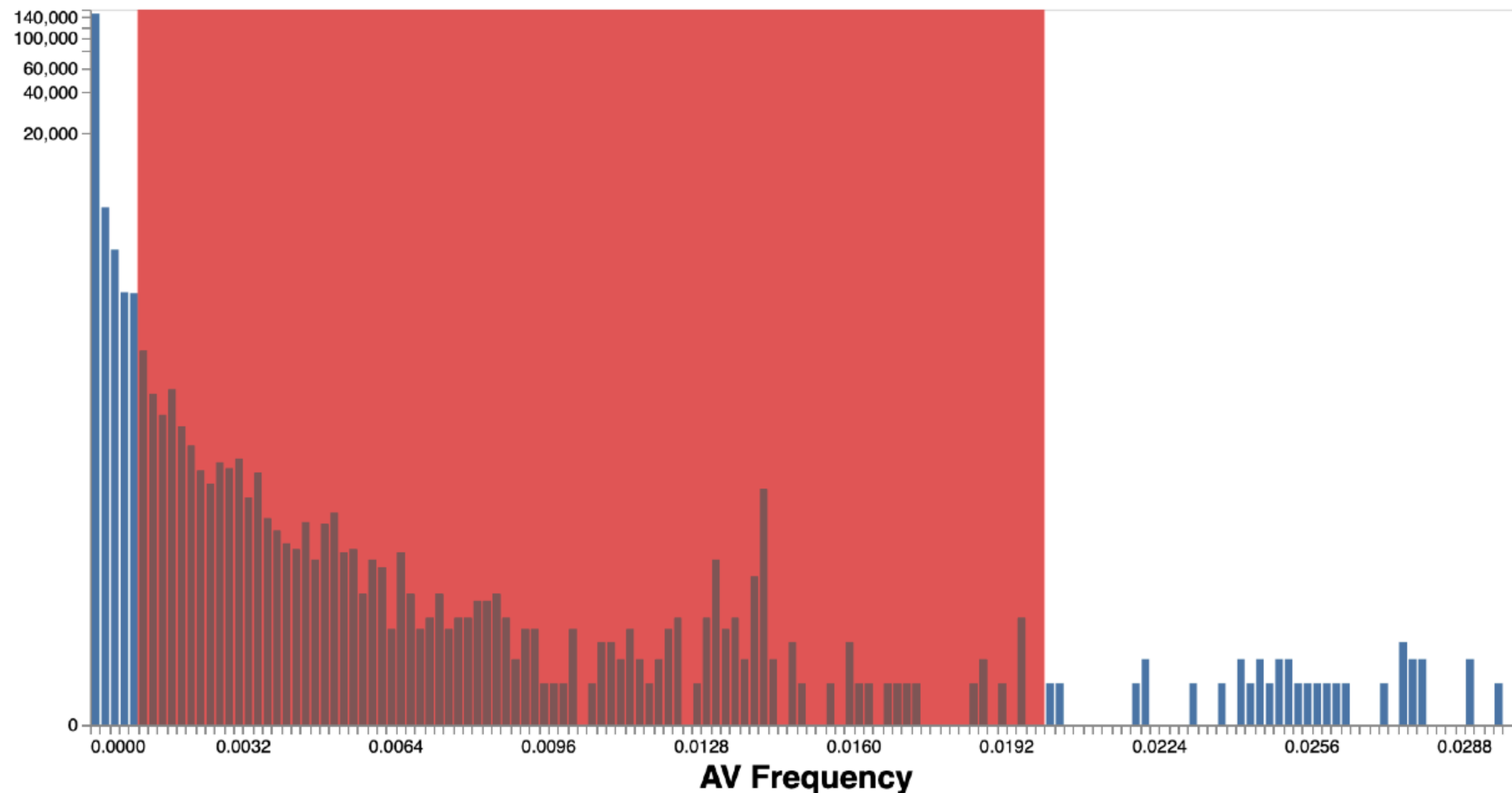


# Which AVs are interesting and non-obvious?

- Less common variants (e.g. 0.1% - 2%) are potentially interesting and non-obvious
- Too many (>2000) to test exhaustively



# Analytical framework

- Key mutations have significant phenotypic and epidemiological significance.
- Analysis of substitution patterns in large sets of SARS-CoV-2 data can reveal genomic sites subject to **selective pressure** => **possibly important**
- Used a suite of stock and modified **dN/dS methods** for **coding sequence evolution** developed over ~2 decades and extensively used in other RNA viruses
- The methods are implemented in the HyPhy software package ([www.hyphy.org](http://www.hyphy.org))
- Our public servers ([datamonkey.org](http://datamonkey.org)) have processed >10,000 complex CoV evolutionary analyses from researchers worldwide.
- Spent considerable effort to scale up analyses and develop data reduction techniques to manage data sizes.
- Collaborative efforts, many other open source tools at [covid-19.galaxyproject.org](http://covid-19.galaxyproject.org)