



Pathogen evolution under natural selection: the influenza A case study

Ph.D. Advisor: Simon Levin

Career Path: Postdoctoral research, University of Pennsylvania

Contact Information: skryazhi@gmail.com

Sergey Kryazhimskiy

Recent advances in molecular biology and medicine have provided us with a previously unavailable opportunity to study large amounts of genetic data from various organisms. Sequences of microbial pathogens are subject to a particular interest because the analysis of this data can help achieve a twofold goal. First, insights gained from such analysis help us mitigate the impact of the rapidly evolving pathogens on people's health. Second, microbial pathogens evolve at rates much higher than those typical for mammals, which allows us to observe evolution in real time and better understand the evolutionary processes in general. My dissertation contributes to the understanding of the role of natural selection in the evolution of rapidly evolving pathogens, in particular of the influenza A virus.

In the dissertation, I develop a statistical framework for studying the fitness landscape of an organism based on genetic sequence data. The analysis of the influenza A hemagglutinin sequences reveals that positive selection occurs with a strong preference with regard to the target amino acid. I also show that synonymous nucleotide substitutions in the influenza A genome evolve under natural selection as opposed to neutrality that is commonly assumed. Finally, I develop a theoretical framework for modeling epidemiology and evolution of multi-strain pathogens. The suggested approach allows us to construct tractable multi-strain pathogen models under a wide variety of assumptions. Using this approach, I suggest a simple mechanism of how frequency-dependent selection shapes the evolution of a virus with two epitopes that elicit independent immune responses.

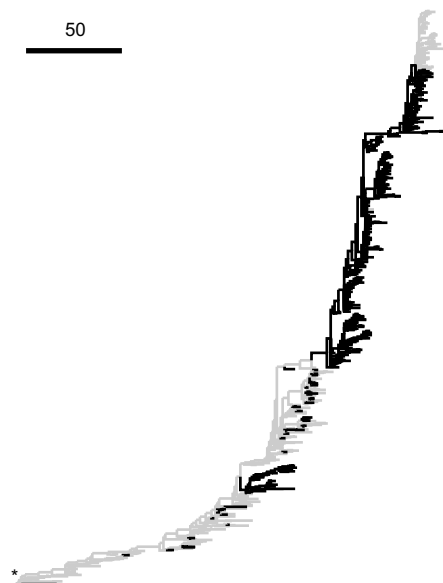


Figure 1. The phylogenetic tree of the hemagglutinin (HA) protein of influenza A is shown, with all amino acid substitutions mapped that resulted in lysine at site 145. The leaf nodes represent the HA sequences sampled between 1968 and 2005. The root of the tree, i.e., the most recent common ancestor of all observed variants, is marked by an asterisk. Branches where the sequence at the descendent node has lysine at site 145 are in black. On this tree, lysine appeared 33 times at site 145. Branch lengths are measured in nucleotide substitutions.