The Ability of SARS-CoV-2 to Transmit to New Hosts Evolved in Bats

Barry Khim
Lake Forest College
Lake Forest, Illinois 60045

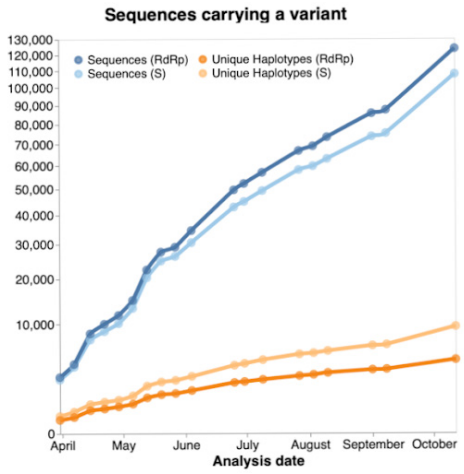
In December of 2019, a novel coronavirus was first reported in Wuhan, China. Evolutionary analysis identified the virus as a severe acute respiratory syndrome-related virus (SARS) with many similarities to the first SARS-CoV (Wu et al. 2020). It was subsequently named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Since then, SARS-CoV-2 has spread around the world, greatly disrupting social and economic interactions. The direct origin of the pandemic is unknown. While it is known that these viruses are found in horseshoe bats, the sudden transmission to human hosts represents decades of evolution (Boni et al. 2020). How exactly did this occur? What extent of evolution is required for a bat virus to transmit to humans? Researchers at MRC-University of Glasgow Centre for Virus Research set out to answer these questions. The ability to replicate efficiently and spread successfully is something most RNA viruses acquire after switching to the new host species. However, SARS-CoV-2 comes from the family of *Sarbecoviruses* which already transmit frequently between bat species because of their generalist properties. These viruses have evolved spike proteins to latch onto cells. They then bind to angiotensin-converting enzyme receptors for cell entry. Coincidentally, this allows them to successfully infect non-bat species, including humans, by binding to the human angiotensin-converting enzyme receptors. The researchers hypothesized that because of these prior adaptations that occurred in bats, SARS-CoV-2 did not need additional adaptations to transmit to humans.

To determine how SARS-CoV-2 came to be, the researchers investigated the evolutionary history of bat *Sarbecoviruses*. Historical evidence of virus adaptation in bats was contrasted with any evidence of virus adaptation that occurred in humans since the outbreak began in 2019. To do this, the researchers used an array of selection detection methods on two groups. One group consisted of 133,741 SARS-CoV-2 genomes sampled during the first year of the outbreak. The other group was 69 *Sarbecovirus* genomes that were separated and examined phylogenetically (MacLean et al. 2021). The researchers also analyzed the *Sarbecovirus* tree for changes in the patterns of CpG (site of a cytosine followed by a guanine in the 5′ to 3′ direction) because shifts in the CpG sites are an indicator of evolutionary changes.

The 133,741 SARS-CoV-2 samples, taken during the first 11 months of the pandemic, were examined for evidence of virus adaptation (MacLean et al. 2021). Each sample's genome was sequenced. The occurrence of unique sequences was plotted over time. The researchers noted that the rate of increase for unique sequences compared to the total number of sequenced genomes was quite slow [1] (MacLean et al. 2021). This indicated that the virus was evolving relatively slowly within the human population. Furthermore, statistical analysis revealed that selection was negative. The few mutations that were occurring were deleterious and failed to persist. Only after increasing levels of host immunity, due to vaccines, and SARS-CoV-2 circulation was it expected to see evidence of adaptation. However, they noted that any subsequent evolution in humans was not relevant for determining the origin of the virus' ability to efficiently spread between human hosts.

The researchers then turned to analyzing the 69 *Sarbecovirus* genomes found in bats. Viruses were separated and organized photogenically. The clade containing SARS-CoV-2 is referred to as new coronavirus (nCoV). To separate and organize the virus, they used an array of selection detection methods: BUSTED, hidden Markov model, aBSREL, mixed effects model of evolution, etc. Examination of the differences between virus phylogeny revealed evidence of historical positive selection (MacLean et al. 2021). In the nCoV clade, Orf1ab, a gene associated with SARS-CoV-2, had many sites that were subjected to positive and diversifying selection. Furthermore, analysis of CpG patterns showed that there was an adaptive shift in viruses of the nCoV clade (MacLean et al. 2021). These findings support the idea that the generalist properties of SARS-CoV-2 evolved in bats and not humans. This is further supported by the fact that SARS-CoV-2 can also transmit to other mammals (pangolins, mink, cats, etc.).

The implication of this research presents a concerning reality. The diversity and generalist nature of *Sarbecoviruses* allows it to transmit to new hosts frequently. This suggests that there are species of wild mammals infected with yet to be identified nCoV-like viruses. The threat of a new SARS-CoV that is genetically divergent enough to evade current, acquired immunity could possibly emerge in the future. For this reason, the researchers stress the importance of ramping up surveillance to better prepare ourselves against future outbreaks of SARS-CoV.



[1] The researchers looked at two traits of interest: spike proteins (S) and RNA-dependent RNA polymerase (RdRp). Total number of S and RdRp gene sequences (blue) and total number of unique S and RdRp variations (orange) were plotted on a graph.