EEOB 563

Final Project

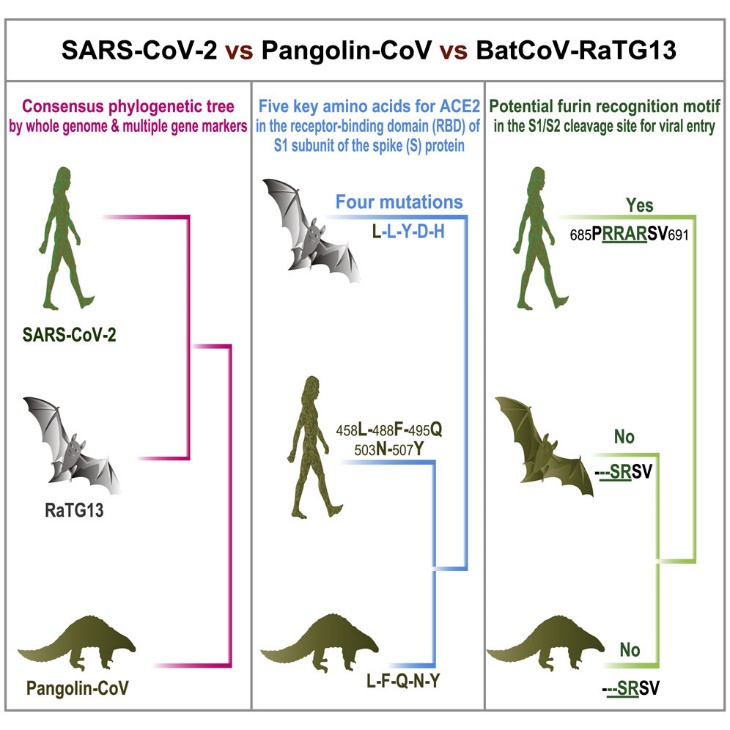
Mihir Kharate

Establishing a Phylogenetic relationship between Zoonotic Coronaviruses:

Introduction

Since last two years, Coronavirus is a topic in Kairos. There has been a lot of speculation going on with how a virus that infects animals suddenly evolved and started infecting humans. There have been some researches that speculate of a possible intermediate animal host that allowed the coronavirus to mutate and adapt to a more complex organism, human. While the COVID-19 pandemic had just started, Malayan pangolins that were being smuggled in china were rescued and tested. It was found that they too were infected by a Coronavirus (Which they possibly got from bats). In this paper I have attempted building a phylogenetic relationship between these viruses. The Spike protein is THE most essential component for the virus to attack and infect the host. Zoonotic viruses have taken the leap to jump from attacking animals to attacking humans. Hence, looking at the Spike protein would give us a glimpse into the evolution and divergence of these coronaviruses.

To look at the evolution of SARS-Cov-2 from other coronaviruses I have constructed phylogenetic trees with the whole genomes of the Coronaviruses of interest. I have also created a phylogenetic tree based on the Spike Protein of the given Coronaviruses. The Coronavirus strains that I have used in this phylogenetic analysis are: SARS-CoV, SARS-CoV-2, MERS-CoV, Pangolin-CoV and Bat-CoV.

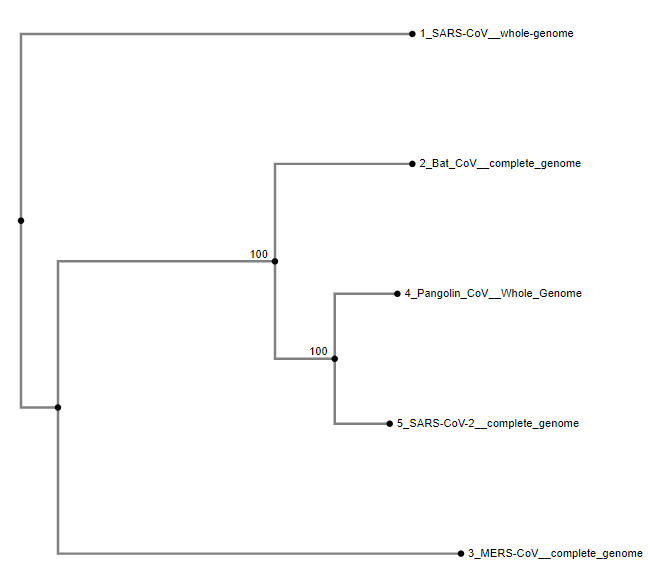


The figure above taken from Zhang et al., 2020 proposes a hypothesis of a relation of the novel coronavirus (SARS-CoV-2) with an intermediate host (pangolin).

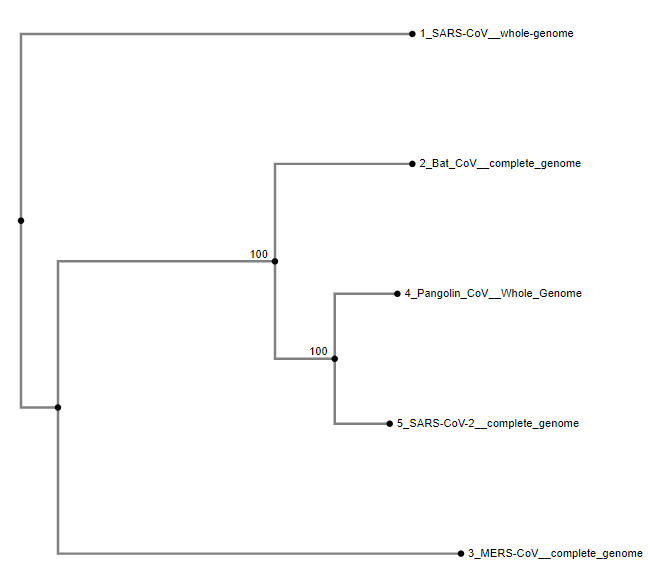
Methods

For constructing my phylogenetic trees, I first used Mafft module to align my multiple sequences for Coronaviruses. I used maximum parsimony alignment on the genome and amino acid trees to construct a phylogenetic tree. I performed Maximum Likelihood analysis on the trees and retrieved various trees. The amino acid tree was ran under 100 bootstraps. I observed the trees using iTOL and have included the two trees that matched for both the whole coronavirus genomes as well as the Spike protein sequences.

Results



Above is the phylogenetic tree constructed from the whole genome analysis of the five different Zoonotic coronaviruses.



The tree above represents the relationship between the five coronaviruses based on their Spike protein composition.

Discussion

The trees for whole genome analysis and the Spike protein match. With Pangolin CoV grouped together in the clade with SARS-CoV-2, their relationship is evident. However, I have also come across some trees in my literature study that show that the Bat CoV is more closely related to the SARS-CoV-2. The possibility of this differentiation could be the analysis of different variants or strains of the viruses. For my analysis I downloaded the genome of the reference SARS-CoV-2. Since this is a reference genome, it is generalized across all other SARS-CoV-2 strains. This might be providing the variance in my results.

Citation

Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Current Biology*, *30*(7), 1346–1351.e2. <https://doi.org/10.1016/j.cub.2020.03.022>

Huang, Y. (2020, August 3). *Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19*. Acta Pharmacologica Sinica. <https://www.nature.com/articles/s41401-020-0485-4?error=cookies_not_supported&code=86145b2c-cfe7-4444-b264-db0d1ff4b40c>

Li, T., Liu, D., Yang, Y., Guo, J., Feng, Y., Zhang, X., Cheng, S., & Feng, J. (2020). Phylogenetic supertree reveals detailed evolution of SARS-CoV-2. *Scientific Reports*, *10*(1), 1–5. <https://doi.org/10.1038/s41598-020-79484-8>