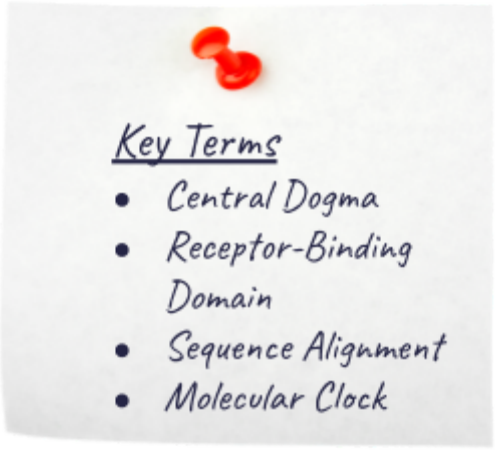
D4P Discussion Guide

Andrés Mansisidor presents on when the new coronavirus became deadly to humans

D4P Video: https://bit.ly/2Szd7Ky

The proximal origin of SARS-CoV-2 · Kristian G. Andersen, Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes, and Robert F. Garry· *Nature Medicine* · March 17, 2020
https://doi.org/10.1038/s41591-020-0820-9 · PDF

This paper reviews and synthesizes over 30 different studies on coronavirus genomes. Through this comparative genomic analysis, researchers utilize several styles of reasoning to determine when the SARS-CoV-2 virus became deadly to some humans, and discuss possible scenarios of how SARS-CoV-2 could have arisen.



questions for learners	discussion points for educator
1. *What is central dogma, and how does this relate to the paper discussed here?*	The central dogma of molecular biology refers to the flow of genetic information (DNA → mRNA → Protein, or for viruses RNA → mRNA → Protein) in a biological system. Here, researchers leaned on the central dogma to understand how a change in the sequence of a protein is a reflection of a mutation in DNA. More specifically, they used sequence alignment to learn about mutations in the receptor binding domain of coronavirus spike proteins, and how these changes altered its function.
2. *What is the spike protein receptor binding domain and why did researchers look at it?*	The receptor binding domain (RBD) in the spike protein is the most variable part of the SARS-CoV-2. Six of the amino acid residues of the RBD are necessary for the virus to bind to the ACE-2 receptor and gain access to cells. Five of these six residues are different between SARS-CoV-2 (2019) and SARS-CoV (2003). The RBD for SARS-CoV-2 is able to effectively bind to the ACE-2 receptor and enter host cells. Based on computational and structural predictions, however, the RBD is a good fit, but not the best fit for binding to ACE-2. This discrepancy between the actual data of binding efficacy and the predictive data suggests that SARS-CoV-2 has undergone natural selection on human or human-like ACE-2 receptors.
3. What is the cleavage site of the spike protein and why did researchers look at it?	The polybasic cleavage site of the spike protein is the junction between the two subunits of the spike protein (S1 and S2). This site is a key spot where host proteases can cleave the spike protein to allow the virus to enter host cells, so this site is important for determining the range of hosts this virus can infect and how infectious it is.
4. Why do researchers think that SARS-CoV-2 could have come from a bat?	Bats are known to contain many coronaviruses. For example, bat coronavirus (RaTG13) was isolated from bat droppings in 2013. The genetic sequence of this virus shows 96% homology (similarity) to the SARS-CoV-2 virus. However, the RaTG13 bat coronavirus sequences encoding for the spike protein differ, which suggests that RaTG13 bat

	coronavirus is a closely related ancestor to the SARS-CoV-2.
5. Why do researchers think that SARS-CoV-2 could have come from a pangolin?	The genetic sequence of the SARS-CoV-2 virus is ~91% similar to coronaviruses found recently in 2 pangolins that died of Covid-19 like symptoms. Although these viruses don't match the sequence as highly as the bat coronavirus, the RBD for SARS-CoV-2 spike protein is the same across all six key residues to the spike protein in coronaviruses isolated from pangolins. Mutations from different coronaviruses can mix together forming hybrid viruses. Thus, SARS-CoV-2 likely evolved from coronaviruses with the RBD mutations found in pangolins and from the highly related bat coronavirus.
6. What is the most likely origin of SARS-CoV-2 given the evidence the researchers present in this paper?	The rate at which mutations are incorporated into genomic sequences is the molecular clock, which researchers can then use to map viral lineages. Given the high sequence homology between SARS-CoV-2 and the bat-CoV, and the exact match between the SARS-CoV-2 and the pangolin-CoV, the most likely origin of SARS-CoV-2 is that these two viruses mixed in either an animal host or a human host to give rise to the current SARS-CoV-2. The cleavage site could have been acquired in either animal hosts or human-to-human transmission. More sequence evidence could support either of these possibilities, for example, animal or human samples that were collected before the pandemic outbreak that show virus sequences that match more highly to SARS-CoV-2.
7. What evidence do we have that SARS-CoV-2 was not engineered or escaped from a lab?	There are some lines of evidence that argue against the virus being engineered in or escaping from a lab. 1) While the sequence of the RBD could have been engineered to be a close fit with the human ACE-2 receptor in the lab, the more plausible scenario is that the RBD sequence was acquired through mutation or recombination, given the existence of the identical pangolin-CoV RBD. 2) The development of the cleavage site of SARS-CoV-2 (which increases the virus' pathogenicity) is something that researchers have seen in other viruses only after ongoing low-pathogenicity transmission (spreading of infection). This scenario has never been replicated in a lab for coronaviruses. 3) The cleavage site in the spike protein is modified by carbohydrates called O-linked glycans. Researchers are not exactly sure what these do, but they could be used by the virus to make a shield to "hide" itself from the host's immune system. Thus, these O-linked glycans are likely evidence of viral evasion of the host's immune system, which would have occurred overtime through mutation and natural selection in host populations, not in a lab. 4) Finally, when researchers study viruses in labs, they often "tag" them with a molecular marker so that they can keep track of them. These tags would stay in the viruses and there is no evidence of this kind of tagging in any of the SARS-CoV-2 viruses isolated from humans.