# hw_06

Haley Dylewski

9/30/2020

This document discusses two plots from the same scientific article and examines why one is effective and the other ineffective.

```
Effective Plot Example
```

This plot represents an effective example.

Description: This plot comes from a paper analyzing and comparing two genomic regions of different sarbecoviruses, in the context of SARS-CoV-2 lineage. At the top, the geographic regions the different viral samples originated in are shown, color, color coded, on a map. Below the samples are mapped on two phylogenetic trees, one for each genomic region, and are color coded according to the geographic region. The dotted lines indicate samples that fall at different phylogenetic positions in the two trees.
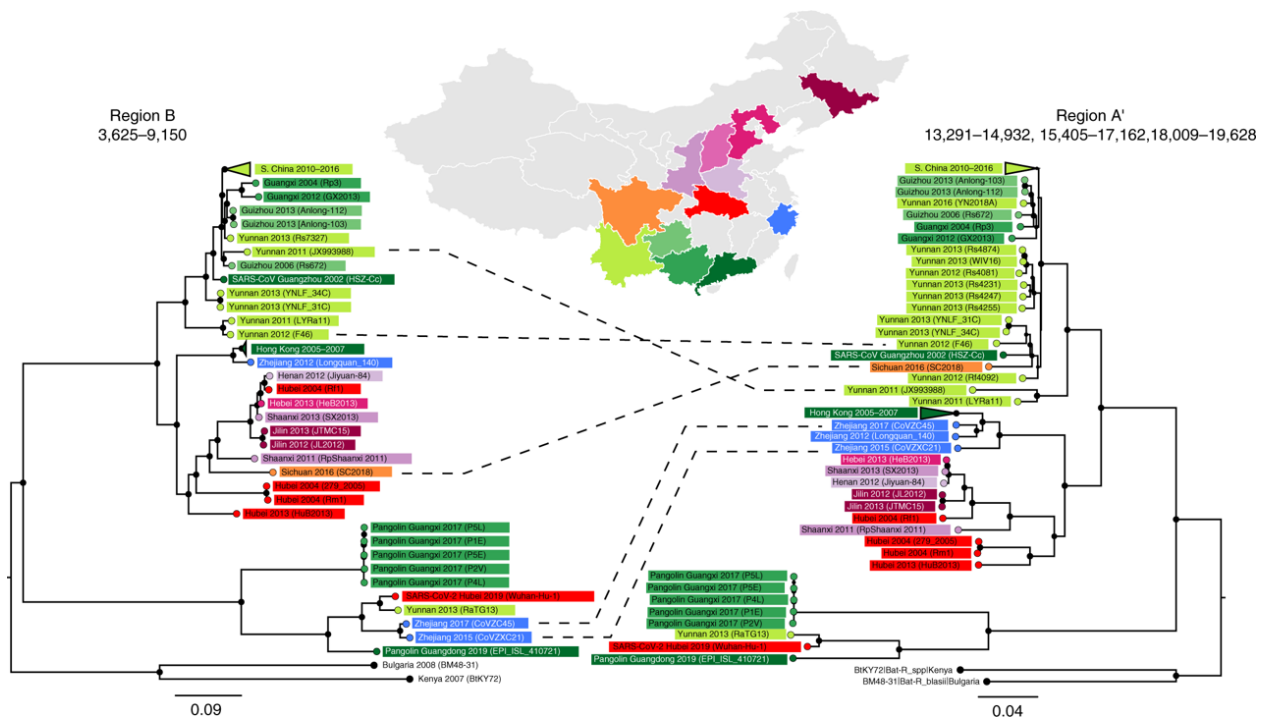
From Boni et al. Nat Microbiol. 2020 Jul 28. "Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic"

Goal: Shows that viruses from different geographic regions group into similar phylogenetic clades. Some samples fall at different phylolgentic positions between the two trees, indicating potential recombination events. Assessment: I think this plot is successful in its communication. I think through the combination

of colors and mapping, the point is gathered very quickly. Each part of the plot serves a purpose, with no extra fluff. The colors are clean and distinct.

I'm sure there is room for improvement but I have no immediate suggestions.

Ineffective Plot Example

Description: This figure shows the general gene structure of the sarbecovirus spike protein gene, S. The genome sequences of 6 sarbecoviruses were compared based on nucleotide sequence similarity and mapped phyologenetically for each S gene region. The variable loop region within the C-terminal domain of this gene was emphasized. The dotted lines in the phylogenetic trees connect broken branches so that differences in the tree for each region can be shown by one continuous tree. Then the genetic distance (aka average substitutions per site) between SARS-CoV-2 and 3 of the sarbecoviruses was determined and graphed for each of th segments of the S gene.
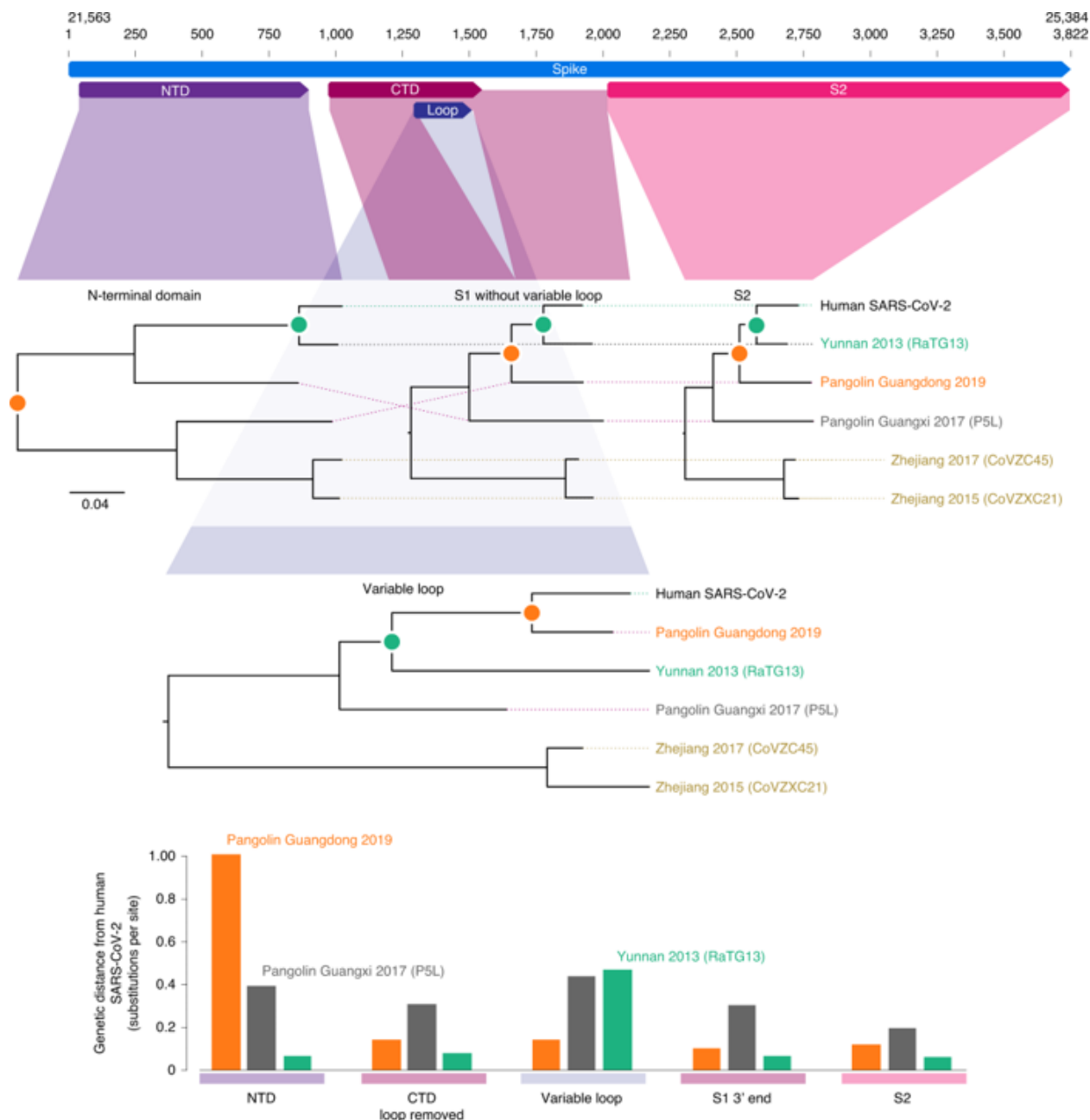
From Boni et al. Nat Microbiol. 2020 Jul 28. "Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic"

Goal: To understand which virus each SARS-CoV-2's S-gene likely came from. This plot is trying to communicate, that within the S protein (that exists in various sarbecoviruses), the variable loop region of SARS-CoV-2 (responsible for host cell receptor binding) is most closely related phylogenetically to the pangolin Guangdon 2019 virus, and that the genetic distance is low for this region.

Assessment: I find this plot unsuccessful in clear information presentation. The flow is unclear, what order I should be looking at things. My biggest critique concerns the phylogenetic trees int he center of the figure with dotted lines. The trees are not easily visualized. Further, the dotted lines are too pale to see easily and the shaded triangle over these lines obscures the colors further. It is also hard to directly compare the values across genetic distance graphs. As its stands, it is very difficult to visually determine the y-values beyond the NTD graph.

Suggestions: Break this figure up. Show the phylogentic trees stacked on below the other so they can still line up with their genetic segment, but are not overlapping. The genetic distance graphs either each need a y axis or there should be guidelines spanning them so that the genetic distance can be precisely judged. At the top, where the shaded regions correspond to the three domains, the two pink shades are very close. I would change one of the colors.