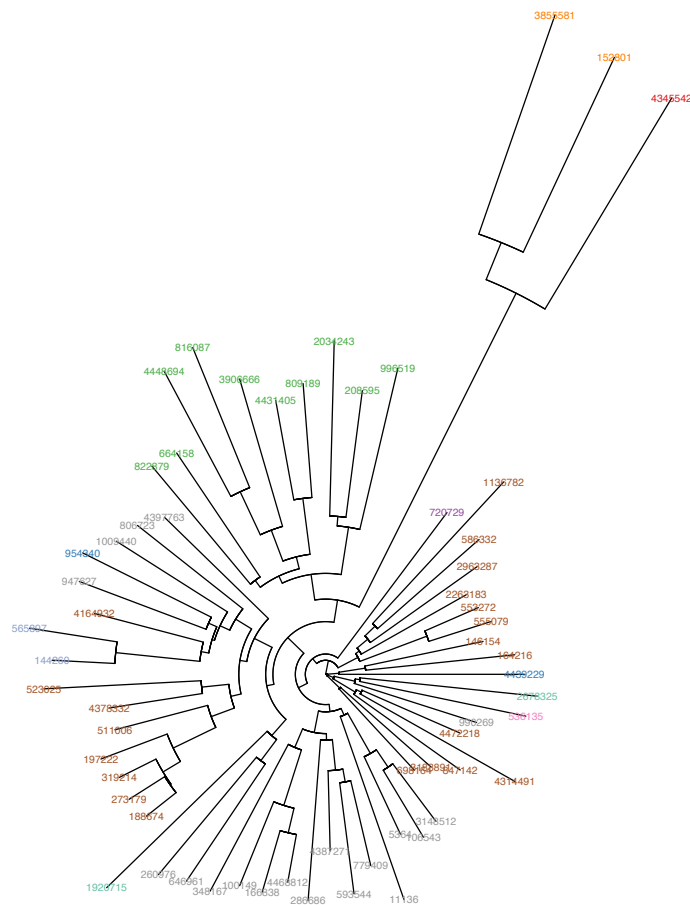3.

4. **Why are tips of similar color mostly clustered together?**

Tips of similar color are mostly clustered together in the phylogenetic tree because the tree is constructed based on genetic sequence similarity, where more similar sequences are placed closer together. Since sequences from the same phylum (represented by the same color) share significant genetic similarity due to their shared evolutionary history and common ancestry, they naturally cluster together in the tree. This clustering pattern when we color the tips by phylum demonstrates that taxonomic classification (phylum level) generally aligns well with genetic relationships, reflecting the fundamental principle that organisms within the same phylum have diverged more recently from each other than from organisms in other phyla. The 16S rRNA gene sequences used here are relatively conserved within phyla while showing greater variation between phyla, making them useful markers for phylogenetic classification at this taxonomic level.

5. Name two distinct likely reasons why the clustering by color is not perfect.

1) Horizontal gene transfer (HGT): Genetic material can be transferred between different species/phyla through mechanisms like transformation, conjugation, or transduction. This lateral transfer of genetic information means that some bacterial species may have acquired 16S rRNA gene segments from distantly related phyla, leading to sequences that show greater similarity to members of other phyla than would be expected from their vertical evolutionary relationships. This phenomenon can disrupt the clean clustering of phyla in the phylogenetic tree.

2) Limitations of the 16S rRNA gene as a phylogenetic marker: While 16S rRNA is a widely used marker for bacterial phylogeny, it has inherent limitations. The gene may evolve at different rates in different lineages, and some regions may be too conserved to effectively resolve relationships between closely related taxa while others may be too variable to reliably reconstruct deep phylogenetic relationships. Additionally, the presence of multiple, slightly different copies of the 16S rRNA gene within a single genome can complicate phylogenetic inference and lead to imperfect clustering by phylum.

7.
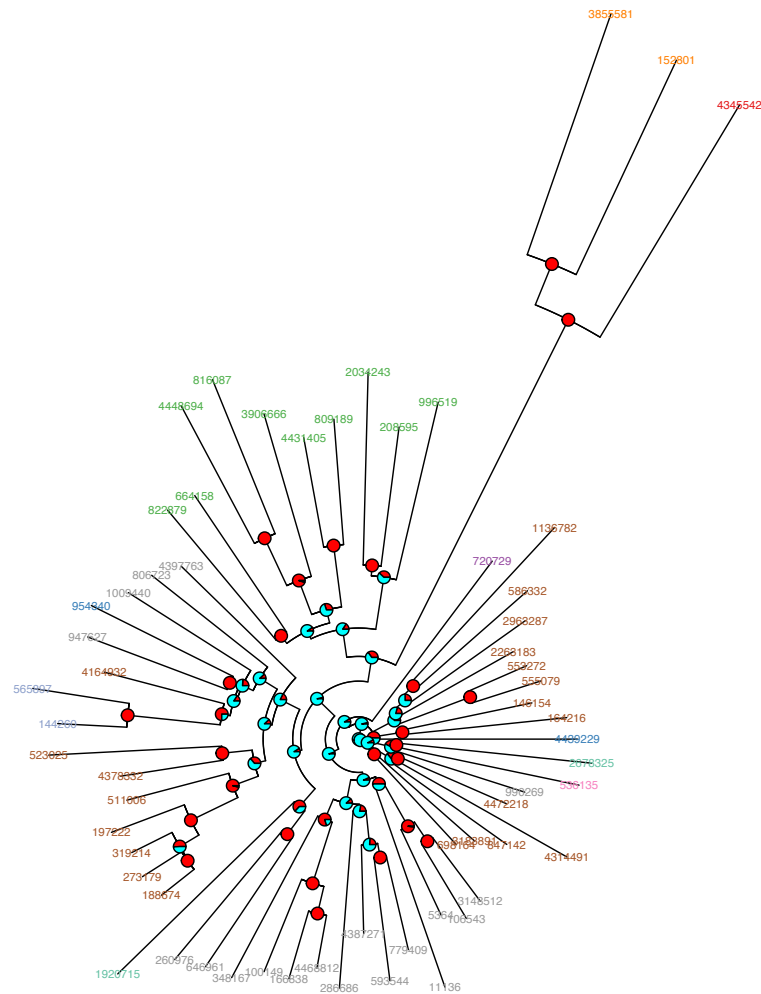
8. Based on the bootstrap tree pictured above, why is bootstrap support generally higher for the internal nodes closest to the tips, and lower for the internal nodes closest to the root? You do not need to complete questions 6 and 7 to answer this question.

reason1:

Higher support near tips: Nodes near the tips represent recent evolutionary relationships between closely related sequences. These sequences share many distinctive, well-preserved genetic signatures with relatively few differences between them. During bootstrap resampling, these clear genetic patterns are consistently recovered because even when some sites are not sampled, enough diagnostic sites typically remain to reconstruct these close relationships. This consistency across bootstrap replicates naturally leads to high support values for these near-tip nodes.

Reason2:

Lower support near root: Nodes near the root represent ancient evolutionary divergences that

have been obscured by multiple rounds of mutation over evolutionary time. The accumulation of substitutions at the same sites (saturation) and the occurrence of parallel or convergent evolution (homoplasy) have masked many of the original genetic signals. During bootstrap resampling, these weak and complex phylogenetic signals are easily disrupted, as random sampling of sites can lead to different tree topologies for these deep branches. This inherent uncertainty in reconstructing ancient relationships results in lower bootstrap support values for nodes near the root.