

SUPPLEMENTARY FIGURES

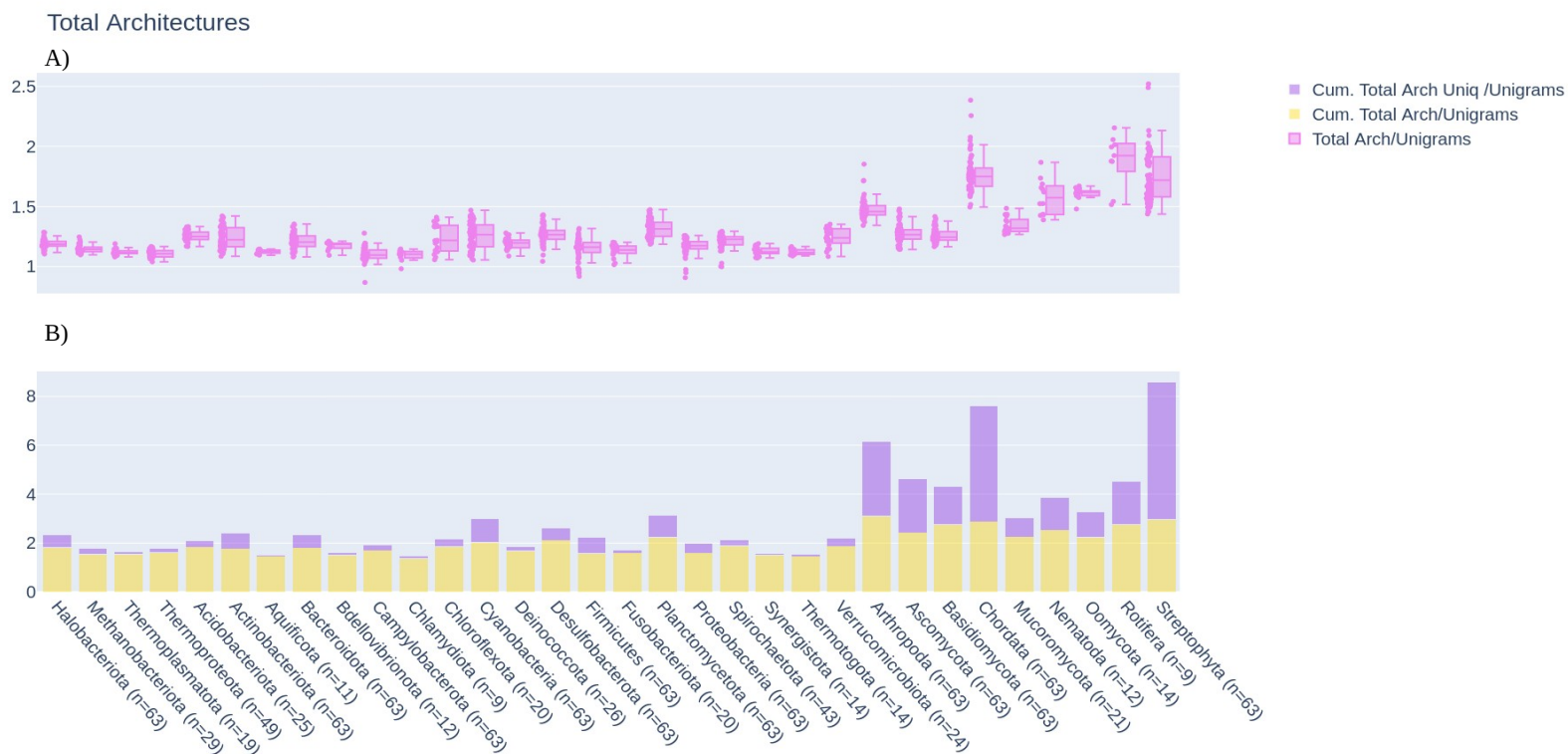


Figure 1. A) Distribution of number of total architectures per organisms per phylum, normalized by number of uni-grams. **B)** Cumulative number of different architectures (yellow) and unique architectures in each phylum, normalized by number of uni-grams.

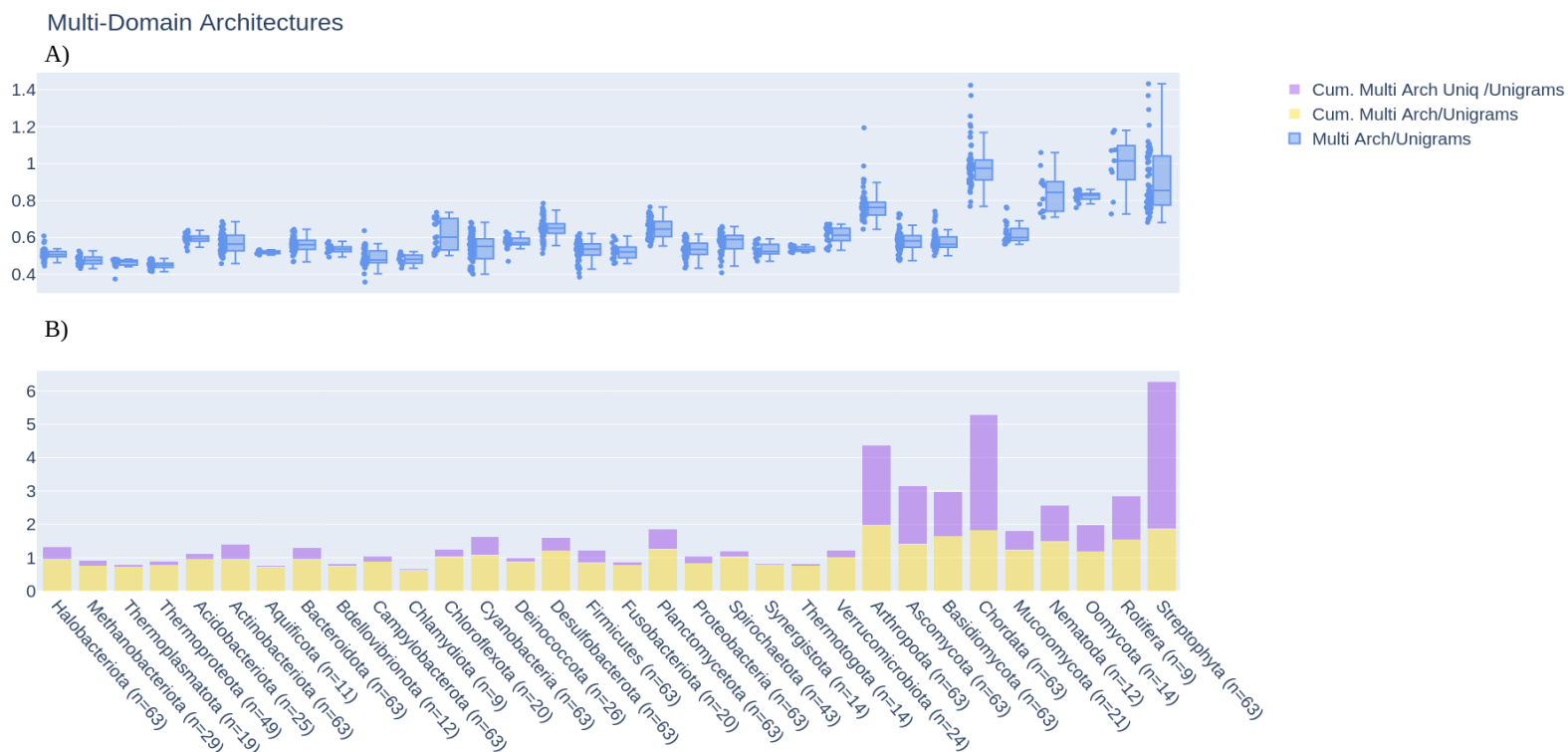
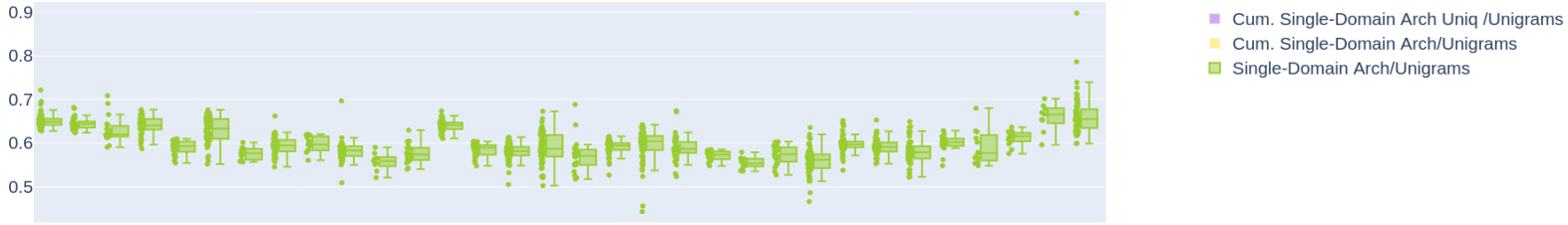


Figure 2. A) Distribution of number of multi-domain architectures per organisms per phylum, normalized by number of uni-grams. **B)** Cumulative number of different multi-domain architectures (yellow) and unique ones in each phylum, normalized by number of uni-grams.

Single-Domain Architectures

A)



B)

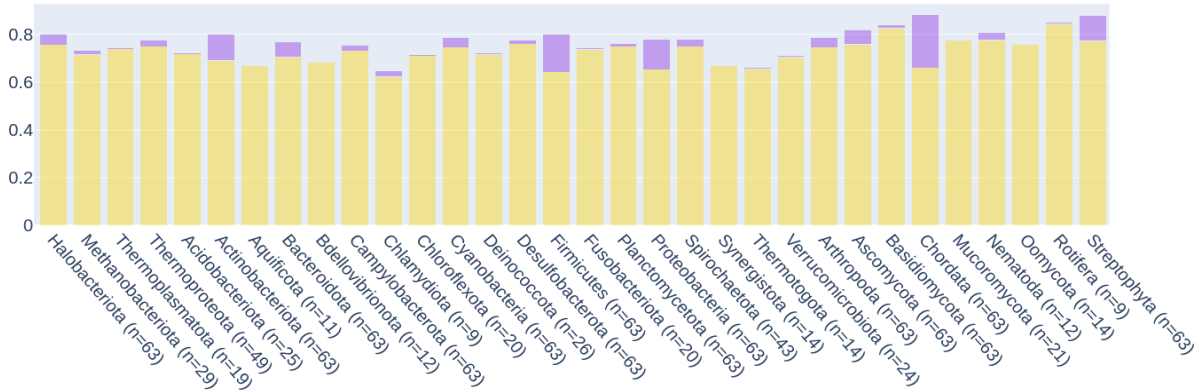


Figure 3. A) Distribution of number of single-domain architectures per organisms per phylum, normalized by number of uni-grams. **B)** Cumulative number of different single-domain architectures (yellow) and unique ones in each phylum, normalized by number of uni-grams.

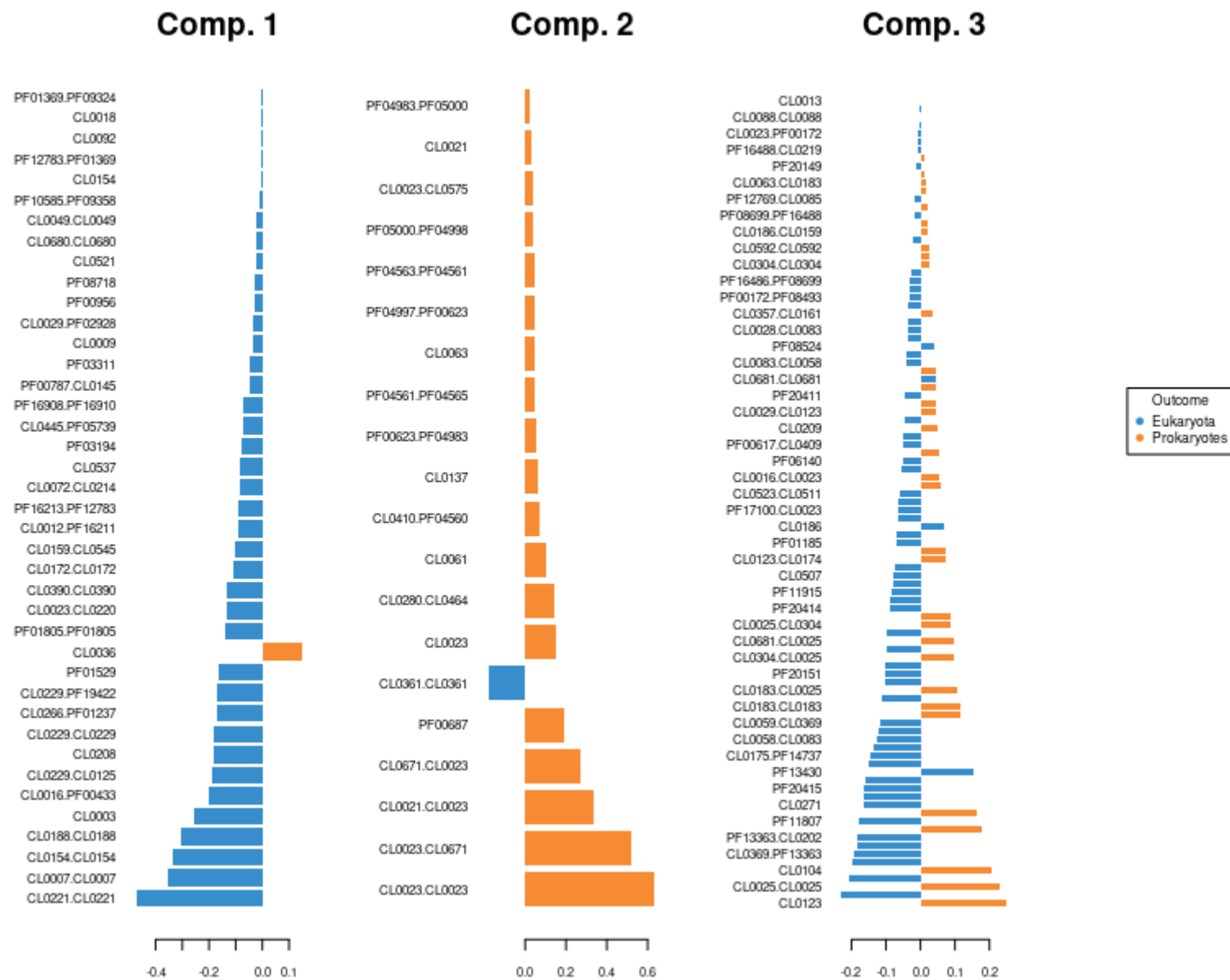


Figure 4. Variables with respective importance of each component of sPLS-DA from classifying Eukaryota and Prokaryota.

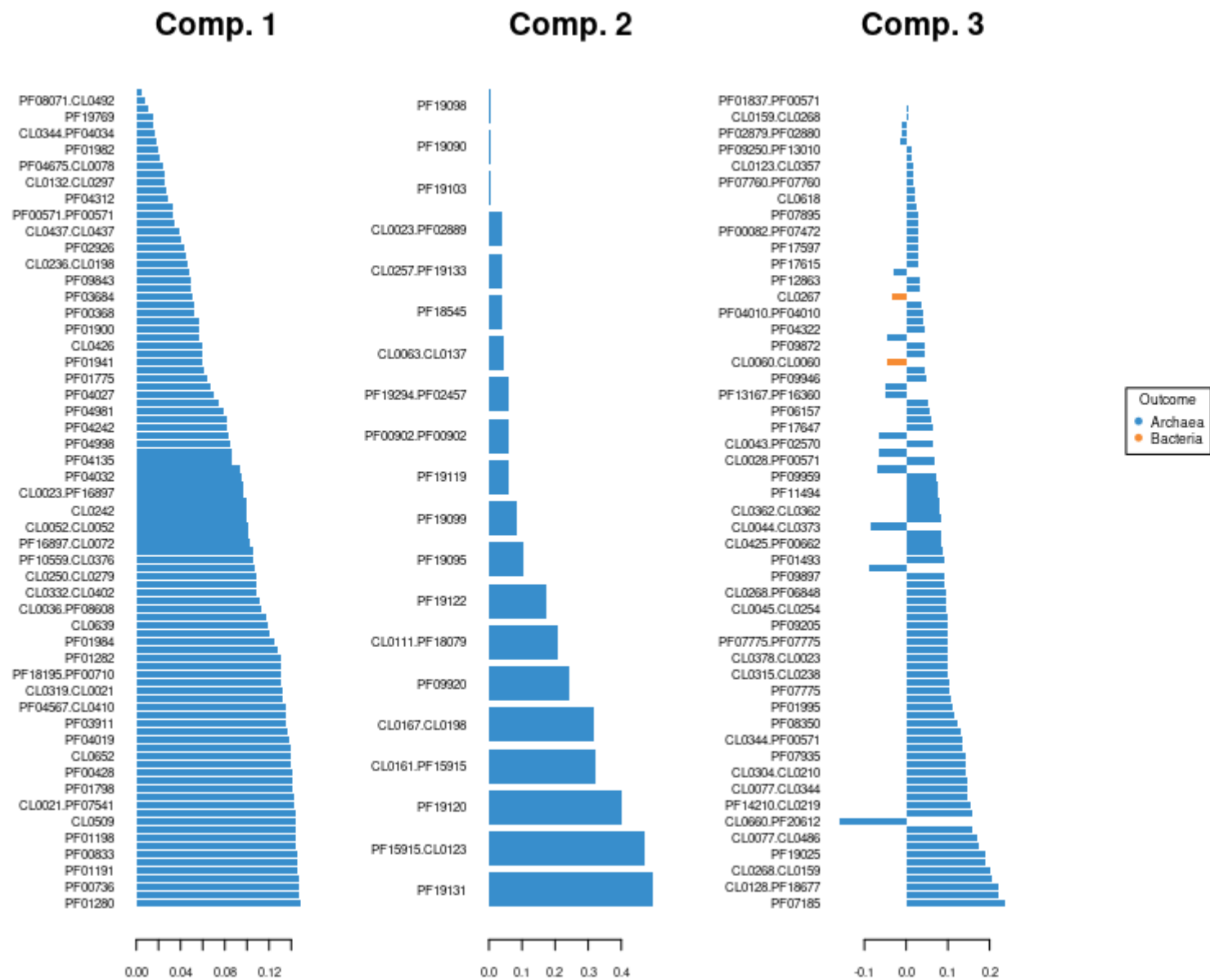


Figure 5. Variables with respective importance of each component of sPLS-DA from classifying Archaea and Bacteria.

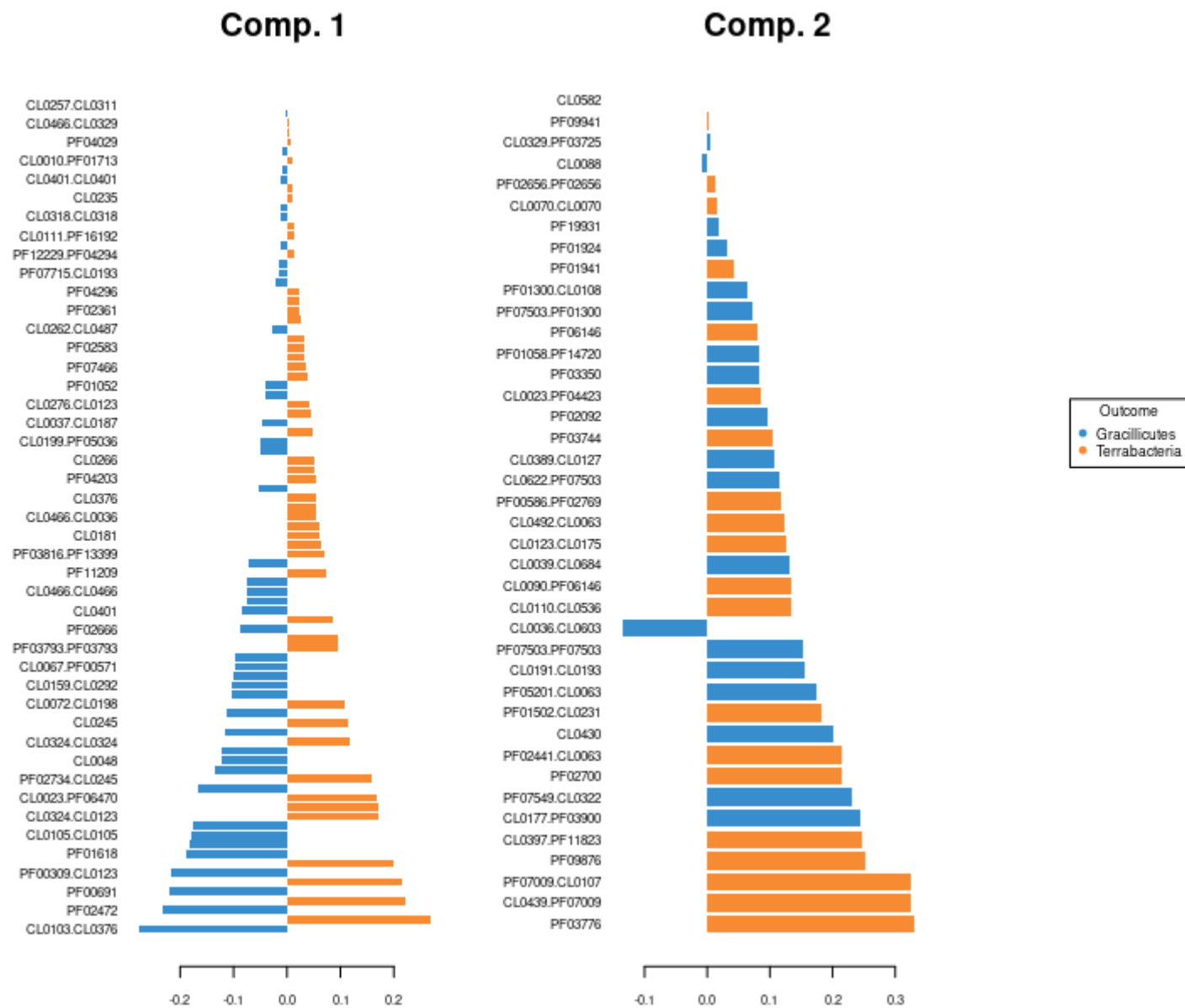


Figure 6. Variables with respective importance of each component of sPLS-DA from classifying *Gracillicutes* and *Terrabacteria*.

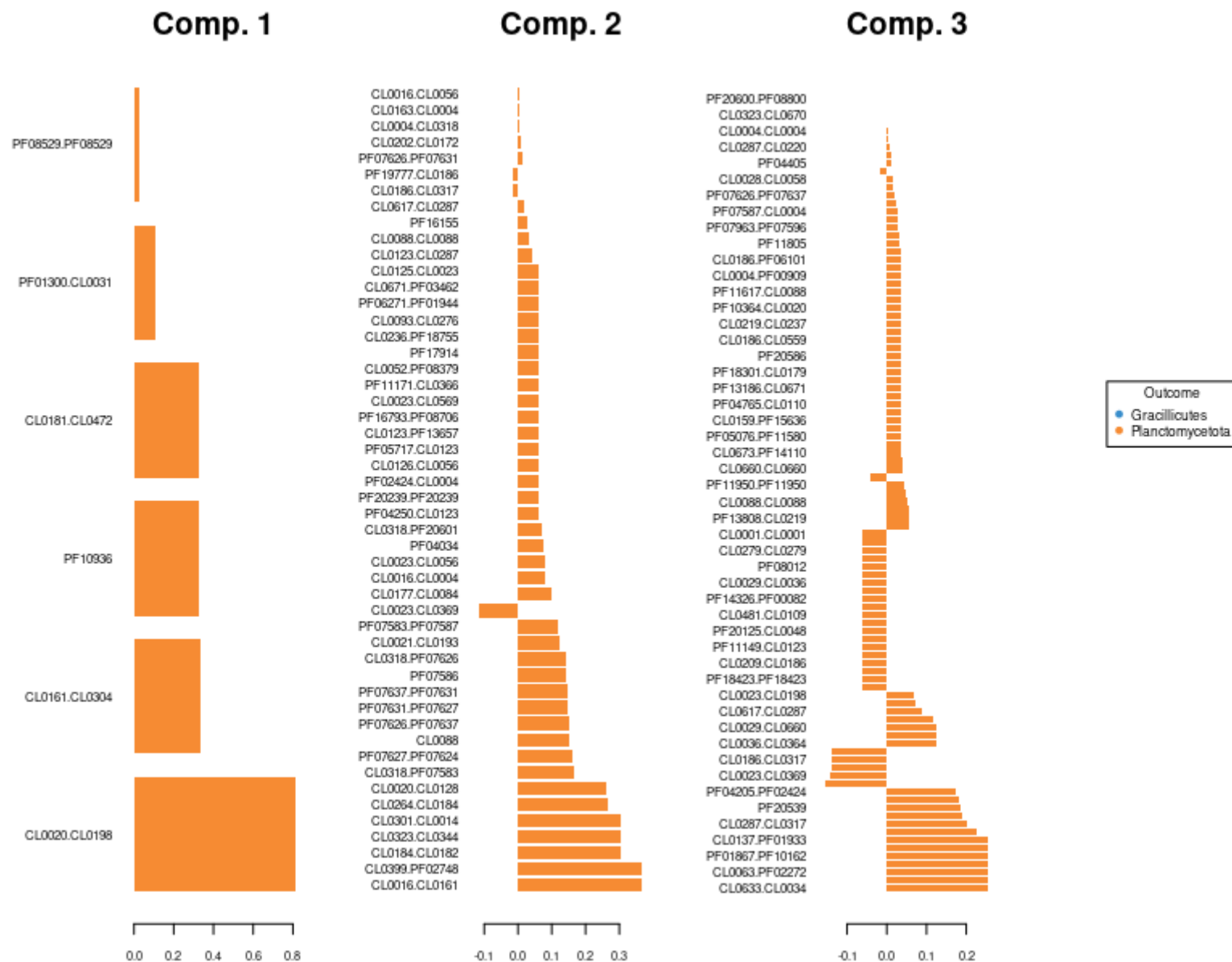


Figure 7. Variables with respective importance of each component of sPLS-DA from classifying *Planctomycetota* from the rest of *Gracillicutes*.

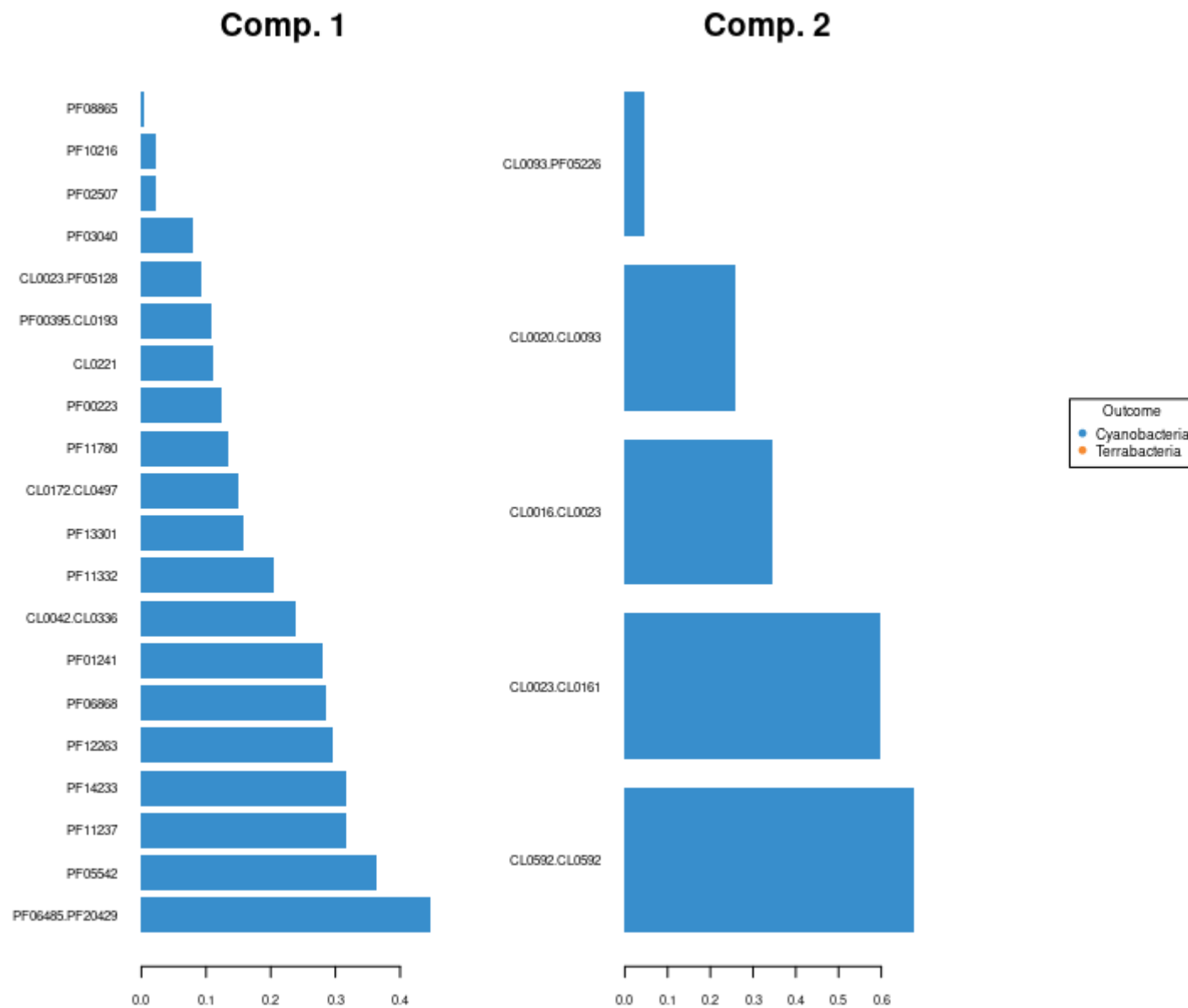


Figure 8. Variables with respective importance of each component of sPLS-DA from classifying Cyanobacteria and the rest of Terrabacteria.

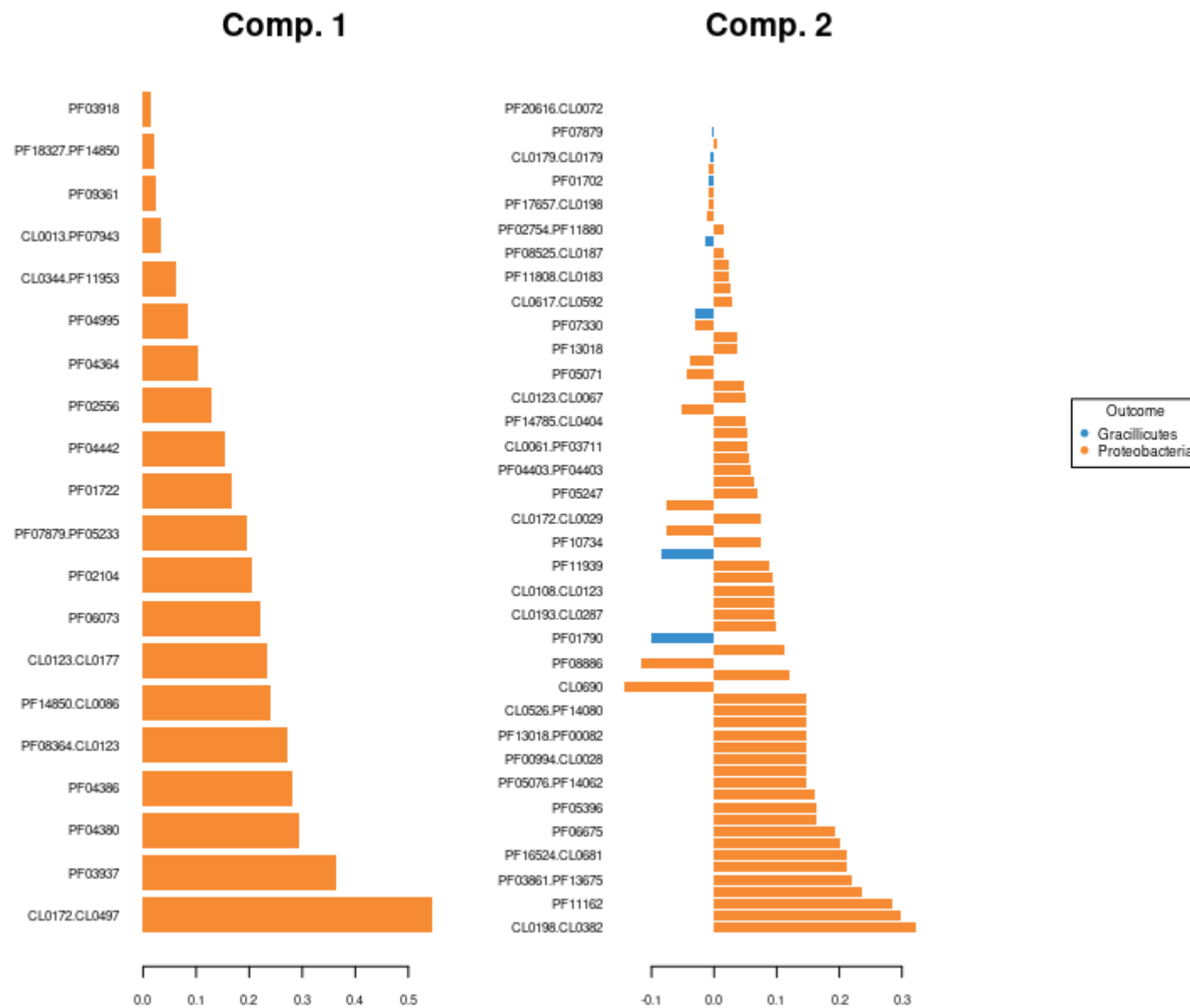


Figure 9. Variables with respective importance of each component of sPLS-DA from classifying *Proteobacteria* from the rest of *Gracillicutes*.