We downloaded two databases that differ in number of sequences and character length. The tufA dataset [13, 14] was >1900 sequences with a length of 480 bp while the 16S rDNA PhytoREF [15] was >4000 sequences with a length of 3380 bp (about 2000 sequences that were short, with introns, or of lower quality, i.e. with IUAPC characters, were removed). We colored and colored+annotated 10 phyletic clades for each datasets (Chlorarachniophyta - CYAN, Chlorophyta - GREEN, Euglenozoa - BLUE, Cryptophyta - BROWN, Rhodophyta - RED, Glaucophyta - MAGENTA, Ochrophyta - ORANGE, Haptophyta - YELLOW, Cyanophyta - BLACK, Streptophyta/Charophyta – PURPLE).

# Color

![16S](#) ![tufA](#)

# Color + Annotation

![16S](#) ![tufA](#)