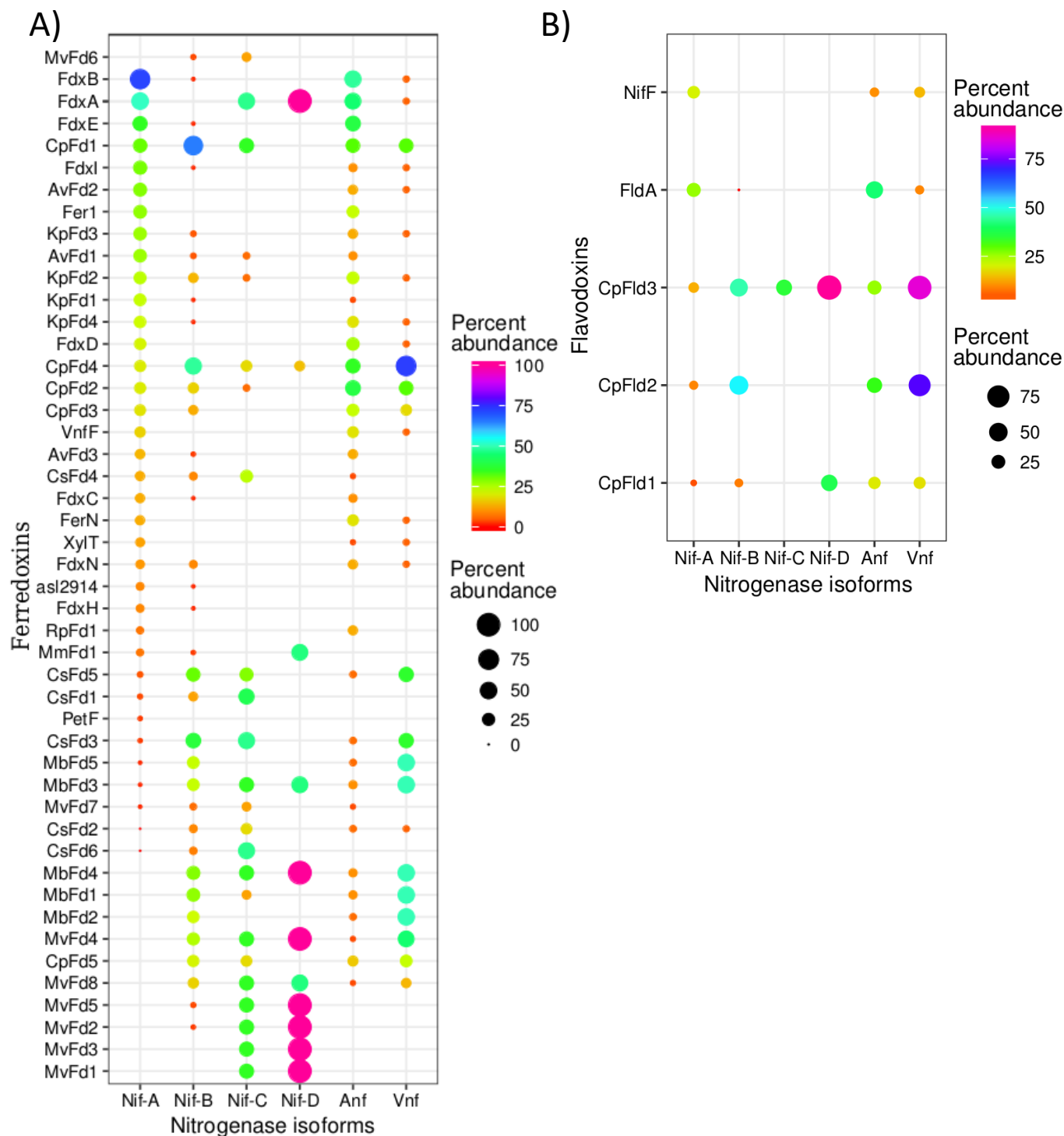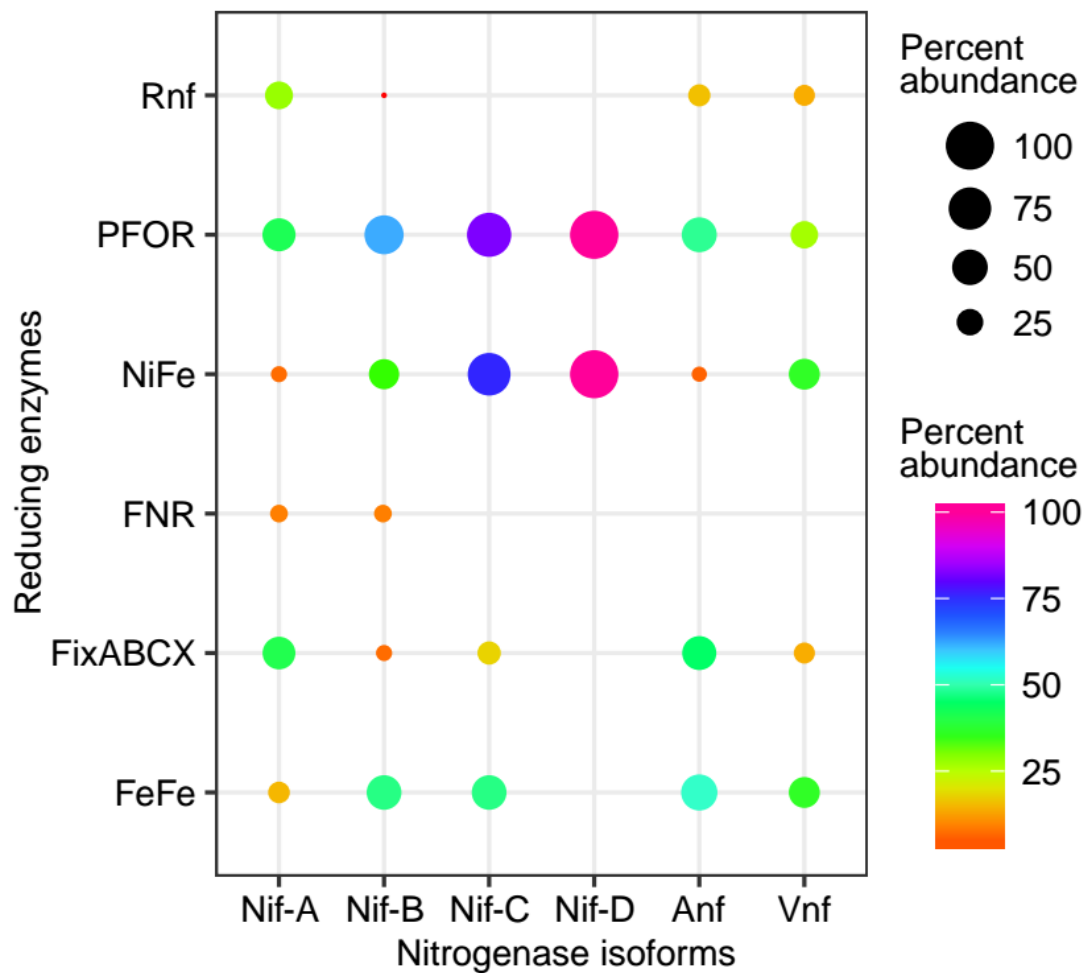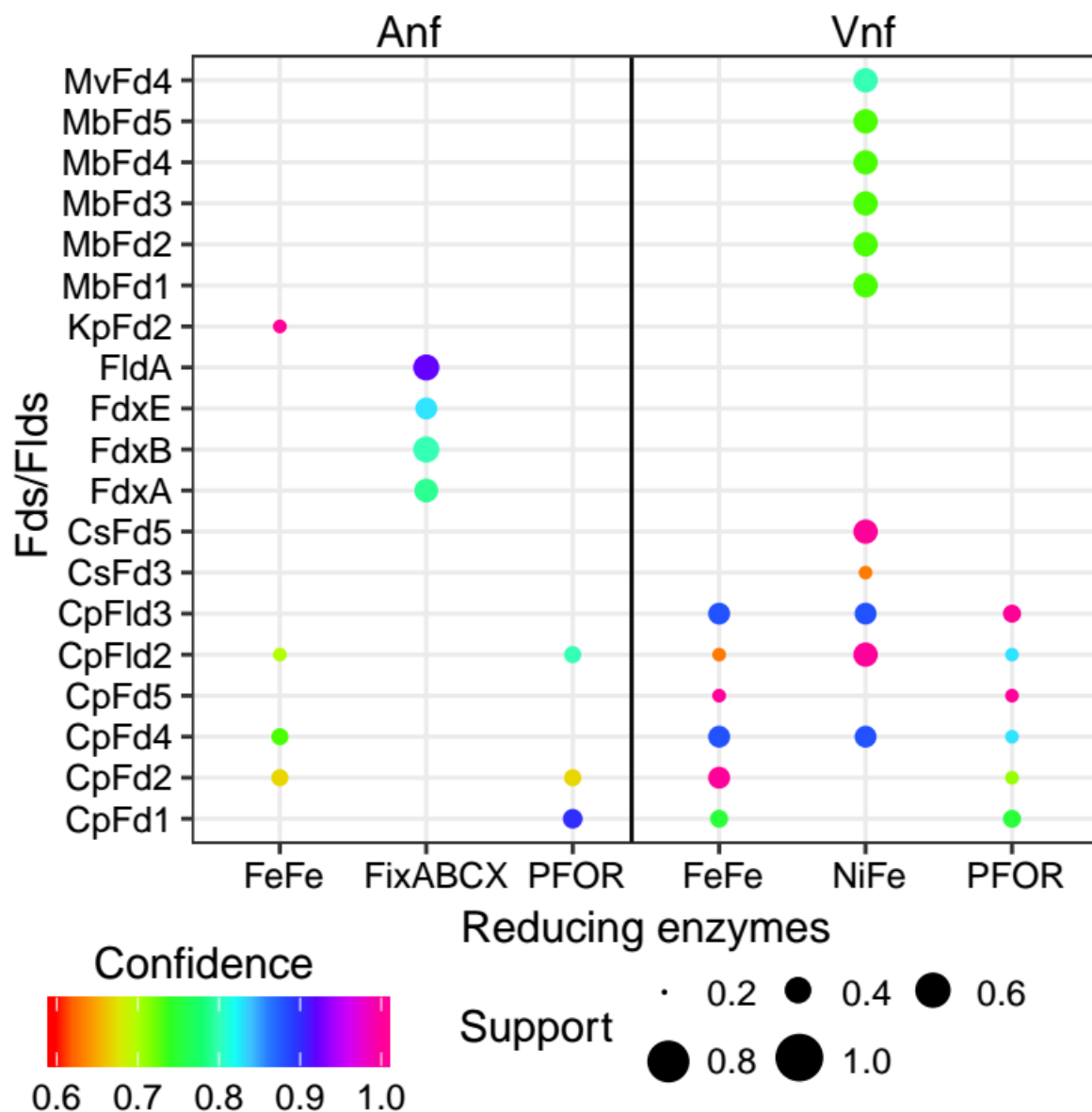**Supplemental Figure 1.** Histogram depicting the percentage of diazotrophs within each specified Nif lineage (see **Fig. 1**) that are phototrophic or chemotrophic, as determined from literature surveys of cultivated organisms.

Supplemental Figure 2. Bubble plot representing the relative abundance of specified Fds (A) and Flds (B) in diazotrophic genomes as a function of the specific lineage/isoform of nitrogenase encoded in those genomes. The size and color of the bubble represents the relative frequency of the protein (in percentage) in the genome that codes for specified isoforms of nitrogenase. Abbreviations for Fds and Flds are presented in **Tables 1** and **2**, respectively.

Supplemental Figure 3. Bubble plot depicting the co-occurrence of enzyme homologs putatively involved in reducing Fds/Flds and the specified nitrogenase isoforms. The size and color of the bubbles represents the relative frequency (in percent) of the enzyme encoded in genomes that also code for specific isoforms of nitrogenase. Abbreviations: FNR, ferredoxin-NADP$^+$ oxidoreductase; PFOR, pyruvate-flavodoxin oxidoreductase; Rnf, *Rhodobacter* nitrogen fixation protein; FeFe, iron only hydrogenase; NiFe, nickel-iron hydrogenase; FixABCX, electron transfer flavoprotein involved in nitrogen fixation.

Supplemental Figure 4. Bubble plot depicting the dominant patterns in the distribution of putative electron carrier proteins (Fds, Flds), enzyme homologs putatively involved in the reduction of Fds/Flds, and Anf and Vnf in genomes coding for specified Anf and Vnf lineages, as determined by the Apriori algorithm (1). Each unique pattern is given as a bubble and the color represents the confidence value or the statistical significance in the co-distribution of specified proteins (only confidence values of ≥ 0.6 are presented). The size of the bubble represents the support value (>=0.2) or the frequency of that two proteins are identified in the same genome. For simplicity, only the proteins that were present in >20% of the diazotrophic genomes for each specified nitrogenase lineage were considered in this analysis. Abbreviations: PFOR, pyruvate-flavodoxin oxidoreductase; Rnf, *Rhodobacter* nitrogen fixation protein; FeFe, iron-only hydrogenase; NiFe, nickel-iron hydrogenase; FixABCX, electron transfer flavoprotein that is involved in nitrogen fixation. Abbreviations for Fds and Flds are presented in **Tables 1** and **2**, respectively.
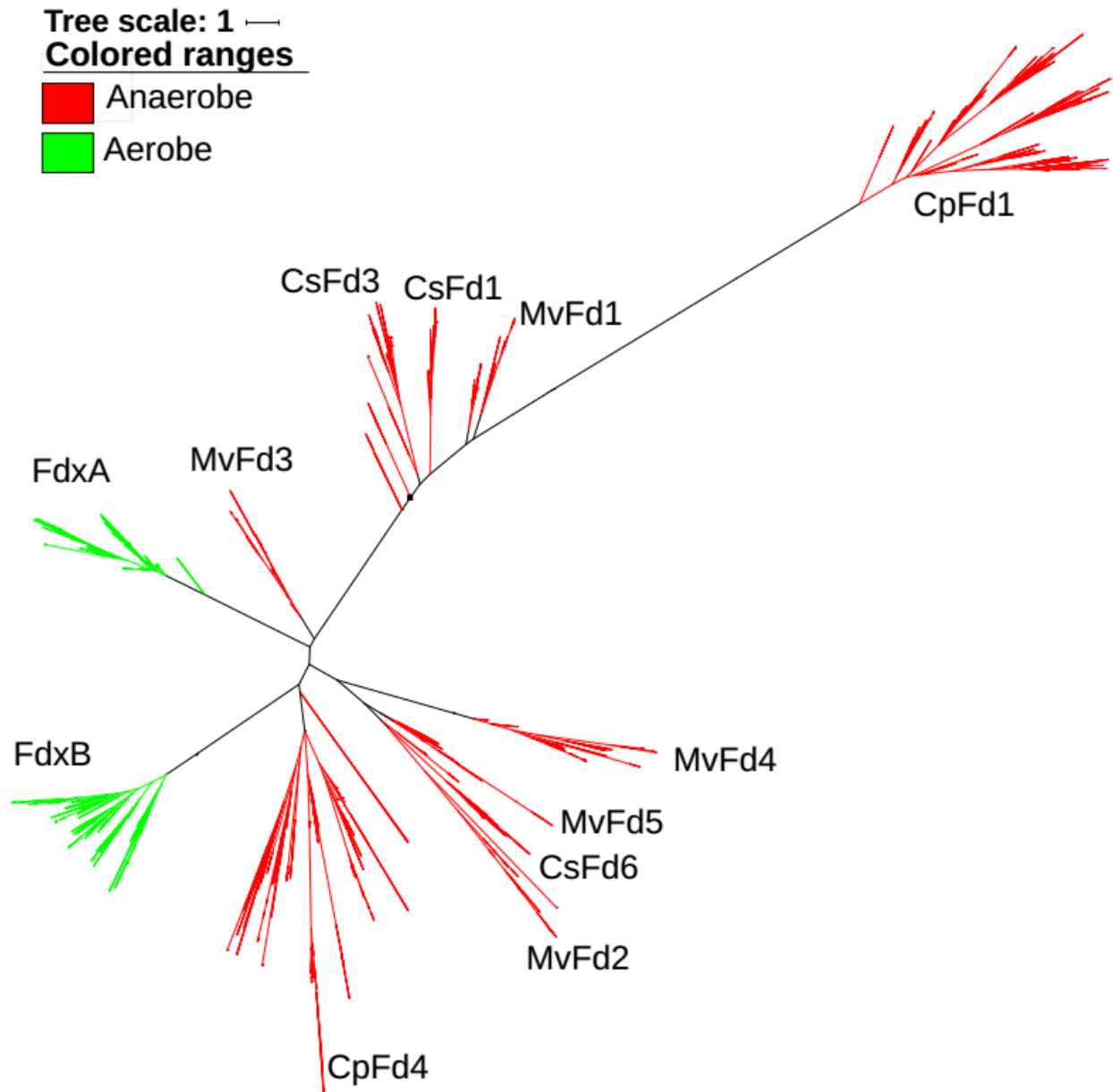
Supplemental Figure 5. Maximum likelihood phylogenetic reconstruction of abundant (present in >40% of the genomes) homologs of Fd (n=917) shown in **Fig. 3A** with an overlay of the aerobicity of the diazotrophic organism that the Fd is primarily associated with: Aerobe (green) and anaerobe (red). All nodes shown exhibited bootstraps >90% out of 1000 bootstrap replicates except where black boxes (>80%) are shown.
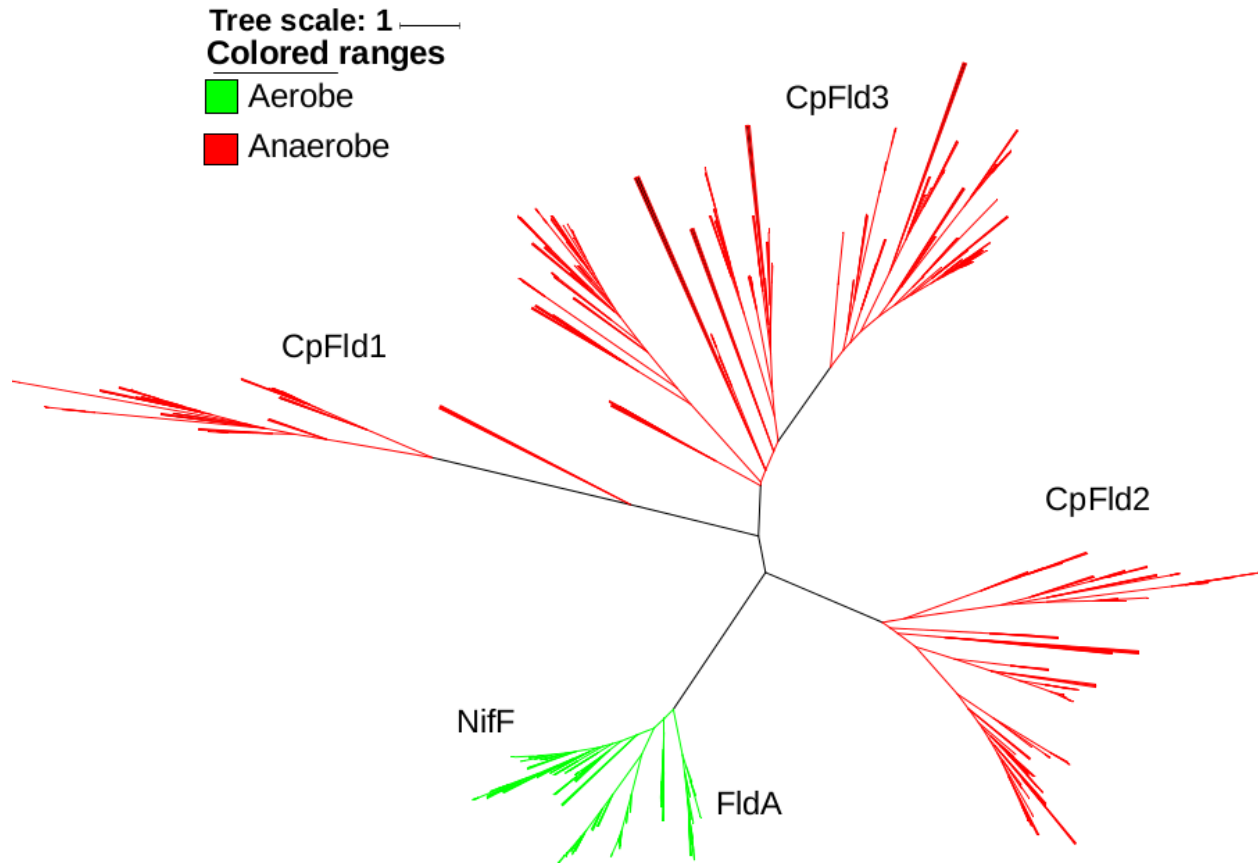
Supplemental Figure 6. Maximum likelihood phylogenetic reconstruction of 405 homologs of Flds given in **Table 2** and **Fig. 3B** with an overlay of the aerobicity of the diazotrophic organism that the Fld is primarily associated with: Aerobe (green) and anaerobe (red). All nodes shown exhibited bootstrap values >90% out of 1000 bootstrap replicates.

REFERENCES

1.   Agrawal R, Srikant R. Fast algorithms for mining association rules, p 487-499. *In* (ed),