Figure S1. The citation record of all published research articles containing reference to *Caulobacter* displayed as a network diagram. Nodes are scaled by the total number of times a paper has been cited and coloured by the number of citations of all neighbouring nodes. The most cited articles were named and labeled by whether they attributed oligotrophic (yellow) or aquatic (light blue) characteristics to *Caulobacter*, and the general area of *Caulobacter* research: cell biology (brown), genomics (purple) and review articles (burgundy).

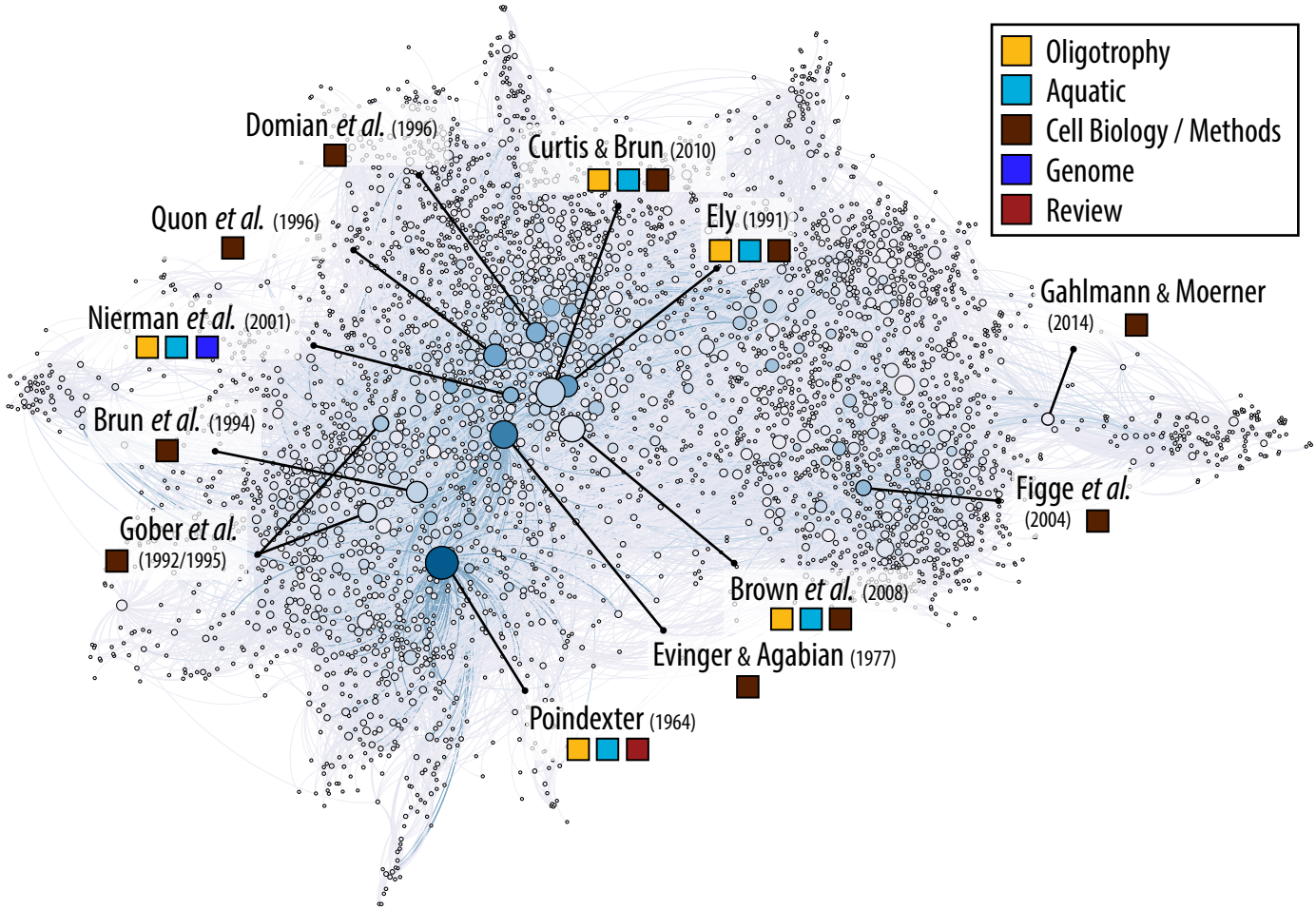

Figure S2. The relative abundance of *hfaB* in shotgun metagenome assemblies from terrestrial and aquatic sources according to geographical coordinates.

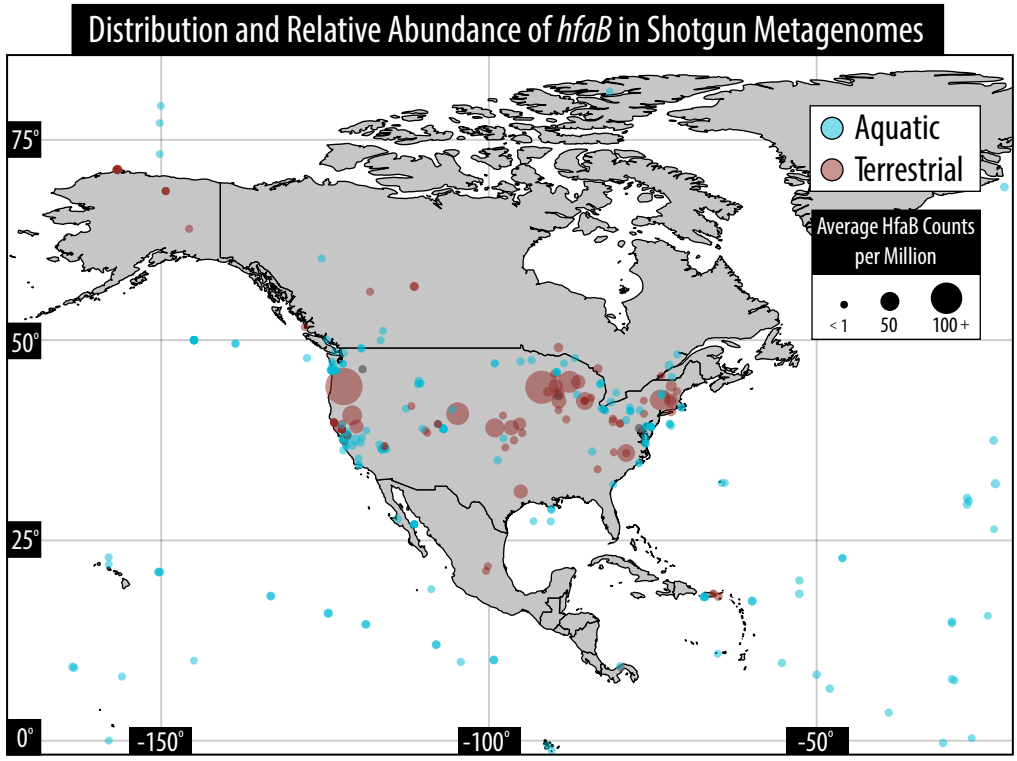

Figure S3. A maximum-likelihood phylogenetic tree based on a multi-locus sequence alignment of all *Caulobacter* genomes and their closest relatives in *Asticcacaulis*, *Phenylobacterium* and *Brevundimonas*. The tree is based on 49 highly conserved Clusters of Orthologous Groups (COG) families. For methodological details consult the Supplementary Methods.

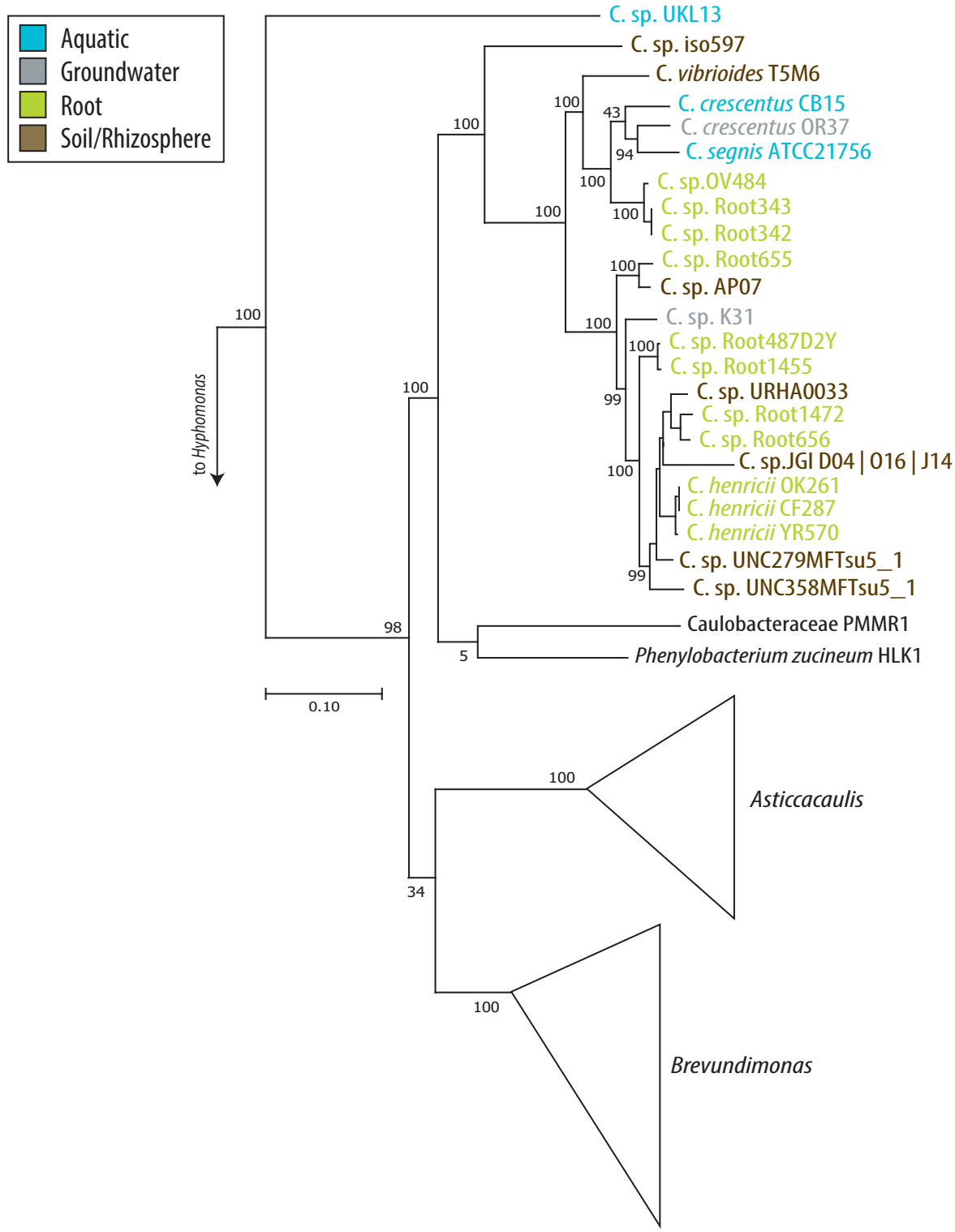

Figure S4. Complete-linkage clustering of *Caulobacter* genomes based on the Bray-Curtis dissimilarity of carbohydrate-active gene content.

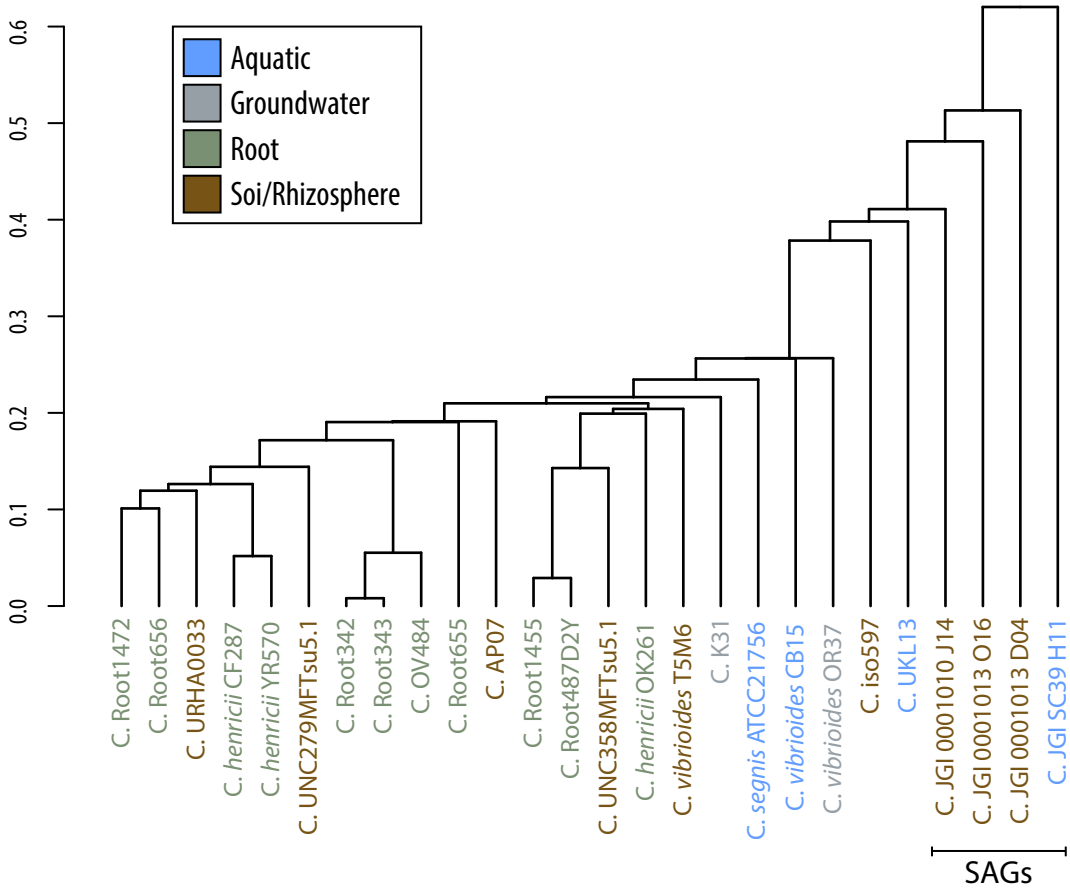

Figure S5. A clustered heatmap showing the similarity of *Caulobacter* genomes (x-axis) based on signatures of oligotrophy and copiotrophy (i.e. a group of or an individual COG) as defined by Lauro *et al*., (2009). Genomes were designated above (red) or below (white) the median for each trait, according to the methods of Lauro *et al*., (2009). The generalized gene content associated with oligotrophy (aqua-marine) or copiotrophy (purple) are shown.

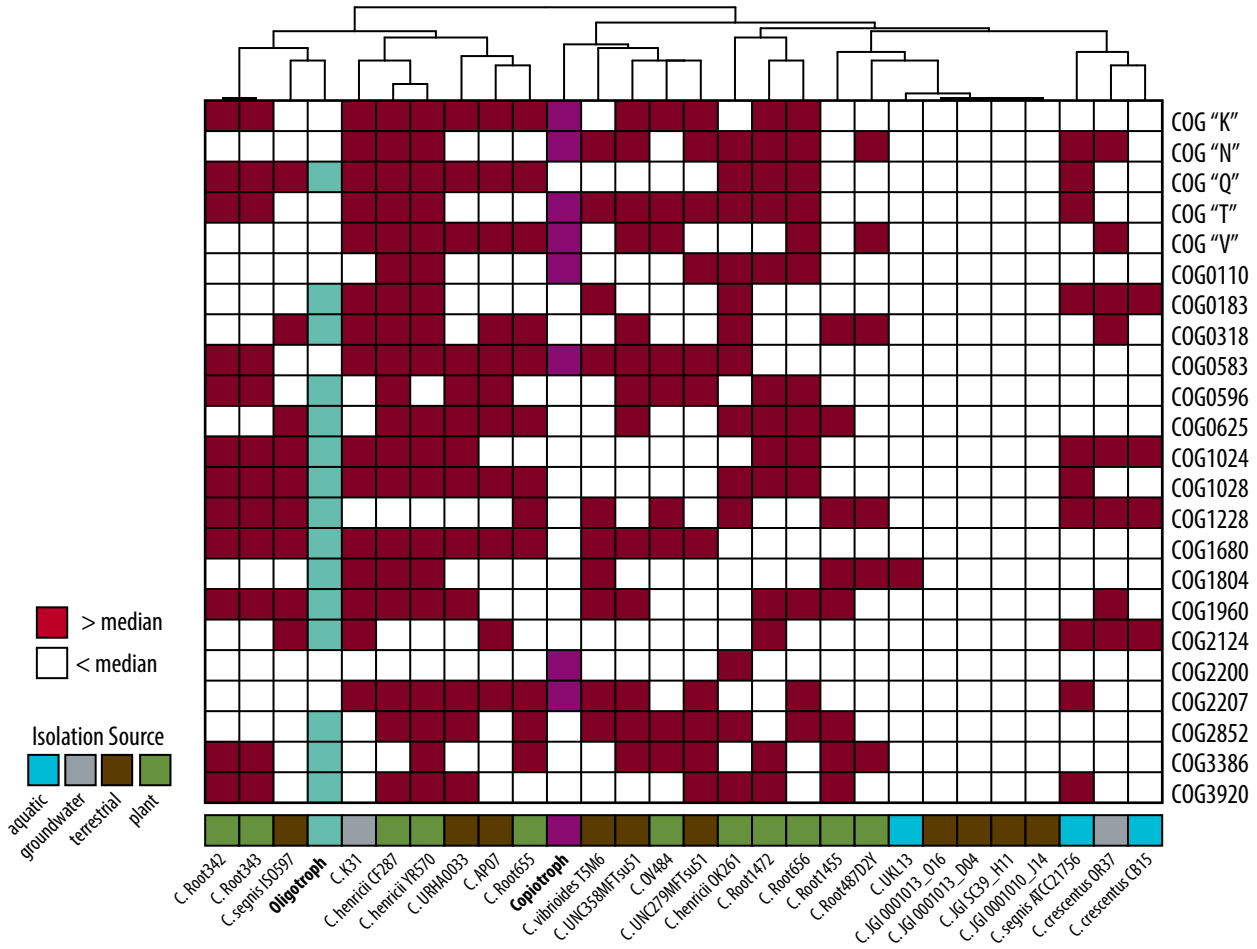

Figure S6. Maximum-likelihood phylogenetic tree of *lovK*, a photo-responsive histidine kinase, encoded in *Caulobacter* genomes. Twenty-three of the twenty-six genomes encoded the gene, missing in two SAGs and one isolate genome (groundwater isolate *C sp.* UKL13). The sequence alignment was based on 370 amino acids and branches were supported by 250 bootstrap permutations.

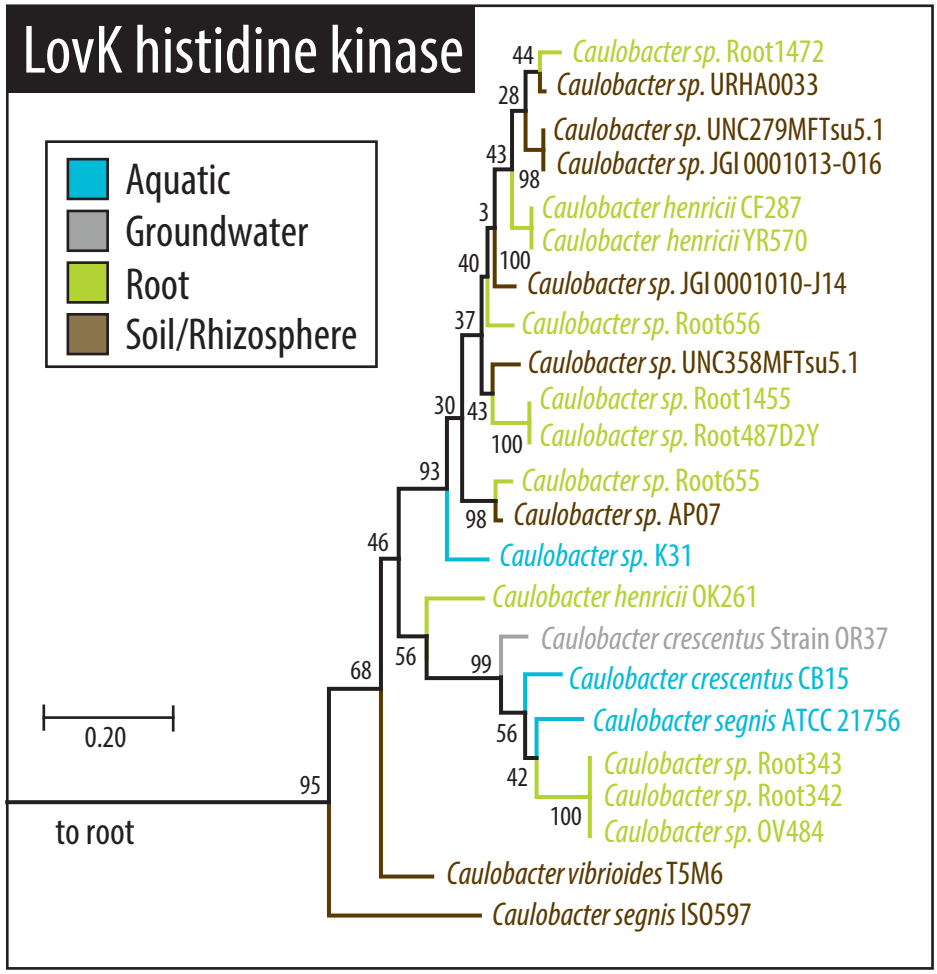

Figure S7. A series of case studies that demonstrate the associations of *Caulobacter* with (a) the capacity to decompose a variety of organic compounds, (b) the habitat distribution of near 100% identical *Caulobacter* sequences along a slope from forest soils to adjacent stream and ocean water and (c) a sensitivity to soil drying. Panel (a) displays data from a stable isotope probing experiment (Youngblut *et al.*, 2018), in which sequencing was performed on ^{13}C-enriched DNA from organisms which incorporated carbon from ^{13}C-labeled compounds. The differential abundance of *Caulobacter* "OTU69" in 'heavy' DNA pools (post CsCl density centrifugation) derived from soil microcosms fed every substrates without ^{13}C-label (i.e. ^{12}C-control) versus those amended with substrates containing only ^{13}C. Panel (b) shows the rank abundance of *Caulobacter* sequences in neighbouring soil (forest and wetland) and aquatic (stream and ocean) environments to assess the predominance of species overlap (Kellog *et al.*, in preparation). No OTU clustering was performed on sequences, though pre-clustered was performed by mothur ('pre.cluster') to reduce sequencing error; thus, each sequence may be representative of aggregate of sequences that differ by 1-2 bp (~ 0.5% of 250 bp trimmed reads). Panel (c) demonstrates the significant reduction in relative abundance of *Caulobacter* in dried agricultural soil (*t*-test; BioProject: PRJNA347493). Two additional studies in the metagenome collection assessed the influence of moisture regimes, one moderating precipitation (PRJNA243310) and one soil wetting (PRJNA176825), but failed to provide sufficient metadata to be of use.