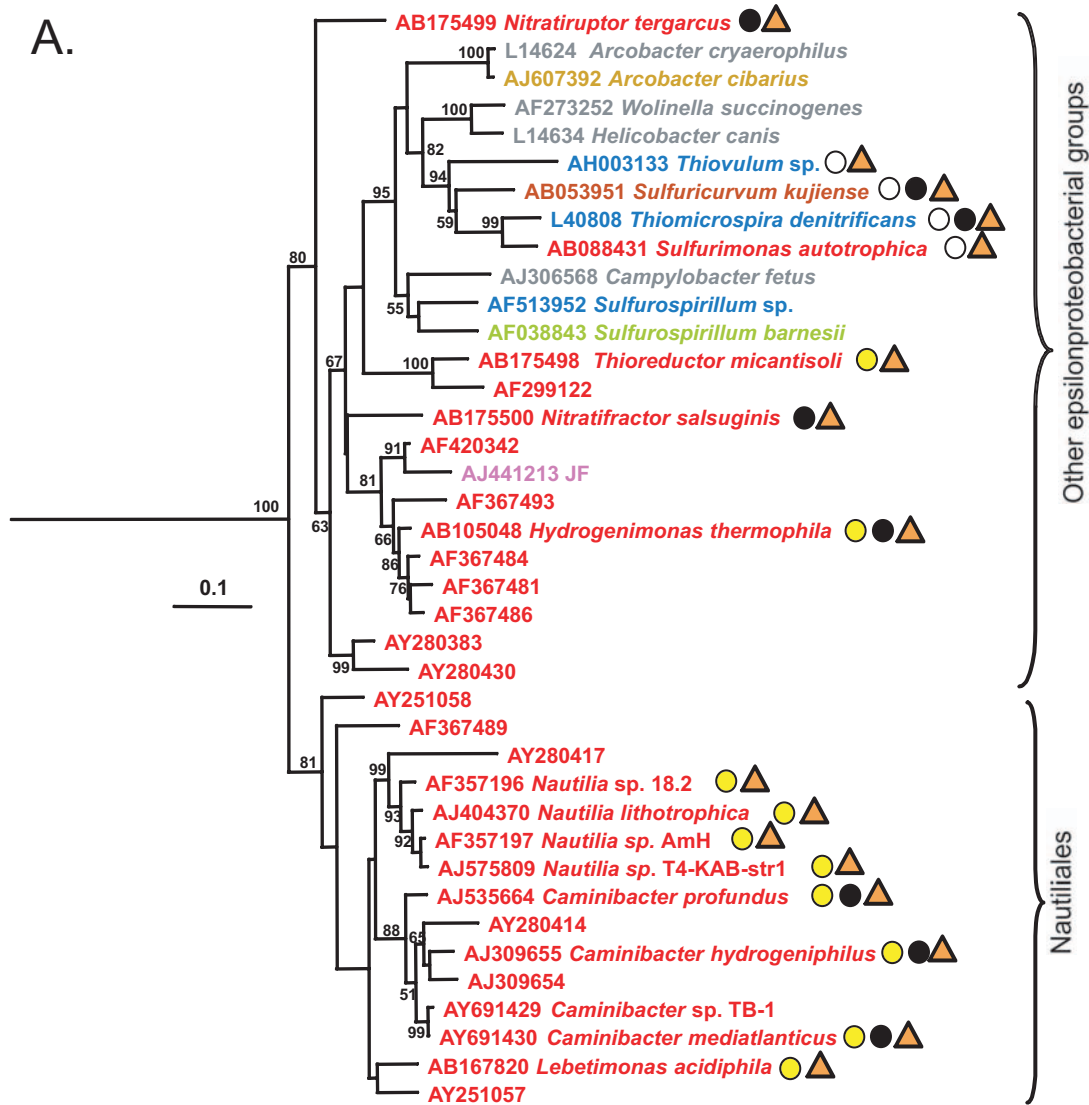
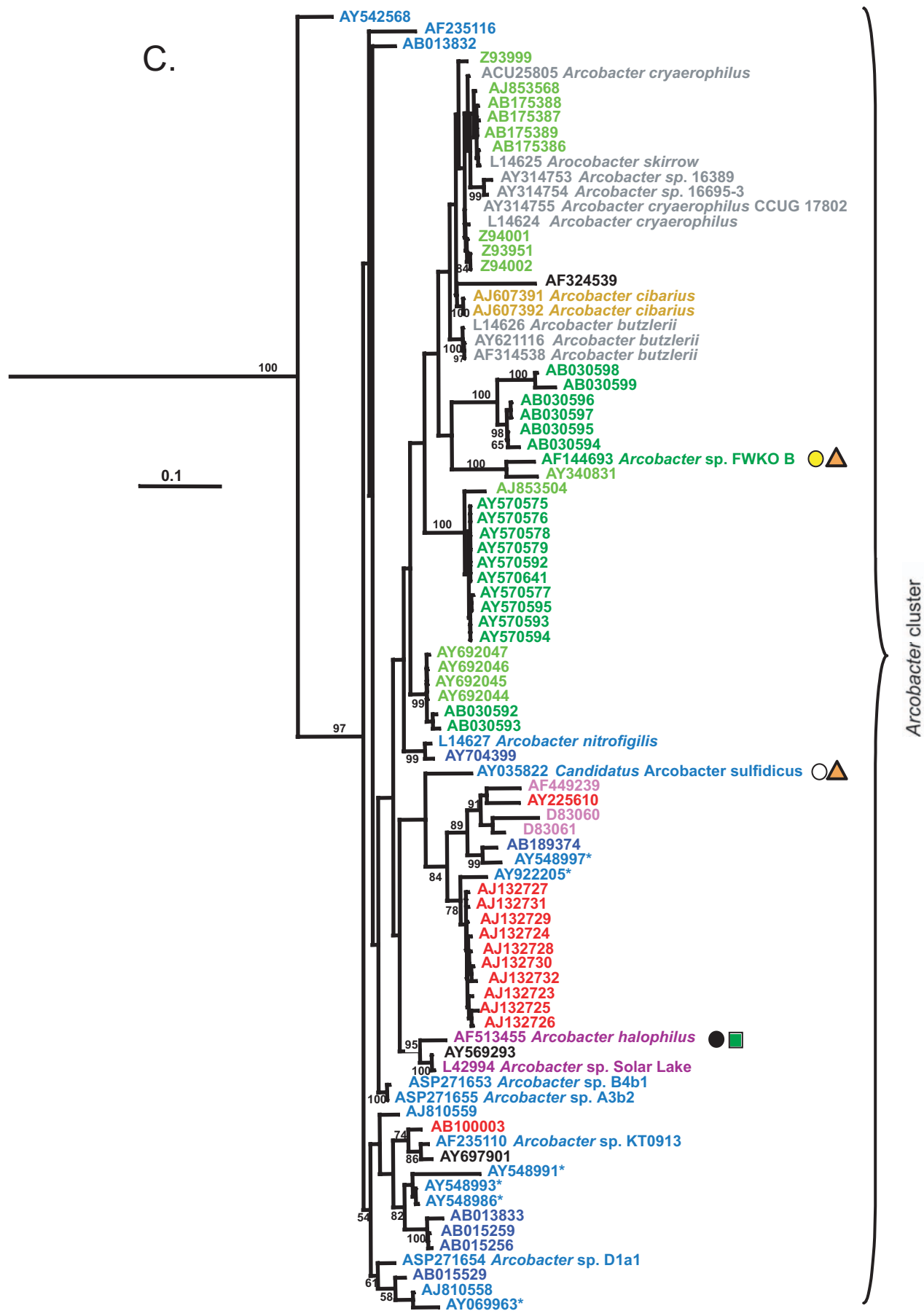


S2 (figure) | Maximum likelihood analyses of the four main phylogenetic clusters of environmental sequences as determined in FIG. 1

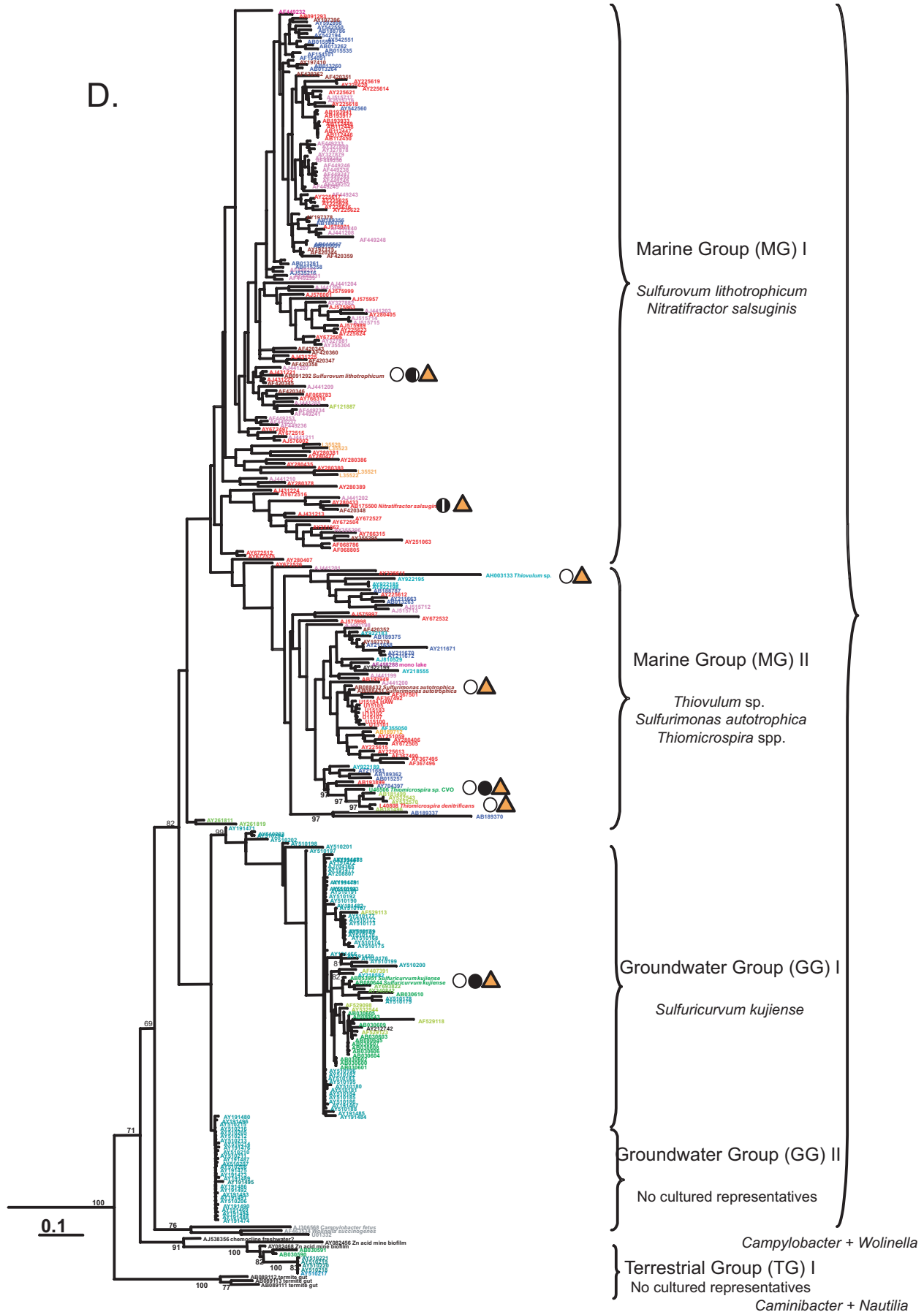


B.





D.



## Supplemental Figure 2a-d Legend

<u>Habitat Key</u>	<u>Metabolic Key</u>
red=hydrothermal vent-associated	▲ Autotrophy
orange=symbiont	■ Heterotrophy
purple=vent fauna associated	● Nitrate reduction
blue=marine/marine sediments	● Sulfur reduction
dk. blue=deep sea sediments	○ Sulfur oxidation
dk. purple=hypersaline water	◇ Alternate e- acceptors
lime=groundwater (+/- contaminated)	
green=oil-contaminated groundwater	
dk. brown = mud	
bright green=bioreactor	
teal = cave	
black=unknown	
gray= mammalian-associated	
brown=chicken-associated	

**S2 (figure) | Maximum likelihood analyses of the four main phylogenetic clusters of environmental sequences as determined in FIG. 1** a) deeply branching lineages, including the Nautiliales and Hydrogenimonaceae; b) *Sulfurospirillum* clade; c) *Arcobacter* clade; d) provisional Thiovolgaceae fam. nov. clade. Trees were reconstructed using PHYML<sup>8</sup>, using *Thermotoga maritima* (AJ401017) and *Aquifex aeolicus* (AJ309733) as outgroups for all analyses. Numbers on branches represent bootstrap proportions from 100 replicates. Taxa labels are Genbank accession numbers and are coloured according to habitat (see key): Symbols following taxa labels indicate metabolic capabilities (see key).