

Supporting Figure S2. Phylogenetic trees showing abundance of clustered archaeal 16S rRNA sequences from (A) unassembled reads and (B) PCR-amplified clone libraries. A maximum likelihood archaeal reference tree was constructed using FastTree [1], based on full-length 16S genes from isolate genomes and environmental clones from Genbank nt, supplemented with sequences obtained from Lake Tyrrell assembled scaffolds. Unassembled 16S rRNA sequences from Lake Tyrrell were clustered and inserted into the reference tree using pplacer version v1.1 (model GTR, fig-eval-all) [2] and visualized using Archaeopteryx 0.968 [3]. Part A shows placement of un-amplified Lake Tyrrell raw reads containing 16S gene sequences > 350 nt. Part B shows placement of PCR-amplified 16S rRNA clones (> nt). Numbers at nodes indicate confidence values estimated by FastTree for the reference database. Red lines indicate branches where Lake Tyrrell sequences were observed. The thickness of each red line is proportional to the number of Lake Tyrrell sequences associated with that branch, ranging from one in the thinnest line to 74 in the thickest line.

References

1. Price, M.N., P.S. Dehal, and A.P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One, 2010. 5(3): p. e9490.
2. Matsen, F., A., Kodner, R., Armbrust, E.V. (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences on a fixed reference tree. BMC Bioinformatics, 11:538
3. Han M.V. and Zmasek C.M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics, 10:356.

A



