

Figure S1. Maximum likelihood phylogenetic tree based on 16S rRNA sequences.

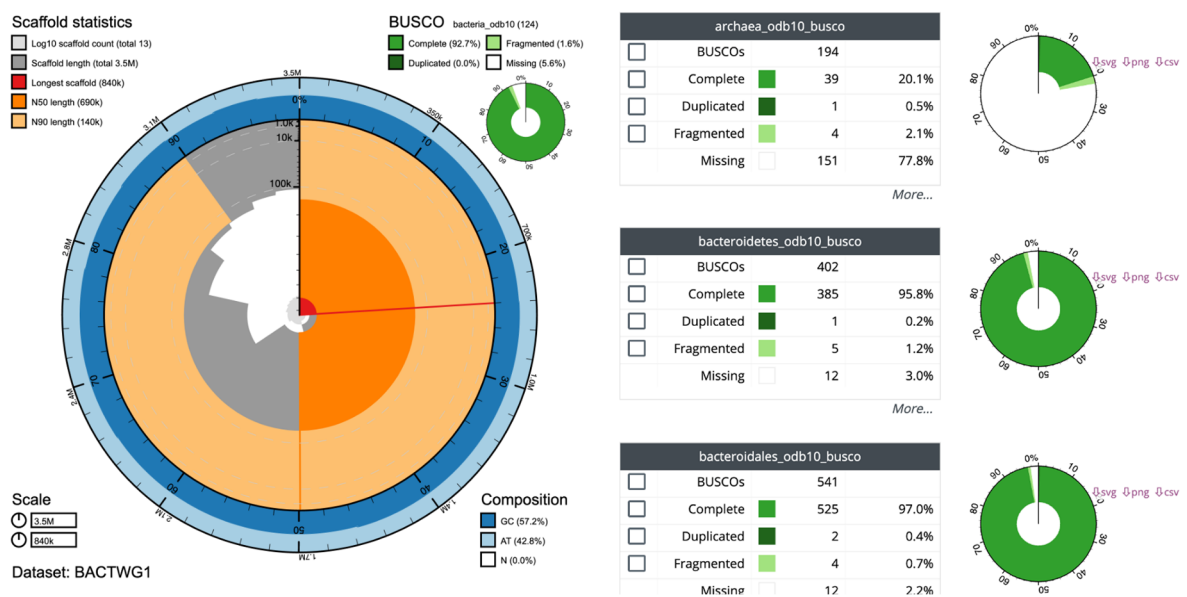


Figure S2. Snail plot summary of assembly statistics for the genome assembly. (A) The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 3 484 450 bp assembly. The distribution of record lengths is shown in dark grey with the plot radius scaled to the longest record present in the assembly (839 650 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 record lengths (691 467 and 139 949 bp), respectively. The pale grey spiral shows the cumulative record count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the bacteria_odb10 set is shown in the top right. (B) A summary of complete, fragmented, duplicated and missing BUSCO genes in the archaea_odb10, bacteroidetes_odb10 and bacteroidales_odb10 set.