Supplementary Information S2 (box) | Metagenome Analysis of Terminal Oxidase Genes

- **Using MG-RAST, we selected each metagenome and used the analysis function to generate a data table using subsystems ontology and a cutoff of e=1e-05.**
- **We used this data table to enumerate each gene of interest (Table 2) and a set of housekeeping genes.** Our housekeeping set was *rpoA, rpoB, rpoC,* and *recA*. These genes were chosen because they are specifically annotated as bacterial housekeeping genes in MG-RAST. Oxidase gene counts attributable to archaea were excluded from our analysis using BLAT alignments to judge phylogeny.
- **The genes were then normalized to their gene length.** The length of the genes we used in our analysis ranged from 984-4059 base pairs. Gene length may influence the number of sequences that are found in a metagenome, so we divided each of the counts by the average length of the gene to account for the influence of gene length on the number of genes detected.
- **The mean of the normalized housekeeping gene values was then calculated.** Theoretically, each genome should have one copy of each housekeeping gene, so the mean of the housekeeping gene set is used as an estimate of the number of genomes/metagenome. In order to determine that this estimate was as robust as possible, the coefficient of variation for the normalized housekeeping gene counts was calculated, and metagenomes with CV >15% were excluded from this part of the study.
- **Then to estimate the relative abundance of high affinity oxidases, gene counts for terminal oxidases with both high and low affinities for oxygen were then normalized to the mean of the set of bacterial housekeeping genes to obtain the number of oxidase genes/average genome which is shown in Figures 2 and Figure 3.**