Description:

For taxonomic categories shared between groups of samples with very low abundances, there is a probability that their recovery, at any abundance, might not be possible if the sampling and sequencing process was re-performed. Thus, it is necessary to take into account, not only the abundance of taxonomic categories determined, but also the variation across samples from within the same group. Because the criteria for recovery is only to detect its presence at a non-zero abundance, the probability of recovering a taxon may be considerably greater than its proportion, when the taxon's observed count is moderate and its within group sampling variance is low. To compute probability-based set overlap between groups, bootstrapping was performed using resampling with replacement for both the samples in each group and the reads acquired, to take into account inter-sample and intra-sample variation, respectively. Across the bootstrap iterations, the number of set relationships (Shallow only, Deep only, Shallow and Deep, neither Shallow nor Deep) that was assigned for each taxonomic category between groups was accumulated. The number of each of the set relationships was divided by the number of bootstrap iterations. Resultant values represent the probability of that set relationship being declared upon re-performing the sampling and sequencing process for each taxonomic category.

Figure Caption

**Fig. S7.** **Probability-based set relationship analysis of indigenous bacterial communities at Site C0020.** Due to the low abundances of reads in many taxonomic categories it is possible that if sampling and sequencing of Shallow (0-0.365 kmbsf, CK0-06) and Deep (1.2-2.4 kmbsf, Expedition 337) sediment samples were repeated, some categories might not be recovered. Therefore, to account for not only the abundance of taxonomic categories recovered but also the variation across samples, inter-sample and intra-sample probability-based set overlaps between sets (Shallow, Deep, Both or Neither) was estimated with bootstrapping by using resampling with replacement both the samples in each group and the reads acquired. In **panel A**, the length of the colored proportion of each bar represents the probability that a set relationship would be re-established upon re-performing the experiment. The colors red, blue, purple, and black, represent the probability of recovering the taxon in the Shallow Only, Deep Only, Both, or Neither samples, respectively. For example, in spite of the significantly higher abundance of Chloroflexi in the Shallow samples compared to in the Deep samples, there was still a consistent and high enough abundance of that taxon in both groups for it to be recovered mutually with a probability of 99.6%. The abundance of Verrucomicrobia was relatively low in both sample groups, resulting in only a 32.4% chance that the taxon would be recovered in both groups, and 17.5% chance it would not be recovered in either group. **Panel B**, represents the overall proportion of phyla (out of 50), and their set relationships, assuming that the recovery category selected for each taxon was its most likely set assignment. For example, Verrucomicrobia would contribute to the Shallow only counts overall, because it's Shallow Only chances (34.5%) was greater than Both (32.4%), Deep Only (15.7%), and Neither (17.5%).