Mitchell Sogin
Josephine Bay Paul Center for Comparative Molecular Biology and Evolution
Marine Biological Laboratory
sogin@mbl.edu

Who are we overlooking in microbial diversity surveys and does it really matter?

Microbial communities exhibit long-tail distributions in which very high-frequency populations precede low-frequency populations that gradually "tail off" asymptotically. For microbes and other biological forms, the long tail corresponds to a "rare biosphere" composed of many low abundance taxa. The International Census of Marine Microbes first described a "rare biosphere" that accounts for an estimated 500,000 kinds of microbes contained within less than one part out of $10^{18}$ parts of the world's oceans. New algorithms capable of discriminating between closely related taxa show that the sheer size of the microbial "rare biosphere" may be much greater than previously reported. The recently introduced techniques of oligotyping and minimum entropy decomposition permit the description of microbial diversity at finer scales for closely related but subtly distinct 16S rRNA gene amplicon sequences that represent distinct genomes in a microbial community. These analytical paradigms take advantage of Shannon entropy calculations to identify information rich nucleotide positions that differentiate between closely related taxa within the same genus. When applied to studies of microbial communities in aquatic, marine and animal microbiomes, we detect increased diversity and identify ecologically important differences between closely related taxa. These analyses of next generation DNA sequencing data sets reveal the rare biosphere has temporal and spatial dimensions that impact our perceptions of microbial ecology and how anthropogenic activities influence human health and well-being. Some members of the rare biosphere might always represent low-abundance populations. Others might normally persist in very low numbers but have the capacity to become more abundant in response to environmental change. Yet other members of the rare biosphere might disperse over long distances from their yet to be discovered endemic sites. The extraordinary length of the long tail is consistent with a model where large numbers of very rare and highly divergent taxa stably coexist because they do not compete for niche space.