Metagenomic Analysis of Uncultured Microorganisms

John F. Heidelberg, Ph.D.
The Institute for Genomic Research (TIGR), USA

Recently, one could imagine that the genes that had been discovered in cultured microorganisms embodied a representation of the total microbial genomic potential. However, complete genome projects have shown us what a small part of this world was really known to us, and how much of this world remains to be explored. The group of organisms that have perhaps the least explored genomes are the uncultured microorganisms; however, contained within these organisms are likely many hitherto unsuspected biogeochemical relevant pathways that could be found. To date, studies of uncultured bacteria have been primarily limited to analysis based on 16S rRNA, and these methods have greatly increased our knowledge of the phylogenetic diversity within many ecosystems. However, such studies do not allow accurate determination of the functional niche these microorganisms occupy.

The "metagenome" is the combined genetic information of an entire ecosystem. Currently, "metagenomic" analysis is one of the most exciting future steps for genomics. These types of study confer the ability to examine the biogeochemical capabilities of uncultured bacteria in a culture independent manner and have found several examples of such discoveries within uncultured organisms are now being published (e.g., rhodopsins, antibiotics, etc). Currently, typical metagenomics involves isolation of DNA from an environmental sample (eDNA) and cloning this eDNA into a vector. Once in the vector, the clones are either screened for specific genes (e.g., rRNA, or a specific function) and interesting clones sequenced entirely or the clones are subjected to high throughput random end sequencing and assembly. Therefore, we can examine the physiological properties of entire populations of bacteria by interrogating this eDNA.

We have applied "whole genome shotgun sequencing" to eDNA collected in the Sargasso Sea near Bermuda. A total of 1.045 billion basepairs of non-redundant sequence was generated, annotated and analyzed to elucidate the gene content, and diversity and relative abundance of the organisms within these environmental samples. We have identified over 1.2 million new genes represented in these samples, including more than 782 new rhodopsin-like photoreceptors. Additionally, variation clone frequency stoichiometry suggests substantial patchy disruption oceanic microorganisms. From these analysis, we believe there is considerable merit in continued metagenomic studies.