Welcome Address and Introduction to Okazaki Biology Conferences "Genomic Insights into Terra Microbiology"

James M. Tiedje, Kostas Konstantinidis, Jae Chang Cho, Alban Ramette, James Cole (Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA)

Terra Microbiology is possible because of new insights from genomics and other molecular methods, new understanding about the extent and novelty of microbial diversity, general recognition of the vital role that terra microbes play in sustaining a livable Earth and the rapid development of DNA sequencing and other high throughput experimental and computational methods that make it possible to explore Earth's oldest and most extensive biological entity. Microbial communities, which carry out the biogeochemical processes of



nature, are made up of populations (often assumed to be species) which in turn are made up of ecotypes, and then of genotypes. To understand the composition and function of communities we must also understand the diversity within species since many functional differences in populations can occur at this level. We have used the greater than 150 sequenced microbial genomes to explore how much variation in gene content occurs among closely related organisms since this indicates how similar their phenotypes should be. Using average nucleotide identity (ANI) of the conserved genes between any pairs of related strains as a measure of evolutionary distance, we found that members of some species may vary by up to 25% in common gene content. This high degree of gene diversity within a species suggests that a common phenotype cannot be reliably predicted for such species. Other examples show that evolutionary divergence as measured by ANI to be quite high among members of these species but that the common gene content is unusually high. In this case the niche of the organism appears to have provided strong selection for the same gene content. Together these results indicate that an organism's ecology must be considered in the species concept and that the current species definition is too liberal if one reasonably expects a species to reflect phenotype and where the organism lives. Advancing the prokaryotic species concept from genomic insights will help understand the extensive terra diversity and aid in predicting behavior and function of natural populations.

Biogeography of terra populations is important for understanding functional capacities of the landscape and reveals the organism's ecology. It gives the characterization of diversity a spatial component. Any important aspect of biogeography is to understand whether populations are cosmopolitan or endemic, and if endemic to what degree, i.e. at the level of species, ecotype or

genotype. Endemism is to be expected when the rate of genetic change in a local environment is greater than the rate of dispersal of the organism. This is expected to be different for different organisms, reflecting their different biology, and for different habitats, reflecting their accessibility, resources and stability. We have used fluorescent *Pseudomonas* and *Burkholderia cepacia* complex as two models with different biology and genome features to explore this question on a global scale. As expected, populations are not endemic at the level of the 16S rRNA gene, unless dispersal has been blocked, because this is a highly conserved gene. However, genome structure, as measured by *rep*-PCR, shows endemism at local scales, i.e. under 200m, and a high degree of endemism at continental scales. As intermediate measures of endemism we have use whole genome DNA-DNA hybridization, microarrays of 1kb genome fragments and multi-locus sequence typing of faster evolving core genes such as *rec*A and *gyr*B. All fill the gap between the above two methods with the microarrays providing the most resolution. All methods indicate moderate to weak endemism in these two target populations.

To advance gene-based infrastructure for exploring key genes in the unknown terra microbiology, we have developed a Functional Gene Database whose gene retrieval is based on protein searches using a Hidden Markov Model (HMM). This has the major advantage over BLAST in that it is a model-based search and it is know to be the best indicator of common function. Online alignment and NJ-joining treeing functions are the currently available analytical tools with a probe match underdevelopment. The retrieved sequences are scored relative to the model for full length and conserved motifs. This tool will be helpful in characterizing metagenome sequences and targeted functional genes retrieved from soil DNA, and may help in subdividing genes families according to functional differences.

Challenges for advancing terra microbiology center on understanding the microbe in relation to its habitat, or in the genomic context, in defining the interaction of the genome with the organism's ecology. The following topics are important to advancing this goal: a) determining biogeography of populations, b) defining density, dynamics of those populations, c) defining gene - phenotype linkages, d) understanding "genes" of unknown function, e) characterizing environmental control of expression, f) conceptualizing models of "species" evolution. These goals will be greatly aided by more genome sequence, of both organisms and metagenomes. For ascribing gene function to ecology, we would be greatly aided by more sequences from closely related organisms that are successful in different niches and/or exhibit different phenotype. With such information we stand a better chance of interpreting genome information in an ecological context.