

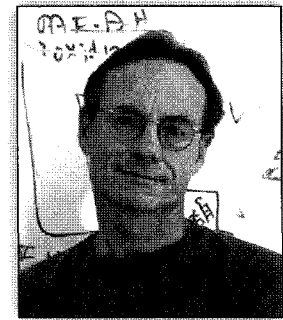
Comparative Metagenomics of Microbial Communities

**Edward M. Rubin¹, Christian von Mering², Jay M. Short³,
Eric J. Mathur³, John C. Detter¹, Peer Bork²,
Susannah Green Tringe¹**

¹DOE Joint Genome Institute, U.S.A.

²European Molecular Biology Laboratory, Germany;

³Diversa Corporation, U.S.A.



Assembled genomes of environmental microbes are informative but difficult to obtain due to the resistance of most organisms to cultivation and the diversity of microbial communities. Here we show that information encoded within fragmented sequence data alone can be used to characterize distinctive metabolic capabilities of microbial communities occupying different environmental niches. We sequenced microbial DNA isolated from a phylogenetically complex soil sample and three separate deep ocean whale skeletons and compared them to each other and to previously generated environmental genomic data. Despite the resulting low sequence coverage of individual microbial genomes in the samples, protein functions predicted from the fragmented sequence data for each environment revealed habitat-specific fingerprints. Binning of these proteins into either gene families, operons or cellular processes produced distinct patterns that correlate with the known metabolic demands presented by the different environments. The identification of environment-specific genes through a gene-centric comparative analysis presents new opportunities for defining and diagnosing environments.