

INVITED SPEAKER PRESENTATION

Open Access

The rare biosphere: sorting out fact from fiction

Mitchell L Sogin*, Hilary Morrison, Sandra McLellan, David Mark Welch, Sue Huse

From *Beyond the Genome: The true gene count, human evolution and disease genomics*
Boston, MA, USA. 11-13 October 2010

Over the past 25 years, microbiologists have employed the occurrence of DNA sequences as proxies for the presence of different kinds of organisms in microbial communities. These culture-independent investigations described new dimensions of diversity, identified novel candidate phyla, and redefined habitable ranges for single-cell organisms. The recent introduction of massively-parallel sequencing technology significantly increased estimates of microbial diversity from molecular-based studies. Matching SSU pyrotags to a reference rRNA database or clustering tags in a taxon-independent manner to identify Operational Taxonomic Units (OTUs) suggests that taxonomic richness in marine, terrestrial, and both the human and mouse microbiomes exceeds all prior estimates of microbial diversity. The occurrence of rare sequences in these data sets correspond to low abundance taxa that comprise the 'rare biosphere'.

Numerous theories and mechanisms that could account for the existence and persistence of rare biosphere members compete with explanations that invoke sequencing or clustering artifacts. Even with sequencing error rates below 0.005 per nucleotide position, the common method of generating OTUs (i.e. multiple sequence alignment and complete-linkage clustering) significantly increases the number of predicted OTUs and inflates richness estimates. The use of a novel Single Linkage Preclustering (SLP) strategy applied to short hypervariable regions of ribosomal RNAs accurately identified the predicted complexity of 'mock' microbial communities with a known number of rRNA operons. The strategy initially identifies sequences that are likely to have arisen by error using nearest neighbor clustering of pairwise sequence distances. The most abundant sequence for each precluster and the number of sequences in the precluster define inputs to average neighbor clustering using MOTHUR. When applied to

sequences obtained from multiple microbial communities, the OTU-based descriptions of microbial population structures under different ecological regimes, and the global distribution patterns of OTUs reinforce credibility of the 'rare biosphere' as revealed through deep sequencing efforts.

Published: 11 October 2010

doi:10.1186/gb-2010-11-S1-119

Cite this article as: Sogin et al.: The rare biosphere: sorting out fact from fiction. *Genome Biology* 2010 11(Suppl 1):119.

Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
Marine Biological Laboratory, Woods Hole, MA 02543, USA

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

