

Poster A-20
**Comparative analysis of
environmental sequences**



Authors:

Konrad U Foerstner (*European Molecular Biology Laboratory*)
Christian von Mering (*European Molecular Biology Laboratory*)
Sean D Hooper (*European Molecular Biology Laboratory*)
Peer Bork (*European Molecular Biology Laboratory*)

Short Abstract: The availability of large-scale metagenomics data will deeply improve our understanding of microbial communities in their natural environment. A comparative analysis of different habitats will provide the context sufficient for understanding each individual sample. Here we present such an investigation of different features of environmental sequence data.

Long Abstract:

Environmental sequencing, also dubbed metagenomics, is increasingly being used to obtain insights into organismal communities in diverse habitats, and has a variety of potential applications foreseeable in biotechnology and medicine. The first public large-scale data provide already a wealth of information hidden in vast amounts of fragmented pieces of DNA from unknown species residing in these environments. Comparative sequence analysis is essential for the interpretation of such data. However, different layers of complexity that are intrinsic to each sample require the establishment of some baselines for comparison. Here we discuss a few of these items and delineate some simple discriminative sequence properties for some distinct habitats [1, 2, 3, 4]. We showed [5] for example that the GC content of complex microbial communities seems to be globally and actively influenced by the environment. The observed nucleotide compositions cannot be easily explained by distinct phylogenetic origins of the species in the environments; the genomic GC content may change faster than was previously thought, and is also reflected in the amino-acid composition of the proteins in these habitats.

[1] Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 37-43

[2] Venter JC et al (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66-74

[3] Tringe SG et al (2005) Comparative metagenomics of microbial communities. *Science* 308: 554-557

[4] Foerstner KU, von Mering C, Bork P (2006) Comparative analysis of environmental sequences: potential and challenges. *Philos Trans R Soc Lond B Biol Sci.* 361(1467): 519-523

[5] Foerstner KU, von Mering C, Hooper SD, Bork P (2005) Environments shape the nucleotide composition of genomes. EMBO Rep. 6(12): 1208-1213