

Making sense of the Metagenome: Novel approaches to bioinformatics for metagenome analysis

Frank Oliver Glöckner, Hanno Teeling, Thierry Lombardot for the Microbial Genomics Group

Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen and International University Bremen, 28759 Bremen, Germany

The possibility to sequence DNA samples from natural environments without prior cultivation has revealed new and unprecedented insights into the microbial community composition and function. A central problem with community sequencing approaches (metagenomics) is the classification of disjoint sequence fragments into bins that most probably resemble an organism or a set of closely related ones. Measures such as the average G+C content of the fragments, best BLAST hits, and the codon usage of the corresponding coding regions have often been often applied to solve this problem. Since the resolution of these approaches is limited, we developed methods based on the analysis of intrinsic DNA signatures e.g. using Markov model-based statistics on oligonucleotide abundances or chaos game representations. When tested on real metagenomic data sets they clearly show an increased sensitivity and specificity, see www.megx.net/tetra. To address environmentally relevant questions like organism adaptations to oceanic provinces and regional differences in the microbial cycling of nutrients using ecological (meta)genomics, it is necessary to couple sequence data with geospatial information and supplement them with contextual information like physical, chemical and biological data. This is the task of Megx.net - a set of specialized databases resources and tools for marine ecological genomics see www.megx.net.