

## **Einladung zur Vortragsreihe** ***Algorithmische Bioinformatik***

**Frau Dr. Uta Bohnebeck,**  
**Bremerhavener Institut für biologische Informationssysteme**

spricht über

### **MetaFunctions: Environmental- and meta-genomics – a bioinformatics system to detect and assign functions to habitat-specific gene patterns**

Datum: Dienstag, 6. Januar 2009  
Zeit: 11:00 Uhr s.t.  
Ort: B-IT, Dahlmannstr. 2, Rheinsaal

Scientists are becoming increasingly interested in metagenomes – the collection of genes from all micro-organisms living in a particular environment. The MetaFunctions project, funded within the European Commissions FP6 NEST Adventure programme, is pooling expertise in bioinformatics, computer science, geographical information systems and marine sciences to develop a data-mining system that correlates genetic patterns in these metagenomes with contextual environmental data. The underlying database contains prokaryotic genome and metagenome sequences of marine origin together with information about their environmental context. Habitat parameters like water and sediment depths, temperature, salinity, and other physical-chemical properties are extracted from the literature or extrapolated based on global ocean data sets and remote sensing information.

Modern sequencing technologies allow rapid sequencing and bioinformatic analysis of genomes and metagenomes. With every new sequencing project a vast number of new proteins becomes available with many genes remaining functionally unclassified based on evidences from sequence similarities alone. Extending similarity searches with gene pattern approaches, defined as genes sharing a distinct genomic neighbourhood, have shown to significantly improve the number of functional assignments. Further functional evidences can be gained by correlating these gene patterns with prevailing environmental parameters. MetaMine was developed to approach the large pool of unclassified proteins by searching for recurrent gene patterns across habitats based on key genes.

MetaMine is an interactive data mining tool which enables the detection of gene patterns in an environmental context. The gene pattern search starts with a user defined environmentally interesting key gene. With this gene a BLAST search is carried out against the MetaStorage database containing marine genomic and metagenomic sequences. This is followed by the determination of all neighbouring genes within a given distance and a search for functionally equivalent genes. In the final step a set of common genes present in a defined number of distinct genomes is determined. The gene patterns found are associated with their individual pattern instances describing gene order and directions. They are presented together with information about the sample and the habitat. The system was evaluated with environmentally relevant genes related to the methane-cycle and carbon monoxide oxidation. Prevailing biological processes associated with a key gene can then be used to infer new annotations and shape hypotheses to guide further analyses.