



MetaLook & MetaMine: Data Mining and Analysis of Gene Patterns in an Environmental Context



Uta Bohnebeck*, Thierry Lombardot+, Renzo Kottmann+# and Frank O. Glöckner+#

*Technology Transfer Centre Bremerhaven, Germany, bohnebeck@ttz-bremerhaven.de

+ Max Planck Institute for Marine Microbiology, Germany, {tlombard | rkottman | fog}@mpi-bremen.de

#Jacobs University Bremen, Germany

Abstract:

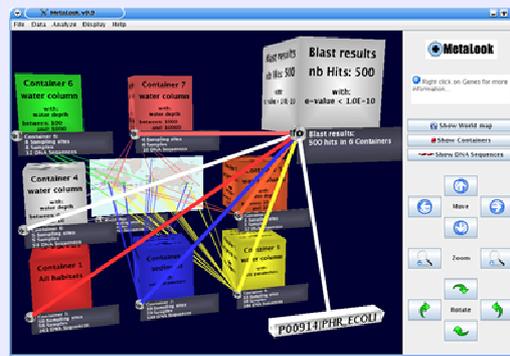
MetaLook and MetaMine are Java-based software tools developed in the context of the EU project MetaFunctions [1]. The goal is to detect and assign functions to **habitat-specific gene patterns**. The backend is a **geographic based** database which contains prokaryotic genome and metagenome sequences 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 [2] based on global ocean data sets (WOA [3]).

Two basic types of gene patterns are currently under investigation: (1) genes which are present or over/under-represented under specific environmental conditions (MetaLook [4]) and (2) patterns consisting of a set of genes occurring in specific genomic neighbourhoods (MetaMine). If such gene patterns are found the respective genomic context helps in assigning a potential function. In addition, if the gene patterns correlate with distinct environmental parameters or processes potential functions may again be inferred.

MetaLook:

- An interactive 3D visualization software for ecological genomics
- Relies on a geographic-centric database
- Allows gene analysis in ecological context (e.g. ocean water depth, phosphate, nitrate concentration)
- Input: gene of interest
- Output: precise distribution of the gene in the environment, according to environmental parameters

Screenshot of MetaLook



Here, a gene encoding photolyase (white gene in foreground) shows BLASTP hits in the top layers of the ocean, as expected, but also some unexpected hits in sediments (blue container).

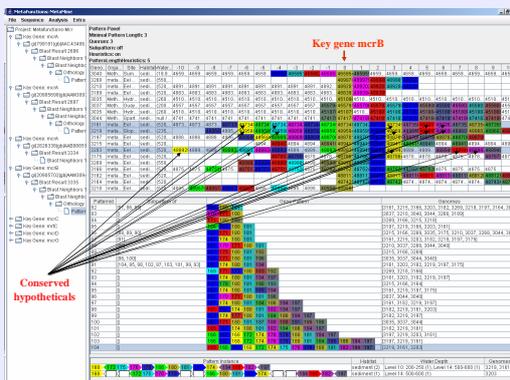
Application:

- Solar UV-light induces mutations in DNA. Photolyases (Phr) are proteins involved in a light-dependant DNA repair mechanisms, which protect microorganisms against the destructive effect of sunlight.
- Phr is expected to be only present where sunlight occurs (e.g. top layers of the ocean).
 - An analysis of Phr with MetaLook easily reveals the unexpected presence of this gene in the deep sea, leading to new hypotheses.

MetaMine:

- An interactive tool for gene pattern detection in ecological context
- Bottom-up approach
 - Prior biological knowledge is used to select key genes.
 - Key genes play the role of seeds to search for related gene patterns within the genomic neighbourhood.
- Input: user-selected key gene sequence
- Output: genomic patterns consisting of a set of common genes which are related to environmental parameters

Screenshot of MetaMine



Using mcrB as key gene, conserved hypothetical genes in genomic context could be found.

Application:

- McrB is a strongly conserved subunit of the Mcr protein and only encoded in (meta-) genomes which are able to carry out methanogenesis or AOM.
- Until now methanogenesis was only observed in isolates in contrast to AOM which was only observed on metagenomes. If it is possible to show different patterns associated with methanogenesis and AOM respectively, it would be additional evidence to prove the hypothesis that AOM is a reverse process of methanogenesis.

Conclusion:

- The integrative geographic-centric approach for genomics chosen by the MetaFunctions project opens new opportunities for data-mining.
- MetaLook and MetaMine are innovative software tools dedicated to ecological genomics that will help biologists to turn DNA information into actual knowledge.

References:

- [1] <http://www.metafunctions.org>
- [2] <http://www.megx.net/gms>
- [3] World Ocean Atlas and World Ocean Database, <http://www.nodc.noaa.gov>
- [4] <http://www.megx.net/metatool>

Acknowledgements:

The MetaFunctions project is supported by the European Commission within FP6, under the contract number NEST-511784