

The Genomes Mapserver: an integrative tool for ecological genomics and metagenomics

Thierry Lombardot, Renzo Kottmann and Frank Oliver Glöckner

Today, more than 20 genomes of marine prokaryotic organisms are available in the public databases. This number is going to increase dramatically since the Moore Foundation (www.moore.org) is starting to release around 90 genome sequences of marine origin. Furthermore, large-scale metagenomic studies focus on marine habitats, revealing a plethora of new genes and a high prokaryotic diversity (e.g. the Sargasso Sea project).

These sequencing efforts lead to the emergence of “marine ecological genomics”, defined as the application of genomic sciences to understanding the structure and function of marine ecosystems. The exponentially growing genomic data source represents a huge potential to gain more insights into marine ecosystems by exploring possible correlations between the genetic potential and the corresponding habitat for prokaryotic key-organisms. Therefore, specialised bioinformatic platforms are needed to systematically acquire, store, analyse and visualise this deluge of sequence data within the ecological context.

The “Genomes Mapserver” is a Geographic Information System (GIS) to store and systematically analyse (meta-) genomic data in correlation with ecological information. It is developed within the framework of the EU-funded project MetaFunctions. Currently the focus is set to marine systems, but our framework can be expanded to integrate all kinds of ecological systems.

GIS are commonly used in the field of geology for data integration. A GIS is a combination of elements designed to store, retrieve, analyse and display geographic data. In our newly developed Genomes Mapserver, the sampling sites of marine (meta)-genomic studies are displayed within a browseable world map. Each sampling site can be selected to display the corresponding sequences and additional contextual information. The underlying database is designed to enable future data mining tasks to reveal possible gene patterns associated with a particular environmental context.

This integrative system is intended to support researchers to generate a better understanding of the functioning of ecosystems based on genomic and metagenomic sequence data.