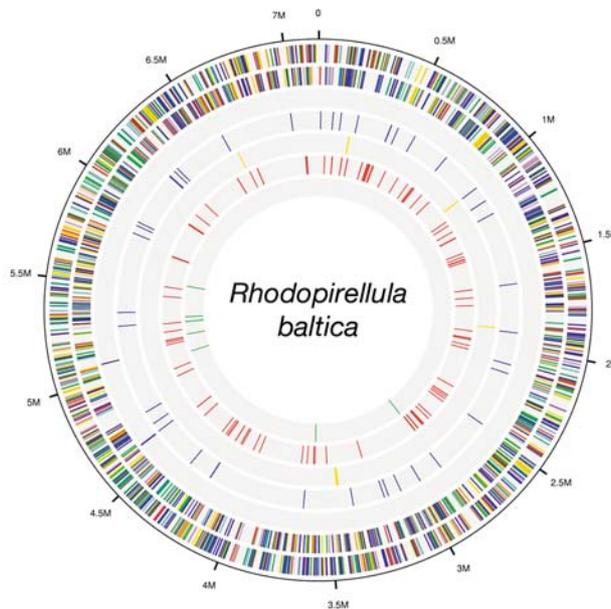


REPORT

NoE Marine Genomics Europe - Exploratory Workshop: Marine Genomics Meets Marine Diversity

Max Planck Institute for Marine Microbiology
June 7 – 9, 2006



Report on the Exploratory Workshop

Documentation

From 8 to 9 June 2006 sixty participants from thirteen countries presented state of the art results on marine microbial diversity and genome research and held lively discussions at the Max Planck Institute for Marine Microbiology in Bremen, Germany. The sessions covered the issues diversity, genomics, ecology, technology, and bioinformatics, including databases and portals. This exploratory workshop was funded by the European Commission as part of the activities of the large Network of Excellence Marine Genomics Europe (NoE MGE).

The participants and speakers came from Austria, Belgium, France, Germany, the Netherlands, Italy, Israel, Poland, Spain, Sweden, Switzerland, United Kingdom, and the United States of America. They included representatives of the two NoEs Marine Genomics Europe and Marine Biodiversity and Ecosystems Functioning (MarBEF) as well as the Global Biodiversity Information Facility (GBIF) and International Census of Marine Microbes (ICoMM) initiatives on marine biodiversity.

Introduction and Objectives of the Workshop

The ecosystem-centred exploration of marine biodiversity has already a long tradition. Over the past two decades an unexpectedly large diversity of marine microbes (defined as microscopic pro- and eukaryotes) has been disclosed by molecular biological techniques, often based on comparative sequence analysis of single genes, such as the small subunit rRNA. Since a few years high-throughput sequencing technologies are adding a new genomic perspective to the research on marine microbes. Already the first preliminary insights into these blueprints of life revealed unexpected findings which provide hints on niche adaptations and key functions of these microorganisms in the global cycling of matters.

Now, it is time to bridge the gap between marine diversity research and marine genomics. Integration can be achieved for: (i) data on diversity and spatio-temporal abundance of microorganisms with (ii) genomic data and (iii) further oceanographic information on the chemistry, physics, geology and biology of the habitat based on geographic location and sampling time. Whereas it is the task of computer scientists to provide searchable integrated datasets, it is up to the ecologists to raise and answer questions that have been out of reach so far, e.g.: “How does the microbial community change from a coastal to an open ocean site, how when looking at different oceanic provinces?”, “What is the seasonal variability of microorganisms and the influence of changes in water temperature and nutrient availability?” “How does the interface between biosphere and geosphere – the biological carbon pump - work on the molecular level?”

Large geographically integrated datasets will undoubtedly change the way marine biology is done today and it's now time to define the questions, expectations and technology that needs to be addressed.

Description

The workshop was organized into five sessions, three on the first and two on the second day, plus two longer discussions at the end of each day. All abstracts are available in a bound folder and on the web (see www.metafunctions.org -> workshops), including information on the speakers like contact details and short curricula vitae and contact information for all participants.

A highlight at the end of the first day was a 30-min talk on the novel technology of pyrosequencing and its application to environmental genomics, clearly presenting its advantages and limitations. It opened the discussions on emerging opportunities, which were continued over the joint dinner on the premises of the Max Planck Institute.

The final discussion was held at the end of the second day, for two hours, with approximately 30 participants present.

Thursday, 8 June

Session 1: Diversity Chair: Ramon Rosselló-Mora

Carlos Pedrós-Alió (CSIC, Institut de Ciències del Mar, Barcelona, Spain)

Marine microbial diversity: can it be determined?

Ramon Rosselló-Mora (Institut Mediterrani d'Estudis Avançats, Esporles, Spain)

The impact of genomics in prokaryotic taxonomy

Wiebe Kooistra (Stazione Zoologica Anton Dohrn, Naples, Italy)

The success of the diatoms in the modern marine plankton

Bernhard Fuchs (Max Planck Institute for Marine Microbiology, Bremen, Germany)

Bacterioplankton composition in the Namibian Upwelling system as revealed by 16S rRNA cloning and fluorescence in situ hybridisation (FISH)

Session 2: Genomics Chair: David Scanlan

David Scanlan (Dept. of Biological Sciences, University of Warwick, UK)

Niche partitioning of picocyanobacteria - an ecological and genomic perspective

Klaus Valentin (Alfred Wegener Institute for Polar & Marine Research, BHV, Germany)

Algal genomics - present and future trends - an overview about finished, ongoing and planned genome projects and their most important outcomes

Uwe John (Alfred Wegener Institute for Polar & Marine Research, BHV, Germany)

Functional genomics meets harmful algal bloom research: insights into toxin synthesis and growth control

Gurvan Michel (Station Biologique de Roscoff, France)

Structural and functional census of the polysaccharidases from *Rhodopirellula baltica*: a key polymer degrader in marine environment

Hanno Teeling (Max Planck Institute for Marine Microbiology, Bremen, Germany)

Novel aspects of marine ecological metagenomics – clustering of metagenomic fragments by intrinsic DNA signatures

Oded Béjà (Technion-Israel Institute of Technology, Haifa, Israel)

The quest for viral 'photosynthesis' using both metagenomics and environmental genomics

Session 3: Ecology Chair: Ian Joint

Klaus Jürgens (Baltic Sea Research Institute (IOW), Warnemünde, Germany)

Diversity of microbial communities and their biogeochemical functions in marine pelagic redoxclines

Ian Joint (Plymouth Marine Laboratory, UK)

Microbial metagenomics and marine biogeochemical cycles

Jakob Pernthaler (Limnological Station, University Zurich, Switzerland)

Diversity of marine Bacteroidetes (with a special focus on coastal North Sea waters)

Gerhard J. Herndl (Royal Netherlands Institute for Sea Research, NIOZ, Netherlands)

Bacterial and archaeal diversity and function in the major deep water masses of the North Atlantic

Bridging talk and discussion

Feng Chen (Livermore Biolaboratory, Walnut Creek, USA)

454 sequencing technology and its application in microbial and metagenomic community sequencing at US DOE Joint Genome Institute

Friday, 9 June:

Session 4: Technology Chair: Jean Weissenbach

Jean Weissenbach (Genoscope, Evry, France)

A metagenomic approach of wastewater processing

Alexander Loy (Universität Wien, Austria)

Functional probing of microbial communities with rRNA-targeted oligonucleotide microarrays

Thomas Schweder (EMA-Universität Greifswald, Germany)

Proteome analyses of marine bacteria

Phillip Neal (Marine Biology Laboratory (MBL), Woods Hole, USA)

Estimating marine microbial diversity: lots of data, lots of challenges, a few suggestions

Session 5: Bioinformatics Chair: Ed van den Berghe

Edward van den Berghe (VLIZ, Oostende, Flanders Marine Institute, Belgium)

MarBEF: lessons learnt from data integration

Francisco Rodriguez-Valera (Universidad Miguel Hernández, Alicante, Spain)

Biogeography of marine bacteria, genes, spacers and genomes

Andrea de Bono (*UNEP/DEWA/GRID-Europe, Châtelaine*, Switzerland)

GRID: The Global Resource Information Database

Anton Güntsch (Botanic Garden and Botanical Museum, FU Berlin, Germany)

Integration of heterogeneous biodiversity data sources using BioCASE and ABCD

Thierry Lombardot (Max Planck Institute for Marine Microbiology, Bremen, Germany)

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



The international group of workshop participants at the Max Planck Institute for Marine Microbiology, Bremen, 8 and 9 June 2006

Points of Discussion

Highlights from all discussions showed where new ecological knowledge is emerging and where research is urgently needed in order to advance our understanding of marine ecosystems.

Session 1

- **rare microorganisms** form a large part of (bio)diversity in the oceans, molecular and culture techniques need to be improved to find out under which conditions they thrive – on the other hand **core species**, having abundances larger than 1 per mill, can be defined
- expression profiling on the transcriptome and proteome level has to be used to further knowledge, but the percentage of genes which escapes from investigation is currently unknown

Session 2

- intrinsic DNA-signatures, cluster analysis, machine learning techniques and self-organizing maps can be used to classify (bin) and assemble fragmented metagenomic data
- currently best results are obtained from low diversity environments, such as open ocean systems, therefore, these environments could function as **models for microbe driven geochemistry**
- viruses play an important role in controlling phytoplankton blooms, possibly directly influencing photosynthesis, and triggering horizontal gene transfer

Session 3 and bridging talk

- new approaches to address the function of microbial communities comprise in situ stimulation to obtain rates of key functional genes, using MICRO-CARD-FISH and SIP (stable isotope probing)
- there is a clear necessity to take into account water mass characteristics (spatial and temporal variability!) when investigating microbial communities
- importance of **sample archiving**, i.e. for later sequencing with new and different methods
- novel **pyrosequencing** (454 technology) increases sequencing speed about 100 fold but has several limitations, especially short reads which demands for new bioinformatics approaches



The host, Prof. Frank Oliver Glöckner of the Max Planck Institute for Marine Microbiology (left) and Dr. Feng Chen from Livermore Biolaboratory, USA, (right) during the discussion of the bridging talk on pyrosequencing

Session 4

- Genoscope mentioned that in-house projects mainly deal with environmental microbiology and that there is an on-going interest in European marine genomics and metagenomics
- promising approaches are
 - to integrate phylochips with the full cycle rRNA approach,
 - to use isotope arrays with radioactive labelled substrates and fluorescent RNA to achieve **habitat specific arrays** to be analysed in a highly parallel manner
- Metaproteomics is already an option but need further (meta)genomic data for mapping
- an efficient bioinformatic pipeline is needed to keep up with the output of pyrosequencing

Session 5

- in a few years microbiologist can prove whether “everything is everywhere”
- Open GIS Consortium (OGC) standards can be used to map existing georeferenced databases into each other, just linkages are necessary between distributed data
- the **data conversion** of not standardised data collections is the time-consuming step, it is a legacy from the past, which will create value for all, but claimed to be a not scientifically rewarding job to be done
- more communication between data managers is necessary, i.e. how to link genomic and oceanographic data automatically
- there is the need to have institutional archives and proper documentation of the datasets, to avoid loss of data
- exposing the discovery of metadata from these archives and other sources, will increase knowledge about existing data, and facilitate exchange of data.



The international group of workshop participants during one of the many fruitful discussions in the lecture hall of the Max Planck Institute for Marine Microbiology, Bremen, 8 and 9 June 2006

Final overall discussion

Needs:

- more genome sequences of the same species and of closely related species for in depth comparison
- derive knowledge out of data and information, e.g. confirm predicted gene functions in the lab
- address hypothetical genes problem
- technological needs for metaproteomics:
 - mixed cultures for simulations to avoid non-reproducible results
 - a good model system
- an interdisciplinary approach including computer scientists, bioinformaticians and biologists to integrate, analyse and interpret existing and upcoming data
- standardisation of data acquisition and consistent data storage
- a standardised minimal set of contextual data for genomic and metagenomic samples as well as for any kind of other sequence data (like rRNA)

Open questions

- How can data from the species-centric world be integrated with sequence-centric information (today 6.000 species vs 220.000 SSU rRNA gene sequences in the microbial world)?
- How to determine the importance of data archiving and database curation compared to generating new field data?
- How large is the unknown fraction of biodiversity of microorganisms in marine environments?
- How to close the gap between known genome sequences and unknown gene functions? Here lies the “true gold of biodiversity”!
- Should the focus lie on open ocean (e.g. mixing areas and ridges as hot spots) or on coastal priority sites (e.g. 5 ATBI sites) to reach next level of understanding?

- What is more important or more urgent: to study functions of genes relevant in natural cycles or to study genetic systems in pure culture?

Options

- **create and maintain sample archives and DNA banks** with geographic coordinates. This could be done in a decentralized way (with a US component (Ocean Genome Legacy?)) with the derived in silico information available via a central portal to secure long-term data storage (e.g. MarBEF)
- using high quality sequence data and data curation to answer ecological questions instead of copying the “let’s sequence the whole ocean” approach of the USA
- **use the open ocean as model system** for genome-enabled biogeochemistry
- meta-biology can help to understand the globe and can be used for predictions and models
- combine metagenomics with emerging project ideas like
 - LifeWatch: a super network aiming for long-term infrastructure (MarBEF, EurOcean, maybe MGE)
 - biodiversity observatories (HERMES, ESONET?)

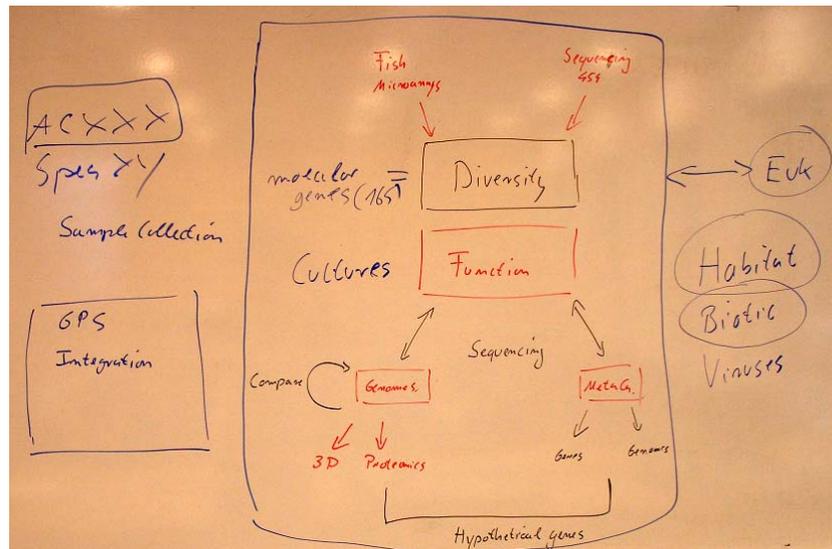
Emerging Project Ideas

There is an urgent need for action in the field of data **integration of (meta)genome- and diversity data** in order to understand ecosystems as a whole. First steps are to integrate molecular diversity and abundance data into the MarBEF diversity database to make them accessible via a central portal. Experience with the integration of diversity data is available, however, details of this integration needs to be jointly developed within a new project.

To bridge the gap between diversity and function in marine systems, all kind of meta(genomic) data have to be collected and integrated in conjunction with diversity and environmental data. This offers a tremendous potential to better understand the functioning of natural ecosystems. Researchers are then able to consider each sequence in their ecological context. On a long run this will help to extend the systems biology approach into an ecosystems approach. This will not only give a detailed description of the current status of the marine ecosystem but help to predict the consequences of ongoing global changes an all life-forms on earth.

Proteomics and metaproteomics approaches are important next steps. They can confer a function to an unknown protein, identify environmentally relevant proteins and help predicting physiological as well as ecological states of a habitat under investigation. Until now, metaproteomic approaches are few and results are hard to interpret. Therefore, there is a clear need for a low diversity model system as a starting point for further method development.

It takes about three years to understand the function of one single protein. Therefore, high-throughput methods are needed as well as projects with several partners to elucidate the functions of genes relevant in environmental biogeochemical cycles.



Schema of an integrated approach to address “Who is out there (diversity)?” and “What are they doing (function)?” developed together with the workshop participants in the final discussion on Friday 9 June

Conclusions and Vision

It has been stressed by the participants of the workshop that biology has moved on from a single experiment endeavour to a high throughput cross-linked science. The amount of data produced every day by molecular and especially genomic techniques exceeds the ability of a single researcher to process and use them in a meaningful way. Thus, only a tight connection between experts in biology and ecology with experts in technology and bioinformatics will make it possible to transform this wealth of information into knowledge, i.e. to identify the key players as well as the basic structural and functional elements of marine ecosystems, in which microorganisms play an important role. A better understanding of the forces driving marine ecosystems is the ultimate goal of marine microbial ecology and will help to manage protection and exploitation of natural resources and services in a sustainable way.

Therefore, **a vision for marine research** during the next decade (or for the seventh Framework Programme of the European Commission (FP7)) based on this workshop is:

Marine research will concentrate on the understanding of large scale geochemical processes in marine environments that are controlled by (micro-)biota to obtain predictive models over a variety of time scales and geographical areas.

This research focus will contribute substantially to pressing current environmental issues of global relevance such as climate change, regulation of geochemical functions, eutrophication, overfishing and reduction of biodiversity of marine organisms.

In the light of currently ongoing activities in the US it has to be stressed that a European, large scale integrated activity is urgently needed. The following paragraph has been taken from the press release in January 2006 on the Website of the Venter Institute:

The Gordon and Betty Moore Foundation has awarded \$24.5 million over seven years to create the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA). Scientists will use CAMERA for metagenomics research -- analyzing microbial genomic sequence data in the context of other microbial species, as well as in comparison to a variety of other "metadata" such as the chemical and physical conditions in which microbes are sampled.

In this context it should be clear that the European project should not try to repeat the efforts in the US, but has to be complementary. The main strategy will be to focus on the quality of data rather than on high-throughput. This means also to constantly act on the standardisation of data acquisition and storage. Ongoing examples are the Genomic Standards Consortium in the UK or the ICoMM initiative. This will allow building a solid basis to start modelling and predicting the complex interplay of marine organism with their environment.

Annexes

List of databases of interest

(Meta)genome database for the integration of genomic and environmental data

www.megx.net (MPI-Bremen, Germany)

www.metafunctions.org (MPI-Bremen, Germany)

<http://camera.calit2.net> (CAMERA)

Geoscientific & Environmental Data

www.grid.unep.ch

www.pangaea.de

Marine Genomics

www.marine-genomics-europe.org (NoE Marine Genomics Europe)

Biodiversity databases

www.biocase.org (GBIF/BioCASE GBIF related project on biological collections)

www.iobis.org (Ocean Biogeographic Information System)

www.marbef.org (NoE MarBef)

Proteome database

www.marine2D.de (Uni Greifswald, Germany)

ProbeBase

www.microbial-ecology.net (DOME, Austria)

Oceanographic databases:

www.iode.org (IODE system of National Oceanographic Data Centres)

www.nodc.noaa.gov (NODC in Silver Spring, US)

www.bodc.ac.uk (British BODC)

www.bsh.de/en/Marine%20data/Observations/DOD%20Data%20Centre/index.jsp (German DOD)

Genomic Standards Consortium

www.genomics.ceh.ac.uk/genomecatalogue/gsc.php

Workshop folder including

1. agenda
2. abstract
3. speaker cvs
4. list of participant
5. contact details for both