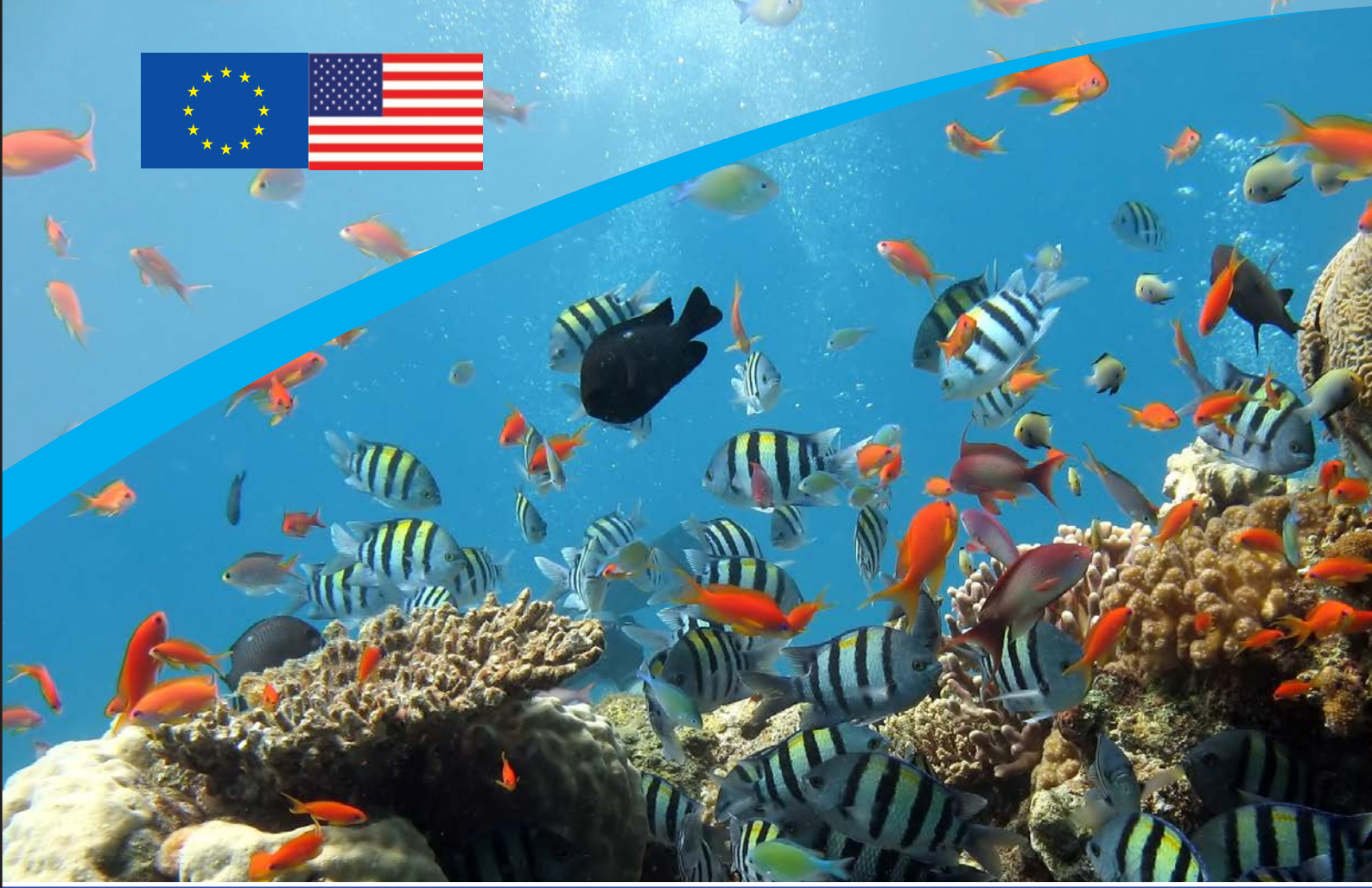
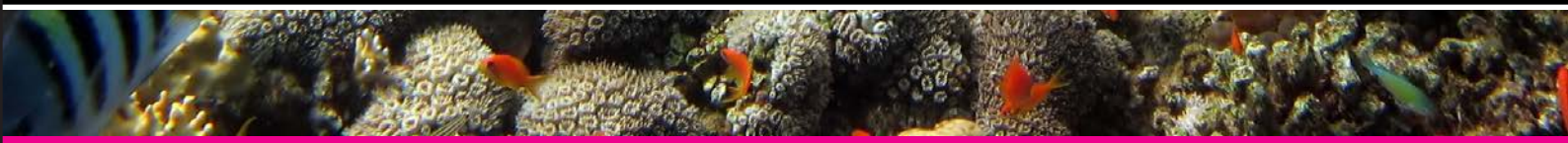


EC-US Task Force on Biotechnology Research

Joint EC-US and CIESM workshop on Marine Genomics



CONFERENCE PROCEEDINGS

Principality of Monaco
12 - 14 October, 2008

Interested in European research?

Research*eu is our monthly magazine keeping you in touch with main developments (results, programmes, events, etc.). It is available in English, French, German and Spanish. A free sample copy or free subscription can be obtained from:

European Commission
Directorate-General for Research
Communication Unit
B-1049 Brussels
Fax (32-2) 29-58220
E-mail: research-eu@ec.europa.eu
Internet: http://ec.europa.eu/research/research-eu

EUROPEAN COMMISSION

EUROPEAN COMMISSION
Directorate-General for Research
Directorate E — Biotechnologies, Agriculture, Food
Unit E2 — Biotechnologies
Contact: Maurice LEX

European Commission
Office SDME 08/35
B-1049 Brussels
Tel:+32 2 2965619
Fax: +32 2 2991860

E-mail: Maurice.LEX@ec.europa.eu

EUROPEAN COMMISSION

Joint EC-US CIESM Workshop on Marine Genomics

At the Interface of Marine Microbial Ecology and Biotechnological Applications

Principality of Monaco
12-14 October, 2008

Directorate-General for Research
2009 EC-US TASK FORCE ON BIOTECHNOLOGY RESEARCH 23908 EN

EUROPE DIRECT is a service to help you find answers to your questions about the European Union

Freephone number (*):
00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed

Cataloguing data can be found at the end of this publication.

Luxembourg: Office for Official Publications of the European Communities, 2009

ISBN 978-92-79-12136-4

ISSN 1018-5593 1831-2322

TABLE OF CONTENTS

P R E F A C E

Since 1990, the **EC-US Task Force on Biotechnology Research** has been coordinating transatlantic efforts to guide and exploit the ongoing revolution in biotechnology and the life sciences. The Task Force was established in June 1990 by the European Commission and the White House Office of Science and Technology Policy. This mandate has been renewed three times. The Task Force has acted as an effective forum for discussion, for coordination and for developing new ideas for the last 18 years.

Task Force members are European Commission and US Government science and technology administrators who meet annually to enhance communication across the Atlantic, and to encourage collaborative research. Through sponsoring ***workshops***, and other activities, the Task Force also brings together scientific leaders and early career researchers from both sides of the Atlantic to forecast research challenges and opportunities and to promote better links between researchers.

Over the years, by keeping a focus on the future of science, the Task Force has played a key role in establishing a diverse range of emerging scientific fields, including biodiversity research,bioinformatics neuroinformatics, plant and animal biotechnology, environmental biotechnology, nanobiotechnology, neonatal immunology, transkingdom molecular biology and systems biology. At Task Force workshops, a small number of leading scientists, each operating in different but related areas, come together for a few days in informal surroundings. These workshops seek to look into the future of emerging fields of science and answer the question of whether international collaboration in a certain field would be useful. Workshop participants represent different disciplines, which need to be integrated in order to move forward in a new area of science. Drawing on these differences in research backgrounds, EC-US Task Force workshops are full of surprising conclusions and they can produce some inspiring thinking.

This report summarizes the presentations and discussions held at the joint EC-US / CIESM Workshop on Marine Genomics: at the Interface of Marine Microbial Ecology and Biotechnological Applications held at Monaco on 12-14 October 2008. The Mediterranean Marine Science Commission (CIESM) jointly sponsored and hosted this workshop which brought together 34 scientists from Member States of the European Union and the United States including those CIESM members from neighbouring Mediterranean countries .The workshop provided a forum for a technical review of the newest research activities using genomics and metagenomics in marine microbial ecological science and biotechnological applications. Discussions were held about possible future research collaborations between US, EC and CIESM scientists.

Dr. Doug Bartlett of Scripps Institution of Oceanography, University of California, San Diego and Dr. Frank Oliver Glöckner of the Max Planck Institute for Marine Microbiology, Bremen were the science co-chairs of the workshop. Dr. Laura Giuliano and Dr. Michele Barbier represented the CIESM and Dr. Lita Proctor from the National Science Foundation, Arlington (VA) and Dr. Maurice Lex of the European Commission, Brussels were co-organisers. These have all compiled and edited this report, which is also available on the EC-US Task Force web site, http://ec.europa.eu/research/biotechnology/ec-us/index_en.html

We would like to thank all participants, particularly the chairs and our CIESM hosts, for their outstanding efforts. The views expressed in this document are those of the workshop participants, and do not necessarily reflect the views of the sponsors or governments.

Dr. Timothy Hall, Acting Director
European Commission

Dr. Judy St. John, Deputy Administrator
Agriculture Research Service, US
Department of Agriculture

Executive Summary

Genomics is defined as the study of the DNA sequences of organisms and metagenomics (also environmental genomics, ecogenomics or community genomics) as the study of genetic material recovered directly from environmental samples. In the past, advances in microbiology, including marine microbiology, depended mostly upon culturing.

The new age of metagenomics enables now the study of the vast majority of microbial species which are as yet unable to be cultivated in the laboratory. These technologies and the analyses they enable (comparative (meta)genomics, (meta)transcriptomics, (meta)proteomics, metabolomics, high throughput gene disruptions, etc…) have ushered in a new era of biology with fundamental implications for basic research and biotechnological advances. But, they also pose challenges in areas of intellectual property and patent law as well as in interdisciplinary training of the next generation of scientists.

The power of these methods and the continuing decreases in sequencing costs ensure that they will be at the forefront of all of biology for the foreseeable future. One example for the advances of these technologies is that the new Roche 454 Genome Sequencer is now capable of producing more than 1,000,000 reads of 400 bases per 10-hour instrument run.

Projects like the Sorcerer II Global Ocean Sampling Expedition have highlighted the fact that much of the phylogenetic and biochemical diversity of life on Earth is present in its marine microbes. Already marine (meta)genomic projects have uncovered information on new biogeochemical cycles and energy-coupling mechanisms, xenobiotic and complex polymer catalysis, secondary metabolite and bioactive compound biosynthesis, enzyme function under diverse conditions, symbiotic and syntrophic interactions, and the role, abundance and diversity of viruses.

Within the EC much of the support for these advances has come from the Networks of Excellence 'Marine Genomics Europe' and 'Marine Biodiversity and Ecosystem Functioning' In the United States funding for basic research in marine (meta)genomics has come from the National Science Foundation, the United States Department of Agriculture and the Department of Energy.

Workshop reports like this one provide a useful opportunity for reflection on the best ways to prioritize and standardize marine (meta)genomics in order to glean as much meaningful information to as many scientists as possible. This field is impossibly immense and this short report cannot do it justice. Our hope is simply that it will provide succinct strategic advice in key areas of its management and optimization. The accompanying session reports provide a number of recommendations for future growth and maturation of this field.

Ten of these are as follows:

1. Where possible (meta)genomic research programs needs to consider all components of the microbial fraction (bacterial, archaeal, eukaryotic and viral fractions).
2. (Meta)genomic efforts needs to continue to explore under-sampled marine habitats.
3. (Meta)genomics can be used as a tool to explore marine microbial ecology, but this needs to be done on scales and gradients relevant to the ecosystem of interest.
4. Further development of single cell genomics ('getting more from less') needs to be encouraged.

5. There is an urgent need for more culturing and subsequent genome as well as post-genomic analyses of more marine model organisms of all domains of life including viruses.
6. More efforts need to be applied to defining the growing constellation of new, and often unknown, protein families using newer high throughput approaches and more standard biochemical and genetic techniques.
7. Bioinformatics is currently a bottleneck to many scientific advances. Improved bioinformatics tools and specific infrastructures are needed to better proceed from (meta)genome sequence information to biochemical and physiological function prediction to finally reach a holistic ecosystem understanding.
8. More cooperation between developed and developing countries needs to be established to accomplish research and development using marine genomic resources (for some examples see U.S. 'International Cooperative Biodiversity Groups' program, EU FP6 'Marine Biodiversity and Ecosystem Functioning' and 'Marine Genomics Europe' Network of Excellence projects).
9. The standards for (meta)genomic data and contextual data formats proposed by the Genomic Standards Consortium need to be universally established.
10. Additional training courses in bioinformatics need to be established.

2008 Joint EC-US/CIESM Workshop on Marine Genomics: at the Interface of Marine Microbial Ecology and Biotechnological Applications

Principality of Monaco
12-14 October 2008

Agenda

Sunday, October 12

13:30-13:45	Welcoming remarks **Frédéric Briand**, Director General, Mediterranean Science Commission (CIESM), Monaco
13:45-14:00	Goals of the workshop **Douglas Bartlett**, Scripps Institution of Oceanography, San Diego, CA, USA**; Frank Oliver Glöckner**, Max Planck Insitute for Marine Microbiology (MPI), Bremen, Germany

Session 1: Marine Molecular Microbiology - the Great Questions

14:00 -14:25	Session introduction and presentation, **Karla Heidelberg**, Univ. Southern Calif., Los Angeles, CA, USA. Marine and Environmental Microbiology and Ecology; Protistan Diversity; Environmental Genomics and Transcriptomics.
14:30 – 14:45	**Andy Allen** "Comparative and functional Genomics of marine microalgae" J. Craig Venter Inst., La Jolla,CA, USA
14:50 – 15:05	**Ramunas Stepanauskas** "Single-cell approaches in microbial diversity, ecology, and evolution" Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME, USA
15:10 - 15:25	**Fitnat Yildiz** "Genetic and functional genomic analyses of the mechanisms of ersistence and survival of *vibrio cholerae*". Univ. California at Santa Cruz, Santa Cruz, CA, USA
15:30 – 15:45	**Alison Murray** "Genomic approaches to descriptions of the diversity of life with an emphasis on Antarctica" Desert Research Inst., Reno, NV, USA
15:50 – 16:20	*Coffee break*

16:20 – 16:35	**Matthew Sullivan** "(Meta)genomic analyses of marine viruses" Univ. Arizona, Tucson, AZ, USA
16:40-16:55	**Michail Yakimov** "Deep-sea hypersaline anoxic basins: chemosynthetic view on ecosystem functioning" Institute for Coastal Marine Environment/The National Research Council (IAMC/CNR), Messina, Italy
17:00 – 18:00	Session Discussion moderated by **Karla Heidelberg**
18:30 – 20:00	*Cocktail dinatoire offert par la CIESM au Musée Océanographique de Monaco*

Monday, October 13

Session 2: Functional (Meta)Genomics of Marine Microorganisms

8:00-8:25	Session introduction and presentation **Ian Joint**, Plymouth Marine Laboratory Plymouth, UK
8:30 – 9:45	**Gurvan Michel** "Bioconversion of algal polysaccharides by marine bacteria: from complete genomes to novel biocatalysts for blue biotechnology" Station Biologique de Roscoff, National Center for Scientific Research/Pierre & Marie Curie University (CNRS/UPMC), Roscoff, France
8:50 – 9:05	**Manuel Ferrer**, "Metagenomics approaches in systems microbiology" Instituto de Catalisis y Petroleoquimica. The Spanish National Research Council (CSIC), Madrid, Spain
9:10 – 9:25	**Christine Klockow** "Functional genome analysis of *Rhodopirellula baltica* SH1^T" Max Plank Institute for Marine Microbiology, Bremen, Germany
9:30 – 9:45	**Granger Sutton** "Computational approaches for functional metagenomics" J. Craig Venter Institute, Rockville, MD, USA.
9:50 – 10:20	*coffee break*
10:20 – 10:35	**Cinzia Verde** "Molecular adaptations in fish – an integrative and evolutionary view," CNR - Institute of Protein Biochemistry, Naples, Italy
10:40 – 10:55	**Thomas Schweder** "Proteomics - an important platform technology

for functional genome analyses of marine microorganisms" Institute for Marine Biotechnology, Greifswald, Germany

11:00 -11:15 **Chris Bowler "**The Tara-Oceans circumnavigation project, National Center for Scientific Research" (CNRS) UMR 8186, Biologie Moléculaire des Organismes Photosynthétiques, Paris, France

11:20- 12:00 Session Discussion
moderated by **Ian Joint**

12:00-13:30 *Lunch, offered by CIESM*

Session 3: Biocatalysis, Drug Discovery and Industrial Production

13:30-13:55 Session introduction and presentation
Dermot Hurst, Marine Institute, Ireland

13:55-14:10 **Wolfgang Aehle** "Industrial enzymes - discovery and development" Brain GmbH, Zwingenberg, Germany

14 :15-14:30 **Kevin Sowers** "Applications of marine microbes for bioremediation and bioenergy" Univ. of Maryland Biotechnology Institute, Baltimore, MD, USA

14:35-14:50 **Peter Golyshin** "Activity-based mining of new proteins from genomes and metagenomes" University Bangor, Bangor, UK

14:55-15:10 **Fernando de la Calle Verdu** "Marine biodiversity as source of new drugs" PharmaMar, Madrid, Spain

15:15-15:30 **Diaa Youssef** "Recent Studies on Red Sea marine organisms" Suez Canal Univ., Ismailia, Egypt

15:35-16:00 *coffee break*

16:00-17:45 Session Discussion
moderated by **Dermot Hurst**

Session 4: Intellectual Property Rights of Marine Genomic Resources

16:45-17:10 — Session introduction and presentation
Joel Querellou, Ifremer, Brest, France

17:15-17:30 — **Amar Mohamed** "The role of biological resource centres for Morocco's bioeconomy" Laboratory of Microbiology and Molecular Biology, Centre National pour la Recherche Scientifique et Technique (CNRST), Rabat, Morocco

17:35-17:50 — **Jean François Bloch** "Intellectual property rights of marine genomic resources" Protéus, Nimes, France

17:55-18:10 — **Margo Haygood** "Bioactive metabolite symbiosis" Oregon Health and Science Univ., Beaverton, OR, USA

18:10-19:00 — Session Discussion
moderated by **Joel Querellou**,

Evening is free

Tuesday, October 14

Session 5
Computing Power and Bioinformatics - Catching Up with the Genomics Revolution

8:30-8:35 — Session introduction
Michèle Barbier

8:35-8:50 — **Frank Oliver Glöckner** "The (meta-)genomic revolution in the marine environment" Max Planck Institute for Marine Microbiology, Bremen, Germany

8:55-9:10 — **Monia El Bour** "Marine biotechnology in Tunisia: the present state and forward proposals" National Institute of Marine Sciences and Technologies, Salammbo, Tunisia,

9:15-9:30 — **Balkiss Bouhaouala-Zahar** "Bioinformatics in Tunisia: current situation and biotechnological application on marine genomics" University of Tunis-El Manar, Tunis, Tunisia

9:35-9:50 — **Granger Sutton** "Sequence fragment assembly, sample comparison, and visualization of metagenomic data" J. Craig Venter Institute, Rockville, MD, USA

9:50-11:00 Session Discussion
moderated by **Frank Oliver Glöckner** and **Michèle Barbier**

including coffee break

11:00 -12:00 Workshop General Discussion :conclusions and recommendations
moderated by **Frank Oliver Glöckner** and **Douglas Bartlett**,

12:00-12:20: Workshop report - the next steps
Lita Proctor, National Science Foundation, Arlington, VA, USA
Maurice Lex, European Commission, Brussels, Belgium

S E S S I O N - 1

Marine Molecular Microbiology - The Great Questions

Karla Heidelberg[1], Andy Allen[2], Ramunas Stepanauskas[3], Fitnat Yildiz[4], Alison Murray[5], Matthew Sullivan[6], Michail Yakimov[7]

[1]University of Southern California, Los Angeles, CA, USA, [2]J. Craig Venter Institute, La Jolla, CA, USA, [3]Bigelow Laboratory for Ocean Sciences, ME, USA, [4]University of California, Santa Cruz, CA, USA, [5]Desert Research Institute, Reno, NV, USA, [6]University of Arizona, Tucson AZ, USA, [7]Institute for Coastal Marine Environment/The National Research Council (IAMC/CNR) Messina, Italy.

Background

The field of marine microbiology has an interesting history. In the last century marine microbiology was dominated primarily by two extremes, microbial autecology and microbial synecology. At one end of the spectrum, autecologists studied controlled systems in laboratory experiments to determine metabolic potential and elucidate biochemical pathways within pure cultures of microbes isolated from the environment.

At the other end of the spectrum, the microbial syncologists focused on the major biochemical processes occurring within an ecosystem, irrespective of what microorganisms were responsible for the processes, a field of study that became known as "black box ecology." By the later part of the 20th century, the application of molecular biology methods in marine ecology started with the sequence analysis of taxonomic conserved genes (i.e. SSU rRNA genes). This technology yielded a major breakthrough for surveying environmental microbial communities that could not be cultured and fundamentally changed our ability to study marine microbial populations in their natural environment.

By the mid 1990s, the advent of a method for whole genome sequencing of microbes provided a mechanism to evaluate a cultivated microbe's full genomic potential (Fleischman et al. 1995) while at the same time, new approaches to access uncultivated microbial genomes from natural environments were being developed (Stein et al. 1996, Schleper et al. 1997). These accomplishments were followed early in this century by complex community-wide 'metagenomic' surveys of entire multispecies communities of microorganisms (Fig. 1).

Each of these technological advances provided insights into the diversity and/or metabolic potential of naturally occurring microbes and challenged traditional paradigms and conceptual models. The more recent addition of transcriptomics, metabolomics, proteomics and other '-omic' approaches has continued to reshape our understanding of ocean biological diversity, biogeography, biogeochemical transformations, biosynthesis of natural products, and evolutionary biology. From the humble beginnings of microbiologists using observational microscopy to the head-spinning advances over the past three decades in the application of molecular technologies to characterize microbial communities, the field of marine microbiology has advanced at an astounding rate.

We have now reached a crossroad in which it is time to reflect not only upon the successes, but to also identify the important questions, including areas of unrealized discovery and future needs. Here, we summarize a workshop discussion around three major areas:

How much do we know about marine microbial diversity?

At no other time in our history have we so completely realized the immensity of microbial diversity or its significance. Even after two decades of surveying microbial diversity the discovery of novel forms is common. Even within a given habitat, microbial communities can vary between locations separated by millimeters to thousands of kilometers. Both large and small-scale patterns on biogeography have been observed over temporal and special scales, and the reporting of a very ancient and diverse 'rare biosphere' provides evidence that microbial populations may contain a nearly inexhaustible source of genomic innovation (Sogin et al. 2006). Although tremendous variation exists between microbes, it remains challenging to evaluate this diversity because there is still no agreement within the scientific community about the definition of a microbial species. A consequence of this is that different studies employ different bioinformatics similarity cutoff values in sample analysis (reviewed by Little et al 2008).

Our current view of microbial diversity is largely shaped and interpreted by the abundant sequence information available for the Bacteria and Archaea. In contrast, the evaluation of diversity and genomic assessments of the single-celled eukaryotic microbes (the protists) and the fungi into the genomic era have been much slower despite their important role in marine microbial communities and in expanding our understanding of evolution of multicellular taxa (Baldauf 2003, Caron et al. 2008). Protistan lineages represent one and a half billion years of evolution and comprise the bulk of eukaryotic phylogenetic diversity as well as myriad life forms. They also constitute several essential components of global food webs. Despite this, the genomic study of eukaryotic microbial organisms lags behind the prokaryotes by at least a decade due to lack of resources and bioinformatic tools to evaluate large and complex genomes.

At the other end of the microbial spectrum, marine viruses have also, until very recently, been overlooked for assessments of diversity and abundance despite their significant role in bacterial and phytoplankton mortality (see Breitbart et al. 2007; Sullivan et al. 2005). Viruses play significant roles in regulation of metabolism genes and have the potential to alter global photosynthetic rates through mortality. Increasing genetic signatures of environmentally representative phages in public databases will undoubtedly help move this field forward.

How much do we know about microbes living in group assemblages?

Outside of laboratory cultures and well-defined symbiotic relationships, virtually all microbes exist in complex consortia, sharing nutrients and other resources and exchanging genetic material. What forces shape these communities, and how they maintain functional capabilities in changing environments is not fully understood. Environmental metagenomic techniques are allowing us to begin to look at genetic and functional relationships. The comparative architectures of different ecotypes of the marine unicellular cyanobacterium, *Prochlorococcus,* revealed genomes that are specifically adapted for living in different light regimes (Rocap et al. 2003). Venter et al. (2004) and Rusch et al. (2006) describe extensive bacterial and archaeal population-level heterogeneity and high levels of environmental adaptation by communities of microbes. Tyson et al. (2004) used a metagenomic reconstruction of a biofilm community to show that microbial communities can function collectively. In fact, similar findings have been

reported for highly constrained episymbionts living in association with a hydrothermal vent polychaete, *Alvinella pompejana* (Grzymski et al. 2008). Genes that occur more frequently in a particular community may be conferring attributes beneficial for maintenance of the function of that particular ecological niche. We have also seen novel strategies of adaptation for different environments in eukaryotic microbes (Allen et al. 2008). Together these findings suggest that marine microbial genomes both free-living and symbiotic, are complex, highly dynamic, and adaptive.

How much do we know about bioenergetic, adaptive, and evolutionary strategies that marine microorganisms utilize to thrive in the variety of habitats that exist in the worlds' ocean?

Microbes regulate key nutrient and biogeochemical cycles in the ocean, yet little is known about microbial community genomic variability, especially along temporal and special scales relevant to ecosystem functions. The ability to now study microbial assemblages that provides information on taxonomic identification, genetic potential and functional activity with environmental parameters provides a powerful tool to characterize ocean function, and the field is now poised to start routinely incorporating comprehensive scientific and ecological approaches to study marine systems.

Accessibility of metagenomics and development of tools for within and between ecosystem comparisons (Tringe et al. 2005, Rusch et al. 2006) now enables questions to be framed in environmentally relevant scales and across important gradients of space and time. DeLong et al. (2006) have provided an exciting glimpse of the potential of this approach from a comparative genomic analyses of a vertically stratified microbial community at the Hawaiian Ocean Time Series Station. Current metagenomic studies in high latitude environments at both poles also promise to address numerous questions concerning variation over the annual cycle (e.g. Murray and Grzymski, 2007), biogeography and bipolar distribution (Staley and Gosink, 1999), that will help determine the effects of global change in these sensitive ecosystems. These are all good examples of how we are building understanding of the causes of microbial diversity change across space and time. It is approaches such as these that will allow for refining understanding of basic relationships between community diversity and ecosystem function and provide important opportunities to gain a predictive understanding of the response of ecosystems in the face of environmental change.

Until recently, one of our additional limitations in evaluating the genomic potential and functional roles of microbes in a given environment was that we were unable to bring into culture, and then sequence, any of the organisms sampled. In fact, most environmentally important microbial groups have never had a representative reference genome (Figure 2; Joint et al. 2009 this report). Even very deeply sequenced metagenome projects have rarely produced complete genome assembly of even the most abundant microorganisms. The lack of completed genomes poses an obstacle in determining ecological roles of the uncultured microorganisms and in deconstructing metabolic pathways in a given environment. However, recent advances in single-cell sorting and single-cell amplification and sequencing from environmental samples hold tremendous promise for increasing our abilities to study ecology and potential function of uncultured microorganisms (Ishoey et al. 2008) collected from targeted environments.

Microbial biodiversity in Southern Ocean, *Dave Caron*

Recommendations

1. Considering all components of the microbial fraction (bacterial, archaeal, eukaryotic and viral fractions) in assessments of diversity, so that we stand the best chance of evaluating specific aspects of the structure and function of marine microbial assemblages. Taking this approach is sure to yield unprecedented discoveries and an alteration in our perception of community function in response to environmental change.
2. Expand genomic and metagenomic efforts to continue to explore undersampled marine habitats (e.g. brines, deep sea and high latitude ecosystems) and organisms (e.g. viruses, protists and marine microbial symbionts).
3. Use metagenomic and other 'omic' approaches to empower understanding of marine ecology through sampling on scales and gradients relevant to the ecosystem of interest. Workshop participants emphasized the need to apply integrated approaches including metagenomics, transcriptomics, and metabolomics to studies of oceanographic, biogeochemical, and environmental processes (see also Karl 2007). Use of these tools in carefully designed experimental and hypothesis-driven programs that incorporate sampling on appropriate temporal and spatial scales are necessary for understanding the complex interplay between genes, organisms, communities and the environment, as well as the properties revealed that regulate global biogeochemical cycles. These approaches will advance our general perspective on microbial ecology and evolution and allow us to determine the biological dynamics that mediate the flux of matter and energy in the world's oceans.
4. Further development of single cell genomics ("getting more from less") to access ecologically important community constituents that have so far resisted cultivation efforts. These technologies will be beneficial in increasing our sequenced genome reference libraries. Along side of efforts to culturing organisms that we think are important in marine ecosystems (e.g. Konnecke et al. 2005; summarized by Joint et al. 2009 this report), directly isolating and sequencing ecologically relevant (abundant, or those that perform key ecosystem services) microbes is a promising advancement that provides significant opportunities to directly study genomic potential of unculturable microbes.

Literature Cited

Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ,
Finazzi G, Fernie AR, and Bowler C. 2008. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. Proc. Natl. Acad. Sci. U. S. A. 105: 10438–10443.

Baldauf, SL. 2003. The deep roots of eukaryotes. Science 300: 1703–1706.

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN, DeLong EF. 2000. Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea. Science 289: 1902-1906
Breitbart, M, Thompson LR, Suttle CA, and Sullivan M. 2007. Exploring the vast diversity of marine viruses. Oceanogr 20: 135-139.

Caron, DA, Worden AZ, Countway PD, Demir E, Heidelberg KB. 2008. Protists are microbes too: a perspective. Int. Soc. Microbial. Ecology. ISME J. advance online publication, November 13, 2008; doi:10.1038/ismej.2008.101

Daffonchio D, Borin S, Brusa T, Brusetti L, van der Wielen PW, Bolhuis H, Yakimov MM, D'Auria G, Giuliano L, Marty D, Tamburini C, McGenity TJ, Hallsworth JE, Sass AM, Timmis KN, Tselepides A, de Lange GJ, Hübner A, Thomson J, Varnavas SP, Gasparoni F, Gerber HW, Malinverno E, Corselli C, Garcin J, McKew B, Golyshin PN, Lampadariou N, Polymenakou P, Calore D, Cenedese S, Zanon F, Hoog S; Biodeep Scientific Party. Stratified prokaryote network in the oxic-anoxic transition of a deep-sea halocline. Nature 440:203-7.

DeLong, EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U, Martinez A, Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl, DM. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science 311: 496-503.

Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, Bult C, Tomb J, Dougherty B, Merrick J (1995). Whole-gemone random sequencing and assembly of Haemophilus influenzae Rd. Science 269: 496–512.

Giovannoni, SJ, Stingl U. 2007. The importance of culturing bacterioplankton in the 'omics' age. Nat. Rev. Microbiol. 5:820-6.

Grzymski, JJ, Murray AE, Campbell B, Kaplarevic M, Gao G, Lee C, Daniel R, Ghadiri, A, Feldman R, and Cary S (2008). Metagenome analysis of an extreme microbial symbiosis reveals eurythermal adaptation and metabolic flexibility. Proc. Natl. Acad. Sci U. S. A. 105, 17516-17521.

Ishoey, T. Woyke T, Stepanauskas R, Novotny M, and Lasken RS. 2008. Genomic sequencing of single microbial cells from environmental samples. Current Opinion Microbiol. 11: 198-204.

Karl, DM. 2007. Microbial oceanography: paradigms, processes and promise. Nature Rev. Microbiol. 5: 759-769.

Konneke, M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, and Stahl DA. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437: 543-546.

Lane RS, Burgdorfer W, Hayes SF, and Barbour AG. 1985. Isolation of a spirochete from the soft tick, *Ornithodoros coriaceus* a possible agent of epizootic bovine abortion. Science 230: 85–87.

Little, AEF, Robinson CJ, Peterson SB, Raffa KF and Headelsman J. 2008. Rules of Engagement: Interspecies interactions that regulate microbial communities. Annu. Rev. Microbiol. 62: 375-401.

Murray, AE, and Grzymski JJ. 2007. Diversity and genomics of Antarctic marine micro-organisms. Phil. Trans. Royal Soc. Biol. Sci. B 362: 2259-2271.

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS, Tolonen A, Webb EA, Zinser ER, and Chisholm SW. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424: 1042-1047.

Rusch, DB, AL Halpern, G. Sutton, KB Heidelberg, et al. 2007. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. 5: e77.

Schleper, C, Swanson, RV, Mathur EJ, and DeLong, EF. 1997. Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon Cenarchaeum symbiosum. J Bacteriol 179: 7803-7811.

Sogin, ML, Morrison HG, Huber JA, DM Welch, Huse SM, Neal PR, Arrieta JM, and Herndl JG. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Nat. Acad. Sci. U. S. A. 103: 12115-12120.

Staley, JT, and Gosink JJ. 1999. Poles apart: Biodiversity and biogeography of sea ice bacteria. Annu. Rev. Microbiol. 53: 189-215.

Stein, JL, Marsh TL, Wu KY, Shizuya, H, and DeLong, EF. 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J. Bacteriol. 178: 591-599.

Sullivan, MB, Coleman ML, Weigele P, Rohwer F , and Chisholm, SW. 2005. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol 3:e144.

Tringe, SG, Mering Cv, Kobayashi A, SAalamov, AA., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., et al. (2005). Comparative metagenomics of microbial communities. Science 308: 554-557.

Tyson GW, J Chapman, P Hugenholtz , EE Allen, RJ Ram, PM Richardson, VV Solovyev, EM Rubin, DS Rokhsar, JF Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37-43

Venter, JC, K Remington, JF Heidelberg, AL Halpern, et al. 2004. Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science 304: 66-74.

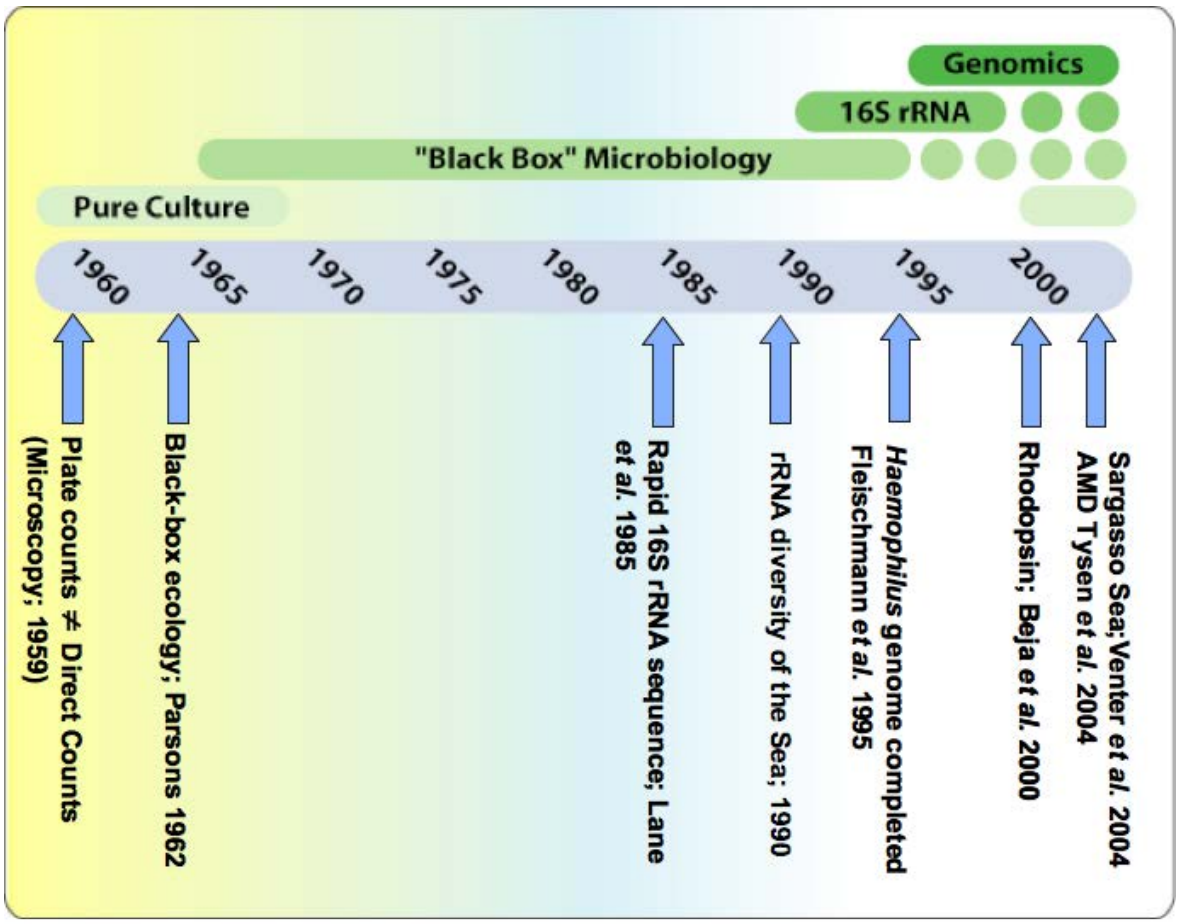


Figure 1.

Selected milestones in marine microbial ecology. The time line of a few of the advances and discoveries that have influenced marine microbiology. Many important contributions could not be included simply owing to space limitations (Figure J. Heidelberg).

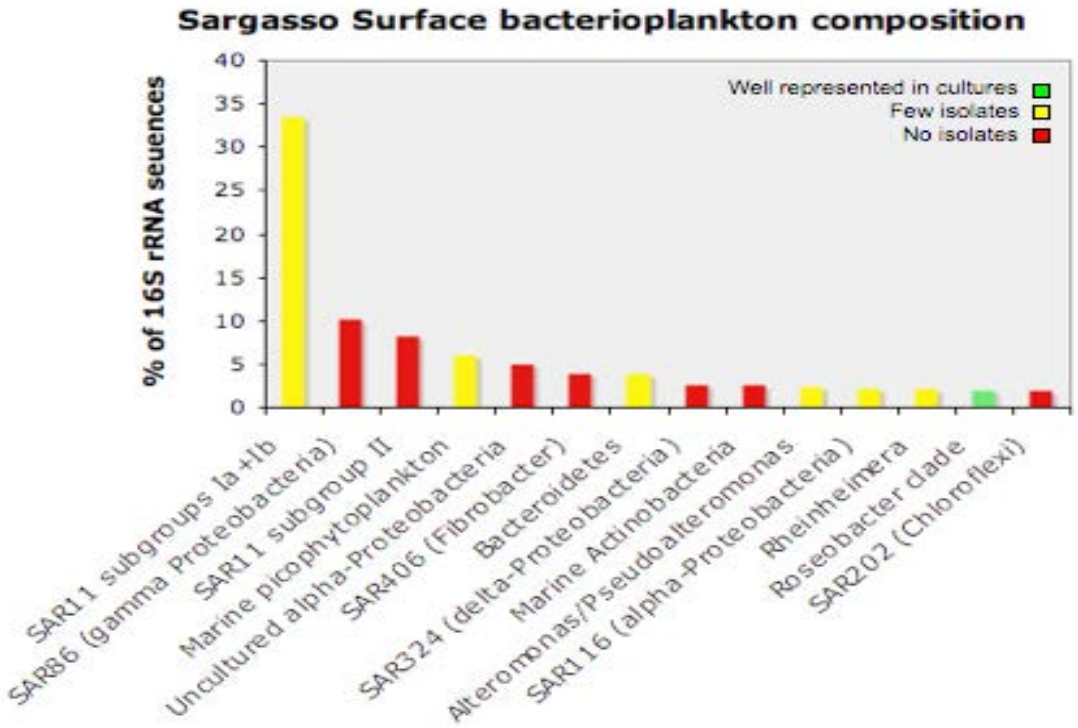


Figure 2.

A representation of organisms in culture compared to occurrence in a natural assemblage found in Sargasso Sea (Venter et al. 2004). Advances in the capabilities for single cell sorting and amplification offer significantly improved opportunities to sequence naturally occurring cells which cannot be brought into culture. (Figure modified from Giovannoni and Stingle 2007)

S E S S I O N - 2

Functional (Meta) Genomics of Marine Microorganisms

Ian Joint[1], Gurvan Michel[2], Christine Klockow[3], Granger Sutton[4], Cinzia Verde[5], Thomas Schweder[6] and Chris Bowler[7].

[1]Plymouth Marine Laboratory, UK, [2]Station Biologique de Roscoff, National Center for Scientific Research/ Pierre & Marie Curie University (CNRS/UPMC), Roscoff, France, [3]Max Planck Institute for Marine Microbiology, Bremen, Germany, [4] Rockville, MD, USA, [5]CNR-Institute of Protein Biochemistry, Naples, Italy, [6]Institute for Marine Biotechnology, Greifswald, Germany, [7]National Center for Scientific Research (CNRS), UMR 8186, Biologie Moléculaire des Organismes Photosynthétiques , Paris, France

Background

Marine microbes are important for the Earth System because they control the cycling of elements in the oceans. Autotrophic processes fix carbon and release oxygen; heterotrophic processes result in the recycling of nitrogen, sulfur and phosphorus and other elements. Bacterial metabolism is involved in the chemical transformation of most elements. About half of the annual primary production of the planet occurs in the ocean so the marine ecosystem plays a very important role in maintaining the wellbeing of our global environment.

Despite the obvious importance of marine microbes (which include both prokaryotic and eukaryotic plankton), very little is known of their diversity, how many species are present in the oceans, and what each individual species does – i.e. its ecological function. Until recently, there were no appropriate techniques available to answer these important questions. The vast majority of these organisms cannot be cultured in the laboratory and so were not amenable to study by the methods that had proved so successful with medically-important microorganisms throughout the 20th century. It was only with the development of high-throughput technology to sequence DNA from the natural marine environment that information began to accumulate that demonstrated the exceptional diversity of microbes in the oceans – in fact, most marine microbes are entirely novel and have not previously been described. Even less is known about their function in the ecosystem or metabolic activity since no function can be assigned to the major part of their genes. Marine microbial assemblages are diverse and unique and the challenge is to discover what functions are played by these microorganisms.

Metagenomics

It is generally expected that genomics and metagenomics will provide answers to these questions. Genomics can be defined as the study of the genetic complement of a single organism. Metagenomics refers to all of the genetic information of a natural assemblage – i.e. equivalent to the genomes of all of the organisms in the sample (NAS, 2007). There have been rapid advances in the technology of DNA sequencing which has resulted in an explosion of information on marine microbes. For example, the first part of the Global

Ocean Survey (GOS), which sampled the North Atlantic, Caribbean and a small part of the Pacific Ocean, added DNA sequence information that was equivalent to 50% of all protein-encoding sequences that had previously been deposited in GenBank. GOS confirmed that marine microbes are diverse; indeed it revealed how little is known about the genetic information of natural assemblages. This study highlighted the difficulties of making sense of metagenomic sequence data. A significant proportion of the open reading frames (ORFs, which are presumed to equate to genes) could not be characterized because there were no similar sequences in the databases.

This difficulty of interpreting the GOS sequence data exists despite the large number of marine microbes whose whole genome sequences are already in the databases. Largely as a consequence of funding from the Gordon and Betty Moore Foundation (www.moore.org), 155 marine bacterial genomes have so far been sequenced.

So, although marine bacteria are well represented in the genomic databases, this information was still not sufficient to decipher the metagenomic information coming from the GOS. The situation is even more complex for eukaryotic microorganisms, which have larger and more complex genomes than bacteria. Also, fewer genome sequences are available for phytoplankton than bacterial species, which increases the difficulties of ascribing function to genetic sequence for eukaryotic microbes.

Metatranscriptomics

Nevertheless, there is optimism in the oceanographic community that metagenomics will provide new insights into the microorganisms that are present in the oceans. In order to obtain information on microbial function, and especially the ways in which microbial assemblages might respond to changing environmental conditions, researchers are applying the same high-throughput sequencing techniques that have worked so well with metagenomics, to the study of metatranscriptomics. This involves sequencing mRNA isolated from complex communities, and synthesizing cDNA that can then be sequenced. Metatranscriptomics has the potential to describe how metabolic activity of an assemblage will change under different conditions in the ocean by revealing differences in both known and previously unknown transcripts in natural communities. Methods have now been published (Frias-Lopez et al. 2008: Gilbert et al., 2008) that allow synthesis of high quality cDNA from mRNA extracted from natural assemblages. The cDNA can then be sequenced to indicate how the transcription profile (the metatranscriptome) of communities differs and allows the immediate response to be determined of an assemblage to environmental change. Furthermore, the use of metatranscriptomics to explore gene function in eukaryotes is preferable to metagenomics because the method focuses on expressed gene repertoires rather than whole genomes, which typically contain large amounts of non-coding and therefore difficult to interpret sequences. Coupled with time-series experiments, preferential at long term ecological research sites, metatranscriptomics can help to unravel the functionality of microbial communities and to monitor seasonal changes.

Proteomics

Another key technique to investigate functional genome analyses of marine microorganisms is proteomics (Schweder et al. 2008). In contrast to metagenomics and

metatranscriptomics, proteomics has so far been most useful with bacteria that can be cultured under defined environmental conditions. It gives valuable information on the physiology of individual species and has been widely used to investigate how bacteria respond to stress and starvation conditions. The approach has improved understanding of, for example, regulatory networks and physiological strategies which ensure the survival under life-threatening environmental conditions. For example, proteomics has been successfully applied to physiological analysis of an uncultured bacterial endosymbiont from a deep sea tube worm (Markert et al. 2007). New proteomic techniques allow direct determination of the physiology of key marine bacteria and are thus valuable tools for future functional genome analyses. A proteomic view of cell physiology reaches beyond the mere prediction of putative metabolic functions as coded in the genome sequence. The extreme conditions of the polar regions also provide examples of environments in which proteins have evolved to operate efficiently at very low temperatures, ensuring that microorganisms survive in extreme habitats. However "polar" genomics and proteomics studies are still in their infancy, and there is a very small database of DNA sequences from polar regions, but there is much information on protein structure and function. Acquiring data on the genome, gene expression, protein structure and function in polar species is the basis for understanding the evolutionary forces operating at sub-zero temperature. Any prediction of the physiological costs and evolutionary consequences of global warming is strictly dependent on the knowledge gained on the structure and functioning of polar ecosystems. More important, life sciences are not the only area gaining key insights from studying biological communities inhabiting the poles, because of the strong linkage between organisms and the oceans and atmosphere.

Current challenges in marine genomics, metagenomics and metatrascriptomics

1. *Lack of data from relevant model organisms* - The wealth of sequence data from both marine microbes and diverse oceanic provinces is presenting considerable challenges. There are huge numbers of putative genes, the function of which is often unknown and at best only deduced from sequence comparisons. Because more is often known about the genetics and physiology of terrestrial organisms, the number of unknown/putative genes is overwhelming for marine samples because there is so little experimental data on marine model organisms. For example, all phyla of marine algae synthesise sulphated polysaccharides that have no equivalent in land plants and most of these enzymes constitute completely new protein families: i- and l-carrageenases (Michel et al. 2003), a-agarases (Flament et al. 2007) or fucanases (Colin, 2006). It is not possible to gain any useful information on these proteins by genomics approaches because the sequence data do not exist. It was only through the application of standard biochemical approaches that these enzymes have been identified, otherwise, they would have been annotated as "conserved hypothetical proteins" or given incorrect substrate specificities in genome annotation.
2. *The need for more relevant marine model organisms* - There is an urgent need for more cultures of marine bacteria, archaea, viruses, protozoa and phytoplankton. Most culture collections are based on readily cultivated microbes. When these organisms were isolated, there were no techniques to establish if the isolate was abundant in the natural environment or even if it had any relevant function. Molecular biology has

changed that and the isolation of new cultivable microbes can now be based on their abundance and relevance in defined marine habitats. There are a number of novel and innovative approaches to the isolation of new potential-model microorganisms. For example, Rappé et al. (2002) used a dilution approach to isolate SAR11 – the bacterium whose 16S sequence has world-wide distribution (the isolates are now referred to as "*Candidatus Pelagibacter ubique*"). Zengler et al. (2002) described a method of encapsulation of individual bacterial cells, which meant that slow-growing cells could be cultured without being overgrown by rapidly dividing species. So methods exist for isolating useful model microbes from the natural environment; but these are not high-throughput systems and are labour-intensive. Nevertheless, they are probably the only way in which relevant bacteria can be brought into culture since classical microbiology methods have not proved to be useful for difficult-to-cultivate microbes. There is also a need to develop forward and reverse genetics techniques and other molecular resources for relevant marine model organisms. There are still too few examples of phytoplankton that can be manipulated in this way, and without such methods it will be difficult to explore the function of the thousands of genes found only in these organisms.

3. *Genomics of novel model microbes - Rhodopirellula baltica* provides an example of an environmentally relevant marine bacterium whose genome has recently been sequenced (Glöckner et al., 2003). *Rhodopirellula baltica* is a marine planctomycete isolated from the water column in the Baltic Sea. Genomic analysis has revealed many fascinating and rare features, such as a high number of sulphatases, genes for a C1 metabolism and a global mechanism of gene regulation. But, as with all genomes so far sequenced, function is unknown for a large proportion (~50%) of the genes. Being in culture, it is possible to change growth conditions in a very defined way and investigate how the transcriptome changes in response. Hence it is possible to unravel gene functions and add valuable information about how the microorganism adapts to changing environmental influences.
4. *Bioinformatics is currently a bottleneck* - Although the giga-base amounts of microbial DNA sequences and other high-throughput approaches have made fundamental improvements to our understanding of uncultivated marine microbes, bioinformatics is often the limiting factor in metagenomic and metatranscriptomic studies. The major hurdles are still (1) the computational aspects of data archiving, analysis and visualization of thousands of millions of DNA sequences which are released to databases and (2) integrating sequences from environmental samples with experimental studies so that unknown genes can be assigned a function. Novel techniques are required that would allow a numerical description of the specific biological functions unique to specific niches and acting against particular elements.

Since we are primarily concerned with establishing the function of microorganisms in the oceans, it is important to be able to characterise protein function from environmental DNA sequence data. The primary method to assign function to a predicted open reading frame (ORF) is by establishing homology to a protein or protein family whose function has been well characterised experimentally, usually in non-marine systems such as mouse, yeast or *Arabidopsis*. Once function has been assigned, it can be mapped to metabolic

pathways or proteins involved in a particular process, to determine the functional activity in the environmental sample. To this end, it is helpful to generate longer contiguous DNA sequences – and third generation sequencing technologies should be capable of generating sequence reads of tens of kilobases. Nevertheless, it is a challenge for present computational methods to assemble the metagenomic shotgun sequences, particularly in environments that support a high level of sequence diversity. That is, a natural assemblage is much more diverse than a clonal isolate and assembly of longer DNA sequences is difficult. New computational techniques are also needed to aid in assigning functions to the millions of marine genes that currently have no known function.

Visualization tools are being developed to display fragment recruitment, genomic context, functional annotation, scaffold characteristics for binning, metabolic pathway overlays, and sample comparisons. These tools will be crucial if biologists are to utilise the large and rapidly expanding datasets that potentially hold the key to understanding microbial function in the oceans.

Antartica sunset, *Karla Heidelberg*

Literature Cited

Colin, S, Deniaud E, Jam M, Descamps V, Chevolot Y, Kervarec N, Yvin JC, BarbeyronT, Michel G, and Kloareg B .2006. Cloning and biochemical characterization of the fucanase FcnA: definition of a novel glycoside hydrolase family specific for sulfated fucans. Glycobiology 16: 1021-1032.

Flament D, Barbeyron T, Jam M, Potin P, Czjzek M, Kloareg B, and Michel G 2007. Alpha-agarases define a new family of glycoside hydrolases, distinct from beta-agarase families. Appl. Environ. Microbiol. 73: 4691-4694.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, et al. 2008 Microbial community gene expression in ocean surface waters. Proc. Natl. Acad. Sci. U. S. A. 105: 3805-3810.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P, and Joint I .2008. Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. PLoS ONE 3(8): e3042. doi:10.1371/journal.pone.0003042.

Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, *et al*. .2003.. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc Natl. Acad. Sci. U. S. A. 100: 8298-8303.

Markert S, Arndt C, Felbeck H, et al. .2007. Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. Science. 315: 247-50.

Michel G, Helbert W, Kahn R, Dideberg O, and Kloareg B . 2003. The structural bases of the processive degradation of iota-carrageenan, a main cell wall polysaccharide of red algae. J. Mol. Biol. 334: 421-433.

NAS. 2007. The new science of metagenomics: revealing the secrets of our microbial planet. ISBN: 0-309-10677-X 170pp http://www.nap.edu/catalog/11902.html.

Rappé MS, Connon SA, Vergin KL, and Giovannoni SJ. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418: 630-633.

Schweder T, Markert S, and Hecker M. 2008. Proteomics of marine bacteria. Electrophoresis. 29, 2603-16.

Zengler K, Toledo G, Rappé M, Mathur EJ, Short JM, and Keller M. 2002. Cultivating the uncultured. Proc. Natl. Acad. Sci. U. S. A. 99: 15681-15686.

S E S S I O N - 3

Biocatalysis, Drug Discovery and the Commercial Application of Marine Organisms

Dermot Hurst[1], Wolfgang Aehle[2], Kevin Sowers[3], Peter Golyshin[4], Fernando de la Calle Verdu[5], Diaa Youssef[6]

[1]Marine Institute, Ireland; [2]Brain GmbH, Zwingenberg, Germany; [3]University of Maryland Biotechnology Institute, Baltimore, MD, USA; [4]University of Bangor, Bangor, UK; [5]PharmaMar, Madrid, Spain; [6]Suez Canal University, Ismailia, Egypt

Background

The use of living organisms to perform chemical transformations predates historical records. The action of a eukaryotic microorganism, yeast, in converting the sugars found in grape juice, forms alcohol and hence wine. Likewise in bread baking, the carbon dioxide formed during the fermentation of sugars in the dough, causes the dough to rise. The use of enzymes or whole cell systems to convert a readily available material into a value added product is termed biocatalysis. Microorganisms are widely recognized and used as a source of novel enzymes, biocatalysts and even bioactive compounds. Until quite recently, terrestrial microorganisms were the most widely used in industrial processes. The search process for novelty has expanded; target habitats spreading from soil, desert sand and the rumen of cattle and other ruminants, to the marine.

Marine environments are diverse, and organisms that occupy the many ocean niches are exposed to various extremes of pressure, temperature, salinity and available nutrients. Evolutionary development has equipped marine organisms with mechanisms to help them survive essentially hostile environments. Not only are such niches likely to yield diverse microbial communities: these organisms are also likely to possess uniquely diverse genetic, biochemical and physiological characteristics.

Commercial interests in marine micro-organisms exist because of the potential ability to exploit the enzymes they produce. There are expectations that enzymes produced by marine microbes might have novel bio-catalytic activity that allows them to function under extreme conditions. Such properties account for the interest shown in marine enzymes by different industry sectors; where they are recognized as potentially beneficial and as likely candidates to replace more traditional terrestrial organisms as the next generation catalysts. The food and detergent sectors tend to concentrate on a limited number of enzyme reactions and fewer substrates for their products. The market and production challenges for the fine chemical and pharmaceutical sectors, faced with far more chemically and structurally diverse molecules, are greater. Firms in these sectors require customized, individual enzymatic solutions to produce their products. Optimism that marine microbes are a potential source of novel enzymes is high following the identification of a range of enzymatic activity from cultured marine microbes. However, this optimism is somewhat tempered, since as in the case of terrestrial organisms, few marine organisms can be cultured by conventional processes.

The advent of molecular tools has substantially changed the classical view of biotechnology and opened up a new vista of marine biotechnology. Faced with a universe of DNA sequences, and organisms that cannot be laboratory cultured, alternative processing techniques are required. Today we are beginning to explore and expand the use of metagenomic approaches to isolate and identify novel enzymes from unculturable marine microbial communities. As a result of using this innovative genetic based technology, the number of new compounds with applications in human health, chemicals, food and an array of industrial products is set to increase.

The cultivation challenge

Most enzymes currently used in biotechnology are of microbial origin. The microbial world contains the greatest fraction of biodiversity in the biosphere. It is thought that the marine environment, covering more than 70 percent of the earth's surface, contains ~4 x 10^{30} microorganisms. Commercial expectations are that microbes will deliver the greater part of enzyme diversity and the majority of new applications. However, the well-known dilemma of microbes, whatever their origin – that the majority cannot be cultivated – limits the application of the traditional means of enzyme discovery. There are few options at present that can effectively overcome the cultivation challenge. It is always possible to devote more effort to try and cultivate these organisms; however, experience indicates a low probability of any success in following this route. The current alternatives focus on exploiting surrogate culturable microbes and genomics to express functions of interest. This is achievable to varying degrees by mining sequenced genomes; generating and screening genomic libraries; and by generating and screening libraries of the genetic resources of the microbial biosphere (metagenomics).

The concept of an "activity-first" approach in mining genomic resources from individual organisms and their communities complements and extends the "genome-gazing" approach in the search for new enzymes. This approach was successful in discovering new enzymes from marine and terrestrial species that *in silico* analysis alone could not predict. The main sources were the deep hypersaline anoxic basins of the Mediterranean Sea; the rumen of cows; and the ubiquitous marine hydrocarbon oil degrading microbes *Alcanivorax borkumensis*, *Oleispira antarctica*, acidophilic archaeon and *Ferroplasma acidiphilum*. A wealth of unique enzymes were discovered and characterised from these microbes and from cellulose-degrading microbial communities found in the earthworm, including carboxylesterases, glycosylhydrolases, dehalogenases, and polyphenoloxidases. The results from experiments such as these indicate good potential for microorganisms and microbial communities from extreme environments for chiral synthesis of drug precursors, biomass conversion, and fine chemical production.

Industrial enzymes

Enzymes are important ingredients of analytical systems for the chemical and food industry or in health care, where enzymes act as therapeutics as well. They also have many other industrial applications including in the textile, grain milling, pulp and paper, food and detergents industry sectors. Applications for enzymes are wide and span the pH-scale from 1 to 14. The temperature range of enzyme usage in aqueous solution starts at the melting point and goes above boiling point in pressurized systems. Some enzymes have to operate in organic solvents when being used as catalysts for the synthesis of fine chemicals and others, as in the case of detergents, have to work at low temperatures.

The discovery of a new enzyme or enzyme functionality for industrial applications is only the beginning of a time consuming development process. Almost every new enzyme fails to completely fulfill the various requirements of industrial processes. In many cases protein engineering techniques are applied in order to improve stability, catalytic efficiency, stereo selectivity or process compatibility in general. In recent times, the use of biotransformation in an industrial environment increased rapidly as the benefits of biocatalysts were realized.

Increasingly, different industry sectors look to exploit uncultivated organisms, and generally, their origin is of little interest to industrial customers and the end user. The pharmaceutical sector needs only minute quantities from which to synthesise kilograms of chiral synthons as building blocks in the production of new drugs. Bulk industrial products such as high-performance detergents require kilo tons per annum to meet market demand.

Metagenomics opens the possibility for industrial enzyme producers to exploit and capture the diversity of marine organisms and make use of them in a wide range of applications. Different industries have widely varying requirements concerning enzyme function. There are however, characteristics that users of uncultivated microorganisms seek out in their search for the ideal biocatalyst. In addition to wanting an enzyme that matches process requirements perfectly, they seek novelty, freedom from any intellectual property restrictions which might limit or otherwise scale back the use of the enzyme and multiple and diverse biocatalysts on which to base their own biotransformation tool boxes.

To allow economically feasible production, more than 90% of the production of industrial enzymes takes place in highly optimized recombinant microbial expression hosts. Industry is in the possession of very few of these hosts; production plants are therefore tailor-made for enzyme production in those hosts. As is the case in the process of drug discovery, the development processes employed in the enzyme sector are also complex and costly. Key technical production goals for new enzymes include the maintenance of high yield and purity and ensuring production processes compatibility with existing fermentation equipment.

Drug discovery processes

Irrespective of the source material, modern drug discovery is a hugely complex and costly process that has no certainty regarding successful commercial or clinical outcome. The typical drug discovery path is serial, starting with the identification of a molecule that promises pharmaceutical activity. Subject to successful development, it can take up to 12 years to reach the stage of having a commercially viable drug. With the advent of high throughput screening it is possible to assess the potential of novel compounds from a wide range of biological sources. To expand the range of source organisms and discover new compounds, the marine environment is the target of bioprospecting activity by elements of the international pharmaceutical sector. Commercial entities and public research institutions are now active in the study of extracts from marine invertebrates, microorganisms and recently, exploring DNA from non-cultivable organisms.

The conversion of a bioactive molecule into a medicine is also a long and risky process. Success rates are low and typically just one molecule in 10,000 will complete all development stages to reach the market. This obviously involves an astronomic investment in research and development. The processes involved include the identification and validation of new targets, drug discovery, medicinal chemistry and drug delivery. To ensure the future supply, chemical synthesis and

biotechnology are the preferred sources for manufacturing. This was the case for the anticancer medicine Yondelis® (PharmaMar, Spain) isolated from a marine tunicate (*Ecteinascidia turbinata*), that is currently manufactured using a hemi-synthetic process. Natural collections are typically used only for drug discovery purposes, where only small amounts (milligram scale) of the pure compound are enough to elucidate the chemical structure and the biological activity.

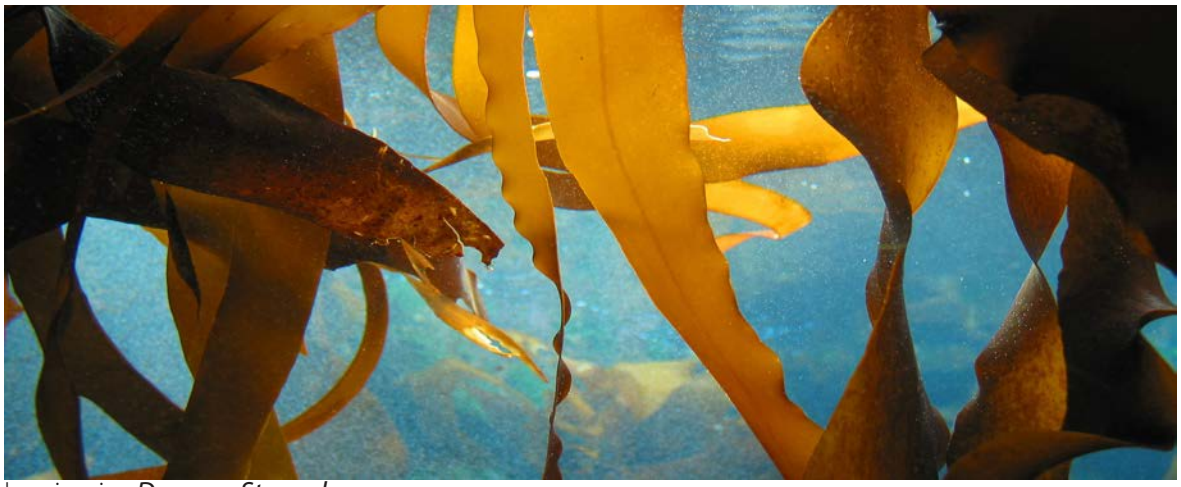
Pharmaceutical companies constantly seek new compounds to assess for drug potential. Some firms like PharmaMar target cancer care using marine derived drugs. Other firms seek to exploit marine compounds that show diverse biological activities including, antiangiogenesis, antiproliferation, anti-inflammatory, antimicrobial, anti-PLA2, and anti-tuberculosis.

Marine microbes in action

In addition to natural organohalides prevalent in marine fauna and flora, the oceans are the ultimate global sinks for persistent organic pollutants such as polychlorinated biphenyls (PCBs) that threaten the health of wildlife and humans due to their accumulation in the food chain and the potential diseases that they may cause. The process of reductive dehalorespiration of PCBs is known to occur in the environment, but the microbial catalysts eluded identification.

An approach that combined classical enrichment protocols with molecular polymerase chain reaction (PCR)-based monitoring of the microbial 16S rDNA communities identified these anaerobic PCB dehalorespiring microorganisms for the first time. Within Chloroflexi, exists the microbial catalyst, *Dehalobium chlorocoercia* DF-1, which is closely related to another species *Dehalococcoides ethenogenes*, that dehalorespires other organohalides.

Using PCR primers developed specifically for these microorganisms it is possible to detect PCB dechlorinating bacteria in sediments. Being able to examine these microbes gives an understanding of their distribution in the environment and of their role in the marine halogen cycle. From a biotechnology perspective, anaerobic dechlorination of anthropogenic PCBs is a critical step in the biodegradation of highly chlorinated congeners commonly detected in the environment. A dearth of isolates for physiological studies, and the inability to detect and monitor the microorganisms in the environment, impeded the development of in situ treatment. The development of a tractable *in situ* biotreatment system is now plausible with the ability to grow these anaerobic PCB transforming bacteria in culture and selectively detect dehalogenating populations in the environment. Development of new approaches for culturing and functional characterisation of marine microbes are critical for identifying biocatalysts with novel capabilities from the environment.



Laminaria, *Dagmar Stengel*

CHALLENGES

The need for new genomic tools

High throughput environmental metagenomics, proteomic and transcriptomics combined with bioinformatics have generated the most comprehensive knowledge of marine microbial communities and processes to date. However, the extent of information generated has exceeded our ability to assign physiological function to microorganisms that have not been isolated and characterized. We must be cognizant also of instances where microorganisms will elude detection by metagenomic approaches or subsequent high throughput isolation methods may be ineffective.

Small versus large molecules

The exploration of marine meta/genomics for small molecules is in its infancy. Whilst important advances have been made, such as the use of BACs as a vector to clone large size of environmental DNA, improved microorganisms for the heterologous expression (apart from *Escherichia coli*, *Streptomyces*, *Pseudomonas*, *Bacillus subtilis* and *Rhodococcus*), and new "specific" primers designed to look for polyketide synthase (PKS) sequences in genomes, other issues remain to resolved. There is a need for the design of a robust process for screening genomic libraries, both at the level of functionality-based screening and sequence-based screening (PCR). A particular problem regarding small molecules is that for the biosynthesis of a peptide or polyketide (400-2000 mw) several large proteins (nonribosomal peptide synthase or PKS enzymes) are necessary. Problems also frequently exist with regard to the heterologous transcription of the large gene clusters (30-200 kb) that encode these enzymes.

Awareness of intellectual property

Most research institutions have formal policies and procedures concerning the management of intellectual property (IP). Differences exist between the USA and the EU in relation to patent law. In general, raw products of nature are not patentable. DNA products usually become patentable when they have been isolated, purified, or modified to produce a unique form not found in nature. Working from a common understanding concerning international patent law as it applies to biological materials would be helpful in building international research partnerships and in dealings with industrial partners/clients.

Addressing constraints on the collection and use of marine organisms

International conventions exist regarding marine scientific research (MSR) and there are also local country conventions. The principal conventions are those established under the International Council for the Exploration of the Seas (ICES) and the OSPAR Commission. As efforts increase to collect biological and other samples from the sea, MSR will expand. Researchers need to develop a greater awareness of the obligations and rights provided for in international agreements and to be aware of individual country rights when planning cruises and or participating in sample collection.

Greater collaboration between researchers across the Atlantic

Policy that seeks to encourage the development of research partnerships between the USA and Europe exists. There is a divergence of policy from reality when it comes to supporting research projects that involves researchers that are outside the jurisdiction. A higher level of coordination

between the EU and the USA in developing research priorities for marine biotechnology would help align both policy and research funds to address common research challenges. Mechanisms that allow for co-funding between the USA and EU could help in enabling the exchange of research staff, training and joint research projects.

Knowledge of factors that influence industrial activity

Despite initiatives designed to encourage interaction between industry and institutional based researchers, there are gaps in understanding concerning the goals of each. Whilst the development of a new drug can take up to 12 years of more from basic discovery to market launch, industry may also have shorter term scientific goals. Industry priorities vary from sector to sector. However, there are common issues which affect relationships between research partners, these relate to publication and ownership of research results; commercialisation rights; how projects are managed and ethical issues etc. A possible solution to potential conflict is the use of formal agreements, drawn up in advance of commencing research projects.

Overcoming industry concerns

Justifying, and investing in molecular technology to explore the genomic structure of marine organisms (whether they can be cultivated or not) is both complex and costly; even more so if the goal is to obtain results in the short term. Very few research groups are focused on short-term rapid evaluation of marine organisms. Environmental analysis, bioenergy and other enzyme seeking disciplines are driving the application of marine metagenomics; however, (bio) pharmaceutical companies remain to be convinced to maximise the use of this tool in support of their drug discovery programs. Theoretically metagenomics has great potential in supporting marine origin drug discovery; however, the private sector needs to see real examples of success before committing to make major investments in research in this area.

Pelvetia, *Dagmar Stengel*

Literature Cited

Beloqui A et al. 2006. Novel Polyphenol Oxidase Mined from a Metagenome Expression Library of Bovine Rumen**:** Biochemical properties, structural analysis, and phylogenetic relationships. J. Biol. Chem. 281: 22933-22942.

Ferrer et al. .2005. Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environmental Microbiology, 7: 1996-2010.

Ferrer et al. .2005. Microbial Enzymes Mined from the Urania Deep-Sea Hypersaline Anoxic Basin. Chem. Biol. 12: 895-904.

Lorenz P and Eck J . 2005. Metagenomics and industrial applications. Nature Rev. Microbiol. 3:510-516.

Ferrer et al..2007. The cellular machinery of *Ferroplasma acidiphilum* is iron-protein-dominated. Nature, 445: 91-94.

Ferrer et al. 2008. A purple acidophilic di-ferric DNA ligase from *Ferroplasma*. Proc. Natl. Acad. U. S. A. 105:8878-8883.

Kennedy J, Marchesi J R and Dobson A. 2008. Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. Microb. Cell Factories 7: 27.

Langer M et al. .2006. Metagenomics: An inexhaustible access to nature's diversity. J. Biotechnol. 1: 815-821.

Schneiker et al. 2006. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. Nature Biotechnol. 24: 997-1004.

S E S S I O N - 4

Intellectual Property Rights of Marine Genomic Resources

Joel Querellou[1], Mohamed Amar[2],
Jean-François Bloch[3] and Margo Haygood[4]

[1]UMR6197, Ifremer Brest, France ; [2]Centre National pour la Recherche Scientifique et Technique (CNRST), Rabat, Morocco ; [3]Proteus, Nimes, France ; [4]Oregon Heath and Science University, Beaverton OR, USA.

Background

Property rights on marine genomic resources (MGR) are based on a set of international and national laws, conventions and agreements taking notably into account territories. Historically they relied on the ability to characterize the marine organisms by the name of exploited species. This approach was and is still efficient for macroscopic species directly collected in marine environments or produced by aquaculture techniques. However, for microbes, bacteria, archaea and picoeukarya, the vast majority are still uncultured and not amenable to culture with current knowledge. The use of metagenomics opens novel avenues to accessing corresponding genetic resources, while ignoring species names of the microorganisms present in the original samples and bypassing culture techniques. Metagenomics for biotechnological applications was initially developed in the USA in the mid nineties and applied to various types of environments including marine. The huge amount of novel genetic resources potentially available for biotech applications, represent a unique and historical opportunity to support the promotion of knowledge-based bio-economy. The recent technological advances give access, not only, to an unprecedented exploration of deep-sea ecosystems but allow revisiting genetic resources of coastal ecosystems all around the world.

Nevertheless, securing a bright future for marine genetic resources (Zewers, 2008) suppose that ownerships rights on crude samples and their species and gene contents as well as intellectual property on resulting products are well defined and subjected to international agreements. The management of marine genetic resources is therefore under high scrutiny at international level and the sovereignty over these resources, the extent of ownership rights and the patentability of inventions derived from them are important issues currently negotiated within the frame of the United Nations Informal Consultative Process on the Oceans and the Law of the Sea (UNICPOLOS).

It is worth mentioning that there is an increasing gap between the world of scientists undergoing permanent revolution (molecular>genomics>metagenomics> x-omics) and the world of lawyers subjected to international negotiations.

Sovereignty and ownership rights: the legacy of the Convention on Biological Diversity (CBD)

Although the issues related to oceans and marine resources are a constant concern of the United Nations (for more information: http://www.un.org/depts/los/index.htm), the 1982 United Nations Convention on the Law of the Sea did not address the issue of ownership

rights beyond areas of national jurisdictions. However, it addressed the role of research and stipulated (art. 241) that "Marine scientific research activities shall not constitute the legal basis for any claim to any part of the marine environment or its resources". The property of MGR collected within the frame of scientific cruises is therefore undefined.
The CBD, while clearly defining the rights under national jurisdictions, provided only guidelines for commercial fishing and research beyond national jurisdictions (International Zone or IZ). It left the issue of marine genetic resources in the IZ open for discussions and debates. In this field, the present situation is characterized by divergent approaches and interests between developing and developed countries (Zewers, 2008). Briefly summarized, developing countries argue that bioprospecting of marine genetic resources should be regulated and their exploitation carried out for the benefit of the common heritage of mankind, including benefit sharing. Obviously, mechanisms to implement such principles would require extensive negotiations and the establishment of an international body on the model of the International Seabed Authority (www.isa.org.jm/en/home). In contrast, developed countries have pleaded for freedom in bioprospecting based on the same principle applied to fishing: free access to marine genetic resources and ownership of the collected genetic resources and their derived products. This clearly reflects the current imbalance in fleets and equipment available for sampling the sea between the two groups of countries.

The objectives of the CBD are the conservation of biological diversity, the sustainable use of its components and the fair and equitable sharing of the benefits arising out of the utilization of genetic resources, including by appropriate access to genetic resources and by appropriate transfer of relevant technologies, taking into account all rights over those resources and to technologies, and by appropriate funding. States have the sovereign right to exploit their own resources pursuant to their own environmental policies. Each State shall endeavor to create conditions to facilitate access to genetic resources for environmentally sound uses by other States and not to impose restrictions that run counter to the objectives of this Convention. Each Contracting Party shall take legislative, administrative or policy measures with the aim of sharing in a fair and equitable way the results of research and development and the benefits arising from the commercial and other utilization of genetic resources with the Contracting Party providing such resources.

It is also important to note that the CBD has not been ratified by several countries, including the United States of America. In that case, exploitation of marine genetic resources in a third country is ruled on the basis of a strictly bilateral agreement.

Interestingly and in contradiction with the spirit of the agreement, certain signatory States impose limitations on the access to the marine genetic resources. In practice, it is very difficult for researchers or industrials to identify the relevant administration and agent of the state who put in practice the rules of the CBD, and especially the "fair benefit sharing" despite the fact that this point is crucial. Often, products issued from marine biotechnology are the results of research and development projects that include not only access to the genetic resource, but further processing steps like library screening, directed evolution, product characterization and optimization, conditioning, etc. Assigning the right value to each step, including access to the genetic resources, return on investment

in industrial processes, is not an easy task and requires case by case negotiations and guaranties of not infringing third parties patents.
Marine genetic resources accessible through metagenomics projects will represent a huge reservoir of novel genes, biomolecules and probably products in different type of applications ranging from bioremediation to pharmaceuticals. The CBD, though imperfect in its practical applications, offers an operational frame to define relations between the resource owner and the party in charge of product development and commercialization. Its adoption by nonsignatory countries is therefore recommended along with its ratification.

Patentability

The amount of litigation regarding patents related to marine genetic resources is unknown and probably negligible compared to the amount of litigation associated with human gene patents (Holman, 2008). Does that mean that everything is perfect in the field of patentability of marine genetic resources? Certainly not. Many questions are are either unanswered or the answers differ between Europe and the US.

European patents are granted for inventions in all fields of technology, provided that they are new, involve an inventive step and have an industrial application (Art. 52(1) EPC). Moreover for a European patent concerning a biotechnological invention including that associated with marine biotechnology, the patent shall be applied and interpreted in accordance with the provisions of Directive 98/44/EC of 6 July 1998 on the legal protection of biotechnological inventions. Under these provisions, genes can be patented (*Art. 3: "Biological material which is isolated from its natural environment or produced by means of a technical process may be the subject of an invention even if it previously occurred in nature*."). If genes have been isolated or technically produced, they are considered as chemical substance and therefore patentable, provided that (i) they are sufficiently structurally defined, (ii) they are neither known nor obvious and (iii) the inventor discloses how they can be obtained, and (iv) the inventor specifies the purpose of the invention. It is worth mentioning that no patent can be granted for an organism as it exists in nature.

The US patent system, although slightly different from the EP system, is also effective at protecting MGR inventions. The patent application must disclose the invention in a manner sufficiently clear and complete for it to be carried out by a person skilled in the art. In the past, this applied to metagenomics data and their derived products as demonstrated by the list of patents granted to Diversa Corporation (i.e. US patent n° 5763239) and others. In the future, this might change. The US Congress is considering legislation that would prospectively ban the patenting of not only human genes, but any nucleotide sequence, or its functions or correlations, or the naturally occurring products its specifies (Genomic Research and Accessibility Act, H.R. 977, 110th Congr. 2007, Fisher case). In a global world, when a new restrictive law is enforced in only one country, getting around the law is possible simply by working in another country and importing the resulting biotechnology product.

Both the EU and US systems protect, with minor variations, the IPR related to marine genetic resources whether they are based on genes, genomic regions or their derived products. However, one aspect of environmental data in general and marine-derived

data in particular is still a matter of concern. When disclosed, an invention should be fully described and display enough information so that others can reproduce it. This is not exactly the case with environmental samples used for metagenomic projects because it is almost impossible to describe all the organisms that were the source of the DNA or RNA. This has not prevented past patent applications in metagenomics, leaving open the way for future disputes.

Biological resource centres (BRCs)

BRCs are an essential part of the infrastructure underpinning life sciences and biotechnology. They consist of service providers and repositories of the living cells, genomes of organisms. They contain collections of cultivable organisms and replicable parts of these (e.g. genes and genomes, plasmids, viruses, cDNAs), viable but not yet cultivable organisms, cells and tissues. They often curate databases containing molecular, physiological and structural information relevant to these collections and related to bioinformatics. Considering the rapid development of metagenomic data in the recent years and notably the contribution of marine environments to this flood of data (for example, see the CAMERA website http://camera.calit2.net/), it is useful to consider the development of dedicated biological resource centers for metagenomes Retrieving the genes from the environment and assigning their origin is not an easy task since ecosystems are dynamic and their microbial species compositions vary both temporally and spatially . The resulting situation is a lack of standardized procedures to store and retrieve duplicates of original environmental samples used in metagenomic projects. This point should be addressed in the future so that conflicts about genetic resources can be avoided or minimized.

Benefit sharing example - The Morocco BRC

In 1998 the Moroccan Coordinated Collections of Microorganisms (CCMM) was created. Since that time the Moroccan National Centre for Scientific and Technical Research (CNRST) has provided support for the CCMM as a sustainable BRC in the fields of genomics, bioinformatics, capacity building and strain validation . CCMM activities have included the following: - enrichment of the CCMM with new strains from different Moroccan ecosystems,

- publication of catalogue of microorganisms in print and online (www.cccc.ma);
- distribution of strains to stakeholders following the guidelines articulated in a Material Transfer Agreement (MTA);
- operation of workshops and international courses on genomics and bioinformatics (www.cccc.ma) for the purpose of enhancing technology transfer and improving capacity building;
- registration of the CCMM at WFCC-MIRCEN World Data Centre for Microorganisms (WDCM) (CCMM WDCM883: Moroccan Coordinated Collections of Microorganisms, and also integrated into the StrainInfo.net bioportal site (www.straininfo.net);

It is anticipated that the Moroccan BRC will be increasingly used by the local industry. It will thus not only fulfill its role as conservation centre but also as a local service centre for

biotechnology. In the framework of the CBD, the Moroccan BRC will be the legal body for all exchanges of biological material between Morocco and other countries. By using certificates of origin and material transfer agreements (MTA), the traceability of the biological material will be guaranteed.

Discovery and Benefits-Sharing: the US International Cooperative Biodiversity Groups (ICBG) Program

Meeting the objectives of the CBD while cooperating between developed and developing countries to accomplish research and development from marine genomic resources of developing countries is a challenge. Since 1992, in order to foster projects that can serve as test cases and models for the complex agreements and activities required, the United States National Institutes of Health Fogarty International Center has administered the ICBG program with contributions from other US government agencies (Rosenthal et al., 1999). Of the nine current projects two of them, one in Fiji and another in the Philippines, are exclusively marine in focus, whereas four others, in Papua New Guinea, Costa Rica, Panama and Madagascar, have marine components. The remaining three focus on terrestrial plants.

These projects are required to explicitly address the objectives of the CBD, sustainable use of biological diversity, access and benefits sharing, research capability building and technology transfer. In addition to local governmental review, prior informed consent of local communities where samples are collected is mandatory. Permits and agreements are reviewed and approved by the ICBG program before activities begin.

These projects are diametrically opposed to the twentieth century attitude of scientists from developed countries regarding developing countries only as sources of samples. The ICBG projects are true partnerships in which as much research as possible, with attendant immediate and downstream economic benefits, is conducted in the country of origin of the samples (Kursar et al., 2007). Ultimately, through these and other well-structured projects, these countries will become self-sufficient in the discovery and development of their own resources, resulting in a strong local incentive to use resources sustainably. Programs that fully embrace the principles of the CBD must become the norm (and be generously funded), so that science will move forward through collaboration of partners; impatience with the process of nurturing science in the developing world will only lead to restriction of access to samples for all scientists.

Recommendations

1. EC and US scientists working on marine (meta)genomic projects should conduct their work in a manner consistent with the objectives of the CBD.
2. Banning the patenting of nucleotide sequences, their functions and products should be avoided.
3. The patenting of inventions derived from metagenome material requires further clarification.
4. Dedicated Biological Resource Centres for metagenome material should be considered.

Literature Cited

Holman, C.M. 2008. Trends in human gene patent litigation. Science 322: 198-199.

Rosenthal JP, Beck D., Bhat Ar, J. Biswas J, Brady L., Bridbord K, Collins S, Cragg G, Edwards J, Fairfield A, Gottlieb M, Gschwind L A, Hallock Y., Hawks Rd, Hegyeli R, Johnson G, Keusch GT, Lyons EE, Miller R, Rodman J, Roskoski J and Siegel-Causey D. 1999. Combining high risk science with ambitious social and economic goals. Pharm. Biol. 37: 6-21.

Kursar, TA, Caballero-George CC, Capson TL, Cubilla-Rios L, Gerwick WH, MV Heller WH, Ibañez A, Linington RG, McPhail KL,. Ortega-Barria E, Romero L.I., and Coley PD.. 2007. Linking bioprospecting with sustainable development and conservation: the Panama case. Biodivers. Conserv. 16: 2789-2800.

Zewers, KE. 2008. Debated heroes from the deep sea - marine genetic resources. WIPO Magazine, 2.

S E S S I O N - 5

Computing Power and Bioinformatics – Catching up with the Genomics Revolution

Michèle Barbier[1], Frank Oliver Glöckner[2], Granger Sutton[3], Monia El Bour[4], Balkiss Bouhaouala[5], Laura Giuliano[1]

[1]Mediterranean Science Commission (CIESM), Monaco [2]Max Plank Institute for Marine Microbiology, Bremen, Germany, [3]J. Craig Venter Institute, Maryland, USA, [4]National Institute of Marine Sciences and Technologies, Salammbo, Tunisia, [5]University of Tunis-El Manar, Tunis, Tunisia.

Bioinformatics, a worldwide overview

New massive parallelisation and high throughput techniques (the so called "-omics"), which allows work at the same time on many genes, transcripts or proteins, have contributed to an exponential growth in DNA and protein sequence databases. Such a huge set of compartmentalised information must be correctly managed and interrogated in order to produce real knowledge. Computational biologists from all over the world are therefore facing an ongoing, and growing, need to collect, store and manage huge data resources in a comprehensive, easy-to-interface manner.

While tremendous progress has been made over the years, many of the fundamental problems in bioinformatics remain unsolved including protein structure prediction, connecting gene function to the environment, and regulatory pathway mapping. Within this context, a critical issue is the efficiency of centres, which coordinate efforts to collect and disseminate biological data.

Europe and the US have always been at the forefront of bioinformatics research[1], but as we need to manage increasing quantities of ever more diversified data, the development of dedicated cyber-architectures to carry out high-performance specific, complex analyses becomes ever more important.

To quote Dr. J. Craig Venter, President of the J. Craig Venter Institute: "The explosion of data from the collection and sequencing of marine microbes requires a completely novel approach to storing, accessing, mining, analyzing, and drawing conclusions from this rich new wealth of information. The goal is to create a community resource to house all metagenomic data that will facilitate and advance knowledge of marine microbial ecology, other natural environments, and evolutionary biology." A prototype cyber-infrastructure, the CAMERA System[2] (Seshadri et al. 2007) has been developed by a group of American scientists with private foundation backing (www.more.org) to meet the challenge of studying marine life and ecosystems to examine, in an unprecedented manner, the genomic complexities of natural communities of microorganisms as they have evolved in their local environments.

1. For more information, see details on the objectives of the European Bioinformatics Institute (EMBL-EBI) (http://www.ebi.ac.uk/Information/About_EBI/about_ebi.html), and the activities of recent EU-funded Networks of Excellence, namely: (i) BioSapiens (http://www.biosapiens.info/page.php?page=home), (ii) EMBRACE (http://www.embracegrid.info/page.php?page=home), ENFIN (http://www.enfin.org/page.php?page=home).
2. For details see CAMERA homepage at the address: http://camera.calit2.net/

Europe has a set of world-class researchers in microbial ecology and in annotating marine genomic data. The availability of adequate computing and storage complexes will accelerate its advances in the knowledge of evolutionary biology and microbial ecology in marine and other natural environments, as well as strengthen"-omics" projects in the pharmacy, health, food, and agro research fields, thereby enhancing its competitiveness.

Within the new perspective of enlarging access of non-EU Member States to European facilities so as to better develop new partnerships, the development of a complex marine cyber-infrastructure system should take into consideration the huge variability of technological constraints within different countries. The basal architecture should rely on the development of standards accessible to countries lacking in certain technologies. A common interface with standards for data file formats, contextual data acquisition as recently proposed by the Genomics Standards Consortium[3] (Field et al. 2008), multilateral networking, multidisciplinary training, and upgrading of technology infrastructure for bioinformatics tools is destined to gain momentum (Chicurel, 2008).

The Marine Institute in Tunisia, *Monia El Bour*

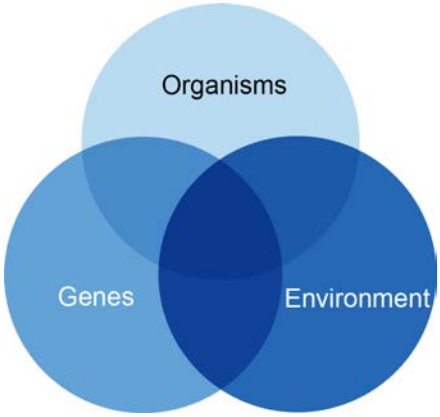
Bioinformatics in the marine environment

Molecular biology has undergone a paradigm shift, moving from a single-experiment science to a high throughput endeavour. Especially in the field of genomics, the introduction of automated sequencing technologies has lead to a massive increase of sequence data. It should be noted that the amount of sequence data in the public data repositories doubles every 18 months, and it is expected that this will significantly increase with the routine application of the next generation of sequencing technologies, like the 454/Roche GS FLX Titanium, ABI SOLiD, Illumina or Helicos HeliScope, systems. Although the genomic revolution is rooted in sequencing humans, model organisms, pathogens and production strains, it is specifically the marine sector that currently delivers the highest

3. www.gensc.org

data quantities (Yooseph et al. 2007). With more than 180 marine genomes and ongoing oceanic global metagenomic sampling campaigns, a gold mine has been opened for new enzymes and processes with a high potential for red and white biotechnology.

Role and function of genes in their environment- need for contextual data



Sequence data are a good starting point, but strings of even billions of AGCTs do not provide much information when it comes to understanding the function of organisms, which is a prerequisite for the inference of new catalytic mechanisms and enzymes. In every new genome or metagenome sequenced, approximately 40 percent of the potential protein encoding genes still lack any functional assignments! To be able to address these novel genes, contextual data (metadata) like habitat parameters (Environment), cell counts (Organisms) , genomic information as well as gene expression information at the transcriptome and proteome level (Genes) are urgently needed. To integrate the data, georeferencing has been shown to be extremely useful, especially for the open ocean, where any kind of genomic data can be easily linked with measured and remote sensing parameters based on location, time and depth (x, y, z, t) (Lombardot et al. 2006).

Although a rich set of metadata is desirable to get a holistic view of the ecosystem and the functions therein, they further increase the work load of the researchers to organize, store, process as well as to analyse and interpret the data. But an integrated view is a prerequisite to be able to screen and filter the flood of data and to nail down the set of genes that are primary targets for in-depth functional analysis in the wet lab.

Computational Functional metagenomics - need for standardization

There is a need for computational resources to provide this functional annotation to the community. In order to facilitate the common understanding and exchange of the computational data generated, standards for data formats, a functional ontology, evidence and confidence for predictions are also needed.

A variety of annotation pipelines have been developed for submission of data in the USA. These include the aforementioned CAMERA (http://camera.calit2.net/) as well as IMG/M (http://img.jgi.doe.gov/cgi-bin/m/main.cgi), and MGRAST (http://metagenomics.nmpdr.org/).pipelines use a variety of approaches to assign function to the often partial protein sequences encoded on the shotgun metagenomic fragments, usually via some method of homology determination to a protein or protein family of experimentally determined function (Markowitz 2007).

Unfortunately, while some form of evidence is usually provided for predictions from all of these systems, no standard for this evidence has been established and no uniform confidence measure defined. In fact, while the assumption is made that there are clear deductive steps linking a protein with a experimentally determined function to the predicted protein, it is not clear that this is always the case and the deductive steps are often not indicated..

This points out the need for a database of all proteins with experimentally determined function, including the methods used to determine functionality and their limits, standardization of acceptable deductive steps for linking proteins of known function to predicting function of unknown proteins, and standards for presenting this evidence chain in computer and human parseable formats.

A number of groups are working towards these goals: notable among them is GO (gene ontology) (Ashburner et al. 2001) which captures experimentally determined functional evidence, and provides an ontology (controlled vocabulary) for protein function as well as a set of evidence codes for how predicted function is determined. Another approach to functional annotation is to cluster similar proteins together, based on full length alignments which preferably indicate orthologous relationships (Yooseph et al. 2007). A variant of this kind of clustering which allows for the partial proteins prevalent in metagenomic data is available from CAMERA.

For this approach to be successful, the clusters themselves must be carefully curated. The recent eggNOG effort is notable for its attempt to automatically assign a "name" to each cluster by looking for consistent similarities between the names of the proteins within the cluster. While the automatic nature is admirable, it relies on the protein "names" being somewhat accurate and presumably somewhat independent since very little confidence is derived if all of the "names" simply derive directly from similarity to the first named protein. The automatic functional assignment for the clusters would benefit from knowing which proteins have experimentally determined function and a controlled vocabulary for that function.

The field of protein functional assignment should be mature enough to use a controlled vocabulary for protein function and protein "names" should be disambiguated from protein function instead of being used interchangeably in some cases. The needed protein functional ontology must be hierarchical, as protein function cannot always be determined or predicted to the ultimate specificity but often only to a broader category. Metagenomic and genomic protein computational functional assignment is still an active and evolving area of research but should now be mature enough to benefit from a more consistent application of standards.

From basic to applied biotechnological research in the Mediterranean

Due to its huge variety of habitats, from near shore areas (including surface hydrothermal vents) to largely unexplored deep-sea regions of astonishing heterogeneity, the Mediterranean Sea offers a promising, yet mostly untapped, reservoir of bioactive compounds with vast potential industrial applications.

While Mediterranean scientific institutions have a long, solid reputation in the field of marine ecology, studies on the industrial applications of marine biological resources are

in their infancy and quite fragmentary. The thematic distribution of biotechnological research across the Basin shows a definite geographic partitioning. In particular, apart from some domains of shared interest (e.g. antibiotic, antifouling, related action of new biological compounds[4]), southern institutions largely opt for "bioremediation" or "agricultural biotechnology" (e.g. Chahad et al 2007), whereas northern institutions have moved on a massive scale to the search for active molecules with potential applications in health, and energy related sectors.

It is generally agreed that the success of these types of studies will rely on the newly adopted approach extending "from proteins structure/function to protein-encoding genes" as well as on the optimization of computational analysis facilities[5]. The latter will enable scientists to run 'omics-style' software programs, and grant them access to a huge variety of specialized databases and services. Development of common tools for inter-computer operations, particularly for what concerns design of "international standards", should give due consideration to differences, especially pronounced in the Mediterranean region, of facilities for data storage and processing systems[6].

Bioinformatics capacities in the Mediterranean

Expertise related to bioinformatics in Mediterranean countries is especially advanced within the "Life Sciences and Biophysics" sector, whereas bioinformatics applied to ecological studies and to marine sciences lags behind. In countries on the south shore a large part of bioinformatics research activities is based in the universities, many of them characterized by a recent upgrading of their computing infrastructures [7], while north shore countries exhibit a larger variety of informational technology centres. In any case, national strategies across the whole Basin show a general tendency to reinforce bioinformatics infrastructure and to improve their networking with European reference Institutions. Within this context a coordinated action would help identify and localize the geographic distribution of infrastructures and related expertise, so as to facilitate exchanges between neighbouring countries and enhance cross-sector interactions.

Current challenges in bioinformatics development

"-Omics" data proliferation reduces the possibility of an individual researcher or even a research group to perform all the necessary steps needed to convert data into biological knowledge. Besides processing the data, standardized pipelines and quality management will become crucial to keep track of the integrity and quality of the information provided.

Without this, it will be impossible to perform sound comparisons 'out of the box' between

4. Bioactive molecules with such applications, which are usually used as defence mechanisms in the source ecosystems, justify the shared interest on molecular mechanisms that drive inter-species interactions (including symbiotic ones).
5. The creation and effective inter-connection (i.e. by web systems) of datasets containing multiple information on gene expression, catalysed chemical reactions, regulatory interactions, protein assembly, as well as metabolic and signal transduction pathways.
6. Current efforts in this direction(e.g. EMBRACE, NoE, FP6), shall be extended to a wider range of scientific users (i.e. those from less equipped Institutions on the southern shore).
7. Among others: Centre of Biotechnology in Sfax, Pasteur Institute of Tunis, Centre of Biotechnology in Borg Cedria, Technopole de Sidi Thabet – for what concerns Tunisia.

datasets, resulting in a waste of time and resources to process the data over and over again. This significantly slows down the productivity and economy of Europe in the field of biotechnology and environmental research. Therefore appropriate cyber-infrastructure is needed to store, analyse, and integrate all the "-omics" data for marine organisms

Need for European marine cyber-infrastructure

To be able to keep pace with the data challenges in Europe, a two tier bioinformatics infrastructure needs to be implemented consisting of:

1. The establishment of a dedicated marine cyber-infrastructure in Europe, along the lines of CAMERA (Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis) in the USA but taking into account the latest advances and European specificities. This is a prerequisite to process and integrate the elaborate and heterogeneous data streams in marine molecular biology. In cooperation with the European Bioinformatics Institute ELIXIR infrastructure, the LIFEWATCH project (www.lifewatch.eu) and the planned European Marine Biological Resource Centre (EMBRC) a follow up research infrastructure of the Network of Excellence "Marine Genomics Europe", an integrated data processing infrastructure for the marine realm should be established. It will acts as a central facility to handle the specific requirements in standardized data processing and data integration for the marine system.
2. Strengthening of local research facilities to enable them to work with the pre-processed and integrated datasets. Effective knowledge generation can only be achieved if the bioinformatics skills needed are combined with the specific biological knowledge of the "on site" research groups. To this end, reduced datasets must become mobile again after the first round of standardized large scale date processing and integration. Easy-to-use software tools and systems, disseminated by the infrastructure providers, are crucial to cover the specific scientific questions and biotechnological applications raised by the investigators. A major component to move forward on these issues will be to offer training capacities that help to reach a common ground in data analysis, data integration and modelling for marine researchers.

A marine cyberinfrastructure bridging across the seas, a technical challenge

An ideal system for a cyberinfrastructure would include an internet-based client / server architecture to allow remote and local access to the system. While end users can manipulate the data over the web, the cyberinfrastructure should permit scientists to connect their local laboratory computers directly to the database and tools. An essential feature would also include a quality assurance process, to allow quick re-annotation of previous results in light of new data. Integration and / or development of visualization tools to allow multiple views of data, annotation, comparison, and comprehension in a graphical environment is highly desirable to allow researchers to get a better picture of their analysis and results (Carter et al., 2000).

The needs of scientists studying the complexity of organisms and the way they function in their natural environment vary with the ecosystems and geographical location. The Mediterranean Sea harbours different ecosystems and is surrounded by many countries. Bioinformatics provides a unique framework for dialogue and cooperation between the

two shores of the Mediterranean, underlining the importance of a partnership for supporting the development and modernisation processes in research highlighted in the EC strategic framework for international and technology cooperation (COM(2008)588).

Major Regional technopoles would benefit from welcoming outstations of this cyber-infrastructure to adapt the service to researchers according to their needs. And the ability to expand the system, via simple addition of modules, would allow the system to evolve as new biotechnology.

Efforts have to be made to continuously develop adapted training programs for research institutions and universities in Europe and Mediterranean countries. Exchange of know-how, expertise and experience through implementing joint PhD programmes, organising joint workshops and conferences will contribute to a new generation of bioinformaticians to establish a Mediterranean bioinformatics network.

Recommendations

1. A European Bioinformatics (Cyber-) Infrastructure specific for the marine realm should be established
2. Standards for data processing and contextual (meta)data acquisition, storage and exchange should be established to facilitate data integration and comparative analysis
3. Knowledge databases with quality management and expert biocuration should be established to provide a well documented and reliable reference datasets for academia and industry
4. Efforts have to be taken to homogenize the technology and knowledge base between all European countries. To foster technology transfer, multinational and multidisciplinary infrastructure and research projects should be implemented,
5. Regional technopoles should be encouraged to become outstations of the European cyberinfrastructure for adapted technology and service in emerging countries.
6. European and international training courses and workshops in bioinformatics and related disciplines should be established gathering biologists and bioinformatician together.

Black smokers chimney, original habitat for extreme organisms, *Joel Querellou*

Literature Cited

Ashburner M et al. 2001. Creating the gene ontology resource: design and implementation. Genome Res. 11:1425-1433.

Carter K, Schibeci D, Bellgard M., .2000. WWW issues for conducting sophisticated Bioinformatics analysis, , the Seventh Australian World Wide Web Conference, http://ausweb.scu.edu.au/aw01/papers/refereed/carter/paper.html.

Chahad BO, El Bour M, Miraouana R, Abdennaceur H, Boudabous and A. 2007. Preliminary selection study of potential probiotic bacteria from aquacultural area in Tunisia, Annals of microbiology, 57: 185-190.

Chicurel M. 2008. Bioinformatics: Bringing it all together technology feature, Nature 419: 751-757.

Field D et al. .2008. The minimum information about a genome sequence (MIGS) specification. Nat. Biotechnol. 26:541-547.

Lombardot T, Kottmann R, Pfeffer H, Richter M, Teeling H, Quast C, Glöckner FO .2006. Megx.net - database resource for marine ecological genomics. Nucleic Acid Res. 34: D390-D393.

Markowitz VM . 2007. Microbial genome data resources. Curr. Opin. Biotechnol. 18:267-272.

Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. 2007. CAMERA: A Community Resource for Metagenomics. PLoS. Biol. 5:e75.

Yooseph S et al. 2007. The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. PLoS. Biol. 5:e16.

LIST OF PARTICIPANTS

Dr Wolfgang Aehle
B.R.A.I.N Aktiengesellschaft
wa@brain-biotech.de
GERMANY

Dr Andy Allen J. Craig
Venter Institute
aallen@jcvi.org
USA

Prof. Mohamed Amar
CNRST- Laboratoire de microbiologie
et biologie moléculaire
amar@cnrst.ma
MOROCCO

Dr Michèle Barbier
CIESM
mbarbier@ciesm.org
MONACO

Dr Doug BartlettScripps
Inst. Oceanography
dbartlett@ucsd.edu
USA

Dr Jean-François Bloch
PROTEUS
jfbloch@proteus.fr
FRANCE

Dr Balkiss Bouhaouala
Univesité de Tunis El Manar
balkiss.bouhaouala@pasteur.rns.tn
TUNISIA

Dr Chris Bowler
ENS - UMR 8186
cbowler@biologie.ens.fr
FRANCE

Dr Monia El Bour
Institut National des Sciences et
Technologies de la Mer,
monia.elbour@instm.rnrt.tn
TUNISIA

Dr Fernando De La Calle
PharmaMar
fdelacalle@pharmamar.com
SPAIN

Dr Manuel Ferrer
Instituto de Catalisis y Petroleoquimica-CSIC
mferrer@icp.csic.es
SPAIN

Prof. Frank Oliver Glöckner
Max Planck Institute of Marine Microbiology
fog@mpi-bremen.de
GERMANY

Dr Laura Giuliano
CIESM
lgiuliano@ciesm.org
MONACO

Prof. Peter Golyshin
School of Biological Sciences,
Bangor University
p.golyshin@bangor.ac.uk
UK

Dr Karla Heidelberg.
University of Southern California
kheidelb@usc.edu
USA

Dr Margo Haygood.
Oregon Health and Science University
haygoodm@ebs.ogi.edu
USA

Dr Dermot Hurst
Marine Institute
hurst@marine.ie
IRLAND

Prof. Ian Joint
Plymouth Marine Laboratory
IRJ@pml.ac.uk
UK

Dr Christine Klockow
Max Planck Institute for
Marine Microbiology
cklockow@mpi-bremen.de
GERMANY

Dr Maurice_ Lex
European Commission
Research Directorate General
maurice.lex@ec.europa.eu
BELGIUM

Dr Gurvan Michel
CNRS-Station Biologique de Roscoff
gurvan@sb-roscoff.fr
FRANCE

Dr Alison Murray
Desert Research Institute
Alison.Murray@dri.edu
USA

Dr Lita_ Proctor
National Science Foundation
lproctor@nsf.gov
USA

Dr Joel Querellou
IFREMER
joel.querellou@ifremer.fr
FRANCE

Prof. Thomas Schweder
Institut für Marine Biotechnologie
schweder@uni-greifswald.de
GERMANY

Dr Kevin Sowers
University of Maryland
Biotechnology Institute
sowers@comb.umbi.umd.edu
USA

Dr Granger Sutton
J. Craig Venter Institute
gsutton@jcvi.org
USA

Dr Ramunas Stepanauskas
Bigelow Laboratory for Ocean Sciences
rstepanauskas@bigelow.org
USA

Dr Matthew Sullivan
University of Arizona
mbsulli@email.arizona.edu
USA

Dr Cinzia Verde
CNR- Institute of Protein biochemistry
c.verde@ibp.cnr.it
ITALY

Dr Michail M. Yakimov
Institute for Coastal Marine Environment
michail.yakimov@iamc.cnr.it
ITALY

Dr Fitnat Yildiz
University of California Santa Cruz
yildiz@etox.ucsc.edu
USA

Prof. Diaa Youssef
Faculty of Pharmacy,
Suez Canal University
diaa22@yahoo.com
EGYPT

How to obtain EU publications

Publications for sale:
- via EU Bookshop (http://bookshop.europa.eu);
- from your bookseller by quoting the title, publisher and/or ISBN number;
- by contacting one of our sales agents directly. You can obtain their contact details on the internet (http://bookshop.europa.eu) or by sending a fax to +352 2929-42758.

Free publications:
- via EU Bookshop (http://bookshop.europa.eu);
- at the European Commission's representations or delegations. You can obtain their contact details on the Internet (http://ec.europa.eu) or by sending a fax to +352 2929-42758.

European Commission

EUR 23908 EN — **EC-US Task Force on Biotechnology Research, Joint EC-US and CIESM Workshop on Marine Genomics, 2008**

2009 — 53 pp. — A4

ISBN 978-92-79-12136-4

ISSN 1018-5593 1831-2322

KI-NA-23908-EN-N

EC-US Task force on Biotechnology Research
The Marine Genomics Working Group

The new age of metagenomics enables the study of the vast majority of marine microbial species which are as yet unable to be cultivated in the laboratory. These technologies and the analyses they enable have ushered in a new era of biology with fundamental implications for basic research and biotechnological advances. But, they also pose challenges in areas of intellectual property and patent law as well as in interdisciplinary training of the next generation of scientists.

http://ec.europa.eu/research/biotechnology/ec-us/index_en.html

ISBN 978-92-79-12136-4