Project no.**GOCE-CT-2004-505403**

Marine Genomics Europe

« Implementation of high throughput genomics approaches to investigate the functioning of marine ecosystems and the biology of marine organisms"

Network of excellence

Thematic Priority 1.1.6 - Sustainable Development, Global Change and Ecosystems

THE PUBLISHABLE FINAL ACTIVITY REPORT OF MARINE GENOMICS EUROPE PROJECT

Period covered from March 2004 to August 2008
Started on 1st March, 2004 to the 31rst of August 2008 Duration: 54 months

Scientific Project coordinator: Catherine Boyen
CNRS-Station Biologique de Roscoff

Financial administrative coordinator : FIST SA

List of MGE partners

Partners	**Short name**
Partner 1 France Innovation Scientifique et Transfert - SA	**FIST**
Partner 2 Centre National de la Recherche Scientifique	**CNRS**
Partner 3 Institut Français de Recherche pour L'Exploitation de la Mer	**IFREMER**
Partner 4 Max Planck Institute for Marine Microbiology Bremen	**MPIMM**
Partner 5 Max Planck Institute for Chemical Ecology Iena	**MPICE**
Partner 6 Max Planck Institute for Molecular Genetics Berlin	**MPIMG**
Partner 7 Bielefeld University	**Bielefeld Univ.**
Partner 9 German Research Centre for Biotech Braunschweig	**GBF**
Partner 10 Alfred Wegener Institute Bremerhaven	**AWI**
Partner 11 European Molecular Biology Laboratory Heidelberg	**EMBL**
Partner 12 Natural Environment Research Council	**NERC**
Partner 13 University of Birmingham	**Birmingham Univ**
Partner 14 University of Whales, Bango	**UWB**
Partner 15 University of Oxford	**Oxford Univ**
Partner 16 Marine Biological Association, Plymouth	**MBA**
Partner 17 University of Warwick	**UoW**
Partner 18 University of Newcastle	**UNEW**
Partner 19 University of Wales Swansea	**UWS**
Partner 20 University of Cardiff	**UWC**
Partner 21 Centre of Marine Sciences Faro	**CCMAR**
Partner 22 Stazione Zoologica "Anton Dohrn" Napoli	**SZN**
Partner 23 National Research Council, Palermo	**IBIM-CNR**
Partner 24 Università degli Studi di Padova	**UNIPD**
Partner 25 Israel Oceanographic & Limnological Research Haifa	**IOLR**
Partner 26 Technion Israel Institute of Technology	**TECHNION**
Partner 27 Katholieke University Leuven	**K.U.Leuven**
Partner 28 University of Groningen	**RUG**
Partner 29 Sars International Centre for Marine Molecular Biology, Bergen	**UiB**
Partner 30 Norwegian School of Veterinary Sciences, Oslo	**NVH**
Partner 31 Institut of Oceanology Gdynia	**IO PAS**
Partner 32 Danish Institut for Fisheries research	**DIFRES**
Partner 33 Consejo Superior de Investigaciones Científicas	**CSIC**
Partner 34 Centre of AquacultureTarragona	**IRTA**
Partner 35 University of Barcelona	**UB**
Partner 36 University of Crete	**UoC**
Partner 37 Hellenic Centre for Marine Research	**HCMR**
Partner 38 Royal Swedish Academy, Kristineberg Marine Station	**KMRS**
Partner 39 University of Göteborg	**UGOT**
Partner 40 Parco Technologico Padano s.r.l.	**PTP LODI**
Partner 41 Plymouth Marine Laboratory	**PML**
Partner 42 Prokaria ltd Reykjavik Iceland	**Prokaria**
Partner 43 Pontificia Universidad Catolica de Chile	**PUCCH**
Partner 44 Universidad de Conception de Chile	**COPAS**
Partner 45 University of Freiburg	**Univ. Freiburg**

CONTENT

http://www.marine-genomics-europe.org/

I - Introduction

The key purpose of Networks of excellence is to tackle fragmentation of existing research capacities in a particular thematic area by integrating, at a European level, the critical mass of resources and expertise needed to provide European leadership in that topic. The initial observation behind the development of “Marine genomics Europe” NoE was that, although research on the biology of marine organisms has been a field of excellence in Europe for more that a century and a half, the related community remained fragmented. Because of that, European marine biologists have benefited less from the genomics revolution than researchers in other fields such as health, plant and animal sciences.

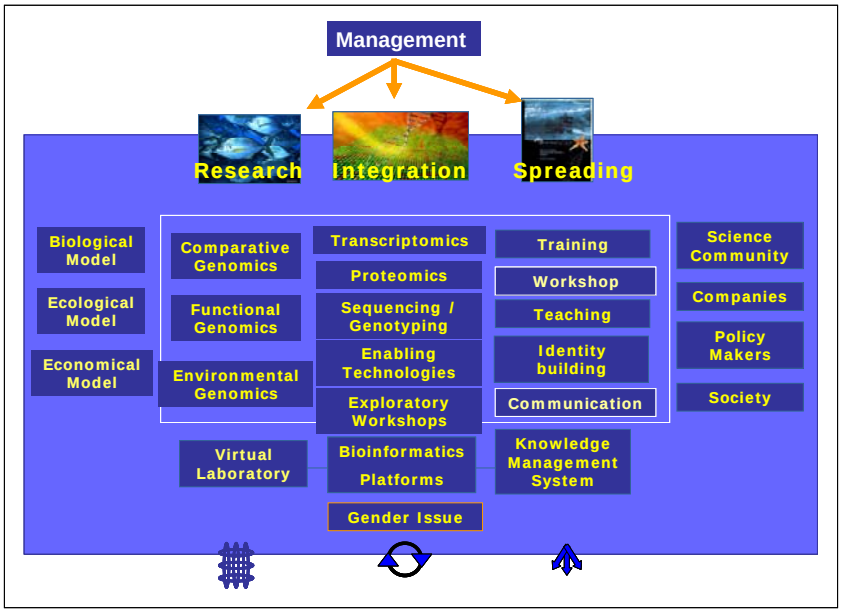
The major goal of "Marine Genomics Europe" (MGE) Network of Excellence was thus to reshape European Research in marine biology by networking groups and hence permitting the necessary critical mass in scientific expertise and technological resources to become a leading force in Marine Genomics. In this way, MGE aimed to promote, develop and spread a better understanding of the functioning of marine ecosystems and the biology of marine organisms throughout the European Union. Our long-term target was to establish a durable European network capable of implementing high throughput genomic approaches in the field of marine biology.

The MGE Network of Excellence involved 45 institutions from 16 countries (14 state members and 2 non European states, ca 450 persons implicated)

MGE NoE concentrated on four major types of activity:

- **Management** of the Consortium
- **Jointly Executed Research** programs
- A large **integration program** involving sharing of large-scale core facilities (technological platforms), development of enabling technologies, creation of a common Bioinformatics centre, organization of Exploratory workshops and Gender actions.
- **Spreading activities** including teaching and training, web site development, brochure publishing and popularisation projects.

The figure below presents the general architecture of the project



Because of the novelty of Networks of Excellence, which are instruments that did not exist before FP6, implementation of the integrating strategy has required a great deal of education and determination for convincing MGE members of the great benefit of such a project. It is thus quite gratifying for all of us within MGE to realize at the end of the project that a strong and real identity has been built up and that most of us are very satisfied and even proud to have being part of this ambitious and challenging initiative.

Some numbers: More than 245, 000 different persons visited the MGE website, which has regularly been enriched with new data and information during the 54 months of the project. About 150 externally funded collaborations have been initiated and carried out and MGE members have presented their data or the activities of the consortium in about 350 conferences. In total **more than 220 articles** acknowledging MGE, were published in International peer reviewed journals, including 1 in Nature, 2 in Cell, 2 in Science,1 in Nature Review Genetics, 3 in PLOS Biology and 11 in P.N.A.S. and our **average impact factor is 4. 557.**

In conclusion after more than four years of existence, one can claim that Marine Genomics Europe network of Excellence has very much improved integration and has strongly promoted interaction and collaboration within the scientific community. Numerous indicators demonstrate that the momentum initiated in 2004 was regularly stoked over the four and a half years and that major integrating and scientific milestones have been attained.

In order to maintain these efforts, a plan for ensuring long-term integration upon termination of the project was established and priorities were identified to structure and strengthen the European Research Area in a durable way in the field of Marine Genomics.

A position paper "**The European Flagship in Marine Sciences for a Sustainable Future - Creating a strong Marine R&D leadership for Europe to benefit society and industry"** resulting from the outputs of several working meetings organized in the frame of MGE and involving many scientists from the consortium was published. This document identifies some new strategic priorities and needs over the next decade to understand and explore the complexity of marine ecosystems.

Three priorities were thus identified to structure and strengthen the European Research Area in the field of Marine Genomics

- Maintain Scientific Coordination and establish an Educational Program
- Support common Infrastructures and common Databases
- Support Joint Research Programmes

The major objectives of the NoE MGE for achieving long term integration are **i)** to sustain the coordination of the consortium (web site, newsletter, workshops and meetings) **ii)** to establish a European PhD programme in Marine System Biology. This initiative will capitalize on the effort done by MGE through the teaching and training action. The ultimate objective is to implement a true independent PhD programme recognized by several Universities in Europe and abroad **iii)** to promote support for common infrastructures and Joint Research Programmes in Marine Genomics.

It was agreed by the Scientific Steering Committee that the "GDRI" (International Research Group) status proposed by CNRS for international collaborations will be considered for maintaining the MGE community integrated at the end of the contract with the EC.

II- Integrating Activities

II-1 GAINING COLLECTIVELY ACCESS TO MAJOR GENOMICS CENTRES

One major objective of MGE was to get support from major European Genomics Centres in several ways: offer a competitive price for sequencing, provide the consortium with expertise for the sequencing and annotation of whole genomes and partially fund a few projects, should their scientific committees agree. The action of MGE has been then to contribute seed money and to fund fellows for coordinating the project. During the contract, several large sequencing projects were accepted and initiated by Genoscope, the French National Sequencing Center.

- **2004 - A large-scale genome project for the filamentous brown algae Ectocarpus** siliculosus

  Coordinator: M. Cock, CNRS, Roscoff, Partner 2 + International consortium (http://www.genoscope.cns.fr/spip/Ectocarpus-siliculosus,740.html)

- **2004 - Marine Cytophagales project, Bacteria specialized for polymer-degradation**

  Coordinator : R. Amann, MPI Bremen, Partner 4

  (http://www.genoscope.cns.fr/spip/Marine-cytophagales-specialized.html)

- **2005 - Generation of EST libraries and sequencing of 230 000 EST from embryonic stages of the sea urchin *Paracentrotus lividus*.**
  Coordinator C. Gache, CNRS, Villefranche, Partner 2

- **2005 - The chaetognaths: a key position in the phylogenetic tree of the bilaterians**. Sequencing a collection of expressed sequence tags (~150 000 ESTs)
  Coordinator : Y. Le Parco, CNRS, Marseille, Partner 2

- **2005 - *Phaeodactylum tricornutum* ESTs library Sequencing of 100,000 new "expressed sequence tags" (ESTs) from cells grown under 10 different conditions.**
  Coordinator: C. Bowler, CNRS, Paris (Partner 2 and 22)
  http://www.genoscope.cns.fr/spip/Phaeodactylum-tricornutum,463.html

- **2006 - Comparative and environmental genomics in eukaryotic marine phytoplankton: the genome of *Bathycoccus* (Prasinophyceae) for environmental and evolutive genomics**
  Coordinator: H. Moreau, CNRS, Banyuls, Partner 2

- **2006- Pilot project for sequencing the genome of a red alga: Chondrus crispus (1X coverage)**
  Coordinators: J. Collen & C. Boyen, CNRS Roscoff, Partner 2 + International consortium

- **2006- Comparative genomics for five marine cornerstone species. Sequencing of BAC clones (300,000 reads) for the tunicate *Ciona intestinali,* the amphioxus *Branchiostoma lanceolatum,* the sea urchin *Strongylocentrotus purpuratu,* the polychaete *Platynereis dumeri* and the acoel *Symsagittifera roscoffensis*.**
  Coordinator: **D. Arendt**, EMBL, Heidelberg, Partner 11 and Partners 1, 6, 22, 35 & 38

- **2007- Sequencing of the complete genome of a red alga: *Chondrus crispus* (11X coverage)**
  Coordinators: J. Collen & C. Boyen, CNRS Roscoff, Partner 2 + International consortium

- **2008 – The oyster *Crassostrea gigas*, an economically important mollusc Sequencing of about 30,000 ESTs from normalized libraries from different organs of Crassostrea gigas**
  Coordinator: P. Favrel, Ifremer, Partner 3 & University of Caen.
  http://www.genoscope.cns.fr/spip/crassostrea-gigas-an-economically.html

II- 2 TECHNOLOGICAL PLATFORM PROJECT (WP 22)

Project Leader : Richard Reinhardt, MPI Berlin, #6

Project assistant: Tamara Safari, MPI Berlin, #6

TP project committee:

- Klaus Valentin, AWI, Bremerhaven, DE, Partner 10
- Rudi Amann, MPI Bremen, DE, Partner 4
- Jeanine Olsen, RUG, Groningen, NL, Partner 28
- Richard Reinhardt, MPI Berlin, DE, Partner 6
- Alexander Goesmann , Bielefeld University, DE, Partner 7
- Ina Arnone, SZN, Napoli, IT, Partner 22
- Tomaso Patarnello, (replaced by Benjamin Piña after one year)

Summary description of project objectives

Giving access to the diverse technology platforms available within the MGE consortium was one of the main objectives of the Network of Excellence. The idea was to allow NoE partners to access medium to high-throughput facilities not usually available to them to generate baseline data by sequencing of genomes or libraries, generation of genomic tools such as EST, cDNA and BAC-libraries or accessing organisms or facilities not available at most partner's own institutes. Access limitations were especially valid for most advanced technologies like 454 sequencing which just had been introduced to a few European high throughput sequencing institutes like the MPI for Molecular Genetics (#6) and and was made available to MGE partners since the end of 2007.

WP 22 fulfilled its main objectives:

- to provide access to genomic, proteomic and bioinformatic Technological Platforms (TP) that had already been developed by larger research institutes in the EU;
- to further develop high-throughput genomics and proteomics tools for integrated research projects;
- to maximise efforts on key areas of marine science in the fields of comparative, functional and environmental genomics.
-

Summary description of contractors involved

The TP evaluation committee developed the schemes described below and assessed applications for the Access Schemes in the 4,5 year of MGE duration. The TP providers allowed access to their facilities at the marine station CNRS Roscoff (#2), the Max Planck Institute for Molecular Genetics in Berlin (#6), the University of Birmingham (#13) and the Stazione Zoologica Anton Dohrn , Napoli (#22). In the first years of MGE more TP providers were part of this group but one had to withdraw due to internal changes and the others were not large and/or popular enough for the needs of MGE. Furthermore a link to the bioinformatics platform at University Bielefeld was established, which worked according to their own work packages and funding schemes, together with other MGE partners providing bioinformatics expertise.

Summary description of work performed and end results (including methodologies, approaches, main achievements and relation to the state of the art)

Four schemes were made available since early in the project:

- Scheme to pay for consumables to allow partners to use Technology Platforms for specific projects (consumables scheme, up to 5000 Euro).
- Short-term visiting fellowships (short travel to TP platform, 200 Euro).
- Longer-term visiting fellowships (travel to TP platform for 1 month, up to 1000 Euro).
- Short term bioinformatics training (short travel to bioinfo platform, up to 400 Euro)..

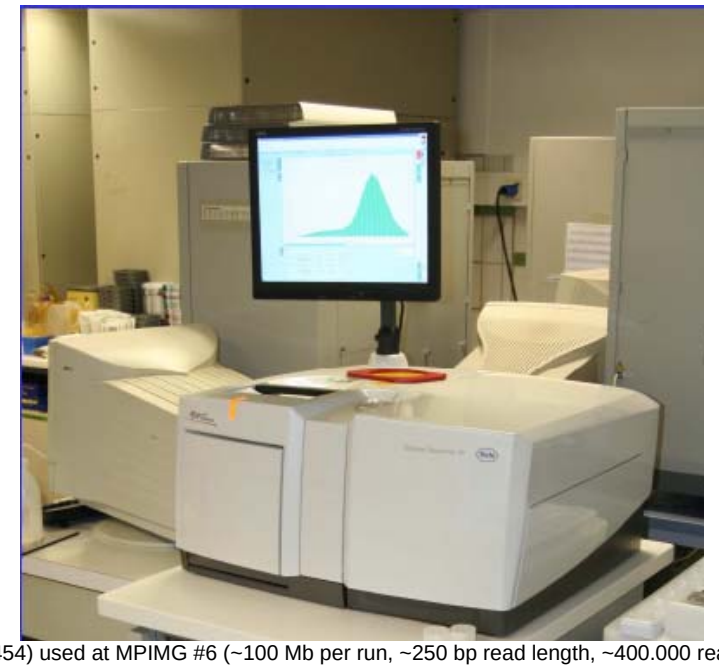
These schemes were set up to allow all participants in the network to utilise the NoE's Technology and Bioinformatic Platforms. One main condition was that two partners applied together for each grant to further integration. The Platforms also continued to spread knowledge through direct contact with NoE partners while

undertaking joint activities with several nodes. A main innovation in the TP access schemes in year four had been the introduction of 454 sequencing as a novel option to use this new and interesting technique for MGE partners. To enable use of this technology the grant size for the consumables scheme was raised to 5750 Euro (for half a run). Many of the applications in the last year encompassed 454 sequencing.

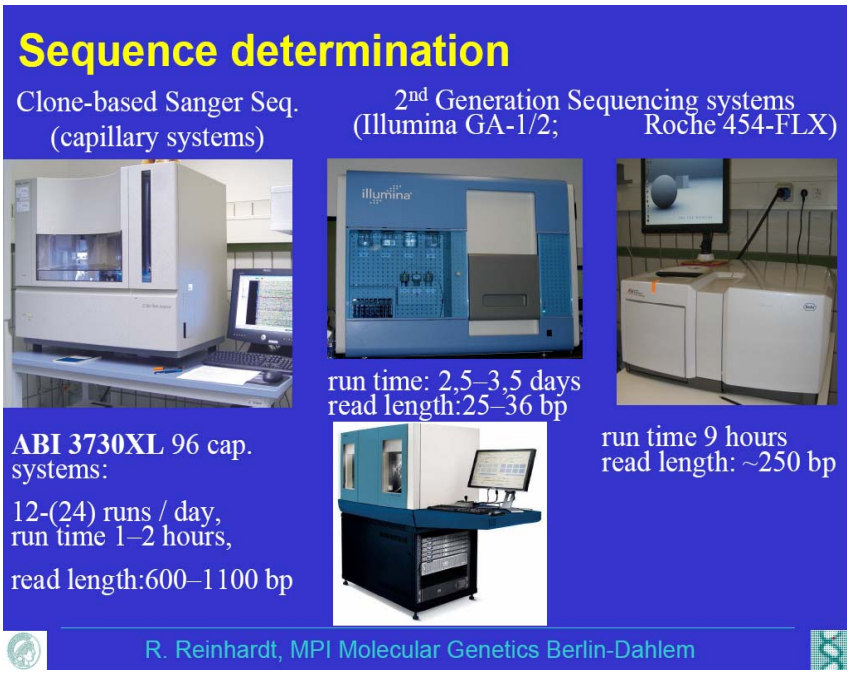
Main achievements: Until the end of the Network of Excellence a total of 65 proposals had been received through the TP secretariat with only seven not selected for funding. Thus, a total of 58 projects have now been funded overall, 10 using the new 454 sequencing technique, thereof eight in the last six months of MGE. A report on the results of using the scheme was a mandatory requirement within the application. Most of the TP progress reports are now available. Overall 40 EST libraries were prepared, sequenced and processed (15x seabass, 14x seabream, 4x oyster, 4x Ruditapes manila clam, 1 Mytilus, 1x Myxine European eel, 1x Octopus vulgaris). A few projects which started in the last six months have not yet been fully finished but are well underway. An administrative assistant at #6 provided full service for all partners needing help and information on the TP access schemes.

Period	**TP proposals**	**TP projects**	**Travel**	**Consumables/ thereof using 454**
4,5 years	65	58	7	51 / 10
Reports submitted		40		
Platform used:				
MPI Berlin	**SZN**	**Uni Birmingham**	**CNRS Roscoff**	**Bielefeld**
37	10	3	2	1

* the SBT scheme was never used, in spite of updates and changes.



Roche GS FLX System (454) used at MPIMG #6 (~100 Mb per run, ~250 bp read length, ~400.000 reads)



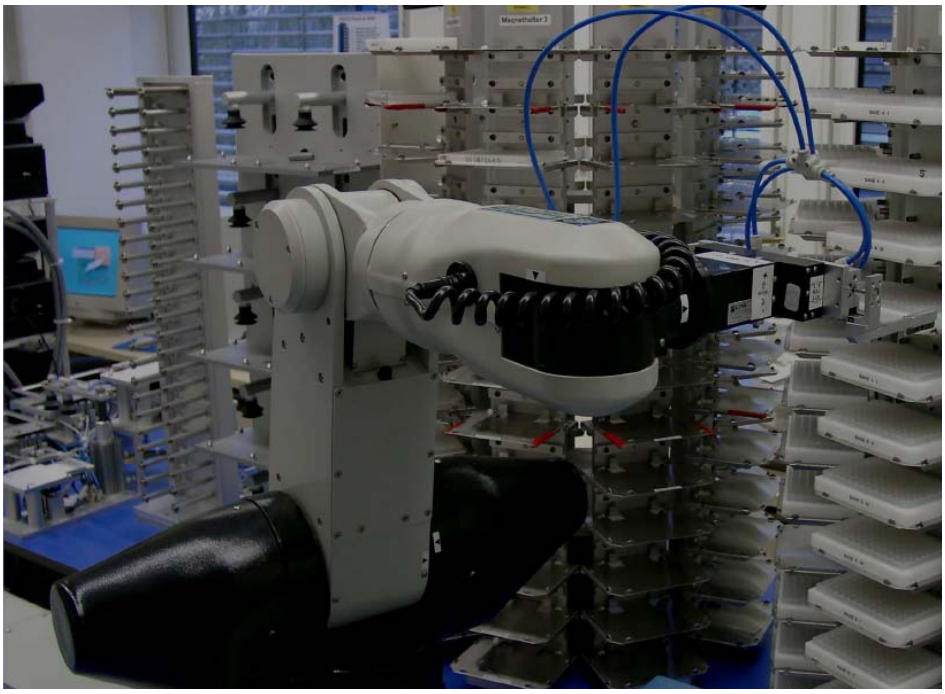
Overview of sequencing technologies available at the MGE technological platforms

All progress reports on access projects show the usefulness of the access projects. Several publications are planned for 2008 and 2009, almost all written in conjunction with the node activities. They can be found in the publication list made for the Plan for Use and Dissemination. In this document, the full list of TP Access Schemes granted can also be found. The reports show that using TP access was often well integrated with work within the nodes of MGE and coordinated with further funding. Reference to publications is also made in several reports; a few examples are given below:

1) "The striped sea bream, *Lithognathus mormyrus* was chosen as a suitable sentinel fish species probing the Mediterranean coastal environment. The cDNA assemblage was fully sequenced and characterized. The gene-specific results are located in the devoted website http://est.molgen.mpg.de/FishShellfish/ The characterized microarray is continuously used for determining pollutant-affected differential expression of genes, including effects of cadmium, estradiol, and tert-butyl hydro peroxide. Some of the results are in preparation for submission of a publication. The project may be considered fully accomplished." (Moshe Tom, IOLR, Partner # 25; Consumable Grant; MPI MG Berlin).

2) "The present project has been successful in providing the basis for the development of a set of microsatellites useful in family assignment and the first steps of mapping work in the Manila clam (*Ruditapes philippinarum*), a highly valuable commercial bivalve. Over 40 primer pairs have been developed and are now being tested for polymorphisms. A second library is under construction and screening. These results, obtained in a relatively short period of time, underline the utility of the enrichment technique used, and the high potential of collaborations established inside the MGE initiative in producing genomic tools in species that are poorly characterized at the genetic level. (Tomaso Patarnello CCMAR, Partner # 21; Consumable Grant; SZN, Naples).

3) "The results from the study show extremely low levels of neutral genetic differentiation between sampled populations, but at the same time large population specific differences in gene expression profiles of flounder originating from the North Sea and the Baltic Sea. Several of the differentially regulated genes could be directly linked to fitness traits. These findings demonstrate that European flounders, despite little neutral genetic divergence between populations, are differently adapted and imply that adaptation in gene expression may be common in other marine organisms with similar low levels of population subdivision. At the moment papers are written based on the data and MGE will naturally be acknowledged for financial support and Technology Platform access." (Peter Foged Larsen, DIFRES, Partner # 32; Consumable Grant; UoB Birmingham).



Sample preparation by a high throughput robotic system as applied for TP projects at #6

Researchers were able to enjoy the benefits of individual training by experts in the use of new genomic technologies allowing transfer of expertise, and, where appropriate, technology across the network. Here an example for one of seven travel & training grants (RUG, Partner # 28): “The stay was excellent, Dr. Hoarau having learned about cDNA library trouble shooting, HTsequencing strategies, EST database management and establishment of the pipeline interfaces between MPI Berlin, University of Bielefield and MGE for the Fucus project”. Furthermore several training courses were held at the platforms; e.g. two times the hands-on course on “Generation of cDNA Libraries by Primer Extension” at the TP of partner # 6 and several courses at partner # 22 (see WP 19 for details).

Impact of the project on its research sector

The TP Berlin had elicited a structured response from the MGE community on the diverse microarrays done within several nodes and work packages. An overview table on all microarrays developed within the project including information on future potential uses is available on request from the coordinator.

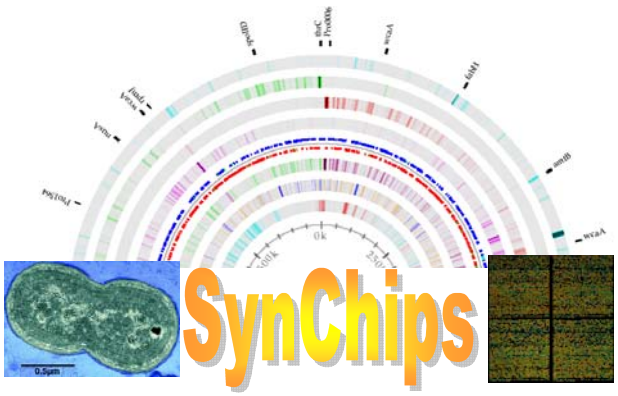
During the last General Assembly meeting in Faro in May in 2008 a final TP meeting was held within the NoE duration. All node leaders and two TP providers were present. Further collaboration between the TP providers and the MGE partners were discussed, mainly within the context of the planned GDRI. It is planned to continue giving access to TPs, when funding can be secured, i.e. via bilateral agreements or through new larger projects. Funding for several new projects could already be secured i.e. three new FP7 projects, one between JGI and two MGE partners on the full Zostera genome and a bilateral one between France and Germany (more details can be found in the PUDK).

Overall it can be said that giving TP access to a reasonably large number of marine biologist partners within MGE was a success story speeding up technology transfer and application from high throughput genomics facilities to the marine science community. Future collaborations are greatly facilitated, due to transfer of the know how to prepare samples and apply these new technologies, i.e. high throughput ‘omics techniques in general and 454 pyrosequencing in particular, for marine research questions. The GDRI will provide a platform for continuous exchange and facilitate joint planning of small and large genomic projects.

II- 3 FLAGSHIP PROJECT SYNCHIP (WP27)

Project leader: W. Hess, Freiburg University, Partner 45

"Combined microarray hybridization and transposon mutagenesis to study the impact of environmental stresses on marine Synechococcus" (SynChips)



Partners:

Albert-Ludwigs-University Freiburg, Faculty of Biology, Schänzlestr. 1, D79104 Freiburg, Germany (#45)

Station Biologique de Roscoff, CNRS UPR 9042 & UPMC (U. Paris 6), BP74, 29682 Roscoff, France (#2)

University of Warwick, Department of Biological Sciences phytoplankton Group, Gibbet Hill Road, Coventry, CV4 7AL, UK (#17)

Interuniversity Institute for Marine Science, Hebrew University
Coral Beach, POB 469, 88103 Eilat, ISRAEL (#26)

II.3.1 – Objectives of the project

Marine picophytoplankton which comprises all the photosynthetic organisms smaller than 2 µm, represents more than 80 % of the chlorophyll biomass in the intertropical, nutrient-depleted (oligotrophic) regions of the world's ocean. It includes small eukaryotes phylogenetically very diverse and two main genera of cyanobacteria: *Prochlorococcus* and *Synechococcus*. Although phylogenetically closely related, these organisms differ by their pigmentation and light-harvesting complexes. *Prochlorococcus* is very abundant in the intertropical, nutrient-depleted (oligotrophic) regions of the world's ocean. In contrast, *Synechococcus* is more ubiquitous and is abundant in nutrient-richer (mesotrophic) areas, such as near upwelling or coastal areas. The very large distribution of these two genera suggests that they possess efficient strategies to respond to environmental stress (high light, UV radiations, nutrients deficiency, etc). However, the fine-tuning of these responses was largely unknown at the beginning of the project.

The recent availability of complete genomes from genotypes adapted to different ecological niches provided an unprecedented opportunity to identify and characterize the expression of niche-specific genes, many of which may be novel and/or unique. Three full genome sequences of *Prochlorococcus* adapted to different light intensities and of two *Synechococcus* strains, isolated from an oligotrophic and a mesotrophic environment respectively, were available at the beginning of the project. Already these genomes constituted a huge amount of information which made the systematic study of gene functions in these organisms possible. Genome comparisons have shown that *Synechococcus* is physiologically much more versatile than *Prochlorococcus*. However, functions were assigned only to about 60% of the genes of any organism and there were neither functional genomics tools nor genetic analysis tools available at the beginning of the project. Thus, methods that can be used to study the function of thousands of genes in parallel needed to be developed to make efficient use of genomic information. The Flagship project SYNCHIPS focused on *Synechococcus* WH7803 as a new model for the marine *Synechococcus* group. The total genome sequence of this strain was sequenced by the Genoscope at our request and has been annotated with substantial effort (Dufresne et al., 2008). Within the frame of this project a total genome oligonucleotide DNA array was constructed for expression analysis, a high density tiling array for interrogating intergenic spacer regions established, several mutant lines constructed and characterized and substantial biocomputational tools developed. In combination with a newly set up cyclostat system (a computer-controlled illumination system allowing to finely modulate growth irradiance), it had been our major objective to use technologies to simulate quasi-natural growth conditions in the laboratory, to study global gene expression with an emphasis on environmental stress conditions and to characterize functions of key regulatory genes that had remained unknown so far. Moreover, we have been aiming at a much stronger integration of the participating groups and laboratories in order to maximize the outcome from the available funding.

II.3.2 – Work performed in the SynChips project
II.3.2.1 – Scientific aspects

The SYNCHIPS project consisted of three workpackages. Two had a focus on methodological development (trancriptomics and mutagenesis) and within the third one these new technologies were applied to study the kinetics of response of *Synechococcus* cells to environmental stresses naturally encountered by cells in the field, namely high light, UV stress and/or nutrient deficiency (N and/or P). Stresses were studied alone or in combination and these studies were conducted on continuous cultures as well as by simulating natural conditions using the cyclostat. As a precondition for all functional genomics work we put emphasis on pushing the analysis, annotation and interpretation of *Synechococcus* total genome sequences to a new level.

II.3.3 – Integrative aspects

Major emphasis was put on the integrative aspects among the participating scientists and laboratories. A lasting impact for significantly improved integration was sparked by the "Marine Picocyanogenomics Annotation and Genome Data Interpretation Jamboree" that was held in February 2006 in Freiburg (partner #45). At this occasion all scientists (and not only the PIs or one representative per lab) participating in SynChips gathered for four full days. During the whole reporting period, representatives from the different labs have been meeting or visiting each other on several occasions: a PhD student from partner #45 has been travelling for experimental work to Roscoff (#2), whereas a post-doc from partner #2 has been for an extended period at partner #17's laboratory to strengthen their bioinformatics capabilities, in return two scientists from partner #17 have been for experimental work in Roscoff (#2). The PI from the lab at Hebrew University (#26) spent 3 days in Freiburg (#45) in December 2006 for joint work on data analysis. Similarly, three of the four PIs and additional scientists from all SynChips partner labs participated in the MGE Exploratory workshop "New approaches for functional genomics of marine microorganisms" Berlin, Germany, June 22-23 2007, which was organized by partner #45. Finally, two important and central meetings during the project period were used to

discuss progress and work within SynChips among partners; these were the MGE Marine Genomics meeting in Sorrento, Italy, in late October 2006 and the ISPP world congress in Pau, France, in August 2006.

II.3.4. Major results of the SynChips project

The database of well-annotated *Synechococcus* genomes as a fundamental precondition to all work in this field was significantly expanded by the work of our consortia towards the analysis, annotation and interpretation of altogether 11 *Synechococcus* total genome sequences (Dufresne et al., 2008). In the natural environment, phytoplankton cells have often to cope with different co-occurring physico-chemical (high light, UV radiations, nutrient deficiency, temperature changes) and biological (viral infections, predation) stresses. To be able to better understand physiological responses to stress, it is therefore necessary to subject cells to these stresses separately as well as in association. Light shift experiments of *Synechococcus* sp. WH7803 have shown the importance of the culture light history on the strength of the physiological response to stress. Two components of the photosynthetic apparatus (phycobilisomes complexes and reaction center II) seem to play a crucial role in the response to both high light and UV stresses (Six et al., 2007a; Garczarek et al., 2008). Furthermore, a cyclostat was developed and successfully used within SynChips which allowed us to expose cells to modulated visible light alone or together with UV-A and/or UV-B radiations. Simulating a true light:dark (L:D) cycle rather than using the classical ON:OFF L:D cycle allowed us to better synchronize cells and study fine changes in the timing of individual gene expression, together with the analysis of genome replication and cell cycle (Holtzendorff et al., 2008; Boutte et al., in prep; Kolowrat et al., in prep). Within the frame of this project a total genome oligonucleotide DNA array was constructed for expression analysis, a high density tiling array for interrogating intergenic spacer regions established, several mutant lines constructed and characterized and substantial biocomputational tools developed (Axmann et al., 2007; Voss et al., 2007).

Synechococcus spp. cells can virtually be found in any seawater sample. Some *Synechococcus* strains are able to change their pigmentation and adapt to blue, nutrient-poor waters of the open ocean or the green colours of nutrient-rich waters but the majority of *Synechococcus* strains is genetically adapted to one or the other condition. By comparing the various *Synechococcus* genome sequences, a well-focused study was performed to understand the biochemical and genetic basis of these processes (Six et al., 2007b).

The markedly different nutrient regimes experienced *in situ* by marine *Synechococcus* strains inhabiting oligotrophic *versus* mesotrophic or coastal environments suggest that they have developed differential nutrient acquisition and regulatory networks, implying different responses to nutrient stress. These stress responses fall into two categories: adaptation and survival. Adaptation involves the induction of high affinity nutrient uptake systems facilitated by e.g. *nrtP* and *pstS* expression which permit improved acquisition of nitrate and phosphate respectively. Moreover the adaptive process may involve induced acquisition of alternative nutrient sources (e.g. urea, amino acids, organic phosphate, phosphonates) which support continued growth, often at maximal rates. The survival process involves induced degradation of intracellular compartments like phycoerythrin and polyphosphate which serve as nutrient pools for those cell constituents that are essential during growth arrest and survival. We expect genetic/physiological differences among marine cyanobacterial strains to be consistent with the niches in which these organisms are found. In particular, the structure of the N and P regulons, which are largely uncharacterized in photosynthetic organisms, will yield new information on the molecular bases of niche adaptation.

The obtained data ultimately help to be able to determine the physiological status of natural populations submitted to specific niches in the ocean. Understanding these processes will be of critical value to assess the effects of environmental changes in marine systems. A major part of the results obtained within SynChips will be published within a comprehensive paper highlighting the genomic basis of ecological success of these marine microorganisms scheduled to appear early next year (Scanlan et al., 2009).

II- 4 FLAGSHIP PROJECT FICELL (WP28)
Functional genomics using marine fish cell lines and embryos

Project leader: M. Leonor Cancela, CCMAR, EDGE, Partner N°21

II.4.1. Objectives

The main objective of work package #28 was to develop molecular tools towards the identification of gene function in fish. The large pool of available genome sequence analysis, obtained through Marine Genomics Europe NoE requires functional genomics to assign a function to specific genes based upon experimental rather than in silico only evidences, thus requiring: (1) in vitro cellular models from various tissues to answer specific questions on gene function and regulation, (2) in vivo fish models suitable to verify the functional significance of selected genes, and (3) molecular tools for large scale analysis of gene function. In vitro studies on specific gene function and regulation were done using gilthead seabream [Sparus aurata] cell lines/cultures already available (bone, branchial arch and fin) or developed within the scope of this work package. In vivo/ in vitro analysis of specific gene function was performed by silencing splicing/translation using miRNA, RNA interference (short hairpin RNA technology) or morpholinos or by up-regulation using expression vectors functional in both fish cells. Embryos from Atlantic killifish [Fundulus heteroclitus] and zebrafish [Danio rerio] were used for in vivo studies. Global analysis of gene expression was performed using seabream microarrays developed within the scope of MGE. (Summarized in Figure 1)

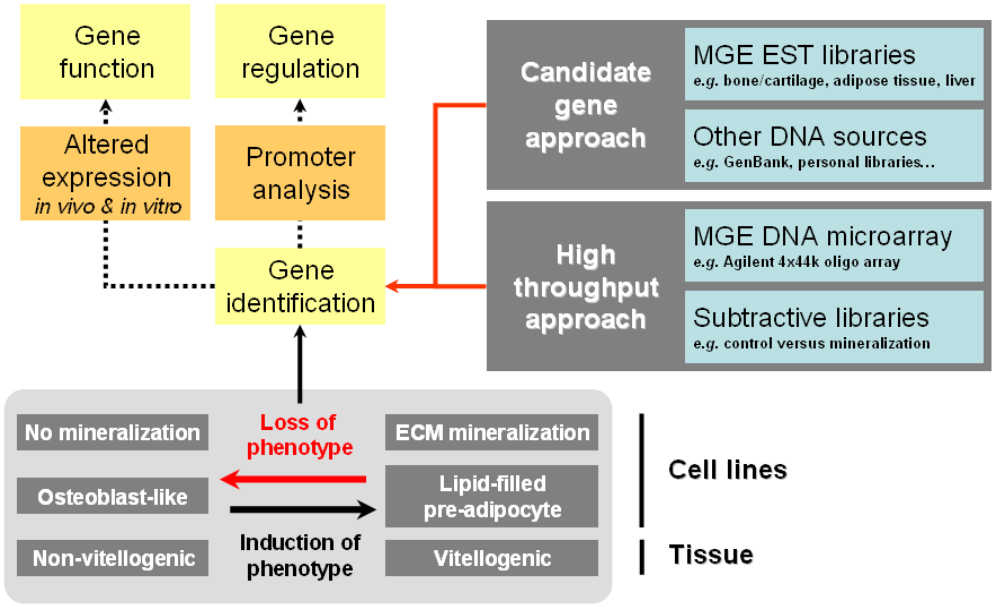


Figure 1: High throughput approach according to WP28. FICEL-WP28 is a pilot project to initiate development/testing of high throughput molecular tools towards the identification of gene function in fish.

II.4.2. Partners involved in WP28

#21 Centre of Marine Sciences (CCMAR), Faro, Portugal
#25 Israel Oceanographic and Limnological Research (IOLR), Haifa, Israel
#34 Institut for Food and Agricultural Research & Technology (IRTA), Barcelona, Spain
#35 University of Barcelona (UB), Barcelona, Spain

II.4.3. Most significant achievements of WP28-FICEL

II.4.3.1 Development/characterization of S. aurata cell cultures/lines from selected tissues

S. aurata cell lines derived from bone, branchial arch and fin tissues were already available in partner #21 laboratory (Figure 2). Additional ones were developed: (1) hepatocyte primary culture (partner #25) (2) adipocyte from transdifferentiation of osteoblast-like cell line (partner #21; Figure 2), (3) embryonic stem cell

line (partner #21), (4) osteoclast primary cultures (partner #21), and (5) osteoclast progenitor cell culture (partner #21). Characterization included optimization of culture conditions and transfection methods, as well as expression and production of specific genes and proteins in control and specific treatment conditions.

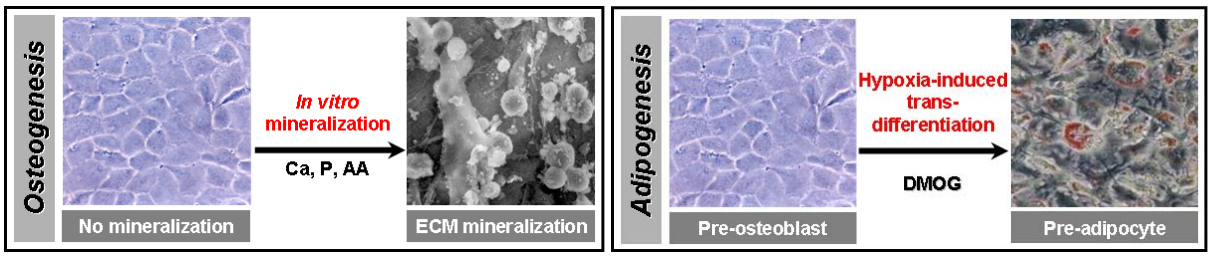


Figure 2: Seabream bone-derived cell undergoing mineralization (*left panel*) or adipocytic transdifferentiation (*right panel*)

II.4.3.2 Development of molecular tools and functional analysis of selected gene promoters

Identification of genes involved in a particular physiological process: Several strategies were used to identify genes involved in physiological processes studied within the scope of WP28-FICEL project (i.e. osteogenesis, adipogenesis and vitellogenesis). Genes were identified from (i) ESTs generated within the scope of MGE, annotated sequence databases or standard cDNA cloning, (ii) from subtractive libraries (e.g. mineralizing versus control cDNA), and (iii) from global expression analysis using DNA microarray. Seabream oligo-array developed within the scope of Fish & Shellfish node in collaboration with Agilent (where each unique sequence is represented by two non-overlapping probes (60mers) and each array is composed by 39,399 probes) has been made available to WP28-FICEL partners in September 2007. Preliminary results on genes differentially expressed during in vitro mineralization (partner #21), vitellogenesis (partner #25) and adipogenesis (partner #35) were obtained recently (Figure 3) and are being processed. Tissue distribution of a number of genes involved in osteogenesis, vitellogenesis and adipogenesis has been characterized by quantitative real-time PCR in adult fish and during development by partners #21, #25 and #35.

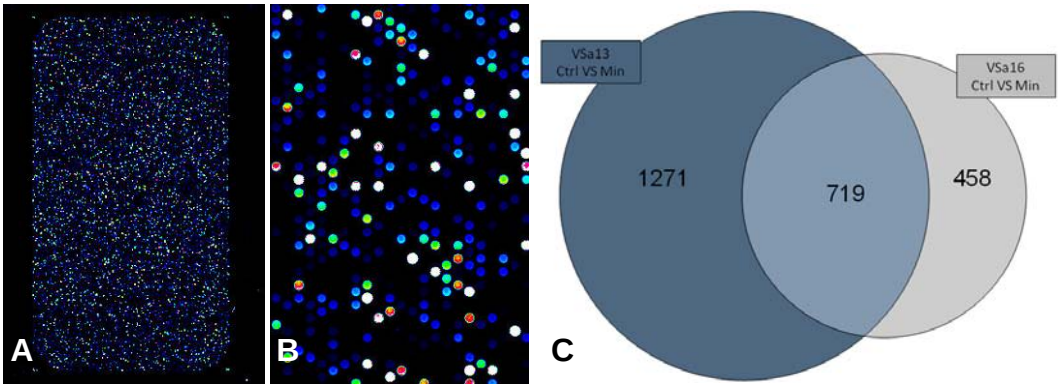


Figure 3: Seabream oligo-array hybridized (collaboration with L. Bargelloni & S. Ferrareso; partner #24) with RNA prepared from bone-derived cells grown under control or mineralizing conditions in order to identify genes differentially expressed and therefore likely to be involved in mineralization mechanisms. **A**, 44K array (each slide contains four 44K arrays) hybridized with control RNA; **B**, zoom in the array showing a selection of spots hybridized with labeled probes (differences in color intensity represent differences in gene expression level) ; **C**, Venn diagram illustrating the number of genes differentially expressed in VSa13 and VSa16 cells, as well as in both cell lines.

microRNAs involved in E2 regulation of vitellogenesis: Differentially expressed microRNA were identified by partner #25 upon estrogen exposure (vitellogenesis is an estrogen-regulated physiological process) using a zebrafish microRNA array. Results provided valuable information to analyze corresponding effects in seabream.

Altered expression of selected genes: Overexpression / silencing of mineralization-related genes was achieved by partner #21 and effect on growth performance & ability to mineralize extracellular matrix characterized using in vitro cell systems. These strategies appeared to be successful and thus suitable for high-throughput approaches.

Functional promoter analysis for mineralization related genes: A number of promoter sequences were identified and cloned into pGL2basic vectors by partner #21 to study regulation of mineralization-related genes. The molecular characterization and functional analysis of multiple promoters (e.g. MGP, BMP2, FHL2, SDR, IGF-1 and OP, Figure 4) were performed and many more promoters (of genes identified through microarray analysis) are in WP28-FICEL pipeline. Data allowed identification of key regulators/regulatory motifs. For example, MGP was shown to be controlled by M concentrations of retinoic acid, a metabolite of vitamin A (Figure 5).

Functional promoter analysis for genes related to adipose tissue function: Partner #35: The promoters of glucose transporter GLUT4, considered a marker for adipogenesis, and interleukin-6 (IL-6), an adipokine (cytokine secreted by adipose cells), were cloned into pGL3basic vector, transfected into L6 cells (although not piscine in origin, this cells are appropriate for expression of various fish adipose or muscle-related promoters) and characterized in terms of regulation by hormones and cytokines.

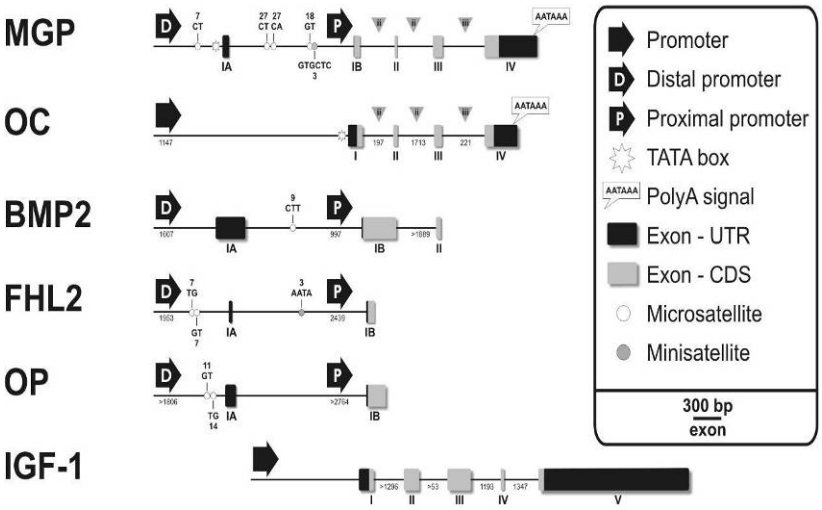
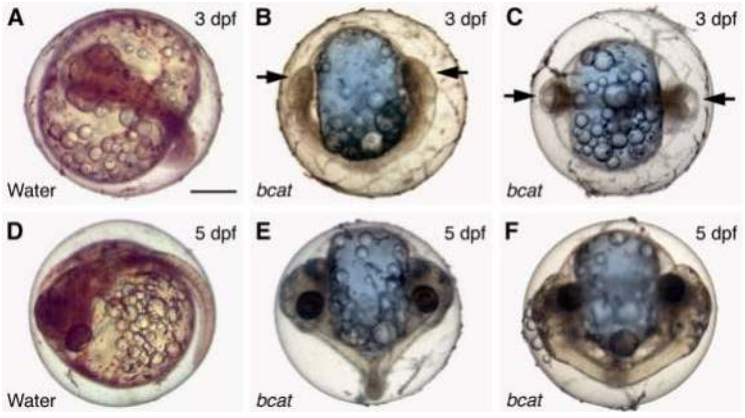


Figure 4: Organization of seabream mineralization-related genes for which promoter activity is analyzed. MGP, matrix Gla protein; OC, osteocalcin; BMP2, bone morphogenetic protein 2; OP, osteopontin; IGF-1, insulin-like growth factor 1.

II.4.3.3 In vivo studies of gene function using marine fish embryos

Methods for over-expression of specific genes, RNAi and dominant-negative mutation were developed for killifish embryos by partner #34 (Figure 5). The analysis of gene expression of different water channel aquaporins (AQPs), during killifish embryo development revealed that aquaporin-3 (AQP3) was highly expressed during gastrulation. This expression appeared to be down-regulated during removal of water from embryos. Development of dominant-negative methods based on engineered AQP3 with altered permeability through site-directed mutagenesis to identify specific amino acids implicated in water and solute permeability showed that changes in one single amino acid significantly reduced AQP3 permeability at three different pH.



II.4.3.4 Development of informatic tools to disseminate achievements and provide new source of information

FICEL webpage, a webpage describing the WP28-FICEL project within the scope of the MGE, and FICEL database (FICELdb) were created and made available to the scientific community through the institutional webpage of partner #21 (fcma.ualg.pt/EDGE/Ficel/main.htm) with link in MGE site.

II.4.3.5 Dissemination and education/training

Results obtained within the scope of FICEL-WP28 resulted in 12 publications in peer-reviewed journals (more are expected, e.g. from microarray data analysis, cell line characterization and functional promoter analysis) and 14 presentations in meetings within and outside MGE (including 2 oral presentations). Nine students (including 3 PhD students partially financed through FICEL) have been trained within the scope of this project.

II.4.4. Conclusions

This pilot project, in addition to its obvious scientific achievements, contributed to develop important molecular and cellular tools and develop new information suitable to promote (i) use of state of the art technology, (ii) interactions among different partners/institutions from different countries, (ii) training of younger scientists/students, (iv)dissemination of information through web site and data base created within this project, and (v) contribution to establish a new European Platform recently funded by EU through FP7 (Project ASSEMBLE) where part of the knowledge and tools obtained during FICEL will be shared with scientific community through services and future collaborations.

II- 5 FLAGSHIP PROJECT "GARNET" (WP38)

Genomic analysis of regulatory networks in evolutionary time

Project leader : D. Arendt, EMBL, Partner 11

Objectives

Key changes in the evolution of animal body plans are thought to be caused by changes in the regulatory genomic networks that govern animal development. To detect such changes in the developmental "hardwiring" of cell types, organs, and tissues, it is crucial to identify the key players of these regulatory networks and their respective interactions in different species, followed by a comparison of the resulting network topologies. In-depth analysis of terrestrial invertebrate model species (such as fruit flies or nematodes) has generated detailed knowledge about several regulatory cascades. However, apart from strongly conserved processes, many regulatory networks that act on a higher organisational scale exhibit very large differences between these 'classical' invertebrate model species and vertebrates, often precluding the simple reconstruction of the ancestral states and the major evolutionary steps that link these divergent systems.
The aim of the GARNET project was to develop techniques and generate resources to close the gap of regulatory model systems between insects, nematodes and vertebrates. The major objectives were:

(i) the isolation of orthologous gene loci from each of the cornerstone species, as well as a related reference species, to allow comparisons on the level of non-coding DNA

(ii) the establishment and use of functional tools to dissect gene-regulatory interactions and assess the regulatory capacity of candidate regulatory elements.

Organization and progress of the project

The project brought together scientists from seven different European locations, and with expertise in five different animal groups (see **Table 1**).

Name	Institution	MGE #	Coord.	Main system
D. Arendt	EMBL	#11	WP1, O	annelids
F. Raible	EMBL / MFPL	#11	WP3, O	annelids
P. Lemaire	CNRS	#2	WP2, S	ascidians
I. Arnone	SZN	#22	WP4	echinoderms
R. Reinhardt	MPI-MG	#6		resource center
A. Poustka	MPI-MG	#6		echinoderms
G. Panopoulou	MPI-MG	#6		cephalochordates
J. Garcia-Fernandez	UB	#35		cephalochordates
P. Martinez	UB	#35		acoels
M. Thorndyke	KMRS	#38		echinoderms

Table 1. Affiliation, role, and interest of participating laboratories. 'MGE#' : partner number of affiliated institution ; 'Coord' : Coordinator roles : Work packages (WP1-WP4), Sequencing (S), Overall coordination (O).

The project was gouped into four different work packages :

WP1: Identifying conserved players in regulatory networks

The goal of the first work package was to identify and study a set of genes that would serve as entry points into studying gene-regulatory interactions in each of the cornerstone species. As there was a shared interest between the groups in the evolution of nervous systems, this gene list primarily focused on regulator and effector genes with a likely role in the establishment of functional nervous systems.

WP2: Prediction of gene regulatory interactions by phylogenetic footprinting

The aim of the second work package was the identification of candidate regulatory elements by analyzing – for a given gene locus – genome alignments between each cornerstone species and a suitable reference species.

To approach this goal, major efforts were dedicated to the selection of suitable reference species for the cornerstone species. This involved manual isolation of related gene loci from 10 different annelid species, as well as generation and partial sequencing of two ascidian BAC libraries. Likewise, identification of orthologous genes in the reference species required the generation of several new EST datasets by the Génoscope, and two different screening procedures (multiplex PCR screening and filter hybridization) were tried to find a suitable strategy for mass identification of BAC clones.

WP3: Development of transgenesis and functional interference techniques

The aim of the third work package was to establish functional interference as well as robust gene delivery protocols in species lacking functional approaches, and to prepare cross-species enhancer tests to detect evolutionary conservation and evolutionary change in gene-regulatory networks.
As functional techniques were already established for the ascidian and echinoderm cornerstone species, this work primarily focused on the annelid and cephalochordate systems. Major progress was made in the injection of morpholino antisense oligomers and DNA reporter constructs into Platynereis dumerilii eggs. Similar efforts proved less successful in the amphioxus Branchiostoma lanceolatum, primarily due to the restriction of experiments to the spawning season of the animal.

WP4: Functional analysis and assembly of regulatory networks

The focus of the fourth work package was on the use of functional tools to assay candidate regulatory interactions, thereby providing insights into the gene regulatory networks for each of the analyzed groups. As it became necessary to dedicate more time and effort than initially planned on the selection of proper reference species and generation of the corresponding resources (WP2), these comparisons could not be performed at the intended scale, but remained restricted to case studies in sea urchins, ascidians and annelids. However, progress was made in the extraction and visualization of regulatory interactions in the context of several data warehouse solutions developed by the partners, thereby providing a conceptual framework for storage and analysis of this type of data.

Perspective

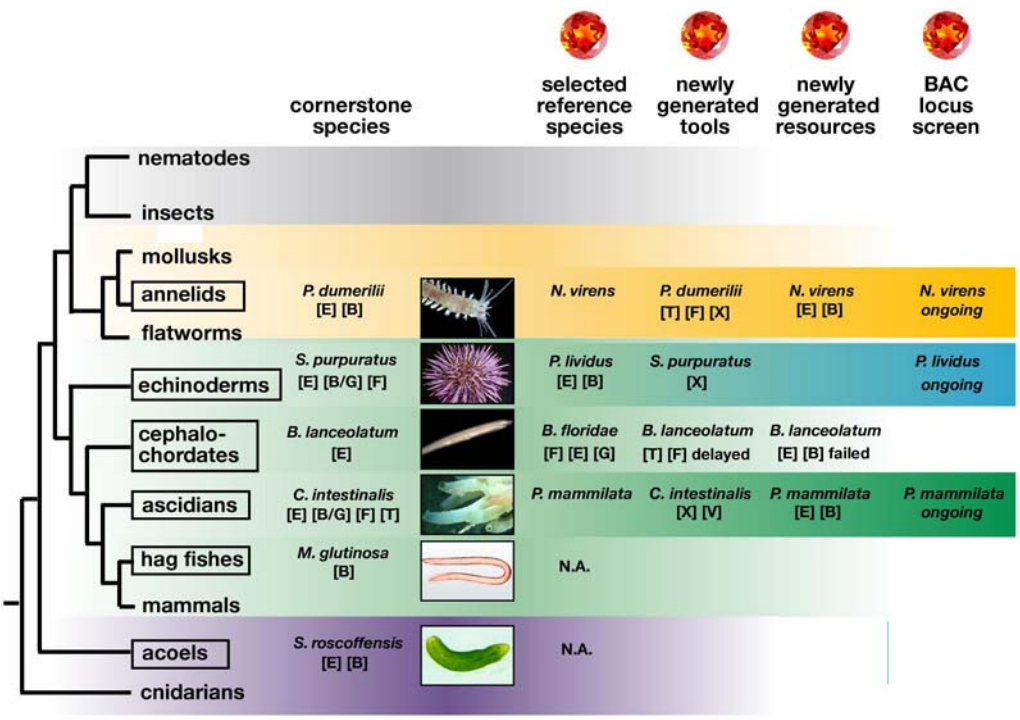


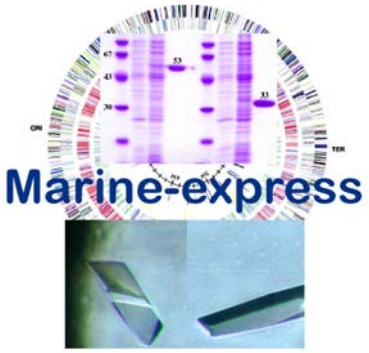
Fig. 1. Overview over the tools and resources generated in the context of the GARNET project. Cornerstone species belong to six different animal groups (boxed), occupying important positions for the analysis of regulatory network evolution. Major outputs of the GARNET project (highlighted in red) were the identification of suitable reference species for phylogenetic footprinting, and the establishment of new tools and genomic resources in cornerstone or reference species. For three of the animal systems, these efforts are being continued by a BAC screen for orthologous gene loci. [E]: EST data, [B]: BAC resources; [G]: Genome sequence; [T]: Gene delivery/Transgenesis; [F]: Functional Interference; [X]: Expression warehousing; [V]: Gateway vector system for cross-species experiments.

Overall, the GARNET project has served to generate key tools and resources that will facilitate successful gene-regulatory analyses in at least three of the cornerstone species, while also providing examples for the other species (Fig, 1). As the identification and sequencing of the correct BACs in the respective reference species is just starting, major parts of the genomic comparisons and the subsequent analyses of gene-regulatory interactions in the cornerstone species are still to be completed, and answers to higher-order questions as to the evolution of gene-regulatory interactions are still elusive. However, the commitment of the Génoscope to identify and sequence the full set of reference species BACs provides a solid basis for the continuation of the started research as a collaborative project among the involved partners. Moreover, research in the respective model systems makes progress in the implementation of phylogenetic footprinting into the framework of cross-species databases. The development of such tools will ensure optimal use and interpretation of the additional genomic data, once they become available.

II- 6 FLAGSHIP PROJECT MARINE EXPRESS (WP39)

Project leader: M. Czjzek, CNRS, Partner2

'Post-genomic validation of marine genomics through the development of medium throughput tools for expression, crystallization and functional screening of target genes' (Marine-express)



Partners :

1. Station Biologique de Roscoff, 'Végétaux Marins et Biomolécules' UMR7139, CNRS-UPMC. Place George Teissier, BP 74, 29682 Roscoff cedex, France **(SBR) microbial, algal node partner 2** together with Max-Planck-Institut für Marine Mikrobiologie, Department of Molecular Ecology, Celsiusstrasse 1 D-28359 Bremen, Germany **(MPI-MM) microbial node partner 4**

2. Lab. Biochem. and Molecular Biology Stazione Zoologica A. Dohrn Villa Comunale 80121 Napoli, Naples, Italy **(SZN) EDD node partner 22**

3. IFREMER, Microbiology of Extreme Environments UMR 6197, IFREMER-CNRS-UBO, Technopole Brest-Iroise, BP 70, 29280 Plouzané, France (**IFREMER) microbial node partner 3**

4. Centro de Ciências do Mar, Universidade do Algarve, Comparative & Molecular Endocrinology, 8005-139 Faro, Portugal **(CCMAR) Fish and Shellfish node partner 21**

5. Observatoire de Banyuls-sur-mer, Models in Cellular and Evolutive Biology, CNRS-UPMC, UMR 7628, Banyuls-sur-mer, France **(OBS-Banyuls) algal node partner 2**

6. Marine Biological Association, The Laboratory, Citadel Hill Plymouth PL1 2PB **(MBA) algal node partner 16**

II.6.1. Objectives of the project

The marine environment is highly diverse and contains the vast majority of known and unknown biodiversity. It is also the last frontier to understand the control of the global climate and hides a wealth of biological resources still to be tapped for food, health and energy. Up until very recently, little genomic data were available for oceanic organisms, but the panorama is rapidly changing with a number of genomic projects now finished or still underway which focus on marine organisms, ranging from microbes[1,2] to multicellular eukaryotes including vertebrates or macro-algae[3], as well as the generation of resources and access to genomes or EST libraries for various eukaryotic systems[4,5] (Table 1).

After this great success of genome sequencing through MGE programs, researchers are more and more confronted with a lack of experimental data to match the wealth of bioinformatic information; data that is however crucial for the post-genomic validation of the genome data. This lack is due in the first place to the introduction of a bias by numerous model organisms such as yeast, certain pathogenic organisms or higher plants like Arabidopsis and the fact that marine organisms are far less studied than these. Automatic annotation is therefore also biased and up to 60% of the marine genome data have no or only poor equivalents with experimentally confirmed biological and/or biochemical functions. The objectives of the Flagship program 'Marine express' was to cope with the unbalance between experimental information and genomic data by applying and adapting medium throughput protein expression methods to produce a rather large number of proteins that are of interest to MGE members. Medium/high throughput methods are known to help get around the bottlenecks of soluble protein expression and crystallization with the aim of obtaining sufficient material to perform the biological, biochemical and/or structural characterization by the partners interested in the respective protein families.

Table 1. Availability of genomic data to the project consortium

Organism	Sequencing center	Taxonomy	Gene families of interest	Partner
Pyroccocus abyssi	Genoscope	Bacteria, Euryarchaeota	Proteins from the DNA replication system	D.Flament J.Querelou IFREMER
Rhodopirellula baltica	MPI Berlin	Bacteria, Planctomycetales	Polysaccharidases, sulfatases	M.Czjzek SBRoscoff R.Amann MPI Bremen
Zobellia galactanivorans	Genoscope	Bacteria, Flavobacteriales	Polysaccharidases, sulfatases	M.Czjzek SBRoscoff
Ectocarpus siliculosus	Genoscope	Eucaryote, Brown algae	Stress related genes, innate defense response, Polysaccharide synthases	M.Cock C.Boyen SBRoscoff
Ostreococcus	Genopole Montpellier	Eucaryote, green micro algae		H.Moreau Banyuls
Ciona intestinalis	JGI	Eucaryote, Urochordate	Hox-genes, Ci-msx, Ci-Rx	M.Brano SZNapels
Sparus aurata	MPI Berlin	Eucaryote, teleost fish	Hormones, calcium and musculo-skeletal development, stress related genes	D.Power A.Canario CCMAR

1 Cohen G, Barbe V, Flament D, Galperin M, Heilig R, Lecompte O, Poch O, Prieur D, Quérellou J, Ripp R, Thierry JC, Van der Oost J, Weissenbach J, Zivanovic Y and Forterre P. (2003) Mol Microbiol. **47**, 1495-512
2 Glockner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, Gade D, Beck A, Borzym K, Heitmann K, Rabus R, Schlesner H, Amann R, Reinhardt R. (2003) Complete genome sequence of the marine planctomycete Pirellula sp. strain 1. *Proc Natl Acad Sci* U S A**. 100**, 8298-303.
3 Peters AF, Marie D, Scornet D, Kloareg B, Cock JM. (2004) Proposal of Ectocarpus siliculosus as a model organism for brown algal genetics and genomics. *J. Phycol*. **40**, 1079-88
4 Satou Y, Satoh N. Draft genome sequence of Ciona intestinalis and its meaning. Tanpakushitsu Kakusan Koso. 2003 Jul;**48**(9), 1282-6. Review
5 Collén J, Guisle-Marsollier I, Léger JJ, Boyen C. (2007) Response of the transcriptome of the intertidal red seaweed Chondrus crispus to controlled and natural stresses. *New Phytol*. **176**, 45-55.

II.6.2. Results of 'Marine express'

'Marine express' has proven a valuable tool for the heterologous production of a large diversity of functional proteins originating from various organisms (Table 2). Two major expression rounds have been performed over a time period from June 2006 to August 2008 and a third, so called "rescue round", is currently under way. Overall, the success rate of 'Marine express' is **84** solubly expressed proteins out of **192** selected and cloned targets, which corresponds to **43.7%** soluble expression. The expression of about **40** of the successful targets has since been up-scaled for production of mg quantities of the proteins of interest. Numerous of these proteins have then been tested for their functionality and the biochemical and structural analysis have successfully been initiated thanks to the expression protocols provided by 'Marine express'.

Table 2. Results of the 2 major protein expression trials as a function of origin

				Combined expression	pFO4 + pGEX **soluble**	
Organism	No of genes R1	No of genes R2	**total**	Round 1	Round 2	Round 1+2
Eukaryotes	**37**	**41**	**78**	**10**	**14**	**24(30%)**
S. auratus (+ other fish)	**16**	**25**	**41**	**5**	**8**	**13**
Ciona intestinalis	**8**	**2**	**10**	**1**	**1**	**2**
Macro algae	**7**	**10**	**17**	**2**	**4**	**6**
Strongylocentrotus	**4**	**0**	**4**	**2**	**0**	**2**
Ostreococcus	**2**	**3**	**5**	**0**	**1**	**1**
Bacteria	**6**	**31**	**37**	**2**	**23**	**25(67%)**
Zobellia galactanivorans	**6**	**28**	**34**	**2**	**22**	**24**
P. haloplanctis	**0**	**2**	**2**	**0**	**1**	**1**
P. carrageenovora	**0**	**1**	**1**	**0**	**0**	**0**
Archae	**53**	**24**	**77**	**23**	**12**	**35(45%)**
Archae	**40**	**13**	**53**	**19**	**7**	**26**
Virus	**13**	**2**	**15**	**4**	**0**	**4**
Plasmids	**0**	**9**	**9**	**0**	**5**	**5**

The following presents the scientific context of a **selection of 3** of these targets to illustrate the impact of soluble protein expression for the biochemical and/or functional characterization:

Partner #2 Functional and structural analysis of target no. 30, Z3597. The target protein **Z3597** has been selected as a putative glycoside hydrolase with an interesting gene context from the genome of the marine bacterium *Zobellia galactanivorans*. It was overexpressed and purified following the protocol established in the medium throughput strategy round 1, and led to a final yield of about **15 mg of pure protein**. Interestingly, two distinct multimeric forms (a dimer and a trimer) could be separated by gel filtration. In view of the genome context, activity was tested on various algal polysaccharides and a **novel carrageenolytic activity** (that potentially may lead to a patent) could be identified for **Z3597**. In parallel to the activity essays, the pure protein was subjected to medium throughput screening of crystallisation conditions and first crystals were produced. A thorough study of the structure-function relationship is therefore actually under way. The protein solution was also analysed by small angle X-ray scattering (SAXS) at the DESY laboratory (EMBL, Hamburg). And a first global structural envelope of the protein in solution has been obtained (Figure 1).

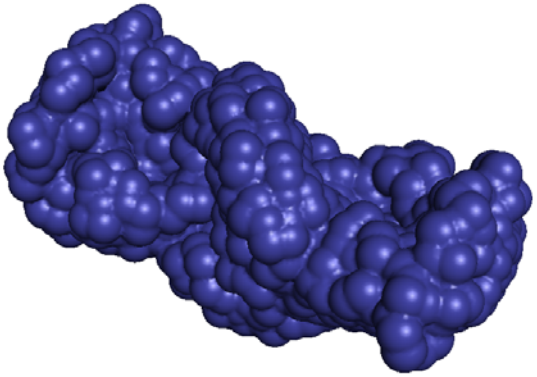


Figure 1: Envelope of the dimeric form of Z3597 in solution as determined by small angle X-ray scattering (SAXS). The experiments were performed in november 2007 at the Hambourg EMBL-Desy SAXS Beamline X11.

Partner #3 Structural and functional characterisation of protein complexes involved in genomic maintenance from a hyperthermophilic archaea *Pyrococcus abyssi* targets no. 34, 37, 45-47, 95

In order to address the question of the interactome involved in DNA repair mechanisms of hyperthermophilic Archaea, we intend to look for the interacting partners of most of the proteins of the genomic maintenance and chromosome packaging repertoire from *Pyrococcus abyssi*. To achieve this medium throughput approach, we have selected about 70 targets, from the genome content, that are likely to be involved in genomic maintenance in *P. abyssi*. At the end of marine express round 1, 16 soluble proteins were produced and purified. Among this set of now available proteins, we have scaled up production and purification for 6 proteins (RPA14, RPA32, RecJ, RadA, Gins14, Gins23). In order to identify additional proteins that interact *in vivo* with these proteins, we have used a combinatorial approach that couples pull-down analyses in the cellular extract of *P. abyssi* with mass spectrometry. In brief, the recombinant tagged proteins, immobilized on magnetic beads, are incubated with the cellular extract of *P. abyssi* and the partners of these molecular baits are then separated on SDS-PAGE, as exemplified on Figure 1. The bands corresponding to potential partners are cut out from the gel and identified using mass spectrometry analyses (MALDI-TOF, MALDI-TOF/TOF and LC-MS/MS). The interaction detected will be then confirmed using a second independent method (SPR, Co-immunoprecipitation, ..).

We are currently producing the 10 remaining proteins and the identification of the interacting partners of the already produced proteins is under investigation. Fig1 shows the results obtained with RadA that tend to indicate that, as for the Eukariotes, the RadA filament could be regulated via proteasome degradation and that the reverse gyrase, unique to hyperthermophile organisms, could be involved in homologous recombination repair mechanisms.

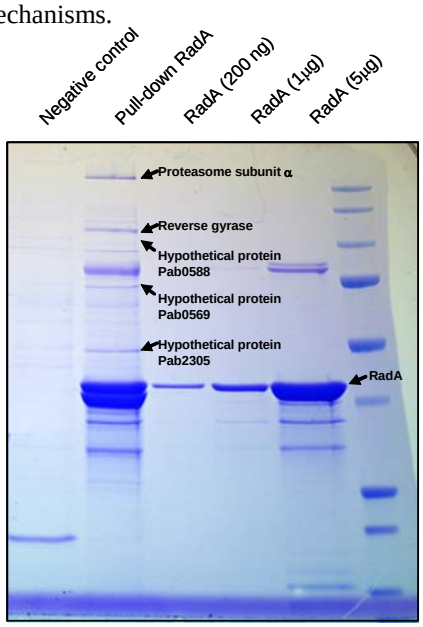


Figure 2: 12% SDS-PAGE analysis of a *P. abyssi* RadA pull-down experiment. Proteins identified using an Ultraflex MALDI-TOF/TOF instrument (Bruker Daltonics) are indicated with arrows.

Partner #21 Functional characterization of proteins from *Sparus aurata* involved in handling and regulation of calcium; targets no. 1, 4, 13 and 14

A number of fish cDNAs were provided to Station Biologique de Roscoff for generation of expression constructs and expression of fusion protein. Several categories of fish cDNA were provided and included hormones,

hormone responsive transcription factors, chaperones and matrix proteins involved in hormone driven calcium turnover. On receipt of the expression constructs giving good expression of either soluble or insoluble fusion protein stocks were prepared and stored. Several constructs were then used to transfect BL21 bacteria and induced to produce recombinant protein. Target constructs used for further studies included sea bream ASPIP, sea bream Osteonectin, Dax 1 and Sox 9.1.

ASPIP: was first identified in human cartilage in 1998 and so the function of the protein still remains to be established. Antisera previously produced against sea bream ASPIP, was tested using the recombinant protein generated using the Roscoff expression constructs. This revealed that the polyclonal antisera produced was high titre (1/12,000) and specific for ASPIP. Further studies are underway to characterise the tissue distribution and processing of ASPIP.

Dax 1 and Sox 9.1: are transcription factors involved in sex determination in fish although their mode of action is unknown. In order to establish if these proteins interact, a pull down assay was developed using the recombinant proteins. To this end expression constructs of Dax 1 with a His tag and Sox 9.1 with a GST tag were used to transfect bacteria BL21and protein expression was induced with IPTG. Recombinant proteins were then purified using a His-Bind Resin Kits (Novagen, Germany) or GST bind kit (Invitrogen, Portugal). In brief, the recombinant proteins were extracted in a soluble fraction from pellets of *E. coli* large scale cultures by mechanical disruption and then extracts passed trhough appropriate equilibrated columns and the fusion proteins purified. Protein:protein interactions were studied using pull-down assays followed by acrylamide gel electrophoresis and Western blotting. This revealed that Dax1 and Sox 9.1 bind to each other. Furhter studies are underway to characterise the interaction and in the future a similar approach will be deployed using Dax 1 or Sox 9.1 as the bate to isolate proteins present in cellular extracts which interact with the proteins.

II-7 THE BIOINFORMATIC PROJECT (WP15-18)

Project leader: Dr. Alexander Goesmann, Bielefeld University, Partner 7

Bioinformatic Project managers: Jomuna Choudhuri & Virginie Mittard

Bioinformatic project committee:

- Olivier Collin, CNRS Roscoff, FR, Partner 2
- Chris Bowler, CNRS, Paris, FR, Partner 2
- Frank-Oliver Glöckner, MPI Bremen, DE, Partner 4
- Patrick Lemaire, CNRS Marseille, FR, Partner 2
- Richard Reinhardt, MPI Berlin, DE, Partner 6
- Thomas Schlitt, NERC-BAS, Cambridge, UK Partner 12 (replaced by Nicolas Bierne after one year)

The major goal of this project was to establish the Marine Genomics Europe (MGE) bioinformatics platform in order to provide general hardware and software support to all research groups of the Marine Genomics Europe consortium within the manifold genome and post genome projects. As a central resource, the Bioinformatics Resource Facility (BRF) of the Center for Biotechnology (CeBiTec) at Bielefeld University in Germany maintains the required hardware infrastructure including large scale storage and compute facilities. Altogether, the current BRF storing capacities add up to approx. 60 Terabytes disk storage and approx. 240 Terabytes tape storage. These capacities were substantially extended by MGE funding and extensively used within the MGE network to store the experimental data sets (DNA sequences, microarray images, and resulting data files) in a centralized and systematic way (LIMS functionality). A high-performance compute cluster with more than 500 CPUs providing an overall capacity of approx. 1,4 Teraflops is also available for large scale computations such as EST analyses or whole genome annotations. In addition to more than 200 publicly available bioinformatics tools that are locally installed within the Bielefeld platform, several *in house* developed software packages (see Figure 1) are provided for genome and transcriptome analyses. All tools were developed under an open source license and are accessible to MGE users via user friendly web based front-ends (see also http://www.cebitec.uni-bielefeld.de/brf/software/brfsoftware.html).

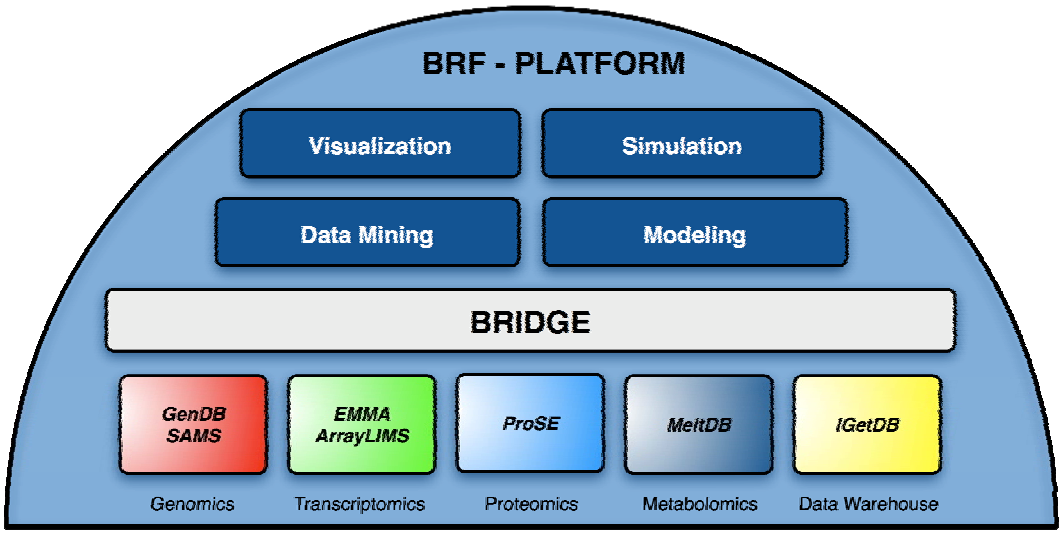


Figure 1: The BRF platform currently comprises software for high-throughput sequence analysis (SAMS), automated and manual genome annotation (GenDB), and transcriptome data analysis (ArrayLIMS, and EMMA). New software applications for proteome (ProSE) and metabolome (MeltDB) data analysis as well as a Data Warehouse (IgetDB) for efficient queries on large data sets are also being developed. All components are linked via the BRIDGE integration layer providing an interface for further data mining, modeling, simulation, and visualization.

In order to facilitate access to the bioinformatics software platform for roughly 500 MGE users, we developed the MGE Bioinformatics portal web site (http://www.cebitec.uni-bielefeld.de/groups/brf/software/mge-portal) serving as a central entry point to all MGE data sets and software tools. All data sets are only accessible with individual login and password authentication and all data transfer is secured via https. Furthermore, the role-based access to individual data sets is controlled via a general project management system (GPMS). A screenshot of the MGE bioinformatics portal is shown in Figure 2.

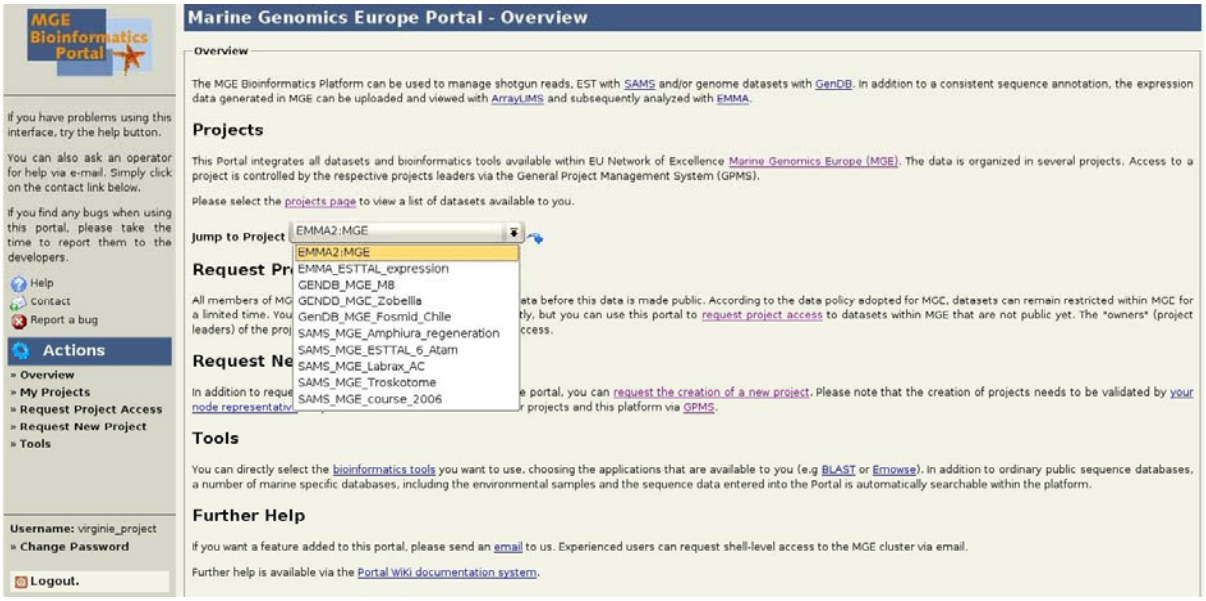


Figure 2: Screenshot of the central MGE bioinformatics portal. After login, users can request the setup of new projects or request access to already existing projects. By simply selecting a project from the drop down list existing projects can be accessed directly.

During the whole duration of the project, all data sets generated within the various nodes were systematically stored and analyzed in a centralized way at the Bielefeld bioinformatics platform. Finally, a total of 62 projects (3 ArrayLIMS, 3 EMMA, 12 GenDB, 44 SAMS projects) is now available via the MGE bioinformatics portal as listed in Table 1.

Node	**# SAMS projects**	**# GenDB projects**	**# EMMA projects**
Algal	28	3	3
EDD	5	-	-
Fish & Shellfish	11	-	-
Microbial	-	9	-

Table 1: List of all MGE projects that are currently maintained by the Bielefeld bioinformatics platform. While the majority of EST sequencing projects was performed by groups of the algal and fish & shellfish node, most of the whole genome projects were conducted by members of the microbial node.

As already mentioned above, within this project we continuously developed several *in house* software applications with a strong focus on DNA sequence analysis, genome annotation, and transcriptomics and adapted these bioinformatic tools to the specific needs of the MGE users. A brief overview of the main software releases and the corresponding major new features is presented in Table 2.

Software	**Release**	**Main features**
GenDB	2.2	Modules for genomic plots, genome comparison, synteny. Management of project members via GPMS web interface.
	2.4	Support for more types of regions, new tools such as Gismo, TransportDB, RFAM, TransTerm.
SAMS	1.2	Support for shotgun reads and EST sequence analysis.
	1.4	Improved import, new filtering option in sequence data viewer.
	1.6	New feature for joining two or more existing sequence libraries, option to filter the sequence data according to their E-value, option to export the chosen sequence data into (multiple) FASTA files, or tab-separated files.

	1.8	Job and annotation statistics, SteN - Statistical electronic Northern blot, Pie-chart views of functional classification distributions.
	2.0	TC update feature, new tools for functional analysis such as HAMAP and PRIAM.
EMMA	2.2	Clustering viewer and LIMMA pipeline.
	2.4	MAGE importer/exporter, SAM and Cyber-T as new statistical test methods.
	2.6	Support for Affymetrix arrays.
	2.8	Support for NimbleGen arrays.

Table 2: List of major software releases and major new features that were made available to the MGE community. Only some major features are listed here, further details are on the homepage of each software application.

All EST data sets were imported, clustered, assembled and automatically annotated using the novel Sequence Analysis and Management System (SAMS).

In the field of whole genome analysis, the development of the GenDB system was successfully continued in order to provide a most efficient, high-quality automated and manual genome annotation. After an initial automatic and manual annotation, all genomes under investigation within this project were maintained and function assignments of single genes were continuously refined with respect to new experimental results, e.g. from microarray analyses.

For the systematic acquisition of all microarray experiments we developed a new Laboratory Information Management System (ArrayLIMS) that stores all microarray data sets in a MIAME standard compliant way and facilitates data submission to the public EBI ArrayExpress database which is a prerequisite for publication of results from microarray experiments. While all raw data sets and experimental descriptions are stored within ArrayLIMS the subsequent data analysis and evaluation is performed with the EMMA 2 software. Therefore, this system includes several automated pipelines for data normalization, statistical tests, clustering, and other advanced analysis tools.

Combining the various MGE data sets within the platform, we implemented a number of new analysis methods, such as viewing gene expression data projected onto a KEGG metabolic pathway, which is cross-linked with all available sequence annotations for that contained enzyme. Using the BRIDGE architecture (Goesmann et al, Nucleic Acids Res 33, 2005), a domain spanning query software, direct links between different data sets from the various specialized components (SAMS, GenDB, EMMA) could be established for interactive navigation and enhanced data analysis approaches (see also Figure 1).

Last but not least, the MGE bioinformatics platform also provided detailed documentation for all tools and offered several practical courses for training users in working with each system.

II- 8 INTEGRATING THE GENDER DIMENSION (WP24)

Project leader: Olsen, Jeanine; University of Groningen, Partner 28

GAP Committee:

- Catherine Boyen, CNRS Roscoff, partner 2
- Michèle Barbier, CNRS Roscoff, partner 2
- Melody Clark, BAS, NERC, partner 12
- Richard Reinhardt, MPIMG Berlin, partner 6

The MGE consortium brought together some 450 persons with a gender balance of 130 women (30%). Although there was the usual turnover in PhD students and post-docs over the course of the 4.5 years, the breakdown by rank was roughly 50 PhD students and 81 PhD holders. Among these, 19 (23%) held senior positions as research group leaders, directors or professors. It was also a big boost to have Dr. Catherine Boyen as overall coordinator of the network. Our target was also the other 220 men and so, from the beginning, we made sure that GAP activities were incorporated into the mainstream programming of MGE General Assemblies, the Scientific Steering Committee meetings and the larger workshops and summer courses.

The **GAP team** included: Groningen-28/Jeanine Olsen (Algal Node and GAP-WP coordinator); Roscoff-2/Michèle Barbier (Web master and administrative assistant for the GAP) and Catherine Boyen (MGE coordinator). Assisting the three of us were MPI-Berlin-6/Richard Reinhardt (Tech Platforms), BAS-Cambridge-12/ Melody Clark (Fish & Shellfish Node), CNRS-Paris-2/Chris Bowler (Algal Node), IOL-Haifa-25/Esther Lubzens (Fish & Shellfish Node) and KMRS-Kristineberg-38/Ulrike Hjelm (Education).

Objectives

The under-representation of women in the natural sciences is well documented. However, the causes of the problem fall into several categories related to underlying cultural norms, academic traditions and practices—some general and some country-specific. The MGE Gender Action Plan (GAP) had four objectives:

1. Provision of **Fellowships for Women** for research stays of 1-3 months. The goal was to ensure that gifted young women scientists were actively promoted.

2. Provision of **Work/Life Balance Travel Grants (=crèche grants).** The goal was to remove short-term, participation obstacles for women (and men) with small children.

3. Provision of a two-tiered **Mentoring Programme** for junior and senior women. The goal was to strengthen understanding, extend networking both laterally and vertically, and to develop effective strategies for the advancement of women, i.e., bottom up and top down.

4. Provision of **Gender Education for Men *and* Women Scientists** of all ranks. The immediate goals of this part of the GAP were to raise awareness at the individual level within the fabric of one's own country/cultural institutions as well as in the larger cross-cultural context which would involve discussion groups/workshops at the larger meetings and a set of web-page learning modules. The long-range goal was to foster permanent structural change, which by itself, made this component of the action plan the most challenging but also the most important in the long run.

5. This objective was added in Year 2 and consisted of two €5,000 awards for a **Jr. and Sr. Outstanding Woman Scientist** in marine genomics. The award was advertised in *Nature* and on various websites.

Results/Deliverables

The **Fellowships for Women** program was a big success with 26 total awards spread over five rounds. In addition to opportunities of working in another lab group for 1-3 months, applicants were required to motivate their applications both scientifically and in terms of how they saw themselves as women scientists. The applications not only provided insights for the GAP selection committee but allowed us to provide feedback to applicants about why they were not selected (e.g., poorly constructed cv or failure to complete the application).

Everyone thought that the **crèche grants** were a great idea. Ironically, they were never actually used. In interviews with would-be users, they reported that making arrangements at home was ultimately easier and cheaper than bringing the child to the meeting. Our general conclusion, is that this objective was not really necessary.

The **Mentoring Program** was originally conceived as a formal structure in which mentors and mentees would be matched following a questionnaire that was circulated to all MGE members. Information was provided on the MGE GAP webpage about what mentoring was and how it differed from regular supervision. Although most people returned the questionnaire, many of our initial target group (PhDs and Post-docs) tended to think of the mentoring programme as "correctional" and that it would take "too much time. We thus spent time talking to persons on an individual basis. It is also true that making a good connection takes time and a bit of luck. Not knowing exactly how to proceed, we decided to approach the mentoring issue through discussions and workshops, which ultimately worked out quite well. This was, of course, linked to our last objective of **Gender Education**. We had originally envisaged a series of monthly modules on the website covering a range of topics. This was not formally realized due to lack of time and, in hindsight, probably would not have been very successful for the same reason—people not taking the time to read them. Informed presentations and discussions ultimately won the day in terms of mentoring and education.

At the end of the second year we decided to offer two awards of €5,000 each for **Outstanding Research by Women in Marine Biological Sciences with an Emphasis on Genomic Processes**. Criteria for selection included not only excellence in research but also excellence in teaching, service and outreach to young women in science. The senior category was open to nominees with ten or more years of experience following their PhDs; and the junior award to researchers with <10 years of experience. From a field of 18 top candidates, the committee chose Dr. Melody S. Clark, British Antarctic Survey, Cambridge (Sr. award) and Dr. Kristin Tessmar-Raible, EMBL-Heidelberg (Jr. award).

Conclusions

As WP coordinator I have had the pleasure to see the GAP mature and can look back on the past 4.5 years with a sense of real satisfaction. Back in 2004, the whole idea of a GAP was met with considerable skepticism by quite a few participants—both men and women—and especially younger researchers (PhD and post-doc level). Although it was not politically correct to have a negative opinion, more than a few people asked, "How goes it with the gender police?" or "Do we really need this?" Fortunately, the GAP team paid no mind to these comments and within the first year, we had pretty well convinced everyone that a GAP was genuinely necessary. Catherine Boyen made sure that there was always time for a GAP presentation and little by little, MGE members actually looked forward to the latest installment.

We started the GAP by providing factual information through comparative country-statistics (e.g. *SheFigures*). Most people were genuinely shocked at the differences in women's participation, promotional chances and differential salaries—especially when it turned out to be their own country. But we also kept the tone positive and proactive with the success stories. We also spent time in discussion groups involving junior and senior researchers in which experiences were shared. For example, women are often reluctant to put themselves forward for a higher position until they are absolutely sure that they are ready, whereas most men will jump at the chance if they have only one of the ten requirements. Encouraging women to take chances, through our own examples proved very successful. Another successful strategy was the use of smaller break-out groups in a series called, "Talk with the Profs", where young men and women PhDs/post-docs had the opportunity to obtain advice about their particular areas of expertise, as well as asking about individual experiences, set-backs and strategies. These sessions were informal and lasted an entire morning—sometimes sitting on the floor in an empty lecture hall or taking a stroll outside. We were very pleased with the turn-out for this series, which involved an equal number of senior and junior, men and women. Later, a workshop was held in Crete, dealing with more specialized topics, notably the "Two-Body Problem" (Nirvana solutions for scientific couples), "getting your first post-doc" (the perfect cv, cover letter and approach to the interview), "career trajectories"(How do I ultimately get to the rank of full professor?) and "childcare" (having babies and effects on career development). By the end of the network, we were not only receiving e-mails from young scientists asking about specific problems and ideas, but several of the more senior post-docs were, in turn, helping more junior researchers. Now that's education and mentoring! Even among the recalcitrant men (Yes, there were a few diehards), there was acknowledgement that their initial stereotypic view of "women in science" issues had softened.

So what worked and what didn't, and what would I do differently next time? The results of the GAP's five objectives speak for themselves (for detailed information with names, award levels and other administrative issues, please see previous annual reports). In the beginning, we tried to set-up too rigid of a program that was basically unrealistic given the very limited time that everyone had. The fact is, special workpackages like GAPs, tend to fall by the wayside as other, more urgent priorities wiggle their way to the forefront. Who would write the modules, who would read the modules, who would assign mentors and mentees? These are not really command performance items. Experience has taught us that most of these things can actually be self-organizing IF the basic foundation and atmosphere are laid in a neutral manner We actually stumbled on this a bit by chance. Because of the emotional nature of the subject, it took time to gain trust and create the necessary neutral but serious atmosphere in which people felt comfortable discussing the various issues.

There was also an over estimation of what people's basic knowledge actually was regarding gender issues in science, cultural perspectives (one-size does not fit all) and to what extent their own ideas reflected popular mythology as opposed to actual fact. Most people had some pretty strong opinions, most of which were not based on either fact or actual experiences. Nurturing change in this regard was perhaps the single most important accomplishment of the GAP.

The MGE GAP has been used as a model for other NoEs. EDIT (European Distributed Institute for Taxonomy 2006-2010) is one of these. As a member of the Scientific Advice Committee for EDIT I have the opportunity to see how their GAP is progressing. So far, the results are most heartening.

One final word. The decision to require GAPs in FP6 was a bold move. The decision to discontinue them on a mandatory basis in FP7 is unfortunate. Gender issues in science are ongoing with the need to educate each new generation of men and women. Using the collective best practices gained from the many FP6 NoEs, it would be easy to provide workable guidelines for future calls.

Finally, on a personal level, I have enjoyed running this workpackage and have learned a great deal about the practical implementation of gender awareness through this experience and in greater involvement with the European Platform of Women Scientists.

II- 9 EXPLORATORY WORKSHOPS (WP26)

Project leader: Catherine Boyen, CNRS, Fr, Partner 2

The organisation of exploratory workshops for promoting cross-node interaction was initiated during the 2nd year of MGE on the strong recommendation of external evaluators of the 1st year activity report with the purpose of fostering bridges between the various node projects of the network. These workshops were aimed at favouring integration within researchers from the four nodes, through the arrangement of meetings where all interested partners were invited to present their work and data related to the topics. These events have also facilitated discussions and meetings of people from various nodes or countries. The key purposes were to initiate and/or facilitate potential collaborations, to identify innovative scientific topics, to promote the preparation of joint projects in marine genomics analysis and to take full advantage of the tools generated by the consortium.

These workshops were open in priority to all MGE members but other scientists were allowed to join. Students and post-docs were especially encouraged to attend these events.

A maximum MGE funding for each workshop has been set: 10 000 €. This amount was aimed at covering accommodation and meals fees for participants as well as the hosting cost. Travel expenses were to be supported by attendees.

For all workshops, a call describing the topics of the workshop, as well as a preliminary programme, was published several months before the event on MGE website and all interested people were able to register and send abstracts.

In total, five calls were open and twelve Exploratory Workshops were organized.

"**Stress responses to environmental variations**"
Organizers: Alex Rogers and Melody Clark (partner 12) in **Cambridge in September 2005.**
The workshop brought together 31 participants from 15 different partners belonging to 3 nodes: EDD, F&S and Algal.

"**Population Diversity and Species Identity**".
Organizer: Paolo Sordino (Partner 22) in the **Stazione Zoologica Anton Dohrn, Napoli, on 4 and 5 November 2005**
The workshop brought together 15 participants from 7 different partners, belonging to 2 different nodes, EDD and algal as well as three external scientists (Portugal, Greece and USA).

"**New Year Genboree in Naples**: **Genome Annotation strategies with special reference to *Strongylocentrotus purpuratus***"

Organizers: Mike Thordndyke (Partner 38) and Ina Arnone (partner 22) in **Stazione Zoologica Anton Dohrn, Napoli 16 – 20 January 2006.**

The workshop brought together 40 participants from 8 different partners belonging to 3 nodes EDD, Microbe and Algal as well as ten external scientists (1 Greece, 3 UK, 4 USA and 1 Italy).

"Evolving gene Networks"

Organizers: Pedro Martinez (University of Barcelona, partner 35) and Xavier Bailly (CNRS, Station Biologique, Partner 2) in **Station Biologique, Roscoff, France, 24-26 April 2006.**

The meeting gathered 30 participants from 8 different partners belonging mainly to the EDD node.

"Marine Genomics meets Marine Diversity"

Organizers: Frank Oliver Glöckner, Rudolf Amann & Johanna Weisnigk from **the Max Plank Institute in Marine Microbiology in Bremen, Partner 4 , in Bremen, Germany from 8 to 10 June 2006**.

The meeting gathered 60 participants from 13 countries, presented state of the art results on marine microbial diversity and genome research. Participants include representatives of the two NoEs "Marine genomics Europe" and MARBEF.

"Stem cells in Marine Organisms"

Organizers: Valeria Matranga from CNR Palermo, Partner 23 and Buki Rinkevitch, National Institute for Oceanography, Israel, Partner 25. The workshop took place **in Palermo, Italy 27-28 November 2006**

The workshop brought together 44 participants attended the workshop, including 30 MGE members (EDD and Fish and shellfish node) and 14 non MGE persons.

"Transcriptomics approaches for the analysis of marine Systems"

Contrarily to all the previous exploratory workshops, this one resulted from an initiative of the Scientific Steering Committee addressing a recommendation of the evaluating panel in June 2006. An organizing committee was set up comprising representatives from each node who had been already implicated in transcriptomic studies:

- EDD node: Florian Raible from EMBL (Partner 11) & Christian Gache from Villefranche (Partner 2)
- Algal node: François-Yves Bouget from Banyuls (Partner 2)
- Microbial node: Wolfgang Hess from Freiburg (Partner 45)
- Fish and Shellfish node: Josep Planas from Barcelona (Partner 35)

The workshop took place **Heidelberg, Germany in May 2007**

The meeting gathered 53 participants from11 different MGE partners and **belonging to all nodes**.

"High throughput sequencing applied to marine organisms, a European perspective"

Organizers: Mike Thorndyke (Kristineberg Marine Station, N°38) and Mark Cock (CNRS, Station Biologique, N°2) **in Station Biologique, Roscoff, France, 15-17 April 2007.**

The meeting gathered 43 participants from 16 different countries (including USA, China and India) from 13 different partners related to the four nodes. The participants included 3 invited speakers and 11 non MGE people.

"**Application of genomics tools and techniques to the understanding of Ocean acidification"**

Organizers: Jon Havenhand and Mike Thorndyke from the Kristineberg Marine Research Station (Partner 38) in **Kristineberg, Sweden, September 16th-18th 2007.**

The meeting gathered 25 participants from 9 different countries including USA, japan and Czech Republic and from 7 different partners related to the nodes EDD, Fish & Shellfish and Algal.

"Marine Genomics: An Ocean of Techniques"

The Organizing Committee: Michèle Barbier (SBR-CNRS, Partner 2) , Marco Borra (SZN, Partner 22), Marc Heijde (ENS-CNRS, Partner 2), Ulrika Hjelm, chair (KMRS, Partner 38), Julia Morales (SBR-CNRS, Partner 2), Ángel E. Pérez Diz (Univ of Wales Swansea, Partner 19) and Patricia Wecker, (MPI-Bremen, Partner 4). **Orthodox Academy of Crete, Greece October 8-11**

The workshop gathered 70 participants (attendees plus speakers) from 12 different countries, from 24 different Partners related to the four nodes.

"**Marine Environmental Genomics and Gene Expression from Genomics to Ecosystem Understanding**"

Organizers: Klaus Valentin (AWI Partner 20) and Uwe John (AWI Partner 20), with assistance from Frank Oliver Glöckner (MPI Bremen, Partner 4) and Allan Cembella (AWI, Partner 20). **Alfred Wegener Institute Bremerhaven, Germany, June 4-7, 2008**

The workshop gathered 20 participants from 5 different countries, 7 different partners related to 3 nodes.

III Jointly Executed Reports

III-1 MICROBIOLY NODE (WP1-3)

Project leader : Name Rudi Amann, MPIMM, Partner 4 for WP 1, Wolfgang Hess, Uni Freiburg, Partner 45 for WP 2,
Project assistant: Johanna Wesnigk, EMPA- sub-contractor via MPI MM, Bremen
Microbial node committee:

- Joël Quérellou, IFREMER, Brest, FR, Partner 3
- Carlos Pedros-Alio, CSIC, Barcelona, ES, Partner 33
- Wolfgang Hess, Freiburg University, DE, Partner 45

Overall Introduction and Objectives

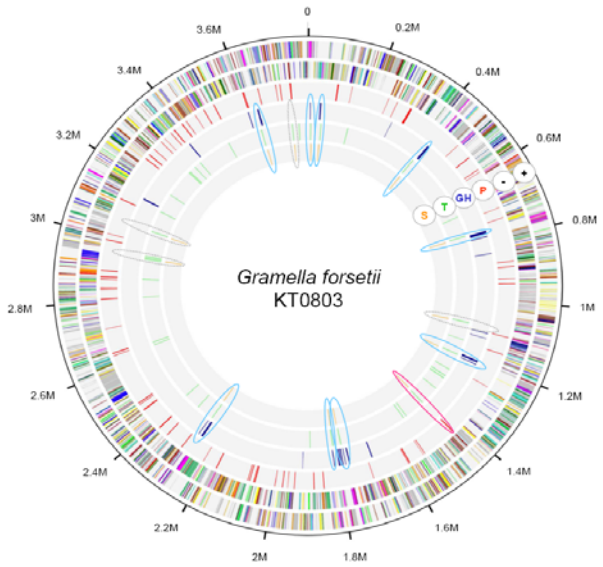
The microbial node within MGE has started out with three work packages, one each in comparative (WP1), functional (WP2) and environmental genomics (WP3). Over the course of the work we merged WP 3 with WP 1 to compare metagenomic with full genome data. Also many of the sub-tasks of the starting phase were phased out or merged into larger ones to promote networking and collaboration. In this report we describe only our major achievements. More detail can be found in the yearly periodic activity reports on MGE covering the microbial node work, from year 1 to 4.
Comparative genomics in the microbial node of MGE has benefitted tremendously from the rapid progress of sequencing technology. Scientists could actually analyze many more genomes of marine heterotrophic bacteria and cyanobacteria than anticipated at the start of the NoE. In the final years of MGE pyrosequencing facilities became available both within MGE, due to the technological platform at MPI Bremen (#6) and through cooperations, with European and US institutions. This opened a new door for innovative experimental setups (for example see Mussmann et al., 2007).
Based on a clear focus on two well defined groups of marine bacteria, Bacteroidetes and Cyanobacteria, mainly of the genus *Synechococcus*, our joint analyses have advanced knowledge on the phylogeny and ecology of these important groups of marine microorganisms which are both involved in global carbon cycling. The cyanobacteria fix carbon dioxide in photosynthesis, and the heterotrophic bacteria mineralize organic carbon.

III.1.1 Comparative and Environmental Genomics focussing on heterotrophic bacteria:

Contractors (in numerical order): Joel Querellou, Julien Briffotaux, Gurvan Michel, Tristran Barbeyron, Didier Flament, Purificacion Lopez-Garcıa, Christian Jeanthon #2; Rudolf Amann, Frank Oliver Glöckner, Mark Mussmann, Marga Schüler (Bauer), Michael Richter #4; Michael Kube, Richard Reinhardt (TP Berlin) #6; Bioinformatics platform Bielefeld #7; Oded Beja, Baruch Karniol #26; Carlos Pedrós-Alió, Ramon Massana, Ramon Rossello-Mora, Beatriz Fernández-Gómez, Thomas Pommier #33; Viggo Marteinsson #42.

Task 1 Comparative Genomics of heterotrophic bacteria

The whole genome analysis of a representative of the marine Bacteroidetes group, *Gramella forsetii* KT0803, a carbon degrading bacterium was performed at the TP Berlin (#6). This genome is the first fully sequenced and annotated of a marine Bacteroidetes, a taxon involved in the degradation of marine snow. The analysis revealed adaptations to the degradation of polymeric organic matter (Bauer et al. 2006).
Comparative analyses of genomes from the ecologically important bacterial group of marine *Bacteroidetes* were initiated, focusing firstly on four flavobacterial strains from two different coastal habitats. Their orthologous gene set was established, paralogous families were delineated and functionally classified, and a size ranking list of paralogous families was compiled in order to compare the “physiological focus” of the strains. In addition, their glyco- and proteolytic potential as predicted from the presence of various families of polysaccharide and protein degradation enzymes in their genomes was compared. Moreover, the question of the validity of a hypothesis on putative polymer transport singularities in marine Bacteroidetes involving the family of TonB-dependent outer membrane receptors was addressed by a comparative analysis of the number of gene tandems encoding SusCD-like proteins and the characterization of their genetic context.
The results obtained so far are being scrutinized in more detail and substantiated with complementary analyses within the next months. The outcome of the comparative work is summarized in a manuscript including members of the Bremen and Barcelona groups as co-authors (#4, #33); the submission is foreseen in the second half of 2008 with the title “Ecological insights of Marine Bacteroidetes genome comparison”.



Full genome sequencing done at the Technological Platform, MPI Berlin (#6) in Oktober 2004

Genome characteristics:

One circular replicon	3,798,864 bp
GC content	36.61 %
ORFs	3,592

Legend: Rings from outside to inside contain genes on plus strand (+), genes on minus strand (-) (both color coded according to COG functional classes), peptidase genes (P), glycoside hydrolase genes (GH), TonB-dependent receptor genes (T), genes encoding SusD-like proteins (S). Genome sites with co-localization of glycoside hydrolase/peptidase genes and a tandem of genes encoding a TonB-dependent receptor family protein and a SusD-like protein are marked with blue/red ovals. Reminiscent of the *Bacteroides thetaiotaomicron* starch utilization system (*sus*) where a TonB-dependent receptor family protein and the SusD protein are involved in starch-binding, the co-localization of analogous genes with genes encoding hydrolytic activities in *G. forsetii* suggests a similar mechanism for biopolymer-binding and degradation.

Final results of this task were:

- Bacteroidetes are specialists for degradation of high molecular weight compounds (HMW) in dissolved and particulate fractions of the marine organic matter pool;
- They play a major role in the marine carbon cycle, especially in remineralisation.
- Publication: Whole genome analysis of the marine Bacteroidetes "*Gramella forsetii*" reveals adaptations to degradation of polymeric organic matter. Bauer, M., M. Kube, H. Teeling, M. Richter, T. Lombardot, E. Allers, C.A. Würdemann, C. Quast, H. Kuhl, F. Knaust, D. Woebken, K. Bischof, M. Mussmann, J.V. Choudhuri, F. Meyer, R. Reinhardt, R.I. Amann, and F.O. Glöckner. Environ. Microbiol. 2006, 8:2201-2213; including the MGE partners #4, 6, and 7.

Task 2 Genome analysis of the proteorhodopsin-containing marine Bacteroidetes *Polaribacter sp. MED152*

Analysis of marine cyanobacteria and proteobacteria genomes has provided a profound understanding of the life strategies of these organisms and their ecotype differentiation and metabolisms. A comparable analysis of the Bacteroidetes, the third major bacterioplankton group has been promoted by genome analysis of *Polaribacter* sp. strain MED152 (Gonzalez et al. 2008). On the one hand, MED152 contains a substantial number of genes for attachment to surfaces or particles, gliding motility, and polymer degradation. This agrees with the currently assumed life strategy of marine Bacteroidetes. Furthermore, it contains the proteorhodopsin gene together with a remarkable suite of genes to sense and respond to light, which may provide a survival advantage in the nutrient-poor sun-lit ocean surface when in search of fresh particles to colonize. An increase in CO_2 fixation in the light suggests that the limited central metabolism is complemented by anaplerotic inorganic carbon fixation. Taken together, the findings suggest that light-stimulated anaplerotic inorganic carbon fixation could be a means allowing proteorhodopsin-containing flavobacteria to efficiently use organic matter for biosynthesis.
This is mediated by a unique combination of membrane transporters and carboxylases. This suggests a dual life strategy that, if confirmed experimentally, would be notably different from what is known of the two other main bacterial groups (the autotrophic cyanobacteria and the heterotrophic proteobacteria) in the surface oceans. The *Polaribacter* genome provides insights into the physiological capabilities of proteorhodopsin-containing bacteria. The genome will serve as a model to study the cellular and molecular processes in bacteria that express proteorhodopsin, their adaptation to the oceanic environment, and their role in carbon-cycling.

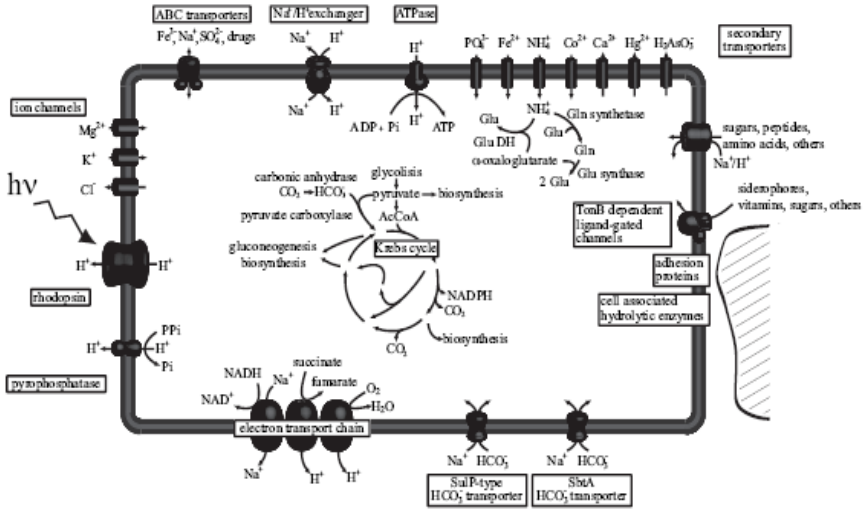


Fig. 2. Diagram of a MED152 cell. Processes that generate a H^+ or Na^+ gradient are shown *Lower* and *Left*. Transporters are indicated *Upper* and *Right*, except bicarbonate transporters that are shown *Lower*; not all types of transporters are shown.

Graph from: González, José M., Beatriz Fernández-Gómez, Antoni Fernàndez-Guerra, Laura Gómez-Consarnau, Olga Sánchez, Montserrat Coll-Lladó, Javier del Campo, Lorena Escudero, Raquel Rodríguez-Martínez, Laura Alonso-Sáez, Mikel Latasa, Ian Paulsen, Olga Nedashkovskaya, Itziar Lekunberri, Jarone Pinhassi, and Carlos Pedrós-Alió (2008). Genome analysis of the proteorhodopsin-containing marine bacterium Polaribacter sp. MED152 (Flavobacteria). **Proceedings of the National Academy of Sciences of the USA**, 105: 8724-8729, incl. MGE partner #33

In another study related to MGE whole-genome sequences of three bacteria belonging to the class Flavobacteria, phylum Bacteroidetes (*Dokdonia sp. strain MED134, Polaribacter sp. strain MED152* and *Leeuwenhoekiella blandensis strain MED217T*) were examined. These bacteria were isolated from surface water from the Mediterranean Sea and were successfully cultured. It was shown, from an analysis of the complete genomes of three marine Flavobacteria, that cultivated bacteria in the phylum Bacteroidetes contain proteorhodopsin. Moreover, growth experiments in both natural and artificial seawater establish that exposure to light results in a marked increase in the cell yield of one such bacterium (Dokdonia sp. strain MED134) when compared with cells grown in darkness. An interesting consequence is that, for a given amount of DOM, bacterial assemblages dominated by species expressing genes encoding PR would produce more biomass than assemblages of bacteria lacking PR, making it available to higher trophic levels in the marine microbial food web (Gómez-Consarnau, L., J.M. González, M. Coll-Lladó, P. Gordon, T. Pascher, R. Neutze, C. Pedrós-Alió & J. Pinhassi. (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. **Nature**, 445, 210-213. with MGE partner #33).

Task 3 Environmental Genomics

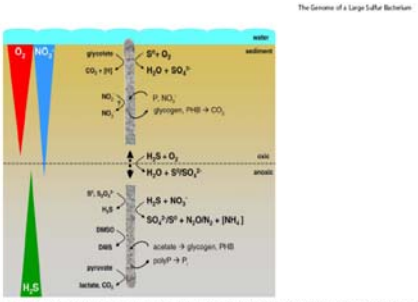
This task started out with several parallel activities, covering habitats of low and high diversity, from hydrothermal vents to Wadden Sea sediments. Six publications were written involving the five groups working in WP 3 on metagenomics; and a further nine from the combined WP 1 and WP 3 results. The groups were consolidated into one working on genome data from heterotrophic bacteria, one working on extreme environments, and one focussing on photosynthetic micro-organisms, incl. eukaryotes. Thus we could compare metagenomic data to well annotated full genome data.

One comparison resulted in the publication "Fosmids of novel marine planctomycetes from the Namibian and Oregon coast upwelling systems and their cross-comparison with planctomycete genomes" which also contains a link to work done within functional genomics on *Rhodopirellula baltica* (not further described here) for which recent genome and microarray data could be utilized (Woebken et al. ISME Journal 2007, 1; MGE partner #4).

Task 4 Insights into the Genome of Large Sulfur Bacteria Revealed by Analysis of Single Filaments

Marine sediments are frequently covered by mats of the filamentous *Beggiatoa* and other large nitrate-storing bacteria that oxidize hydrogen sulfide using either oxygen or nitrate, which they store in intracellular vacuoles. Despite their conspicuous metabolic properties and their biogeochemical importance, little is known about their genetic repertoire because of the lack of pure cultures. A unique approach to access the genome of single filaments of *Beggiatoa* could be implemented by combining whole genome amplification, pyrosequencing, and

optical genome mapping (Mussmann et al. 2007). Comparative genomics indicates substantial horizontal gene transfer of storage, metabolic, and gliding capabilities between Beggiatoa and cyanobacteria. These capabilities enable Beggiatoa to overcome non-overlapping availabilities of electron donors and acceptors while gliding between oxic and sulfidic zones. The first look into the genome of these filamentous sulfur-oxidizing bacteria substantially deepened the understanding of their evolution and their contribution to sulfur and nitrogen cycling in marine sediments.



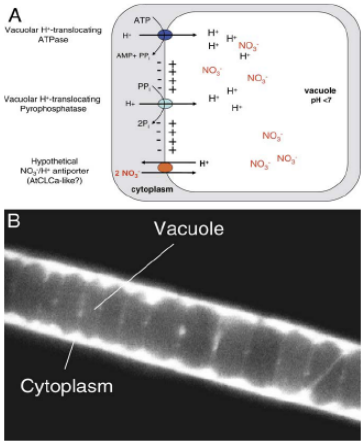


Figure 4. Energetic Vacuolar Concentration of Nitrate by *Beggiatoa* Sp.
(A) Hypothetical model of nitrate accumulation in the vacuole of *Beggiatoa*.
(B) *Beggiatoa* filament stained with the cationic, lipophilic dye rhodamine 123. Rhodamine 123 accumulated in the cytoplasm but was excluded from vacuoles, indicating the presence of an electric potential (inside positive) over this membrane.
doi:10.1371/journal.pbio.0050230.g004

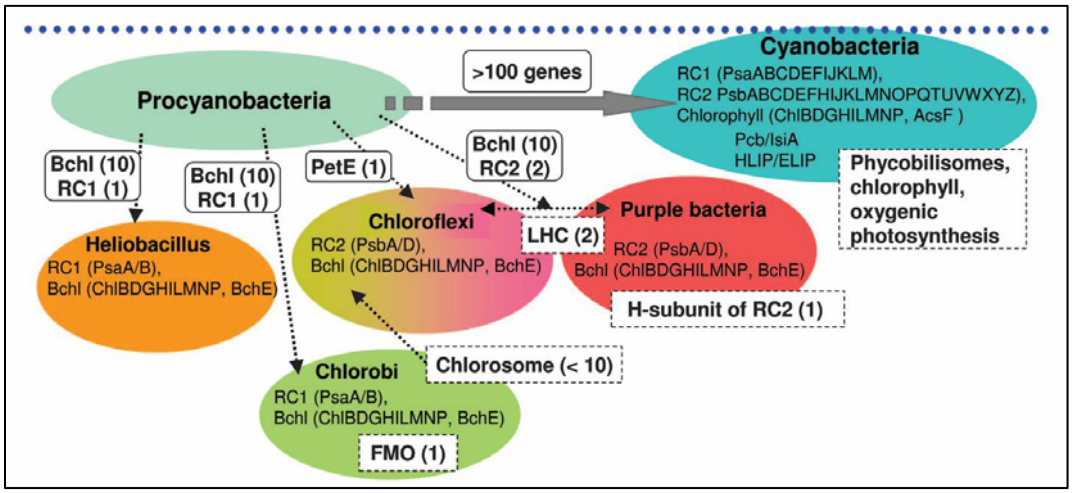
Both graphs come from: Mussmann, M., F. Z. Hu, M. Richter, D. de Beer, A. Preisler, B. B. Jorgensen, M. Huntemann, F. O. Glöckner, R. Amann, W. J. H. Koopman, R. S. Lasken, B. Janto, J. Hogg, P. Stoodley, R. Boissy, and G. D. Ehrlich (2007) Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. **PLoS Biology** 5: 1923-1937., incl. MGE partner #4.

Overall, 23 publications were written involving the five groups working in WP 3 on metagenomics and those working in comparative genomics (WP1 and WP3). Those not quoted are listed below.

III.1.2 Comparative and Functional Genomics focussing on autotrophic bacteria

Contractors (in numerical order):, Frédéric Partensky, Alexis Dufresne, Julia Holtzendorff #2; Richard Reinhardt, Heiner Kuhl #6; Thomas Börner #8; David J. Scanlan, Martin Ostrowski #17; Anton Post #26; Wolfgang Hess, Ilka Axmann #45.

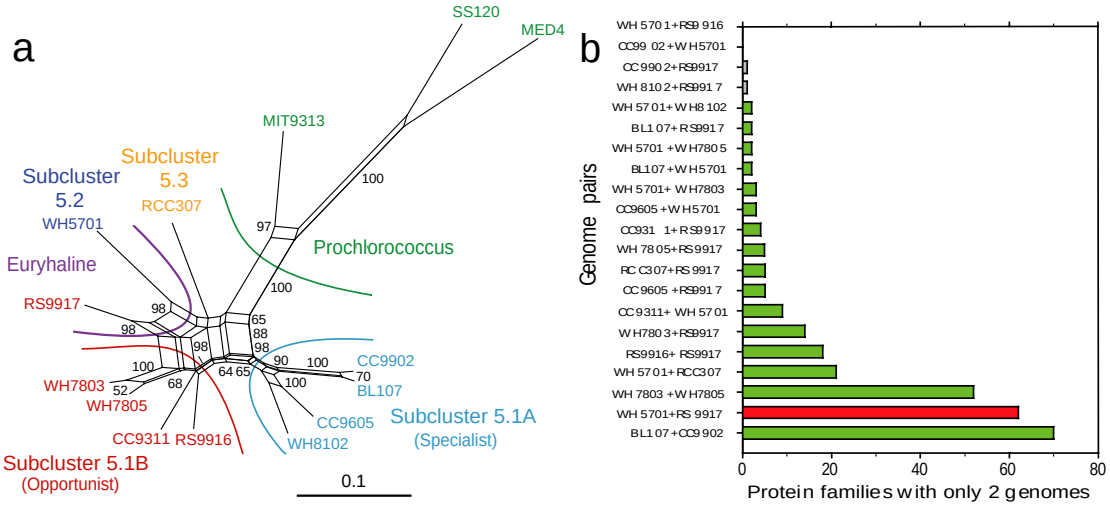
Bioinformatics tools were developed for extensive comparative genomics studies, in particular to better understand the evolution of genes and functions within this key phylogenetic group. This includes a public web site with manual annotation of gene clusters in 14 marine cyanobacteria (http://www.sb-roscoff.fr/Phyto/cyanorak/). Sequence analysis of the first four published total genome sequences of closely related marine cyanobacteria showed an unprecedented number of differences in their genome size, the number of annotated protein-coding genes and GC content (Dufresne et al., 2005). The recent availability of 11 marine *Synechococcus* and 12 *Prochlorococcus* genome sequences enable the use of mathematical modelling. Basically three parameters were determined (1) size of the core genome; (2) size of the dispensable genome; and (3) number of genes constituting the pan genome (Hess et al., unpublished data).
Work was performed on the identification of niche-defining protein-coding genes based on the comparative analysis of marine cyanobacteria genomes. Findings included (i) the contribution of genomic islands to the accessory genome and their role in lateral gene transfer; (ii) a better organismal phylogeny of the *Synechococcus* genus; and (iii) identification of potential niche-defining genes, including phycobilisome genes, shown to be laterally transferred between lineages; regulators, more numerous in coastal than offshore-adapted strains; and nutrient uptake and regulation genes, which showed high variability between strains.
The evolution of the circadian clock genes was also studied in marine cyanobacteria, revealing that deletion of the *kaiA* gene in *Prochlorococcus* results from genome streamlining. The effects of this deletion on circadian rhythms of gene expression and cell cycle was compared between a *kaiA*-lacking *Prochlorococcus* strain and a *kaiA*-containing *Synechococcus* strain (Holtzendorff et al., 2008).
MGE scientists contributed key genome data on *Prochlorococcus* and *Synechococcus* to a paper on the origin of photosynthesis (Mulkidjanian et al., 2006). Among 84 genes shared by cyanobacteria and plants, 35 families of yet uncharacterized proteins were identified which could be involved in photosynthesis. These proteins will soon be subjected to structural biology analyses in order to better predict their potential function. By comparing the cyanobacterial genomes and genomes of other non-oxygenic prokaryotes it is proposed that the first phototrophs were anaerobic ancestors of 'pro'cyanobacteria that conducted anoxygenic photosynthesis using a photosystem I-like reaction center (see graph below).



Distribution of photosynthetic genes in different lineages of prokaryotic phototrophs and the directions of proposed lateral gene transfer. Graph from Mulkidjanian, A. Y., Koonin, E. V., Makarova, K. S., Mekhedov, S. L., Sorokin,A., Wolf, Y.I., Dufresne, A., Partensky, F., Burd, H., Kaznadzey, D.,Haselkorn, R. & Galperin, M. Y.(2006) The cyanobacterial genome core and the origin of photosynthesis. **Proceedings of the National Academy of Sciences of the USA** 103: 13126-13131., incl. MGE partner #2.

A recent MGE paper (Dufresne et al., 2008) brought light into the genetic basis of niche partitioning. It concluded that while members of a given marine *Synechococcus* lineage may have the same broad geographical

distribution, local niche occupancy is facilitated by lateral gene transfers, a process in which genomic islands play a key role as a repository for transferred genes. One of these gene islands contains most genes involved in the phycobilisome assembly and hence the pigmentation of *Synechococcus* strains (Six et al., 2007; Dufresne et al., 2008). Also, it was shown that two *Synechococcus* strains (RS9917 and WH5701) have an unusually high number of genes in common given their large phylogenetic distance, based on 16S rRNA. The set of common genes (mostly uncharacterized), probably includes many genes involved in the adaptation to variable salinity, since these strains are euryhaline. This work highlights the need for developing picocyanobacterial systematics based on genome-derived parameters combined with ecological and physiological data.



Relationships between the picocyanobacterial genomes based on shared accessory gene content. Graph from Dufresne A., Ostrowski M., Scanlan D.J., Garczarek L., Mazard S., Palenik B., Paulsen I., Tandeau de Marsac N., Wincker P., Dossat C., Ferriera S., Johnson J., Post A.F., Hess W.R. and Partensky F. (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. **Genome Biology** 9: R90 incl. MGE partners #2, 17, 26, 45.

Overall, 15 publications were written involving the three groups working in WP 1/2 on cyanobacteria and those working on functional genomics of heterotrophic bacteria. Those not fully quoted are listed below.

Impact on the research sector

From genes to ecosystems: the ocean's new frontier – genome enabled biogeochemistry

Preliminary ecosystem model simulations predict large changes in regional productivity and marine community structure (Boyd and Doney 2002). However, the impacts on ocean biogeochemistry and carbon uptake, not to mention fisheries and ecosystem health, are not yet clear. The oceanic genome encapsulates a complete blueprint of potential biological function, and should help to provide answers for such questions as how marine biogeochemistry will respond to global warming (Doney et. al. Front. Ecol. Environ 2004; 2(9): 457–466). Work within the microbial node of the NoE Marine Genomics Europe has contributed important knowledge in the area of bacterial catalysis of biogeochemical processes in the ocean, and on the evolution of life (Carlos Pedrós-Alió, Science 2007). New research teams formed that will continue to use genomic tools for a better understanding of the consequences of global change.

III-2 ALGAL NODE (WP 4, 5, 6, 31, 32, 33, 41)

Project leader : Hervé Moreau, CNRS, Observatoire Océanologique de Banyuls, Fr, Partner2
Algal node Committee:

- Mark Cock, CNRS, Roscoff, FR, Partner 2
- Gareth Pearson, CCMAR, Faro, PT, Partner 21
- Jeanine Olsen, RUG, Groningen, NL, Partner 28
- Chris Bowler, CNRS, Paris, FR, Partner 2
- Klaus Valentin, AWI, Bremerhaven, DE, Partner 10

Algae are a rather ill-defined assemblage that includes all aquatic photosynthetic thallophytes. This definition has no phylogenetic significance and includes groups as distant as uni- and multi-cellular eukaryotes and prokaryotes. After many discussions, photosynthetic prokaryotes have been excluded from the algal node and have been treated in the microbial node, with many interactions between the two nodes on these micro-organisms. Among eukaryotes, algae are present in four of the six main lineages of the eukaryotic crown, including the brown, green and red algae, dinoflagellates and some other smaller algal groups. This very broad phylogenetic diversity represents many different evolutionary steps so that comparisons between algae and with other lineages can provide new insights into the general mechanisms of evolution. In spite of its broad dimension, the term “algae” is still widely used and keeps its operational significance. Micro-algae, i.e. unicellular or loosely colonial organisms, constitute the vast majority of both pelagic (phytoplankton) and benthic oceanic primary producers and are at the base of the marine food chain. They represent one of the most important components of the global marine biosystem. They have a major impact on the environment as a source of oxygen for many other aquatic organisms and, although marine phytoplankton constitutes only about 1 percent of the entire planet’s biomass, it accounts for over 50% of the planet’s photosynthesis. In coastal ecosystems, red, green and brown macroalgae are usually the most obvious components, sometimes forming dense forests that cover an impressive area of the seabed.

However, when the NoE “Marine Genomics Europe” started, one important consequence of this diversity was that, although Europe had a scientific leadership in various aspects of algal biology, European phycologists used a very wide range of species. This fragmentation of resources strongly impaired the development of algal genomics in Europe and the main objective of the “algal node” was to develop an ambitious genomics integrating program, structuring the community around specific research priorities. Furthermore, there is a growing interest across the globe in the application of genomic approaches to marine algal models and several large-scale genomics programs in Japan and US have been initiated or were nearing completion when this project started (March 2004).
The first year of the contract, several calls for proposal have been launched among the different partners of the algal node (JPA1, JPA2 and 3), with the double objective i) to define one or two flag sequencing project(s), and ii) to identify several integrative research programs on a limited number of models. Following this call and the selection of few proposals, a clear evolution happened, from highly dispersed projects for which almost all partners presented their own topic on their own model organism, towards more and more integration.

A unique flagship sequencing project has carried out for the multi-cellular brown alga *Ectocarpus siliculosus* (Cock et al, 2004). This project was submitted to Genoscope with strong support from the node and was accepted . The sequencing of the genome has now been completed and about a quarter of the 18,000 genes have been annotated manually. The annotation process federated a broad range of laboratories from around the world, including many partners of the node. http://www.genoscope.cns.fr/spip/Ectocarpus-siliculosus,740.html.

Additional tools that have been established within the context of this project include a collection of more than 90,000 EST sequences, whole genome tiling array data, large-scale expression microarray data, proteomic techniques and a collection of developmental mutants. A genetic map is also being constructed for this alga.

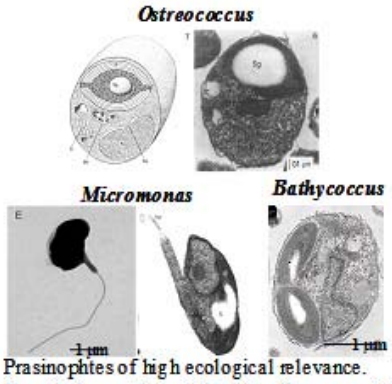
Ectocarpus siliculosus
Credit D. Scornet CNRS

More recently, a new sequencing project of the genome of the red alga *Chondrus crispus* has been accepted by the Genoscope. This project is coordinated by members of the algal node (J. Collen and C. Boyen) and has been strongly supported by the node.

Chondrus crispus
Credit J. Collen

Genomes of some other micro-algae have also been supported by the node (such as *Phaeodactylum tricornutum*, *Emiliana Hux*, *Fragilariopsis cylindrus* and *Ostreococcus tauri*), allowing comparative analysis of some cellular functions among distantly related organisms. A big effort has also been devoted to the development of molecular tools on a restricted number of algal models. These studies led recently to the successful transformation of at least two algal species (*O. tauri* and *P. tricornatum*) and a third species is still under development (*E. siliculosus*). Micro-arrays have been developed for all three of these species. Environmental genomics approaches allowed also the development of new markers to follow the evolution of populations among several macro- and micro-algal species.

Preparative work on some unicellular green algal species (Prasinophyceae) allowed the identification of cultivated strains adapted to different marine environments as low- versus high-light strains.
Genomes of several of these strains have been sequenced, both in Europe and in the US, and teams of the algal node have been strongly involved in these projects. All these studies now allow to start exhaustive studies of the gene content in relation to the ecological niches of these strains.



Prasinophtes of high ecological relevance.
Keyphylogenetic position Small genomes (15-30 Mb)

It has been decided to fund some integrative projects, involving at least two partners and for a limited budget (max 5,000 Euros). The submission and selection processes were simplified to obtain a maximum of flexibility. This programme was very successful and up to 10 projects have been funded.
A total of 51 scientific publications where “Marine genomics Europe” has been acknowledged have been published by partners of the algal node. Many other publications are still in progress, either in preparation or submitted but not yet accepted. Several other outputs as genome browsers/projects, a book chapter on algal genomics (coordinator M. Cock), bioinformatic software and many presentations in various international meetings have been released from the node.

A final scientific meeting was held in May 2008 in Faro, Portugal. During this meeting, most of the results obtained by partners from projects funded by the node were presented (see attached schedule of the meeting).

At the end of “Marine Genomics Europe”, it appears that the work done in the algal node has been successful both for the definition of a restricted number of algal models where efforts have been emphasized, and on integration of partners on few projects developed on these models. The support of the node to big sequencing projects on macro-algae gave a leadership to European teams on these organisms. For unicellular algae, the support of the algal node was also essential, both for sequencing several projects on European platforms (*Bathycoccus* in the Genoscope for example) and to stimulate cooperation with US teams and platforms (JGI). As a conclusion, the Algal Node has been an invaluable tool to maintain and improve competitiveness of European teams in the international competition which developed in this field over the last four years.

III – 3 EVOLUTION DEVELOPMENT DIVERSITY -EDD NODE (WP35-36-42)

Project leader: Mike Thorndyke, Kristineberg, Marine Station, Se, Partner 38
EDD node committee:

- Ina Arnone, SZN, Napoli, IT, Partner 22
- Alex Rogers, NERC-BAS, Cambridge, UK Partner 12
- Pedro Martinez, Barcelona University, ES, Partner 34
- Patrick Lemaire, CNRS Marseille, FR, Partner 2
- Frédérique Viard , CNRS Roscoff, FR, Partner 2
- Francesco Patti, SZN Napoli, IT, Partner 22

The aims and objectives of the EDD node were to integrate "evo-devo" research with modern ecological genetics and biodiversity approaches to ecosystem functional studies. In particular to exploit the study of developmental mechanisms at the molecular level in a way that informs our understanding of plasticity, functional biodiversity and adaptive responses to environmental change. This ahs been extraordinarily successful and has provided us with a suite of molecular biological tools necessary to evaluate quantitatively the biotic and abiotic impacts on the world's ecosystems generated by our industrial and societal activities. It is important to note that these tools are LASTING and many for the consortia established will be durable and continue after the end of MGE, united by common goals and activities as well as new and novel applications.

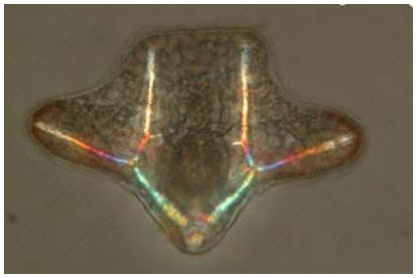
The EDD node was organised as several Work Packages listed below that evolved over the course of the project. The initial topics and areas covered included:

Comparative genomics
Functional Genomics
Environmental Genomics: Molecular Identification of species richness
Environmental Genomics: Adaptive variation
Environmental Genomics: Natural and anthropogenic effects on genetic diversity

These initial activities generated good deal of data and interactions that led to innovative applications later in the programme.

Highlights from the first period:

Generation of EST library from key embryonic stages of the European sea urchin *Paracentrotus lividis.* This united more than 10 laboratories across Europe and over 50 scientists. It also allowed us to contribute significantly to the sea urchin genome project that was headed by Baylor College (Houston, Texas) Human Genome Centre. In fact EDD/MGE members were leaders of several annotation groupings and as a result published high profile papers in Science and Developmental Biology. This established European "Evo-Devo" research at centre stage worldwide. As a development from this, our group is continuing with an agreed European urchin genome project.



Early pluteus larva *P. lividis*

Adult *P. lividis*

Larval to adult transition in the invasive species *Crepidula fornicata*

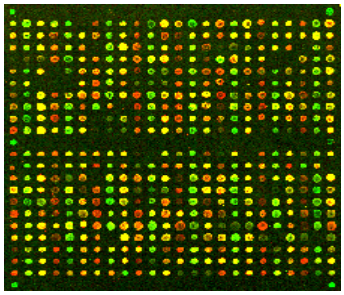
As a result of climate change, the number of alien, invasive species finding new homes in the European coastal zone is increasing dramatically, with for example, the Pacific Oyster *Crassostrea gigas* now being found as far north as Sweden.

On of important areas for focus of the EDD node environmental genomics activities was to employ the power of genomics to identify and track invasive species including the bivalve mollusc *Crepidula fornicata*. One of the key issues here is to be able to identify larvae and larval stages. This is because it is the plank tonic larval stages that represent the greatest threat as they are transported through the warming coastal seas. The adults are easy to identify when they are found on the shore, but by then it is usually too late, they have arrived and have become established. It is the small invading larval stages that we need to monitor. Here we produced new coda libraries from both adults and larval stages and used these to monitor planktonic stages. This will enable predictive models for estimation of spread and invasive potential of this species which is likely to be a threat to native blue mussel (*Mytilis edulis)* populations. This is but one example and there is increasing concern about the danger of other invading alien species. Here genomic screening of total plankton cDNA can provide us with an "early warning system" for "aqua-aliens" that threaten our European coastal seas.

Larval stage *C. fornicate*

Adult *C. fornicate* (Roscoff)

cDNA array for screening



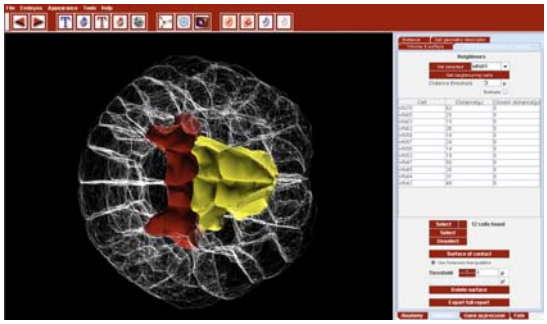
Genomic identity of species

It is becoming increasingly important to be able to identify accurately what a certain species actually is. With the availability of our new genomics tools it is now possible to realise the ambition of detailed analyses of "new species" and "cryptic species" in populations hitherto considered homogeneous. Nowhere has this become more important than in the area of "Evo-Devo" and in particular the community world wide using the models species *Ciona intestinalis.* The Ciona genome was sequenced some years ago now by colleagues in the USA and Japan but researchers in Europe use the "local" Ciona that has extensive populations from the Mediterranean to Scandinavia. Clearly in studies using genomic sequences it is vital to know what species and indeed what sequences are being used. Novel and groundbreaking work by EDD MGE partners have established that
Intensive research with this marine invertebrate is based on the assumption that natural
populations globally belong to a single species. Therefore, understanding the true taxonomic classification may have implications for experimental design and data management. Phylogenies inferred from mitochondrial and nuclear DNA markers accredit the existence of two cryptic species: *C. intestinalis* sp. A, genetically homogeneous, distributed in the Mediterranean, northeast Atlantic and Pacific, and *C. intestinalis* sp. B, geographically structured and encountered in the North Atlantic. Species-level divergence is further entailed by crossbreeding estimates. *C.intestinalis* A and B from allopatric populations cross-fertilize but hybrids remain infertile due to defective gametogenesis. Although anatomy illustrates an overall inter-specific similarity lacking in diagnostic features, we have now provided consistent molecular tools for in-field and in-lab species discrimination. Finding of two cryptic taxa in *C. intestinalis* raises interest in a new tunicate genome as a gateway to studies in speciation and ecological adaptation of Chordates.

Native adult Ciona in Sweden

Ciona larva, Naples

Another way in which MGE EDD activities have ensured community integration and indeed long term sustainability, durability and global presence has been the development of the Aniseed data base for model tunicates that includes a ground breaking facility for working with “virtual embryos (fig right) See also the web site:
http://aniseed-ibdm.univ-mrs.fr/



This has allowed.

- Development of organisms along 24 developmental stages covered (egg to larva)
- Over 4000 anatomical territories described
- 3D representation of embryonic development up to the early gastrula stage.
- 30000 gene models (two genomes)
- WT expression pattern described for 30% of all genes in Ciona (24000 in situ hybridisation patterns)
- >3000 Expression patterns in deregulated conditions (loos of function/gain of function)
- 125 published articles described using a standardised, structured format

Overall EDD has created a successful and sustainable globally recognised community

III -4 FISH AND SHELLFISH NODE (WP12- 14)

Project Leader : Adelino Canário, Centre of Marine Sciences, Faro, Portugal,
F&S node committee : Filip Volckaert, Elhefterius Zouros, François Bonhomme

III.4.1 Genomic toolbox: cDNA libraries, molecular markers, gene maps, genome sequencing and microarrays

A main objective was to increase the volume of sequence data from key fish and shellfish in order to develop a comprehensive toolbox containing neutral and adaptive markers, sequences of characterised functional genes and derived tools such as linkage and physical maps and microarrays. It was also an objective, which revealed very successful, to unite a diverse group of partners around a common focus of resource generation which could be used for most scientific purposes, from genetics to physiology and ecology. Two fish species were chosen as main targets, the sea bream *Sparus auratus* and sea bass *Dicentrarchus labrax*. Eight shellfish species were selected. Two of these, the Japanese oyster *Crassostrea gigas* and the European blue mussel *Mytilus galloprovincialis* were designated as major species. The other species were the clams *Ruditapes decussatus* and *Ruditapes philippinarum*, the deep-sea hydrotherm vent mussels *Bathymodiolus azoricus* (Atlantic species) *and Bathymodiolus thermophilus* (Pacific species), the oyster *Crassostera angulata* and the mussels *Mytilus edulis* and *Mytilus trossulus*. Over the course of the project a total of approximately 50 cDNA libraries were prepared and sequenced generating over 100,000 EST.

Efficient and high-throughput genotyping necessitates a large and reliable source of markers. EST sequences qualify readily; they carry some redundancy, when prepared from several outbred individuals, and include a range of repeat sequences, facilitating the detection of polymorphisms. Large-scale search for markers focused initially on microsatellites (and to a lesser degree exon primed – intron crossing (EPIC) primers and amplified fragment length polymorphisms (AFLP)) but shifted to SNPs in key species such as the shellfish Pacific oyster, carpet-shell and Manila clam, *Mytilus* mussel, and the fishes European sea bass, sea bream and cod. Microsatellite and single nucleotide polymorphism (SNP) *in silico* mining of EST compared available software followed by laboratory validation and large-scale genotyping. A consensus emerged that SNPs are highly suitable for high-throughput analysis because of their common occurrence throughout the genome (at a density of one every 100 bp or less) and their comparability (facilitating calibration). There is a tendency to upscale SNP genotyping for the number of loci and individuals.

As the genomes of very few marine organisms were known, MGE has mapped microsatellite and SNP markers through dedicated mapping panels for a number of organisms (Manila clam, Pacific oyster, mussel, gilthead seabream, European sea bass). Medium-density linkage maps now available which make feasible quantitative treat loci (QTL) studies. End-sequencing of a BAC library of European sea bass and the consecutive assemblage of these sequences lead to the preparation of a physical BAC map. This has show that genome sizes of fishes such as sea bass and seabream are compact (e.g. genome size of European sea bass is 1.6 fold larger than pufferfish) and revealed some conserved and divergent features among perciform fishes. An additional advantage of the large-scale sequencing of BAC ends and a growing number of full-sized BACs has been their integration to the linkage maps.

A further development of BACend sequencing was the draft sequencing of the European seabass genome led and financially supported by the MGE platform Max Planck Institute of Molecular Genetics (Berlin). Currently at 2 fold coverage, the assembly has benefited from comparative mapping to available sequenced genomes of other fish.

Finally, cDNA clones have been printed into glass slides to develop microarrays for European sea bass (~17,000 unigenes), sea bream (~17,000 unigenes), Japanese oyster (~10,000 unigenes), mussels (~7,500 unigenes). Two oligonucleotide arrays were developed for sea bass and sea bream each corresponding to ~20,000 unigenes. Currently 15 projects inside and outside MGE are applying these microarrays.

III.4.2. Applications to evolution, fisheries and aquaculture

Genotyping of a large number of individuals at a large number of markers has considerably increased the potential of applications in environmental genomics. Neutral markers allow for detecting genome-wide events largely linked to population dynamics, while selective markers allow for detecting adaptive events. Basin-wide and regional studies are in progress to understand how neutral and non-neutral (adaptive) variation has shaped population structure in marine (shell)fish. Results from cod and European sea bass point to a largely neutral background of evolution, but at some loci significant signatures of selection.

Nuclear and mitochondrial genomes are used to probe for the mechanisms of speciation and adaptation in

the marine environment. Detailed analyses of SNPs in Pacific oyster revealed a level of polymorphism an order of magnitude larger than reported in animals so far. Non-synonymous substitutions contributed substantially to the polymorphism in the coding region, which tends to show that purifying selection is not that strong as could be expected from species with large population sizes. Codon bias might provide a useful indicator of the level of constraints upon proteins in the oyster genome.

The proliferation of raw sequence data and the large number of microsatellite and SNP markers now available for commercially exploited fishes has opened realistic perspectives for **fisheries** management, including traceability. For example distinct populations ("stocks") of cod were observed with large numbers of SNPs both in terms of genetic differentiation and levels of variability. However, inferences on the population of origin of a fish or individual fish can rarely rely on fixed genetic differences among populations, since they are connected by gene flow. Instead statistical methods have to be used for the calculation of the most likely population of origin. Most commonly the genotype is used, but the transcriptome can be also valuable to identify populations. The newly developed cDNA and oligonucleotide microarrays of European sea bass, sea bream, Pacific oyster and other shellfish should assist with stock identification.

The benefits of the novel genomic tools for aquaculture are diverse. High-throughput genotyping of microsatellites and SNPs has made paternity identification and broodstock diversity management feasible at the commercial level. For example annually summer mortality, in Pacific oyster of the East Atlantic coast, which is related to virus OSHV1, has been linked to four QTL. Growth in European sea bass has been linked to two major QTL, while handling stress has been linked to two QTL in males and one QTL in females. Such results will be substantiated through the incorporation of additional families combined with the analysis of genomic architecture and transcriptomics.

Inheritance of Mitochondrial DNA in Bivalves

When we refer to an organism's genome, we usually mean the organism's nuclear DNA. This is justified given that cytoplasmic genomes, either in the form of mitochondrial or plastid DNA, represent only 10^{-5} of the cell's genetic information. Another important difference between nuclear and organelle DNA is that the first is inherited biparentally, according to Mendel's laws, and the second generally uniparentally. This appears to be one of biology's most general rules.

However, in mussels, mitochondrial DNA (mtDNA) inheritance is biparental: the species carries two mtDNA genomes, not just one as is the cases with all other organisms. The female or F mtDNA is transmitted from mothers to daughters and the the male or M mtDNA fathers to sons. In a sense, the mussel species is a vehicle through which two different mtDNAs travel through time, each following a uniparental inheritance, either maternal or paternal. For this reason, the phenomenon has been known as doubly uniparental inheritance (DUI). Even though known since 1994, DUI remains largely unexplored. MGE has made significant contributions to the understanding of the mechanistic and evolutionary basis of the phenomenon.

1. Complete mitochondrial genomes. The complete nucleotide sequence of the F and the M genomes of the sea blue mussel (*Mytilus edulis/galloprovincialis*) was obtained and compared to each other. The main finding was that these two conspecific genomes differ in their primary sequence by more than 20%, a molecular distance normally found between classes of vertebrates, suggesting that DUI is a very old phenomenon. But the gene order has remained identical. The major difference is confined in a region that most likely contains the elements for the control of the molecule's replication and transcription. A third type of genome, genome C, was also sequenced (Figure 1).This is a genome that reversed its transmission route from female to male. The C genome is almost identical to F, except in the control region, which is a mosaic containing multiple sequences of the M genome inserted within the F genome.

2. mtDNA recombination. Whether animal mtDNA in general recombines has been hotly debated. One reason for this is that life-history inference from mtDNA sequence comparisons rests on the assumption of no mtDNA recombination. Several studies produced direct evidence of recombination between mussel mtDNA genomes and others extended it to animal mtDNA in general. The identification of genomes with recombined control regions have also led to the hypothesis that the control region may contain elements that affect the genomes transmission route.

3. Mussel egg proteomics. In mussels sperm mitochondria enter the egg. In adult females, the paternal mtDNA is practically absent and in males is the sole resident of the germ line. It was found that in female-destined embryos the sperm mitochondria follow a random assortment after early cell division, but in male-destined embryos they form an aggregate that keep them in the same cell. This might be the way through which sperm mtDNA is delivered into the male embryo's primordial germ cells, a finding that may lead to an understanding of the developmental pathway of DUI. That sperm mitochondria may have different fate depending on the type of the egg into which they are delivered (i.e. whether the egg originates form a daughter-producing or a son-producing mother) suggests that factors in the unfertilized egg may interact with sperm mitochondria.

4. How wide-spread is DUI? DUI was first discovered in mussels and was subsequently found in several other bivalve families. At present, the phenomenon is known to occur in some 34 species, yet its distribution remains

sporadic due to problems of detection and possible re-invasion of the male line by the female. As a result, the question of one-versus-multiple origins of DUI and the possibility of secondary loss of DUI remains also unanswered.

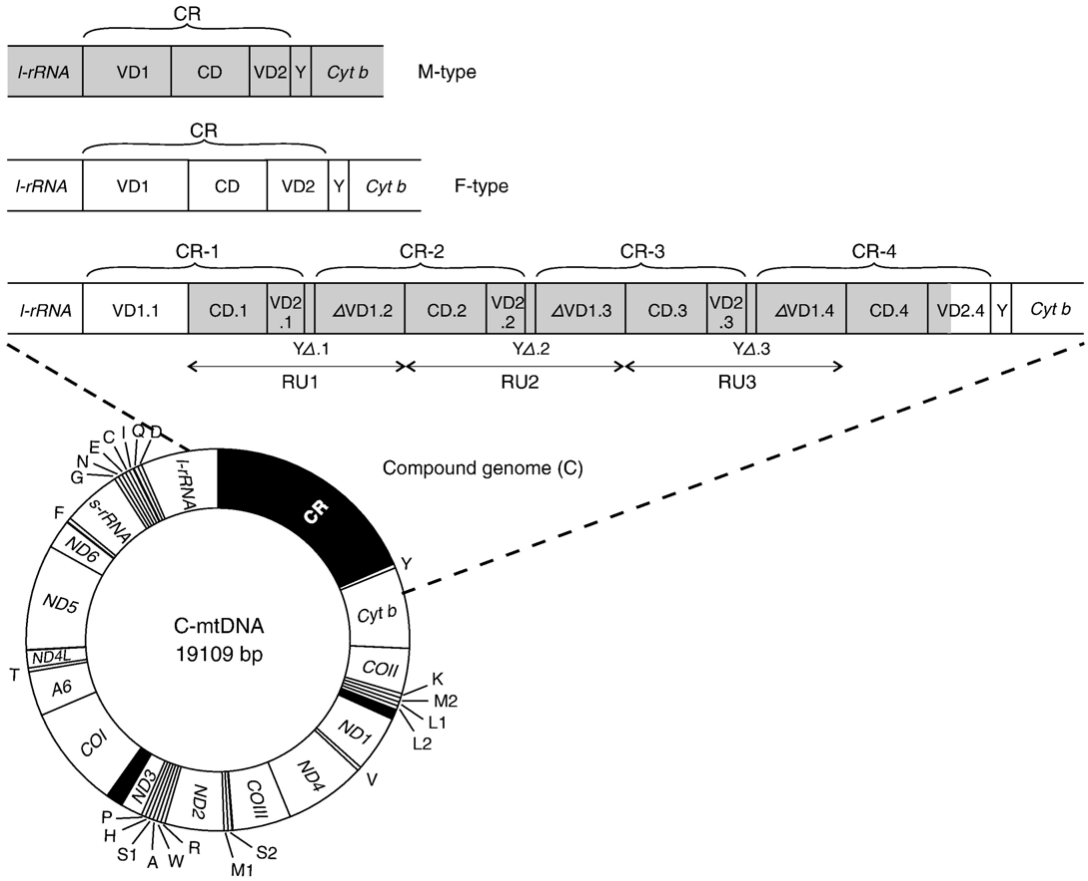


Fig. 1. Circular map of mitochondrial genome C.

The number inside the circle indicates the size of the genome. All genes are transcribed clockwise. The one-letter amino acid code is used for tRNA designation; s-rRNA and l-rRNA, small and large subunits of ribosomal RNA genes; other symbolisms refer to protein coding genes. CR, main control region. The magnified region depicts the complex CR of the genome in linearized form: VD1, first variable region; CD, conserved domain; VD2, second variable domain. Parts shown in grey are of M origin, other parts are of F origin. The CR of the F and the M genome are given for comparison.

IV Spreading Activities

IV-1 TRAINING AND EDUCATION ACTIVITIES (WP19)

Project leader: Mike Thorndyke, Kristineberg Marine Research Station, Sweden, Partner 38
Teaching and Education Programme Manager: Ulrika Hjelm, Kristineberg Marine Station Partner 38
Teaching and Education Committee:

- Johanna Wesnigk, MPI Bremen, DE, Partner 4
- Alex Rogers, NERC-BAS, Cambridge, UK, Partner 12 (2004 -2005)
- Euan Brown, SZN, Napoli, IT, Partner 22
- Patrick Cormier, CNRS Roscoff, FR, Partner 2 (2004-2005)
- Filip Volckaert, Katholieke University, Leuwen, BE, Partner 27
- Frédérique Viard, CNRS Roscoff, FR, Partner 2 (2006 -2008)

The three objectives for the Training & Education Programme have been i) to integrate MGE Network members by implementing training courses in Marine Genomics (Activity Planning), ii) to identify long-term training and teaching activities, needs and provisions/supply (Strategy Planning), iii) to further European Mobility and Integration in the fields of marine biology and genomics by offering accredited courses to European graduate student programmes. To reach these objectives a T & E Council was set up with members from the following contractors: #38, KMRS (chair & manager); #2, CNRS; #4, MPI; #11, EMBL; #22, SZN; #27, KULeuven

The main focus of the programme has been to provide two types of courses:

Short training courses

- 4-5 days
- practical, technical approach
- concentrating on a specific technique
- locally organized
- aimed at the MGE community

Summer courses

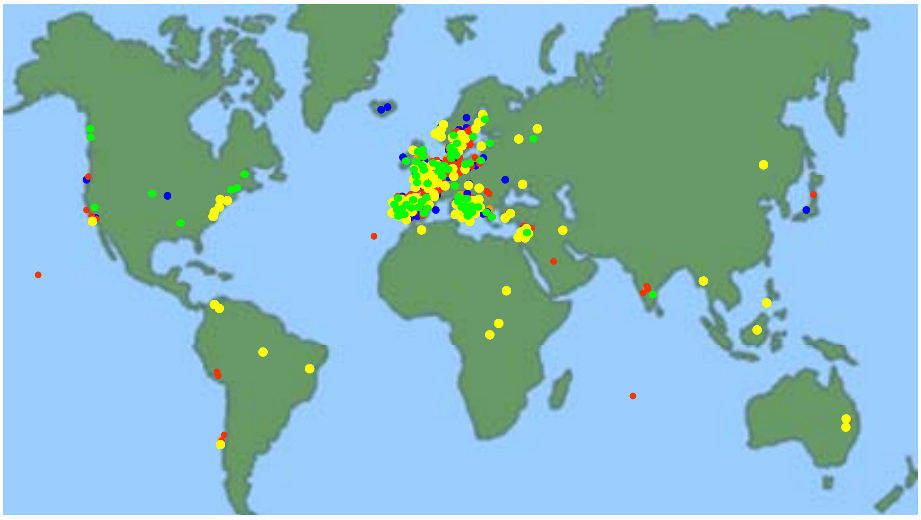
- 2 weeks
- a wide scope mixing theory and practice
- MGE organizing committee
- large number of invited teachers
- primarily for the MGE community and secondarily
- for outside the network
- MGE homemade products! Major investments made!

Over the 4.5 years of the project 16 short training courses and 10 summer courses have been organized (table 1). The different subjects were decided after communication and discussion within the consortium. Questionnaires have been circulated to all members asking about the interest in established courses as well as inviting suggestions for new topics. The questionnaire result has been complemented by input from involved scientists and their opinions on what is needed to fill any gaps. The majority of courses have been oversubscribed and hence repeated. The participants were mainly consortium members during the first year but after that the fraction of non-members and even non-Europeans increased over the course of the project (fig 1). In total 640 applications were received for 402 places. In addition to the course participants there have been 220 teachers and organizers involved in the course programme.

Table 1. MGE courses 2004-2008

Short training courses	*Summer courses*
2004	*2005*
Bioinformatics I, Bremen	Marine Evolutionary & Ecological Genomics, Naples
2005	*2006*
Bioinformatics II, Bielefeld	Marine Evolutionary & Ecological Genomics, Naples
Proteomics, Rennes	Analysing Biodiversity & Life History Strategies, Kristineberg
Microarrays, Bremerhaven	*2007*
Microarrays, Bielefeld (twice)	Marine Evolutionary & Ecological Genomics, Roscoff
2006	Plankton Bloom Dynamics, Barcelona
Bioinformatics I, Bremen	Evo-Devo meets Marine Genomics, Naples
Bioinformatics II, Bielefeld	Analysing Biodiversity & Life History Strategies, Kristineberg
Proteomics, Rennes	*2008*
cDNA libraries, Berlin	Marine Evolutionary & Ecological Genomics, Roscoff
2007	Plankton Bloom Dynamics, Bremerhaven
Bioinformatics I, Bremen	Analysing Biodiversity & Life History Strategies, Kristineberg
Bioinformatics III, Ghent	
qPCR, Heidelberg	
2008	
Bioinformatics I, Bremen	
cDNA libraries, Berlin	
qPCR, Heidelberg	

Overall the T & E Programme has been very successful with key techniques and theories being spread to a large group of mainly junior scientists. The participants evaluated all courses between good and excellent, indicating that the level of genomic knowledge has indeed been raised across the members of the consortia and beyond. One of the major outcomes from the courses has been the excellent team spirit of the groups. The participants have had not only good learning experiences together during the courses but also continued to keep in touch after the course; developing their professional and personal networks. Their academic careers should most surely benefit from these new contacts, increasing the integration of MGE members.



2004, 29 applications
2005, 139 applications
2006, 138 applications
2007, 209 applications
2008, 125 applications
TOTAL, 640 APPLICATIONS

Figure 1: Applications for MGE courses 2004-2008

The T &E Council also had the task to specifically integrate the group of MGE pre and post doctoral fellows. For this purpose four different activitieswere organized. (i) A specific workshop was held for all MGE *funded* grad students and post docs during the SSC meeting in Berlin in October 2005. We aimed to get them to know each

otherand to learn about each other's project and to learn more about Marine Genomics Europe and this was certainly achieved to a large extent. The International Marine Genomics Conference in Sorrento (October 2006) was an excellent occasion for MGE PhD students & post docs (and others!) to interact among their own group and with others. The T & E Council took the opportunity to stimulate integration by organizing (ii) a poster competition and by inviting the MGE funded students and post docs to the Banquet. (iii) A larger workshop for junior marine scientists was organized in Crete (Greece) in October, 2007. 80 participants from 28 MGE partners took part (see figure 2). The aim was to provide PhD students and post docs working in the field of marine genomics an arena for discussions of scientific and non-scientific subjects. The workshop was set up with four different sessions; three with scientific focus and one session dealing with career issues (Genome analysis; Functional analysis; Image analysis; Career development)

The research oriented themes were covered by invited speakers and by talks selected from abstracts. The starting point for these sessions was the techniques and how they are used for studying different organisms and answering a wide range of biological questions. The career session was divided into two parts, the first part with talks starting with Dr Sohail Luka from the European Commission (Facilitating Researchers' Career Development through Mobility) followed by speakers covering different post docs fellowships (Marie Curie, HFSP and EMBO); giving evidence about their experience on moving from academicresearch to industry as well as from academic work to starting up a company; outlining the "Two Body Problem" and other gender issues in career development. A poster session followed that moved slowly into a Cretan evening with traditional food and dance. The next day the second part of the career session was set up as an informal "chat-with-the-Prof" session where the participants gathered in smaller groups over a cup of coffee and talked with a senior scientist about any career issues they had an interest in.

On one evening, the participants were split up in sub-groups for round table discussions on different issues of 'genomics today & in the future'. The discussions were lively and continued during the days after in a smaller working group and resulted in which put together a written statement reflecting the views of the larger group ("The declaration of Kolombari") that was published on the website.

Figure 2. The participants of the MGE Workshop "Marine Genomics: An Ocean of Techniques"

As part of the final assembly in Faro (Portugal) in May 2008 (iv) a Junior Scientist Session was organized comprising a talk on Open Access and its possibilities and follow-up group discussion.

In summary: Training is an excellent tool for integrating people! 26 MGE courses have provided new networking opportunities for 402 junior marine scientists!

IV-2 COMMUNICATION (WP37)

Project leaders: Michèle Barbier and Anne Saisi, Roscoff, CNRS, Partner 2

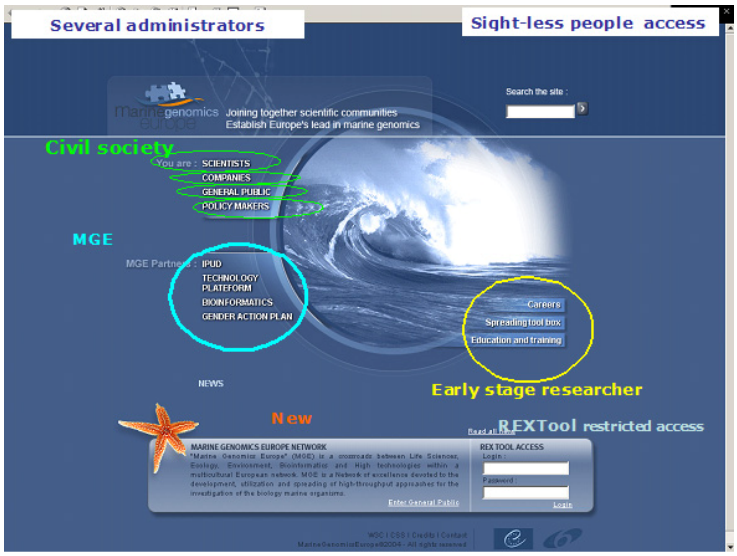
Objectives of communication activities

MGE developed a dedicated knowledge portal to marine genomics for the civil society organisations as well as to researchers, and generated a specific interface between researchers, their Institutions and Society. Based on an intuitive portal solution (http://www.marine-genomics-europe.org), the visitor could find an access to scientific information, pedagogic documentation and background material.

This resource centre, linking information resources, knowledge and communication tools, provided support to a) the integration of distinct research communities and b) the spreading of knowledge to the ERA participants by e-communication activities.

The Marine Genomics Europe website

http://www.marine-genomics-europe.org



The web site dedicated to MGE has been developed in September 2004 and is accessible since October 2004 and is still running. The MGE web site home page has been designed to be intuitive for :

- Civil society end users (with 4 different accesses reflecting the main civil society end-users. The scientist pages have been completed and updated once a month. The company and general public have been both under development and will be completed in the next 18 month.
- MGE members (The web site comprises short cut access for MGE members who are looking for precise information regarding (application form, TP access scheme, Bioinformatic platform..)

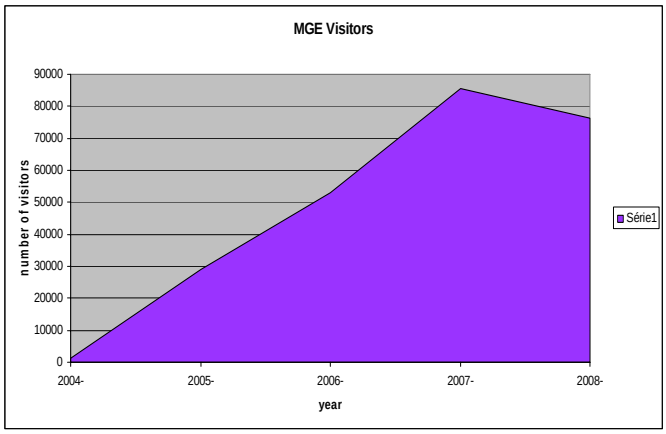
Early stage researchers (MGE members or not) for easy access to jobs opportunity (Career) and the MGE Training & Education Programme (courses, application Form), open to both internal and external MGE participants and is updated in average once a week.

MGE website

The objective of the WP Communication was to improve, on a regular basis, the virtual tools set up during the first year: the General MGE website and the Rextool. This resource communication provided a link to know the improvement of the dissemination of MGE knowledge throughout the communication activities.

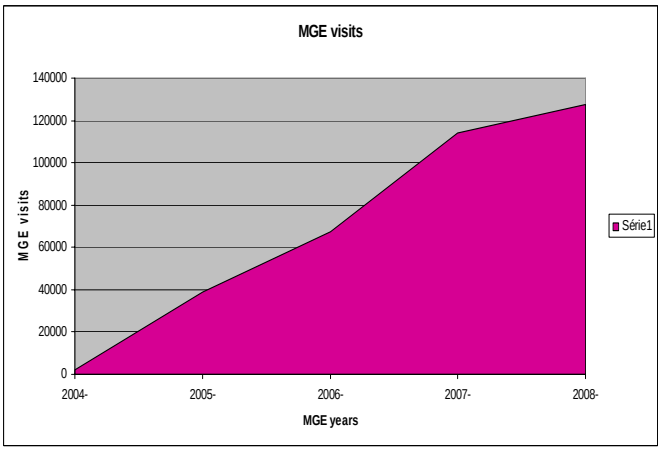
Starting from the month following installation, the web site access has been monitored thus providing statistics as shown in the figures below. In 2004 there were about 1343 visitors and since, this number has increased regularly and notably.

Year	Visitors number
2004-	1343
2005-	29182
2006-	52840
2007-	85387
2008-	76278,5



Year 2008 takes only into account the data for 10 first months

MGE years	Number of visits
2004-	2083
2005-	38725
2006-	67166
2007-	114283
2008-(jan-oct)	127298



The web site is also referenced by all the major search engines (Google, MSN, Yahoo, etc.), which gave a better visibility for the Marine Genomics NoE. A noticeable fact is also the appearance of numerous links pointing at the Marine Genomics Website from external institutions (Eurocean, Ademe, institution's partner, Marbef, Eur-Ocean, etc).

As shown in the figure below, visitors from European and non European countries paid a visit to MGE web site. Here is a table listing of the Top-7 countries from 2004 until 2008.

MGE 2004	MGE 2005	MGE 2006	MGE 2007	MGE 2008
France	France	France	France	Israel
Germany	Germany	Germany	Czech Republic	Brazil
United Kingdom	United Kingdom	Canada	Germany	France
Portugal	Italy	Italy	United Kingdom	Argentina

Italy	Netherlands	Belgium	Canada	Italy
Spain	Belgium	United Kingdom	Italy	United Kingdom
Belgium	Canada	Netherlands	USA Educational	Germany

The Rextool furnished of sets of information relating to participants' needs for knowledge within the NoE REX Tool management Platform.

The REXTool management platform was aimed to assist the different partners in their activities by integrating their documentation within a shared knowledge base. It has been used for providing access to confidential document restricted to the MGE community (minutes of meeting, report on the training and education programme) and deliverables. The platform was very useful during the reporting periods and allowed also to update the list of persons actively working in the programme.

Communication towards Companies

Additional actions have been developed in collaboration with FIST in charge of Intellectual Property Use and Dissemination. A brochure formatted for companies has been design and one sample has been distributed to each team leader in MGE (118 team leaders) in 2004 for the 1rst year.

Communication towards General Public

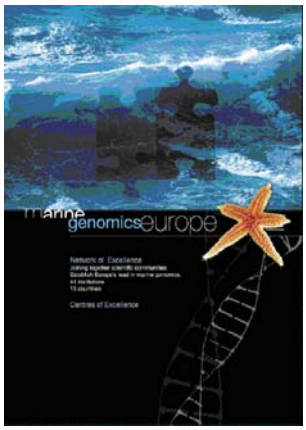
Communication towards the general public included Popularization Conferences **(e.g.** during the week of Science, the night of Science) involvement in various events and the publishing of a specific brochure describing MGE activities.

An Experimental school project aiming at describing Research and marine Genomics to teenagers (11-13 year-old) through school program was implemented. This project entitled "When Prévert tells European the Genomics" was a partnership between CNRS-Roscoff) and a local high school (college Jacques Prévert). This project involved School teachers, scientists and children and allowed to create tools that have been used for the design of a school booklet and for the development of general public enter-web site. Moreover, all MGE team leaders received a package containing pedagogic material: the booklet & slide shows for to school (conference).


Communication towards Scientists

Beside the website, different tools have been developed to reach MGE scientists as well as other scientists:

- Posters (size A0, A4) describing the whole MGE network, or more specific to an institution have been designed, posted on the web site (available for MGE partners), printed and hang in labs belonging to the consortium.
- Several dedicated Brochures
- Electronic newsletters (on a monthly basis) and one paper-printed newsletter



Communication towards Policy makers

The Management team attended different conferences in order to promote Marine Genomics towards policy makers (e.g. the European Conference Marine Science & Ocean Technology, in Galway, May 2004; the ESF marine Board autumn meeting in Lisboa, Pt, November 2004; Meetings of the EraNet MarinEra, etc.,)

ImaGene, the Photo Contest

The Science and Society Committee implemented two rounds of a photo contest t: ImaGene, the first photo contest on marine genomics. This photo contest was subdivided into 4 main categories: ecosystem (landscape), Organisms (animals, invertebrates), Cell (electronic or photonic microscopy images) and genomics (open to any new representation of what is genomics). A specific committee, including three professional photographs met in fall 2005 to award the best pictures. The winners received an allocation of €1000 for travelling to a European laboratory (partner of MGE) for their career development. The images of this photo contest have been used for setting up a travelling exhibition that was exposed in 2006 in several aquaria, laboratories and museum through out Europe.

IV – 3 GENERAL ASSEMBLIES AND INTERNATIONAL CONFERENCE

A **kick off meeting** was held in Rennes (France) on the 24th of June 2004 and about 160 participants coming from all countries involved in the project were present. A full-day conference was dedicated to a general presentation of the project's content (integration, spreading and research activities, management scheme and budget allocation). It was also the occasion to remind the missions and rules of a Network of Excellence and to arrange a General Assembly. Numerous joined sub-project meetings were also organised in the following days. The kick off meeting really gave an opportunity for all participants to meet altogether to discuss and to truly come into a concrete stage of the project.

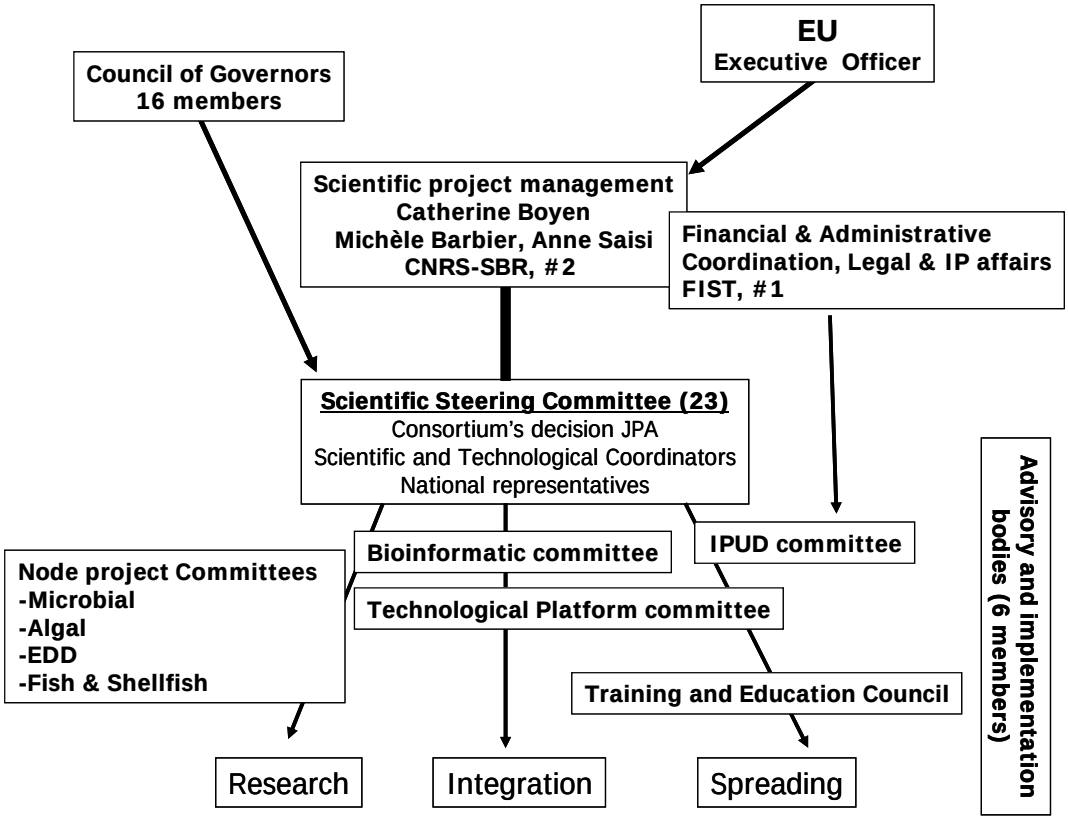
A **one-week General Assembly** was held in Banyuls (France) in February 2005. The event brought together 150 participants and comprised many parallel and successive meetings (SSC meeting, eight sub-project's committees and 4 general node's assemblies). This rather large meeting was mainly devoted to the first year's state of advancement of the project, to the validation of the project deliverables and to the preparation of the 2nd Joint Programme of Activities (JPA).

The International Conference in Marine Genomics took place in Sorrento, Italy, October 28 – November 1 2006. The event was jointly organised by Marine Genomics Europe, the Stazione Zoologica (partner 22) and the SARS Institute (partner 29).
In total 284 participants from 30 different countries attended the conference. The program included 10 plenary lectures (45 min), 25 keynote lectures (30 min) and 50 short talks (15 min). About 125 posters were displayed during the poster sessions. A special issue of the Journal GENE was devoted to manuscripts related to the talks or posters presented in the IMGC in Sorrento. This event was a great success and above all, allowed to bring together scientists from very different fields (evolution, development, population genetics, environmental microbiology, oceanology and model biology) demonstrating thus the strong interest of merging these areas. It also revealed that the huge biodiversity of marine organisms allows researchers to investigate in depth developmental and evolutionary problems as well as questions related to global warming and sustainable development. Marine genomics clearly appears to be a **new emerging discipline** showing great potential.

A **Final General Assembly** took place in Faro, Portugal in May 14-16. The objective was to gather the project partners in order to present an overview of the activities of the network (Integrating, spreading and scientific activities). This event was a showcase of the achievements of Marine Genomics Europe. In total there were 133 participants, of which more than 15 postdocs and 25 PhD students. Attendees came from 13 different European Countries. Twenty four MGE partners were represented and members from seven other Institutes joined the event.

V - CONSORTIUM MANAGEMENT (WP23)

The scheme below gives an overview of the management of MGE Consortium. The central decision-making body was the **Scientific Steering Committee** (SSC) that, as specified in the Consortium Agreement, was composed of scientific coordinators of one or several work package(s) of the project, of the national representatives and of the chairperson of the SSC (Scientific Manager, namely Dr. Catherine Boyen). A detailed composition of the SSC including the allotment of responsibilities is provided in Table A.



Nine SSC meeting were convened the course of the project

- 16-17 March 2004 in Leuwen, Belgium
- 23-24 June 2004 in Rennes, France
- 9-10 February 2005 in Banyuls, France
- 2-3 October 2005 in Berlin, MPI, Germany
- 19-21 February 2006 in Barcelona, CSIC, Spain
- 28 October 2006 in Napoli, SZN, Italy
- 2 March 2007 in Paris, CNRS, France
- 11-12th October 2007 in Chania, Creta.
- 13 May 2008 in Faro, Portugal.

During these meetings, the SSC discussed and decided on:

- The strategic and scientific orientations of the Project and its implementation.
- The Joint Programme of Activities (Node projects, Teaching & Education, Bioinformatic platform, Technological Platforms, Spreading, Intellectual property, Gender Action Plan, Workshops and Conference, Web site...)

- The project's budget and the allocation of the EC financial contribution to the different activities and among Contractors
- The reporting activities

Several other people were invited to the SSC meetings on a regular basis:

- The managers appointed exclusively for the Network:
  - Michèle Barbier [2004-2007] and Anne Saisi [2008]: Assistants to the scientific project manager (located in Roscoff, Partner 2) who were in charge of drafting the minutes of the meetings
  - Jomuna Choudhuri [2004-2006] and Virginie Mittard [2007-2008]: Coordination of the Bioinformatic Platform (located in Bielefeld, Partner 7)
  - Ulrika Hjelm [2004-2008]: Teaching & Education Manager (located in Kristineberg, Partner 38)
- Frédéric Legros [2004-2005], Séverine Iffland [2006-2007], Joseph de Macedo [2007-2008] and Samia Halloui [2008] as the representative of FIST (Partner 1), the administrative and financial coordinator
- Cathy Eccles [2004 - 2006] and Ana-Teresa Caetano [2007 - 2008] Scientific Executive Officers, European Commission.

As the **Scientific Project Manager** Catherine Boyen chaired the SSC and was also in charge of implementing the decisions taken by the SSC. She was assisted in that task by Michèle Barbier and Anne Saisi. Along with the SSC meetings, a substantial part of the coordination activities was realized via e-mails and phone calls. For instance, numerous calls for proposals and votes for budget allocation were organised using electronic mails. A dedicated website (http://www.marine-genomics-europe.org/) was created in October 2004 and was used to provide access to calls for proposals (such as the Technological Platform access schemes), to questionnaires (such as the Gender action Plan questionnaire) as well as to various news articles and announcements. In February 2005 a web-based knowledge management tool, REXtool, was created for centralizing documents and partner's information. The REXtool was extensively used for the preparation of the periodic reports and of the Joint Program of Activities.

The Scientific Project Manager also maintained a tight link with the European Commission and with the Financial Coordinator (FIST), again by means of e-mails, meetings and phone calls.

Eight **sub-project committees** were implemented (the four node committees, the Bioinformatic committee, the IPUD committee, the technological platforms committee and the Training & Education Council). These committees prepared proposals on programmes to be conducted and on the orientations of the sub-projects and of the Programme of Activities. They also made recommendations on the allocation of sub-project tasks, financial needs and allocation among the contractors and on the need for sub-contractors for the establishment and the fulfilment of the Programme of Activities. All these discussions and decisions were monitored and validated by the SSC. These committees have met at least once a year in the course of the project and records were made available on each occasion.

The **Council of Governors**, composed of high-level representatives from the 16 participating countries met in Berlin on the 4th of October 2005. Contractors from each country decided on a single spokesperson from their organisations. The composition of the Council of Governors was established during year 1 of the project and was approved by the Commission staff. The Council of Governors is an advisory body which is aimed at providing "Marine Genomics Europe" NoE with a discussion forum that will help to coordinate the network's policy with that of national research agencies.

Annex I : list SSC members and their responsibilities

Name	Partner's number	Institution	Country	WP leader	responsibilities
Catherine Boyen	2	CNRS-Roscoff	FR	**23, 26**	Scientific coordinator
Bernard Kloareg	2	CNRS-Roscoff	FR		National representative
Hervé Moreau	2	CNRS-Banyuls	FR	**4,41**	Algal node leader
Chris Bowler	2	CNRS	FR, IT	**32**	Functional Genomics
Joël Quérellou	3	Ifremer Brest	FR	**2**	
Rudi Amann	4	Max Planck Institute for Marine Microbiology Breme	DE	**3** Microbial EG	Microbial node leader, National representative
Richard Reinhardt	6	Max Planck Institute for Molecular Genetics Berlin	DE	TP/Integration	TP leader
Alexander Goesmann	7	Bielefeld University	DE	**15-18** Bioinformatics	
Folker Meyer	7	Bielefeld University	DE	**15-18** Bioinformatics	
Wolfgang Hess	45	University of Freiburg	DE	**2** Microbial FG **27** Flagship	Comparative Genomics
Melody Clark	12	B.A.S NERC	UK	**9** F&SF EG **22**	
Alex Rogers	12	B.A.S NERC	UK	**9** EDD	National representative
Adelino Canario	21	Centre of Marine Sciences Faro	PT	**13** F&S FG	F&S node leader, National representative
Giorgio Bernardi	22	Stazione Zoologica di Napoli	IT		National representative
Ina Arnone	22	Stazione Zoologica di Napoli	IT	**8** EDD CG	
Esther Lubzens	25	Israel Oceanographic and Limnological Research Haifa	IS	**13** F&S FG	National representative
Filip Volckaert	27	Katolieke University Leuven	BE	**14** F&S EG	National representative
Jeanine Olsen	28		NL	**24**	Environmental

		University of Groningen		Gender dimension	Genomics, National representative, GAP
Roman Wenne	31	Institut of Oceaonology Gdynia	PL		National representative
Carlo Pedrós-Alió	33	CMIMA, CSIC	ES	**1** Microbial CG	National representative
Elefteri Zouros	36	University of Crete	GR	**12** F&S CG	National representative
Mike Thorndyke	38	Royal Swedish Academy, Kristineberg Marine Station	SE	**19** Education and Training	EDD node leader, National representative
S-H Chan	29	SARS Institute	NO		
Daniel Chourrout	29	SARS Institute	NO		National representative
Einar Eg Nielsen	32	Technical University of Denmark	DK		National representative
Mickael Möller Hansen	32	Technical University of Denmark	DK		National representative