The Exploration of Marine Biodiversity

Scientific and Technological Challenges

Carlos M. Duarte (ed.)

Offprint of the Chapter

5. MARINE GENOMICS AND THE EXPLORATION OF MARINE BIODIVERSITY

by

J. Mark Cock, Delphine Scornet, Susana Coelho, Bénédicte Charrier, Catherine Boyen and **Akira F. Peters**
Marine Plants and Biomolecules Laboratory (UMR 7139, CNRS-UPMC)
Roscoff Biological Station, Roscoff, France

www.fbbva.es
ISBN: 978-84-96515-27-7

5.1. INTRODUCTION

THE OCEANS COVER approximately 70 percent of the Earth's surface. Apart from providing a third of the oxygen that we breathe and acting as moderators of global climatic change with a significant influence on the human population's terrestrial environment, these vast expanses are also an important source of high-protein food. Marine and coastal environments include many diverse pelagic and benthic habitats such as open-ocean ecosystems, deep-sea communities including thermal vent ecosystems, kelp forests, mangroves, coral reefs, etc. Although these varied environments support a rich abundance of life, marine biodiversity has received much less attention than its terrestrial counterpart. This may be because the oceans have historically been thought of as regions of low biodiversity and because of difficulties with accessing marine environments. In fact, by some measures, biodiversity in the oceans is greater than on land. Marine biosystems have been evolving for an additional 2.7 billion years compared to terrestrial environments, and almost all the currently described phyla are represented in the ocean while only about half have terrestrial members. The phylogenetic diversity of marine organisms is, therefore, much broader than that of their terrestrial counterparts (Ray 1988). Marine biodiversity may also be greater on a functional level, in the sense that marine organisms have adopted many novel survival strategies for which there is no equivalent amongst their terrestrial counterparts, such as microbes and animals associated with hydrothermal vents.

There is a pressing need for a more detailed understanding of marine biodiversity in oceans, and particularly in coastal areas, as they come under increasing threat from pollution, over-exploitation and badly planned development programmes. These threats take many forms such as the direct effects of chemical pollution, eutrophication, over-fishing and physical alterations to the coastline, together with the indirect effects of global climate change and the intro-

◀ **Photo 5.1: Giant kelp *(Macrocystis pyrifera)*.** This alga has gas-filled bladders known as pneumocysts that keep it floating upright near the surface. The genome of the species is currently being explored by a consortium of US and European pharmaceutical companies.

duction of exotic species. As a result of this sort of activity many areas have been degraded and over-exploited beyond repair, with coral reefs and mangroves being particularly at risk. The mounting concern about these problems has led to an increase in the number of international instruments aimed at addressing the threats to marine and coastal biodiversity, and at protecting and using marine resources sustainably. For these instruments to be effective, however, the threatened ecosystems need to be understood in more detail. Studies need to be more than censuses of the organisms present in individual biosystems, and should include information about the genetic structure of the populations of organisms that make up a biosystem, about functional aspects of interactions within ecosystems and about the ability of populations to adapt to changing conditions. Several tools are available for this type of study. This article looks at how techniques developed within the new discipline of genomics can be applied to the study of marine diversity, concentrating particularly on coastal biodiversity. We focus primarily on eukaryotic organisms, which, because of their often large genome sizes, represent the greatest challenge for the application of these techniques.

5.2. GENOMIC PROGRAMMES AND MARINE BIOLOGY

Genomic approaches are expected to provide essential information for studying, monitoring and exploiting biodiversity in the oceans. In this respect, the remarkable diversity of life in the sea can be viewed both as an advantage and as a disadvantage for marine biologists. On the positive side, this diversity holds the promise of a great richness at several levels, from ecosystems down to genes. From a more practical point of view, however, the problem arises as to how methods can be developed for studying such a wide range of ecosystems and organisms. This problem becomes particularly acute when the aim is to apply genomic approaches, because large-scale analyses of this sort are difficult to apply across a broad range of organisms.

The field of genomics was initially developed by biologists working on the biology of terrestrial species, and one key factor in the emergence of this discipline was the existence of well-defined and intensely studied model organisms such as baker's yeast *(Saccharomyces cerevisiae)*, the fruit fly *(Drosophila melanogaster),* a nematode worm *(Caenorhabditis elegans),* mouse ear cress *(Arabidopsis thaliana)* and, more recently, the mouse *(Mus musculus*; Davis 2004). These model organisms were developed to study animal and terrestrial plant biology and, although much of the information obtained from them is

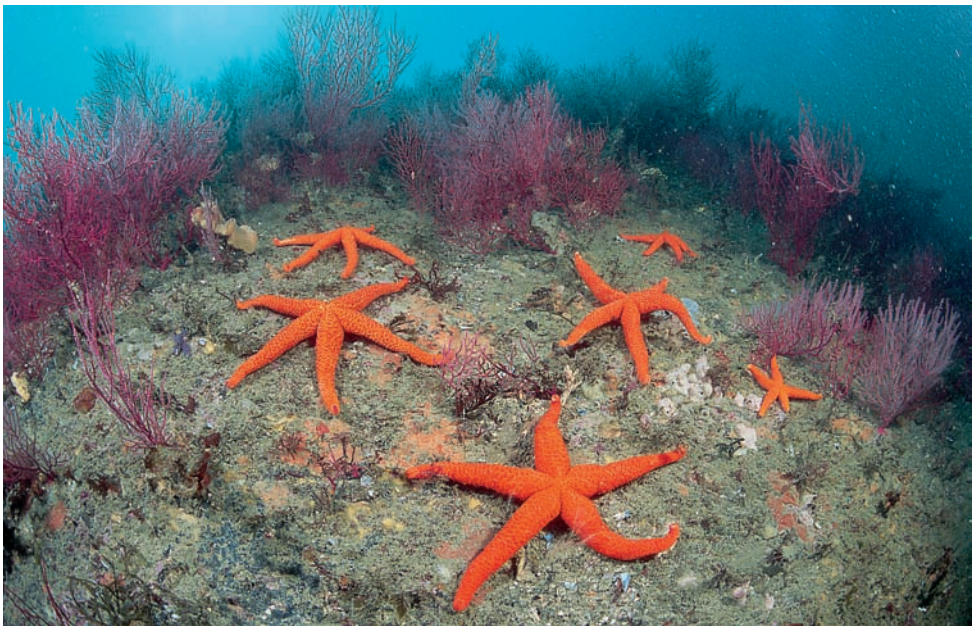


Photo 5.2: Red starfish *(Echinaster sepositus)*. Starfish feed on animals, like certain varieties of sponge, that use chemical defence, and have developed adaptations allowing their consumption. This "chemical warfare" has served as an evolutionary motor for the production of bioactive compounds.

of fundamental interest, their study has often been justified and driven by their use as tools to address concrete, "applied" problems. These problems are essentially of two types, both directly relevant to the human population: disease (including both infectious diseases and diseases of a developmental nature such as cancer) for the animal models, and food production (in a wide sense, including the effects of both development factors and disease on plant production) for plant models. In order to facilitate the transfer of knowledge to humans or crop plants, work on these model organisms has concentrated primarily on conserved traits that are, in many cases, understood in considerable depth. This approach allowed the establishment of large research communities and the development of extensive resources around these model systems, and this was a key factor in the transition to genome-scale biology.

The context for marine biology is significantly different, the accent being more on understanding how organisms function in the context of their particular ecosystem than on asking general questions about their biology. This does not mean that genomic approaches are not relevant to marine biology but rather that they need to be applied in a different way. For example, for marine biologists, the concept of a model organism is used in a much more flexible

manner. In some contexts it could be useful to have a very complete model organism for which both genome sequence data and functional genomics tools are available, whereas for other questions models may not need to allow such in-depth analysis and, for example, a genome sequence or even a large-scale EST collection might suffice.

In some situations, even the organism-level approach itself is not relevant. Hence the development of metagenomic approaches in which marine biosystems are directly sampled and sequenced (Beja et al. 2000; see also the study by Venter et al. 2004 who carried out high-throughput sequencing on DNA from microplankton obtained by filtering water from the Sargasso Sea through a 3 µm filter). This type of approach not only represents a very interesting method of obtaining a "snapshot" of the genetic complexity of a particular biosystem, but also obviates the need for culture methodologies for the constituent organisms. Metagenomics was pioneered in marine biology and provides a good example of how genomic approaches can be adapted to address the questions posed by marine biologists. However, whilst metagenomics provides a broad overview of the genetic composition of ecosystems, more detailed analyses will require organism-level approaches. The question therefore remains as to how genomic approaches can be developed for the diverse marine biosystems.

5.3. GENOMIC MODELS FOR MARINE BIOLOGY: THE NEED FOR MODEL ORGANISMS DISTRIBUTED ACROSS THE TREE OF LIFE

Because of the vast phyletic diversity of marine organisms, existing genomic model organisms are often of limited relevance, because there is an enormous evolutionary distance separating these models from an organism of interest. To provide adequate tools for marine biologists, therefore, one important aim will be to develop genomic approaches, such as whole genome sequencing and functional genomics, for key species across the evolutionary tree. These key species can then act as "local" models for phylogenetically related organisms in the same way that, for example, the wealth of genomic information available for *Arabidopsis* has been exploited by researchers working on economically important crop plant species within the angiosperms. The first step towards the establishment of such models is whole genome sequencing (or, in some cases, extensive EST sequencing). As described below, the current collection of fully sequenced genomes will provide a starting point for such a project, but a concerted effort will be required from the marine community to attain this aim and convincing arguments will have to be put forward to support such a programme.

Until recently, genome projects have concentrated on model organisms or on organisms, such as pathogens or plant crops, that are of direct importance to the human community. Despite the fact that the organisms' phylogenetic position was not a major argument for most of these projects, they already provide a sampling of many diverse phylogenetic groups. Most of the model organisms (in the sense of being amenable to laboratory manipulation) with sequenced genomes are members either of the opistokont (animals or fungi) or the viridiplantae (green plant and algae) lineages, with *Dictyostelium discoideum* (a slime mold in the amoebozoa lineage) being a notable exception. Sequencing of human pathogens, however, has provided sequenced genomes from several other major eukaryotic groups including another amoeba *(Entamoeba histolytica)* and members of the apicomplexa (e.g., *Plasmodium falciparum)* and the euglenozoa (e.g., *Trypanosoma brucei).*

As genome projects have become cheaper, it has been possible to finance more diverse projects including, for example, the sequencing of the genomes of environmentally important organisms, such as the diatom *Thalassiosira pseudonana*, which provided the first complete genome from the heterokont lineage. Of course environmental importance was not the only argument put forward for *Thalassiosira pseudonana*, and the phylogenetic argument itself was also important in addition to other factors such as the biotechnological potential of silicate metabolism in this species *(Thalassiosira pseudonana,* like most diatoms, constructs a silicate exoskeleton, the frustule, and the processes involved in the production of this structure are of great interest for applications in nanotechnology). In this respect, *Thalassiosira pseudonana* is an interesting example for marine biologists of how phylogenetic arguments can be combined with other arguments, for example of a biotechnological or environmental nature, to convince funding bodies of the interest of sequencing the genome of a particular organism.

Table 5.1 lists the eukaryotic organisms for which complete genome sequences have been published. From this table it is clear that existing genome projects are gradually covering many of the major lineages that make up the evolutionary tree of the eukaryotes. Genome projects for additional key species are in progress, including quite a number of marine species, such as *Emiliania huxleyi* (a pelagic coccolithophore), *Hydra magnipapillata, Strongylocentrotus purpuratus* (purple sea urchin), *Litopenaeus vannamei* (the pacific white shrimp), and Amphioxis (the closest living invertebrate relative of the vertebrates), together with key species from other environments such as *Phytophthora infectans* (an oomycete) and the unicellular green alga *Chlamydomonas*

Table 5.1: Eukaryote species for which complete genome sequences have been published

Species	Classification	
Homo sapiens	Metazoa, Chordata, Vertebrata	
Pan troglodytes	Metazoa, Chordata, Vertebrata	
Rattus norvegicus	Metazoa, Chordata, Vertebrata	
Mus musculus	Metazoa, Chordata, Vertebrata	
Danio rerio	Metazoa, Chordata, Vertebrata	
Tetraodon nigroviridis	Metazoa, Chordata, Vertebrata	
Takifugu rubripes	Metazoa, Chordata, Vertebrata	
Ciona intestinalis	Metazoa, Chordata	
Drosophila melanogaster	Metazoa, Arthropoda	
Bombyx mori	Metazoa, Arthropoda	
Anopheles gambiae	Metazoa, Arthropoda	
Caenorhabditis briggsae	Metazoa, Nematoda	
Caenorhabditis elegans	Metazoa, Nematoda	
Neurospora crassa	Fungi, Ascomycota	
Aspergillus fumigatus	Fungi, Ascomycota	
Saccharomyces cerevisiae	Fungi, Ascomycota	
Schizosaccharomyces pombe	Fungi, Ascomycota	
Kluyveromyces lactis	Fungi, Ascomycota	
Candida glabrata	Fungi, Ascomycota	
Ashbya (Eremothecium) gossypii	Fungi, Ascomycota	
Yarrowia lipolytica	Fungi, Ascomycota	
Debaryomyces hansenii var. *hansenii*	Fungi, Ascomycota	
Phanerochaete chrysosporium	Fungi, Basidiomycota	
Cryptococcus neoformans	Fungi, Basidiomycota	
Encephalitozoon cuniculi	Fungi, Microsporidia	
Entamoeba histolytica	Entamoebidae	
Dictyostelium discoideum	Mycetozoa, Dictyosteliida	
Oryza sativa L. ssp. *indica*	Viridiplantae	
Oryza sativa ssp. *japonica*	Viridiplantae	
Arabidopsis thaliana	Viridiplantae	
Cyanidioschyzon merolae	Rhodophyta, Bangiophyceae	
Thalassiosira pseudonana	Stramenopiles, Bacillariophyta	
Plasmodium falciparum	Alveolata, Apicomplexa	
Plasmodium yoelii yoelii	Alveolata, Apicomplexa	
Cryptosporidium hominis	Alveolata, Apicomplexa	
Cryptosporidium parvum	Alveolata, Apicomplexa	
Theileria parva muguga	Alveolata, Apicomplexa	
Trypanosoma brucei	Euglenozoa, Kinetoplastida	
Trypanosoma cruzi	Euglenozoa, Kinetoplastida	

	Description	Genome size	Main criterion	Marine species?
	Human	3300 Mbp	M	
	Chimpanzee	3100 Mbp	C	
	Model vertebrate	2800 Mbp	M	
	Model vertebrate	3454 Mbp	M	
	Zebrafish, vertebrate model species	1700 Mbp	M	
	Fish genomic model	342 Mbp	C	
	Fish genomic model	400 Mbp	C	yes
	Sea squirt, basal chordate	160 Mbp	M	yes
	Model organism	122 Mbp	M	
	Silkworm	530 Mbp	I	
	Malaria mosquito	26 Mbp	P	
	Comparative genomics model	104 Mbp	C	
	Model organism	97 Mbp	M	
	Model organism	38 Mbp	M	
	Mold, opportunist human pathogen	30 Mbp	M	
	Model organism	12.1 Mbp	M	
	Model organism	12.4 Mbp	M	
	Yeast, genetic studies and industrial applications	10.6 Mbp	C	
	Opportunistic human pathogen	12.3 Mbp	C/P	
	Pathogen of cotton and citrus fruits in the tropics	9.2 Mbp	P	
	Commonly found e.g. on food, industrial applications	20.5 Mp	C/I	
	Halotolerant marine yeast	12.2 Mbp	C	yes
	White rot fungus, wood decay	30 Mbp	I	
	Opportunistic human pathogen	24 Mbp	C/P	
	Microsporidian pathogen affects nervous system	2.8 Mbp	P	
	Enteric parasite	20 Mbp	P	
	Slime mold, model organism	34 Mbp	M	
	Food crop	390 Mbp	F	
	Food crop	390 Mbp	F	
	Plant model species	157 Mbp	M	
	Unicellular red alga from hot, acidic springs	16.5 Mbp	E	
	Planktonic diatom	34.5 Mbp	E	yes
	Human malaria parasite	22 Mbp	P	
	Rodent malaria parasite	23 Mbp	C	
	Intestinal parasite	9 Mbp	C/P	
	Causes human cryptosporidiosis	10.4 Mbp	C/P	
	Tick-borne parasite (East Coast fever)	8.3 Mbp	P	
	Causes African sleeping sickness	35 Mbp	P	
	Causes Chagas' disease	108 Mbp	P	

1 Main criterion presumably informing the choice of each species for a genome programme: C: comparative genomics; E: environmental or phylogenetic importance; F: crop plant; I: industrial applications; M: model species; P: pathogen of humans or important crop species.

reinhardtii (a green alga, for which the genome sequence has been completed). For the prokaryotes, progress is even more rapid and many sequenced genomes are available including genomes of several marine organisms such as multiple strains of the pelagic photosynthetic bacteria *Synechococcus* and *Prochlorococcus*. Hence progress is being made towards coverage of all the major eukaryotic and prokaryotic groups. However, it will be important to actively channel this process in the future, to ensure that coverage extends to all the most important groups and especially to key groups for marine biologists, in particular the eukaryotes, many of which have large genomes. Initiatives such as the white paper "Frontiers in Genomics: Insights into Protist Evolutionary Biology" generated by an international workshop organised by Debashish Bhattacharya at the University of Iowa in 2004 (http://www.biology.uiowa.edu/workshop/Genomics_of_Eukaryotic_Microbes.html) are important in this respect, because the arguments they put forward are based on a wide phylogenetic perspective. This white paper proposed target protist species for whole genome sequencing from across the eukaryotic evolutionary tree based on a combination of phylogenetic and other criteria. These target species would not only fill in major gaps in the coverage of the eukaryotic tree, but would also include some marine groups such as chlorarachniophytes and foraminifers along with other groups, such as chytrids and paraphysomonads, that include some marine species.

The availability of a complete genome sequence is, of course, important if a particular organism is to be developed as a model, but the usefulness of the genome sequence is significantly enhanced if tools are available for the analysis of gene function. Functional genomics approaches can then provide insights into the novel biological characteristics of a particular group of organisms that are not attainable simply by analysis of the genome sequence. Moreover, the two aspects, genome sequence and functional tools, are related in as far as the availability of tools for gene function analysis provides an additional argument for genome sequencing. This was the case for classical models such as *Drosophila* and *Arabidopsis*, and has been an important argument for the selection of *Phaeodactylum tricornutum* for the second diatom genome-sequencing project (this project is nearing completion at the JGI).

Of course, it is difficult to imagine a full-scale model organism (permitting functional genomics) in all the major groups of the eukaryotic tree at this stage. It would, therefore, be useful to define a list of minimal requirements for a genomic model as an initial target. The Aquaculture Genome Coordinating Committee recently proposed such a list of requirements in a white paper

Photo 5.3: Rocky bottom community including various types of sponges. Sponges produce a range of chemical substances to fend off predators. Some recently extracted compounds have proved to be of pharmaceutical interest.

aimed at promoting genomics for aquacultured species (http://www.animalgenome.org/aquaculture/updates/NRSP8/White_Paper_2005_s.html). The following is a suggestion as to how these requirements could be adapted to the wider context of marine biology in general.

Proposed standards for the definition of a genetically enabled model species:

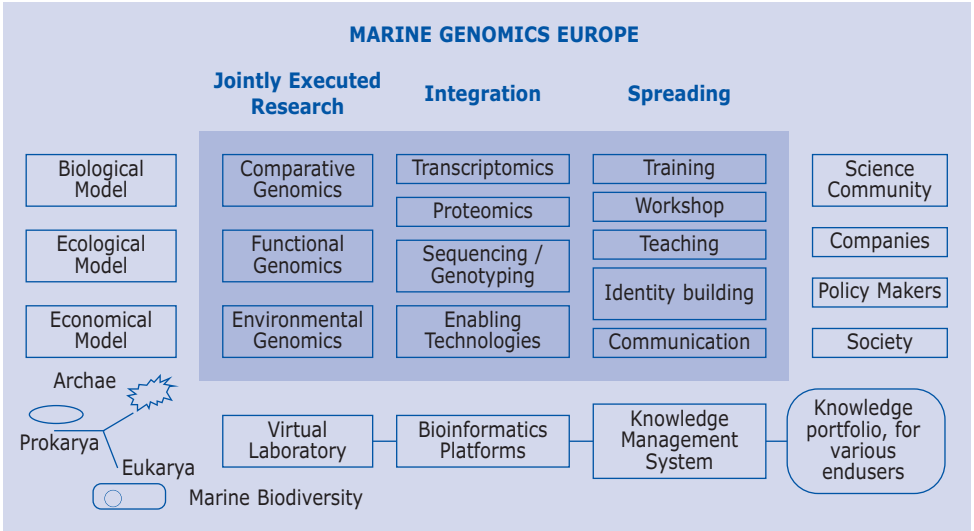
- An EST database of sufficient breadth (tissue and developmental/life cycle stages) and depth to represent most of the organism's transcriptome.
- Large insert, deep coverage BAC libraries.
- A microarray with the maximum set of unigenes identifiable.
- Sample sequencing of the genome sufficient to provide initial resources for gene identification, repeat content and polymorphism.
- Where classical genetic approaches can be applied, a linkage map with a resolution of <1 centiMorgan.
- A stable infrastructure, both physical and bioinformatic, to ensure the continued maintenance and public availability of genomic resources.

Another important objective in promoting the application of genomic approaches to marine systems is to federate the marine biology community with the aim of creating common projects and focusing efforts on a manageable number of target species. This type of activity is important for the creation of interest groups with a sufficient critical size around emerging model systems, and should also bring groups with key biological expertise but little experience of genomic approaches into contact with groups that possess technical expertise with genomic methodologies. In Europe a major effort is being made towards this end via the EU funded Network of Excellence "Marine Genomics Europe". This network will be described in more detail in the following section.

5.4. THE MARINE GENOMICS EUROPE NETWORK OF EXCELLENCE

The European Network of Excellence, "Marine Genomics Europe", is composed of 450 researchers from 45 institutions (118 laboratories or research groups) from 16 countries (website: http://www.marine-genomics-europe.org/). The aim of the network is to promote the application of high-throughput genomic approaches to the study of marine organisms. The network focuses on federating marine biology laboratories around common projects with the aim of

Figure 5.1: Marine Genomics Europe Network of Excellence

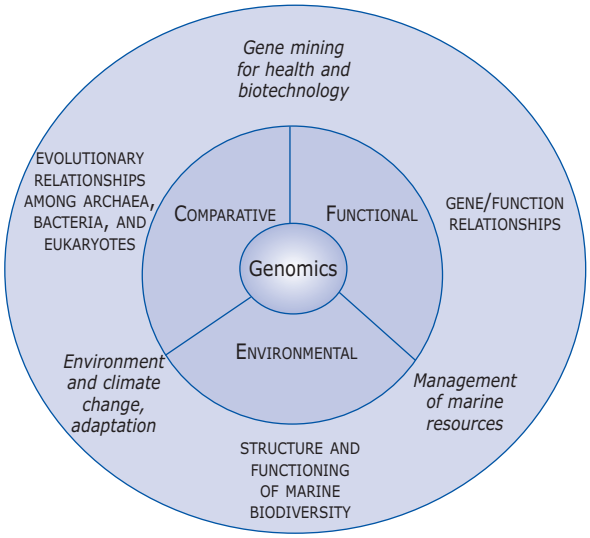


Schematic representation of the Marine Genomics Europe Network of Excellence showing how the networking objective ("integration") articulates with the research programme ("jointly executed research") and other activities such as interactions with policy makers and society ("spreading").

creating the critical mass necessary for the development of genomic approaches (figure 5.1). Examples of this type of integration include the networking of several genomics platforms, the establishment of a common bioinformatics facility, the launching of several large-scale "flagship" projects and an education and training programme for young scientists. Within the network, genomic approaches are being used to investigate a wide range of questions related to the functioning of marine ecosystems and to the biology of marine organisms.

The collaborative research (the jointly executed research programme) between the partners in the network is broken down into comparative, functional and environmental genomic approaches with emphasis, respectively, on comparison between genomes in a phylogenetic context, on high-throughput analysis of gene function, and on the application of genomic methodologies to the study of marine biodiversity (figure 5.2). These approaches are applied across four "nodes" that associate laboratories interested in particular groups of organisms: the microbial node, the algal node, the evolution, development and diversity node, and the fish and shellfish node. The jointly executed research is intended to generate data that can be exploited both by marine resource management programmes (prediction of global changes in marine populations, conservation of biodiversity, fisheries management and the improvement of aquacultured species), and by gene mining projects for health and

Figure 5.2: Structure of the joint research activity of the Marine Genomics Europe Network of Excellence[1]



1 The end point objectives, which include problems relevant to the management and exploitation of coastal biodiversity, are shown in italics.

Photo 5.4: View of the DNA sequencing laboratory at the Institute for Genomic Research in Gaithersburg, Maryland, USA. Rapid progress in DNA sequencing technology is accelerating the capacity to resolve the genome of marine organisms. The challenge ahead is to interpret the information these genomes contain.

biotechnology. Some additional details about the aims of the three different types of jointly executed research are given in the following paragraphs.

The comparative genomics programme aims to identify and focus on representative marine model organisms from across the different phyla of the tree of life. Some of the species identified are already entering the post-genomic stage, with work now concentrating on understanding the functions of the genes that make up the genome, and the list of new candidates is growing rapidly. The latter include organisms of major evolutionary importance either because of their phyletic novelty or because they possess gene families that are of particular interest for comparative analysis.

The functional genomics programme is exploring the complex relationships between endogenous and exogenous, biotic and abiotic stimuli and gene expression using a wide range of methodologies, including microarrays and proteomic and metabolomic approaches. These approaches are being developed for selected model organisms.

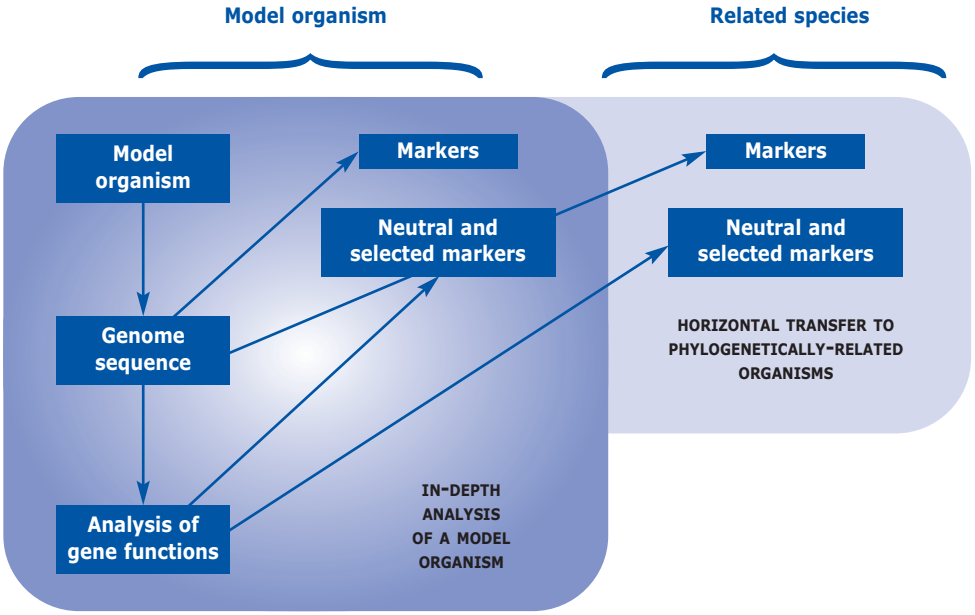
The environmental genomics programme aims to delineate the structure and dynamics of biodiversity in marine ecosystems. Particular efforts are being made to associate these studies with the functional genomics programme, to allow the

data obtained for these organisms to be exploited in the context of the different biosystems that are being studied. This articulation between work on model organisms and the direct application of genomic approaches in an ecological context is a very important feature of the work within the network. The following section will discuss this point in more detail, using work on the brown algae and, in particular, on the model species *Ectocarpus siliculosus* as an example.

5.5. MODEL ORGANISMS AS A MEANS TO APPLY GENOMIC METHODS TO MARINE QUESTIONS

The underlying causes of changes in marine biosystems can only be understood in the light of knowledge about the biology of the organisms making up those ecosystems. The most powerful way of establishing such knowledge is to obtain a deep (vertical) understanding of the biology of selected model organisms, and to use this knowledge as a base for the study of related organisms in the field (horizontal transfer). Hence, as discussed above, characterisation of a component organism of an ecological biosystem can be greatly facilitated by the availability of a well-characterised model organism within the same phylogenetic group (figure 5.3). The brown algae, for example, are dominant components of

Figure 5.3: Example of how in-depth analysis of a model organism can be exploited to develop tools for the analysis of biodiversity in ecosystems



rocky shore ecosystems. They are often the most abundant organisms in these ecosystems in terms of biomass, and in some coastal areas undersea kelp forests can rival terrestrial forests in extent and density. Our understanding of the biology of brown algae, however, is limited, particularly at the molecular level. Moreover, existing model systems, for example in the opistokont or green plant lineages, are of limited use because of the great evolutionary distances separating them from the brown algae (more than a billion years). Based on these and other arguments, we decided to develop a model brown alga that would be amenable to genomic and functional genomic approaches. To choose the model organism, we used selection criteria that focused both on genome size and on characteristics that allow genetic analysis, such as small size and the possibility to complete the life cycle and to carry out genetic crosses in the laboratory.

At the time that this project was initiated, several brown algae had been used as models to study certain aspects of brown algal biology. *Fucus* for example had been used extensively for cell biology approaches (Berger, Taylor and Brownlee 1994; Bouget, Berger and Brownlee 1998; Corellou et al. 2000; Goddard et al. 2000; Corellou et al. 2001; Brownlee, Bouget and Corellou 2001; Coelho et al. 2002) and expressed sequence tag (EST) data were available (Roeder et al. 2004) for the economically important *Laminaria digitata* (McHugh 2003). However, both of these organisms produce large thalli, and it is very difficult to complete their life cycles in the laboratory. Moreover, the genome of *L. digitata* was known to be very large (650 Mbp; Le Gall et al. 1993) and we determined that the genomes of *Fucus* spp. were even larger (more than 1000 Mbp, Peters et al. 2004; see also Kapraun 2005). In contrast, the genome sizes of members of the Ectocarpales have been shown to be significantly smaller (Stache 1993), as we were able to confirm (Peters et al. 2004). Also, the members of this order are smaller and more easily cultivated in the laboratory. Following a comparative study of several different members of the Ectocarpales, we proposed *Ectocarpus siliculosus* as model organism for the brown algae.

5.6. *ECTOCARPUS SILICULOSUS:* A MODEL ORGANISM FOR THE BROWN ALGAE

All stages of the *Ectocarpus* life cycle can be cultured in the laboratory in Petri dishes in natural or artificial seawater. The sexual life cycle, which involves an alternation between two separate generations (the sporophyte and the gametophyte), can be completed in three months and sexual crosses can be made by mixing gametes from male and female gametophytes. Other advantages of *Ectocarpus* as a model organism include its high fertility, the fact that large col-

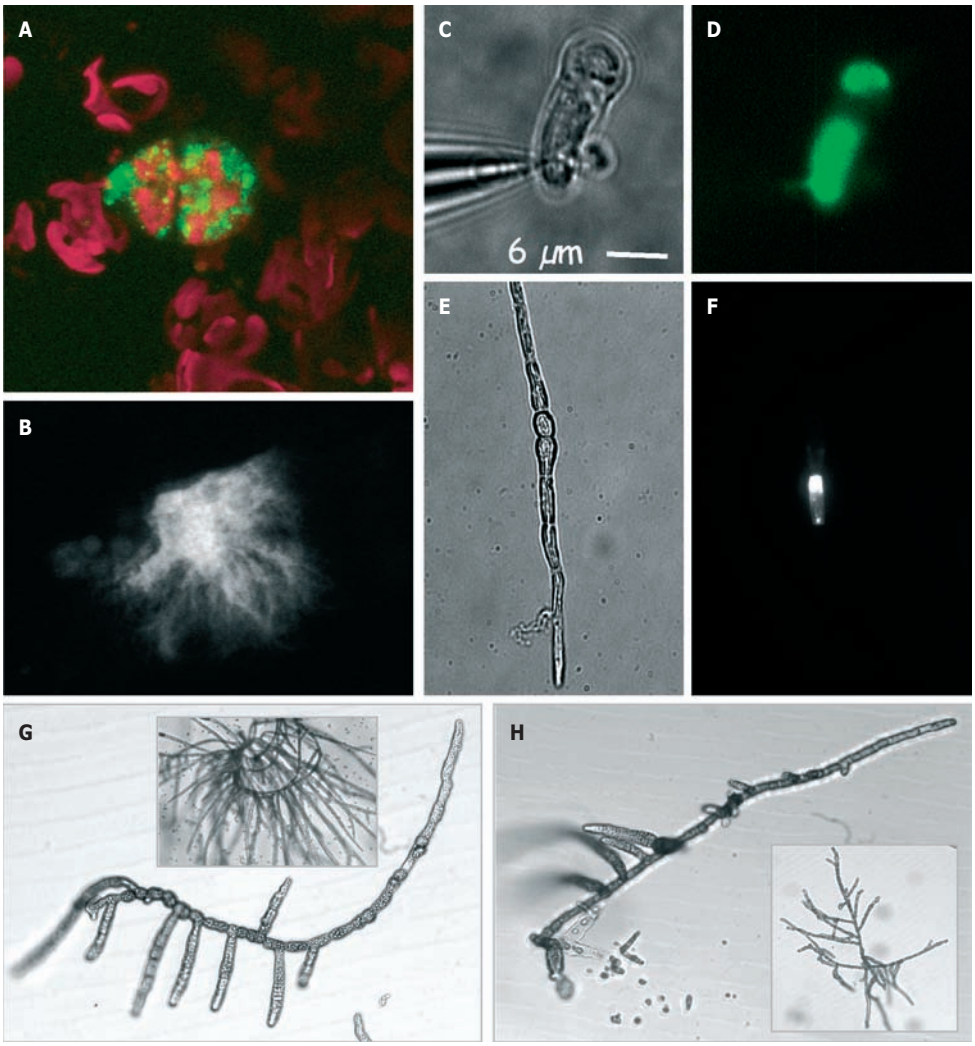


Photo 5.5: Illustration of some of the molecular tools being developed for the brown algal model species *Ectocarpus siliculosus*

A: Fluorescent markers loaded into an *Ectocarpus* filament cell using biolistics. The two fluorescent dyes, FITC (green) and Texas Red (red), were detected by confocal microscopy. The autofluorescence of the chloroplasts is also visible (violet). This method allows the introduction of fluorescent dyes for cell biology analyses and is also being used to optimise biolistic loading for the development of a transformation protocol (collaboration with Colin Brownlee, Marine Biological Association, Plymouth, UK). **B:** Example of a cell biology technique applied to *Ectocarpus.* Confocal microscope image of the actin cytoskeleton in a filament cell of a sporophyte. Cells were fixed and then stained with Alexa Fluor Phalloidin**. C and D:** Microinjection of a germinating *Ectocarpus* gamete with the fluorescent dye FITC (C, bright field; D, fluorescence). **E and F:** Development of an RNA interference protocol. Microinjection of a young *Ectocarpus* sporophyte (8 cells) with double stranded RNA and a fluorescent marker dye (Alexa Fluor 488). **G and H:** Example of a mutant *Ectocarpus* strain isolated from a UV irradiated population. The immature mutant alga (H) is compared with a wild type sporophyte at the same stage (G), later stages of development are shown inset.

lections of strains exist from temperate regions throughout the world, the close phylogenetic distance between the Ectocarpales and economically important seaweed within the Laminariales and the fact that it is, historically, one of the best studied brown algae (see Peters et al. 2004 and references therein).

In June 2004 a consortium of 35 laboratories submitted a proposal for complete sequencing of the *Ectocarpus* genome to the French sequencing centre Genoscope. This project, which was accepted in September 2004, proposed a 10x shotgun coverage of the genome (4,280,000 reads) plus 100,000 reads on cDNA sequences. The cDNA sequencing will aim to obtain a maximum of full-length cDNA sequences. The sequencing part of the genome project is expected to be completed in 2006. Access to *Ectocarpus* sequence data is currently available via a password-accessed website (http://genomer.sb-roscoff.fr/Ectocarpus/).

In parallel, considerable effort is going into the development of molecular tools for *Ectocarpus* including genetic transformation and RNAi technology (figure 5.4). Protocols have been established for both UV and chemical mutagenesis (EMS and MMS) and a pilot microarray has been produced and tested.

In April 2005, an international *Ectocarpus* meeting was held in Roscoff that attracted some 50 scientists from a large number of countries including Japan, Korea, the USA, Australia, Chile, France, Germany and Great Britain (http://www.sb-roscoff.fr/Esil2005prog.pdf). This meeting provided a forum for the coordination of the genome project, and allowed discussion of *Ectocarpus*-related research in a broad range of fields including developmental biology, cell biology, physiology, ecology and systematics, chemical ecology and biochemistry.

5.7. THE *ECTOCARPUS* GENOME PROJECT AND COASTAL BIODIVERSITY

How will the application of genomics to *Ectocarpus* help us to understand coastal biodiversity? Firstly, *Ectocarpus* will serve as a model to study how populations of brown algae adapt to their environment and, secondly, structural and functional knowledge about the *Ectocarpus* genome will be exploited for the study of species (such as *Fucus*) that play more important roles in coastal ecosystems. Work has already begun to learn more about the ecology of *Ectocarpus*, both at the level of the worldwide distribution and relatedness of *Ectocarpus* strains and at a more local, ecological level, building on earlier studies (see for example Stache 1989). This work involves several member laboratories of the *Ectocarpus* Genome Consortium.

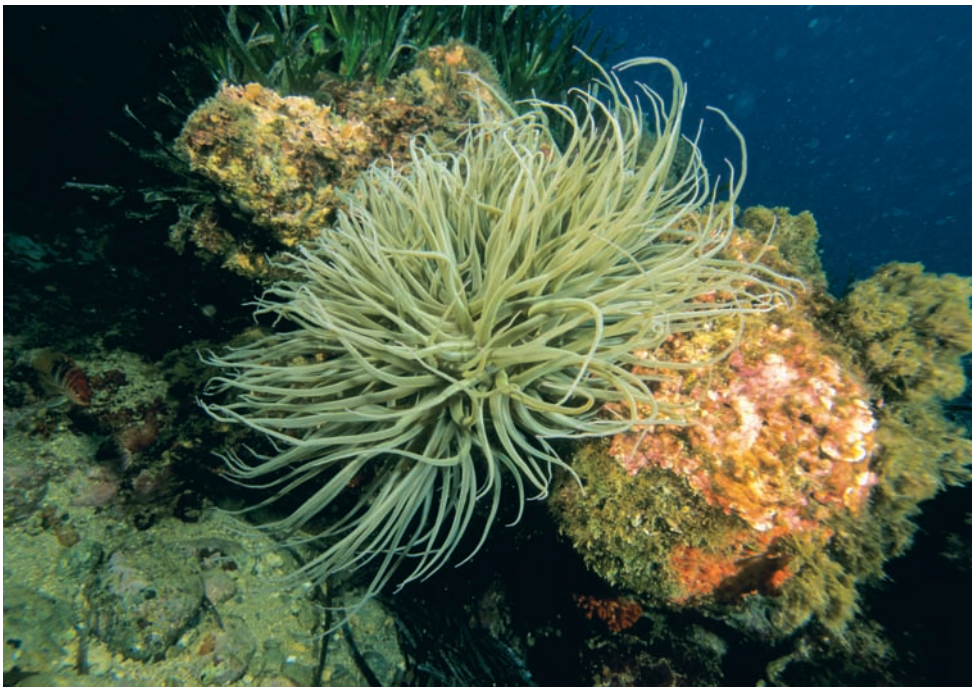


Photo 5.6: Actinia or sea anemones. Like other sessile organisms, sea anemones are among the marine creatures richest in bioactive substances and a potential source of molecules for pharmaceutical purposes.

Concerning the transfer of information from *Ectocarpus* to other brown algae, it is important to note that our understanding of the relative influence of factors such as selection, mutation, genetic drift and gene flow on the genetic composition of coastal biosystems has been significantly hampered by the limited availability of appropriate genetic markers, particularly nuclear markers. From a general point of view, the application of genomics to population genetics is providing new insights into the genetic and evolutionary processes affecting nuclear variation in numerous animal and plant models (Mitchell-Olds and Clauss 2002; Nordberg and Innan 2002; Maloof 2003). In this context, the *Ectocarpus* genome project has attracted the interest of several groups interested in brown algal microevolution, population genetics and systematics, including groups that will apply information from the *Ectocarpus* genome to studies of key brown algae within coastal ecosystems.

Indeed, the *Ectocarpus* genome will provide much needed genomic tools for the study of microevolutionary processes in brown algae, at both inter- and intraspecific scales. Through the analysis of the conservation and polymorphism of homologous loci in *Ectocarpus* and other brown algae, genome-wide surveys of

neutral and coding sequences in populations across species distributions will provide information about locus-specific evolutionary and genetic processes (e.g., selection, mutation, assortative mating, and recombination) as well as about demographic processes that affect the entire genome (e.g., genetic drift, gene flow and inbreeding; Luikart et al. 2003). The availability of homologous protein-coding sequences will provide a means to study adaptive molecular variation in this group. For example, comparing allele genealogies for multiple protein-coding loci among several species, differences in levels of polymorphism both within and among species may provide clues to the selective effects of functional processes involved in speciation (Mitchell-Olds and Clauss 2002). In addition, comparison of phenotypic and genotypic variation among individuals within a species will elucidate the genetic controls and limitations that influence not only organism distributions but also environmental interactions (Jackson et al. 2002). Genome-wide comparison of sequence variation within and among populations allows the identification of truly neutral loci (Luikart et al. 2003). These neutral loci are essential for robust, accurate estimation of population genetic parameters, such as effective population size and effective migration rates. In haploid-diploid species such as *Ectocarpus*, these parameters are of particular interest for the study of the consequences of the co-occurrence of free-living haploid and diploid stages on population genetic structure (e.g., Engel, Destombe and Valero 2004).

5.8. ADDITIONAL EMERGING AND FUTURE GENOMIC MODEL ORGANISMS FOR MARINE BIOSYSTEMS

The important constituents of the flora of coastal biosystems, in terms of biomass, include not only brown but also red and green algae and seagrasses. The seagrasses are angiosperms, so work on terrestrial angiosperms such as *Arabidopsis* and rice can potentially be exploited to study these organisms. The red and green algae, on the other hand, are very distantly related both to brown algae and to other existing model species (even if the situation is slightly better for marine green algae because they are members of the broader group of green plants, the viridiplantae, that also includes the angiosperms). The genome sequence of *Cyanidioschyzon merolae*, a unicellular red alga, has recently been reported (Matsuzaki et al. 2004), but this species is distantly related to the multicellular red algae and its genome is unusual and highly compact, perhaps as a result of its habitat in sulphate-rich, hot acidic springs. There is therefore a need to develop model organisms for the other algal groups, particularly the red algae, in a manner analogous to the development of *Ectocarpus* as a model for the brown algae.

A recent survey of potential macroalgal models proposed *Porphyra yezoensis* as a candidate model organism (Waaland, Stiller and Cheney 2004). The choice of *Porphyra yezoensis* was based on many of the same criteria that were behind the choice of *Ectocarpus* as a model for the brown algae. These included the size of the genome (approximately 300 Mbp; Kapraun et al. 1991), the facility with which cultures can be handled in the laboratory, the existence of mutants (Ohme and Miura 1988; Mitman and van der Meer 1994; Yan, Fujita and Aruga 2000), the development of methods for preparing and regenerating from protoplasts (Waaland et al. 1990) and the large body of information available in the literature concerning its biochemistry, physiology and culture. Additional arguments included the economic importance of *Porphyra* (which is the basis of the multi-billion dollar nori industry), its ecological importance in some coastal habitats, the existence of EST collections (Nikaido et al. 2000; Asamizu et al. 2003) and advances being made towards the development of genetic transformation (Cheney, Metz and Stiller 2001; He et al. 2001; Lin et al. 2001). Taken together, the arguments for developing *Porphyra yezoensis* as a genomic model are strong. and it is likely that a genome project will emerge for this organism in the near future. However, genomic sampling of the red algae should not be limited to the Bangiophyceae (both *P. yezoensis* and *C. merolae* are members of this class) but should also include at least one member of the other major class, the Floridiophyceae, which includes economically important agarophytes (agar producers such as *Graccilaria* spp.) and carageenophytes (carageenan producers such as *Kappaphycus* spp. and *Chondrus crispus*).

Genomics of green macroalgae currently appears to be less of a priority, probably because of the less obvious potential of these organisms for industrial applications compared to the red and the brown algae. Probably the best candidate from among the green macroalgae is *Ulva* (including taxa previously called *Enteromorpha*), because we have an extensive literature describing work on this species (Bryhni 1974; Fjeld and Løvle 1976; Reddy, Iima and Fujita 1992) as well as a collection of EST sequences, and because they can multiply rapidly to cause eutrophication-based coastal blooms (green tides). However, there is currently no genome programme for this organism.

The above discussion has been limited to macroalgae, as the most conspicuous constituents of coastal biosystems, but obviously there are other emerging model organisms of relevance to marine ecosystems. Among photosynthetic organisms in pelagic habitats, prokaryotes such as *Synechococcus* and *Prochlorococcus* have already been studied in some detail using genomic approaches. More recently, the eukaryotic prasinophyte *Ostreococcus tauri,* a picoplanktonic green alga widely distributed in the oceans, has emerged as a

model organism. *Ostreococcus tauri* possesses the smallest genome known for a free-living photosynthetic eukaryote (11.5 Mbp), housed within a very small cell (1.5 µm in diameter) with one chloroplast and one mitochondrion. The genome of *Ostreococcus* has been sequenced recently at the Laboratoire Arago (Banyuls, France, a member of the MGE network). It is highly compact with very few introns and short intergenic sequences. An important factor for genetic analysis is the genome's low level of genetic redundancy; gene families being very small, often consisting of a single gene. Genetic tools such as transformation and microarray analysis of gene expression are currently being developed for this organism.

In the metazoan lineage, a number of novel model species are emerging at key phylogenetic positions, as a result of the application of evo-devo approaches to understanding the evolution of the developmental complexity in this phylum. Many of these models are from the marine environment.

5.9. CONCLUSION

The needs of the marine community in terms of genomics differ somewhat from those of terrestrial biologists, and broad sampling approaches such as metagenomics may prove to be more relevant in this context. However, the development of in-depth genomics, applied to model organisms, is also important for the progression from a descriptive to a functional understanding of biosystems. One of the difficulties with regard to the development of such approaches is the phylogenetic diversity of marine biosystems. As a result of this diversity, existing model organisms are often too distantly related to be of use for a particular species under study. To address this problem, genome sequencing and other genomic approaches will need to be applied to selected species from across the tree of life to provide a better coverage of its inherent biodiversity. Further development of some of these species as full-blown genomic model species will then provide in-depth functional knowledge that can be applied to related species from the same phylogenetic group.

ACKNOWLEDGEMENTS

The Marine Genomics Europe Network of Excellence is funded by the European Union. The *Ectocarpus* genome is being sequenced at Genoscope, Evry, France. Work on *Ectocarpus* in the Roscoff laboratory is supported by Marine

Genomics Europe, the national programme (Groupement d'Intérêts Scientifiques) Marine Génomique, the Centre National de Recherche Scientifique and the Université Pierre et Marie Curie, Paris.

REFERENCES

ASAMIZU, E., M. NAKAJIMA, Y. KITADE, N. SAGA, Y. NAKAMURA, and S. TABATA. "Comparison of RNA expression profiles between the two generations of *Porphyra yezoensis* (Rhodophyta), based on expressed sequence tag frequency analysis". *J. Phycol.* 39 (2003): 923-930.

BEJA, O., L. ARAVIND, E. V. KOONIN, M. T. SUZUKI, A. HADD, L. P. NGUYEN, S. B. JOVANOVICH, et al. "Bacterial rhodopsin: evidence for a new type of phototrophy in the sea". *Science* 289 (2000): 1902-1906.

BERGER, F., A. TAYLOR, and C. BROWNLEE. "Cell fate determination by the cell wall in early *Fucus* development". *Science* 263 (1994): 1421-1423.

BOUGET, F. Y., F. BERGER, and C. BROWNLEE. "Position dependent control of cell fate in the *Fucus* embryo: role of intercellular communication". *Development* 125 (1998): 1999-2008.

BRYHNI, E. "Control of morphogenesis in the multicellular alga *Ulva mutabilis*. Defect in cell wall production". *Dev. Biol.* 37 (1974): 273-277.

BROWNLEE C., F. Y. BOUGET, and F. CORELLOU. "Choosing sides: establishment of polarity in zygotes of fucoid algae". *Semin. Cell Dev. Biol.* 12 (2001): 345-351.

CHENEY, D., B. METZ, and J. STILLER. "*Agrobacterium*-mediated genetic transformation in the macroscopic marine red alga *Porphyra yezoensis*". *J. Phycol.* 37, suppl. (2001): 11.

COELHO S. M., A. R. TAYLOR, K. P. RYAN, I. SOUSA-PINTO, M. T. BROWN, and C. BROWNLEE. "Spatiotemporal patterning of reactive oxygen production and Ca(2+) wave propagation in *Fucus* rhizoid cells". *Plant Cell* 14 (2002): 2369-2381.

CORELLOU F., S. R. BISGROVE, D. L. KROPF, L. MEIJER, B. KLOAREG, and F. Y. BOUGET. "A S/M DNA replication checkpoint prevents nuclear and cytoplasmic events of cell division including centrosomal axis alignment and inhibits activation of cyclin-dependent kinase-like proteins in fucoid zygotes". *Development* 127 (2000): 1651-1660.

CORELLOU F., C. BROWNLEE, B. KLOAREG, and F. Y. BOUGET. "Cell cycle-dependent control of polarised development by a cyclin-dependent kinase-like protein in the *Fucus* zygote". *Development* 128 (2001): 4383-4392.

DAVIS, R. H. "The age of model organisms". *Nat. Rev. Genet.* 5 (2004): 69-76.

ENGEL, C. R., C. DESTOMBE, and M. VALERO. "Mating system and gene flow in the red seaweed *Gracilaria gracilis*: Effect of haploid-diploid life history and intertidal rocky shore landscape on fine-scale genetic structure". *Heredity* 92 (2004): 289-298.

FJELD, A., and A. LØVLE. "Genetics of multicellular algae". In R. A. Lewin, ed. *The Genetics of Algae*. Oxford: Blackwell Scientific Publications, 1976, 219-235.

GODDARD H., N. F. MANISON, D. TOMOS, and C. BROWNLEE. "Elemental propagation of calcium signals in response-specific patterns determined by environmental stimulus strength". *Proc. Natl. Acad. Sci. USA* 97 (2000): 1932-1937.

HE, P., Q. YAO, Q. CHEN, M. GUO, A. XIONG, W. WU, and J. MA. "Transferring and expression of glucose oxidase gene – gluc in *Porphyra yezoensis*". *J. Phycol.* 37, suppl. (2001): 22.

JACKSON, R. B., D. R. LINDER, M. LYNCH, M. PURUGGANAN, S. SOMERVILLE, AND S. S. THAYER. "Linking molecular insight and ecological research". *Trends Ecol. Evol.* 17 (2002): 409-414.

KAPRAUN, D. F. "Nuclear DNA content estimates in multicellular green, red and brown algae: phylogenetic considerations". *Annals of Botany* 95 (2005): 7-44.

KAPRAUN, D. F., T. K. HINSON, and A. J. LEMUS. "Karyology and cytophotometric estimation of inter-and intraspecific nuclear DNA variation in four species of *Porphyra* (Rhodophyta)". *Phycologia* 30 (1991): 458-66.

LE GALL, Y., S. BROWN, D. MARIE, M. MEJJAD, and B. KLOAREG. "Quantification of nuclear DNA and G-C content in marine macroalgae by flow cytometry of isolated nuclei". *Protoplasma* 173 (1993): 123-132.

LIN, C. M., J. LARSEN, C. YARISH, and T. CHEN. "A novel gene transfer in *Porphyra*". *J. Phycol.* 37, suppl. (2001): 31.

LUIKART, G., P. R. ENGLAND, D. TALLMON, S. JORDAN, and P. TABERLET. "The power and promise of population genomics: from genotyping to genome typing". *Nature Rev. Genet.* 4 (2003): 981-994.

MALOOF, J. "Genomic approaches to analyzing natural variation in *Arabidopsis thaliana*". *Curr. Opin. Genet. Dev.* 13 (2003): 576-582.

MATSUZAKI M., O. MISUMI, T. SHIN-I, S. MARUYAMA, M. TAKAHARA, S. Y. MIYAGISHIMA, T. MORI, et al. "Genome sequence of the ultrasmall, unicellular red alga *Cyanidioschyzon merolae* 10D". *Nature* 428 (2004): 653-657.

MCHUGH, D. J. *A guide to the seaweed industry. FAO Fisheries Technical Paper*, t. 441. Rome: FAO, 2003.

MITCHELL-OLDS, T., and M. J. CLAUSS. "Plant evolutionary genomics". *Curr. Opin. Plant Biol.* 5 (2002): 74-79.

MITMAN, G. G., and J. P. VAN DER MEER. "Meiosis, blade development, and sex determination in *Porphyra purpurea* (Rhodophyta)". *J. Phycol.* 30 (1994): 147-159.

NIKAIDO, I., E. ASAMIZU, M. NAKAJIMA, Y. NAKAMURA, N. SAGA, and S. TABATA. "Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*". *DNA Res.* 7 (2000): 223-227.

NORDBERG, M., and H. INNAN. "Molecular population genetics". *Curr. Opin. Plant Biol.* 5 (2002): 69-73.

OHME, M., and A. MIURA. "Tetrad analysis in conchospore germlings of *Porphyra yezoensis* (Rhodophyta, Bangiales)". *Plant Sci.* 57 (1988): 135-140.

PETERS, A. F., D. MARIE, D. SCORNET, B. KLOAREG, and J. M. COCK. "Proposal of *Ectocarpus siliculosus* as a model organism for brown algal genetics and genomics". *J. Phycol.* 40 (2004): 1079-1088.

RAY, G. C. "Ecological diversity in coastal zones and oceans". In E. O. Wilson, ed. *Biodiversity*. Washington, D.C.: National Academy Press, 1988, 36-50.

REDDY, C. R. K., M. IIMA, and Y. FUJITA. "Induction of fastgrowing and morphologically different strains through intergeneric protoplast fusions of *Ulva* and *Enteromorpha* (Ulvales, Chlorophyta)". *J. Appl. Phycol.* 4 (1992): 57-65.

ROEDER, V., J. COLLÉN, S. ROUSVOAL, E. CORRE, L. LEBLANC, and C. BOYEN. "Identification of stress genes transcripts in *Laminaria digitata* (Phaeophyceae) protoplast cultures by expressed sequence tag analysis". *Journal of Phycology* 41 (2005): 1227-1235.

STACHE, B. "Sexual compatability and species concept in *Ectocarpus siliculosus* (Ectocarpales, Phaeophyceae) from Italy, North Carolina, Chile and New Zealand". In D. J. Garbary and G. R. South, eds. *NATO ASI Series* G-22. *Evolutionary biogeography of the marine algae of the North Atlantic*. Berlin/Heidelberg: Springer-Verlag, 1989.

STACHE, B. "Kreuzungsexperimente bei Braunalgen. Vergleich von lokalpopulationen des kosmopoliten *Ectocarpus siliculosus*". Ph. D. Thesis. Konstanz: Hartung-Gorre, 1993.

VENTER, J. C., K. REMINGTON, J. F. HEIDELBERG, A. L. HALPERN, D. RUSCH, J. A. EISEN, D. WU, et al. "Environmental genome shotgun sequencing of the Sargasso Sea". *Science* 304 (2004): 66-74.

WAALAND, J. R., L. G. DICKSON, and B. A. WATSON. "Protoplast isolation and regeneration in the marine red alga *Porphyra nereocystis". Planta* 181 (1990): 522-8.

WAALAND, J. R., J. W. STILLER, and D. P. CHENEY. "Macroalgal candidates for genomics". *J. Phycol.* 40 (2004): 26-33.

YAN, X., Y. FUJITA, and Y. ARUGA. "Induction and characterization of pigmentation mutants in *Porphyra yezoensis* (Bangiales, Rhodophyta)". *J. Appl. Phycol.* 12 (2000): 69-81.

LIST OF PHOTOS AND ILLUSTRATIONS

ABOUT THE AUTHORS

J. Mark Cock is a group leader at the Roscoff Biological Station and a member of the French National Centre for Scientific Research (CNRS). He is working on brown algal developmental biology using *Ectocarpus siliculosus* as a model organism.

e-mail: cock@sb-roscoff.fr

Delphine Scornet is a research technician with the French National Centre for Scientific Research (CNRS). She is currently working at the Roscoff Biological Station on the *Ectocarpus siliculosus* genome project.

e-mail: scornet@sb-roscoff.fr

Susana Coelho is a research scientist working for the National Centre for Scientific Research (CNRS) at the Roscoff Biological Station, France. She is currently studying aspects of the cell biology of the filamentous brown alga *Ectocarpus siliculosus* with particular attention to the genetic control of haploid-diploid life cycles.

e-mail: coelho@sb-roscoff.fr

Bénédicte Charrier is a research scientist at the French National Centre for Scientific Research (CNRS), currently working at the Roscoff Biological Station. Her research focuses on the developmental mechanisms leading to the final architecture of the filamentous brown alga *Ectocarpus siliculosus*, using mainly molecular and cellular approaches.

e-mail: charrier@sb-roscoff.fr

Catherine Boyen is a National Centre for Scientific Research (CNRS) scientist and is both director of the Marine Plants and Biomolecules Laboratory in Roscoff, France, and coordinator of the Marine Genomics Europe Network of Excellence. Her work focuses on stress responses in both brown and red macroalgae.

e-mail: boyen@sb-roscoff.fr

Akira F. Peters is a phycologist with a particular interest in brown algae. His most recent work, at the Roscoff Biological Station, has centred on developing *Ectocarpus siliculosus* as a model organism for this group.

e-mail: akirapeters@hotmail.fr