THE TARA OCEANS EXPEDITION

2009-2012

Scientific Project

Eric Karsenti Director

Proposal summary

It is now unequivocal that global warming is occurring at a rate 100 times greater than documented in geological studies. The major part (>50%) of global primary biomass production occurs in the upper 200 m surface water layer of the oceans (the photic zone) and this also drives most of the global elemental cycling involved in climate regulation. Key actors in these elemental cycling processes are thought to be plankton protists, which produce oxygen and recycle carbon dioxide. However, protists are embedded in a network of plankton organisms that range from viruses to fish larvae and the complex, dynamic food webs that these organisms form remain largely un-investigated.

TARA OCEANS will analyze plankton ecosystems in relation to physicochemical conditions throughout the world's oceans, assessing their adaptation to and feedback on a rapidly changing earth system. Large scale genomics and metagenomics projects coupled to newly developed high-resolution and high-throughput imaging methods will allow quantitative studies of plankton ecosystems and identification of the plankton composition within these ecosystems. Integrative bio-informatics analyses of the imaging, genomes, physicochemical and climate data generated in the project, will be used to assess plankton biodiversity and activity in the diverse ocean ecosystems sampled. All of the data generated through the project will go to form an open-source multidimensional bio-oceanographic database that will allow generating predictive models of the spatio-temporal evolution of plankton ecosystems.

The results of this project will have outstanding implications for our understanding of early life evolution, global biogeochemical cycles and spatio-temporal evolution of the earth climate.

Extended Synopsis of the scientific project

Models that calculate the evolution of global warming are entirely based on geophysical data[1], and incorporate strong assumptions about the carbon cycle [2]. The carbon cycle is determined by complex processes involving the evolution and turnover rate of living organisms. The assumptions in the geophysical models are very basic, simply because there are no global biological data about the most important oceanic ecosystems (qualitatively and quantitatively) contributing to the regulation of gaseous equilibria in the atmosphere.

Indeed, the early Earth atmosphere was almost devoid of O_2 and its origin has probably begun with the appearance of a single clade of bacteria in the oceans that became the progenitor of all photosynthetic eukaryotes from algae to higher plants [3] [4]. However, over long geological periods, respiration produces CO_2 out of the oxygen. Therefore, stable accumulation of oxygen in the atmosphere cannot occur without removal of the CO_2. This is achieved by the burial of organic matter in the Earth's crust, mostly in the form of sediment rocks, through the sedimentation properties of a class of single cell eukaryotic algae present in the phytoplankton: the diatoms.

There is a clear correlation between oxygen and CO_2 concentration fluctuations in the atmosphere and global earth climate changes over geological times. This is also strongly correlated with the evolution of non-motile, dense phytoplankton taxa like coccolithophorids and diatoms over the past 205 million years [5]. Today, there are about 25000 morphologically defined forms of phytoplankton distributed among at least 8 phyla. Among these, diatoms account for approximately 40% of oceanic primary productivity and over 50 % of organic carbon burial in marine sediments [6]. Figure 1 shows a few examples of phytoplankton cells (from Falkowski et al. 2004).

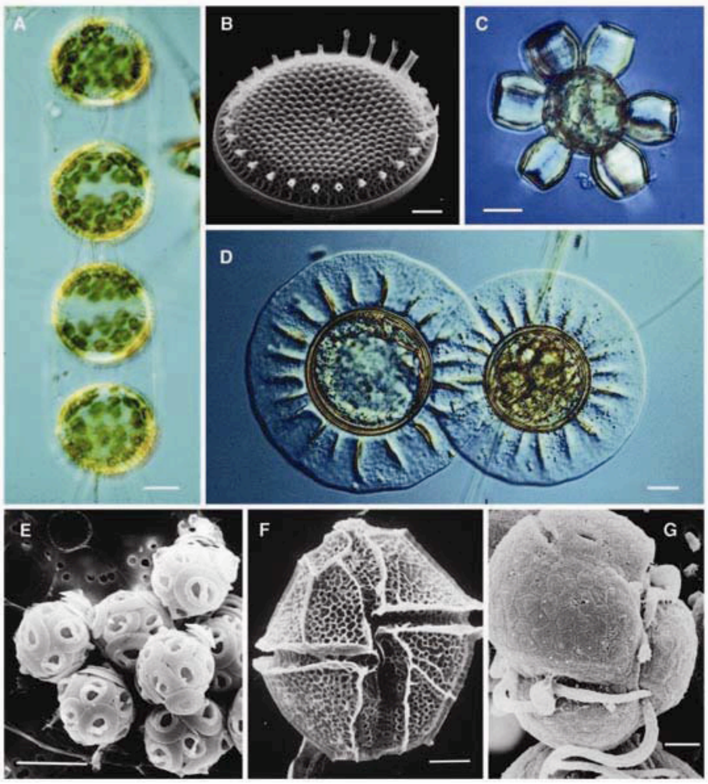

Figure 1. Some examples of unicellular protists.

The key problem associated with global warming is therefore CO2 kinetics: how much CO2 can the phytoplankton bury in a given amount of time, where in the sea such organisms will proliferate, which ones will dominate and how fast such systems can react when exposed to increased water temperature, pH changes or other effects associated with increasing temperatures of the oceans? [7]. Because climate warming will not occur everywhere simultaneously or with the same intensity, it is hard to know how the ecosystems will react as a whole and how much mixing may occur between various oceanic provinces. Some mathematical models of the spatio-temporal evolution of marine micro-organism ecosystems are being built, but they are limited by data about oceanic microscopic ecosystems composition [8]. Indeed, even if phytoplankton is the dominant actor in CO2 removal from the ocean, it is embedded in a complex ecosystem including viruses, bacteria, copepods and fish larvae. Phytoplankton is eaten by copepods as well as many other large predators, and bacteria concentrations are subject to regulation by virus attacks [9] just like protists [10].

Recent experiments in which an entire ecosystem from the North Sea has been maintained for almost 7 years in a mesocosm under constant conditions, clearly show that under such conditions the system has several characteristics of chaos, the ability of mathematical models to predict the quantitative evolution of species compositions in the future being limited to 30 days [11]. The authors conclude from their study that, like weather forecast, it will probably be impossible to predict the dynamic evolution of planktonic ecosystems beyond a couple of weeks. However, because of environmental constraints which are not present in this mesocosm it is unlikely that this will be the case in a wider context. Chaos is a good way to generate dynamic systems on which regulatory parameters can impose long-term stable behaviour (emergent stability). Therefore this study, rather than showing that the evolution of marine ecosystems is unpredictable, indicates that we should look into the mathematics of these systems using *in situ* data, combining quantitative oceanographic and biological data [8]. Moreover, short term dynamics may not be a proper reflection of the long term stability of an ecosystem. We clearly need models describing relatively large and stable ecosystems evolving on time scales of years or more [12]. Another recent study on the causes and projections of abrupt climate-driven ecosystem shifts in the North Atlantic illustrates this point quite well [13]. In fact, ecologists know very well that in a given oceanic domain the ecosystem is relatively stable. In their recent study, Beaugrand et al. [13] used data from 1958 to 2005 on phytoplankton abundance (using a phytoplankton color index obtained both from satellite data and a continuous plankton recorder), zooplankton abundance and diversity (copepod measurements) and cod abundance, which is well documented in this area. This was correlated with sea temperature and O2 concentration changes. They found that, changes in ecosystem structures depend largely on whether they are close to critical thermal boundaries. In their case (the North Atlantic Ocean) this boundary is about 9-10oC and they could follow spatial changes in species distribution (including cod abundance) associated with a temporal northward shift in the position of this boundary. This study also addresses directly another classical question of Ocean Biology: is everything everywhere? The answer is clearly not at the very least in terms of ecosystems structures, but probably also in terms of absolute species distributions [14].

If we are to make progress towards the modeling of ocean ecosystems dynamics to couple them with oceanic fluxes and climate dynamics, we clearly need extremely good integrated worldwide oceanographic metadata. But we also need the right biological data. Most studies are being done on relatively short expeditions involving large oceanographic vessels and targeted to the study of a restricted number of organism. These expeditions bring extraordinarily important data, but they are not homogeneous and difficult to use in mathematical modeling of global world-wide processes. Most of the time the biological data used originates from global satellite observations that report mostly about pigment and therefore photosynthetic organisms present on the surface of the oceans, precluding the incorporation of data about viruses and zooplankton and life in deeper oceans.

A question therefore arises: What kind of data do we really need, at which resolution, at which level of complexity and on which geographical scale, to begin to build models of the dynamics of ocean plankton ecosystems in relation to climate dynamics?

It is essential in the coming years to develop new high throughput quantitative methods of imaging and to couple them with the adequate genomic tools in order to characterize all the various species present in the oceanic ecosystems and quantify them. It will also be important to correlate these detailed results with more coarse-grained data obtained from floats and satellites in order to extend detailed snap shot results to dynamics over several years.

Another important question deals with the regions of the oceans that should be sampled. Oceans at the planetary level represent a huge territory and cannot be sampled exhaustively. It is important to carefully choose the areas to be sampled so that a maximum amount of information can be obtained. Regions of strong interest at the moment would include places where there are large temperature or pH gradient or regions in which the climate is changing rapidly versus regions where global warming is less obvious at the moment. There is also the question of whether coastal versus open ocean waters be sampled? These are just a few of the considerations that need to be made.

The Tara Oceans Project

The expedition voyage will cross the major oceanic gyres in the Indian, Atlantic and Pacific oceans, it will sail through areas that are warming up at a high rate in the north west passage while the southern hemisphere warming shows a slower rate. TARA OCEANS will sample the Portuguese, South African coast, South American coast and North American coast upwellings, the Indonesian biodiversity gradient and oligotrophic waters of the Pacific Ocean. In addition, the cruise will have legs dedicated to poorly studied coral reef ecosytems (studies not detailed in this application). The first part of the cruise in the Mediterranean Sea will be used to test the equipments and sampling routine on board (figure 2).

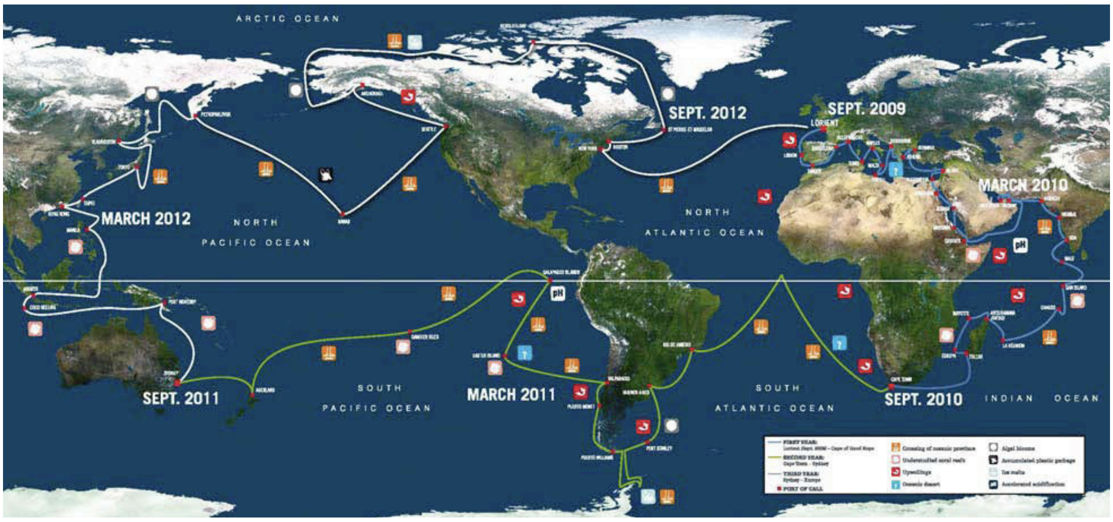

Figure 2. Path of the Tara Oceans expedition. The vessel will leave Lorient, France, in September 2009 and sail across all of the oceans and major hydrographic provinces of the world over three consecutive years.

In order to obtain the relevant data required to analyze the diversity of planktonic ecosystems it was necessary to develop a concept including a sampling plan regime, the nature of the data to be recorded, sample collection storage and processing methods, tools to analyze the samples and most importantly, tools to archive and analyze the data in an integrative way. A group of scientists led by E. Karsenti have met regularly over the past year to develop a new integrative, systemic approach to Ocean life study that could be named “Bio-Oceanography”. This involved a specific sampling plan and data acquisition on board and an organization on land store and analyze the samples, treat and analyze the oceanographic and biological data, in an integrated and synthetic way.

Sample collection, processing, and storage.

At each of the sampling stations during the 25 legs of the cruise a recorded volume of *sub-surface* (SS) and *Deep Chlorophyll Maximum* (DCM) water, as well as other depths whenever possible down to 1000 m, will be sampled using nets, pumps, and bottles. The organisms will be split into 4 size-fractions, 1-3; 3-20; 20-200; and 200-2000 µm, using a newly designed *Gravity Plankton Sieving System* (GPSS) developed by one of the team members (figure 3).

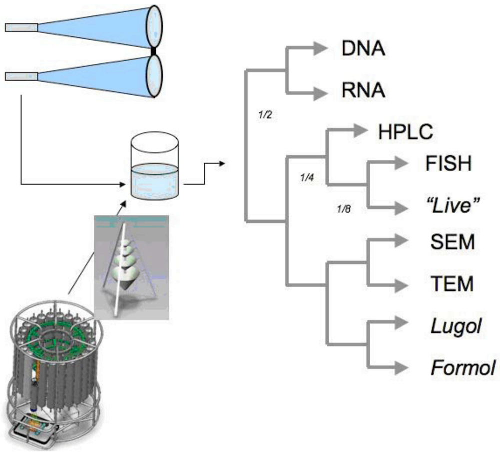

Figure 3: ***sampling strategy for morpho-genetic analyses of pelagic plankton species during the Tara-Oceans expedition*** (global ocean photic zone, ~375 stations). Top: The "*Gravity Plankton Sieving System*" recently developed will be used to gently separate the cells from large volumes of seawater sampled using Niskin bottles, pumping system, and plankton nets into 4 size-fractions covering the entire range of unicellular eukaryotes. Samples will then be processed for various genetic and morphological analyses (on the right).

A complete suite of chemical and physical data characterizing the sampled water will be recorded in parallel as well as the photosynthetic activity level in the subsurface water (FRRF) and its organism composition (Seaflow). Each size-fraction will be divided and processed to generate material for further laboratory DNA, RNA, and morphological analyses, as well as for archive material. After collection, samples will be immediately labelled, recorded in the onboard database, appropriately stored (liquid-N, freezer, fridge, dried), and as soon as possible shipped via a courier (most often directly from each port of call) to TARA OCEANS headquarters at EMBL. The uniqueness of this approach consists in rigorously applying identical sampling protocols which target the entire plankton diversity from viruses to zooplankton, and using DNA, RNA, and phenotypic data circum-globally.

Imaging

Since Tara Oceans will bring back species ranging from viruses to metazoans, one of the challenges is to image all of them using state of the art microscopy techniques. Viruses and bacteria will be imaged by electron microscopy and Cryo-EM. Already existing low resolution underwater imaging (UVP, VPR and CETO) will be used for relatively large organisms (in the mm-cm size range), and live imaging will be done on board (FlowCam, µm-mm) and higher resolution images taken with the new 3D microscope (Single Plane Illumination Microscope, SPIM) developed at EMBL. Moreover, samples will be imaged using two high throughput screening platforms on land [15, 16]. The first one will do metazoan morphological screening and recognition using a library of morphotypes (obtained with the SPIM [17] and X-ray microtomography). The second one will do protist morphological screening and genotype analysis by combining FISH and high throughput microscopy [15, 16]. In this way, we will fulfill most of the requirements that marine biologists were previously requesting for high throughput species identification [18] [19]. Finally, the use of modern image segmentation algorithms in 2D and 3D will be used to analyze on board and on land image databases.

Sequencing

Recent advances in high-throughput sequencing technologies make projects of cutting-edge nature like TARA OCEANS feasible for the first time. However, certain key methodological and strategic issues must first be addressed. Therefore although samples will be taken for sequencing from the outset of the expedition, the first two years will be dedicated to the establishment of a firm experimental basis for the subsequent analysis of all of the samples. Currently reference genomes of protists, are being sequenced. The specific technical objectives are to:

-Develop standard protocols for unbiased extraction of nucleic acids from mixed communities of pelagic protists;

-Test and optimize different high-throughput sequencing strategies (group-specific PCR, whole DNA and whole RNA shotgun sequencing);

-Generate cutting-edge bioinformatics tools for efficient processing of huge quantities of sequence data;

-Provide a framework for interpretation of massive environmental sequencing datasets by sequencing and automatically annotating draft genomes of a diverse set of taxonomically defined protists from culture collections. There are already several virus and bacterial reference genomes sequenced. Metazoan genomics will not be done, at least initially.

Once this is done massive sequencing will start of all the samples collected. All this work will be coordinated and largely done by the Genoscope in France as well as other laboratories in Spain and the United States.

A new Integrated Ocean Database

Integrating this wealth of data ranging from oceanographic and satellite data to images and sequences is a challenge. The database will have to be accessible to researchers around the world as well as the public and be able to establish correlations between the various sets of data it hosts.

The data management infrastructure will be called the Integrated Ocean Database (IOD, Figure 4). It will be based on a “Central Metadata Registry” (CMR), on “Designated Data Archives” (DDA) and an “Integrated Information System” IIS The flow of data [from direct measurements made onboard Tara and onboard Satellites, and from the large and diversified suite of analyses that will be made back on-land using material from samples collected onboard Tara will be structured around “unique sample identifiers”, and stored in the CMR. Raw data will be stored in DDAs, like NASA and ESA for satellite data, LEGOS-DB for near real-time ThermoSalinoGraph (TSG) data, EBI for genetic and molecular data, EMBL for microscopic imagery, and WDC-MARE for plankton taxonomy, physical oceanography and biogeochemistry.

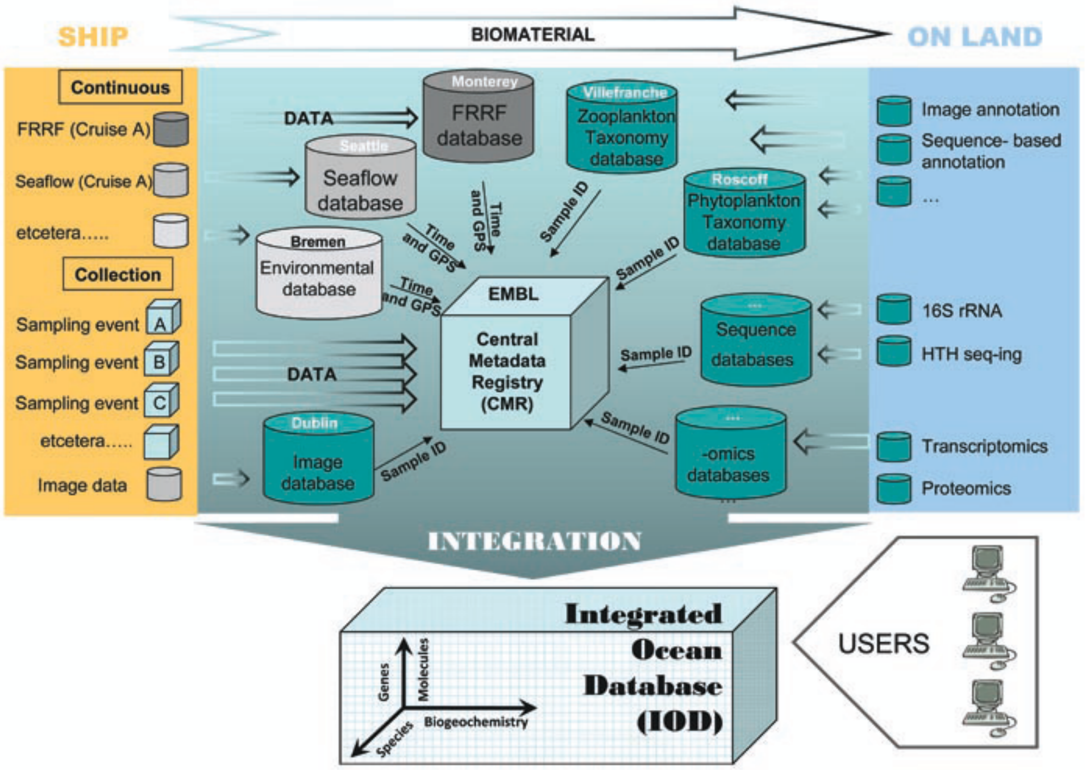

Figure 4. Data flow scheme. Data collected on-board are regularly transferred to the on-shore sample database and to the Dedicated Data Archives (DDA). The DDAs are named after the data collected (image, FRRF, etc). The web-based sample index (Sample DB), later complemented by the experiment database, will interlink all datasets using time, GPS and sample ID and will be accessible via internet. Sample and experiment components will constitute the Central Metada Registry (CMR). Participating databases will store and make available raw and processed data (e.g. via ftp). Data processing services will be provided either locally, by a data generator, or distantly, by a specialized service provider (e.g. EBI). Upon completion of processing and annotation of individual data types, the results will be incorporated and brought into the Integrated Ocean Database either by a standard format or through a pipeline.

The IOD should include a tool to interlink and group all the different datatypes to make a unique resource available for the scientific community and general public. The IIS will do that by linking all data aggregated and processed by the DDAs, based on the common georeferences (latitude, longitude, depth, date and time) provided by the CMR. The IIS will provide online data discovery tools providing references to the original datasets, meta-analysis tools to filter data, based on queries, and data warehouses designed for mass downloads of data. The whole system described above will form the IOD that will serve also as an inventory of knowledge gained upon analysis of the data as well as providing references to the original datasets.

This IOD will allow to determine ecosystems structures at various stations. Because we will have data from a given station at a given depth, we will be able to determine the complexity of this particular water volume with unprecedented data density along the water column and model it around the globe by extending analysis and modeling to all the stations. This will provide a baseline of ocean biodiversity and ecosystem dynamics of the world Oceans.

Expected Integrated Outcome

The approach and methods described above will first lead to an unprecedented discovery of new plankton species. This will also generate quantitative data about the relative abundance of organisms like viruses, bacteria, protozoa and small metazoans in various oceanic environments. Because we will use genomic methods, we will have access to the quantification of species in these ecosystems. Combined with imaging, the molecular data will bring new information about the correlation between shape and genome structures and most importantly about complex symbiotic events that occur in the sea as well as about horizontal gene transfer processes. Also, correlations between the presence of specific viruses and potential target organisms within various physico-chemical oceanic environments will provide important informations about the way oceanic ecosystems adapt to physico-chemical constraints.

The voyage will cross oceanic regions with very different pH (figure 5). We plan to study first the samples from ecosystems of areas having important pH gradients under various latitudes. Presently we do not know how pH variations may affect the susceptibility of protists or bacteria to virus attacks or how these pH variations may affect protists health. Because we will have access to recording of many other physico-chemical parameters, it is almost certain that some very interesting ecosystem patterns (e.g. in low oxygen systems or according to pH) will emerge. This should also tell a lot about the principles governing the evolution of microscopic life in the oceans. On the relatively long term (over 3 to 5 years from the beginning of the expedition) enough data should have been accumulated in the IOD to build dynamic models of these ecosystems and begin to explore the relationship between climate and oceanic ecosystems evolution.

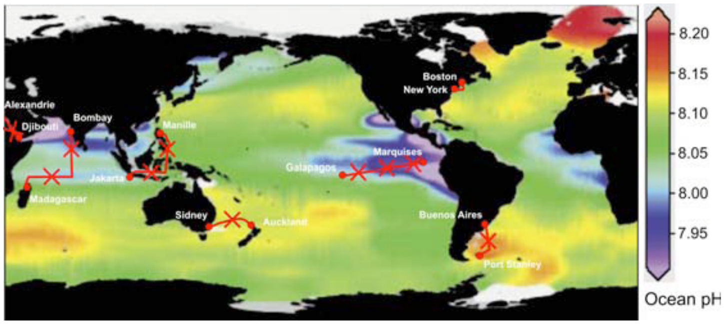

Figure 5: ***Location of 6 Tara-Oceans legs and 10 stations proposed, on a map of modern surface ocean pH.*** The mosaic of various pH domains is related principally to ocean circulation patterns.

Finally, because new species will be discovered in large numbers, unexpected spinoff results may be obtained in terms of new systems for cell and developmental biology and products for pharmaceutical applications.

Beyond the scientific research that will be carried out, the TARA OCEANS expedition will have a very large public outreach program. It will include worldwide frequent press releases, an interacting website as well as several high-density prime time television broadcasts (contracts already signed with France Television, under negotiation with German and Italian televisions).

Previous expeditions during the 19th century have had outstanding and long-lasting impacts on science and society (the expedition of the Beagle, and that of the Challenger for example). We expect the output of TARA OCEANS to become one of the most important oceanic metadata resources of the 21st century both for the scientists and through the importance of its legacy in terms of public awareness of the functioning of our planet.

References

1.Stainforth, D.A., et al., *Uncertainty in predictions of the climate response to rising levels of greenhouse gases.* Nature, 2005. 433(7024): p. 403-6.

2.Allen, M.R., et al., *Warming caused by cumulative carbon emissions towards the trillionth tonne.* Nature, 2009. 458(7242): p. 1163-6.

3.Falkowski, P.G., et al., *The evolution of modern eukaryotic phytoplankton.* Science,

2004. 305(5682): p. 354-60.

4.Falkowski, P.G. and Y. Isozaki, *Geology. The story of O2.* Science, 2008. 322(5901): p. 540-2.

5.Falkowski, P.G., et al., *The rise of oxygen over the past 205 million years and the evolution of large placental mammals.* Science, 2005. 309(5744): p. 2202-4.

6.Rabosky, D.L. and U. Sorhannus, *Diversity dynamics of marine planktonic diatoms across the Cenozoic.* Nature, 2009. 457(7226): p. 183-6.

7.Huisman, J., et al., *Reduced mixing generates oscillations and chaos in the oceanic deep chlorophyll maximum.* Nature, 2006. 439(7074): p. 322-5.

8.Follows, M.J., et al., *Emergent biogeography of microbial communities in a model ocean.* Science, 2007. 315(5820): p. 1843-6.

9.Hewson, I. and J.A. Fuhrman, *Characterization of lysogens in bacterioplankton assemblages of the southern California borderland.* Microb Ecol, 2007. 53(4): p. 631-8.

10.Suttle, C.A., *Viruses in the sea.* Nature, 2005. 437(7057): p. 356-61.

11.Beninca, E., et al., *Chaos in a long-term experiment with a plankton community.* Nature, 2008. 451(7180): p. 822-5.

12.Azaele, S., et al., *Dynamical evolution of ecosystems.* Nature, 2006. 444(7121): p. 926-8.

13.Beaugrand, G., et al., *Causes and projections of abrupt climate-driven ecosystem shifts in the North Atlantic.* Ecol Lett, 2008. 11(11): p. 1157-68.

14.Medlin, L.K., *If everything is everywhere, do they share a common gene pool?* Gene, 2007. 406(1-2): p. 180-3.

15.Pepperkok, R. and J. Ellenberg, *High-throughput fluorescence microscopy for systems biology.* Nat Rev Mol Cell Biol, 2006. 7(9): p. 690-6.

16.Neumann, B., et al., *High-throughput RNAi screening by time-lapse imaging of live human cells.* Nat Methods, 2006. 3(5): p. 385-90.

17.Reynaud, E.G., et al., *Light sheet-based fluorescence microscopy: more dimensions, more photons, and less photodamage.* HFSP J, 2008. 2(5): p. 266-75.

18.Benfield, M.C., et al., *RAPID Research on Automated Plankton Identification.* Oceanography, 2007. 20: p. 172-186.

19.Culverhouse, P.F., et al., *Automatic Image Analysis of Plankton: Future Perspectives.* Marine Ecology Progress Series, 2006. 312: p. 297-309.