

Chapter 4

Infrastructures for generating and sharing knowledge

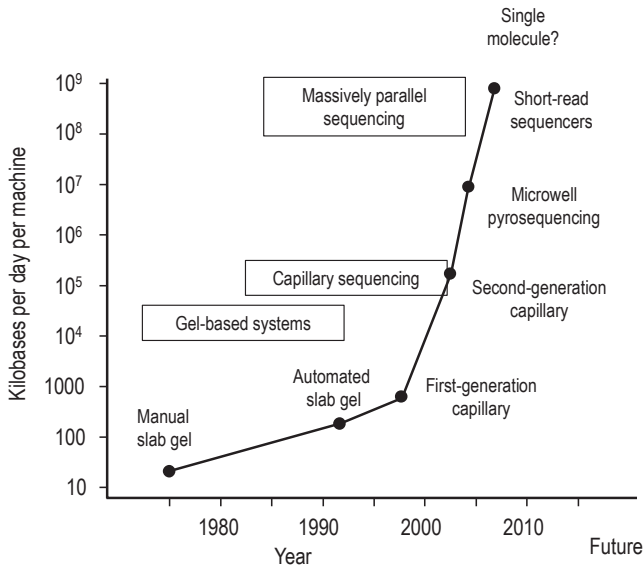
This chapter discusses the knowledge-based, scientific and technological infrastructure required to reap the benefits of marine biotechnology. In other fields, large national or international projects have drawn financial and political attention to the infrastructures required to meet their goals. The chapter asks: “What type of infrastructure is necessary to drive development of the field?” and “What policies might be required to achieve this goal?”

The extent to which the benefits of marine biotechnology are realised will depend, in large part, on how well marine bioresources and the marine ecosystems are understood and conserved. As the preceding chapter has argued, this will require the collection and analysis of new scientific data and comparison of the data with existing knowledge. Realising the full potential of marine biotechnology will also require appropriate research and development (R&D) infrastructures.

Research infrastructures

Heightened interest in marine biotechnology is linked to recent advances in “omic” technologies (e.g. genomics, proteomics) and the new insights into marine bioresources they have made possible. These technologies are fundamental to many marine biotechnology R&D activities and are producing a wealth of genetic data which can lead to better understanding of ocean life and its potential for biotechnological development. Genomic sequencing technology, once a technical and financial stumbling block, has matured and produces data at rates unheard of a decade ago. Illumina HiSeq technology, for example, produces some 10 terabits (Tb) of sequence data per machine annually and output from future technologies is expected to increase to an annual 112 Tb per machine by 2015. Meanwhile, the cost per nucleotide of sequencing is dropping rapidly (Figure 4.1).

Figure 4.1. Progression of sequencing rates and cost



Source: Stratton, Campbell and Futreal (2009), “The Cancer Genome”, *Nature* 458:719-724.

Despite the advances in “omics”,¹ very little is known as yet about marine biodiversity and its potential in terms of bioresources. If, as some suggest (Mora et al., 2011; COML, 2010), something has been learnt about 10% of the ocean’s species, this is undoubtedly the most accessible fraction, encompassing most of the larger mammals, fish and plants and the most common and easy-to-culture microbes. Of the rest – the largest share of marine organisms by number and weight – very little is known.

Metagenomics, or plurality sequencing, of marine microbial communities is opening an unprecedented window on biodiversity. The recent advent of single-cell sequencing (SCS) will further improve the capacity to link the structure and functions of microbial communities in the coming decade (Zhang et al., 2006; Su et al., 2012). Past investment in genomics has led to a marine biotechnology renaissance, which is leading to a new infrastructure bottleneck that threatens to limit the rate at which its benefits can be realised. The widespread application of marine biotechnology will clearly require additional infrastructure.

Tools and processes to collect, culture and catalogue samples

Tools and processes for collecting samples from the marine environment are improving all the time. However, exploration and sampling are still difficult in areas of environmental extremes which offer great potential for discovering organisms with novel functionalities. The Mariana Trench in the western Pacific is a good example. At its deepest point, it is over 10 km deep and the water column exerts a pressure of 1 086 bars (15 750 psi), over a thousand times the standard atmospheric pressure at sea level. As of mid-2012, four expeditions had been made to the bottom of the trench and several more are planned, but they are technically difficult, expensive and very risky. Sampling is limited and, although living organisms have been collected, culturing and studying them remains a challenge. In such extreme environments, technologically advanced vessels are needed for marine biotechnology R&D.

Culturing provides a means of undertaking in-depth analysis of biochemical networks and systems and of preserving marine resources in biobanks. The complex, symbiotic nature of marine organisms means that new culture methods will be required for basic research (e.g. model organisms, screening). In particular, new methods and media, perhaps developed with knowledge from metagenomic studies (and associated microbial communities), that enable cell-to-cell communication or signalling, may be required to culture as yet “unculturable” organisms and symbiotic organisms. New culturing methods based on co-metabolism between community members represent a radical change from the conventional “isolate and enrich” approach to cell culture.

At the production level, culturing may provide a solution to the problem of unsustainable harvesting and may be required for sustainable production of many new compounds/molecules/enzymes. New processes range from the optimisation of biorefineries to produce algal biofuels, to the culturing of genetically modified organisms (GMOs) for pharmaceuticals, to the culturing of bacteria and viruses in the laboratory, to the development of new cell lines.

Collections of biological specimens (or parts thereof) can be used to analyse and preserve biodiversity, to facilitate the exchange of resources and to develop model organisms. Biobanks containing living cultures, nucleic acid archives, extract libraries (such as those obtained with new chromatographic instruments and media) and compound libraries (which enable the study of structures, function and origin) can facilitate the development of new molecules, compounds and bioactives. Development of novel cryopreservation techniques and capacity can also support biobanking efforts and commercial-scale culturing of larvae and marine organisms.

Databases

Databases are an integral part of the study of marine bioresources and biodiversity. A number of databases containing different types of information exist. The World Register of Marine Species (WORMS)², hosted at the Flanders Marine Institute, VLIZ, was established as a global effort to register the names of all marine species. The project involved 270 expert taxonomists from 185 institutions in 38 countries, and the database describes 215 000 species. WORMS is accessible on the Internet and is broken down into sub-portals for different taxonomic groups.

The UNESCO-IOC/IODE Ocean Biogeographic Information System (OBIS)³ is the largest source of information on the distribution of marine species and is a data legacy of the ten-year Census of Marine Life (COML) programme. OBIS contains 32 million records from 1 000 datasets and over 100 000 marine species. Its geographical coverage is good for highly populated regions but less so for remote regions such as the open ocean, the deep sea and the polar regions. The database is an excellent resource for the study of marine organisms and biodiversity in the marine environment.

Both of these databases are publicly accessible and have an international base. However, to remain current, they will require access to networks of stations, seagoing platforms and observatories. They will also need to draw on a large team of taxonomists to ensure that the data are of high quality. This work offers opportunities for synergies. For instance, the establishment of marine observatories will provide an opportunity to collect excess material and bulk samples of water and ocean floor for preservation *in-situ* or *ex-situ* biobanks for later use and for as yet non-existent technologies. There is also

a need to be able to couple the information in these databases with genetic, species and habitat information in other databases and to contextualise this information in terms of ecosystem parameters. Finally, the standardisation of protocols for sample collection and cataloguing will facilitate the assimilation of the work of other research groups, making for easier sharing of information

Platforms for screening and analysis

Genome sequence data (from whole genome data to metagenomics data) from the marine environment is certainly the main type of information used in marine biotechnology R&D. The data are produced rapidly and increasingly inexpensively and this is challenging current R&D infrastructures. For instance, the rate of sequence submissions to archival databases, which are fundamental for reproducible science, is outstripping the rate of growth of storage capacity (Kodama et al., 2012).⁴

Perhaps the most immediate challenge to marine biotechnology is the development of tools and platforms to facilitate high throughput screening of new “omics”-related information. Screening seeks to compare sequence data with information about known genes (including gene products and gene expression profiles) to infer the structure, function or identity of the sequence or organism of interest. These screens have historically relied primarily on comparison to annotated DNA sequences (and related information) in databases such as Genbank (Bensen et al., 2012) or the Marine Genomics Europe (MGE) Bioinformatics Portal.⁵ However, given the relative lack of information about marine bioresources, and the speed at which new sequence data are being generated, these approaches are proving insufficient for accurate annotation⁶ of genomic sequences from novel marine organisms. The existing infrastructure is also challenged by the biodiversity and complexity of the marine environment, which will require the development of new approaches and platforms to link genotype with phenotype from single cells to ecosystems.

Model systems, including *in silico* models (Lerman et al., 2012), have been developed for many organisms in an attempt to bridge the gap between genotype and phenotype. Model organisms provide a means of obtaining a better understanding of biochemical processes and thus of identifying pathways for modification (targets) for further development or production. Such models exist for a number of marine species of medical, industrial or commercial importance (e.g. salmon, sea urchin) or for evolutionary or developmental study (e.g. marine annelid worm).

Model systems support both basic research and the development of advanced marine biotechnology and thus facilitate a systems-based approach to genome annotation and elucidation of new gene functions. In the development phases, model systems can be used to access or harvest organisms or derivatives of interest. Synthetic biology may eventually find application in this field but until then, model organisms can help identify and refine the most cost-effective routes to harvesting or producing functional compounds or organisms of interest. Model systems will be especially useful for the study of new phyla or classes of organisms and extremophiles, which are considered a significant source of new functions. The definition of model systems will need to be expanded to include ecosystems and ocean observatories in order to evaluate community responses to environmental perturbation.

The concept of the “minimal genome” provides a useful way of linking genes to functions and has been useful for identifying minimal metabolic pathways and linking genes with function (Mushegian, 1999). For instance, Dufresne et al., (2003) published the sequence of *Prochlorococcus marinus*, one of the ocean’s dominant photosynthetic organisms. It is one of the smallest photosynthetic organisms, and its genome approximates the minimal gene complement for a photosynthetic organism. Using both model organisms and minimal genomes, it is possible to identify and assign functions to unknown genes.

In recent years, a number of databases have been developed to facilitate the comparative analysis of species. These databases are the workhorse of marine biotechnology and contain molecular and genomic data from microbial communities and individual genomes. Like the Integrated Microbial Genomes (IMG) system (Markowitz et al., 2006, 2012), they serve as a community resource for comparative analysis and annotation of all publicly available genomes. They are free and publicly available.

The complexity of metagenomic data brings with it a further challenge for annotation linking genotype and phenotype. This challenge is concisely articulated by Chisholm and Cary (2001): “Our genetic and biochemical understanding of metabolism, and other cell functions, is based largely on the study of complete pathways within cells. However, microbial communities are a collection of gene functions distributed amongst its individual members which form distributed metabolic pathways directing matter and energy exchange among and between microbes. No single organism contains all the genes necessary to perform the diverse biogeochemical reactions that make up ecological community function.” This realisation has important implications for how gene functions are defined or classified, within symbiotic or related groups of microorganisms, into meaningful units of selection, utilisation or conservation. The complexity of microbial communities is

driving the development of a new wave of e-infrastructures (see Wright et al., 2012, for a review of the microbial ecology of expanding oxygen minimum zones).

Of particular importance will be the development of interactive services that allow for uploading user information for analysis and visualisation and enabling the study of comparative genomics and metagenomics. These e-infrastructures should take into account the multidimensionality of marine genomics data, which include physical and chemical properties and molecular information, in order to integrate metadata with sequence information in the taxonomic and metabolic pathway context.

The development of novel data management platforms and information services⁷, and the generation of data products such as visual analytics and web services, are of paramount importance as growth opportunities for marine genomics and biotechnology. Developing the required infrastructures will be a significant undertaking, financially, structurally and operationally. It will involve considerations unique to marine bioresources and may therefore benefit from specific policy attention. Traditionally, national or multilateral collaboration on research infrastructures has been justified in terms of cost sharing or the need to generate economies of scale and scope. These considerations are even more important for the shared resources, and the associated data, of the very large and highly complex marine environment.

International partnerships to drive innovations in R&D infrastructure

The OECD Innovation Strategy recognises the impact that fully functioning knowledge networks can have on the efficiency and effectiveness of the innovation process, both stimulating innovation and improving its efficiency by reducing transaction costs (OECD, 2010). As marine biotechnology becomes a focus of investment and innovation strategies, it will be important to ensure that mechanisms are in place to generate, share and give value to knowledge in order to enable innovation. Given the global nature of marine bioresources, it will also be useful to consider international, trans-boundary approaches that can help to drive innovation in R&D infrastructure. The existence of several marine biotechnology funding programmes and initiatives suggests that the value of international partnerships and investment for fostering knowledge development is recognised.

EU Joint Programming Initiative: A regional co-ordinated approach to investment

Within the European Union (EU), the Joint Programming Initiative (JPI) provides one model for resource sharing and the co-creation of knowledge. The EU JPI grew from the recognition that “Europe not only needs to invest more in research, but also needs to invest it to better effect if it is to achieve its declared vision: a balanced and sustainable development.”⁸

The purpose of the JPI Oceans (www.jpi-oceans.edu)⁹ is to increase the value of national R&D investments in ocean research in the Europe Union in order to avoid fragmentation and unnecessary duplication of R&D, to look for synergies, and to facilitate different types of co-operation to meet policy objectives and global challenges. The benefits of JPI Oceans include a long-term perspective (10-15 years), a high level of commitment and voluntary participation in different actions. JPI Oceans will focus on: the development of new bioactive products for health treatments; biotechnology applied to aquaculture; biofuel from marine algae; screening of marine genetic diversity; development of marine biosensors to monitor the environment; and mitigation of human and climate change impacts on the ocean. This initiative is an indication of renewed attention to international research partnerships based on shared goals, understanding and challenges.

The EU’s ERA-NET (the European Research Area Networks scheme) was also designed to foster the co-ordination of national and regional research programmes. ERA-NET activities related to marine biotechnology under the EU’s 7th Framework Programme (FP7) aim at transnational co-operation in this area.

Regional partnership with a regional focus

The much older Mediterranean Science Commission (CIESM), established in 1908, is an international partnership with a regional focus. The Commission has grown to 22 member states which support the work of several thousand marine researchers who seek to understand, monitor and protect the rapidly changing, highly stressed Mediterranean Sea. The CIESM takes a multidisciplinary approach. It encourages the exchange of scientific standards and ideas and maintains a constructive dialogue around the Mediterranean Sea and the Black Sea basins, so that the region benefits from collaboration among researchers and populations.

The European marine biological resource centre: A distributed infrastructure

The European marine biological resource centre (EMBRC) is a collaboration of 12 leading marine stations and the EMBL (European Molecular Biology Laboratory) which uses the latest technologies to study marine organisms (microbes, plants, animals). Through a network of distributed research infrastructure with state-of-the-art research and training facilities, EMBRC enables the scientific community at large, including universities and industry, to access marine organisms, aquaria facilities, and dedicated platforms for genomics, structural and functional biology, microscopy, and bioinformatics. Through its network, EMBRC aims to provide comprehensive support (including interdisciplinary training) for the intelligent and sustainable exploitation of marine resources.

The Red Sea Research Centre: International collaboration with a regional focus

The Red Sea Research Centre at KAUST is another example of international collaboration with a regional focus. The Centre, with collaborators from the United Kingdom, the United States, the Netherlands, Hong Kong (China) and the Arabian Peninsula, is working to develop a scientific basis for sustaining and conserving coral reef environments along the Red Sea coast of Saudi Arabia. The group takes a multidisciplinary approach to learning about coral reef ecosystems and their oceanographic context, including the physical, chemical, biological and geological environment and the stresses arising from natural as well as anthropogenic factors such as overfishing, pollution, coastal development and global climate change.

BioMarks: A midsize international collaboration

In contrast to the other partnerships described, BioMarks (Biodiversity of Marine Eukaryotes) is a relatively small international collaboration of eight EU research institutions in four countries. Its goal is to develop “metagenetics” or “metabarcoding” to facilitate the study of global biodiversity. The group is tackling a specific challenge: the development of biomarkers (metabarcodes) that describe the taxonomic, genomic and metabolic diversity of natural environments. The 30 members of this multidisciplinary team examine coastal marine protist biodiversity using massive rDNA sequencing integrated into a network of taxonomic expertise and comprehensive contextual phenotypic and environmental metadata. If successful, metabarcoding and metagenomic data could be used for the biomonitoring of sites affected by human activities, for bioprospecting, for

understanding the functioning of the global ecosystem and for reconstructing past environmental changes. The work will also provide a foundation for the future annotation of genomic sequences and taxonomically controlled eukaryotic metagenomics.

As governments consider investment in and development of research infrastructures for marine biotechnology, it will be important to take account of the lessons learned from these and other initiatives. It will be useful to look at synergies that can be achieved across initiatives and how these synergies can be exploited to best effect. It will also be essential to consider ways for developed countries to share scientific and technological infrastructures with emerging economies and developing countries. If tackled in the right way, these infrastructures could transform the way and rate at which knowledge is generated and used.

Marine biotechnology megaprojects

In other scientific, technical or engineering fields, large national or international projects have focused financial and political attention on developing the necessary infrastructures. This approach has been critical to success in fields such as the Human Genome Project (HGP). More than two decades after it began, the HGP is often used to illustrate the benefits of internationally co-ordinated, or harmonised, development of infrastructures: reduced duplication, easier data sharing, larger and more cohesive infrastructures; etc. From 1990 to 2003, the HGP concentrated considerable investment (USD 3.8 billion from the US government alone) and infrastructure on efforts to determine the sequence of the chemical base pairs that make up DNA and to identify and map the approximately 20 000-25 000 genes of the human genome.

Completed in 2003, the HGP left a legacy of genomics and bioinformatics infrastructure which continues to be used and developed by the international research community in both the public and private sectors. The enduring nature of this legacy contrasts with some other major “big science” projects with a much more finite life span, such as the Superconducting Super Collider (USD 11 billion, estimated life of 30 years) and the Hubble Space Telescope (USD 1.5 billion, estimated life of 15–20 years).

The HGP was also transformational in that it illustrated the potential for biology to be an economic driver. The US federal government investment in the HGP, through to its completion in 2003 (USD 5.6 billion in 2010 terms), was crucial to the generation of an economic output of USD 796 billion, giving a return on investment to the US economy of 141 to 1 (Battelle, 2011). In 2010 alone, the genomics-enabled industry generated over USD 3.7 billion in federal taxes and USD 2.3 billion in state and local taxes. In

other words, in one year the government received revenue nearly equal to the entire 13-year investment in the HGP. Also in 2010, human genome sequencing projects and associated research and industry activity directly and indirectly created 310 000 jobs.

A more recent example is the 1000 Genomes Project, an international research effort to establish a detailed catalogue of human genetic variation by sequencing the genomes of 1 000 individuals.¹⁰ At its peak, this international megaproject sought to generate over 8.2 billion bases per day (the equivalent of two genomes a day), a feat unthinkable a few years ago. Over its three-year lifespan, the 1000 Genomes Project will have generated 60 times the amount of sequence data that had been deposited in public DNA databases over the past quarter century. It will constitute a major bioinformatics and statistical resource for researchers and will inevitably drive innovations in data analysis and interpretation.

These examples show that large-scale, international initiatives and investment can create critical infrastructure and provide tangible and quantifiable returns on investment. Government funding and support will necessarily play a large role in the development of research infrastructures for marine biotechnology.

The success of the human genome megaprojects inspires hope that marine biotechnology megaprojects may also be successful. One of the first of these, sequencing of the tiger pufferfish, was initiated in 1989 to inform the functional genomics research for the human genome (Aparicio et al., 2002). More recently, commercially important aquaculture species have catalysed large-scale international collaboration and resulted in the mapping or sequencing of salmon (Davidson et al., 2010), trout (Palti et al., 2011) and cod (Star et al., 2011). Marine microbes have also been the focus of significant, often international, initiatives.¹¹

Questions about the merit and feasibility of marine biotechnology megaprojects nonetheless remain. “What question or questions should drive such projects?” “What costs and project duration should be anticipated, with what risks?” “How might international collaboration be organised?” “Are the barriers to such megaprojects too great at present, owing to concerns about the complexity of the marine environment or the structuring of international collaboration?”

Several considerations increase the need to understand the essence of the marine biotechnology challenge and will affect the scope and focus of a potential megaproject. The challenge is first to understand better the huge, largely unstudied, complexity of marine organisms and marine ecosystems. It is close to impossible to understand the totality of marine organisms and ecosystems. At the same time, it is not possible to focus on one organism –

this is not a “human genome” project – so it will be necessary to consider the type and volume of information required for a marine biotechnology project. It will then be necessary to consider the type of research infrastructure required and how it might be developed, for example in relation to infrastructure being developed in other fields or by other megaprojects.

The development of the infrastructure brings its own questions, such as: “Who will champion the project?” or “How can the project garner the support it needs on an international level?” Many organisations have some jurisdiction in terms of monitoring and managing the ocean, some of them with significant resources and funding and some with very little. However, there is no single authoritative body able to focus the interests and resources of these fragmented, often competing, organisations and states on a megaproject.

In the absence of such an entity, states and stakeholders will need to work together to define project scope and concentrate resources. Successful megaprojects have focused first and foremost on knowledge generation at a fundamental, application-neutral level (e.g. sequencing the human genome, mapping human genetic variation). For a marine-based project, biodiversity – the core of productivity and sustainability – might be a suitable analogy.

The project would also need to compete for resources and financial support with other government initiatives and international megaprojects. Here again, the benefits of a sole administrator championing the need for research and international attention are clear. In the absence of a single champion, stakeholders will need to raise awareness of the potential opportunities of a project.

Conclusions

The renewal of interest in marine biotechnology in many OECD countries today owes much to advances in science and technology over the last decade. These advances are enabling good access to marine organisms, especially the wealth of microscopic marine microbes that seem to hold so much potential for marine biotechnology. DNA sequence information about these microorganisms can be generated faster and more cheaply than ever before, at a rate which exceeds the capacity to analyse it. In the most general sense, two main challenges are shaping the infrastructure needs of the field.

The first concerns the paucity of data about marine organisms, which makes it very difficult to characterise or classify community structure and function. Without taxonomic information or the association of genetic data and phenotype data, it is difficult to assess the potential of an organism or its genetic repertoire. For invisible microbes and larger multicellular organisms, there is a need for model systems that will allow top-down and bottom-up

studies of gene expression at the individual, population, community and ecosystem levels. The resulting data will require infrastructure for coupling genetic, species and habitat information. To be most effective, global data and information infrastructures to develop and distribute data are needed. Such databases must be permanent and provide all users free and open access to all data.

The second challenge relates to the lack of analytical platforms to process data efficiently and effectively and keep pace with upstream data generation. This, along with the complexity of samples (e.g. metagenomes) to be processed, is creating the need for new e-infrastructure (e.g. novel data management platforms and information services, generation of visual analytics and web service products) to allow for better analysis across databases, especially those containing different types of information. This meta-analysis will be critical for achieving a better understanding of the biodiversity of marine ecosystems. Properly integrated into time-variable forecasts and monitoring strategies, it might also prove useful in diagnosing and predicting environmental responses to natural and anthropogenic disturbances.

Such resources will require significant investment and will benefit from multi-country and multi-stakeholder collaboration. While various international partnerships address some aspects of marine biotechnology, few international initiatives are working on these challenges. Current efforts in this regard are generally led by small groups or individual countries. There is a need to look for synergies among these projects to reduce the likelihood of duplication.

Fortunately, it is likely that scientific progress in other fields will diffuse to marine biotechnology. This has been the case for next-generation sequencing technology arising from investment in the Human Genome Project, and may well be the case for advances arising from the 1000 Genomes project. However, there are some unique features of marine biotechnology (its size, its complexity and the untapped diversity of the majority of marine organisms) that may argue for a megaproject centred on marine organisms. Details of such a project remain to be defined but could be of significant value to governments wanting to improve the productivity and sustainability of marine biotechnology.

Notes

1. For a review of the state of the art, see Liu et al. (2012), Shokralla et al. (2012) and Pareek et al. (2011).
2. www.marinespecies.org
3. www.iobis.org/
4. This problem is the focus of much attention by researchers, but the private sector has recognised the challenge (and perhaps an opportunity): in 2012 Amazon's cloud computing unit, Amazon Web Services, announced it would store the entire contents of the National Institutes of Health's 1000 Genomes Project, and provide free access to everyone. See <http://bits.blogs.nytimes.com/2012/03/29/amazon-web-services-big-free-genetic-database/?scp=1&sq=Genomes&st=cse>.
5. www.cebitec.uni-bielefeld.de/groups/brf/software/portal/portal/mge/?cookie_test=1.
6. Annotation of genomic sequences, or DNA annotation, is the process by which supplementary information about the sequence (e.g. genomic position to intron-exon boundaries, regulatory sequences, repeats, gene names and protein products) is added prior to inclusion in a database.
7. Information services include:
 - IMG: <http://img.jgi.doe.gov/cgi-bin/m/main.cgi>
 - CAMERA: <http://camera.calit2.net/>
 - GenGIS: http://kiwi.cs.dal.ca/GenGIS/Main_Page
 - COML: www.coml.org/
 - OBIS: www.iobis.org/
 - MetaCyc: <http://metacyc.org/>
 - ORION: www.orionprogram.org/OOI/
 - GEO: www.earthobservations.org/index.shtml
 - Gensc: http://gensc.org/gc_wiki/index.php/Main_Page
 - Network Workbench: <http://nwb.cns.iu.edu/about.html>
 - CIRCOS: <http://circos.ca/>
 - MIZBEE: www.cs.utah.edu/~miriah/mizbee/
 - Microbial Ecological Genomics DataBase (MegDB) www.megx.net
8. Commission of the European Communities (2008), "Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Towards Joint Programming in Research: Working together to

- tackle common challenges more effectively”, COM(2008) 468 final, http://ec.europa.eu/research/press/2008/pdf/com_2008_468_en.pdf, accessed March 2012.
9. Full title: Joint Programming Initiative “Healthy and Productive Seas and Oceans”.
 10. In the first phase, 2008-10, the genomes of 1 000 individuals were sequenced to 4x coverage. Throughout all phases of the project, the genomes of about 2 500 unidentified individuals from about 25 populations will be sequenced (The 1000 Genomes Project Consortium, 2010).
 11. For example, the Microbial Genome Sequencing Project is sponsored by the Gordon and Betty Moore Foundation’s Marine Microbiology Initiative, <http://camera.calit2.net/microgenome/>, accessed September 2012.

References

- Aparicio, S., J. Chapman, E. Stupka, N. Putnam, J.M. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon et al. (2002), “Whole genome shotgun assembly and analysis of *Fugu rubripes*”, *Science* 297(5585):1301-10, 23 August.
- Battelle Technology Partnership Practice (2011), “The economic impact of the human genome project. How a \$3.8 billion investment drove \$796 billion in economic impact, created 310,000 jobs and launched the genomic revolution”, Battelle, Columbus OH, May.
- Benson, D.A., I. Karsch-Mizrachi, K. Clark, D.J. Lipman, J. Ostell and E.W. Sayers (2012), “GenBank”, *Nucleic Acids Research* 40 (D1): D48-D53. doi: [10.1093/nar/gkr1202](https://doi.org/10.1093/nar/gkr1202).
- Chisholm, S.W. and C. Cary (2001), “Ecological genomics: the application of genomic sciences to understanding the structure and function of marine ecosystems”, In Report, NSF workshop on Marine Microbial Genomics 1–20 (Univ. Press).
- COML (Consensus on Marine Life) (2010). “First Census of Marine Life 2010: Highlights of a Decade of Discovery”, COML, Washington DC.
- Davidson, W.S., B.F. Koop, S.J.M. Jones, P. Iturra, R. Vidal, A. Maass, I. Jonassen, S. Lien and S.W. Omholt (2010), “Sequencing the genome of the Atlantic salmon”, *Genome Biology* 11:403. <http://genomebiology.com/2010/11/9/403>.
- Dufresne, A., M. Salanoubat, F. Partensky, F.M. Artiguenave, I. Axmann, V. Barbe, S. Duprat, M.Y. Galperin, E.V. Koonin and F. Le Gall (2003), “Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome”, *Proceedings of the National Academy of Sciences*, 100(17): 10020–10025, 19 August. doi: [http://dx.doi.org/10.1073/pnas.1733211100](https://doi.org/10.1073/pnas.1733211100).
- Kodama, Y., M. Shumway and R. Leinonen (2012), “The sequence read archive: explosive growth of sequencing data”, *Nucleic Acids Research* 40 (D1): D54-D56. doi: [http://dx.doi.org/10.1093/nar/gkr854](https://doi.org/10.1093/nar/gkr854).

- Lerman, J.A., D.R. Hyduke, H. Latif, V.A. Portnoy, N.E. Lewis, J.D. Orth, A.C. Schrimpe-Rutledge, R.D. Smith, J.N. Adkins, K. Zengler and B.O. Palsson (2012), “*In silico* method for modelling metabolism and gene product expression at genome scale”, *Nature Communications* 3;3:929, July. doi: <http://dx.doi.org/10.1038/ncomms1928>.
- Liu L., Li Y., Li S., Hu N., He Y., Pong R., Lin D., Lu L. and M. Law (2012), “Comparison of next-generation sequencing systems”, *Journal of Biomedicine and Biotechnology* 2012:251364. Epub 2012 Jul 5.
- Markowitz, V.M., F. Korzeniewski, K. Palaniappan, E. Szeto, N.N. Ivanova, N.C. Kyrpides et al. (2006), “An Experimental Metagenome Data Management and Analysis System”, Lawrence Berkeley National Laboratory: Lawrence Berkeley National Laboratory, <http://escholarship.org/uc/item/1r9637xm/>.
- Markowitz, V.M., I.M.A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, B. Jacob, J. Huang, P. Williams, M. Huntemann, I. Anderson, K. Mavromatis, N.N. Ivanova and N.C. Kyrpides (2012), “IMG: The integrated microbial genomes database and comparative analysis system”, *Nucleic Acids Research* 40 (D1): D115-D122. doi: <http://dx.doi.org/10.1093/nar/gkr1044>.
- Mushegian, A. (1999), “The minimal genome concept”, *Current Opinion in Genetics and Development* 9(6):709-14.
- Mora C., D.P. Tittensor, S. Adl, A.G.B. Simpson and B. Worm (2011), “How Many Species Are There on Earth and in the Ocean?”, *PLoS Biology* 9(8): e1001127. doi: <http://dx.doi.org/10.1371/journal.pbio.1001127>.
- OECD (2010), *The OECD Innovation Strategy: Getting a Head Start on Tomorrow*, OECD Publishing. doi: <http://dx.doi.org/10.1787/9789264083479-en>.
- Palti, Y., C. Genet, M.C. Luo, A. Charlet, G. Gao, Y. Hu, C. Castano-Sanchez, K. Tabet-Canale et al. (2011). “A first generation integrated map of the rainbow trout genome”, *BMC Genomics* 12:180. doi: <http://dx.doi.org/10.1186/1471-2164-12-180>.
- Pareek, C.S., R. Smoczynski, A. Tretyn (2011), “Sequencing technologies and genome sequencing”, *Journal of Applied Genetics*. 52(4):413-35, November. doi: <http://dx.doi.org/10.1007/s13353-011-0057-x>.
- Shokralla, S., J.L. Spall, J.F. Gibson and M. Hajibabaei (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* 21(8):1794-805, April. doi: <http://dx.doi.org/10.1111/j.1365-294X.2012.05538.x>.

- Star, B., A.J. Nederbragt, S. Jentoft, U. Grimholt, M. Malmstrom, T.F. Gregers, T.B. Rounge, J. Paulsen, M.H. Solbakken, A. Sharma et al. (2011), “The genome sequence of Atlantic cod reveals a unique immune system”, *Nature* 477, 207–210, 8 September. doi: <http://dx.doi.org/10.1038/nature10342>.
- Stratton, Campbell and Futreal (2009), “The Cancer Genome”, *Nature* 458:719-724.
- Su C., Lei L., Duan Y., Zhang K.Q. and Yang J. (2012) “Culture-independent methods for studying environmental microorganisms: methods, application, and perspective”, *Applied Microbiology and Biotechnology*, Feb;93(3):993-1003. Epub 2011 Dec 22.
- The 1000 Genomes Project Consortium (2010), “A map of human genome variation from population-scale sequencing”, *Nature*, 467(7319):1061-1073.
- Wright, J.J., K.M. Konwar and S.J. Hallam (2012), “Microbial ecology of expanding oxygen minimum zones”, *Nature Reviews Microbiology* 10, 381-394, June. doi: <http://dx.doi.org/10.1038/nrmicro2778>.
- Zhang, K., A.C. Martiny, N.B. Reppas, K.W. Barry, J. Malek, S.W. Chisholm and G.M. Church (2006), “Sequencing genomes from single cells by polymerase cloning”, *Nature Biotechnology* 24, 680-686, doi: <http://dx.doi.org/10.1038/nbt1214>.



From:

Marine Biotechnology

Enabling Solutions for Ocean Productivity and Sustainability

Access the complete publication at:

<https://doi.org/10.1787/9789264194243-en>

Please cite this chapter as:

OECD (2013), “Infrastructures for generating and sharing knowledge”, in *Marine Biotechnology: Enabling Solutions for Ocean Productivity and Sustainability*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264194243-6-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.