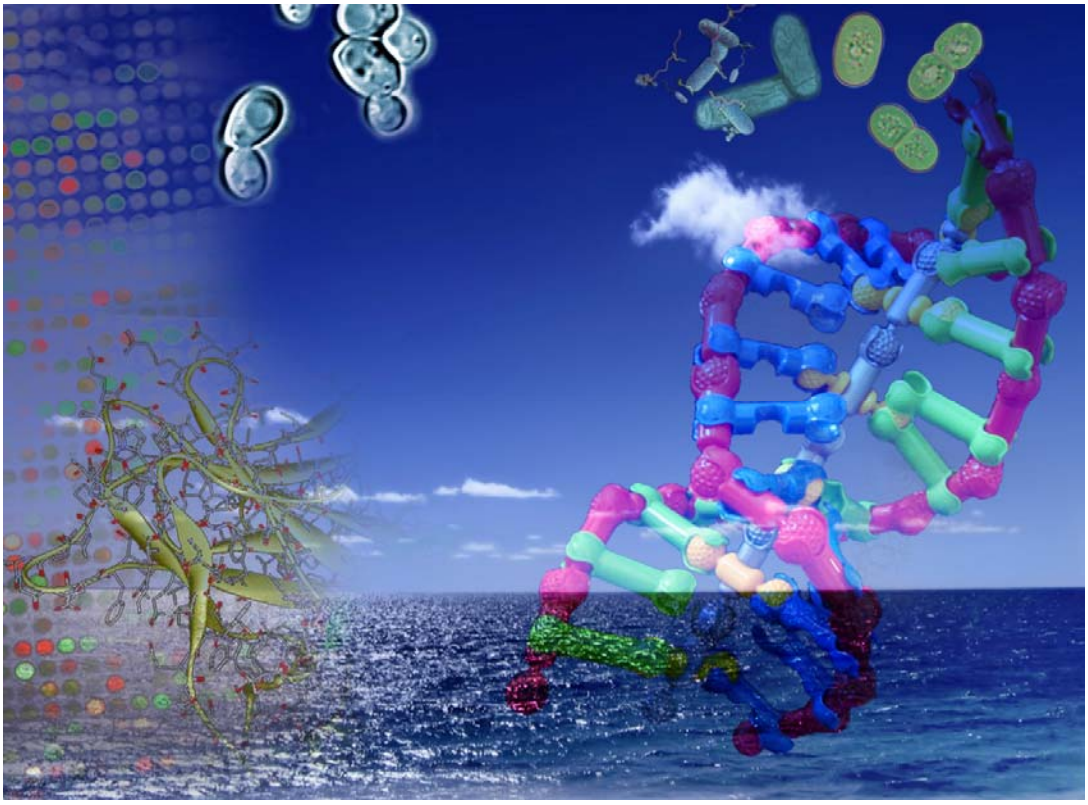
JRC Scientific and Technical Reports

Microbial Biodiversity and Molecular Approach

Aquatic microbial world and biodiversity: Molecular Approach to improve the knowledge

Elena Nardini, Veljo Kisand and Teresa Lettieri



EUR 24243 EN - 2010

JRC EUROPEAN COMMISSION

ies Institute for Environment and Sustainability

The mission of the Institute for Environment and Sustainability is to provide scientific-technical support to the European Union's Policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability

Contact information: Teresa Lettieri
Address: Via Enrico Fermi, 2749
21027 Ispra (VA), Italy TP270
E-mail: teresa.lettieri@jrc.ec.europa.eu
Tel.: +39 0332789868
Fax: +39 0332786645

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

EUR 24243 EN

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

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

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu/

JRC 56843

EUR 24243 EN
ISBN 978-92-79-14990-0
ISSN 1018-5593
DOI 10.2788/60582
Luxembourg: Office for Official Publications of the European Union

Printed in Italy

Contents Pg

Figures and Tables

Introduction

Biological diversity - or biodiversity - is the term given to the variety of species on Earth resulting from billions of years of evolution. Biodiversity is shaped by natural processes and, increasingly, by the influence of humans. It forms the web of life of which we are an integral part and upon which we so fully depend. This diversity depends on the wide variety of plants, animals and microorganisms.

"Microbes orchestrate life on earth" seems more a philosophy meaningful than a scientific assert. They live on this planet since more than three billion of years (the appearance of *Homo* species is dated about 1.8-2 million ago) and their capability to adapt to several environments make them still the "highlanders". Why? Because they possess enormous metabolic versatility which ensure them a key-role in biogeochemical processes such as the carbon, nitrogen cycles (terrestrial and aquatic ecosystem) and, being producers and decomposers, in food webs. At the same time microbes in natural ecosystems receive little attention; they are largely ignored even by most professionals and are virtually unknown to the public except in pathogens plus few other minor areas. Yet, the functioning of whole biosphere depends absolutely on the activities of the microbial world. We are living in a "microbial planet", at the same time public knowledge, awareness and political actions do not deal with microbes when biodiversity and its decrease is in focus.

Behind the awareness of microbiologists for their crucial role, until now only few microbes, and mainly involved in human diseases, had been characterized e.g. the genome sequence, metabolic pathways. The explosion of genomic tools in 2000s such as the automatation of sequencing platform has accelerated the possibility to sequence many other microorganisms relevant in environmental field. However the milestone of environmental genomics era have been achieved by J. Craig Venter in 2004 when it challenged to reveal the unraveling with metagenomics. Indeed, he and his coworkers sequenced the entire sample from Sargasso Sea characterizing the diversity, the abundance of several microorganisms, deciphering the interaction among the communities. Especially they overcome with this method to the difficulties to characterize the uncultivable microbes. From that time thanks to drop in price of the

sequencing cost, many papers have been published on metagenomics. These studies applied metagenomics to several environmental samples and discovered a huge biodiversity of microbial world.

In News of journal Nature published in January the 7th, the United Nations has proclaimed 2010 as the International Year of Biodiversity, to pose more attention on hopes to establish strategies to prevent biodiversity loss. Related to this issue, in the vision 2020 (Opinion, Nature, 7 January 2010), one key step should be to mitigate this loss of biodiversity and the degradation of ecosystem function. It underlines that one step would be to gauge the resilience of ecological networks such as food webs — in particular, their capacity to withstand disturbance and species loss.

This concept would then imply two new concepts, i) a deeper knowledge of microbial communities and their interaction within the ecological network; ii) preservation of microbial biodiversity because of their role in the food webs, would mean preservation of the other species and their niches.

This review has been conceived in this context. Indeed, behind the description of the microbial role into ecosystem, some chapters are dedicated to the importance of microbial diversity and the impact of anthropogenic pressures (e.g. climate change, pollutants) on its loss. The coral reefs degradation is a relevant example. The characterization of microbial biodiversity is still far away to be completed but on the way thanks to the new post-genomics technologies which are highlighted in two chapters e.g. pyrosequencing; single microbial cell sequencing, metatranscriptomics, and metaproteomics.

Particularly metatranscriptomics, and metaproteomcis are new approaches which in the coming years will be more and more routinely used in environmental studies. These refer to the collection and analysis of transcriptition (mRNA) and protein profile information from microbial communities. Such data will help to identify activity of metabolic pathways, so far unknowns and, then, to build up a knowledge of the microbial ecosystem functioning as a global network.

Which will be, then, the next step? In the future the acquired knowledge of microorganisms could open new avenues such as the potential use of microbial processes

to support the sustainable development. It is the future, but it will reinforce that the microbes orchestrate the life on the earth. It is a scientific assert.

Chapter 1

Microbial world in water ecosystems

Various aquatic ecosystems such as World Oceans, lakes, rivers, springs, ponds, and ground water provide most of the living space on Earth. A multitude of environmental niches are present in the various – some hot (even extreme hot with temperatures over 100˚C), some cold, at high pressures in the deep ocean. Despite the undisputed and vital roles of microorganisms in the global ecosystems –driving biogeochemical cycles and basic part of foodwebs that affect climate and the cycling of elements and nutrients to other organisms – the most of aquatic environments remain under-explored both and therefore represent to huge pool of unexplored biodiversity. Until recently, the study of biodiversity was hampered significantly by infancy in technical methodologies. These included the difficulties in collecting representative samples from some of the more difficult-to-reach niches and methodological approaches to characterize high and uncultivable diversity of microbes. For example, deep-sea microbiology is relatively new because it is both difficult and expensive to recover material either from water many thousands of meters deep, or from beneath the seafloor. Or most of metagenomic studies are still far from basing on highly replicated and systematic sampling campaigns.

Earth's ocean is estimated to contain 10^{29} bacteria (1), a number larger than the estimated 10^{21} stars in the universe. They are typically in size of 0.2–0.6 μm in diameter - size defined as picoplankton (0.2 – 2 μm in diameter). Most of organisms in picoplankton size class are bacterial – both auto- and heterotrophic and archaeal (2). While comparing the biomass of picoplankton to the next size class – nanoplankton (2 – 20 μm in diameter), the biomasses are equal. Considering the order of magnitude higher abundance (number of individual organisms) the surface area of picoplanktonic organisms is huge compared to other groups (Figure 1) and therefore has major importance in elemental fluxes.

The emphasis on the organizational level of biodiversity responsible for ecosystem processes is shifting from a species-centered focus to include genotypic

diversity. The relationship between biodiversity measures at these two scales remains largely unknown. Communities with intermediate species richness show high genotypic diversity while species-poor communities do not (3). Disturbance of these communities disrupts niche space, resulting in lower genotypic diversity despite the maintenance of species diversity.

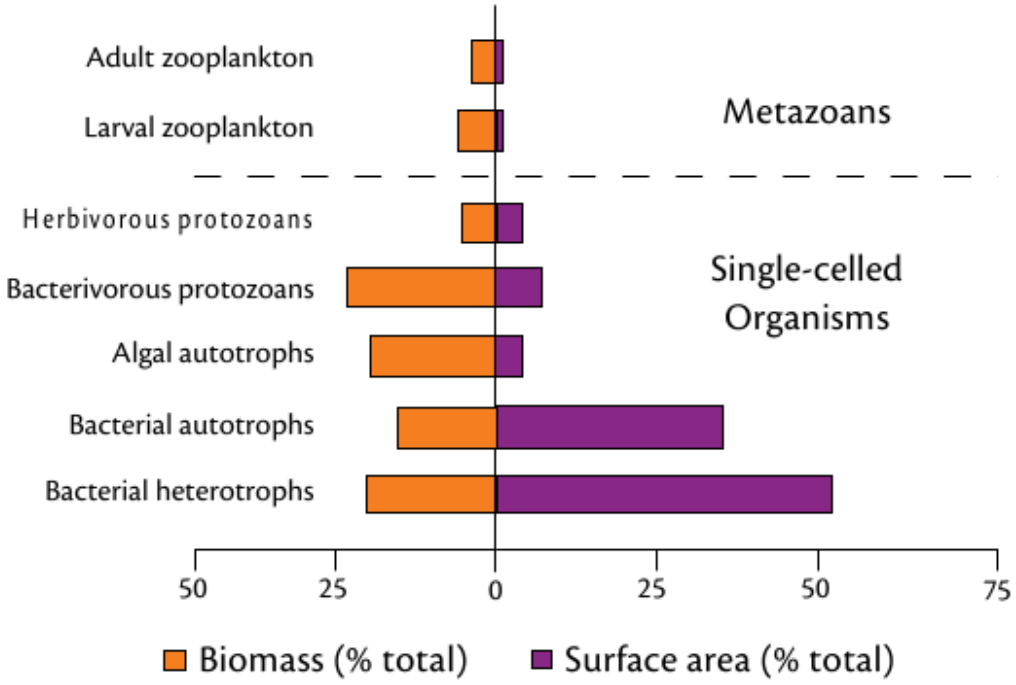


Figure 1. Distribution of biomass and cell surface area between major groups of plankton organisms (expressed as percentage of total).

1.1 Diversity of bacterial heterotrophs

Heterotrophic bacteria dwelling in aquatic environments are highly diverse. At coarse level the gram-positive bacteria, the Verrucomicrobiales and the Alpha- and Gamma-Proteobacteria are distributed throughout a range of aquatic habitats including marine and fresh water systems. Some phylogenetic groups appear to be adapted to more narrowly defined niches such as anoxic water and sediments (Delta-Proteobacteria) or aggregates (Bacteroidetes). Beta-proteobacteria have been detected throughout freshwater

habitats, but these organisms are largely absent from open ocean environments. At narrower level of identification some phylotypes are probably globally distributed as they have been detected in geographically disparate environments. High diversity of heterotrophic bacteria in aquatic environment is explained by high variety of ecological niches occurring and wide spectrum of substrates these organisms utilize.

1.2. Diversity of bacterial autotrophs

Although the biomass of autotrophic prokaryotes in the euphotic zone of the ocean is of the same order of magnitude as that of heterotrophic ones (4), their diversity is strikingly low. There are only four genera of marine picocyanobacteria that are globally significant: *Prochlorococcus*, *Synechococcus*, *Trichodesmium* and N_2-fixing *Synechocystis* (5). Relatively low diversity is explained by the fact that the number of ecological niches available to autotrophs are probably much fewer discriminated mostly by different levels of light and nutrients. Therefore only a small number of taxa are adapted to given levels of light and nutrients. However, recent studies revealed that abundance and diversity of facultative autotrophs is spread into major groups of bacteria: most of sub-classes of Proteobacteria, Bacteroidetes and Actinobacteria (6).

1.3. Diversity of viruses

The abundance of viruses exceeds that of Bacteria and Archaea by approximately 15-fold in the world ocean. However, because of their extremely small size, viruses represent only approximately 5% of the prokaryotic biomass because their content of matter is low. It is estimated that viruses kill approximately 20% of bacterial biomass per day. As well as being agents of mortality, viruses are one of the largest reservoirs of unexplored genetic diversity on the Earth besides Bacteria and Archea.

More importantly, viruses are proposed to be major regulators of the enormous molecular diversity of prokaryotic communities. The host-specific, often strain-specific, nature of viral infection might specifically control the community composition of prokaryotes. According to this model the diversity of the microbial community is maintained by viral infection and microbial abundance is controlled by the nonspecific (or barely size specific) nature of protozoan grazing. This model in known as ‘killing the

winner', how viruses regulate microbial diversity in nature remains ambiguous due to missing studies, and it is unclear whether the differences in viral cellular receptors between natural microbial communities is strain specific or has broader measures of host genotypic diversity. Most abundant groups of viruses found in aquatic environments are bacterio- and cyanophages. Majority of aquatic marine viruses seem to have genome sizes of 25–50 kilobases (kb), whereas less-abundant virus types possess genome sizes that lie between approximately 60 kb and 150 kb. The first metagenomic studies of viral communities have revealed that viral communities contain large amounts of sequences with very low homology to any described sequences available in public database (7).

1.4. Microbial loop to aquatic food webs

Picosize, diverse in phylogeny and metabolic functions microbes form the most abundant and highest in biomass compartments in aquatic foodwebs. The microbial loop is a term coined to general understanding about food webs in aquatic ecosystems and highlighted the importance of osmotrophic heterotrophs as producers (8). Microbial loop is a trophic pathway where dissolved organic matter (DOM) and particulate organic matter (POM) are reintroduced to the food web via bacteria (Figure 2). Bacteria as major osmotrophs incorporate DOM, they are also involved in degradation (mainly hydrolysis) of POM. Bacteria are mostly consumed by eukaryotic protists such as flagellates and ciliates. These protists, are consumed by larger aquatic organisms (metazooplankton), thereafter by fish. Rough estimate is that about 90% of the biomass in World ocean is in microbial loop (pico- and nanoplankton) and even more accounts on the flux of organic material.

The DOM and POM as a sink of the organic matter originates from several sources, such as the leakage of fixed carbon from algal cells or the excretion of waste products by aquatic animals and microbes. Part of POM is transformed to DOM by degradations. In inland waters and coastal environments terrestrial ecosystems (i.e. terrestrial plants and soils) are significant source of DOM and POM. Most of DOM and partly POM is unavailable to aquatic organisms other than heterotrophic bacteria. Because microbes are the base of the food web in most aquatic environments, the trophic efficiency of the microbial loop (microbial food web) has a profound impact on important aquatic processes. Such processes include the productivity of fisheries and the amount of carbon exported to the ocean floor. Trophic efficiency in turn is regulated by biodiversity of the microbial loop.

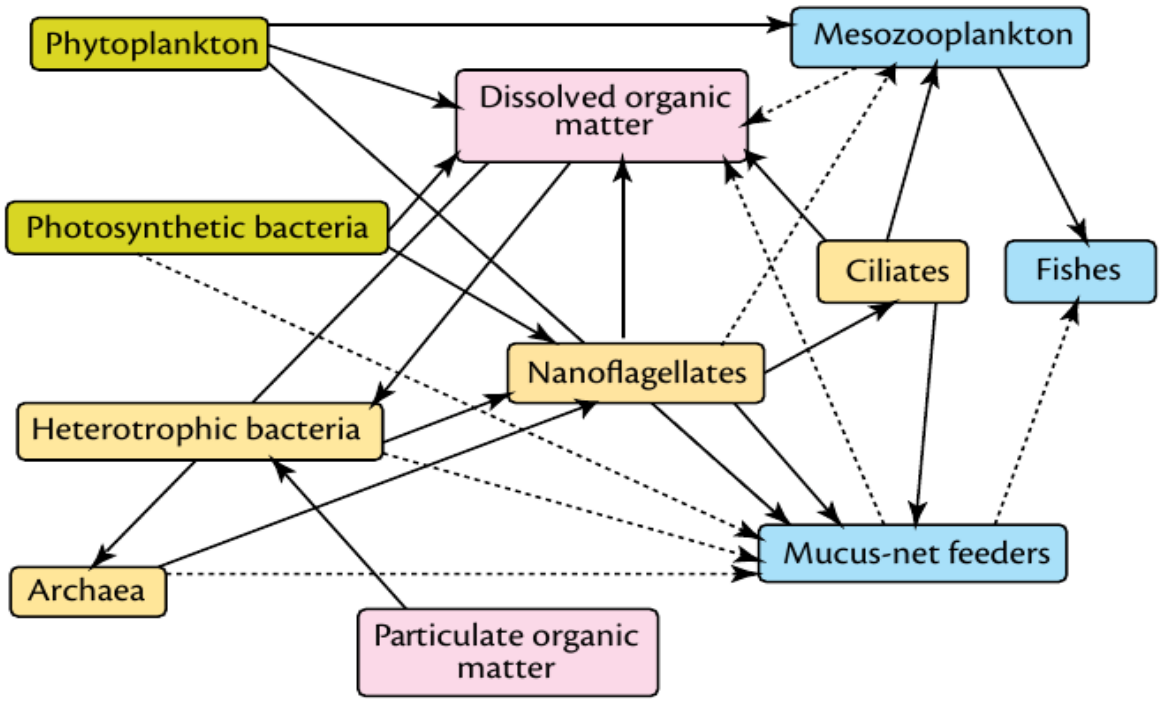


Figure 2. Microbial loop and aquatic food web.

1.5. Microbial world and geochemical cycles

The Earth's biosphere is shaped by geochemical activities of microbes that have

provided conditions both for the evolution of plants and animals and for the continuation of all life on Earth including human activities. Many microorganisms carry out unique geochemical processes critical to the operation of the biosphere (9) and there are no geochemical cycles were they are not involved. Metabolic variety of microbes is enormous ranging from photo- and chemosynthesis and to degradation various anthropogenic xenobiotic compounds. For example, the global nitrogen cycle in nature is dependent on microorganisms. Unique processes carried out by microorganisms include nitrogen fixation, oxidation of ammonia and nitrite to nitrate, and nitrate reduction with formation of dinitrogen and nitrous oxide gases (9). Similar important and unique roles are played in other cycles, such as the sulfur and in carbon cycles. In addition, microbes run less visible elemental cycles of metals, carrying out oxidation/reduction of metals (e.g. manganese, iron). Carbon cycle in aquatic ecosystems has peculiar character due to microbial loop, mentioned above, via which dead and non-accessible DOM is reintroduced into food web at nearly primary producers manner. Microorganisms are the primary organisms responsible for degradation of a great variety of natural organic compounds, including cellulose, hemi-cellulose, lignin, and chitin which are the most abundant organic matter on Earth (10).

Due to their versatility microbes are the major natural providers of ecological services as well play major role in semi-artificial systems such as sewage treatment plants, landfills, and in toxic waste bioremediation. To mention few examples in which microbes are responsible for degradation of toxic chemicals derived from anthropogenic sources such as PAH (polyaromatic hydrocarbons), PCBs (polychlorinated biphenyls), dioxins, pesticides etc.. In most cases these microbes are genuine members of natural communities, not always abundant when specific chemical compounds are released into the system might become dominating (references about examples). Some organisms are obligatory degraders, frequently switching their metabolism on degradation and consumption to acquire carbon and/or energy.

Chapter 2

Pressures and drivers causing decrease of microbial biodiversity

The diversity of microscopic life forms (including viruses, Archaea, Bacteria, and small Eukarya) are recently coming to light, and their varieties, abilities, distributions, ecosystem functions and conservation status need to be further investigated. The principal pressure is habitat fragmentation, degradation and destruction due to land use, change arising from conversion, intensification of production systems, abandonment of traditional (often biodiversity–friendly) practices, construction and catastrophic events including fires. Other key pressures are excessive exploitation of the environment, pollution and the spread of invasive alien species.

Commonly used measures of biodiversity, such as the number of species present, are strongly scale-dependent and only reveal a change after species have been lost. There is no widely accepted and globally available set of measures to assess biodiversity. The problem lies in the diversity of the data and the fact that it is physically dispersed and unorganized. The solution is to organize the information, and to create systems whereby data of different kinds, from many sources, can be combined. This will improve our understanding of biodiversity and will allow the development of measures of its condition over time. The links between loss of species diversity in nature and the health of human populations are less well understood. Of course there are, and will be, substantial impacts from species loss, although our understanding of them is rudimentary. SEBI (Streamlining European Biodiversity Indicators) 2010 is a pan-European initiative, launched in January 2005 to develop appropriate indicators to assess achievement of the 2010 biodiversity target at European level. The SEBI 2010 process proposed 26 indicators annexed to the COM (2008).864 Final.

2.1 Impact of anthropogenic pressures

Several publications document the effect of chemical pollutants such as Polycyclic Aromatic Hydrocarbons (PAHs) on microbial community structure. PAHs are present in oil and coal and produced by incomplete combustion of wood, coal; they are wide spread over the world and they are considered heavy pollutants due to their toxic

carcinogenic, mutagenic effects on the organisms. The study of bacterial communities in PAH contaminated soils at an electronic-waste processing center in China (11) shows that different levels of PAHs might affect the bacterial community by suppressing or favoring certain groups of bacteria, for instance, uncultured *Clostridium* sp. and *Massilia* sp., respectively. Taxonomic analysis indicated Beta-proteobacteria and Firmicutes were abundant bacterial lineages in PAHs-polluted soils. The study of the effects of temperature and fertilization on total versus active bacterial communities exposed to crude and diesel oil pollution in NW Mediterranean Sea (12) shows that fertilization reduced diversity index of both total and active bacterial communities.

A comparison of two distinct large-scale field bioremediation experiments, located at the Canadian high-Arctic stations of Alert (*ex situ* approach) and Eureka (*in situ* approach) demonstrates a rapid reorganization of the bacterial community structure and functional potential as well as rapid increases in the expression of alkane monooxygenases and polyaromatic hydrocarbon-ring-hydroxylating dioxygenases 1 month after the bioremediation treatment commenced in the Alert soils (13)

Even the level of atmospheric pollution influences the structure of the microbial communities in three differently polluted sites rural, urban, and industrial (14). Microalgae, bacteria, rotifers, and testate amoebae biomasses were significantly higher in the rural site. Cyanobacteria biomass was significantly higher at the industrial site. Fungal and ciliate biomasses were significantly higher at the urban and industrial sites for the winter period and higher at the rural site for the spring period. These results suggest that microbial communities are potential bioindicators of atmospheric pollution.

Biodiversity of prokaryotic communities in sediments of different sub-basins of the Venice lagoon (15) has been demonstrated by the dominance of Gammaproteobacteria clones (84% with a high proportion of Vibrionaceae (Photobacterium)**,** indicator of urban pollution in the station adjacent to industrial and metropolitan areas. The relative importance of these pressures varies from place to place and very often, several pressures act in concert.

However, understanding ecosystem function, and predicting Earth's response to global changes such as warming and ocean acidification, calls for much better knowledge than we have today about microbial processes and interactions.

2.2. Reef Ecosystem

Most coral reefs are moderately to severely degraded by local human activities such as fishing and pollution as well as global change, hence it is difficult to separate local from global effects. Sandin *et al*. (16) surveyed coral reefs on uninhabited atolls in the northern Line Islands to provide a baseline of reef community structure, and on increasingly populated atolls to document changes associated with human activities. The authors found that top predators and reef-building organisms dominated unpopulated Kingman and Palmyra, while small planktivorous fishes and fleshy algae dominated the populated atolls of Tabuaeran and Kiritimati. Sharks and other top predators overwhelmed the fish assemblages on Kingman and Palmyra so that the biomass pyramid was inverted (top-heavy). In contrast, the biomass pyramid at Tabuaeran and Kiritimati exhibited the typical bottom-heavy pattern. Reefs without people exhibited less coral disease and greater coral recruitment relative to more inhabited reefs. Thus, protection from overfishing and pollution appears to increase the resilience of reef ecosystems to the effects of global warming.

Reef-building corals associate with many microbes. Best known are dinoflagellates in the genus *Symbiodinium* ("zooxanthellae"), which are photosynthetic symbionts. They are a large, genetically diverse group of which there is little information on the ecology of free-living stages and how different zooxanthellae perform as partners. Other microbial associates of reef corals are much less well known, but studies indicate that individual coral colonies host diverse assemblages of bacteria, some of which seem to have species-specific associations. This diversity of microbial associates has important evolutionary and ecological implications. Environmental stresses that incapacitate the ability of partners to reciprocate can destabilize associations by eliciting rejection by their hosts. Moreover, coral bleaching (the loss of zooxanthellae) and coral diseases, both increasing over the last several decades, may be examples of stress-related mutualistic instability (17).

Coral reefs, in one taxonomic and evolutionary guise or the other, have graced the Earth for about 500 million years and have survived several major extinction events. Most of these mass extinctions had a climatic component. Rapid climatic changes have

always caused major extinctions. Thus, given the currently observed rates of climatic change, there is reason to worry about the future of coral reefs. The greatest global-scale threats currently faced by coral reefs appear to be all linked to man-made or man-mediated changes of climate: 1. Bleaching, a heat- and light-mediated loss of symbiotic algae within the corals, has increased markedly in impact and severity over the past decades and affects virtually every reef worldwide. 2. Diseases have increased in incidence and diversity and caused severe population declines of corals. 3. Predator outbreaks have recurred repeatedly and have caused severe degradation on affected reefs. 4. Losses in keystone predators and herbivores have created phase shifts away from corals and to the establishment of stable states dominated by algae. 5. Ocean acidification is an emergent problem. 6. Runoff, sedimentation, and nutrient enrichment 7. Coastal construction leading to smothering of habitat and creation of high turbidity around coasts. 8. Overfishing and destructive fishing techniques. Experts in the field believe that, if the current trend of coral-reef degradation continues unabated, we will remain on the path of a mass coral extinction event. Corals will not likely go completely extinct, but the coral-reef ecosystems that currently harbor immense biodiversity, provide the necessities of life for millions of people, and produce valuable global economic services will disappear (18)

The study by Graham N.A.J *et al.* (19) shows for the first time the long-term impact of sea temperature rises on reef coral and fish communities. The results suggest that global warming may have had a more devastating effect on some of the world's finest coral reefs than previously assumed.

2.3. Climate Change

To date the climate change effects on biodiversity (such as changing distribution, migration and reproductive patterns) are already observable. In Europe, average temperatures are expected to rise by between 2^{o}C and 6.3^{o}C above 1990 temperatures by the year 2100. Predicted impacts associated with such temperature increase include a further rise in global mean sea level of 9 to 88 cm, more precipitation in temperate regions and Southeast Asia, associated with a higher probability of floods, less precipitation in Central Asia, the Mediterranean region, Africa, parts of Australia and New Zealand, associated with a greater probability of droughts, more frequent and

powerful extreme climatic events, such as heat waves, storms, and hurricanes, an expanded range of some dangerous “vector-borne diseases”, such as malaria, and further warming of the Arctic. Pollution from nutrients such as nitrogen, introduction of invasive species, over harvesting of wild animals can all reduce resilience of ecosystems. In the atmosphere, greenhouse gases such as water vapor, carbon dioxide, ozone, and methane act like the glass roof of a greenhouse by trapping heat and warming the planet. The natural levels of greenhouse gases are being supplemented by emissions resulting from human activities, such as the burning of fossil fuels, farming activities and land-use changes. As a result, the Earth's surface and lower atmosphere are warming. This will have profound effects on biodiversity.

UN Climate Change Conference (COP15), Copenhagen, Denmark, 7-18 December 2009. The Copenhagen climate conference ended by taking note of the 'Copenhagen Accord', which was supported by a large majority of Parties, including the European Union, but opposed by a small number. The conference also mandated the two *ad hoc* working groups on long-term cooperative action under the UN Framework Convention on Climate Change (UNFCCC) and on further commitments for developed countries under the Kyoto Protocol to complete their work at the next annual climate conference, to be held in Mexico City in November 2010. Though disappointing, the Copenhagen outcome is however a step in the right direction. The EU secured key elements of the Copenhagen Accord, which was negotiated among some 30 parties – many of them represented by their heads of state or government – from all UN regional groups during the course of 18 December and into the early hours of 19 December. These parties collectively represent more than 80% of global emissions. The Accord endorses for the first time at global level the objective of keeping warming to less than 2°C above the pre-industrial temperature. Another positive element is that it requires developed countries to submit economy-wide emission reduction targets, and developing countries to submit their mitigation actions, by 31 January 2010 so that they can be listed as part of the document. The Accord also lays the basis for a substantial ‘fast start’ finance package for developing countries approaching $30 billion for the period 2010-12, and medium-term financing of $100 billion annually by 2020. However, the Accord does not refer to the conclusion of a legally binding agreement, a key objective for the EU, or set the goal

of at least halving global emissions by 2050 compared to 1990 levels in order to keep warming below 2°C. The EU will continue to push for these. The European Commission's goal is now to ensure that a legally binding treaty is agreed in November 2010 in Mexico.

2.4. Effect of temperature on microbial communities

The effects of factors such as temperature, nutrient availability, grazing, salinity, seasonal cycle and carbon dioxide concentration have each been demonstrated to affect bacterial community structure in polar and alpine ecosystems (20). The results suggest that the spatial distribution of genetic variation and, hence, comparative rates of evolution, colonization and extinction are particularly important when considering the response of microbial communities to climate change. Although the direct effect of a change in e.g. temperature is known for very few Antarctic microorganisms, molecular techniques and genomic techniques are starting to give us an insight into what the potential effects of climate change might be at the molecular/cellular level.

In bacterial systematics species frequently contain unnamed and unrecognized populations defined ecotypes differing in physiology, genome content, and ecology. Bacterial responses to global warming can be better tracked with an ecotype-based systematics than current systematics (21). DNA sequence surveys are well suited to discovering ecologically distinct bacterial populations ('ecotypes'). Species have been demarcated for decades under the guidance of a universal criterion of genome content similarity, as quantified by DNA–DNA hybridization. More recently, species demarcation has been guided by divergence at the 16S rRNA locus, first with a 3% cut-off and more recently with a 1% cut-off. However, there is no theoretical rationale for these cut-offs to correspond to biologically significant clades (a group consisting of a single common ancestor and all its descendants with species-like properties), nor is it clear that any particular cut-off should apply to all bacteria. A theory-based approach has been proposed called ecotype simulation to derive cut-offs that are appropriate for demarcating a particular clade's ecotypes, allowing that different bacterial groups may have different cut-offs. Details of ecotype simulation may be found in a previous work (22), and the software may be downloaded from http://fcohan.web.wesleyan.edu/ecosim/.

The suggested ecotype simulation algorithm has proved capable of supporting investigation of replacements of one ecotype by another due to global warming and it has detected temperature-distinguished ecotypes invisible to the present bacterial systematics. Therefore, creating an ecotype-based systematics could help to identify the units of diversity to track to observe the early microbial responses to global warming.

Nemergut *et al*. (23) examined the diversity of bacterial, eucaryal, and archaeal 16S rRNA genes in tundra and talus soils across seasons in the alpine. This work has provided support for spatial and seasonal shifts in specific microbial groups, which correlate well with previously documented transitions in microbial processes. These preliminary results suggested that the physiologies of certain groups of organisms may scale up to the ecosystem level, providing the basis for testable hypotheses about the function of specific microbes in this system.

Results by Castro *et al*. (24) illustrate the potential for complex community changes in terrestrial ecosystems under climate change scenarios that alter multiple factors simultaneously. The authors measured the direct and interactive effects of climatic change on soil fungal and bacterial communities (abundance and composition) in a multi-factor climate change experiment that exposed a constructed old-field ecosystem to different atmospheric CO_2 concentration (ambient, +300 ppm), temperature (ambient, +3ºC), and precipitation (wet and dry). Fungal abundance increased in warmed treatments; bacterial abundance increased in warmed plots with elevated atmospheric CO_2, but decreased in warmed plots under ambient atmospheric CO_2; the phylogenetic distribution of bacterial and fungal clones and their relative abundance varied among treatments as indicated by changes in 16S rRNA and 18S rRNA genes; changes in precipitation altered the relative abundance of Proteobacteria and Acidobacteria where Acidobacteria decreased with a concomitant increase in the Proteobacteria in wet relative to dry treatments; changes in precipitation altered fungal community composition, primarily through lineage specific changes within a recently discovered group known as Soil Clone Group I (SCGI). These results indicate that climate change drivers and their interactions may cause changes in bacterial and fungal overall abundance; however changes in precipitation tended to have a much greater effect on the community composition.

Robador et al. (25) showed the impact of temperature on decline of specific groups of sulfate-reducing bacteria and confirmed a strong impact of increasing temperatures on the microbial community composition of arctic sediment (Svalbard). Conversely, in seasonally changing sediment (German Bight, North Sea) sulfate reduction rates and sulfate-reducing bacterial abundance changed little in response to changing temperature.

Using recent advances in molecular ecology, metagenomics, remote sensing of microorganisms and ecological modeling, it is now possible to achieve a comprehensive understanding of marine microorganisms and their susceptibility to environmental variability and climate change (26).

Chapter 3

New sequencing technologies in characterization of microbial diversity

Microbial diversity was revealed by exploring phylogenetic markers such as the rRNA genes. Such work revealed that the vast majority of microbial diversity had been missed by cultivation-based methods and that natural diversity was far more complex than was known. It is estimated that about 95-99% of microorganisms observable in nature are typically not cultivated using standard techniques (27). A bar-coded pyrosequencing approach targeting some hypervariable region of the bacterial 16S rRNA gene (e.g. V3, V6 regions) has allowed studies of the genetic diversity at significantly higher resolution compared to traditional fingerprinting methods (28), (29). However, single phylogenetic marker does not allow studies of whole genetic diversity as phylogeny based on a single gene is not directly associated with the metabolism. Today the aim to characterize complete microbial ecosystems by combining metagenomics, meta-transcriptomics and meta-metabolics to study microbial systems at the ecosystem level (eco-systems biology) is approaching (30). Above mentioned approaches are largely facilitated by ongoing revolution in sequencing technologies allowing already today massive sequencing producing millions of bases in a single day (31). The increased throughput makes possible to increase the sampling frequency for metagenomics, even sequence quickly several environmental microbial genomes. Moreover it is suggested that in the nearest future sequencing on the individual organism level will be available.

3.1. Massive throughput sequencing technologies

"Next-generation" sequencing (NGS) technologies aim to sequence genomes in a shorter time and a lower cost than traditional Sanger sequencing. These methods have different underlying biochemistries. They bypass the cloning of DNA fragments before sequencing, a necessary step for most Sanger sequencing, and this has resulted in the discovery of new microorganisms that previously had been missed because of cloning difficulties and biases (32).

454 –Roche pyrosequencing was among the first of so-called "next-generation" sequencing developed by 454 Life Sciences (33) (Figure 3). 454 pyrosequencing generates 1 million fragments (reads) which are shorter than conventional Sanger technique but compared to most of other technologies produce the longest read length (presently up to 400 bp).

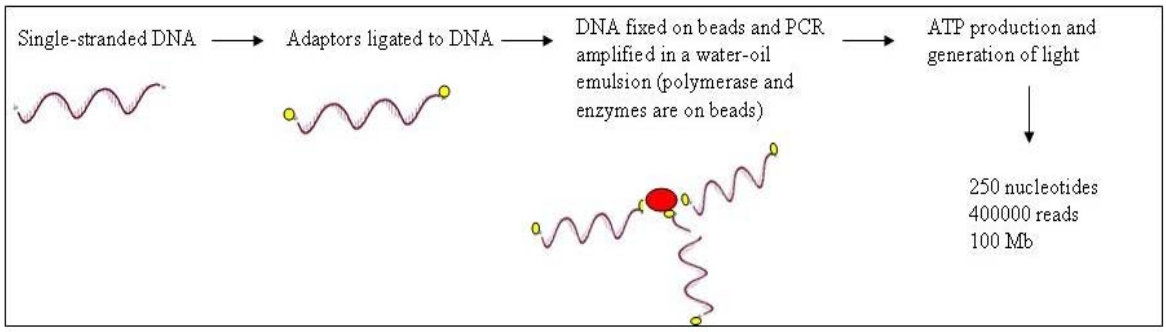


Figure 3. **Pyrosequecing method**. In the first step, oligonucleotide adaptors are ligated to fragmented DNA and immobilized to the surface of microscopic beads to perform PCR amplification in an oil-droplet emulsion. In the next step, beads are isolated in picolitre wells and incubated with dNTPs, DNA polymerase and beads bearing enzymes for the chemiluminescent reaction. Indeed, incorporation of a nucleotide into the complementary strand releases pyrophosphate, which is used to produce ATP. This, in turn, provides the energy for the generation of light. The light emitted recorded as an image for analysis.

Solexa GA – developed by Illumina was released in 2007. Solexa GA technology produces more nucleotides per run (1 Gbp data) with better accuracy (more than 99%) compared to pyrosequencing but with read length of 30-35 bp (Figure 4).

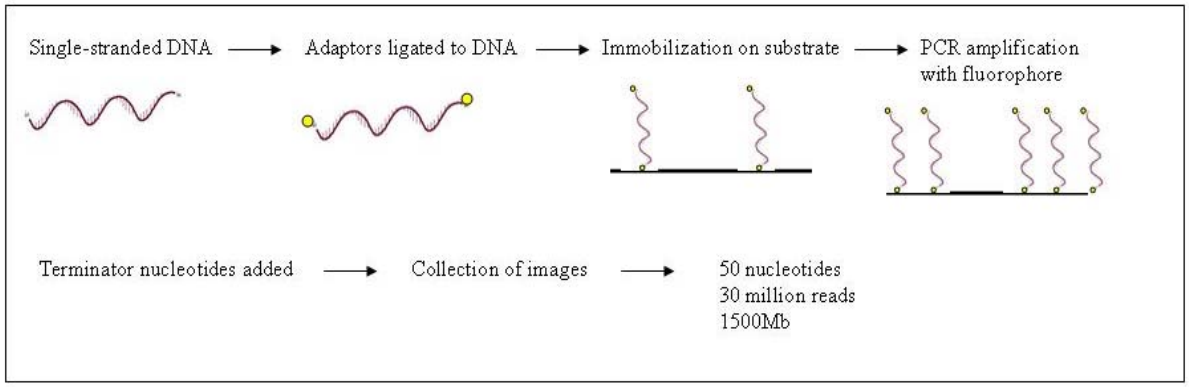


Figure 4. **Solexa GA sequencing**. The adaptors are ligated onto DNA and used to anchor the fragments to a prepared substrate. Fold-back PCR results in isolated spots of DNA of a large enough quantity that the amassed fluorophore can be detected. Terminator nucleotides and DNA polymerase are then used to create complementary-strand DNA. Images are collected at the end of each cycle before the terminator is removed.

SOLiD - this methodology is based on sequential ligation of oligonucleotides labeled with fluorochromes (Figure 5). SOLID generates up to 3 Gbp data (30-50 bp).

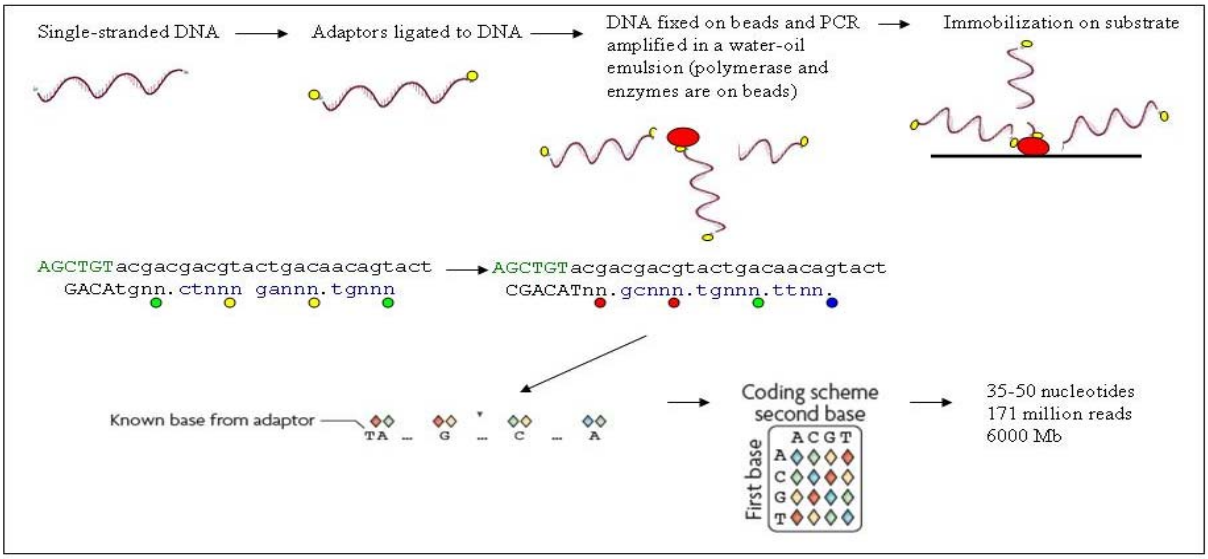


Figure 5. **SOLiD sample preparation**. After amplification, the beads are immobilized onto a custom substrate. A primer that is complementary to the adaptor sequence (green), random oligonucleotides with known 3' dinucleotides (blue) and a corresponding fluorophore (colored circles) are hybridized sequentially along the sequence and image data collected. After five repeats, the complementary strand is melted away and a new primer is added to the adaptor sequence, ending at a position one nucleotide upstream of

the previous primer. Second-strand synthesis is repeated, allowing two-color encoding and double reading of each of the target nucleotides. Repeats of these cycles ensure that nucleotides in the gap between known dinucleotides are read. Knowledge of the first base in the adaptor reveals the dinucleotide using the color-space scheme.

Helioscope - Applied Biosystems released its' own technology Heliscope (Figure 6) that sequences single molecules. The output consists of 50 nucleotides, 30 - 90 million reads and 500Mb with high accuracy (99.4%).

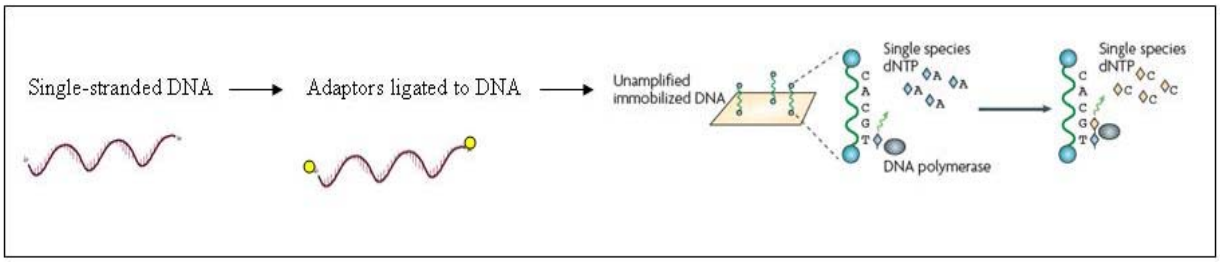


Figure 6. **Heliscope sequencing**. Unamplified DNA is immobilized with ligated adaptors to a substrate. Each species of dNTP with a bright fluorophore attached is used sequentially to create second-strand DNA; a 'virtual terminator' prevents the inclusion of more than one nucleotide per strand and cycle, and background signal is reduced by removal of 'used' fluorophore at the start of each cycle.

Pacific Biosciences is developing not yet commercially available sequencing method (Figure 7). The method is expected to be commercially released in 2010. Output (read length is expected to be several thousands bps) and data quality are not know.

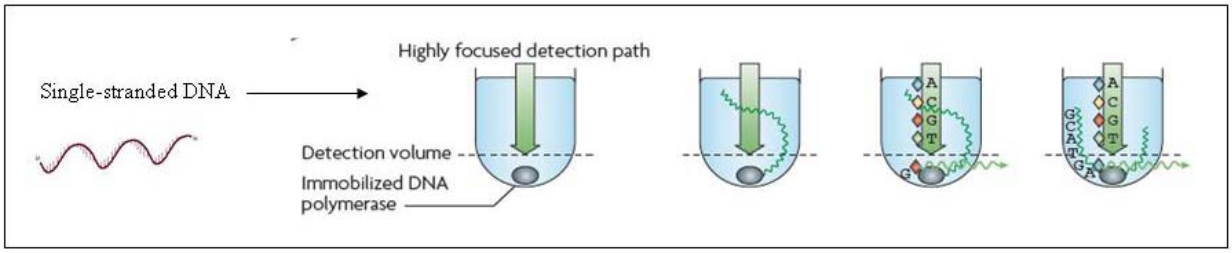


Figure 7. **Pacific Biosciences sequencing** occurs in zeptolitre wells that contain an immobilized DNA polymerase. DNA and dNTPs are added for synthesis. Fluorophores are cleaved from the complementary strand as it grows and diffuse away, allowing single nucleotides to be read. Continuous detection of fluorescence in the detection volume and high dNTP concentration allow extremely fast and long reading.

3.2 Data, databases and data analysis

"Next-generation" sequencing technologies bring up a huge amount of sequenced information. Until recently such genome or metagenome sequencing was almost entirely restricted to large genome centers, now it is feasible for individual laboratories. Next to computational resources, uncharacterized gene products with unknown function are likely to be the biggest bottleneck for the foreseeable future.

The major public database of genome nucleotide sequences is maintained by NCBI Entrez. Sequence data are stored in Entrez Genome (as complete chromosomes, plasmids, organelles, and viruses) and Entrez Nucleotide (as chromosome or genomic fragments such as contigs). The Genome Project database provides an umbrella view of the status of each genome project, links to project data in the other Entrez databases, and links to a variety of other NCBI and external resources associated with a defined genome project. Sequences associated with a defined organism can also be retrieved in the taxonomy browser. Due to massive release of NGS data (sort read sequences, SRSs) the major databases needed to be restructured and new databases appeared. The Table 1 include, behind the NCBI database, a list of other databases available for NGS data.

Table 1. **NGS related databases.**

Acronym	Full name/Description	Web site
Entrez Genome Project	NCBI subdatabase	http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj
Entrez Genome	NCBI subdatabase	http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome
Genomes OnLine Database	List of completed genome-sequencing projects	http://www.genomesonline.org
EMBL-EBI	Genome/gene expression database	http://www.ebi.ac.uk/Databases/genomes.html
MG-RAST	Tools for Rapid Annotations using Subsystems Technology	http://metagenomics.nmpdr.org
MEGAN	MEtaGenome ANalyzer	http://www-ab.informatik.uni-tuebingen.de/software/megan
MIGS	Minimum Information about a Genome Sequence	http://gensc.sf.net
MGED	Microarray and Gene Expression Database.	http://www.mged.org http://mbgd.genome.ad.jp
JGI	The Joint Genome Institute database	http://genome.jgi-psf.org/

Assembly (genomes or other genetic information) refers to the process of compiling a large number of short DNA sequences, and putting them back together to create a representation of the original whole sequence or partial fragments from which the DNA originated. A genome assembly algorithm works by taking all the pieces and aligning them to one another, and detecting all places where two of the short sequences, or reads, overlap. These overlapping reads can be merged together, and the process continues. Genome assembly is an extensive computational exercise, made more difficult because many genomes contain large numbers of identical sequences, known as repeats (Table 2). In addition, each sequencing technology has specific sources of biases and errors therefore dealing with data differing slightly. However, computational exercise is similar, new technologies produce shorter reads compared to Sanger sequencing. Fortunately, new assembly algorithms have been developed during the past few years that perform remarkably well even with reads as short as 30-35 nucleotides (34). On the other

hand, new technologies are improving in length of reads and therefore assembly may be eased even compared to Sanger technology.

Table 2. **Major assembly algorithms**

Algorithm	**Methods description**	**Software**
Greedy assemblers	The assembler greedily joins together the reads that are most similar to each other.	No used extensively
Overlap-layout-consensus	The relationships between read are represented as a graph, where the nodes represent each of the reads and an edge connects two nodes if the corresponding reads overlap. The assembly problem thus becomes the problem of identifying a path through the graph that contains all the nodes.	Roche Newbler
Eulerian path	Graph-based approach called de Bruijn graph. each edge is a k-mer that has been observed in the input data and implicitly represents a series of overlapping k-mers that overlap by a length of k–1	Velvet, Euler-SR, MIRA, Edena
Align-layout-consensus or assembly to reference	The overlap stage of short reads is replaced by an alignment step to a reference sequence.	Most of major assembly programs

For many applications, a draft genome-sequence assembly is sufficient and there is no need to invest in finishing. In addition, finished *de novo* assembly of sequence reads is not always necessary when comparing closely related strains; cataloguing polymorphisms relative to a reference genome sequence is often a satisfactory goal. The main goal of resequencing projects is generally to identify SNPs and other types of polymorphism, such as short insertions and deletions (collectively called indels). SNP discovery is essential for genetic mapping in eukaryotic organisms as they possess large genomes. However, SNP approach might be useful in ecological studies of microbes which otherwise need a vast sequencing due to high number of individual organisms.

Comparisons of microbial genomes widen possibilities to identify chromosomal rearragement events such as gene acquisition, duplications, deletions. On the other hand

using the complete genomes in phylogenetic analysis might lead to loss of phylogenetic signal – mainly due to lateral gene transfer (LGT). LGT results in variable phylogenetic histories across genes and is suggested to lead complicated or even completely defeating attempts to reconstruct bacterial evolution. High level of LTG may cause elusive phylogeny at organism level because we do not know which genes represent the true history of the cell lineages. However, the existence of core genes resistant to LGT has been proposed and is supported by some studies. Using complete genomes for phylogeny needs sufficient taxon sampling within a clade – yet rapidly increasing number of fully sequenced microbial genomes enables such taxon sampling (35).

Chapter 4

Applications for new sequencing technologies

Approaches, which are driven by whole genome sequencing and high-throughput functional genomics data, are revolutionizing studies on microbial biology. High-throughput sequencing technologies are the base for many applications e.g. metagenomics, metatranscriptomics, metaproteomics, and single amplified genomes.

4.1. Meta-approaches: metagenomics, metatranscriptomics, metaproteomics

Metagenomics is the analysis of genomic DNA obtained directly from whole community of organisms inhabiting environment (36). To date, the approach has been applied mostly to microbial communities (37). Metagenomics provides a view not only of the community structure (species phylogeny, richness, and distribution) but also of the functional (metabolic) potential of a community because virtually about all genes are captured and sequenced.

Metagenomic protocols begin with the extraction of genomic DNA from cellular organisms and/or viruses in an environmental sample; the DNA is then randomly sheared, these many short fragments are cloned, sequenced in either a random or targeted fashion and reconstructed into a consensus sequence (Figure 8).

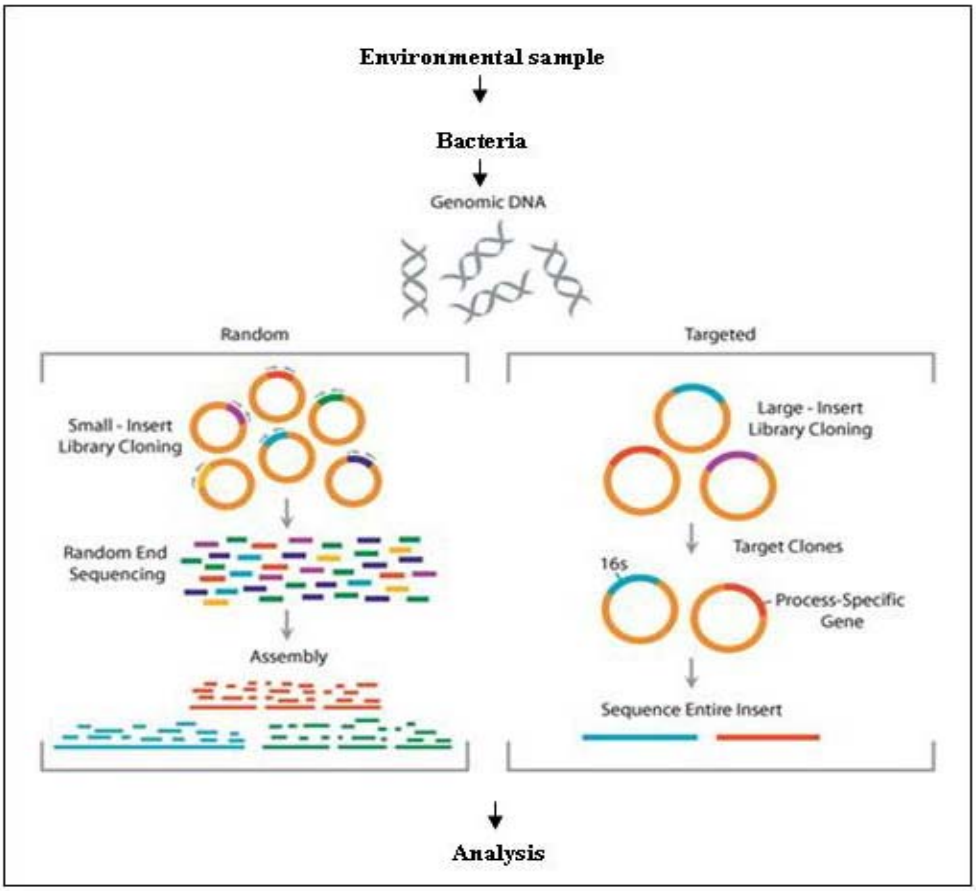


Figure 8. Model of a traditional metagenomics project by Sanger sequencing. The first step consists in the extraction of genomic DNA from an environmental sample. DNA is then sheared into fragments that are used in construction of a DNA clone library. Clone libraries are either small- or medium-insert (2-15 kb insert size) libraries or large-insert bacterial artificial chromosome (BAC) or fosmid libraries (up to 150 kb insert size), that may be sequenced in either a random or targeted fashion. In a "**random"** sequencing approach, the clones are randomly chosen and end-sequenced, and the resulting sequences are assembled into larger contiguous pieces ("contigs") by matching up overlapping sequences. Genes are then predicted from these sequence data using various methods**.** In a "**targeted**" sequencing approach, clones are first screened for the presence of a desirable gene (e.g., by PCR amplification) or a gene function (by functional assay).

Many genes may go unnoticed due to their "unclonability" in a heterologous or non-native host like *Escherichia coli* (most commonly used host for cloning libraries). Failure to produce clones representing these novel genes arises primarily due to their toxicity in *E. coli.* Basically, these genes may be too "foreign," and their expressed protein may cause failures in the operation of their host cell. New sequencing

technologies like 454 pyrosequencing can address this problem because they eliminate the cloning step by direct sequencing of extracted DNA.

In principle, any environment is amenable to metagenomic analysis provided that nucleic acids can be extracted from sample material and that they are of good quality. Most interest, however, has centered on the marine environment: the largest metagenomic study to date is the Global Ocean Sampling Expedition, which follows the voyage of Darwin's ship HMS Beagle. Of particular note is an international initiative, the Human Microbiome Project, which aims to map human-associated microbial communities (including those of the gut, mouth, skin and vagina) (37). Other metagenomics projects (also not related to microbial communities) include the study of the air over NY, the construction of metagenomics libraries from glacial ice (38), isolation of metalloproteases (39), the characterization of denitrification gene clusters (40), the cloning of a new cold-active lipase from a deep-sea sediment metagenome (41), the detection of pathogens in nasal and fecal specimens (42).

Viral communities were among the earliest to be studied using metagenomic approaches (37). A recent interesting paper by Dinsdale *et al*. (43) reports a metagenomic comparison of almost 15 million sequences from 45 distinct microbiomes and, for the first time, 42 distinct viromes. It shows that there are strongly discriminatory metabolic profiles across environments and that the magnitude of the metabolic capabilities encoded by the viromes is extensive. This suggests that metabolic profiles serve as a repository for storing and sharing genes among microbial hosts influencing global evolutionary and metabolic processes. This and other studies of viral communities point to a central role of viruses in microbial evolution and ecology. Initially, only double-stranded DNA viruses were accessible through cloning, but the newer cloneless sequencing technologies allow access to all types (such as single-stranded and RNA viruses). The results have revealed that between 65 and 95% of virus-derived sequences are unique to each metagenomic study, suggesting that virus-derived sequence is still massively under-represented in our databases.

Eukaryotes in general have much larger genomes and a higher proportion of DNA that doesn't code for proteins. As sequencing costs continue to fall, particularly with the development of higher-throughput technologies, eukaryotes should become a tractable

component of a metagenomic analysis. In the case of large multicellular eukaryotes such as humans, the equivalent of metagenomics is to sequence the genomes of many individuals (37).

High-throughput approaches may be used to analyze bioremediation of sites contaminated with hazardous and/or recalcitrant wastes (44). The strategy and outcome of bioremediation in open systems or confined environments depend on a variety of physico-chemical and biological factors that need to be assessed and monitored. In particular, microorganisms are key players in bioremediation applications, yet their catabolic potential and their dynamics *in situ* remain poorly characterized.

Metatranscriptomics refers to the analysis of the collective transcriptomes of a given habitat. Poretsky et al. (45) developed an environmental transcriptomic approach based on the direct retrieval and analysis of microbial transcripts from marine and freshwater bacterioplankton communities. They suggested that their environmental transcriptomic procedure may be a promising tool for exploring functional gene expression within natural microbial communities without bias toward known sequences. However this approach has not been tested yet for the analysis of microbial communities in contaminated sites.

Other recent papers have described new protocols for environmental metatranscriptome analysis using DNA microarrays (46) (47). While microarray-based metatranscriptome analysis undoubtedly provides valuable information about the response of microorganisms to environmental parameters, the information remains restricted to the number and nature of the probes spotted on the array. Frias-Lopez *et al.* (48) report a global analysis of expressed genes in a naturally occurring microbial community. Although many transcripts detected were highly similar to genes previously detected in ocean metagenomic surveys, a significant fraction (approximately 50%) were unique. Microbial community transcriptomic analyses revealed not only indigenous gene and taxon-specific expression patterns but also gene categories undetected in previous DNA-based metagenomic surveys.

Recently, a metatranscriptomic analysis of microbial communities during day/night in the North Pacific subtropical gyre has provided detailed information on metabolic and biogeochemical responses of a microbial community to solar forcing (49).

Environmental **metaproteomics** i.e. the study of the entire protein content of a given habitat is still in its infancy and faces great challenges in terms of protein extraction procedures (50), protein separation and identification, and bioinformatic tools to archive and analyze the huge amount of data generated by this approach (51, 52). Moreover the interpretation of protein expression levels in environmental organisms is a challenge due to the high genetic variability, the dependence on the nutritional and reproductive state of the organisms, as well as climatic and seasonal variations in the environment itself (52).

In metaproteomics, complex mixtures of proteins from an environmental sample are typically separated with two-dimensional (2D) gel electrophoresis or high performance liquid chromatography. Following protein separation, fractions of interest (e.g., protein spots on a 2D gel) are analyzed by high-throughput mass spectrometry based analytical platforms (53). Protein prediction and subsequent identification are greatly facilitated by available relevant metagenomic sequence data. So far only a few environmental metaproteomic studies have been achieved (54) (55) (56) (57).

4.2 Single amplified genomes

Direct sequencing from community DNA (i.e. metagenomics) is unsuitable for genome assemblies and metabolic reconstruction of the members of complex (i.e. most of natural communities (i.e. most of environmental communities) even with very large sequencing efforts. Luckily, DNA from individual cells can be amplified and analyzed by various means. Such new emerging strategy is called single amplified genomes (SAGs) approach (58). The multiple displacement amplification (MDA) method generates micrograms of DNA from the several femtograms present in a typical bacterial cell. MDA is based on isothermal (at 30°C) strand displacement synthesis in which the highly productive phi29 DNA polymerase repeatedly extends random primers on the template as it concurrently displaces previously synthesized copies (59).

Depending on desired throughput and the environment and organisms targeted, single cells have been isolated for use in MDA reactions by dilution, fluorescence activated cell sorting (FACS), micromanipulation, and microfluidics). Sorting by FACS has the best potential for high throughput technologies as using FACS one can isolate thousands of cells in minutes. Potentially, single-cell sorting can be combined with

fluorescent in situ hybridization (FISH) to enrich for specific taxa (58). Cells can be sorted into micro-plates facilitating automation.

Thereafter 2 basic approaches can be applied, downstream PCR (60) or SAGs genome sequencing and assembly (61):

(i) Downstream PCR of SAG is the direct analysis of multiple genes in individual marine bacteria cells, demonstrating the potential for high-throughput metabolic assignment of yet-uncultured taxa. The protocol uses a combination of high-speed fluorescence-activated cell sorting, whole-genome multiple displacement amplification (MDA), and subsequent PCR screening in metabolic mapping of taxonomically diverse uncultured marine bacterioplankton. A pilot library of 11 single amplified genomes (SAGs) was constructed from Gulf of Maine bacterioplankton as proof of concept. Proteorhodopsin genes were detected in two of five flavobacteria, providing evidence that Flavobacteria are major carriers of this photometabolic gene. It was also determined that Flavobacteria were a major component of HNA bacterioplanktonin the analyzed coastal sample. Fewer than 1% of the analyzed cells carried *nas*A, *puf*M, and *nif*H. Single-cell MDA provided access to the genomic material of numerically dominant but yet-uncultured taxonomic groups. Compared with metagenomics, the power of this approach lies in the ability to detect metabolic genes in uncultured microorganisms directly, even when the metabolic and phylogenetic markers are located far apart on the chromosome. Finally MDA and subsequent PCR sequencing is significantly less expensive than metagenomic sequencing.

(ii) whole genome sequencing from SAGs needs special approaches which are necessary to work with amplified DNA. MDA may not recover the entire genome from the single copy present in most bacteria. Also, some sequence rearrangements can occur during the DNA amplification reaction. Efforts continue to improve the MDA reaction enzymology to reduce bias and chimeric rearrangements. However, even with current limitations, single-cells sequencing will enable rapid progress identifying metabolic properties and ecological adaptations in the great numbers of uncultivated microorganisms. In addition, it provides a new method to examine patterns in interspecies and intraspecies genetic variation in evolutionary, phylogenetic and epidemiological studies. Single-cell sequencing, combined with metagenomics, will be a powerful tool for

addressing the complexity when species encompass a broad range of sequences and distinct boundaries between species are unclear. Single cell sequencing was adopted to obtain high-quality genome assemblies of two uncultured, numerically significant marine microorganisms (61).

Chapter 5

Legislation

Scientific research demonstrated that microbes have a very important role in natural ecosystems and that microbial biodiversity is also very important. Notwithstanding this microbial biodiversity is not even contemplated into legislation. The only indirect mention is included in coral reefs' protection regulations (Council Regulation (EC) N^{o} 1568/2005 of 20 September 2005 amending Regulation (EC) N^{o} 850/98 regarding the protection of deep-water coral reefs in certain areas of the Atlantic Ocean). Indeed, reef-building corals associate with many microbes and environmental stresses that incapacitate the ability of reef's partners to reciprocate can destabilize associations by eliciting rejection by their hosts (17).

As reported in the Proposal for a Directive of the European Parliament and of the Council establishing a Framework for Community Action in the field of Marine Environmental Policy (Marine Strategy Directive), **"**Increased knowledge of marine biochemical processes provides a wide range of opportunities, notably for the development of biotechnology." Therefore research should help to understand the biochemical part and obviously this should be also accounted into future legislation. "Indeed an improved knowledge base will be indispensable for guiding the development of policy actions and remediation measures".

In contrast to microbial biodiversity, general biodiversity legislations have been already developed as demonstrated by the numerous **European and International provisions** (Table 3). The most recent available report on progress is the Biodiversity Action Plan Report 2008 where a summary of progress in each Member State is included for the first time. The main conclusion is that the EU is highly unlikely to meet its 2010 target of halting biodiversity decline. The 2008 assessment highlights priority measures for the coming years. These range from more action to manage and restore sites to restoring ecosystem health and services in the wider EU countryside and in freshwater and marine environments. New issues, such as expansion of the agricultural sector to meet increasing demand for food, and the emergence of alternative market outlets such as biofuels, have emerged as major challenges. The seventh annual report (to end 2013) will

provide a similar evaluation, addressing also all post–2010 targets in the Action Plan. 2010 will be a major milestone for biodiversity policy both in the EU and globally. It will be the year of the full evaluation of the delivery to the EU Biodiversity Action Plan and as well the UN International Year for Biodiversity.

Table 3. European and International Provisions on biodiversity

European provisions

Year	Name	Content
1992	Birds and Habitats Directives "nature directives"	http://ec.europa.eu/environment/nature/legislation/birdsdirective/index_en.htm http://ec.europa.eu/environment/nature/legislation/habitatsdirective/index_en.htm
2001-2005	Millennium Ecosystem Assessment (MEA)	The most comprehensive evaluation of the ecological health of the planet ever undertaken: it assessed the damage and it presented ways to reverse ecosystem degradation and biodiversity loss.
2002	6th Environment Action Programme (6th EAP)	Ten-year program dedicated to biodiversity conservation, climate change, nature, flora, fauna, environment, health use of natural resources. It reduced impacts of point–source pollutants, such as those of urban waste waters on ecological status of rivers.
2003	CAP reform	Pro–biodiversity measures
2003?	Reformed Common Fisheries Policy (CFP)	It reduced fishing pressure and better protected non–target species and habitats.
2006	Biodiversity Action Plan	It faced the challenge of integrating biodiversity concerns into other policy sectors by planning of priority actions and responsibility of community institutions and Member States. It also contained indicators to monitor progress and a timetable for evaluations.

International provisions

Year	Name	Location	Content
1992	Convention on Biological Diversity (CBD)	Earth Summit in Rio de Janeiro	International framework for the conservation and sustainable use of biodiversity and the equitable sharing of its benefits. http://www.cbd.int/convention http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/04/28&format=HTML&aged=1&language=EN&guiLanguage=en
2002	Strategic Plan	Conference of the Parties (COP)	Three objectives: the conservation of biological diversity, the sustainable use of its components, and the fair and equitable sharing of the benefits from the use of genetic resources. Global headline indicators list.
2008		Bonn Biodiversity Meeting	Major financing mechanism for protected areas (forests), a fair sharing of genetic resources and for biofuels.

Conclusions

Although microorganisms are very important for the functioning of the whole biosphere, public knowledge, awareness and political actions do not deal with microbes when biodiversity and its decrease are in focus. Europe should focus on microbial biodiversity for important reasons. First of all the functioning of whole biosphere depends absolutely on the activities of the microbial world. Microbes have a fundamental role in the environment and in human health. In addition they have a potential role in key interdisciplinary areas like alternative energy/renewable energy (biofuels), and semi-artificial systems (sewage treatment plants, landfills, and in toxic waste bioremediation).

Molecular approach is fundamental in determining microbial biodiversity and new technologies contribute to explore biodiversity. As for other organisms many pressures and drivers are causing a decrease of microbial biodiversity. Microbes are complex and dynamic organisms and to better understand them it will be necessary to integrate knowledge in different fields and at at different levels molecular(genomic, transcriptomic and proteomic), cellular, population, ecosystem.

Acknowledgements

We would like to thank José Joaquín Blasco-Muñoz for the picture presented on cover page.

References

1. Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* 95: 6578-83
2. Fogg GE. 1986. *Picoplankton*. 1-30 pp.
3. Parnell JJ, Crowl TA, Weimer BC, Pfrender ME. 2009. Biodiversity in microbial communities: system scale patterns and mechanisms. *Mol Ecol* 18: 1455-62
4. Campbell L, Nolla HA, Vaulot D. 1994. The importance of Prochlorococcus to community structure in the central North Pacific Ocean. *Limnology and Oceanography* 39: 954-61
5. Zehr JP, Waterbury JB, Turner PJ, Montoya JP, Omoregie E, Steward GF, Hansen A, Karl DM. 2001. Unicellular cyanobacteria fix N2 in the subtropical north Pacific Ocean. *Nature* 412: 635-8
6. Giovannoni JSaRMS. 2000. Evolution, diversity, and molecular ecology of marine prokaryotes. *Microbial Ecology of the Oceans, Wiley-Liss Inc* 47–84
7. Suttle CA. 2007. Marine viruses - Major players in the global ecosystem. *Nature Reviews Microbiology* 5: 801-12
8. Azam F, Fenchel, T., Field, I., Gray, I., Meyer-Reil, L. and Thingstad, F. 1983. The ecological role of water column microbes in the sea. *Mar. Ecol. Prog. Ser* 10145: 257-63
9. Gruber N, Galloway JN. 2008. An Earth-system perspective of the global nitrogen cycle. *Nature* 451: 293-6
10. Ogawa H, Amagai Y, Koike I, Kaiser K, Benner R. 2001. Production of refractory dissolved organic matter by bacteria. *Science* 292: 917-20
11. Zhang W, Wang H, Zhang R, Yu XZ, Qian PY, Wong MH. Bacterial communities in PAH contaminated soils at an electronic-waste processing center in China. *Ecotoxicology* 19: 96-104
12. Rodriguez-Blanco A, Antoine V, Pelletier E, Delille D, Ghiglione JF. 2009. Effects of temperature and fertilization on total vs. active bacterial communities exposed to crude and diesel oil pollution in NW Mediterranean Sea. *Environ Pollut*
13. Yergeau E, Arbour M, Brousseau R, Juck D, Lawrence JR, Masson L, Whyte LG, Greer CW. 2009. Microarray and real-time PCR analyses of the responses of high-arctic soil bacteria to hydrocarbon pollution and bioremediation treatments. *Appl Environ Microbiol* 75: 6258-67
14. Meyer C, Gilbert D, Gaudry A, Franchi M, Nguyen-Viet H, Fabure J, Bernard N. 2009. Relationship of Atmospheric Pollution Characterized by Gas (NO(2)) and Particles (PM10) to Microbial Communities Living in Bryophytes at Three Differently Polluted Sites (Rural, Urban, and Industrial). *Microb Ecol*
15. Borin S, Brusetti L, Daffonchio D, Delaney E, Baldi F. 2009. Biodiversity of prokaryotic communities in sediments of different sub-basins of the Venice lagoon. *Res Microbiol* 160: 307-14
16. Sandin SA, Smith JE, Demartini EE, Dinsdale EA, Donner SD, Friedlander AM, Konotchick T, Malay M, Maragos JE, Obura D, Pantos O, Paulay G, Richie M, Rohwer F, Schroeder RE, Walsh S, Jackson JB, Knowlton N, Sala E. 2008.

Baselines and degradation of coral reefs in the Northern Line Islands. *PLoS One* 3: e1548
17. Knowlton N, Rohwer F. 2003. Multispecies microbial mutualisms on coral reefs: the host as a habitat. *Am Nat* 162: S51-62
18. Riegl B, Bruckner A, Coles SL, Renaud P, Dodge RE. 2009. Coral reefs: threats and conservation in an era of global change. *Ann N Y Acad Sci* 1162: 136-86
19. Graham NA, Wilson SK, Jennings S, Polunin NV, Bijoux JP, Robinson J. 2006. Dynamic fragility of oceanic coral reef ecosystems. *Proc Natl Acad Sci U S A* 103: 8425-9
20. Pearce DA. 2008. Climate change and the microbiology of the Antarctic Peninsula region. *Sci Prog* 91: 203-17
21. Cohan FM. 2009. Tracking bacterial responses to global warming with an ecotype-based systematics. *Clin Microbiol Infect* 15 Suppl 1: 54-9
22. Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, Rooney AP, Brambilla E, Connor N, Ratcliff RM, Nevo E, Cohan FM. 2008. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci U S A* 105: 2504-9
23. Nemergut DR, Costello EK, Meyer AF, Pescador MY, Weintraub MN, Schmidt SK. 2005. Structure and function of alpine and arctic soil microbial communities. *Res Microbiol* 156: 775-84
24. Castro HF, Classen AT, Austin EE, Norby RJ, Schadt CW. 2009. Soil Microbial Community Responses to Multiple Experimental Climate Change Drivers. *Appl Environ Microbiol*
25. Robador A, Bruchert V, Jorgensen BB. 2009. The impact of temperature change on the activity and community composition of sulfate-reducing bacteria in arctic versus temperate marine sediments. *Environ Microbiol*
26. Karl DM. 2007. Microbial oceanography: paradigms, processes and promise. *Nat Rev Microbiol* 5: 759-69
27. Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180: 4765-74
28. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* 103: 12115-20
29. Miller SR, Strong AL, Jones KL, Ungerer MC. 2009. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Applied and Environmental Microbiology* 75: 4565-72
30. Raes J, Bork P. 2008. Molecular eco-systems biology: Towards an understanding of community function. *Nature Reviews Microbiology* 6: 693-9
31. Metzker ML. Sequencing technologies the next generation. *Nature Reviews Genetics* 11: 31-46
32. Blow N. 2008. Metagenomics: exploring unseen communities. *Nature* 453: 687-90
33. Ronaghi M. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Research* 11: 3-11

34. Chaisson MJ, Brinza D, Pevzner PA. 2009. De novo fragment assembly with short mate-paired reads: Does the read length matter? *Genome Research* 19: 336-46
35. Lerat E, Daubin V, Moran NA. 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. *PLoS Biol* 1: E19
36. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5: R245-9
37. Hugenholtz P, Tyson GW. 2008. Microbiology: metagenomics. *Nature* 455: 481-3
38. Simon C, Herath J, Rockstroh S, Daniel R. 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol*
39. Waschkowitz T, Rockstroh S, Daniel R. 2009. Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl Environ Microbiol* 75: 2506-16
40. Demaneche S, Philippot L, David MM, Navarro E, Vogel TM, Simonet P. 2009. Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach. *Appl Environ Microbiol* 75: 534-7
41. Jeon JH, Kim JT, Kim YJ, Kim HK, Lee HS, Kang SG, Kim SJ, Lee JH. 2009. Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Appl Microbiol Biotechnol* 81: 865-74
42. Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T. 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS ONE* 4: e4219
43. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F. 2008. Functional metagenomic profiling of nine biomes. *Nature* 452: 629-32
44. Stenuit B, Eyers L, Schuler L, Agathos SN, George I. 2008. Emerging high-throughput approaches to analyze bioremediation of sites contaminated with hazardous and/or recalcitrant wastes. *Biotechnol Adv* 26: 561-75
45. Poretsky RS, Bano N, Buchan A, LeCleir G, Kleikemper J, Pickering M, Pate WM, Moran MA, Hollibaugh JT. 2005. Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol* 71: 4121-6
46. Gao H, Yang ZK, Gentry TJ, Wu L, Schadt CW, Zhou J. 2007. Microarray-based analysis of microbial community RNAs by whole-community RNA amplification. *Appl Environ Microbiol* 73: 563-71
47. Parro V, Moreno-Paz M, Gonzalez-Toril E. 2007. Analysis of environmental transcriptomes by DNA microarrays. *Environ Microbiol* 9: 453-64
48. Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW, Delong EF. 2008. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* 105: 3805-10

49. Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol* 11: 1358-75
50. Maron PA, Ranjard L, Mougel C, Lemanceau P. 2007. Metaproteomics: a new approach for studying functional microbial ecology. *Microb Ecol* 53: 486-93
51. Wilke A, Ruckert C, Bartels D, Dondrup M, Goesmann A, Huser AT, Kespohl S, Linke B, Mahne M, McHardy A, Puhler A, Meyer F. 2003. Bioinformatics support for high-throughput proteomics. *J Biotechnol* 106: 147-56
52. Nesatyy VJ, Suter MJ. 2007. Proteomics for the analysis of environmental stress responses in organisms. *Environ Sci Technol* 41: 6891-900
53. Domon B, Aebersold R. 2006. Mass spectrometry and protein analysis. *Science* 312: 212-7
54. Wilmes P, Bond PL. 2004. The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* 6: 911-20
55. Wilmes P, Bond PL. 2006. Towards exposure of elusive metabolic mixed-culture processes: the application of metaproteomic analyses to activated sludge. *Water Sci Technol* 54: 217-26
56. Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, 2nd, Shah M, Hettich RL, Banfield JF. 2005. Community proteomics of a natural microbial biofilm. *Science* 308: 1915-20
57. Benndorf D, Balcke GU, Harms H, von Bergen M. 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *Isme J* 1: 224-34
58. Ishoey T, Woyke T, Stepanauskas R, Novotny M, Lasken RS. 2008. Genomic sequencing of single microbial cells from environmental samples. *Current Opinion in Microbiology* 11: 198-204
59. Dean FB, Nelson JR, Giesler TL, Lasken RS. 2001. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research* 11: 1095-9
60. Stepanauskas R, Sieracki ME. 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci U S A* 104: 9052-7
61. Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H, Saw JH, Senin P, Yang C, Chatterji S, Cheng JF, Eisen JA, Sieracki ME, Stepanauskas R. 2009. Assembling the marine metagenome, one cell at a time. *PLoS One* 4: e5299

European Commission

EUR 24243 EN – Joint Research Centre – Institute for Environment and Sustainability

Title: Microbial Biodiversity and Molecular Approach

Author(s): Elena Nardini, Veljo Kisand and Teresa Lettieri

Luxembourg: Office for Official Publications of the European Communities

2010 – 47 pp. 21 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1018-5593

ISBN 978-92-79-14990-0

DOI 10.2788/60582

Abstract

Biodiversity is given by the variety of species on Earth resulting from billions of years of evolution. Molecular-phylogenetic studies have revealed that the main diversity of life is microbial and it is distributed among three domains: Achaea, Bacteria, and Eukarya. The functioning of whole biosphere depends absolutely on the activities of the microbial world. Due to their versatility microbes are the major natural providers of ecological services as well play major role in semi-artificial systems such as sewage treatment plants, landfills, and in toxic waste bioremediation.

As for other organisms many pressures and drivers are causing decrease of microbial biodiversity. Several publications document the effect of chemical pollutants e.g. Polycyclic Aromatic Hydrocarbons (PAHs), of atmospheric pollution, of temperature change and of fertilization on microbial community structure.

These studies are now possible because sequencing technologies are in ongoing revolution allowing massive de novo sequencing producing millions of bases in a single day. Metagenomics, metatranscriptomics, metaproteomics and single-cell sequencing are approaches providing a view not only of the community structure (species phylogeny, richness, and distribution) but also of the functional (metabolic) potential of a community because virtually about all genes are captured and sequenced.

Unfortunately, although microrganisms are very important for the functioning of whole biosphere public knowledge, awareness and political actions did not yet deal with microbes when biodiversity and its decrease are highlighted.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-24243-EN-C

ISBN 978-92-79-14990-0

Publications Office
Publications.europa.eu