Genomics in marine monitoring: New opportunities for assessing marine health status

Sarah J. Bourlat[1], Ángel Borja[2], Jack Gilbert[3], Martin I. Taylor[4], Neil Davies[5,6], Stephen B. Weisberg[7], John Griffith[7], Teresa Lettieri[8], Dawn Field[6,9], John Benzie[10,11], Frank Oliver Glöckner[12], Naiara Rodríguez-Ezpeleta[2], Daniel P Faith[13], Tim P. Bean[14] and Matthias Obst[1]

Abstract

This viewpoint paper explores the potential of genomics technology to provide accurate, rapid, and cost efficient observations of the marine environment. The use of such approaches in next generation marine monitoring programs will help achieve the goals of marine legislation implemented world-wide. Genomic methods can yield faster results from monitoring, easier and more reliable taxonomic identification, as well as quicker and better assessment of the environmental status of marine waters. A summary of genomic methods that are ready or show high potential for integration into existing monitoring programs is provided (e.g., qPCR, SNP based methods, DNA barcoding, microarrays, metagenetics, metagenomics, transcriptomics). These approaches are mapped to existing indicators and descriptors and a series of case studies is presented to assess the cost and added value of these molecular techniques in comparison with traditional monitoring systems. Finally, guidelines and recommendations are suggested for how such methods can enter marine monitoring programs in a standardized manner.

[1]*University of Gothenburg, Department of Biological and Environmental Sciences, Gothenburg, Sweden*

[2]*AZTI-Tecnalia, Marine Research Division, Pasaia, Spain*

[3]*University of Chicago and Argonne National Laboratory, Argonne, IL*

[4]*Bangor University, School of Biological Sciences, UK*

[5]*University of California Berkeley, Gump South Pacific Research Station, Moorea, French Polynesia*

[6]*University of Oxford, Biodiversity Institute, Department of Zoology, Oxford, UK*

[7]*Southern California Coastal Water Research Project, Costa Mesa, CA*

[8]*European Commission, Institute for Environment and Sustainability, DG Joint Research Centre, Ispra, Italy*

[9]*Centre for Ecology and Hydrology, Oxford, UK*

[10]*University College Cork, School of Biological Earth and Environmental Sciences, Cork, Ireland*

[11]*WorldFish Centre, Jalan Batu Maung, Penang, Malaysia*

[12]*Jacobs University, Bremen and Max Planck Institute for Marine Microbiology, Bremen, Germany*

[13]*The Australian Museum, Sydney, Australia*

[14]*Cefas Weymouth Laboratory Centre for Environment Fisheries and Aquaculture Science, Weymouth, UK*

Introduction

In response to the increasing human impact on our oceans (Pew Oceans Commission 2003, Ban and Alder 2008, Halpern *et al.* 2008, Claudet and Fraschetti 2010, Lotze 2010), legislation has been implemented world-wide to protect, conserve or enhance marine ecosystems, proposing integrative tools and methods to assess ecological integrity and marine health status (Borja *et al.* 2008).

The United Nations Convention on the Law of the Sea (UNCLOS 1982) is the international basic legal framework that governs the use of the oceans and seas, establishing an international obligation to protect and use the resources of the marine environment sustainably; it is further supported by the 1992 Convention on Biological Diversity (CBD 2000). At a national or regional level, several initiatives have been developed (for details, see Borja *et al.* 2008), such as: i) Oceans Policy, in Australia; ii) Oceans Act and Oceans Strategy, in Canada; iii) Oceans Act, in the USA; iv) the Water Framework Directive (WFD; 2000/60/EC), and the Marine Strategy Framework Directive (MSFD; 2008/56/EC), in Europe; v) the National Water Act, in South Africa; and vi) several laws on water and ocean quality, in the People's Republic of China.

These initiatives try to make sustainable use of the seas compatible with the conservation of marine ecosystems and the maintenance of a good status for marine waters, habitats and resources. Status is assessed in an integrative way including measurement of many components of the ecosystem together with physico-chemical parameters and elements of pollution. This approach is intended to provide an 'ecosystem-based management' of marine waters (Apitz *et al.* 2006, Barnes and McFadden 2008, Lester *et al.* 2010). This concept takes into account the structure, function and processes of marine ecosystems bringing together natural physical, chemical, physiographic, geographic and climatic factors, and integrating them with anthropogenic impacts and activities in the area concerned (Borja *et al.* 2008).

To undertake such an assessment, the above-mentioned marine legislation requires adequate and rigorous monitoring at different spatial and temporal scales. Despite the importance of monitoring, in terms of non-compliance with a threshold and the subsequent need for (expensive) policy and managerial actions, the current global economic crisis, and especially cuts in government spending, are leading many countries (and industries) to try and save on their monitoring budgets (Borja and Elliott 2013). This has added further motivation for investigating new, more cost-effective methods to monitor and assess marine waters (Frolov *et al.* 2013), and the innovative application of recent scientific advances.

Genomics, the science that uses nucleotide sequences (DNA or RNA) to analyze biological systems, represents perhaps the most likely source of innovation in marine monitoring techniques. There is great potential for the development of genomic techniques for *in situ* detection and monitoring of the biodiversity, abundance and activity of organisms (Minster and Connolly 2006), and novel sequencing technologies (Mardis 2008) have led to an enormous increase in the amount of genetic data available on organisms, communities, and habitats over the last decade (Hajibabaei *et al.* 2011, Bik *et al.* 2012, Radom *et al.* 2012). As a result of this development, the assembly and analysis of nucleotide data has become routine methodology in most biological disciplines, including marine biodiversity (e.g., De Long 2005, Karsenti *et al.* 2011, Glöckner 2012, Roger *et al.* 2012, Teeling and Glöckner 2012). Following this trend, the methods of genomic analysis are being continuously modified and refined in order to serve new purposes and applications in conservation biology and monitoring programs (e.g., the projects FishPopTrace (https://fishpoptrace.jrc.ec.europa.eu/) and DEVOTES (www.devotes-project.eu)). This process is closely coordinated with the development of bioinformatic and e-science tools that integrate genomic information into conventional data streams (e.g., BiSciCol (http://biscicol.blogspot.com); BioVeL (http://www.biovel.eu)), and has opened up enormous opportunities for analysing patterns, functions, and processes in marine environments.

This collaborative viewpoint paper explores the potential of genomics to provide accurate, rapid, and cost efficient observations of the marine environment. These approaches are likely to be especially useful in next generation marine monitoring programs currently designed to help achieve the goals of marine legislation being implemented world-wide.

The Need to Monitor: An Example from Legislation

The MSFD in Europe provides a good example of the policy approaches developed using current concepts of ecosystem-based management, and can be used to illustrate a framework for the discussion of genomic technologies in relation to marine environmental assessment. The MSFD aims to achieve or maintain 'good environmental status' (GES) in EU waters by 2020. The status is defined by 11 descriptors (e.g., alien species, fishing, eutrophication, seafloor integrity, etc.), and the maintenance of biodiversity is a cornerstone of GES (Cochrane *et al.* 2010). A series of associated 'criteria' and 'indicators' for each descriptor will be used to decide on the status of marine ecosystems (Table 1). Expert groups have defined 29 criteria and 56 indicators to determine this status (Cardoso *et al.* 2010).

There are still significant gaps in the understanding of marine ecosystems, and in the knowledge required to achieve an ecosystem-based management policy that integrates all of the above MSFD indicators (Borja *et al.* 2010). For example, in many cases, important baseline knowledge needed to define GES of European marine ecosystems is missing, although several attempts to assess status have been published (HELCOM 2010, Borja *et al.* 2011).

Bottlenecks in Assessing Marine Health using Current Marine Monitoring Methods

Marine environmental monitoring is highly 'station oriented' (focused on a few permanent/regular sampling sites) and usually limited to observations of specific groups of organisms (e.g., benthic macroinvertebrates, phytoplankton, or fish) with little consistency in observation methods across ecosystems (de Jonge *et al.* 2006, Elliott 2011). As a consequence, policy decisions are often based on limited and/or biased data, which may significantly constrain policy development. In particular, traditional methods for species identification have a number of shortfalls, listed in Table 2.

Many inventories used in monitoring are difficult to compare and are often of low and/or unverifiable taxonomic precision. In addition, the targeting of selected taxa means that the relevance of these data to other groups (e.g., planktonic, meiofaunal, microorganisms), other life stages (e.g., larvae), and to ecological processes in general, is not always clear. Ideally, an informed choice of what to monitor would be based on studies that include all taxa (including animals, plants, fungi, protists and bacteria) and life stages. In particular, microbial community interactions and their metabolic pathways are emerging as essential components of any comprehensive estimate of ecosystem function.

Currently, there are no genomic methods implemented for the assessment of MSFD indicators, and few genetic methods are considered for contribution to the MSFD. Yet, some of the indicators of biodiversity (e.g., species distribution, population genetic structure; see table 1 for a comprehensive list) could benefit from DNA-based techniques. All molecular approaches that could improve monitoring programs are informed by the increasing knowledge of the variation found among whole genomes within and between species across the tree of life. The emerging science of 'biodiversity genomics' addresses this issue, which was a major theme in a recent Genomic Observatories Network (http://genomicobservatories.org/) meeting (Davies *et al.* In press). Examples of the application of this knowledge includes DNA-based tools for the identification of species, and the ratio between alien and native species in samples, providing useful information for the non-indigenous species descriptor in the MSFD. The accuracy and comprehensiveness of other indicators, related to human-induced eutrophication and seafloor integrity descriptors, might also be assisted by the use of genomic tools (see Table 3).

Genomic Methods Relevant to Assess Marine Health

New tools based on genomic methods could be used to address the bottlenecks in assessing marine health, and can therefore be applied to improve current practices; see examples from case-studies world-wide in Table 3.

DNA Barcoding

DNA barcoding consists in assigning a specimen or sample (e.g., a piece of tissue or contents of a gut) to species by sequencing a standardized short DNA fragment (the 'DNA barcode') and comparing it against a reference database (Hebert *et al.* 2003). This technique has the advantage of being independent of the user's taxonomic expertise and makes it possible to assign species names

Table 1. Qualitative descriptors and different criteria and indicators, to be used in environmental status assessment, within the Marine Strategy Framework Directive, selected by the European Commission (2010). Asterisks show the indicators for which genomics could be used in monitoring and assessment.

Descriptor	Criteria	Indicator
1: Biological diversity	1.1 Species distribution	1.1.1 Distributional range*
		1.1.2 Distributional pattern within the latter*
		1.1.3 Area covered by the species (for sessile/benthic species)
	1.2 Population size	1.2.1 Population abundance and/or biomass
	1.3 Population condition	1.3.1 Population demographic characteristics
		1.3.2 Population genetic structure*
	1.4 Habitat distribution	1.4.1 Distributional range
		1.4.2 Distributional pattern
	1.5 Habitat extent	1.5.1 Habitat area
		1.5.2 Habitat volume, where relevant
	1.6 Habitat condition	1.6.1 Condition of the typical species* and communities
		1.6.2 Relative abundance and/or biomass, as appropriate
		1.6.3 Physical, hydrological and chemical conditions
	1.7 Ecosystem structure	1.7.1 Composition and relative proportions of ecosystem components (habitats, species)
2: Non-indigenous species	2.1 Abundance and state of non-indigenous species, in particular invasive species	2.1.1 Trends in abundance, temporal occurrence and spatial distribution of non-indigenous species*
	2.2 Environmental impact of invasive non-indigenous sp.	2.2.1 Ratio between invasive non-indigenous species and native species*
		2.2.2 Impacts of non-indigenous invasive species at the level of species, habitats and ecosystem*
3: Exploited fish and shellfish	3.1 Level of pressure of the fishing activity	3.1.1 Fishing mortality (F)
		3.1.2 Catch/biomass ratio
	3.2 Reproductive capacity of the stock	3.2.1 Spawning Stock Biomass (SSB)
		3.2.2 Biomass indices
	3.3 Population age and size distribution	3.3.1 Proportion of fish larger than the mean size of first sexual maturation
		3.3.2 Mean maximum length across all species found in research vessel surveys
		3.3.3 95% percentile of the fish length distribution observed in research vessel surveys
		3.3.4 Size at first sexual maturation
4: Food webs	4.1 Productivity of key species or trophic groups	4.1.1 Performance of key predator species using their production per unit biomass
	4.2 Proportion of selected species at the top of food webs	4.2.1 Large fish (by weight)
	4.3 Abundance/distribution of key trophic	4.3.1 Abundance trends of functionally important selected groups/species*
5: Human-induced eutrophication	5.1. Nutrients levels	5.1.1 Nutrients concentration in the water column
		5.1.2 Nutrient ratios (silica, nitrogen and phosphorus)
	5.2 Direct effects of nutrient enrichment	5.2.1 Chlorophyll concentration in the water column
		5.2.2. Water transparency related to increase in suspended algae
		5.2.3 Abundance of opportunistic macroalgae*
		5.2.4 Species shift in floristic composition such as diatom to flagellate ratio, benthic to pelagic shifts, as well as bloom events of nuisance/toxic algal blooms caused by human activities*
	5.3 Indirect effects of nutrient enrichment	5.3.1 Abundance of perennial seaweeds and seagrasses impacted by decrease in water transparency*
		5.3.2 Dissolved oxygen changes and size of the area concerned
6: Seafloor integrity	6.1 Physical damage, having regard to substrate characteristics	6.1.1 Type, abundance, biomass and areal extent of relevant biogenic substrate
		6.1.2 Extent of the seabed significantly affected by human activities for the different substrate types
	6.2 Condition of benthic community	6.2.1 Presence of particularly sensitive and/or tolerant species*
		6.2.2 Multi-metric indices assessing benthic community condition and functionality, such as species diversity and richness, proportion of opportunistic to sensitive species*
		6.2.3 Proportion of biomass or number of individuals in the macrobenthos above specified length/size
		6.2.4 Parameters describing the characteristics of the size spectrum of the benthic community

Table 1. Continued

Descriptor	Criteria	Indicator
7: Hydrographical conditions	7.1 Spatial characterisation of permanent alterations	7.1.1 Extent of area affected by permanent alterations
	7.2 Impact of permanent hydrographical changes	7.2.1 Spatial extent of habitats affected by the permanent alteration
		7.2.2 Changes in habitats, in particular the functions provided due to altered hydrographical conditions
8: Contaminants	8.1 Concentration of contaminants	8.1.1 Concentration of the contaminants measured in matrices such as biota, sediment and water
	8.2 Effects of contaminants	8.2.1 Levels of pollution effects on the ecosystem components concerned, having regard to the selected biological processes and taxonomic groups where a cause/effect relationship has been established
		8.2.2 Occurrence, origin, extent of significant acute pollution events and their impact on biota physically affected by this pollution
9: Contaminants in fish and seafood	9.1 Levels, number and frequency of contaminants	9.1.1 Actual levels of contaminants that have been detected and number of contaminants which have exceeded maximum regulatory levels
		9.1.2 Frequency of regulatory levels being exceeded
10: Litter	10.1 Characteristics of litter in the marine and coastal environment	10.1.1 Trends in the amount of litter washed ashore and/or deposited on coastlines, including analysis of its composition, spatial distribution and source
		10.1.2 Trends in the amount of litter in the water column and deposited on the seafloor
		10.1.3 Trends in the amount, distribution and composition of micro-particles
	10.2 Impacts of litter on marine life	10.2.1 Trends in the amount and composition of litter ingested by marine animals
11: Energy and noise	11.1 Distribution in time and place of loud, low and mid frequency impulsive sounds	11.1.1 Proportion of days and their distribution within a calendar year over areas of a determined surface, as well as their spatial distribution, in which anthropogenic sound sources exceed levels that are likely to entail significant impact
	11.2 Continuous low frequency sound	11.2.1 Trends in the ambient noise level within the 1/3 octave bands 63 and 125 Hz (centre frequency) measured by observation stations and/or with the use of models

Table 2. Shortfalls in traditional monitoring methods.

Shortcomings in Current Monitoring Methods	Explanation for Shortcomings
Maintaining a consistent and high quality standard of species-level identification	Monitoring programs vary in their spatial, temporal and qualitative taxonomic coverage
Providing a good estimate based on all biodiversity in monitored marine waters	Biological monitoring relies on the identification of selected groups of species and the relative abundance of individuals belonging to the 'vulnerable' or 'disturbance tolerant' species because those are the species (and life stages) it is easy to count
Providing, rapid, cost-efficient and scalable species identification for monitoring and traceability purposes	Species identification relies on the specialized knowledge of taxonomic experts, which for many species (in fact, for virtually all species if one considers the full range of taxa in a system) is time consuming, costly, unreliable, low throughput, and difficult to use for large scale monitoring programs
Discovering species genetic diversity at the population level	The lack of precise population estimates makes it difficult to apply management and policy actions
Providing a comprehensive estimate of ecosystem function	Measurement of physical and chemical parameters alone is insufficient to give a good estimate of ecosystem function
Proving an estimate of trophic interactions in the ecosystem	Current methods based on morphological analysis of gut contents of selected demersal fish, seabirds, and benthic macrofauna to uncover the diversity of prey items are not comprehensive enough to facilitate an analysis of food web structure.

Table 3. Mappings of indicators, as described in Table 1, against genomic methods with an application to monitoring.

Indicators (*New Indicators)	Traditional Method	Genomic Methods	Application to Monitoring	Case Study/Example	Requirements	Limitations
1.1.1 Distributional range 1.1.2 Distributional pattern 1.1.3 Area covered by species 1.6.1 Condition of typical species and communities 1.7.1 Composition of ecosystem components * Composition of meiofaunal/planktonic community assemblages	Morphological species identification	DNA barcoding/ Metagenetics	Species level identification (includes larval stages, microscopic organisms and tissue fragments)	1. Moorea Biocode (http://mooreabiocode.org/) 2. Ocean sampling day (http://www.microb3.eu/work-packages/wp2) 3. DEVOTES (www.devotes-project.eu)	Build a reference library	The method is not yet quantitative
1.3.2 Population genetic structure		SNPs	Assignment of individuals, or collections of individuals, to population of origin based on their genotypes	FishPopTrace (https://fishpoptrace.jrc.ec.europa.eu/; Nielsen et al. 2012)	Whole genome scan for SNP discovery in new species	
2.1.1 Trends in abundance, occurrence, and spatial distribution of NIS	Morphological species identification	DNA barcoding/ Metagenetics/ Microarrays	Species level identification (includes larval stages of NIS)			
3.1.1 Fishing mortality		SNPs	Assignment of individuals, or collections of individuals, to population of origin based on their genotypes	FishPopTrace (https://fishpoptrace.jrc.ec.europa.eu/; Nielsen et al. 2012)	Whole genome scan for SNP discovery in new species	
4.3.1 Abundance trends of of functionally important selected groups/species *assess predator/prey interactions at all trophic levels	Morphological species identification	DNA barcoding/ Metagenetics	Species level identification (includes gut contents/ highly digested prey)	Moorea Biocode (http://mooreabiocode.org/) study of predator-prey interactions (Leray et al. 2012)	Build a reference library	
5.2.3 Abundance of opportunistic macroalgae 5.2.4 bloom events of nuisance/toxic algal blooms	Morphological/ Laboratory culture	Microarrays	Species identification through probe-target hybridization	Detection of HAB's, MIDTAL (Lewis et al. 2012)		
*Occurence of water borne pathogens	Laboratory culture	qPCR	Quantitative detection of species/genes	California beach water quality case study (Griffith and Weisberg 2011)		
6.2.1 Presence of particularly sensitive and/or tolerant species 6.2.2 Multi-metric indexes assessing benthic community condition *assessment of microorganism community function	Morphological species identification	DNA barcoding/ Metagenetics/ Metagenomics	1. Environmental gene surveys for community analysis 2. Shotgun sequencing of community genomes for ecosystem function	1. Earth microbiome project (http://www.earthmicrobiome.org) 2. DEVOTES (www.devotes-project.eu)	Build a reference library	The method is not quantitative, but can provide relative abundances

Table 3. Continued

Indicators (*New Indicators)	Traditional Method	Genomic Methods	Application to Monitoring	Case Study/Example	Requirements	Limitations
8.2.1 Levels of pollution effects on the ecosystem components 8.2.2 Occurrence, origin, extents and impact on biota	Toxicological analyses	Transcriptomics/ Molecular ecotoxicology	Gene expression response to chemical exposure	1. Exposure of diatoms to hydrocarbons (Carvalho et al. 2011) 2. Shifts in benthic microbial communities following Deep water Horizon oil spill (Bik et al. 2012)		
*Origin of contaminants in fish			Assignment of contaminated individuals, or collections of contaminated individuals, to population of origin based on their genotypes	FishPopTrace (https://fishpoptrace.jrc.ec.europa.eu/; Nielsen et al. 2012)		

to specimens or samples that are challenging (or impossible) to identify any other way. Importantly, this applies not only individual organisms (or tissues from those organisms, like a fin clip from a fish or leg from a crab), but also to environmental or 'bulk' samples, from which the target gene/barcode can be sequenced. The approach consisting in sequencing a DNA fragment from a whole environmental sample is sometimes called metagenetics or metabarcoding (for example, see: Taberlet *et al.* 2012).

The essential prerequisite for DNA barcoding (and metabarcoding) is the creation of a reference database consisting of a library of species names linked to the DNA barcodes. Building the reference library requires an expert taxonomist to name a representative specimen for each species (usually deposited in a natural history museum or herbarium) and to sequence the specimen for the appropriate barcode gene (or genes) designated by the international Consortium for the Barcode of Life (CBOL). The reference library (usually created from adult life stages) serves as a tool for robust and reproducible species identification for assigning biological material (any sample with DNA) to species so long as the DNA barcode can be sequenced from the sample and is present in the reference library. The BOLD platform (www.barcodinglife.com), which is one of the largest existing DNA barcode libraries, contains over two million sequences (as of February 2013), of which almost 130,000 are formally described animals, over 42,000 are formally described plants and about 2,500 are formally described fungi and protists (Hajibabaei 2007).

DNA barcoding techniques have the potential to contribute to a large number of MSFD indicators (Table 3) and other legislation worldwide, wherever species identification is required, such as indicators of biological diversity, non-indigenous species, and food webs. DNA barcoding and metabarcoding have a high priority for marine monitoring and assessment, and more pilot studies and cost-benefit analyses are needed to test the general applicability of this method.

Costs of DNA Barcoding

In 2006, the cost of DNA barcoding was estimated at about $5 per sample (Cameron *et al.* 2006), including: DNA extraction, US$1.90; PCR, US$0.37; PCR purification, US$0.28; and Sanger sequencing, US$2.36, plus minor laboratory supplies such as buffers, gels, etc. Note that this does not include the

collection or transport of the specimen or sample and it assumes that the species is already present in a reference library. Six years later, building such a reference library still requires a voucher-based approach and each individual organism is sequenced using essentially the same method evaluated by Cameron *et al.* (2006). Consequently, initiatives that aim to build reference libraries (e.g., Moorea Biocode Project) still face a similar cost per specimen sequenced. Even if the costs of sequencing fall substantially, other costs associated with building a reference library are relatively incompressible, including labor costs, the collection of the specimens, their shipping to museum and molecular laboratories, and their identification by an expert taxonomist. The investment for building DNA barcode reference libraries will therefore remain quite significant, with the cost per reference barcode highly dependent on the taxon being studied (cost of identification/description, primer efficacy), the location of the study (cost of collection, cost of permits, etc.), the availability of software and informatics resources (cost of data management), and the nature of the project (cost of small team versus larger efforts with economies of scale). Approximately $100 to $200 per sample might be needed for biotic inventories seeking to create a reference barcode library for a biota containing thousands of species across all taxonomic groups, but even this could underestimate the full costs in some situations.

While the costs of building a reference library for DNA barcoding might be relatively uncompressible (at least if one employs the current standard for Linnaean species names), the revolution in DNA sequencing technologies has slashed the cost of screening samples against a reference library once it has been built. Thus, there is a high initial investment in characterizing a biota of interest, but once done and the elements for a 'genomic observatory' are in place, biodiversity dynamics can be monitored for just a few cents per identification. All the advantages of DNA barcoding then apply and DNA based identification can be carried out rapidly and reliably, irrespective of the taxonomic group or available taxonomic expertise, by sending samples to any laboratory capable of carrying out genetic sequencing (which is increasingly a commodity product).

Opportunities Offered by DNA Barcoding

Molecular approaches can be used to identify species at all life cycle stages, including highly digested tissue (Carreon-Martinez *et al.* 2011). Identifying the species involved in food webs is one of the main limitations in trophic-chain analyses, and mapping ecological food webs by analyzing the stomach contents of commercially important fish species is likely to be critical in the future management of fish stocks. In a case study on coral reefs, DNA barcoding of gut contents using the ecosystem-level Moorea Biocode reference barcode library enabled the identification of a large proportion of semi-digested fish, crustaceans and molluscs found in the guts of three hawkfish and two squirrelfish species (Leray *et al.* 2012).

Another opportunity for DNA barcoding involves taxa where species identification by morphological means is only possible for one sex (e.g., in arthropods, the defining characters are sometimes associated with male genitalia) because DNA barcoding works equally well for both sexes (Cook and Mostovski 2002).

Additional benefits of DNA barcoding stem from the ease with which these data are incorporated into population genetic and phylogenetic analyses, thus providing added value to the DNA barcode beyond the species name (e.g., historical biogeography, demographic trends etc.), especially if additional molecular markers are available. For example, we referred above to analyses based on species, but the use of phylogenetic estimates derived from this same information offer a way to side-step species while potentially increasing predictive power. Studies are now exploring the application of measures extending the "phylogenetic diversity" measure (PD; Faith 1992). PD analyses of the information from large-scale DNA barcoding programs can provide a range of biodiversity assessment and monitoring applications (Faith and Baker 2006). Smith and Fisher (2009) demonstrated that PD applied to phylogenetic patterns derived from DNA barcoding provided good estimates of species richness and species-level "complementarity" values – measures of biodiversity gains or losses (see also Zhou *et al.* 2009, Krishnamurthy and Francis 2012).

Finally, DNA sequences are 'born digital' and are easily (and freely) retained in public databases where they can be retrieved and reinterpreted as necessary (e.g., if a group is subject to taxonomic revision). Traditional approaches to species identification, by contrast, often rely on specialist knowledge and it can be hard to verify the decisions made even when detailed records (photographs and specimens) are

kept. DNA barcoding is also able to leverage many web-based tools (including those generated originally for biomedical purposes) that can greatly increase its potential usage. While informatics challenges remain in the tracking of DNA sequences and retaining linkage to related biodiversity data and metadata (e.g., photos, specimens, species names) across projects and institutions, and public repositories, pipelines are becoming increasingly robust and advances in semantic web technology are helping to improve tracking and discoverability of specimens and digital biodiversity data (e.g., the BiSciCol project).

Technical Challenges of DNA Barcoding

DNA based species identification can take quite a long time unless the field collections happen in close proximity to a suitably equipped laboratory for carrying out PCR and sequencing. Typically samples need to be shipped to a laboratory but once there the turn-around time can be a matter of hours. High throughput laboratories are able to process a huge number of samples very rapidly, with the bottleneck remaining the speed at which samples can be moved from field to lab.

Furthermore, recent work by Zhou *et al.* has demonstrated the potential for directly sequencing DNA barcodes using the Illumina NGS platform without the need for the prior step of PCR amplification (Zhou *et al*. 2013). This PCR-independent metagenomics approach requires a mitochondrial enrichment step and uses computational bioinformatics to then determine which DNA barcodes are present in the sample. While still at a relatively early stage of development, this technique even offers the possibility of determining the relative abundance (relative biomass) of species in a mixed (bulk) sample, a requirement in the assessment of many biological indices such as the Benthic Quality Index (Leonardsson *et al.* 2009). Such projects and many others show the speed at which new DNA based technologies are evolving and offering exciting opportunities for biodiversity monitoring (Baird and Hajibabaei 2012).

The Moorea Biocode Project

The Moorea Biocode Project (Check 2006) is a textbook example of a comprehensive DNA barcoding project. It compiles voucher specimens, digital photographs, high-quality DNA extractions, and genetic sequences (minimally DNA barcodes) for almost all species (adult stage >1 mm) in marine, freshwater, and terrestrial habitats on the island of Moorea (136 km^2) French Polynesia. So far, the project has amassed >42,000 specimens and >18,000 sequences from >7,000 species: this is already an unparalleled database for a tropical ecosystem. Moorea Biocode is also developing an IT platform to support this research: a standards-based informatics infrastructure connecting scientific data, and tracking Access and Benefit Sharing (ABS) agreements, across disparate sites, research teams, laboratoriess, collections, and data repositories. As the Moorea reference database is populated, researchers are carrying out innovative projects (e.g., on marine plankton and food web dynamics) to demonstrate the applications of DNA barcoding in a system with a comprehensive reference library. Increasingly, these studies employ next generation sequencing technologies and metagenomics (e.g., in gut content analyses). They also connect to microbial surveys and the physical and ecological time-series data collected on Moorea's coral reefs (e.g., by CNRS-EPHE CRIOBE since 1971 and the NSF MCR-LTER since 2004). Model ecosystems, like Moorea, are thus becoming 'Genomic Observatories', contributing to the emerging field of biodiversity genomics and mainstreaming genetic data into Earth Observing Systems (see GEO BON http://www.earthobservations.org/geobon.shtml).

Metagenomics

Metagenomics is, simply put, an extension of traditional genomics designed to encompass analysis of all genetic material in a community or assemblage of organisms, and is most often used to survey microbial species, the majority of which are recalcitrant to the culturing techniques that would provide enough DNA for genomic sequencing of an individual isolate. Since the mid 1990's this technique has relied on isolation and cloning (into heterologous expression vectors) fragments of DNA from an environmental sample, followed by sequence or functional assay screening. However, since 2005 next-generation sequencing approaches (454-pyrosequencing, Illumina GAIIx/HiSeq/MiSeq, etc.) have enabled sequencing of the isolated DNA without cloning.

Costs of Metagenomics

Metagenomics is infinitely scalable, and so it is difficult to know if it is cheaper than traditional methods. To process the first 10,000 samples from

the Earth Microbiome Project (see below) using 16S rRNA amplicon metagenomics has cost approximately $576,000. This is significantly cheaper than existing methods for typing samples; although there are cheaper methods available, they often lack taxonomic or sample resolution. It is currently possible using the EMP's pipeline to process (amplify, sequence, analyze and publish data online) ~500 environmental samples in under 5 days. So this technique is considerably faster than anything used before. The method achieves higher longitudinal, cross-sectional and taxonomic/functional resolution than ever achieved previously.

Opportunities Offered by Metagenomics

Potential advantages of PD-based biodiversity analyses discussed earlier for DNA barcodes also extend to metagenomics contexts. A recent review of microbial ecology applications, by McDonald *et al.* (2013) notes the advantages of the phylogenetic diversity framework: "Phylogenetic diversity calculations allow us to determine the relative similarity of microbial communities, using similarity of the fragment of the marker gene as a proxy for the relatedness of the organisms represented by those marker genes….in practice the difference in gene content between two organisms closely tracks the differences in marker genes such as the 16S rRNA gene." McDonald *et al.* (2013) also noted a contrast in the weaknesses of operational taxonomic units or OTUs: "…this definition is known to be problematic for several reasons. One is that the rate of evolution of the 16S rRNA gene differs among taxonomic lineages."

The PD-based measures of similarity among samples or communities open the door to a range of strategies for assessment and monitoring. Indeed, many methods conventionally employed at the species level (e.g., analyses based on ordinations) extend directly to PD analyses (Faith *et al.* 2009). These offer fresh prospects for the toolbox for marine monitoring, including assessments of marine health.

Technical Challenges of Metagenomics

While shotgun metagenomics has considerable advantages over amplicon metagenetics (e.g., it does not involve PCR amplification or primer biases), it also has some notable limitations. First, some studies have reported that the abundance of taxa and their functional genes in a metagenomic library do vary depending on the DNA extraction protocol used to acquire the nucleic acid from the environmental sample. Second, metagenomic datasets are often only sequenced to a low depth compared with the quantity of DNA in a sample, which results in only the extremely dominant populations being observed. Third, it is difficult to annotate the function or taxonomy of a short sequence fragment, resulting in a large portion of data lacking an appropriate annotation.

The Earth Microbiome Project

The Earth Microbiome Project (EMP; www.earthmicrobiome.org; Gilbert *et al.* 2010, 2011) is a massively multidisciplinary and collaborative international study aimed at characterising the Earth's microbial diversity and function. The study is predicated on crowd-sourcing environmental samples from researchers across the planet, extracting these samples with a single DNA extraction technology (MoBio's PowerSoil extraction kit), and then processing these samples initially for 16S rRNA amplicon metagenetics, and then processing a subset for shotgun metagenomics. The study has processed and sequenced more than 20,000 environmental samples in the last 2 years, and aims to complete 50,000 by the end of 2013. The study is using metagenomics to explore how microbial communities are structured along environmental parameter gradients.

The EMP is an ideal example of a pilot study that became a standard way of analyzing and working with communities. It has spawned a number of other initiatives (including the Brazilian Microbiome Project; www.brazilianmicrobiome.org) and the model is now being emulated by other studies. Three key things to make sure of are that samples are prepared in the same way, sequenced in the same way and analyzed in the same way to enable comparison. To overcome major issues it is often necessary to include standard samples in processing pipelines at multiple sites, so that irregularities that may occur due to site specific bias can be dealt with.

Ocean Sampling Day

Ocean Sampling Day (OSD; www.oceansamplingday.org, http://oceansamplingday.blogspot.se/) is an initiative to undertake, through global collaborations, the simultaneous sampling of the microbial communities in the world's oceans. OSD is part of the 9 million Euro Ocean of Tomorrow grant Micro B3 - Marine Microbial Biodiversity, Bioinformatics and Biotechnology. Coordinated by

Jacobs University Bremen, Germany, and consisting of 32 European partners, Micro B3 (Jan 2012 - Dec 2015) is designed for bioinformatic capacity building in Europe. Ocean Sampling Day takes place on the June and December solstices each year with pilot events happening in 2012 and 2013 and ramping up to a full scale sampling campaign on June 21st 2014. The solstices were selected because six-years of metagenomic studies at the 'L4' site in the Western Channel Observatory (UK) have shown that there is a predictable 'dip' in microbial diversity on the summer solstice, while the 'peak' of microbial diversity occurs on the December solstice at L4, with the variability largely explained by differences in day length between seasons (8 hours at this latitude). DNA-sequencing of the microbial communities as part of Micro B3's OSD will provide insights into the fundamental rules describing microbial diversity and function and will contribute to the blue economy through the identification of novel, ocean-derived biotechnologies. Micro B3's OSD is working closely with the Genomic Observatories Network, the Earth Microbiome Project and the Smithsonian's Global Genome Initiative to take this project forward. The long-term aim is to build an OSD Consortium to continue building a global time-series data set as part of the world's Ocean Observatories.

Microarrays

DNA microarrays are coated solid surfaces onto which a large number of fluorescently labelled DNA probes can be spotted. Each probe is specific for a species, and when the probe hybridizes with a sample, the sample/probe complex fluoresces in UV light. Microarrays are used for *in situ* monitoring of multiple harmful algal bloom (HAB) species using DNA probe arrays coupled with enzyme-linked immunosorbent assays (ELISA) to simultaneously detect algal toxins.

This method is especially useful for the rapid identification of HABs, toxic algae that can have serious health consequences (Bricker *et al.* 2007). As an example, the European project MIDTAL (Microarrays for the detection of toxic algae) has developed a microarray to target major HAB species including toxic dinoflagellates, raphidophytes, prymnesiophytes, Dichtyocophyceae and the diatom *Pseudo-nitzschia* (Lewis *et al.* 2012). Another study (Doucette *et al.* 2009) introduced the Environmental Sample Processor (ESP) which was developed for the autonomous detection of HAB species using DNA probe arrays, as well as their associated toxins. The algal toxin domoic acid (DA) was extracted and detected *in situ* from *Pseudo-nitzschia* cells onboard the ESP within 3h (Doucette *et al.* 2009).

Costs of Microarrays

Although the custom nature of the ESP makes purchasing and maintaining one of these instruments expensive, since no ship or laboratory time is involved in collecting and analyzing samples once the instrument is deployed, per sample cost compared with ship and laboratory time may actually be less. Standardization/commercialization of reagents and other consumable items is likely to make this system more cost effective than collecting samples by ship and returning them to the lab on a routine basis.

Opportunities Offered by Microarrays

Because this instrument relies on DNA probes for detection of HAB species, the potential for new indicators is nearly unlimited. The cELISA-based assay used to detect and quantify algal toxins is similarly adaptable, as all one would need to develop is a set of antibodies for the desired toxin. HABs can have potentially devastating socioeconomic, public health and ecosystem impacts (Bricker *et al.* 2007). The ability to monitor for and detect these organisms in real time is an extremely high priority.

Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

This method consists in the amplification and quantification a gene sequence specific to the organism(s) of interest. The correlation of the amount of DNA obtained with the number of individuals will allow quantification of the organisms of study in a given sample. This is only possible for unicellular organisms that contain a single or a known number of copies of the gene under study. Exponential amplification of the target sequence is followed in real-time by means of a fluorescent dye or fluorescently labeled DNA probe and detected by the optics of the qRT-PCR instrument. Quantification is generally via comparison to a standard curve, which is run concurrently with samples using reference material consisting of pre-enumerated cells or DNA.

California Beach Water Quality

Beach water quality monitoring currently employs culture-based methods to measure fecal indicator bacteria. These methods require 24 hours for sample processing, which is too slow to provide warning against water-borne pathogens, with the majority of contamination events dissipating by the time results become available. In a case study of California beach water quality (Griffith and Weisberg 2011), qPCR (quantitative PCR) methods are used to reduce the sample processing time to two hours. A pilot study was conducted in 2010 led by the Southern California Coastal Water Research Project. Three public agencies that perform routine microbiological monitoring of marine waters using traditional growth-based methods (Orange County Sanitation District, Orange County Public Health Laboratory, South Orange County Wastewater Authority) performed the rapid qPCR measurement method for *Enterococcus* for an eight-week period at nine beaches. Samples were collected at 8:00 a.m. each morning and returned to the lab for processing. Results were provided to beach managers by 11:00 a.m. on average. Public notification of water quality advisories was relayed to beach-goers by noon via electronic signs at the beach, the County Health Department website and Twitter.

Costs of qPCR

The rapid method for qPCR as implemented in the pilot study was approximately three times the cost of traditional methods. Higher costs included both labor and assay materials. Additional labor was required for dedicated samplers to bring water samples to the laboratory sooner than they would have arrived under usual circumstances. Supplies to conduct the qPCR analysis were approximately $35 vs. about $12 for the traditional method. The cost of supplies is expected to drop as reagents are produced on a commercial scale, but additional labor to return samples to the lab in a timely manner will still be required if answers are expected in time to warn potential swimmers of poor water quality before they enter the water.

Opportunities Offered by qPCR

The qPCR method can be performed in about one and one-half hours. The fastest culture method takes 24 hours. In terms of protecting public health from poor water quality, the rapid qPCR method far surpasses growth-based methods. This method is highly amenable to new indicators and has already been adapted to host associated fecal markers. Implementation of this methodology is a priority in many locals where beach tourism drives the economy. Managers and swimmers want to know when health risks to swimmers are elevated.

Technical Challenges of qPCR

The primary limitations to the widespread use of this methodology for producing same-day water quality information are cost and logistics. Although the method produces results in approximately one and one-half hours, it may not be possible to collect and return samples from distant or numerous beaches and still produce results in time to notify swimmers before they are exposed to contaminated water. A straightforward solution is to send individual samplers to each beach, but the additional labor and vehicle costs in employing this strategy may limit the use of the method to high priority locations.

Short Nucleotide Polymorphisms (SNPs)

Short Nucleotide Polymorphisms are DNA sequence variations occurring when a single DNA nucleotide in the genome (A,G,C,T) differs among individuals of the same species. For example the change of one nucleotide cytosine (C) to another nucleotide thymine (T) in a certain stretch of DNA would be a single SNP. SNPs can be used as biological markers to demarcate populations of individuals within a species. Recent improvements in the speed, cost and accuracy of next generation sequencing and associated bioinformatic tools are revolutionizing the discovery of single nucleotide polymorphisms (SNPs). Some SNPs can have very high information content for population structure analysis. Population genetic applications, such as conservation management, product traceability and forensic genetic analysis involve the assignment of individuals, or collections of individuals, to population of origin based on their genotypes (Helyar *et al.* 2011).

Costs of SNPs

The cost of developing and genotyping large numbers of samples is still relatively high and likely to be beyond the means of many labs. However, sequencing costs are falling rapidly, and genotyping by sequencing (GBS) rather than using other SNP genotyping methods (e.g., Taqman, GoldenGate arrays, etc.) is close to general implementation.

In the case of traceability of fish to population of origin (see FishPopTrace case study below), it is not a matter of whether the technology is cheaper, but whether the technology is capable of answering the question being asked. Short nucleotide polymorphisms are the first markers capable of assigning fish back to population of origin at all stages of the food chain at relatively fine geographic scales. Previous DNA-based markers such as microsatellites provide some resolution for assignment, but often at larger geographic scales. Genotyping SNP markers will become progressively cheaper over the next few years as new technologies are developed and existing technologies become more efficient.

Opportunities Offered by SNPs

Genotyping using SNP markers is clearly more rapid than previous DNA based technologies such as microsatellites. High numbers of SNPs can be genotyped simultaneously using array based methods. Current custom SNP arrays can simultaneously genotype 1 million individual SNPs.

First, using SNP markers that are putatively under selection allows populations to be delineated on much smaller scales than were previously possible. Second, a big advantage of SNP markers over size-based DNA methods (e.g microsatellites) is the digital nature of the outputs (presence or absence of a particular allele). This means extensive cross-calibration among labs is not necessary and results from published research can be easily compared. Moreover, a database can be established that contains baseline allele frequencies of different populations. Any laboratory can then compare their own genotypes to the baseline to assist in assigning individuals to population. Given the number of SNP markers found in eukaryotic genomes, the potential to develop targeted SNP assays for specific traceability issues is good. This is particularly the case in many commercially exploited marine species where population sizes are large meaning selection is relatively powerful in comparison to genetic drift.

FishPopTrace

The FishPopTrace project has developed and tested a range of traceability tools for assigning fish and fish products back to population of origin (SNPs, otolith shape and microchemistry, gene expression, proteomics). Short nucleotide polymorphisms were identified as the only tool that could be used at every stage of the food chain, from freshly caught fish though to processed fish products such as canned or other processed products. Short nucleotide polymorphisms were developed and tested in three species (herring, sole, and hake) and existing SNP markers were tested in cod. Short nucleotide polymorphisms allowed high levels of assignment to population of origin - with a small subset of SNP markers providing 'maximum power for minimum cost' (Nielsen *et al.* 2012). Moreover, all protocols were forensically validated. In this study, SNPs for herring, sole and hake were identified through 454 sequencing (Roche 454 GS FLX sequencer) of the transcriptome. By using gene-associated single nucleotide polymorphisms, it was shown that individual marine fish can be assigned back to population of origin with unprecedented high levels of precision. By applying high differentiation single nucleotide polymorphism assays, in four commercial marine fish, on a pan-European scale, 93 to 100% of individuals could be correctly assigned to origin in policy-driven case studies. The authors show how case-targeted single nucleotide polymorphism assays can be created and forensically validated, using a centrally maintained and publicly available database. The results demonstrate how application of gene-associated markers will likely revolutionize origin assignment and become highly valuable tools for fighting illegal fishing and mis-labeling worldwide (Nielsen *et al.* 2012).

Transcriptomics

Transcriptomics comprises, amongst other methods, the analysis of gene expression changes (as measured by the amount of RNA from a particular gene) of either an entire organism or part of it (e.g., cells, tissues) under different conditions (eg., at different developmental stages or upon exposure to chemicals or stressors). The most common technologies used to investigate gene expression changes are DNA microarrays, quantitative real time PCR (qRT-PCR; Lettieri 2006) and RNAseq (Montgomery, 2010).

A DNA microarray is a glass or a nylon membrane on which parts of gene sequences (oligonucleotide probes) are spotted; the fluorescently labeled RNA extracted from organisms, organs (e.g., liver), or cells exposed to a pollutant/stressor is hybridized against the array. After image scanning analysis, RNA abundance is obtained, and the relative gene expression of the treated sample compared to the untreated control can be measured. Quantitative Real

Time PCR (qRT-PCR) for measuring gene expression is based on detecting and quantifying RNA from a particular gene (Heid *et al.* 1996). The main differences between DNA microarray and qRT-PCR techniques are: i) the number of transcripts analyzed in one step (experiment): more in a DNA microarray; and ii) the intensity of the signal: higher for qRT-PCR than for the microarray.

The RNAseq technique utilizes recent advances in sequencing technologies that allow large quantities of high-throughput sequencing data to be produced for relatively low levels of capital; RNA sequencing essentially allows gene transcription to be quantified by sequencing and counting the number of individual transcripts that are present for each gene. Unlike miocroarrays, RNAseq is open-ended (without constraints on the number of targets), requires little prior knowledge of the target organism's genome and can be directly scaled according the level of sequencing required. It is thus ideally suited to developing techniques in non-model species, or in systems where choice of sentinel species is limited, as is common in the marine environment.

Applications of transcriptomic experiments in aquatic toxicology have already been described mainly in freshwater ecosystems (Falciani *et al.* 2008, Garcia-Reyero *et al.* 2008). There are fewer studies in marine organisms (Carvalho *et al.* 2011a, Shrestha *et al.* 2012). Transcriptomics offer: i) discovery of molecular biomarkers of exposure as early signals to predict the effects first at a physiological level, and later at a population level; and ii) provide the mode of action (MOA) of the chemicals or a stressor, i.e., the mechanism of toxicity or the mechanism of adaptation or response to the environmental changes. The MOA could reduce the uncertainty in chemical risk assessment by providing, for example, a basis for the extrapolation of the effects across species; iii) the possibility of integrating MOA data with a deleterious outcome and in this way understand the impact on the ecosystem more than only on a single organism or species; and iv) discovery of gene expression pattern for complex mixtures or complex stressors.

Costs of Transcriptomics

Costs have dropped in the last year, although the DNA microarray technique requires a dedicated instrument for scanning which is still costly. However, core facilities are available from several academic institutes and the service price has decreased roughly 20 to 25% in the last five years. In terms of time, the analysis requires one night and half a day. qRT-PCR runs in only 1 hour, with an additional 30'-60' if RNA has to be extracted prior to running.

Opportunities Offered by Transcriptomics

Transcriptomics can provide information about the effects of complex mixtures on organisms, effects which cannot be accounted for through classical chemical analytical methods. Transcriptomics also provides information on complex stressors which include additional parameters such as temperature changes, nutrient depletion, and pollutants. Transcriptomics represents the shift from a merely chemical monitoring to an early warning system based on biological monitoring. Transcriptomics is a priority for the regulations and can, together with other "omics" approaches, provide a global scenario of multiple stressors on marine ecosystems. Standardization is required and an inter-calibration exercise for the validation of selected molecular biomarkers can be the first step.

Technical Challenges of Transcriptomics

Limitations for the microarray include the lack of standardization of data collection and analysis. Currently, a wide variety of approaches are used to generate data and different platforms would require a formal standardization and validation to be considered for a regulatory test. Unfortunately, research for method standardization is expensive and often too routine and tedious (Ankley *et al.* 2006). The standardization process for qRT-PCR for transcriptomics may be considered more promising and cheaper.

Exposure to Benzo(a)pyrene and Gene Expression in Diatoms

Carvalho *et al.* (2011b) exposed the marine diatom *Thalassiosira pseudonana* to benzo(a)pyrene (BaP), a polyclic aromatic hydrocarbon (PAH). They investigated whether the gene expression profile compared to the untreated cells could provide molecular biomarkers linked to a physiological status change due to the pollutant effects. They showed that the silicification process was affected under these conditions, particularly the down regulation of silicon transporter encoding gene, ST1, thus compromising the silica uptake from the media. The same result was confirmed also when the diatoms were exposed to marine PAH-extracted sediment samples (Carvalho *et al.* 2011b). In a pilot study, surface sediments

were collected at the port of Genoa in Italy, an environmentally contaminated site, to validate the gene expression changes identified by transcriptomic analysis in marine diatoms upon exposure to the PAH benzo(a)pyrene. This part of the Italian coastline is a densely populated area with intense industrial activity, where high PAH concentrations have been previously measured in surface sediments, in particular close to the urban centers and the port of Genoa. Cultures of the marine diatom *T. pseudonana* were exposed to the complex mixture of PAHs extracted from the samples. Expression of several genes was analyzed by qRT-PCR confirming their suitability as molecular biomarkers of phytoplankton species exposed to PAHs in contaminated aquatic environments. Furthermore the gene expression changes of two genes suggest that they could specifically target BaP contamination, and retrieve information on the BaP:PAHs ratio of a monitored site (Carvalho *et al.* 2011b).

Infrastructures for Genomic Monitoring

Marine biodiversity is not only changing at large scales of time and space, but also at smaller scales relevant for local or regional management (e.g., Marine Spatial Planning; Ehler and Douvere 2009). To understand these changes effectively, a major effort is required to build biodiversity monitoring and research infrastructures in the future (Basset and Los 2012). Such infrastructures will consist of three principal components: the data generation layer (including sensors, monitoring programs, research, etc.), the data storage layer (including databases, data curation, archives, and repositories), and the analytical layer (including interoperability systems, analytical resources). The genomic components will be integrated simultaneously on all three levels, and this process is coordinated by the Genomic Observatories infrastructure initiative. Here leading genomic scientists are working together to introduce the technology, data, standards, and analytical resources from the genomics sector into ecosystem and conservation research (Davies *et al.* 2012, In press). This initiative is a powerful contribution to the next generation of marine monitoring programs, because it has the potential to add a very cost efficient technology and information rich data source to existing marine monitoring activities.

Data Generation

On the first level, contents are generated by current marine monitoring activities world-wide (e.g., in the context of the MSFD in Europe). These activities are increasingly supported by the marine research community, such as the pan-European Marine Biodiversity Observatory Network (www.embos.eu), to be used for research as well as monitoring. This system will consist of a network of observatories in carefully selected geographical locations that generate biological observation data based on common protocols, quality control and free access to data, where biodiversity measurements are combined with environmental measurements. Here, genomics technology can almost instantly contribute with the standardized generation of sequencing data from conventional samples (Baird and Hajibabaei 2012), while the Genomics Standards Consortium (http://gensc.org/) will safeguard the adoption of the appropriate standards for sample and data collection (Field *et al.* 2011). On the long-term, fast evolving observation platforms such as ecogenomic sensor systems (Scholin 2010) will be introduced in either marine observatory networks or national monitoring programs.

Data Storage and Curation

The link between genomic data and national, regional or commercial data centers for marine monitoring data is relatively straightforward, as genomics databases, due to their large data volumes, are very well structured. In the future, all genetic data generated by monitoring activities will be deposited in one of the existing archives. The databases for genetic information are: the European Nucleotide Archive (ENA), an open access, annotated collection of publicly available nucleotide sequences and their protein translations; the US National Center for Biotechnology Information (NCBI); and the DNA Data Bank of Japan (DDBJ). The link among them is maintained via the International Nucleotide Sequence Database Collaboration (INSDC), the central authority that manages genetic libraries globally. Through INSDC also a large number of specific archives can be accessed, such as the dbSNP for single nucleotide polymorphisms and short tandem repeats, the dbEST for expressed sequence tags, or the SRA for raw sequence reads. All INSDC databases are furthermore coupled to NCBI's Taxonomy database. An elaborate service set of BLAST and alignment functions is coupled to these libraries allowing for

initial data inspection, exploration, and some basic analytical functions.

Data Analysis

Efforts towards improved coordination of biodiversity observations, data and research tools are already underway, with strong efforts to integrate genetic data in conservation and ecosystem research (Heip and McDonough 2012). As an example, the European Strategy Forum on Research Infrastructures (ESFRI) program LifeWatch (www.lifewatch.eu) and its pilot implementation program BioVeL (www.biovel.eu) are currently interconnecting primary data repositories to create e-Services as well as virtual laboratories on top of these (Hardisty and Roberts 2013). Here, bioinformatics tools are currently developed to analyze complex marine data sets (including ecological, taxonomic, climatic, and genetic data) across large geographic distances and time scales. Examples are DNA identification tools to identify fish stomach contents and larval stages, and these methods can be customized to match current or future indicators for marine health assessment.

Analytical Approaches

Workflows -- powerful analytical pipelines which access distributed computing resources -- are being constructed through the BioVeL project to address the needs of the biodiversity research community. Micro B3 and BioVel have agreed to join forces to develop metagenome workflows of OSD.

Additional workflows are being designed to process metagenetic data from environmental samples (e.g., DNA metabarcoding), to enable identification of species from a metagenetic sample by matching them to databases and reference libraries, and to provide measures of phylogenetic or alpha and beta diversity between samples.

These analysis pipelines are complementary to tools that translate genomic data into indicator metrics that can be used for decision making, which are being developed through the DEVOTES project.

Stakeholders and End-Users

The entry point for new methods into regular monitoring programs is at the national level and therefore the envisaged methods have to meet the requirements of the national and regional programs. In order to be effective, all of the important partners in this innovation process have to be identified beforehand. The scientific network representing genomics methods and standards is the Genomic Standards Consortium (http://gensc.org/). The network of end users may be represented by some European regional sea convention programs, such as HELCOM (www.helcom.fi/) and OSPAR (www.ospar.org/), the national environmental agencies, as well as the national scientific institutes that currently implement the MSFD. The coordination activity between these partner groups should also connect and assign responsibilities to related European wide initiatives working with marine observations, as for example EMBOS (embos.eu), Micro B3's Ocean Sampling Day (www.oceansamplingday.org), DEVOTES (devotes-project.eu), STAGES (marineboard.eu/external-projects/stages), and European marine GEO-BON initiatives. The primary objective of this communication activity between these networks should be to disseminate the potential of genomic tools, specify the requirements for these methods to enter national programs, and to design national and regional pilots. This activity should produce precise utility descriptions to the end, such as guidelines, protocols and analytical tools for the application of this new technology. A global "Marine Genomics for Users Network" has been proposed under the Genomic Observatories Network initiative, which is a collaboration of the GSC and GEO BON.

In order to stimulate the uptake of these new technologies also by the industrial sector, the coordination activity should include local and regional SME partners. Marine biotechnology has been identified as one of the key areas on the European roadmap for blue growth (http://ec.europa.eu/maritimeaffairs/policy/blue_growth/index_en.htm), and this technology transfer will provide an excellent opportunity to stimulate the development of tools by industrial partners and to contribute to securing environmental health.

The technology transfer from the scientific sector to national monitoring programs can be regarded as an 'innovation' project. For that purpose recently, a number of wider 'innovation' strategies have been developed at various scales, such as the OECD Innovation Strategy (www.oecd.org/site/innovationstrategy/), or the EU Innovation Union (http://ec.europa.eu/research/innovation-union/). These common policies offer helpful support instruments for leveraging such new methods at European and national levels, in addition to the traditional support

strategies for Research and Development (http://cordis.europa.eu/).

Summary

Currently, there is an increasing need worldwide for monitoring in real time to feed into management (it is no good if the data takes a year to obtain but a management decision is needed quickly or if the final data will not be fit-for-purpose, as stated by Borja and Elliott 2013). Many of the genomic tools described above can assist in achieving this near real time information for management, e.g., barcoding, qPCR, etc.

Borja and Elliott (2013) also emphasize that whereas recent legal initiatives focus on a 'structural' approach (i.e., numbers of taxa, abundance data, level of a pollutant, etc.), others are suggesting a more functional approach (e.g., the MSFD, the Ocean's Act, etc.). This 'holistic' approach could help determine whether an ecological system is working well and functioning rather than merely what organisms it contains. For this purpose, genomic technologies are a valuable resource and can assist in producing rapid and rigorous information about ecosystem functioning, at a lower cost than traditional approaches. In this context, we propose the following steps towards the implementation of molecular methods in marine monitoring:

Pilot studies and cost-benefit analyses comparing molecular with traditional methods.

1. Standardized manuals and protocols for sampling and sample processing.

2. Analytical pipelines and technologies integrating genomic data with other data sources (remote sensing, mapping tools, taxonomy databases).

3. Molecular indicators and /or methods to translate the results from molecular analyses into indicator metrics for use in policy and decision making (eg. status assessment).

4. Dissemination to and facilitation of adoption by monitoring bodies.

Literature Cited

Ankley, G.T., G.P. Daston, S.J. Degitz, N.D. Denslow, R.A. Hoke, S.W. Kennedy, A.L. Miracle, E.J. Perkins, J. Snape, D.E. Tillitt, C.R. Tyler and D. Versteeg. 2006. Toxicogenomics in regulatory ecotoxicology. *Environmental Science & Technology* 40:4055-4065.

Apitz, S.E., M. Elliott, M. Fountain and T.S. Galloway. 2006. European Environmental Management: Moving to an Ecosystem Approach. *Integrated Environmental Assessment and Management* 2:80-85.

Baird, D.J. and M. Hajibabaei. 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology* 21:2039-2044.

Ban, N. and J. Alder. 2008. How wild is the ocean? Assessing the intensity of anthropogenic marine activities in British Columbia, Canada. *Aquatic Conservation-Marine and Freshwater Ecosystems* 18:55-85.

Barnes, C. and K.W. McFadden. 2008. Marine ecosystem approaches to management: challenges and lessons in the United States. *Marine Policy* 32:387-392.

Basset, A. and W. Los. 2012. Biodiversity e-Science: LifeWatch, the European infrastructure on biodiversity and ecosystem research. *Plant Biosystems* 146:780-782.

Bik, H.M., D.L. Porazinska, J.G. Caporaso, R. Knight and W.K. Thomas. 2012. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology and Evolution* 27:233-234.

Borja, Á. and M. Elliott. 2013. Marine monitoring during an economic crisis: The cure is worse than the disease. *Marine Pollution Bulletin* 68:1-3.

Borja, Á., S.B. Bricker, D.M. Dauer, N.T. Demetriades, J.G. Ferreira, A.T. Forbes, P. Hutchings, X. Jia, R. Kenchington, J.C. Marques and C. Zhu. 2008. Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. *Marine Pollution Bulletin* 56:1519-1537.

Borja, Á., M. Elliott, J. Carstensen, A.-S. Heiskanen and W. van de Bund. 2010. Marine management - Towards an integrated implementation of the European Marine Strategy Framework and the Water Framework Directives. *Marine Pollution Bulletin* 60:2175-2186.

Borja, Á., I. Galparsoro, X. Irigoien, A. Iriondo, I. Menchaca, I. Muxika, M. Pascual, I. Quincoces, M. Revilla, J. Germán Rodríguez, M. Santurtún, O.

Solaun, A. Uriarte, V. Valencia and I. Zorita. 2011. Implementation of the European Marine Strategy Framework Directive: A methodological approach for the assessment of environmental status, from the Basque Country (Bay of Biscay). *Marine Pollution Bulletin* 62:889-904.

Bricker, S., B. Longstaff, W. Dennison, A. Jones, K. Boicourt, C. Wicks and J. Woerner. 2007. Effects of Nutrient Enrichment in the Nation's Estuaries: A Decade of Change. NOAA Coastal Ocean Program Decision Analysis. Series No. 26. National Centers for Coastal Ocean Science. Silver Spring, MD.

Cameron, S., D. Rubinoff and K. Will. 2006. Who will actually use DNA barcoding and what will it cost? *Systematic Biology* 55:844-847.

Cardoso, A.C., S. Cochrane, H. Doemer, J.G. Ferreira, F. Galgani, C. Hagebro, G. Hanke, N. Hoepffner, P.D. Keizer, R. Law, S. Olenin, G.J. Piet, J. Rice, S.I. Rogers, F. Swartenbroux, M.L. Tasker and W. van de Bund. 2010. Scientific Support to the European Commission on the Marine Strategy Framework Directive. Management Group Report. EUR 24336 EN. Joint Research Centre. Office for Official Publications of the European Communities. Luxembourg.

Carreon-Martinez, L., T.B. Johnson, S.A. Ludsin and D.D. Heath. 2011. Utilization of stomach content DNA to determine diet diversity in piscivorous fishes. *Journal of Fish Biology* 78:1170-1182.

Carvalho, R.N., S.K. Bopp and T. Lettieri. 2011a. Transcriptomics responses in marine diatom *Thalassiosira pseudonana* exposed to the polycyclic aromatic hydrocarbon Benzo[a]pyrene. *PLOS One* 6:11.

Carvalho, R.N., A.D. Burchardt, F. Sena, G. Mariani, A. Mueller, S.K. Bopp, G. Umlauf and T. Lettieri. 2011b. Gene biomarkers in diatom *Thalassiosira pseudonana* exposed to polycyclic aromatic hydrocarbons from contaminated marine surface sediments. *Aquatic Toxicology* 101:244-253.

CBD. 2000. Ecosystem Approach. Fifth Conference of the Parties to the Convention on Biodiversity. Nairobi, Kenya, May 2000. www.iisd.ca/biodiv/cop5/.

Check, E. 2006. Treasure island: Pinning down a model ecosystem. *Nature* 439:378-379.

Claudet, J. and S. Fraschetti. 2010. Human-driven impacts on marine habitats: A regional meta-analysis in the Mediterranean Sea. *Biological Conservation* 143:2195-2206.

Cochrane, S.K.J., D.W. Connor, P. Nilsson, I. Mitchell, J. Reker, J. Franco, V. Valavanis, S. Moncheva, J. Ekebom, K. Nygaard, R. Serrao Santos, I. Naberhaus, T. Packeiser, W. van de Bund and A.C. Cardoso. 2010. Marine Strategy Framework Directive – Task Group 1 Report Biological Diversity. EUR 24337 EN. Joint Research Centre, Office for Official Publications of the European Communities. Luxembourg.

Cook, C.E. and M.B. Mostovski. 2002. 16S mitochondrial sequences associate morphologically dissimilar males and females of the family Phoridae (Diptera). *Biological Journal of the Linnean Society* 77:267-273.

Davies, N., D. Field, L. Amaral-Zettler, K. Barker, M. Bicak, S. Bourlat, J. Coddington, J. Deck, A. Drummond, J. Gilbert, F.O. Glöckner, R. Kottmann, C. Meyer, N. Morrison and M. Obst. In press. Report of the 14th Genomic Standards Consortium Meeting, Oxford, UK, September 17-21, 2012. *Standards in Genomic Sciences*.

Davies, N., D. Field and the Genomic Observatories Network. 2012. Sequencing data: A genomic network to monitor Earth. *Nature* 481:145.

De Jonge, V.N., M. Elliott and V.S. Brauer. 2006. Marine monitoring: its shortcomings and mismatch with the EU Water Framework Directive's objectives. *Marine Pollution Bulletin* 53:5-19.

De Long, E.E. 2005. Microbial community genomics in the ocean. *Nature Reviews Microbiology* **3**:459-469.

Doucette, G.J., C.M. Mikulski, K.L. Jones, K.L. King, D.I. Greenfield, R. Marin III, S. Jensen, B. Roman, C.T. Elliott and C.A. Scholin. 2009. Remote, subsurface detection of the algal toxin domoic acid onboard the Environmental Sample Processor: Assay development and field trials. *Harmful Algae* 8:880-888.

Ehler, C. and F. Douvere. 2009. Marine Spatial Planning: a step-by-step approach toward ecosystem-based management. Intergovernmental Oceanographic Commission and Man and the

Biosphere Programme. IOC Manual and Guides No. 53, ICAM Dossier No. 6. UNESCO. Paris, France.

Elliott, M. 2011. Marine science and management means tackling exogenic unmanaged pressures and endogenic managed pressures – A numbered guide. *Marine Pollution Bulletin* 62:651-655.

Faith, D.P. 1992. Conversation evaluation and phylogenetic diversity. *Biological Conservation* 61:1-10.

Faith, D.P. and A.M. Baker. 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evolutionary Bioinformatics Online* 2:121-128.

Faith, D.P., C. Lozupone, D. Nipperess and R. Knight. 2009. The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to environmental gradients and supports broad applications of microbial ecology's "phylogenetic beta diversity" framework. *International Journal of Molecular Sciences* 10:4723-41.

Falciani, F., A.M. Diab, V. Sabine, T.D. Williams, F. Ortega, S.G. George and J.K Chipman. 2008. Hepatic transcriptomic profiles of European flounder (*Platichthys flesus*) from field sites and computational approaches to predict site from stress gene responses following exposure to model toxicants. *Aquatic Toxicology* 90:92-101.

Field, D., L. Amaral-Zettler, G. Cochrane, J.R. Cole, P. Dawyndt, G.M. Garrity, J. Gilbert, F.O. Glöckner, L. Hirschman, I. Karsch-Mizrachi, H.-P. Klenk, R. Knight, R. Kottman, N. Kyrpides, F. Meyer, I.S. Gil, S.-A. Sansone, L.M. Schrinl, P. Sterk, T. Tatusova, D.W. Ussery, O. White and J. Wooley. 2011. The Genomic Standards Consortium. *PLOS Biology* 9:e1001088.

Frolov, S., R.M. Kudela and J.G. Bellingham. 2013. Monitoring of harmful algal blooms in the era of diminishing resources: A case study of the U.S. West Coast. *Harmful Algae* 21-22:1-12.

Garcia-Reyero, N., I. Adelman and N. Denslow. 2008. Gene expression profiles of fathead minnows exposed to surface waters above and below a sewage treatment plant in Minnesota. *Marine Environmental Research* 66:134-136.

Gilbert, J.A., M. Bailey, D. Field, N. Fierer, J.A. Fuhrman, B. Hu, J. Jansson, R. Knight, G.A. Kowalchuk, N.C. Kyrpides, F. Meyer and R. Stevens. 2011. The Earth Microbiome Project: The Meeting Report for the 1st International Earth Microbiome Project Conference, Shenzhen, China, June 13th-15th 2011. *Standards in Genomic Science* 5:2.

Gilbert, J.A., F. Meyer, J. Jansson, J. Gordon N. Pace, J. Tiedje, R. Ley, N. Fierer, D. Field, N. Kyrpides, F.-O. Glöckner, H.P. Klenk, K.E. Wommack, E. Glass, K. Docherty, R. Gallery, R. Stevens and R. Knight. 2010. The Earth Microbiome Project: Meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6th 2010. *Standards in Genomic Science* 3:249-253.

Glöckner, F.O., L.J. Stal, R.-A. Sandaa, J.M. Gasol, F. O'Gara, F. Hernandez, M. Larenz, E. Stoica, M.M. Varela, A. Bordalo and P. Pitta. 2012. Marine Microbial Diversity and Its Role in Ecosystem Functioning and Environmental Change NPG Scientific Reports. ESF Marine Board Position Paper 17. European Science Foundation Marine Board. Ostend, Belgium.

Griffith, J.F. and S.B. Weisberg. 2011. Challenges in implementing new technology for beach water quality monitoring: Lessons from a California demonstration project. *Marine Technology Society Journal* 45:65-73.

Hajibabaei, M. 2007. The Barcode of Life Initiative. *Journal of Phycology* 43:20-20.

Hajibabaei, M., S. Shokralla, X. Zhou, G.A.C. Singer and D.J. Baird. 2011. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* 6:e17497.

Halpern, B.S., S. Walbridge, K.A. Selkoe, C.V. Kappel, F. Micheli, C. D'Agrosa, J.F. Bruno, K.S. Casey, C. Ebert, H.E. Fox, R. Fujita, D. Heinemann, H.S. Lenihan, E.M.P. Madin, M.T. Perry, E.R. Selig, M. Spalding, R. Steneck and R. Watson. 2008. A global map of human impact on marine ecosystems. *Science* 319:948-952.

Hardisty, A. and D. Roberts. 2013. A decadal view of biodiversity informatics: Challenges and priorities. *BMC Ecology* 13:16.

Hebert, P.D.N., S. Ratnasingham and J.R. de Waard. 2003. Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species.

Proceedings of the Royal Society of London Series B-Biological Sciences 270:S96-S99.

Heid, C.A., J. Stevens, K.J. Livak and P.M. Williams. 1996. Real time quantitative PCR. *Genome Research* 6:986-994.

Heip, C. and N. McDonough. 2012. Marine Biodiversity: A Science Roadmap for Europe. Marine Board Future Science Brief 1. European Science Foundation Marine Board. Ostend, Belgium.

HELCOM. 2010. Ecosystem health of the Baltic Sea 2003-2007: HELCOM Initial Holistic Assessment. *Baltic Sea Environmental Proceedings* 122:68.

Helyar, S.J., J. Hemmer-Hansen, D. Bekkevold, M.I. Taylor, R. Ogden, M.T. Limborg, A. Cariani, G.E. Maes, E. Diopere, G.R. Carvalho and E.E. Nielsen. 2011. Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources* 11:123-136.

Karsenti, E., S.G. Acinas, P. Bork, C. Bowler, C. De Vargas, J. Raes, M. Sullivan D. Arendt, F. Benzoni, J.-M. Claverie, M. Follows, G. Gorsky, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels-Lewis, U. Krzic, F. Not, H. Ogata, S. Pesant, E.G. Reynaud, C. Sardet, M.E. Sieracki, S. Speich, D. Velayoudon, J. Weissenbach, P. Wincker and the Tara Oceans Consortium. 2011. A holistic approach to marine eco-systems biology. *Plos Biology* 9:e1001177.

Krishnamurthy, P.K. and R.A. Francis. 2012. A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation* 21:1901-1919.

Leonardsson K., M. Blomqvist and R. Rosenberg. 2009. Theoretical and practical aspects on benthic quality assessment according to the EU-Water Framework Directive - examples from Swedish waters. *Marine Pollution Bulletin* 58:1286-96.

Leray, M., J.T. Boehm, S.C. Mills and C.P. Meyer. 2012. Moorea BIOCODE barcode library as a tool for understanding predator prey interactions: Insights into the diet of common predatory coral reef fishes. *Coral Reefs* 31:383-388.

Lester, S.E., K.L. McLeod, H. Tallis, M. Ruckelshaus, B.S. Halpern, P.S. Levin, F.P. Chavez, C. Pomeroy, B.J. McCay, C. Costello, S.D. Gaines, A.J. Mace, J.A. Barth, D.L. Fluharty and J.K. Parrish. 2010. Science in support of ecosystem-based management for the US West Coast and beyond. *Biological Conservation* 143:576-587.

Lettieri, T. 2006. Recent applications of DNA microarray technology to toxicology and ecotoxicology. *Environmental Health Perspectives* 114:4-9.

Lewis, J., L.K. Medlin and R. Raine. 2012. MIDTAL (Microarrays for the Detection of Toxic Algae): A Protocol for a Successful Microarray Hybridisation and Analysis. ARG Gantner. Ruggell, Liechtenstein.

Lotze, H.K. 2010. Historical reconstruction of human-induced changes in U.S. estuaries. *Oceanography and Marine Biology: An Annual Review* 48:267-338.

Mardis, E.R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24:133-141.

McDonald, D., Y. Va´zquez-Baeza, W.A. Walters, J.G. Caporaso and R. Knight. 2013. From molecules to dynamic biological communities. *Biology and Philosophy* 28:241-259.

Minster, J.F., N. Connolly, A. Carbonnière, J.W. de Leeuw, M. Evrard, J. Mees, K. Nittis, G. Sullivan and N. Walter. 2006. Navigating the Future - III Updated Synthesis of Perspectives on Marine Science and Technology in Europe. ESF Marine Board Position Paper 8. 71 pp. European Science Foundation Marine Board. Strasbourg, France.

Montgomery, S.B., M. Sammeth, M. Gutierrez-Arcelus, R.P. Lach, C. Ingle, J. Nisbett, R. Guigo and E.T. Dermitzakis. 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**:773-777.

Nielsen, E.E., A. Cariani, E. Mac Aoidh, G.E. Maes, I. Milano, R. Ogden, M. Taylor, J. Hemmer-Hansen, M. Babbucci, L. Bargelloni, D. Bekkevold, E. Diopere, L. Grenfell, S. Helyar, M.T. Limborg, J.T. Martinsohn, R. McEwing, F. Panitz, T. Patarnello, F. Tinti, J.K.J. Van Houdt, F.A.M. Volckaert, R.S. Waples, the FishPop Trace Consortium and G.R. Carvalho. 2012. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications* 3:851.

Pew Oceans Commission. 2003. America's Living Oceans: Charting a Course for Sea Change.

Summary Report, May 2003. Pew Oceans Commission. Arlington, VA.

Radom, M., A. Rybarczyk, R. Kottmann, P. Formanowicz, M. Szachniuk, F.O. Glöckner, D. Rebholz-Schuhmann and J. Blazewicz. 2012. Poseidon: An information retrieval and extraction system for metagenomic marine science. *Ecological Informatics* 12:10-15.

Roger, F., A. Godhe and L. Gamfeldt. 2012. Genetic diversity and ecosystem functioning in the face of multiple stressors. *PLOS One* 7:e45007.

Scholin, C.A. 2010. What are "ecogenomic sensors?" A review and thoughts for the future. *Ocean Science* 6:51-60.

Shrestha, R.P., B. Tesson, T. Norden-Krichmar, S. Federowicz, M. Hildebrand and A.E. Allen. 2012. Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics* 13:499.

Smith, M.A. and B.L. Fisher. 2009. Invasions, DNA barcodes, and rapid biodiversity assessment using ants of Mauritius. *Frontiers in Zoology* 6:31.

Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann and E. Willerslev. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* 21:2045-2050.

Teeling, H. and F.O. Glöckner. 2012. Current opportunities and challenges in microbial metagenome analysis-a bioinformatic perspective. *Briefings in Bioinformatics* **13**:728-742.

UNCLOS. 1982. United Nations Convention on the Law of the Sea, signed at Montego Bay, Jamaica, on 10 December 1982. www.un.org/Depts/los/index.htm.

Zhou, X., S.J. Adamowicz, L.M. Jacobus, R.E. DeWalt and P.D.N. Hebert. 2009. Towards a comprehensive barcode library for arctic life - Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in Zoology* 6:30.

Zhou, X., Y. Li, S. Liu, Q. Yang, X. Su, L. Zhou, M. Tang, R. Fu, J. Li and Q. Huang. 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* **2**:4.

Acknowledgements

Sarah J. Bourlat and Matthias Obst are funded by the Marine Genomics for Users EU FP7 project (Coordination and support action, call FP7-KBBE-2010-4) grant no. 266055. Special thanks go to Bernard Kloareg and Damien Guiffant for coordinating the project. Ángel Borja and Naiara Rodríguez-Ezpeleta are supported by the project DEVOTES (DEVelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status) funded by the EU 7th FP 'The Ocean for Tomorrow' Theme (grant agreement no. 308392), www.devotes-project.eu". David Murphy, Jan-Bart Calewaert, Andris Andrusaitis, Nikolaos Zampoukas, Gert Verreet, Gunnar Gerdts, and Chuck Cook are thanked for their input in the project and/or their comments on the manuscript.