Deep-Sea Research II ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Deep-Sea Research II

journal homepage: www.elsevier.com/locate/dsr2

Ocean microbial metagenomics

Lee J. Kerkhof [a,*], Robert M. Goodman [b]

[a] Institute of Marine and Coastal Sciences, School of Environmental and Biological Sciences, Rutgers University, NJ, USA
[b] Department of Ecology, Evolution, and Natural Resources, School of Environmental and Biological Sciences, Rutgers University, NJ, USA

ARTICLE INFO

Keywords:
Metagenomics
Ocean microbiology

ABSTRACT

Technology for accessing the genomic DNA of microorganisms, directly from environmental samples without prior cultivation, has opened new vistas to understanding microbial diversity and functions. Especially as applied to soils and the oceans, environments on Earth where microbial diversity is vast, metagenomics and its emergent approaches have the power to transform rapidly our understanding of environmental microbiology. Here we explore select recent applications of the metagenomic suite to ocean microbiology.

1. Introduction

The direct analysis of nucleic acids from marine samples has made studies like the Census of Marine Life (www.coml.org) and the characterization of uncultured oceanic microorganisms possible. Easy access to hereditary material from marine organisms collected in the field has revolutionized our views of genetic variation and cryptic speciation. As technology for DNA recovery, sequencing, assembly, and annotations have advanced, the potential to generate complete genome data, re-construct large DNA fragments, and infer functions of marine organisms has greatly improved. For example, at the writing of this manuscript, the National Center of Biotechnology Information (NCBI) reports 863 centers are producing data on 5275 genomes for bacteria alone. Another sign that an emerging scientific field has a broad audience and a large amount of interest is the number of studies and reviews that are generated in a short period of time. For the period of 2005–2008, there were 18 review articles and 45 original research papers published. For example, Abby and Daubin (2007) detailed large-scale genome structuring and horizontal gene transfer; Bansal (2005) describes the approaches and assumptions in the tools used in bioinformatics; Devereux et al. (2006) discuss the use of metagenomics and molecular tools in microbial source tracking; and Dunlap et al. (2006) present the potential and pitfalls of using large DNA fragments to generate reliable sources of marine secondary metabolites for developing new pharmaceuticals. Other recent reviews include genomic comparisons to identify the genes associated with Roseobacters (Moran et al., 2007; Brinkhoff et al., 2008). Furthermore, in June of 2005, Nature Reviews—Microbiology published a special issue on metagenomics including reviews by DeLong and Karl (2005), and Allen and Banfield (2005). Finally, Medini et al. (2008) describe the emerging paradigms regarding the pan genome and horizontal gene transfer in microbial ecology resulting from the new sequencing technologies. The reader is encouraged to seek out these additional compilations for a more thorough understanding of how metagenomics is re-structuring our understanding of ocean microbiology. Here we confine our focus to the marine microbial genomic data for the last few years. We discuss the findings, present some caveats to the assumptions using metagenomics, and point to some future directions for research.

Many of the papers we cite in this review use public databases of genes or genomes. In particular, there are 2 datasets that are prominently featured in many recent marine metagenomic studies. One is the Sargasso Sea dataset (Venter et al., 2004), representing 1.5 Gbp of compiled sequence from the BATS site, off the coast of Bermuda. This one study nearly doubled the known database of protein genes by adding 1.2 million new sequences. The other important dataset involves surface seawater samples collected along the first stages of a worldwide transect. The study is called the Global Ocean Survey (GOS) and the first reports by Rusch et al. (2007) indicate the combined dataset includes 6.25 Gbp of sequence data from 41 different locations, stretching from the Gulf of Maine, down the western North Atlantic coast, through the Panama Canal, and into the eastern tropical South Pacific. These two enormous datasets are only the beginning of the deluge of genomic sequence information from the environment that will soon be available to ocean scientists. Indications are that there is a great interest in using genomics to address fundamental questions regarding microbial ecology, metabolic potentials, genome structure, and the nature of adaptations to the marine environment.

* Corresponding author.
E-mail address: lkerkhof@rutgers.edu (L.J. Kerkhof).

0967-0645/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dsr2.2009.05.005

2. Approaches to analyzing genomic data

Metagenomic analyses rely on a large database of DNA/RNA sequences. The sequence can be generated utilizing two different approaches. One method utilizes bacterial artificial chromosomes (BACs), which are used to isolate large DNA fragments (20–400 kb; Liu et al., 2006) in an *Escherichia coli* recombinant library. These fragments are sub-cloned into smaller pieces (<2 kb) and subjected to Sanger sequencing. This methodology can yield reads up to 1 kb from a single reaction. The various contiguous sequences are then assembled in the computer and the assembly can be verified on the large cloned insert in the BAC library. Another approach utilizes rapid, direct sequencing methods, such as pyrosequencing, which can provide enormous amounts of short reads (<200 bp). The system can generate 400 million base pairs in a 10-h run. All pyrosequence is then assembled in the computer without the ability to directly verify using large cloned DNA fragments. An overview of the various metagenome sequencing technologies is presented by Medini et al. (2008).

2.1. Linkage to ribosomal RNA genes

The methods used to analyze this vast array of sequence data generally employ a series of targeted searches in a metagenomic dataset or the large DNA fragments of a BAC clonal library. For example, a highly successful and widely used strategy utilizes ribosomal RNA genes to isolate and characterize other genes physically linked and associated with the phylogenetic targets. This approach was first applied in the Stein et al. (1996) study searching of Archaeal genes by probing for ribosomal RNA genes in a library of over 3500 fosmid clones. In this pioneering work, the authors describe the identification of a single clone containing the ribosomal target as well as genes involved in elongation, amino-transfer, and heat shock. This tactic (targeting rRNA genes) was also followed by Beja et al. (2000) in the discovery of proteorhodopsin. Here, the authors were searching for fragments associated with the SAR 86 group, a largely uncharacterized branch of the *Proteobacteria*. The fortuitous result of finding a gene homologous to a light-driven pump from Archaea was exquisitely verified by cloning the proteorhodopsin gene in a heterologous host, demonstrating light-driven activity in *E. coli* in the presence of retinal, and ushering in a new understanding of the role of light in ATP formation in oceanic environments for heterotrophic bacteria.

In a recent example of the SSU gene pursuit by Grzymski et al. (2006) the authors describes the use of a fosmid library from Antarctica to identify large fragments from representatives of the α and γ Proteobacteria, *Bacteroidetes*, *Gemmatimonadetes* and high-G C Gram-positive bacteria. The aim was to detect protein genes associated with specific Antarctic bacterial groups that were ecologically important in polar-regions or were underrepresented in the genetic databases. Additionally, the authors wanted to glean information on cold adaptation for these microorganisms by comparison of the sequences from polar genes to mesothermic homologues. The results demonstrated changes in arginine+lysine ratios, decreases in aliphatic amino acids, increased stretches of disordered protein structure, and reductions in glutamic acids. Although not directly tested, the changes observed in polar amino acid content suggested reduced protein rigidity and provides a basis for monitoring cold adaptation in deep ocean environments.

Jensen and Lauro (2008) mined the GOS dataset for 16S rRNA genes associated with the Actinobacteria. Unlike other approaches listed above, the authors did not focus on genes linked to the SSU genes, but rather sought to expand the understanding of actinobacterial diversity at the subclass and order level in conjunction with a ribosomal RNA gene variable region pyrosequencing effort (e.g. V6 deep sequencing; Sogin et al., 2006). The authors found 33 OTUs from nearly 240 scaffolds with best matches to 70 sequences in Genbank. Most of the new Actinobacteria signatures came from estuarine settings (Delaware and Chesapeake Bay) or from tropical locals (Panama and Ecuador). The research indicates a large number of marine actinobacteria exist in a variety of oceanic settings. This finding has major implications for drug discovery from the seas, since natural products from the oceanic actinobacteria represent novel structures with pharmacological or bacteriostatic properties (for reviews see, Dunlap et al., 2006; Moore et al., 2005).

2.2. Targeted searches using characterized functional genes from Genbank

In addition to ribosomal RNA targets, a large variety of functional genes from environmental samples have been isolated and sequenced as PCR technology has improved. These genes associated with specific processes can also be used to characterize microbial communities in metagenomic studies. For example, McDonald and Vanlerberghe (2005) screened the Sargasso Sea protein database for alternative oxidation genes involved in the terminal reduction of O_2 to H_2O. This process represents a pathway to regenerate NAD+ without the proton translocation associated with the cytochrome pathway. As such, it is a non-energy-conserving branch of electron transport that does not lead to the production of ATP. The authors searched using the AOX gene from plant mitochondria and a plastoquinone terminal oxidase from cyanobacteria. This search yielded 37 homologous genes of both prokaryotic and eukaryotic origins. These findings indicated there were 7 different synteny groups of AOX genes and that this non-respiratory process may contribute significantly to O_2 consumption in the ocean.

In a similar vein, Badger et al. (2006) investigated the genes of the carbon-concentrating mechanism (CCM) in cyanobacteria. Analysis included the genes involved in inorganic carbon transport, formation of the proteinaceous coat of the carboxysome, Rubisco, and carbonic anhydrases. The findings demonstrated that 2 types of carboxysomes have evolved in the α and β cyanobacteria. It appears that the α cyanobacteria are restricted to open-ocean environments while the β cyanobacteria are present in both freshwater and saltwater systems. Similar findings were elucidated for the inorganic carbon (Ci) uptakes systems, although more physiological studies are required before it will be possible to interpret the possession of specific genes in an ecological context.

In an effort to better elucidate the genes involved in secondary metabolite biosynthesis, Fieseler et al. (2006) identified the polyketide synthase (PKS type I) genes in a metagenomic library from over 20 marine sponges from the Mediterranean, the Pacific, and the Caribbean for a bioprospecting effort. These genes are responsible for synthesis of many of the novel pharmacologically active metabolites obtained from the marine environment. Nearly 90,000 cosmid clones (representing 3.2 Gb of DNA from the Pacific sponge, *Theonella swinhoei*) and a library of 30,000 fosmid clones from *Alplysina aerophoba* were screened for PKS genes by PCR. The authors describe the discovery of 150 marine PKS genes in which 127 genes forming an independent clade from all previously described PKS or FAS gene sequences in Genbank. The effort yielded 3 intact PKS operons with a ketide synthase, an acyltransferase, a dehydratase, a methyltransferase, an enoylreductase, a ketoreductase, and an acyl carrier protein that may allow for heterologous synthesis of specific secondary metabolites

Please cite this article as: Kerkhof, L.J., Goodman, R.M., Ocean microbial.... Deep-Sea Research II (2009), doi:10.1016/j.dsr2.2009.05.005

from marine sources for improved pharmaceutical development (e.g. Dunlap et al., 2006).

Likewise, in an effort to define the mechanisms controlling genome structure, Kagan et al. (2008) explored the tryptophan pathway in the Sargasso Sea genomic dataset. The authors used the *Bacillus subtilis* operon as a probe and found over 4000 homologous genes associated with almost 3000 scaffolds from this anabolic pathway. The data suggested most tryptophan pathways contained either full or split operons. A phylogenetic analysis demonstrated the full operons were more closely related to known microorganisms with wide metabolic potentials (e.g., *E. coli*, *Vibrio parahaemolyticus*, *Listeria monocytogenes*, etc.) while the split operons were more closely related to marine oligotrophs, such as *Pelagibacter ubique*. These studies may help elucidate how genes fuse and mini operons are created in oceanic settings.

Krupovic and Bamford (2007) employed the sequence from a major protein coat protein of the marine lytic phage PM2 to identify homologous regions in complete bacterial genomes in Genbank to perform comparative genomics on viruses. Thirteen prophages from 11 bacterial genomes (mostly *Vibrio* and *Photobacterium*) were identified using this approach. The authors were focusing on identifying regions of "self" and "non-self", which are comparable to the core and the pan genomes of bacteria.

Finally Monier et al. (2008) probed the GOS dataset with the *polB* gene, a phage polymerase to re-construct the phylogenetic association of over 800 viral sequences. The researchers were able to relate phage type polymerases with water temperature along the transect lines. For example, the phage *polB's* showed a higher abundance with respect to eukaryotic *polB's* in temperatures $>20\,^{\circ}C$ while an inverse relationship was demonstrated at lower temperatures. The researchers also concluded that the *Mimiviridaea* are the most abundant and widespread of the large eukaryotic DNA viruses.

2.3. Bulk homologies searches to Genbank

As the ability to compile random sequences into contiguous fragments by computer became feasible, the search for bulk homologies using assembled data from large insert environmental libraries in Genbank became more robust. For example, Delong et al. (2006) created fosmid clonal libraries from the upper/lower euphotic zone (10, 70, 130 m), the mesopelagic (200, 500, 770 m) and the seafloor (4000 m). Nearly 5000 fosmids from each depth were sequenced with approx. 64 Mbp of DNA and a total of 4.5 Gbp. The authors found depth-dependent occurrence of many genes within their dataset, with chemotactic and photosynthetic pathways predominating the surface samples. In contrast, transposases and respiratory dehydrogenases were more abundant in the deeper samples. Viral genes were found throughout much of the sample set. Although this study represents a extensive characterization of a single site, the approach of comparisons along environmental gradients may lead to a better understanding of the genetic adaptations that enable specific microorganisms to survive and thrive in various marine habitats.

Biddle et al. (2008) investigated subsurface microbial communities at 1, 16, 32, and 50 m below the seafloor. Using pyrosequencing techniques, the authors generated nearly 62 Mb of genomic information with $>6\times10^5$ reads of approximately 100 bp each. Standard BLAST searches with expectancy matches $>10^{-6}$ were able to identify less than 15% of the total identified ORFs. Few of these identified genes demonstrated any depth relationship, including genes involved with amino acid, carbohydrate, or nitrogen metabolism. Only genes involved with aromatic or phosphorous metabolism, cell communication, or locomotion showed changes in abundance with depth of the sediment. The authors of the study concluded, "because the majority of the metagenome did not match database sequences, little information about the dominant metabolic functions can be deduced for these sites." That limitation should change with time as more model organisms from deep-sea sediments, whose physiologies are well characterized, as a reference framework to interpret similar metagenome datasets.

Frias-Lopez et al. (2008) used the samples collected in the euphotic zone off Hawaii to create a metagenome transcript library. Message RNA was extracted and linearly amplified by polyadenylation using a T7 RNA polymerase system. The fidelity of this amplification was tested using a *Prochlorococcus* system in culture compared with unamplified mRNA ($r^2>0.85$). This approach allowed the authors to retrieve and analyze 14 Mbp of transcripts by pyrosequencing. Interestingly, the most expressed sequences were associated with hypothetical open reading frames. Frias-Lopez found nearly 43% of the cDNAs did not match anything in either Genbank or the Global Ocean Survey (GOS) database [bit scores <40]. The majority of the expressed genes that were detected could not be related to specific phylogenetic groups. However, many of the characterized transcripts were associated with groups such as *Prochlorococcus*.

Grzymski et al. (2008) investigated the genome of the *Alvinella pompeijana* episymbiont using a 38.5 Mb library of 1–4 kb plasmid clones with 270,000 reads. The authors were able to delineate the genes involved in carbon fixation, denitrifiction, and sulfide oxidation for the epibionts. Additionally, they reported the episymbionts had a more acidic, more hydrophilic proteins that used significantly more Asp, Gly, Ile, Arg, and Thr in comparison with the *Epsilonproteobacteria* in the database. The presence of epibiont genes involved in vitamin production also provided a testable hypothesis for the benefit the host receives from this unique relationship.

2.4. Verifications and testing of first principles

Some studies of marine metagenomes rely on the identification of unique features that are then verified with respect to the original findings or can be used to generate first principle understandings of genome structure. For example, Chen et al. (2006) found sequences homologous to viral genes in the *Silicibacter* genome and tested for the ability of prophage to become lytic by treating cultures with mitomycin C. Five different phage particles were identified, after induction, which shared little homology, but accounted for nearly 5% of the host genome. This research approach isolated the inducible prophages in *Silicibacter* and may provide a means to study the mechanisms of horizontal gene transfer for this organism in the ocean.

Foerstner et al. (2005) investigated the variation in GC content for 4 metagenomic databases (Sargasso Sea (Venter et al., 2004); Whale carcass (Tringe and Rubin, 2005); acid mine biofilm (Tyson et al., 2004); and Minnesota farm soil (Tringe and Rubin, 2005)). The authors found a surprisingly narrow range of GC content for the overall sites, which was more restricted than the average GC content of the various microbial phyla that comprised the samples. The results suggested that an unknown mechanism existed which focuses the overall microbial community GC content with respect to location.

Kettler et al. (2007) used a comparative genome approach to establish the genetic basis for differences between *Prochlorococcus* and *Synechococcus* and high-light/low-light adaptation. The researchers investigated 12 *Prochlorococcus* genomes and 4 *Synechococcus* genomes. Of the 1855–3017 genes represented by organisms within these two genera, only 33 genes were found to be unique to all *Prochlorococcus*. A core genome for

Please cite this article as: Kerkhof, L.J., Goodman, R.M., Ocean microbial.... Deep-Sea Research II (2009), doi:10.1016/j.dsr2.2009.05.005

Prochlorococcus of 1250 genes was observed while the pan genome was nearly 5*x* larger with over 5700 genes. A similar study by Palenik et al. (2006) used the genomic variation in *Synechococcus* to understand the genetic differences between coastal and open-ocean cyanobacteria. The findings indicated a near doubling of response regulators in coastal genomes with a significant increase in metallo-enzymes and metal storage proteins and a large decrease in phosphorous uptake genes.

Martinez et al. (2007) searched nearly 13,000 clones from a library of fosmids obtained using Hawaiian surface water samples as described in Delong et al. (2006) for the presence of proteorhodopsin genes that could be heterologously expressed in *E. coli*. The authors searched for a characteristic color shift that was exhibited in the original report of proteorhodopsin (Beja et al., 2000), i.e. a red or orange pigmentation in *E. coli* when exposed to retinal. Interestingly, only 3 clones from this library exhibited a change in pigmentation during the screening process. Two of these clones were further characterized and found to contain the genes necessary for retinal formation as well as a functional proteorhodopsin that could generate a proton motive force in the light.

3. Caveats

Although great strides in isolating and characterizing genes from marine organisms have been made in recent years, a few limitations should be mentioned that directly impact the interpretation of marine metagenomic data. For example, many studies rely on the ability to clone DNA fragments [large (20–400 kb) and small (<2 kb)] to supply the fluorescent Sanger sequencing machines. This approach raises the potential for clonal bias to influence the interpretation of metagenomic data. The observation that clonal libraries and more direct measures [such as terminal restriction length polymorphism analysis (T-RFLP), denaturing gradient gel electrophoresis (DGGE), or FISH hybridizations] do not agree has been detailed in the literature (Phelps et al., 1998; Kerkhof et al., 2000; Cottrell and Kirchman, 2000; Vetriani et al., 2003). Although some possible reasons for generating clonal bias have been postulated (Phelps et al., 1998), the exact mechanisms contributing to the selective pressures within *E. coli* recombinant libraries affecting particular PCR products remains elusive. Although it is well known that increased numbers of transformants are obtained with increasing amount of DNA during *E. coli* transformation, these studies are typically done with a single plasmid, such as PBR322 or PUC19. There have not been sufficient studies of transformation bias looking at mixtures of plasmids or cloned genes. This assumption that shotgun libraries represent actual abundance in the transformation mixture may not be accurate and all randomly sheared DNA fragments may not clone with equal efficiency. It is conceivable that some random DNA fragment disrupt *E. coli* gene regulation and either trigger a significant lowering of growth rate or possibly cause death and not be well represented in random libraries.

A second concern in creating clonal libraries is that the *E. coli* cells used for transformations can become saturated. For example, when DNA concentrations less than 1 ng are used to transform hexamine cobalt-treated cells, a linear response to DNA concentration is seen (Hanahan 1983). Concentrations above 3.5 ng lead to saturation of this type of competent cell. Calcium chloride-treated cells saturate at higher DNA concentrations [approx. 500 ng]. In contrast, a linear response to DNA concentration is seen with electro-competent cells transformed with up to 300 ng of DNA (Dower et al., 1988). However, most cloning/transformation kits that are widely used in metagenomic studies, utilize 50 ng of plasmid and chemically competent cells. Often researchers do not transform with the lower amounts of ligated plasmid (e.g., 1 ng) that are required to remain in the linear range of the transformation curve. Interestingly, this notion of clonal bias in recombinant libraries is also supported by the results by Grzymski et al. (2008) where the most abundant class of cloned genes from the *Alvinellid* symbiont library were associated with plasmid replication initiation. Yet, there was no evidence of plasmid DNA in genomic DNA preps from the samples.

Other problems stem from the shotgun sequence and assembly approach and are discussed in Rusch et al. (2007). A lowering of assembly stringency from 98% to 80% increases the size of assembled scaffolds in 10 and 100 Kb contigs. However, this may lead to "highly similar sequences being lumped within a single read or to reads often recruiting to distantly related sets of genomes". Another issues stems from a "guilt by association" that is inherent in homology searches through Genbank. Specifically, imagine that a protein gene is found to be 60% homologous to another protein whose function has been experimentally determined. This new protein is assigned the same function and annotated based on the homology. All subsequent searches using new DNA sequences that are found to be homologous to this newly annotated protein will also be assumed to have the same function. However, the actual homology to the experimentally validated protein could be quite low, casting doubt on the annotation. Therefore, whenever possible, the annotation of protein-encoding genes detected in a metagenome should be confirmed by comparative analyses of the deduced amino acid sequence of key regions of the protein, such as its active site, with known protein homologues in the database. Finally, there are great difficulties in presenting such large metagenomic datasets to the scientific community. For example, a few of the manuscripts cited in this review contain more supplemental figures than actual figures in papers (e.g., Delong, Frias-Lopez). There will be a tendency for the scientific community to ignore the supplemental data because of the additional efforts needed to acquire the additional figures and tables. Furthermore, the supplemental data will only be available as long as journals and their publishers are willing to curate.

4. Future directions

Although the amount of information generated in these genomic studies is staggering, many of the open reading frames that are identified have no homologies to known proteins in Genbank. Therefore, there is a need to develop an approach to characterize these hypothetical ORFs that comprise nearly 50% of the genomic data that is currently being collected. Furthermore, as data and testable hypotheses are generated from genomic data, it is essential that a verification scheme be established using additional molecular tools. For instance, Giovannoni et al. (2005) presented an excellent example of verifying genomic data using proteomics. In this study, the researchers found a proteorhodopsin gene in the genome from *Pelagibacter ubique*. Maldi-TOF analysis of cultures and surface seawater off the Oregon coast demonstrated a 10 amino acid peptide, consistent with the expression of the *Pelagibacter* proteorhodopsin in the lab and in the field. Another approach to link genomes to specific activities was demonstrated by Mou et al. (2008). The manuscript describes the use of a thymidine analogue (bromodexoyuridine) to capture newly synthesized DNA associated with the utilization of 100 nM dimethyl sulphoniopropionate or vanillate in coastal waters after a 12-h incubation. Analysis of the immunocaptured DNA yielded 28 Mbp of genomic data from bacteria replicating their chromosomes and capable of BrDU uptake. Finally, the possibility of

Please cite this article as: Kerkhof, L.J., Goodman, R.M., Ocean microbial.... Deep-Sea Research II (2009), doi:10.1016/j.dsr2.2009.05.005

single-cell genome amplification from oceanic microorganisms combined with PCR screening has been demonstrated by Stepanauskas and Sieracki (2007). Here, the researchers physically separate the cells using fluorescence activated cell sorting and amplify the chromosome by multiple displacement amplification. They used the method to screen 11 different microorganisms from the α/γ *Proteobacteria*, *Flavobacteria*, and the *Sphingobacteria* for SSU rRNA, proteorhodopsin, bacteriochlorophyll, nitrogenase, and assimilative nitrate reductase genes. This approach also could be used to verify *in silico* assemblies on a single-cell basis.

In conclusion, the use of genomic sequence data from marine samples is proceeding at an exponential pace. At present, only a few, well-funded laboratories or institutes are capable of generating the huge amount of raw sequence that is represented by efforts such as the GOS dataset. However, as technology and computer tools improve, the routine use of genomic data by nearly every member the oceanographic community is sure to become possible.

References

Abby, S., Daubin, V., 2007. Comparative genomics and the evolution of prokaryotes. Trends in Microbiology 15, 135–141.

Allen, E.E., Banfield, J.F., 2005. Community genomics in microbial ecology and evolution. Nature Reviews Microbiology 3, 489–498.

Badger, M.R., Price, G.D., Long, B.M., Woodger, F.J., 2006. The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. Journal of Experimental Botany 57, 249–265.

Bansal, A.K., 2005. Bioinformatics in microbial biotechnology—a mini review. Microbial Cell Factories, 4.

Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C., Jovanovich, S.B., Feldman, R.A., DeLong, E.F., 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environmental Microbiology 2, 516–529.

Biddle, J.F., Fitz-Gibbon, S., Schuster, S.C., Brenchley, J.E., House, C.H., 2008. Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. Proceedings of the National Academy of Sciences of the United States of America 105, 10583–10588.

Brinkhoff, T., Giebel, H.-A., Simon, M., 2008. Diversity, ecology, and genomics of the *Roseobacter clade*: a short overview. Archives of Microbiology 189, 531–539.

Chen, F., Wang, K., Stewart, J., Belas, R., 2006. Induction of multiple prophages from a marine bacterium: a genomic approach. Applied and Environmental Microbiology 72, 4995–5001.

Cottrell, M.T., Kirchman, D.L., 2000. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. Applied and Environmental Microbiology 66, 5116–5122.

DeLong, E.F., Karl, D.M., 2005. Genomic perspectives in microbial oceanography. Nature 437, 336–342.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., Martinez, A., Sullivan, M.B., Edwards, R., Brito, B.R., Chisholm, S.W., Karl, D.M., 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science 311, 496–503.

Devereux, R., Rublee, P., Paul, J.H., Field, K.G., Santo Domingo, J.W., 2006. Development and applications of microbial ecogenomic indicators for monitoring water quality: report of a workshop assessing the state of the science, research needs and future directions. Environmental Monitoring and Assessment 116, 459–479.

Dower, W.J., Miller, J.F., Ragsdale, C.W., 1988. High-efficiency transformation of *Escherichia coli* by high-voltage electroporation. Nucleic Acids Research 16, 6127–6145.

Dunlap, W.C., Jaspars, M., Hranueli, D., Battershill, C.N., Peric-Concha, N., Zucko, J., Wright, S.H., Long, P.F., 2006. New methods for medicinal chemistry-universal gene cloning and expression systems for production of marine bioactive metabolites. Current Medicinal Chemistry 13, 697–710.

Fieseler, L., Quaiser, A., Schleper, C., Hentschel, U., 2006. Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum Poribacteria by environmental genomics. Environmental Microbiology 8, 612–624.

Foerstner, K.U., von Mering, C., Hooper, S.D., Bork, P., 2005. Environments shape the nucleotide composition of genomes. Embo Reports 6, 1208–1213.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., DeLong, E.F., 2008. Microbial community gene expression in ocean surface waters. Proceedings of the National Academy of Sciences of the United States of America 105, 3805–3810.

Giovannoni, S.J., Bibbs, L., Cho, J.C., Stapels, M.D., Desiderio, R., Vergin, K.L., Rappe, M.S., Laney, S., Wilhelm, L.J., Tripp, H.J., Mathur, E.J., Barofsky, D.F., 2005. Proteorhodopsin in the ubiquitous marine bacterium SAR11. Nature 438, 82–85.

Grzymski, J.J., Carter, B.J., DeLong, E.F., Feldman, R.A., Ghadiri, A., Murray, A.E., 2006. Comparative genomics of DNA fragments from six Antarctic marine planktonic bacteria. Applied and Environmental Microbiology 72, 1532–1541.

Grzymski, J.J., Murray, A.E., Campbell, B.J., Kaplarevic, M., Gao, G.R., Lee, C., Daniel, R., Ghadiri, A., Feldman, R.A., Cary, S.C., 2008. Metagenome analysis of an extreme microbial symbiosis reveals eurythermal adaptation and metabolic flexibility. Proceedings of the National Academy of Sciences of the United States of America 105, 17516–17521.

Hanahan, D., 1983. Studies on tranformation in *Escherichia coli* with plasmids. Journal of Molecular Biology 166, 557–580.

Jensen, P.R., Lauro, F.M., 2008. An assessment of actinobacterial diversity in the marine environment. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 94, 51–62.

Kagan, J., Sharon, I., Beja, O., Kuhn, J.C., 2008. The tryptophan pathway genes of the Sargasso Sea metagenome: new operon structures and the prevalence of non-operon organization. Genome Biology, 9.

Kerkhof, L., Santoro, M., Garland, J., 2000. Response of soybean rhizosphere communities to human hygiene water addition as determined by community level physiological profiling (CLPP) and terminal restriction fragment length polymorphism (TRFLP) analysis. Fems Microbiology Letters 184, 95–101.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G.M., Richardson, P., Chisholm, S.W., 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. Plos Genetics 3, 2515–2528.

Krupovic, M., Bamford, D.H., 2007. Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. Bmc Genomics, 8.

Liu, W., Zhao, Y., Liu, Z., Zhang, Y., Lian, Z., Li, N., 2006. Construction of a 7-fold BAC library and cytogenetic mapping of 10 genes in the giant panda (*Ailuropoda melanoleuca*). BMC Genomics 7, 294.

Martinez, A., Bradley, A.S., Waldbauer, J.R., Summons, R.E., DeLong, E.F., 2007. Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. Proceedings of the National Academy of Sciences of the United States of America 104, 5590–5595.

McDonald, A.E., Vanlerberghe, G.C., 2005. Alternative oxidase and plastoquinol terminal oxidase in marine prokaryotes of the Sargasso Sea. Gene 349, 15–24.

Medini, D., Serruto, D., Parkhill, J., Relman, D.A., Donati, C., Moxon, R., Falkow, S., Rappuoli, R., 2008. Microbiology in the post-genomic era. Nature Reviews Microbiology 6, 419–430.

Monier, A., Claverie, J.M., Ogata, H., 2008. Taxonomic distribution of large DNA viruses in the sea. Genome Biology, 9.

Moran, M.A., Belas, R., Schell, M.A., Gonzdlez, J.M., Sun, F., Sun, S., Binder, B.J., Edmonds, J., Ye, W., Orcutt, B., Howard, E.C., Meile, C., Palefsky, W., Goesmann, A., Ren, Q., Paulsen, I., Ulrich, L.E., Thompson, L.S., Saunders, E., Buchanlo, A., 2007. Ecological genomics of marine Roseobacters. Applied and Environmental Microbiology 73, 4559–4569.

Moore, B.S., Kalaitzis, J.A., Xiang, L.K., 2005. Exploiting marine actinomycete biosynthetic pathways for drug discovery. Antonie van Leeuwenhoek 87, 49–57.

Mou, X.Z., Sun, S.L., Edwards, R.A., Hodson, R.E., Moran, M.A., 2008. Bacterial carbon processing by generalist species in the coastal ocean. Nature 451, 708–714.

Palenik, B., Ren, Q.H., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., Madupu, R., Nelson, W.C., Brinkac, L.M., Dodson, R.J., Durkin, A.S., Daugherty, S.C., Sullivan, S.A., Khouri, H., Mohamoud, Y., Halpin, R., Paulsen, I.T., 2006. Genome sequence of *Synechococcus* CC9311: insights into adaptation to a coastal environment. Proceedings of the National Academy of Sciences of the United States of America 103, 13555–13559.

Phelps, C.D., Kerkhof, L.J., Young, L.Y., 1998. Molecular characterization of a sulfate-reducing consortium which mineralizes benzene. FEMS Microbiology Ecology 27, 269–279.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D.Y., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.H., Falcon, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M., Venter, J.C., 2007. The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical pacific. Plos Biology 5, 398–431.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences of the United States of America 103, 12115–12120.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., DeLong, E.F., 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment front a planktonic marine Archaeon. Journal of Bacteriology 178, 591–599.

Stepanauskas, R., Sieracki, M.E., 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proceedings of the National Academy of Sciences of the United States of America 104, 9052–9057.

Tringe, S.G., Rubin, E.M., 2005. Metagenomics: DNA sequencing of environmental samples. Nature Reviews Genetics 6, 805–814.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F., 2004. Community

structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D.Y., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., Smith, H.O., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.

Vetriani, C., Tran, H.V., Kerkhof, L.J., 2003. Fingerprinting microbial assemblages from the oxic/anoxic chemocline of the Black Sea. Applied and Environmental Microbiology 69, 6481–6488.

Please cite this article as: Kerkhof, L.J., Goodman, R.M., Ocean microbial.... Deep-Sea Research II (2009), doi:10.1016/j.dsr2.2009.05.005