Reference: *Biol. Bull.* **227:** 107–116. (October 2014)

Deciphering Diversity and Ecological Function From Marine Metagenomes

HOLLY M. BIK*

UC Davis Genome Center, University of California—Davis, One Shields Ave, Davis, California 95616

Abstract. Metagenomic sequencing now represents a common, powerful approach for investigating diversity and functional relationships in marine ecosystems. High-throughput datasets generated from random fragments of environmental DNA can provide a less biased view of organismal abundance (*versus* PCR-based amplicon sequencing) and enable novel exploration of microbial genomes by recovering genome assemblies from uncultured species, identifying ecological functions, and reconstructing metabolic pathways. This review highlights the current state of knowledge in marine metagenomics, focusing on biological insights gained from recent environmental studies and detailing commonly employed methods for data collection and analysis.

Introduction

High-throughput sequencing technologies have dramatically and fundamentally altered our ability to characterize global biodiversity and explore ecosystem function. The generation of millions of DNA sequences is now becoming a routine part of biological studies across diverse habitats (Dinsdale *et al.*, 2008; Fierer *et al.*, 2012; Yatsunenko *et al.*, 2012; Mason *et al.*, 2014). This phenomenon is particularly apparent in marine ecosystems, where the use of "shotgun metagenomics" (sequencing random fragments of environmental DNA, herein referred to solely as "metagenomics") was pioneered (Breitbart *et al.*, 2002; Venter *et al.*, 2004; Angly *et al.*, 2006; DeLong *et al.*, 2006), and where the application of new sequencing technologies continues to transform our understanding of oceanic habitats.

Perhaps the most ubiquitous application of high-throughput sequencing has been the amplification of conserved, phylogenetically informative marker genes from environmental DNA (typically genes encoding ribosomal RNA subunits or spacer regions, such as 16S for studies of Bacteria and Archaea, 18S for eukaryotic assemblages, and ITS for fungi). Although rRNA studies are sometimes mistakenly referred to as "metagenomic" approaches, their reliance on conserved PCR primer sets often precludes the capture of the total community diversity in an environmental sample (Logares *et al.*, 2013). One striking example is the recent discovery of the Nanohaloarchaea from metagenomic data; this group of Archaea is abundant in hypersaline habitats but was previously undescribed because it fails to amplify with standard 16S environmental primer sets (Narasingarao *et al.*, 2012). Nonetheless, marker gene studies have provided an unprecedented view of microbial communities: deep sequencing has enabled the discovery of new lineages (Jones *et al.*, 2011), provided insight into taxa comprising the "rare biosphere" (Sogin *et al.*, 2006), identified temporal and depth-related differentiation across bacterial SAR11 ecotypes (Vergin *et al.*, 2013), revealed previously unknown dominance of microbial eukaryote groups in marine sediments (Platyhelminthes; Fonseca *et al.*, 2010), and served as a basis for model-based predictions about global species distributions in bacteria (Ladau *et al.*, 2013; Sul *et al.*, 2013). A wealth of investigations focusing on environmental rRNA amplicons have been published, and a number of authors have reviewed relevant findings from marine marker gene studies (Gilbert and Dupont, 2011; Caron *et al.*, 2012; Richards *et al.*, 2012) and detailed marker gene workflows and analyses (Zaneveld *et al.*, 2011; Bik *et al.*, 2012b; Lladser and Knight, 2013). This review focuses solely on shotgun metagenomics and will exclude discussion of amplicon-based approaches.

There exists a suite of review articles covering various aspects of metagenomic studies, from sampling to downstream bioinformatics workflows (Meyer *et al.*, 2008; Simon

Received 16 March; accepted 10 July 2014.

* To whom correspondence should be addressed. E-mail: holly.bik@gmail.com

and Daniel, 2011; Sun *et al.*, 2011; Thomas *et al.*, 2012; Hunter *et al.*, 2014). Gilbert and Dupont (2011) even offer a review of metagenomic reviews, of which there exists a substantial list. This review does not cover specific methodological workflows or computational tools in detail, but instead focuses on the biological insights that have been gained using metagenomic approaches across different marine ecosystems. In addition, since the field of metagenomics itself has been heavily reviewed in past years, this review focuses only on the most recent studies, generally from 2011 onward; for an overview of earlier work, Gilbert and Dupont (2011) present another comprehensive review of marine metagenomics, although note that the authors lump both rRNA amplicon studies and shotgun environmental sequencing under the umbrella term of "metagenomics."

What Is Possible From Metagenomics?

Workflows for analyzing metagenomic data are much less standardized than those for environmental rRNA amplicons. Most environments harbor a complex and diverse array of microbial species; gaining adequate sequencing depth to cover all genomes remains a challenge, and the heterogeneous nature of metagenomes (in terms of population and strain-level genomic variation) complicates the process of assembling raw sequence reads into longer contigs (Scholz *et al.*, 2012). Nonetheless, metagenomes can provide an important source of information, particularly for studies looking to analyze a subset of data for a specific purpose. For example, metagenomic data can provide an alternative, PCR-free method (*versus* amplicon sequencing) for assessing the taxonomic composition and diversity of microbial communities. Pipelines developed for this approach typically mine phylogenetically informative marker genes from shotgun data (rRNA genes or single-copy protein coding genes, usually representing $<1\%$ of a metagenomic dataset (Kembel *et al.*, 2011; Darling *et al.*, 2014). Other studies focus their analyses on specific genes known to be involved in important biochemical or metabolic processes. For example, in the context of the *Deepwater Horizon* oil spill, the identification of benzylsuccinate synthase-like genes (*bssA)* in metagenomic datasets has been used to assess the microbial community potential for degrading hydrocarbon compounds (alkylbenzenes; Kimes *et al.*, 2013). However, the value of metagenomic datasets goes beyond the questions asked in individual studies. Shotgun data deposited in public archives can be downloaded and analyzed by any researcher, genomes reconstructed from environmental datasets help to expand the phylogenetic diversity contained in online databases, and metagenomes from diverse habitats can be used for fragment recruitment approaches that complement genome sequencing projects.

Metagenomes as a community resource

The *Sorcerer II* Global Ocean Sampling Expedition dataset (herein abbreviated as GOS (Rusch *et al.*, 2007; Yooseph *et al.*, 2007) has been credited with kickstarting the metagenomic revolution. Despite common gripes about poor experimental design and a lack of accompanying metadata, such early environmental studies have proven their long-term utility (Gilbert *et al.*, 2011b). Exploratory biodiversity surveys have accelerated our knowledge of marine microbiology, greatly expanded the field of microbial ecology, and served as a reference dataset for comparative metagenomics and annotation (Gilbert and Dupont, 2011; Gilbert *et al.*, 2011b).

This trend of leveraging public metagenomic datasets continues. Elo*e et al.* (2011) used comparative analyses to identify enrichments of transporter protein classes in deep-sea metagenomes, suggesting that this distinction from shallow-water habitats is driven by a cellular need to cope with increasing hydrostatic pressure. In the same study, metagenomes from the Puerto Rico Trench showed significant enrichment of transporter proteins related to heavy metal efflux compared to shallow-water samples. This raises the possibility that deep-sea organisms must maintain the capacity for heavy metal resistance, potentially due to the rapid depletion of organic matter in sinking marine snow. Published shotgun data can also provide an important source of evidence for assessing the diversity and distribution of uncultured lineages across marine habitats. Lloyd *et al.* (2013) used published metagenome data from marine sediments to search for homologs of Archaeal peptidases identified through single-cell genome sequencing. This investigation found such peptidases to be widespread, thus hinting that Archaea have a previously unknown role in remineralizing proteins in anoxic marine sediments. Finally, public metagenome datasets can be repeatedly re-analyzed to ask different questions and assess different fractions of the community. Ma *et al.* (2012) focused on plasmid recovery from published pelagic metagenomic datasets (including GOS data), identifying different levels of complexity in plasmid assemblages recovered between samples. The authors identified a higher diversity of plasmids in GOS metagenomes, suggesting greater host (microbial) diversity in these samples. In addition, another marine dataset appeared to contain a high level of small, cryptic plasmids—such plasmids do not present an obvious evolutionary advantage but may lead to increased viral resistance for hosts. In both cases, analysis of the plasmid fraction within metagenomes proved useful for making inferences about microbial communities and hypotheses regarding potential evolutionary dynamics in marine environments.

Genomes from metagenomes

In recent years, metagenomic sequencing has provided a novel approach for isolating and reconstructing genome sequences from uncultured microbial taxa. An increasing number of researchers are using "binning" techniques to assemble draft genomes from mixed, metagenomic community samples (Sharon and Banfield, 2013). Reliance on this approach is partly driven by the fact that single-cell genomics is a difficult process that often results in highly fragmented genomes, and even putatively complete microbial genomes can remain as non-contiguous sequences in tens to hundreds of contigs (Rinke *et al.*, 2013). In addition, whole-gene amplification steps are often necessary to produce sufficient material for single-cell library construction (*e.g.,* multiple displacement amplification), which can skew sequence data to the point where a significant part of the genome is underrepresented (Marine *et al.*, 2014). One of the most commonly employed binning methods is the construction of emergent self-organizing maps (ESOMs; Dick *et al.*, 2009), a visual map of genome signatures based on tetranucleotide frequencies of sequence fragments. Due to the nature of the bioinformatics methods involved, genome reconstruction is limited by the complexity of the metagenomic sample, and so far has been restricted to low-diversity environments—acid mine drainage habitats (Tyson *et al.*, 2004); hypersaline lake ecosystems (Podell *et al.*, 2013)—or samples in which it is comparatively easy to isolate sequences from the target genome (*e.g.,* bacterial symbiont genomes from a eukaryotic host, where genome signatures based on %GC and read coverage are distinct).

A variety of computational techniques can be used to assess metagenomic datasets, identify targets for genome reconstruction, and isolate subsets of data for assembling full or partial genomes. Iverson *et al.* (2012) constructed assembly graphs to link contigs putatively representing the same genome, using information from mate-pair sequencing, read coverage, and nucleotide composition. This resulted in a closed genome from an uncultured representative of the marine group II *Euryarchaeota.* Other studies have been able to recover a large number of putatively complete viral genomes: 208 phage genomes (Mizuno *et al.*, 2013b); 608 ssDNA viruses (Labonté and Suttle, 2013). Currently, it is feasible to use such reconstruction approaches only to recover the small, compact genomes of Bacteria, Archaea, and viruses; separating larger, more complex eukaryotic genomes from environmental metagenomes is likely to remain a significant challenge.

In marine ecosystems, genome binning has also been used to identify and isolate symbiont genomes from host species. Kwan *et al.* (2012) utilized shotgun metagenomic reads from a tunicate species (*Lissoclinum patella*) to reconstruct the complete genome of an endosymbiotic Alphaproteobacteria. This symbiont, *Candidatus Endolissoclinum faulkneri*, was shown to have undergone a high degree of genome reduction, but maintained pathways for patellazole biosynthesis (a secondary metabolite that is typically toxic to eukaryotic cells). In this study, maintenance of the patellazole pathway suggested that the bacterial compound may play an important role in host defense. Metagenomic data have also been used to recover the genome sequences of chemosynthetic microbial symbionts, most notably from the gutless oligochaete *Olavius algarvensis* (Dubilier *et al.*, 2008).

Fragment recruitment: Linking metagenomes with genomes and single-cell contigs

Fragment recruitment is another important analytical method in metagenomic studies. In this approach, raw or assembled sequence reads from metagenomic datasets are compared to available genome sequences or collections of contigs and scaffolds (*e.g.,* partially assembled genomes generated from single-cell sequencing). Regardless of whether a published genome exists as a complete or draft sequence, the taxonomic identity or phylogenetic position of the sequenced organism is known, thus making it possible to link proteins of interest with an organismal ID. Sequence similarity or protein homology searches are used to assess whether a sequenced organism (parts of its genome) is represented in metagenomic datasets collected from diverse environments. Recruitment approaches have been notable for demonstrating the environmental abundance of particular marine viruses, including SAR11 pelagiphage viruses (Zhao *et al.*, 2013) and the HMO-2011 phage infecting a bacterium in the SAR116 clade (Kang *et al.*, 2013). In another application, Swan *et al.* (2013) used metagenome fragment recruitment to identify correlations between environmental parameters (seawater temperature and latitude) and the abundance of specific genomes. Read recruitment is also critical for interpreting the relevance of hypothetical proteins, which would otherwise be ignored in metagenomic datasets due to the lack of functional annotation. In this scenario, fragment recruitment also enables researchers to determine the source (taxonomic ID) of such unannotated environmental reads or contigs. For example, metagenomic fragment recruitment was able to identify a number of subclade-specific hypothetical proteins corresponding to the SAR11 Ic bathytype, shown to be abundant in deepwater metagenomes (Thrash *et al.*, 2014). Combining metagenomics and single-cell genome sequencing has proven to be especially powerful in understudied habitats such as the deep sea. Due to the paucity of existing genome sequences for uncultivated taxa from this remote environment, even a handful of contigs from highly fragmented single-cell genomes can provide valuable information, facilitating deeper exploration of microbial communities and biogeographic patterns inferred from shotgun metagenomes (Eloe *et al.*, 2011).

Metagenomic Studies in Marine Environments

In marine systems, metagenomic data is most commonly used to compare patterns of diversity across samples (taxonomic, phylogenetic, and functional), and to identify the presence or absence of specific metabolic pathways that may help to explain the ecological roles of communities in a given habitat. Broad-scale diversity assessments and targeted analyses of functional genes are most often carried out in parallel, since both aspects are important for gaining knowledge of ecosystem function.

Studies of polluted marine sites provide some of the more obvious (but nonetheless powerful) insights from metagenomic sequencing. Plewniak *et al.* (2013) used shotgun sequencing to compare microbial metabolic potential between two sites on the French Mediterranean coast, one with heavy arsenic pollution and another with arsenic concentrations well below toxic levels. Genes involved in arsenic resistance and sulfate reduction were significantly enriched in the more polluted site, suggesting that the microbial communities have the capacity to disperse environmental arsenic and potentially lower its biological toxicity. Other studies have similarly used metagenomic sequencing to characterize the functional profiles of microbial communities in polluted habitats, identifying pathways for heavy metal resistance, stress response, and horizontal gene transfer (Kisand *et al.*, 2012; Thureborn *et al.*, 2013). An increased capacity for horizontal gene transfer appears to be consistent with harsh environmental conditions; microbial taxa often show enrichment of genes such as integrases or transposases that are involved in genetic recombination events, enabling species to cope with and adapt to extreme or competitive habitats such as hydrothermal vents (Tang *et al.*, 2013) and particle-associated microbial communities (Ganesh *et al.*, 2014).

Sample comparisons obtained through metagenomic sequencing are also helping to define previously unknown community distinctions and habitat boundaries in marine environments. Different microbial assemblages were shown to be abundant in vent water in shallow hydrothermal systems compared to in the overlying surface water, with further functional differentiation in the pathways for microbial carbon fixation and sulfur metabolism (Tang *et al.*, 2013). Li *et al.* (2014) identified an obvious distinction between host-associated communities and surrounding seawater, identifying symbiotic microbial lineages that were specific to *Lamellomorpha* sp. sponge tissues. The methodological scope of a study can also reveal biological partitions in marine ecosystems. In a study by Ganesh *et al.* (2014), the filter size used for sampling was strongly correlated with microbial community composition, much more so than depth of sample collection. In marine environments, the authors suggest that particle-associated microbial communities (larger filter size fractions) represent an important microhabitat with a distinct profile of functional genes related to particle adhesion (Ganesh *et al.*, 2014). Other evidence suggests that subsurface deep-sea sediments have a distinct metagenomic community profile compared to top sediment layers collected at the interface of water and sediment (Kimes *et al.*, 2013), and that organic carbon content of sediments may be important in structuring microbial communities, more so than geography. The functional capacity of marine metagenomic samples has also been linked to specific environmental parameters such as oxygen, salinity, and temperature (Thureborn *et al.*, 2013). Deep-sea metagenomic profiles have so far indicated that microbial taxa maintain an expanded metabolic repertoire indicative of an opportunistic lifestyle, and that taxa are subjected to relaxed purifying selection, in contrast to shallow-water communities (Eloe *et al.*, 2011).

Analysis of the phylogenetic diversity within metagenomic datasets suggests that in the water column, microbial assemblages in shallower depths consist of taxa that are more closely related than those in deep-sea communities, which show more phylogenetic breadth and whose species are more distantly related (Kembel *et al.*, 2011). However, patterns of phylogenetic diversity are not always straightforward—microbial assemblages appear to vary significantly according to the size fraction sampled, and these fractions can show contrasting patterns of diversity across environmental gradients (Allen *et al.*, 2012; Ganesh *et al.*, 2014).

Metagenomics has challenged our view of community ecology, suggesting that hypotheses derived from larger, eukaryotic taxa do not always hold true at the microbial level. Burke *et al.* (2011) reported that microbial communities associated with the macroalga *Ulva australis* were conserved at the level of functional genes, rather than comprising a specific set of taxa. The authors hypothesize that microbial colonization events are in fact a lottery, such that species with the necessary functional capacities who are the first to claim space on a host are the ones that are ultimately successful. Metagenomic studies are also shedding light on functional ecology in marine habitats by describing metabolic pathways and energy utilization by microbial communities, particularly in regard to pathways related to nitrogen and sulfur cycling (Allen *et al.*, 2012; Ugalde *et al.*, 2013; Ganesh *et al.*, 2014; Li *et al.*, 2014).

Case study: Deepwater Horizon

The 2010 *Deepwater Horizon* (DWH) oil spill represents a unique case study for the application of new sequencing technologies in marine environments. At the time of the spill, sampling methods and computational workflows were robust enough to facilitate widespread use of 454 and Illumina sequencing to investigate post-spill environmental

impacts in the Gulf of Mexico. Unsurprisingly, high-throughput rRNA marker gene surveys were a common approach for describing microbial communities as the oil spill progressed, identifying *Colwellia* and *Oceanospirillales* taxa as hydrocarbon-degrading bacterial taxa that became enriched in the deepwater "oil plume" (Hazen *et al.*, 2010; Baelum *et al.*, 2012). Other rRNA amplicon studies assessed the potential impact of oil on human gut microbiota (Kim *et al.*, 2012), the effects of COREXIT dispersants on bacteria (Hamdan and Fulmer, 2011), the bacterial community structure and response in Gulf of Mexico sediments (Kostka *et al.*, 2011; Liu and Liu, 2013), and shifts in the microbial eukaryote community between pre-spill and post-spill beach sediments (Bik *et al.*, 2012a).

Another group of studies focused on shotgun metagenomic sequencing, either alone or in conjunction with rRNA amplicons. Kimes *et al.* (2013) used an approach that combined metagenomic sequencing, clone library construction to search for functional genes, and metabolite profiling to assess the impact of the DWH oil spill on deep-sea sediments in the Gulf of Mexico. Although only three samples were included in this analysis, the results showed clear differentiation between samples on the basis of distance from the wellhead. Anaerobic metabolism was identified as a putatively important mechanism for hydrocarbon degradation in deep-sea sediments, compared to predominantly aerobic processes that governed oil-plume and surface oil degradation. Furthermore, the authors noted parallels between the microbial response in oil-contaminated sediments and communities observed in natural hydrocarbon seeps, where in both cases Deltaproteobacteria appear to assume a dominant role in degrading hydrocarbon compounds. Although oil-degrading *Oceanospirillales* bacteria appeared to be present in deep-sea sediments, the abundance of this taxon was quite low (<2% of sequences (Kimes *et al.*, 2013)) in contrast to the overwhelming dominance observed in the deepwater oil plume (>90% of sequences; Hazen *et al.*, 2010).

In another investigation of deep-sea sediments, Mason *et al.* (2014) generated an unprecedented amount of metagenomic data, obtaining 1 Tb of sequence data across 14 samples. In sediments that exceeded EPA standards for polycyclic aromatic hydrocarbon (PAH) toxicity, enrichment analyses found that genomic pathways for metabolism of hydrocarbon compounds were significantly more abundant in these samples. Metagenomics showed that sediment communities possessed specific pathways for cyclohexane degradation, and other general capabilities for degrading aliphatics and simple aromatics. Genes for PAH degradation were not predominant or enriched for between oiled and non-oiled samples, supporting previous evidence that PAHs have been recalcitrant and persistent in marine sediments after the DWH oil spill (Montagna *et al.*, 2013). Predictive relative metabolomic turnover (PRMT) analysis (Larsen *et al.*, 2011) was additionally used to demonstrate significant shifts in nitrogen cycling pathways resulting from hydrocarbon contamination. In terms of microbial community response, *Colwellia* bacteria appeared to play an important role in degrading hydrocarbon in sediments, potentially settling onto sediments *via* marine snow flocs originating from the deepwater oil plume. Joint deposition of bacteria and hydrocarbons may thus promote breakdown of some oil compounds in deep-sea sediments.

In an investigation of the deepwater oil plume, a combination of metagenomics, metatranscriptomics, and single-cell sequencing was used to explore the pathways for hydrocarbon degradation present in *Oceanospirillales* genomes (Mason *et al.*, 2012). This integrated approach was successful in reconstructing the near-complete microbial pathway for cyclohexane oxidation in *Oceanospirillales*, shedding light on the rapid response by these taxa during the initial deepwater dispersion of oil. In addition, metagenomic data were able to illustrate biases in the metabolism of various hydrocarbon compounds. Pelagic microbial communities in the deepwater oil plume showed an enrichment of genes involved in alkane degradation and a much lower capacity to degrade aromatic compounds, suggesting that the biodegradation of certain hydrocarbon classes is likely to be much slower in some marine ecosystems (Mason *et al.*, 2012).

Taken together, these results provide significant insights into the biological responses following the 2010 oil spill in the Gulf of Mexico. Our knowledge of oil spill impacts continues to expand, with new data on DWH impacts emerging as new studies are published.

Case study: Viral metagenomics

Metagenomic approaches have transformed our understanding of viruses in marine ecosystems, significantly advancing our knowledge of diversity, abundance, and virus-host interactions. Viral metagenomics represents another prominent case study due to the methodological challenges involved—viral fractions are difficult to work with because of low nucleic acid yields, often requiring amplification of environmental DNA (multiple displacement amplification, which may preferentially recover single-stranded DNA viruses (Mizuno *et al.*, 2013b; Yoshida *et al.*, 2013). Given the predicted diversity of viruses and the lack of any common gene families across the taxonomic classification, viral databases are extremely sparse (even compared to those for eukaryotes) and known sequences likely represent a very small fraction of global viral diversity (Hurwitz and Sullivan, 2013). Metagenomic studies routinely turn up putative viral sequences with no known homologs in public databases (Williamson *et al.*, 2012; Hurwitz and Sullivan, 2013). To circumvent some of these challenges, viral metagenomics often invokes novel methods to narrow the target

genomes being sequenced (comparing size fractionations (Williamson *et al.*, 2012); pulse field gel electrophoresis fractions (Ray *et al.*, 2012); reliance on physical linkage information to facilitate bioinformatic analyses, *via* metagenomic sequencing of individually barcoded fosmid libraries (Ghai *et al.*, 2010; Mizuno *et al.*, 2013a, b; Zhao *et al.*, 2013)). However, current environmental sequencing methods remain biased toward DNA viruses, and the diversity and role of RNA viruses in marine environments remain poorly investigated (Lang *et al.*, 2009).

Viral metagenomic studies have investigated a variety of habitats. A significant number of studies have characterized viral assemblages in ocean surface waters and throughout the water column (Breitbart, 2012; Williamson *et al.*, 2012; Hingamp *et al.*, 2013; Hurwitz and Sullivan, 2013; Labonté and Suttle, 2013; Mizuno *et al.*, 2013b), while others have focused sequencing efforts on benthic sediments (Yoshida *et al.*, 2013), hydrothermal vent plumes (Ray *et al.*, 2012), and metazoan hosts (Dunlap *et al.*, 2013). Viruses in the water column appear to exhibit a high level of local diversity, with community structure potentially driven by small-scale selective pressures; evidence, however, suggests the existence of a shared, global gene pool for marine viruses that seems to constrain overall diversity (Breitbart, 2012). Breitbart (2012) provides a comprehensive review on the current state of knowledge for marine phages in the pelagic environments (spatial patterns of abundance and diversity), including a discussion of emerging work related to the role of adaptive microbial immune systems (CRISPRs and Cas-associated genes) that govern virus-host interactions (see also Kang *et al.*, 2013). As a result of lower sampling efforts, much less is known about viral assemblages in marine sediments, although benthic communities appear to be distinct from pelagic viromes (Yoshida *et al.*, 2013). In addition to the differentiation seen in viral assemblages across habitat types, size fractionation and serial filtration of samples has revealed significant differences in taxonomy and metabolic pathways between viral communities recovered across different filter sizes (Williamson *et al.*, 2012). Such differentiation between size fractions likely reflects the partitioning of hosts during sampling, with larger fractions containing a high proportion of viruses infecting eukaryotic taxa such as phytoplankton and protists. Viruses infecting eukaryotic hosts remain less well characterized, although recent metagenomic efforts have begun to elucidate the distribution of nucleo-cytoplasmic large DNA viruses (NCLDVs) and Mimivirus gene homologs in pelagic samples as part of the Tara Oceans project (Hingamp *et al.*, 2013).

Integrating "–omic" Approaches

The use of multiple "–omic" datasets is being increasingly adopted by researchers looking to deeply characterize marine ecosystems, as high-throughput sequencing fields move more toward "systems ecology." Metagenomic datasets are useful for assessing the functional potential of a community, as determined by the presence or absence of genes encoding specific biochemical pathways. However, shotgun sequencing of environmental DNA does not relay information about the active members of the community, nor does it always produce a comprehensive list of the taxa within a sample (particularly for eukaryote-focused studies or diverse environmental communities in which metagenomic sequencing depth per genome is low). Generating parallel, complementary datasets from environmental samples is an effective approach for overcoming the inherent limitations of any given data type. For example, marker gene amplicons (typically rRNA) provide a thorough assessment of biodiversity, especially in regard to "rare biosphere" taxa (Sogin *et al.*, 2006) that may not be recovered from shotgun metagenomic sequencing. Metatranscriptomics, environmental sequencing of mRNA gene transcripts, returns a snapshot of gene expression and enables the characterization of active species and their metabolic processes (*e.g.,* Marchetti *et al.*, 2012; Sanders *et al.*, 2013). In addition, expressed transcripts can be used as a database-independent method for annotating and validating genes present in accompanying metagenomic datasets (Moran, 2009). Furthermore, the relative abundance of sequence reads can vary between metagenomic and metatranscriptomic datasets and provide a way to assess variance in transcriptional activity across taxa, particularly for dominant species in the microbial community (Shi *et al.*, 2011).

Gilbert *et al.* (2010) were the first authors to obtain multiple data types to describe microbial community assemblages, utilizing a time series of pelagic samples collected from the western English Channel. The integration of 16S rRNA amplicons, shotgun metagenomics, and metatranscriptomics facilitated a deep exploration of temporal and seasonal patterns in this marine system. Metagenomic data and 16S amplicons were most useful for assessing seasonal shifts in microbial assemblages, while metatranscriptome data appeared to better reflect diurnal rhythms seen across functional gene categories. Notably, higher diversity in amplicon datasets was also supported by higher diversity of gene transcripts in metatranscriptomes. Mason *et al.* (2012) used a similar integrated –omic approach to investigate the bacterial response to the deepwater oil plume following the *Deepwater Horizon* oil spill, as detailed previously.

Beyond the generation of parallel –omic datasets, modeling approaches can also be applied to predict patterns in the absence of data. Software tools such as PICRUST allow users to predict the functional potential of a microbial community solely on the basis of 16S rRNA amplicon data, *via* simulated metagenomic analysis based on known microbial genome sequences (Langille *et al.*, 2013). Another modeling method that has been applied in the western

English Channel is PRMT (Larsen *et al.*, 2011), which uses metagenomic data to generate putative metabolite profiles of microbial communities, helping to generate biologically testable hypotheses and identify correlations between metabolite turnover and measured environmental parameters. Artificial neural networks have additionally been used to predict microbial communities and species interactions from environmental parameters (Larsen *et al.*, 2012b). Since data generation and analysis can be a slow process, some authors have even argued for the large-scale development of computational statistical methods (general ecosystem models, Purves *et al.*, 2013) for use in generating and testing predictions about vast, undescribed ecosystems such as the deep-sea. Predictive ecosystem modeling is an emerging research area, which will no doubt benefit by the continuing expansion of environmental sequencing, an increased emphasis on detailed metadata collection, and experimental characterization of microbial metabolism (Larsen *et al.*, 2012a).

Integrating "–omic" datasets remains a relatively new concept, and thus many challenges remain. For both metatranscriptomic studies and metagenomic studies, it can be difficult to obtain sufficient sequencing depth and read coverage for diverse environmental community assemblages (needed for characterizing rare taxa and raising statistical confidence in unique, low-coverage reads), and the ubiquitous occurrence of "hypothetical protein" annotations makes it difficult to infer the function of genes with potentially important ecological relevance (Shi *et al.*, 2011). Some authors have argued for an even wider expansion of integrated –omics, advocating the systematic analysis of proteomes, metabolomes, and other measurements of phenotypic traits alongside nucleotide sequence data (Muller *et al.*, 2013).

Challenges and Future Outlook

Although metagenomic approaches are being increasingly applied to marine ecosystems, a number of challenges remain. First, the interpretation of shotgun data is still limited by public sequence databases (sequenced genomes and their corresponding annotation), and the identification and interpretation of eukaryotic data from metagenomes remains especially difficult (Gilbert and Dupont, 2011). For microbial eukaryotes in particular, the lack of consistency in taxonomic ranks and nomenclature schemes has further hampered the use of automated classification pipelines for environmental sequence data (Yilmaz *et al.*, 2013). To increase our capacity to conduct meaningful analyses, Gilbert *et al.* (2011a) make a plea for infrastructure investment in database resources and the expansion and annotation of these resources by the scientific community. Secondly, carrying out integrated, broad-scale taxonomic analyses of metagenomes currently requires disparate workflows; analyses of Bacteria, Archaea, Eukaryotes, and Viruses must be conducted in separate pipelines, despite the need to concurrently assess patterns across all taxa (Scholz *et al.*, 2012). In addition, automated pipelines themselves come with many limitations: software workflows can offer a brief overview of functional differences across samples, but manual steps and deeper analysis are almost certainly required to tease out more subtle biological patterns from metagenomic datasets (Plewniak *et al.*, 2013). Finally, future environmental sequencing studies must emphasize better experimental design and more comprehensive metadata collection (Knight *et al.*, 2012). This goal is already being realized in some projects, although replication and measurement of environmental parameters is likely to remain challenging for some inaccessible and remote marine habitats. As the price of sequencing (per base) continues to drop, future metagenomic studies will likely see the cost of bioinformatic analyses outstrip that of data generation. Nonetheless, the field of metagenomics continues to advance rapidly, and this rate of progress is steadily reflected in our deepening knowledge of marine ecosystems.

Acknowledgments

HMB gratefully acknowledges funding from the Alfred P. Sloan foundation, the North Pacific Research Board, and the National Science Foundation (Research Coordination Network award DBI-1262480; http://eukhits.wordpress.com).

Literature Cited

Allen, L. Z., E. E. Allen, J. H. Badger, J. P. McCrow, I. T. Paulsen, L. D. Elbourne, M. Thiagarajan, D. B. Rusch, K. H. Nealson, and S. J. Williamson. 2012. Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *ISME J.* **6:** 1403–1414.

Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, and H. Liu. 2006. The marine viromes of four oceanic regions. *PLoS Biol.* **4:** e368.

Baelum, J., S. Borglin, R. Chakraborty, J. L. Fortney, R. Lamendella, O. U. Mason, M. Auer, M. Zemla, M. Bill, and M. E. Conrad. 2012. Deep-sea bacteria enriched by oil and dispersant from the Deepwater Horizon spill. *Environ. Microbiol.* **14:** 2405–2416.

Bik, H. M., K. M. Halanych, J. Sharma, and W. K. Thomas. 2012a. Dramatic shifts in benthic microbial eukaryote communities following the Deepwater Horizon oil spill. *PLoS ONE* **7:** e38550.

Bik, H. M., D. L. Porazinska, S. Creer, J. G. Caporaso, R. Knight, and W. K. Thomas. 2012b. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.* **27:** 233–243.

Breitbart, M. 2012. Marine viruses: truth or dare. *Annu. Rev. Mar. Sci.* **4:** 425–448.

Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. USA* **99:** 14250–14255.

Burke, C., P. Steinberg, D. Rusch, S. Kjelleberg, and T. Thomas. 2011. Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. USA* **108:** 14288–14293.

Caron, D. A., P. D. Countway, A. C. Jones, D. Y. Kim, and A.

Schnetzer. 2012. Marine protistan diversity. *Annu. Rev. Mar. Sci.* **4:** 467–493.

Darling, A. E., G. Jospin, E. Lowe, F. A. Matsen IV, H. M. Bik, and J. A. Eisen. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2:** e243.

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

Dick, G. J., A. F. Andersson, B. J. Baker, S. L. Simmons, B. C. Thomas, A. P. Yelton, and J. F. Banfield. 2009. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10:** R85.

Dinsdale, E. A., R. A. Edwards, D. Hall, F. Angly, M. Breitbart, J. M. Brulc, M. Furlan, C. Desnues, M. Haynes, and L. Li. 2008. Functional metagenomic profiling of nine biomes. *Nature* **452:** 629–632.

Dubilier, N., C. Bergin, and C. Lott. 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6:** 725–740.

Dunlap, D. S., T. F. F. Ng, K. Rosario, J. G. Barbosa, A. M. Greco, M. Breitbart, and I. Hewson. 2013. Molecular and microscopic evidence of viruses in marine copepods. *Proc. Natl. Acad. Sci. USA* **110:** 1375–1380.

Eloe, E. A., D. W. Fadrosh, M. Novotny, L. Z. Allen, M. Kim, M.-J. Lombardo, J. Yee-Greenbaum, S. Yooseph, E. E. Allen, and R. Lasken. 2011. Going deeper: metagenome of a hadopelagic microbial community. *PLoS ONE* **6:** e20388.

Fierer, N., J. W. Leff, B. J. Adams, U. N. Nielsen, S. T. Bates, C. L. Lauber, S. Owens, J. A. Gilbert, D. H. Wall, and J. G. Caporaso. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci. USA* **109:** 21390–21395.

Fonseca, V. G., G. R. Carvalho, W. Sung, H. F. Johnson, D. M. Power, S. P. Neill, M. Packer, M. Blaxter, P. J. D. Lambshead, W. K. Thomas, and S. Creer. 2010. Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat. Commun.* **1:** 98.

Ganesh, S., D. J. Parris, E. F. DeLong, and F. J. Stewart. 2014. Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J.* **8:** 187–211.

Ghai, R., A.-B. n. Martin-Cuadrado, A. G. Molto, I. G. Heredia, R. l. Cabrera, J. Martin, M. Verdu, P. Deschamps, D. Moreira, P. Lopez-Garcia, A. Mira, and F. Rodriguez-Valera. 2010. Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J.* **4:** 1154–1166.

Gilbert, J. A., and C. L. Dupont. 2011. Microbial metagenomics: beyond the genome. *Annu. Rev. Mar. Sci.* **3:** 347–371.

Gilbert, J. A., D. Field, P. Swift, S. Thomas, D. Cummings, B. Temperton, K. Weynberg, S. Huse, M. Hughes, I. Joint, P. J. Somerfield, and M. Mühling. 2010. The taxonomic and functional diversity of microbes at a temperate coastal site: A 'multi-omic' study of seasonal and diel temporal variation. *PLoS ONE* **5:** e15545.

Gilbert, J. A., F. Meyer, and M. J. Bailey. 2011a. The future of microbial metagenomics (or is ignorance bliss?). *ISME J.* **5:** 777.

Gilbert, J. A., R. O'Dor, N. King, and T. M. Vogel. 2011b. The importance of metagenomic surveys to microbial ecology: or why Darwin would have been a metagenomic scientist. *Microb. Inform. Exp.* **1:** 5.

Hamdan, L. J., and P. A. Fulmer. 2011. Effects of COREXIT 9500A on bacteria from a beach oiled by the Deepwater Horizon spill. *Aquat. Microb. Ecol.* **63:** 101.

Hazen, T. C., E. A. Dubinsky, T. Z. DeSantis, G. L. Andersen, Y. M. Piceno, N. Singh, J. K. Jansson, A. Probst, S. E. Borglin, and J. L. Fortney. 2010. Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* **330:** 204–208.

Hingamp, P., N. Grimsley, S. G. Acinas, C. Clerissi, L. Subirana, J. Poulain, I. Ferrera, H. Sarmento, E. Villar, and G. Lima-Mendez. 2013. Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* **7:** 1678–1695.

Hunter, S., M. Corbett, H. Denise, M. Fraser, A. Gonzalez-Beltran, C. Hunter, P. Jones, R. Leinonen, C. McAnulla, and E. Maguire. 2014. EBI metagenomics—a new resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res.* **42:** D600–D606.

Hurwitz, B. L., and M. B. Sullivan. 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* **8:** e57355.

Iverson, V., R. M. Morris, C. D. Frazar, C. T. Berthiaume, R. L. Morales, and E. V. Armbrust. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* **335:** 587–590.

Jones, M. D., I. Forn, C. Gadelha, M. J. Egan, D. Bass, R. Massana, and T. A. Richards. 2011. Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474:** 200–203.

Kang, I., H.-M. Oh, D. Kang, and J.-C. Cho. 2013. Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc. Natl. Acad. Sci. USA* **110:** 12343–12348.

Kembel, S. W., J. A. Eisen, K. S. Pollard, and J. L. Green. 2011. The phylogenetic diversity of metagenomes. *PLoS ONE* **6:** e23214.

Kim, J. N., B.-S. Kim, S.-J. Kim, and C. E. Cerniglia. 2012. Effects of crude oil, dispersant, and oil-dispersant mixtures on human fecal microbiota in an in vitro culture system. *mBio* **3:** e00376–00312.

Kimes, N. E., A. V. Callaghan, D. F. Aktas, W. L. Smith, J. Sunner, B. Golding, M. Drozdowska, T. C. Hazen, J. M. Suflita, and P. J. Morris. 2013. Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front. Microbiol.* **4:** 50.

Kisand, V., A. Valente, A. Lahm, G. Tanet, and T. Lettieri. 2012. Phylogenetic and functional metagenomic profiling for assessing microbial biodiversity in environmental monitoring. *PLoS ONE* **7:** e43630.

Knight, R., J. Jansson, D. Field, N. Fierer, N. Desai, J. A. Fuhrman, P. Hugenholtz, D. van der Lelie, F. Meyer, and R. Stevens. 2012. Unlocking the potential of metagenomics through replicated experimental design. *Nat. Biotechnol.* **30:** 513–520.

Kostka, J. E., O. Prakash, W. A. Overholt, S. J. Green, G. Freyer, A. Canion, J. Delgardio, N. Norton, T. C. Hazen, and M. Huettel. 2011. Hydrocarbon-degrading bacteria and the bacterial community response in Gulf of Mexico beach sands impacted by the Deepwater Horizon oil spill. *Appl. Environ. Microbiol.* **77:** 7962–7974.

Kwan, J. C., M. S. Donia, A. W. Han, E. Hirose, M. G. Haygood, and E. W. Schmidt. 2012. Genome streamlining and chemical defense in a coral reef symbiosis. *Proc. Natl. Acad. Sci. USA* **109:** 20655–20660.

Labonté, J. M., and C. A. Suttle. 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J.* **7:** 2169–2177.

Ladau, J., T. J. Sharpton, M. M. Finucane, G. Jospin, S. W. Kembel, J. O'Dwyer, A. F. Koeppel, J. L. Green, and K. S. Pollard. 2013. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7:** 1669–1677.

Lang, A. S., M. L. Rise, A. I. Culley, and G. F. Steward. 2009. RNA viruses in the sea. *FEMS Microbiol. Rev.* **33:** 295–323.

Langille, M. G., J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. V. Thurber, and R. Knight. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31:** 814–821.

Larsen, P. E., F. R. Collart, D. Field, F. Meyer, K. P. Keegan, C. S. Henry, J. McGrath, J. Quinn, and J. A. Gilbert. 2011. Predicted

Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. *Microb. Inform. Exp.* **1:** 4.

Larsen, P., Y. Hamada, and J. Gilbert. 2012a. Modeling microbial communities: Current, developing, and future technologies for predicting microbial community interaction. *J. Biotechnol.* **160:** 17–24.

Larsen, P. E., D. Field, and J. A. Gilbert. 2012b. Predicting bacterial community assemblages using an artificial neural network approach. *Nat. Meth.* **9:** 621–625.

Li, Z.-Y., Y.-Z. Wang, L.-M. He, and H.-J. Zheng. 2014. Metabolic profiles of prokaryotic and eukaryotic communities in deep-sea sponge *Lamellomorpha* sp. indicated by metagenomics. *Sci. Rep.* **4:** 3895.

Liu, Z., and J. Liu. 2013. Evaluating bacterial community structures in oil collected from the sea surface and sediment in the northern Gulf of Mexico after the Deepwater Horizon oil spill. *MicrobiologyOpen* **2:** 492–504.

Lladser, M. E., and R. Knight. 2013. Mathematical approaches for describing microbial populations: practice and theory for extrapolation of rich environments. Pp. 85–104 in *The Human Microbiota: How Microbial Communities Affect Health and Disease*, John Wiley, Hoboken, NJ.

Lloyd, K. G., L. Schreiber, D. G. Petersen, K. U. Kjeldsen, M. A. Lever, A. D. Steen, R. Stepanauskas, M. Richter, S. Kleindienst, and S. Lenk. 2013. Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496:** 215–218.

Logares, R., S. Sunagawa, G. Salazar, F. M. Cornejo-Castillo, I. Ferrera, H. Sarmento, P. Hingamp, H. Ogata, C. Vargas, and G. Lima-Mendez. 2013. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* doi: 10.1111/1462-2920.12250.

Ma, Y., I. T. Paulsen, and B. Palenik. 2012. Analysis of two marine metagenomes reveals the diversity of plasmids in oceanic environments. *Environ. Microbiol.* **14:** 453–466.

Marchetti, A., D. M. Schruth, C. A. Durkin, M. S. Parker, R. B. Kodner, C. T. Berthiaume, R. Morales, A. E. Allen, and E. V. Armbrust. 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl. Acad. Sci. USA* **109:** E317–E325.

Marine, R., C. McCarren, V. Vorrasane, D. Nasko, E. Crowgey, S. W. Polson, and K. E. Wommack. 2014. Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* **2:** 3.

Mason, O. U., T. C. Hazen, S. Borglin, P. S. Chain, E. A. Dubinsky, J. L. Fortney, J. Han, H.-Y. N. Holman, J. Hultman, and R. Lamendella. 2012. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J.* **6:** 1715–1727.

Mason, O. U., N. M. Scott, A. Gonzalez, A. Robbins-Pianka, J. Baelum, J. Kimbrel, N. J. Bouskill, E. Prestat, S. Borglin, and D. C. Joyner. 2014. Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *ISME J.* **8:** 1464–1475.

Meyer, F., D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, and R. A. Edwards. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Mizuno, C. M., F. Rodriguez-Valera, I. Garcia-Heredia, A.-B. Martin-Cuadrado, and R. Ghai. 2013a. Reconstruction of novel cyanobacterial siphovirus genomes from Mediterranean metagenomic fosmids. *Appl. Environ. Microbiol.* **79:** 688–695.

Mizuno, C. M., F. Rodriguez-Valera, N. E. Kimes, and R. Ghai. 2013b. Expanding the marine virosphere using metagenomics. *PloS Genet.* **9:** e1003987.

Montagna, P. A., J. G. Baguley, C. Cooksey, I. Hartwell, L. J. Hyde, J. L. Hyland, R. D. Kalke, L. M. Kracker, M. Reuscher, and A. C. Rhodes. 2013. Deep-sea benthic footprint of the deepwater horizon blowout. *PLoS ONE* **8:** e70540.

Moran, M. A. 2009. Metatranscriptomics: eavesdropping on complex microbial communities. *Microbe* **4:** 329–335.

Muller, E. E., E. Glaab, P. May, N. Vlassis, and P. Wilmes. 2013. Condensing the omics fog of microbial communities. *Trends Microbiol.* **21:** 325–333.

Narasingarao, P., S. Podell, J. A. Ugalde, C. l. Brochier-Armanet, J. B. Emerson, J. J. Brocks, K. B. Heidelberg, J. F. Banfield, and E. E. Allen. 2012. De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J.* **6:** 81–93.

Plewniak, F., S. Koechler, B. Navet, E. Dugat-Bony, O. Bouchez, P. Peyret, F. Seby, F. Battaglia-Brunet, and P. N. Bertin. 2013. Metagenomic insights into microbial metabolism affecting arsenic dispersion in Mediterranean marine sediments. *Mol. Ecol.* **22:** 4870–4883.

Podell, S., J. A. Ugalde, P. Narasingarao, J. F. Banfield, K. B. Heidelberg, and E. E. Allen. 2013. Assembly-driven community genomics of a hypersaline microbial ecosystem. *PLoS ONE* **8:** e61692.

Purves, D., J. P. W. Scharlemann, M. Harfoot, T. Newbold, D. P. Tittensor, J. Hutton, and S. Emmott. 2013. Ecosystems: time to model all life on earth. *Nature* **493:** 295–297.

Ray, J., M. Dondrup, S. Modha, I. H. Steen, R.-A. Sandaa, and M. Clokie. 2012. Finding a needle in the virus metagenome haystack—micro-metagenome analysis captures a snapshot of the diversity of a bacteriophage armoire. *PLoS ONE* **7:** e34238.

Richards, T. A., M. D. Jones, G. Leonard, and D. Bass. 2012. Marine fungi: their ecology and molecular diversity. *Annu. Rev. Mar. Sci.* **4:** 495–522.

Rinke, C., P. Schwientek, A. Sczyrba, N. N. Ivanova, I. J. Anderson, J.-F. Cheng, A. Darling, S. Malfatti, B. K. Swan, and E. A. Gies. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499:** 431–437.

Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, and K. Remington. 2007. The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5:** e77.

Sanders, J., R. Beinart, F. Stewart, E. Delong, and P. Girguis. 2013. Metatranscriptomics reveal differences in *in situ* energy and nitrogen metabolism among hydrothermal vent snail symbionts. *ISME J.* **7:** 1556–1567.

Scholz, M. B., C.-C. Lo, and P. S. Chain. 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr. Opin. Biotechnol.* **23:** 9–15.

Sharon, I., and J. F. Banfield. 2013. Genomes from metagenomics. *Science* **342:** 1057–1058.

Shi, Y., G. W. Tyson, J. M. Eppley, and E. F. DeLong. 2011. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J.* **5:** 999–1013.

Simon, C., and R. Daniel. 2011. Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.* **77:** 1153–1161.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the unexplored "rare biosphere." *Proc. Natl. Acad. Sci. USA* **103:** 12115–12120.

Sul, W. J., T. A. Oliver, H. W. Ducklow, L. A. Amaral-Zettler, and M. L. Sogin. 2013. Marine bacteria exhibit a bipolar distribution. *Proc. Natl. Acad. Sci. USA* **110:** 2342–2347.

Sun, S., J. Chen, W. Li, I. Altintas, A. Lin, S. Peltier, K. Stocks, E. E.

Allen, M. Ellisman, and J. Grethe. 2011. Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Res.* **39:** D546–D551.

Swan, B. K., B. Tupper, A. Sczyrba, F. M. Lauro, M. Martinez-Garcia, J. M. Gonzalez, H. Luo, J. J. Wright, Z. C. Landry, and N. W. Hanson. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl. Acad. Sci. USA* **110:** 11463–11468.

Tang, K., K. Liu, N. Jiao, Y. Zhang, and C.-T. A. Chen. 2013. Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system. *PLoS ONE* **8:** e72958.

Thomas, T., J. Gilbert, and F. Meyer. 2012. Metagenomics—a guide from sampling to data analysis. *Microb. Inform. Exp.* **2:** 3.

Thrash, J. C., B. Temperton, B. K. Swan, Z. C. Landry, T. Woyke, E. F. DeLong, R. Stepanauskas, and S. J. Giovannoni. 2014. Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J.* **8:** 1440–1451.

Thureborn, P., D. Lundin, J. Plathan, A. M. Poole, B.-M. Sjöberg, and S. Sjöling. 2013. A metagenomics transect into the deepest point of the Baltic Sea reveals clear stratification of microbial functional capacities. *PLoS ONE* **8:** e74983.

Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37–43.

Ugalde, J. A., M. J. Gallardo, C. Belmar, P. Muõz, N. Ruiz-Tagle, S. Ferrada-Fuentes, C. Espinoza, E. E. Allen, and V. A. Gallardo. 2013. Microbial life in a fjord: metagenomic analysis of a microbial mat in Chilean Patagonia. *PLoS ONE* **8:** e71952.

Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, and W. Nelson. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Vergin, K. L., B. n. Beszteri, A. Monier, J. C. Thrash, B. Temperton, A. H. Treusch, F. Kilpert, A. Z. Worden, and S. J. Giovannoni. 2013. High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *ISME J.* **7:** 1322–1332.

Williamson, S. J., L. Z. Allen, H. A. Lorenzi, D. W. Fadrosh, D. Brami, M. Thiagarajan, J. P. McCrow, A. Tovchigrechko, S. Yooseph, and J. C. Venter. 2012. Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS ONE* **7:** e42047.

Yatsunenko, T., F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, and A. P. Anokhin. 2012. Human gut microbiome viewed across age and geography. *Nature* **486:** 222–227.

Yilmaz, P., L. W. Parfrey, P. Yarza, J. Gerken, E. Pruesse, C. Quast, T. Schweer, J. Peplies, W. Ludwig, and F. O. Glöckner. 2013. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **42:** D643–D648.

Yooseph, S., G. Sutton, D. B. Rusch, A. L. Halpern, S. J. Williamson, K. Remington, J. A. Eisen, K. B. Heidelberg, G. Manning, and W. Li. 2007. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5:** e16.

Yoshida, M., Y. Takaki, M. Eitoku, T. Nunoura, and K. Takai. 2013. Metagenomic analysis of viral communities in (hado) pelagic sediments. *PLoS ONE* **8:** e57271.

Zaneveld, J. R., L. W. Parfrey, W. Van Treuren, C. Lozupone, J. C. Clemente, D. Knights, J. Stombaugh, J. Kuczynski, and R. Knight. 2011. Combined phylogenetic and genomic approaches for the high-throughput study of microbial habitat adaptation. *Trends Microbiol.* **19:** 472–482.

Zhao, Y., B. Temperton, J. C. Thrash, M. S. Schwalbach, K. L. Vergin, Z. C. Landry, M. Ellisman, T. Deerinck, M. B. Sullivan, and S. J. Giovannoni. 2013. Abundant SAR11 viruses in the ocean. *Nature* **494:** 357–360.