Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Stabilizing the foundation of the house that 'omics builds: the evolving value of cultured isolates to marine microbiology

Michael S Rappé

The value of cultivating microbial strains that are representative of abundant microorganisms *in situ* is generally acknowledged amongst marine microbial ecologists, primarily because they provide the means to determine phenotypic properties and detailed physiological characteristics of living cells in a controlled setting. In the shadow of the rapid, ongoing expansion in environmental genomic, transcriptomic, and proteomic surveys of marine systems, a minor resurgence in experiments designed to isolate and grow free-living marine microorganisms has met some success. Interestingly, the most immediate impact that many of the resulting strains have had on our understanding of marine microbial communities has not resulted from experiments aimed to interrogate cellular physiology, but rather from their sequenced genomes. It is predicted, however, that their prolonged impact on marine ecology will result from basic laboratory research that links cellular physiology with its molecular underpinnings.

Addresses
Hawaii Institute of Marine Biology, University of Hawaii at Manoa, Kaneohe, HI 96744, USA

Corresponding author: Rappé, Michael S (rappe@hawaii.edu)

Current Opinion in Microbiology 2013, **16**:618–624

This review comes from a themed issue on **Genomics**

Edited by **Barbara A Methé** and **Iñigo Lasa**

For a complete overview see the Issue and the Editorial

Available online 29th October 2013

http://dx.doi.org/10.1016/j.mib.2013.09.009

Introduction

There is little doubt that marine microbial ecology has entered into an era of big science, ushered in by surveys, experiments and datasets of sometimes breathtaking magnitude. For example, scientists are mounting field expeditions to traverse major tracts of Earth's global ocean [1,2], observing thousands of microbial taxa ebb and flow with striking reproducibility over multi-annual temporal scales [3], sequencing the genomes from dozens of single planktonic microbial cells at once in order to illuminate their genomic characteristics and potential [4••], and monitoring the transcriptional pulse of the ocean at staggeringly high depth [5] and temporal resolution [6]. One major impetus of this work is to understand the systems biology and biogeochemical functioning of the global ocean by quantifying taxa, genes, and gene products in the context of their natural environment, and ultimately to build a predictive framework that is both relevant and consistent with forecasted global climate change. While a diverse array of technological innovations have contributed to the creation of many of these *magna opera*, a major factor has been the continued evolution of DNA sequencing cost and capacity [7••,8]. Indeed, marine microbial ecologists are riding a swell of DNA, RNA, and protein sequence data that challenges existing computational infrastructure and capabilities [7••].

Concurrent with the emergence of metagenomics, transciptomics and proteomics as mainstream approaches in marine microbial ecology has been a renewed interest in the isolation and laboratory cultivation of marine microorganisms [9,10]. While the two types of approaches are fundamentally different, their motivation originates from the same set of observations: phylogenetic marker (primarily ribosomal RNA gene) based surveys have revealed that the vast majority of planktonic, putatively heterotrophic (i.e. non-cyanobacterial) marine microbes have managed to evade laboratory domestication for much of the history of marine microbiology [11]. Thus, environmental genomics, transcriptomics and proteomics are logical and natural extensions of the ribosomal RNA approach in that they continue evading the necessity of cultivating cells in order to expand the exploration of natural communities of marine microbes from a simple phylogenetic or taxonomic assessment into functional gene, expressed transcript, and expressed protein space. Some recently emerged cultivation methodologies attempt to utilize the information gleaned from cultivation independent surveys while also incorporating new technologies, relying on the premise that at least some portion of cells in seawater will grow axenically if provided conditions for growth that mimic as closely as possible their natural environment [12–15].

While the comparatively qualitative act of isolating and growing a marine microbe in the laboratory may seem on the surface to be quite disconnected from the sometimes heavily quantitative, computationally intensive environmental sequence-based studies, this review strives to highlight the important ways in which environmental studies depend on these strains. The link most often thought of is simply a coarse association between the presence of a particular phylogenetic or taxonomic group and a functional category; 'photoheterotroph', for example. However, genome sequences directly and concretely link cultivated marine microbes with

environmental studies, providing a firm foothold from which to interrogate environmental sequence datasets. In this context, this review primarily focuses on strains resulting from an approach, termed 'high throughput culturing', for isolating and propagating cells free-living in seawater developed by Giovannoni and colleagues since it has generated a number of isolates who's genomes have significantly enhanced environmental studies. High throughput culturing relies on the dilution of cells in culture media based on natural seawater, incubation under environmental conditions that mimic the natural environment, a highly replicated experimental design, and high throughput and highly sensitive screening methodology [12]. In light of recent advances in single cell genomics and the reconstruction of whole genome sequences directly from metagenomic datasets, this review will also consider other ways it is expected that isolated marine microorganisms will provide value to environmental sequence-driven studies of the marine environment in the future.

Breadth of genome-sequenced marine bacteria from extinction cultures

Beginning with the *Candidatus* Pelagibacter ubique HTCC1062 (SAR11) genome reported in 2005 [16], 52 whole genome sequences have been generated from bacterial strains isolated by high throughput dilution to extinction and publicly released (Table 1). Over half of these were sequenced by the J. Craig Venter Institute as part of the Gordon and Betty Moore Foundation's Marine Microbial Genome Sequencing Project (http://www.moore.org/microgenome/) [17], while the others have been sequenced through a variety of smaller, individual investigator-led projects and collaborations including several via the Department of Energy Joint Genome Institute Community Sequencing Program. In addition to SAR11, whole genomes have been sequenced from extinction cultures belonging to other abundant marine bacterial lineages such as SAR116 [18,19], the *Rhodobacteraceae* or marine Roseobacter clade [20–23], OM60/NOR5 [24–26], SAR92 [27], and marine *Methylophilales* or OM43 [28,29] (Table 1). The list also includes the genome of the first cultivated representative from the bacterial phylum *Lentisphaerae* [30].

With regard to ocean ecology, a certain measure of success for the high throughput culturing approach is that most of the genomes listed in Table 1 originate from phylogenetic lineages that contain no other isolated or genome sequenced representatives, despite potentially high abundance in seawater. Many of these genomes are currently serving as useful anchors for environmental sequence based studies. However, it is important to note that it is a tenuous assumption that features such as genome architecture and functional gene complement will be ubiquitously and uniformly distributed throughout the clade of interest. This is based on the recurring observation that many of these lineages emerge as highly diverse clusters of related sequences in ribosomal RNA gene-based studies of seawater, among others. However, as additional isolated strains have become available their whole genome sequences have been used to investigate the genomic features common to all members of a particular clade versus those features specific to a particular subclade or lineage. To highlight one prominent example, genome sequences are now available for 15 isolates of the SAR11 clade that span a broad range of phylogenetic depth. A subset have been used to investigate the evolutionary history of the SAR11 lineage within the *Alphaproteobacteria* using phylogenomics and other approaches, spurred on by the potential for shared ancestry between SAR11 and mitochondria [31–35]. The same subset was also used to identify shared versus variable characteristics of the major SAR11 sub-lineages in a detailed comparative genomics approach [36•].

Reference genomes from seawater extinction cultures have strongly influenced our interpretation of environmental sequence datasets

Generally speaking, two types of output are frequently derived from environmental sequence datasets: taxonomic identification of the environmental sequence, and biochemical function if one has been assigned to a gene with recognizable homology in a reference database. One method that has been successfully used to assess the prevalence of a particular microorganism in an environmental sequence dataset involves employing its whole genome sequence as a template in a similarity-based search, and then counting the number of environmental sequence reads that 'hit' the genome sequence above some predetermined threshold cutoff value. This type of binning has been referred to previously as fragment recruitment [2], and has been used to quantify the importance of reference genomes in environmental genomic and transcriptomic datasets (Figure 1). Several recent studies have used subsets of the whole genome sequences derived from high throughput extinction culturing listed in Table 1 to recruit sequencing reads from seawater, revealing these genomes to dominate the recruited reads, particularly for the non-cyanobacterial fraction of the microbial community. In one comprehensive study that employed many of the genome sequences listed in Table 1 to recruit against the Sorcerer II Global Ocean Sampling (GOS) expedition metagenomic dataset [2], the three most highly recruiting genomes were from strains of SAR11 isolated by high throughput extinction culturing and, of the 34 most highly recruiting genomes, 12 of the 21 non-cyanobacterial representatives were also obtained by this method of culturing [17]. In a separate study, deep metatranscriptomic sequencing was used to characterize protein-coding gene expression patterns in coastal seawater of the southeastern United States, with the goal of investigating niche diversification between