Towards a Global Perspective of the Marine Microbiome

8

Silvia G. Acinas, Marta Sebastián, and Isabel Ferrera

Abstract

Marine microbes play fundamental roles in nutrient cycling and climate regulation at a planetary scale. The field of marine microbial ecology has experienced major breakthroughs following the application of high-throughput sequencing and culture-independent methodologies that have pushed the exploration of the marine microbiome to an unprecedented scale. This chapter overviews how the advances in gene- and genome-centric approaches as well as in culturing and single cell physiological methodologies in conjunction with global oceanographic circumnavigation expeditions and long-term time series are fueling our understanding of the biogeography, temporal dynamics, functional diversity, and evolutionary processes of microbial populations. We discuss how the joint effort of all those integrative approaches will help to boost our knowledge of the marine microbiome to reach a predictive understanding of how it is going to evolve in future scenarios.

S. G. Acinas (✉) · M. Sebastián
Department of Marine Biology and Oceanography, Institut de Ciències del Mar (ICM-CSIC), Barcelona, Spain
e-mail: sacinas@icm.csic.es

I. Ferrera
Centro Oceanográfico de Málaga, Instituto Español de Oceanografía, Málaga, Spain

8.1   Marine Microbial Ecology: Opening the Black Box

8.1.1   Major Breakthroughs before the -Omics Revolution

As Sydney Brenner, Nobel Prize in Physiology or Medicine 2002 said, "*Progress in science depends on new techniques, new discoveries and new ideas, probably in that order.*" The field of marine microbial ecology has been no exception to this quote as it has seen major breakthroughs in the last 60 years following the application of new technologies. For example, the use of epifluorescence microscopy for the estimation of bacterial abundance unveiled that traditional plate counting methods were underestimating the real values by several orders of magnitude (Jannasch and Jones 1959) because most bacterial species do not form colonies on solid media (what was later coined as the "Great Plate Count Anomaly", Staley and Konopka 1985). Likewise, epifluorescence microscopy was also crucial to uncover the presence of abundant and ubiquitous unicellular cyanobacteria in the surface ocean (Johnson and Sieburth 1979). Initial studies using radioisotopes to estimate respiration and primary production also revealed that much of the respiration, dissolved organic matter (DOM) processing, and primary production were carried out by picoplanktonic organisms (Li et al. 1983; Williams 1970, 1981). The application of techniques commonly used in the biomedical field like flow cytometry also constituted a tipping point in our understanding of marine microbial ecology. For example, it allowed the discovery of *Prochlorococcus*, the most abundant primary producer in the ocean (Chisholm et al. 1988), and enabled the delineation of two populations of heterotrophic bacteria based on their nucleic acid content (Li et al. 1995; Robertson and Button 1989). This separation in the DNA content was initially attributed to a difference in the physiological state with the high nucleic acid (HNA) cells representing the active fraction of the community (Lebaron et al. 2001; Servais et al. 2003). However, this view was later challenged when it was shown that low nucleic acid (LNA) cells were also important contributors to bacterial metabolism in the sea (Longnecker et al. 2005; Sherr et al. 2006), and that the nucleic acid content may partially reflect the genome size of the cells (Vila-Costa et al. 2012). The addition of flow cytometry to the microbial ecologist's set of tools fueled a series of studies combining different stains probing cellular activity or growth (see Del Giorgio and Gasol 2008 for a review on this topic). These studies were crucial to open the black box of microbial communities unveiling the large heterogeneity in the physiological status of cells in the environment, which has an impact on their contribution to ecosystem function as we will expand on below.

The application of molecular tools like the 16S rRNA gene PCR amplification and sequencing from environmental DNA enabled access to the diversity of the "uncultured majority" (Amann et al. 1995; Rappé and Giovannoni 2003), which profoundly revolutionized the field of marine microbial ecology. These techniques opened a new venue to understand and classify marine microbial diversity and resulted in, among others, the discovery of the most abundant bacteria in the ocean, the SAR11 clade (Giovannoni et al. 1990), elusive for culturing using traditional methods. The rRNA approach also fueled the implementation of

fluorescent in situ hybridization (FISH) methods to visualize and enumerate specific phylogenetic groups in natural samples. The combination of FISH with radioisotopic methods such as microautoradiography (MAR) was pivotal in unraveling niche partitioning in regard to the use of organic substrates between the different components of bacterial communities (Alonso-Sáez et al. 2012a; Cottrell and Kirchman 2000). Likewise, the estimation of the net growth rates of microorganisms belonging to various phylogenetic groups by using FISH has provided evidence for the important role that numerically unremarkable microbes can play in ecosystem functioning (Kirchman 2016). The advent of high-throughput sequencing approaches represented another revolution in microbial ecology that shed light on the composition of marine microbial communities, the relative abundance of their components, and the discovery of the "rare biosphere" (Sogin et al. 2006). The latter comprises low abundant taxa that act as reservoir for most phylogenetic and functional diversity (Pedrós-Alió 2012).

8.1.2 It Is Not Always Black and White: The Discovery of Photoheterotrophs

One of the most important findings in microbial oceanography in the last decades was the discovery of photoheterotrophs as an important component of the bacterioplankton. This discovery challenged the classic portrayal of bacterioplankton composed of photoautotrophic microorganisms as primary producers and of chemoheterotrophic microorganisms as consumers. Early genomic analyses of natural occurring marine bacterioplankton reported that an uncultured bacterium harbored a gene coding for proteorhodopsin (PR), a light-dependent proton pump, unveiling a new type of phototrophy in the ocean (Béjà et al. 2000). That same year, using infrared fluorometry Kolber et al. (2000) detected high signals of bacteriochlorophyll a (BChla) in the surface oligotrophic ocean suggesting that aerobic anoxygenic phototrophic (AAP) bacteria are a substantial component of the marine microbiome. These two reports represented the beginning of a change of paradigm in the field of marine microbial ecology that demanded to consider the direct effects of light on heterotrophic processes and, consequently, to rethink the models of organic carbon fluxes in the ocean.

Metagenomics and other molecular approaches have revealed an unsuspected large diversity among PR and AAP-containing bacteria and have shown that the genes responsible for photoheterotrophy are widespread among the most abundant microbial taxa in the surface ocean (DeLong and Béjà 2010; Koblížek 2015). Proton-pumping rhodopsins are found in marine Proteobacteria (including the SAR11 clade), Bacteroidetes, and Euryarchaeota (Pinhassi et al. 2016). These rhodopsins consist of only one opsin protein with a covalently bound pigment (retinal) and this simple structure has favored their lateral gene transfer and spreading among distant taxa, even across the domains Bacteria and Archaea (Frigaard et al. 2006). Contrarily, the machinery for light harvesting and energy synthesis in AAP bacteria consists of several pigments and proteins. Thus, even though the genes

involved in aerobic anoxygenic photosynthesis have also been laterally transferred in marine bacteria, these events of gene transfer has been constrained to the Proteobacteria, mostly to the Alpha- and Gammaproteobacteria. Besides the size of coding operons, rhodopsins and bacteriochlorophyll systems display differences in the cost of biosynthesis and benefits of phototrophy (Kirchman and Hanson 2013). While the broad occurrence of marine photoheterotrophs has been well described, less is known about their physiology and ecology. Some axenic cultures of AAP and PR-containing bacteria have shown that they can use light to grow more efficiently (i.e., Gómez-Consarnau et al. 2007; Hauruseu and Koblížek 2012) and this has also been confirmed for natural populations of AAPs (Ferrera et al. 2017). In contrast, other PR-containing isolates do not grow better with light (González et al. 2008), and it has been reported that proteorhodopsins can also promote survival during starvation (Gómez-Consarnau et al. 2010). In addition to proton-pumping rhodopsins, other types of rhodopsins have been discovered in marine bacteria and archaea including chloride- and sodium-pumping rhodopsins (see Pinhassi et al. 2016), xanthorhodopsins (proton pumps associated with a particular carotenoid molecule named salinixanthin, Balashov et al. 2005), or sensory rhodopsins such as the newly discovered heliorhodopsins (Pushkarev et al. 2018). Further, rhodopsins, including heliorhodopsins, have been reported in marine viruses and single-celled eukaryotes (Needham et al. 2019). In short, following the discovery of photoheterotrophs it seems now clear that, in the marine environment, photoheterotrophy is not just an exception but probably the rule. In fact, there is growing evidence that bacteriochlorophyll- and rhodopsin-based dual phototrophy may have evolved in nature and is awaiting discovery (Zeng et al. 2020).

8.1.3   Are all Microorganisms Equally Active in the Ocean?

When single cell approaches started to be used in the late 70s (e.g., Hoppe 1976) it became evident that the activity of marine microbes is not homogeneous but on the contrary is tremendously heterogeneous. Within a given microbial community cells can be dead, injured, dormant (i.e., in a non-growth state), active but constrained by the availability of a certain resource, slow- or fast growing. Any of these physiological states will have strong implications for the role that such cells play in the environment. Yet, one of the major challenges that microbiologists face is how to define the physiological state of a microorganism.

    A variety of methods has been developed in order to differentiate active- from non-active cells, probing cellular division, membrane integrity, respiratory activity, substrate uptake, or protein synthesis (see Del Giorgio and Gasol 2008; Sebastián and Gasol 2019; Singer et al. 2017 for reviews on this topic). These methods target different processes and differ in resolution, and thus the delineation of active cells is not always consistent (see Del Giorgio and Gasol 2008 for a comparison of methods). Therefore, the categorization within the various physiological states is purely operational. For example, using the tetrazolium salt 5-cyano-2,3-ditolyl tetrazolium chloride (CTC) as an indicator of activity Del Giorgio and Scarborough

(1995) found that the percentage of active cells represented only 5% of the cells in the oligotrophic ocean, which is consistent with the microautoradiography (MAR) values that were obtained using ^{3}H-thymidine as substrate (Longnecker et al. 2010). However, when these authors used ^{33}P-phosphate instead of thymidine the percentage of active cells increased up to 50%. This indicates that the categorization of active cells using substrate uptake methods like MAR strongly relies on the substrate of choice. Despite the variability found in the proportion of active cells, these single cell techniques have been highly informative, particularly when coupled with other approaches that allow taxonomic characterization such as fluorescence in situ hybridization (MAR-FISH). These techniques have shown that some members of the community are more active than others and that there is heterogeneity in the levels of activity even within a given population (e.g., Alonso-Sáez et al. 2007, 2012a; Cottrell and Kirchman 2000).

The delineation of dormant cells is also important because it has been hypothesized that these cells constitute a seed bank of taxonomic and functional diversity that guarantees the persistence and long-term maintenance of the diversity and function of the community (Lennon and Jones 2011). Yet, we still lack a method to assess whether or not a cell is dormant. Some attempts have been made to characterize the active and inactive members of bacterial and archaeal communities through the joint analysis of ribosomal RNA and DNA of individual taxa, using RNA:DNA ratios below 1 as a threshold to delineate "inactivity" or dormancy (Bowsher et al. 2019; Campbell et al. 2011; Jones and Lennon 2010; Kearns et al. 2016). However, the delineation of the inactive members of a community is sometimes problematic (Steven et al. 2017) because some taxa accumulate ribosomes during dormancy in order to be able to respond swiftly to favorable conditions (see Blazewicz et al. 2013 for a review on this topic). Despite these caveats, the sequencing of ribosomal RNA has provided valuable information about marine taxa that have or lack the potential for protein synthesis and how this changes over space or time (Campbell and Kirchman 2013; Campbell et al. 2009, 2011; Ghiglione et al. 2009; Hugoni et al. 2013; Hunt et al. 2013; Zhang et al. 2014).

Labeling with the thymidine substitute 5-bromo-2′-deoxyuridine (BrdU) has also been used to define the active members of the community (Galand et al. 2013; Pernthaler et al. 2002). This and other approaches that link activity and identity, such as stable isotope probing, have been utilized to identify the major players in a certain biogeochemical process or follow the uptake of specific compounds (Bryson et al. 2017; Mou et al. 2008; Nelson and Carlson 2012; Orsi et al. 2016; Taubert et al. 2017). Other methods that are used in marine microbial studies but are still in their infancy include Bioorthogonal non-canonical amino acid tagging (BONCAT) (Couradeau et al. 2019; Hatzenpichler et al. 2016; Leizeaga et al. 2017; Samo et al. 2014) and Raman micro-spectroscopy (Berry et al. 2015; Huang et al. 2007; Lorenz et al. 2017). These methods could have a tremendous potential to address the physiological status of individual cells as will be expanded on below.

Altogether these studies have shed light on microbial processes in the ocean and identified the major players driving them. Yet, accurate knowledge is lacking of how many cells are active, how dynamic is the transition between activity and inactivity

of individual cells, and what are the factors driving these transitions. Furthermore, most studies are temporally and geographically restricted and are particularly focused on the sunlit ocean. When inexpensive techniques like BONCAT will be broadly applied by the research community these gaps of knowledge may be filled. This is crucial to understand microbial function and to predict how it is going to evolve in future scenarios.

8.2   The Marine Microbiome over Space and Time

8.2.1   The Beginning of the Global Exploration of the Marine Microbiome

Microbial biogeography relies on the description of how microbial communities are distributed in space, in the vertical as well as in the horizontal dimension. While multiple studies that were restricted to particular biogeographical areas provided hints on the spatial distribution of marine microorganisms (e.g., Agogué et al. 2011; DeLong et al. 2006; Pham et al. 2008; Pommier et al. 2010) it was not until the global oceanographic circumnavigations took place (Fig. 8.1) when the exploration of the worldwide distribution of marine microorganisms became feasible. The pioneering large-scale survey was the Global Ocean Sampling Expedition (GOS) that was launched in 2003 in the Sargasso Sea. This survey continued as a several years expedition across the globe although most of the data were generated between 2004 and 2006 from a transect running from the North Atlantic to the South Pacific through the Panama Canal. The GOS expedition indeed represents the first approximation to the microbial diversity of the global surface ocean (Rusch et al. 2007; Venter et al. 2004). The GOS unveiled ~1300 different 16S rRNA gene sequences in surface seawater samples from the Sargasso Sea (Venter et al. 2004) and thousands of gene families in the surface ocean (Rusch et al. 2007). Contemporary to the GOS, the
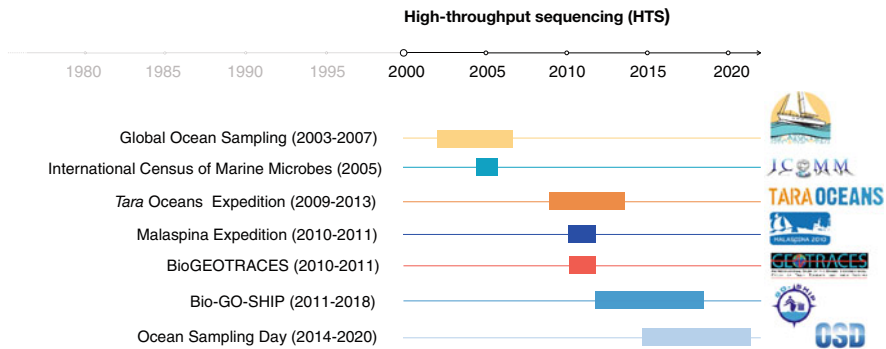

Fig. 8.1 The most important global marine circumnavigations for the exploration of marine microbiomes and when they took place. Figure courtesy of Dr. Marta Royo-Llonch from SHOOK Studio

International Census of Marine Microbes (ICoMM) was launched in 2005 aiming at using standardized sampling and data analysis procedures to study microbial diversity in a multitude of marine habitats including pelagic and benthic systems using 454-pyrosequencing of the 16S rRNA gene (Amaral-Zettler et al. 2010). At that time, the ICoMM represented the most comprehensive diversity picture of global ocean bacterial and archaeal communities and reported remarkable horizontal and vertical large-scale segregation showing contrasting patterns for the different explored habitats (pelagic, benthic, and anoxic ecosystems, or vents, Zinger et al. 2011). This study showed that alike yet remote habitats can harbor similar communities in the pelagic realm supporting the Baas-Becking and Beijerinck hypothesis that "*everything is everywhere, but the environment selects*" (Baas Becking 1934; Beijerinck 1913). Hence, marine planktonic bacteria display an unlimited potential for dispersal while abiotic environmental filtering is responsible for their different distributions in the global ocean. Furthermore, it was the ICoMM initiative that allowed the discovery of the abovementioned "rare biosphere" (Pedrós-Alió 2012; Sogin et al. 2006).

After the GOS and the ICoMM initiatives the main surveys with focus on the worldwide exploration of the marine microbiome have been the *Tara* Oceans Expedition (2009–2013) (Karsenti et al. 2011), the Malaspina 2010 Expedition (Duarte 2015), the Ocean Sampling Day, which is a simultaneous sampling campaign of the world's coasts on the summer solstice that has been carried out since 2014 (Kopf et al. 2015; Tragin and Vaulot 2018), the BioGEOTRACES (Biller et al. 2018), the Bio-GO-SHIP programs (Larkin et al. 2021) (Fig. 8.1), and other initiatives that are in the planning phase. Between 2009 and 2013, the *Tara* Oceans Expedition sampled the global ocean from surface waters to the mesopelagic layer. It used standardized sampling procedures to obtain seven size fractions of planktonic diversity, from viruses to small metazoans, and comprised a large sequencing effort (>30 Gb/sample) (Pesant et al. 2015). One of the main legacies of this expedition was the generation of the Ocean Microbial Reference Gene Catalog (OMR-GC) containing a total of >40 million-non-redundant genes of the global marine microbiome of which over 80% of the sequences were new (Sunagawa et al. 2015). This catalog was updated with the OMR-GCv2 integrating the metagenomes of the Arctic region and the metatranscriptomes of the global ocean (Salazar et al. 2019). The multidisciplinary and extensive effort of the *Tara* Oceans consortium has resulted in many important outcomes (see Sunagawa et al. 2020 for a review) such as the genetic repertoire of bacteria and archaea (Sunagawa et al. 2015) including more than 500 bacterial and archaeal metagenome assembled genomes (MAGs) from the polar Arctic Ocean (Royo-Llonch et al. 2021) and eukaryotes (Carradec et al. 2018) in the sunlit and mesopelagic global ocean, the diversity of eukaryotic plankton (De Vargas et al. 2015) and viruses (Brum et al. 2015; Gregory et al. 2019), and the planktonic interactions occurring in the photic ocean (Lima-Mendez et al. 2015). Other relevant achievements from the *Tara* Oceans Expedition were studies on the potential contribution of unexpected components such as bacteria, archaea, and viruses to carbon export in the nutrient-depleted oligotrophic ocean (Guidi et al. 2016), the potential impact of ocean warming on community composition and gene

expression (Salazar et al. 2019), or the discovery of latitudinal gradients of diversity for most planktonic groups (Ibarbalz et al. 2019). Furthermore, data from the *Tara* Oceans Expedition confirmed that the geographic distance only plays a subordinate role in determining the taxonomic and functional microbial community composition in the photic open ocean whereas environmental selection seems to be an important driver of microbial biogeography (Sunagawa et al. 2015).

Between 2010 and 2011, the Malaspina 2010 Circumnavigation Expedition (Duarte 2015) sampled the marine microbiome in the tropical and subtropical oceans from the surface down to bathypelagic waters (~4000 m depth). This expedition showed that shifts towards communities enriched in rare taxa in the sunlit ocean reflect environmental transitions (Ruiz-González et al. 2019) and it explored the role of dispersion on planktonic and micro-nektonic organisms (Villarino et al. 2018). The Malaspina Expedition also contributed with an assessment of the diversity and biogeography of deep-sea pelagic bacteria and archaea (Salazar et al. 2016) as well as provided an account of the diversity of heterotrophic protists in the deep ocean, particularly unveiling the special relevance of fungal taxa (Pernice et al. 2015). This expedition also unraveled that the particle-association lifestyle is a phylogenetically conserved trait in bathypelagic microorganisms (Salazar et al. 2015), and provided a metabolic characterization of the deep ocean microbiome based on Metagenome Assembled Genomes (Acinas et al. 2021). Nonetheless, the Malaspina -omics datasets have not yet been fully exploited and new inputs are expected in the coming years, such as those derived from ongoing detailed analyses of vertical profiles and of specific water masses or insights in the microbiome associated with the deep scattering layer in the ocean.

The datasets of the BioGEOTRACES and Bio-GO-SHIP programs have recently been made available (Biller et al. 2018; Larkin et al. 2021) and these will surely push further our knowledge of the marine microbiome. The BioGEOTRACES initiative spans 610 metagenomes collected from diverse regions of the Pacific and Atlantic oceans (Biller et al. 2018). It also adds a temporal dimension to the marine microbiome by providing metagenomes collected every month during two years at the stations HOT (North Pacific) and BATS (North Atlantic), which are part of long-term time series programs (Karl and Church 2014; Steinberg et al. 2001) and for which a suite of physicochemical and biological data are available. The Bio-GO-SHIP program aims at providing high-resolution spatiotemporal sampling of the marine microbiome in order to link microbial traits with ecosystem function and biochemical fluxes. This program has released 720 globally distributed surface ocean metagenomes from samples collected every 4–6 h representing a median distance between sampling stations of only 26 km (Larkin et al. 2021) and thus providing a much higher spatial resolution than other global expeditions such as *Tara* Oceans (~700 km) or bioGEOTRACES (~200 km).

Overall, these global expeditions together with other local and regional studies have unveiled some general microbial patterns across horizontal and vertical scales in the ocean. These general trends include among others: (i) the key concept that microbial communities are formed by a few abundant taxa and a long tail of low abundant taxa (the rare biosphere) (Sogin et al. 2006), (ii) that temperature is one of

the main drivers to predict the taxonomic and functional gene composition of microbial communities in epipelagic waters of the open ocean (Sunagawa et al. 2015), (iii) that phosphorus exerts a strong selective pressure in the surface ocean (Coleman and Chisholm 2010; Grote et al. 2012), (iv) that the latitudinal gradient of diversity observed in macroorganisms, meaning that species richness increases in latitudes closer to the equator and mid-latitudes, also applies to most marine planktonic microorganisms (Amend et al. 2013; Ibarbalz et al. 2019; Sul et al. 2013), (v) that the microbial community composition of contrasting polar zones, even though being distant from each other, is more similar than that of microbial communities of temperate or tropical latitudes (Cao et al. 2020; Ghiglione et al. 2012; Royo-Llonch et al. 2021; Sul et al. 2013), (vi) a vertical segregation between photic and aphotic microbial communities (Amaral-Zettler et al. 2010; DeLong et al. 2006; Sunagawa et al. 2015), and (vii) the existence of a vertical connectivity between surface and deep ocean communities (Cram et al. 2015; Mestre et al. 2018; Parada and Fuhrman 2017; Ruiz-González et al. 2020).

Despite the relevant information generated by all these studies, the myriad of microbial processes ocurring in the ocean are far from understood. For instance, there is an increasing recognition of the role of sub-mesoscale hydrographic features in key processes such as the carbon pump (Boyd et al. 2019; Resplandy et al. 2019) or the dispersion of marine microbes, pollutants, and microplastics, but studies on the changes of the marine microbiome at fine spatial scales are still missing. Current initiatives looking at this heterogeneity like the EXPORTS program (https://oceanexports.org/about.html) and further implementation of genomic sensors (Scholin et al. 2017) will help to address this issue. Similarly, the aim of the recently launched AtlantECO EU project (https://www.atlanteco.eu) is a better understanding and integration of the marine microbiome in the context of ocean circulation and the presence of pollutants, e.g. plastics. AtlantECO also pursues to assess the role of the marine microbiome in driving the dynamics of the Atlantic ecosystem at basin and regional scales. The joint effort of all those initiatives will help to boost our knowledge of the marine microbiome.

8.2.2 Seasonality and Temporal Dynamics of Marine Microbial Communities

Besides studies aiming at understanding how microbial communities vary across space, increasing efforts are being invested towards exploring how they change over time. The establishment of microbial observatories across the globe has allowed the monitoring of microbial communities over time from short- to long-term scales (see reviews by Bunse and Pinhassi 2017; Buttigieg et al. 2018). Defining seasonality is essential to understand how microbes react to changes in environmental conditions or perturbations. Time series of marine microbiomes also allow addressing fundamental ecological questions such as which patterns of biodiversity are present in an ecosystem, how are these patterns governed, how stable and predictable are microbial communities, how do species interact or what is the ecological niche of a given

taxon. Seasonality had been observed in phytoplankton blooms but it was only after the molecular revolution that it could be investigated in bacterioplankton communities. The pioneer application of fingerprinting methods and clone libraries to samples collected from marine microbial observatories over 1–2 year periods elucidated community shifts over seasons and demonstrated the existence of temporal niches for specific organisms (Alonso-Sáez et al. 2007; Brown et al. 2005; Ghiglione et al. 2005). Nevertheless, time series of many consecutive years are required to test the robustness of seasonal patterns. Such long-term time series with large sampling efforts have been undertaken in oceanic and coastal monitoring stations; the San Pedro Ocean Time Series (SPOT) and the Hawaii Ocean Time Series (HOT) in the Pacific Ocean, the Bermuda Atlantic Time Series (BATS) in the Atlantic Ocean, the Western Channel Observatory in the English Channel, or the Service d'Observation du Laboratoire Arago (SOLA Station; Banyuls-sur-Mer, France) and the Blanes Bay Microbial Observatory (BBMO) in the Mediterranean Sea are some examples of such long-term programs (for a detailed list, see Bunse and Pinhassi 2017; Buttigieg et al. 2018).

Over a decade ago the application of high-resolution molecular fingerprinting methods over monthly samples from SPOT revealed remarkably repeatable and predictable seasonal patterns in the distribution and abundance of microbial taxa (Fuhrman et al. 2006). High-throughput sequencing of bacterioplankton communities confirmed this observation at higher resolution in different locations (Cram et al. 2015; Eiler et al. 2011; Fuhrman et al. 2015; Gilbert et al. 2012; Lambert et al. 2019) and unveiled that also the rare members of the bacterioplankton showed seasonality (Alonso-Sáez et al. 2015). Moreover, long-term time series captured the occasional blooming of some rare members of the community which became dominant when the conditions were favorable (Gilbert et al. 2012). Likewise, microbial eukaryotes showed recurrent seasonal patterns (Giner et al. 2019; Lambert et al. 2019). In addition, changes in community composition were accompanied by repeatable shifts in alphadiversity (Gilbert et al. 2012; Giner et al. 2019). Besides, the seasonal patterns of relevant phylogenetic groups (Díez-Vives et al. 2019; Salter et al. 2015; Vergin et al. 2013) or of certain functional groups (AAPs, Auladell et al. 2019) have been studied, revealing remarkably similar patterns to those of the whole communities (Fig. 8.2).

Current methodologies allow to investigate the dynamics of individual taxa, unveiling that closely related populations can represent distinct ecotypes that temporally occupy different niches (Auladell et al. 2019, 2021; Chafee et al. 2018). Further, analysis of finely resolved taxonomic units in combination with high frequency sampling over multiple years has shown that regardless of interannual variation in phytoplankton blooms recurrent modules of co-varying microbes exist (Chafee et al. 2018). Beyond long-term studies, high frequency sampling over a phytoplankton bloom has shown that biological interactions among bacteria, archaea, and eukaryotic microorganisms may play critical roles in controlling plankton diversity and dynamics (Needham and Fuhrman 2016), contradicting the traditional view that blooms are mainly controlled by physical- and chemical processes. Daily sampling over periods of several months has also revealed that
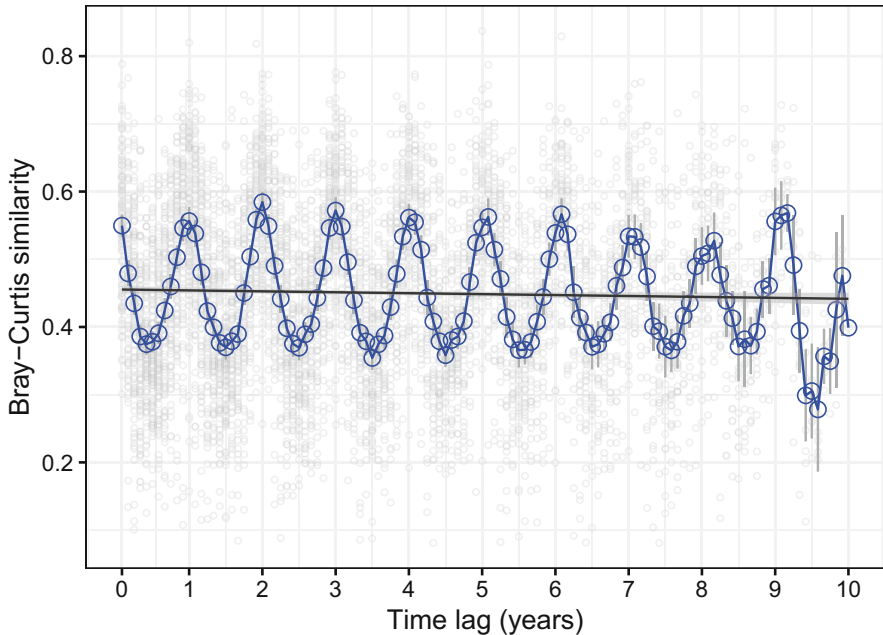
Fig. 8.2 Time-decay plot showing the recurrence of aerobic anoxygenic phototrophic communities in the Blanes Bay Microbial Observatory (NW Mediterranean). Bray–Curtis similarity between samples is plotted against the time lag between each comparison. Blue dots represent mean values for each time lag and gray vertical bars the standard error (background gray dots show each comparison). A linear regression with 95% confidence intervals is shown (modified from Auladell et al. 2019)

microbial plankton is organized in clearly defined but ephemeral communities whose turnover is rapid, mirroring environmental variability (Martin-Platero et al. 2018). On shorter time scales, metatranscriptomics has shown evidence of diel transcriptional oscillations of both phototrophic and chemotrophic microorganisms (Ottesen et al. 2013, 2014) as well as of viruses (Aylward et al. 2017; Kolody et al. 2019).

   Altogether, long time series have provided evidence for seasonal and interannual recurrence of some microbial taxa, highly resolved time series have demonstrated that communities fluctuate on a daily and monthly scale along with changes in environmental conditions, and sampling on an hourly scale has unveiled diel periodicity of gene transcription. These studies demonstrate that the temporal scale of sampling is directly linked to the scale of temporal variability that we are able to capture. Highly resolved sampling over long periods will be eased by the development of automated samplers. This, in combination with the decreasing analytical and computing costs, will provide further insights into the stability and reproducibility of the short-term changes over longer periods of time. Moreover, increasing efforts on obtaining data from high and low latitudes will help defining a more global picture of

the temporal dynamics of marine microorganisms. All these data will be useful to feed model-based analyses aimed at better predicting the temporal dynamics of marine microbial communities in future scenarios.

8.3 Approaches to Link Taxonomy and Function of Marine Bacteria and Archaea

8.3.1 The Genome-Centric Approaches: Single Amplified Genomes (SAGs) and Metagenome Assembled Genomes (MAGs)

Whereas direct analysis of metagenomes can provide a community overview, other strategies have been developed to either access individual environmental genomes without the need for cultivation (Single Amplified Genomes, SAGs) or group the community's metagenomic information into meaningful genomic units reflective of a population of close taxa as in Metagenome Assembled Genomes (MAGs). Moreover, properly assigning function to taxonomy, which has been an essential goal in the microbial ecology of uncultured microorganisms, requires the genetic information to be considered in a genomic context.

8.3.1.1 Single-Amplified Genomes (SAGs)

SAGs are generated from the direct amplification of DNA from previously sorted individual cells, its sequencing and assembly (Fig. 8.3). SAGs may represent
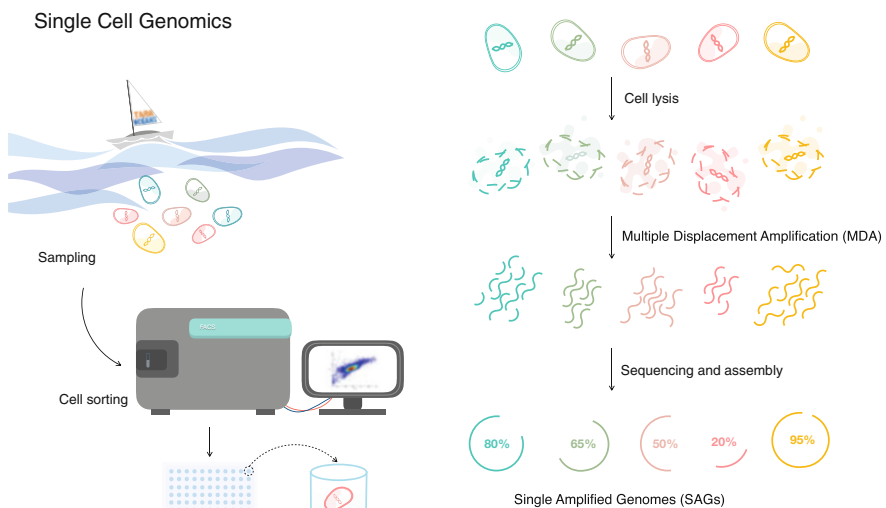

Fig. 8.3 Simplified workflow for SAGs generation from seawater samples. After sample collection cells are sorted in a flow cytometer. Individual cells are then lyzed and nucleic acids undergo Multiple Displacement Amplification (MDA). After that, genomes are sequenced and assembled resulting in SAGs of variable quality, from low to high coverage. Figure courtesy of Dr. Marta Royo-Llonch from SHOOK Studio (https://www.instagram.com/shookstudio/?hl=en)

environmental genomes sequenced from the most fundamental units of life (Blainey 2013; Stepanauskas 2012; Sieracki 2007; Woyke et al. 2009). Individual cells are selected from a sample by fluorescence-activated cell sorting (FACS) or microfluidics. Subsequently, the cells undergo lysis and whole genome amplification through various methodologies: degenerate oligonucleotide-primed PCR, multiple displacement amplification (MDA), or WGA-X, an MDA method that utilizes a thermostable mutant of the phi29 polymerase (Stepanauskas et al. 2017). This is followed by sequencing and genome assembly. SAGs retrieves all the DNA molecules of a cell, unveiling microbial intimate interactions in their natural environment otherwise overlooked like infections, symbioses, and predation (Castillo et al. 2019; Labonté et al. 2015; Martinez-Garcia et al. 2014; Roux et al. 2014; Yoon et al. 2011). Because SAGs circumvents the taxonomic binning used in metagenome assembly, it improves the understanding of microevolutionary processes in the environment (Kashtan et al. 2014), providing as well unique reference genome datasets from uncultured microbes. Indeed, the analysis of 2715 partial SAGs from the tropical and subtropical euphotic ocean enabled the functional and taxonomic annotation of about 80% of metagenomic reads from diverse oceanographic cruises and marine stations (Pachiadaki et al. 2019).

Some of the valuable lessons learnt from SAGs studies include: (i) genome streamlining is a prevalent feature in the oligotrophic surface ocean (Swan et al. 2013), (ii) the extensive microdiversity and co-existence of hundreds of genomes within *Prochlorococcus* populations (Kashtan et al. 2014), (iii) the reconstruction of novel uncultured bacterial species (Royo-Llonch et al. 2020), (iv) the ubiquity of light harvesting and secondary metabolite biosynthetic pathways across microbial lineages (Pachiadaki et al. 2019), (v) the impact of chemolithoautotrophic microorganisms such as the SAR324 clade and the Gammaproteobacteria ARCTIC96BD-19 (Swan et al. 2011), (vi) the overlooked role of the nitrite-oxidizing bacteria (Pachiadaki et al. 2017), or (vii) the adaptation to anoxic niches such as Oxygen Minimum Zones (OMZ) of unique SAR11 lineages with capacity for nitrate respiration (Tsementzi et al. 2016).

8.3.1.2 Metagenome Assembled Genomes (MAGs)

Metagenomic reads can be assembled into contigs and later binned into the so-called Metagenome Assembled Genomes (Fig. 8.4). MAGs are composite genomes of closely related populations from natural communities. The first attempts to reconstruct genomes from metagenomic DNA sequences of environmental communities started in the early 2000s (Martín et al. 2006; Tyson et al. 2004) but more reliable methods and larger-scale results emerged during the last decade (Albertsen et al. 2013; Alneberg et al. 2014; Parks et al. 2017; Sharon and Banfield 2013; Wrighton et al. 2012). Nowadays, thousands of genomes have been reconstructed from marine metagenomes both from discrete sampling events and from global circumnavigations (Acinas et al. 2021; Delmont and Eren 2018; Delmont et al. 2018; Royo-Llonch et al. 2021; Tully et al. 2017). Just recently, a large-scale study used 10,450 metagenomes sampled from a variety of habitats, including marine environments, and recovered 52,515 medium- and high-quality MAGs,
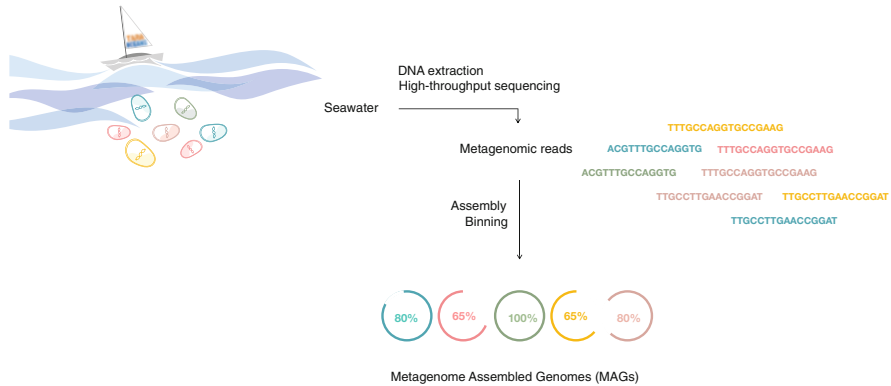
Metagenome Assembly Genomes

Fig. 8.4 Simplified workflow for MAGs generation from seawater samples. Community DNA is extracted from seawater samples and undergoes high-throughput sequencing, which yields large amounts of metagenomic reads. Reads are then assembled into contigs which are later binned to form Metagenome Assembled Genomes (MAGs) of varying coverage levels. Figure courtesy of Dr. Marta Royo-Llonch from SHOOK Studio (https://www.instagram.com/shookstudio/?hl=en)

which constitute the Genomes from Earth's Microbiomes (GEM) catalog, of which 8578 represent marine microbial MAGs (Nayfach et al. 2021).

There has been an increase in the development of tools for estimating genome completion and contamination (Eren et al. 2015; Parks et al. 2015) improving the approaches to retrieve completed (circularized curated, no gaps) Metagenome Assembled Genomes (CMAGs) (Chen et al. 2020) and guidelines on genome quality standards and complementary analyses for the correct deposition in public databases (Bowers et al. 2017; Konstantinidis et al. 2017). However, a community consensus on the pipeline for MAG reconstruction is missing since different strategies in each step of the process (e.g., assemblies, binning, or annotation) are still used.

The reconstruction of bacterial and archaeal MAGs has provided insights into the ecology and evolution of marine microbial taxa unveiling, among other things: (i) the prevalence of diazotrophs in the surface ocean belonging to the Proteobacteria and Planctomycetes (Delmont et al. 2018), (ii) the potential for primary productivity in a globally distributed AAP bacterium (Graham et al. 2018), (iii) the metabolic diversity within Marine Group II Euryarchaea (Tully 2019), (iv) the biogeography and evolutionary processes within the SAR11 clade (Delmont et al. 2019), (v) the widespread potential for mixotrophy found in the genomes of uncultured bacteria and archaea in the bathypelagic ocean (Acinas et al. 2021) or (vi) transcriptional patters of unique bacteria and archaea polar Arctic MAGs (Royo-Llonch et al. 2021).

Both SAGs and MAGs have been successfully combined to derive conclusions from organisms and their ecosystems and to improve the binning quality of single cell assemblies or metagenomes. Finally, the emergence of new pipelines for analysing bacterial and archaeal individual genomes, metagenome assembled

genomes, and single-amplified genomes such as MetaSanity (Neely et al. 2020) will facilitate genome quality evaluation, phylogenetic and functional annotation through a variety of integrated programs. It is clear that these two genome-centric approaches will enhance the understanding of functional and evolutionary processes of the prevalently uncultured marine microbes and will as well serve as a starting point for future experimental validation processes of their potential metabolisms, for in situ taxonomic quantification using CARD-FISH, or for providing essential information in order to design strategies for isolation of key microorganisms into culture.

8.3.2   The Relevance of Culturing Marine Bacteria in the -Omics Era

Despite the fact that gene- and genome-centric approaches have allowed the description of marine microbial diversity from diverse habitats at an unprecedented scale, as overviewed in this chapter, isolates are still a necessary and complementary resource of knowledge. Isolating bacteria and archaea in the laboratory is a fundamental requirement to investigate their physiology under different scenarios, to test ecological hypotheses raised from metagenomics and genome-centric studies, to have access to their complete genomes, to assess the function of novel genes (Muller et al. 2013), to interpret multi-species interactions in co-culture experiments (Stomp et al. 2004, 2008), or to investigate evolutionary principles and population dynamics in long-term monitoring efforts (Good et al. 2017; Rosenzweig et al. 1994). At the same time, genomes from isolates are important to update and improve the existing databases that are needed for the correct annotation of sequencing data (Giovannoni and Stingl 2007; Gutleben et al. 2017). Moreover, isolation is still the only current option for the official procedures for classification and characterization of novel prokaryotic species (Parker et al. 2019) although new efforts to create guidelines for nomenclature of uncultured microorganisms are being developed (Murray et al. 2020). Finally, the short generation times and the nearly 4 billion years of evolution of marine microorganisms have resulted in an enormous biodiversity and a plethora of metabolic pathways and thus, having access to pure cultures, represents an excellent opportunity for biotechnology research (Luna 2015), including bioremediation of polluted ecosystems.

There is little overlap between taxa retrieved by molecular techniques and those retrieved by isolation (Crespo et al. 2016; Lekunberri et al. 2014). This is mainly due to the fact that molecular techniques usually recover the abundant bacteria present in a given environment, while cultures often retrieve those taxa that belong to the rare biosphere (Pedrós-Alió 2012; Sogin et al. 2006). Isolation is thus still essential to decipher the full spectra of diversity of the marine ecosystem (Sanz-Sáez et al. 2020). New culture-dependent techniques have been developed to expand the range of bacteria that can be cultured like microfluidics (Boitard et al. 2015; Ma et al. 2014), culturing chips (Gao et al. 2013; Hesselman et al. 2012; Ingham et al. 2007), manipulation of single cells (Ben-Dov et al. 2009; Park et al. 2011), high-throughput culturing techniques termed "culturomics" (Giovannoni and Stingl 2007; Lagier et al. 2012), or culturing following large-scale dilution to extinction (Henson et al.

2020). A common feature of all these new techniques is that they are based on the same principles of recreating the nutrient conditions of natural environments and overcoming the tendency of rapidly growing cells to outcompete species that reproduce more slowly. The use of these methodologies allowed the isolation of previously uncultured groups such as the alphaproteobacterium *Candidatus* Pelagibacter ubique within the SAR11 clade (Morris et al. 2002; Rappé et al. 2002), or the chemolithotrophic ammonia-oxidizing archaeon *Nitrosopumilus maritimus* (Könneke et al. 2005). The existence of these and other cultures has been essential to test hypotheses derived from genomic data and confirm the role of some proteins in the cell's physiology. For example, experiments with isolates have unveiled light-stimulated growth in some bacteria harboring proteorhodopsin (Gómez-Consarnau et al. 2007), whereas in other bacteria this protein is involved in cell survival (Gómez-Consarnau et al. 2010) or in improving cell fitness (González et al. 2008; Steindler et al. 2011). Likewise, experiments with *Candidatus* Pelagibacter ubique have made possible to understand the growth requirements of bacteria with a limited genetic repertoire (Carini et al. 2013; Tripp 2013; Tripp et al. 2008). Altogether, these culturing depending approaches are fundamental to microbial ecologists to fully understand the ecology, function, and biotechnological potential of microorganisms in marine ecosystems.

8.3.3 Shedding Light on the Active Microbiome

Genome-centric approaches such as SAGs and MAGs have increased our understanding of the metabolic capabilities of uncultured bacteria and archaea and culturing efforts are increasing the array of model organisms to be used for carrying out physiological studies. However, the ultimate goal of microbial ecologists is to understand the activity and function of the different microbes in situ and how they are affected by changes in environmental conditions. Only then we will be able to have a predictive understanding of microbial processes in the ocean.

Different approaches have been used to unravel the role of key taxa driving microbial processes (Berry et al. 2015; Hall et al. 2011; Musat et al. 2012; Singer et al. 2017). Microautoradiography (MAR) coupled with fluorescence in situ hybridization (FISH) has been undoubtedly the most widely used technique to assess the groups of bacteria or archaea that are taking part in the uptake of a given substrate (Alonso-Sáez et al. 2012a; Cottrell and Kirchman 2000; Sintes and Herndl 2006), but it suffers from the poor taxonomic resolution of FISH. Stable isotope probing (SIP), which involves the incubation with a stable isotope-labeled substrate and the downstream analysis of heavy-isotope enriched cellular components such as DNA, RNA, or proteins has also been widely used (see Dumont and Murrell 2005; Musat et al. 2012 for reviews on this technique). Nano-scale secondary ion mass spectrometry (nanoSIMS) is a SIP-based technique that enables the quantification of stable isotopes with high spatial resolution. NanoSIMS has been crucial for unveiling the interactions between individual microbial cells and biochemical processes (see Mayali 2020 and references therein) and particularly relevant for the study of

metabolic fluxes between symbionts and their hosts (Foster et al. 2011; Thompson et al. 2012). However, NanoSIMS is usually coupled with FISH and it is therefore also limited by its phylogenetic resolution. An alternative approach is applying nanoSIMS to an isotope-labeled whole community RNA hybridized to a phylogenetic microarray. This method is called Chip-SIP and it increases the taxonomic resolution to the level of individual taxa (Mayali et al. 2012; Bryson et al. 2017).

Given the short half live of mRNA, metatranscriptomics provides information about the gene expression profiling in near-real time conditions (Moran et al. 2013). This approach has been pivotal to elucidate daily transcriptional oscillations in heterotrophic bacterioplankton (Ottesen et al. 2014) and in viral genes in host cell assemblages (Aylward et al. 2017; Kolody et al. 2019). Besides, degradation pathways may be identified based on shifts in the proportion of certain transcripts upon substrate additions (Li et al. 2014; McCarren et al. 2010; Mou et al. 2011; Vila-Costa et al. 2010), and the fluctuations of transcripts in the mRNA pool are highly informative for how cells sense shifts in environmental conditions and the machinery involved in the response to these shifts (Moran et al. 2013). Nevertheless, high sequencing depth is required to detect transcripts of functional genes that are not involved in core metabolic pathways or transcripts of rare microbes. Metaproteomics is an emergent field in ocean studies (see Saito et al. 2019 for a review), and it has been useful to shed light on transport functions and microbial nutrient utilization (Morris et al. 2010; Sowell et al. 2009), nutrient stresses (Saito et al. 2014) as well as on changes on substrate utilization by microbial communities in the water column (Bergauer et al. 2018). However, metaproteomics is limited to well described proteins produced by the abundant fraction of microbial communities (Saito et al. 2019). Targeted meta-omics with SIP (Chen and Murrell 2010; Coyotzi et al. 2016; Grob et al. 2015) or targeted Single Cell Genomics with fluorescently labeled substrates (Doud et al. 2020; Martinez-Garcia et al. 2012) is a powerful alternative to circumvent the limitation of metatranscriptomics and metaproteomics to only the abundant fraction of the community. They allow the detection of rare taxa participating in a given biochemical process and facilitate the identification of the different enzymes involved. Similarly, other approaches like Bioorthogonal non-canonical amino acid tagging (BONCAT) or Raman micro-spectroscopy (see Hatzenpichler et al. 2020 and references therein) present great prospect for targeted meta-omics. BONCAT is a sensitive technique that uses a synthetic amino acid that upon incorporation can be fluorescently detected via copper-catalyzed alkyne–azide click chemistry (Dieterich et al. 2006). It has been applied to environmental samples to identify protein-synthesizing cells and the taxonomic identification of these cells has been performed by FISH (Hatzenpichler et al. 2014; Sebastián et al. 2019) or by 16S rRNA tagging after fluorescence-activated cell sorting (Couradeau et al. 2019; Hatzenpichler et al. 2016; Reichart et al. 2020). BONCAT has also the potential to study short-term proteomic responses of bacteria (Bagert et al. 2016) and opens new avenues of research to study the proteins involved in relevant biochemical processes. Likewise, Raman activated cell sorting after incubation with heavy water or other isotope-labeled substrates (Berry et al. 2015; Lee et al. 2019; Wang et al. 2013;

Zhang et al. 2015) can be used for subsequent single cell genomics or mini-metagenomics of selected populations (Yu et al. 2017).

Although single cell RNA-Seq has not made its way yet into environmental studies it has the potential to unravel transcriptional heterogeneity of individual bacterial cells (Imdahl et al. 2020) or to discern different stages of viral–host interactions among a single population and monitor how the host metabolism is rewired during viral infection (Ku et al. 2020). All these emerging techniques will surely contribute to a better understanding of microbial processes in the ocean and the key organisms involved, which is central to assess how changes in diversity in the future will impact global biogeochemical cycles.

8.4 What Have we Learnt from the Exploration of the Marine Microbiome?

8.4.1 The Unknown Marine Microbial Diversity

Current predictions indicate that the ocean may be home to $\sim 10^{10}$ microbial species of which only $\sim 10^4$ have been cultured (Locey and Lennon 2016). Moreover, the few cultured microbial species often represent rare members of microbial communities while the most abundant taxa remain largely elusive and begun only to be elucidated after the development of culturing-independent methodologies (Rappé and Giovannoni 2003). Certainly, although the "Great Plate Count Anomaly" was known at the time (Staley and Konopka 1985), the pioneer studies that investigated bacterioplankton diversity in the 90s led among others to the discovery of SAR11 and were groundbreaking because they brought to light that the most abundant organisms in the ocean were in fact unknown (Fuhrman et al. 1993; Giovannoni et al. 1990). Ever since, sequencing DNA from the ocean has continuously unveiled hitherto unknown microorganisms, which has expanded the tree of life to a great extent. Only a negligible fraction of the diversity detected in molecular surveys has been eventually cultured such as the SAR11 (*Candidatus* Pelagibacter ubique) (Rappé et al. 2002), while most of the observed diversity remains uncultured and therefore largely unknown (the so-called "microbial dark matter"). Indeed, most microbial isolates belong to only a few phyla (Hug et al. 2016) and most phyla do not have a cultured representative. Yet, metagenomics and single cell genomics have provided genetic information of many uncultured lineages including those represented by many of the first clones reported in the 90s. This has helped to get insights on the potential functional and ecological roles of these uncultured lineages (Parks et al. 2017; Rinke et al. 2013). For example, the SAR86, an abundant marine clade belonging to the Gammaproteobacteria and detected for the first time in clone libraries constructed almost 30 years ago (Britschgi and Giovannoni 1991) and for which no isolate exists, shares traits with SAR11 such as the presence of proteorhodopsin (Béjà et al. 2000) and metabolic streamlining, but also displays distinct carbon compound specialization that might possibly avoid competition with SAR11 (Dupont et al. 2012). Moreover, analyses of the SAR86 pangenome indicate

that the clade is composed of different ecotypes with unique geographic distributions (Hoarfrost et al. 2020). The SAR406 clade, first reported by Gordon and Giovannoni (1996) and now referred to as the "candidate" phylum *Marinimicrobia*, represents a deeply branching lineage of bacteria that is abundant in the aphotic zone. *Marinimicrobia* possesses the potential to degrade complex carbohydrate compounds as well as performing nitrate reduction, dissimilatory nitrite reduction to ammonia, and sulfur reduction (Thrash et al. 2017; Wright et al. 2014). The SAR202, also common in the aphotic zone, appears to metabolize multiple organosulfur compounds, oxidize sulfite, and oxidize recalcitrant organic compounds and thus are predicted to play major roles in the sulfur and carbon cycles in the aphotic water column (Landry et al. 2017; Mehrshad et al. 2018; Saw et al. 2020).

In addition, SAGs and MAGs have unveiled the existence of many new "candidate" phyla (Rinke et al. 2013; Parks et al. 2017). These "candidate" phyla are often detected in marine environments other than seawater, for example, in hydrothermal vents, blue holes, marine sediments, or associated with marine animals (Dudek et al. 2017; He et al. 2020; Parks et al. 2017; Rinke et al. 2013). However, analyses of the *Tara* Oceans metagenomes have revealed the presence of diverse members of "candidate" archaeal and bacterial lineages in the pelagic realm (Delmont et al. 2018; Lannes et al. 2019; Parks et al. 2017; Royo-Llonch et al. 2021; Tully et al. 2018). Moreover, many unknown clades within the "known" phyla are constantly being discovered in the marine water column (Parks et al. 2017; Yilmaz et al. 2016) and the publication of the genomic catalog of Earth's microbiomes has disclosed a breadth of phylogenetic diversity from multiple biomes including the marine aquatic biome (Nayfach et al. 2021), highlighting the ocean as a reservoir of hidden diversity.

8.4.2 Insights into New Metabolic Capacities of Uncultured Microorganisms

Although the wealth of genomic data of the "uncultured majority" in the ocean is exponentially increasing, there is still a long way to go to be able to interpret these data (Ferrera et al. 2015). Approximately 50% of the predicted genes detected in the ocean have an unknown function (Sunagawa et al. 2015) and the other half has an assigned putative function based on sequence homology to a gene from a distantly related cultured isolate for which the function has been experimentally demonstrated. Indeed, experiments with isolates have demonstrated that homologous genes yield different phenotypes in distinct microbes as is the case with proteorhodopsins mentioned earlier in this chapter. Despite this, homology searches in global metagenomic datasets have been pivotal to explore the potential relevance of different processes once the genes involved in a particular process are identified.

Like the discovery of proteorhodopsin changed the concept of phototrophy in the ocean, the discovery of ammonia oxidation genes in a genomic fragment belonging to Thaumarchaeota in the marine metagenome from the Sargasso Sea (Venter et al.

2004) changed our understanding of the global nitrogen cycle. Until then, nitrification was thought to be performed by a few low abundant bacterial genera. The abundant Thaumarchaeota were recognized as the organisms that exerted the primary control on ammonia oxidation in oligotrophic waters (Martens-Habbena et al. 2009; Wuchter et al. 2006). Genomic data of the "uncultured majority" was also fundamental to unveil other aspects of the nitrogen cycle such as that cyanate and urea may serve as potential substrates for nitrification (Alonso-Sáez et al. 2012b; Pachiadaki et al. 2017; Shi et al. 2011; Yakimov et al. 2011), which was subsequently tested by using cultures or single cell culture-independent approaches (Kitzinger et al. 2019, 2020; Palatinszky et al. 2015).

Likewise, our perspective of the life of microorganisms in the aphotic ocean has changed with the increase in genomic data from the deep-sea realm. In addition to heterotrophic and chemolithotrophic microorganisms, the deep ocean contains also mixotrophs that may play an important role in the carbon cycle (Acinas et al. 2021; Anantharaman et al. 2013; Pachiadaki et al. 2017; Sheik et al. 2014; Swan et al. 2011; Tang et al. 2016). These mixotrophic microorganisms may obtain energy from the oxidation of a plethora of compounds including ammonia, nitrite, CO, sulfur, hydrogen, and recalcitrant DOM (Acinas et al. 2021; Anantharaman et al. 2013; Landry et al. 2017; Martin-Cuadrado et al. 2009; Mehrshad et al. 2018; Pachiadaki et al. 2017; Sheik et al. 2014; Swan et al. 2011). Besides Thaumarchaeota (Reinthaler et al. 2010), other key players participating in carbon fixation in the dark ocean have been identified such as the ubiquitous SAR324 (Swan et al. 2011), the Thiomicrospirales (SUP05 cluster, Mattes et al. 2013), and *Nitrospina* (Pachiadaki et al. 2017). Moreover, a novel pathway for carbon fixation, the reductive glycine pathway, has just been described (Sánchez-Andrea et al. 2020), and the wealth of genomic data available allows the assessment of the distribution and potential relevance of this pathway in the ocean (Fig. 8.5).

After the discovery that the cosmopolitan SAR11 clade produces methane in phosphorus deficient waters as a by-product of the decomposition of methylphosphonate (Carini et al. 2014), the role of the vast oligotrophic gyres as a source of methane to the atmosphere became clear. The genes coding for the enzymes of this pathway of methanogenesis is widespread in marine bacterial genomes (Villarreal-Chiu et al. 2012) and in marine metagenomes (Sosa et al. 2019). Similarly, a relevant strategy to cope with phosphorus stress was unveiled after the identification of a phospholipase C responsible for lipid remodeling in a soil bacterium (Zavaleta-Pastor et al. 2010). This led to the discovery that lipid substitution is a widespread strategy to decrease the phosphorus demand of heterotrophic bacteria in the vast phosphorus depleted waters of the ocean (Carini et al. 2015; Sebastián et al. 2016) similarly to what had been previously observed in phytoplankton (Van Mooy et al. 2009). The discovery of the strategy of lipid substitution to overcome phosphorus depletion shows the need to implement the flexible stoichiometry of planktonic cells in biogeochemical budgets in which a fixed stoichiometry is assumed to be the rule.

Strategies to deal with the vast number of genes with unknown function are constantly being developed. Some approaches involve the functional screening of
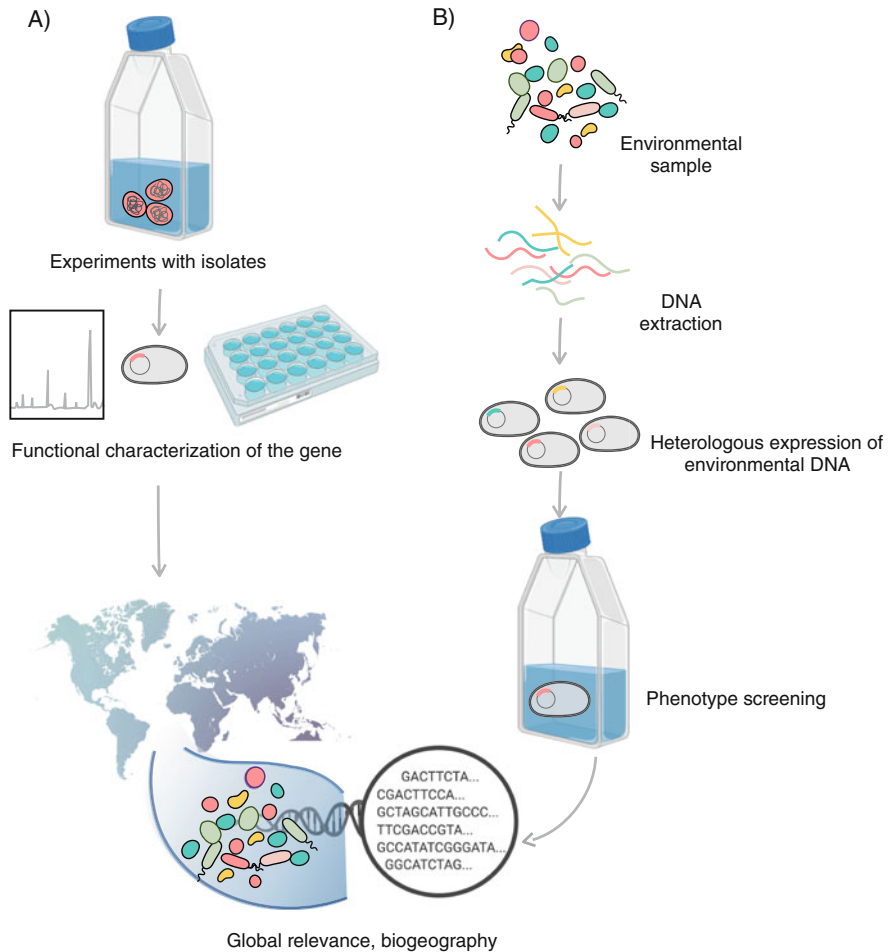
Fig. 8.5 Approaches to unveil new functions in uncultured microorganisms. (**a**) Novel genes are identified and functionally characterized in experiments with isolates from diverse environments (e.g., ocean, soil, gut). Search for these genes in global environmental -omics datasets provides information about the relevance and biogeography of the novel genes, their taxonomic distribution, and the conditions under which these genes are expressed. (**b**) Genes with unknown function can also be identified through functional metagenomics in which environmental DNA fragments are cloned into a plasmid and expressed in a surrogate host. Then, the phenotype can be screened (e.g., enzymatic activity assays, development of color in rhodopsin containing cells). After the function is identified its biogeography and global relevance can be assessed in the global -omics databases. Some items have been created with BioRender.com

metagenomic libraries (e.g., Colin et al. 2015; Pushkarev et al. 2018) while others rely on computational workflows that narrow down the potential role of genes based on the clustering of their coding sequence spaces and their contextualization with genomic and environmental information (Vanni et al. 2020). A large fraction of the

unknown genes is phylogenetically conserved and may be relevant for niche adaptation (Vanni et al. 2020). Thus, identifying these unknown genes is crucial to gain a deeper insight of microbial processes in the ocean. Given that sequencing effort has been directedly linked to the increase in the rate of gene discovery (Duarte et al. 2020), it is expected that many new genes will be uncovered in the near future. This calls for ways to unveil the function of these "unknown" genes.

8.4.3 Delineation of Ecological Meaningful Units of Uncultured Microorganisms

Bacterial and archaeal populations, and by extension species, are fundamental units of ecology and evolution (VanInsberghe et al. 2020). Identifying genetically and ecologically congruent units remains a great challenge in microbial ecology because bacteria and archaea reproduce asexually. Their genomes are subjected to homologous recombination events (Fraser et al. 2007; Papke et al. 2007) and lateral gene transfer between similar or distant relatives frequently occurs (Doolittle and Papke 2006; Fernández-Gómez et al. 2012). Furthermore, even within delineated species, there is no homogeneity in the genetic content or in the total nucleotide composition of bacteria and archaea. This variability is known as microdiversity (Acinas et al. 2004; Fuhrman and Campbell 1998) and it can be seen as "bushy tips" of distinct sequence clusters in phylogenetic reconstructions (Cohan 2001; Giovannoni 2004), by clusters of gene orthologous groups from genomes (Thompson et al. 2019), or by the identification of gene-flow discontinuities derived from recent genetic exchanges (Arevalo et al. 2019). Microdiversity may persist thanks to forces like periodic selection (Cohan 2001) and homologous recombination (Fraser et al. 2007; Konstantinidis and DeLong 2008; Shapiro et al. 2012; Whitaker et al. 2005) in which gene gain and losses are often involved and may promote divergence between populations favoring the delineation of ecotypes (Cordero et al. 2012). The "ecotype concept" describes a collection of strains that show some ecological distinctiveness within its species. Ecotypes preserve nearly the full phenotypic and ecological potential of the species with slight changes in their genetic repertoire that enables them to exploit a slightly different ecological niche. However, microbial speciation occurs in a continuous spectrum in which microbial populations are in constant evolutionary tradeoffs between gene flow and natural selection (Shapiro and Polz 2015; Shapiro et al. 2012) and therefore the delineation of ecological meaningful microbial populations or units remains challenging.

Despite the "species concept" still remains highly controversial, a widely accepted view of microbial species is the "pangenome concept" that classifies the genetic repertoire of a species into the core genome and the flexible genome. The former includes the shared genes between all individuals categorized as the same species and the latter includes the gene pool that is partially shared or strain-specific (Mira et al. 2010; Tettelin et al. 2005). Thus, the genomes of multiple representatives of a species are needed to accurately define the genetic potential and size of the pangenome. Even though the core genome contains the essence of the species and is
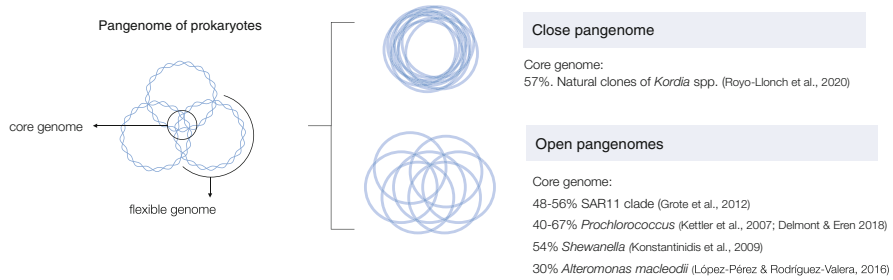
Fig. 8.6 Visual representation of the pangenome concept for genomes of bacteria and archaea. The pangenome classifies de genetic repertoire of a species into the core genome, that includes all genes shared between all individuals categorized as the same species, and the flexible genome, which includes the gene pool that is partially shared or strain-specific. Depending on the genome size and genetic diversity of each species, the pangenome can be open or close (some examples from the literature are shown). Figure courtesy of Dr. Marta Royo-Llonch from SHOOK Studio (https://www.instagram.com/shookstudio/?hl=en)

indispensable, it is the flexible genome that can confer selective advantages like niche adaptation, new host colonization, or antibiotic resistance, and contributes to the species diversity (Tettelin et al. 2005, 2008). The size of the pangenome is dynamic and pangenomes can be more open or close depending on the multitude of niches in which the species is able to live (Medini et al. 2005) (Fig. 8.6).

In marine ecosystems, the pangenome exploration was first performed with genomes from bacterial isolates including *Prochlorococcus marinus* (Kettler et al. 2007), *Synechococcus* (Dufresne et al. 2008), *Shewanella* (Konstantinidis et al. 2009), *Alteromonas macleodii* (López-Pérez and Rodriguez-Valera 2016), and *Vibrio alginolyticus* (Chibani et al. 2020) but generally using a limited number of genomes. A large-scale analysis of pangenomes using ~7000 high-quality cultured genomes from 155 phylogenetically diverse species belonging to ten phyla revealed the important role of environmental preferences and phylogeny in explaining the majority of variation of pangenome features across different species (Maistrenko et al. 2020). The advent of Single Cell Genomics, however, has enabled the study of the pangenome concept in uncultured microbial taxa (Fig. 8.6). Some pioneer studies have explored the pangenomes of uncultured *Prochlorococcus* (Kashtan et al. 2014; Thompson et al. 2019), SAR11 lineages (Grote et al. 2012; Haro-Moreno et al. 2020; Thrash et al. 2014; Tsementzi et al. 2016), or uncultured Bacteroidetes relatives of *Kordia* sp. (Royo-Llonch et al. 2020).

Further, the combination of pangenomes and fragment recruitment analyses on marine metagenomes has led to the "metapangenomes concept" that uses the delineation of single amino acid variants to explore the biogeography of distinct populations. This has been successfully applied on single cell genomes of *Prochlorococcus* (Delmont and Eren 2018) and the SAR11 clade (Delmont et al. 2019) (Fig. 8.6). Similarly, the "genomospecies concept" represents a species that can be differentiated from others based on the average nucleotide identity (ANI) value by comparative genomics, phylogenomic and fragment recruitment analyses (Haro-Moreno et al. 2020). Alternatively, the reverse ecology approach (Arevalo

et al. 2019; Shapiro and Polz 2015) used on isolate's genomes or single cell genomes relies on the identification of microbial populations as gene-flow units using network analyses. This approach enables the delineation of ecologically relevant populations and the identification of genes that have been under recent positive selection (Arevalo et al. 2019; Shapiro et al. 2012). All these alternatives will become more powerful when applied at large scale using uncultured genomes from different oceanic regions spanning a wider array of phylogenetic taxa, which will altogether enhance our knowledge on the evolution processes driving microbial speciation.

8.5   Future Perspectives

Despite last years' integrative research has boosted the knowledge of the marine microbiome to an unprecedented scale, as overviewed in this chapter, it is still insufficient for understanding the functioning of the microbiome and how it will respond to climate change-driven alterations. Here, we discuss some of the issues that in our opinion will become increasingly important in marine microbiome research.

- *Large-scale sequencing and inputs from new technologies.* Analyses of marine microbial communities from global circumnavigations have pointed out that only the "tip of the iceberg" has been unraveled and that even larger sequencing efforts are necessary to cover the immense genetic diversity of marine microbiomes. The generation of high-quality genomic references (using long read sequencing such as Nanopore) should be a priority to enhance the genome completeness and quality of SAGs and MAGs.
- *Large-scale single cell genomics studies.* Extending the analyses of SAGs at large scale from a wide array of phylogenetic taxa combined with cutting-edge analyses of population genomics will enhance our knowledge of the evolutionary processes driving microbial speciation.
- *Increasing efforts in high-throughput culturomics.* Large marine microbial culture collections will improve the microbial reference gene catalogs and will also be a fundamental tool to investigate the physiology of marine bacteria and archaea and to test physiological traits inferred from -omics data.
- *The scale matters.* Choosing the right spatial and temporal scales is essential to fully comprehend microbial processes in the ocean and to understand the triggers driving changes in community composition and ecosystem function. Despite the increasing effort to move from local to global scales, analyses at the fine scale are still lacking, such as how sub-mesoscale oceanographic features drive changes in the composition and function of the marine microbiome. Other unique features like the deep scattering layer, which is a hotspot for microbial activity, deserves further investigation. Likewise, at the temporal scale, highly resolved sampling over long periods must be carried out to be able to investigate the stability and repeatability of short-term changes.

- *Expand environmental genomic datasets, particularly on the temporal dimension.* Global expeditions have an incalculable value. Yet, they provide a snapshot of the marine microbiome at the moment of sampling. Repeated global expeditions at different times of the year in combination with time series from different latitudes will be necessary to have a complete view of the marine microbiome.

- *Beyond DNA sequencing.* DNA-based approaches have been highly informative on the diversity, potential metabolic capabilities, and evolution of the uncultured majority, but more effort should be invested towards understanding how this genetic information translates into function. The extended use of other -omics approaches, such as metatranscriptomics, metaproteomics, and metabolomics, will help to identify which metabolic processes are actually occurring in the environment and their magnitude. Likewise, the application of next generation physiological approaches such as BONCAT and Raman will be pivotal to experimentally validate information inferred from -omics data. A multifaceted approach in microbiome research is needed to grasp the relevance that the observed diversity has for ecosystem functioning.

References

Acinas SG, Klepac-Ceraj V, Hunt DE et al (2004) Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430:551–554

Acinas SG, Sánchez P, Salazar G et al (2021) Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. Commun Biol 4:604

Agogué H, Lamy D, Neal PR et al (2011) Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. Mol Ecol 20:258–274

Albertsen M, Hugenholtz P, Skarshewski A et al (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol 31:533–538

Alneberg J, Bjarnason BS, De Bruijn I et al (2014) Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146

Alonso-Sáez L, Balagué V, Sà EL et al (2007) Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. FEMS Microbiol Ecol 60:98–112

Alonso-Sáez L, Díaz-Pérez L, Morán XAG (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. Environ Microbiol 10:3766–3780

Alonso-Sáez L, Sánchez O, Gasol JM (2012a) Bacterial uptake of low molecular weight organics in the subtropical Atlantic: are major phylogenetic groups functionally different? Limnol Oceanogr 57:798–808

Alonso-Sáez L, Waller AS, Mende DR et al (2012b) Role for urea in nitrification by polar marine archaea. Proc Natl Acad Sci 109:17989–17994

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

Amaral-Zettler L, Artigas LF, Baross J et al (2010) A global census of marine microbes. Life in the world's oceans: diversity, distribution and abundance. Wiley-Blackwell, Oxford, UK, pp 221–245

Amend AS, Oliver TA, Amaral-Zettler LA et al (2013) Macroecological patterns of marine bacteria on a global scale. J Biogeogr 40:800–811

Anantharaman K, Breier JA, Sheik CS, Dick GJ (2013) Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. Proc Natl Acad Sci 110: 330–335

Arevalo P, VanInsberghe D, Elsherbini J et al (2019) A reverse ecology approach based on a biological definition of microbial populations. Cell 178:820–834

Auladell A, Barberán A, Logares R et al (2021) Seasonal niche differentiation among closely related marine bacteria. ISME J. https://doi.org/10.1038/s41396-021-01053-2

Auladell A, Sánchez P, Sánchez O et al (2019) Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. ISME J 13:1975–1987

Aylward FO, Boeuf D, Mende DR et al (2017) Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. Proc Natl Acad Sci 114:11446–11451

Baas Becking LGM (1934) Geobiologie of inleiding tot de milieukunde. Van Stockum WP and Zoon NV, Den Haag, The Netherlands

Bagert JD, van Kessel JC, Sweredoski MJ et al (2016) Time-resolved proteomic analysis of quorum sensing in *Vibrio harveyi*. Chem Sci 7:1797–1806

Balashov SP, Imasheva ES, Boichenko VA et al (2005) Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. Science 309:2061–2064

Beijerinck MW (1913) De infusies en de ontdekking der bakterien. Johannes Müller

Béjà O, Aravind L, Koonin EV et al (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science 289:1902–1906

Ben-Dov E, Kramarsky-Winter E, Kushmaro A (2009) An *in situ* method for cultivating microorganisms using a double encapsulation technique. FEMS Microbiol Ecol 68:363–371

Bergauer K, Fernandez-Guerra A, Garcia JAL et al (2018) Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. Proc Natl Acad Sci 115:E400–E408

Berry D, Mader E, Lee TK et al (2015) Tracking heavy water (D2O) incorporation for identifying and sorting active microbial cells. Proc Natl Acad Sci 112:194–203

Biller SJ, Berube PM, Dooley K et al (2018) Marine microbial metagenomes sampled across space and time. Sci Data 5:180176

Blainey PC (2013) The future is now: single-cell genomics of bacteria and archaea. FEMS Microbiol Rev 37:407–427

Blazewicz SJ, Barnard RL, Daly RA, Firestone MK (2013) Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J 7:2061–2068

Boitard L, Cottinet D, Bremond N et al (2015) Growing microbes in millifluidic droplets. Eng Life Sci 15:318–326

Bowers RM, Kyrpides NC, Stepanauskas R et al (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731

Bowsher AW, Kearns PJ, Shade A (2019) 16S rRNA/rRNA gene ratios and cell activity staining reveal consistent patterns of microbial activity in plant-associated soil. mSystems 4:e00003–e00019

Boyd PW, Claustre H, Levy M et al (2019) Multi-faceted particle pumps drive carbon sequestration in the ocean. Nature 568:327–335

Britschgi TB, Giovannoni SJ (1991) Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. Appl Environ Microbiol 57:1707–1713

Brown MV, Schwalbach MS, Hewson I, Fuhrman JA (2005) Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. Environ Microbiol 7:1466–1479

Brum JR, Ignacio-Espinoza JC, Roux S et al (2015) Patterns and ecological drivers of ocean viral communities. Science 348:1261498

Bryson S, Li Z, Chavez F et al (2017) Phylogenetically conserved resource partitioning in the coastal microbial loop. ISME J 11:2781–2792

Bunse C, Pinhassi J (2017) Marine bacterioplankton seasonal succession dynamics. Trends Microbiol 25:494–505

Buttigieg PL, Fadeev E, Bienhold C et al (2018) Marine microbes in 4D–using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. Curr Opin Microbiol 43:169–185

Campbell BJ, Kirchman DL (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. ISME J 7:210–220

Campbell BJ, Yu L, Heidelberg JF, Kirchman DL (2011) Activity of abundant and rare bacteria in a coastal ocean. Proc Natl Acad Sci 108:12776–12781

Campbell BJ, Yu L, Straza TR, Kirchman DL (2009) Temporal changes in bacterial rRNA and rRNA genes in Delaware (USA) coastal waters. Aquat Microb Ecol 57:123–135

Cao S, Zhang W, Ding W et al (2020) Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. Microbiome 8:47

Carini P, Steindler L, Beszteri S, Giovannoni SJ (2013) Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. ISME J 7:592–602

Carini P, Van Mooy BAS, Thrash JC et al (2015) SAR11 lipid renovation in response to phosphate starvation. Proc Natl Acad Sci 112:7767–7772

Carini P, White AE, Campbell EO, Giovannoni SJ (2014) Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. Nat Commun 5:4346

Carradec Q, Pelletier E, Da Silva C et al (2018) A global ocean atlas of eukaryotic genes. Nat Commun 9:373

Castillo YM, Mangot J, Benites LF et al (2019) Assessing the viral content of uncultured picoeukaryotes in the global-ocean by single cell genomics. Mol Ecol 28:4272–4289

Chafee M, Fernàndez-Guerra A, Buttigieg PL et al (2018) Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J 12:237–252

Chen LX, Anantharaman K, Shaiber A et al (2020) Accurate and complete genomes from metagenomes. Genome Res 30:315–333

Chen Y, Murrell JC (2010) When metagenomics meets stable-isotope probing: progress and perspectives. Trends Microbiol 18:157–163

Chibani CM, Roth O, Liesegang H, Wendling CC (2020) Genomic variation among closely related Vibrio alginolyticus strains is located on mobile genetic elements. BMC Genomics 21:1–14

Chisholm SW, Olson RJ, Zettler ER et al (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature 334:340–343

Cohan FM (2001) Bacterial species and speciation. Syst Biol 50:513–524

Coleman ML, Chisholm SW (2010) Ecosystem-specific selection pressures revealed through comparative population genomics. Proc Natl Acad Sci 107:18634–18639

Colin PY, Kintses B, Gielen F et al (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. Nat Commun 6:10008

Cordero OX, Ventouras LA, DeLong EF, Polz MF (2012) Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. Proc Natl Acad Sci 109: 20059–20064

Cottrell MT, Kirchman DL (2000) Natural assemblages of marine proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-weight dissolved organic matter. Appl Environ Microbiol 66:1692–1697

Couradeau E, Sasse J, Goudeau D et al (2019) Probing the active fraction of soil microbiomes using BONCAT-FACS. Nat Commun 10:2770

Coyotzi S, Pratscher J, Murrell JC, Neufeld JD (2016) Targeted metagenomics of active microbial populations with stable-isotope probing. Curr Opin Biotechnol 41:1–8

Cram JA, Chow C-ET, Sachdeva R et al (2015) Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. ISME J 9:563–580

Crespo BG, Wallhead PJ, Logares R, Pedrós-Alió C (2016) Probing the rare biosphere of the North-West Mediterranean Sea: an experiment with high sequencing effort. PLoS One 11:e0159195

De Vargas C, Audic S, Henry N et al (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348:1261605

Del Giorgio PA, Gasol JM (2008) Physiological structure and single-cell activity in marine bacterioplankton. In: Kirchman DL (ed) Microbial ecology of the oceans, 2nd edn. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 243–298

Del Giorgio PA, Scarborough G (1995) Increase in the proportion of metabolically active bacteria along gradients of enrichment in freshwater and marine plankton: implications for estimates of bacterial growth and production rates. J Plankton Res 17:1905–1924

Delmont TO, Eren AM (2018) Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. PeerJ 6:e4320

Delmont TO, Kiefl E, Kilinc O et al (2019) Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. elife 8:e46497

Delmont TO, Quince C, Shaiber A et al (2018) Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol 3:804–813

DeLong EF, Béjà O (2010) The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. PLoS Biol 8:e1000359

DeLong EF, Preston CM, Mincer TJ et al (2006) Community genomics among stratified microbial assemblages in the ocean's interior. Science 311:496–503

Dieterich DC, Link AJ, Graumann J et al (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). Proc Natl Acad Sci 103:9482–9487

Díez-Vives C, Nielsen S, Sánchez P et al (2019) Delineation of ecologically distinct units of marine Bacteroidetes in the Northwestern Mediterranean Sea. Mol Ecol 28:2846–2859

Doolittle WF, Papke RT (2006) Genomics and the bacterial species problem. Genome Biol 7:116

Doud DFR, Bowers RM, Schulz F et al (2020) Function-driven single-cell genomics uncovers cellulose-degrading bacteria from the rare biosphere. ISME J 14:659–675

Duarte CM (2015) Seafaring in the 21st century: the Malaspina 2010 Circumnavigation Expedition. Limnol Oceanogr Bull 24:11–14

Duarte CM, Ngugi DK, Alam I et al (2020) Sequencing effort dictates gene discovery in marine microbial metagenomes. Environ Microbiol 22:4589–4603

Dudek NK, Sun CL, Burstein D et al (2017) Novel microbial diversity and functional potential in the marine mammal oral microbiome. Curr Biol 27:3752–3762

Dufresne A, Ostrowski M, Scanlan DJ et al (2008) Unravelling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biol 9:1–16

Dumont MG, Murrell JC (2005) Stable isotope probing–linking microbial identity to function. Nat Rev Microbiol 3:499–504

Dupont CL, Rusch DB, Yooseph S et al (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J 6:1186–1199

Eiler A, Hayakawa DH, Rappé MS (2011) Non-random assembly of bacterioplankton communities in the subtropical North Pacific Ocean. Front Microbiol 2:140

Eren AM, Esen ÖC, Quince C et al (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319

Fernández-Gómez B, Fernàndez-Guerra A, Casamayor EO et al (2012) Patterns and architecture of genomic islands in marine bacteria. BMC Genomics 13:347

Ferrera I, Sánchez O, Kolářová E, Koblížek M, Gasol JM (2017) Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. ISME J 11:2391–2393

Ferrera I, Sebastian M, Acinas SG, Gasol JM (2015) Prokaryotic functional gene diversity in the sunlit ocean: stumbling in the dark. Curr Opin Microbiol 25:33–39

Foster RA, Kuypers MMM, Vagner T et al (2011) Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. ISME J 5:1484–1493

Fraser C, Hanage WP, Spratt BG (2007) Recombination and the nature of bacterial speciation. Science 315:476–480

Frigaard NU, Martinez A, Mincer TJ, DeLong EF (2006) Proteorhodopsin lateral gene transfer between marine planktonic bacteria and archaea. Nature 439:847–850

Fuhrman JA, Campbell L (1998) Microbial microdiversity. Nature 393:410–411

Fuhrman JA, Cram JA, Needham DDM (2015) Marine microbial community dynamics and their ecological interpretation. Nat Rev Microbiol 13:133–146

Fuhrman JA, Hewson I, Schwalbach MS et al (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. Proc Natl Acad Sci 103:13104–13109

Fuhrman JA, McCallum K, Davis AA (1993) Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific oceans. Appl Environ Microbiol 59:1294–1302

Galand PE, Alonso-Sáez L, Bertilsson S, Lovejoy C, Casamayor EO (2013) Contrasting activity patterns determined by BrdU incorporation in bacterial ribotypes from the Arctic Ocean in winter. Front Microbiol 4:118

Gao W, Navarroli D, Naimark J et al (2013) Microbe observation and cultivation array (MOCA) for cultivating and analyzing environmental microbiota. Microbiome 1:4

Ghiglione JF, Conan P, Pujo-Pay M (2009) Diversity of total and active free-living vs. particle-attached bacteria in the euphotic zone of the NW Mediterranean Sea. FEMS Microbiol Lett 299:9–21

Ghiglione JF, Galand PE, Pommier T et al (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. Proc Natl Acad Sci 109:17633–17638

Ghiglione JF, Larcher M, Lebaron P (2005) Spatial and temporal scales of variation in bacterioplankton community structure in the NW Mediterranean Sea. Aquat Microb Ecol 40:229–240

Gilbert JA, Steele JA, Caporaso JG et al (2012) Defining seasonal marine microbial community dynamics. ISME J 6:298–308

Giner CR, Balagué V, Krabberød AK et al (2019) Quantifying long-term recurrence in planktonic microbial eukaryotes. Mol Ecol 28:923–935

Giovannoni S (2004) Oceans of bacteria. Nature 430:515–516

Giovannoni S, Stingl U (2007) The importance of culturing bacterioplankton in the 'omics' age. Nat Rev Microbiol 5:820–826

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63

Gómez-Consarnau L, Akram N, Lindell K et al (2010) Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. PLoS Biol 8:e1000358

Gómez-Consarnau L, González JM, Coll-Lladó M et al (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. Nature 445:210–213

González JM, Fernández-Gómez B, Fernàndez-Guerra A et al (2008) Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). Proc Natl Acad Sci 105:8724–8729

Good BH, McDonald MJ, Barrick JE et al (2017) The dynamics of molecular evolution over 60,000 generations. Nature 551:45–50

Gordon DA, Giovannoni SJ (1996) Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific oceans. Appl Environ Microbiol 62:1171–1177

Graham ED, Heidelberg JF, Tully BJ (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. ISME J 12:1861–1866

Gregory AC, Zayed AA, Conceição-Neto N et al (2019) Marine DNA viral macro-and microdiversity from pole to pole. Cell 177:1109–1123

Grob C, Taubert M, Howat AM et al (2015) Combining metagenomics with metaproteomics and stable isotope probing reveals metabolic pathways used by a naturally occurring marine methylotroph. Environ Microbiol 17:4007–4018

Grote J, Thrash JC, Huggett MJ et al (2012) Streamlining and core genome conservation among highly divergent members of the SAR11 clade. MBio 3:e00252–e00212

Guidi L, Chaffron S, Bittner L et al (2016) Plankton networks driving carbon export in the oligotrophic ocean. Nature 532:465–470

Gutleben J, Chaib M, Mares D et al (2017) The multi-omics promise in context: from sequence to microbial isolate. Crit Rev Microbiol 44:212–219

Hall EK, Singer GA, Pölzl M et al (2011) Looking inside the box: using Raman microspectroscopy to deconstruct microbial biomass stoichiometry one cell at a time. ISME J 5:196–208

Haro-Moreno JM, Rodriguez-Valera F, Rosselli R et al (2020) Ecogenomics of the SAR11 clade. Environ Microbiol 22:1748–1763

Hatzenpichler R, Connon SA, Goudeau D et al (2016) Visualizing in situ translational activity for identifying and sorting slow-growing archaeal−bacterial consortia. Proc Natl Acad Sci 113: E4069–E4078

Hatzenpichler R, Krukenberg V, Spietz RL, Jay ZJ (2020) Next-generation physiology approaches to study microbiome function at single cell level. Nat Rev Microbiol 18:241–256

Hatzenpichler R, Scheller S, Tavormina PL et al (2014) *In situ* visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. Environ Microbiol 16:2568–2590

Haruuseu D, Koblížek M (2012) Influence of light on carbon utilization in aerobic anoxygenic phototrophs. Appl Environ Microbiol 78:7414–7419

He P, Xie L, Zhang X et al (2020) Microbial diversity and metabolic potential in the stratified Sansha Yongle blue hole in the South China Sea. Sci Rep 10:5949

Henson MW, Lanclos VC, Pitre DM et al (2020) Expanding the diversity of bacterioplankton isolates and modeling isolation efficacy with large-scale dilution-to-extinction cultivation. Appl Environ Microbiol 86:1–23

Hesselman MC, Odoni DI, Ryback BM et al (2012) A multi-platform flow device for microbial (co-) cultivation and microscopic analysis. PLoS One 7:e36982

Hoarfrost A, Nayfach S, Ladau J et al (2020) Global ecotypes in the ubiquitous marine clade SAR86. ISME J 14:178–188

Hoppe HG (1976) Determination and properties of actively metabolizing heterotrophic bacteria in the sea, investigated by means of micro-autoradiography. Mar Biol 36:291–302

Huang WE, Stoecker K, Griffiths R et al (2007) Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. Environ Microbiol 9:1878–1889

Hug LA, Baker BJ, Anantharaman K et al (2016) A new view of the tree of life. Nat Microbiol 1: 16048

Hugoni M, Taib N, Debroas D et al (2013) Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. Proc Natl Acad Sci 110:6004–6009

Hunt DE, Lin Y, Church MJ et al (2013) Relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. Appl Environ Microbiol 79:177–184

Ibarbalz FM, Henry N, Brandão MC et al (2019) Global trends in marine plankton diversity across kingdoms of life. Cell 179:1084–1097

Imdahl F, Vafadarnejad E, Homberger C et al (2020) Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. Nat Microbiol 5:1202–1206

Ingham CJ, Sprenkels A, Bomer J et al (2007) The micro-petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. Proc Natl Acad Sci 104:18217–18222

Jannasch HW, Jones GE (1959) Bacterial populations in sea water as determined by different methods of enumeration. Limnol Oceanogr 4:128–139

Johnson PW, Sieburth JMN (1979) Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototrophic biomass. Limnol Oceanogr 24:928–935

Jones SE, Lennon JT (2010) Dormancy contributes to the maintenance of microbial diversity. Proc Natl Acad Sci 107:5881–5886

Karl DM, Church MJ (2014) Microbial oceanography and the Hawaii Ocean time-series programme. Nat Rev Microbiol 12:699–713

Karsenti E, Acinas SG, Bork P et al (2011) A holistic approach to marine eco-systems biology. PLoS Biol 9:e1001177

Kashtan N, Roggensack SE, Rodrigue S et al (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science 344:416–420

Kearns PJ, Angell JH, Howard EM et al (2016) Nutrient enrichment induces dormancy and decreases diversity of active bacteria in salt marsh sediments. Nat Commun 7:12881

Kettler GC, Martiny AC, Huang K et al (2007) Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet 3:2515–2528

Kirchman DL (2016) Growth rates of microbes in the oceans. Annu Rev Mar Sci 8:285–309

Kirchman DL, Hanson TE (2013) Bioenergetics of photoheterotrophic bacteria in the oceans. Environ Microbiol Rep 5:188–199

Kitzinger K, Marchant HK, Bristow LA et al (2020) Single cell analyses reveal contrasting life strategies of the two main nitrifiers in the ocean. Nat Commun 11:767

Kitzinger K, Padilla CC, Marchant HK et al (2019) Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. Nat Microbiol 4:234–243

Koblížek M (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. FEMS Microbiol Rev 39:854–870

Kolber ZS, Van Dover CL, Niederman RA, Falkowski PG (2000) Bacterial photosynthesis in surface waters of the open ocean. Nature 407:177–179

Kolody BC, McCrow JP, Allen LZ et al (2019) Diel transcriptional response of a California current plankton microbiome to light, low iron, and enduring viral infection. ISME J 13:2817–2833

Könneke M, Bernhard AE, de la Torre JR et al (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437:543–546

Konstantinidis KT, DeLong EF (2008) Genomic patterns of recombination clonal divergence and environment in marine microbial populations. ISME J 2:1052–1065

Konstantinidis KT, Rosselló-Móra R, Amann R (2017) Uncultivated microbes in need of their own taxonomy. ISME J 11:2399–2406

Konstantinidis KT, Serres MH, Romine MF et al (2009) Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. Proc Natl Acad Sci 106:15909–15914

Kopf A, Bicak M, Kottmann R et al (2015) The ocean sampling day consortium. Gigascience 4:27

Ku C, Sheyn U, Sebé-Pedrós A et al (2020) A single-cell view on alga-virus interactions reveals sequential transcriptional programs and infection states. Sci Adv 6:eaba4137

Labonté JM, Swan BK, Poulos B et al (2015) Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. ISME J 9:2386–2399

Lagier JC, Armougom F, Million M et al (2012) Microbial culturomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 18:1185–1193

Lambert S, Tragin M, Lozano JC et al (2019) Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. ISME J 13:388–401

Landry Z, Swa BK, Herndl GJ et al (2017) SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. mBio 8:e00413–e00417

Lannes R, Olsson-Francis K, Lopez P, Bapteste E (2019) Carbon fixation by marine ultrasmall prokaryotes. Genome Biol Evol 11:1166–1177

Larkin AA, Garcia CA, Garcia N et al (2021) High spatial resolution global ocean metagenomes from Bio-GO-SHIP repeat hydrography transects. Sci Data 8:107

Lebaron P, Servais P, Agogué H, Courties C, Joux F (2001) Does the high nucleic acid content of individual bacterial cells allow us to discriminate between active cells and inactive cells in aquatic systems? Appl Environ Microbiol 67:1775–1782

Lee KS, Palatinszky M, Pereira FC et al (2019) An automated Raman-based platform for the sorting of live cells by functional properties. Nat Microbiol 4:1035–1048

Leizeaga A, Estrany M, Forn I, Sebastián M (2017) Using click-chemistry for visualizing *in situ* changes of translational activity in planktonic marine bacteria. Front Microbiol 8:2360

Lekunberri I, Gasol JM, Acinas SG et al (2014) The phylogenetic and ecological context of cultured and whole genome-sequenced planktonic bacteria from the coastal NW Mediterranean Sea. Syst Appl Microbiol 37:216–228

Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol 9:119–130

Li WKW, Rao DVS, Harrison WG et al (1983) Autotrophic picoplankton in the tropical ocean. Science 219:292–295

Li M, Jain S, Baker BJ et al (2014) Novel hydrocarbon monooxygenase genes in the metatranscriptome of a natural deep-sea hydrocarbon plume. Environ Microbiol 16:60–71

Li WKW, Jellett JF, Dickie PM (1995) DNA distributions in planktonic bacteria stained with TOTO or TO-PRO. Limnol Oceanogr 40:1485–1495

Lima-Mendez G, Faust K, Henry N et al (2015) Determinants of community structure in the global plankton interactome. Science 348:1262073

Locey KJ, Lennon JT (2016) Scaling laws predict global microbial diversity. Proc Natl Acad Sci 113:5970–5975

Longnecker K, Lomas MW, Van Mooy BA (2010) Abundance and diversity of heterotrophic bacterial cells assimilating phosphate in the subtropical North Atlantic Ocean. Environ Microbiol 12:2773–2782

Longnecker K, Sherr BF, Sherr EB (2005) Activity and phylogenetic diversity of bacterial cells with high and low nucleic acid content and electron transport system activity in an upwelling ecosystem. Appl Environ Microbiol 71:7737–7749

López-Pérez M, Rodriguez-Valera F (2016) Pangenome evolution in the marine bacterium *Alteromonas*. Genome Biol Evol 8:1556–1570

Lorenz B, Wichmann C, Stöckel S, Rösch P, Popp J (2017) Cultivation-free Raman spectroscopic investigations of bacteria. Trends Microbiol 25:413–424

Luna GM (2015) Biotechnological potential of marine microbes. In: Kim S (ed) Springer handbook of marine biotechnology. Springer, Berlin, Heidelberg, pp 651–661

Ma L, Kim J, Hatzenpichler R et al (2014) Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in human microbiome Project's Most wanted taxa. Proc Natl Acad Sci 111:9768–9773

Maistrenko OM, Mende DR, Luetge M et al (2020) Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity. ISME J 14:1247–1259

Martens-Habbena W, Berube PM, Urakawa H et al (2009) Ammonia oxidation kinetics determine niche separation of nitrifying archaea and bacteria. Nature 461:976–979

Martín HG, Ivanova N, Kunin V et al (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotechnol 24:1263–1269

Martin-Cuadrado AB, Ghai R, Gonzaga A, Rodriguez-Valera F (2009) CO dehydrogenase genes found in metagenomic fosmid clones from the deep Mediterranean Sea. Appl Environ Microbiol 75:7436–7444

Martin-Platero AM, Cleary B, Kauffman K et al (2018) High resolution time series reveals cohesive but short-lived communities in coastal plankton. Nat Commun 9:266

Martinez-Garcia M, Brazel DM, Swan BK et al (2012) Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of *Verrucomicrobia*. PLoS One 7:e35314

Martinez-Garcia M, Santos F, Moreno-Paz M et al (2014) Unveiling viral–host interactions within the 'microbial dark matter'. Nat Commun 5:4542

Mattes TE, Nunn BL, Marshall KT et al (2013) Sulfur oxidizers dominate carbon fixation at a biogeochemical hot spot in the dark ocean. ISME J 7:2349–2360

Mayali X (2020) NanoSIMS: microscale quantification of biogeochemical activity with large-scale impacts. Annu Rev Mar Sci 12:449–467

Mayali X, Weber PK, Brodie EL et al (2012) High-throughput isotopic analysis of RNA microarrays to quantify microbial resource use. ISME J 6:1210–1221

McCarren J, Becker JW, Repeta DJ et al (2010) Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. Proc Natl Acad Sci 107:16420–16427

Medini D, Donati C, Tettelin H et al (2005) The microbial pan-genome. Curr Opin Genet Dev 15:589–594

Mehrshad M, Rodriguez-Valera F, Amoozegar MA et al (2018) The enigmatic SAR202 cluster up close: shedding light on a globally distributed dark ocean lineage involved in sulfur cycling. ISME J 12:655–668

Mestre M, Ruiz-González C, Logares R et al (2018) Sinking particles promote vertical connectivity in the ocean microbiome. Proc Natl Acad Sci 115:E6799–E6807

Mira A, Martín-Cuadrado AB, D'Auria G, Rodríguez-Valera F (2010) The bacterial pan-genome: a new paradigm in microbiology. Int Microbiol 13:45–57

Moran MA, Satinsky B, Gifford SM et al (2013) Sizing up metatranscriptomics. ISME J 7:237–243

Morris RM, Nunn BL, Frazar C et al (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J 4:673–685

Morris RM, Rappé MS, Connon SA et al (2002) SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420:806–810

Mou X, Sun S, Edwards RA, Hodson RE, Moran MA (2008) Bacterial carbon processing by generalist species in the coastal ocean. Nature 451:708–711

Mou X, Vila-Costa M, Sun et al (2011) Metatranscriptomic signature of exogenous polyamine utilization by coastal bacterioplankton. Environ Microbiol Rep 3:798–806

Muller EEL, Glaab E, May P et al (2013) Condensing the omics fog of microbial communities. Trends Microbiol 21:325–333

Murray AE, Freudenstein J, Gribaldo S et al (2020) Roadmap for naming uncultivated archaea and bacteria. Nat Microbiol 5:987–994

Musat N, Foster R, Vagner T et al (2012) Detecting metabolic activities in single cells, with emphasis on nanoSIMS. FEMS Microbiol Rev 36:486–511

Nayfach S, Roux S, Seshadri R et al (2021) A genomic catalog of Earth's microbiomes. Nat Biotechnol 39:499–509

Needham DM, Fuhrman JA (2016) Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat Microbiol 1:16005

Needham DM, Yoshizawa S, Hosaka T et al (2019) A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. Proc Natl Acad Sci 116:20574–20583

Neely CJ, Graham ED, Tully BJ (2020) MetaSanity: an integrated microbial genome evaluation and annotation pipeline. Bioinformatics 36:4341–4344

Nelson CE, Carlson CA (2012) Tracking differential incorporation of dissolved organic carbon types among diverse lineages of Sargasso Sea bacterioplankton. Environ Microbiol 14:1500–1516

Orsi WD, Smith JM, Liu S et al (2016) Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. ISME J 10:2158–2173

Ottesen EA, Young CR, Eppley JM et al (2013) Pattern and synchrony of gene expression among sympatric marine microbial populations. Proc Natl Acad Sci 110:E488–E497

Ottesen EA, Young CR, Gifford SM et al (2014) Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. Science 345:207–212

Pachiadaki MG, Brown JM, Brown J et al (2019) Charting the complexity of the marine microbiome through single-cell genomics. Cell 179:1623–1635

Pachiadaki MG, Sintes E, Bergauer K et al (2017) Major role of nitrite-oxidizing bacteria in dark ocean carbon fixation. Science 358:1046–1051

Palatinszky M, Herbold C, Jehmlich N et al (2015) Cyanate as an energy source for nitrifiers. Nature 524:105–108

Papke RT, Zhaxybayeva O, Feil EJ et al (2007) Searching for species in haloarchaea. Proc Natl Acad Sci 104:14092–14097

Parada AE, Fuhrman JA (2017) Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. ISME J 11:2510–2525

Park J, Kerner A, Burns MA, Lin XN (2011) Microdroplet-enabled highly parallel co-cultivation of microbial communities. PLoS One 6:e17019

Parker CT, Tindall BJ, Garrity GM (2019) International code of nomenclature of prokaryotes. Int J Syst Evol Microbiol 69:S1–S111

Parks DH, Imelfort M, Skennerton CT et al (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055

Parks DH, Rinke C, Chuvochina M et al (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542

Pedrós-Alió C (2012) The rare bacterial biosphere. Annu Rev Mar Sci 4:449–466

Pernice MC, Forn I, Gomes A et al (2015) Global abundance of planktonic heterotrophic protists in the deep ocean. ISME J 9:782–792

Pernthaler A, Pernthaler J, Schattenhofer M, Amann R (2002) Identification of DNA-synthesizing bacterial cells in coastal North Sea plankton. Appl Environ Microbiol 68:5728–5736

Pesant S, Not F, Picheral M et al (2015) Open science resources for the discovery and analysis of *Tara* oceans data. Sci Data 2:150023

Pham VD, Konstantinidis KT, Palden T, DeLong EF (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific subtropical gyre. Environ Microbiol 10:2313–2330

Pinhassi J, DeLong EF, Béjà O, González JM, Pedrós-Alió C (2016) Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. Microbiol Mol Biol Rev 80:929–954

Pommier T, Neal P, Gasol J et al (2010) Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. Aquat Microb Ecol 61:221–233

Pushkarev A, Inoue K, Larom S et al (2018) A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. Nature 558:595–599

Rappé MS, Connon SA, Vergin KL, Giovannoni SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418:630–633

Rappé MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394

Reichart NJ, Jay ZJ, Krukenberg V et al (2020) Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. ISME J 14:2851–2861

Reinthaler T, van Aken HM, Herndl GJ (2010) Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. Deep Sea Res Part II Top Stud Oceanogr 57:1572–1580

Resplandy L, Lévy M, McGillicuddy DJ (2019) Effects of eddy-driven subduction on ocean biological carbon pump. Global Biogeochem Cycles 33:1071–1084

Rinke C, Schwientek P, Sczyrba A et al (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437

Robertson BR, Button DK (1989) Characterizing aquatic bacteria according to population, cell size, and apparent DNA content by flow cytometry. Cytometry 10:70–76

Rosenzweig RF, Sharp RR, Treves DS, Adams J (1994) Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. Genetics 137:903–917

Roux S, Hawley AK, Torres Beltran M et al (2014) Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. elife 3:e03125

Royo-Llonch M, Sánchez P, González JM et al (2020) Ecological and functional capabilities of an uncultured *Kordia* sp. Syst Appl Microbiol 43:126045

Royo-Llonch M, Sánchez P, Ruiz-González C et al (2021) Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. Nat Microbiol. https://doi.org/10.1038/s41564-021-00979-9

Ruiz-González C, Logares R, Sebastián M et al (2019) Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean. Mol Ecol 28:1930–1945

Ruiz-González C, Mestre M, Estrada M et al (2020) Major imprint of surface plankton on deep ocean prokaryotic structure and activity. Mol Ecol 29:1820–1838

Rusch DB, Halpern AL, Sutton G et al (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through eastern tropical Pacific. PLoS Biol 5:e77

Saito MA, Bertrand EM, Duffy ME et al (2019) Progress and challenges in ocean metaproteomics and proposed best practices for data sharing. J Proteome Res 18:1461–1476

Saito MA, McIlvin MR, Moran DM et al (2014) Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. Science 345:1173–1177

Salazar G, Cornejo-Castillo FM, Benítez-Barrios V et al (2016) Global diversity and biogeography of deep-sea pelagic prokaryotes. ISME J 10:596–608

Salazar G, Cornejo-Castillo FM, Borrull E et al (2015) Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. Mol Ecol 24:5692–5706

Salazar G, Paoli L, Alberti A et al (2019) Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. Cell 179:1068–1083

Salter I, Galand PE, Fagervold SK et al (2015) Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest Mediterranean Sea. ISME J 9:347–360

Samo TJ, Smriga S, Malfatti F, Sherwood BP, Azam F (2014) Broad distribution and high proportion of protein synthesis active marine bacteria revealed by click chemistry at the single cell level. Front Mar Sci 1:48

Sánchez-Andrea I, Guedes IA, Hornung B et al (2020) The reductive glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. Nat Commun 11:5090

Sanz-Sáez I, Salazar G, Sánchez P et al (2020) Diversity and distribution of marine heterotrophic bacteria from a large culture collection. BMC Microbiol 20:1–16

Saw JHW, Nunoura T, Hirai M et al (2020) Pangenomics analysis reveals diversification of enzyme families and niche specialization in globally abundant SAR202 bacteria. MBio 11:e02975–e02919

Scholin C, Birch J, Jensen S et al (2017) The quest to develop ecogenomic sensors: a 25-year history of the environmental sample processor (ESP) as a case study. Oceanography 30:100–113

Sebastián M, Estrany M, Ruiz-González C et al (2019) High growth potential of long-term starved deep ocean opportunistic heterotrophic bacteria. Front Microbiol 10:760

Sebastián M, Gasol JM (2019) Visualization is crucial for understanding microbial processes in the ocean. Philos Trans R Soc B Biol Sci 374:20190083

Sebastián M, Smith AF, González JM et al (2016) Lipid remodelling is a widespread strategy in marine heterotrophic bacteria upon phosphorus deficiency. ISME J 10:968–978

Servais P, Casamayor EO, Courties C et al (2003) Activity and diversity of bacterial cells with high and low nucleic acid content. Aq Microb Ecol 33:41–51

Shapiro BJ, Friedman J, Cordero OX et al (2012) Population genomics of early events in the ecological differentiation of bacteria. Science 336:48–51

Shapiro BJ, Polz MF (2015) Microbial Speciation. Cold Spring Harb Perspect Biol 7:a018143

Sharon I, Banfield JF (2013) Genomes from metagenomics. Science 342:1057–1058

Sheik CS, Jain S, Dick GJ (2014) Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. Environ Microbiol 16:304–317

Sherr EB, Sherr BF, Longnecker K (2006) Distribution of bacterial abundance and cell-specific nucleic acid content in the Northeast Pacific Ocean. Deep Sea Res Part I Oceanogr Res Pap 53: 713–725

Shi Y, Tyson GW, Eppley JM, DeLong EF (2011) Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. ISME J 5:999–1013

Singer E, Wagner M, Woyke T (2017) Capturing the genetic makeup of the active microbiome in situ. ISME J 11:1949–1963

Sintes E, Herndl GJ (2006) Quantifying substrate uptake by individual cells of marine bacterioplankton by catalyzed reporter deposition fluorescence in situ hybridization combined with microautoradiography. Appl Environ Microbiol 72:7022–7028

Sogin ML, Morrison HG, Huber JA et al (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci 103:12115–12120

Sosa OA, Repeta DJ, DeLong EF et al (2019) Phosphate-limited ocean regions select for bacterial populations enriched in the carbon–phosphorus lyase pathway for phosphonate degradation. Environ Microbiol 21:2402–2414

Sowell SM, Wilhelm LJ, Norbeck AD et al (2009) Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J 3:93–105

Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Ann Rev Microbiol 39:321–346

Steinberg DK, Carlson CA, Bates NR et al (2001) Overview of the US JGOFS Bermuda Atlantic time-series study (BATS): a decade-scale look at ocean biology and biogeochemistry. Deep Sea Res Part II Top Stud Oceanogr 48:1405–1447

Steindler L, Schwalbach MS, Smith DP et al (2011) Energy starved Candidatus Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. PLoS One 6: e19725

Stepanauskas R (2012) Single cell genomics: an individual look at microbes. Curr Opin Microbiol 15:613–620

Stepanauskas R, Fergusson EA, Brown J et al (2017) Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat Commun 8:84

Stepanauskas R, Sieracki ME (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci 104:9052–9057

Steven B, Hesse C, Soghigian J, Dunbar J (2017) Simulated rRNA/DNA ratios show potential to misclassify active populations as dormant. Appl Environ Microbiol 83:AEM.00696–AEM.00617

Stomp M, Huisman J, de Jongh F et al (2004) Adaptive divergence in pigment composition promotes phytoplankton biodiversity. Nature 432:104–107

Stomp M, van Dijk MA, van Overzee HMJ et al (2008) The timescale of phenotypic plasticity and its impact on competition in fluctuating environments. Am Nat 172:E169–E185

Sul WJ, Oliver TA, Ducklow HW et al (2013) Marine bacteria exhibit a bipolar distribution. Proc Natl Acad Sci 110:2342–2347

Sunagawa S, Acinas SG, Bork P et al (2020) Tara Oceans: towards global ocean ecosystems biology. Nat Rev Microbiol 18:428–445

Sunagawa S, Coelho LP, Chaffron S et al (2015) Structure and function of the global ocean microbiome. Science 348:1261359

Swan BK, Martinez-Garcia M, Preston CM et al (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science 333:1296–1300

Swan BK, Tupper B, Sczyrba A et al (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci 110:11463–11468

Tang K, Yang Y, Lin D et al (2016) Genomic, physiologic, and proteomic insights into metabolic versatility in Roseobacter clade bacteria isolated from deep-sea water. Sci Rep 6:35528

Taubert M, Grob C, Howat AM et al (2017) Methylamine as a nitrogen source for microorganisms from a coastal marine environment. Environ Microbiol 19:2246–2257

Tettelin H, Masignani V, Cieslewicz MJ et al (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci 102:13950–13955

Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 11:472–477

Thompson AW, Foster RA, Krupke A et al (2012) Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. Science 337:1546–1550

Thompson LR, Haroon MF, Shibl AA et al (2019) Red Sea SAR11 and *Prochlorococcus* single-cell genomes reflect globally distributed pangenomes. Appl Environ Microbiol 85:1–18

Thrash JC, Seitz KW, Baker BJ et al (2017) Metabolic roles of uncultivated bacterioplankton lineages in the northern Gulf of Mexico "dead zone". mBio 8:e01017–e01017

Thrash JC, Temperton B, Swan BK et al (2014) Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. ISME J 8:1440–1451

Tragin M, Vaulot D (2018) Green microalgae in marine coastal waters: the ocean sampling day (OSD) dataset. Sci Rep 8:14020

Tripp HJ (2013) The unique metabolism of SAR11 aquatic bacteria. J Microbiol 51:147–153

Tripp HJ, Kitner JB, Schwalbach MS et al (2008) SAR11 marine bacteria require exogenous reduced Sulphur for growth. Nature 452:741–744

Tsementzi D, Wu J, Deutsch S et al (2016) SAR11 bacteria linked to ocean anoxia and nitrogen loss. Nature 536:179–183

Tully BJ (2019) Metabolic diversity within the globally abundant marine group II Euryarchaea offers insight into ecological patterns. Nat Commun 10:271

Tully BJ, Graham ED, Heidelberg JF (2018) The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data 5:170203

Tully BJ, Sachdeva R, Graham ED, Heidelberg JF (2017) 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. PeerJ 5:e3558

Tyson GW, Chapman J, Hugenholtz P et al (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43

Van Mooy BAS, Fredricks HF, Pedler BE et al (2009) Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. Nature 458:69–72

VanInsberghe D, Arevalo P, Chien D, Polz MF (2020) How can microbial population genomics inform community ecology? Philos Trans R Soc B Biol Sci 375:20190253

Vanni C, Schechter MS, Acinas SG et al. (2020) Light into the darkness: unifying the known and unknown coding sequence space in microbiome analyses. bioRxiv 2020.06.30.180448

Venter JC, Remington K, Heidelberg JF et al (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74

Vergin KL, Beszteri B, Monier A et al (2013) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic time-series study site by phylogenetic placement of pyrosequences. ISME J 7:1322–1332

Vila-Costa M, Rinta-Kanto JM, Sun S et al (2010) Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. ISME J 4:1410–1420

Vila-Costa M, Gasol JM, Sharma S, Moran MA (2012) Community analysis of high- and low-nucleic acid-containing bacteria in NW Mediterranean coastal waters using 16S rDNA pyrosequencing. Environ Microbiol 14:1390–1402

Villarino E, Watson JR, Jönsson B et al (2018) Large-Scale Ocean connectivity and planktonic body size. Nat Commun 9:142

Villarreal-Chiu JF, Quinn JP, McGrath JW (2012) The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. Front Microbiol 3:19

Wang Y, Ji Y, Wharfe ES et al (2013) Raman activated cell ejection for isolation of single cells. Anal Chem 85:10697–10701

Whitaker RJ, Grogan DW, Taylor JW (2005) Recombination shapes the natural population structure of the hyperthermophilic archaeon *Sulfolobus islandicus*. Mol Biol Evol 22:2354–2361

Williams PJL (1970) Heterotrophic utilization of dissolved organic compounds in the sea I. Size distribution of population and relationship between respiration and incorporation of growth substrates. J Mar Biol Assoc United Kingdom 50:859–870

Williams PJL (1981) Microbial contribution to overall marine plankton metabolism: direct measurements of respiration. Oceanol Acta 4:359–364

Woyke T, Xie G, Copeland A et al (2009) Assembling the marine metagenome, one cell at a time. PLoS One 4:e5299

Wright JJ, Mewis K, Hanson NW et al (2014) Genomic properties of marine group a bacteria indicate a role in the marine sulfur cycle. ISME J 8:455–468

Wrighton KC, Thomas BC, Sharon I et al (2012) Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337:1661–1665

Wuchter C, Abbas B, Coolen MJL et al (2006) Archaeal nitrification in the ocean. Proc Natl Acad Sci 103:12317–12322

Yakimov MM, La Cono V, Smedile F et al (2011) Contribution of crenarchaeal autotrophic ammonia oxidizers to the dark primary production in Tyrrhenian deep waters (Central Mediterranean Sea). ISME J 5:945–961

Yilmaz P, Yarza P, Rapp JZ, Glöckner FO (2016) Expanding the world of marine bacterial and archaeal clades. Front Microbiol 6:1524

Yoon HS, Price DC, Stepanauskas R et al (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science 332:714–717

Yu FB, Blainey PC, Schulz F et al (2017) Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples. elife 6:e26580

Zavaleta-Pastor M, Sohlenkamp C, Gao JL et al (2010) Sinorhizobium meliloti phospholipase C required for lipid remodeling during phosphorus limitation. Proc Natl Acad Sci 107:302–307

Zeng Y, Chen X, Madsen AM et al (2020) Potential rhodopsin-and bacteriochlorophyll-based dual phototrophy in a high Arctic glacier. MBio 11:e02641–e02620

Zhang Q, Zhang P, Gou H et al (2015) Towards high-throughput microfluidic Raman-activated cell sorting. Analyst 140:6163–6174

Zhang Y, Zhao Z, Dai M, Jiao N, Herndl GJ (2014) Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. Mol Ecol 23:2260–2274

Zinger L, Amaral-Zettler LA, Fuhrman JA et al (2011) Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. PLoS One 6:e24570