Chapter 12

A Global Census of Marine Microbes

Linda Amaral-Zettler¹, Luis Felipe Artigas², John Baross³, Loka Bharathi P.A.⁴, Antje Boetius⁵, Dorairajasingam Chandramohan⁶, Gerhard Herndlⁿ, Kazuhiro Kogure⁶, Phillip Neal¹, Carlos Pedrós-Alió⁶, Alban Ramette⁶, Stefan Schouten⁷, Lucas Stal¹⁰, Anne Thessen¹, Jan de Leeuw⁷, Mitchell Sogin¹

¹Marine Biological Laboratory, Woods Hole, Massachusetts, USA

²Université Lille Nord de France, Université du Littoral Côte d'Opale, CNRS UMR 8187 LOG, Wimereux, France

³University of Washington, Seattle, Washington, USA

⁴Microbiology Laboratory, National Institute of Oceanography, Dona Paula, Goa, India

⁵Max Planck Institute for Marine Microbiology, Bremen, Germany

⁶Ratnapuri Colony, J.N. Salai, Koyambedu, Chennai, India

⁷The Royal Netherlands Institute for Sea Research, Den Burg, Texel, The Netherlands

8Ocean Research Institute, University of Tokyo, Tokyo, Japan

⁹Institut de Ciències del Mar, Barcelona, Spain

¹⁰Netherlands Institute of Ecology, Yerseke, The Netherlands

12.1 Introduction

12.1.1 Importance

The oceans abound with single cells that are invisible to the unaided eye, encompassing all three domains of life – Bacteria, Archaea, and Eukarya – in a single drop of water or a gram of sediment (Figs. 12.1A, B, C, and D). The microbial world accounted for all known forms of life for more than 80% of Earth's history. Today, microbes continue to dominate every corner of our biosphere, especially in the ocean where they might account for as much as 90% of the total biomass (Fuhrman et al. 1989; Whitman et al. 1998). Even the most seemingly inhospitable marine environments

host a rich diversity of microbial life (Figs. 12.1E and H). For the past six years, microbial oceanographers from around the world have joined the effort of the International Census of Marine Microbes (ICoMM, Box 12.1) to explore this vast diversity. In this chapter we provide a brief history of what is known about marine microbial diversity, summarize our achievements in performing a global census of marine microbes, and reflect on the questions and priorities for the future of the marine microbial census.

From the time of their origins, single-cell organisms – initially anaerobic and later aerobic – have served as essential catalysts for all of the chemical reactions within biogeochemical cycles that shape planetary change and habitability. Marine microbes carry out half of the primary production on the planet (Field *et al.* 1998). Microbial carbon re-mineralization, with and without oxygen, maintains the carbon cycle. Microbes account for more than 95% of the respiration in the oceans (Del Giorgio & Duarte 2002). They control global utilization of nitrogen through

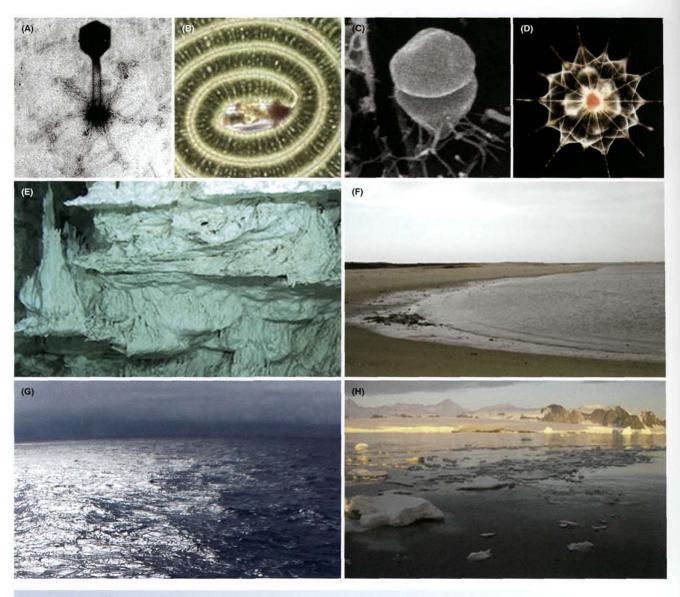


Fig. 12.1

Microbial life spans all three domains of life inclusive of Bacteria, Archaea, and Eukarya and their associated viruses. This collage shows examples of the types of marine microbes and diverse habitats included in the microbial census. Photograph credits are given in parentheses. From the leftmost panel, (A) a Synechococcus phage (John Waterbury), (B) filaments of the marine cyanobacterium Lyngbya (David Patterson, used under license), (C) the hyperthermophilic archaeon "GR1" (Melanie Holland), and (D) a single-celled eukaryote called an acantharian (Linda Amaral-Zettler, used under license). Examples of diverse environments sampled as part of the microbial census include the following: (E) the Lost City Hydrothermal Vent flange actively venting heated hydrogen and methane rich fluids, (IFE, URI-IAO, UW, Lost City science party, and NOAA); (F) the sandy coastline from the North Sea island Sylt (Angélique Gobet); (G) the open ocean waters of the South Pacific Ocean (Katsumi Tsukamoto), and (H) the waters off the Antarctic Peninsula (Hugh Ducklow).

N₂ fixation, nitrification, nitrate reduction, and denitrification, and drive the bulk of sulfur, iron, and manganese biogeochemical cycles (Kirchman 2008; Whitman *et al.* 1998). Marine microbes regulate the composition of the atmosphere, influence climate, recycle nutrients, and decompose pollutants. Without microbes, multicellular

animals on Earth would not have evolved or persisted over the past 500 million years.

Measuring microbial diversity in a broad range of marine ecosystems (see, for example, Figs. 12.1E, F, G, and H) will facilitate quantification of the magnitude and dynamics of the microbial world and its stability through space and

Box 12.1

A Brief History of ICoMM

ICoMM is one of 14 Census of Marine Life ocean realm projects that explores the diversity, distribution, and abundance of microbial life in the oceans. ICoMM's leadership represents a collaborative effort between the Royal Netherlands Institute for Sea Research (NIOZ), in Texel, The Netherlands, and the Marine Biological Laboratory (MBL) in Woods Hole, MA, USA. Collectively ICoMM has provided a means to galvanize the microbial oceanographic community in conducting a global census of marine microorganisms. The goal of ICoMM is to determine the range of genetic diversity and relative numbers of different microbial organisms at sampling sites throughout the world's oceans.

Since 2004, ICoMM has provided support for training workshops and meetings including five primary working groups (Benthic, Open Ocean and Coastal Systems, Technology, Informatics and Data Management, and Microbial Eukaryotes), and its Scientific Advisory Council that engage the international community of marine microbiologists. In 2006, ICoMM served as the coordinating body that helped to secure funding from the W. M. Keck Foundation for a 454 DNA pyrosequencing system dedicated to DNA tag sequencing projects. Additional information about ICoMM's membership, scope and activities can be found on the ICoMM website: icomm.mbl.edu.

time. The phylogenetic and physiological diversities of microbes are considerably greater than those of animals and plants, and microbial interactions with other life-forms are correspondingly more complex (Pace 1997). Measuring marine microbial diversity and determining corresponding associated functions will thus provide a wealth of information about specific microbial processes of great significance such as wastewater treatment, industrial chemical production, pharmaceutical production, bioremediation, and global warming. Examining the relationships between microbial populations and whole communities within their dynamic environment will allow us to formulate better the definition of what constitutes an ecologically relevant species in the microbial world. Molecular methods rely upon measures of genetic similarity to describe operational taxonomic units (OTUs). Statistical treatments can use the relative number of distinct OTUs to estimate diversity, but these inferences do not translate directly into numbers of microbial species. Microbiologists have not reached consensus on the definition of microbial species using either molecular or phenotypic approaches. However, ecological concepts of microbial species based upon molecular data will inform theoretical applications and guide solutions to major challenges facing science and human society.

12.1.2 Microbial diversity and abundance

The reliance upon traditional cultivation and staining techniques led to gross underestimates of microbial abundance and species richness in both oceanic and terrestrial

environments (Jannasch & Jones 1959; Zimmermann & Mever-Reil 1974; Hobbie et al. 1977) (Fig. 12.2). The application of fluorescence-based microscopy coupled with DNA staining methods revealed the great "plate count anomaly", which posits that microbiologists have underestimated microbial abundances by at least three orders of magnitude. Instead of a mere 100 cells per milliliter of seawater, nucleic-acid staining technology showed the number of bacteria in the open ocean exceeds 10²⁹ cells, with average cell concentrations of 106 per milliliter of seawater (Whitman et al. 1998). In marine surface sediments, cell abundances are 108-109 per gram, and even in the greatest depths of the subsurface seabed, more than 105 cells per gram are encountered (Jørgensen & Boetius 2007). The ocean also hosts the densest accumulations of microbes known on Earth, reaching 1012 cells per milliliter, like the photosynthetic mats thriving in hypersaline environments, and the methanotrophic mats of anoxic seas, resembling ancient microbial assemblages before the advent of eukaryote grazers (Knittel & Boetius 2009). Archaeal cell abundances rival those of bacteria in certain parts of the ocean and the seabed, and microbial eukaryotic (protistan) densities vary widely from tens of cells per liter to bloom conditions that can surpass 106 cells per milliliter of seawater.

As of 2010, cultivation-based studies have described more than 10,000 bacterial and archaeal species (http://www.bacterio.cict.fr/number.html) and an estimated 200,000 protistan species (Corliss 1984; Lee *et al.* 1985; Patterson 1999; Andersen *et al.* 2006). Cultivation-independent studies that rely upon molecular methods such

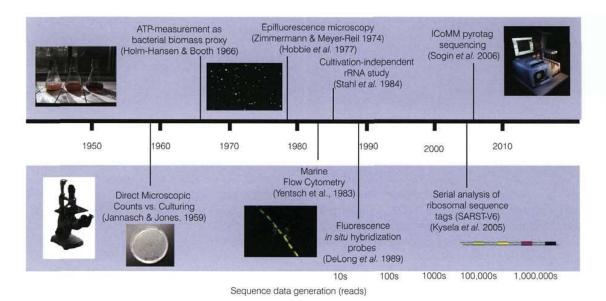


Fig. 12.2

A timeline showing milestones in advances in technology that have enabled the microbial census (Jannasch & Jones 1959; Holm-Hansen & Booth 1966; Zimmermann & Meyer-Reil 1974; Hobbie et al. 1977; Yentsch et al. 1983; Stahl et al. 1984; DeLong et al. 1989; Kysela et al. 2005; Sogin et al. 2006). Upper right photograph by Tom Kleindinst, Woods Hole Oceanographic Institution.

as the sequencing of 16S ribosomal RNA (rRNA) genes show microbial diversity to be approximately 100 times greater (Pace 1997). With each new molecular survey, this window on the microbial world has increased in size.

12.2 Challenges of a Microbial Census

The ocean covers 70% of the Earth's surface (an estimated volume of about $2 \times 10^{18} \, \mathrm{m}^3$) and has an average depth of 3,800 m. Strategies for conducting a census must consider the enormous geographical area to be surveyed, an almost unimaginable number of cells, and the impact of spatial gradients and temporal shifts on microbial assemblages. In fact, before ICoMM, little was known about global patterns in microbial communities. Basic questions such as "is there a distinct difference between pelagic and benthic microbial communities?" or "what is the temporal turnover in microbial cells between two sampling dates?" profoundly influenced our sampling strategies.

Contemporary molecular approaches typically use rRNA sequences as proxies for the occurrence of different microbial genomes in an environmental DNA sample (coding regions for functional genes can also provide information about microbial population structures). However, the expense of conventional DNA sequencing has constrained the number of homologous sequences that microbial

ecologists typically collect to describe community composition. Relative to the number of microbes in most samples, these surveys superficially describe microbial community structures. There are more than 10⁸ microorganisms in a liter of seawater or a gram of soil (Whitman *et al.* 1998). Few studies collect more than 10⁴ sequences, which correspond to 0.01% of the cells in a liter of seawater or a gram of soil. The detection of organisms that correspond to the most abundant OTUs or species equivalents requires minimal molecular sampling, whereas the recovery of sequences from rare taxa that constitute the "long tail" of low abundance organisms in taxon rank–abundance curves demands surveys that are orders of magnitude larger.

As an alternative to analyzing nearly full-length polymerase chain reaction (PCR) amplicons of rRNA genes from environmental DNA samples, short sequence tags from hypervariable regions in rRNAs (pyrotags) can provide measures of diversity (species or OTU richness) and relative abundance (evenness) of OTUs in microbial communities. When combined with the massively parallel capacity of "next generation" DNA sequencing technology that allows for the simultaneous sequencing of hundreds of thousands of templates (Margulies et al. 2005), it becomes possible to increase the number of sampled gene sequences in an environmental survey by orders of magnitude (Huber et al. 2007; Sogin et al. 2006). Enumerating the number of different rRNA pyrotags provides a first-order description of the relative occurrence of specific microbes in a population. The highly variable nature of the tag sequences and paucity

of positions do not allow direct inference of phylogenetic frameworks. However, when tag sequences are queried against a comprehensive reference database of hypervariable regions within the context of full-length sequences, it is possible to extract information about taxonomic identity and microbial diversity. Initial tests of this innovative technology examined the microbial population structures of samples from the meso- and bathypelagic realm of the North Atlantic Deep Water Flow and two diffuse flow samples from Axial Seamount on the Juan de Fuca ridge off the west coast of the United States (Sogin et al. 2006). These initial data sets led ICoMM investigators to the discovery of the "rare biosphere", a rich diversity of novel, low-abundance populations and dormant or slow growing microbes. A single liter of seawater, on average containing 10⁸–10⁹ bacteria, represents about 20,000 "species" of bacteria, a number that is one or two orders of magnitude higher than estimated earlier (Venter et al. 2004). When plotted on a two dimensional x-y microbial rank distribution diagram, this species-richness shows an extraordinarily long tail, the long tail including low-abundance taxa, many of which represent types of microbes that have never been seen before. Huber et al. (2007) extended this approach to the Archaea, also targeting the V6 16S rRNA hypervariable region and reported species richness estimates to be on the order of 3,000 "species" per liter of seawater. Amaral-Zettler et al. (2009) developed a tag sequencing strategy

for the V9 hypervariable region of the 18S rRNA gene in eukaryotes and determined that estimates of microbial eukaryotic (protist) species richness can be on the order of magnitude seen in the archaeal domain but may be an order of magnitude lower in more extreme environments such as Antarctic waters.

The International Census of Marine Microbes subsequently adopted this pyrotag strategy in a coordinated microbial census of samples from globally distributed marine environments. A study of lipid molecular structures from marine microbes complements the pyrotag survey. The database MICROBIS (http://icomm.mbl.edu/microbis) serves information to ICoMM, and its website provides access to this information including the capacity to retrieve contextual data information for all samples (Fig. 12.3).

The database VAMPS (Visualization Analysis of Microbial Population Structures, http://vamps.mbl.edu) and its links to MICROBIS provide full access to the pyrotag sequences, the contextual data, analytical and graphical tools for comparing microbial population structures for different sites, search tools for locating sequences in each of our samples, descriptions of community composition at taxonomic ranks of phyla, class, order, family, or, when possible, genus for all samples, and rarefaction and diversity analyses for all of ICoMM's data. Figure 12.4 depicts the geospatial breadth of pyrotag and lipid data for this global study of microbes in the world's oceans. It includes

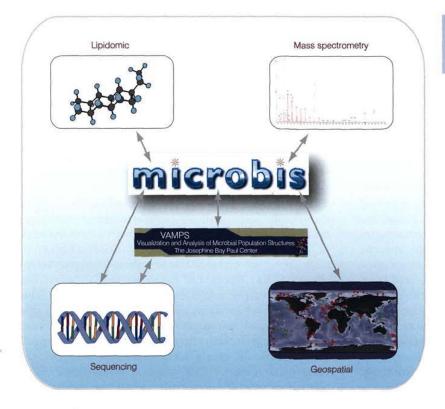
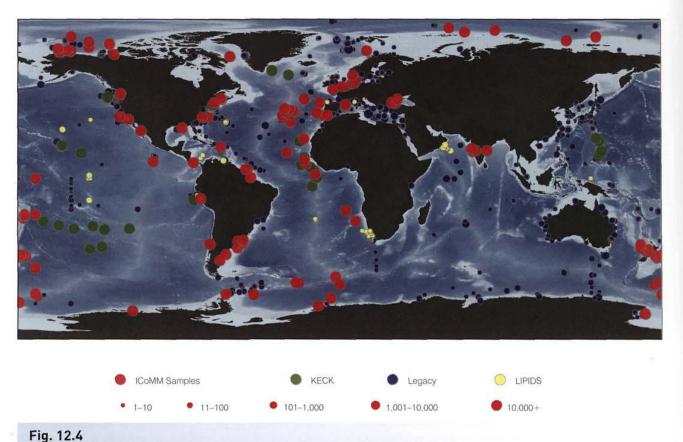


Fig. 12.3

An overview of MICROBIS and its relationship to VAMPS and the microbial lipid database.



The global distribution of observations gathered/recorded in the MICROBIS database. These include 454-pyrosequenced DNA pyrotag data generated during the ICoMM community sequencing project (red) and the Keck core sequencing projects (green), as well as legacy molecular data observations compiled from the literature (blue) and lipid-based analyses (yellow). The diameter of the circle represents the log₁₀ of the sample size.

a subset of more than 18 million DNA sequence reads distributed among 583 bacterial, 120 archaeal, and 59 eukaryotic datasets from a larger dataset of >25 million sequences from >1,200 samples. The samples represent all major oceanic systems including the Atlantic, Pacific, Arctic, Southern, and Indian Oceans, and sediment and water samples from estuaries to deep-water environments including vents and seeps, seamounts, corals, sponges, microbial mats and biofilms, and polar regimes. Table 12.1 describes the origin of samples, targeted domains, project descriptions, and relevance to other Census ocean-realm projects. Here we present a broad-brush synthesis of our data emphasizing the most abundant pyrotags recovered from our surveys. Although a comprehensive synthesis of these data lies beyond the scope of this chapter, Figures 12.5, 12.6, 12.7 and 12.8 and the highlights that follow offer a glimpse into novel insights that will soon emerge from this international study of microbial community structures of the world's oceans. More detailed meta-analyses will frame the bulk of ICoMM's working groups during 2010.

12.3 Highlights of ICoMM Investigations

12.3.1 The "Abundant Biosphere"

The pie charts in Figures 12.5, 12.6, and 12.7 summarize the most abundant tags in our bacterial, archaeal, and microeukaryotic datasets respectively. As expected from the work of S.J. Giovannoni in the Sargasso Sea (Giovannoni *et al.* 1990), pyrotags corresponding to α -Proteobacteria and specifically SAR11 represented the most abundant organisms (primarily in planktonic samples) in our global survey. This heterotrophic α -Proteobacterial lineage plays a critical role in the cycling of carbon, nitrogen, and sulfur and accounts for approximately 25% of the biomass and 50% of the cell abundance in the ocean. More recently, researchers discovered that members of this group of bacteria contain proteorhodopsin, which potentially enables the harvesting of

Table 12.1
ICoMM microbial population structures of the world's oceans projects

| Primary Investigator | PI First Name | Code | Project description | Domain | Examples of relevant projects |
|----------------------|---------------|------|------------------------------|--------|-------------------------------|
| Aguiar | Paula | ASV | Azorean Shallow Vents | В | ChEss/MAR-ECO |
| Amaral-Zettler | Linda | мнв | Mount Hope Bay | BE | |
| Andersen | Robert | SAB | Surreptitious Algal Bacteria | В | |
| Artigas | Felipe | LCR | LaCAR Cooperative Run | BAE | NaGISA/CMarZ |
| Bharathi | Loka | ICR | Indian Ocean Cooperative Run | В | COMARGE/CeDAMar |
| Bertilsson | Stefan | BSP | Baltic Sea Proper | В | |
| Bolhuis | Henk | СММ | Coastal Microbial Mats | ВА | NaGISA |
| Brazelton | William | LCY | Lost City | BA | ChEss |
| Caron | David | GPS | Global Protist Survey | E | CAML/CMarZ |
| Chistoserdov | Andrei | CAR | Cariaco Basin | В | COMARGE |
| Chistoserdov/Artigas | Andrei/Felipe | AGW | Amazon-Guianas Water | В | NaGISA |
| Coolen | Marco | WBS | Black Sea | E | HMAP/CMarZ |
| Dennett | Mark | DOF | Deep Ocean Flux | E | HMAP/CMarZ |
| D'Hondt | Steven | KNX | Station KNOX | BA | |
| Epstein | Slava | SSD | Spatial Scaling Diversity | В | NaGISA |
| Franklin | Rima | AOT | Atlantic Ocean Transect | В | MAR-ECO |
| Gaidos | Eric | CRS | Coral Reef Sediment | BA | CReefs |
| Gallardo | Victor | VAG | Humboldt Marine Ecosystem | В | COMARGE/ChEss |
| Gerdts | Gunnar | MPI | Helgoland | В | |
| Gilbert | Jack | PML | English Channel | В | |
| Hamasaki | Koji | ABR | Active but Rare | BA | CAML |
| Herndl | Gerhard | NADW | North Atlantic Deep Water | BA | |
| Huber | Julie | EEL | Eel Pond Winter Pilot Study | В | |
| Huber | Julie | SMT | Seamounts | BA | ChEss, CenSeam |
| Kirchman | David | ACB | Arctic Chukchi Beaufort | BA | ArcOD |
| Lovejoy | Connie | DAO | Deep Arctic Ocean | ВА | ArcOD |
| Maas | Els | NZS | New Zealand Sediment | В | |
| Martins | Ana | AWP | Azores Waters Project | BAE | MAR-ECO/CMarZ |
| Murray | Alison | CAM | Census Antarctic Marine | В | CAML |
| Pawlowski | Jan | DSE | Deep Sea Eukarya | E | CeDAMar/CMarZ |

| Primary Investigator | PI First Name | Code | Project description | Domain | Examples of relevant projects |
|----------------------|---------------|------|-------------------------------------|--------|-------------------------------|
| Polz | Martin | CNE | Coastal New England | В | NaGISA |
| Pommier | Thomas | ВМО | Blanes Bay Microbial Observatory | В | |
| Post | Anton | GOA | Gulf of Aqaba | В | |
| Prosser | James | SMS | Station M Sediments | В | CeDAMar |
| Ramette | Alban | FIS | Frisian Island Sylt | В | NaGISA |
| Rappé | Michael | нот | Hawaii Ocean Time Series | В | |
| Reysenbach | Anna-Louise | ALR | Lau Hydrothermal Vent | BA | ChEss |
| Rocap | Gabrielle | HCW | Hood Canal Washington | В | NaGISA |
| Rooney-Varga | Juliette | JRV | Gulf of Maine | E | GoMA/CMarZ |
| Sogin | Mitchell | LSM | Little Sippewissett Salt Marsh | В | NaGISA |
| Staley | James | BSR | Black Sea Redox | В | |
| Stoeck | Thorsten | APP | Anaerobic Protist Project | E | COMARGE/CMarZ |
| Sunagawa | Shinichi | CCB | Caribbean Coral Bacteria | BA | CReefs |
| Teske | Andreas | GMS | Guaymas Methane Seeps | BA | ChEss |
| Teske | Andreas | ODP | Ocean Drilling Project | BA | CeDAMar |
| Wagner | Michael | SP0 | Sponges | В | CReefs |
| Webster | Gordon | CFU | Deep Subseafloor Sediment | BA | CeDAMar |
| Yager | Patricia | ASA | Amundsen Sea Antarctica | В | CAML |

B, Bacteria; A, Archaea; E, Eukarya.

energy from light (Fuhrman et al. 2008; Giovannoni et al. 2005). The presence of this clade in different habitats (Fig. 12.8) including coastal waters, seamounts, polar waters, and the open ocean (not shown) reflects its ubiquity in the marine pelagic environment. The 20 most abundant tags in our bacterial analyses also include members of the Rhodobacteraceae. One member of this group, Roseobacter sp., is cultivable by adding extracts of algal secreted organic matter to the medium (Mayali et al. 2008). The worldwide association of Roseobacter with algal blooms suggests it has a role in controlling bloom outbreaks.

The most abundant tag sequence derived from a photosynthetic bacterium belonged to a member of the Prochlorales (Cyanobacteria) and shares 100% V6 rRNA region sequence identity with the cultivar *Prochlorococcus marinus*. The picocyanobacteria (smaller than 2 μm) *Prochlorococcus* spp. along with *Synechococcus* spp. dominate the oceans with cell numbers of up to 10⁵–10⁶ per milliliter (Heywood *et al.* 2006; Scanlan *et al.* 2009). Collectively they contribute up to 50% of oceanic primary production (Li 1994). Cyanobacteria represent an ancient group of organisms. These inventors of oxygenic photosynthesis drove the oxygenation of the Earth's atmosphere 2.5 billion years ago. The evolution of aerobic Bacteria and Archaea made possible the origins of plants and animals about 0.5 billion years ago when the oxygen concentration in the atmosphere reached its present-day level. Today, Cyanobacteria produce about 50% of the oxygen on Earth. Most Cyanobacteria occur in marine communities (Garcia-Pichel *et al.* 2003).

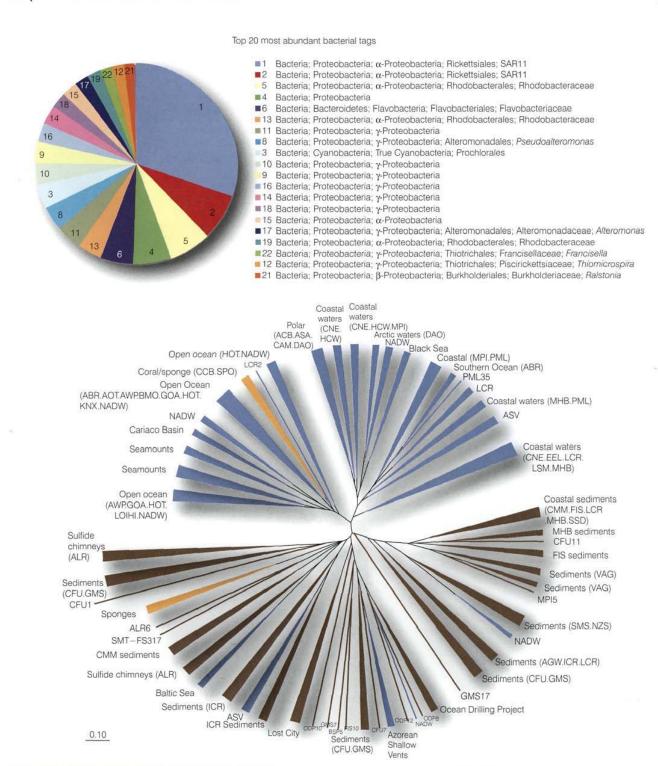
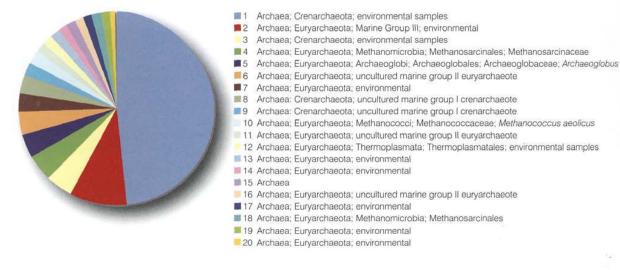


Fig. 12.5

A summary of results from pyrotag bacterial projects. Top, the taxonomic breakdown of the top 20 most abundant bacterial sequences found across 583 bacterial datasets. The rankings are based on the sum of the relative abundances of individual sequences from each sample. Taxonomies are based on the Global Alignment for Sequence Taxonomy (GAST) procedure (Huse *et al.* 2008). The numbering has been adjusted to match the tag sequence numbering in Figure 12.8 and is ordered in descending order of abundance. Bottom, a radial dendrogram of clustered bacterial datasets. Clusters are based on similarity calculations of presence/absence data of the most abundant pyrotag sequences. Brown, benthic samples; blue, water-column samples; orange, sponge- or coral-associated samples. See Table 12.1 for descriptions of project abbreviations.



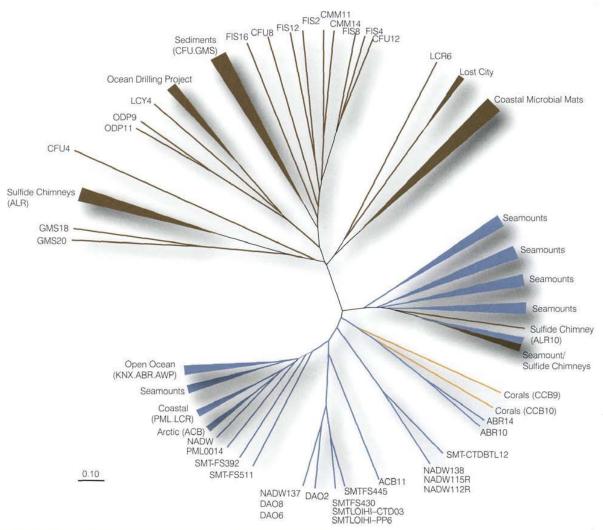
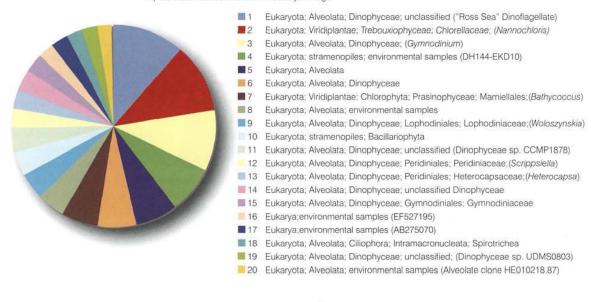


Fig. 12.6

A summary of results from pyrotag archaeal projects. Top, the taxonomic breakdown of the top 20 most abundant archaeal tag sequences found in 120 archaeal datasets. The rankings are based on the sum of the relative abundances of individual sequences from each sample. Taxonomies are based on the GAST procedure. Bottom, a radial dendrogram of clustered archaeal datasets. Clusters are based on similarity calculations of presence/absence data of the most abundant tag sequences. Brown, benthic samples; blue, water column samples; orange, coral-associated samples. See Table 12.1 for descriptions of project abbreviations.



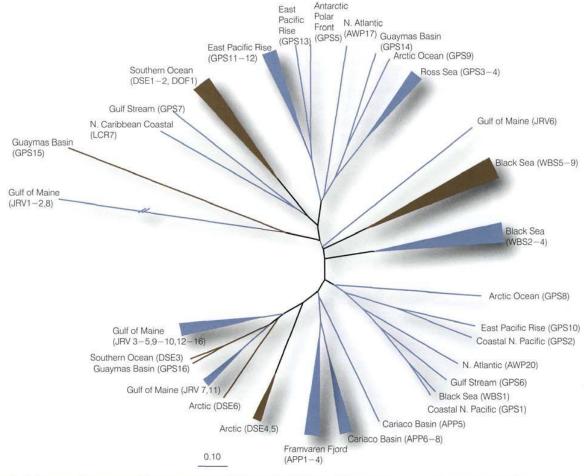


Fig. 12.7

A summary of eukaryotic pyrotag projects. Top, the taxonomic breakdown of the top 20 most abundant microbial eukaryotic tag sequences found in 59 eukaryotic datasets. All tags with metazoan-associated taxonomy have been removed from the analysis. The rankings are based on the sum of the relative abundances of individual sequences from each sample. Taxonomies are based on a combination of the GAST procedure and BLAST. Tags with a 100% match to a representative sequence/taxon in GenBank have that representative affiliation in parentheses. Bottom, a radial dendrogram of clustered eukaryotic datasets. Clusters are based on similarity calculations of presence/absence data of the most abundant tag sequences. Brown, benthic samples; blue, water column samples. See Table 12.1 for descriptions of project abbreviations.

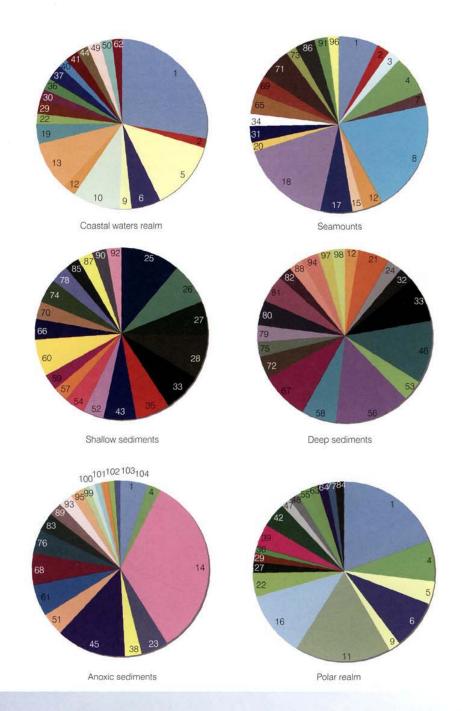


Fig. 12.8

A summary of the major ocean realms sampled showing the top 20 most abundant bacterial tag sequences for each habitat. Realms (abbreviations from Table 12.1) shown include Coastal Waters (CNE, HCW, LCR, MPI, PML, EEL, LSM, MHB), Seamounts (SMT), Shallow Sediments (AGW, CMM, CRS, FIS, GMS, ICR, LCR, SSD, VAG, MHB), Deep Sediments (CFU, ICR, NZS, SMS, ODP), Anoxic Sediments (BSR, CAR), and Polar Regions (ABR, ACB, ASA, CAM, DAO). Numbers facilitate comparisons between samples. The lowest possible taxonomic rank assigned for each tag follows the number designation: (1) SAR11; (2) SAR11; (3) Prochlorales; (4) Proteobacteria; (5) Rhodobacteraceae; (6) Flavobacteriaceae; (7) Sulfurovum; (8) Pseudoalteromonas; (9) y-Proteobacteria: (10) γ-Proteobacteria; (11) γ-Proteobacteria; (12) *Thiomicrospira*; (13) Rhodobacteraceae; (14) γ-Proteobacteria; (15) α-Proteobacteria; (16) γ-Proteobacteria; (17) Alteromonas; (18) γ-Proteobacteria; (19) Rhodobacteraceae; (20) α-Proteobacteria; (21) Ralstonia; (22) Francisella; (23) Actinobacteria; (24) γ-Proteobacteria; (25) Bacillus; (26) Clostridium; (27) Flavobacteriaceae; (28) γ-Proteobacteria; (29) Flavobacteriaceae; (30) α-Proteobacteria; (31) Alteromonas; (32) Ectothiorhodospiraceae; (33) γ-Proteobacteria; (34) Marinobacter; (35) Clostridium; (36) Methylophilus; (37) Roseovarius; (38) Pseudomonas; (39) \(\gamma\)-Proteobacteria; (40) Flavobacteriaceae; (41) Rhodobacteraceae; (42) \(\gamma\)-Proteobacteria; (43) Bacillus; (44) Glaciecola; (45) Bacteria; (46) Erythrobacter; (47) Flavobacteriaceae; (48) α-Proteobacteria; (49) γ-Proteobacteria; (50) Rhodospirillales; (51) Proteobacteria; (52) Paenibacillus; (53) Diaphorobacter; (54) Bacillus; (55) γ-Proteobacteria; (56) JS1; (57) Clostridium; (58) γ-Proteobacteria; (59) Tepidibacter; (60) Flavobacteriaceae; (61) Bacteria; (62) γ-Proteobacteria; (63) Bacteria; (64) Sulfitobacter; (65) Pseudomonas; (66) Methylophaga; (67) Actinobacteria; (68) Bacteria; (69) Thiomicrospira; (70) Bacillus; (71) Burkholderia; (72) Gemmatimonadetes; (73) Comamonadaceae; (74) γ-Proteobacteria; (75) γ-Proteobacteria; (76) Bacteria; (77) Flavobacteriaceae; (78) γ-Proteobacteria; (79) γ-Proteobacteria; (80) Hyphomicrobium; (81) Clostridia; (82) γ-Proteobacteria; (83) γ-Proteobacteria; (84) Methylophaga; (85) Desulfosarcina; (86) Caminibacter; (87) Flavobacteriales; (88) γ-Proteobacteria; (89) Bacteria; (90) Chromatiales; (91) Thioreductor; (92) Desulfobulbaceae; (93) γ-Proteobacteria; (94) Dehalococcoidetes; (95) Desulfobacterium; (96) Thioreductor micantisoli; (97) γ-Proteobacteria; (98) γ-Proteobacteria; (99) Proteobacteria; (100) γ-Proteobacteria; (101) Deferribacteres; (102) γ-Proteobacteria; (103) Bacteria; (104) Deferribacteres.

Members of the phylum Crenarchaeota dominated the Archaeal pyrotag surveys. Microbial ecologists originally thought that all Crenarchaea represented extremophiles, until the discovery of their ubiquity in everyday marine and terrestrial environments (DeLong 1992; Fuhrman et al. 1992; Simon et al. 2000). Crenarchaeotal abundances can exceed bacterial abundances below 100 m depth in the ocean where they are metabolically active and can contribute to the oceanic carbon cycle (Herndl et al. 2005). In the Arctic, Marine Group III Euryarchaeota can dominate the deep water masses such as the deep Atlantic Layer in the central Arctic Ocean (Galand et al. 2009b). Sequences related to this group represented the second most commonly encountered pyrotags in our global archaeal dataset. In many cases, we only detected other abundant archaeal pyrotags in specific samples. For example, the methanogens Methanosarcinales occurred primarily in Lost City Hydrothermal Vents (see below), whereas Archaeglobus- and Methanococcus-related tags specifically associated with sulfide chimneys.

The pyrotag studies showed that dinoflagellates dominated most eukaryotic microbial communities. Members of the dinoflagellates include phototropic and heterotrophic representatives and many co-occur with or may be responsible for harmful algal blooms, making them commercially and ecologically important. The high frequency of dinoflagellate tags likely reflects bias introduced by the very large copy number of rRNA genes in the genomes of most dinoflagellates (as many as 12,000 copies in species such as Akashiwo sanguinea (Zhu et al. 2005)). Indeed, many of the tags recovered among the top 20 most abundant microbial eukaryotes included members of the picoeukaryotes (0.2-2 µm in size) that numerically dominate, but tend to have lower copy numbers of their 18S rRNA genes. The diversity of these Lilliputians of the protist world was only first recognized at the beginning of the twenty-first century (Díez et al. 2001; López-García et al. 2001; Moon-van de Staay et al. 2001).

The top most abundant tag among our eukaryotic datasets displayed 100% identity with the sequence of an unclassified dinoflagellate within the *Karenia/Karlodinium* group that Gast et al. (2006) first identified in the Ross Sea, Antarctica. In some cases these cells occur at densities up to 29,000 cells per liter. Equally intriguing, our global analyses of eukaryotic tags revealed this pyrotag also occurs in the Arctic, Pacific, and Atlantic Oceans (from the Caribbean to the Gulf of Maine), the Framvaren Fjord in Norway and the Black Sea. Whether this tag represents the same cosmopolitan species or closely related ecotypes that extend over the globe remains unknown.

Because of differences in their gene copy number in different taxa, the relative abundance of eukaryotic pyrotags does not reflect the number of cells in a sampled environment. However, these data provide important taxonomic information at the species level for many morphologically rich eukaryotic microbes including dinoflagellates (for example *Scrippsiella*, *Heterocapsa*, *Woloszynskia*) and for members of the picoeukaryotes such as *Bathycoc*cus that are harder to distinguish morphologically.

12.3.2 The "Rare Biosphere"

In the microscopic realm, ICoMM's sampling of many diverse marine ecosystems has reinforced the concept of a ubiquitous rare biosphere (Pedrós-Alió 2006; Sogin et al. 2006). This forces us to reconsider the potential feedback mechanisms between shifts in extremely complex microbial communities and global change, as well as how microbial communities and the genomes of their constituents change over evolutionary timescales. Minor population members may serve as functional keystone species in microbial consortia, or they might be the products of historical ecological change with the potential to become dominant in response to shifts in environmental conditions, for example when local or global change favors their growth. The absence of information about the global distribution of members of the rare biosphere makes it impossible to ascertain if they represent specific biogeographical distributions of bacterial taxa, functional selection by particular environments, or cosmopolitan distribution of all microbial taxa - the "everything is everywhere" hypothesis. Data from recent pyrosequencing efforts, however, are beginning to shed light on this topic. Galand et al. (2009a), for example, compared pyrotags from Arctic Ocean samples. These samples included both surface and deep waters, as well as winter and summer samples from different locations. When they clustered the samples, they observed nearly identical patterns for all abundant sequences (more than 1% of all tags), or only rare sequences (less than 0.01% of all tags). This indicates that in this system, the rare OTUs have the same biogeography as the abundant OTUs. This opens up two possibilities: either the pyrosequencing approach is only targeting the most abundant of the rare biosphere and the actual tails are even longer; or the rare OTUs must be a dynamic lot, able to grow and experience losses due to predation and viral attack in some intriguing and unknown way.

To explore whether high-abundance taxa represent physiologically active populations whereas low-abundance pyrotags represent less active or dormant microbes, Hamasaki et al. (2007) applied the bromodeoxyuridine (BrdU, thymidine analogs for detecting de novo DNA synthesis) magnetic bead immunocapture method to examine both the abundant and rare members of the microbial community. They applied this technique to surface seawater samples from four stations along a north–south transect in the South Pacific taken from November 2004 to March 2005 during the KH-04-5 cruise of the R/V Hakuho-maru (Ocean Research Institute, the University of Tokyo and

JAMSTEC) (Fig. 12.1G). They incubated subsamples on board after adding BrdU and then compared the pyrotag microbial community structures in treated and untreated samples. Their results indicated that some high-abundance taxa incorporated BrdU whereas others did not, and some rare taxa represented only by singletons in the resulting tag dataset also took up BrdU. More importantly, some BrdU-labelled taxa were detected only in the BrdU-labelled fraction but were absent in the untreated fractions. These results suggested that the rare biosphere is not restricted to physiologically inactive populations but can include taxa involved in biogeochemical cycles. This study further illustrates that there may be dynamic exchange between abundant and rare microbial populations.

12.3.3 High archaeal microdiversity in the Lost City Hydrothermal Field

The Lost City Hydrothermal Field on the Mid-Atlantic Ridge is the first deep-sea environment discovered where exothermic water-rock reactions in the sub-seafloor and not magmatic sources of heat drive hydrothermal fluid flow. These reactions create a combination of extreme conditions never before seen in the marine environment: the venting of high-pH (from 9 to 11), warm (40-91°C) hydrothermal fluids with high concentrations of hydrogen, methane, and other hydrocarbons of low molecular mass. The Lost City Hydrothermal Field may thus represent a new type of lifesupporting system in the deep sea. Mixing of the warm, high-pH fluids with seawater precipitates carbonate that drives the growth of chimneys that tower up to 60 m above the seafloor (Fig. 12.1E). These carbonate towers, some of which remain active for thousands of years, house extensive microbial biofilms that are dominated by a single group of Archaea, the Methanosarcinales. However, multiple lines of evidence including morphology, and evidence for diversity of specific genes such as the genes involved in N2 fixation and methane production and anaerobic oxidation, indicate physiological diversity within this Methanosarcinales-dominated biofilm.

Brazelton and colleagues (Brazelton et al. 2009, 2010) correlated archaeal and bacterial V6 tag distributions with isotopic ages of carbonate chimneys collected from the Lost City Hydrothermal Field spanning a 1,200 year period. Clear shifts in the archaeal and bacterial communities were evident over time, and many of the shifts featured rare sequences that dramatically increased in relative abundance to become dominant in older chimneys. These results indicate that some organisms can remain rare at a location for many years before "blooming" and becoming dominant when the environmental conditions allow. Furthermore, the very low overall diversity of the Lost City chimneys revealed that each of the dominant archaeal and bacterial

sequences represented one member of a large pool of similar but much rarer sequences. For example, the most abundant archaeal sequence was more than 90% similar to 1,771 different sequences clustering into 517 operational taxonomic units at 97% sequence similarity, all of which were too rare to be detected by clone library sequencing. Further work in the Baross laboratory has shown that this microdiversity in the V6 region is correlated with microdiversity in a more variable marker, the intergenic transcribed spacer (ITS) region, indicating that it is not generated by pyrosequencing error and that the archaeal population contains even more microdiversity than represented by the 1,771 variants detected in the V6 region. These results confirm that there are many rare species of Methanosarcinales that had not been previously identified from 16S rRNA gene clone libraries that likely represent multiple "ecotypes" within the biofilm and that the ecotype composition changes depending on the age of the carbonate structure.

12.3.4 Rapid temporal turnover in sands of the North Sea island of Sylt

Permeable sandy sediments play a critical role in the recycling of carbon and nitrogen, and act as natural filters that may concentrate microorganisms, nutrients, and organic matter on the extensive continental shelves. Despite the importance of such ecosystems, the extent of microbial diversity and how microbes respond to environmental, spatial, and temporal changes (such as global warming, ocean acidification, and various anthropogenic effects) are still mostly unknown.

Using a 454 massively parallel pyrotag sequencing strategy to describe microbial diversity in temperate sandy sediments from the North Sea island of Sylt (Fig. 12.1F), Ramette and colleagues obtained between 5,000 to 19,000 unique types of bacteria in each gram of sand (A. Gobet, S.I. Böer, J.E.E. van Beusekom, A. Boetius and A. Ramette, unpublished observations). Rarefaction analyses suggest that the OTU richness of sand-associated bacterial communities significantly exceeds the diversity of water column communities from the same environment. The OTU richness also changed dramatically over a few centimeters of sediment depth or between any two consecutive sampling times, with up to 70-80% community turnover. Those remarkable, non-random shifts in community composition may reflect responses to variation of environmental/biogeochemical parameters (for example temperature, nutrients, pigments, production of extracellular enzymes) at the study site. The reservoir of highly diverse low abundant bacterial types might include taxa that become abundant in response to environmental differences in this system. The comparison of diversity

patterns at different taxonomic levels indicated that community shifts occurred at broad taxonomic levels, but that fine-scale patterns in community composition were mostly responsible for the large community turnover observed over sediment depth and sampling time. This study demonstrates the dynamic nature of coastal sandy sediments in terms of microbial diversity, allowing for the formulation of strong ecological hypotheses to explain this phenomenon: strong vertical shifts in nutrient, organic matter, and oxygen availability create a large range of microbial niches, which may support a high turnover of bacterial types in sandy sediments.

12.3.5 Diversity varying with oxygen availability in the Black Sea

The Black Sea, a permanently anoxic basin connected to the ocean through the Bosphorus Sea, has well defined redox gradients and known microhabitats for different metabolic groups of bacteria. C.A. Fuchsman and colleagues (unpublished observations) obtained bacterial pyrotags from four low-oxygen Black Sea water samples: a low-oxygen sample (30 µM oxygen), a sample from the middle of the suboxic zone (2 µM oxygen), a sample from the bottom of the suboxic zone with no detectable oxygen or sulfide, and a sinking particle obtained from the middle of the suboxic zone. The bottom-of-the-suboxic-zone sample (0 µM oxygen) and the particle-attached bacterial sample were more diverse than the 30 µM oxygen and 2 µM oxygen samples. Although all three samples contained low oxygen and no measurable sulfide, only the microbial community structures for the 0 or 2 µM oxygen samples had similar community structures (51% similarity by Bray Curtis) whereas neither resembled the 30 µM oxygen samples (11%). Micro-aerophilic heterotrophs and nitrate reducers dominated the 30 µM and 2μM oxygen samples. In contrast, the 0μM oxygen sample and the particle-attached bacterial samples were more diverse and contained strikingly different taxonomic groups of bacteria. Enriched populations of Deferribacter, δ-Proteobacteria, Lentisphaera, ε-Proteobacteria and Planctomycetes occurred in the particle-attached fraction. These taxonomic groups of bacteria are not normally identified as part of the particle-attached community from oxic waters. Pyrotags for the particle-associated ε-Proteobacteria resemble rRNA sequences from epsilon species that oxidize sulfide. The Deferribacter species are known to reduce metals including manganese and iron oxides, or nitrate and elemental sulfur. Lentisphaera occur in anaerobic environments but little is known about their metabolism. The δ-Proteobacteria, which include known sulfate reducers, were found in the 0 µM oxygen sample and in the particle-attached fraction. However, the OTUs of δProteobacteria differed between the samples, with *Desulfobacteraceae* dominating the 0 µM oxygen sample whereas the particle-attached group could not be assigned to a cultured species.

These results showed a clear correlation between the fluxes and depth of the chemical species such as O₂, NO₃-, NH₄+, CH₄, MnO₂, H₂S and the inferred metabolisms of the bacterial OTUs. Manganese and sulfate reducers and sulfide oxidizers dominated metabolic groups associated with sinking particles whereas microaerophilic and nitrate reducers dominated the water column. This study provides insights into the importance of the particle-attached bacterial communities and points to the potential biases on bacterial diversity estimation when researchers pre-filter samples for microbial diversity studies.

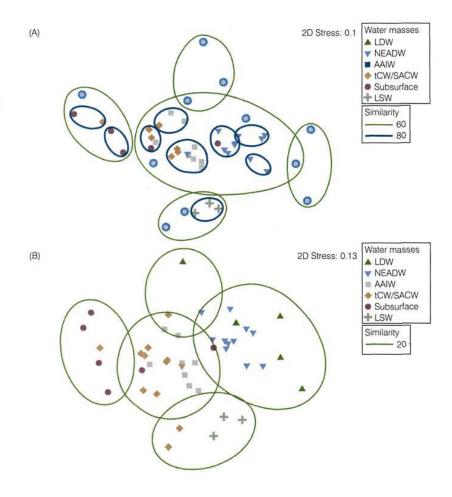
12.3.6 Community signatures of the North Atlantic Deep Water masses

Small size suggests that microbes have high dispersal and high immigration rates, leading to a ubiquitous distribution in the marine environment. However, recent studies demonstrate that microbes can have biogeographic distributions corresponding to individual water masses. Distinct salinity, temperature, and nutrient characteristics differentiate several deep water masses separated by thousands of kilometers of thermohaline ocean circulation. Using bacterial pyrotag sequencing, Herndl and colleagues (unpublished observations) tested this hypothesis by determining the biogeography of bacterioplankton communities following the flow of the North Atlantic Deep Water over a stretch of 8,000 km in the North Atlantic. They focused on the distribution of the abundant versus rare phylotypes to decipher whether rare phylotypes exhibit a similar distribution pattern as the abundant phylotypes or whether they occur ubiquitously. If the rare phylotypes represent a seed bank for the few abundant phylotypes, then the community structure of the rare phylotypes should be fairly uniform across water masses.

Cluster analysis (Fig. 12.9) showed that abundant bacterial phylotypes clustered according to the water masses (Fig. 12.9A). The samples partitioned into one cluster containing bacterial communities from the subsurface zone, two clusters from the mesopelagic waters, three deep water clusters, and one cluster of bacterial communities from the deep Labrador Seawater. Bacterial community composition of deep waters was less similar than samples from subsurface and intermediate waters. Bacterial communities from the same water mass but separated by thousands of kilometers resembled each other more than communities separated by a few hundred meters at individual sites but originating from different water masses.

Fig. 12.9

Non-metric multi-dimensional scaling analysis based on relative abundance of (A) abundant tags (frequency greater than 1% within a sample) and (B) rare tags (frequency less than 0.01% within a sample). Discrimination among samples by water mass. Superimposed circles represent clusters of samples at similarity values of 60% and 80% (A) and 20% (B) (Bray-Curtis similarity). LDW, Lower Deep Water; NEADW, Northeast Atlantic Deep Water, AAIW, Antarctic Intermediate Water, tCW- transitional Central Water, SACW, South Atlantic Central Water; LSW, Labrador Sea Water.

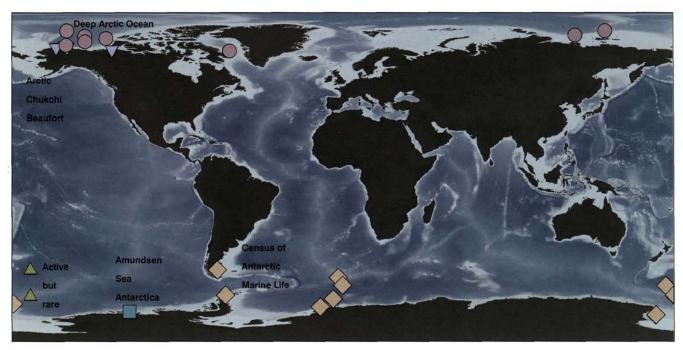


The clustering of the rare sequences (frequency less than 0.01% within a sample, including the singletons; Fig. 12.9B) was similar to the clustering of the abundant sequences (frequency greater than 1% within a sample; Fig. 12.9A), albeit with a generally lower percentage of similarity. Proteobacteria constituted more than 85% of the 1,000 most abundant tags and of these, 52% were α-Proteobacteria, mostly composed of the SAR 11 cluster. The bathypelagic zone had the highest proportion of unassigned bacteria, but showed the highest tag richness and evenness compared to overlying waters. y-Proteobacteria increased with depth, with higher proportions of Chromatiales and Alteromonadales (8.7%) in the bathypelagic zone than in the subsurface zone. The distinct clusters of bacterial communities in specific water masses reflected the presence of unique phylotypes specific to distinct water masses. The variability in the abundance of tags increased with decreasing overall abundance. The most abundant pyrotags exhibited a ubiquitous distribution pattern whereas the representation of low abundance pyrotags seemed to be specific to different

In summary, the bacterial rare biosphere in the North Atlantic is water-mass-specific and hence not ubiquitously distributed as previously suggested. Thus, in this example, it is likely that the rare biosphere originates from the more abundant and/or more active bacterial community through mutation. The biogeochemical role of the high richness of the rare microbial biosphere remains enigmatic and deserves further investigation.

12.3.7 A bipolar distribution of the most abundant bacterial pyrotags

Among the globally distributed samples sequenced as part of the ICoMM Community Sequencing Project, Polar Regions contributed 56 datasets divided among five projects (ABR, ACB, ASA, CAM, and DAO; Table 12.1). The Polar Realm pie chart in Figure 12.8 shows the taxonomic affiliation for the top 20 most abundant tags from these samples. One of the striking results from our comparative study is that the most abundant SAR 11 pyrotag corresponds to the most abundant polar pyrotag. When we map the distribution of the top 20 most abundant tags, we find that all 20 occur in both the Arctic and Southern



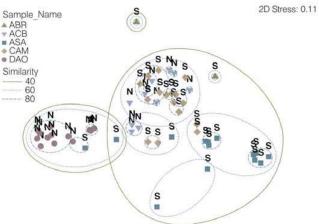


Fig. 12.10

Top, a map detailing the location of all the Arctic and Antarctic datasets examined in this synthesis. Bottom, a non-metric multidimensional scaling plot of 56 Arctic (N) and Antarctic (S) samples based on standardization by total and square-root transformed pyrotag abundances and Bray-Curtis similarities (Clarke & Warwick 2001).

Oceans. Our non-metric multidimensional scaling analysis (Fig. 12.10) confirms the similarity of certain Arctic and Antarctic samples by way of similarity envelopes that encircle datasets that cluster at 80% similarity levels. Although these tags show a "bipolar" distribution using the V6 region as a metric for comparison, we do not know how these populations compare when looking across the entire genome. This question, as well as the source and persistence of bipolar species, awaits the next decade of Census research.

12.4 A Census of Microbial Lipids

12.4.1 MICROBIS and lipid maps

ICoMM has also considered phenotypic characters ranging from microbial physiology to metabolic capability in its global marine microbial census. Intact polar lipids provide a case in point. Lipids can be a powerful tool

for deciphering microbial communities in present and past environments. However, in contrast to genetic data, lipids lack a large database linking identification, chemical structure, and biological or environmental sources. Such a linked database would substantially increase the potential for lipids to be used as a tool for relating environmental microbial communities to cultivated microbes and provide support for phylogenetic relationships. Our approach to constructing a lipid database follows MICRO-BIS (http://icomm.mbl.edu/lipids) in which a central database/program/website links several databases including a database of microbial organisms, lipid structures (Lipid Maps), and mass spectra.

The Lipid Maps database (www.lipidmaps.org; Fahy et al. 2005) represents the most extensive lipid library, but it currently targets biomedical applications and lacks a comprehensive collection of marine microbial lipids. Collaboration with Lipid Maps to generate a universal lipid collection that includes marine microbial lipids enabled the deposition of more than 200 marine microbial lipid structures in the Lipid Maps database. The Lipid Maps database uses a systems biology approach for the categorization, nomenclature, and chemical representation of lipids (Fahy et al. 2005). The classification scheme of Lipid Maps distributes lipids into eight defined categories that are divided into classes and subclasses. Each lipid in the Lipid Maps library can be retrieved using different search criteria including its lipid identification, classification, systematic name, synonym, and chemical name and structure. Furthermore, Lipid Maps assigns unique numbers to lipid structures comparable to the unique accession numbers within GenBank for genomic sequences. New lipid structures can be continuously submitted to Lipid Maps with a proposed lipid identification and thus can evolve continuously if support within the biogeochemical community is strong.

Typically, mass spectrometry identifies lipids in the laboratory. The mass spectra of lipids are generally very comparable between laboratories and thus well suited for database purposes. Unfortunately, at present there are no publicly available mass spectral databases and only commercial libraries such as those from the National Institute of Standards and Technology exist but do not contain a large number of typical marine microbial lipids. Within ICoMM's database MICROBIS we therefore developed a mass spectrometry library that contains lipid data derived from microbes from modern and ancient environments. This library can run under National Institute of Standards and Technology SEARCH software, which is the most commonly used software to search with mass spectra in mass spectral libraries. At this point we have assembled more than 200 mass spectra of the most common and diagnostic lipids of marine microbes.

MICROBIS has laid the foundation for an integrated database for searching and relating lipid data with other molecular and geospatial data. Once further developed, the MICROBIS website will cross-reference lipidomic, taxonomic, DNA sequence, and geospatial data (Fig. 12.3). Currently, ICoMM is designing a search-engine-supported mass spectrometry library wherein cross-referencing between Lipid Maps and the mass spectrometry database will proceed by the LIPID identifications that will also be linked to geospatial data. In the future, MICROBIS will provide a user-friendly interface that will allow searching for taxonomic, DNA sequence, and phylogenetic information, enabling biogeochemists to link lipid data to genomic and geospatial data.

12.4.2 Looking back in time with lipids

ICoMM has also attempted to explore the unknowable: a glimpse at microbial diversity in the past. Dating of evolutionary events within phylogenetic clusters is problematic as it mostly relies on the morphological identification of fossilized remains of microbes, which are usually limited to microbes having inorganic skeletons such as diatoms and coccolithophorids. A new approach is to use fossilized organic molecules. Indeed, fossil DNA occurs in several selected cases (Fish et al. 2002; Coolen et al. 2004), but findings such as these are rare and controversial. In contrast, fossilized lipids commonly occur in sediments of up to 2 billion years old and thus may be, as long as they are diagnostic for certain microbial phylogenetic clusters, suitable to trace the evolutionary history of microbes.

The usefulness of this approach lies in a detailed study of 18S rRNA genes and lipid biomarkers of more than 100 representative marine diatoms (Sinninghe Damsté *et al.* 2004). This study revealed that several lipid biomarkers are quite specific for phylogenetic clusters within the diatoms. For example, the biosynthesis of so-called highly branched isoprenoid alkenes is restricted to two specific phylogenetic clusters, which independently evolved in the centric and pennate diatoms (Sinninghe Damsté *et al.* 2004; Fig. 12.11). The molecular record of C_{25} highly branched isoprenoid chemical fossils in a large suite of well-dated marine sediments and petroleum reveals that the older cluster, comprising rhizosolenoid diatoms, evolved 91.5 \pm 1.5 million years ago (Upper Turonian), enabling an unprecedented accurate dating of diatom evolution.

12.5 Viewing Microbial Diversity through a Community Lens

ICoMM's systematic and high-throughput analyses of the sequence variation of rRNA genes have provided a wealth of microbial community sequence data for numerous,

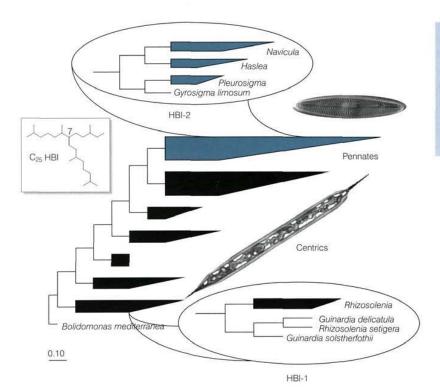


Fig. 12.11

A phylogenetic tree showing how highly branched isoprenoid (HBI) alkenes are restricted to two specific phylogenetic clusters of diatoms (HBI-1 and HBI-2), which independently evolved in the group of centric and pennate diatoms, respectively. Photomicrographs used under license, courtesy of Robert Andersen and David Patterson (http://microscope.mbl.edu).

poorly understood marine environments. When contextual parameters are recorded together with diversity data, it is now possible to assess the impact of space, time, and complex environmental gradients on microbial communities, and to quantify interactions among factors. The integration of laboratory-developed microbiological sensors into observing platforms that track changes at high temporal and spatial resolution will enable autonomous observation of changes in marine microbial diversity in the field (Paul et al. 2006). Here we can find answers to as different questions as the following: Why do specific communities flourish in one environment and not in another? Which microbial populations are more successful than others in the competition for energy or space? Which environments host those seed populations that only temporarily dominate communities? The same tools that have been used in classical community ecology are available for the analysis of changes in microbial community patterns, because ultimately standard sample-by-species matrices can be obtained with any high-throughput method. Our next challenge will be the generation of microbial diversity theories that will allow further comparisons with established ecological theories for macroorganisms or that can be tested across various ecosystems. For instance, recent developments in the study of microbial biogeography may be seen as a prelude to a more dramatic revolution in better understanding microbial communities in their complex environments.

12.6 Marine Microbes and Their Roles in a Changing Ocean

The importance of marine microbes to our biosphere cannot be overstated (Box 12.2). Since the microbial census began, several major scientific breakthroughs in microbial diversity and microbial ecology have occurred. Owing to the rapid developments in high-throughput and relatively cost-effective sequencing technologies like massively parallel DNA sequencing, it has become possible to deeply explore microbial (that is, bacterial, archaeal, and eukaryotic) genetic diversity of environmental samples in both qualitative and quantitative ways. Over the past five to ten years, spectacular findings have highlighted new and unexpected roles of microbes in biogeochemical cycling of carbon, nitrogen, sulfur, iron, and many other (trace) elements owing to interdisciplinary research based on the integration of sequencing, membrane lipid research, and isotope techniques. Fascinating examples of new and important microbial shunts in biogeochemical cycles include the following: the existence of anaerobic oxidation of methane by bacterial-archaeal consortia oxidizing methane with sulfate operating in (sub)oxic environments in the ocean and on land (Knittel & Boetius 2009 and the literature cited therein); the anaerobic oxidation of methane by

Box 12.2

Marine Microbial Diversity and Abundance Highlights

- The number of bacteria in the open ocean exceeds 10²⁹ cells and microbes in total contribute as much as 90% of the biomass in the ocean.
- Microbes may be more than 100 times more diverse than plants and animals.
- A single liter of seawater can represent approximately 20,000 "species" of bacteria.
- A gram of sand can contain between 5,000 and 19,000 "species" of bacteria.
- Archaeal cell numbers can rival those of bacteria in the ocean but their diversity is 10% that of bacteria.
- Protist diversity rivals that of Archaea in some parts of the ocean.

- Most of marine microbial diversity is represented by low abundance populations.
- Each metazoan may have its own unique microbiome population structure.
- Different water masses possess signature microbial community structures.
- Some microbes are everywhere: ICoMM's most abundant type of bacterial pyrotag matched sequences from the SAR11 marine bacterial group which accounts for 25% of the biomass and 50% of the cell abundance in the pelagic ocean.
- Lipids allow us to look back in time at ancient microbial populations and delve into the unknown.

bacteria oxidizing methane with nitrate (Raghoebarsing et al. 2006); anaerobic ammonium oxidation (Anammox), whereby anammox bacteria use specific membrane components to oxidize ammonia with nitrite to form nitrogen gas that escapes from the oceans (Strous et al. 1999; Sinninghe Damste et al. 2002; Kuypers et al. 2003); the discovery that some Crenarchaea use ammonia as an energy source (Konneke et al. 2005); crenarchaeotal CO₂ fixation (Wuchter et al. 2003); and the incredible phenotypic adaptation of the largest known bacterial cells on Earth, the giant sulfide-oxidizing bacteria, to their environment (Gallardo 1977; Teske & Nelson 2006).

These and many other examples (Giovannoni & Stingl 2005; Karl 2007; Azam & Malfatti 2007; Bowler et al. 2009; DeLong 2009; Fuhrman 2009) clearly indicate that marine microbial biogeochemical cycling of elements is even more important than traditionally thought and that microbes dominate these cycles in many known and recently discovered ways. This increased recognition of microbial importance in biogeochemical cycling of elements combined with the discovery of vast microbial diversity, the discovery of the rare biosphere, and the dominance of just a few microbial taxa in environmental settings, makes us aware of the crucial importance of microbes in climate and climate change. In ICoMM we are aware that ongoing human-induced global climate change will and probably already has impacted microbial diversity. Owing to rising seawater temperatures, ocean acidification, and salinity changes, dominant marine microbial taxa may become dormant and completely unknown taxa present in the environment but extremely rare, may become dominant. Because we have little idea about which marine microbial taxa will become dormant and which will become dominant, predictions of changes in biogeochemical cycles are very difficult to make. Thereby predictions of not whether but rather how the climate will change and how to ameliorate such change present even greater challenges to scientists and policy-makers.

The above can be illustrated by the following. The far greater part of N₂ fixation in the marine environment is currently performed by just a very few bacteria, Trichodesmium and an uncultivated 'Group A' putative unicellular cyanobacterium lacking oxygenic photosystem II, whereas the endosymbiotic cyanobacterium Richelia intracellularis, as well as other endo- and ectosymbiotic N2fixing cyanobacteria, are less important globally (Arrigo 2005). Very preliminary laboratory experiments with artificially acidified seawater indicate that N2-fixing bacteria react very strongly to lowered pH values, thereby changing the rate and nature of nitrogen and carbon fixation (Hutchins et al. 2007). This will impact the complete marine nitrogen and carbon cycles and thereby other biogeochemical cycles (for example, Crenarchaeota use ammonium to fix carbon; phosphate may become the omnipresent limiting nutrient). Similar experiments with the major carbon-fixing organisms in oligotrophic water, Prochlorococcus and Synechococcus, also indicated that acidification and temperature rise will severely affect their

metabolism, leading to changes in the rate of CO₂ fixation (Fu et al. 2007).

There are many factors that add to the uncertainty in the estimation of the pathways and the scale of the interaction of marine microbe communities with anthropogenically driven global climate change. Second only to human impacts on the environment, the interplay between environmental parameters and shifts in marine microbial diversity will dominate the course of climate change. The paucity of research on underlying mechanisms severely constrains the ability of policy-makers to make informed decisions about mitigating strategies. Trying to understand microbial diversity and functioning in biogeochemical and nutrient cycling is of major importance for future research on a worldwide scale.

12.7 New Questions

Despite the great diversity we see in microbial communities, there is a high degree of structure and non-random patterns in temporal and spatial scaling, and biological associations. Hence, new questions have emerged from ICOMM studies:

- What is the turnover of microbial populations and communities across various scales in space and time?
- Why are some groups dominating marine habitats globally?
- Why is there such a division between the community structure of pelagic and benthic habitats?
- Are the most diverse taxa also the most numerically abundant and why?
- What kinds of taxa are associated with plants and animals, and to what extent are they unique to each species?
- Why are there so many rare populations?
- Is the "rare biosphere" the result of 3.5 billion years of evolution, of massive horizontal gene transfer, of dispersion in the oceans, of life strategies, or combinations of the above?

To unravel the origin of these phenomena and to address these questions are challenges for the future.

12.8 Outlook

Looking forward to the next decade of microbial census research, we see many opportunities and challenges. As massively parallel DNA sequencing brings an unprecedented volume of data, it also brings challenges in analyzing these data. In this chapter we have chosen to highlight the general findings of our efforts so far, but the necessary computer algorithms and models required to bring us closer

to more robust estimates of microbial diversity are still being developed and the required computational power still being sought. Improving the taxonomy attached to pyrotags is another area that will need attention in the future. Much of this will likely come though improved annotations of the vast amount of full-length or nearly full-length sequences presently housed in public databases, as well as next-generation sequencing providing much longer reads that will enable better taxonomic assignment. Yet, we find that even when definitive taxonomic assignment is not possible, the approach is still very powerful for comparisons of assemblage composition and diversity. The technique reveals substantial diversity undetected by previously used techniques. Even ICoMM formalin-preserved samples belonging to the Joint Global Ocean Flux Study have been successfully sequenced, unveiling the potential of pyrosequencing to become a powerful tool for paleobiological and paleoenvironmental studies. The massive amount of data from pyrosequencing also enables predictions of functions for many OTUs. In short, large-scale patterns could emerge not only of phylogenetic diversity but also functional diversity along geochemical gradients. As the 454-pyrotag-based technology does not distinguish between dead and living microbial taxa, it could be complemented by other techniques for assessing the level of viability within a sample such as starting with RNA samples and reverse transcribing it to reveal the most active populations. Deciphering ecological signals linked to the definition of rare or abundant OTUs at different taxonomic levels is crucial. This definition will influence how diversity patterns are measured. Furthermore, this new paradigm will aid researchers to better understand marine microbial communities and their impact on planetary biogeochemical cycles.

Future endeavors must pay closer attention to the temporal dimension of changes in microbial community structures. As sequencing costs decline, it will be possible to monitor microbial populations and their changes over time in appropriate marine environments on different timescales from minutes to centuries. Developing such monitoring strategies through existing observing systems, time-series stations (for example Bermuda Atlantic Time-Series, Hawaii Ocean Time-Series), and Long Term Ecological Research Sites, we might be better able to predict changes in microbial populations as a consequence of natural and anthropogenic climate change. The ICoMM team recommends that this kind of monitoring approach may thus help substantially to improve our ability to predict climate change, harmful algal blooms, and ultimately our own impact on biodiversity in the ocean.

Acknowledgments

We thank all the project principal investigators who participated in the ICoMM 454 community pyrosequencing

effort, as well as all the scientists, technicians, and students who participated in sample collection, processing, and analysis of data in each regional laboratory and especially at the Marine Biological Laboratory, Woods Hole. ICoMM is a project of the Census of Marine Life program funded by the Sloan Foundation. A grant from the W. M. Keck Foundation supported the 454 pyrotag sequencing conducted at the Marine Biological Laboratory, Woods Hole. Local and international funding facilitated project participation in this global initiative. GenBank sequence read archive data include SRP000903, SRP000912, SRP001108, SRP001172, SRP001206-31, SRP001242-45, SRP001259, SRP001265-66, SRP001268-70, SRP001273, SRP001309, and SRP001542-47.

References

- Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. & Huse, S.M. (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA enes. *PLoS ONE*, 4(7), e6372.
- Andersen, R.A., Beakes, G., Blackburn, S., et al. (2006) Report on the Alfred P. Sloan Foundation Workshop on Protistan Barcoding, Reference Material and Cultures, Portland.
- Arrigo, K.R. (2005) Marine microorganisms and global nutrient cycles. Nature 437, 349–355.
- Azam, F. & Malfatti, F. (2007) Microbial structuring of marine ecosystems. Nature Reviews Microbiology 5, 782–791.
- Bowler, C., Karl, D.M. & Colwell, R.R. (2009) Microbial oceanography in a sea of opportunity. *Nature* 459, 180–184.
- Brazelton, W.J., Sogin, M.L. & Baross, J.A. (2009) Multiple scales of diversification within natural populations of Archaea in hydrothermal chimney biofilms. *Environmental Microbiology Reports* 2, 236–242.
- Brazelton, W.J., Ludwig, K.A., Sogin, M.L., Andreishcheva, E.N., Kelley, D.S., Shen, C.-C., Edwards, R.L. & Baross, J.A. (2010) Archaea and bacteria with surprising microdiversity show shifts in dominance over 1,000-year time scales in hydrothermal chimneys. Proceedings of the National Academy of Sciences of the USA 107, 1612–1617.
- Clarke, K.R. & Warwick, R.M. (2001) Change in Marine Communities: An Approach to Statistical Analysis and Interpretation, 2nd edn. Plymouth Marine Laboratory, UK: PRIMER-E.
- Coolen, M.J.L., Muyzer, G., Rijpstra, W.C., et al. (2004) Combined DNA and lipid analyses of sediments reveal changes in Holocene phytoplankton populations in an Antarctic lake. Earth and Planetary Science Letters 223, 225–239.
- Corliss, J.O. (1984) The kingdom Protista and its 45 phyla. Bio-Systems, 17, 87–126.
- Del Giorgio, P.A. & Duarte, C.M. (2002) Respiration in the open ocean. Nature 420, 379–384.
- DeLong, E.F. (1992) Archaea in coastal marine environments. Proceedings of the National Academy of Sciences of the USA 89, 5685–5689.
- DeLong, E.F. (2009) The microbial ocean from genomes to biomes. Nature 459, 200-206.
- DeLong, E.F., Wickham, G.S. & Pace, N.R. (1989) Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. Science 243, 1360–1363.
- Díez, B., Pedrós-Alió, C. & Massana, R. (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions

- by small-subunit rRNA gene cloning and sequencing. Applied and Environmental Microbiology 67, no. 7, 2932–2941.
- Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., et al. (2005) A comprehensive classification system for lipids. *Journal of Lipid Research* 46, 839–861.
- Field, C.B., Behrenfeld, M.J., Randerson, J.T. & Falkowski, P. (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281, 237–240.
- Fish, S.A., Shepherd, T.J., McGenity, T.J. & Grant, W.D. (2002) Recovery of 16S ribosomal RNA gene fragments from ancient halite. *Nature* 417, 432–436.
- Fu, F.-X., Warner, M.E., Zhang, Y., et al. (2007) Effects of increased temperature and CO₂ on photosynthesis, growth, and elemental ratios in marine Synechococcus and Prochlorococcus (Cyanobacteria). Journal of Phycology 43, 485–496.
- Fuhrman, J.A. (2009) Microbial community structure and its functional implications. *Nature*, 459, 193–199.
- Fuhrman, J.A., McCallum, K. & Davis, A.A. (1992) Novel major archaebacterial group from marine plankton. *Nature* 356, 148–149.
- Fuhrman, J.A., Schwalbach, M.S. & Stingl, U. (2008) Proteorhodopsins: an array of physiological roles? *Nature Reviews Micro*biology 6, 488–494.
- Fuhrman, J.A., Sleeter, T.D., Carlson, C.A. & Proctor, L.M. (1989) Dominance of bacterial biomass in the Sargasso Sea and its ecological implications. *Marine Ecology Progress Series* 57, 207–217.
- Galand, P.E., Casamayor, E.O., Kirchman, D.L. & Lovejoy, C. (2009a) Ecology of the rare microbial biosphere of the Arctic Ocean. Proceedings of the National Academy of Sciences of the USA 106, 22427–22432.
- Galand, P.E., Casamayor, E.O., Kirchman, D.L., et al. (2009b) Unique archaeal assemblages in the arctic ocean unveiled by massively parallel tag sequencing. The International Society of Microbial Ecology Journal 3, 860–869.
- Gallardo, V.A. (1977) Large benthic microbial communities in sulphide biota under Peru-Chile Subsurface Countercurrent. *Nature* 268, 331–332.
- Garcia-Pichel, F., Belnap, J., Neuer, S. & Schanz, F. (2003) Estimates of global cyanobacterial biomass and its distribution. Archive für Hydrobiologie/Algological Studies 109, 213–228.
- Gast, R.J., Moran, D.M., Beaudoin, D.J., et al. (2006) Abundance of a novel dinoflagellate phylotype in the Ross Sea, Antarctica. Journal of Phycology 42, 233–242.
- Giovannoni, S.J., Britschgi, T.B., Moyer, C.L. & Field, K.G. (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345, 60–63.
- Giovannoni, S.J. & Stingl, U. (2005) Molecular diversity and ecology of microbial plankton. *Nature*, 437, 343–348.
- Giovannoni, S.J., Tripp, H.J., Givan, S., et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. Science 309, 1242–1245.
- Hamasaki, K., Taniguchi, A., Tada, Y., et al. (2007) Actively growing bacteria in the inland Sea of Japan, identified by combined bromodeoxyuridine immunocapture and denaturing gradient gel electrophoresis. Applied and Environmental Microbiology 73, 2787–2798.
- Herndl, G.J., Reinthaler, T., Teira, E., et al. (2005) Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. Applied and Environmental Microbiology 71, 2303–2309.
- Heywood, J.L., Zubkov, M.V., Tarran, G.A., et al. (2006) Prokaryoplankton standing stocks in oligotrophic gyre and equatorial provinces of the Atlantic Ocean: evaluation of inter-annual variability. Deep-Sea Research II 53, 1530–1547.
- Hobbie, J.E., Daley, R.J. & Jasper, S. (1977) Use of nuclepore filters for counting bacteria by fluorescence microscopy. Applied and Environmental Microbiology 33, 1225–1228.

- Holm-Hansen, O. & Booth, C.R. (1966) The measurement of adenosine triphosphate in the ocean and its ecological significance. Limnology and Oceanography 11, 510–519.
- Huber, J.A., Mark Welch, D.B., Morrison, H.G., et al. (2007) Microbial population structures in the deep marine biosphere. Science 318, 97–100.
- Huse, S.M., Dethlefsen, L., Huber, J.A., et al. (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genetics 4(11), e1000255.
- Hutchins, D.A., Fu, F.-X., Zhang, Y., et al. (2007) CO₂ Control of Trichodesmium N₂ fixation, photosynthesis, growth rates, and elemental rations: implications for past, present and future biogeochemistry. Limnology and Oceanography 52, 1293–1304.
- Jannasch, H.W. & Jones, G.E. (1959) Bacterial populations in sea water as determined by different methods of enumeration. *Limnology and Oceanography* 4, 128–139.
- Jørgensen, B.B. & Boetius, A. (2007) Feast and famine microbial life in the deep-sea bed. Nature Reviews Microbiology 5, 770–781.
- Karl, D.M. (2007) Microbial oceanography: paradigms, processes and promise. Nature Reviews Microbiology 5, 759–769.
- Kirchman, D.L. (ed.) (2008) Microbial Ecology of the Oceans. Hoboken: John Wiley.
- Knittel, K. & Boetius, A. (2009) Anaerobic oxidation of methane: progress with an unknown process. Annual Review of Microbiology 63, 311–334.
- Konneke, M., Bernhard, A.E., de la Torre, J.R., et al. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437, 543–546.
- Kuypers, M.M.M., Sliekers, A.O., Lavik, G., et al. (2003) Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. Nature 422, 608–611.
- Kysela, D.T., Palacios, C. & Sogin, M.L. (2005) Serial analysis of V6 ribosomal sequence tags (SARST-V6): a method for efficient, highthroughput analysis of microbial community composition. *Envi*ronmental Microbiology 7, 356–364.
- Lee, J.J., Hutner, S.H. & Bovee, E.C. (eds.) (1985) An Illustrated Guide to the Protozoa. Lawrence, Kansas: Society of Protozoologists.
- Li, W.K.W. (1994) Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnology and Oceanography* 39, 169–175.
- López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C. & Moreira, D. (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409, 603–607.
- Margulies, M., Egholm, M., Altman, W.E. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.
- Mayali, X., Franks, P.J.S. & Azam, F. (2008) Cultivation and ecosystem role of a marine Roseobacter clade-affiliated cluster bacterium. Applied and Environmental Microbiology 74, 2595–2603.
- Moon-van der Staay, S.Y., De Wachter, R. & Vaulot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409, 607–610.
- Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. Science 276, 734–740.

- Patterson, D.J. (1999) The diversity and diversification of protists. American Naturalist 154, S96–S124.
- Paul, J., Scholin, C., Van den Engh, G. & Perry, M.J. (2006) In situ instrumentation. Oceanography 20, 70–78.
- Pedrós-Alió, C. (2006) Marine microbial diversity: can it be determined? Trends in Microbiology 14, 257–263.
- Raghoebarsing, A.A., Pol, A., van de Pas-Schoonen, K.T., et al. (2006) A microbial consortium couples anaerobic methane oxidation to denitrification. Nature 440, 918–921.
- Scanlan, D.J., Ostrowski, M., Mazard, S., et al. (2009) Ecological genomics of marine Picocyanobacteria. Microbiology Molecular Biology Reviews 73, 249–299.
- Simon, H.M., Dodsworth, J.A. & Goodman, R.M. (2000) Crenarchaeota colonize terrestrial plant roots. *Environmental Micro*biology 2, 495–505.
- Sinninghe Damsté, J.S., Muyzer, G., Abbas, B., et al. (2004) The rise of the rhizosolenid diatoms. Science 304, 584–587.
- Sinninghe Damste, J.S., Strous, M., Rijpstra, W.I.C., et al. (2002) Linearly concatenated cyclobutane lipids form a dense bacterial membrane. Nature 419, 708–712.
- Sogin, M.L., Morrison, H.G., Huber, J.A., et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences of the USA 103, 12115–12120.
- Stahl, D.A., Lane, D.J., Olsen, G.J. & Pace, N.R. (1984) Analysis of hydrothermal vent-associated symbionts by ribosomal RNA sequences. Science 224, 409–411.
- Strous, M., Fuerst, J.A., Kramer, E.H.M., et al. (1999) Missing lithotroph identified as new planctomycete. *Nature* 400, 446–449.
- Teske, A. & Nelson, D.C. (2006) The genera Beggiatoa and Thioploca.
 In: The Prokaryote (eds. M. Dworkin, M. Falkow, K.H. Schleifer.
 & E. Stackebrandt), pp.784–810. New York: Springer.
- Venter, J.C., Remington, K., Heidelberg, J.F., et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.
- Whitman, W.B., Coleman, D.C. & Wiebe, W.J. (1998) Prokaryotes: the unseen majority. Proceedings of the National Academy of Sciences of the USA 95, 6578–6583.
- Wuchter, C., Schouten, S., Boschker, H.T.S. & Sinninghe Damsté, J.S. (2003) Bicarbonate uptake by marine Crenarchaeota. Federation of European Microbiological Societies Microbiology Letters 219, 203–207.
- Yentsch, C.M., Horan, P.K., Muirhead, K., Dortch, Q., et al. (1983) Flow cytometry and cell sorting: a technique for analysis and sorting of aquatic particles. Limnology and Oceanography 28, 1275–1280.
- Zhu, F., Massana, R., Not, F., et al. (2005) Mapping of picoeukaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. Federation of European Microbiological Societies Microbiology Letters 52, 79–92.
- Zimmermann, R. & Meyer-Reil, L. (1974) A new method for fluorescence staining of bacterial populations on membrane filters. Aus dem Institut fur Meereskunde an der Uniiversitat Kiel 30, 24-27.