REVIEWS

MICROBIAL COMMUNITY GENOMICS IN THE OCEAN

Edward F. DeLong

Abstract | Marine microbial communities were among the first microbial communities to be studied using cultivation-independent genomic approaches. Ocean-going genomic studies are now providing a more comprehensive description of the organisms and processes that shape microbial community structure, function and dynamics in the sea. Through the lens of microbial community genomics, a more comprehensive view of uncultivated microbial species, gene and biochemical pathway distributions, and naturally occurring genomic variability is being brought into sharper focus. Besides providing new perspectives on oceanic microbial communities, these new studies are now poised to reveal the fundamental principles that drive microbial ecological and evolutionary processes.

EUPHOTIC ZONE
The uppermost stratum of the water column that receives sufficient light for photosynthesis.

BENTHIC
Living in, or on the bottom of, a body of water.

Division of Biological Engineering and Department of Civil and Environmental Engineering, Room 48-427, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.
e-mail: delong@mit.edu
doi:10.1038/nrmicro1158
Published online 10 May 2005

Molecular approaches for characterizing microbial species and assemblages have significantly influenced our understanding of microbial diversity and ecology. In particular, ribosomal RNA (rRNA) gene sequence comparisons have provided a revolutionary approach for interpreting microbial evolutionary relationships[1]. In a logical extension of this technique, extraction of phylogenetically informative genes (like rRNAs) directly from naturally occurring microorganisms represented another important development in microbial biology, opening up the natural microbial world to closer scrutiny[2,3]. Resulting discoveries include the recognition of new phylogenetic lineages[4–6], distributional mapping of taxa in sometimes unsuspected habitats[7,8] and the fundamental realization that most extant microbial diversity had eluded detection by traditional cultivation approaches[3,5].

Advances in genome sequencing technologies have similarly had great impact on microbial biology, providing new insights into microbial evolution, biochemistry, physiology and diversity. Genomic technologies are now also extending their influence to microbial ecology and environmental science. With respect to ocean science, several marine microbial whole genome sequences have been completed (TABLE 1). The first sequenced marine microorganism was that of a marine archaeon, *Methanocaldococcus jannaschii*, which was isolated from deep-sea hydrothermal vents[9], and the list of sequenced marine microorganisms now includes marine viruses (see also the article by R.A. Edwards and F. Rohwer in this issue), cyanobacteria[10–12], bacteria[13–22], archaea[9,23–30] and protists[31] from the EUPHOTIC ZONE, deep water, BENTHIC habitats and hydrothermal vents. Many more marine bacterial, archaeal and protistan genome sequencing projects are now underway. As these new sequencing initiatives progress (see Online links box), the database of reference sequences from ecologically relevant organisms will expand rapidly. This is a crucial consideration in cultivation-independent genomic projects, as whole genome sequences from relevant organisms provide the foundations for interpretation and annotation of environmental genomic data.

Most recently, the merging of cultivation-independent gene sequences with contemporary genomic approaches (such as whole-genome shotgun sequencing) is providing a more comprehensive picture of the structure and function of indigenous microbial communities. Genomic approaches for studying natural microbial assemblages have been variously dubbed environmental genomics, population genomics, metagenomics or ecogenomics. Regardless of the moniker, all these approaches involve cultivation-independent genomic

Table 1 | **Published whole genomes from marine microorganisms**

Species	Genome size	References
Bacteria		
Photobacterium profundum	6,400 kb	21
Vibrio fischeri	4,284 kb	22
Silicibacter pomeroyi	4,109 kb	14
Idiomarina loihiensis	2,839 kb	13
Desulfotalea psychrophila	3,659 kb	18
Vibrio vulnificus	5,211 kb	15
Prochlorococcus marinus subsp. Pastoris MED4	1,657 kb	11
Prochlorococcus marinus MIT9313	2,410 kb	11
Synechococcus sp.	2,434 kb	12
Prochlorococcus marinus SS120	1,751 kb	10
Rhodopirellula baltica	7,145 kb	19
Vibrio parahaemolyticus	5,165 kb	16
Oceanobacillus iheyensis	3,630 kb	20
Archaea		
Methanococcus maripaludis	1,661 kb	25
Methanosarcina acetivorans	5,751 kb	24
Methanopyrus kandleri	1,694 kb	30
Pyrococcus furiosus	1,908 kb	29
Pyrobaculum aerophilum	2,222 kb	113
Pyrococcus abyssi	1,765 kb	23
Thermotoga maritima	1,860 kb	17
Aeropyrum pernix	1,669 kb	27
Pyrococcus horikoshii	1,738 kb	26
Archaeoglobus fulgidus	2,178 kb	28
Methanocaldococcus jannaschii	1,664 kb	9
Eukarya		
Thalassiosira pseudonana	25,000 kb	31

analysis of DNA extracted from naturally occurring microbial biomass. These techniques were first applied to marine plankton to characterize uncultivated marine bacterial and archaeal species[32,33], and are now becoming a common method to characterize microbial assemblages. Applications include the genome analysis of uncharacterized taxa[34–49], expression of novel genes or pathways from uncultured environmental microorganisms[34,48,50–53], elucidation of community-specific metabolism[54,55] and comparison of different community gene contents. Several recent reviews and commentaries have summarized the history and potential applications of cultivation-independent genomic surveys[50,51,56–62]. Here, the development, application and potential of ocean-going microbial genomics is explored.

Practical approaches

Large-insert bacterial artificial chromosome and fosmid libraries. Several strategies for cultivation-independent genomic survey of marine microbial communities have been used (FIG. 1). One of the earliest examples used bacteriophage-λ cloning techniques to produce genomic libraries from marine picoplankton[32]. More recently, fosmids and bacterial artificial chromosomes[63,64] (BACs) have been applied in genomic analyses of naturally occurring marine microorganisms[34,35,38–40,49,54,65] (FIG. 1a). These vectors are particularly useful for stable, high-fidelity propagation of large DNA inserts[63,64]. DNA fragments of up to 200 kb can be stably cloned in these vectors; therefore, one clone could represent 5–10% of the entire genome of a small bacterium. BACs prepared from microbial assemblage DNA can be easily screened to identify and characterize cloned gene fragments for functions or to evaluate phylogeny. The first example of characterization of a microorganism using this approach examined an abundant but uncultivated group of planktonic marine archaea[33]. Several studies have expanded the characterization of uncultivated archaeal species using this general approach (see also the article by C. Schleper and colleagues in this issue). Large-fragment cloning and sequencing has also led to several fundamental discoveries about the genome content, functional properties and potential ecological significance of bacteria in the ocean environment. Finally, BAC libraries are repositories of genomic material, and can also serve as a valuable reference resource for further sequencing and *in vitro* biochemical experimentation.

Small-insert whole-genome shotgun libraries. Another approach for cultivation-independent microbial genome characterization is a variant of whole-genome shotgun (WGS) sequencing[66] (FIG. 1b). For pure bacterial cultures, the WGS approach has been important for obtaining complete genome sequences, including those of several marine bacteria and archaea (TABLE 1). WGS sequencing has also been used to sequence microbial symbionts[67,68] and, in one case, an extremely simple microbial biofilm assemblage[69]. WGS sequencing relies on the preparation and end-sequencing of small-DNA-insert libraries and subsequent sequence assembly *in silico*. The high throughput nature of this approach makes it extremely attractive. Variations on this theme, using linker ligation and subsequent amplification techniques, have also been used to generate shotgun libraries from naturally occurring viral populations[36,37].

To date, it seems that WGS approaches alone cannot adequately deconvolute whole genome sequences from complex microbial assemblages. As with the human genome sequencing effort[70,71], the most complete and reliable datasets will probably result from a combination of sequencing and analysis strategies. These will also probably include front-end cell purification strategies to reduce inherent complexity, followed by combined WGS and large-insert sequencing strategies. In combination, these approaches could enhance the accuracy, coverage and reliability of genomics-based efforts to understand complex

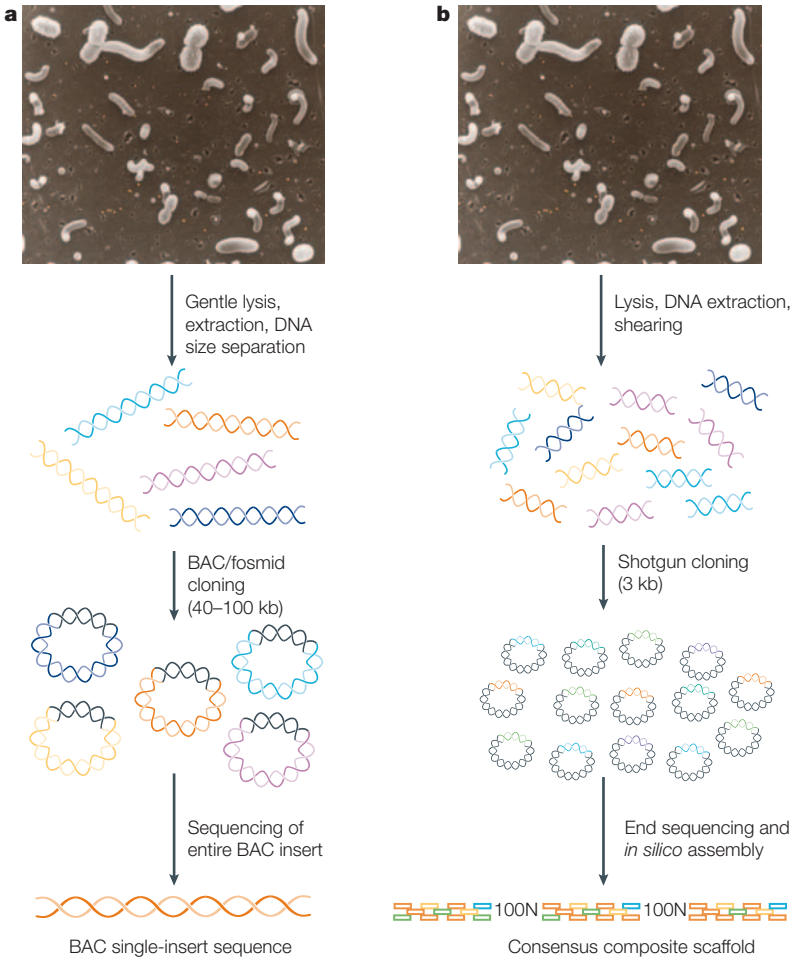

Figure 1 | **Microbial community DNA sequencing.** Schematic diagram of common approaches for retrieving genomic sequence information from natural microbial populations. **a** | One approach uses large DNA inserts recovered in bacterial artificial chromosomes (BACs) that are each derived from an individual cell. Subsequent sequencing and assembly results in a contiguous DNA sequence that is derived from a single cell in the original population. **b** | Another approach is based on recovery of small inserts, and attempts subsequent assembly from cloned DNA derived from a genetically heterogeneous population. The end result is an assembly of DNA sequence that is derived from many different cells.

microbial communities. Nevertheless, WGS sequencing of microbial communities represents a powerful, if expensive, approach for high-volume, single-pass gene survey and sampling.

Marine microbial case studies

Several studies have used either large-insert DNA cloning techniques, WGS approaches or both to characterize marine microbial assemblages (TABLE 2). The outcomes of these studies include the discovery of unsuspected mechanisms of light-driven energy generation in the ocean[34,35,38,39,72–75], a massive survey of the gene complement of Sargasso Sea microorganisms[76,77] and the characterization of metabolic pathways of methane-oxidizing archaea in deep-sea sediments[54,78,79]. A brief review of these studies illustrates the potential, progress and pitfalls of cultivation-independent genomic analyses.

Photobiology of marine picoplankton. Early forays into environmental genomics demonstrated the feasibility of obtaining informative genomic 'snapshots' from uncultivated marine microorganisms[32,40,49,52]. Recent surveys of genome fragments from bacterioplankton that were archived in BAC libraries have led to several surprising discoveries. One approach was to identify genome fragments containing phylogenetic markers (for instance, rRNAs) and sequence the flanking genomic regions — a type of phylogeneticially anchored chromosome walking[33,65]. Using this method, a 130 kb BAC clone was isolated from an uncultivated SAR86 bacterium[34] (an abundant component of α-proteobacteria in ocean surface waters). Sequencing of the 130 kb fragment revealed a new class of genes of the rhodopsin family (named proteorhodopsin) that had never before been observed in bacteria. When the bacterial proteorhodopsin was expressed in *Escherichia coli*, it functioned as a light-driven proton pump[34]. Before these genome-enabled studies, these photoproteins had not been known to occur in bacteria or the ocean. So, this genomic survey of uncultivated marine bacteria led directly to the discovery of a new type of light-driven energy generation in oceanic bacteria. Later studies confirmed the presence of retinal-bound proteorhodopsin in the ocean, and showed that optimized spectral 'tuning' of bacterial rhodopsins matches depth-specific light availability[75]. Shotgun sequencing in the Sargasso Sea has now verified both the abundance and diversity of this new class of photoproteins. The emerging understanding of proteorhodopsin taxonomic and environmental distributions is providing new insights into gene and genome evolution in microbial populations[38,75,80].

As in other areas of science, discoveries encountered in environmental genomic surveys can influence different fields in unexpected ways. In terms of biophysics, rhodopsins are possibly the most studied and well-understood membrane proteins. Before the discovery of proteorhodopsin, however, the only known prokaryotic rhodopsins were from extremely halophilic archaea, and these were difficult to express functionally in *E. coli*. By contrast, proteorhodopsin is readily expressed and functionally active in *E. coli*, and has therefore become a valuable biophysical tool for elucidating the biochemical and biophysical properties of these transmembrane proteins[81–91]. The ability to genetically and functionally manipulate natural and engineered proteorhodopsin variants indicates their potential for use in biotechnological applications[92].

Other modes of bacterial phototrophy have also recently been shown to be common in ocean surface waters. These represent alternative strategies of light utilization compared with that of well-known oxygenic photoautotrophs, like *Synechococcus* sp. and *Prochlorococcus* spp. Abundant bacteriochlorophyll-containing aerobic, anoxygenic phototrophic (AAnP) bacteria were first identified in seawater through bio-optical and biophysical measurements[93,94]. Following these reports, several different types of AAnP were found in Monterey Bay microbial BAC libraries[35].

Table 2 | **Published marine microbial environmental genomic studies**

Year	Site	Depth (meters)	Community type	Library type	References
1991	Subtropical Pacific	0	Open ocean picoplankton	λ	32
1996	Oregon coast	200	Coastal picoplankton	Fosmid	33,41
1996	Santa Barbara, California	10	Sponge symbionts	Fosmid	40,49
1999	Delaware Bay, Delaware	0	Estuarine picoplankton	λ	53
2000	Monterey Bay, California	0	Coastal picoplankton	BAC	34,35,65
2002	Antarctic peninsula	0	Coastal picoplankton	Fosmid	101
2002	Mission Bay, California	0	Coastal planktonic virus	Shotgun*	36,37
2003	Subtropical Pacific	0	Open ocean picoplankton	BAC	38
2004	Antarctic Polar front	500	Picoplankton	Cosmid	43,45
2004	Sargasso Sea	0	Open ocean picoplankton	Shotgun	77
2004	Mediterranean Sea	0	Open ocean picoplankton	BAC	39,72
2004	Eel River Basin, California	550	Deep-sea sediment microorganisms	Shotgun and fosmid	54
2005	Pacific Ocean	1,674	Deep-sea whale fall	Shotgun	55
2005	Antarctic peninsula	560	Deep-sea whale fall	Shotgun	55

*Modified linker-ligation shotgun libraries. BAC, bacterial artificial chromosome.

Some of these BAC clones contained 'photosynthetic superoperons', which encode the photosynthetic reaction centre, and carotenoid and bacteriochlorophyll biosynthetic genes[35]. Both the genes and genomic organization indicated that some of these photosynthetic superoperons were peripherally related to more familiar planktonic AAnP bacteria of the α-proteobacteria group. Other planktonic phototrophs, containing photosynthetic superoperons that seem unrelated to α-proteobacteria, were also prevalent among the Monterey Bay BAC clones. The photoprotein gene similarity and gene order of these clones indicated a possible relationship to β-proteobacteria or γ-proteobacteria. Planktonic AAnP bacteria that are more closely affiliated with *Roseobacter* species have also been reported using genomic techniques[72,95,96]. In total, the combined biophysical, genomic and culture-based studies have indicated that bacteriochlorophyll-containing AAnP bacteria are broadly distributed throughout the world's oceans, both taxonomically and geographically. Together, these genome-enabled analyses of proteorhodopsin and AAnP marine bacteria are changing our views about the nature and prevalence of light-utilization strategies in ocean surface waters.

Shotgun sequencing of Sargasso Sea microorganisms. One of the more remarkable applications of environmental genomics to date is a WGS sequencing effort of the microbial community in the Sargasso Sea[77]. This study produced a total of 1,987,936 DNA sequence reads, yielding approximately 1,625 Mb of DNA sequence. In their first analyses, Venter *et al.* pooled shotgun DNA sequences from four out of seven discrete samples (~1.36 Gb of sequence total) for subsequent *in silico* DNA assembly. The dataset is remarkable in terms of its sheer magnitude and total gene content. This data collection effort demonstrated the raw power of environmental genomics for exploring natural microbial gene content and diversity. The study also reveals some of the pitfalls that can be encountered in sampling, contig assembly, data analyses and interpretation.

The gene complement of the assembled Sargasso Sea dataset consisted of about 1,214,207 identified protein-encoding genes, encompassing ~1,412 individual rRNA genes. Counts of phylogenetically informative genes were used to estimate SPECIES RICHNESS. The observed species number within these protein-encoding genes was estimated by counting 'genomic species', defined as clusters of assemblies or individual sequence reads with ≥94% similarity at the DNA level. Using two separate approaches, Venter and colleagues inferred a minimal total number of 1,800 species in their sample — corresponding with the total unique rRNA counts.

The type and distribution of microorganisms present in the Sargasso Sea samples were fairly consistent with known oceanic microbial species and distributions. As expected, bacterial groups known to be abundant in marine surface waters[4], such as the α-proteobacteria SAR11 clade (~400 different SAR11 rRNA sequences were detected), the γ-proteobacteria SAR86 clade (known to contain the photoprotein proteorhodopsin), cyanobacteria and *Cytophaga* sp. were well represented. Other typical marine planktonic microorganisms, such as *Roseobacter* spp., *Alteromonas* spp., the SAR116 clade and marine crenarchaeota were also detected.

Among the native microbial Sargasso Sea sequence assemblies, it is clear that there are still difficulties in assembling large, accurate DNA contigs and

SPECIES RICHNESS
The number of different species in a given habitat, biotope, community or assemblage.

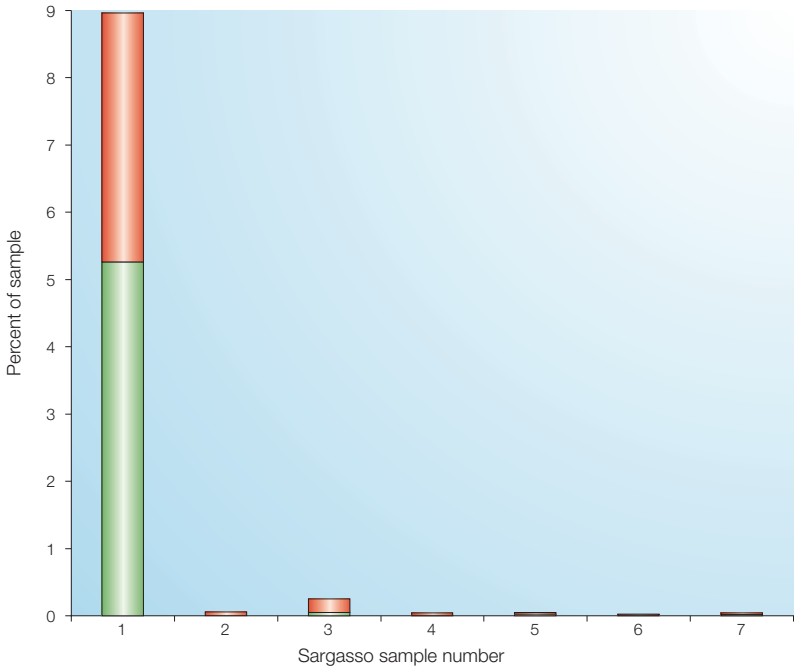

Figure 2 | ***Burkholderia* and *Shewanella* gene content in Sargasso Sea samples.** BLAST[112] analyses of the percentage of *Burkholderia cepacia* or *Shewanella oneidensis* top hits (blastn DNA local alignment hits with expectation cutoff values less than 1 x 10^{-20}), using the Sargasso Sea tracefiles deposited in NCBI. Contaminating vector and repetitive sequence was removed from the tracefile data before analyses. Green bars represent the percentage of top BLAST hits to *B. cepacia* recovered in individual Sargasso Sea sample datasets. Red bars represent the percentage of top BLAST hits to *S. oneidensis* recovered in individual Sargasso Sea sample datasets.

scaffolds from complex mixed microbial populations. In relatively few cases, large contigs could be assembled from known abundant population constituents, like *Prochlorococcus* spp., SAR11 and SAR86 bacterial types. For SAR11, the largest contig that could be assembled was only 21 kb (REF. 77). For open ocean, SAR86-like, proteorhodopsin-containing DNA fragments, the largest assemblies were around 10 kb, compared with the 70 to 180 kb of genome sequence information in BAC clones from these and other proteorhodopsin-containing microorganisms[39]. These problems probably originate from the high levels of 'microheterogeneity' known to exist within SYMPATRIC bacterioplankton species. The first observations of the SAR11 clade originally revealed substantial rRNA sequence microheterogeneity[97]. It is now known that many bacterioplankton species, including the SAR11 clade, marine crenarchaeota, *Prochlorococcus* spp. and *Vibrio* spp., have high levels of sympatric sequence variation[97–104]. Recent studies indicate that increased sympatric sequence variation in naturally occurring microbial populations cannot be explained either by intracellular variation in rRNA operons or by sequencing errors or PCR artefacts[98,103,105]. Instead, sympatric, nonclonal microbial populations that harbour vast amounts of allelic-sequence diversity seem to be the rule and not the exception in co-occurring species populations. This inherent intra-species genetic complexity, combined with variable species richness and SPECIES EVENNESS, poses challenges for current shotgun sequencing and *in silico* assembly algorithms.

The difficulties encountered in assembling well-known and abundant bacterioplankton genomes serves as a backdrop for a surprising result reported in the Sargasso Sea work. Successful assembly of near-complete genome scaffolds from two different bacterial genera — one *Burkholderia* species and two *Shewanella* types — was reported. The 8.5 Mb scaffold of *Burkholderia* contigs contained several small subunit rRNA genes, each matching sequences from terrestrial *Burkholderia cepacia* strains at ≥98% sequence similarity. The *Burkholderia* assemblies contained low levels of single nucleotide polymorphisms (SNPs) — about 1 polymorphism per 10,000 bases, which is quite unusual considering known levels of sympatric sequence variation in marine bacterioplankton populations. The two other sets of large scaffolds seemed to be derived from *Shewanella* species. These *Shewanella* contig sets were similar in gene order and showed sequence similarity to the genome sequence of *Shewanella oneidensis*, a freshwater, manganese-reducing bacterium. Ribosomal RNA genes in these contigs were ≥98% similar to those of *S. oneidensis* and other closely related terrestrial *Shewanella putrefaciens* strains.

Although the *Shewanella* and *Burkholderia* DNA sequence assembly results are impressive, there remain doubts as to whether these sequences truly represent indigenous Sargasso Sea microorganisms. First, both the *S. oneidensis* and *B. cepacia* genome sequences reported are more similar, in terms of DNA sequence similarity, to terrestrial and freshwater strains. These terrestrial bacterial species have not been reported in significant numbers in open ocean marine bacterioplankton, even though they are readily detected in other habitats. Second, the level of SNPs found in the *B. cepacia* sequence assemblies (1 in 10,000) is astoundingly low, especially for naturally occurring bacterioplankton. Almost every cultivation-independent gene survey of marine microorganisms conducted over the past decade has revealed the presence of extensive sympatric sequence microheterogeneity, even among highly conserved rRNA genes[4,97,98,100–104]. By contrast, the Venter *et al. Burkholderia* assembly reflected a dominant clonal population of a single *B. cepacia* strain[77]. Third, the assembled *Shewanella* and *Burkholderia* genomes (and their rRNAs in individual sequence reads) were significantly represented in only one of the seven Sargasso Sea samples sequenced[14,77], with over 5% of the samples bearing more than 80% DNA sequence homology to *B. cepacia* (FIG. 2). Together, these two groups comprised >50% of all the data in this single sample[77]. Finally, the Sargasso Sea results seem internally inconsistent, as an identical seawater sample, collected on the same day at the same site, contained no *Shewanella* or *Burkholderia*[14,77]. Although the authors suggested that high nutrient sources in marine snow and/or marine mammal vectors might explain the presence of these terrestrial-like *Burkholderia* and *Shewanella* strains, their extraordinarily high

SYMPATRIC
Populations, species or taxa occurring in the same geographical area.

SPECIES EVENNESS
The relative abundance of species in a given habitat, biotope, community or assemblage.

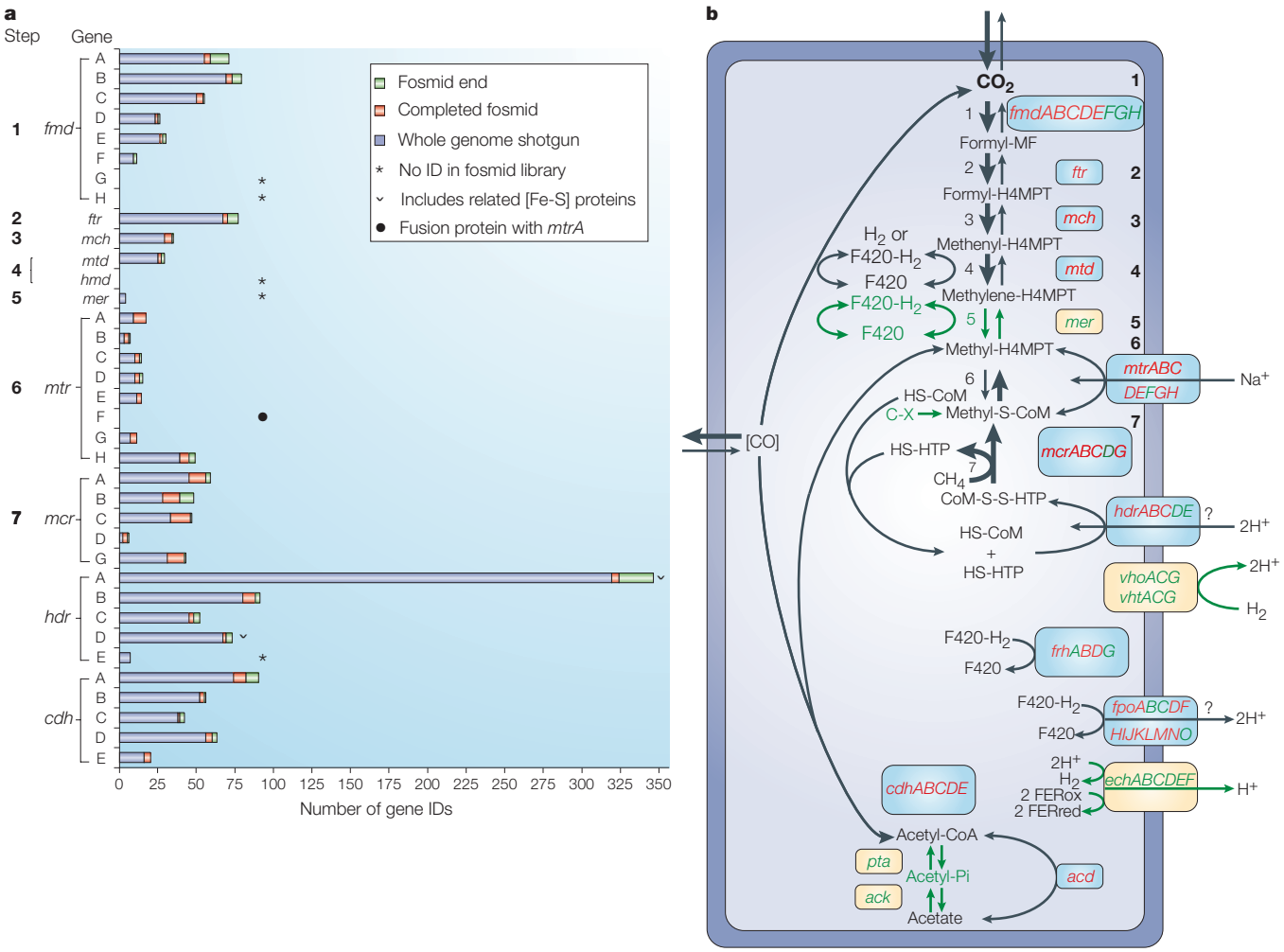

Figure 3 | **Environmental genomics of deep-sea archaeal methanotrophs. a** | Key enzymes of methane metabolism identified in environmental genome datasets of deep-sea archaeal methanotrophs. Steps 1–7 represent the key transformation in methane production[54]. **b** | Postulated pathway of methane oxidation, as inferred from gene distributions in the environmental genomic datasets. Missing genes or pathway steps are shown in green. All but one of the steps (step 5) were shown to be present in and associated with the ANME-1 archaeal methanotroph group[54]. *acd*, ADP-forming acetyl-CoA synthetase; *ack*, acetate kinase; *cdh*, CO dehydrogenase acetyl-CoA synthase; *ech*, nickel–iron-bound hydrogenase; *fmd*, formylmethanofuran dehydrogenase; *fpo,* phenazine oxidoreductase; *frh*, iron–sulphur protein coenzyme F420-reducing hydrogenase; *ftr*, formylmethanofuran-tetrahydromethanopterin cyclohydrolase; *hdr*, heterodisulphide reductase; *mch*, methenyltetrahydromethanopterin cyclohydrolase; *mcr*, methyl-coenzyme M reductase; *mer*, methylenetetrahydromethanopterin reductase; *mtd*, methylenetetrahydromethanopterin dehydrogenase; *mtr*, methyltetrahydromethanopterin S-methyltransferase; *pta*, phosphate acetyltransferase.

biomass and anomalous near-clonal nature render these explanations unlikely. Instead, shipboard contamination of the sample (specifically, sample 1, see FIG. 2) with ALLOCHTHONOUS *Shewanella* and *Burkholderia* seems the most probable explanation. These results emphasize the need for careful sampling and verification procedures, and coordination with field experts in such large-scale sequencing studies. Ecologically oriented population, phylogenetic or quantitative studies using the currently deposited Venter *et al.* dataset should proceed with caution, and might best focus on the (uncontaminated) Sargasso Sea samples 2–7 (REFS 14,77) (FIG. 2). Currently, these individual sample data are only available as sequence trace files in the GenBank database.

Within the Sargasso Sea sequence assemblies, there are also clear misplacements of genes within scaffolds of contigs organized in 'organism bins' (see Online links box). As an example, two large Sargasso Sea scaffolds labelled as Archaea (gi 44893855 and gi 44893849) both contain large portions of bacterial 5S and 23S rRNA genes, embedded within the largely archaeal-protein-encoding gene scaffolds (see also the article by C. Schleper and colleagues in this issue). These bacterial rRNA genes are not examples of lateral gene transfer, but instead represent systematic *in silico* assembly errors within scaffolds. Whereas misassemblies of bacterial rRNAs placed within an archaeal scaffold might be easy to correct, more subtle errors will prob-

ALLOCHTHONOUS
Non-native. Found somewhere other than the place of origin.

ably not be so readily detected and corrected. For example, a misplaced *Vibrio* spp. rRNA assembled within a *Pseudomonas* spp. scaffold would be almost impossible to identify and correct with any certainty. Considering the above, WGS assembly data from mixed microbial communities needs to be approached more cautiously than do sequences derived from individual BACs or whole genome sequences of pure cultures. In analogy to results from the Human Genome Project[70,71], the Sargasso Sea effort demonstrates that shotgun sequencing efforts alone are insufficient to accurately deconvolute contiguous genome sequences from complex, naturally occurring microbial populations.

Nevertheless, the Sargasso Sea dataset represents a very useful resource when placed in the appropriate context and interpreted properly. The value of the dataset is evident from the large number of new genes that it contains. For example, among previously known genes like those encoding proteorhodopsins, the Sargasso Sea shotgun dataset has revealed many new variations on a theme. In addition, newer comparative analyses that take a more general 'gene-centric' approach[55], as opposed to a more 'genome-centric' assembly-based focus[76,77], highlight the value and power of the approach. Important considerations that should guide future studies include well-considered choice of sampling sites and the availability of ancillary data, the cost-to-benefit ratio of deep versus shallow sequencing, and optimizing strategies that leverage the most biological information for the cost, effort and ecological range of communities studied. In many ways, future discussions in this area will not be so different from previous whole-genome sequencing prioritization, which focused on enhancing phylogenetic breadth and depth.

Genomic approaches for reconstructing* in situ *metabolism. In deep-sea marine sediments near CONTINENTAL MARGINS, large quantities of methane are stored in reservoirs of solid gas hydrates that congeal in low-temperature, high-pressure environments. The fact that little of this methane escapes into the overlying water column has been a geochemical mystery for some time. Recently, it has been discovered that sediment-dwelling deep-sea microorganisms can consume this methane anaerobically and couple this methane oxidation to sulphate reduction. Although no anaerobic methane-oxidizing microorganisms have been cultivated so far, rRNA surveys[106] and stable isotope analyses[107,108] have revealed the presence of methanogen-related archaea that are responsible for this process. Consortia of these archaeal methanotrophs, along with sulphate-reducing bacteria, can oxidize methane to CO_2 with concomitant sulphate reduction in these deep-sea, methane-rich habitats. As they have not been cultivated, however, the biochemical pathway for this unusual methane metabolism remained undescribed.

Hallam *et al.* recently applied community genomic approaches to characterize these archaeal methanotrophs from complex sediment microbial communities in the deep-sea[54,109]. This study used an alternative strategy intended to reduce the overall complexity of the microbial community before sequencing and provide more reliable and accurate coverage. As each gram of sediment contains a diverse array of microorganisms, the deep-sea-sediment microorganisms were first subjected to density-gradient separation, followed by size fractionation by filtration. Genomic DNA from the purified methanotroph-enriched cells was then used to produce small-insert (3 kb)WGS libraries and large-insert (average insert size ~36 kb) fosmid libraries. Shotgun sequences, fosmid-end sequences and full-length fosmid sequences were then produced and analysed with a specific focus on the putative methane metabolism. A total of 111 Mb of small-insert shotgun sequence, 4.6 Mb of fosmid-end sequence and 7.4 Mb of complete fosmid sequence was produced. The 191 fosmids selected for complete sequencing were chosen based on paired-end sequencing combined with information on rRNA and functional gene content. Phylogenetic analysis of the shotgun and fosmid sequence data verified strong enrichment in the sample for archaeal methanotrophs in the purified cell preparation. Out of 114 rRNA genes found among shotgun sequences, 66% were derived from archaeal methanotrophs, with 46% from one specific lineage, ANME-1. The ANME-1-derived fosmids were identified based on gene-content phylogenetic analyses, as well as GC content versus read-depth correlations. This approach combined most fosmids sequenced into one group of high-read-depth fosmids with an average GC content of 45%. On the basis of independent evidence, this bin of 51 fosmids originated from the ANME-1 group of archaeal methanotrophs.

Homologues of known genes in the methanogenesis pathway mapped specifically to the ANME-1 genome-sequence bin, supporting previously hypothesized mechanisms of archaeal methane oxidation. In support of the 'reverse methanogenesis' hypothesis, expected components of the methanogenic pathway were present at similar frequencies and stoichiometry in both shotgun and fosmid sequence datasets (FIG. 3), and were specifically linked to the highly represented ANME-1 genome bin[54]. Genes encoding proteins involved in all but one of the seven steps of the methanogenic pathway were found to be associated with the ANME-1 archaeal methanotrophs[54] (FIG. 3a). Only one central gene in the methane-metabolizing pathway (*mer*, methylenetetrahydromethanopterin reductase), which has been found in all other methanogens examined to date, seemed to be missing. The *mer* gene encodes an enzyme that catalyses a key reductive step in the methanogenic pathway. Alteration of this step could regulate the directional flux of carbon for archaeal methanotrophs in the oxidative direction (FIG. 3b). Additionally, the absence of this key gene in the ANME-1 group might indicate that archaeal methanotrophs have lost the ability to generate methane and are now committed to a methane-consuming lifestyle. These genome-based results are consistent with other genetic and biochemical studies conducted on archaeal methanotrophs[78,79].

CONTINENTAL MARGIN
Shallow submarine extension of the continents, generally tens of meters deep, that extends seaward to the continental slope and the deep ocean.

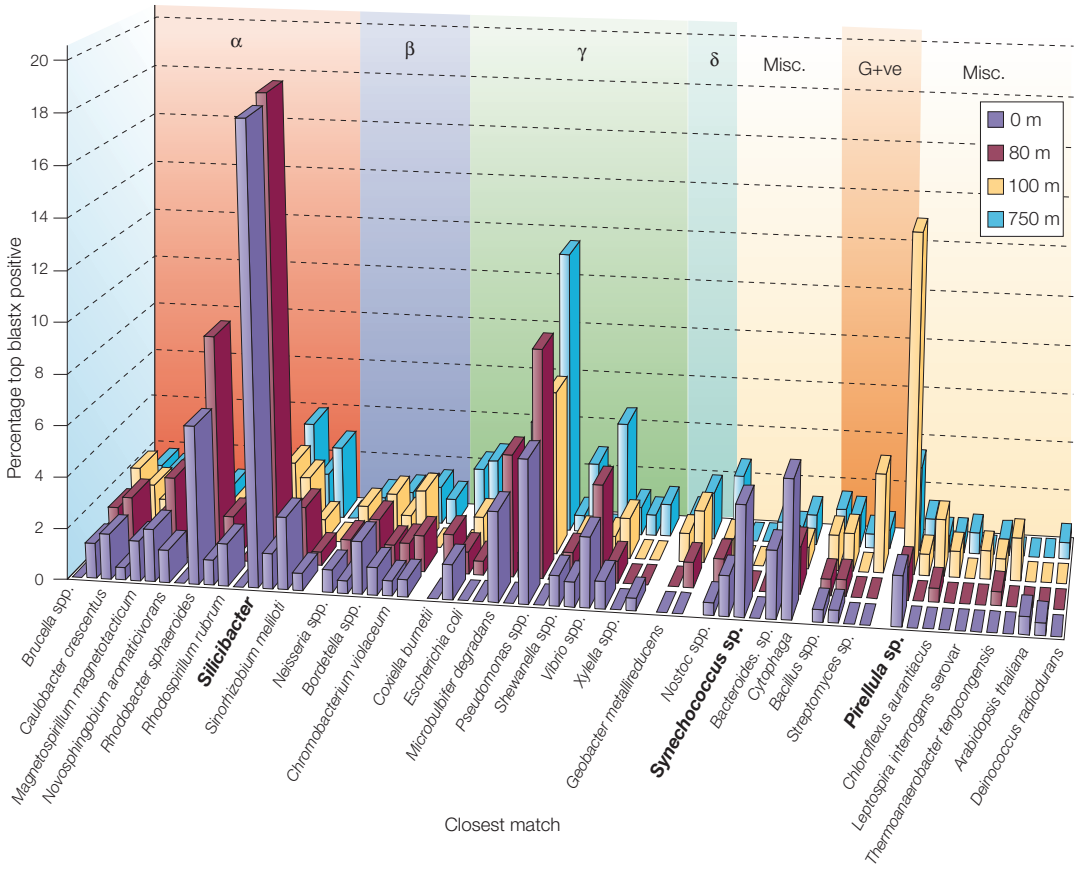

Figure 4 | **Apparent taxonomic affiliation of protein-encoding genes from different depths in Monterey Bay.** The percentage of predicted open reading frames that most closely match the indicated taxon present in bacterial artificial chromosomes (BACs) collected at each different depth. BLAST[112] analyses were carried out using blastx and an expectation cutoff value of 1×10^{-8}. For a description of the BAC libraries and associated metadata see REF. 111. The primary sequence data are available at http://www.tigr.org/tdb/MBMO. *Silicibacter* sp.-like protein sequences represented 18% of the total sequences in both 0 m and 80 m BAC libraries from Monterey Bay. Ribosomal-RNA gene sequences from *Silicibacter* sp.-like bacteria represented 21.1%, and 23.6% of the total rRNA genes in the same 0 m and 80 m BAC libraries, respectively[111]. G+ve, Gram-positive.

This research demonstrates that more detailed analysis of multicomponent metabolic pathways of uncultured microorganisms from complex microbial assemblages is an achievable goal (FIG. 3b). In addition to providing new information on the potential metabolic pathways in archaeal methanotrophs, this study highlights several promising directions for future environmental genomic studies[54]. First, reduction of microbial community complexity before DNA extraction can prove extremely useful. Targeted microbial groups can be physically separated before genomic characterization, which greatly simplifies downstream analyses. Other potential techniques to separate DNAs on the basis of size include the use of laser tweezers or flow cytometry. Single-cell isolation techniques coupled with genome amplification strategies, such as rolling-circle amplification, offer promising future alternatives for deconvoluting single genome sequences from complex microbial populations[110]. Such purification strategies complement, but do not replace, whole-community genome surveys and analyses. Second, this study shows that combined use of small-insert and large-insert libraries (in this case, shotgun and fosmid libraries) is an obvious and important application that has been essential to most whole-genome projects conducted to date. Combined WGS and large-genome-fragment analyses provides more complete information than either strategy does alone.

From systems biology to systems ecology

In the near future, ocean microbial genomics will continue to mine complex community datasets to better understand how community gene content maps onto taxonomic composition, metabolic repertoire and phenotypic expression. Current efforts are starting to define the 'parts list'. Systematic sampling, together with other environmental physical and chemical data, promises to increase our current knowledge of microbial genotypes and phenotypes in the oceans. Important issues to tackle include assessing the relative quality and usefulness of different data types such as WGS assemblies, completed BACs and whole genomes. Additionally, organizing appropriate community data centres that can accommodate environmental metadata not currently present in sequence databases, such as sampling location and physical and chemical parameters, will be an

important development. In the future, careful coordination of sampling, genome sequence analyses and downstream field efforts will increase the value and usefulness of environmental genomic datasets.

One exciting new direction includes comparative genomics approaches within microbial community genomics. Comparisons among environmental genomic datasets originating from different microbial habitats are now starting to emerge and will probably become a common strategy for comparing different microbial communities and their genomic and metabolic potential. Tringe *et al.*[55] recently compared the acid mine drainage, Sargasso Sea, whale fall and silage shotgun datasets derived from associated microbial communities. Microbial gene content in the Sargasso Sea was fundamentally different from the acid mine drainage microbial community, the anaerobic 'whale bone communities' and silage, in fairly predictable ways[55]. Whereas these vastly different microbial habitats and communities can be readily discriminated, could useful biological information be extracted from more similar assemblages that have more subtle differences, using comparative community genomic approaches? Are specific adaptations detectable along the depth, light or redox gradients, for example? Can lateral gene transfer be 'caught in the act', along such gradients of varying population composition and selective pressures? Can such gene flow and the selective pressures that enrich for particular metabolic features be detected without any *a priori* knowledge of the specific genes or selection pressures involved?

Genomic comparisons of bacterioplankton along temporal and spatial gradients are beginning to indicate that even modest sequencing efforts can yield important information about microbial distributions, population structure and function. For example, by randomly sequencing only 500 bp from the ends of approximately 1,000 BAC clones from libraries prepared at different depths, insight can be obtained on population structure, taxa distribution and functional gene content in the water column (FIG. 4). Important differences that reflect microbial population structure in the water column and identify differences between depth-stratified microbial populations are evident at the genomic level. In general, these genomically inferred profiles map reasonably well onto other independent analyses. For example, the concentrations of *Synechococcus* sp. inferred from the BAC-derived shotgun sequence data are consistent with typical depth distributions determined by epifluorescence microscopic counts or flow cytometric data. In addition, the depth distribution and abundance of *Silicibacter*-like genome sequences is internally consistent with rRNA gene counts within the same BAC libraries[111], as well as with predicted abundances of *Silicibacter*-like microorganisms in coastal surface waters[14].

Ecologically and biogeochemically relevant information can also be extracted from comparative analyses of gene and taxon distributions inferred from shotgun sequence distributions. One advantage to analyses conducted with BACs is that, attached to the 500–1,000 bp of shotgun sequence information, 35–200 kb more of detailed, contiguous genome sequence information can be obtained in association with each and every sequence. Similar correlative gene surveys and comparative community genomic analyses will undoubtedly shape the future of how we analyse and interpret natural microbial community structure and function.

Microbial community genomics can now provide a deeper view of the network of genes, genomes, organisms and communities present in the natural world. These studies cross the traditional disciplinary boundaries of microbial ecology, evolution, biogeochemistry and Earth science. Integrated studies linking multiple hierarchical levels of biological organization, from metabolic pathways to ecosystem modelling, represent one possible outcome of current research. Systems biology seeks to interpret the connections and interactions between various parts of complex biological networks, usually at the intracellular level. By contrast, systems ecology seeks to define higher-order relationships between complex biotic and abiotic Earth systems at the ecosystem level. The application and 'postgenomic' use of microbial community genome datasets might help to bridge the gap between highly reductionist molecular approaches and more holistic views of biotic systems at the ecosystem level. New analyses and theoretical developments, spanning from genomes to biomes, will soon begin to more closely approximate how the complex Earth system, largely driven by biological processes and evolution, operates. The oceans represent an outstanding environment in which to begin , in part because microbial populations there are stably stratified on spatial scales (from meters to tens of meters) that can be readily and reproducibly sampled. Future ocean-going microbial genomic surveys will undoubtedly have a lot to teach us over the coming years.

1. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
2. Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. & Stahl, D. A. Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* **40**, 337–365 (1986).
3. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
4. Rappe, M. S. & Giovannoni, S. J. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**, 369–394 (2003).
5. Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774 (1998).
6. DeLong, E. F. Microbial seascapes revisited. *Curr. Opin. Microbiol.* **4**, 290–295 (2001).
7. Wright, T. D., Vergin, K. L., Boyd, P. W. & Giovannoni, S. J. A novel δ-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl. Environ. Microbiol.* **63**, 1441–1448 (1997).
8. DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
9. Bult, C. J. *et al.* Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073 (1996).
10. Dufresne, A. *et al.* Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl Acad. Sci. USA* **100**, 10020–10025 (2003).
11. Rocap, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042–1047 (2003).
This paper reports on the comparative genomic analysis of one of the most abundant photosynthetic organisms in the oceans. The authors report on depth-specific adaptations and genomic differences to gradients of light and nutrients found in the water column.
12. Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* **424**, 1037–1042 (2003).

13. Hou, S. *et al.* Genome sequence of the deep-sea γ-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc. Natl Acad. Sci. USA* **101**, 18036–18041 (2004).
14. Moran, M. A. *et al.* Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**, 910–913 (2004).
This paper reports on the genome sequence of a marine bacterium related to *Roseobacter* spp., an α-proteobacterial group that is abundant in marine surface waters. The authors suggest the presence of specific adaptations to the marine environment with respect to carbon and sulphur metabolism, and test these hypotheses in physiological experiments.
15. Chen, C. Y. *et al.* Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome Res.* **13**, 2577–2587 (2003).
16. Makino, K. *et al.* Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet* **361**, 743–749 (2003).
17. Nelson, K. E. *et al.* Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329 (1999).
18. Rabus, R. *et al.* The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ. Microbiol.* **6**, 887–902 (2004).
19. Glockner, F. O. *et al.* Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc. Natl Acad. Sci. USA* **100**, 8298–8303 (2003).
20. Takami, H., Takaki, Y. & Uchiyama, I. Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. *Nucleic Acids Res.* **30**, 3927–3935 (2002).
21. Vezzi, A. *et al.* Life at depth: *Photobacterium profundum* genome sequence and expression analysis. *Science* **307**, 1459–1461 (2005).
This paper reports the first DNA sequence of a piezophilic bacterium adapted to high hydrostatic pressures encountered in the deep sea. The sequence was used in microarray analyses to infer regulatory responses to changing hydrostatic pressure.
22. Ruby, E. G. *et al.* Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. *Proc. Natl Acad. Sci. USA* **102**, 3004–3009 (2005).
23. Chinen, A., Uchiyama, I. & Kobayashi, I. Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction–modification genes with large genome polymorphisms. *Gene* **259**, 109–121 (2000).
24. Galagan, J. E. *et al.* The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* **12**, 532–542 (2002).
25. Hendrickson, E. L. *et al.* Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. *J. Bacteriol.* **186**, 6956–6969 (2004).
26. Kawarabayasi, Y. *et al.* Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**, 55–76 (1998).
27. Kawarabayasi, Y. *et al.* Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* **6**, 83-101, 145–152 (1999).
28. Klenk, H. P. *et al.* The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370 (1997).
29. Robb, F. T. *et al.* Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods Enzymol.* **330**, 134–157 (2001).
30. Slesarev, A. I. *et al.* The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl Acad. Sci. USA* **99**, 4644–4649 (2002).
31. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
This report describes the draft sequencing of the 34-Mb genome of marine diatom *Thalassiosira pseudonana*, which has 24 diploid nuclear chromosomes. Metabolic features inferred from the genome sequence were related to the growth strategies and ecology of this diatom. Evidence for the secondary endosymbioses with a red algal endosymbiont was found in the nuclear genome.
32. Schmidt, T. M., DeLong, E. F. & Pace, N. R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**, 4371–4378 (1991).
33. Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178**, 591–599 (1996).
34. Béjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
35. Béjà, O. *et al.* Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**, 630–633 (2002).
36. Breitbart, M. *et al.* Genomic analysis of uncultured marine viral communities. *Proc. Natl Acad. Sci. USA* **99**, 14250–14255 (2002).
37. Breitbart, M. *et al.* Diversity and population structure of a near-shore marine-sediment viral community. *Proc. Biol. Sci.* **271**, 565–574 (2004).
This report describes the genetic make-up of two viral shotgun libraries from phage populations collected in coastal waters of San Diego. The authors show that double stranded DNA tailed phages and algal phages were well represented in the sample. Additionally, the most abundant viral group detected by shotgun cloning could represent upwards of 3% of the total phage.
38. de la Torre, J. R. *et al.* Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl Acad. Sci. USA* **100**, 12830–12835 (2003).
39. Sabehi, G., Beja, O., Suzuki, M. T., Preston, C. M. & DeLong, E. F. Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ. Microbiol.* **6**, 903–910 (2004).
40. Schleper, C. *et al.* Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **180**, 5003–5009 (1998).
41. Vergin, K. L. *et al.* Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. *Appl. Environ. Microbiol.* **64**, 3075–3078 (1998).
42. Rondon, M. R. *et al.* Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547 (2000).
43. Lopez-Garcia, P., Brochier, C., Moreira, D. & Rodriguez-Valera, F. Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ. Microbiol.* **6**, 19–34 (2004).
44. Treusch, A. H. *et al.* Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ. Microbiol.* **6**, 970–980 (2004).
45. Moreira, D., Rodriguez-Valera, F. & Lopez-Garcia, P. Analysis of a genome fragment of a deep-sea uncultivated group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. *Environ. Microbiol.* **6**, 959–969 (2004).
46. Quaiser, A. *et al.* Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol. Microbiol.* **50**, 563–575 (2003).
47. Quaiser, A. *et al.* First insight into the genome of an uncultivated crenarchaeote from soil. *Environ. Microbiol.* **4**, 603–611 (2002).
48. Schleper, C., Swanson, R. V., Mathur, E. J. & DeLong, E. F. Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **179**, 7803–7811 (1997).
49. Preston, C. M., Wu, K. Y., Molinski, T. F. & DeLong, E. F. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc. Natl Acad. Sci. USA* **93**, 6241–6246 (1996).
50. Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669–685 (2004).
51. Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38**, 525–552 (2004).
52. Stein, J. L., Haygood, M. & Felbeck, H. Nucleotide sequence and expression of a deep-sea ribulose-1,5-bisphosphate carboxylase gene cloned from a chemoautotrophic bacterial endosymbiont. *Proc. Natl Acad. Sci. USA* **87**, 8850–8854 (1990).
53. Cottrell, M. T., Moore, J. A. & Kirchman, D. L. Chitinases from uncultured marine microorganisms. *Appl. Environ. Microbiol.* **65**, 2553–2557 (1999).
54. Hallam, S. J. *et al.* Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**, 1457–1462 (2004).
55. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308**, 554–557 (2005).
56. DeLong, E. F. in *Microbial Genomics* (eds Fraser, C. M., Nelson., K. E. & Read, T. D.) 419–442 (Human Press Inc., Totowa, New Jersey, 2004).
57. DeLong, E. F. Microbial population genomics and ecology. *Curr. Opin. Microbiol.* **5**, 520–524 (2002).
58. DeLong, E. F. Towards microbial systems science: integrating microbial perspective, from genomes to biomes. *Environ. Microbiol.* **4**, 9–10 (2002).
59. Rodriguez-Valera, F. Approaches to prokaryotic biodiversity: a population genetics perspective. *Environ. Microbiol.* **4**, 628–633 (2002).
60. Rodriguez-Valera, F. Environmental genomics, the big picture? *FEMS Microbiol. Lett.* **231**, 153–158 (2004).
61. Beja, O. To BAC or not to BAC: marine ecogenomics. *Curr. Opin. Biotechnol.* **15**, 187–190 (2004).
62. Doney, S. C., Abbott, M. R., Cullen, J. J., Karl, D. M. & Rothstein, L. From genes to ecosystems: the ocean's new frontier. *Frontiers in Ecology and the Environment* **2**, 457–466 (2004).
63. Kim, U.-J., Shizuya, H., Dejong, P., Birren, B. & Simon, M. Stable propagation of cosmid sized human DNA inserts in an F-factor based vector. *Nucleic Acids Res.* **20**, 1083–1185 (1992).
64. Shizuya, H. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA *in Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794–8797 (1992).
65. Béjà, O. *et al.* Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* **2**, 516–529 (2000).
66. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995).
67. Tamas, I. *et al.* 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**, 2376–2379 (2002).
68. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81–86 (2000).
69. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
70. Waterston, R. H., Lander, E. S. & Sulston, J. E. On the sequencing of the human genome. *Proc. Natl Acad. Sci. USA* **99**, 3712–3716 (2002).
71. Waterston, R. H., Lander, E. S. & Sulston, J. E. More on the sequencing of the human genome. *Proc. Natl Acad. Sci. USA* **100**, 3022–3024; author reply 3025–3026 (2003).
72. Oz, A., Sabehi, G., Koblizek, M., Massana, R. & Beja, O. *Roseobacter*-like bacteria in Red and Mediterranean Sea aerobic anoxygenic photosynthetic populations. *Appl. Environ. Microbiol.* **71**, 344–353 (2005).
73. Sabehi, G. *et al.* Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ. Microbiol.* **5**, 842–849 (2003).
74. Man, D. *et al.* Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**, 1725–1731 (2003).
75. Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
76. Falkowski, P. G. & de Vargas, C. Shotgun sequencing in the sea: a blast from the past? *Science* **304**, 58–60 (2004).
77. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
78. Kruger, M. *et al.* A conspicuous nickel protein in microbial mats that oxidize methane anaerobically. *Nature* **426**, 878–881 (2003).
79. Teeling, H., Meyerdierks, A., Bauer, M., Amann, R. & Glockner, F. O. Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ. Microbiol.* **6**, 938–947 (2004).
80. Bielawski, J. P., Dunn, K. A., Sabehi, G. & Beja, O. Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc. Natl Acad. Sci. USA* **101**, 14824–14829 (2004).
81. Krebs, R. A., Alexiev, U., Partha, R., DeVita, A. M. & Braiman, M. S. Detection of fast light-activated H+ release and M intermediate formation from proteorhodopsin. *BMC Physiol.* **2**, 5 (2002).
82. Friedrich, T. *et al.* Proteorhodopsin is a light-driven proton pump with variable vectoriality. *J. Mol. Biol.* **321**, 821–838 (2002).
83. Varo, G., Brown, L. S., Lakatos, M. & Lanyi, J. K. Characterization of the photochemical reaction cycle of proteorhodopsin. *Biophys. J.* **84**, 1202–1207 (2003).
84. Lakatos, M., Lanyi, J. K., Szakacs, J. & Varo, G. The photochemical reaction cycle of proteorhodopsin at low pH. *Biophys. J.* **84**, 3252–3256 (2003).

85. Dioumaev, A. K. *et al.* Proton transfers in the photochemical reaction cycle of proteorhodopsin. *Biochemistry* **41**, 5348–5358 (2002).
86. Wang, W. W., Sineshchekov, O. A., Spudich, E. N. & Spudich, J. L. Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. *J. Biol. Chem.* **278**, 33985–33991 (2003).
87. Lakatos, M. & Varo, G. The influence of water on the photochemical reaction cycle of proteorhodopsin at low and high pH. *J. Photochem. Photobiol. B* **73**, 177–182 (2004).
88. Man-Aharonovich, D. *et al.* Characterization of RS29, a blue-green proteorhodopsin variant from the Red Sea. *Photochem. Photobiol. Sci.* **3**, 459–462 (2004).
89. Bergo, V., Amsden, J. J., Spudich, E. N., Spudich, J. L. & Rothschild, K. J. Structural changes in the photoactive site of proteorhodopsin during the primary photoreaction. *Biochemistry* **43**, 9075–9083 (2004).
90. Imasheva, E. S., Balashov, S. P., Wang, J. M., Dioumaev, A. K. & Lanyi, J. K. Selectivity of retinal photoisomerization in proteorhodopsin is controlled by aspartic acid 227. *Biochemistry* **43**, 1648–1655 (2004).
91. Dioumaev, A. K., Wang, J. M., Balint, Z., Varo, G. & Lanyi, J. K. Proton transport by proteorhodopsin requires that the retinal Schiff base counterion Asp-97 be anionic. *Biochemistry* **42**, 6582–6587 (2003).
92. Kelemen, B. R., Du, M. & Jensen, R. B. Proteorhodopsin in living color: diversity of spectral properties within living bacterial cells. *Biochim. Biophys. Acta* **1618**, 25–32 (2003).
93. Kolber, Z. S., Van Dover, C. L., Niederman, R. A. & Falkowski, P. G. Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407**, 177–179 (2000).
94. Kolber, Z. S. *et al.* Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* **292**, 2492–2495 (2001).
95. Pradella, S. *et al.* Genome organization and localization of the *pufLM* genes of the photosynthesis reaction center in phylogenetically diverse marine α-proteobacteria. *Appl. Environ. Microbiol.* **70**, 3360–3369 (2004).
96. Allgaier, M., Uphoff, H., Felske, A. & Wagner-Dobler, I. Aerobic anoxygenic photosynthesis in *Roseobacter* clade bacteria from diverse marine habitats. *Appl. Environ. Microbiol.* **69**, 5051–5059 (2003).
97. Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60–63 (1990).
98. Acinas, S. G. *et al.* Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**, 551–554 (2004).
99. Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**, 464–467 (1998).
100. Field, K. G. *et al.* Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl. Environ. Microbiol.* **63**, 63–70 (1997).
101. Béjà, O. *et al.* Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* **68**, 335–345 (2002).
102. Garcia-Martinez, J. & Rodriguez-Valera, F. Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I. *Mol. Ecol.* **9**, 935–948 (2000).
103. Klepac-Ceraj, V. *et al.* High overall diversity and dominance of microdiverse relationships in salt marsh sulphate-reducing bacteria. *Environ. Microbiol.* **6**, 686–698 (2004).
104. Thompson, J. R. *et al.* Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**, 1311–1313 (2005).
This report examined the diversity of *Vibrio splendidus* isolates using rRNA and Hsp60 to examine the genetics of a single naturally occurring bacterial species. The extent of sympatric, co-occurring genetic diversity was remarkable, with this *Vibrio* species showing extensive allelic and genome size variation.
105. Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. & Polz, M. F. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J. Bacteriol.* **186**, 2629–2635 (2004).
106. Hinrichs, K. U., Hayes, J. M., Sylva, S. P., Brewer, P. G. & DeLong, E. F. Methane-consuming archaebacteria in marine sediments. *Nature* **398**, 802–805 (1999).
107. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc. Natl Acad. Sci. USA* **99**, 7663–7668 (2002).
108. Orphan, V. J., House, C. H., Hinrichs, K. U., McKeegan, K. D. & DeLong, E. F. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293**, 484–487 (2001).
109. Hallam, S. J., Girguis, P. R., Preston, C. M., Richardson, P. M. & DeLong, E. F. Identification of methyl coenzyme M reductase A (*mcrA*) genes associated with methane-oxidizing archaea. *Appl. Environ. Microbiol.* **69**, 5483–5491 (2003).
110. Detter, J. C. *et al.* Isothermal strand-displacement amplification applications for high-throughput genomics. *Genomics* **80**, 691–698 (2002).
111. Suzuki, M. T. *et al.* Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. *Microb. Ecol.* **48**, 473–488 (2004).
112. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
113. Fitz-Gibbon, S. T. *et al.* Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc. Natl Acad. Sci. USA* **99**, 984–989 (2002).

Acknowledgements

I am indebted to all my students and collaborators, past and present, for their dedication, inspiration and perspiration. Thanks to S. Hallam for providing FIG. 3 and to my oceanographic collaborators and colleagues, especially F. Chavez at the Monterey Bay Aquarium Research Institute and D. Karl at the University of Hawaii, for ongoing collaborative field efforts. The author's work is supported by the National Science Foundation, the Gordon and Betty Moore Foundation, and sequencing support from the Department of Energy (DoE) carried out at the DoE Joint Genome Institute.

Competing interests statement

The author declares no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez: http://www.ncbi.nlm.nih.gov/Entrez
gi44893849 | gi44893855 | *Methanocaldococcus jannaschii* | *Shewanella oneidensis* | *Synechococcus* sp.

FURTHER INFORMATION

Edward F. DeLong's homepage:
http://web.mit.edu/be/people/delong.htm
DoE Joint Genome Institute:
http://genome.jgi-psf.org/microbial
The Gordon and Betty Moore Foundation:
http://www.moore.org/program_areas/science/initiatives/marine_microbiology/initiative_marine_microbiology.asp
NSF Microbial Sequencing Program FY 2005:
http://www.nsf.gov/pubs/2005/nsf05512/nsf05512.htm
The organism bins assembled from the Sargasso Sea WGS environmental sample dataset:
http://www.ncbi.nlm.nih.gov:80/books/bv.fcgi?rid=coffeebrk.table.634
Access to this interactive links box is free online.