

# Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time

Ramunas Stepanauskas\* and Michael E. Sieracki

Bigelow Laboratory for Ocean Sciences, P.O. Box 475, West Boothbay Harbor, ME 04575-0475

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved April 11, 2007 (received for review January 18, 2007)

The identification of predominant microbial taxa with specific metabolic capabilities remains one of the biggest challenges in environmental microbiology, because of the limits of current metagenomic and cell culturing methods. We report results from the direct analysis of multiple genes in individual marine bacteria cells, demonstrating the potential for high-throughput metabolic assignment of yet-uncultured taxa. The protocol uses high-speed fluorescence-activated cell sorting, whole-genome multiple displacement amplification (MDA), and subsequent PCR screening. A pilot library of 11 single amplified genomes (SAGs) was constructed from Gulf of Maine bacterioplankton as proof of concept. The library consisted of five flavobacteria, one sphingobacterium, four alphaproteobacteria, and one gammaproteobacterium. Most of the SAGs, apart from alphaproteobacteria, were phylogenetically distant from existing isolates, with 88–97% identity in the 16S rRNA gene sequence. Thus, single-cell MDA provided access to the genomic material of numerically dominant but yet-uncultured taxonomic groups. Two of five flavobacteria in the SAG library contained proteorhodopsin genes, suggesting that flavobacteria are among the major carriers of this photometabolic system. The *pufM* and *nasA* genes were detected in some 100-cell MDA products but not in SAGs, demonstrating that organisms containing bacteriochlorophyll and assimilative nitrate reductase constituted <1% of the sampled bacterioplankton. Compared with metagenomics, the power of our approach lies in the ability to detect metabolic genes in uncultured microorganisms directly, even when the metabolic and phylogenetic markers are located far apart on the chromosome.

Flavobacteria | flow cytometry | multiple displacement amplification | proteorhodopsin | single-cell genomics

The PCR- and direct cloning-based sequencing of environmental DNA extracts has revealed the enormous previously unknown phylogenetic and metabolic diversity of prokaryotes (1–9). Although yet-uncultured taxa are believed to comprise >99% of all prokaryotes, their metabolic capabilities and ecological functions remain enigmatic, largely because of methodological limitations. For example, PCR-based clone libraries are intrinsically limited to the analysis of one gene at a time, with no direct way of linking libraries of diverse genes. Large-scale environmental shotgun sequencing, although extremely productive for finding novel genes, is prohibitively expensive, and so far is limited to only partial genome assembly of the most numerically dominant taxa in complex marine microbial communities (5, 9). Genomic analyses of large environmental DNA inserts can lead to remarkable discoveries, such as proteorhodopsin genes in bacteria (2). However, large insert-based function assignment is limited to situations where the metabolic gene of interest is located near phylogenetic markers (e.g., ribosomal genes). Thus, currently available culture-independent research tools are poorly suited for identification of microorganisms with specific metabolic characteristics. This significantly limits the progress in such diverse fields as biogeochemistry, microbial ecology and evolution, and bioprospecting.

We propose sequencing multiple DNA loci in individual bacterial cells rather than environmental DNA extracts, as a

more productive alternative for metabolic mapping of uncultured microorganisms. This strategy has been gaining momentum, with recent implementations of single-cell multiplex PCR in termite gut microbiota by Ottesen *et al.* (10) and partial genome sequencing of single cells of *Prochlorococcus* by Zhang *et al.* (11). Here we show important improvements in single-cell separation and DNA analysis protocols and demonstrate a proof-of-concept metabolic mapping of taxonomically diverse marine bacterioplankton.

One of the main challenges for single-cell studies is efficient and contamination-free separation of individual cells from other microorganisms and extracellular DNA in an environmental sample. Prior studies of DNA in individual prokaryote cells used serial sample dilution (11), dilution using microfluidics (10), or micromanipulation (12). Here we used high-speed droplet-based FACS. Compared with alternative methods, FACS offers several critical advantages, including high-throughput rates and the ability to sort targeted plankton groups, based on cell size and fluorescence signals of natural cell components and fluorochromes (13). Furthermore, cell separation by FACS creates microsamples containing the target cell and only 3–10 pl of sample around it (13). This reduces the codeposition of extracellular DNA, which in marine waters occurs at concentrations similar to cell-bound DNA (14, 15).

Current sequencing technologies require nanogram-to-microgram DNA templates and are not capable of direct sequencing of individual DNA molecules. Thus, DNA preamplification is necessary to sequence genes or genomes from individual cells. For the analysis of up to two loci per cell, single cell multiplex PCR has been used in medical research since the 1980s (16) and was recently used in an environmental microbiology study (10). As a more versatile alternative, allowing for analysis of an unlimited number of loci, several methods have been suggested for whole-genome amplification, including degenerated oligonucleotide-primed PCR, primer extension preamplification, ligation-mediated PCR, and multiple displacement amplification (MDA) using phi29 or Bst DNA polymerases (17, 18). Among them, phi29-based MDA appears the most suitable for efficient whole-genome amplification with low error and bias (17, 18) and is capable of generating micrograms of genomic DNA from nanogram-sized samples (19–21). Recently, Phi29-based MDA was used on single human (18, 22, 23), *Escherichia coli* (24), and *Prochlorococcus* (11) cells. Direct

Author contributions: R.S. and M.E.S. designed research; R.S. and M.E.S. performed research; R.S. analyzed data; and R.S. and M.E.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: MDA, multiple displacement amplification; SAG, single amplified genome; SSU rRNA, small-subunit rRNA; HNA, high nucleic acid; T-RFLP, terminal restriction fragment length polymorphism.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EF202334–EF202347 and EF508145–EF508148).

\*To whom correspondence should be addressed. E-mail: rstepanuskas@bigelow.org.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0700496104/DC1](http://www.pnas.org/cgi/content/full/0700496104/DC1).

© 2007 by The National Academy of Sciences of the USA

**Table 1. Phylogeny of bacterial SSU rRNA genes obtained from single amplified genomes**

SAG ID	Lysis protocol	Genus*	Closest isolate <sup>†</sup>	Closest sequence <sup>‡</sup>	T-RFL (HhaI), bp	T-RFL (HaeIII), bp
<b>Flavobacteria/Flavobacteriaceae</b>						
MS021-5C	A	<i>Kordia</i> , 26	Flavobacterium sp. 3034 AM110988, 91	Clone NorSea37 AM279169, 96	90	283
MS024-2A	B	<i>Kordia</i> , 36	Flavobacterium sp. 3034 AM110988, 91	Clone NorSea43 AM279191, 99	94	No cut
MS024-3C	B	<i>Cellulophaga</i> , 80	<i>Cellulophaga</i> sp. CC12 DQ356487, 93	Clone 1D10 AY274838, 99	96	32
MS024-1F	B	<i>Tenacibaculum</i> , 98	Sponge bacterium Zo9 AY948376, 97	Clone WLB13-197 DQ015841, 96	90	281
MS056-2A	C	<i>Ulvibacter</i> , 99	<i>Ulvibacter litoralis</i> AY243096, 95	Clone PB1.23 DQ071072, 99	94	284
<b>Sphingobacteria/Saprospiraceae</b>						
MS190-1F	B	<i>Heliscomenobacter</i> , 55	Saprospiraceae bacterium MS-Wolf2-H AJ786323, 88	Clone SanDiego3-A7 DQ671753, 100	92	407
<b>Alphaproteobacteria/Rhodobacteraceae</b>						
MS056-3A	C	<i>Sulfitobacter</i> , 98	<i>Roseobacter</i> sp. AY167254, 99	Clone F3C24 AY794157, 100	55	32
MS024-1C	B	<i>Jannaschia</i> , 60	<i>Ophiopholis aculeata</i> symbiont U63548, 99	Clone EB080-L11F12 AY627365, 100	55	32
MS190-2A	B	<i>Jannaschia</i> , 55	<i>O. aculeata</i> symbiont U63548, 99	Clone EB080-L11F12 AY627365, 100	55	32
MS190-2F	B	<i>Loktanelia</i> , 41	<i>Octadecabacter orientus</i> KOPRI 13313 DQ167247, 97	Rhodobacteraceae bact. 183 AJ810844.1, 99	55	32
<b>Gammaproteobacteria/Oceanospirillaceae</b>						
MS024-3A	B	<i>Balneatrix</i> , 24	Marine gammaproteobacterium HTCC2120 AY386340, 90	Clone Ant4D3 DQ295237, 99	575	413
<b>Gammaproteobacteria/Comamonadaceae</b>						
MS024-2C	B	<i>Delftia</i> , 100	<i>Delftia acidovorans</i> AM180725, 99	<i>D. acidovorans</i> AM180725, 99	203	197

\*Determined by RDP Classifier. Both type and nontype "good"  $\geq 1,200$ -bp sequences were used. Numbers indicate confidence level (%) for genus identification.

<sup>†</sup>Determined by RDP Seqmatch. Provided are sequence ID, GenBank accession no., and sequence identity to the SAG (%).

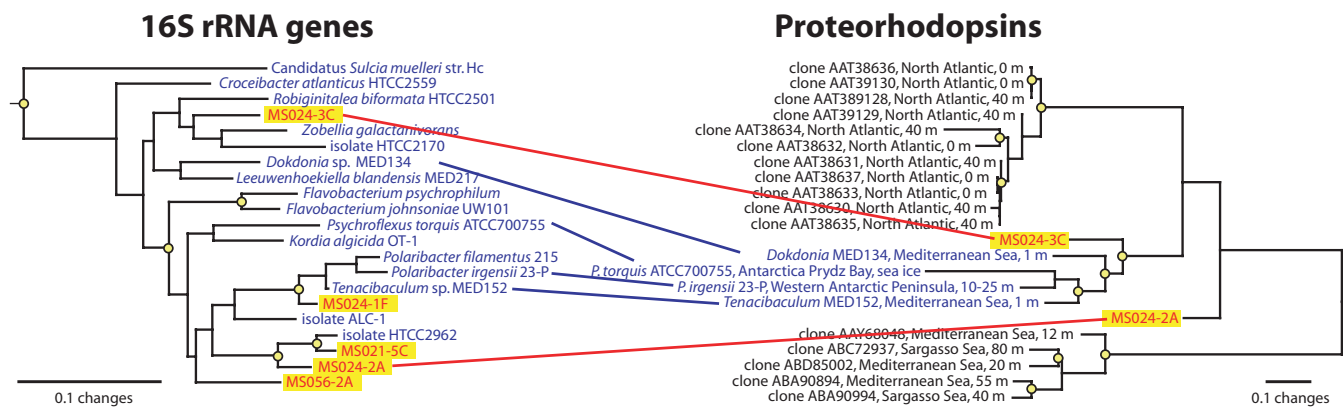
<sup>‡</sup>Determined by National Center for Biotechnology Information BLAST-N. Provided are sequence ID, GenBank accession no., and sequence identity to the SAG (%).

cloning and whole-genome sequencing of single-cell MDA products remains technically challenging because of template fragmentation, amplification biases, self-primed DNA synthesis, and the multibranching nature of MDA products (11, 25). However, downstream PCR appears insensitive to these MDA artifacts (24), is simple and inexpensive, and can be effectively used in the analysis of multiple genes.

The aim of this study was a proof-of-concept metabolic mapping of marine bacterioplankton. We developed procedures for clean single-cell separation by FACS, cell lysis, and subsequent single-cell whole-genome MDA for downstream PCR-based analyses of multiple loci. We successfully constructed a pilot library of single amplified genomes (SAGs) from a temperate coastal marine site. We analyzed them for the presence and DNA sequences of genes representing phylogenetic markers [small-subunit rRNA (SSU rRNA)] and several significant biogeochemical functions in marine ecosystems (proteorhodopsin, bacteriochlorophyll, nitrogenase, and assimilatory nitrate reductase). Bacterial proteorhodopsins (2) and bacteriochlorophylls (26) are photometabolic systems, recently recognized for their ubiquity and likely significance in the global carbon and energy fluxes. Nitrogenase is a key enzyme in the fixation of N<sub>2</sub>, effectively controlling primary production in vast areas of the ocean, and appears to be possessed by some heterotrophic bacterioplankton (27). Assimilatory nitrate reductase enables some heterotrophic bacteria to use nitrate and in this way compete with phytoplankton for the upwelled nitrogen (28). So far, little is known about the taxonomic composition of microorganisms carrying these genes in marine environments.

## Results and Discussion

**Taxonomic Composition of the SAG Library.** The SSU rRNA gene was successfully PCR-amplified and sequenced from 12 of 48 single-cell MDA reactions (Table 1). The SAG MS024-2C was identified as a contaminant and excluded from further analyses [see [supporting information \(SI\) Text](#)]. The remaining SAG library consisted of five flavobacteria, one sphingobacterium, four alphaproteobacteria of the *Roseobacter* lineage, and one gammaproteobacterium, all most closely related to marine isolates and clones. Diverse representatives of the *Roseobacter* lineage are readily isolated and are relatively well studied (29, 30). Accordingly, SSU rRNA genes of the four alphaproteobacterial SAGs were 99% identical to existing isolates. In contrast, all flavobacterial, sphingobacterial, and gammaproteobacterial SAGs were phylogenetically distant from established cultures, with 88–97% identities in the SSU rRNA gene. Flavobacteria as a group are proficient degraders of complex biopolymers, including cellulose, chitin, and pectin (31). Thus, certain Flavobacteria taxa may play important and specialized roles in microbial food webs and may be attractive for bioprospecting. Single-cell MDA provided access to the unique genomic material of these yet-uncultured taxa at the individual organism level. Different from single-cell multiplex PCR (10, 16), which enables analysis of up to two loci per cell, our approach generated a large quantity of high molecular-weight whole-genome amplification products. This material can be used in a virtually unlimited number of downstream PCRs (see below) and hybridization analyses and may be suitable for genomic sequencing (11).



**Fig. 1.** Maximum-likelihood phylogenetic trees of bacterial SSU rRNA genes and proteorhodopsins. The SSU rRNA tree includes SAGs from this study and all Flavobacteriaceae isolates with completed or currently undergoing whole-genome sequencing. The proteorhodopsin tree includes SAGs from this study and the most closely related sequences in GenBank, based on a BLASTP search. Colors indicate SAGs (yellow background), isolates (blue text), and environmental clones (black text). GenBank accession nos. are provided for each clone. Nodes marked with circles have  $>70\%$  neighbor-joining bootstrap support.

Only high nucleic acid (HNA) bacterioplankton was analyzed in this study, which comprised 57% of all heterotrophic prokaryotes in the sample. This may have biased the taxonomic composition of the SAG library, likely explaining the lack of SAR11 and some other ubiquitous groups (32). Nevertheless, the predominance of Bacteroidetes in the SAG library was unexpected. Alphaproteobacteria typically dominate SSU rRNA PCR clone libraries of marine surface bacterioplankton, whereas Bacteroidetes constitute  $<3\%$  of all marine clones (33). In contrast, studies employing fluorescent *in situ* hybridization (34, 35), quantitative PCR (36), and metagenomics (5, 8, 9) suggest a higher proportion of Bacteroidetes (particularly Flavobacteria), in some cases  $\geq 70\%$  of the total bacterioplankton (31). This contradiction may be caused by PCR and/or cloning biases against Flavobacteria (31). Interestingly, the ratio of Alphaproteobacteria vs. Bacteroidetes in our SAG library was 0.7 (Table 1), whereas the corresponding ratio of community terminal restriction fragment length polymorphism (T-RFLP) analyses peak areas at 55 bp (assumed *Roseobacter* lineage) and at 90–96 bp (assumed Bacteroidetes) was 1.0 (100 cell MDA-PCR) and 3.5 (1,000-cell seminested PCR) (SI Fig. 5). This discrepancy is supportive of a PCR bias against Bacteroidetes, which would affect T-RFLP profiles, especially those based on two rounds of seminested PCR. On the other hand, the construction of the SAG library was insensitive to PCR biases and did not involve cloning. Thus, scaled-up SAG libraries may become an ultimate tool for quantitative bacterioplankton analyses at high phylogenetic resolution. The advantage of SAG screening over fluorescent *in situ* hybridization, another taxon-specific quantification method, is demonstrated by the extraction of high-resolution phylogenetic information through sequencing of the entire SSU rRNA gene, as well as protein-encoding loci (see below).

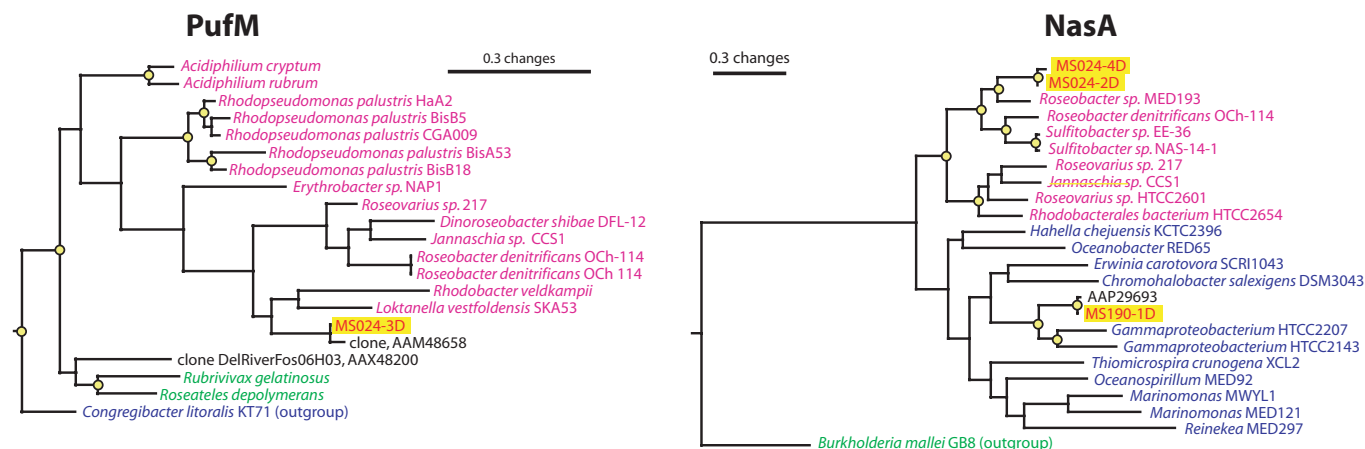
Proteorhodopsin genes were detected by PCR and confirmed by sequence analysis in 2 of 11 SAGs (Fig. 1). In addition, PCR-screening detected proteorhodopsin genes in all 12 100-cell MDA reactions. Accordingly, Sabeji *et al.* (37) estimate that 13% bacterioplankton in the photic zone of the Mediterranean Sea and the Red Sea carried proteorhodopsin genes. Our study provides further evidence that proteorhodopsin-containing microorganisms comprise a significant fraction of marine bacterioplankton.

Interestingly, both proteorhodopsin-positive SAGs were Flavobacteria, providing evidence that proteorhodopsins are common in the numerically abundant representatives of this taxonomic group (Fig. 1). The presence and photometabolic functionality of proteorhodopsins in Flavobacteria were recently

confirmed by genome sequencing of four isolates (38). The first indication of proteorhodopsins in Flavobacteria was obtained from shotgun sequencing of Sargasso Sea microbes, where a proteorhodopsin gene was found on a scaffold also containing a DNA-directed RNA polymerase sigma subunit (*rpoD*) typical of Bacteroidetes (5). Bacterial proteorhodopsins were first discovered by screening environmental BAC libraries (2). Using this technique, several Gammaproteobacteria, Alphaproteobacteria, and Euryarchaea were identified as proteorhodopsin hosts (39, 40). However, this approach has so far not indicated the presence of proteorhodopsins in Flavobacteria, possibly because the proteorhodopsin and SSU rRNA genes are too far apart. Studies based on community proteomics (41), community PCR (42, 43), community shotgun sequencing (5, 9), and PCR screening of metagenomic BAC libraries (39) demonstrated high diversity of proteorhodopsins in the ocean, although the vast majority of their hosts remain unknown. So far, only five marine isolates have been reported to contain proteorhodopsin genes, including alphaproteobacterium *Pelagibacter ubique* (44) and four Flavobacteria (38). Here we demonstrate how single-cell MDA-PCR can provide a powerful and relatively inexpensive tool for the phylogenetic mapping of this biogeochemically important gene, independent of the gene's position on the chromosome or host cultivability.

The two SAG proteorhodopsins were most closely related (up to 71% identity) to proteorhodopsins from four Flavobacteria isolates and to a group of environmental clones from the North Atlantic (Fig. 1). Consistent with the flavobacterial isolates and near-surface environmental sequences, both SAGs had methionine at amino acid position 105 (eBAC31A08 numbering), indicative of absorption maxima near 530 nm (green light) (38). In general, phylogenetic relationships among proteorhodopsins and SSU rRNA genes mirrored each other, providing no evidence for recent cross-taxa horizontal transfer events like those observed in Archaea (40). On the other hand, the presence of proteorhodopsin genes was inconsistent among some closely related Flavobacteria, e.g., *Polaribacter filamentus* 215 and *P. irgensii* 23-P (Fig. 1), suggesting recent proteorhodopsin gene losses. Interestingly, proteorhodopsin genes closely related to Flavobacterial SAGs and isolates were present among environmental clones from the North Atlantic, Mediterranean Sea, and Red Sea, indicating that Flavobacteria may be major carriers of proteorhodopsin genes in diverse marine environments.

**Other Genes.** The *pufM* and *nasA* were not detected in any of the single-cell MDA products. However, they were present in six



**Fig. 2.** Maximum-likelihood phylogenetic trees of the *PufM* and *Nasa*. Included are protein sequences obtained from 100-cell MDA reactions and most closely related sequences in GenBank, based on BLASTP searches. Colors indicate sequences from this study (yellow background), isolates of Alphaproteobacteria (purple text), isolates of Betaproteobacteria (green text), isolates of Gammaproteobacteria (blue text), and environmental clones (black text). GenBank accession nos. are provided for each clone. Nodes with yellow circles have >70% neighbor-joining bootstrap support.

(*pufM*) and three (*nasA*) 100-cell MDA reactions (of a total of 12), indicating that <1% of bacterioplankton in the sample carried either of these genes. Accordingly, bacteriochlorophyll was previously found to be expressed (infrared fluorescence) in  $\approx$ 1% of bacteria in coastal Maine waters at this time of year (45). All six *pufM* were 100% identical to each other and were most closely related to bacteriochlorophylls from the *Roseobacter* lineage (Fig. 2A). Thus, it appears that a single *Roseobacter* taxon dominated bacteriochlorophyll-containing bacterioplankton in the studied sample. Two *nasA* were most closely related to assimilatory nitrate reductases in *Roseobacter* lineage, whereas one *nasA* was most closely related to marine Gammaproteobacteria (Fig. 2B). The pilot SAG library failed to unambiguously identify these relatively rare but biogeochemically important microorganisms. Screening of a larger SAG library would be an ideal tool for this task. Alternative community genomics-based analyses have proven less effective to match SSU rRNA, and functional genes in such rare taxa.

Genes encoding archaeal SSU rRNA and nitrogenases were not detected in any of the sorted wells, suggesting that Archaea and nitrogen-fixing organisms were extremely rare or absent in the analyzed heterotrophic HNA bacterioplankton. Eukaryote SSU rRNA genes were also not detected, confirming effective separation of prokaryotes from protists by FACS.

## Conclusions

We demonstrate how a combination of single-cell FACS, MDA, and PCR can be used in metabolic mapping of taxonomically diverse uncultured marine bacterioplankton. Large quantities of high molecular-weight whole-genome amplification products were obtained from individual cells, allowing for a virtually unlimited number of downstream analyses. In this proof-of-concept study, we detected proteorhodopsin genes in two of five flavobacteria, providing evidence that Flavobacteria are major carriers of this photometabolic gene. We also determined that Flavobacteria were a major component of HNA bacterioplankton in the analyzed coastal sample. Fewer than 1% of the analyzed cells carried *nasA*, *pufM*, and *nifH*.

We used standard configuration flow cytometry instrumentation that is available on most major research campuses and increasingly used aboard oceanographic research vessels. Working at the single-cell level requires especially stringent instrument cleaning, sample handling, and quality-control methods to prevent DNA contamination. We show that our methods were

able to achieve sufficiently low DNA blank controls (see *SI Text*). The cost of MDA and subsequent PCR sequencing is on the order of tens of U.S. dollars per cell and thus is significantly less expensive than metagenomic sequencing. In addition to high-throughput screening by PCR or hybridization, SAG libraries may provide material for genomic sequencing of selected uncultured microorganisms. Two of our SAGs are currently in the process of whole-genome sequencing.

## Materials and Methods

**Sample Collection and Single-Cell Sorting.** The coastal water sample was collected from Boothbay Harbor, Maine, from 1-m depth at the Bigelow Laboratory dock (43°50'40''N, 69°38'27''W) on March 28 at 9:45 a.m. during high tide (water temperature 7.0°C). The unmanipulated sample was 10-fold diluted with filtered (0.2  $\mu$ m pore size) sample water and stained with 5  $\mu$ M (final concentration) SYTO-13 nucleic acid stain (Invitrogen, Carlsbad, CA) for prokaryote detection as in delGiorgio *et al.* (46). Individual bacterioplankton were sorted into 96-well plates containing 5  $\mu$ l per well of PBS. Only HNA cells were sorted to reduce the probability of depositing dead cells with partially degraded genomes. Single cells were sorted into four of the eight rows on each plate. Of the remaining rows, two were dedicated to background controls, consisting of single drops generated from a sort gate drawn in the “noise” area in the lower left corner of the side scatter/green fluorescence plot. One row of 12 wells was dedicated to blanks with no drop deposition, and one row received 100 HNA bacterioplankton cells per well. Sorting was done with a MoFlo (Dako Cytomation, Carpinteria, CA) flow cytometer equipped with the CyClone robotic arm for sorting into plates, using a 488-nm argon laser and a 70- $\mu$ m nozzle orifice. The cytometer was triggered on side scatter, the sort gate was based on side scatter and SYTO-13 fluorescence, and the “purify 0.5 drop” sort mode was used for maximal sort purity. Extreme care was taken to prevent sample contamination by any nontarget DNA. New sheath fluid lines were installed before each sort day. Sheath fluid and sample lines were cleaned by a succession of warm water, 5% bleach solution, and an overnight flush with DNA-free deionized water. Sheath fluid was prepared by dissolving combusted (2 h at 450°C) NaCl in DNA-free deionized water for a final concentration of 1%. Sorted plates were stored at  $-80^{\circ}\text{C}$  until MDA.

**Lysis and MDA.** We compared three protocols for cell lysis, DNA denaturing, and MDA in this study.

**Protocol A.** Three cycles of heating to 97°C and cooling to 8°C were used for cell lysis and DNA denaturing, after which 18-h MDA was performed by using REPLI-g Mini (Qiagen, Chatsworth, CA) Phi29 polymerase and reaction buffer. For each well containing 5  $\mu$ l of PBS, we used 0.5  $\mu$ l of polymerase, 14.5  $\mu$ l of buffer, and 5  $\mu$ l of DNA-free deionized water.

**Protocol B.** Alkaline lysis on ice and 18-h MDA were performed by using REPLI-g Mini (Qiagen) kit reagents and following the manufacturer's protocol for blood samples.

**Protocol C.** As protocol B, except that REPLI-g Midi kit (Qiagen) was used, and PicoGreen DNA stain (Invitrogen) was added to the reaction at 0.5 $\times$  (final concentration). DNA synthesis was monitored with IQ5 real-time PCR system (Bio-Rad, Hercules, CA). Duplicate standards containing 0.05, 5, 500, and 50,000 fg of human genomic DNA (Promega, Madison, WI) were amplified simultaneously with the sort samples.

Initially, each of the three protocols were applied on 24 wells: 12 with single cells, 3 no-drop controls, 6 background controls, and 3 with 100 cells. Protocols A, B, and C were used after 7, 8, and 94 days of sorted cell storage at  $-80^{\circ}\text{C}$ , respectively. An additional 24 wells were analyzed by using protocol B after 350 days of storage. The DNA concentration in MDA reactions was determined by using a ND-1000 spectrophotometer (Nanodrop) after a cleanup with MinElute PCR Purification Kit (Qiagen).

**PCR-Based Analyses of MDA Products.** The MDA products were diluted 10-fold (protocols A and B; REPLI-g Mini kit products) or 200-fold (protocol C; REPLI-g Midi kit products). Two microliters of the dilute products served as templates in 25  $\mu$ l of PCR. Previously described primers and PCR conditions were used to amplify genes encoding bacterial, archaeal, and eukaryal SSU rRNA, proteorhodopsin, bacteriochlorophyll, nitrogenase, and assimilative nitrate reductase (SI Table 2). The PCR products were cleaned with QIAquick (Qiagen). For the T-RFLP of bacterial SSU rRNA genes, PCR amplicons obtained with 27F-FAM and 907R primers were digested with either HhaI or

BsuRI (HaeIII) restriction endonucleases (Fermentas, Hanover, MD). Sequencing and fragment analyses were performed with a 3730xl analyzer (Applied Biosystems, Foster City, CA) at the W. M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana—Champaign, IL. For T-RFLP, GeneScan 1000 ROX Size Standard (Applied Biosystems) was used.

Taxonomic identification of SAGs was achieved by SSU rDNA gene analysis with GenBank BLASTN (47) and Ribosomal Database Project (RDP) Classifier and Seqmatch search tools (48). The SSU rRNA sequences were checked for chimeras by using the RDP Chimera Check tool. Protein-encoding sequences were translated with National Center for Biotechnology Information ORF Finder and their identities verified by GenBank BLASTP searches for closest relatives. Evolutionary trees were constructed by using PHYLIP (49) after an automatic sequence alignment with ClustalX (50).

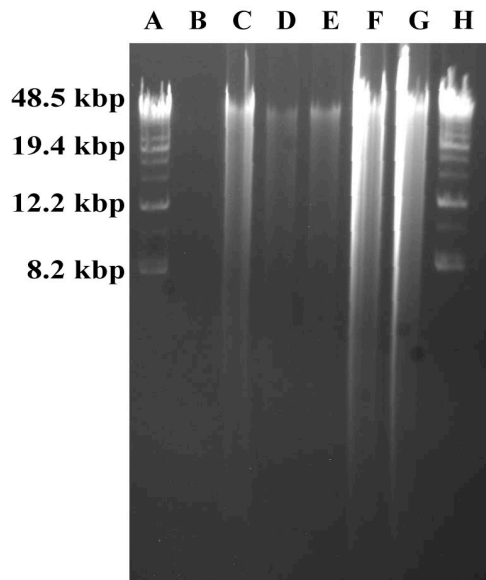
**T-RFLP Profiling of Bacterioplankton Communities.** Triplicate 1,000 cell aliquots of HNA bacterioplankton were sorted as above into microcentrifuge tubes preloaded with 5  $\mu$ l of Lyse-N-Go (Pierce, Rockford, IL) and then stored at  $-80^{\circ}\text{C}$ . Cell lysis was performed according to Lyse-N-Go instructions. Entire lysate volumes were used as templates in 50- $\mu$ l 30-cycle PCRs by using primers 27F and 1492R (SI Table 2). Two-microliter aliquots of these PCR products served as templates in a second semi nested 25- $\mu$ l and 30-cycle PCR with primers 27F-FAM and 907R. PCR products were cleaned and digested, and fragment analyses were performed as above.

We thank Nicole Poulton and Wendy Bellows (Bigelow Laboratory) for technical help and Janet Hart (Duke University, Durham, NC) for early work on single-cell PCR optimization. Arumugham Raghunathan (Qiagen) is acknowledged for expert advice on whole-genome amplification by using REPLI-g kits. This study was funded by the State of Maine Technology Institute, Marine Research Initiative, and National Science Foundation SGER award no. EF-0633142.

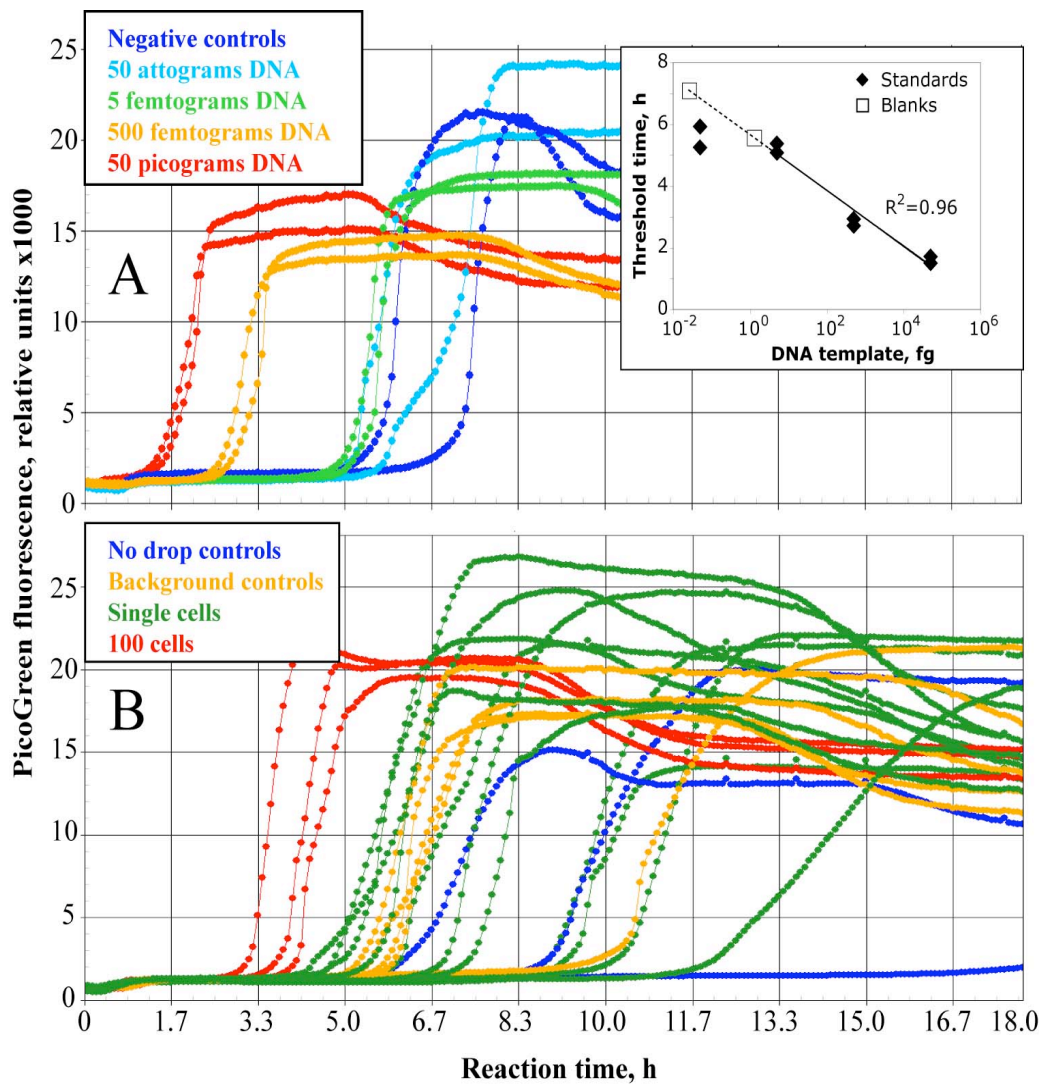
- Pace NR, Stahl DA, Lane DJ, Olsen GJ (1986) *Adv Microb Ecol* 9:1–55.
- Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich S, Gates CM, Feldman RA, Spudich JL, et al. (2000) *Science* 289:1902–1906.
- Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, et al. (2000) *Appl Environ Microbiol* 66:2541–2547.
- Schloss PD, Handelsman J (2003) *Curr Opin Biotechnol* 14:303–310.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu DY, Paulsen I, Nelson KE, Nelson W, et al. (2004) *Science* 304:66–74.
- Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, et al. (2005) *Science* 308:554–557.
- DeLong EF, Karl DM (2005) *Nature* 437:336–342.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N-U, Martinez A, Sullivan MB, Edwards R, Brito BR, et al. (2006) *Science* 311:496–503.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, et al. (2007) *PLoS Biol* 5:e77.
- Ottesen EA, Hong JW, Quake SR, Leadbetter JR (2006) *Science* 314:1464–1467.
- Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, Church GM (2006) *Nat Biotechnol* 24:680–686.
- Frohlich J, Konig H (1999) *Syst Appl Microbiol* 22:249–257.
- Sieracki M, Poulton N, Crosbie N (2005) in *Algal Culturing Techniques*, ed Andersen R (Elsevier Academic, New York), pp 101–116.
- Karl DM, Bailiff MD (1989) *Limnol Oceanogr* 34:543–558.
- Brum JR (2005) *Aq Microb Ecol* 41:103–113.
- Li HH, Gyllensten UB, Cui XF, Saiki RK, Erlich HA, Arnheim N (1988) *Nature* 335:414–417.
- Pinar D, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH (2006) *BMC Genomics* 7:216.
- Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K (2006) *Hum Mutat* 27:496–503.
- Dean FB, Hosono S, Fang LH, Wu XH, Faruqi AF, Bray-Ward P, Sun ZY, Zong QL, Du YF, Du J, et al. (2002) *Proc Natl Acad Sci USA* 99:5261–5266.
- Hutchison CA, Smith HO, Pfannkoch C, Venter JC (2005) *Proc Natl Acad Sci USA* 102:17332–17336.
- Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Chang HW, Watson D, Brodie EL, Hazen TC, et al. (2006) *Appl Environ Microbiol* 72:3291–3301.
- Hellani A, Coskun S, Benkhalifa M, Tbakhi A, Sakati N, Al-Odaib A, Ozand P (2004) *Mol Hum Reprod* 10:847–852.
- Jiang ZW, Zhang XQ, Deka R, Jin L (2005) *Nucleic Acids Res* 33:e91.
- Raghunathan A, Ferguson HR, Bornarth CJ, Song WM, Driscoll M, Lasken RS (2005) *Appl Environ Microbiol* 71:3342–3347.
- Hutchison CA, Venter JC (2006) *Nat Biotechnol* 24:657–658.
- Kolber ZS, Plumley FG, Lang AS, Beatty JT, Blankenship RE, VanDover CL, Vetriani C, Koblizek M, Rathgeber C, Falkowski PG (2001) *Science* 292:2492–2495.
- Zani S, Mellon MT, Collier JL, Zehr JP (2000) *Appl Environ Microbiol* 66:3119–3124.
- Allen AE, Booth MG, Frischer ME, Verity PG, Zehr JP, Zani S (2001) *Appl Environ Microbiol* 67:5343–5348.
- Buchan A, Gonzalez JM, Moran MA (2005) *Appl Environ Microbiol* 71:5665–5677.
- Moran MA, Buchan A, Gonzalez JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, et al. (2004) *Nature* 432:910–913.
- Kirchman DL (2002) *FEMS Microbiol Ecol* 39:91–100.
- Mary I, Heywood JL, Fuchs BM, Amann R, Tarran GA, Burkhill PH, Zubkov MV (2006) *Aq Microb Ecol* 45:107–113.
- Giovannoni SJ, Rappe MS (2000) in *Microbial Ecology in the Oceans*, ed Kirchman DL (Wiley-Liss, New York), pp 47–84.
- Mary I, Cummings DG, Biegala IC, Burkhill PH, Archer SD, Zubkov MV (2006) *Aq Microb Ecol* 42:119–126.
- Jaspers E, Nauhaus K, Cypionka H, Overmann J (2001) *FEMS Microbiol Ecol* 36:153–164.
- Abell GCJ, Bowman JP (2005) *FEMS Microbiol Ecol* 51:265–277.
- Sabehi G, Loy A, Jung KH, Partha R, Spudich JL, Isaacson T, Hirschberg J, Wagner M, Beja O (2005) *PLoS Biology* 3:1409–1417.
- Gomez-Consarnau L, Gonzalez JM, Coll-Llado M, Gourdon P, Pacher T, Neutze R, Pedros-Alio C, Pinhassi J (2007) *Nature* 445:210–213.

39. Sabehi G, Beja O, Suzuki MT, Preston CM, DeLong EF (2004) *Environ Microbiol* 6:903–910.
40. Frigaard NU, Martinez A, Mincer TJ, DeLong EF (2006) *Nature* 439:847–850.
41. Beja O, Spudich EN, Spudich JL, Leclerc M, DeLong EF (2001) *Nature* 411:786–789.
42. Man DL, Wang WW, Sabehi G, Aravind L, Post AF, Massana R, Spudich EN, Spudich JL, Beja O (2003) *EMBO J* 22:1725–1731.
43. Sabehi G, Massana R, Bielawski JP, Rosenberg M, DeLong EF, Beja O (2003) *Environ Microbiol* 5:842–849.
44. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, *et al.* (2005) *Science* 309:1242–1245.
45. Sieracki ME, Gilg IC, Thier EC, Poulton NJ, Goericke R (2006) *Limnol Oceanogr* 51:38–46.
46. delGiorgio P, Bird DF, Prairie YT, Planas D (1996) *Limnol Oceanogr* 41:783–789.
47. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) *Nucleic Acids Res* 25:3389–3402.
48. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM (2003) *Nucleic Acids Res* 31:442–443.
49. Felsenstein J (1989) *Cladistics* 5:164–166.
50. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) *Nucleic Acids Res* 31:3497–3500.

## Supporting Information

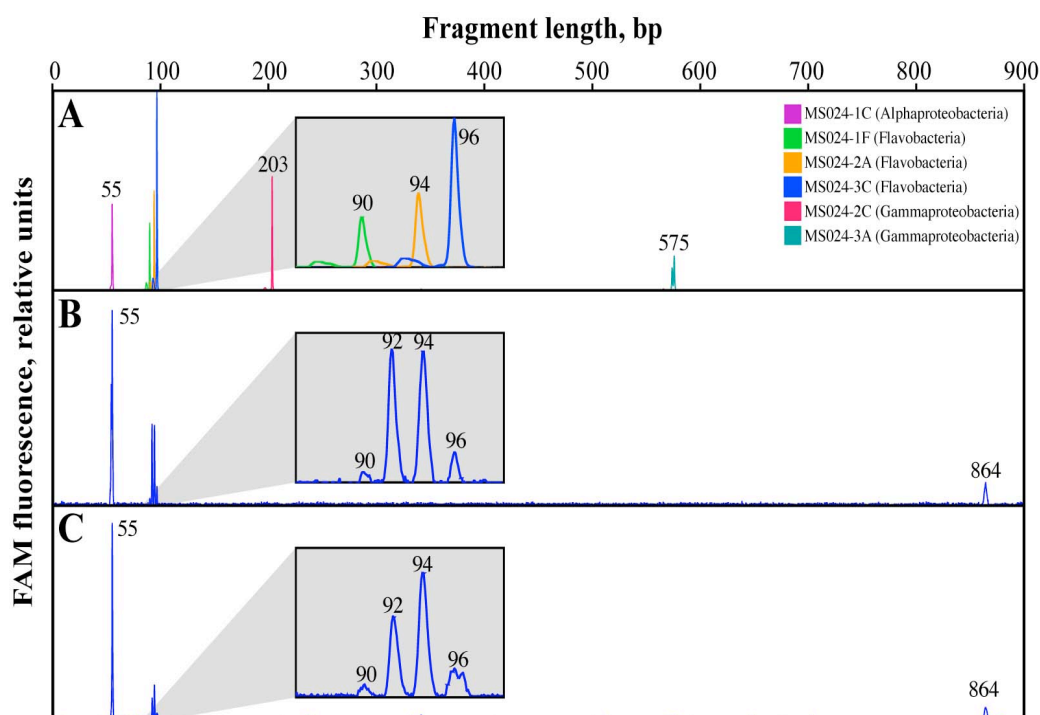


**Fig. 3.** Gel electrophoresis of multiple displacement amplification (MDA) products obtained using Protocol C: molecular weight ladder (lanes A and H), no-drop controls (lanes B-C), single bacterioplankton cells (lanes D-E), and 100 bacterioplankton cells (lanes F-G). The 0.75% TAE agarose gel was loaded with 3 ml of high molecular weight ladder (Invitrogen) and 10x diluted products of MDA reactions and then electroporated at 0.5 V/cm for 5 h.



**Fig. 4.** Real-time monitoring of multiple displacement amplification reactions of (A) standards prepared from human genomic DNA and (B) flow-cytometrically sorted material. *Inset* in A shows the log-linear standard curve for DNA concentration.





**Fig. 5.** T-RFLP profiles of bacterial SSU rRNA genes obtained by (A) single-cell MDA-PCR, (B) 100-cell MDA-PCR, and (C) 1,000-cell seminested PCR. Panel A is a composite of six single-cell profiles. Insets show discrimination of the peaks near 90 bp with an expanded scale. All profiles were obtained using 27F-FAM and 907R primers and HhaI restriction endonuclease.

**Table 2. PCR primers**

<u>Gene</u>	<u>Primers</u>	<u>Product, bp</u>	<u>References</u>
Bacterial SSU rRNA	27F, 519F, 907R, 1492R	Various	1, 2
Archaeal SSU rRNA	S-D-Arch-0344-a-S-20, 907R	550	3, 4
Eukaryote SSU rRNA	EUK328f, EUK329r	1500	5, 6
Proteorhodopsin	o-PR2, o-PR3	330	7, 8
Bacteriochlorophyll, <i>pufM</i>	pufM_228F, pufM_228R	228	9
Nitrogenase, <i>nifH</i>	nifUP, nifDN, NifH3, NifH4	450	10, 11
Assimilatory nitrate reductase, <i>asa</i>	nas22, nas1933, nas964, nasA1735	771	12, 13

1. Lane DJ (1991) in *Nucleic Acid Techniques in Bacterial Systematics*, eds Stackebrandt E, Goodfellow M (Wiley, Chichester, U.K.).

2. Page KA, Connon SA, Giovannoni SJ (2004) *Appl Environ Microbiol* 70:6542-6550.

3. Vetriani C, Jannasch HW, MacGregor BJ, Stahl DA, Reysenbach AL (1999) *Appl Environ Microbiol* 65:4375-4384.
4. Wilms R, Sass H, Kopke B, Koster H, Cypionka H, Engelen B (2006) *Appl Environ Microbiol* 72:2756-2764.
5. Zhu F, Massana R, Not F, Marie D, Vaultot D (2005) *FEMS Microbiol Ecol* 52:79-92.
6. Marie D, Zhu F, Balague V, Ras J, Vaultot D (2006) *FEMS Microbiol Ecol* 55:403-415.
7. Frigaard NU, Martinez A, Mincer TJ, DeLong EF (2006) *Nature* 439:847-850.
8. Sabehi G, Loy A, Jung KH, Partha R, Spudich JL, Isaacson T, Hirschberg J, Wagner M, Beja O (2005) *PloS Biology* 3:1409-1417.
9. Schwalbach MS, Fuhrman JA (2005) *Limnol Oceanogr* 50:620-628.
10. Zani S, Mellon MT, Collier JL, Zehr JP (2000) *Appl Environ Microbiol* 66:3119-3124.
11. Church MJ, Short CM, Jenkins BD, Karl DM, Zehr JP (2005) *Appl Environ Microbiol* 71:5362-5370.
12. Allen AE, Booth MG, Verity PG, Frischer ME (2005) *Aq Microb Ecol* 39:247-255.
13. Allen AE, Booth MG, Frischer ME, Verity PG, Zehr JP, Zani S (2001) *Appl Environ Microbiol* 67:5343-5348.

## SI Text

### Quality control of single-cell sorting and whole-genome amplification

**Multiple displacement amplification.** Approximately 0.5 and 5 mg of genomic DNA was synthesized in single cell REPLI-g Mini (protocols A and B) and REPLI-g Midi (protocol C) reactions, respectively, with the apparent dominant product size >10 kbp (SI Fig. 3). Most MDA reactions, including all but two blank treatments, resulted in DNA synthesis. The synthesis of DNA in MDA negative controls has been observed before and is likely caused by Phi29 self-priming, DNA contamination, or both (1-3). Real-time monitoring of MDA demonstrated that reaction speed depended upon template amount when template was above 5 fg DNA (SI Fig. 4A). The template-dependent dynamics of MDA observed in this study suggested that 6-8 h were necessary to complete MDA of fg-level templates using REPLI-g Midi kit (QIAGEN). This agrees with a similar analysis by Zhang *et al.* (2) but contradicts Spits *et al.* (4), who suggested 2-h reactions for MDA of single-cell genomes.

A log-linear standard curve was produced from the 5-fg, 500-fg, and 50-pg DNA treatments (SI Fig. 4A *Inset*), and used to calculate tentative DNA concentrations in the sorted samples. The average amount of DNA in the no-drop controls, background controls, single-cell, and 100-cell treatments were estimated to be 0.07 (range 0.0-0.2), 0.78 (0.0-1.8), 2.0 (0.0-10.3) fg (SI Fig. 4B). The average for single cells was 4.8 fg, if wells with virtually no amplification (<0.3-fg DNA content estimate) were excluded from the calculation. The estimated amount of DNA in the single- and 100-cell treatments was in agreement with published marine bacterioplankton genome size estimates based on fluorometry (2.5 fg) (5), flow cytometry (1.5 fg) (6), and whole-genome sequencing data (7, 8). The low-DNA content in no-drop controls (mean, 0.07 fg) suggests that sample contamination from handling and reagents was below 4% of an average single-cell genome. The background "noise" controls were sorted drops from the low side scatter and fluorescence region. They showed significantly higher DNA (mean, 0.78 fg)

than the no-drop controls. This DNA could have originated from large viruses, DNA debris, and small (low nucleic acid) bacteria. No DNA was detected in sheath fluid collected at the end of instrument lines. Thus, our sample handling stringency was adequate to prevent significant sample contamination with extraneous genomic DNA.

**PCR of the bacterial SSU rRNA gene.** Initially, 12 wells with single cells were subjected to each of the lysis-MDA protocols A, B, and C. Using MDA products as templates in PCR, SSU rRNA genes were successfully amplified from one, six, and two of the wells, respectively (Table 1). Thus, protocol B (cold alkaline lysis and MDA using REPLI-g Mini kit) resulted in the highest success rate (50%) of MDA-PCR. According to manufacturer's statements, we expected similar cell lysis and MDA success rate with REPLI-g Mini and Midi kits, with the Midi kit producing higher DNA yields per sample. The lower success rate of protocol C may be in part explained by possible DNA degradation during prolonged sample storage (94 vs. 8 days). After a 350-day storage, the application of protocol B on 12 additional single-cell wells resulted in a 25% success rate. Diverse factors may have contributed to the less than 100% success rate in single-cell MDA-PCR, including failed deposition of some cells during FACS, DNA degradation during post-FACS storage, incomplete cell lysis, incomplete MDA, and mismatches to some bacterioplankton groups in the "universal" PCR primers used in this study. It is noteworthy that the 25-50% success rate in MDA-PCR achieved with protocol B was similar to the single cell PCR success rate achieved by Ottesen *et al.* (9) and to FISH success rate using "universal" SSU rRNA probes (10).

The SSU rRNA PCR products were obtained from all 12 100-cell wells and from one background control (of 24 total). No PCR products were obtained from the 12 no-drop controls. High-quality nonchimerical sequences of near-complete SSU rRNA genes were obtained from all single-cell MDA-PCR products. The SSU rRNA genes of 11 of the 12 SAGs were most closely related to marine isolates and clones (Table 1). One exception, MS024-2C, was identified as *Delftia acidovorans*, characteristic for soils, freshwaters, and anthropogenic environments. An identical sequence was also retrieved from one of the background controls. This suggests that the sequence is likely a contaminant originating from handling or the reagents. Thus, the overall rate of apparent contamination with bacterial SSU rRNA gene was two of 84 wells (2%) of single cells, background controls, and no-drop controls combined.

For additional quality control, T-RFLP profiles were generated for the SSU rRNA gene of each PCR-positive single cell and control MDA reaction, using two alternative restriction endonucleases. All profiles contained single peaks (Table 1, SI Fig. 5A), further confirming that only one cell type was deposited and amplified in each well with no apparent contamination with DNA from other bacterial taxa. It is possible that in some cases aggregates of genetically identical cells, which were optically similar to single cells (e.g., partially divided cells), were sorted into the same well. However, this would not have any adverse effect on the downstream molecular analyses or interpretation.

The HhaI-based T-RFLP profiles generated from 100-cell MDA reactions were dominated by a peak at 55 bp and a group of peaks at around 90 bp (SI Fig. 5B). Similar T-RFLP profiles were also obtained from 1,000 cell aliquots lysed with Lyse-N-Go protocol and amplified using seminested PCR (SI Fig. 5C). This suggests that the diverse cell lysis and DNA amplification methods used in this study were targeting the same bacterioplankton taxa. The T-RFs found in 100- and 1,000-cell profiles corresponded to those found in SAG profiles of Alphaproteobacteria (55 bp) and Bacteroidetes (around 90 bp; SI Fig. 5A). The 203-bp *Delftia* fragment was absent in community profiles, further suggesting its origin as a sample handling contaminant. Conspicuously, the 575-bp T-RF, characteristic to the marine gammaproteobacterium MS024-3A, was also absent in community profiles, either due to PCR biases or rarity of the taxon. On the other hand, 92- and 864-bp peaks, present in all 100- and 1,000-cell community profiles, were not represented in the SAG library. Considering the relatively small size of the pilot SAG library and the fact that all but two (MS024-1C and MS190-2A) SAG SSU rRNA gene sequences were unique, it is clear that this library covered only a fraction of biodiversity in the bacterioplankton community.

1. Raghunathan A, Ferguson HR, Bornarth CJ, Song WM, Driscoll M, Lasken RS (2005) *Appl Environ Microbiol* 71:3342-3347.
2. Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, Church GM (2006) *Nat Biotechnol* 24:680-686.
3. Hutchison CA, Venter JC (2006) *Nat Biotechnol* 24:657-658.
4. Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, Sermon K (2006) *Hum Mutat* 27:496-503.
5. Fuhrman JA, Azam F (1982) *Mar Biol* 66:109-120.
6. Button DK, Robertson BR (2001) *Appl Environ Microbiol* 67:1636-1645.
7. Moran MA, Buchan A, Gonzalez JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, *et al.* (2004) *Nature* 432:910-913.
8. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, *et al.* (2005) *Science* 309:1242-1245.
9. Ottesen EA, Hong JW, Quake SR, Leadbetter JR (2006) *Science* 314:1464-1467.
10. Kirchman DL (2002) *FEMS Microbiol Ecol* 39:91-100.