Miniprimer PCR, a New Lens for Viewing the Microbial World^{∇}[†]

Thomas A. Isenbarger,^{1,2} Michael Finney,²‡ Carlos Ríos-Velázquez,³ Jo Handelsman,¹* and Gary Ruvkun²

Departments of Bacteriology and Plant Pathology, University of Wisconsin—Madison, 1550 Linden Drive, Madison, Wisconsin 53706-1521¹; Department of Genetics, Harvard University, and Department of Molecular Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114-2790²; and Department of Biology, University of Puerto Rico—Mayagüez, Mayagüez, Puerto Rico³

Received 22 August 2007/Accepted 29 November 2007

Molecular methods based on the 16S rRNA gene sequence are used widely in microbial ecology to reveal the diversity of microbial populations in environmental samples. Here we show that a new PCR method using an engineered polymerase and 10-nucleotide "miniprimers" expands the scope of detectable sequences beyond those detected by standard methods using longer primers and *Taq* polymerase. After testing the method in silico to identify divergent ribosomal genes in previously cloned environmental sequences, we applied the method to soil and microbial mat samples, which revealed novel 16S rRNA gene sequences that would not have been detected with standard primers. Deeply divergent sequences were discovered with high frequency and included representatives that define two new division-level taxa, designated CR1 and CR2, suggesting that miniprimer PCR may reveal new dimensions of microbial diversity.

Characterization of 16S rRNA gene sequences has become a central feature of microbial ecology. Frequently, these analyses are initiated by using PCR to amplify 16S rRNA genes directly from environmental samples without culturing (4). Indeed, these types of studies have transformed our view of the microbial world. However, as the 16S rRNA sequence database has grown, it has become evident that many sequences deviate within the most conserved regions targeted by "universal" 16S rRNA gene PCR primers (8, 9, 38). To accommodate these deviations, commonly used primers have been modified with degenerate positions to enable the primers to target a wider range of 16S rRNA gene sequences. However, because polymerases used for PCR require primers of ~20 to 30 nucleotides (nt), 16S rRNA gene primer design has been constrained to target conserved regions of those lengths. New thermostable polymerases have recently become available (40), opening the possibility of changes in primer design. Preliminary data suggested that some of these polymerases may be able to utilize primers shorter than the minimum length of ~ 20 to 30 nt typically recognized by the standard enzyme used for PCR, Taq (Thermus aquaticus) DNA polymerase.

We posited that decreasing the length of 16S rRNA gene PCR primers would broaden the scope of sequences detectable by PCR. In this study we employ a new polymerase, S-Tbr, and short primers to demonstrate the utility of a "miniprimer" PCR method, both computationally and with environmental samples.

MATERIALS AND METHODS

PCR conditions. The S-Tbr (DyNAmo II; Finnzymes Oy, Espoo, Finland) and Taq (Roche, Indianapolis, IN) DNA polymerases were used for PCR. S-Tbr is an engineered polymerase where the N-terminal 5'-3' exonuclease domain of Thermus brockianus DNA polymerase I has been removed and replaced by the 7-kDa double-stranded DNA-binding protein Sso7d from Sulfolobus solfataricus (40). Oligonucleotide primers were synthesized by Operon (Huntsville, AL) or Integrated DNA Technologies (Coralville, IA). PCR was performed in 20 or 30 µl with sterile, nucleic acid-free water (MO BIO Laboratories, Carlsbad, CA), $1 \times$ Taq polymerase reaction buffer (10 mM Tris-HCl, 1.5 mM MgCl₂, 5 mM KCl, pH 8.3 at 20°C), 200 µM each deoxynucleoside triphosphate, 0.75 U of polymerase, and primers at a concentration of 10 pmol/µl. PCR components were mixed on ice before placing them in a thermocycler preheated to 80°C. PCR samples were incubated at 95°C for 120 s; thermally cycled 30 times at 94°C for 5 s, 40 to 50°C for 40 s, and 72°C for 60 s; and finally incubated at 72°C for 600 s. Unless otherwise noted in the text and figures, an annealing temperature of 40°C was used. MJ Research (Waltham, MA) PTC-100, PTC-200, and Tetrad PTC-225 thermocyclers were used for thermal cycling.

Environmental sample preparation. For PCR from soil, three ~1-g samples were collected from the courtyard outside the Wellman building on the Massachusetts General Hospital campus (Boston, MA) and boiled in 1 ml sterile, nucleic acid-free water for 5 min. After centrifugation to pellet solid material, the supernatants were combined, diluted 1:10 in sterile nucleic acid-free water, and used for PCR. A boiling method was used to prepare DNA quickly to test miniprimers and was not intended to provide material for a comprehensive analysis of a soil sample. The microbial mat samples were collected from a mature mat community of the Candelaria lagoon, which is one of a series of evaporative salterns in Cabo Rojo, southwest Puerto Rico (17°56'N, 67°11'W), and immediately frozen for storage. The mat is a laminated structure of ~ 2.0 to 2.5 cm in thickness and has distinctive green, pink, and black layers. For a detailed physical and geochemical description of similar mats from Cabo Rojo, see reference 11. To prepare microbial mat DNA, mat samples were thawed on ice and three \sim 1-cm-wide vertical sections were cut so that all mat layers were present in the pieces removed. The sections were mixed to form a sample of ~ 20 g and then mixed to homogeneity with 20 ml of sterile, nucleic acid-free water. After centrifugation to collect the solid material, the liquid was removed and three \sim 1-g samples were removed from the top, middle, and bottom of the pellet. The nine samples were processed individually using a bead-beating method (MO BIO PowerSoil kit; MO BIO Laboratories, Carlsbad, CA); the protocol included the optional 70°C initial incubation in solution C1. Equal volumes of the nine DNA samples were combined, and this mixture was used as the template in the PCRs. For each PCR method tested, three separate PCRs were performed with the microbial mat template, and the products of the three

^{*} Corresponding author. Mailing address: Department of Bacteriology, University of Wisconsin—Madison, 1550 Linden Drive, Madison, WI 53706-1521. Phone: (608) 263-8783. Fax: (608) 265-5289. E-mail: joh@bact.wisc.edu.

[‡] Present address: Finney Capital, 489 Douglass St., San Francisco, CA 94114.

[†] Supplemental material for this article may be found at http://aem .asm.org/.

^v Published ahead of print on 14 December 2007.

Forward primer	Sequence $(5'-3')$	Reverse primer	Sequence (5'-3')	Reference or source	
27F-P	AGAGTTTGATCCTGGCTCAG	1492R-P	GGTTACCTTGTTACGACTT	e.g., 18	
27F-L	AGAGTTTGATCMTGGCTCAG	1492R-L	TACGGYTACCTTGTTACGACTT	25	
27F-HT	AGRGTTTGATYMTGGCTCAG	1492R-HT	GGYTACCTTGTTACGACTT	37	
27F-10	AGAGTTTGAT	1505R-10	CCTTGTTACG	This work	

TABLE 1. Common variants of 16S rRNA primers^a

 a R = A or G; Y = C or T; M = A or C.

PCRs were combined for cloning and subsequent sequencing. All sample handling was performed in a clean, sterile laminar flow hood.

Cloning and sequencing. PCR products were extracted from ~1.0% agarose gels and cloned using the TOPO cloning kit for sequencing (Invitrogen, Carlsbad, CA). Escherichia coli genomic DNA was prepared using the QIAamp DNA mini kit (Qiagen, Valencia, CA) or purchased from USB Corporation (Cleveland, OH); to prepare Halobacterium salinarum genomic DNA, H. salinarum strain MPK1 (24) was grown in culture as previously described (12), followed by genomic DNA preparation using the QIAamp DNA mini kit. Automated DNA sequencing of clones was performed by the Massachusetts General Hospital DNA Sequencing Core Facility (Cambridge, MA) or by the University of Wisconsin Biotechnology Center (Madison, WI) using sequencing primers located outside the bounds of the cloned 16S rRNA gene inserts. Sequence reads were assembled using Phred and Phrap as implemented in XplorSeq (D. Frank, unpublished data) or the Staden package, release 1.6.0. Soil study clones A01 and A04 were fully sequenced with two- to fivefold coverage by independent sequence reads and had at least threefold coverage of all highly conserved regions. Mallard was used as the primary tool to identify and remove anomalous or chimeric 16S rRNA gene sequences (7); for some ambiguous cases, Bellerophon (20), Pintail (6), BLAST, and manual checking were used to supplement results from Mallard analysis.

Informatics and computation. Sequences were aligned using the NAST aligner (16) and imported into Arb version 05.10.26 (27), using the full greengenes database (see below) as a 16S rRNA reference database (15). The Arb database of Cabo Rojo and reference sequences is available at plantpath.wisc.edu/~isen/ CR. Distance matrices calculated in Arb using the Olsen correction and Lane mask (25) were used to cluster sequences into operational taxonomic units (OTUs) by pairwise identity and a furthest-neighbor algorithm (33) and for calculating similarity indices to compare libraries (34). Microbial mat sequences were inserted into the Arb dendrogram using the parsimony tool and Lane mask. Primary taxonomic assignments were informed by the resulting Arb dendrogram and the Lawrence Berkeley National Laboratory (LBNL) greengenes classifier. Novel taxa identified with colloquial names and for which references are not otherwise indicated in the text include BD5-13 (23), BRC1 (13), FW129 (10), NB23 (28), OP11 (21), WS2, and WS6 (17); publications for AC1 and Eub6 were not identified by our search. Similarities to database records were calculated by a maximum likelihood method as implemented in PHYLIP, using F84 correction, a transition:transversion ratio of 2.0, empirically determined base frequencies, and the Lane mask (25). Sequences that were not placed into previously described taxa or that formed monophyletic groups of greater than 0.15 distance unit from the closest previously isolated sequence in the full LBNL greengenes database were subjected to additional characterization as follows. First, novel sequences were aligned with reference sequences representing major defined lines of bacterial descent (logical intersection of the "core" and "GOLD stamp" sequence sets from greengenes.arb) and with their nearest-neighbor sequences. Then, the alignments were used to infer phylogenies by neighbor joining using evolutionary distances corrected by the Kimura two-parameter or maximum-likelihood correction using empirically determined gamma distribution and base frequencies (PAUP*4b10), maximum parsimony using heuristic search (PAUP*4b10), and maximum likelihood using the general time-reversible model (35). Bootstrap analysis was used to assess the robustness of inferred topologies. Finally, taxonomic identifiers CR1 to CR6 were assigned to monophyletic clusters that were reproducibly unaffiliated with previously described taxa in all topologies and which also contained at least three sequences greater than 1,000 nt from independent libraries.

Database searches. The environmental sequence database "env_nt" was downloaded from NCBI on 9 May 2006 (1,047,779 sequences and 1,063,283,128 total nucleotides; the list of accession numbers is available on plantpath.wisc.edu/ ~isen/CR). The following 16S rRNA gene sequence databases were downloaded from LBNL (14): the nonredundant, nonchimeric ("revered") database of full-length rRNA gene sequences (11 November 2005 version) and the full 16S rRNA

gene database (greengenes.arb; 27 April 2007 version). For searches of the environmental sequence database, putative PCR primer binding sites were identified with BLAST and putative amplicons were identified by Perl scripts (available at plantpath.wisc.edu/~isen/CR). For each primer search, pertinent BLAST parameters were as follows: no gaps, M = 1, N = -3, V = 1,000,000, and B = 1,000,000, and both the word size (W) and score threshold (S) were set to the length of the primer sequence used for the search. Putative 16S rRNA gene amplicons found in the environmental sequence database were confirmed by a BLAST score ratio (31) of greater than 0.80, computed for each sequence by dividing the BLAST score for its best match to the revered 16S rRNA gene database by the BLAST score of the sequence compared to itself; default BLAST parameters were used for these searches. All BLAST searches were implemented using WU-BLAST (blastn 2.0MP-WashU [10 May 2005]; W. Gish, personal communication).

Final note on primer sequences, nomenclature, and citation. In reviewing the literature to prepare this work, we found many instances of ambiguity regarding 16S rRNA gene PCR primers. These ambiguities resulted primarily from using the same name (e.g., "27F") to refer to primers having different sequences. Because of this, we feel it is important to be cautious in reporting 16S rRNA gene primer sequences, especially as more variations of commonly used primers are developed and published. Though there may eventually be a need to agree on a more systematic method of naming primers, most ambiguities can easily be minimized in publications by reporting the sequences of primers instead of designating them by a common name and reference.

Nucleotide sequence accession numbers. These sequence data have been submitted to GenBank under accession numbers EU245047 to EU246330.

RESULTS

Minimum primer length for Taq and S-Tbr polymerases. We conducted studies to determine the minimum primer length required for PCR amplification by the S-Tbr and Taq polymerases. For initial testing, primers were designed based on the nondegenerate versions of the 27F (sometimes called 8F) and 1492R primers used to amplify bacterial 16S rRNA genes (designated 27F-P and 1492R-P in this work) (Table 1) (18). Beginning with a 20-nt forward primer sequence (Fig. 1A), we designed a series of progressively shorter primers with lengths from 20 to 8 nt (Fig. 1A). To determine the lower limit for primer length, PCRs were performed using a forward primer from the series of decreasing lengths paired with a 19-nt reverse primer; E. coli genomic DNA was used as the template (Fig. 1B). With S-Tbr polymerase, successful amplification resulted with a primer as short as 10 nt. In a parallel experiment, Taq required the primer to be longer than 14 nt for detectable amplification (Fig. 1B). PCR products were produced with forward and reverse primers of 10 nt at annealing temperatures as high as $\sim 49^{\circ}$ C (Fig. 1C). To distinguish this new PCR method from traditional Taq PCR, we refer to the shortened primers as "miniprimers" and to PCR using them as "miniprimer PCR".

Miniprimer design and specificity. Several more 16S rRNA gene miniprimer candidates of 9 to 11 nt were designed (Fig. 2A) and tested to assess their performance and specificity (Fig.



FIG. 1. Establishing conditions for miniprimer PCR. (A) Minimum primer length. Beginning with a 20-nt primer, a series of shortened primers having a common 5' end was designed. These primers were used in combination with a 19-nt reverse primer (B) and a 10-nt primer (C) to demonstrate that 10-nt primers can be used for PCR. (B) Amplification using the forward primer series. When electrophoresed on an agarose gel, PCR with S-Tbr produces detectable amplification using primers as short as 10 nt (F10/R19). (C) Miniprimer PCR. Agarose gel electrophoresis of PCR products with S-Tbr and a range of annealing temperatures demonstrates robust amplification for a pair of 10-nt primers (F10/R10). In panels B and C, agarose gels are stained with ethidium bromide and arrows mark the expected size of PCR products.

2B). Miniprimers were designed by consulting 16S rRNA gene sequence alignments and published primer sequences to choose the most conserved \sim 10-nt regions within several long 16S rRNA gene primers. We based our designs on variants

most often reported in the literature (Table 1). The 788F-10 and 797R-10 primers (Fig. 2A) were designed to target the 10-nt region found to be the longest string of totally conserved bases in an alignment of 500 bacterial 16S rRNA gene sequences (8). By use of several combinations of miniprimers, PCRs were performed to amplify 16S rRNA genes from Escherichia coli or Halobacterium salinarum genomic DNA (from the domains Bacteria and Archaea, respectively). First, we selected miniprimer pairs that produced amplicons of the expected sizes with a minimum of spurious products. Then, amplicons from several PCRs were sequenced to verify that they were regions of 16S rRNA genes. All sequences from amplicons of the expected sizes were from the correct regions of the 16S rRNA gene targeted by each particular miniprimer pair. This analysis yielded several pairs of miniprimers that performed well for the amplification of bacterial and archaeal 16S rRNA genes (Fig. 2); in particular, 27F-10, 524F-10, and 2F-10 paired with 1505R-10 have reproducibly performed well in our analyses.

In silico evaluation of miniprimer targets in environmental sequences. To determine whether miniprimer PCR might expand the phylogenetic breadth of 16S rRNA gene sequences detected in environmental samples, the environmental sequence database was searched for putative 16S rRNA gene amplicons. It is important to note that this database compiles sequences obtained not by PCR but by random shotgun cloning of DNA directly extracted from environmental samples. A modeled PCR search identified putative 16S rRNA gene amplicons delimited by forward and reverse primer sequences separated by an appropriate distance (Fig. 3). Our goal was not to determine the fraction of 16S rRNA sequences in the database identified by the primers but to compare experimental methods for 16S rRNA gene discovery. Each database sequence and its reverse complement were searched using several of the best-performing primer pairs (Fig. 2); additional primer pairs were also designed to identify smaller amplicons

	В				
sequence (5'-3')	forward	reverse	Ec	Hs	16S
TTCCGGTTGA	2F-10	797R-10	-	+	+
TCCGGTTGAT	2F-10	1505R-10	-	+	+
AGAGTTTGAT	3F-10	1505R-10	-	+/	nd
GTTGATCCTG	27F-10	338R-10	+	nd	nd
GCCGCGGTAA	27F-10	922R-9	+	nd	nd
TTAGATACCC	27F-10	1505R-10	+	-	+
AAGGAATTG	12F-10	1403R-10	-	+/	+
CTGCTGCCTC	12F-10	1505R-10	-	+	+
GGGTATCTAA	524F-10	1403R-10	+	-	+
CAATTCCTT	524F-10	1407R-11	+	-	+
GGCGGTGTGT	524F-10	1505R-10	+	-	+
GACGGGCGGTG	788F-10	1505R-10	+	+	+
CCTTGTTACG	914F-9	1403R-10	+/	nd	nd
	Sequence (5'-3') TTCCGGTTGA TCCGGTTGAT AGAGTTTGAT GTTGATCCTG GCCGCGGTAA TTAGATACCC AAGGAATTG CTGCTGCCTC GGGTATCTAA CAATTCCTT GGCGGTGTGT GACGGGCGGTG CCTTGTTACG	B Forward TTCCGGTTGA 2F-10 TTCCGGTTGAT 2F-10 AGAGTTTGAT 2F-10 AGAGTTTGAT 27F-10 GTTGATCCTG 27F-10 GCCGCGGGTAA 27F-10 AAGGAATTG 27F-10 AAGGAATTG 12F-10 CTGCTGCCTC 12F-10 GGGTATCTAA 524F-10 CAATTCCTT 524F-10 GGACGGGCGTGTGT 788F-10 CCTTGTTACG 914F-9	B forward reverse TTCCGGTTGA 2F-10 797R-10 TCCGGTTGAT 2F-10 1505R-10 AGAGTTTGAT 2F-10 338R-10 GTTGATCCTG 27F-10 338R-10 GCCGCGGTAA 27F-10 922R-9 TTAGATACCC 27F-10 1505R-10 AAGGAATTG 12F-10 1505R-10 CTGCTGCTCC 12F-10 1505R-10 GGGTATCTAA 524F-10 1403R-10 CAATTCCTT 524F-10 1407R-11 GGCGGGTGTGT 524F-10 1505R-10 GACGGGCGGTG 788F-10 1505R-10 GACGGGCGGTG 914F-9 1403R-10	B forward reverse Ec TTCCGGTTGA 2F-10 797R-10 - TCCGGTTGAT 2F-10 1505R-10 - AGAGTTTGAT 3F-10 1505R-10 - GTTGATCCTG 27F-10 338R-10 + GCCGCGGTAA 27F-10 922R-9 + TTAGATACCC 27F-10 1505R-10 - AAGGAATTG 12F-10 1505R-10 + AAGGAATTG 12F-10 1505R-10 + CTGCTGCCTC 12F-10 1403R-10 - GGGTATCTAA 524F-10 1407R-11 + GGCGGTGTGT 524F-10 1407R-11 + GGCGGGTGTGT 788F-10 1505R-10 + GACGGGCGGTG 788F-10 1505R-10 + GACGGGCGGTG 914F-9 1403R-10 +/-	B Forward reverse Ec Hs TTCCGGTTGA 2F-10 797R-10 - + TCCGGTTGAT 2F-10 1505R-10 - + AGAGTTTGAT 3F-10 1505R-10 - +/- GTTGATCCTG 27F-10 338R-10 + nd GCCGCGGTAA 27F-10 922R-9 + nd TTAGATACCC 27F-10 1505R-10 - +/- AAGGAATTG 12F-10 1403R-10 + - AAGGAATTG 524F-10 1403R-10 + - GGCGGGTGTGT 524F-10 1407R-11 + - GACGGGCGGTG 788F-10 1505R-10 + + GACGGGCGGTG 914F-9 1403R-10 +/- nd

FIG. 2. Miniprimers designed and tested. (A) Sequences of miniprimers tested. (B) Performance of miniprimer pairs. Forward and reverse miniprimers were used to PCR amplify 16S rRNA genes from bacterial (*Escherichia coli* [Ec]) or archaeal (*Halobacterium salinarum*[Hs]) genomic DNA. PCR was determined successful if a band of the correct size was detected with minimum background. Success or failure with each template is indicated with + or -, respectively, in the appropriate column; a mark of +/- indicates that the product was in low abundance or contained a moderate amount of background but may be clonable. In several cases, amplicons were sequenced to confirm them to be the expected 16S rRNA sequences; confirmation is indicated by a + in the column labeled "16S." nd, not determined.

Forward primer		Reverse primer		Separation (nt)	Sequences	Unique	16S
Name	Sequence (5'-3')	Name	Sequence (5'-3')				
27F - P	AGAGTTTGATCCTGGCTCAG	338R	CTGCTGCCTCCCGTAGGAGT	230-430	84	68	54
27F-10	AGAGTTTGAT	338R-10	CTGCTGCCTC	230-430	292	242	226
27F - P	AGAGTTTGATCCTGGCTCAG	1492R	GGTTACCTTGTTACGACTT	1284-1684	0	na	na
27F-10	AGAGTTTGAT	1505R-10	CCTTGTTACG	1284-1684	32	31	30
4Fa	TCCGGTTGATCCTGCCRG	805R	GACTACCAGGGTATCTAAT	701-901	0	na	na
12F-10	GTTGATCCTG	797R-10	GGGTATCTAA	701-901	4	4	4
4Fa	TCCGGTTGATCCTGCCRG	1492R	GGTTACCTTGTTACGACTT	1288-1688	0	na	na
12F-10	GTTGATCCTG	1505R-10	CCTTGTTACG	1288-1688	2	2	2
356F	ACTCCTACGGGAGGCAGCAG	805R	GACTACCAGGGTATCTAAT	349-549	112	87	86
366F-10	GAGGCAGCAG	797R-10	GGGTATCTAA	349-549	232	169	165
515F	GTGCCAGCMGCCGCGGTAA	805R	GACTACCAGGGTATCTAAT	190-390	191	131	127
524F-10	GCCGCGGTAA	797R-10	GGGTATCTAA	190-390	357	222	217
515F	GTGCCAGCMGCCGCGGTAA	1492R	GGTTACCTTGTTACGACTT	877-1077	6	6	6
524F-10	GCCGCGGTAA	1505R-10	CCTTGTTACG	877-1077	68	65	62
805F	ATTAGATACCCTGGTAGTC	1492R	GGTTACCTTGTTACGACTT	587-787	12	10	10
788F-10	TTAGATACCC	1505R-10	CCTTGTTACG	587-787	165	147	141
1391F	TGYACWCACCGCCCGTC	1492R	GGTTACCTTGTTACGACTT	50-150	43	18	18
1403F-10	ACWCACCGCC	1505R-10	CCTTGTTACG	50-150	496	247	221
				long primer total	448	320	301
				miniprimer total	1648	1129	1068

FIG. 3. Comparison of putative miniprimer and long-primer amplicons in the GenBank environmental sequence database. Pairs of long primers and miniprimers were used to identify putative 16S rRNA gene amplicons in the GenBank environmental sequence database. Horizontal lines separate aligned, matched pairs of long primers (gray shading) and miniprimers (no shading) that target the same 16S rRNA gene region. Sequences delimited by forward and reverse primer sequences separated by an appropriate distance were identified. The numbers of total sequences found (Sequences), unique sequences present within the total (Unique), and 16S rRNA gene sequences present within the unique sequences (16S) are indicated in the appropriate columns. 16S rRNA gene sequences were identified by a BLAST score ratio of >0.80 (see Materials and Methods). Degenerate positions are indicated using R for A or G, M for A or C, and W for A or T; na, not applicable.

because the average sequence length of the environmental sequence database entries is \sim 1,000 nt, lowering the probability of identifying long amplicons. To compare the abundances of putative amplicons defined by pairs of long primers and miniprimers, matched primer pairs were used to target the same regions of the 16S rRNA gene (Fig. 3). Miniprimers identified 1,648 putative amplicons and the standard primers identified 448; surprisingly, of the 1,648 miniprimer sequences, 971 were not identified by any long-primer pair tested. By BLAST comparison against a database of 30,312 nonredundant, nonchimeric bacterial and archaeal 16S rRNA gene sequences of at least 1,350 nt, 1,068 miniprimer and 301 longprimer sequences were verified to be regions of 16S rRNA genes (Fig. 3). Importantly, each miniprimer pair identified 2to 10-fold more putative 16S rRNA gene amplicons (Fig. 3). This analysis predicted that miniprimer pairs can amplify more 16S rRNA gene sequences in environmental samples than can standard long primers.

Next, the phylogenetic breadth represented by the putative 16S rRNA gene amplicons was determined. In sum, the miniprimer and long-primer pairs identified 1,068 and 301 total putative 16S rRNA gene sequences, respectively. To build the best phylogenies and remove redundant sequences, the longest sequence identified from each database entry was used in instances when the same database entry was identified by multiple primer pairs. After removing replicated sequences, the remaining 685 miniprimer and 205 long-primer sequences were aligned to a comprehensive 16S rRNA gene sequence database (16). One long-primer sequence and 14 miniprimer sequences could not be satisfactorily aligned to the database. A phylogenetic tree was constructed using the remaining sequences, and the resulting taxonomic distribution indicated that almost all the sequences originated from *Bacteria*, with Proteobacteria sequences in the majority (see S1 in the supplemental material). For almost every taxonomic class, miniprimers identified more sequences than long primers. Also, miniprimers identified sequences in some classes for which long primers identified no sequences (see S1 in the supplemental material); this generally occurred for classes in which few members were identified but also occurred for Sargasso Sea group 11 (SAR11), the most abundant class of marine bacteria currently known (29). In addition, miniprimers identified nine sequences from Archaea, a domain for which long primers identified no sequences. This suggests that miniprimer PCR can amplify more sequences in more phylogenetic groups than can standard long primers.

Evaluation of miniprimer PCR with a soil sample. To determine whether the computational analysis reflected experimental results, miniprimer PCR was used in a study to amplify 16S rRNA genes from soil by use of the primer pairs 27F-10/1505R-10, 12F-10/1505R-10, and 524F-10/1505R-10 (Table 1 and Fig. 2). In this study, we did not attempt a comprehensive analysis of a soil sample but only an assessment of the miniprimers and whether they could detect sequences with mismatches to the standard primers. After PCR and cloning, 32 clones from each primer pair were partially sequenced from one end to examine the primer binding regions. Products from the 27F-10/1505R-10 PCR were chosen for further study because 27F-10 comprises the first 10 nt from the 5' half of the 27F suite of primers (Table 1) and thus enabled the identifi-

TABLE 2. Candelaria microbial mat 16S rRNA gene libraries constructed

Library	Forward primer	Reverse primer	Polymerase	No. of sequences
M1	27F-10	1505R-10	S-Tbr	255
P1	27F-P	1492R-P	Taq	147
H1	27F-HT	1492R-HT	Taq	150
M2	27F-10	1505R-10	S-Tbr	343
P2	27F-P	1492R-P	S-Tbr	202
H2	27F-HT	1492R-HT	S-Tbr	184

cation of amplicons whose sequences had mismatches to the 3' half of 27F (see S2 in the supplemental material). In addition, the 27F-10/1505R-10 miniprimer pair amplifies nearly the entire 16S rRNA gene, and it performed well in pilot experiments. Very-high-quality sequences were obtained for 10 of the 32 27F-10/1505R-10 sequences; of these 10 end sequences, 5 began with the 27F-10 sequence and 5 began with the 1505R-10 sequence. Two of the sequences anchored by the miniprimer 27F-10, A01 and A04, contained mismatches to the 27F-P binding sequence and also to the sequence of 27F-HT, a degenerate primer designed to be more general (Table 1; also see S2 in the supplemental material).

After A01 and A04 were fully sequenced, it was noted that the A01 sequence also had a mismatch within the 4 nt at the 3' end of the 1492R binding site that are not shared by 1505R-10 (see S2 in the supplemental material). The A01 and A04 sequences were aligned to the comprehensive 16S rRNA gene sequence database to assess the sequences within conserved regions (see S2 in the supplemental material). The alignment revealed that A01 and A04 contained mismatches within the 27F-P and 1492R-P primer binding regions and within other regions of the 16S rRNA gene that are very highly conserved across the Bacteria domain (see S2 in the supplemental material). Importantly, predictions of rRNA secondary structure based on the Arb alignments supported the presence of the divergent bases in the A01 and A04 sequences-of the nine divergent bases observed (see S2 in the supplemental material), seven are located in base-paired regions, and for five of these bases, the expected complementary changes are present at the proper locations to preserve Watson-Crick pairing, and in the remaining two cases, G:G and G:A base pairs are present for all members of the phylogenetic clusters in which A01 and A04 reside, respectively. An examination of 20 randomly selected divergent bases within conserved helical regions also revealed the expected complementary bases at the correct locations to preserve Watson-Crick pairing for all 20 sites chosen (data not shown). Thus, given the covariation present in the sequences, it is exceedingly unlikely that the divergent bases arose from PCR errors or chimeric sequences.

Miniprimer PCR characterization of the Candelaria microbial mat community. We applied the method to a mature microbial mat community from an extreme environment, the hypersaline Candelaria lagoon of the Cabo Rojo salterns. This mat community was chosen for study because it was expected to have intermediate diversity and interesting phylotypes. To identify members of the mat community for this study, we constructed six 16S rRNA gene sequence libraries with either the miniprimer or standard long-primer techniques.

We constructed libraries with the miniprimer pair 27F-10/ 1505R-10 and two different long-primer pairs: 27F-P/1492R-P and 27F-HT/1492R-HT (Table 1). The primers 27F-P and 1492R-P are "first-generation" nondegenerate universal bacterial primers that remain in widespread use today (1, 5, 18, 19, 39), and 27F-HT and 1492R-HT are more recently published primers that are based on the original first-generation primers and include several degenerate positions to broaden their scope of targets (37). These primers produce nearly full-length 16S rRNA gene sequences and thus provide a fair assessment of the miniprimer method relative to standard methods in common use; duplicate libraries were constructed with each of the primer pairs (Table 2).

The Candelaria microbial mat sequence libraries comprised over 40 bacterial divisions, with the *Chloroflexi, Bacteroidetes, Halanaerobiales*, and *Planctomycetes* divisions having the greatest representation (Fig. 4). All of the most highly populated divisions identified were represented in each of the libraries, though in different proportions (Fig. 4). Notably, miniprimers amplified a greater proportion of sequences that could not be classified at or below the division level (see below).

The miniprimers amplified more sequences with poor matches to previously isolated 16S rRNA gene sequences (Fig. 5). Miniprimer libraries contained a larger fraction than did long-primer libraries of sequences that matched the database at distances greater than 0.10 (Fig. 5); conversely, the miniprimer libraries matched many fewer sequences at distances less than or equal to 0.05. The distributions of database matches for the two long-primer libraries were very similar and



FIG. 4. Distribution of microbial mat library sequences in major bacterial divisions identified. The 10 taxonomic divisions with the highest representation in the libraries are shown as their relative proportions within the miniprimer (M) and two types of long-primer (P, 27F-P/1492R-10; H, 27F-HT/1492R-HT) libraries. The aggregate distribution of all libraries is also shown (all). Phylogenetic assignments were performed in Arb; the "other divisions" category comprises 31 divisions whose representation is less than \sim 3 to 5% in each library. Proper *Proteobacteria* class names with standing have been shortened using Greek letters.



FIG. 5. Distribution of microbial mat best database matches. The fractions of sequences whose best database matches fall into the designated distance intervals are shown for the long-primer (P, 27F-P/1492R-10; H, 27F-HT/1492R-HT) and miniprimer (M) libraries. Distances of 0.05, 0.10, and 0.20 approximately define the taxonomic levels of genus, order, and division, respectively (32). Plotted values and errors are the averages and standard deviations of six maximum like-lihood calculations using six different taxonomies (15).

more similar to each other than either distribution was to the miniprimer distribution (Fig. 5). Thus, the miniprimer method appears to amplify more novel sequences than the long-primer methods.

A small fraction of sequences matched the database at distances greater than 0.20 (Fig. 5). Most of these sequences were not placed into defined taxa or were placed into recently defined taxa, many of which include sequences isolated from similar environments (Fig. 6). A few groups contain clusters of miniprimer sequences, in particular a subgroup within WS6 and the CR1 and CR2 groups (Fig. 6). The clusters CR1 and CR2 (Fig. 6; also see S3 in the supplemental material) formed at distances of 0.26 and 0.24, respectively, from previously isolated sequences and thus are likely to be representatives of new division-level taxa. Another four clusters formed at distances slightly below the division level: CR3 to CR5 at 0.18 and CR6 at 0.17 (see S3 in the supplemental material). Interestingly, seven of the eight sequences defining CR1 and CR2, and 25 of the 32 sequences defining CR3 to CR6, are miniprimer sequences. Many other sequences also branched deeply at distances of 0.29 to 0.16 to their nearest neighbors but did not meet all the criteria for defining new division-level taxa. A small fraction of library sequences accounted for these putative novel groups: approximately 3.5% of library sequences comprised monophyletic groups at distances of at least 0.17, with 5.9% of the miniprimer sequences contributing and 1.5% of the two long-primer libraries combined contributing (1.7%) from the 27F-P/1492R-P and 1.2% from the 27F-HT/ 1492R-HT libraries, respectively).

Other Candelaria sequences expanded the membership of previously defined taxonomic groups. The *Chloroflexi* group Eub6 is almost entirely composed of sequences cloned in this study and from the Guerrero Negro microbial mats (26) and was the largest single taxonomic group from any division to which miniprimer sequences contributed; the other *Chloroflexi* subtaxa were largely populated by long-primer sequences, suggesting the existence of a primer bias for particular groups of Chloroflexi. Similar subdivision bias is demonstrated by the Halanaerobiales sequences: most Halobacteroidaceae family sequences were from miniprimer libraries, whereas most Halanaerobiaceae family sequences were from long-primer libraries. A large fraction of miniprimer sequences was clustered within the Planctomycetes group and sequences from all libraries were distributed over many Deltaproteobacteria groups. Five sequences contributed to the Deltaproteobacteria group GN04, which was recently defined based on sequences isolated from the Guerrero Negro microbial mat. Several miniprimer sequences populated the Spirochaetes group GN05-1, which was first defined as a group of Guerrero Negro sequences and is currently exclusive to sequences cloned from microbial mats. Very few miniprimer sequences were from the Bacteroidetes and Firmicutes divisions, though large fractions of long-primer sequences were classified in these divisions (Fig. 4). The remaining divisions into which library sequences were placed each contributed less than 5% to their respective libraries.

To evaluate the reproducibility of the three library construction methods and to compare the methods, the 16S rRNA gene sequence libraries were compared to each other by calculating similarity indices (34) using 97% sequence similarity to define OTUs. Here, we focus on the classic incidence-based Sørenson similarity index, a measure of membership, and the Clayton θ , a nonparametric maximum likelihood estimator of community structure that considers abundance (41). These calculations demonstrate two trends (Table 3): first, duplicate libraries constructed with the same primers show high levels of similarity in both membership and structure; and second, libraries constructed with long primers are typically more similar to each other than to libraries constructed with miniprimers. These trends are particularly evident when relative abundances are considered (θ , Table 3). Importantly, the data suggest no significant dependence of library membership or structure on the particular polymerase used, as demonstrated by the high similarities of duplicate libraries constructed with the same primer pairs, 27F-P/1492R-P or 27F-HT/1492R-HT, but amplified with different polymerases (Tables 2 and 3). Moreover, the miniprimer method is as reproducible as either of the other methods tested here using long-primer pairs. Thus, these data suggest that differences in compositions of the libraries constructed with miniprimers and long-primer libraries are not due to the enzyme used for amplification.

The coverage and richness of the 1,281 Candelaria sequences were estimated by calculating rarefaction curves and the nonparametric diversity estimators Chao1 and Ace1 (see S4 in the supplemental material). Rarefaction analysis suggested the sequences represented a diversity of at least 423 phylotypes defined at an identity threshold of 97% but indicated that sampling was far from complete at this threshold. At an 80% identity threshold, the estimates appear to have reached plateaus, suggesting that sampling is more complete at approximately the division level. Chao1 and Ace1 suggest the presence of 91 to 124 taxa (95% confidence intervals) at the 80% threshold, and the rarefaction analysis indicates that the libraries represent 87 taxa at this level. Thus, though most division-level taxa have probably been identified in the microbial mat sample, these analyses suggest a high degree of diver-



Library	Similarity to indicated library						
	M1	M2	P1	P2	H1	H2	
M1		0.64 ± 0.07	0.07 ± 0.02	0.08 ± 0.03	0.06 ± 0.02	0.08 ± 0.03	
M2	0.60 ± 0.09		0.06 ± 0.02	0.08 ± 0.02	0.03 ± 0.01	0.06 ± 0.02	
P1	0.58 ± 0.16	0.49 ± 0.14		0.52 ± 0.08	0.51 ± 0.11	0.56 ± 0.10	
P2	0.50 ± 0.12	0.37 ± 0.10	0.76 ± 0.15		0.58 ± 0.10	0.64 ± 0.08	
H1	0.42 ± 0.15	0.49 ± 0.16	0.88 ± 0.13	0.74 ± 0.15		0.73 ± 0.07	
H2	0.51 ± 0.12	0.55 ± 0.13	0.72 ± 0.10	0.60 ± 0.09	0.65 ± 0.11		

TABLE 3. Estimation of library similarities^a

^{*a*} The Sørenson (S, lower left) and Clayton (θ, upper right) indices were computed for each pair of libraries using an OTU definition of 97% sequence identity. For both indices, values can range from 0 to 1, with 1 indicating identity. Values are reported with their standard errors. See Table 2 for details regarding the six libraries.

sity at a greater taxonomic resolution and that many organisms have yet to be identified.

Lastly, we examined the miniprimer clone sequences to quantify mismatches within the regions that would be targeted by the long primers 27F-P, 27F-HT, 1492R-P, and 1492R-HT (Fig. 7). For the 27F binding region, 309 and 266 of the 598 miniprimer sequences deviated from the 27F-P and 27F-HT primer sequences, respectively, at one or more nucleotides; sequences with two mismatches to 27F-P and 27F-HT numbered 20 and 10, respectively, and 1 sequence had three mismatches to the 27F-P sequence. Though the miniprimer sequence consensus largely agreed with the sequences of the long 27F primers, the nucleotide frequency at position 12 did not agree with the degeneracy present at this position in 27F-HT. Position 12 of 27F-HT is designed to match A or C; however, while in the cloned miniprimer sequence libraries C was present in the majority of sequences, G was present more than four times as often as was A, and T was observed twice. Similar analysis of the 1492R binding region identified one sequence from the miniprimer clone libraries having a single mismatch to 1492R-P and 1492R-HT (data not shown).

DISCUSSION

In this work, we show that 10-nt miniprimers can be used to identify divergent 16S rRNA gene sequences from environmental samples. Miniprimer PCR amplified a greater proportion than did standard primers of sequences that were novel or that poorly matched a database of previously isolated 16S rRNA gene sequences. This confirmed a prediction based on a computational search of the NCBI environmental sequence database. Importantly, we show that miniprimer PCR is as reproducible as other methods and provides a different portrait of microbial communities than is attained with current methods.

Potential amplification of nontarget sequences. One may expect that the shorter size of miniprimers could increase promiscuous binding to targets other than the 16S rRNA gene.

However, analysis of environmental sequences suggests this may not be problematic. Of the 1,129 predicted miniprimer amplicons in the NCBI environmental sequence database (Fig. 3), 61 (~5.4%) were found not to be 16S rRNA gene sequences; a similar calculation for long primers yields a similar false-positive rate (19/320 \approx 5.9%). In addition, the rates of identifying mitochondrial and chloroplast sequences (see S1 in the supplemental material) were also the same (8/685 \approx 3/205 $\approx 1\%$) for both methods. In support of this, the microbial mat libraries contained no eukaryote, chloroplast, or mitochondrial sequences, though eukaryotes are known to associate with the Candelaria mats (11). Overall, the results suggest that miniprimers can identify more 16S rRNA gene sequences without an increased rate of amplifying false positives. However, much additional testing and sequencing of amplicons isolated from environmental samples should be completed to test this prediction.

Benefits of miniprimer PCR. Adding miniprimer PCR to the tools used for analyzing microbial communities may enable a more accurate measure of 16S rRNA gene sequences in environmental samples by expanding the sequences detectable by PCR. For example, in the 1 Gbp of randomly cloned sequence in the environmental sequence database, if one assumes a microbial genome size of 2 to 3 Mbp and an average of 2.5 16S rRNA genes per genome (2), it is expected that ~ 800 to 1,200 16S rRNA genes are present in the sample. When totaled, the miniprimer computational searches identified 1,068 putative 16S rRNA gene amplicons (Fig. 3), in close agreement with the estimate, while the searches with the longer primers totaled 301 (Fig. 3). If one assumes that decreasing the length of a PCR primer by 1 nt can increase by fourfold the number of perfectly matched sequence targets, then reducing the length of a primer from 20 nt to 10 nt could theoretically increase the possible targets by a factor of a million ($\approx 4^{10}$). Sequence variation within the regions targeted by these particular miniprimers may render the theoretical limit meaningless in this

FIG. 6. Microbial mat sequences with low-scoring database matches. Miniprimer (red), long-primer (green and blue), and reference (black) sequences were assembled into a phylogenetic tree by use of maximum likelihood (35). The best-scoring tree of 20 independent calculations is shown; branches supported by bootstrap values of greater than 50 (100 iterations) are labeled. Some bootstrap values near closely spaced terminal clusters have been removed for clarity; none of the removed values were less than 50 and most were greater than 80. Uncultured isolates are identified by their accession numbers and by the environment from which the sequences were cloned; cultured isolates are in italics. Taxonomic groups are indicated for clusters of three or more sequences that could be classified. Seven archaeal 16S rRNA gene sequences were used to root the tree.



FIG. 7. Mismatches of 27F-P and 27F-HT primers to microbial mat miniprimer library sequences. (A) Portions of long primers 27F-P and 27F-HT that bind past the 3' end of miniprimer 27F-10. (B) Bases represented by degenerate codes in 27F-HT. (C and D) The relative (C) and absolute (D) base frequencies observed within the 598 sequences of the miniprimer sequence libraries at each position within the 3' halves of the 27F-P and 27F-HT binding sites. Note that at some positions the quantities of bases shown in panel D are not large enough to be resolved by the representation shown in panel C.

instance; nonetheless, this upper limit could inform other applications of the method.

In addition to expanding the range of detectable targets, combining miniprimer PCR with standard techniques might increase the accuracy of environmental sampling by enabling estimates of microbial diversity to reflect sampled communities more closely. When PCR is used to analyze samples containing multiple heterogeneous templates, many mechanisms can result in a biased distribution of amplicons that does not accurately represent the distribution of templates in the sample (3, 22, 30, 36). Miniprimers may reduce these kinds of biases in two ways. First, because miniprimers are shorter and can target the most conserved regions of the 16S rRNA gene, the potential for mismatch is reduced compared with long primers. The degeneracy of 27F-HT was not appropriate for the nucleotide frequencies of the miniprimer sequences (Fig. 7), suggesting that particular communities may be better analyzed with a miniprimer such as 27F-10. Second, because miniprimers target smaller regions, the requirement for degenerate positions in primers can be eliminated or minimized, thus removing or decreasing preferential binding of particular primer species within a degenerate primer mixture. Obviously, the use of shortened primers may introduce different biases into PCRbased analyses of microbial communities, and more study is required to understand that potential, perhaps using known mixtures of bacterial DNAs. However, despite the biases of any PCR method, miniprimer PCR may be useful to supplement traditional methods either to reduce or make apparent the biases that occur in analysis of environmental samples and allow a more accurate description of microbial communities.

Miniprimer design. Though general rules have been established for *Taq* PCR primer design, it is not yet evident what parameters require optimization when primers are substantially shortened. Successful amplification was obtained using miniprimers whose G+C content covered a broad range from 30 to 90% (Fig. 2). Thus, product formation appeared unrelated to the G+C content of miniprimers. In our tests, a high rate of success has been obtained with 1505R-10 paired with many different forward primers (Fig. 2). Of the putative primer binding sequences identified in the environmental sequence database, the fewest sites were found for 1505R-10, suggesting that a productive miniprimer maximizes specificity for the target sequence in the context of the total template sequence present in the PCR. This may be a good rule with which to begin miniprimer design. It should be noted that the miniprimers used here are initial versions and that they should be optimized further. For instance, neither 27F-10 nor 27F perfectly matches Chlamydia, and 788F-10 does not match Deha*lococcoides*. Thus, some problems associated with long primers remain concerns in miniprimer design, and regularly checking, curating, and updating miniprimer sequences are as required for miniprimers as for the longer primers.

Future prospects for miniprimer PCR. The use of miniprimers in PCR can benefit many areas of biological research in several ways. PCR with miniprimers may enable PCR strategies that are not possible when primer design is restricted to a minimum length of ~ 20 nt. Some strategies might benefit from using even shorter primers-preliminary data indicate that miniprimers can be as short as 8 nt when paired with a primer of 15 to 20 nt (data not shown). Also, miniprimer sequences provide a small amount of additional sequence data that may be helpful for making fine discriminations among closely related sequences. Finally, the combination of highly processive enzymes and shorter primers may decrease the cost of PCR. The increased processivity enables PCR to be performed with smaller amounts of enzyme (40), and shorter primers are less costly to synthesize. Although cost savings may be minimal for small projects, these considerations become important for projects requiring very large numbers of PCRs.

ACKNOWLEDGMENTS

We gratefully acknowledge Tuomas Tenkanen (Finnzymes Oy) for providing DNA polymerases and for helpful technical discussions, Todd DeSantis (Lawrence Berkeley National Laboratory) and Phil Hugenholtz (DOE Joint Genome Institute) for assistance with 16S rRNA gene sequence databases, Kirk Harris (University of Colorado) and Joe Felsenstein (University of Washington) for help with constructing phylogenetic trees, and members of the Ruvkun laboratory (Massachusetts General Hospital) and the Handelsman laboratory (University of Wisconsin) for helpful discussion of this work. We also acknowledge the SETG team, especially George Church (Harvard Medical School) and Maria Zuber (Massachusetts Institute of Technology), for thoughtful discussions regarding this research and the Massachusetts Institute of Technology Center for Space Research for guidance.

This work was supported by grant number F32AI065067 from the National Institute of Allergy and Infectious Diseases, grant number NRA-01-01-ASTID-020 from the National Aeronautics and Space Administration, and grant numbers MCB-0132085 and MCB-0455620 from the National Science Foundation Microbial Observatory Program.

The content is solely our responsibility and does not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases, the National Institutes of Health, the National Aeronautics and Space Administration, or the National Science Foundation.

REFERENCES

- Abulencia, C. B., D. L. Wyborski, J. A. Garcia, M. Podar, W. Chen, S. H. Chang, H. W. Chang, D. Watson, E. L. Brodie, T. C. Hazen, and M. Keller. 2006. Environmental whole-genome amplification to access microbial populations in contaminated sediments. Appl. Environ. Microbiol. 72:3291– 3301.
- Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz. 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430:551–554.
- Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl. Environ. Microbiol. 71:8966–8969.
- Amann, R. I., W. Ludwig, and K.-H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59:143–169.
- Angenent, L. T., S. T. Kelley, A. St. Amand, N. R. Pace, and M. T. Hernandez. 2005. Molecular identification of potential pathogens in water and air of a hospital therapy pool. Proc. Natl. Acad. Sci. USA 102:4860–4865.
- Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. 71:7724–7736.
- Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman. 2006. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. Appl. Environ. Microbiol. 72:5734–5741.
- Baker, G. C., J. J. Smith, and D. A. Cowan. 2003. Review and re-analysis of domain-specific 16S primers. J. Microbiol. Methods 55:541–555.
- Blackwood, C. B., A. Oaks, and J. S. Buyer. 2005. Phylum- and class-specific PCR primers for general microbial community analysis. Appl. Environ. Microbiol. 71:6193–6198.
- Brofft, J. E., J. V. McArthur, and L. J. Shimkets. 2002. Recovery of novel bacterial diversity from a forested wetland impacted by reject coal. Environ. Microbiol. 4:764–769.
- Casillas-Martinez, L., M. L. Gonzalez, Z. Fuentes-Figueroa, C. M. Castro, D. Nieves-Mendez, C. Hernandez, W. Ramirez, R. E. Sytsma, J. Perez-Jimenez, and P. T. Visscher. 2005. Community structure, geochemical characteristics and mineralogy of a hypersaline microbial mat, Cabo Rojo, PR. Geomicrobiol. J. 22:269–281.
- Cline, S. W., and W. F. Doolittle. 1987. Efficient transfection of the archaebacterium *Halobacterium halobium*. J. Bacteriol. 169:1341–1344.
- Derakshani, M., T. Lukow, and W. Liesack. 2001. Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. Appl. Environ. Microbiol. 67:623–631.
- DeSantis, T. Z., I. Dubosarskiy, S. R. Murray, and G. L. Andersen. 2003. Comprehensive aligned sequence construction for automated design of effective probes (CASCADE-P) using 16S rDNA. Bioinformatics 19:1461–1468.
- DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72:5069–5072.
- DeSantis, T. Z. J., P. Hugenholtz, K. Keller, E. L. Brodie, N. Larsen, Y. M. Piceno, R. Phan, and G. L. Andersen. 2006. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res. 34:W394–W399.
- Dojka, M. A., P. Hugenholtz, S. K. Haack, and N. R. Pace. 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. Appl. Environ. Microbiol. 64:3869– 3877.
- Harris, J. K., S. T. Kelley, and N. R. Pace. 2004. New perspective on uncultured bacterial phylogenetic division OP11. Appl. Environ. Microbiol. 70:845–849.
- Hayashi, H., M. Sakamoto, M. Kitahara, and Y. Benno. 2006. Diversity of the *Clostridium coccoides* group in human fecal microbiota as determined by 16S rRNA gene library. FEMS Microbiol. Lett. 257:202–207.
- 20. Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to

detect chimeric sequences in multiple sequence alignments. Bioinformatics **20**:2317–2319.

- Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. Impact of cultureindependent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. 180:4765–4774.
- Ishii, K., and M. Fukui. 2001. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Appl. Environ. Microbiol. 67:3753–3755.
- Kato, C., and L. Li. 1999. Bacterial diversity in deep-sea sediments from different depths. Biodivers. Conserv. 8:659–677.
- Krebs, M. P., T. Hauss, M. P. Heyn, U. L. RajBhandary, and H. G. Khorana. 1991. Expression of the bacterioopsin gene in *Halobacterium halobium* using a multicopy plasmid. Proc. Natl. Acad. Sci. USA 88:859–863.
- Lane, D. J. 1991. 16S/23S rRNA sequencing, p. 115–147. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics, John Wiley & Sons Ltd., Chichester, United Kingdom.
- Ley, R. E., J. K. Harris, J. Wilcox, J. R. Spear, S. R. Miller, B. M. Bebout, J. A. Maresca, D. A. Bryant, M. L. Sogin, and N. R. Pace. 2006. Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. Appl. Environ. Microbiol. 72:3685–3695.
- 27. Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32:1363–1371.
- Miyoshi, T., T. Iwatsuki, and T. Naganuma. 2005. Phylogenetic characterization of 16S rRNA gene clones from deep-groundwater microorganisms that pass through 0.2-micrometer-pore-size filters. Appl. Environ. Microbiol. 71:1084–1088.
- Morris, R. M., M. S. Rappe, S. A. Connon, K. L. Vergin, W. A. Siebold, C. A. Carlson, and S. J. Giovannoni. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420:806–810.
- Polz, M. F., and C. M. Cavanaugh. 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. 64:3724–3730.
- Rasko, D. A., G. S. Myers, and J. Ravel. 2005. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics 6:2.
- Schloss, P. D., and J. Handelsman. 2004. Status of the microbial census. Microbiol. Mol. Biol. Rev. 68:686–691.
- Schloss, P. D., and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. 71:1501–1506.
- Schloss, P. D., and J. Handelsman. 2006. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. Appl. Environ. Microbiol. 72:6773–6779.
- Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
- Suzuki, M. T., and S. J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62:625–630.
- 37. Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428:37–43.
- 38. Vergin, K. L., E. Urbach, J. L. Stein, E. F. DeLong, B. D. Lanoil, and S. J. Giovannoni. 1998. 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.
- Vickerman, M. M., K. A. Brossard, D. B. Funk, A. M. Jesionowski, and S. R. Gill. 2007. Phylogenetic analysis of bacterial and archaeal species in symptomatic and asymptomatic endodontic infections. J. Med. Microbiol. 56:110–118.
- Wang, Y., D. E. Prosen, L. Mei, J. C. Sullivan, M. Finney, and P. B. Vander Horn. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. 32:1197– 1207.
- Yue, J. C., M. K. Clayton, and F. C. Lin. 2001. A nonparametric estimator of species overlap. Biometrics 57:743–749.