APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Feb. 2004, p. 845–849 Vol. 70, No. 2
0099-2240/04/$08.00+0 DOI: 10.1128/AEM.70.2.845–849.2004

New Perspective on Uncultured Bacterial Phylogenetic Division OP11

J. Kirk Harris,[1,2] Scott T. Kelley,[1]† and Norman R. Pace[1]*

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309-0347,[1] *and Graduate Group in Microbiology, University of California, Berkeley, California 94720*[2]

Received 25 July 2003/Accepted 23 October 2003

Organisms belonging to the OP11 candidate phylogenetic division of *Bacteria* have been detected only in rRNA-based sequence surveys of environmental samples. Preliminary studies indicated that such organisms represented by the sequences are abundant and widespread in nature and highly diverse phylogenetically. In order to document more thoroughly the phylogenetic breadth and environmental distribution of this diverse group of organisms, we conducted further molecular analyses on environmental DNAs. Using PCR techniques and primers directed toward each of the five described subdivisions of OP11, we surveyed 17 environmental DNAs and analyzed rRNA gene sequences in 27 clonal libraries from 14 environments. Ninety-nine new and unique sequences were determined completely, and approximately 200 additional clones were subjected to partial sequencing. Extensive phylogenetic comparisons of the new sequences to those representing other bacterial divisions further resolved the phylogeny of the bacterial candidate division OP11 and identified two new candidate bacterial divisions, OP11-derived 1 (OD1) and Sulphur River 1 (SR1). The widespread environmental distribution of representatives of the bacterial divisions OD1, OP11, and SR1 suggests potentially conspicuous biogeochemical roles for these organisms in their respective environments. The information on environmental distribution offers clues for attempts to culture landmark representatives of these novel bacterial divisions, and the sequences are specific molecular signatures that provide for their identification in other contexts.

Sequence analyses of rRNA genes from natural microbial communities have identified a broad diversity of previously unknown microorganisms. Currently, about half of the >70,000 rRNA sequences in the public databases represent uncultured organisms (3). More than one-third of the 40 to 50 main relatedness groups, natural divisions, of the phylogenetic domain *Bacteria* are known only from detection of rRNA sequences and have no described cultivated representatives. These division-level clades with no cultured representatives, typically known from only limited numbers of rRNA sequences, have been termed candidate divisions to reflect the limited documentation that describes them (12).

One of the most widely distributed phylogenetic candidate divisions in environmental molecular surveys is the division OP11, first encountered in the Yellowstone hot spring Obsidian Pool (11). Sequences representing the OP11 group have since been found in marine and freshwater sediments (16, 21, 22; P. Hugenholtz, K. L. Hershberger, J. L. Flanagan, B. Kimmel, and N. R. Pace, Abstr. 97th Gen. Meet. Am. Soc. Microbiology, abstr. N-23, 1997) geothermal pools (12; Hugenholtz et al., 97th Gen. Meet. Am. Soc. Microbiol., 1997), subsurface bedrock (18), soil (2, 13), the human oral cavity (17), and a hydrocarbon-polluted aquifer (6). Based on the extent of rRNA sequence divergence within the group, 33% overall sequence divergence, the OP11 clade, if truly monophyletic, would represent the most diverse division within the domain *Bacteria*. This extent of rRNA sequence divergence is substantial in comparison to other bacterial relatedness groups. For instance, the extent of rRNA sequence variation in the *Proteobacteria* division is about 23% excluding mitochondria (6). Before the present study, the OP11 division appeared to comprise five monophyletic subdivisions. However, the early analyses were based on a relatively small selection of OP11 sequences, and some of the subdivisions contained only a few sequences (12). Consequently, phylogenetic calculations relating these OP11 subgroups to one another and to other bacterial divisions were uncertain.

In order to analyze further the phylogenetic diversity and environmental distribution of the candidate phylogenetic division OP11, we have determined novel OP11-related sequences from a variety of chemically distinct environments. Phylogenetic analyses of the novel sequences provide new perspective on the former candidate bacterial division OP11.

MATERIALS AND METHODS

Sample collection and DNA extraction. Samples included oxidized-iron deposits from a (nominal) *Gallionella* community located on Boulder Creek (Boulder, Colo.); a marine tidal mat located near Bolinas, Calif.; and sediment with microbial streamers from Sulphur River, Parkers Cave, Ky. DNA was extracted using a modified bead-beating and solvent extraction protocol as previously described (6). Other DNAs used in this study, as specified below, were obtained from Michael Dojka (5). These included DNAs isolated from two marine sediments (CA), two terrestrial sediments (IN), two human fecal samples, two hydrocarbon-damaged aquifer samples (CA and MI), four thermal features from Yellowstone National Park, landscaped topsoil (CA), and Pavillion Lake (Canada).

PCR and cloning. Community ribosomal DNA sequences (rDNAs) were amplified in 100-μl PCRs, which included the following (final concentrations): 1 to 50 ng of DNA in reaction mixtures, 1× PCR buffer II (Perkin-Elmer), 1.5 mM $MgCl_2$, 200 μM (each) deoxynucleoside triphosphates, 200 μM (each) forward and reverse primers, 800 μg of bovine serum albumin per ml, and 2.5 U of AmpliTaq Gold DNA polymerase (Perkin-Elmer) per ml. Reaction mixtures were incubated in a PTC-100 programmable thermal cycler (MJ Research) at 94°C for 12 min (for initial denaturation and activation of AmpliTaq Gold),

* Corresponding author. Mailing address: Department of Molecular, Cellular and Developmental Biology, University of Colorado, Campus Box 347, Boulder, CO 80309-0347. Phone: (303) 735-1864. Fax: (303) 492-7744. E-mail: nrpace@colorado.edu.

† Present address: San Diego State University, San Diego, CA 92182.

TABLE 1. Summary of clone libraries and phylogenetic groups detected

Source of DNA	Library name	Primer	PCR annealing temp (°C)	Target sequences detected[c]
Marine sediment (California)	BMS[b]	OP11-1090R/27F	52	Y
Marine sediment (California)	BoI[b]	OP11-1090R/27F	55	Y
Marine microbial mat (California)	BTM[b]	OP11-1090R/27F	52	Y
Landscaped topsoil (California)	DA	OP11-1090R/27F	52	Y
Landscaped topsoil (California)	DB	11_1_502F/1525R	60	Y
Sulphur river (Kentucky)	SRA	11_3_1368R/27F	60	Y
Sulphur river (Kentucky)	SRB	11_5_1380R/27F	60	Y
Sulphur river (Kentucky)	SRC	OP11-1090R/27F	52	Y
Sulphur river (Kentucky)	SRD	11_2_1236R/27F	56	Y
Contaminated aquifer (Michigan)	WSA	11_2_1236R/27F	56	Y
Marine sediment (California)	BMSA	11_3_1368R/27F	60	N
Marine sediment (California)	BMSB	11_1_502F/1525R	60	N
Marine sediment (California)	BolA	11_3_1368R/27F	60	N
Marine sediment (California)	BolB	11_1_502F/1525R	60	N
Fairfax swamp (Indiana)	FSA	11_3_1368R/27F	60	N
Oxidized iron deposits (Colorado)	GalA	11_1_502F/1525R	60	N
Gallionella Mat (Colorado)	GalB	11_3_1368R/27F	60	N
Lake Lemon sediments (Indiana)	LLB	11_3_1368R/27F	60	N
Pool N10, Norris Geyser Basin (YNP)[a]	N10A	11_1_502F/1525R	60	N
Green Finger Pool, (YNP)	O1aA	11_3_1368R/27F	60	N
Green Finger Pool, (YNP)	O1aB	OP11-1090R/27F	52	N
Obsidian Pool (YNP)	OBPB	11_3_1368R/27F	60	N
Pavillion Lake (British Columbia, Canada)	PLTA	11_3_1368R/27F	60	N
Quenn's Laundry (YNP)	QL2A	11_3_1368R/27F	60	N
Quenn's Laundry (YNP)	QL2B	OP11-1090R/27F	52	N
Sulphur river (Kentucky)	SRE	11_1_502F/1525R	60	N
Contaminated aquifer (Michigan)	WSB	11_1_502F/1525R	60	N

[a] YNP, Yellowstone National Park.
[b] Hemi-nested PCR.
[c] Y, yes; N, no.

followed by 30 cycles at 92°C for 1 min, various annealing temperatures (Table 1) for 1 min, and 72°C for 1 min, followed by a final extension period of 15 min at 72°C. The primers used in this study were 27F (bacterium specific, 5′-AGAGTTTGATCCTGGCTCAG-3′), 1492R (universal, 5′-GGTTACCTTGTTACGACTT-3′), 1525R (bacterium specific, 5′-AAGGAGGTGATCCAGCC-3′) (14), 11_1_502F (5′-GCCGCTAACTGCGTGCCA-3′), 11_2_1286R (5′-CCCACTGAGAAGCCGTTT-3′), 11_3_1368R (5′-GCGAGAACGTATTCACCG-3′), 11_5_1380R (5′-TGAGTGCAAGGAACAGGG-3′), and OP11-1090R (5′-TCGTTGTCCCACTTAA-3′). For libraries Bol, BMS, and BTM, a hemi-nested amplification was employed that consisted of a first-round PCR amplification using 27F and 1492R primers, with the product of first amplification serving as the template for a second PCR using 27F and OP11-1090R primers. PCR products were cloned using the TOPO TA cloning kit (Invitrogen Corp., Carlsbad, Calif.) as specified by the manufacturer. Plasmid DNAs containing unique inserts were sorted by restriction fragment length polymorphism (RFLP) analysis (6) and sequenced using an ABI 373 stretch or ABI 377 DNA sequencer (BigDye Terminator ready reaction mixture; PE Applied Biosystems, Inc.) according to the manufacturer's instructions. Primers used for sequencing included vector primers (T3 and T7) and the small-subunit (SSU) rDNA primers 515F (5′-GTGCCAGCMGCCGCGGTAA-3′), 907R (5′-CCGTCAATTCCTTTRAGTTT-3′) and DASEQ (5′-SGGAGWATCGACCCT-3′).

For each library, eight randomly chosen RFLP types were screened by sequence to test primer specificity. All RFLP types in libraries with target sequences among the initially screened clones were sequenced. Libraries with no target rRNA sequences among the screened samples were not analyzed further.

Phylogenetic analysis and chimera detection. Sequences were compared to the GenBank database using the Basic Local Alignment Search Tool (BLAST) network service to determine their approximate phylogenetic relationships (1). Partial sequences were edited and compiled in AutoAssembler 2.1 (PE Applied Biosystems), and the sequences were aligned using the ARB database (Strunk and Ludwig [20]). We used the methods described by Hugenholtz et al. to identify chimeras (12), and none were identified. Sequence alignments used to infer phylogenetic relationships were created using the Lane mask (15), which excludes hypervariable regions of the SSU-rRNA sequence. Phylogenetic relationships were tested using reference taxa from previously described phylogenetic divisions (11). Heuristic searches were performed using the minimal evolution (ME) (number of repetitions [nreps] = 100 or 200), maximum parsimony (MP) (nreps = 100 or 200), and maximum likelihood (ML) (nreps = 1 or 10) methods to find the best tree possible. For trees constructed using ME and MP methods, data sets contained 181 or 245 taxa. In order to test the robustness of the inferred topologies in phylogenetic trees, Bootstrap analyses were performed using ME (with K2P, HKY85, GTR, and ML models of evolution [10]) and MP (100 bootstrap replicates with 10 random addition sequence replicates per bootstrap) in the PAUP* phylogenetic package (version 4.0) compiled on a dual-processor PC running Linux. Seventy percent or greater bootstrap support was considered to identify supported nodes (9). Based on the ME and MP analyses, a core of 15 sequences (5 each) from candidate divisions OP11, OD1, and SR1 was selected to be included in all subsequent ML analyses. For trees constructed using ML methods, multiple data sets (7, with 38 to 64 taxa each) were constructed to sample all of the reference taxa. ML analyses were used to test the monophyly of clades that resulted from the analyses and to identify potential problems of long-branch attraction that arise with distance and MP methods. All ML analyses used the HKY85 model of evolution due to computational constraints of the analyses.

About 240 complete or partial sequences not belonging to the target phylogenetic divisions were determined in this study. Phylogenetic information for these clones can be obtained from http://pacelab.colorado.edu/publications/publications.html.

Accession numbers. The sequences of all the rDNA clones analyzed are available under GenBank accession no. AY192989 to AY193296.

RESULTS

SSU-rDNA libraries were developed using primers targeted to each of the five subdivisions of the previously described phylogenetic division OP11. As summarized in Table 1, 6 of the 17 community DNAs tested yielded target sequences. Three environments (two human fecal samples and one hydrocarbon-damaged aquifer sample) produced no PCR product with any of the primer sets tested. In general, the environments

that contained OP11-related sequences were anoxic: marine sediments, a contaminated aquifer, and the Sulphur River streamers (Table 1). The qualitatively more oxic environments examined did not yield OP11-related sequences. This result is consistent with previous encounters of representatives of the OP11 division, mainly in anoxic zones.

Approximately 2,600 rDNA clones from 27 libraries were screened by RFLP, and about 300 unique sequences were determined completely or partially. The primers used for this study were designed to be specific for their target groups based on known sequences, and stringent conditions (maximum annealing temperature that produced the product) were used for PCR to maximize specificity of the amplified product. Nonetheless, libraries constructed from these primers contained numerous nontarget rRNA gene sequences. In general, nontarget sequences showed low sequence identity (typically 90% or less) to known 16S rRNA sequences (see Materials and Methods for accession numbers).

OP11-related rRNA sequences were used in extensive phylogenetic analyses to determine the relationships of the new sequences to known ones. Multiple large data sets (181 or 245 taxa) representative of the full diversity of the bacterial domain were analyzed using ME and MP analyses to identify preliminary phylogenetic placements. Bootstrap analyses were used to test the robustness of the trees determined. The results indicate that the OP11 clade, previously thought to be monophyletic, in fact is a polyphyletic group. Inclusion of the new sequences in tree calculations broke the former OP11 clade into three division-level groups, without specific relationship to any other bacterial division. We term these groups candidate divisions OP11 (three of the five former clades of the OP11 group), OD1, and SR1. Based on the initial results, multiple small data sets (38 to 64 taxa) covering the diversity of *Bacteria* were assembled for heuristic and/or bootstrap ML analyses (nreps = 10 or 100, respectively) to test for potential artifacts in the ME and MP analyses. Specifically, we tested whether the monophyly of the OP11, OD1, and SR1 clades resulted from long-branch attraction (LBA).

Table 2 summarizes the phylogenetic analyses performed to test the relationships of sequences belonging to the candidate divisions OP11, OD1, and SR1 in the context of diverse sequences from the domain *Bacteria*. The monophyletic nature of each of the candidate divisions (OP11, OD1, and SR1), as well as the relationship between these candidate divisions, was tested in the context of the original articulation of the candidate division OP11 (12). The results show that each of the candidate divisions discussed here is monophyletic in all analyses, with good bootstrap support. Heuristic searches using ML methods did recover the original OP11 clade in several analyses (three of four) with approximately 40 taxa. Although the ML criterion is generally less prone to problems of LBA, we suspected that the limited sampling of 40-taxa data sets can lead to such problems even with ML. To test this possibility, four data sets, three containing approximately 60 taxa and one containing 181 taxa, were constructed and analyzed (Table 2). With better sampling, the aggregate OP11 clade, including OP11, OD1, and SR1, from the original articulation was not obtained in two of three 60-taxa data sets or the 181-taxa data set. These results indicate that LBA and limited sample size were responsible for the earlier result of the monophyly of the original OP11 group.

TABLE 2. Summary of phylogenetic analyses

Analysis[a]	Model	No. of taxa	nreps[b]	Monophyly/bootstrap support[c]			
				OP11	OD1	SR1	All[d]
Heuristic search							
ME	K2P	181	100	Y	Y	Y	N
	HKY85	181	100	Y	Y	Y	N
	GTR	181	100	Y	Y	Y	N
	ML	181	100	Y	Y	Y	N
	K2P	245	100	Y	Y	Y	Y
	HKY85	245	100	Y	Y	Y	Y
	GTR	245	100	Y	Y	Y	Y
	ML	245	100	Y	Y	Y	Y
MP		181	200	Y	Y	Y	N
		245	200	Y	Y	Y	Y
ML	HKY85	42	10	Y	Y	Y	Y
	HKY85	40	10	Y	Y	Y	N
	HKY85	40	10	Y	Y	Y	Y
	HKY85	38	10	Y	Y	Y	Y
	HKY85	61	10	Y	Y	Y	N
	HKY85	61	10	Y	Y	Y	Y
	HKY85	64	10	Y	Y	Y	N
	HKY85	181	1	Y	Y	Y	N
Bootstrap							
ME	K2P	181	100	95	75	100	U
	GTR	181	100	92	73	100	U
	HKY85	181	100	94	78	100	U
	ML	181	100	96	73	100	U
	K2P	245	100	94	79	100	U
	GTR	245	100	89	80	100	U
	HKY85	245	100	84	86	100	U
	ML	245	100	95	90	100	U
MP		181	100	81	U	100	U
		245	100	86	89	100	U
ML	HKY85	44	100	91	81	100	U
	HKY85	40	100	91	98	100	U
	HKY85	40	100	95	85	100	U
	HKY85	38	100	94	89	100	U

[a] For ML analyses, seven smaller independent data sets were used to test the monophyly of the target groups. For ME and MP all data sets were the same for heuristic searches and bootstrap analyses.

[b] Numbers denote the numbers of repititions (nreps) performed in heuristic searches and bootstrap analyses.

[c] Monophyly of the groups of interest in heuristic searches are denoted as yes (Y) or no (N) and if the clade is supported or unsupported (U) by bootstrap analysis. Bootstrap support (percent) for nodes that demark the monophyly of the indicated clade is shown. Seventy percent was used as a cutoff for convincing bootstrap support (10).

[d] The "All" category denotes monophyly of OP11, OD1, and SR1 as a single clade, as seen in the original articulation of OP11.

DISCUSSION

These results clarify the phylogenetic tree of the former candidate division OP11. The expanded sampling of diversity afforded finer phylogenetic resolution than previously possible and split the former candidate division OP11 into three separate division-level clades. The new candidate phylogenetic divisions SR1 (Sulphur River 1) and OD1 (OP11-derived 1) formerly were subdivisions 4 and 5, respectively, of the candidate division OP11. The extent of diversity, as indicated by pairwise rRNA sequence differences (5) in each of these clades, is still very high. The revised candidate division OP11

remains the most diverse of the known bacterial divisions, with 29.8% overall sequence difference within the clade. The new candidate phylogenetic division OD1 is the second most diverse at 27.2%, while the phylogenetic division SR1 (both described in this work) is represented so far by sequences with only 15.7% overall divergence. The extent of SSU-rRNA sequence divergence in these clades is comparable to those of well-established bacterial divisions with cultivated representatives, such as the *Proteobacteria* (with 23% sequence divergence excluding mitochondria), the low-G+C gram-positive bacteria (24%), and *Cyanobacteria* (13% excluding chloroplasts).

The split of the former OP11 clade into multiple division-level clades is not surprising considering the extent of sequence divergence previously noted for the group, coupled with the small number of sequences (limited sampling) that represented the original OP11 group. Problems in phylogenetic results due to incomplete sampling are well documented (7, 8, 19). The subdivisions of the former OP11 clade (12) were comprised of only a few sequences, yet the extent of sequence variation between the subdivisions was as great as that between any of the primary bacterial divisions. The former subdivisions 1 to 3, comprised of 24 sequences in the original study (12), are now represented by 49 diverse sequences and remain the candidate phylogenetic division OP11. The smallest subdivision of the former OP11 division (number 4) was indicated by only 2 sequences in the original study (12), but that representation is expanded to 11 sequences in this work. The phylogenetic analyses indicate that this former subdivision constitutes a novel division-level clade, candidate division SR1. Subdivision 5 of the former OP11 division contained 10 sequences in the original study (12) and is now comprised of 53 sequences. This larger data set resolves into still another novel candidate phylogenetic division, OD1.

The split of the former phylogenetic division OP11 into multiple division-level clades highlights both the difficulty of resolving the relationships among the bacterial divisions and the importance of adequate sampling for the identification of relatedness groups. Any pair of rRNA sequences from a particular phylogenetic division of *Bacteria* can be as much as 30% different. This level of sequence divergence is a primary reason why adequate sampling is critical: greater sampling of deeply divergent groups breaks up long branches in phylogenetic tree calculations and increases the accuracy of phylogenetic inferences (7, 8, 19).

The new candidate phylogenetic division SR1 provides a good example of the sampling problem. The original phylogenetic analysis of the OP11 division found the two available SR1 sequences to be a strongly supported subgroup of OP11. With increased sampling, the present analyses clearly distinguish SR1 sequences as forming a distinct and strongly supported (100% bootstrap in all analyses) monophyletic clade distinct from all other bacterial sequences (Fig. 1; Table 2). The original clustering of SR1 sequences with those of the OP11 group apparently was a consequence of LBA and insufficient data to resolve the deep branchings even using robust phylogenetic methods, such as ML. Indeed, our results bring into question the practice of using subsets of the data in ML analyses. Although ML is technically more robust for problems of LBA, we found that LBA can be problematic with ML when sample diversity is limited. Dalevi et al. found similar behavior of ML when studying the effects of different outgroups on ML analyses of bacterial divisions (4).

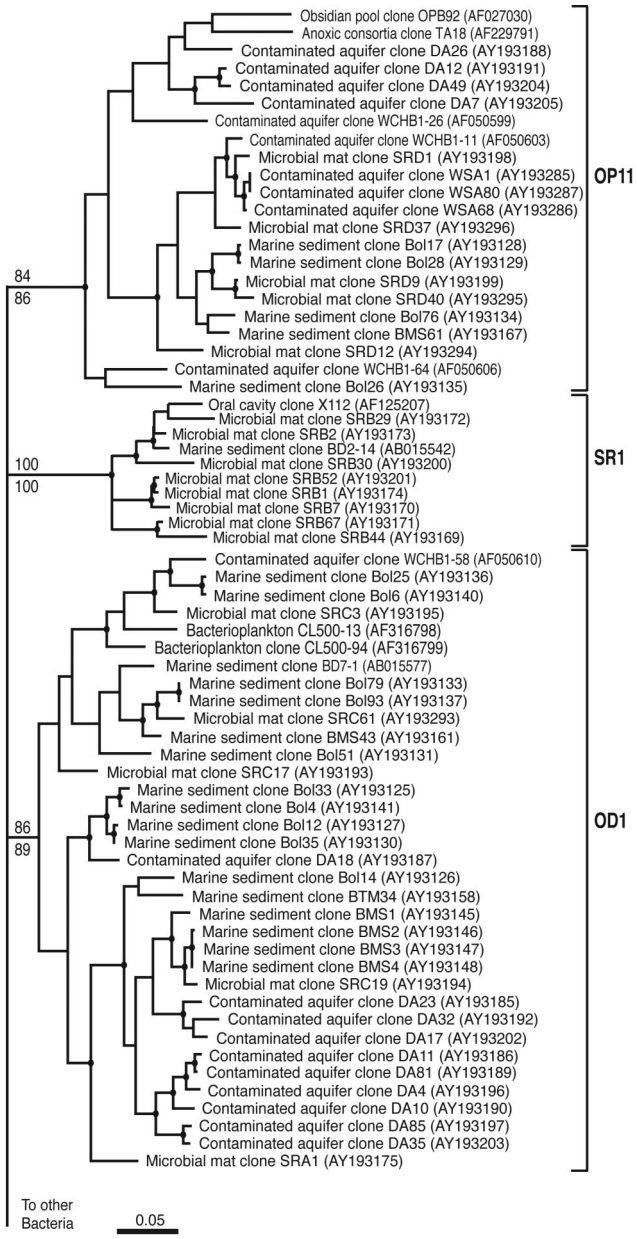

FIG. 1. Bootstrap consensus tree showing the well-supported phylogenetic relationships for the bacterial divisions OP11, OD1, and SR1, with division names listed outside the brackets. Nodes within the tree that are supported (bootstrap values of >70%) (9) in all analyses (distance, MP, and ML [see Results]) are indicated by filled circles. Nodes without circles were not strongly supported (bootstrap of <70%) in all analyses. Bootstrap support for the division-level divergences (Table 2) are given for ME (top, 245-taxa data set, HKY85 model of evolution) and MP (bottom, 245-taxa data set). The bar represents 5% sequence divergence.

The PCR primers designed for this study were nominally specific for the targeted groups based on available bacterial rRNA sequences and screens for specificity with BLAST. Nonetheless, numerous nontarget sequences were obtained with the subdivision 1 and 3 primers, even under highly strin-

gent conditions (reduced magnesium concentrations and elevated annealing temperatures). The number of environmental DNAs that were PCR positive was a good indication of the relative specificities of the primers. The most-specific primers amplified from only a small number of environmental DNAs, whereas primers that exhibited amplification from a large number of the environmental DNAs showed lower levels of specificity. The lower specificities of some of these primers is not necessarily a negative characteristic, since novel sequences were obtained. Since the sequences were obtained by PCR, we do not know if amplification of nontarget sequences was due to low fidelity of primer binding or the actual occurrence of the target sequence in the nontarget environmental rDNAs. Most of the nontarget sequences were highly divergent (>10%) from the target group and from known bacterial rRNA sequences. The relatively low phylogenetic specificities of the primers may be a consequence of the limited diversity of rRNA sequences that are in the databases and are used to design specific primers. The relaxed specificities of the primers used in this study also point out a potential pitfall for methods that rely on nucleic acid hybridization assays based on sequences in the public databases. Probes may appear highly specific in the context of database sequences but actually hybridize to target sequences from nucleic acids of phylogenetic groups that remain uncharacterized.

The new candidate phylogenetic divisions reported here have no cultivated members, so their physiological properties remain unknown beyond the expectations of the general properties of organisms that belong to the phylogenetic domain *Bacteria*. However, the environmental distribution of the sequences suggests hypotheses about the environmental roles of the organisms identified by the sequences. In general, representatives of the OP11, OD1, and SR1 groups are found in reduced environments with a conspicuous presence of sulfur compounds, such as sulfate and sulfide in the marine environments, the contaminated aquifer sample, and the Sulphur River sample. This may indicate that the organisms detected engage in sulfur cycling in the environment. The sequences determined in this study are specific identifiers of the corresponding organisms which can be used in further studies.

ACKNOWLEDGMENTS

This work was supported by grants from the National Science Foundation and the NASA Astrobiology Institute.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Borneman, J., and E. W. Triplett.** 1997. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. Appl. Environ. Microbiol. **63:**2647–2653.
3. **Cole, J. R., B. Chai, T. L. Marsh, R. J. Farris, Q. Wang, S. A. Kulam, S. Chandra, D. M. McGarrell, T. M. Schmidt, G. M. Garrity, and J. M. Tiedje.** 2003. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res. **31:**442–443.
4. **Dalevi, D., P. Hugenholtz, and L. Blackall.** 2001. A multiple-outgroup approach to resolving division-level phylogenetic relationships using 16S rDNA data. Int. J. Syst. Evol. Microbiol. **51:**385–391.
5. **Dojka, M. A., J. K. Harris, and N. R. Pace.** 2000. Expanding the known diversity and environmental distribution of an uncultured phylogenetic division of bacteria. Appl. Environ. Microbiol. **66:**1617–1621.
6. **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.
7. **Graybeal, A.** 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. **47:**9–17.
8. **Hillis, D. M.** 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. **47:**3–8.
9. **Hillis, D. M., J. J. Bull, M. E. White, M. R. Badgett, and I. J. Molineux.** 1992. Experimental phylogenetics: generation of a known phylogeny. Science **255:**589–592.
10. **Hillis, D. M., and C. Moritz.** 1990. Molecular systematics. Sinauer Associates, Inc., Sunderland, Mass.
11. **Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. **180:**4765–4774. (Erratum, **180:**6793.)
12. **Hugenholtz, P., C. Pitulle, K. L. Hershberger, and N. R. Pace.** 1998. Novel division level bacterial diversity in a Yellowstone hot spring. J. Bacteriol. **180:**366–376.
13. **Kuske, C. R., S. M. Barns, and J. D. Busch.** 1997. Diverse uncultivated bacterial groups from soils of the arid southwestern United States that are present in many geographic regions. Appl. Environ. Microbiol. **63:**3614–3621.
14. **Lane, D. J.** 1991. 16S/23S rRNA sequencing, p. 115–175. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. John Wiley & Sons, New York, N.Y.
15. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. USA **82:**6955–6959.
16. **Li, L., C. Kato, and K. Horikoshi.** 1999. Bacterial diversity in deep-sea sediments from different depths. Biodivers. Conserv. **8:**659–677.
17. **Paster, B. J., S. K. Boches, J. L. Galvin, R. E. Ericson, C. N. Lau, V. A. Levanos, A. Sahasrabudhe, and F. E. Dewhirst.** 2001. Bacterial diversity in human subgingival plaque. J. Bacteriol. **183:**3770–3783.
18. **Pedersen, K., J. Arlinger, S. Ekendahl, and L. Hallbeck.** 1996. 16S rRNA gene diversity of attached and unattached bacteria in boreholes along the access tunnel to the Äspö hard rock laboratory, Sweden. FEMS Microbiol. Ecol. **19:**249–262.
19. **Poe, S.** 1998. Sensitivity of phylogeny estimation on taxonomic sampling. Syst. Biol. **47:**18–31.
20. **Strunk, O., and W. Ludwig.** 1999. ARB: a software environment for sequence data. [Online.] http://www.mikro.biologie.tu-muenchen.de.
21. **Ueda, T., Y. Suga, and T. Matsuguchi.** 1995. Molecular phylogenetic analysis of a soil microbial community in a soybean field. Eur. J. Soil Sci. **46:**415–421.
22. **Wise, M. G., J. V. McArthur, and L. J. Shimkets.** 1997. Bacterial diversity of a Carolina bay as determined by 16S rRNA gene analysis: confirmation of novel taxa. Appl. Environ. Microbiol. **63:**1505–1514.