

Empirical Testing of 16S rRNA Gene PCR Primer Pairs Reveals Variance in Target Specificity and Efficacy Not Suggested by In Silico Analysis[∇]

Sergio E. Morales¹ and William E. Holben^{1,2*}

Microbial Ecology Program, Division of Biological Sciences,¹ and Montana—Ecology of Infectious Diseases Program,²
The University of Montana, Missoula, Montana

Received 18 September 2008/Accepted 20 February 2009

Phylogenetic and “fingerprinting” analyses of the 16S rRNA genes of prokaryotes have been a mainstay of microbial ecology during the last two decades. However, many methods and results from studies that rely on the 16S rRNA gene for detection and quantification of specific microbial taxa have seemingly received only cursory or even no validation. To directly examine the efficacy and specificity of 16S rRNA gene-based primers for phylum-, class-, and operational taxonomic unit-specific target amplification in quantitative PCR, we created a collection of primers based solely on an extensive soil bacterial 16S rRNA gene clone library containing ~5,000 sequences from a single soil sample (i.e., a closed site-specific library was used to create PCR primers for use at this site). These primers were initially tested in silico prior to empirical testing by PCR amplification of known target sequences and of controls based on disparate phylogenetic groups. Although all primers were highly specific according to the in silico analysis, the empirical analyses clearly exhibited a high degree of nonspecificity for many of the phyla or classes, while other primers proved to be highly specific. These findings suggest that significant care must be taken when interpreting studies whose results were obtained with target specific primers that were not adequately validated, especially where population densities or dynamics have been inferred from the data. Further, we suggest that the reliability of quantification of specific target abundance using 16S rRNA-based quantitative PCR is case specific and must be determined through rigorous empirical testing rather than solely in silico.

The use of 16S rRNA gene-based primers for the detection, description, and enumeration of bacterial targets has figured very prominently into modern microbial ecology. Microbial community-level studies based on the 16S rRNA gene are assumed to fall victim to a variety of so-called “PCR biases” (25). These include, but may not be limited to, differential DNA extraction efficiencies, idiosyncrasies of the amplification method (e.g., the specific thermostable polymerase used, other reaction components, or cycling conditions) (22), inhibition of PCR amplification by contaminants, sequence-based differences in PCR amplification efficiency (20), PCR artifacts (e.g., chimeras, heteroduplexes, or point mutations) (1, 2), or artifactual 16S rRNA sequence variation due to *rrn* operon heterogeneity (8). Indeed, such biases are likely exacerbated by the high degree of similarity in primer binding sites across taxa (especially where group-level specificity is desired) combined with potentially broad differences in GC content within the variable regions being amplified. Collectively, these biases are assumed to result in uneven amplification of the various phylotypes in a mixed template pool that is not representative of individual template abundance in total microbial community DNA. These considerations represent serious potential pitfalls when relying on PCR as a tool for community or population analysis. However, where conditions are standardized across samples, reasonable comparative, but likely not comprehensive, analyses can be performed (22).

Additional concerns regarding primer specificity and primer-template mismatch come into play where quantitative analyses or comprehensive surveys are desired. Because of the massive numbers but uneven representation of 16S rRNA gene sequences in databases (i.e., generally low representation of cultivated, well described, environmental organisms), the feasibility of comprehensive primer testing is limited. This has been partly dealt with by performing in silico testing using freely available software such as PRIMROSE (3, 7). Alternatively, researchers have simply tested specificity against a subset of potential targets (5, 6, 11, 15, 16, 19, 21, 23), which in most cases likely represent only a fraction of the biodiversity present in their respective study sites.

The purpose of this study was to directly and rigorously assess the validity and efficacy of using 16S rRNA-based primers for phylum-, class-, and operational taxonomic unit (OTU)-specific target amplification in support of bacterial population studies at our site. We utilized a large 16S rRNA gene clone library (~5,000 sequences) and the PRIMROSE program (3) to develop primers and to assess (in silico) the specificity of primers targeting multiple taxonomic levels across several phyla. Each potential primer set was then subjected to primer and reaction condition optimization for PCR and subsequently tested in vitro for specificity against genomic DNA and PCR-derived templates.

MATERIALS AND METHODS

Genomic DNAs. Genomic DNAs prepared from *Bradyrhizobium japonicum* USDA 110d, *Streptomyces griseus*, *Acidovorax facilis*, *Pseudomonas putida*, *Acidobacterium capsulatum*, and *Pedobacter heparinus* using the method of Doi (9) were used in positive control reactions for primer testing.

* Corresponding author. Mailing address: Microbial Ecology Program, Division of Biological Sciences, The University of Montana, Missoula, MT 59812-1006. Phone: (406) 243-6163. Fax: (406) 243-4184. E-mail: bill.holben@mso.umt.edu.

[∇] Published ahead of print on 27 February 2009.

TABLE 1. Primer targets and sequence representation within libraries

Taxonomic level and taxon	No. of sequences ^a	
	KBS	ARB
Phylum		
<i>Acidobacteria</i>	955	516
<i>Actinobacteria</i>	491	5,360
<i>Bacteroidetes</i>	453	2,683
CD OD1	68	73
CD OP10	43	55
<i>Chlorobi</i>	22	142
<i>Chloroflexi</i>	27	66
<i>Gemmatimonadetes</i>	251	111
<i>Nitrospira</i>	59	169
<i>Planctomycetes</i>	243	699
<i>Proteobacteria</i>	1,690	20,166
<i>Thermomicrobia</i>	323	361
<i>Verrucomicrobia</i>	103	159
Class		
<i>Alphaproteobacteria</i>	374	5,065
<i>Betaproteobacteria</i>	485	4,527
<i>Deltaproteobacteria</i>	348	1,217
<i>Gammaproteobacteria</i>	478	8,479
OTU^b		
Genus <i>Aeromonas</i>	99	
<i>Acidobacteria</i> group 4	81	
Genus <i>Lysobacter</i>	81	
<i>Thermomicrobia</i> no. 4	63	
<i>Nitrosomonadales</i>	62	
<i>Acidobacteria</i> group 6	61	
<i>Thermomicrobia</i> no. 7	46	
Genus <i>Bradyrhizobium</i>	39	
Genus <i>Pseudomonas</i>	38	
Genus <i>Comamonas</i>	38	

^a Total number of sequences found in the database.^b Closest classification based on $\geq 97\%$ sequence similarity at the 16S rRNA gene.

16S rRNA gene sequence library. A 16S rRNA gene sequence library containing 4,889 sequences that was generated as described previously (17) from soil at the Kellogg Biological Station Long Term Ecological Research site (KBS-LTER) by targeting hypervariable regions V4 and V5 of the 16S rRNA gene

(GenBank accession no. EU352912 to EU357802) was used for primer development and in silico testing as described below. An additional 16S rRNA gene sequence library containing >50,000 partial, aligned, and annotated sequences was downloaded from the ARB website (<http://www.arb-home.de/>) (14) for in silico testing of primer specificity.

Primer design and in silico testing. The PRIMROSE software (3) was used to design 16S rRNA gene primer sets for several major bacterial groups identified in the KBS-specific library based on taxonomic assignments developed using ARB (Table 1) (17). Each primer pair was comprised of a phylum-, class-, or OTU-specific forward primer targeting the V4 and V5 hypervariable regions and the generally conserved reverse primer 907r (5'-CCGTC AATTCMTTTRAGT TT-3') (13). Primer sets were generated for predominant phyla, as well as for some of the most abundant individual phylotypes or OTUs that were found at the site, based on $\geq 97\%$ sequence similarity. All primer sets were designed using the default settings of the PRIMROSE program.

Primer sets were tested in silico against all target and nontarget sequences in the KBS library using PRIMROSE. Primers were judged as acceptable by the following criteria: (i) most ($\geq 85\%$) of sequences contained within its target taxon (within the KBS library) were indicated as detected; (ii) concomitantly, a low proportion ($\leq 17\%$) of nontarget phylotypes were indicated as detected; and (iii) the least number of degeneracies within the primer was required to achieve the first two criteria. Primers that qualified under these criteria were then tested in silico using PRIMROSE against target and nontarget taxa from the ARB-generated library containing over 50,000 16S rRNA gene sequences, including archaeal and eukaryotic sequences not specific to the KBS site.

Primer optimization and in vitro testing. Genomic DNA preparations from *Bradyrhizobium japonicum* USDA 110d, *Streptomyces griseus*, *Acidovorax facilis*, *Pseudomonas putida*, *Acidobacterium capsulatum*, and *Pedobacter heparinus* were initially used to assess and optimize amplification from each respective set of phylum- and class-level primers. These isolates are not from the KBS site but are known through sequence analysis to have 100% complementarity to their respective primers. Temperature gradient ($\pm 10^\circ\text{C}$) PCRs were run for each phylum- or class-specific primer (Tables 2 and 3), using primer 907r as the second primer in the pair and the predicted melting temperature (T_m) for the specific primer as the central point in the temperature gradient. Reactions were performed in a final volume of 50 μl , and each reaction mixture contained 1 \times PCR buffer (containing 1.5 mM MgCl_2), 200 μM deoxynucleoside triphosphates, 10 pmol of each primer, 20 μg bovine serum albumin, 1 U HotStar Taq polymerase (Qiagen, Valencia, CA), and 5 ng of genomic DNA as the template. Cycling conditions included an initial denaturation step of 15 min at 95°C ; 30 cycles of amplification consisting of denaturation (25 s at 96°C), primer annealing (30 s at the requisite T_m), and primer extension (30 s at 72°C); and a final extension step of 1.5 min at 72°C for all primers tested.

In order to generate positive controls for phylogenetic groups with no cultured representative available, plasmid DNA encoding the desired target sequence

TABLE 2. Phylum- and class-level primer sequences and target specificities as tested by PRIMROSE

Primer	Target	Sequence	Degenerate bases	% Specificity					
				KBS		ARB			
				Target ^a	Nontarget ^b	Target	Prokaryotes	Archaea	Eucarya
688-706fAB	<i>Acidobacteria</i>	GCGGTGAAATGCGTASAT	1	95.29	11.42	83.10	19.40	ND ^c	ND
691-709fAP	<i>Alphaproteobacteria</i>	GTGAAATDCGTAGAKATT	2	95.19	3.76	87.20	8.70	ND	ND
697-715fAT	<i>Actinobacteria</i>	TGCGCAGAKATCRGGARG	3	94.30	0.09	91.00	0.10	ND	ND
715-733fBP	<i>Betaproteobacteria</i>	AAYACCRATGCGAAGGC	2	96.70	1.83	92.60	7.60	ND	ND
685-703BT	<i>Bacteroidetes</i>	GTAGCGGTGAAATGCWTA	1	96.25	0.02	90.50	0.20	ND	ND
554-572fCH	<i>Chlorobi</i>	TCCGGAWTYACTGGGTRT	3	95.45	0.08	93.00	3.00	ND	ND
687-705fCX	<i>Chloroflexi</i>	AGTGGTGAAATGCGTWGA	1	96.30	4.19	45.50	0.40	ND	ND
542-560fDP	<i>Deltaproteobacteria</i>	GTGCNARCGTTGYTCGGA	3	86.21	2.59	55.90	0.60	ND	ND
680-698fGP	<i>Gammaproteobacteria</i>	CMKGTGTAGCRGTGAAAT	3	93.31	16.85	78.20	10.80	ND	ND
677-695fGT	<i>Gemmatimonadetes</i>	TTCCSGGTGTAGCGGTGG	1	97.21	ND	73.90	0.40	ND	ND
706-724fNP	<i>Nitrospira</i>	ATCGGGAGGAASRCKGT	3	93.22	3.94	43.80	0.70	ND	ND
785-802fOD1	OD1	GGATTAGATACYCYWGT	3	85.29	0.70	63.00	4.10	6.60	ND
889-907fOP10	OP10	ASTACGGCCGCAAGGTTG	1	83.72	6.22	78.20	5.60	0.10	ND
681-699fPM	<i>Planctomycetes</i>	NRGTGRAGCGGTGAAATG	3	94.24	0.62	83.80	0.60	ND	ND
767-785fProt	<i>Proteobacteria</i>	AAGCGTGCGGAGCAAACA	0	89.08	7.55	73.30	23.70	ND	ND
555-573fTM	<i>Thermomicrobia</i>	CCGGAKTAYTGGGCGTA	3	91.02	8.30	74.50	8.20	ND	ND
562-580fYM	<i>Verrucomicrobia</i>	YAYTGGGCGTAAAGGGWG	3	92.23	4.38	80.50	7.50	ND	ND
853-871fWS3	WS3	GTGCCGAGCYAASSCAT	3	94.12	5.42	67.90	5.50	ND	ND

^a Target sequences are those belonging to the phylogenetic group intended for amplification.^b Nontarget sequences are sequences not belonging to the phylogenetic group intended for amplification.^c ND, not detected.

TABLE 3. Primer sequences and target specificities to the KBS library as tested by PRIMROSE for the top 10 most abundant OTUs ($\geq 97\%$ sequence similarity)

Primer	Target	Sequence	Degenerate bases	% Specificity	
				Target ^a	Nontarget ^b
Coma851-869f	Genus <i>Comamonas</i>	YCAGTRMCGAAGCTAACG	3	97.37	4.99
Pseudo573-591f	Genus <i>Pseudomonas</i>	AAGSGCKCGTAGGYGGTT	3	94.74	2.47
Aero851-869f	Genus <i>Aeromonas</i>	GGSTKCCGGMGCTAACGC	3	86.87	1.94
Acido (#4)599-617f	<i>Acidobacteria</i> group 4	CGAYTGTGAAATCTCCGG	1	96.30	0.04
Lyso726-744f	Genus <i>Lysobacter</i>	CGAAGGCGGYTSYCTGGA	3	92.59	1.83
Thermo (#4)735-753f	<i>Thermomicrobia</i> no. 4	CTTCCTGGCCTGTTCTTG	0	95.24	0.14
Nitro813-831f	<i>Nitrosomonadales</i>	TAAACGATGTCGACTGGT	0	95.16	0.10
Acido (#6)654-672f	<i>Acidobacteria</i> group 6	GAGDTYGGGAGAGGGATG	2	93.44	0.50
Thermo (#7)658-676f	<i>Thermomicrobia</i> no. 7	GCAGGAGAGGGTAGTGGA	0	95.65	0.08
Brady850-868f	Genus <i>Bradyrhizobium</i>	CTWGTGGCGMAGCTAACG	2	94.87	0.19

^a Target sequences are those belonging to the phylogenetic group intended for amplification.

^b Nontarget sequences are sequences not belonging to the phylogenetic group intended for amplification.

from our library was amplified using the primer pair 536f and 907r as outlined elsewhere (17). Since all specific primers generated in this study are internal to this region, it can be used as template to generate positive controls. To ensure comparable amplification rates, PCR products were used to test amplification efficiency and compared with reactions performed with both cloned and genomic DNA controls (i.e., genomic DNA from cultivable examples). PCR products were thereafter used for positive controls for all amplification reactions. The minimal DNA concentration necessary for effective amplification was determined using 10-fold serial dilutions down to 1 picogram of control DNA and the best empirically determined annealing temperature from prior temperature gradient analyses.

Each primer set was subsequently screened for specificity by testing against its relevant positive control DNA, while using the positive controls for all other primer sets as negative controls. These reactions were run using the optimal experimentally derived annealing temperature and DNA concentration for each primer pair. The first round of specificity screening was conducted by creating four separate negative control groups of DNA, each containing four negative control DNAs (four different nontarget sequences) in equimolar concentration. Candidate primers were first tested against these groups to identify nontarget amplification. Primers that showed positive amplification with more than one negative control group were not tested further. A second round of testing was done with individual templates found within the amplified negative control groups (see above) in order to identify which of the four nontarget DNAs in that group was responsible for nonspecific amplification.

Real-time PCR assays. Quantitative real-time PCR (qPCR) was performed on an iCycler iQ thermocycler (Bio-Rad, Hercules, CA) using 25- μ l reaction mixtures that included 1 μ l of template DNA at the concentration being tested (see Fig. 3), 10 pmol of each primer, 20 μ g bovine serum albumin, and 12.5 μ l of Absolute Blue QPCR Sybr green ROX mix (ABgene, Rochester, NY). Forty cycles of amplification were performed using the same cycling conditions used for primer testing [T_m of 55°C and 53°C for primer pairs 688-706fAB plus 907r and Acido (#6)654-672 plus 907r, respectively]. An additional melting curve analysis, where fluorescence was measured as the temperature increased from 50°C to 100°C, was also performed to test for the amplification of a single target. The efficiency of PCR amplification for both primer pairs was assessed using a standard curve prepared using the PCR product generated from a cloned 16S rRNA gene insert (clone 302-F22 [Table 4]) assigned to the *Acidobacteria* group 6 cluster as determined by sequence alignments performed in ARB (17, 18). Control reactions lacking template DNA were run to ensure that primer dimers were not contributing to the overall signal. Two independent rounds of triplicate reactions were performed for each target, and the results of at least three qPCRs were analyzed. Abundances for all replicate reactions were related to the standard curve by their respective fluorescence intensity values, giving values of relative concentration.

Checking primer specificity using clone libraries. To assess whether primer sets were amplifying only target taxa in total community DNA from the KBS site, clone libraries were made from the amplicons produced and subjected to DNA sequence analysis. To accomplish this, triplicate standard PCRs were performed separately as mentioned above for the primer pairs 688-706fAB plus 907r and Acido (#6)654-672 plus 907r using total community DNA from the KBS-LTER treatment 1 soil extracted using the Mobio PowerSoil DNA isolation kit (Carlsbad, CA). PCR products were purified using the Qiaquick PCR purification kit

(Qiagen, Valencia, CA) and cloned as previously described (17). The sequences obtained were assigned to phylogenetic groups using the RDP Classifier program (26) as previously described (17).

RESULTS

Primer design and in silico testing. Using the library of 4,889 16S rRNA gene sequences generated from KBS-LTER soil (17), we created 28 phylogenetic group-specific forward primers targeting the V4 and V5 hypervariable regions (Tables 2 and 3). Unique regions of sequence were identified and tested in silico using the PRIMROSE program. All primers consisted of 17 or 18 oligonucleotides, had at most three degenerate bases, and had an average GC content of 53.6% (standard deviation, 8.8%).

On average, all higher-order primers (targeting phylum- and class-level groups) (Table 2) were predicted to detect about 92.7% (standard error [SE], 1%) of their specific targets while conversely predicted to amplify 4.6% (SE, 1.1%) of nontarget sequences. In order to assess whether primers designed from our site-specific sequence library might have a reduced ability to detect sequences from other sites (indicating some level of site specificity), an ARB database with over 50,000 sequences (Table 1) was used to assess both specificity and target group detection capabilities in silico. The results indicated a reduced ability of the primers to detect their corresponding phylum- or class-specific target sequences in the ARB database, especially for groups with low sequence representation such as the *Chlorobi*, *Chloroflexi*, and *Gemmatimonadetes*. On average, primer sets were able to detect 75.3% (SE, 3.6%) of their intended phylum or class in the ARB database while being predicted to detect about 6% (SE, 1.6%) of nontarget prokaryotic sequences. No significant detection of archaeal or eukaryotic sequences was noted.

An additional 10 primer sets were designed for detection of the 10 most abundant OTUs (genus-level phylotypes based on $\geq 97\%$ sequence similarity) identified from our KBS-LTER data set (Table 3). On average, these primers detected 94.2% (SE, 0.9%) of their specific targets while detecting only 1.2% (SE, 0.5%) of nontarget sequences in the KBS database.

The effect of primer-target mismatch (theoretical number of mismatched base pairs allowed during annealing) on primer specificity was examined and shown to be highly significant in

TABLE 4. Positive control targets and classification based on ARB and Classifier

Primer	Clone no.	Best match ^a	Confidence threshold (%)	ARB ^b
688-706fAB	298.F6	<i>Acidobacterium</i>	100	<i>Acidobacteria</i> group 5
Acido (#4)599-617f	232.F14	<i>Spirochaetes</i>	24	<i>Acidobacteria</i> group 4
Acido (#6)654-672f	302.F22	<i>Firmicutes</i>	43	<i>Acidobacteria</i> group 6
Aero851-869f	204.F6	<i>Aeromonas</i>	100	<i>Aeromonas</i>
691-709fAP	620.F14	<i>Bradyrhizobium</i>	100	<i>Bradyrhizobiaceae</i>
697-715fAT	637.F4	<i>Streptomyces</i>	99	<i>Streptomyces griseus</i> et rel.
715-733fBP	026.F2	<i>Nitrospira</i>	34	Uncultured <i>Betaproteobacteria</i>
Brady850-868f	613.F4	<i>Bradyrhizobium</i>	100	<i>Bradyrhizobium</i>
685-703BT	093.F10	<i>Flavobacterium</i>	100	<i>Flavobacterium</i>
554-572fCH	814.F14	<i>Bacteroidetes</i>	41	Uncultured <i>Chlorobi</i>
Coma851-869f	131.F2	<i>Firmicutes</i>	62	Uncultured <i>Comamonas</i>
687-705fCX	275.F4	<i>Sphaerobacter</i>	41	<i>Sphaerobacter-Thermomicrobium</i>
542-560fDP	326.F14	<i>Smithella</i>	15	Uncultured <i>Deltaproteobacteria</i>
680-698fGP	560.F24	<i>Rhodanobacter</i>	82	<i>Rhodanobacter</i> et rel.
677-695fGT	345.F22	<i>Gemmatimonas</i>	96	<i>Gemmatimonas</i>
Lyso726-744f	106.F12	<i>Thermomonas</i>	61	Uncultured <i>Lysobacter</i> et rel.
Nitro813-831f	447.F14	<i>Alicyclophilus</i>	35	Uncultured <i>Nitrosomonadales</i>
706-724fNP	747.F14	<i>Nitrospira</i>	100	<i>Nitrospira moscoviensis</i>
785-802fOD1	048.F8	<i>Bacteroidetes</i>	37	OD1
889-907fOP10	476.F8	<i>Firmicutes</i>	67	OP10
681-699fPM	214.F20	<i>Deltaproteobacteria</i>	19	Uncultured <i>Planctomycetes</i>
Pseudo573-591f	577.F4	<i>Pseudomonas</i>	100	<i>Pseudomonas putida</i> et rel.
Thermo (#4)735-753f	072.F4	<i>Firmicutes</i>	62	<i>Thermomicrobia</i>
Thermo (#7)658-676f	390.F2	<i>Firmicutes</i>	57	<i>Thermomicrobia</i>
555-573fTM	100.F30	<i>Gammaproteobacteria</i>	23	Uncultured <i>Thermomicrobia</i>
562-580fVM	261.F14	<i>Verrucomicrobium</i>	57	Uncultured <i>Prostheobacter-Verrucomicrobium</i>
853-871fWS3	245.F30	<i>Firmicutes</i>	42	WS3

^a Best match based on classification using the RDP Classifier program.

^b Classification based on treeing using ARB and $\geq 50,000$ reference sequences.

increasing nonspecific target detection (Fig. 1 and 2). As anticipated, a single mismatch was sufficient to substantially increase detection of nontarget sequences for all primers tested (some data not shown). When tested against our site-specific library, all higher-order primers were predicted to exhibit a ninefold increase in detection of nonspecific targets on average when a single mismatch was allowed (Table 2). For the OTU-level primers, a sixfold increase in detection of nontarget sequences was predicted when a single mismatch was allowed (Table 3).

Primer optimization and in vitro testing. An initial comparison using genomic DNA preparations from pure cultures and PCR products from our clone library as positive controls was performed to compare the results obtained using these two different template types and to validate the use of PCR products where genomic DNA was not available (i.e., from uncultivable taxa). We note that specific genomic DNA samples were used as positive controls at the OTU, class, and/or phylum level. Genomic DNA controls and their respective primer set(s) (following each in parentheses and assuming 907r as reverse primer) were as follows: *Bradyrhizobium japonicum* USDA 110d (691-709AP and Brady850-868f), *Streptomyces griseus* (697-715fAT), *Acidovorax facilis* (715-733-fBP), *Pseudomonas putida* (680-698fGP, 767-785fProt, and Pseudo573-591f), *Acidobacterium capsulatum* (688-706fAB), and *Pedobacter heparinus* (685-703BT). PCR temperature gradients performed using genomic DNA as the template displayed the same trends as those using PCR products from cloned inserts selected as controls based on ARB and Classifier results

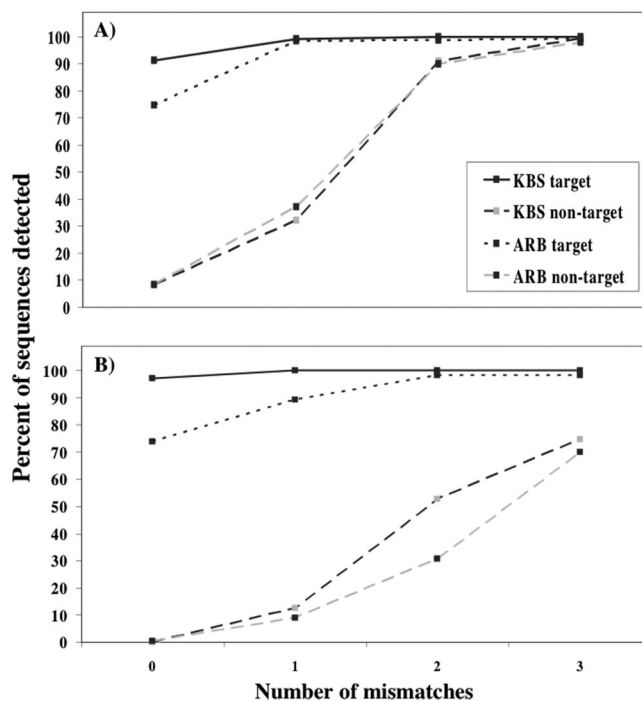


FIG. 1. Effect of mismatched bases on the recovery of target and nontarget sequences using phylum-level primers. Primers were tested in silico against 4,889 sequences from the KBS-LTER library and $>50,000$ sequences from the ARB database. (A) *Thermomicrobia*-specific primer 555-573fTM; (B) *Gemmatimonadetes*-specific primer 677-695fGT.

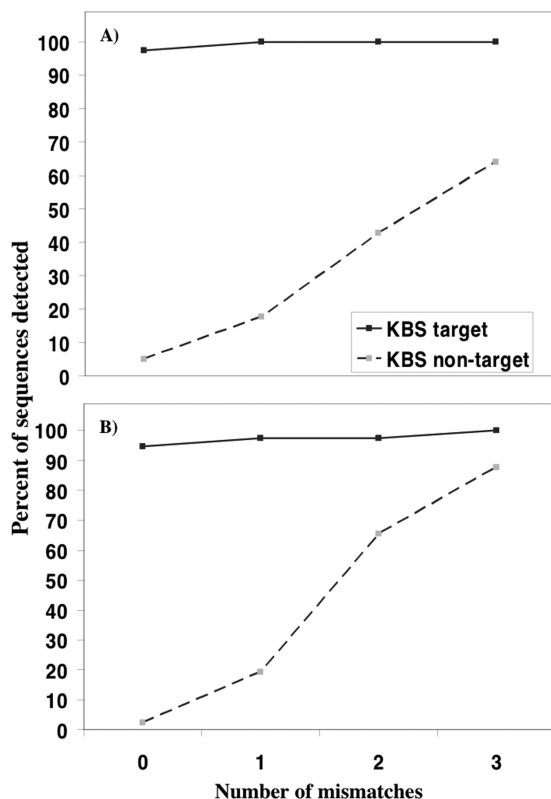


FIG. 2. Effect of mismatched bases on the recovery of target and nontarget sequences using OTU-level ($\geq 97\%$ sequence similarity) primers. Primers were tested in silico against 4,889 sequences from the KBS-LTER library. (A) OTU-specific primer Coma851-869f; (B) OTU-specific primer Pseudo573-591f.

(Table 4 and some data not shown). Readers are directed to reference 17 for a discussion of discrepancies between ARB and Classifier classification as observed for some taxa listed in Table 4. To normalize PCR conditions (excluding DNA sequence variation) and in order to be able to test primers from groups with few or no cultured representatives, PCR products generated from cloned 16S rRNA gene fragments were used as targets in reactions for all subsequent analyses. All positive controls were shown to be perfect matches for their respective primers, and all primers tested amplified their corresponding positive target controls (i.e., produced individual, coherent reaction products readily observed as visible bands in an agarose gel) from 1 pg of target.

Primer specificity was tested using a set of mixed pools of four nontarget DNA samples in PCRs. Of all 28 primers tested, only 5 primers (at their optimal T_m , which is given in parentheses below) were shown to have sufficient specificity to support their use in qPCR, namely, Acido (#4)599-617f (57°C), Acido (#6)654-672f (53°C), Thermo (#4)735-753f (52°C), Thermo (#7)658-676f (52°C), and Nitro813-831f (49°C) (Tables 2 and 3). Of these five, the two primer sets designed to target subgroups within the *Thermomicrobia* were not sufficiently specific to distinguish between these subgroups within the phylum when tested empirically but rather represented useful primer sets for the entire phylum. The other three primer sets exhibited no nonspecific amplification using

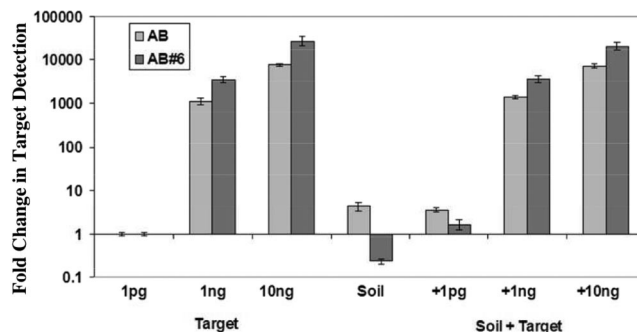


FIG. 3. Specific detection of *Acidobacterium* target DNA (*Acidobacteria* group 6 clone 302.F22 DNA) using the Acido (#6)654-672 plus 907r and 688-706fAB plus 907r primer sets. The given amounts of target DNA were tested alone or after addition to 9 ng of total community DNA isolated from KBS-LTER treatment 1, replicate plot 1. Values indicate the fold change in detection of the target group as a function of the amount of target added. Values for each primer set were normalized to 1 pg of specific target to show fold change in detection. Error bars are one SE of the mean for two rounds of triplicate qPCRs (final $n \geq 3$) Target, *Acidobacteria* group 6 clone 302.F22; soil, 9 ng of total community DNA extracted from soils at the KBS-LTER treatment 1, replicate plot 1.

any of the tested sequences and thus represent validated primers for further quantitative analysis. Additional optimization was attempted to increase the specificity of the remaining primer sets that had not met validation criteria thus far, but conditions that improved specificity invariably resulted in concurrent decrease in target amplification efficiency, thereby limiting their utility as specific qPCR primers for community DNA.

Target detection using real-time qPCR. The target detection and PCR efficiencies of two primer sets, one that passed validation [Acido (#6)654-672 plus 907r, specific for *Acidobacteria* group 6 targets] and one that did not (688-706fAB plus 907r, putatively specific for the phylum *Acidobacteria*), were compared using a real-time qPCR assay. Primers were initially validated for qPCR assays by creating a target standard curve (10^5 to 10 copies of the target). Once the primers were validated, an appropriate standard curve was used to assess and compare the abilities of the primers to detect changes in target concentration. A 10% difference in PCR efficiency was observed between the primer pairs, with Acido (#6)654-672 plus 907r exhibiting 103.05% efficiency while 688-706fAB plus 907r showed 92.9% efficiency. Although both primer sets were able to detect and quantify different target concentrations both alone and in the presence of total community DNA from soil (Fig. 3), the higher PCR efficiency of the primer pair containing Acido (#6)654-672 resulted in ~ 3 -fold more signal for a given target concentration compared to that with the primer pair containing 688-706fAB. Interestingly, despite the lower efficiency of amplification with the phylum-level primer set, quantification of target sequences from an unspiked soil sample from the KBS-LTER (treatment 1, replicate plot 1) indicated a higher abundance (4 pg, or $\sim 10^7$ copies) of phylum-level acidobacterial targets than of genus-level *Acidobacteria* group 6 targets (0.23pg, or 10^5 copies) per 10 ng of soil extracted DNA, indicating a low proportion of *Acidobacteria*

TABLE 5. Phylogenetic distribution of soil clones generated using primers 688-706fAB and Acido (#6)654-672

Phylum ^a	% (no.) of sequences identified with primer:	
	688-706fAB	Acido (#6)654-672
Target ^b	35 (27)	96 (76)
Nontarget	65 (50)	4 (3)
<i>Actinobacteria</i>	13 (10)	
<i>Bacteroidetes</i>	3 (2)	
<i>Proteobacteria</i>	9 (7)	
<i>Chloroflexi</i>	1 (1)	
Unclassified	39 (30)	4 (3)

^a As determined using RDP Classifier with an 80% confidence threshold.

^b The target for 688-706fAB is the phylum *Acidobacteria*, and that for Acido (#6)654-672 is *Acidobacteria* group 6.

group 6 within the total *Acidobacteria* phylum representatives present in that soil.

Recovery of target-specific sequences from soil by cloning.

To experimentally test the target specificities of a primer set that passed validation [Acido (#6)654-672 plus 907r] and one that did not (688-706fAB plus 907r), two independent clone libraries were created from the same soil sample used for the real-time PCR assay. Totals of 79 and 77 clones were analyzed, respectively, for these primer sets (Table 5). As expected from the results of the validation experiments, the phylum-specific *Acidobacteria* primer set (which had failed validation) recovered target-specific phylotypes but also resulted in recovery of a large number of nontarget phylotypes from soil community DNA (35% versus 65%, respectively). These nontarget sequences were distributed among four different phyla, with a large proportion (39%) being unclassified. In contrast, the *Acidobacteria* group 6-specific primer set Acido (#6)654-672 plus 907r (which passed validation) displayed an extremely high level of specificity, resulting in 96% of all sequences recovered from the soil community being classified as *Acidobacteria* group 6 and the remaining 4% not being highly associated with any particular phylum (Table 5).

DISCUSSION

The aim of this study was to develop and test tools for quantitative analysis of specific phyla, classes, or OTUs (phylotypes at 97% similarity) in the KBS-LTER soil bacterial community by qPCR analysis of the 16S rRNA gene using validated primer sets offering appropriate levels of resolution and specificity. The group-specific bacterial primer sets used in this study were designed from a large (~5,000 sequences), site-specific database of bacterial 16S rRNA gene sequences from treatment 1 at the KBS-LTER site. The size of this library allowed a highly robust analysis of specificity within a well-defined microbial community. The rationale for this approach was that phylum-, class-, and OTU-specific primer sets derived solely from this extensive site-specific (i.e., closed) library could be optimally directed against the groups present without concern for nonspecific detection of extraneous sequences from other locations and environments. Primer sets were designed using the widely available and widely utilized PRIMROSE software at its default settings. Given that the

primer sets generated using this approach exhibited a diminished (but still substantial) ability to amplify phylotypes from other sites within the ARB database but belonging to the same phylum or class, they are probably at least to some degree site specific. This site specificity presumably makes them more appropriate for studies at the KBS-LTER site than primer sets generated in the context of all known sequences from everywhere, which was the objective of this work. We suggest that this approach represents a general strategy for development of qPCR primers that will be most effective at a given site.

The utility, efficacy, and specificity of these primer sets were assessed both in silico (using the KBS closed library and a much broader ARB-based library) and empirically using PCR on positive and negative control DNAs. Despite predicted high specificity for targeted sequences in silico ($\geq 83\%$ targets detected versus $\leq 17\%$ nontargets detected for all primer sets [Tables 2 and 3]), empirical testing revealed significant variability in target specificity, where some primer sets worked as predicted and others were highly cross-reactive. Presumably, this is due to the ability of primers with slight internal mismatches to bind sufficiently well to enable an initial elongation event. After the initial round of such "misprimed" elongation, subsequent PCR products would readily accumulate and sustain the exact sequence of the primers being used and thereafter would amplify with high efficiency. Thus, mismatching between primer and DNA template lends high specificity only where there is virtually no initial elongation taking place. Unfortunately, this likely occurs only where multiple, consecutive, or 3' prime mismatches are present or in in silico exercises where this "initial mispriming followed by high-fidelity amplification" phenomenon is not accounted for. We suggest that this phenomenon, at least in part, explains the equivocal nature of reports regarding primer specificity in previously published studies.

We acknowledge that our focus on a partial region of the 16S rRNA gene (corresponding to *Escherichia coli* positions 536 to 907) rather than full-length sequence information limited us to reliance on the V4 and V5 hypervariable regions for identification of signature sequences to provide specificity. However, it has previously been shown by our group and others that the V4-V5 region is suitable for consistent phylogenetic assignment compared to full-length sequence information (17, 26). This was further supported by our in silico analyses, which indicated high ratios of predicted target to nontarget sequence recovery from the large ARB database, suggesting that sufficient sequence information and variation were present to find differences delineating groups at the desired levels. Unfortunately, that indicated degree of specificity was often not sufficient when used experimentally, indicating a need for caution when analyzing data obtained using 16S rRNA gene-derived primers that have not been properly tested and validated empirically.

Our study employed a strategy of maximizing the depth of coverage of diversity by analyzing large numbers of partial (~400-bp) 16S rRNA gene sequences in single reads, which were sufficient for reliable classification. This strategy made it challenging to identify two specific, opposing primers that would generate a readily detectable PCR product, and we therefore paired a target-specific forward primer with the same general reverse primer used in the original cloning exercise.

Studies utilizing complete or larger stretches of 16S rRNA gene sequences might afford greater opportunity to identify targets, even those from pure cultures and with perfect sequence matches, casting doubts on the universality of such primers. Further, other groups have suggested that 16S rRNA gene group-specific primers can produce upwards of 25% non-target amplification, making them unsuitable for qPCR (10). It has also been suggested that reliance on the slowly evolving 16S rRNA gene makes it difficult to recognize recent events in the evolutionary history of a species, such as those associated with incipient speciation (24), which might also be a contributing factor to the high levels of nonspecific binding for some of the primer sets employed here.

Recent reports (4, 10, 12) have shown that so-called "universal" primer sets are incapable of efficiently amplifying all targets, even those from pure cultures and with perfect sequence matches, casting doubts on the universality of such primers. Further, other groups have suggested that 16S rRNA gene group-specific primers can produce upwards of 25% non-target amplification, making them unsuitable for qPCR (10). It has also been suggested that reliance on the slowly evolving 16S rRNA gene makes it difficult to recognize recent events in the evolutionary history of a species, such as those associated with incipient speciation (24), which might also be a contributing factor to the high levels of nonspecific binding for some of the primer sets employed here.

We conclude by recommending that specific care must be taken when interpreting new or previously published results obtained with 16S rRNA gene-based PCR primers that have not been fully validated, especially where population densities or dynamics are being inferred from the data. We further suggest that the reliability of quantification of group abundance using 16S rRNA gene-based qPCR is case specific and must be determined empirically rather than solely in silico.

ACKNOWLEDGMENTS

Funding for this project was provided by the U.S. Department of Agriculture National Research Initiative (USDA-CSREES grant 2004-03501).

Soil samples for this project were graciously provided by the Kellogg Biological Station Long Term Ecological Research project (KBS-LTER). We also gratefully acknowledge Linda Schimmelpfennig and Tara Westlie for technical assistance in creating the clone libraries.

REFERENCES

1. 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.
2. 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.
3. Ashelford, K. E., A. J. Weightman, and J. C. Fry. 2002. PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Res.* **30**:3481–3489.
4. 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.
5. Bathe, S., and M. Hausner. 2006. Design and evaluation of 16S rRNA sequence based oligonucleotide probes for the detection and quantification of *Comamonas testosteroni* in mixed microbial communities. *BMC Microbiol.* **6**:54.
6. Buckley, D. H., and T. M. Schmidt. 2001. Environmental factors influencing the distribution of rRNA from Verrucomicrobia in soil. *FEMS Microbiol. Ecol.* **35**:105–112.
7. Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity, and J. M. Tiedje. 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **33**:D294–D296.
8. Crosby, L. D., and C. S. Criddle. 2003. Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. *BioTechniques* **34**:790–802.
9. Doi, R. H. 1983. Recombinant DNA techniques, p. 162–163. Addison-Wesley Publishing Co., Inc., Reading, MA.
10. Farris, M. H., and J. B. Olson. 2007. Detection of Actinobacteria cultivated from environmental samples reveals bias in universal primers. *Lett. Appl. Microbiol.* **45**:376–381.
11. Fierer, N., J. A. Jackson, R. Vilgalys, and R. B. Jackson. 2005. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl. Environ. Microbiol.* **71**:4117–4120.
12. Frank, J. A., C. I. Reich, S. Sharma, J. S. Weisbaum, B. A. Wilson, and G. J. Olsen. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* **74**:2461–2470.
13. 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.
14. Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Y. Kumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Firster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Kinig, T. Liss, R. L  mann, 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.
15. Manz, W., R. Amann, W. Ludwig, M. Vancanneyt, and K.-H. Schleifer. 1996. Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum Cytophaga-Flavobacter-Bacteroides in the natural environment. *Microbiology* **142**:1097–1106.
16. Meier, H., R. Amann, W. Ludwig, and K.-H. Schleifer. 1999. Specific oligonucleotide probes for in situ detection of a major group of gram-positive bacteria with low DNA G+C content. *Syst. Appl. Microbiol.* **22**:186–196.
17. Morales, S. E., T. F. Cosart, J. V. Johnson, and W. E. Holben. 2009. Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters. *Appl. Environ. Microbiol.* **75**:668–675.
18. Morales, S. E., T. F. Cosart, J. V. Johnson, and W. E. Holben. 2008. Supplemental programs for enhanced recovery of data from the DOTUR application. *J. Microbiol. Methods* **75**:572–575.
19. Overmann, J., M. J. L. Coolen, and C. Tuschak. 1999. Specific detection of different phylogenetic groups of chemocline bacteria based on PCR and denaturing gradient gel electrophoresis of 16S rRNA gene fragments. *Arch. Microbiol.* **172**:83–94.
20. Polz, M. F., and C. M. Cavanaugh. 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**:3724–3730.
21. Reilly, K., V. R. Carruthers, and G. T. Attwood. 2002. Design and use of 16S ribosomal DNA-directed primers in competitive PCRs to enumerate proteolytic bacteria in the rumen. *Microb. Ecol.* **43**:259–270.
22. Sipos, R., A. J. Szekely, M. Palatinszky, S. Revesz, K. Marialigeti, and M. Nikolausz. 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol. Ecol.* **60**:341–350.
23. Stach, J. E. M., L. A. Maldonado, A. C. Ward, M. Goodfellow, and A. T. Bull. 2003. New primers for the class Actinobacteria: application to marine and terrestrial environments. *Environ. Microbiol.* **5**:828–841.
24. Stackebrandt, E., W. Frederiksen, G. M. Garrity, P. A. D. Grimont, P. Kampfer, M. C. J. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H. G. Truper, L. Vauterin, A. C. Ward, and W. B. Whitman. 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* **52**:1043–1047.
25. von Wintzingerode, F., U. B. Gobel, and E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* **21**:213–229.
26. Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**:5261–5267.