System. Appl. Microbiol. 26, 228–235 (2003)

http://www.urbanfischer.de/journals/sam

SYSTEMATIC AND APPLIED MICROBIOLOGY

# Candidate Division BD: Phylogeny, Distribution and Abundance in Soil Ecosystems

Daniel L. Mummey, and Peter D. Stahl

Department of Renewable Resources, University of Wyoming, Laramie, Wyoming, U.S.A.

Received: January 2, 2003

## Summary

Oligonucleotide primers were designed and used to amplify partial 16S rDNA sequences of the recently identified bacterial group BD from four diverse soils. Phylogenetic analysis of 34 BD group sequences supports division-level status for the group and also indicates that the BD group consists of at least 3 subdivision-level groups. Sequence divergence (21%) amongst these BD group sequences was found to be near the average for bacterial division-level lineages. An intercalating dye-based quantitative PCR (qPCR) assay was used to quantify BD phylogenetic group 3 16S rDNA in Wyoming shortgrass steppe soils. Although BD phylogenetic group 3 16S rDNA sequence numbers were high, averaging $3 \times 10^8$ copies per g soil, no significant correlations were found between their abundance and soil organic matter content, inorganic N concentration, or pH. Based on microscopically estimated cell numbers and the range of rRNA operons per genome in the bacterial domain, we estimate that BD group 3 represents between 0.75% and 10.7% of the microbial population in a shortgrass steppe soil. Our results indicate that the BD group is widely distributed in the environment and present in significant numbers in Wyoming shortgrass steppe soils.

**Key words:** *Gemmimonas* – candidate division BD – candidate division KS-B – quantitative PCR – phylogenetic analysis

## Introduction

A number of division-level relatedness groups have been identified in recent years using molecular phylogenetic methods [11]. One such division-level assemblage is the BD group, named for two PCR-amplified 16S rDNA gene fragments recovered from deep-sea sediments [16]. The monophyly of the BD group, and therefore it's status as a candidate division, was originally proposed by Hugenholtz et al. [12], based on the two deep-sea sediment clones, two soil clones (R. M. Goodman, S. B. Bintrim, J. S. Ireland, D. A. Joseph, T. J. Donohue, J. Handelsman, and G. P. Roberts, unpublished data), and a clone isolated from a sludge reactor [12]. Since that time additional 16S rDNA sequences affiliated with the BD group have been cloned from a number of ecosystems and an isolate cultivated from sludge reported (Sekiguchi,Y. personal communication) and tentatively named *Gemmimonas aurantiaca*.

While surveying microbial community structure in shortgrass steppe soils of Wyoming, USA, we PCR amplified three 16S rDNA sequences that clade with the BD group from three different ecosystems using "universal" eubacterial primers. The high incidence of BD group 16S rDNA sequences in our clone libraries (approximately 8%) suggests that this group may be an important and integral component of microbial communities in these semiarid soils. This apparent abundance, contrasting with the paucity of BD group sequences in public databases, prompted us to develop group-specific PCR primers to facilitate further study of this little known division-level bacterial clade.

The present study describes a rigorous phylogenetic analysis of the BD group, expands its known distribution in the environment, and, through quantitative analysis, demonstrates the numerical importance of a BD subdivision-level group in soil of a shortgrass steppe ecosystem.

## Materials and Methods

### Sample sites and soil collection

Soil samples (5–10 cm depth) were collected from four different ecosystems; two shortgrass steppe systems [an Aridisol collected from under homogeneous plant cover (Bouteloua gracilis) in southeastern Wyoming and an Aridisol collected from under

0723-2020/03/26/02-228 $ 15.00/0

heterogeneous plant cover dominated by sagebrush (Artemesia tridentata) in central Wyoming], an oak/magnolia forest in Georgia (Ultisol), and a disturbed prairie reclaimed after surface mining in Mongolia (Mollisol prior to disturbance). The Mongolian soil was stored and transported in non-sterile plastic bags at ambient temperatures prior to DNA extraction. All other soils were collected in sterile bags and frozen (–20 °C) within 24 hrs and maintained at this temperature prior to DNA extraction. Additional soil samples were obtained from the Wyoming sagebrush site for determination of BD 16S rDNA abundance using a quantitative PCR assay (see below). These soil samples were selected to represent a range of pH, and organic matter and inorganic N concentrations. Sample methodologies, site characteristics, and soil chemical measures are described in detail in Mummey et al. [19].

### Soil DNA extraction

Whole community genomic DNA was extracted from 0.4 g of soil from each sample using the Ultraclean Soil DNA kit (MoBio Laboratories, Solano Beach, Calif.) according to the instructions of the manufacturer. DNA yield was estimated using ethidium bromide-stained agarose gels with pUC18 molecular weight calibration standard.

### Primer design

Eight BD group 16S rDNA sequences, five obtained from public databases and three clones isolated from Wyoming Aridisols, were aligned using the ARB database and software (O. Strunk, and W. Ludwig, ARB: a software environment for sequence data. 1999. [http://www.mikro.biologie.tu-muenchen.de]). The ARB probe design utility was used to develop BD group specific primer candidates. Specificity of primer candidates was further assessed by searching both GeneBank and Ribosomal Database Project [17] databases for non-target matches to primer sequences. The primer BD48f (5′-CAAGGGCAACCGGCGAAC-3′) was shown to be specific for all three BD groups sequences we cloned from shortgrass steppe soils and two previously identified BD group sequences (sludge clone SBRH83 and turfgrass soil clone SAB1030, Table 1). A second primer, STBDf, (5′-GCAACCCCGGTTCAGTGGCG-3′) was developed after recognizing the affiliation of Sturt soil clones 0319-7E19 and 0319-7G21 [9] (Table 1) with the BD group (see analyses below) using the ARB probe design utility as above.

### PCR, cloning, and sequencing

Soil community 16S rDNA was PCR amplified using BD48f or STBDf as forward primers and either 907r or 1492r as reverse primers [15]. The 50 µl reaction mixtures contained 1× PCR buffer, 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.5 µM of each primer, 2 U Taq polymerase (Promega, Madison, WI.), and approximately 50 ng template DNA. Thermocycling was conducted on an iCycler thermal cycler (Bio-Rad, Benecia, Calif.). Amplifications utilizing primer BD48f were started with an initial denaturation step of 94 °C for 3 min, followed by 32 to 38 cycles of 94 °C for 30 sec, 56 °C for 45 sec, and 72 °C for 2 min; cycling was completed by a final extension period of 72 °C for 5 min. Amplifications utilizing primer STBDf were started with an initial denaturation step of 94 °C for 3 min, followed by 2 cycles of 94 °C for 30 sec, 58 °C for 45 sec, and 72 °C for 2 min, 2 cycles in which the annealing temperature was lowered to 57 °C, and 24 cycles with 56 °C annealing temperature, cycling was completed by a final extension period of 72 °C for 5 min.

PCR products generated from whole soil community genomic DNA were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, Calif.). Randomly selected positive clones were screened for recombinant plasmids having inserts of expected size using agarose gel electophoresis. Plasmids were purified from overnight cultures using Wizard Plus minipreps (Promega). Putative 16S rDNA inserts were sequenced using approximately 1 µ(g of purified plasmid DNA as the template in cycle sequencing reactions with fluorescent dye-labeled terminators (ABI PRISM dRhodamine Cycle Sequencing Ready Reaction Kit, Applied Biosystems Inc., Fremont, Calif.) according to the manufacturers recommendations. Reactions were determined using a ABI 377 DNA sequencer (Applied Biosystems).

**Table 1.** BD group sequences used in this study.

| Taxon | Accession no. | Length (nt) |
|---|---|---|
| Georgian soil clone BDGEO_1* | AY145851 | 830 |
| Georgian soil clone BDGEO_13* | AY145852 | 830 |
| Georgian soil clone BDGEO_17* | AF545647 | 830 |
| Georgian soil clone BDGEO_53* | AF545653 | 830 |
| Wyoming soil clone STBD1* | AF545654 | 834 |
| Wyoming soil clone STBD7* | AF545656 | 835 |
| Wyoming soil clone STBD13* | AF545655 | 837 |
| Wyoming soil clone STBD14* | AF545657 | 833 |
| Wyoming soil clone BDWYO_4*# | AF545646 | 799 |
| Wyoming soil clone BDWYO_7*# | AF545644 | 831 |
| Wyoming soil clone BDWYO_7.2* | AF545648 | 831 |
| Wyoming soil clone BDWYO_8*# | AF545643 | 831 |
| Wyoming soil clone BDWYO_13* | AF545642 | 824 |
| Wyoming soil clone BDWYO_14* | AF545645 | 830 |
| Wyoming soil clone BDb_2*# | AF545652 | 833 |
| Wyoming soil clone WYO13* | AF545642 | 897 |
| Wyoming soil clone BDfull_1* | AF545649 | 1397 |
| Wyoming soil clone Path12* | AF545658 | 900 |
| Mongolian soil clone pBD_5*# | AF545650 | 830 |
| Mongolian soil clone pBD_7*# | AF545651 | 829 |
| Restored surface mine clone GR12* | AF545640 | 902 |
| Gemmimonas aurantiaca | AB072735 | 1446 |
| Agricultural soil clone SCI2# | AJ252607 | 1046 |
| Turfgrass soil clone SBB1063 | AF009988 | 1469 |
| Turfgrass soil clone SAB1030 | AF009987 | 1465 |
| Sturt soil clone #0319-7E19 | AF234139 | 1301 |
| Sturt soil clone #0319-7F1 | AF234140 | 1440 |
| Sturt soil clone #0319-7G21 | AF234148 | 1349 |
| Rhizosphere clone RSC-II-71 | AJ252692 | 1029 |
| Marine sediment clone TIHP368-34 | AB031649 | 811 |
| Mud clone KS77 | AF328213 | 993 |
| Mud clone KS67 | Af328204 | 873 |
| Aquatic cave clone Wb1_a18 | AF317745 | 1481 |
| Deep sea sediment clone str BD2_11 | AB015540 | 1524 |
| Deep sea sediment clone str BD7_2 | AB015578 | 1524 |
| Sludge clone SBRH63 | AF268993 | 1450 |
| Sludge clone Ebpr21 | AF255632 | 1380 |

*Sequences determined in this study.
# Sequences that consistently clade within phylogenetic group 3 but are unaffiliated with any other sequence.

### Phylogenetic analyses

Sequences were screened for the presence of chimeric artifacts [14] by secondary structure anomalies and using the CHIMERA_CHECK program (version 2.7) of the Ribosomal Database Project (RDP) [17]. Sequences were compared to those in available databases using BLAST (Basic Local Alignment Search Tool) [1] and RDP SIMILARITY_RANK [17] searches to determine potential phylogenetic affiliations. Cloned sequences generated in this study and sequences from public databases having > 85% identity to potential BD sequences were aligned

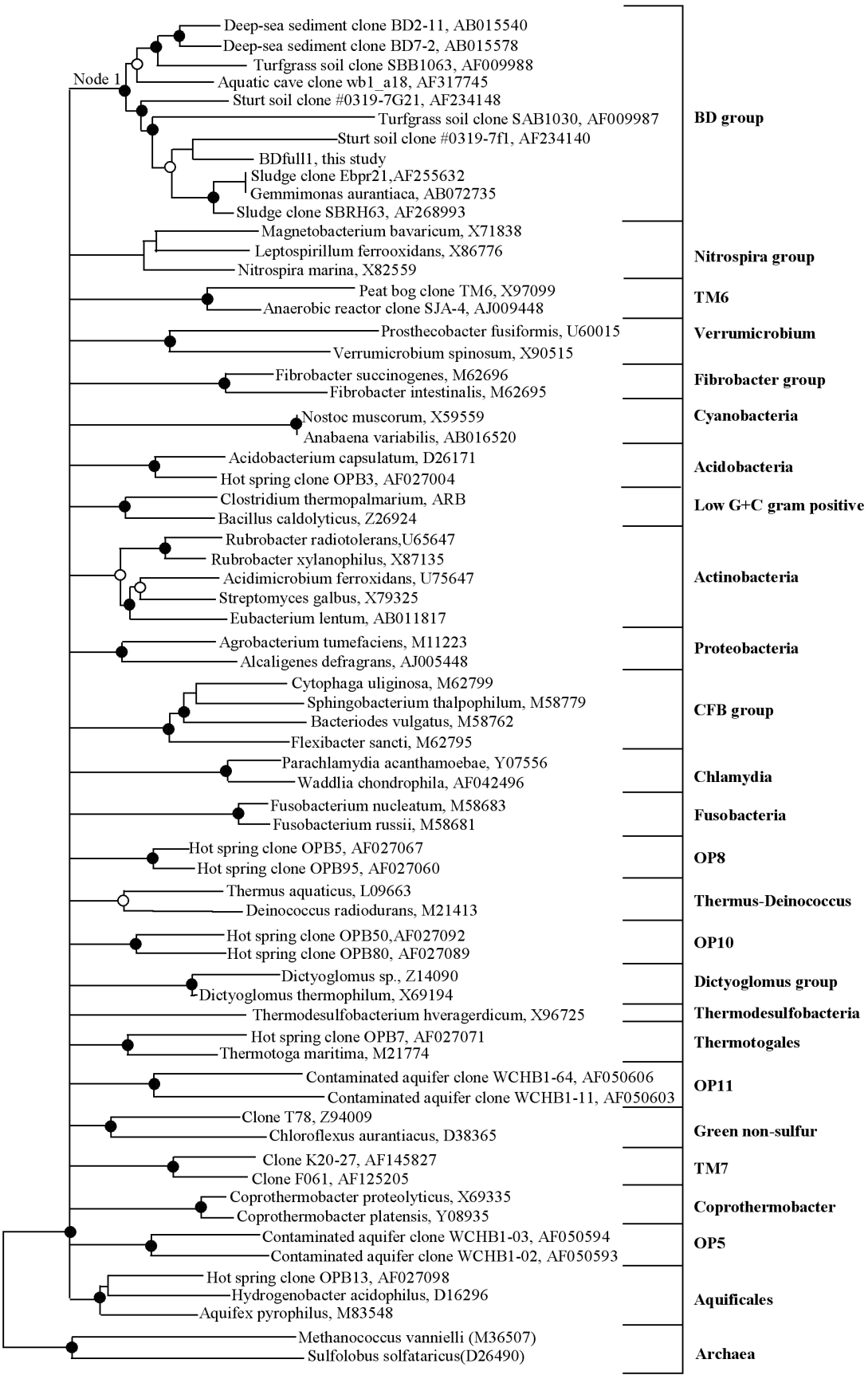
Deep-sea sediment clone BD2-11, AB015540
Deep-sea sediment clone BD7-2, AB015578
Turfgrass soil clone SBB1063, AF009988
Node 1
Aquatic cave clone wb1_a18, AF317745
Sturt soil clone #0319-7G21, AF234148
Turfgrass soil clone SAB1030, AF009987
BD group
Sturt soil clone #0319-7f1, AF234140
BDfull1, this study
Sludge clone Ebpr21,AF255632
Gemmimonas aurantiaca, AB072735
Sludge clone SBRH63, AF268993
Magnetobacterium bavaricum, X71838
Leptospirillum ferrooxidans, X86776
Nitrospira group
Nitrospira marina, X82559
Peat bog clone TM6, X97099
TM6
Anaerobic reactor clone SJA-4, AJ009448
Prosthecobacter fusiformis, U60015
Verrumicrobium
Verrumicrobium spinosum, X90515
Fibrobacter succinogenes, M62696
Fibrobacter group
Fibrobacter intestinalis, M62695
Nostoc muscorum, X59559
Cyanobacteria
Anabaena variabilis, AB016520
Acidobacterium capsulatum, D26171
Acidobacteria
Hot spring clone OPB3, AF027004
Clostridium thermopalmarium, ARB
Low G+C gram positive
Bacillus caldolyticus, Z26924
Rubrobacter radiotolerans,U65647
Rubrobacter xylanophilus, X87135
Actinobacteria
Acidimicrobium ferroxidans, U75647
Streptomyces galbus, X79325
Eubacterium lentum, AB011817
Agrobacterium tumefaciens, M11223
Proteobacteria
Alcaligenes defragrans, AJ005448
Cytophaga uliginosa, M62799
Sphingobacterium thalpophilum, M58779
CFB group
Bacteriodes vulgatus, M58762
Flexibacter sancti, M62795
Parachlamydia acanthamoebae, Y07556
Chlamydia
Waddlia chondrophila, AF042496
Fusobacterium nucleatum, M58683
Fusobacteria
Fusobacterium russii, M58681
Hot spring clone OPB5, AF027067
OP8
Hot spring clone OPB95, AF027060
Thermus aquaticus, L09663
Thermus-Deinococcus
Deinococcus radiodurans, M21413
Hot spring clone OPB50,AF027092
OP10
Hot spring clone OPB80, AF027089
Dictyoglomus sp., Z14090
Dictyoglomus group
Dictyoglomus thermophilum, X69194
Thermodesulfobacterium hveragerdicum, X96725
Thermodesulfobacteria
Hot spring clone OPB7, AF027071
Thermotogales
Thermotoga maritima, M21774
Contaminated aquifer clone WCHB1-64, AF050606
OP11
Contaminated aquifer clone WCHB1-11, AF050603
Clone T78, Z94009
Green non-sulfur
Chloroflexus aurantiacus, D38365
Clone K20-27, AF145827
TM7
Clone F061, AF125205
Coprothermobacter proteolyticus, X69335
Coprothermobacter
Coprothermobacter platensis, Y08935
Contaminated aquifer clone WCHB1-03, AF050594
OP5
Contaminated aquifer clone WCHB1-02, AF050593
Hot spring clone OPB13, AF027098
Hydrogenobacter acidophilus, D16296
Aquificales
Aquifex pyrophilus, M83548
Methanococcus vannielli (M36507)
Archaea
Sulfolobus solfataricus(D26490)

**Table 2.** Outgroup datasets used in analysis 2.

| Division | Taxon (Accession no.) | Outgroup datasets | | | | |
|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 |
| Aquificales | Aquifex pyrophilus (M83548) | | × | | | × |
| | Hydrogenobacter acidophilus (D16296) | | × | | | × |
| | Hot-spring clone OPB13 (AF027098) | | × | | | × |
| Thermotogales | Thermotoga maritima (M21774) | × | × | | | |
| | Hot-spring clone OPB7 (AF027071) | × | × | | | |
| | Geotoga subterranea (L10659) | × | × | | | |
| Nitrospira | Nitrospira marina (×82559) | × | | × | | × |
| | Leptospirillum ferrooxidans (×86778) | × | | × | | × |
| | Magnetobacterium bavaricum (×71838) | × | | × | | × |
| Green sulfur | Hot-spring clone OPB56 (AF027009) | × | | | × | |
| | Chlorobium vibrioforme (M62795) | × | | | × | |
| Low G+C gram positive | Bacillus caldolyticus (Z26924) | × | | | × | |
| | Clostridium botulinum (L37593) | × | | | × | |
| Proteobacteria | Agrobacterium tumefaciens (M11223) | × | | | × | |
| | Alcaligenes defragrans (AJ005448) | × | | | × | |
| | Geobacter metallireducens (L07834) | × | | | × | |
| green non-sulfur | Chloroflexus aurantiacus (D38365) | | × | | × | |
| | Clone T78 (Z94009) | | × | | × | |
| | Clone SAR307 (U20798) | | × | | × | |
| Chlamydia | Parachlamydia acanthamoebae (Y07556) | | × | | | |
| | Simkania negevensis (L27666) | | × | | | |
| | Waddlia chondrophila (AF042496) | | × | | | |
| Actinobacteria | Acidimicrobium ferroxidans (U75647) | | × | | | × |
| | Streptomyces galbus (×79325) | | × | | | × |
| | Eubacterium lentum (Ab011817) | | × | | | × |
| | Rubrobacter radiotolerans (U65647) | | × | | | × |
| Cyanobacteria | Nostoc muscorum (×59559) | | | × | | × |
| | Anabaena variabilis (AB016520) | | | × | | × |
| CFB group | Bacteriodes vulgatus (M58762) | | | × | × | |
| | Cytophaga uliginosa (M62799) | | | × | × | |
| | Flexibacter sancti (M62795) | | | × | × | |
| Acidobacteria | Acidomicrobium capsulatum (D26171) | | | × | | × |
| | Holophaga foetida (×77215) | | | × | | × |
| | Hot-spring clone OPB3 (AF027004) | | | × | | × |
| Fibrobacteria | Fibrobacter succinogenes (M62696) | | | × | × | |
| | Fibrobacter intestinalis (M62695) | | | × | × | |

**Fig. 1.** Modified phylogenetic dendrogram of dataset 1 generated using neighbor joining (Olsen correction) in ARB. GenBank accession numbers are listed parenthetically. Division-level branches were collapsed back to the root node. Branch points within the BD group supported by bootstrap values >75% for ED, MP and ML (F81) methods and resolved by GTR analysis in all analyses are indicated by filled circles; open circles indicate branch points resolved by the GTR method but lacking support (bootstrap values <75%) in one ED, MP and ML (F81) analysis; unmarked branches were either not resolved or lacked bootstrap support for two or more analyses. Filled circles at branch points for all other bacterial groups indicate support (bootstrap values >75% for all ED and MP analyses; open circles indicate lack of bootstrap support in a single analysis; unmarked branches were not supported (bootstrap values <75%) in more than one analysis.

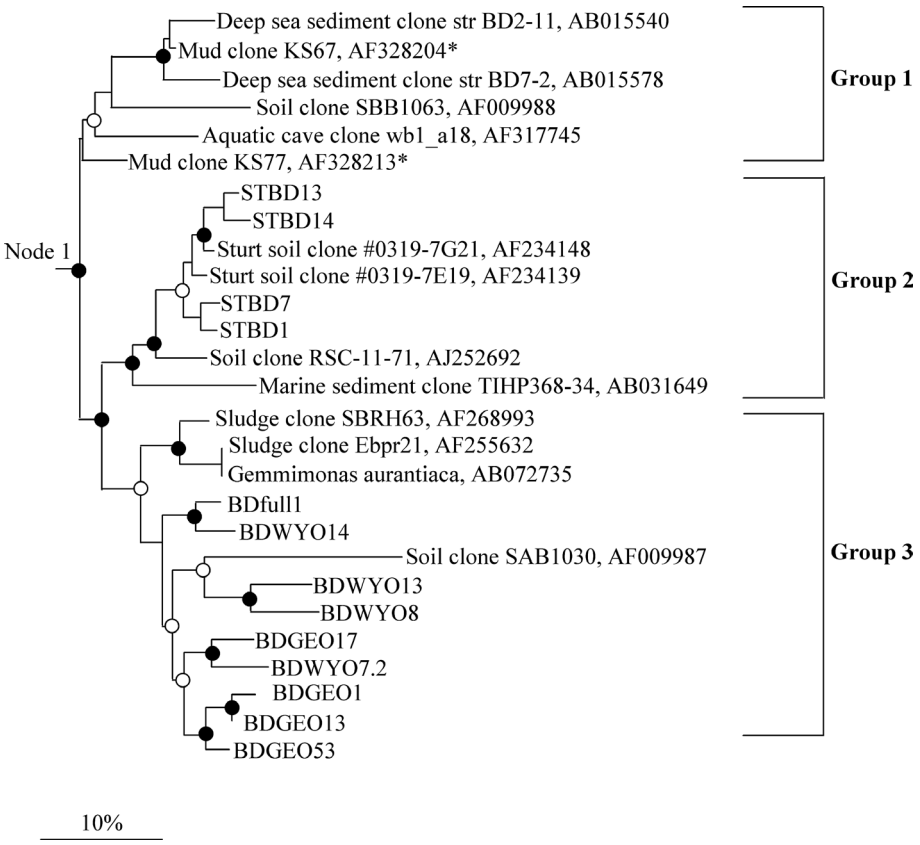


**Fig. 2.** Phylogenetic dendrogram of BD group sequences (pos. 126 to 959, *E. coli* numbering) with outgroup sequences removed. Branch points supported by bootstrap values >75% for MP and both ED analyses are indicated by filled circles; open circles indicate branch points having bootstrap support for two of three MP and ED analyses; unmarked branches lacked bootstrap support for two or more analyses. Group 3 sequences with support only at the group 3 root node and lacking affiliation with at least one other group

with 16S rDNA reference sequences using the ARB database and software and refined manually. Hypervariable regions of 16S rDNA alignments were excluded from similarity calculations using the bacterial mask of Lane [15].

Initial phylogenetic analyses incorporated 16S rDNA sequences previously proposed to be affiliated with the BD group [11], sequences cloned from soil in this study, and database sequences identified using BLAST as having homology to potential BD sequences. Outgroup configurations for these analyses included all or portions of the ARB database (>16000 sequences). Phylogenetic trees were generated in ARB (using neighbor joining with Olsen correction and the Quick add sequence insertion tool using parsimony) in order to determine tentative phylogenetic affiliations. Sequences found to cluster with BD sequences were retained and subjected to more rigorous phylogenetic analysis as follows.

Five methods were employed to generate tree topologies using the software package PAUP* [D. L. Swofford, Phylogenetic Analysis Using Parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Mass., 1998.]. They include two evolutionary distance (ED) methods (neighbor joining with either maximum-likelihood or Kimura two-parameter correction with empirically determined base frequencies and empirically determined gamma distribution models of site-to-site rate variation), maximum parsimony (MP) (default settings, heuristic search) and two maximum likelihood (ML) methods [(Felsenstein 1981 substitution model (f81) [5]; and the general time-reversible (GTR) substitution model [21], estimated shape and proportion of invariable sites, empirically determined base frequencies; heuristic searches].

A number of taxon datasets were constructed to include a diverse and relatively large number of outgroups [7] and to facilitate use of computationally intensive ML methods. Dataset one contained 56 near full-length sequences from members of 25 bacterial divisions (taxa and accession numbers are presented in Figure 1). Only ED and MP analyses were used to analyze this dataset. The ingroup consisted of all potential BD group sequences having length greater than 1325 bp or, in separate analyses, all potential BD group sequences (pos. 126 to 926, *E. coli* numbering) (Table 1).

Five additional taxon datasets were constructed in order to utilize a multiple-outgroup approach [3]. Each dataset contained between 13 to 16 outgroup sequences and 5 or 6 division-level outgroups (Table 2). Ingroup sequences consisted of all potential BD group sequences longer than 1325 bp. These datasets were analyzed using MP, ED, and ML methods. All six datasets used in this study are available at http://w3.uwyo.edu/~dmummey/BD.htm.

The robustness of ED, MP, and ML (f81) tree topologies was determined by bootstrap resampling. Two thousand bootstrap repetitions were performed for ED and MP analyses and 100 for ML analyses using the F81 substitution model. Branches were considered resolved by the GTR method if branch lengths were positive at $p < 0.01$ level of significance by a likelihood ratio test (PAUP*, ZeroLenTest).

### qPCR

Quantification of BD group 3 sequences in total soil DNA was performed with an iCycler system (Bio-Rad Laboratories, Calif.) utilizing SYBR® Green PCR Core Reagents (PE Biosystems, Calif.). Oligonucleotide primer, $MgCl_2$, and SYBR® Green concentrations were optimized according to ABI protocols. Each 25 µl reaction contained 25 pmol each of primers BD48 and 907r, 2 mM $MgCl_2$, and 2.5 µl SYBR® Green mix. Reaction mixtures were heated to 96 °C for 10 min and amplified with 35 cycles at 96 °C for 25 sec, 56 °C for 1 min, and 72 °C for 2 min, with a final 5 min extension at 72 °C. Detection of the fluorescent product was set at the last step in each cycle. To determine the specificity of amplification, product size verification by gel electrophoresis and analysis of product melting was performed after each amplification. Product melting curves were obtained by slow heating with a 0.1 °C/s increment from 60 °C to 95 °C, with fluorescence collection at 0.1 °C intervals. All reaction mixtures were amplified in triplicate PCRs. Dilutions of the 1397 bp clone Bdfull1 were used as calibration standards. Efficiency of calibration standard amplification reactions were determined using the formula $E = (10^{-1/slope}) - 1$.

Cell lysis efficiency was estimated by counting 4′,6′-diamidino-2-phenylindole (DAPI)-stainable cells in soil before and after bead-beating and DNA extraction. Samples were immersed in 2.5 µg/ml DAPI for 5 min on polycarbonate membrane filters (Poretics Corp., Calif., 25 mm diameter) and examined by epifluorescent microscopy using a Nikon Diaphot inverted microscope.

External standards were used to determine soil DNA extraction efficiency, the extent of PCR inhibition, and optimal template concentrations. To determine soil DNA extraction efficiency, DNA encoding *Coxiella burnetii scvA* [8] was added to soils immediately prior to DNA extraction. Extractable *scvA* DNA was estimated by qPCR using primers scvAf (5′-GACAAAT GTCCAACAACAACG-3′) and scvAr (5′-ATTTGGTCGACG TGGGTTAG-3′) in reaction mixtures optimized as above. Reaction mixtures were heated to 96 °C for 5 min and amplified with 35 cycles at 96 °C for 30 sec, 57 °C for 30 sec, and 72 °C for 1 min, with a final extension at 72 °C for 2 min.

Estimation of the extent of PCR inhibition and optimal template concentrations were determined by spiking a series of soil DNA extracts dilutions (0, 10, 100, 500, and 1000 fold dilutions) with known amounts of DNA encoding *scvA*, followed by qPCR analysis as described above. The highest soil DNA extract concentration in each dilution series exhibiting a linear relationship with greater dilutions, and yielding the expected threshold cycle for external standards, were used in quantitative analyses of BD group 3 abundance.

## Results and Discussion

### Phylogenetic analysis

Two BD group-specific primers, BD48f and STBDf (see materials and methods), were developed. Primer BD48f, when used in conjunction with the universal eubacterial primer 907r, was found to be highly specific for members of the BD group. No non-target matches were found between public database sequences and this primer sequence (< 3 mismatches) and, although only a few clones from each of four libraries were sequenced (Table 1), no non-BD group sequences were found.

Using primers BD48f or STBDf, BD group sequences were amplified from whole-community genomic DNA extracted from soil of 2 shortgrass steppe ecosystems in Wyoming, USA, a forest in Georgia, USA, and a Mongolian surface mine reclamation site. Despite the paucity of BD group sequences found in soil 16S rDNA libraries reported in the literature, all four soil environments sampled in this study yielded BD group sequences.

A total of 21 BD group clones were sequenced in this study (Table 1) and checked for the presence of chimeric artifacts as described above. Although no chimeric sequences were identified, the presence of chimeric artifacts cannot be completely ruled out due to the small number of BD group sequences available. Additional potential BD group sequences (Table 1) were selected from public databases based on their similarity to sequences cloned from soil in this study and to sequences previously identified as being affiliated with the BD group [12]. After initial screening, using >16000 bacterial 16S rDNA sequences in the ARB database to determine tentative phylogenetic affiliation, 34 sequences were selected for further analysis to verify BD group membership.

A bacterial division can be defined as a lineage comprised of two or more 16S rRNA sequences that are reproducibly monophyletic and unaffiliated with all other division-level relatedness groups that constitute the bacterial domain [11]. After analyses utilizing MP, 2 ED, and 2 ML methods to generate tree topologies, with six different outgroup configurations representing 25 division-level phylogenetic groups, all 34 BD group sequences were found to be reproducibly monophylytic and distinct from all other sequences used in our analyses. These results are therefore strongly supportive of BD group division-level status.

A broad sampling of bacterial divisions indicates that intradivisional sequence divergence ranges from 13 to 33% [4]. The ten BD group sequences having lengths greater than 1325 bp exhibited near 16% sequence divergence after removal of hypervariable regions (Lane bacterial mask [15]). When regions between positions 126 and 959 (*E. coli* numbering) of all BD group sequences used in this study were included, a sequence divergence of 21% was found, indicating that candidate division BD group sequence divergence is well within the range of established bacterial divisions.

Similar to the above bacterial division definition, a bacterial subdivision can be defined as a lineage comprised of two or more 16S rRNA sequences within a division that are reproducibly monophyletic and unaffiliated with all other representatives of that division [11]. Employing these criteria, analyses using near full-length sequences indicated the presence of least two subgroups within the BD group (Fig. 1). When all BD group sequences (pos. 126 to 959, *E. coli* numbering) were included in the analyses, three groups met the subdivision criteria. Of the 20 sequences comprising phylogenetic group 3, six sequences having consistently high support for group-3 membership (Table 1) were unaffiliated with any other sequence. The presence of these sequences suggests that as more BD group sequence data becomes available, additional subgroupings may be identified or group 3 may coalesce into a more defined phylogenetic entity.

Environments from which the BD group sequences included in this study were obtained are diverse: high-pressure, oligotrophic deep-sea sediments [16], mobile ocean nearshore mud deposits [18], a dense "mantle" of biological materials in an aquatic cave [10], sludge reactors [2, 12], an Antarctic cryoconite hole (Christner, B.C., Kvitko, B.H., Reeve, J.N., unpublished data) and a number of different soil environments (listed below), suggesting that members of the BD group are widespread in the environment.

Including the four geographically dispersed soils from which we amplified BD sequences, at least five additional soil ecosystems have yielded 16S rDNA sequences shown to be affiliated with the BD group. These include arid Australian soils [9], Wisconsin turfgrass soil (Goodman, R.M., Bintrim, S. B., Ireland, J.S., Joseph, D.A., Donohue, T.J., Handelsman, J., Roberts, G. P. unpublished data), two European agricultural soils (Lukow, T. unpublished data) and an agricultural soil near Athens, Georgia [6]. These soils are from four different continents (N. America, Australia, Asia, and Europe) and vary greatly in plant cover, chemical and physical properties, as well as temperature and precipitation regimes. These findings suggest that representatives of the BD group may be a component of surface soil microbial communities worldwide.

## Estimation of BD phylogenetic group-3 abundance

A potential problem associated with quantitative methods based on the PCR of DNA extracted from environmental samples is inhibition of reactions by humic matter and other substances that co-purify with DNA during sample extraction [22]. Although all samples analyzed utilizing undiluted soil DNA extracts exhibited PCR inhibition to some extent, dilution of soil DNA extracts was found to eliminate PCR inhibition, as determined by both the linearity of product formation after serial dilution of soil DNA extracts and lack of inhibition of PCR product formation from external standards (data not shown). Typically, 100 fold dilutions of template soil DNA extracts were required to eliminate PCR inhibition.

Our analyses utilized a specific (BD48f) and a general primer (907r), which yielded amplicons of nearly 800 bp. Although these reactions were highly efficient (efficiency values approximately 102%) and reproducible, the unknown composition of our samples complicates determination of PCR product specificity. While the possibility of non-specific amplification cannot be completely ruled out, no non-BD database sequences were found to have fewer than 3 mismatches to primer BD48f, and sequence analysis revealed no non-BD group clones. In addition, melting curve analysis indicated no nonspecific products, such as primer dimers, having lower *Tm*s, and agarose gel electrophoresis of reaction mixtures consistently showed discrete PCR products of the expected molecular weight for all assays (data not shown).

To determine BD group 3 16S rDNA copies per g soil the efficiency of cell lysis and DNA extraction must first be determined. Cell lysis efficiency, as estimated by epifluorescent microscopically determined cell counts, averaged 96.5% (stdev = 2.1). DNA extraction efficiency, as estimated by qPCR quantification of external standards added to soil before DNA extraction, averaged 89% (stdev = 9.5). After correcting for differences in cell lysis and DNA extraction efficiencies BD group 3 16S rDNA copies per gram soil were found to average 3.05E+08 (95% confidence interval of 2.20E+08 to 3.91E+08).

It is not possible at this time to accurately determine BD group 3 cell numbers from 16S rDNA abundance because rRNA operon copy numbers per cell are unknown. However, within the bacterial domain rRNA operon numbers per cell range from 1 to 15 [13]. Total cell numbers were found to average $2.45 \times 10^9$ per gram soil. Using this average estimate and the range of rRNA operons per cell found in the bacterial domain, BD group 3 cells would, on average, comprise about 0.75% of the total cell number if 15 rRNA operons per cell is used in the calculation, and 10.7% of the total cell number per gram of soil if all BD cells are considered to have a single rRNA operon. These results strongly suggest that representatives of BD group 3 are of numerical importance in these soils.

No significant correlations were found between BD group 3 16S rDNA copy number and soil organic matter content, inorganic N concentration, or soil pH (data not shown). These results suggest that BD group 3 abundance may not be affected by these variables within the ranges found in the soils analyzed. However, a given soil sample contains a multitude of microhabitats which likely constitute a broad range of physical and chemical conditions [20] and our analyses provide little insight into other potential regulators on bacterial abundance, such as soil structural differences. More rigorous analysis of soil properties, or analysis at different scales than those conducted here, may therefore be required to elucidate environmental factors controlling BD group 3 abundance. Additionally, since many division- or subdivision-level lineages are physiologically diverse, controls on BD group 3 abundance may differ significantly between representatives.

In conclusion, this study substantiates the division-level nature of the BD group within the domain *Bacteria*. The group is represented by at least 3 subdivision-level lineages that are widespread in the environment. Our results indicate that BD group 3 is of numerical importance in soil of a shortgrass steppe ecosystem, suggesting these organisms are at least important to the carbon budget of this ecosystem. The wide distribution and numerical importance of the BD group warrants its further study, which the primers and methods described here will facilitate.

**Acknowledgments**

We thank Mark Gomelsky, David White, and William Wilson for critical review of this manuscript. Additional thanks are extended to William Wilson for technical advice and use of equipment, and to Dale Howe for generously supplying scvA DNA and PCR primers. This research was supported by a grant from NSF-EPSCoR (EPS-9983278).

## References


1. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman. D. J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 24, 3389–3402 (1997).
2. Crocetti, G. R., Hugenholtz, P., Bond, P. L., Schuler, A., Keller, J., Jenkins, D., Blackall, L. L.: Identification of polyphosphate-accumulating organisms and design of 16S rRNA-directed probes for their detection and quantification. Appl. Env. Microbiol. 66, 1175–1182 (2000).
3. Dalevi, D., Hugenholtz, P., Blackall, L. L.: A multiple-outgroup approach to resolving division-level phylogenetic relationships using 16S rDNA data. Int. J. Syst. Evol. Microbiol. 51, 385–391 (2001).
4. Dojka, M. A., Harris, J. K., Pace, N. R.: Expanding the known diversity and environmental distribution of an uncultured phylogenetic division of bacteria. Appl. Env. Microbiol. 66, 1617–1621 (2000).
5. Felsenstein, J.: Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376 (1981).
6. Furlong, M. A., Singleton, D. R., Coleman, D. C., Whitman, W. B.: Molecular and culture-based analyses of prokaryotic communities from an agricultural soil and the burrows and casts of the earthworm *Lumbricus rubellus*. Appl. Environ. Microbiol. 68, 1265–1279 (2002).
7. Hillis, D. M.: Inferring complex phylogenies. Nature 383, 130–131 (1996).
8. Heinzen, R. A., Howe, D., Mallavia, L. P., Rockey, D. D., Hackstadt, T.: Developmentally regulated synthesis of an unusually small, basic peptide by *Coxiella burnetii*. Mol. Microbiol. 22, 9–19 (1996).
9. Holmes, A. J., Bowyer, J., Holley, M. P., O'Donoghue, M., Montgomery, M., Gillings, M. R.: Diverse, yet-to-be-cultured members of the Rubrobacter subdivision of the Actinobacteria are widespread in Australian arid soils. FEMS Microbiol. Ecol. 33, 111–120 (2000).
10. Holmes, A. J., Tujula, N. A., Holley, M., Contos, A., James, J. M., Rogers, P., Gillings, M. R.: Phylogenetic structure of unusual aquatic microbial formations in Nullarbor caves, Australia. Environ. Microbiol. 3, 256–264 (2001).
11. Hugenholtz, P. G., Goebel, B. M., Pace, N. R.: Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. 150, 4765–4774 (1998).
12. Hugenholtz, P., Tyson, G. W., Webb, R. I., Wagner, A. M., Blackall, L.L.: Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. Appl. Envir. Microbiol. 67, 411–419 (2001).
13. Klappenbach, J. A., Saxman, P. R., Cole, J. R., Schmidt, T. M.: rrndb: the ribosomal RNA operon copy number database. Nucleic Acids Res. 29, 181–184 (2001).
14. Kopczynski, E. D., Bateson, M. M., Ward, D. M.: Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Appl. Envir. Microbiol. 60, 746–748 (1994).
15. Lane, D. J.: 16S/23S rRNA sequencing, pp. 115-175. In: Nucleic acid techniques in bacterial systematics (E. Stackebrandt, M. Goodfellow, eds.), John Wiley and Sons, New York, N.Y. 1991.
16. Li, L., Kato, C., Horikoshi, K.: Bacterial diversity in deep-sea sediments from different depths. Biodiversity and Conservation 8, 659–677 (1999).
17. Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker, C. T. Jr., Saxman P. R., Farris, R. J., Garrity, G. M., Olsen, G. J., Schmidt, T. M., Tiedje, J. M.: The RDP-II (Ribosomal Database Project). Nucleic Acids Res. 29, 173–174 (2001)
18. Madrid, V. M., Aller, J. Y., Aller, R. C., Chistoserdov, A. Y.: High prokaryote diversity and analysis of community structure in mobile mud deposits of French Guiana: identification of two new bacterial candidate divisions. FEMS Microbio. Ecol. 37, 197–209 (2001).
19. Mummey, D. L., Stahl, P. D., Buyer, J. S. : Soil microbiological properties 20 years after surface mine reclamation: spatial analysis of reclaimed and undisturbed sites. Soil Biol. Biochem.34, 1711–1719 (2002).
20. Nunan, N., Ritz, K., Crabb, D., Harris, K., Wu, K., Crawford, J. W., Young, I. M.: Quantification of the in situ distribution of soil bacteria by large-scale imaging of thin sections of undisturbed soil. FEMS Microbiol. Ecol. 36, 67–77 (2001).
21. Swofford, D. L., Olsen, G. J., Waddell, P. J., Hillis, D. M.: Phylogenetic Inference, p. 407–514. In: Molecular Systematics (D. Hillis, C. Moritz, B. Mable, eds.), Sinauer Associates, Sunderland, Mass. 1999.
22. Wilson, I. G.: Inhibition and facilitation of nucleic acid amplification. Appl. Environ. Microbiol. 63, 3741–3751 (1997).



**Corresponding author:**
Daniel L. Mummey, Department of Renewable Resources, University of Wyoming, P. O. Box 3354, Laramie, Wyoming 82070, U.S.A.
e-mail: Dmummey @uwyo.edu