J Forensic Sci, May 2006, Vol. 51, No. 3
doi:10.1111/j.1556-4029.2006.00131.x
Available online at: www.blackwell-synergy.com

Cheryl R. Kuske,[1] *Ph.D.; Susan M. Barns,*[1] *Ph.D.; Christy C. Grow,*[1] *B.S.; Lori Merrill,*[1] *M.S.; and John Dunbar,*[1] *Ph.D.*

Environmental Survey for Four Pathogenic Bacteria and Closely Related Species Using Phylogenetic and Functional Genes*

ABSTRACT: Bacterial species with high DNA sequence similarity to pathogens could affect the specificity of assays designed to detect biological threat agents in environmental samples. The natural presence of four pathogenic bacteria, *Bacillus anthracis*, *Clostridium perfringens*, *Francisella tularensis*, and *Yersinia pestis* and their closely related species, was determined for a large collection of soil and aerosol samples. Polymerase chain reaction (PCR) and gene sequencing were used using group-specific 16S rRNA primers to identify pathogens and related species, and pathogen-specific virulence genes. Close relatives of *B. anthracis* (*B. cereus* group species) were detected in 37% of the soils and 25% of the aerosol samples. The *B. anthracis* protective antigen (*pag*) gene or a close homolog was detected in 16 of these samples. For the other three pathogen groups, the frequency of detection was much lower, and none of the samples were positive with both the phylogenetic and virulence gene primer sets.

KEYWORDS: forensic science, environmental detection of biothreat agents, *Bacillus anthracis*, *Yersinia pestis*, *Francisella tularensis*, *Clostridium perfringens*, backgrounds, soil, aerosol, select agent

Detection and forensic investigation of events that potentially involve release of biothreat agents require differentiation of a released pathogenic agent from the natural microbial community present as "background" in environmental samples. Microbial background may include naturally occurring pathogen strains present in the environment. The four pathogens surveyed in the current work, *Bacillus anthracis*, *Clostridium perfringens*, *Yersinia pestis*, and *Francisella tularensis*, occur endemically in the United States and are responsible for natural disease outbreaks (1–7). Equally important, the environmental background may include bacterial species that are closely related to target pathogens. Such species can share considerable genomic, proteomic, physiologic, and sometimes pathogenic traits with target pathogens. This is exemplified by recent genomic comparisons among *Bacillus* species (8–12) and *Yersinia* species (13,14). Closely related species that carry genomic and physiologic traits in common with the pathogen may cross-react in pathogen detection assays. Some of these close relatives may indeed be pathogenic.

The polymerase chain reaction (PCR) has been used extensively for over a decade to detect and characterize bacterial diversity in environmental samples. Highly sensitive PCR detection of target species has been attained through the use of specific primers, and this has been applied as the foundation for DNA-based detection systems for biothreat pathogens (15,16). In principle, such systems offer the capability to rapidly and accurately detect the presence of pathogenic organisms in environmental samples. Assays designed for environmental surveillance or forensic activities must be both sensitive and specific for a target pathogen species or strain in the presence of the natural microbial background. Given that the majority of bacterial species in the environment remain unidentified and uncultivated (17), there is potential for cross-reaction in these assays that has not yet been characterized.

For this study, PCR techniques were used in a broad-sweep environmental survey to begin to understand the distribution and diversity of four bacterial pathogens and their close relatives in a variety of environmental samples from across the United States. PCR primers were designed that amplify 16S rRNA genes from four Category A potential biological terrorism agents (18); *B. anthracis*, *C. perfringens*, *Y. pestis*, and *F. tularensis* and their closely related species. This pathogen+close relatives approach allows identification of novel, environmentally relevant strains as well as known cultured isolates. In parallel, PCR primers targeting pathogen-specific virulence genes were used to survey for pathogenic potential. Each primer set was used to amplify DNA extracted from 89 diverse soil samples and over 15,000 aerosol samples from 15 major U.S. cities. Cloning and sequencing of products from positive amplification reactions was conducted to determine the identities of species in the environmental samples and their relationship to the target pathogens.

Materials and Methods

Soil Collection, DNA Extraction, and Quantification

Soil samples (*c.* 10–50 cm^3) were collected from 0 to 10 cm surface soils or shoreline sediments (of ponds/lakes). Single grab samples were collected into 50 mL conical plastic tubes and stored at room temperature or refrigerated until shipment to the laboratory (<5 days). Upon receipt, soil DNA was stored frozen at −70°C before extraction. The 129 sample set included samples from 32 United States and the District of Columbia, and repre-

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545.

*This research was supported by grants to CRK from the U.S. Department of Energy Chemical & Biological Nonproliferation Program and the Department of Homeland Security. Los Alamos Unclassified Report: LAUR # 05-1744 (portions of this work have previously been orally presented at granting agency program reviews and in poster form at the American Society for Microbiology, 2003).

Received 25 June 2005; and in revised form 8 Oct. 2005; accepted 22 Oct. 2005; published 21 April 2006.

sented all major regions of the United States. Soils/sediments were collected from urban, suburban, agricultural, desert and forested areas, as well as lake and river shores and ocean beaches.

Nucleic acids were extracted and purified from duplicate 0.5 g aliquots of each soil sample using 70°C incubation in SDS-containing baffer followed by bead mill homogenization and centrifugation as described by Kuske et al. (19). Nucleic acids were precipitated from the DNA extract solution using 0.1 volume 3 M sodium acetate (pH 5.2) and two volumes ethanol, incubation at −20°C, and centrifugation for 30 min at 12,000 × *g* at 4°C. Precipitated nucleic acids were suspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), and the two replicate samples were pooled. DNA concentration was estimated in ethidium bromide-stained 3% SeaKem agarose gels (FMC Bioproducts, Rockland, ME) using lambda DNA as a calibration standard, and quantified by digital image analysis (Science Lab 99 Image Guage v.3.2, FujiFilm Corp., San Jose, CA). DNA was diluted to *c*. 2 ng/μL in TE, and purified away from contaminants using Sephadex G-200 spin columns (in 96-well coarse polypropylene filter plates; Advanced Genetic Technology Corp., Gaithersburg, MD) equilibrated in TE, as described (19). The clear column eluate containing DNA was diluted 1:10 in TE (to *c*. 200 pg/μL) and used as template for PCR. Blank samples were prepared using extracton buffer alone, with no sample addition, and proceeding with all extraction and purification steps described.

Aerosol Collection, DNA Extraction, and Quantification

Aerosol filters were obtained from U.S. Environmental Protection Agency state air quality monitoring agencies. Each aerosol sample consisted of 3/4 of a 47-mm-diameter teflon filter containing particulates (the "respirable" fraction, ≤ 2.5 μm diameter; $PM_{2.5}$) collected continuously over an *c*. 23 h interval (information on Environmental Protection Agency (EPA) standard collection instrumentation and methods can be found at the EPA Web site: http://www.epa.gov/ttn/amtic/). Blank filters were also provided as negative controls. Filters were collected and stored according to standardized EPA methods (http://www.epa.gov/ttn/amtic/). In most cases, samples were stored dry for *c*. 1 year at 4°C by state air quality monitoring laboratories, then shipped to Los Alamos National Laboratory. A total of 15,108 aerosol filters, from 51 urban air monitoring locations in 15 major U.S. cities were used in this survey. These represented time spans of 1–3 years and three to six sites in each of the cities. As many samples as possible were obtained from each city, but sample sets covering the same calendar period from different cities were not routinely available. Thus, the overall sample set varies by sampling time and sampling frequency among sites (either daily sampling, 1-in-3-day sampling, or 1-in-6-day sampling, depending on the EPA air quality monitoring procedures for that city).

For each sampling location, individual $PM_{2.5}$ filters were pooled into groups representing monthly sampling periods. Locations with a daily sampling schedule therefore had a maximum of 31 filters per monthly period, while locations with a 1-in-3-day or 1-in-6-day sampling schedule had a maximum of 10 or five filters per monthly period. Thirty random quality control blanks were also selected for each city and combined. This pooling resulted in a total of 744 combined samples (729 collected samples + 15 blank sets). Nucleic acids were extracted from the pooled filters by placing them into 50-mL conical tubes with a maximum of 16 filters per tube. Monthly pooled samples consisting of more than 16 filters were extracted as two replicate samples and combined after extraction. To wash cells from the filter surfaces, 30 mL of 100 mM phosphate buffer (pH 7.4) containing 0.05% (v/v) Tween 80 and freshly prepared 10% sodium pyrophosphate solution was added. Tubes were vortexed for 30 s and shaken on a rocking platform shaker for 15 min. The vortexing and shaking were repeated four times for a total of 1 h of washing. To concentrate microbial cells, tubes were centrifuged at 3200 × *g* for 10 min at 5–10°C in a Beckman GS-6KR centrifuge (Beckman Coulter, Inc., Fullerton, CA) using a swinging bucket rotor. Filters were removed with forceps. Excess supernatant liquid was removed by pipet. The cell pellet was suspended in 750 μL of TE buffer and transferred to a sterile 2-mL screw cap tube containing 500 mg each of 106- and 500-μm diameter zirconia-silica beads (Biospec Products, Bartlesville, OK). Cells were disrupted for 2 min in a mini bead-beater (Biospec Products), homogenates were centrifuged at 16,000 × *g* for 4 min, and *c*. 500 μL of supernatant was transferred to a sterile 2-mL tube. To increase DNA recovery, 300 μL of sterile TE buffer was added to the bead beater tube, mixed thoroughly by inversion, centrifuged at 16,000 × *g* for 4 min, and this supernatant was combined with the original supernatant. To the *c*. 800 μL of supernatant, 250 μL of 4 M guanidine thiocyanate and 100 μL of 100 mM phenacyl thiazolium bromide were added to remove PCR inhibitors (20). Samples were mixed by vortexing. Solutions were then incubated at 37°C overnight (about 18 h), centrifuged for 10 min and the supernatant transferred to a fresh tube. DNA was precipitated by addition of 1 μL GenElute LPA (Sigma, Aldrich, St. Louis, MO), sodium acetate (pH 5.2), and isopropanol, and centrifuged. The resulting pellet was washed with cold 70% ethanol, and suspended in 50 μL of 10 mM Tris pH 8.0.

Preliminary results indicated that extracts from single filters were unlikely to yield sufficient DNA for quantification. Therefore, aliquots of extracts were pooled by site, to yield a total of 25 μL of DNA extract per site for each of 12 cities. Aerosol DNA was quantified using PicoGreen intercalating dye. Ten microliters of undiluted pooled extract was analyzed in 250 μL of PicoGreen reagent (Molecular Probes, Eugene, OR), according to manufacturer's protocols, using 0.1–10 ng lambda DNA reference standards, in a spectrofluorometer equipped with a 150 W Xenon source (Spex FluoroLog-2, Jobin Yvon Inc., Edison, NJ; excitation at 502 nm, emission at 520 nm).

Design and Testing of 16S rRNA gene PCR Primers

Primers for amplification of 16S rRNA genes from each of the four pathogen+close relatives groups were designed *de novo* for this study, or were modified from primers described previously in the literature. Gene regions, primer sequences and target groups are listed in Table 1. In all cases, putative primers were compared with manual alignments of all available target sequences obtained from the GenBank (http://www.ncbi.nlm.nih.gov/) and Ribosomal Database Project databases (http://rdp.cme.msu.edu/ (21)) and were screened for specificity using the Probe Match program of the RDP or BLAST 2.0 (NCBI; http://www.ncbi.nlm.nih.gov/blast/ (22)).

In preliminary surveys with a PCR primer set designed for the genus *Bacillus*, we detected *Bacillus* species in nearly every soil sample in the collection (data not shown). For this reason, the phylogenetic 16S rRNA group primer set for *B. anthracis* and closely related species was confined to the *B. cereus* subgroup. Ribosomal RNA gene primers for the *B. cereus* group were predicted to be specific for amplification of the rRNA gene from *B. anthracis*, *B. cereus*, *B. thuringiensis*, *B. medusa*, and *B. mycoides*, to the exclusion of all other *Bacillus* species. The phyloge-

TABLE 1—*PCR primers used in this study.*

Target Group	Target Gene (16S rRNA Gene Target Region)	Forward and Reverse Primers	Product Size (bp)*	Annealing Temperature (°C)	References
All bacteria	16S rRNA (8–806; *E. coli* nts)	27F.1: AGRGTTTGATCMTGGCTCAG 787Rb: GGACTACNRGGGTATCTAAT	798 (*E. coli*)	55	F: Modified from (63) R: This study
Bacillus cereus group	16S rRNA (452–1090)	Bc471F: CAAGTGCTAGTTGAATAAGC Bc1100R: ACCCAACATCTCACGAC	640	60	This study
Bacillus anthracis	*pag* Protective antigen	Pag-110F: GCCGTCCCAGTATTTACATA Pag-110R: ATTAATGGCATTGTCTACGAT	1176	55	This study
Yersinia species	16S rRNA (603–1305)	GA-YP2F: TGTGAAATCCCCGCGCTTAACG GA-YP1R: CCGGACTACGACAGACTTTATGT	704	60°	(64)
Y. pestis	*caf 1* F1 capsular antigen	Caf1-F: CCCGCATCACTCTTACAT Caf1-R: ACGGTTACGGTTACAGCA	388	55	This study
Francisella species	16S rRNA (137–1299)	Fr153F.1: GCCCATTTGAGGGGGATACC Fr1281R.1: GGACTAAGAGTACCTTTTTGAGT	1168	60°	Modified from (26,65)
Francisella tularensis	*tul4* 17-kDa membrane protein precursor	Tul4-435F: GCTGTATCATCATTTAATAAACTGCTG Tul4-863R: *TTGGGAAGCTTGTATCATGGCACT*	407	60	(30)
Clostridium perfringens	16S rRNA (127–666)	Cp145F: GGTAACCTGCCTCATAGAGT Cp647R: CTCTCCTGCACTCTAGATAA	520	56	This study
Enterotoxigenic *Clostridium perfringens*	*cpe* Enterotoxin gene	*cpe*F: GGAGATGGTTGGATATTAGG *cpe*R: GGACCAGCAGTTGTAGATA	233	55	(27)

*Predicted PCR amplicon size from target pathogen species (*B. anthracis*, *Y. pestis*, *F. tularensis*, or *C. perfringens*).

netic group primer sets for the other three pathogens were designed more broadly. Primers for *C. perfringens* were designed using the PRIMROSE program (23), and predicted to be specific for all strains of that species for which sequences are available, including sequences from some uncultivated environmental samples that cluster within the group. *Francisella* 16S rRNA gene primers targeted all *Francisella* species, including *F. philomiragia*, but also were predicted to amplify DNA from *Wolbachia persica*, a closely related species, as well as from related insect symbionts (24–26). Primers for *Yersinia* species were designed to specifically amplify rRNA gene fragments from all *Yersinia* species, except *Y. ruckeri*.

Optimal PCR conditions for each primer pair and detection sensitivity were tested using genomic DNA from the target pathogen species (*B. anthracis* 91-429C-2, *Y. pestis* CO92 biovar *orientalis*, *F. tularensis* subspecies *holarctica* LVS (ATCC 29684) and *F. tularensis* subspecies *holarctica* 99A-9419, *C. perfringens* ATCC 27324), and DNA from related, non-target species (*B. flexus* (ATCC 49095), *B. subtilis*, *C. botulinum* type A (ATCC 17862), *Y. rohdei* (ATCC 43380), *Y. ruckeri* (ATCC 29473), *Hafnia alvei* (ATCC 13337), *Proteus vulgaris* (ATCC 13351), *Escherichia coli* W3110).

Design and Testing of Virulence Gene Primer Sets

PCR primers from characterized virulence genes were designed for *B. anthracis* (protective antigen gene, *pag*) and *Y. pestis* (F1 capsular antigen, *caf1*). These primer sets were tested against a panel of bacterial DNAs that included the four pathogens, two to three closely related species or genera for each pathogen, human DNA and two soil DNAs. A primer set developed by Meer and Songer (27) that has been used for *C. perfringens* (enterotoxin gene, *cpe*) detection in foods (28) was used (Table 1). The virulence mechanisms and corresponding genes of *F. tularensis* are not characterized (29), and we used a primer set developed by Sjöstedt et al. (30) for the *tul4* gene that has been widely used in clinical DNA-based assays (30,31). In addition to the testing mentioned above, and presented in the previous literature, the specificity for each primer set was screened by BLAST analysis to database sequences.

PCR Amplification

The optimal annealing temperature for each primer pair was empirically determined (Table 1). All samples were initially amplified with 27F/787Rb primers, designed to amplify 16S rRNA gene fragments from all bacteria, to test the ability of the soil or aerosol DNA to support PCR. For each primer pair, 25 μL PCR reactions contained 10 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 μM each primer, 0.94 U of Ampli*Taq* LD polymerase (Perkin-Elmer, Applied Biosystems, Foster City, CA), 5 μg BSA (Boehringer Mannheim), and 1 μL (*c.* 200 pg) soil or 1 μL aerosol DNA.

Cycling conditions were as follows: 4 min denaturation at 94°C; annealing at the appropriate temperature (Table 1) for 45 s, 72°C for 60 s, and 94°C for 30 s; for 45 cycles for aerosols/40 cycles for soils, and a final cycle of 55°C for 45 s and 72°C for 5 min (20 min for reactions to be cloned), carried out in a PTC-200 thermal cycler (MJ Research, Waltham, MA). Positive control reactions using DNA from target pathogens (1 and 0.1 pg), and negative control reactions, without template DNA, were included in each PCR set. Pathogen DNAs were added to control reactions only after all experimental reaction tubes were sealed, to prevent false positives due to contamination. A 5-μL aliquot of each reaction mixture was analyzed on 1%, 2%, or 3% (depending on expected product size) SeaKem agarose gels (FMC Bioproducts), and DNA visualized by ethidium bromide staining and UV transillumination. Putatively positive reactions were repeated at least once to confirm the results and obtain products for cloning. Only products matching the size of the positive control fragment were considered "positive" for these assays. Amplification reactions using the *Y. pestis caf1* and *B. anthracis pag* primer pairs were multiplexed, using 0.1 μM of each primer.

The soil collection contained multiple samples for some collection sites and single samples from others. For PCR analysis with group-specific primers, one sample was chosen from each collection site, resulting in 89 soil samples that represented the diversity of geographic distribution, environment and soil type in the collection, that could be analyzed in 96 well format (with controls).

Cloning and Sequencing of PCR Products

PCR products of the correct size were purified by electrophoresis in SeaKem agarose gels. The DNA fragments were excised and purified using a Qiaex DNA Purification Kit (Qiagen Inc., Chatsworth, CA). Products were cloned into the pCR4 vector, using the TOPO-TA cloning kit and manufacturer's protocols (Invitrogen, San Diego, CA). Clones were picked and stored in glycerol medium for sequencing. Thirty-six to 96 clones per amplified PCR product were sequenced. Plasmid DNA was isolated from overnight cultures using a solid phase reverse immobilization procedure (32), and inserts were sequenced using the M13 forward primer or a PCR primer, with the BigDye Terminator cycle sequencing reagents (v. 3.0, Applied Biosystems). Sequencing reactions were analyzed on ABI 3700 and 3730 automated sequencers (Applied Biosystems).

Sequence Data Analysis

Raw data were analyzed using Sequence Analysis (Applied Biosystems) and Sequencher software (Gene Codes Corporation, Ann Arbor, MI). Initial comparisons were made to database sequences using the RDP Sequence Match program (RDP; http://rdp.cme.msu.edu/ (21)) and BLAST 2.0 (NCBI; http://www.ncbi.nlm.nih.gov/blast/ (22)). Translated sequences were analyzed using BLASTX (NCBI). For phylogenetic analyses, sequences were obtained from databases and aligned using Clustal X (33) and/or a Web-based Sequence Aligner Program (RDP), with final alignment accomplished manually using the GDE multiple sequence editor (RDP). Phylogenetic trees were inferred from aligned data using maximum likelihood analysis (fastDNAml version 1.1; distributed by RDP (34)).

Results

DNA Concentrations in Soil and Aerosol Samples

Soil DNA concentrations ranged over at least three orders of magnitude (from 0.2 to 146 μg DNA/g soil; Fig. 1). Most ($n = 75$) samples had yields between 1–10 μg/g soil; the median DNA concentration was 13.2 μg, and 25% and 75% quantiles were 3.7 and 65.1 μg/g soil, respectively. DNA quantity did not appear to be correlated with general soil texture, ecological setting, or geographic region.

Aerosol DNA samples from 12 cities were quantified using fluorometric analysis with PicoGreen intercalating dye. DNA was not detected in Atlanta samples, suggesting that DNA concentrations for these samples were < 5 pg/μL (the detection limit for this assay), although these samples did produce observable products in standard PCR. Estimates from PicoGreen analyses indicate that extracts from aerosol samples contained < 0.25–1.22 ng (mean yield = 0.71 ng) total DNA per pooled monthly sample.

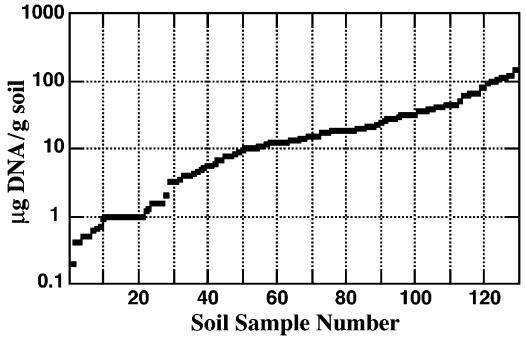

FIG. 1—*Concentration of DNA extracted from 129 soil samples from a wide variety of soil types collected across the United States. Each square corresponds to a single soil sample. The median yield was 13.2* μg *DNA/g soil, and the DNA yield from individual soils varied over three orders of magnitude.*

Detection Sensitivity of PCR Primer Sets

All primer sets amplified detectable PCR products with target DNA quantities of ≥ 0.1 pg/25 μL reaction volume, except the *F. tularensis* Tul4 set, which was only capable of detecting > 10 pg of *F. tularensis* DNA, and thus was 100 times less sensitive. The 0.1 pg template DNA represents 17–46 genome equivalents for the target bacterial species. Spiking experiments with target sequences added to PCRs containing soil and aerosol DNA confirmed this level of sensitivity in our collection of environmental templates (data not shown). In initial trials of DNA extraction and purification methods for these aerosol samples, the phenacyl thiazolium bromide treatment was found to effectively remove inhibitors from the EPA aerosol samples, and improved PCR amplification of bacteria in natural and spiked aerosol DNA samples (e.g., Fig. 2).

PCR Amplification of Soil DNA Extracts with General and Specific Primers

The soil DNAs produced easily detectable PCR amplicons of the correct size using the bacterial 27F/787Rb 16S rRNA gene PCR primers (Table 2), indicating that the extracts were sufficiently purified and contained adequate bacterial DNA to support DNA amplification. Positive reactions in soils with the group-specific 16S rRNA gene and virulence gene primer sets are tallied in Table 2. All primer pairs, with the exception of those targeting *Francisella* species and the enterotoxin gene of *C. perfringens*, amplified products of the correct size with one or more soil samples.

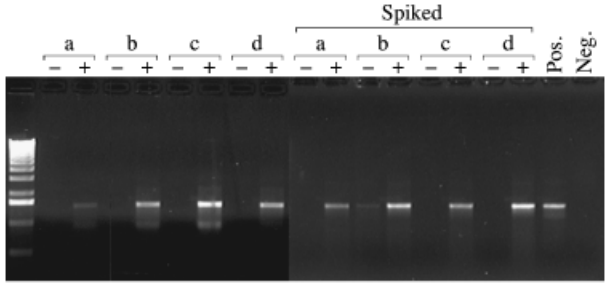

FIG. 2—*Effect of phenacyl thiazolium bromide (PTB) on ability of DNA extracted from aerosol samples to support PCR. Shown in lanes 2–9 are PCR products amplified using general bacterial primers 27F/1492R, from DNA of four EPA aerosol samples (a–d) extracted with or without PTB (−, +). A product was amplified from each sample after PTB precipitation but not before treatment. To verify that the negative results in the untreated samples were due to inhibitors, the same samples were spiked with 100 pg of* Bacillus anthracis *DNA (lanes 10–17). Untreated samples with the* B. anthracis *spike did not amplify. Lanes 18 and 19 are PCR-positive and -negative controls.*

TABLE2—*PCR results for soils and EPA aerosol samples screened with primer pairs for four pathogens and close relatives.*

EPA Aerosol Sample Data					PCR-Positive Results by Target Species and Gene							
City	Sampling Dates	No. of Locations	No. of Filters	No. of Pooled Samples	All Bacteria 16S rRNA Gene	*Bacillus cereus* Group 16S rRNA Gene	*Bacillus anthracis pag*	*Clostridium perfringens* 16S rRNA Gene	Toxigenic *Clostridium perfringens cpe*	*Francisella* Species 16S rRNA Gene	*Francisella tularensis tul4*	*Yersinia* Species 16S rRNA Gene	*Yersinia pestis caf1*
Albuquerque	Jan–Dec 01	2	377	24	20 (83%)	4	0	1	0	0	0	0	0
Atlanta	Jan–Dec 00	3	834	35	28 (80%)	16	0	0	0	0	0	0	0
Chicago	Jan 98–Dec 01	3	2693	108	73 (68%)	48	10	0	0	0	0	0	0
Dallas	June 99–June 00	3	1000	36	34 (94%)	1	0	0	0	0	0	0	0
Denver	Jan 00–Dec 01	5	1388	96	66 (69%)	12	0	2	0	21	0	0	0
Houston	June 99–June 00	3	968	39	26 (67%)	12	0	0	0	0	0	0	0
Los Angeles	Jan–Sept 00	3	773	27	18 (67%)	6	1	0	1	0	0	0	0
Miami	Jan–Dec 00	3	1019	36	31 (86%)	11	0	0	0	0	0	0	0
Nashville	Jan–Dec 00	3	729	36	18 (50%)	2	1	0	0	0	0	0	0
New York	Jan–Dec 00	3	865	42	36 (86%)	26	0	1	0	0	0	0	0
Phoenix	Jan–Dec 99	4	894	46	41 (89%)	3	0	0	2	0	0	0	1
San Diego	Apr 00–June 01	3	773	35	29 (83%)	9	0	1	0	1	0	0	0
San Francisco	Aug 00–Sept 01	4	798	56	20 (36%)	0	0	0	0	0	0	0	0
Seattle	Sept 98–Dec 99	6	1232	67	61 (91%)	28	1	0	1	0	0	0	0
Washington, DC	May 99–Aug 00	3	765	46	11 (24%)	7	1	0	0	0	0	0	0
Total		51	15108	729	512 (70%)	185	14	5	4	22	0	0	1
Soil sample data													
Total	Jan 95–Jan 01	89		89	89 (100%)	32	2	6	0	0	0	3	1

PCR Amplification of Aerosol DNA Extracts with General and Specific Primers

Approximately 70% of the 729 urban aerosol DNA extracts produced amplicons during PCR amplification using the 27F/787Rb PCR primers (Table 2). Failure to support PCR in 30% of the extracts is most likely due to the very low DNA concentrations in these samples, as the phenacyl thiazolium bromide clean-up step was found to effectively remove PCR inhibitors in the aerosol samples (Fig. 2). In the aerosol extracts, six of the eight primer sets (all except the rRNA gene primer pair targeting *Yersinia* species 16S rRNA genes and the *tul4* gene of *F. tularensis*) produced amplicons from one or more samples. Positive PCRs were sporadic across the aerosol sample set, and seasonal or within-city location trends were not identified due to the low frequency of detection. Between-city comparisons are confounded by the different sampling regimes in each city and the different time spans for which we received samples. Therefore, the results in Table 2 are summarized across time and within-city monitoring locations for each city.

PCR Product Cloning and Sequencing

To characterize and confirm the identity of the PCR products obtained from the soil and aerosol samples, clone libraries were constructed from many amplified products from different primer sets and environmental samples (Table 3). Altogether, 1255 sequences, averaging 600 nucleotides in length, were of sufficient quality to be analyzed.

Sequence analysis of clone libraries from single amplified products generally produced sets of highly similar sequences, differing from one another by a single nucleotide change. The frequency and type of most of these changes are consistent with the known, common errors of Taq DNA polymerase (data not shown) (35,36). Additional analyses of these single nucleotide changes with respect to predicted regions of variability of the target gene (hypervariable regions of 16S rRNA gene sequences and third-codon positions of *pag*, *caf1*, and *cpe*; data not shown) showed that the changes did not correspond to expected variable positions. This result indicates that most of these single base changes were probably introduced during amplification, although some may have arisen from heterogeneity between sequences in the sample DNAs (37).

B. cereus *Group in Soils and Aerosols*

Thirty-six percent of the soil samples tested positive for the *B. cereus* group (*B. anthracis*, *cereus*, *thuringiensis*, *medusa*, *mycoides*) using primer sets specific for this group (Table 2). Although *Bacillus* species sequences outside the target group were detected in some samples, all soil clone libraries contained sequences from the *B. cereus* group (Table 3). This primer set produced products with 25% of the aerosol samples (Table 2), and all sequences obtained were from the target group. Our results indicate the presence of close relatives of *B. anthracis* in a wide variety of soil/sediment samples from different geographic regions of the United States, and in urban aerosols from 14 of the 15 cities surveyed. Positive rates for individual cities ranged from 0.5% to 62% of the monthly aerosol samples (Table 2).

Sequences matching the *B. anthracis* protective antigen (*pag*) gene were detected in two soil and 14 aerosol samples (Tables 2 and 3), indicating the presence of either *B. anthracis* or another organism carrying a highly similar gene. The *pag*-positive soil samples were also positive for *B. cereus* group 16S rRNA gene sequences (not shown). Sequences obtained from one soil sample

TABLE 3—*Sequencing results for soil and aerosol samples amplified with 16S rRNA phylogenetic and virulence gene primer sets.*

Primer Target	Gene	Sample Type/ Source*	No. of Sequences Obtained	No. (%) Sequences in Target Group	No. (%) Matching Pathogen Sequence	Comments
Bacillus cereus group	16S rDNA	S-NM	34	34 (100%)	8 (24%)	Twenty-six (76%) of several sequence types, all >99% identical to *B. anthracis*
		S-AZ	68	60 (89%)	2 (3%)	Non-*B. cereus* group sequences were mostly *Cytophagas*
		S-FL	20	19 (95%)	0	One non-*B. cereus* group *Bacillus* sequence
		S-CO	39	6 (15%)	0	Several non-*Bacillus* genera present in library
		S-NY	87	64 (74%)	0	Several non-*Bacillus* genera present in library
		S-DC	70	23 (33%)	0	Several non-*Bacillus* genera present in library
		A-L.A.	19	19 (100%)	19 (100%)	
		A-Chicago	38	38 (100%)	1 (3%)	Thirty-seven (97%) identical to *B. cereus* ATCC 15479
		A-Chicago	43	43 (100%)	0	Thirty-two (74%) identical to *B. cereus* ATCC 15479; 11 (25%) identical to *B. mycoides*
		A-NYC	41	41 (100%)	12 (29%)	Twenty-nine (71%) of several sequence types, all >99% identical to *B. anthracis*
		A-Seattle	40	40 (100%)	0	Thirty-six (90%) identical to *B. cereus* ATCC 15479; 4 (10%) identical to *B. mycoides*
		A-Nashville	38	38 (100%)	0	Thirty-seven (97%) identical to *B. cereus* ATCC 15479; 1 (3%) other *Bacillus* spp.
B. anthracis	*pag*	S-NM	9	9 (100%)	0	All sequences have twp silent mutations vs. *pag*
		A-L.A.	27	27 (100%)	27 (100%)	
		A-Chicago	29	29 (100%)	29 (100%)	
Clostridium perfringens	16S rDNA	S-CO	27	0	0	Four sequence types, all *Clostridia*
		S-AR	38	0	0	Several sequence types, all *Clostridia*
		A-San Diego	26	26 (100%)	8 (31%)	18 (69%) sequences differ from *C. perfringens* by 1 nt
		A-NYC	25	0	0	Four sequence types, all *Clostridia*
		A-Denver	27	27 (100%)	0	One sequence type with four differences from *C. perfringens*
		A-Albuquerque	37	37 (100%)	2 (5%)	Thirty-five (95%) sequences (several types) >99% identical to *C. perfringens*
C. perfringens	*cpe*	A-Phoenix	34	34 (100%)	0	All sequences have three silent mutations vs. *cpe*
		A-L.A.	37	37 (100%)	37 (100%)	
		A-Seattle	39	39 (100%)	0	All sequences have three silent mutations vs. *cpe*
Francisella spp.	16S rDNA	A-San Diego	7	7 (100%)	7 (100%)	
		A-Denver	35	35 (100%)	35 (100%)	
		A-Denver	19	19 (100%)	19 (100%)	
		A-Denver	35	35 (100%)	35 (100%)	
		A-Denver	51	51 (100%)	51 (100%)	
		A-Denver	22	22 (100%)	22 (100%)	
		A-Denver	39	39 (100%)	35 (90%)	Four other types; all >99% identical to *F. tularensis*
		A-Denver	38	38 (100%)	24 (63%)	Two other types; all >99% identical to *F. tularensis*
		A-Denver	35	35 (100%)	21 (60%)	Three other types; all >99% identical to *F. tularensis*
Yersinia spp.	16S rDNA	S-WI	30	30 (100%)	0	All 100% match to *Y. rohdei*
		S-HI	10	0	0	All sequences are low matches to *Alteromonas* spp.
Y. pestis	*caf1*	A-Phoenix	42	42 (100%)	12 (29%)	Thirty (71%) sequences have one silent mutation vs. *caf1*

*S, soil sample; A, EPA aersosol samples; two letter U.S. state abbreviations used for soils.

were all of one type, and differed by two nucleotides from the *B. anthracis* Sterne *pag* sequence (AF065404) (Table 3). These two single nucleotide polymorphisms (SNPs) are "silent" (data not shown), and do not match any of those identified by Price et al. (38) among known isolates of *B. anthracis*. Protective antigen gene PCR products were obtained from aerosol samples from five cities, and clone/sequence analysis of two of the amplified products revealed sequences that exactly matched the *B. anthracis pag* gene sequence. Corresponding analysis of 16S rRNA gene clone libraries from these two samples revealed *B. anthracis*-like rRNA gene sequences in each case (Table 3), suggesting that the pathogen or a very close relative containing all/part of pX01 may be present in these samples.

C. Perfringens *in Soils and Aerosols*

Positive amplification results using the *C. perfringens* 16S rRNA gene phylogenetic group primers were obtained from 11 soils and aerosol samples from four cities (Table 2). Sequencing results from soil and aerosol PCR reactions indicate that the phylogenetic group primer set amplified *C. perfringens* rRNA gene genes, but also amplified other *Clostridium* species in addition to *C. perfringens*. Clone libraries from two soil samples contained *Clostridium* species sequences, but no sequences highly similar to that of *C. perfringens* (Table 3). Three of the clone libraries generated from aerosols produced sequences >99% identical to that of *C. perfringens*, while a fourth sample contained only sequences similar to *C. favososporum*.

The *C. perfringens* enterotoxin gene (*cpe*) sequence was not amplified from any of the soil samples, but was amplified from four aerosol samples from three cites. Sequencing of cloned products from two of these samples produced only sequences identical or highly similar to that of the *cpe* gene (Table 3). Sequences from a Phoenix and a Seattle sample were identical, and differed from the wild-type *cpe* gene sequence at three positions. All are predicted to be silent mutations (not shown).

Francisella Species in Soils and Aerosols

F. tularensis and closely related species were not detected in any of the soil samples in this study, but were found in aerosols from two cities (Table 2). Sequence analysis of 16S rRNA gene clones from eight aerosol samples showed that all were dominated by sequences identical to that in *F. tularensis*, while three also contained two to three additional very similar sequence types, differing by a few nucleotides from each other (Table 3). Despite the presence of 16S rRNA gene sequences matching those of *F. tularensis* in several samples, no products were obtained from any reactions using the Tul4 primer set. The relatively low sensitivity of this primer pair (>10 pg DNA) however, complicates interpretation of the negative results.

Yersinia Species in Soils and Aerosols

Only three soil samples, and none of the aerosols, produced positive PCR products using 16S rRNA gene primers for the *Yersinia* (Table 2). Sequence analysis of cloned products was conducted for two of the samples. Analysis of clones from one soil indicated the presence of strains with 16S rRNA sequences identical to *Y. rohdei*, *Y. mollaretii*, and *Y. kristensenii* (Table 3). However, sequence analysis of clones from a second soil indicated that the 16S rRNA gene primers also amplified rRNA gene genes from other gammaproteobacterial species. No *Yersinia*-like sequences were obtained from this library.

One soil and one aerosol sample produced amplicons of the expected size when amplified using the Caf1 primer set (Table 2). Seventy-one percent of the sequences obtained from the aerosol sample differed from the database *caf1* sequence by one nucleotide (predicted to be a silent mutation; not shown), whereas the remainder matched the database sequence exactly (Table 3). However, neither of these samples produced amplicons when reactions contained the *Yersinia* 16S rRNA gene primers, suggesting that *Yersinia* species are not present or were below detection limits. The *caf1* gene encodes the F1 capsular antigen that comprises the *Y. pestis* capsule and functions as an adhesin that is critical for *Y. pestis* virulence (39). Our results suggest it may be related to capsule proteins from other bacteria in environmental samples. Whether these samples contained the *Y. pestis caf1* gene or a homologous gene in another organism remains in question.

Discussion

This study is an initial survey for four bacterial pathogens and their close relatives in a large number of soil and aerosol samples from diverse environments and regions of the United States. An objective of this study was to develop and test broadly inclusive phylogenetic primer sets, capable of detecting (but not necessarily discriminating) the target pathogen and its closely related species, to determine the presence of natural pathogen strains as well as closely related environmental species that may share genomic traits, but are not currently represented in culture collections. The broadly inclusive 16S rRNA surveys were coupled with clone sequence analysis assays designed to detect virulence factors or other genetic traits that have been used for diagnosis.

Using this approach, novel 16S rRNA gene sequences of potential new species in the target genera were detected for all four pathogen groups. This result was not surprising, since existing culture collections are very biased for clinical isolates, while most environmental bacterial species, even those related to pathogens of interest, have not yet been cultured or characterized. The database of 16S rRNA sequences, although it contains $>100{,}000$ bacterial sequences (RDP; http://rdp.cme.msu.edu/), is still limited in its coverage of the perhaps millions of bacterial species thought to exist (40). Thus, design of specific primers is challenged by the vast diversity of as-yet unknown species present in most environmental samples, for which no sequences are available in the database.

Detection of 16S rRNA gene sequences outside the original design group was more common in soil DNA than in aerosol DNA, which is likely a reflection of the greater diversity and abundance of bacteria in soils than in air. When tested against soil samples that probably contained many thousands of bacterial species, the *Francisella* species 16S rRNA gene primer pair was specific for its target group (26). However, the *B. cereus*, *Yersinia*, and *C. perfringens* group primer sets also detected other sequences in the respective genera that were outside of their original design groups, and the *Yersinia* group primers occasionally detected other genera (Table 3). In contrast to their performance in soils, when tested in aerosols all four 16S rRNA primer sets detected only sequences within the target groups (with one exception, Table 3).

Additional opportunity for cross-reactivity of the 16S rRNA primers arises from the high level of primary sequence conservation of the rRNA molecule in all organisms. Our analysis of environmental sequences obtained with these primer sets suggests that when the target group is present in the sample, it is detected. However, the primers may additionally amplify sequences outside the original design group when target species are not present or are low in abundance.

The virulence gene primer sets for three of the four pathogens amplified products in the environmental sample sets and positive amplification reactions were specific for the target genes in both soils and aerosols. All analyzed sequences were $>99\%$ identical to that of the target pathogen. A few of the libraries were dominated by sequences identical to that of the pathogen gene present in the database. However, the majority of the sequences obtained differed from the pathogen sequences by one to three (silent) mutations, indicating that homologues of the target genes are present in these samples. The presence of the target virulence gene sequences was not necessarily correlated with the presence of the target species (by 16S rRNA detection), suggesting that these genes may be present in an alternate host. Although the virulence gene primer sets specifically amplified target gene sequences in environmental samples, it is not possible to conclude that a pathogenic organism is present in the sample based on environmental detection of a single virulence gene. Virulence genes are defined by their activity during the infection process, and they and their sequence homologues may have other functions unrelated to virulence. In the absence of culture-based assessment or other supporting information, evidence solely from DNA-based detection must be interpreted with caution.

The combined use of PCR with phylogenetic and virulence gene primers followed by clone/sequence analysis revealed the presence of a variety of bacteria with identical or very similar rRNA gene sequences to the four target species, as well as the presence of genes that matched or were very similar to sequences of pathogen virulence genes. In the complex environmental background described here, reliable detection of the pathogen will require multiple assays that target different genome areas and include pathogenicity attributes. In practice, obtaining a culture of the detected organism that can be tested for pathogenicity remains the gold standard. Toward that end, attempts to culture the environmental species detected in this study could provide

improved comparative materials for development of specific detection methods for forensic and surveillance activities.

The ability to detect a target DNA molecule using PCR depends on the abundance of that target DNA relative to nontarget DNA in the sample used at PCR template. In an environmental sample containing large amounts of nontarget DNA, ability to detect the target can be reduced due to the limited amount of material one can use as a template in PCR. Concentrations of DNA obtained from the soil and aerosol samples in this study varied enormously. Most of the soils in this study yielded microgram amounts of DNA, and could be considered high-DNA samples. To achieve optimal *Taq* polymerase kinetics in PCR assays (19), these samples required dilution of the original extracted DNA. In a high-DNA sample, a target sequence that represents a small proportion of the total sample can be missed due to the dilution of the original template but may be detectable in a sample with lower overall DNA concentration. At the other end of the spectrum, most of the aerosol samples in this study contained only picogram amounts of DNA and may have contained insufficient target even though the amount of extraneous DNA was very low. It is possible to miss a target DNA sequence in environmental PCR reactions, due to dilution of high-DNA samples (too much nontarget template DNA), lack of total DNA (not enough template) from the sample, or presence of inhibitors. The aerosol samples had been stored before analysis, and this may have affected cell integrity and DNA yield. We therefore anticipate that our results may be an underestimation of the prevalence of pathogens and their close relatives in soil and aerosol samples.

Potential Impact of Background Species on Assays for Biothreat Agent Detection

The genera *Bacillus*, *Clostridium*, *Francisella*, and *Yersinia* each include species that range from nonpathogenic environmental species, through symbionts and facultative pathogens of vertebrates and invertebrates, to highly virulent human and animal pathogens. Emerging datasets from comparative genomic sequencing and typing studies for these genera indicate that the sequence similarity and gene composition of species having very different lifestyles can be very high (11,13,14,41). Extensive similarity in genome sequence suggests a corresponding protein makeup and similar physiological attributes as well. Our results indicate that close relatives of biothreat pathogens are present in environmental samples, and while the diversity and distribution of these background species is variable, they have the potential to impact interpretation of pathogen detection assay results. Potential for interference by background organisms is higher in soils that in aerosols, where biomass, extracted DNA and diversity of related species are higher. In aerosol samples, where biomass and extractable DNA concentrations are very low (relative to soils), any background species may be below assay detection limits in routine testing, and a detected spike in a pathogen target may have higher reliability.

B. anthracis

B. cereus group species, very closely related to *B. anthracis*, were found in a high percentage of the soil and aerosol samples, indicating background DNA related to *B. anthracis* is potentially abundant and widespread in the environment. Although *B. anthracis* has not been cultured directly from soil, the closely related *B. cereus*, *B. thuringiensis*, *B. mycoides*, and *B. medusa* species detectable by this primer set are common in many types of soils (42). Spore-forming bacteria can easily survive the desiccated conditions in air and on aerosol collection filters, and it is not surprising that members of this group were detected in aerosols. Although the 16S rRNA gene group primer set for *B. anthracis* was designed to be more narrow in phylogenetic scope than primer sets for the other three pathogens, positive results were obtained with many more soil and aerosol samples than with the other pathogen primer sets. This result indicates that the potential for interference due to natural background may be higher for *B. anthracis* than for the other three pathogens. The 16S rRNA gene sequence of *B. anthracis* and those of several closely related *B. cereus* isolates are indistinguishable (43–45). Although sequences matching that of *B. anthracis* were found in several soil and aerosol samples, it is not possible to say if they derived from the pathogen or a closely related *B. cereus* strain.

The widespread distribution of species closely related to *B. anthracis* in many soils and nearly every city examined in this study, and the presence of the *pag* gene sequence in some samples, suggests that specific detection of pathogenic *B. anthracis* may be a significant challenge. Environmental bacilli can confound detection of *B. anthracis* in a number of ways. First, environmental *B. cereus* and *B. thuringiensis* strains that are normal soil organisms, or that cause food poisoning or dental disease, and do not cause anthrax disease, have been found to carry portions of the *B. anthracis* virulence plasmids (9,10,41) and are closely related by genomic AFLP typing (46,47), MLST analysis (48), and MEE (8). The genomic sequence similarity is very high between *B. anthracis* and nonpathogenic *B. cereus* and *B. thuringiensis* species (11). Second, *Bacillus* strains clinically identified as species other than *B. anthracis* have been found to cause inhalation anthrax symptoms in humans and carry a plasmid that is 99.6% similar to the pXO1 plasmid (12). In our study, DNA from aerosol samples testing positive with the *B. cereus* group and *pag* primers were subsequently tested in PCR using primer sets for other virulence genes (*cya*, *lef*), a pXO1 ORF of unknown function, and the *rpoC* gene. Amplification results from this screening were variable, with positive results obtained with some primer sets in some samples, but no single sample was positive for all *B. anthracis* markers (data not shown) (49). Sequence analysis of products amplified with primers designed to the *rpoC* gene identified only the presence of sequences highly similar, but not identical to that of *B. anthracis,* each with one to three SNPs when compared with the *B. anthracis rpoC* sequence. Taken together, the most plausible explanation is the presence of very closely related *Bacillus* species that harbor some but not all of the pXO1 plasmid genes. Recent genomic and physiological studies support this explanation by demonstrating that isolates related to *B. anthracis* can carry pXO1 genes, including the toxin genes (9,12,41). Our results and the recent genomic comparisons highlight the importance of using several DNA loci and a variety of complimentary assays for detection confirmation.

C. perfringens

C. perfringens is widely distributed in the environment. It has been found in soil, dust, feces, and other environmental samples, and is a normal resident of the intestinal tract of many animals including humans (50). Strains have been typed into five groups based on the complement of toxins they produce (51–53). The 16S rRNA gene primers in our survey should detect members of all the types, and the *cpe* gene primers detect strains potentially causing food poisoning (primarly type A strains). This is the first report of detection of *C. perfringens*-like organisms in aerosol samples.

Samples giving positive results with the 16S rRNA gene primers for *C. perfringens* were not positive for the *cpe* gene, suggesting that the 16S rRNA gene-positive samples contained only nonenterotoxin producing *C. perfringens* strains or other *Clostridium* species. Also, our sequence analysis of *cpe*-positive/16S rRNA gene-negative samples suggests that the *cpe* gene, or homologues of it, may be present in other bacterial species outside the *C. perfringens* group in environmental samples.

F. tularensis

16S rRNA gene sequences identical or nearly so to that of *F. tularensis* were detected in aerosol samples from two cities. Although this is the first report of *F. tularensis*-like sequences obtained from aerosol samples, there is epidemiological evidence for aerosol transmission of this pathogen (54). Although most sequences obtained matched those of known *F. tularensis* strains, a number of sequences differed slightly from those available in the database for any cultured isolate. These differences do not correspond to those identified by Forsman et al. (55) as diagnostic of *Francisella* subspecies, nor do they match any of the *F. tularensis*-like sequences we detected in soil samples in a separate study (26). These may derive from different rRNA gene operons within *F. tularensis* cells, or different strains or subtypes of the species. They could also result from *Taq* polymerase errors. The data collected in the present study indicates that this pathogen and/or very closely related species are present in naturally occurring urban aerosols, and may have potential for detection by biothreat monitoring efforts. While the natural reservoir(s) for *F. tularensis* remain largely uncharacterized, it has been detected in water (24,56), and a recent DNA-based survey of soil samples from Houston identified four new groups of *Francisella* rRNA gene sequences, and additional putative *F. tularensis* subspecies (26). This suggests that environmental sources may be important reservoirs for this pathogen.

Y. pestis

Y. pestis was not detected, and *Y. pestis* relatives were rarely detected in the soil and aerosol samples tested here. Only sequences matching those of *Y. rohdei*, *Y. mollaretii*, and *Y. kristensenii*, species that have previously been isolated from water, soil, and animal and human excrement (57–60), were obtained. This result suggests the possibility that *Y. pestis* close relatives are not widespread in soil/sediment and aerosol samples, and may not be of great concern for interference in surveillance activities. This result is consistent with the observation that most known *Y. pestis* strains have been found in association with the animal or flea hosts. Other species of *Yersinia* have been isolated primarily from animal hosts, as well as from water and soil samples (61,62). It is also possible, however, that as nonspore forming bacteria the *Yersinia* do not survive well on EPA filters that for this study were stored for several months before analysis.

Unlike most clinical assays for pathogenic organisms, assays for organisms in environmental samples face the challenge of detecting specific organisms in a very complex, highly dynamic and largely uncharacterized background of bacteria. Although the database of bacterial genomic sequences is growing rapidly, it may never include all environmental DNA sequences. As the sequence database expands, it is becoming clear that high levels of sequence identity exist between pathogens and their closely related species and that our understanding of the specific complement of genetic determinants that distinguishes a highly virulent pathogen from a nonvirulent environmental species is very limited. The results of this study indicate that a variety of such closely related species are present in the environment and validation of such assays against a wide variety of environmental samples is an important component of assay development, to test against this natural background. There remain significant gaps in our knowledge about the ability of these four bacterial pathogens to survive in the environment, and about environmental sources from which outbreaks initiate. Continued study of these species and their close relatives should provide new information on their ecology, epidemiology, and evolution.

Acknowledgments

For the generous contribution of $PM_{2.5}$ aerosol samples to this study, we thank Lee Ann Byrd of the U.S. EPA, regional EPA managers, and the following representatives of local air quality monitoring agencies: Ted Erdman and Bob Day (DC); Mahmood Hossain and Steve Barbosa (CA); Randy Sedlacek and Bev Chenausky (AZ); Edward Michel (TX); Bob La Morte (IL); Dirk Felton (NY); Tracy Daub and Jeffery D. Robbins (CA); and Susan Davis (WA); Patrick McGraw (CO); Dick Arbes (FL); Kit Redmond (GA); and Israel Tavarez (NM). We thank numerous volunteers for collection of soil samples. Pathogen and related species' DNA were provided by Karen Hill, Paul Jackson, and Rich Okinaka. DNA sequencing was provided by the DOE JGI at Los Alamos National Laboratory. The authors thank Erin Williams, Jennifer Richardson, Judith Buckingham, Peter Stark, and Amy Yactor for their technical contributions to this study.

References

1. Petersen JM, Schriefer ME, Carter LG, Zhou Y, Sealy T, Bawiec D, et al. Laboratory analysis of tularemia in wild-trapped, commercially traded prairie dogs, Texas, 2002. Emerg Infect Dis 2004;10:419–25.
2. CDC. Tularemia-United States, 1990–2000. MMWR 2002;51:181–4.
3. Girard JM, Wagner DM, Vogler AJ, Keys C, Allender CJ, Drickamer LC, et al. Differential plague-transmission dynamics determine *Yersinia pestis* population genetic structure on local, regional, and global scales. Proc Natl Acad Sci USA 2004;101:8408–13.
4. Chang MH, Glynn MK, Groseclose SL. Endemic, notifiable bioterrorism-related diseases, United States, 1992–1999. Emerg Infect Dis 2003;9:556–64.
5. Feldman KA, Enscore RE, Lathrop SL, Matyas BT, McGuill M, Schriefer ME, et al. An outbreak of primary pneumonic tularemia on Martha's Vineyard. N Engl J Med 2001;345:1601–6.
6. Dembek ZF, Buckman RL, Fowler SK, Hadler JL. Missed sentinel case of naturally occurring pneumonic tularemia outbreak: lessons for detection of bioterrorism. J Am Board Fam Pract 2003;16:339–42.
7. Olsen SJ, MacKinon LC, Goulding JS, Bean NH, Slutsker L. Surveillance for food-borne-disease outbreaks, United States 1993–1997. MMWR 2000;49:1–51.
8. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, et al. *Bacillus anthracis, Bacillus cereus*, and *Bacillus thuringiensis*–one species on the basis of genetic evidence. Appl Environ Microbiol 2000;66:2627–30.
9. Pannucci J, Okinaka RT, Sabin R, Kuske CR. *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. J Bacteriol 2002;184:134–41.
10. Pannucci J, Okinaka RT, Williams E, Sabin R, Ticknor LO, Kuske CR. DNA sequence conservation between the *Bacillus anthracis* pXO2 plasmid and genomic sequence from closely related bacteria. BMC Genomics 2002;3:34.
11. Read TD, Peterson SN, Tourasse N, Baillie LW, Paulsen IT, Nelson KE, et al. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature 2003;423:81–6.
12. Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, Marston CK, et al. Identification of anthrax toxin genes in a Bacillus cereus associated with an illness resembling inhalation anthrax. Proc Natl Acad Sci USA 2004;101:8449–54.

13. Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, et al. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proc Natl Acad Sci USA 2004;101:13826–31.
14. Hinchliffe SJ, Isherwood KE, Stabler RA, Prentice MB, Rakin A, Nichols RA, et al. Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. Genome Res 2003;13:2018–29.
15. McBride MT, Masquelier D, Hindson BJ, Makarewicz AJ, Brown S, Burris K, et al. Autonomous detection of aerosolized *Bacillus anthracis* and *Yersinia pestis*. Anal Chem 2003;75:5293–9.
16. Davidson JW, Imbro DR. The Biological Aerosol Sentry and Information System (BASIS): A DDAP to provide early detection of bioterrorism incidents. Report No.: LA-UR-02-4545, Los Alamos National Laboratory, Los Alamos, 2002.
17. Rappe MS, Giovannoni SJ. The uncultured microbial majority. Annu Rev Microbiol 2003;57:369–94.
18. Rotz LD, Khan AS, Lillibridge SR, Ostroff SM, Hughes JM. Public health assessment of potential biological terrorism agents. Emerg Infect Dis 2002;8:225–30.
19. Kuske CR, Banton KL, Adorada DL, Stark PC, Hill KK, Jackson PJ. Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. Appl Environ Microbiol 1998;64: 2463–72.
20. Dunbar J, Kuske CR Inventors. Los Alamos National Laboratory, assignee. Chemical agent for selective removal of PCR inhibiting compounds from environmental DNA extracts. Invention disclosure no. 2001-108.
21. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, et al. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res 2003;31:442–3.
22. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
23. Ashelford KE, Weightman AJ, Fry JC. 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 2002;30:3481–9.
24. Forsman M, Nyrén A, Sjöstedt A, Sjökvist L, Sandström G. Identification of *Francisella tularensis* in natural water samples by PCR. FEMS Microbiol Ecol 1995;16:83–92.
25. Weisburg WG, Dobson ME, Samuel JE, Dasch GA, Mallavia LP, Baca O, et al. Phylogenetic diversity of the Rickettsiae. J Bacteriol 1989;171: 4202–6.
26. Barns SM, Grow CC, Okinaka RT, Keim P, Kuske C. Detection of diverse new *Francisella*-like bacteria in environmental samples. Appl Environ Microbiol 2005;71:5494–500.
27. Meer RR, Songer JG. Multiplex polymerase chain reaction assay for genotyping *Clostridium perfringens*. Am J Vet Res 1997;58:702–5.
28. Lin YT, Labbe R. Enterotoxigenicity and genetic relatedness of *Clostridium perfringens* isolates from retail foods in the United States. Appl Environ Microbiol 2003;69:1642–6.
29. Titball RW, Johansson A, Forsman M. Will the enigma of *Francisella tularensis* virulence soon be solved? Trends Microbiol 2003;11:118–23.
30. Sjöstedt A, Eriksson U, Berglund L, Tärnvik A. Detection of *Francisella tularensis* in ulcers of patients with tularemia by PCR. J Clin Microbiol 1997;35:1045–8.
31. Versage JL, Severin DD, Chu MC, Petersen JM. Development of a multitarget real-time TaqMan PCR assay for enhanced detection of *Francisella tularensis* in complex specimens. J Clin Microbiol 2003;41:5492–9.
32. Hawkins TL, O'Connor-Morin T, Roy A, Santillan C. DNA purification and isolation using a solid-phase. Nucleic Acids Res 1994;22:4543–4.
33. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;24: 4876–82.
34. Olsen GJ, Matsuda H, Hagstrom R, Overbeek R. FastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput Appl Biosci 1994;10:41–8.
35. Bracho MA, Moya A, Barrio E. Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. J Gen Virol 1998;79 (Part 12):2921–8.
36. Gunther S, Sommer G, Von Breunig F, Iwanska A, Kalinina T, Sterneck M, et al. Amplification of full-length hepatitis B virus genomes from samples from patients with low levels of viremia: frequency and functional consequences of PCR-introduced mutations. J Clin Microbiol 1998; 36:531–8.
37. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. Genetic diversity in Sargasso Sea bacterioplankton. Nature 1990;345:60–3.
38. Price LB, Hugh-Jones M, Jackson PJ, Keim P. Genetic diversity in the protective antigen gene of *Bacillus anthracis*. J Bacteriol 1999;181: 2358–62.
39. Abramov VM, Vasiliev AM, Khlebnikov VS, Vasilenko RN, Kulikova NL, Kosarev IV, et al. Structural and functional properties of *Yersinia pestis* Caf1 capsular antigen and their possible role in fulminant development of primary pneumonic plague. J Proteome Res 2002;1:307–15.
40. Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci USA 2002;99:10494–9.
41. Rasko DA, Ravel J, Okstad OA, Helgason E, Cer RZ, Jiang L, et al. The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. Nucleic Acids Res 2004;32:977–88.
42. Kolstø AB. Time for a fresh look at the bacterial chromosome. Trends Microbiol 1999;7:223–6.
43. Bavykin SG, Lysov YP, Zakhariev V, Kelly JJ, Jackman J, Stahl DA, et al. Use of 16S rRNA, 23S rRNA, and gyrB gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. J Clin Microbiol 2004;42:3711–30.
44. Ash C, Farrow JAE, Dorsch M, Stackebrandt E, Collins MD. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus* and related species on the basis of reverse transcriptase sequencing of ribosomal RNA. Int J Syst Bacteriol 1991;41:343–6.
45. Barns SM, Hill KK, Jackson PJ, Kuske CR. 16S ribosomal RNA sequence-based phylogeny of *Bacillus* species and the *B. cereus group*. Report No.: LA-UR-99-5628, Los Alamos National Laboratory, Los Alamos, NM, 1999.
46. Hill KK, Ticknor LO, Okinaka RT, Asay M, Blair H, Bliss KA, et al. Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis, Bacillus cereus*, and *Bacillus thuringiensis* isolates. Appl Environ Microbiol 2004;70:1068–80.
47. Ticknor LO, Kolstø AB, Hill KK, Keim P, Laker MT, Tonks M, et al. Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. Appl Environ Microbiol 2001;67:4863–73.
48. Priest FG, Barker M, Baillie LWJ, Holmes EC, Maiden MCJ. Population structure and evolution of the *Bacillus cereus* group. J Bacteriol 2004; 186:7959–70.
49. Kuske CR. Analysis of pathogen backgrounds in soils and EPA aerosol samples. Report No.: LA-UR-04-3191, Los Alamos National Laboratory, Los Alamos, NM, 2004.
50. Hatheway CL. Toxigenic clostridia. Clin Microbiol Rev 1990;3:68–98.
51. Smedley JG III, Fisher DJ, Sayeed S, Chakrabarti G, McClane BA. The enteric toxins of *Clostridium perfringens*. Rev Physiol Biochem Pharmacol 2004;152:183–204.
52. Brynestad S, Granum PE. *Clostridium perfringens* and foodborne infections. Int J Food Microbiol 2002;74:195–202.
53. Rood JI. Virulence genes of *Clostridium perfringens*. Annu Rev Microbiol 1998;52:333–60.
54. Feldman KA, Stiles-Enos D, Julian K, Matyas BT, Telford SR III, Chu MC, et al. Tularemia on Martha's Vineyard: seroprevalence and occupational risk. Emerg Infect Dis 2003;9:350–4.
55. Forsman M, Sandström G, Jaurin B. Identification of *Francisella* species and discrimination of type A and type B strains of *F. tularensis* by 16S rRNA analysis. Appl Environ Microbiol 1990;56:949–55.
56. Anda P, Segura del Pozo J, Díaz García JM, Escudero R, García Peña FJ, López Velasco MC, et al. Waterborne outbreak of tularemia associated with crayfish fishing. Emerg Infect Dis 2001;7:575–82.
57. Falcao JP, Brocchi M, Proenca-Modena JL, Acrani GO, Correa EF, Falcao DP. Virulence characteristics and epidemiology of *Yersinia enterocolitica* and Yersiniae other than *Y. pseudotuberculosis* and *Y. pestis* isolated from water and sewage. J Appl Microbiol 2004;96:1230–6.
58. Wauters G, Janssens M, Steigerwalt AG, Brenner DJ. *Yersinia mollaretii* sp. nov. and *Yersinia bercovieri* sp. nov., formerly called *Yersinia enterocolitica* biogroups 3A and 3B. Int J Syst Bacteriol 1988;38: 424–9.
59. Aleksic S, Steigerwalt AG, Bockemuehl J, Huntley-Carter G, Brenner DJ. *Yersinia rohdei*: new species isolated from human and dog feces and surface water. Int J Syst Bacteriol 1987;37:327–32.
60. Bercovier H, Ursing J, Brenner DJ, Steigerwalt AG, Fanning G, Carter G, et al. *Yersinia kirstensenii*- a new species of Enterobacteriaceae composed

of sucrose-negative strains (formerly called atypical *Yersinia enterocolitica* or *Yersinia enterocolitica*-like). Curr Microbiol 1980;4:219–24.
61. Dolina M, Peduzzi R. Population genetics of human, animal and environmental *Yersinia* strains. Appl Environ Microbiol 1993;59:442–50.
62. Kapperud G. *Yersinia enterocolitica* in food hygiene. Int J Food Microbiol 1991;12:53–65.
63. Lane DJ. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic acid techniques in bacterial systematics. New York: John Wiley and Sons; 1991:115–75.
64. Radnedge L, Gamez-Chin S, McCready PM, Worsham PL, Andersen GL. Identification of nucleotide sequences for the specific and rapid detection of *Yersinia pestis*. Appl Environ Microbiol 2001;67:3759–62.
65. Forsman M, Sandström G, Sjöstedt A. Analysis of 16S ribosomal DNA sequences of *Francisella* strains and utilization for determination of the phylogeny of the genus and for identification of strains by PCR. Int J Syst Bacteriol 1994;44:38–46.

Additional information and reprint requests:
Cheryl R. Kuske, Ph.D.
Bioscience Division, B-N1, M888
Los Alamos National Laboratory
Los Alamos, NM 87545
E-mail: kuske@lanl.gov