Biogeography of Actinomycete Communities and Type II Polyketide Synthase Genes in Soils Collected in New Jersey and Central Asia[∇]

Boris Wawrik,[1] Djumaniyaz Kutliev,[2] Urinova A. Abdivasievna,[2] Jerome J. Kukor,[1,4]
Gerben J. Zylstra,[1,3] and Lee Kerkhof[5]*

*Biotechnology Center for Agriculture and the Environment,[1] Department of Biochemistry and Microbiology,[3] Department of
Environmental Sciences,[4] and Institute of Marine and Coastal Sciences,[5] Cook College, Rutgers University,
New Brunswick, New Jersey 08901-8521, and Uzbekistan Academy of Sciences Institute of
Microbiology, Tashkent, Uzbekistan 700128[2]*

Soil microbial communities are believed to be comprised of thousands of different bacterial species. One prevailing idea is that "everything is everywhere, and the environment selects," implying that all types of bacteria are present in all environments where their growth requirements are met. We tested this hypothesis using actinomycete communities and type II polyketide synthase (PKS) genes found in soils collected from New Jersey and Uzbekistan ($n = 91$). Terminal restriction fragment length polymorphism analysis using actinomycete 16S rRNA and type II PKS genes was employed to determine community profiles. The terminal fragment frequencies in soil samples had a lognormal distribution, indicating that the majority of actinomycete phylotypes and PKS pathways are present infrequently in the environment. Less than 1% of peaks were detected in more than 50% of samples, and as many as 18% of the fragments were unique and detected in only one sample. Actinomycete 16S rRNA fingerprints clustered by country of origin, indicating that unique populations are present in North America and Central Asia. Sequence analysis of type II PKS gene fragments cloned from Uzbek soil revealed 35 novel sequence clades whose levels of identity to genes in the GenBank database ranged from 68 to 92%. The data indicate that actinomycetes are patchily distributed but that distinct populations are present in North American and Central Asia. These results have implications for microbial bioprospecting and indicate that the cosmopolitan actinomycete species and PKS pathways may account for only a small proportion of the total diversity in soil.

The idea that most bacterial species are widely distributed ("everything is everywhere") and that different ecosystems select for the bacteria that are best adapted, which leads to relatively greater abundance of these bacteria ("the environment selects"), has been a staple of microbial ecological theory for almost a century (2, 8, 9). There is now good evidence supporting the notion that different ecosystems harbor unique microbial populations, i.e., that bacterial populations can exhibit biogeographic distribution (7, 8, 10–14, 26, 28, 29). For instance, a *Roseobacter* cluster has been found in polar and temperate regions, especially the southern ocean, but it has not been detected in tropical and subtropical regions (26). *Prochlorococcus marinus*, on the other hand, is known to dominate subtropical and tropical surface waters of the world's oceans but is not found in temperate and polar regions (4–6). Additionally, different ecotypes of *Prochlorococcus* (populations adapted to high-light conditions and populations adapted to low-light conditions) have been found in areas with various light regimens in the oceans (19, 20, 23). The factors leading to such distributions are often not understood, but the data clearly demonstrate that the global distribution of individual microbial species can covary with physical parameters, such as climate and light. In this context, Fierer and Jackson (12) used rRNA gene fingerprinting to compare the microbial communities found in soils collected in areas across the Americas. Their experimental results indicated that there was no direct relationship between microbial diversity and a variety of physical and chemical characteristics, such as temperature, latitude, precipitation, silt and clay content, C/N ratio, and moisture content. Species richness was, however, linked to ecosystem type and soil pH, indicating that certain parameters, such as soil chemistry and ecological context, can affect the distribution of bacteria in the environment.

An important corollary of the "everything is everywhere" hypothesis is that all types of bacteria are found in all environments where their growth requirements are met. This prediction has significant implications for bioprospecting for novel secondary metabolites or enzymatic processes of industrial interest. If most bacterial species are found everywhere (i.e., are cosmopolitan), then only a limited number of samples from a particular type of environment need to be surveyed intensely to obtain a large proportion of all microbes associated with that environment. However, if there are endemic populations in similar environments at different locations, then it would be prudent to survey the greatest possible geographic breadth. In several surveys workers have investigated the levels of endemism of specific groups of microbes. For example, in an investigation of the biodiversity and endemism of cyanobacteria in thermal hot springs researchers discovered that some thermophilic strains from temperate zone North American springs are not present in hot springs in Alaska and Iceland (7). In a similar study, using 150 3-chlorobenzoate-degrading isolates

from six regions on five continents, workers observed that more than 91% of the strains had unique genotypes and were present at a single site (13). In a more in-depth molecular fingerprinting survey of 248 fluorescent *Pseudomonas* strains isolated from 10 locations on four continents, Cho and Tiedje detected 85 unique genotypes for which there was no overlap in the sites and continental regions of the collection sites (8). Nucleic acid-based methods have also revealed geographic structuring of denitrifying bacteria in coastal sediments (25), nitrifying bacteria in the ocean (1), soil microbial populations (12), and sulfate-reducing bacteria (24).

It has been hypothesized that the bacteria that are more easily dispersed are better suited for colonizing new environments and are more cosmopolitan (29). For example, certain gram-positive bacteria, particularly *Bacillus* spp. and *Streptomyces* spp., are exceptionally well adapted for dispersal, because they produce spores that are highly resistant to desiccation and heat. The purpose of this study was to extend our inquiries of actinomycete communities and secondary metabolite gene diversity found in New Jersey soils (34) to soils collected in Central Asia in order to investigate whether different populations of this important group of microorganisms could be found in Uzbekistan. The degree of actinomycete cosmopolitanism was investigated by terminal restriction fragment length polymorphism (TRFLP) analysis of actinomycete rRNA genes. An analogous analysis was performed for an important class of secondary metabolite genes, the type II polyketide synthase (PKS) genes, which are frequently found in actinomycetes. Environmental fingerprint data demonstrated that there are distinct differences between actinomycete communities present in soil collected in Uzbekistan and actinomycete communities present in soil collected in New Jersey. Evidence of globally distributed actinomycete phylotypes, as determined by TRFLP analysis, was obtained, but these phylotypes appeared to constitute only a small proportion of the phylotypes at each location. These findings indicate that it should be useful to sample remote areas of the globe to enhance bioprospecting endeavors.

MATERIALS AND METHODS

Sampling locations. A total of 23 soil samples were collected in New Jersey in March 2004 at several sites in the Greenwood Forest Wildlife Management Area (Ocean County) and the Picatinny Arsenal (Morris County). These samples have been described previously (34). Care was taken to obtain a diverse set of soils ranging from low-organic-matter sand to soils composed of almost entirely of root and leaf litter. Samples were collected in 50-ml sterile plastic tubes, frozen immediately, and thawed for DNA extraction. The Central Asian samples were collected in the vicinity of Tashkent, Uzbekistan, during the summer of 2004. Soil samples were obtained on 4 to 19 May and 17 to 19 July 2004 near Sukok village. Additional samples were obtained on 4 to 7 June 2004 near the villages of Yusuphona, Burchmulla, and Chimyon (Chimgan) and on 27 and 28 June 2004 near Khorezm village. The samples were acquired using a bulb planter, which was cleaned and sterilized with ethanol between individual sampling procedures, at depths of 0 to 60 cm. The small soil cores were placed into plastic whirl-pack bags and stored in a refrigerator for approximately 1 month until shipment to New Jersey, where they were stored at 4°C.

DNA extraction. DNA extraction and PCR were performed as previously described (34). DNA was extracted from 0.5 g of soil using a MO BIO Ultra-Clean soil DNA kit as recommended by the manufacturer. DNA yields were determined by comparison of band intensities to the band intensity for 250 ng of phage λ DNA digested with HindIII on a 1% agarose gel stained with ethidium bromide and photographed using a Kodak EDAS 290 gel imaging system.

PCR and cloning. Actinomycete 16S rRNA genes were PCR amplified from 5 ng of soil community DNA using the actinomycete group-specific forward primer

243F (15) (5'-GGATGAGCCCGCGGCCTA-3') and the eubacterial reverse primer 1401R (5'-CGGTGTGTACAAGACCC-3'). The more frequently used ribosomal reverse primers 1525R and 1492R in combination with primer 243F produce nonspecific banding patterns in agarose gels and were not suitable for TRFLP analysis of actinomycete communities (34). Actinomycete TRFLP patterns were generated by labeling primer 243F with 6-carboxyfluorescein (FAM). The PCR mixtures (50 μl) contained 200 nM of each PCR primer, 2.5 mM MgCl, 5 U *Taq* DNA polymerase (Promega, Madison WI), and 0.4 μl of a 10-mg ml^{-1} bovine serum albumin solution (Promega, Madison WI). The amplification reactions were performed for 30 cycles consisting of 1 min at 95°C, 1 min at 60°C, and 1.5 min at 72°C, followed by extension at 72°C for 15 min. Partial ketoacyl synthase (KS) genes of type II PKS pathways were amplified from soil community DNA by using degenerate PCR primers 540F (5'-GGITGCACSTCIGGIM TSGAC-3') and 1100R (5'-CCGATSGCICCSAGIGAGTG-5') (34). The PCR mixtures (50 μl) contained 5 U of *Taq* DNA polymerase, 1 μM of each primer, 2.5 mM MgCl, and 0.4 μl of a 10-mg ml^{-1} bovine serum albumin solution (Promega, Madison WI). Each PCR cycle consisted of 1 min at 95°C, 1 min at 68°C, and 1.5 min at 72°C. Clone libraries were generated by performing 40 PCR cycles, followed by 15 min of extension at 72°C. PCR products were excised from 1% agarose gels. The agarose was removed by using a QIAquick gel extraction kit (QIAGEN GmbH, Germany), and the DNA was cloned into the pCR4-TOPO vector (Invitrogen, Carlsbad CA) according to the manufacturer's recommendations. The four most diverse soil samples from Uzbekistan, as determined by TRFLP analysis of PKS gene PCR products, were chosen for cloning, and 24 colonies were picked from each transformation into a 96-well microtiter plate (to obtain a total of 96 individual clones). Plasmid DNA was purified using a QIAGEN plasmid extraction kit, and inserts were sequenced by using the M13 forward and reverse primers. PKS TRFLP patterns were generated by first amplifying 5 ng of soil DNA for 30 cycles using unlabeled primers. Five microliters of the reaction mixture was then used for another PCR performed with FAM-labeled primer 540F. Ten additional cycles were performed as described above to label PKS gene PCR products.

TRFLP and data analysis. FAM-labeled actinomycete 16S rRNA and KS gene PCR products were diluted to obtain a concentration of 5 ng μl^{-1}. Samples that produced no PCR product visible in an ethidium bromide-stained 1% agarose gel were not included in subsequent analyses. A total of 30 ng of labeled DNA was digested with 2 U of MnlI (NEB, Beverly, MA) for 4 h at 37°C. Digests were precipitated using glycogen as a carrier (18). The DNA was suspended in 20 μl deionized formamide, denatured at 95°C, and separated with an ABI 310 automated sequencer (Applied Biosystems, Foster City, CA). The data for fragment lengths between 50 and 500 nucleotides were exported in a tabular, binary (presence/absence) format to Matlab (Natick, MA) for analysis. A minimum of 50 fluorescence units was used, and only the peaks accounting for 95% of the area were considered. The data were then analyzed using two different algorithms, one of which binned the data into discrete nucleotide length and one of which used a search window of ±0.5 bp to compare all peaks in the data set. Binning can sometimes be problematic (16, 34) during data analysis (peaks that are only a fraction of 1 bp apart are occasionally binned differently, due to rounding artifacts) but is necessary to generate peak frequency histograms (Fig. 1). Dendrograms were generated using both binning and search window algorithms (Fig. 2). The two methods produced analogous results. Pairwise similarity values were calculated using two different methods, as previously described (33, 34). The Sorenson index was calculated as follows: $C_s = 2N_{ab}/(N_a + N_b)$, where N_{ab} is the number of shared peaks and N_a and N_b are the numbers of peaks in individual samples (25). Alternatively, similarity values were calculated as follows: $P_s = S_a/N_a$, where S_a is the number of peaks in sample A that are also found in sample B and N_a is the total number of peaks in sample A (33). The top left and bottom right portions of similarity matrices were averaged and exported as $1 - C_s$ or $1 - P_s$. Binned data were exported into MEGA (Molecular Evolutionary Genetics Analysis Software, version 2.1) (17) to generate neighbor-joining dendrograms. Similar topologies were obtained regardless of which type of similarity matrix was calculated (data not shown). A bootstrap analysis (10,000 replicates) was performed as described previously (33) to test the robustness of the dendrograms.

Phylogenetic analysis. DNA sequences were trimmed to remove vector and primer sequences. Sequences were then aligned with their closest matches in the GenBank database as determined by BLASTX and BLASTN, as well as with a representative number of other PKS sequences in the GenBank database. Neighbor-joining trees were generated using MEGA by applying the Tajima-Nei distance model. One thousand bootstrap replicates were performed.

Nucleotide sequence accession numbers. PKS gene sequence data obtained in this study have been deposited in the GenBank database under accession numbers EF068379 to EF068414.
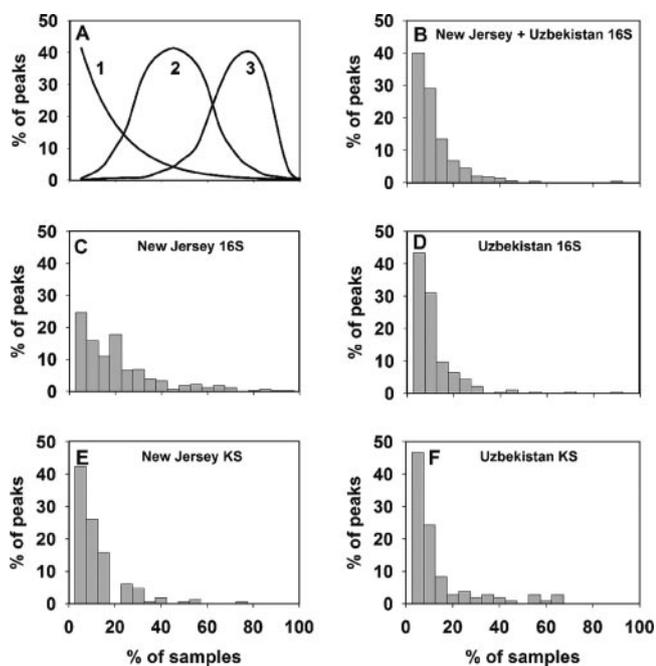
FIG. 1. Histograms showing the frequency distributions of fragment lengths in the data set. The *x* axis shows the percentages of samples (TRFLP fingerprints) that contained different percentages of the fragments detected in the data set (*y* axis). We expected our data to exhibit one of three distributions. (A) Three models for the frequency distribution of TRFLP data. (B) Frequency distribution of all actinobacterial 16S rRNA gene TRFLP fingerprints generated in this study. (C) Frequency distribution of actinobacterial 16S rRNA gene TRFLP fingerprints obtained for samples collected in New Jersey (*n* = 23). (D) Frequency distribution of actinobacterial 16S rRNA gene TRFLP fingerprints obtained for samples collected in Uzbekistan (*n* = 61). (E) Type II PKS TRFLP fingerprints of samples collected in New Jersey (*n* = 16). (F) Type II PKS TRFLP fingerprints of samples collected in Uzbekistan (*n* = 56).

RESULTS

Soils were collected in New Jersey (*n* = 23) and Uzbekistan (*n* = 68), and TRFLP fingerprints were generated to estimate actinomycete and type II PKS gene sequence richness. Conceptually, we expected the data to support one of three models, which were designated models 1, 2 and 3 (Fig. 1A). Model 1 describes a situation where actinomycete populations are highly variable. There are very few TRFLP peaks (species) that are found at a high frequency in the environment (i.e., are distributed over a large geographic range [cosmopolitan]). Model 2 implies that many actinomycete species are found frequently in the environment. Few species are truly endemic (i.e., found in only one or a few samples), and similarly, few species are truly cosmopolitan (i.e., found everywhere). Model 3 illustrates the "everything is everywhere" hypothesis. Most species are found in a large proportion of samples, while only a few species are found to have limited distributions; virtually no species are endemic.

A total of 355 different actinomycete terminal 16S rRNA gene fragments were detected in the TRFLP data set. Most of the peaks observed were present in only a small proportion of the soil samples, regardless of whether the data were compared at a global scale (Central Asia and North America [Fig. 1B]) or

regionally (different locations in New Jersey [Fig. 1C] or in the area near Tashkent [Fig. 1D]). More than two-thirds of the TRFLP peaks (82.5, 69, and 90.3% in Fig. 1B, C, and D, respectively) were observed in less than 20% of all actinomycete profiles. More than one-quarter of the peaks (40, 25, and 43% in Fig. 1B, C, and D, respectively) were observed in <5% of the samples. Sixty-four peaks (18%) were obtained with only one soil sample, while 112 peaks (32%) were obtained with less than three soil samples. Only two peaks were obtained with >50% of the samples (Fig. 1B) (fragment lengths, 96 and 216 bp). The 96-bp fragment was present in nearly all fingerprints. To obtain more information about these two highly cosmopolitan peaks, the TRFLP size and priming site data were used to screen sequences found in the Ribosomal Database Project (RDP) database (http://rdp.cme.msu.edu/). For this purpose all 49,249 16S rRNA sequences in the RDP database at the time of the analysis (sequences from isolates for which good-quality sequences more than 1,200 nucleotides long were available) were downloaded. All sequences that matched our forward primer were identified, which yielded a total of 2,857 16S rRNA gene sequences. Of these, only one sequence produced the 96-bp fragment in silico (*Streptomyces* sp. strain 1A01503; accession no. EF056489). The 216-bp fragment was produced in silico by eight sequences in the database, which originated from *Cryptosporangium* sp. (accession no. AB006166 and AB006168), *Streptomyces lividans* (accession no. X95968, X95969, and X86354), *Streptomyces microstreptospora* (accession no. AB006159), *Streptomyces caelestis* (accession no. AJ508062), and *Streptomyces* sp. strain EN9 (accession no. AY148087). These data indicate that the microorganisms that produce the 96- and 216-bp peaks are not frequently observed in culture or in 16S rRNA gene libraries of actinomycetes, despite their wide geographic distribution.

TRFLP analysis of PKS PCR products (Fig. 1E and F) yielded results similar to the results obtained for the actinomycete 16S rRNA genes. More than 80% of the 191 discernible terminal restriction fragments (TRFs) were detected in less than 20% of the soil samples (84 and 91% in Fig. 1E, and F, respectively). More than 31% of the TRFs (60 peaks) were found in only a single soil sample. These data are consistent with model 1 (Fig. 1A), suggesting that cosmopolitanism of actinomycetes and their secondary metabolite genes may be an exception and not the rule. Most TRFs appeared infrequently, and the distribution of actinomycetes appears to be highly variable.

It has been postulated that bacterial phylogeny should correlate with bacterial biogeography, if species are endemic (29). No geographic clustering of closely related species would be expected for taxa that are cosmopolitan (29). By the same token, it would be expected that the distribution of peaks in TRFLP fingerprints would reflect the geographic origin of samples if bacteria exhibit regional ranges and/or endemism. Geographic clustering of TRFLP fingerprints should not occur if the bacteria examined are cosmopolitan. We compared fingerprint profiles by cluster analysis to determine if geographic clustering of actinomycete communities occurred (Fig. 2). The analysis revealed a strong tendency of actinomycete TRFLP patterns to cluster by country of origin (Fig. 2A). This division was retained whether the data were analyzed by binning or by using a search window approach (see Materials and Methods).
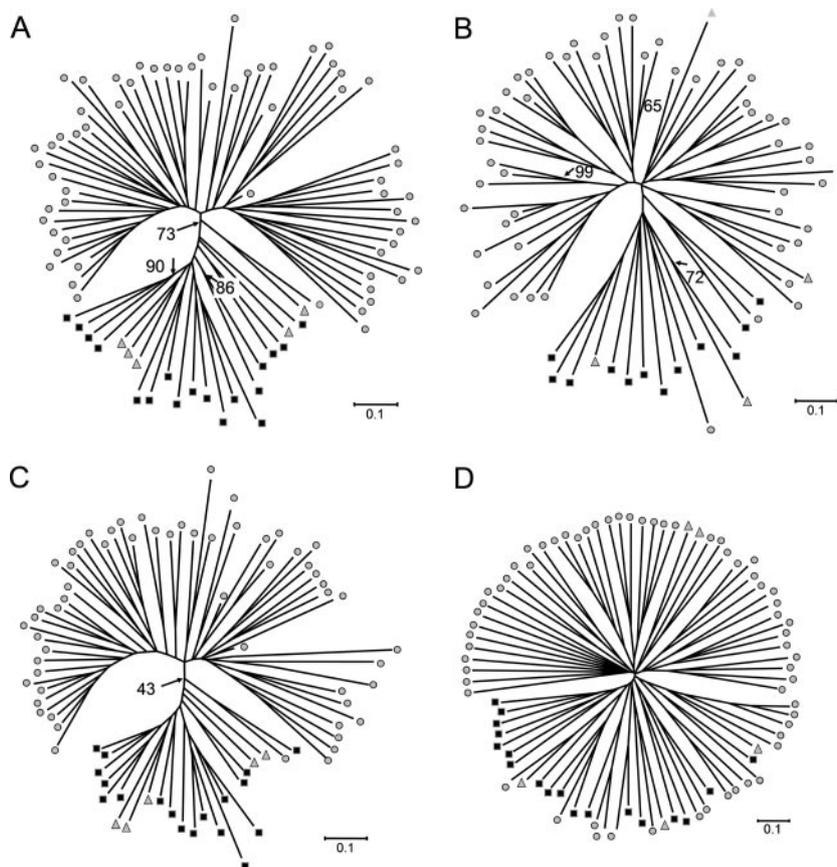
FIG. 2. Cluster analysis of actinobacterial 16S rRNA gene and type II PKS gene TRFLP data. The dendrograms are neighbor-joining dendrograms obtained from a similarity matrix (P_s) by using a sliding search window of ± 0.5 bp. A distance of 0.1 indicates a 10% difference between samples. Bootstrap values were obtained by binning the data and calculating a consensus tree using 1,000 bootstrap replicates. With the exception of the central node in panel C, only the bootstrap values that are >65% are shown. The arrows indicate bootstrap values for specific nodes. ●, soil samples collected in Uzbekistan; ■, soil samples collected in the Greenwood Forest Wildlife Management Area; △, soil samples collected in Picatinny Arsenal. (A) Actinobacterial 16S rRNA gene TRFLP data. (B) Type II PKS gene TRFLP data. (C) Actinobacterial 16S rRNA gene TRFLP data when only the peaks that were found in more 10% of all samples were considered. A bootstrap value is shown for the central node, which divides New Jersey samples from Central Asian samples, even though the value is less than 65%. (D) Actinobacterial 16S rRNA gene TRFLP data when only the peaks that were found in less than 10% of all samples were considered.

Similar dendrograms were obtained using the unweighted-pair group method using average linkage (data not shown) and the neighbor-joining method. A bootstrap analysis was performed to confirm the significance of the branching pattern observed. Only three significant nodes (bootstrap values, >65%) were observed (Fig. 2A). However, the node that indicates a division between North American and Central Asian samples was among the significant branch points of the dendrogram (bootstrap value, 73%). The same analysis revealed no significant clustering of fingerprints generated from samples collected in different parts of New Jersey (northern New Jersey versus southern New Jersey). Weaker regional clustering of PKS TRFLP patterns (Fig. 2B) was observed but was not empirically supported by bootstrap analysis. For more in-depth analysis, we divided the actinomycete 16S rRNA gene TRFLP data into the fragments sizes found frequently in the fingerprints (>10% of all fingerprints) and the fragment sizes found infrequently (<10% of all samples). This allowed us to determine whether the branching patterns shown in Fig. 2A and B were due to differential distribution of abundant or rare peaks in the

data set. Each subset of the data was analyzed independently, as indicated above. The analysis of frequently observed peaks produced a dendrogram that was almost identical to the dendrogram generated by considering the complete data set (Fig. 2C), albeit with slightly weaker bootstrap support. On the other hand, regional clustering was not observed for peaks found infrequently in the data set (74% of all fragments) (Fig. 2D). These data indicate that the distribution of actinomycete phylotypes (TRFs) has a regional imprint, supporting the idea that some actinomycete species can have restricted distributions. However, only 9 of the 91 fragment sizes that were detected in more than 10% of all samples appeared to be present in only one country (all 9 were detected in Uzbekistan). Similar results were obtained when the PKS fingerprints were considered. Of 30 fragments that were detected in more than 10% of all samples, only 2 were present only in New Jersey (none was present only in Uzbekistan). These data suggest that there is significant overlap between the widely distributed actinomycete populations and PKS genes found in Central Asia and New Jersey.
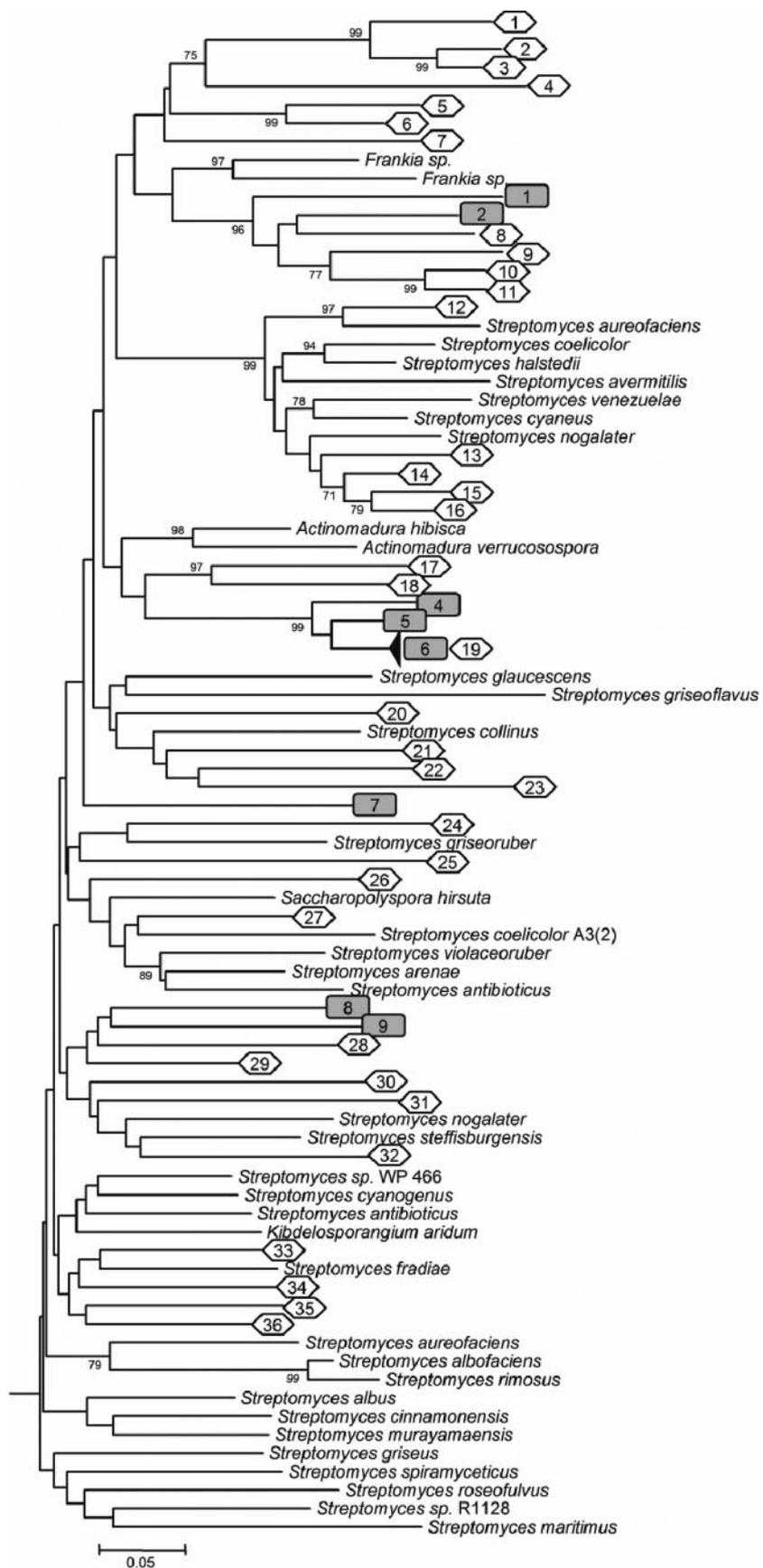
Uzbekistan ⬡   New Jersey ▭

Four Uzbek soil samples were selected based on the number of peaks in their PKS fingerprints (the two samples with the highest total number of peaks and the two samples with the highest number of unique peaks). The TRFLP fingerprints for these four samples contained 54 fragments that were different sizes, accounting for 50.4% of the fragments found in the Uzbek PKS data set. Community DNA was amplified using PKS gene PCR primers, and the PCR products were TA cloned. Twenty-four of the clones from each collection site (a total of 96 clones) were sequenced. All sequences exhibiting more than 95% sequence identity were grouped into clades. Sequence analysis revealed 37 novel sequence clades, all of which exhibited the highest level of similarity to type II KS genes found in the database (Fig. 3). Only one clade (Uzbekistan clade 19) (Fig. 3) matched PKS gene sequences that were obtained from New Jersey soil samples (New Jersey clade 6) (Fig. 3) in our previous study (34). All of the remaining sequences were unique to Uzbekistan and not similar to sequences in the GenBank database. The DNA sequences of cloned PKS genes were between 68 and 92% identical to sequences for the PKS pathways found in public databases. These findings indicate that the majority of type II PKS gene sequences occur infrequently in the environment but that some sequences are present in both Asia and North America, indicating that they have a cosmopolitan distribution.

DISCUSSION

Data presented here suggest that cosmopolitanism among actinomycete species and PKS genes encoding proteins in secondary metabolite pathways is not common (Fig. 1). The ribosomal genes exhibited significant regional clustering, while the type II PKS genes clustered only weakly. This weak clustering may have resulted from more rapid evolution of the PKS genes than of highly conserved genes, such as the 16S rRNA genes. A rapidly evolving gene could be expected to generate restriction fragment length polymorphisms at a higher rate than a highly conserved gene. This notion is supported by the observation that >30% of all PKS terminal fragments detected were unique to just one sample, while only 18% of the 16S rRNA gene terminal fragments were unique to individual samples. Perhaps the generation of so many unique PKS TRFLP peaks for the different geographic sites obscured similarities and obscured regional clustering. It is also possible that the clustering patterns observed (Fig. 1 and 2) were the result of artifacts of the TRFLP method. By nature, all PCR-based approaches for studying microbial communities are limited by inherent biases of the PCR amplification step. These biases, however, should favor increased similarity among samples. Primer biases should lead to preferential amplification of similar sequences from all samples. Alternatively, different targets with different starting

concentrations should be found at similar concentrations after PCR (30). Moreover, the TRFLP method measures the length of only one labeled, terminal fragment and can thus distinguish only a relatively small number of different operational taxonomic units (450 different fragment sizes, as determined in this study). Thousands of different actinomycete strains have been isolated from the environment, and thousands of different bacterial species can be found in just a single soil sample (31, 32). Our PCR-based methods should result in underestimation of the actual number of actinomycetes species in the environment. There is also no a priori reason to expect a PCR bias to affect one sample and not another, leading to altered similarity measurements. It is also possible that small peaks, which may be artifacts or noise in the community profiles, skew frequency distributions and clustering algorithms to generate the patterns which are observed. However, similar results were obtained regardless of the minimal peak height that was used (50 to 500 fluorescence units) or the percentage of the peak area in each TRFLP that was considered (100 to 50%) (data not shown). Another possibility is primer bias resulting from the 243F primer. In the original study describing this primer (15), however, only 13 of the 2,270 nonactinomycete 16S rRNA gene sequences that were analyzed were found to be identical at the 3′ terminus (positions 239 to 243). A more likely bias is introduced by the fact that the 243F primer does not match all actinomycete sequences (15, 34). In our analysis using the RDP database, we found that the primer matched only 2,859 of the 8,478 actinomycete sequences in the database (good quality, from isolates, and longer than 1,200 bp). Again, however, this should have biased our analysis toward greater similarity among TRFLP traces. Thus, we believe that the clustering of Uzbekistan and New Jersey community profiles was not the result of PCR or TRFLP methodological artifacts. A final concern worth noting is the fact that the the "everything is everywhere" hypothesis is, in fact, quite difficult to disprove because of the possibility that many bacteria in each sample are present at very low levels which are below the limit of detection of commonly used techniques.

Our findings imply that actinomycete distributions are exceedingly patchy and/or that the actinomycetes are highly endemic. "Patchy" refers to a pattern of distribution in which species are found over a large geographic range but individuals are concentrated (i.e., occur in groups). Many individuals of a species are present at specific locations, while a large proportion of the range of the species is not populated at any one time. Patchiness leads to a pattern in which there is a high degree of differentiation in the actinomycete communities in soil samples obtained at locations that are relatively close to each other (meters to kilometers). Highly diverse and patchily distributed actinomycetes could thus yield frequency distribu-

FIG. 3. Neighbor-joining dendrogram of DNA sequences showing relationships of sequenced PKS gene fragments from Uzbek soils. The four most diverse PKS PCR products (as determined by TRFLP) were cloned, 96 white colonies (24 colonies from each reaction) were picked, and their inserts were sequenced. All sequences exhibiting more than 95% sequence identity were grouped into clades and numbered. The most closely related sequences found in the GenBank database, as well as sequences that were recovered from the New Jersey soil samples (34), are also included. The numbers indicate bootstrap values calculated using 10,000 bootstrap replicates. Only bootstrap values greater than 65% are shown. Each clade consists of sequences that are >95% identical. Open hexagons, sequences recovered from Uzbek soil samples; shaded boxes, samples recovered from New Jersey soil samples.

tion curves similar to those shown in Fig. 1. The results of studies of the distribution of other microbes in the environment support the notion that bacterial distributions can be highly patchy, particularly in soil and sediment. For example, in a study of sulfate-reducing sediment bacteria collected on four continents, a patchy distribution that had evolved in a homogeneous background was proposed as one mechanism to explain the geographic structure observed for biodegradative populations (24). Similarly, in a study of the spatial distribution of crenarchaeal assemblages in mesophilic soil habitats it was found that different phylotypes dominated patches of soil in a uniform agricultural field, producing a mosaic distribution of different crenarchaeal phylotypes (27). When the biogeography of the purple nonsulfur bacterium *Rhodopseudomonas palustris* was investigated, it was observed that individual genotypes of *R. palustris* were detected only locally and exhibited a lognormal distribution along a 10-m transect (21). The workers concluded that *R. palustris* in freshwater marsh sediment is composed of locally distinct ecotypes with highly patchy distributions. In a subsequent study researchers found that the distribution of *R. palustris* assemblages can be patchy even at the 1-m scale (3). However, patchiness alone cannot entirely account for the distribution of actinomycetes observed in this study. Patchy distributions do not preclude species from having global and cosmopolitan ranges; i.e., everything could still be everywhere. If truly random patchiness were the only factor contributing to a high level of variability in actinomycete communities, then no significant clustering of samples by geographic origin should occur (Fig. 2A). The clustering of TRFLP community profiles by continent implies that certain actinomycete species may be found only in certain regions of the world (endemism). The hypothesis that actinomycetes can be endemic is further supported by the observation that almost one-fifth of all peaks in our data set were detected in only one sample.

The mechanisms leading to bacterial strain differentiation and strain endemism in the environment have not been established, but several hypotheses have been put forth. Processes such as dispersal, physical isolation, and genetic drift (similar to the theory of island biogeography) have been hypothesized for the differentiation of local microbial populations (11, 22). Other authors have suggested that competitive interactions and microhabitat variability might contribute to the high degree of genotype endemism that can be detected in microbial populations (3). The degree to which these factors contribute to the distribution of actinomycetes in the environment is difficult to establish using current methods, even if a relatively large number of soil samples is analyzed (91 samples were analyzed in our study). Analysis of infrequently observed peaks (peaks found in less than 10% of samples) revealed no significant regional signal (Fig. 2D), suggesting that sampling density or the resolution of the method was not sufficient to study the distribution of the infrequently observed actinomycete phylotypes and shed light on the processes structuring the community.

In conclusion, the data shown here suggest that only minor components of actinomycete populations and secondary metabolite genes in the environment are cosmopolitan. Only a single actinomycete 16S rRNA terminal fragment was found in virtually all samples that were analyzed. Likewise, only a single

clade of 36 type II PKS genes that were detected in Uzbek soil was also detected in New Jersey soils. Based on these data, we propose that distinct populations of actinomycetes (and perhaps bacteria as a whole) can be found in soils collected from different regions of the world. Therefore, bioprospecting that targets actinomycetes and their PKS pathways would benefit from sampling soils from a wide range of geographic locations.

REFERENCES

1. **Bano, N., and J. T. Hollibaugh.** 2000. Diversity and distribution of DNA sequences with affinity to ammonia-oxidizing bacteria of the beta subdivision of the class *Proteobacteria* in the Arctic Ocean. Appl. Environ. Microbiol. **66:**1960–1969.
2. **Beijerinck, M. W.** 1913. De infusies en de ontdekking der backteriën. Müller, Amsterdam, The Netherlands.
3. **Bent, S. J., C. L. Gucker, Y. Oda, and L. J. Forney.** 2003. Spatial distribution of *Rhodopseudomonas palustris* ecotypes on a local scale. Appl. Environ. Microbiol. **69:**5192–5197.
4. **Blanchot, J., J. M. Andre, C. Navarette, J. Neveux, and M. H. Radenac.** 2001. Picophytoplankton in the equatorial Pacific: vertical distributions in the warm pool and in the high nutrient low chlorophyll conditions. Deep-Sea Res. Part I **48:**297–314.
5. **Campbell, L., H. Liu, H. A. Nolla, and D. Vaulot.** 1997. Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991-1994 ENSO event. Deep Sea Res. Part I **44:**167–192.
6. **Campbell, L., H. A. Nolla, and D. Vaulot.** 1994. The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. Limnol. Oceanogr. **39:**955–961.
7. **Castenholz, R. W.** 1996. Endemism and biodiversity of thermophilic cyanobacteria. Nova Hedwigia Beih. **112:**33–47.
8. **Cho, J.-C., and J. M. Tiedje.** 2000. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. Appl. Environ. Microbiol. **66:**5448–5456.
9. **Dykhuizen, D. E.** 1998. Santa Rosalia revisited: why are there so many species of bacteria? Antonie Leeuwenhoek **73:**25–33.
10. **Eardly, B. D., L. A. Materon, N. H. Smith, D. A. Johnson, M. D. Rumbaugh, and R. K. Selander.** 1990. Genetic structure of natural populations of the nitrogen-fixing bacterium *Rhizobium meliloti*. Appl. Environ. Microbiol. **56:**187–194.
11. **Escobar-Páramo, P., S. Ghosh, and J. DiRuggiero.** 2005. Evidence for genetic drift in the diversification of a geographically isolated population of the hyperthermophilic archaeon *Pyrococcus*. Mol. Biol. Evol. **22:**2297–2303.
12. **Fierer, N., and R. B. Jackson.** 2006. The diversity and biogeography of soil bacterial communities. Proc. Natl. Acad. Sci. USA **103:**626–631.
13. **Fulthorpe, R. R., A. N. Rhodes, and J. M. Tiedje.** 1998. High levels of endemicity of 3-chlorobenzoate-degrading soil bacteria. Appl. Environ. Microbiol. **64:**1620–1627.
14. **Garcia-Pichel, F., L. Prufert-Bebout, and G. Muyzer.** 1996. Phenotypic and phylogenetic analyses show *Microcoleus chthonoplastes* to be a cosmopolitan cyanobacterium. Appl. Environ. Microbiol. **62:**3284–3291.
15. **Heuer, H., M. Krsek, P. Baker, K. Smalla, and E. M. Wellington.** 1997. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Appl. Environ. Microbiol. **63:**3233–3241.
16. **Kitts, C. L.** 2001. Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics. Curr. Issues Intest. Microbiol. **2:**17–25.
17. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei.** 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:**1244–1245.
18. **McGuinness, L. M., M. Salganik, L. Vega, K. D. Pickering, and L. J. Kerkhof.** 2006. Replicability of bacterial communities in denitrifying bioreactors as measured by PCR/T-RFLP analysis. Environ. Sci. Technol. **40:**509–515.
19. **Moore, L. R., R. Goericke, and S. W. Chisholm.** 1995. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. Mar. Ecol. Prog. Ser. **116:**259–275.
20. **Morel, A., Y.-H. Ahn, F. Partensky, D. Vaulot, and H. Claustre.** 1993. *Pro-*

chlorocuccus and *Synechococcus*: a comparative study of their size, pigmentation and related optical properties. J. Mar. Res. **51:**617–649.

21. **Oda, Y., B. Star, L. A. Huisman, J. C. Gottschal, and L. J. Forney.** 2003. Biogeography of the purple nonsulfur bacterium *Rhodopseudomonas palustris*. Appl. Environ. Microbiol. **69:**5186–5191.

22. **Papke, R. T., N. B. Ramsing, M. M. Bateson, and D. M. Ward.** 2003. Geographical isolation in hot spring cyanobacteria. Environ. Microbiol. **5:**650–659.

23. **Partensky, F., N. Hoepffner, W. K. W. Li, O. Ulloa, and D. Vaulot.** 1993. Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated from the North Atlantic and Mediterranean Sea. Plant Physiol. **101:**295–296.

24. **Perez-Jimenez, J. R., and L. J. Kerkhof.** 2005. Phylogeography of sulfate-reducing bacteria among disturbed sediments, disclosed by analysis of the dissimilatory sulfite reductase genes (*dsrAB*). Appl. Environ. Microbiol. **71:**1004–1011.

25. **Scala, D. J., and L. J. Kerkhof.** 2000. Horizontal heterogeneity of denitrifying bacterial communities in marine sediments by terminal restriction fragment length polymorphism analysis. Appl. Environ. Microbiol. **66:**1980–1986.

26. **Selje, N., M. Simon, and T. Brinkhoff.** 2004. A newly discovered *Roseobacter* cluster in temperate and polar oceans. Nature **427:**445–448.

27. **Sliwinski, M. K., and R. M. Goodman.** 2004. Spatial heterogeneity of cren-archaeal assemblages within mesophilic soil ecosystems as revealed by PCR-single-stranded conformation polymorphism profiling. Appl. Environ. Microbiol. **70:**1811–1820.

28. **Souza, V., T. T. Nguyen, R. R. Hudson, D. Pinero, and R. E. Lenski.** 1992. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? Proc. Natl. Acad. Sci. USA **89:**8389–8393.

29. **Staley, J. T., and J. J. Gosink.** 1999. Poles apart: biodiversity and biogeography of sea ice bacteria. Annu. Rev. Microbiol. **53:**189–215.

30. **Suzuki, M. T., and S. J. Giovannoni.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62:**625–630.

31. **Torsvik, V., F. L. Daae, R. A. Sandaa, and L. Ovreas.** 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. J. Biotechnol. **64:**53–62.

32. **Torsvik, V., J. Goksoyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. **56:**782–787.

33. **Wawrik, B., L. Kerkhof, J. Kukor, and G. Zylstra.** 2005. Effect of different carbon sources on community composition of bacterial enrichments from soil. Appl. Environ. Microbiol. **71:**6776–6783.

34. **Wawrik, B., L. Kerkhof, G. J. Zylstra, and J. J. Kukor.** 2005. Identification of unique type II polyketide synthase genes in soil. Appl. Environ. Microbiol. **71:**2232–2238.