Org. Divers. Evol. 1, 179–191 (2001)

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

ORGANISMS DIVERSITY & EVOLUTION

# Geographical diversity of genomic lineages in *Bacillus subtilis* (Ehrenberg) Cohn sensu lato

**Conrad A. Istock*, Nancy Ferguson, Nancy L. Istock, Kathleen E. Duncan[1]**

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson
[1] Present address: Department of Microbiology, University of Oklahoma, Norman

Received 25 September 2000 · Accepted 26 April 2001

## Abstract

Prior genetic studies of wild isolates of *Bacillus subtilis* (Ehrenberg) Cohn from a single site in Arizona, USA, revealed four deeply separated lineages within this bacterial species traditionally defined through its physiological traits. The present study examines isolates from eight sites at varying degrees of geographical separation to assess the global genomic variation of *B. subtilis*. Sites are located in Arizona, California, and Utah in the USA, and in Mexico, Chile, Chad, Tunisia, and China. Using a random 10-bp PCR primer, RAPD DNA fingerprints were obtained for 106 isolates. From these data a UPGMA dendrogram was constructed. Six major genomic lineages separating at 9–18% similarity were found. Tree topology was virtually identical with an independently derived phylogeny using distinctly different data. Three lineages were the same as previously observed in Arizona, two were new, and one involved association of a small cluster of Tunisian isolates with the single member of a distinctive fourth lineage from Arizona, though this association was ambiguous. Four of the lineages were found on three or four continents; the others were found only on one or two continents depending on the interpretation of the ambiguous association. Within each major lineage there were cascades of sublineages. Some sublineages exhibited geographically local genomic differentiation; others mingled similar genomes from geographically distant locations. The major lineages separated at levels of genomic similarity only slightly different from those observed with random permutations of the data.

**Key words:** *Bacillus subtilis*, genomic variation, genetic cohesion, bacterial evolution, bacterial species

## Introduction

Previous studies have revealed deep subdivisions within large fields of genetic variation in traditionally recognized species of *Bacillus* Cohn and other bacteria (Duncan et al. 1989, 1994; Rossello et al. 1991; Istock et al. 1992, 1996; Welsh et al. 1993; Souza et al. 1994, 1999; Bell & Friedman 1995; Sikorski et al. 1999). For *Bacillus subtilis* (Ehrenberg) Cohn, at least, the depth of these subdivisions or genomic lineages suggests that this species – recognized traditionally by its physiological characteristics (Gordon et al. 1973) – may not be genetically cohesive (Templeton 1989, Istock et al. 1996). Genetic cohesion means that genetic variation grades smoothly among members of a species, without sharp internal boundaries and subdivisions. Since at least some *Bacillus* species undergo genetic exchange through natural transformation, with potential transfer of DNA stretches up to 50 kb (Itaya 1999), they have the potential for such genetic cohesion (Graham & Istock 1978, 1979, 1981).

Initially, we discovered genomic lineages within *B. subtilis* through allozyme allelic differences, but exactly the same major lineages appeared when we used the Randomly Amplified Polymorphic DNA technique (RAPDs) and other molecular biological methods (Duncan et al. 1994, Istock et al. 1996). One subdivision within the traditional concept of *B. subtilis* has subsequently been claimed to be genetically isolated from other *B. subtilis* and named *B. mojavensis* (see Roberts et al. 1994). Another subdivision consisting of five isolates from Death Valley, California, has been named *B. vallismortis* (see Roberts et al. 1996). Several subspecies within *B. subtilis* have also been named (Nakamura et al. 1999). Throughout this paper we use the species name *B. subtilis* in its broad, traditional sense

***Corresponding author:** Conrad A. Istock, Department of Ecology and Evolutionary Biology, Cornell University, 213 Texas Lane, Ithaca, New York 14850, USA; e-mail: cai4@cornell.edu

(Gordon et al. 1973, Sonenshein et al. 1993), for reasons that will be explained in the Discussion section.

Many bacteria may have no geographical boundaries to their dispersal, making the whole earth their potential geographical range: an expansive environmental setting for their evolution. This possibility is particularly strong for spore-forming bacteria such as *Bacillus*. Widespread dispersal can occur because *Bacillus* spores attach to minute, near-surface soil particles that become airborne and travel back to earth in precipitation and dryfall. We have isolated colonies from rain (unpublished data). Roberts & Cohan (1995) provide additional discussion of widespread *Bacillus* dispersal.

In this report we analyze the geographical distribution and structure of *B. subtilis* genomic lineages using RAPD fingerprints, extending earlier studies of the local population structure of *B. subtilis* at Tumamoc Hill, Arizona, to locations in California, Utah, Mexico, Chile, China, Tunisia, and Chad.

The questions are:

1. How are *B. subtilis* genomic lineages distributed around the world? Are they geographically local or cosmopolitan? How many are there?
2. Does genomic variation within lineages diverge locally? Or is migration so great that it swamps local differentiation?
3. To what extent are *B. subtilis* populations clonal, locally and globally?
4. Is *B. subtilis*, as traditionally defined according to physiological characteristics, a single, genetically cohesive species? Or does genetic cohesion best apply to the genomic lineages within it?
5. How might genomic lineages arise? What role may they play in the evolution of bacteria like *B. subtilis*?

## Materials and methods

### Obtaining wild isolates

All of the *B. subtilis* isolates used in this study came from spores in soil samples taken at their respective localities. We extracted isolates from soils sampled at Tucson, Arizona (Tumamoc Hill); Hill Air Force Base, Ogden, Utah; Pinacate, Sonora, Mexico; Atacama Desert, Chile; and the Sahel, Chad. Dr. Frederick Cohan provided isolates from the Mojave, Gobi, and Sahara Deserts. A total of 106 isolates was used. The original, arcane, isolate labels have been retained for consistency with other publications (Table 1).

Species identifications were made using API 20E and Rapid CH System (Analytab Products, Plainview, NY,

**Table 1.** *B. subtilis* sensu lato isolates establishing concordance between genomic lineages of this study and phylogeny obtained by Roberts & Cohan (1995)

| Genomic lineages | | | | | | |
|---|---|---|---|---|---|---|
| B | F | C′ | D | C | E | A |
| T89-48[b] | RO-E-2[a] | TU-F-7[a] | T88-10[b] | T88-8[b] | TU-A-10[a] | RO-C-2[a] |
| T89-10[b] | RO-G-4[a] | TU-B-8[a] | T89-3[b] | | TU-E-9[a] | TU-A-8[a] |
| T89-55[b] | | TU-D-6[a] | TG1-16[b] | | TU-A-7[a] | TG2-42[b] |
| T89-18[b] | | TU-E-6[a] | TT1-23[b] | | TU-E-8[a] | TG3-41[b] |
| T89-49[b] | | TU-B-10[a] | T89-6[c] | | RO-H-1[a] | |
| Tf-32[c] | | | TT1-33[c] | | RO-B-2[a] | T89-14[d] |
| T89-18[c] | | | T-88-11[c] | | RO-QQ-2[a] | |
| TG2-5[c] | | | T-89-6[c] | | IM-A-224[a] | |
| T3A14[c] | | | TU-D-8[a] | | IM-A-312[a] | |
| TU-C-6[a] | | | TU-F-6[a] | | IM-B-35[a] | |
| TU-C-7[a] | | | | | | |
| TU-C-10[a] | | | | | | |
| RO-A-4[a] | | | | | | |
| RO-DD-2[a] | | | | | | |
| RO-CC-1[a] | | | | | | |
| RO-GG-2[a] | | | | | | |
| RO-L-2[a] | | | | | | |
| RO-FF-1[a] | | | | | | |

[a] Wild isolates from Dr. Frederick Cohan's laboratory used in present study
[b] Wild isolates from our laboratory used in present study and correct in Roberts & Cohan phylogeny
[c] Wild isolates from our laboratory correctly classified in Roberts & Cohan phylogeny, but not used in present study
[d] Single misclassified isolate in Roberts & Cohan phylogeny

USA; Logan & Berkeley 1984). The isolation procedure was described in Duncan et al. (1994).

### DNA isolation

DNA from *B. subtilis* isolates was extracted using the rapid procedure of Miller et al. (1988) adapted for Gram-positive bacteria as follows. Each isolate was grown overnight in Penassy broth (Difco). Cells (1.5 ml) were collected in a microfuge tube, washed in Tris EDTA (10 mM Tris pH 8.0, 1 mM EDTA), collected again and resuspended in 567 µl Tris EDTA + 30 µl of 10% SDS + 3 µl Proteinase K at 20 mg/ml. After incubation at 37 ºC for 30 min, the cells were collected, the supernatant discarded, 600 µl of TEN (10 mM Tris pH 8.0, 0.4 M NaCl, 2 mM EDTA) added, followed by 5 min incubation in a water bath at 75–80 ºC. Next, 3 µl RNase A (10 mg/ml) were added, the cells were incubated for 15–60 min at 37 ºC, then cooled to room temperature. Saturated NaCl (0.2 ml) was added, followed by vortexing and centrifugation. The supernatant was added to 0.6 ml cold isopropanol, mixed and centrifuged again. The isopropanol was poured off, and 0.6 ml cold 70% ethanol added to wash the DNA. After gentle mixing and centrifugation, the ethanol was poured off and the tube was allowed to air dry. The DNA was resuspended in 100 µl 1mM Tris (pH 8.0).

### PCR amplification and RAPD-DNA fingerprinting

A single, random 10-mer primer, OPA-03 (sequence = AGTCAGCCAC; Operon Technologies, Alameda, CA, USA), was used to obtain DNA fingerprints for all 106 isolates. This primer produced a single "universal" band of about 2200–2300 bp in all *B. subtilis* isolates, a fragment not yet seen in other *Bacillus* species. It provides a check on isolates identified as *B. subtilis* using the API system. A second, random 10-mer primer, OPA-02 (sequence = TGCCGAGCTG), was used to provide comparisons involving a small number of isolates with results from OPA-03 fingerprinting, but data for all 106 isolates using OPA-02 is not available.

Exclusive of the template DNA, our PCR reaction mixture was made as a single cocktail to provide 15 µl for each reaction. Hence, each tube containing 10 µl of DNA suspension received 4 µl 25 mM $MgCl_2$, 6.3 µl $H_2O$, 1 µl dNTPs, 2.5 µl 10 ( PCR Buffer II, 1 µl primer, 0.2 µl Taq DNA polymerase (5 units/µl), with Taq added last. Thirty µl of mineral oil were overlayed on each reaction mixture in a 400 µl microcentrifuge tube. PCR amplification took place in a thermal cycler (Perkin-Elmer) programmed to denature at 94 ºC for 1 min, anneal at 50 ºC for 1 min, elongate at 72 ºC for 2 min, for 30 cycles, with a final cycle of 5 min at 72 ºC.

Reaction mixtures were held at 4 ºC; 5 µl of loading dye was added prior to loading of 10 µl per gel pocket. Electrophoresis was performed at 45 mA with 1 × TBE buffer in 1.5% agarose/1 × TBE gels containing 0.5 µg/ml ethidium bromide. Each gel had a 1-kb DNA ladder of standards (Life Technologies, Grand Island, NY, USA) in four lanes spaced across the gel. In addition, independent reactions for the same four Arizona isolates representing genomic lineages A–D were included in each gel (indicated by double asterisks in Fig. 1).

### Creating the RAPD data set

RAPD gel images on Polaroid type 55 negatives were scanned into digital form and analyzed using Whole Band Analyzer software (Millipore Co., Bio Image Applications, Bedford, MA, USA). In addition to matching fragments across different gels using the printed results from the image analysis, all scorings were checked visually. With one exception, fragments larger than 3000 bp were not scored because they did not amplify consistently. The exception was a fragment of approximately 3400 bp that consistently marked many lineage D isolates (Fig. 5). With the OPA-03 primer and all 106 isolates, 67 different fragments were recognized, resulting in a total of 1081 individual fragments scored. The average number of fragments per isolate was 10.2, with a variance of 6.26 and range of 4–16. With the OPA-02 primer, about 3–4 more fragments amplified on average.

To maintain consistency in scoring across many gels, the fingerprint data matrix of fragments by isolates was built up progressively. This was done by combining data from gels for one locality at a time with an initial reference set of 13 Arizona isolates (indicated by single or double asterisks in Fig. 1), and then combining data for all the localities. Once isolates from all localities were in the data set, more isolates from Arizona were added to produce a better representation of lineages A–D from that location. Two factors cause geographical variation in the number of isolates obtained: the actual abundance of *B. subtilis* spores in a soil sample, and success in obtaining high quality resolution via PCR and electrophoresis. Information concerning the reproducibility of RAPD fingerprints is presented in the Results section.

### Statistical analyses

Dendrograms were constructed using the Jaccard similarity coefficient (Jaccard 1901, Norusis 1994) and the UPGMA method in the hierarchical cluster routine of SPSS for Windows v. 6.0 (Norusis 1993). The Jaccard coefficient yields fractional values from zero to one that can also be expressed as percentages.

Matrices containing random permutations of the actual RAPD data for 106 isolates were created using Microsoft

Excel 97 (Microsoft Corp.), and dendrograms from these were constructed for comparison with the one from the actual data. The procedures for constructing such random matrices can be found in the Excel documentation. To create these random images a matrix of random rational numbers between 0 and 10 was created having the same dimensions as the actual data. Then a cutoff value among the random numbers was adjusted to produce a 0–1 matrix with a mean and variance for the number of fragments per isolate matching the data. The conversion from the matrix of numbers between 0 and 10 to a matrix of 0s and 1s used the IF-THEN statement provided in Excel. This procedure was performed with and without a single fixed or "universal" fragment (Fig. 5). Other statistical tests were calculated using S-Plus (Mathsoft, Inc.).

## Results

### Genomic lineages

Previous analysis of allozyme data for 60 wild isolates from Arizona, plus two laboratory strains (168 and W23), classified them into four major groups called A, B, C, and D, with group C a single isolate. Among the 60 wild isolates there were 55 different electrophoretic types, and five pairs of clonemates. Southern hybridization data for 25 of these isolates identified the same four groups (Duncan et al. 1994: Fig. 1). [The dendrogram in Duncan et al. (1994) used simplified labels for the isolates and laboratory strains. The same dendrogram with the original labels, ones consistent with the present paper, appears in Istock et al. (1996: Fig. 1)]. Likewise for the same 25 isolates, the OPA-03 or OPA-02 primers yield RAPD data that produce an identical UPGMA classification (Fig. 1) that illustrates the deep subdivision and extensive variation in *B. subtilis* as we understood it at the onset of the present study. Jaccard similarity coefficients for the three highest level branchings separating four major lineages were 0.07, 0.12, and 0.13 when the OPA-03 primer was used, indicating that the lineages are very different. With OPA-02 the values were 0.09, 0.12, and 0.22. Further subdivision within the lineages is extensive. The highest similarity value is 0.83, because the 25 isolates do not include any clonemates.

Figure 2 illustrates the process of combining isolates from a second location, the Atacama Desert of Chile, with the reference set in Fig. 1. Atacama isolates classify with lineages A, B, and D. Similarity values from 0.08 to 0.15 separate the four lineages, including C. Within the three major lineages the Atacama genomes form distinct sublineages, suggesting that local differentiation may have occurred after each lineage became established. Another unusual feature is the extent of clonality among the Atacama organisms. The concentration of *B. subtilis*

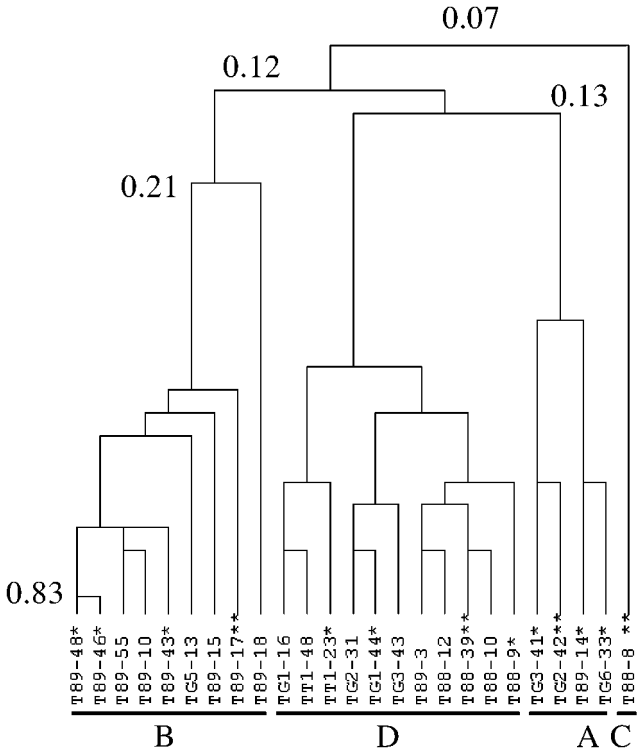


**Fig. 1.** Dendrogram based on RAPD fingerprints, illustrating the deep separation of major genomic lineages found in a sample of 25 isolates from the Tumamoc Hill site, in the Sonoran Desert of Arizona. Cascades of increasing similarity among genomes are observed within lineages A, B, and D. Decimal fractions are Jaccard similarity values. The 13 isolates with single or double asterisks were used as a reference set in the sequential assembly of the complete data set (see 'Materials and methods'). Double asterisks identify the four isolates included in all RAPD gels. The OPA-03 primer was used.

spores in these soils was approximately $10^2$–$10^3$/g, two to three orders of magnitude lower than usual in other deserts. The Atacama Desert is one of the driest places on Earth, frequently experiencing periods of several consecutive years without rain. Yet it has considerable *B. subtilis* genomic diversity, and the unusually high frequency of clonality suggests there has been local reproduction, not just accumulation of spores in precipitation or dryfall from the atmosphere. Previously we suggested that clonality might be common at low density, because even when competent for genetic exchange cells do not encounter each other frequently enough to accomplish extensive mixis (Istock et al. 1992).

### How reproducible are RAPD fingerprints?

Even with good PCR reactions and clear gels, RAPD results are often far from identical when the same isolate is fingerprinted independently. In an earlier paper we showed that pairwise similarity values can range from

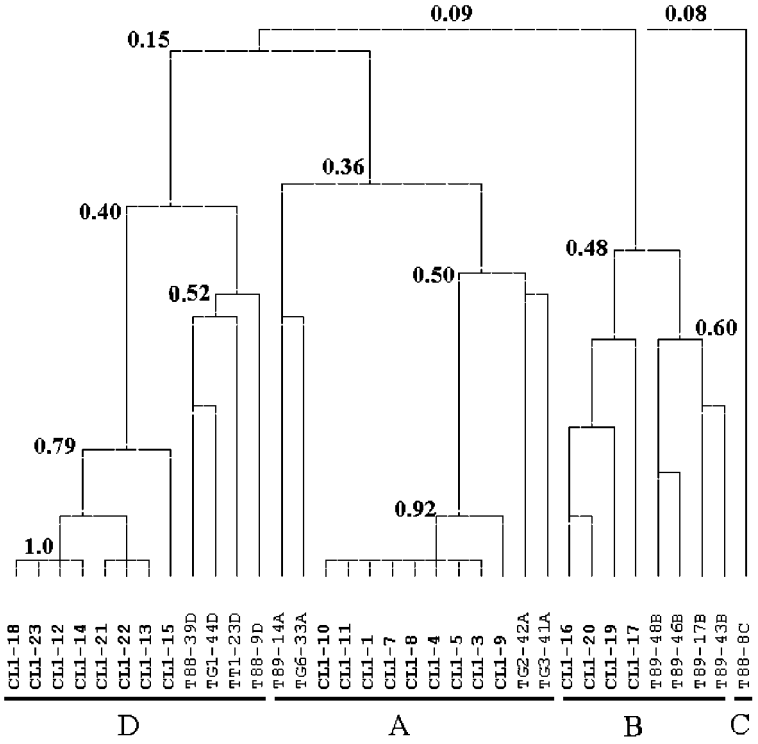


**Fig. 2.** Dendrogram after addition of isolates from Atacama Desert, Chile, to the Tumamoc reference isolates (first letter T with lineage identity appended). Figure shows interim step in process toward consistency in the full data set. See text for unusual features of *B. subtilis* in the Atacama Desert. Chile isolates (prefix CL) fall into genomic lineages A, B, and D, with an unusual degree of clonality in A and D. Decimal fractions are Jaccard similarity values. The OPA-03 primer was used.

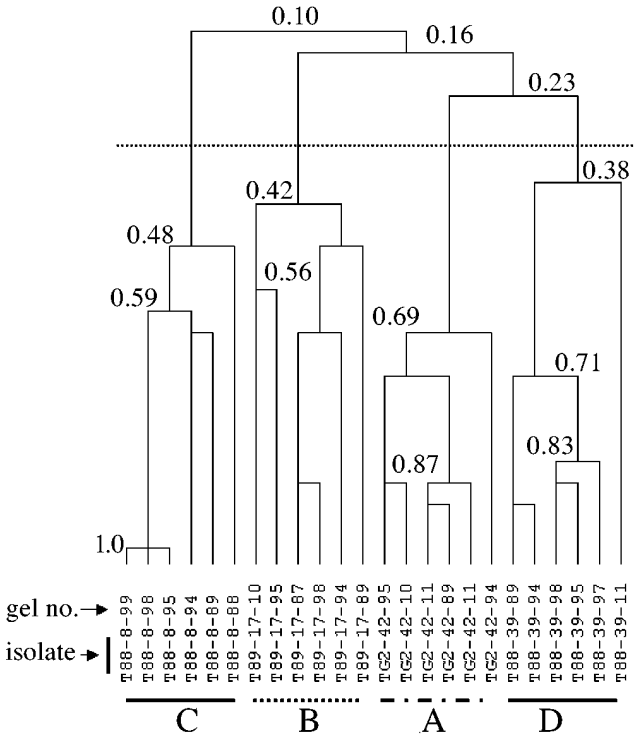


**Fig. 3.** Dendrogram examining reproducibility of RAPD fingerprints. Six independent fingerprints included for each of four isolates representing lineages A–D. Decimal fractions are Jaccard similarity values; those above dotted line establish "lineages" for correctly matched fingerprints of each of the isolates. Representative similarity values shown at various branch points. Mean overall similarity below dotted line is 0.72, with standard error of 0.04. The OPA-03 primer was used.

0.40 to 1.00 for fingerprints from the same isolate (Istock et al. 1996), indicating that on average RAPD fingerprints underestimate genomic similarities. However, this variation never caused isolates to be misclassified as to major genomic lineage because the similarity between major lineages was only 0.10–0.16 (Istock et al. 1996: fig. 3). For the present study, we analyzed six independent RAPD fingerprints from different PCR reaction mixtures and gels for the four isolates included in all gels (T88-8, T89-17, TG2-42, and T89-39 of lineages C, B, A, and D, respectively), and these were used to calculate Jaccard similarity coefficients. When a dendrogram (Fig. 3) was constructed with these data, the results were: (1) repeat fingerprints for each of the isolates clustered separately without misclassification into the four "genomic lineages," with separation of the lineages at similarities of 0.10, 0.16, and 0.23; and (2) within these "lineages" the mean similarity was 0.72, with a standard error of 0.04 and a range from 0.38 to 1.00. Again, it is clear that RAPDs frequently underestimate genomic similarities.

The rapid procedure we used to isolate DNA does not involve quantification of DNA concentrations prior to PCR amplification. Fluctuation in DNA concentrations might have engendered the variation in Fig. 3. Degradation of DNA during storage between electrophoresis runs might also contribute, but storage at 4° C makes this unlikely. Less than perfect matching of primers and template DNA may also occur to a variable degree during amplification. Nonetheless, the procedures we followed

did faithfully classify genomes belonging to deeply branching lineages, the main focus of this paper. Greater reproducibility is achieved when all DNA samples are amplified using a single PCR reaction mixture, and are run on a single gel. When isolates have identical fingerprints it is safe to conclude they are clonemates.

## The complete dendrogram

Figure 4 provides a UPGMA dendrogram for all 106 *B. subtilis* isolates. The Arizona isolates with lineage identifications A–D appended are shown in boldface type. An additional, geographically dispersed lineage E appeared. Two clonemates from the Mojave Desert of California added a distinct lineage F, the first to branch off at the top. The six major lineages separate with similarity values between 0.09 to 0.18, the latter applying to the separation between lineages C and D. The principal result is that four of the lineages (A, B, D, E) have geographically extensive representation, while C, C′, and F may have more limited distribution. Clonemates appear in isolates from Chile (three times), Mojave, Mexico (twice), Tunisia (twice), and in a striking pair from Utah and the Gobi Desert in lineage E. This low level of clonality is likely due to the inflation of differences with the RAPD technique.

The relationships between lineages and locations are summarized in Table 2. Because of wide variation in sample sizes, the present data and Fig. 4 provide only a sampling of the number and geographical distribution of *B. subtilis* lineages. Indeed, there is a strong correlation between the number of isolates for a lineage and the number of locations where it was found (Pearson r = 0.977, p = 0.0004, df = 4; Spearman ρ = 0.971, p = 0.0175). A weaker, but significant, correlation obtains between the number of isolates per location and the number of lineages found per location (Pearson r = 0.618, p = 0.05, df = 6; Spearman ρ = 0.707, p = 0.033).

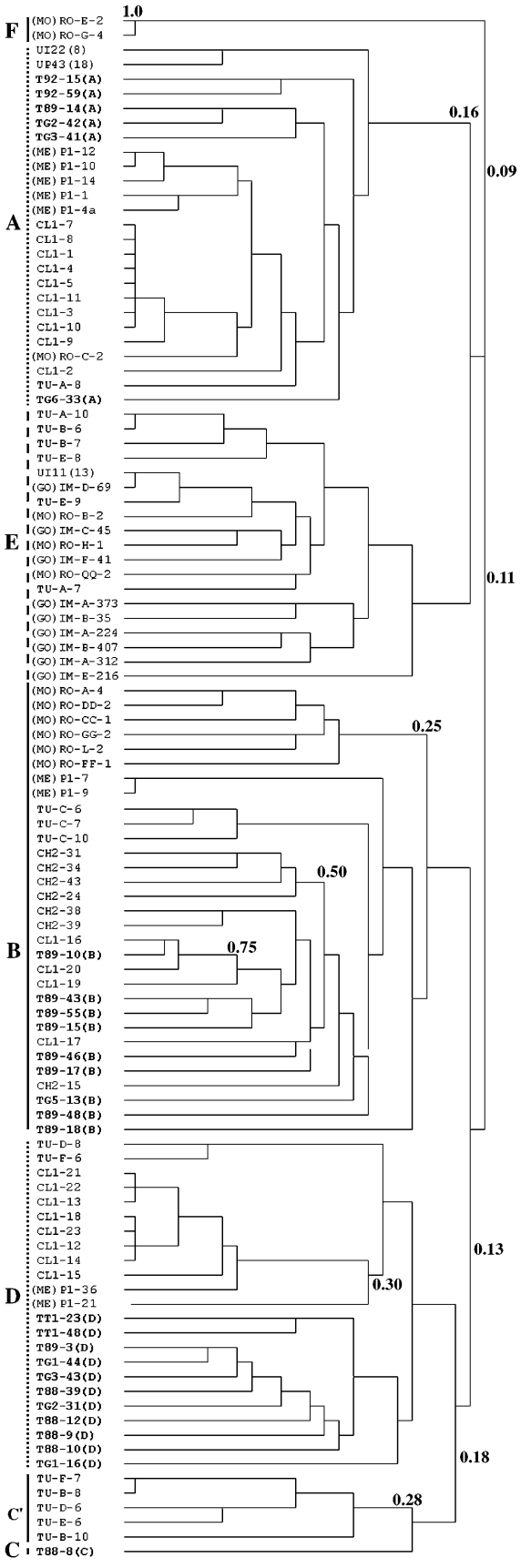


**Fig. 4.** Complete dendrogram for 106 *B. subtilis* isolates from eight areas across the world. Localities are identified by either the first letter or two of their strain labels, or by a prefix in parentheses that is not part of the strain label. Sonoran Desert isolates from Arizona, with initial letters T or TG, shown in boldface type with suffixes (A), (B), (C), and (D) to indicate the original delineation of these major lineages. Thus, Arizona isolates anchor geographically dispersed isolates in the present study to original discovery of major genomic lineages. Genomic lineages are identified by large letters at far left, with C the single Arizona isolate for this lineage, and C′ denoting a potentially related sublineage. The geographical sources of other isolates are indicated as: CH = dry grassland in Chad, CL = Atacama Desert of Chile, (GO) = Gobi Desert in China, (ME) = Sonoran Desert in Pincate, Mexico, (MO) = Mojave Desert in California, TU = Sahara Desert of Tunisia, U = grassland in Utah. Decimal fractions are Jaccard similarity values. The OPA-03 primer was used.

**Table 2.** Numbers of *B. subtilis* s. l. isolates by geographic location and major genomic lineage

| Locations | *Bacillus subtilis* lineages | | | | | | |
|---|---|---|---|---|---|---|---|
| | A | B | C/C′ | D | E | F | Row totals |
| USA, Arizona, Sonoran Desert | 6 | 9 | 1 | 11 | | | 27 |
| USA, California, Mojave Desert | 1 | 6 | | | 3 | 2 | 12 |
| USA, Utah, Hill Air Force Base grassland | 2 | | | | 1 | | 3 |
| Mexico, Sonora, Pinacate, Sonoran Desert | 5 | 2 | | 2 | | | 9 |
| Chile, Atacama Desert | 10 | 4 | | 8 | | | 22 |
| Chad, SubSaharan Sahel | | 7 | | | | | 7 |
| Tunisia, Sahara Desert | 1 | 3 | 5 | 2 | 6 | | 17 |
| China, Gobi Desert | | | | | 9 | | 9 |
| Column totals | 25 | 31 | 6 | 23 | 19 | 2 | 106 |

**Table 3.** Frequency of shared RAPD fragments in the total sample of isolates, within genomic lineages, and in geographically local clusters (see Fig. 4)

| | Number of isolates | | Average percentage of fragments shared[a] | |
|---|---|---|---|---|
| | with clusters | without clusters | with clusters | without clusters |
| Total sample | 104 | 50 | 15 | 18 |
| Major genomic lineages | | | 34.8 | 36.5 |
| A | 25 | 11 | 40 | 42 |
| B | 31 | 21 | 28 | 34 |
| C/C′ | 6 | 1 | 46 | NA[b] |
| D | 23 | 7 | 26 | 37 |
| E | 19 | 10 | 34 | 33 |
| F | 2 | 2 | 100 | 100 |
| Geographically local sublineages | | | 68.4 (all sublineages) | |
| Tunisia cluster in sublineage C′ | 5 | | 62 | |
| Chile cluster in lineage A | 9 | | 99 | |
| Mojave cluster in lineage B | 6 | | 52 | |
| Chad cluster in lineage B | 6 | | 61 | |
| Chile cluster in lineage D | 8 | | 80 | |
| Tumamoc cluster in lineage D | 8 | | 48 | |
| Pinacate cluster in lineage A | 5 | | 82 | |
| Gobi cluster in lineage E | 5 | | 61 | |
| Tunisia cluster in lineage E | 4 | | 71 | |

[a] Wilcoxon rank-sum test of clusters versus lineages without clusters included in their respective lineages, (with C′ included only among clusters, and C/C′ and F not included among lineages); $Z = 2.704$; $p = 0.0065$

[b] NA: not applicable because C has only one isolate.

Within the A, B, D, and E lineages there are instances where isolates from the same geographical location cluster in a sublineage (Fig. 4). These are listed in Table 3. By contrast, in parts of lineages A, B, and E isolates from different parts of the world are mingled, for example the putative clonemates from Gobi and Utah in the E lineage joined by isolates from Tunisia and Mojave. Other sublineages include isolates from disparate places.

A dendrogram perforce establishes a hierarchy of increasing genomic similarity from the entire sample to major lineages to smaller and smaller sublineages including local geographical clusters. We recognize major

lineages branching at similarity values of ~20% or less by the fact that, with the exception of lineage F, and possibly C and C′, they were found in several parts of the world. Sublineages form at similarity values ranging upward from 25% (Fig. 4). Genomic resemblance at various levels in the hierarchy can be measured by first calculating the fraction of isolates in a cluster that have each RAPD fragment, and then by expressing the average of these values as a percentage. This is the "average percentage of fragments shared". Table 3 lists this measure for several levels: the total sample of 106 isolates, each major lineage with and without sublineages, and for sublineages with local geographical clusters. Excluding the value of 100% for lineage F with only two isolates, there is no overlap between values for lineages and geographically local clusters, and the difference is certain to be statistically significant (Wilcoxon test, Table 3). Genomes in geographically local clusters share RAPD fragments about twice as frequently as do genomes within each major lineage taken as a whole. If the value of 100% for F is included the relation is marginally significant ($p = 0.0827$), but then the other four geographical clonal pairs should be included on the sublineage side, and significance returns ($Z = 2.0928$, $p = 0.0364$). Thus, local geographical differentiation occurs in some places and is not swamped by migration from other parts of the world.

### Randomness in divergence of lineages?

Eight random permutations of the data behind Fig. 4 were performed, four with and four without a "universal" band (Fig. 5). Without a universal band the Jaccard values for the first six major subdivisions in the random emulations ranged from 0.05 to 0.08, somewhat lower than the observed values of 0.09 to 0.18 in Fig. 4. On the other hand, with the universal band Jaccard values were 0.11–0.13, starting slightly higher than the observed values. In both cases the random values lie in a narrower range. Thus, the degrees of similarity separating the major genomic lineages in Fig. 4 are close to those expected with random divergence of major lineages, and consistent with the possibility that much of the real divergence is due to random mutation with little recombi-

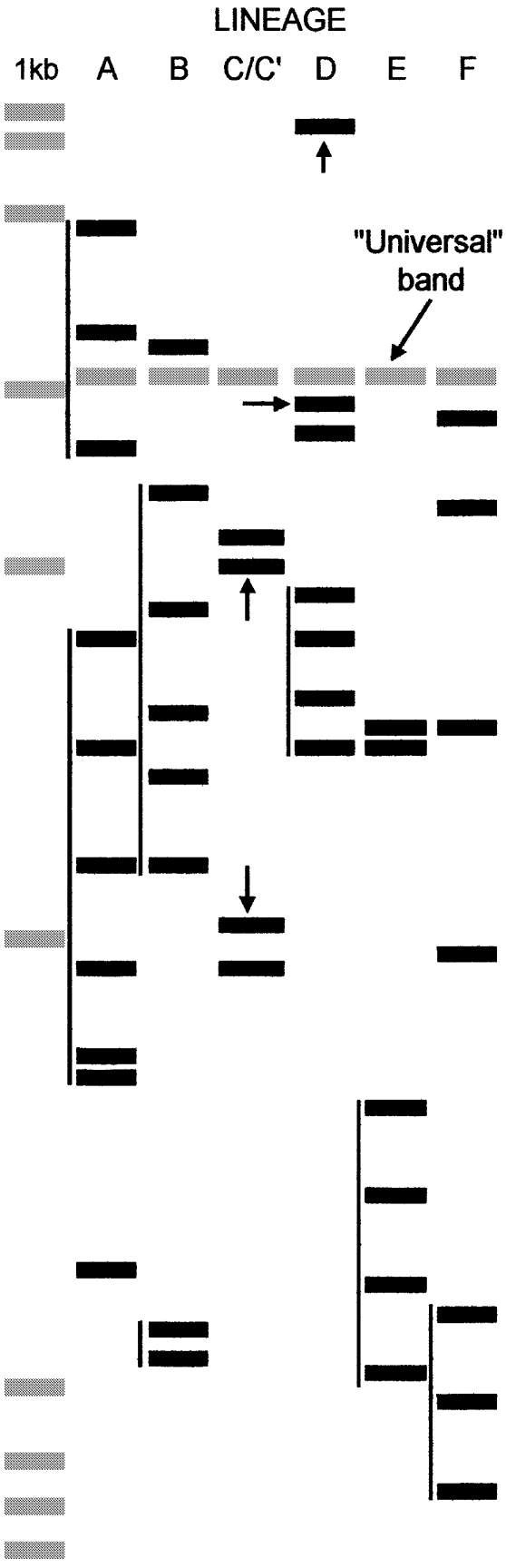


**Fig. 5.** Schematic representation of RAPD signatures of genomic lineages A–F identified with OPA-03 primer. Vertical lines and arrows indicate characteristic parts of genomic signatures. Fragment bands spaced as accurately as possible relative to standard 1-kb ladder in gray at left, but spacing of ladder not meant to be exact. Computer analysis of gel images used exact spacing and size values for the ladder. The 5090-bp band is uppermost one shown for ladder; smallest band from ladder is 236 bp. The *B. subtilis* universal fragment of approximately 2200 bp also shown in gray.

nation between major lineages, possibly augmented by purifying (periodic) selection (Palys et al. 1997).

However, within "major lineages" the random emulations differ markedly from Fig. 4. The highest similarities between any of the 106 mock isolates in random emulations range from 0.23 to 0.54, while similarity values range from above 0.5 to 1.0 within each of the major lineages found in nature. Hence, there is much greater "genetic cohesion" within the major lineages of Fig. 4 than randomness can emulate. The latter observation is also consistent with some recombination within lineages and sublineages, some purifying selection, modest clonal proliferation, or all of these together.

## Genomic lineages and genetic cohesion

Evidence of genetic cohesion within lineages was also found in the genomic signatures depicted in Fig. 5. As we accumulated RAPD gels it became clear there were repeated patterns marking each major lineage. To represent these in a systematic way any fragments that appeared in more than 40% of the isolates of a given lineage, and in at least two geographical locations, were included in Fig. 5. These criteria were used to extract what we actually saw on the gels, because while some isolates in a given lineage had the full pattern, others had parts of it missing. However, even with some diagnostic fragments missing, on visual inspection we could still predict with complete accuracy into which lineage an isolate would fall when the dendrogram was constructed; these are strong patterns. The fact that, on average, the fragments in Fig. 5 were present in 76–86% of the isolates in lineages A–E, was another way of indicating their prominence. This value was necessarily 100% in lineage F.

These signature patterns may represent remnants of the full ancestral genome for each lineage, remnants not yet erased by mutation, selection, or recombination. However, it is possible that some or all of the fragments in these patterns have arisen since the origin of the lineage, have subsequently been shared around the lineage through recombination, and have gone to fixation or near fixation. In the latter case, genomic similarity would be building up and not due simply to retention of parts of an ancestral genome. Possibly, these processes oppose each other to create an evolving signature within each major lineage.

Association of the C′ subcluster of five Tunisian isolates with the single C isolate from Tumamoc, Arizona, in Fig. 4 is problematic, with only 28% similarity between them. This association stems largely from the fact that all six isolates variously share the two fragments indicated by arrows in the C/C′ lane of Fig. 5. With fingerprints obtained using the OPA-02 primer dendrogram (not shown), C′ was not associated with lineages C or D, but with A at a similarity of 30%. When OPA-02 and OPA-03 fingerprint data were combined, the two C′ isolates for which such data were available clustered with C at 33% similarity (not shown). Hence, we use C′ as a designation for these five Tunisian isolates to recognize this weak and ambiguous relationship.

## Genomic lineages and gene phylogeny

Using restriction site variation in PCR-amplified segments of three genes (*gyr*A, a DNA gyrase; *pol*C, the DNA polymerase III; *rpo*B, the β subunit of RNA polymerase) Roberts & Cohan (1995: Fig. 1) constructed a maximum-parsimony phylogeny for 115 wild isolates. Based on previous work (Roberts et al. 1994) and this phylogeny, Roberts & Cohan split *B. subtilis* into two species: *B. subtilis* and *B. mojavensis*. A schematic representation of the branching topology of the Roberts & Cohan phylogeny appears as solid lines in Fig. 6 (their outgroups omitted).

The Roberts & Cohan study included some of our Sonoran Desert isolates and we included some of their isolates from the Gobi, Mojave, and Sahara Deserts in the present study (Table 1, Fig. 4). Although there was no prearranged focus on a common set of isolates in the

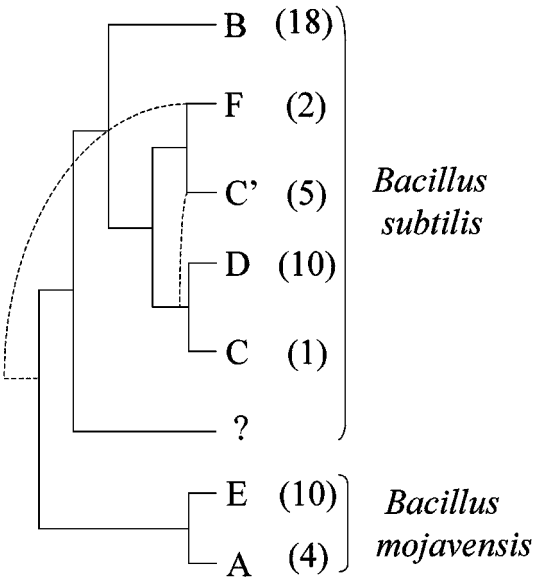


**Fig. 6.** Comparison of phylogenies from Roberts & Cohan (1995) and present study. Solid lines provide simplified version of topology from Roberts & Cohan, using restriction site analysis of portions of three genes. Letters A–F associate exact relationship of genomic lineages found in present study (Fig. 4). Numbers of wild isolates that support concordance are listed in parentheses. Recognition of *B. subtilis* and *B. mojavensis* as new species within *B. subtilis* sensu lato is due to Roberts & Cohan. The "?" indicates distinctive set of isolates from Death Valley, California, possibly an additional major lineage, named *B. vallismortis* by Roberts et al. (1996).

two studies, there was sufficient overlap in the isolates used to result in a remarkable concordance between the genomic lineages of Fig. 4 and the Roberts & Cohan phylogeny (Fig. 6). Lineages A–F of Fig. 4 appear clearly in the Roberts & Cohan phylogeny, along with one additional potential lineage (the "?" in Fig. 6) comprising five distinct isolates from Death Valley, California, subsequently named *B. vallismortis* by Roberts et al. (1996). Nakamura et al. (1999) have given subspecific names to the B and D lineages: *B. subtilis* subsp. *subtilis*, and *B. subtilis* subsp. *spizizenii*, respectively.

There were 50 out of 51 isolates (Table 1) that matched our genomic lineages to the main clades and subclades in the Roberts & Cohan phylogeny. Even the single member of lineage C (T88-8) also appeared alone in that phylogeny. A single mismatch occurred: isolate T89-14 was the most extreme member of the clade within the Roberts & Cohan phylogeny that matches our B lineage. However, it was a member of the A lineage in two earlier genetic analyses using different methods, allozyme electrophoresis and Southern hybridization (Duncan et al. 1994; where T89-14 = S89-13), as it was with RAPDs (Figs. 1, 2, and 4). Figure 6 includes the separation of *B. subtilis* and *B. mojavensis*, but for simplicity we have not added the subspecies names assigned by Nakamura et al. (1999).

Using our notation, each of the major lineages from Fig. 4 is shown in Fig. 6 where they match the Roberts & Cohan phylogeny. Numbers in parentheses are the number of isolates from both analyses concordant for each clade/lineage. The binomial probability for only one mismatch out of 51 trials is $\sim 10^{-41}$ using a probability of 6/7 for a wrong association of an isolate between lineage and clade in each trial.

There remain three discrepancies between our results and those of Roberts & Cohan (1995). (1) In Fig. 4 the most divergent lineage is F, as indicated by the dotted line to the base of the tree in Fig. 6, while it was internal to a "strain 168 group" in the *B. subtilis* section of the Roberts & Cohan phylogeny. (2) C′ is more closely allied with C and D in Fig. 4, not with F as in the Roberts & Cohan phylogeny (in their "strain W23 group" within *B. subtilis*), although the correct relationship remains ambiguous. A short dotted line in Fig. 6 indicates the possible relation of C′ with C and D. (3) Another potential problem lies with lineages A and E. Isolates from both these lineages were included under *B. mojavensis* where they actually form separate subclades in the Roberts & Cohan phylogeny, and A and E have only 16% genomic similarity based on RAPD fingerprints (Fig. 4). Since RAPDs tend to underestimate similarity it will be interesting to obtain additional estimates of genomic similarity between A and E with other methods, because the results will bear directly on the operational problem of delineating *Bacillus* species.

### Phenotypic variation

As in previous work (Duncan et al. 1994), we have been unable to find any correspondence between genomic lineage structure and the distribution of phenotypic traits among isolates assessed using 60 API metabolic/biochemical tests. In the current instance, 144 isolates of *B. subtilis*, identified in the traditional way (Gordon et al. 1973, Logan & Berkeley 1984), were used to construct a phenotypic dendrogram (not shown) similar to the smaller one in Duncan et al. (1994). Nothing corresponding to the underlying genomic lineages was observed. With the exception of a few isolates branching off at 61–78% phenotypic similarity, almost all isolates are 80–100% alike. A similar absence of phenotypic structure related to genetic structure was found in studies of *Pseudomonas stutzeri* (Lehmann & Neumann) Sijderius (see Rossello et al. 1991, Sikorski et al. 1999).

Such results suggest that the large divergences among genomic lineages may be phenotypically and ecologically neutral, including the separation of *B. subtilis* from *B. mojavensis*, and lineages A and E within *B. mojavensis* as well. However, we do not know if the API tests, or other physiological attributes evaluated in the laboratory, provide a meaningful assessment of ecological differentiation in nature. The lives of bacterial cells in natural soils involve many conditions not captured by such tests, e.g., predators, soil chemistry, soil structure, moisture, bacteriophages, antibiotics, and plant root exudates. *B. subtilis* (sensu lato) certainly encounters some ecological restrictions in nature. We failed to isolate it from numerous soil samples from mesic habitats in Costa Rica. Surprisingly, we also failed to isolate it from soils of the Judaean and Negev Deserts.

## Discussion

### Geographical distribution of lineages

Genomic lineages A, B, D, and E each occurred in samples from three or four continents. They have cosmopolitan distributions, and must have dispersed all around the planet from their sites of origin. However, members of these lineages may not be detected everywhere, even if present. Given vagaries in the sampling of soil bacterial populations – the populations are so vast – only the most abundant types are likely to be detected. An exception seems to be lineage C; despite additional sampling at Tumamoc Hill, Arizona, we found no other members of this lineage, and none were closely allied with it in the Roberts & Cohan (1995) gene phylogeny. Lineage C′ is possibly a distant branch of C, but the evidence is not strong, and it might also involve past recombination between C and A. Pending further sam-

pling, it remains uncertain whether C, C′, and F have global distributions. Lineages close in time and space to their origin or extinction will not have global ranges.

## Are the major *B. subtilis* lineages genetically cohesive species?

As it is traditionally recognized, *B. subtilis* turns out to be a set of deeply separated genomic lineages, each with genetic cohesion via shared elements in their genomic architectures. Each lineage could be considered a species and given a binomial designation, as has been done in part (Roberts et al. 1994, 1996). In contrast, the traditional physiological characteristics used to define *B. subtilis* as a single species provide phenotypic cohesion that may or may not mask ecological differences between the lineages in nature.

Our study is clearly preliminary. We do not know how many lineages there are in the world, although the remarkable concordance between our analysis and that of Roberts & Cohan (1995) suggests that the number may not be large. We do not know how rapidly major lineages originate, spread geographically, diversify within themselves, and disappear. Certainly, *B. subtilis* in the traditional sense is a set of highly divergent genomic lineages. However, there remains uncertainty about where the boundaries of genetically cohesive species lie; e.g., the previously undetected major lineages A and E within *B. mojavensis*. Even within the major lineages there are strongly differentiated sublineages that could be viewed as species. The same problem was addressed by Rossello et al. (1991) based on their study of "genomovar" (= major lineage) diversity of *Pseudomonas stutzeri*. On the basis of DNA hybridization data, they were compelled to still agree with the view of Palleroni et al. (1970) that no useful division of *P. stutzeri* into species is yet possible. The more recent analysis of Sikorski et al. (1999) – using RAPDs, other PCR-based fingerprinting methods, and allozyme variation – leads to the same conclusion. It is striking in the latter authors' dendrogram from RAPD data that the major lineages of *P. stutzeri* separate at similarities of 0.08–0.09, values similar to those for the most divergent of *B. subtilis* major lineages in Fig. 4.

We recognize that some subdivisions can be so distinct genetically that naming them would be appropriate and useful, as with the group I and group II subdivisions we found previously in *B. licheniformis* (Weigmann) Chester (Duncan et al. 1994, Istock et al. 1996).

The patterns in Fig. 4 also alert us to the fact that complete genome sequencing of one strain, or even a few isolates, may not adequately explore the broad genomic variation of *B. subtilis* as traditionally defined. The single complete DNA sequence available for *B. subtilis* used strain 168 (Kunst et al. 1997). Based on allozyme variation strain 168 is a member of the B genomic lineage, but it is a fairly extreme outlier showing only about 60% similarity with 27 wild isolates from Arizona that also belong to the B lineage; only one wild isolate was similarly divergent (Duncan et al. 1994, Istock et al. 1996). In addition, strain 168 is derived from the "Marburg" strain that was irradiated to produce metabolic mutants in the 1940s (Burkholder & Giles 1947, Kunst et al. 1997). Hence, it may be an anomalous representative of *B. subtilis*. Future sequencing should involve wild isolates from the major genomic lineages. Similarly, the major lineages may or may not conform to the extensive genetic map for strain 168 (Anagnostopoulos et al. 1993), and it is likely from the RAPD variation within *B. subtilis* that physical maps obtained with restriction enzymes (Itaya 1993) will vary among genomic lineages. Gene presence and ordering may vary as well (for example see Økstad et al. 1999).

## Local genomic differentiation versus global interspersion of lineages

We observe multiple isolates from the same place in the same sublineage, and in geographically exclusive local clusters within sublineages. This is found in every one of the major lineages to some degree. It is also apparent in the Roberts & Cohan (1995) phylogenetic tree wherein isolates from Arizona, Mojave, Death Valley, Gobi, and Tunisia, respectively, form separate subclades in several parts of the tree. Geographical differentiation cannot be entirely the result of the fact that a local sublineage has a more recent ancestor than its entire major lineage, because each of these sublineages shares common ancestors with neighboring ones. Local differentiation is consistent with neutral mutation and genetic drift, or localized genetic exchange of mutational changes, or exchange of variation shaped by natural selection favoring ecological specializations.

Along with some local differentiation there are equally impressive examples of sublineages that combine isolates from different parts of the world (Fig. 4). The modest amount of clonality observed suggests that mutational divergence, and possibly genetic exchange, is common within major lineages and sublineages of *B. subtilis*. However, caution is warranted here because RAPDs tend to seriously underestimate genomic similarity. Because of this, we are reluctant to attempt a more detailed analysis of the patterns among sublineages, using a method such as AMOVA (Schneider et al. 2000).

## Potential modes of evolution of major lineages

Milkman's clonal frame hypothesis asserts that unusually fit bacterial genotypes appear at some place on earth, proliferate rapidly, disperse widely, and subsequently

undergo mutational and recombinational alterations that progressively efface the original genomic frame (Milkman & Stoltzfus 1988, Milkman & Bridges 1990). It is a process of biological diversification that begins in absolute sympatry; the descendant lineage springs from and is initially embedded in its ancestral population until it disperses. The clonal frame hypothesis offers one mode of evolution to explain why Figs. 1 and 4 appear as they do. A different mode posits the steady appearance of new DNA sequence patterns that through drift, positive selection, purifying selection, or all three, along with recombination, fashion new shared sequences within genomic lineages. These opposing modes, perhaps simultaneous and continuous during bacterial evolution, would inexorably create and test new haploid genomes against a global environmental kaleidoscope. Evolution in these organisms may involve a flow of genomic lineages through time and across the earth; each lineage perhaps arising in absolute sympatry within an ancestral population, dispersing, diversifying, possibly giving rise to new lineages, and ultimately disappearing. At this point these ideas merely provide some models for bacterial diversification. Along with other models, such as long-term stability of major and perhaps ancient lineages, they await rigorous testing.

Crucial issues for the future are the extent to which the major lineages still exchange genetic information, the number of such lineages derived from common ancestry that exist, and how enduring and stable they are over short or long stretches of evolutionary time.

## Acknowledgements

Dr. and Mrs. Vas Aposhian took the Atacama Desert soil samples. Mr. Joe Tabor collected the soil samples from the Sahel. Dr. Frederick Cohan provided isolates from Gobi, Sahara, and Mojave Desert soils. We are grateful to each of them for their generous assistance. We thank Drs. Julia Bell, Frederick Cohan, Gregory Krukonis, and two reviewers for exceptionally helpful comments on drafts of this paper. This research was supported by grant DEB-9214040 from the U.S. National Science Foundation. Permission to import foreign soils was granted under a license from the U.S. Department of Agriculture.

## References

Anagnostopoulos, C., Piggot, P. J. & Hoch, J. A. (1993): The genetic map of *Bacillus subtilis*. Pp. 425–461 in: Sonenshein, A. L., Hoch, J. A. & Losick, R. (eds) *Bacillus subtilis* and Other Gram-Positive Bacteria. American Society for Microbiology, Washington, D.C., USA.

Bell, J. A. & Friedman, S. B. (1995): Genetic structure and diversity within local populations of *Bacillus mycoides*. Evolution 48: 1678–1714.

Burkholder, P. R. & Giles, N. H. (1947): Induced biochemical mutations in *Bacillus subtilis*. Am. J. Bot. 33: 345–348.

Duncan, K. E., Istock, C. A., Graham, J. B. & Ferguson, N. (1989): Genetic exchange between *Bacillus subtilis* and *Bacillus licheniformis*: variable hybrid stability and the nature of bacterial species. Evolution 43: 1585–1609.

Duncan K. E., Ferguson, N., Kimura, K., Zhou, X. & Istock, C. A. (1994): Fine-scale genetic and phenotypic structure in natural populations of *Bacillus subtilis* and *Bacillus licheniformis*: implications for bacterial evolution and speciation. Evolution 48: 2002–2025.

Gordon, R. E., Haynes, W. C. & Pang, C. H.–N. (1973): The Genus *Bacillus*. Agricultural Handbook 427. Agricultural Research Service, USDA, Washington, D.C., USA.

Graham, J. B. & Istock, C. A. (1978): Genetic exchange in *Bacillus subtilis* in soil. Mol. Gen. Genet. 166: 287–290.

Graham, J. B. & Istock, C. A. (1979): Gene exchange and natural selection cause *Bacillus subtilis* to evolve in soil culture. Science 204: 637–639.

Graham, J. B. & Istock, C. A. (1981): Parasexuality and microevolution in *Bacillus subtilis* populations in soil. Evolution 35: 954–963.

Istock, C. A. Duncan, K. E., Ferguson, N. & Zhou, X. (1992): Sexuality in a natural population of bacteria – *Bacillus subtilis* challenges the clonal paradigm. Mol. Ecol. 1: 95–103.

Istock, C. A., Bell, J. A., Ferguson, N. & Istock, N. L. (1996): Bacterial species and evolution: theoretical and practical perspectives. J. Ind. Microbiol. 17: 137–150.

Itaya, M. (1993): Physical map of the 168 chromosome. Pp. 463–471 in: Sonenshein, A. L., Hoch, J. A. & Losick, R. (eds) *Bacillus subtilis* and Other Gram-Positive Bacteria. American Society for Microbiology, Washington, D.C., USA.

Itaya, M. (1999): Genetic transfer of large DNA inserts to designated loci of the *Bacillus subtilis* 168 genome. J. Bacteriol. 181: 1045–1048.

Jaccard, P. (1901): Distribution de la flore alpine dans le Bassin des Dranes et dans quelques regions voisines. Bull. Soc. Vaudoise Sci. Nat. 37: 241–272.

Kunst, F., Ogasawara, N., Moszer, I. & 148 co-authors. (1997): The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390: 249–256.

Logan, N. A. & Berkeley, R. C. W. (1984): Identification of *Bacillus* strains using the API system. J. Gen. Microbiol. 130: 1871–1882.

Milkman R. & Stoltzfus, A. (1988): Molecular evolution of the *Escherichia coli* chromosome. II. Clonal segments. Genetics 120: 359–366.

Milkman, R. & Bridges, M. M. (1990): Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics 126: 505–517.

Miller, S. A., Dykes, D. D. & Polesky, H. F. (1988): A simple salting out procedure for extracting DNA from human nucleated cells. Nucl. Acids Res. 16: 1215.

Nakamura, L. K., Roberts, M. S. & Cohan, F. M. (1999): Relationship between the *Bacillus subtilis* clades associated with stains 168 and W23: a proposal for *B. subtilis* subsp. *subtilis* and *B. subtilis* subsp. *spizizenii*. Int. J. Syst. Bacteriol. 49: 1211–1215.

Norusis, M. J. (1993): SPSS for Windows, Release 6.0. SPSS, Chicago, IL, USA.

Norusis, M. J. (1994): SPSS Professional Statistics, Release 6.1. SPSS, Chicago, IL, USA.

Økstad, O. A., Hegna, I., Lindböack, T., Rishovd, A. & Kolstø, A. (1999): Genome organization is not conserved between *Bacillus cereus* and *Bacillus subtilis*. Microbiology 145: 621–631.

Palleroni, N. J., Douderoff, M., Stanier, R. Y., Solanes, R. E. & Mandel, M. (1970): Taxonomy of the aerobic Pseudomonads: the properties of the *Pseudomonas stutzeri* group. J. Gen. Microbiol. 60: 215–231.

Palys, T., Cohan, F. M. & Nakamura, L. K. (1997): Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int. J. Syst. Bacteriol. 47: 1145–1156.

Roberts, M. S., Nakamura, L. K. & Cohan, F. M. (1994): *Bacillus mojavensis* sp. nov., distinguishable from *B. subtilis* by sexual isolation, divergence in DNA sequence and differences in fatty acid composition. Int. J. Syst. Bacteriol. 44: 256–264.

Roberts, M. S. & Cohan, F. M. (1995): Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. Evolution 49: 1081–1094.

Roberts, M. S., Nakamura, L. K. & Cohan, F. M. (1996): *Bacillus vallismortis* sp. nov., a close relative of *Bacillus subtilis*, isolated from soil in Death Valley, California. Int. J. Syst. Bacteriol. 46: 470–475.

Rossello, R., Garcia-Valdes, E., Lalucat, J. & Ursing, J. (1991): Genotype and phenotype diversity of *Pseudomonas stutzeri*. Syst. Appl. Microbiol. 14: 150–157.

Schneider, S., Roessli, D. & Excoffier, L. (2000): Arlequin ver. 2000: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sikorski, J., Rossello-Mora, R. & Lorenz, M. G. (1999): Analysis of genotype diversity and relationships among *Pseudomonas stutzeri* strains by PCR-based genomic fingerprinting and multilocus enzyme electrophoresis. Syst. Appl. Microbiol. 22: 393–402.

Sonenshein, A. L., Hoch, J. A. & Losick, R. (1993): *Bacillus subtilis* and Other Gram-Positive Bacteria. American Society for Microbiology, Washington, D.C., 987 pp.

Souza, V., Eguiarte, L., Avila, G., Capello, R., Gallardo, C., Montoya, J. & Pinero, D. (1994): Genetic structure of *Rhizobium etli* biovar *phasioli* associated with wild and cultivated beans (*Phaseolus vulgaris* and *Phaseolus coccineus*) in Morelos, Mexico. Appl. Environ. Microbiol. 60: 1260–1268.

Souza, V., Rocha, M., Valera, A. & Eguiarte, L. E. (1999): Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. Appl. Environ. Microbiol. 65: 3373–3385.

Templeton, A. R. (1989): The meaning of species and speciation: a genetic perspective. Pp. 3–27 in: D. Otte & Endler, J., (eds) Speciation and its Consequences. Sinauer, Sunderland, MA, USA.

Welsh, J., Pretzman, C., Postic, D., Saint Girons, I., Baranton, G. &. McClelland, M. (1993): Genomic fingerprinting by arbitrarily primed polymerase chain reaction resolves *Borrelia burgdorferi* into three distinct phyletic groups. Intl. J. Syst. Bacteriol. 42: 370–375