High-Density Microarray of Small-Subunit Ribosomal DNA Probes

Kenneth H. Wilson,1 Wendy J. Wilson,2 Jennifer L. Radosевич,2 Todd Z. DeSantis,2 Vijay S. Viswanathan,2 Thomas A. Kuczmański,2 and Gary L. Andersen2*

Veterans Affairs Medical Center and Duke University Medical Center, Durham, North Carolina 27710,1 and Lawrence Livermore National Laboratory, Livermore, California 945502

Received 26 September 2001/Accepted 2 January 2002

Ribosomal DNA sequence analysis, originally conceived as a way to provide a universal phylogeny for life forms, has proven useful in many areas of biological research. Some of the most promising applications of this approach are presently limited by the rate at which sequences can be analyzed. As a step toward overcoming this limitation, we have investigated the use of photolithography chip technology to perform sequence analyses on amplified small-subunit rRNA genes. The GeneChip (Affymetrix Corporation) contained 31,179 20-mer oligonucleotides that were complementary to a subalignment of sequences in the Ribosomal Database Project (RDP) (B. L. Maida et al., Nucleic Acids Res. 29:173-174, 2001). The chip and standard Affymetrix software were able to correctly match small-subunit ribosomal DNA amplicons with the corresponding sequences in the RDP database for 15 of 17 bacterial species grown in pure culture. When bacteria collected from an air sample were tested, the method compared favorably with cloning and sequencing amplicons in determining the presence of phylogenetic groups. However, the method could not resolve the individual sequences comprising a complex mixed sample. Given these results and the potential for future enhancement of this technology, it may become widely useful.

Originally developed to better understand the evolution of life on our planet (11, 29), small-subunit (SSU) rRNA and ribosomal DNA (rDNA) sequence analysis has had many applications. The approach continues to be used to explore the biological diversity of our planet (1, 2, 20) and the biological composition of various ecosystems (23, 27). It has had a profound effect on microbial taxonomy (6, 28), has enabled the identification of uncultured pathogens (22, 25), and has been used in the setting of clinical microbiology (9). The Ribosomal Database Project (RDP) (16) has catalogued around 23,000 of these sequences, classifying them at various hierarchical phylogenetic levels (Fig. 1). The detailed classification can be downloaded from the website http://rdp.cme.msu.edu/download/SSU_rRNA/.

There are several platforms currently used to analyze rRNA and rDNA sequences. By growing a pure culture of an organism or cloning mixed rDNAs, it is possible to obtain a pure template for dideoxy sequencing; sequences obtained in this way can be compared with sequences in large databases. Alternatively, oligonucleotide probes directed at phylogenetically specific sequences can be used to identify the SSU rDNA of phylogenetic clusters of organisms. Both of these approaches are for the most part serial, i.e., they can be used to identify only one organism at a time by detecting one sequence at a time, though it is possible to differentially label oligonucleotide probes for in situ hybridization (4). In addition, the approach of cloning amplicons and then sequencing them is laborious and time-consuming.

It has recently become possible to make arrays of oligonucleotide probes in various formats (10, 19). The photolithography chip is a particularly powerful approach, allowing simultaneous testing of over 250,000 oligonucleotide probes. By confirming the presence of relatively long DNA sequences, these chips are used for various applications, including rapid determination of mRNA expression profiles (15, 18) and detection of the rpoB gene sequence for simultaneous identification of the species and rifampin sensitivity of mycobacteria (12).

We report here the use of a photolithography chip that displays an array consisting of oligonucleotides complementary to rDNA sequences found in a subalignment of the RDP. SSU rDNA sequences are highly conserved, the overall degree of similarity between two sequences correlating with their distance from one another in the RDP hierarchical phylogenetic scheme. Because many organisms with sequences not included in RDP share sequence with organisms that were included, it was possible to glean information about unknown environmental organisms whose SSU rDNA sequences were neither included in RDP version 5.0 nor deliberately placed on the chip.

MATERIALS AND METHODS

Photolithography chip design. RDP version 5.0 contained 3,286 aligned SSU rRNA sequences (17). The subalignment used for this work was defined by Escherichia coli positions 1409 to 1491. There were adequate data in this segment to include 1,945 prokaryotic sequences and 431 eukaryotic sequences on the chip. The region used is bounded on both ends by universally conserved segments that can be used as PCR priming sites to amplify any SSU rDNA (14, 26). Thus, primers amplified both prokaryotic and eukaryotic rDNA. Approximately half of the possible overlapping 20-mers were used, allowing extensive redundancy. Where the database contained an ambiguous base, the oligonucleotides with sites complementary to the ambiguously sequenced nucleotide were synthesized with the base omitted. Each array cell containing an rDNA probe (probe cell) was paired with a control cell in which the oligonucleotide was identical except that the 11th nucleotide consisted of a base that would mismatch the targeted rDNA sequence. Thus, the oligonucleotide in the control cell served as a mismatched control for nonspecific hybridization. The entire array consisted of 62,358 oligonucleotides synthesized on the chip, each with a copy number of around 107. The number of probes for individual sequences contained in RDP
version 5.0 ranged from 0 to 70, depending on the completeness of sequence information between E. coli positions 1409 and 1491. All the synthesized probes that were designed to match one sequence were designated the probe set for that sequence.

Bacterial strains. Most bacterial strains tested were type strains obtained from the American Type culture collection (ATCC). These included Bacteroides vulgatus (ATCC 8462), Brucella abortus (ATCC 23448), Clostridium difficile (ATCC 9864), Acidobacterium capsulatum (ATCC 5196), Pseudomonas aeruginosa (ATCC 10145), Hafnia alvei (ATCC 13337), Legionella pneumophila (ATCC 33152), Campylobacter jejuni (ATCC 33560), Fusobacterium nucleatum (ATCC 25566), Staphylococcus aureus (ATCC 12600), Bacillus thuringiensis (ATCC 10792), Streptococcus salivarius (ATCC 7073), Mycoplasma pneumoniae (ATCC 15531), and Clostridium perfringens (ATCC 13124). The Sterne strain of Clostridium tetani was obtained from the Rocky Mountain Laboratory, Hamilton, Mont. The reaction mixtures contained 10 mM (total) deoxynucleoside triphosphates (dNTPs) and the Advantage 2 PCR enzyme system (Clontech, Palo Alto, Calif.). A reaction mix included Adaptor DNA polymerase and TaqStart according to the 10X Advantage 2 PCR buffer, consisting of 400 mM Tricine-KOH (pH 8.7), 150 mM potassium acetate, 35 mM magnesium acetate, 37.5 μg of bovine serum albumin (BSA) per ml, 0.05% Tween 20, and 0.05% Nonidet P-40 per reaction. PCR conditions were 1 cycle of 1 min at 95°C, followed by 35 cycles of 1 min at 94°C, 30 s at 45°C, and 45 s at 68°C. A 10-min incubation at 68°C was the final step.

Preparation of sample for GeneChip assay. A fragmentation working mix was prepared of 0.25 U of fragmentation reagent (Affymetrix, Santa Clara, Calif.) with equal volume of 20 mM EDTA, 2.5 U of calf intestine alkaline phosphatase (Gibco-BRL Life Technologies, Rockville, Md.), and 0.5 mM Tris-acetate (pH 8.2). Concentrations of PCR products were determined on agarose gels with a DNA mass ladder (Life Technologies, Rockville, Md.). A 45-μl solution of 2 μg of unpurified ampiclon DNA in distilled H2O was prepared, 5 μl of working mix was added, and fragmentation was performed on a PE 9800 thermocycler for 12 min at 25°C and 10 min at 95°C. Then 21.5 μl of fragmented products was biotin labeled for 2 h at 37°C in a reaction mix of 50 U of terminal transferase (Boehringer Mannheim), 1.8 mM CoCl2, and 57 μM Renaissance biotin-N6-ddATP (NEI Life Science Products) in 1X terminal transferase buffer, for a final volume of 35 μl.

GeneChip processing. GeneChip probe arrays were prehybridized with 200 μl of RapidHyb buffer (Amersham, Arlington Heights, Ill.) containing 0.8 M tetramethylammonium chloride (TMAC) in a custom hybridization rotisserie oven (Stovall Life Science, Greensboro, N.C.) for 10 min at 45°C and 60 rpm. Target cocktail [Rapid Hyb/TMAC solution, 0.06 mM control oligonucleotide B1 (Affymetrix, Inc.) and 80 to 770 ng of biotin-labeled target sample] was heated for 5 min at 99°C and centrifuged for 5 min at 14,000 rpm. The cocktail supernatant fraction was equilibrated to 45°C and added to a GeneChip array that was drained of prehybridization solution.

Hybridization was performed in the rotisserie oven for 3 h at 45°C and 60 rpm. The microarray was washed and stained in a GeneChip fluids station 400 (Affymetrix, Inc.). Four washes with 0.9 M NaCl-60 mM NaF-Po4-6 mM EDTA-6X SSPE-0.005% Triton X-100, pH 7.4, were followed by two washes with 1X SSPE-0.005% Triton X-100, pH 7.4. The array was then stained with 2 μg of streptavidin phycoerythrin-R conjugate (SAPE) (Molecular Probes, Inc., Eugene, Oreg.) and 0.95 mg of acetylated BSA per ml in 6X SSPE-0.005% Triton X-100 for 15 min at 40°C, followed by three washes with 1X SSPE-0.005% Triton X-100 at 25°C.

Scanning and data analysis. The array was subsequently scanned by a 488-nm argon-ion laser (GeneChip Scanner 500; Molecular Dynamics) with emission detection above 520 nm. The phycoerythrin emission was 576 nm. The scan was recorded as pixel data. Images were analyzed using standard Affymetrix software (GeneChip Analysis Suite, version 3.3). Background hybridization and signal noise were treated as two separate phenomena and were calculated as suggested by the array manufacturer. Background cells were identified as those producing intensities in the lowest 2% of all intensities. The average intensity of the background cells was subtracted from the fluorescence intensity of all cells. The noise value (N) was the variation in pixel intensity signals observed by the scanner as it read the array surface. The standard deviation of the pixel intensities within each of the identified background cells was divided by the square root of the number of pixels comprising that cell. The average of the resulting quotients was used for N in the calculations described below.

In Affymetrix probe pair scoring, two thresholds must be surpassed for a probe to be scored as positive or negative. The empirically derived criteria for a probe to be scored as positive were as follows: (i) the intensity of fluorescence from the perfectly matched probe cell (PM) was greater than or equal to 1.25 times the intensity from the mismatched control cell (MM), and (ii) the difference in fluorescence from the perfectly matched probe cell (PM) was greater than or equal to 1.25 times the intensity from the mismatched control cell (MM). A negative PM required the MM intensity to exceed the PM according to the same two criteria. When intensities between probes in a pair did not differ sufficiently to exceed the thresholds in either direction, the pair was scored as indeterminate. Thus, a signal could be read as positive, negative, or indeterminate.

Our chip contained 62.558 unique oligonucleotides based on sequences in the RDP 5.0 database. We updated our analysis abilities by reassigning the PM probes using the sequences in the RDP 8.1 database. Positive signals were matched by computer to oligonucleotides found in RNA sequences occurring in RDP version 8.1. As noted above, ambiguities in the database were handled by synthesizing oligonucleotides in which the ambiguous bases were deleted. In practice, unless these deleted bases occurred near the end of the oligonucleotide, many of the probes had a lower signal intensity than other probes in the probe.
set. For this reason, all potential probe sites containing these deletions were omitted from further analyses.

For amplicons derived from pure cultures, a positive result for an RDP sequence required that all probes for that sequence give a positive signal. Because some sequences were incomplete, another requirement was that at least 27 probes for a sequence be present on the chip. The sequences meeting these parameters were empirically loosened. A positive result for an RDP sequence ranked as described above.

All sequences listed on the same line shared the same complete set of probes complementary to their rDNA sequence. For analyses of hierarchical phylogenetic levels corresponding to chip data, the software parameter was used.

The 368 sequenced clones passed quality tests of Phred 20 (base call error probability < 10^-2) and were compared with sequences in the RDP 8.1 database (4,000 rpm). The supernatant fraction was discarded, and the two wash pellets were combined for a final volume of 200 μL. Genomic DNA was extracted directly from the filter concentrate using a derivation of the MoBio UltraClean Soil DNA kit (MoBio Laboratories, Solana Beach, Calif.) modified to use silica beads instead of garnet beads to lyse cells. DNA obtained in this way was amplified as described above.

Analysis of air sample. The resulting air DNA amplicons were split into two samples, and microbial composition was analyzed by two methods. One sample was hybridized to three chips from different production lots on different days. The other sample was cloned using the pGEM-T Easy cloning vector system (Invitrogen, San Diego, Calif.) and transformed into Escherichia coli DH10B.

Individual cloned rDNA fragments were sequenced using ABI 3700 instruments (Applied Biosystems, Foster City, Calif.) and assembled using Phred and Phrap (7, 8). The 368 sequenced clones passed quality tests of Phred 20 (base call error probability < 10^-2) and were compared with sequences in the RDP 8.1 database (16) for phylogenetic identification by measuring sequence similarity with known rDNA sequences as described above. RDP level 3 codes were assigned.

RESULTS

Organisms in pure culture. Table 1 shows the rank order of matching RDP sequences selected by the scanner’s software.

set. For this reason, all potential probe sites containing these deletions were omitted from further analyses.

For amplicons derived from pure cultures, a positive result for an RDP sequence required that all probes for that sequence give a positive signal. Because some sequences were incomplete, another requirement was that at least 27 probes for a sequence be present on the chip. The sequences meeting these criteria were rank-ordered with respect to the mean difference in fluorescence intensity between all probe cells and matched control cells for each sequence. For analyses of hierarchical phylogenetic levels, the RDP sequence ranked first was used.

It was anticipated that in contrast to rDNA sequences from pure cultures, mixed rDNA sequences sampled directly from the environment would often fail to perfectly match an rDNA sequence in RDP. For that reason, software parameters were empirically loosened. A positive result for an RDP sequence required 24 probes to be present on the chip, of which 22 were scored as positive. Because some sequences were incomplete, another requirement was that at least 27 probes for a sequence be present on the chip. The sequences meeting these parameters were empirically loosened. A positive result for an RDP sequence ranked as described above.

All sequences listed on the same line shared the same complete set of probes complementary to their rDNA sequence. For analyses of hierarchical phylogenetic levels, the software parameter was used.

The 368 sequenced clones passed quality tests of Phred 20 (base call error probability < 10^-2) and were compared with sequences in the RDP 8.1 database (4,000 rpm). The supernatant fraction was discarded, and the two wash pellets were combined for a final volume of 200 μL. Genomic DNA was extracted directly from the filter concentrate using a derivation of the MoBio UltraClean Soil DNA kit (MoBio Laboratories, Solana Beach, Calif.) modified to use silica beads instead of garnet beads to lyse cells. DNA obtained in this way was amplified as described above.

Analysis of air sample. The resulting air DNA amplicons were split into two samples, and microbial composition was analyzed by two methods. One sample was hybridized to three chips from different production lots on different days. The other sample was cloned using the pGEM-T Easy cloning vector system (Invitrogen, San Diego, Calif.) and transformed into Escherichia coli DH10B.

Individual cloned rDNA fragments were sequenced using ABI 3700 instruments (Applied Biosystems, Foster City, Calif.) and assembled using Phred and Phrap (7, 8). The 368 sequenced clones passed quality tests of Phred 20 (base call error probability < 10^-2) and were compared with sequences in the RDP 8.1 database (16) for phylogenetic identification by measuring sequence similarity with known rDNA sequences as described above. RDP level 3 codes were assigned.

**RESULTS**

**Organisms in pure culture.** Table 1 shows the rank order of matching RDP sequences selected by the scanner’s software.
for each of the bacterial species tested. Fifteen of 17 species were correctly ranked number one. In several instances, sequences from other species had an equal score with the one being tested. For all such species, the sequences in the sub-alignment were identical to the sequences of the test species. Bacillus anthracis was not included in the RDP version used to design the chip. However, B. anthracis has the same rDNA sequence (GenBank accession number AF290552) as the Bacillus cereus strain selected by the software. Thus, B. anthracis was not miscalled. The sequences of Eubacterium tarantellae was not miscalled. The sequences of Fusobacterium nucleatum was not miscalled. The sequences of Bacillus anthracis has the same rDNA sequence as RDP's entry for Bacillus anthracis/H11032. The software selected the correct species by GeneChip. Thus, the PCR product was cloned, generating this assessment.

**Table 2. Phylogenetic placement of sequences detected by GeneChip**

<table>
<thead>
<tr>
<th>Organism(s)</th>
<th>Phylocode (expected result)</th>
<th>Accuracy of RDP phylogenetic placement at each resolution level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st 2nd 3rd 4th 5th 6th</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single organism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. vulgatus</td>
<td>2.15.1.2.8</td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>A. capsulatum</td>
<td>2.25.3.6</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>B. abortus</td>
<td>2.28.1.6.18</td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>R. prowazekii</td>
<td>2.28.1.5.5</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>L. pneumophila</td>
<td>2.28.3.1.9</td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>C. bartetti</td>
<td>2.28.3.9.2</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>2.28.3.15</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>H. alvei</td>
<td>2.28.3.27.14</td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>C. jejuni</td>
<td>2.28.5.5.3</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>F. nucleatum</td>
<td>2.29.5</td>
<td>+ + + + na</td>
</tr>
<tr>
<td>S. aureus</td>
<td>2.30.7.12.1</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>B. anthracis</td>
<td>2.30.7.12.4</td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>B. thuringiensis</td>
<td>2.30.7.12.4</td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>S. salivarius</td>
<td>2.30.7.12.6</td>
<td>+ + + + + +</td>
</tr>
<tr>
<td>M. pneumoniae</td>
<td>2.30.8.4.3</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>C. difficile</td>
<td>2.30.4.5.8</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>M. pneumoniae</td>
<td>2.30.8.4.3</td>
<td>+ + + + +</td>
</tr>
</tbody>
</table>

* Organism(s) from which DNA was extracted for amplification of SSU rDNA.
* Various levels of phylogenetic information concerning the corresponding sequence(s) rank-ordered by GeneChip software. +, correct call; +*, correct sequence corresponds to more than one phylocode; na, not applicable, taxon does not exist. Where applicable, all were identified correctly to the fourth-level grouping, although in the case of H. alvei and S. salivarius sequences outside of the expected fourth-level grouping gave as strongly positive a result as did sequences within the correct fourth-level grouping.
quences in the database outside of the third-level phylogenetic designation of the test organism met criteria for a positive result, and in all cases, the mean difference in fluorescence was far lower than the mean difference for the correct group (Fig. 3).

In order to evaluate the usefulness of the chip for analysis of more complex samples, a mixture of DNA from *Rickettsia prowazekii* and *Mycoplasma pneumoniae* was tested. As shown in Table 1, each component was detected as accurately as when tested alone.

Testing of microbes filtered from air. Figure 4 shows the reported relative abundance of sequences belonging to the various phylogenetic clusters detected by clone sequencing compared with groups found by chip analysis. As expected, both eukaryotic and bacterial sequences were detected by both methods. Overall, there was close agreement between the chip analysis and the sequencing results in detecting the presence of phylogenetic groups as defined by RDP. Eight of 10 phylogenetic clusters detected by the chip were confirmed as present when clones were sequenced, a result that was reproduced by all three production lots of the chip. Gamma proteobacteria and the group eubacteria were each detected in small amounts on the chip in all three array lots but were not sampled by cloning. Twenty-eight clones (7.6% of total) contained rDNA from organisms that could not be placed into phylogenetic groups based on similarity to any aligned sequence in RDP release 8.1 and were also not detected on the chip. Appropriate phylogenetic clusters for none of these sequences were represented in RDP release 5.0, from which the chip was designed. Most of these sequences had low similarity to all sequences in RDP release 8.1 as well and tended to be most similar to sequences that have yet to be placed into phylogenetic groups. Thus, they appear to be derived from novel organisms.

**DISCUSSION**

rRNA sequence analysis was initially conducted by cataloging short rRNA breakdown products (11). This approach led to startling revelations concerning the phylogeny of life forms on this planet (20, 29). The approach became more powerful with application of the Sanger method to sequence rRNA and rDNA by using conserved oligonucleotide primers (14). Sampling and sequencing were simplified considerably by the application of PCR to rDNA analysis (3, 26). PCR also made it possible to easily obtain complex samples of mixed rDNAs
from natural environments (23, 24). Despite these advances, the use of rDNA analysis for microbial ecology remains problematic.

The complexity of microbial ecosystems and the number of unique ecosystems hinder efforts to explore biodiversity. For instance, it is estimated that one cubic millimeter of earth may contain as many as 10,000 different microbial types, and the microbial content of soil varies considerably from site to site (5). Furthermore, complex microbial ecosystems can change over time as environmental conditions vary. If these ecosystems are to be understood, they must be studied in the dimension of time as well as the dimensions of space. The need for increased throughput is obvious. By making it possible to conduct sequence analyses on many uncloned rDNAs at once, the photolithography chip has the potential to meet this need. The present study investigated the feasibility of this approach.

For samples known to consist of single organisms represented in RDP version 5.0, the approach rapidly determined the identity of most organisms tested, though sometimes more than one organism in the database had the same sequence. Occasionally the sequence of a close relative of the organism being tested was ranked above that of the test organism. In all instances the discrepancy could be explained on theoretical grounds. The chip was highly accurate at identifying the presence of sequences categorized at higher phylogenetic levels. Occasionally a low-amplitude signal was detected for an unrelated phylogenetic cluster. In these instances, it is unclear whether cross-hybridization occurred or there was contamination of the DNA sample or PCR product. When the chip results were compared with cloning and sequencing of rDNAs derived from an air sample, overall the two representations of phylogenetic groups occurring in the natural sample were highly similar. Cloned rDNAs not detected by the chip were all from novel organisms with no close relatives represented in RDP. Phylogenetic clusters detected on the chip but not sampled as rDNA clones appeared as low-amplitude fluorescent signals. In general, there was not a good correlation between the numbers of clones from a phylogenetic group and the intensity of the fluorescent signal from that group. Thus, the two methods may be complementary.

The photolithography chip reported here has several potential applications including the potential for public health applications, clinical microbiological testing, monitoring bioremediation, and monitoring for biological weapons. There is currently interest in microbiological observatories to monitor microbial populations around the world. While the present

FIG. 4. Relative abundance of microbes detected in an air sample as determined by rDNA amplification followed both by cloning with sequencing and by GeneChip analysis. rDNA sequences as determined by each method were sorted into taxa. Vertical bars represent mean difference of replicate hybridizations on three chips from different manufacturers’ production lots. Bar height is proportional to the greatest mean difference in fluorescence found among the sequences in that group. Abbreviations for taxa and RDP phylogenetic numerical codes: Sphingobacterium group, Sphingo (2.15.2); α-proteobacteria subdivision, α-proteo (2.28.1); β-proteobacteria subdivision, β-proteo (2.28.2); γ-proteobacteria subdivision, γ-proteo (2.28.3); gram-positive high G+C bacteria, high G+C (2.30.1); gram-positive Eubacterium and relatives, Eubact (2.30.4); gram-positive Bacillus-Lactobacillus-Streptococcus subdivision, Bacil (2.30.7); gram-positive Clostridium and relatives, Clostr (2.30.9); Ascomycota fungi, Asco (3.9.1); and Basidiomycota fungi, Basidio (3.9.2). The 28 novel sequences did not correspond well to any phylogenetic group in the RDP database.
version of the SSU chip does not appear to be capable of differentiating mixed populations of highly similar organisms, it could still play a useful role in this endeavor. At the very least, it could probably determine whether or not the microbial communities found at a given site had changed over time and, if so, the probable makeup of the changes found.

The efficiency and usefulness of the SSU rDNA chip can be enhanced in several ways. Often, bacterial species have unique rDNA sequences at various sites (13, 21). A strong hybridization signal at such a site would have more than average significance. With further development of the software, it should be possible to weight such sites more highly than phylogenetically less informative sites. The subalignment used for the present chip is bounded by universally conserved sequences and is small enough that the segment can be amplified from even highly degraded DNA. However, it may not be realistic to expect always to identify species based on an 80- to 90-bp region of a conserved gene for which the reference database contains only one or two examples per species. Though the region studied is moderately variable, other areas are more variable and are worthy of consideration. As an alternative approach, it should be possible to amplify large segments of rDNA and probe only the most variable regions. Finally, the rDNA database is now much larger; the sequence information includes many mixed populations of highly similar organisms, and the SSU chip is not capable of differentiating mixed populations of highly similar organisms, including a new genus and five new species, and standards for the identification of organisms. Int. J. Syst. Bacteriol. 49:415–440.


