# Bacterial Species Determination from DNA-DNA Hybridization by Using Genome Fragments and DNA Microarrays

JAE-CHANG CHO<sup>1</sup> AND JAMES M. TIEDJE<sup>1,2\*</sup>

Center for Microbial Ecology<sup>1</sup> and Department of Microbiology and Molecular Genetics,<sup>2</sup> Michigan State University, East Lansing, Michigan 48824

Received 28 November 2000/Accepted 18 May 2001

Whole genomic DNA-DNA hybridization has been a cornerstone of bacterial species determination but is not widely used because it is not easily implemented. We have developed a method based on random genome fragments and DNA microarray technology that overcomes the disadvantages of whole-genome DNA-DNA hybridization. Reference genomes of four fluorescent *Pseudomonas* species were fragmented, and 60 to 96 genome fragments of approximately 1 kb from each strain were spotted on microarrays. Genomes from 12 well-characterized fluorescent *Pseudomonas* strains were labeled with Cy dyes and hybridized to the arrays. Cluster analysis of the hybridization profiles revealed taxonomic relationships between bacterial strains tested at species to strain level resolution, suggesting that this approach is useful for the identification of bacteria as well as determining the genetic distance among bacteria. Since arrays can contain thousands of DNA spots, a single array has the potential for broad identification capacity. In addition, the method does not require laborious cross-hybridizations and can provide an open database of hybridization profiles, avoiding the limitations of traditional DNA-DNA hybridization.

Bacterial identification methods currently used include analysis of morphological, physiological, biochemical, and genetic data. In the last two decades, molecular methods, especially 16S rRNA gene sequencing, have been a reliable aid to the identification of diverse bacteria. Although the 16S rRNA method has served as a powerful tool for finding phylogenetic relationships among bacteria (24), because of its molecular clock properties and the large database for sequence comparison, the molecule is too conserved to provide good resolution at the species and subspecies levels (2, 4, 9, 20, 23). The relationship between 16S rRNA gene similarity and percent DNA-DNA reassociation is a logarithmic function in which the sequence similarity within a species (>70% DNA relatedness) is expected to be >98% (3), and the similarity among different species in a genus, e.g., fluorescent Pseudmonas spp., is 93.3 to  $\sim$ 99.9% (11). Considering the high sequence conservation and relative standard errors at 98 and 90% sequence similarities of 19 and 8%, respectively (5), 16S rDNA analysis results on closely related strains could be inaccurate and inconsistent with the results obtained by other methods. Incongruity between genome structure and 16S rDNA sequence similarity was also reported (8). Since many important ecological and clinical characteristics of bacteria, such as pathogenicity, competitiveness, substrate range, and bioactive molecule production, vary below the species level, methods with higher resolution than that of 16S rDNA sequencing are needed.

DNA-DNA hybridization is one method that provides more resolution than 16S rDNA sequencing, and the 70% criterion (22) has been a cornerstone for describing a bacterial species. In spite of these values, the method is not popular. Major disadvantages are the laborious nature of pairwise cross-hybridizations, the requirement for isotope use, and the impossibility of establishing a central database.

Here, we propose a new approach to identify and type bacteria based on genomic DNA-DNA similarity that eliminates the above disadvantages. The method takes advantage of the capacity provided by microarray technology (6). Bacterial genomes are fragmented randomly, and representative fragments are spotted on a glass slide and then hybridized to test genomes. Resulting hybridization profiles are used in statistical procedures to identify test strains. Importantly, a database of hybridization profiles can be established. This paper describes this method and its evaluation with previously characterized fluorescent *Pseudomonas* strains.

### MATERIALS AND METHODS

**Bacterial strains and DNA extraction.** The *Pseudomonas* strains used in this study were *Pseudomonas fluorescens* ATCC 13525<sup>T</sup>, *P. fluorescens* ATCC 17397, *P. fluorescens* ATCC 17400, *P. fluorescens* ATCC 17467, *P. fluorescens* ATCC 33512, *P. marginalis* LMG 5039, *P. chlororaphis* ATCC 9447, *P. chlororaphis* ATCC 17811, *P. aureofaciens* ATCC 13985<sup>T</sup>, *P. putida* ATCC 12633<sup>T</sup>, *P. aeruginosa* ATCC 15692, and *P. aeruginosa* ATCC 17429. All strains were routinely cultivated at 30°C in nutrient broth medium (Difco, Detroit, Mich.). Genomic DNAs from the strains were extracted and purified using Genomic Tips (Qiagen, Valencia, Calif.) with Genomic DNA Buffer Set (Qiagen). The concentration of DNA was determined by UV spectrophotometry and with SpotCheck (Sigma, St. Louis, Mo.).

**Microarray fabrication.** Genomic DNAs from four fluorescent *Pseudomonas* strains (*P. fluorescens* ATCC 13525<sup>T</sup>, *P. chlororaphis* ATCC 9447, *P. putida* ATCC 12633<sup>T</sup>, and *P. aeruginosa* ATCC 15692, hereafter termed the reference strains) were fragmented by bead beating to ensure randomness, and the fragments were size fractionated (1 to 2 kb) by agarose gel electrophoresis. The QIAquick gel extraction kit (Qiagen) was used to elute and purify DNA from the agarose gel. The genomic DNA fragments were inserted into pPCR-Script Amp vector (Stratagene, La Jolla, Calif.) and then PCR amplified with the T3-T7 promoter primer set. Amplified genomic DNA fragments were purified with a QIAquick 8 PCR purification kit (Qiagen) and quantified with PicoGreen (Molecular Probes, Eugene, Oreg.).

Purified DNAs were resuspended (200 ng/ $\mu$ l) in 3× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and printed (ca. 1 nl/spot) on CMT-GAPS

<sup>\*</sup> Corresponding author. Mailing address: Center for Microbial Ecology, Plant and Soil Sciences Bldg., Michigan State University, East Lansing, MI 48824. Phone: (517) 353-9021. Fax: (517) 353-2917. E-mail: tiedjej@msu.edu.



FIG. 1. Format of the array used in this study. Genome fragments from four reference strains were spotted in duplicate (upper and lower halves). Positive (yeast gene *STE*) and negative (yeast gene *ACT*, lambda DNA, and water) controls are indicated by circles and squares, respectively. Pseudocolors indicate the ratio (R) between the Cy3 channel and Cy5 channel (green, R > 1; yellow, R = 1; red, R < 1).

amino silane-coated slides (Corning Co., Corning, N.Y.). Fragments from *P. fluorescens, P. chlororaphis, P. putida* and *P. aeruginosa* (92, 90, 96, and 60, respectively) were spotted in duplicate (Fig. 1). The yeast *STE* gene (pheromone receptor gene; GenBank accession no. M12239) was spotted as a positive control, and the yeast *ACT* gene (actin gene; GenBank accession no. L00026), lambda DNA, and water were spotted as negative controls. PCR primer pair STE3F1 (CCC CTT CAAAAT TGG AGC TTG C) and STE3R1 (CCC CCT TTA GCA TGG CAT TCA) and pair ACT1F1 (GAT GGA GCC AAA GCG GTG A) and ACT1R1 (GCG CTT GCA CCATCC CAT T) were used to amplify yeast genes *STE* and *ACT*, respectively.

After drying, the slides were processed with the succinic anhydride blocking method according to the manufacturer's protocol and stored at room temperature until use.

Genomic DNA labeling and hybridization. Genomic DNAs  $(1 \mu g)$  from all the strains listed, including reference strains, were labeled with FluoroLink Cy3-dCTP (Amersham Pharmacia, Piscataway, N.J.) by random priming (High Prime; Roche, Indianapolis, Ind.) and used as test DNAs. Mixtures of genomic DNA (1  $\mu g$ ) from the four reference strains (1:1:1) used for microarray fabrication were labeled with FluoroLink Cy5-dCTP (Amersham Pharmacia) and used as reference DNA for signal ratio calculation (Cy3-test/Cy5-reference). Yeast gene *STE* (10 ng) was included in each labeling reaction as a positive control as well as an internal standard (IS; Cy3-IS and Cy5-IS) for labeling-efficiency correction.

The arrays were prehybridized in prehybridization buffer  $(3.5 \times SSC, 0.1\%$  sodium dodecyl sulfate [SDS], 10 mg of bovine serum albumin per ml) for 20 min at 65°C, hybridized with approximately 1  $\mu$ g of Cy3- and Cy5-labeled DNA mixture (1:1) in hybridization buffer (3× SSC, 0.1% SDS, 0.5 mg of yeast tRNA per ml) at 65°C overnight, and then washed once with primary wash buffer (0.1× SSC, 0.1% SDS) at room temperature for 5 min and twice with secondary wash buffer (0.1× SSC) for 5 min.

Scanning and data processing. Hybridized arrays were scanned with a Gene-Pix 4000 laser scanner (Axon, Foster City, Calif.). Laser lights of wavelengths at 532 and 635 nm were used to excite Cy3 and Cy5 dye, respectively. Fluorescent images were captured as multi-image-tagged image file format and analyzed with GenePix Pro 3.0 software (Axon). The ratio (R) of the extent of hybridization between test DNAs and reference DNAs was derived from a median value of pixel-by-pixel ratios. By using this approach to calculate R, nonspecific signals, which appear in both wavelength images, had less of an effect than when the mean values of a whole spot were used.

Hybridization signal ratios (*R*) between test DNA and reference DNA (Cy3-Test/Cy5-Ref) were calculated and corrected with the correction factor (c = Cy5-IS/Cy3-IS) from the internal standard (yeast gene *STE*) (corrected signal ratio  $R' = c \times [Cy3$ -Test/Cy5-Ref]). Spearman correlation coefficients (r) were calculated to find relationships between hybridization patterns and transformed to a percentage scale. Unweighted arithmetic average clustering (UPGMA) was used for hierarchical data ordination. For characterizing the shape of hybridization signal distribution, the evenness (E) value of each spotted genome fragment was calculated based on information theory using  $E = (-\Sigma p \log p)/\log q$  (7, 15),

where *p* is the relative proportion of hybridization signal ratio (*R'*) and *q* is the total number of hybridizations performed. Since the distribution of the calculated *E* values was highly skewed (skewness = -0.855), the *E* values were normalized by arc cosine transformation. The arc cosine-transformed evenness value,  $\theta_E$ , was used to represent the degree of conservation of each genome fragment.

Microsoft Excel, Systat (SPSS, Chicago, Ill.), and NTSYS-pc (Exeter Software, East Setauket, N.Y.) were used for all statistical calculations.

## **RESULTS AND DISCUSSION**

**Reproducibility.** The ratio of Cy5 to Cy3 incorporation (Cy5-IS/Cy3-IS) during the DNA labeling was  $1.04 \pm 0.32$  for all experiments. An incorporation ratio (c = Cy5-IS/Cy3-IS) obtained for each microarray was used as a correction factor for hybridization signal calibration (corrected signal ratio  $R' = c \times [Cy3$ -Test/Cy5-Ref]). The correction factor, however, did not affect the correlation coefficient calculation, since the correlation coefficient is independent of any constant (e.g., c).

In order to test the reproducibility of array hybridization, seven arrays were hybridized to genomic DNAs of *P. fluorescens* ATCC 13525<sup>T</sup> (three times), *P. putida* ATCC 12633<sup>T</sup> (two times), and *P. aeruginosa* ATCC 15692 (two times). Figure 2 shows the scatter plot representation of triplicate hybridization profiles of *P. fluorescens* ATCC 13525<sup>T</sup>. The arrays hybridized to *P. fluorescens* ATCC 13525<sup>T</sup> (triplicate), *P. putida* ATCC 12633<sup>T</sup> (duplicate), and *P. aeruginosa* ATCC 15692 (duplicate) showed similarity values of >97.5% (r > 0.949, P < 0.0001), 95.3% (r = 0.906, P < 0.0001), and 94.1% (r = 0.882, P < 0.0001), respectively.

**Resolution.** Regression analysis showed a good agreement between DNA-DNA reassociation values and the similarity coefficients obtained from this study (Fig. 3). The coefficient of determination  $(r^2)$  was 0.713. Order 1 of linear relationship and the regression coefficient (slope, 0.718) indicate that the microarray method is similar in resolution to the whole-genome DNA-DNA hybridization method. The two methods lost their linear relationships below 50% DNA-DNA similarity, which approximately corresponds to a 60% similarity coefficient obtained by the DNA microarray method.

A similar result was observed with the relationship between



FIG. 2. Scatter plot diagram of hybridization profiles of *P. fluorescens* ATCC  $13525^{T}$ . Results from triplicate hybridization experiments ( $r^2 = 0.94$ ) are displayed, and each axis (x, y, and z) represents the log-transformed hybridization signal ratios from each experiment.

repetitive extragenic palindromic (REP)-PCR genomic DNA fingerprint similarity and percent DNA similarity values (16). REP-PCR fingerprinting (17) lost resolution when applied to strains of less than 70% DNA similarity, indicating that REP-PCR fingerprinting is capable of resolving relationships only among very closely related strains. The DNA chip method used here showed linearity over a broader span of DNA similarity values (50 to 100%) but provided slightly less resolution at >70% DNA similarity values than for the REP-PCR finger-printing method. The microarray method, however, can still resolve closely related strains and, more importantly, provides resolution over the gap between REP-PCR fingerprinting and 16S rDNA analysis (1).

We have considered the case where different strains of the same species have differences in genome size, e.g., *Escherichia* 



FIG. 3. Relationship between previously reported whole-genome DNA similarity values (12, 13) and similarity values obtained by the microarray method. The solid line and dotted lines indicate the regression curve and 95% prediction interval, respectively.



FIG. 4. Similarity dendrogram (UPGMA) of microarray hybridization profiles of fluorescent *Pseudomonas* strains. The solid line indicates a cutoff value at which all different strains tested were resolved. The dashed line indicates species-level resolution that corresponds to 70% whole-genome DNA hybridization.

*coli* K-12 versus O157 (GenBank accession no. U00096 and AE005174, respectively). This scale of difference (1 of 5 Mb) should not invalidate our approach, although our percent similarity should be slightly higher than the average percent similarity from whole-genome DNA-DNA hybridization.

**Cluster analyses.** The overall topology of the dendrogram based on cluster analysis of similarity coefficient matrix was consistent with the phylogenetic tree obtained from 16S rDNA sequence data (11) except for *P. putida* and *P. aeruginosa* clusters (Fig. 4). The *P. aeruginosa* group clustered with the *P. fluorescens* and *P. chlororaphis* groups at a higher similarity (67.9%) than for the *P. putida* group (39.0%), of which generally shows greater 16S rDNA similarity to *P. fluorescens* and *P. chlororaphis* than to *P. aeruginosa* (11). However, a similar result to our array data was reported by Palleroni et al. (13) using DNA-DNA similarity values, where the *P. aeruginosa* group was found to be a closer relative to *P. fluorescens* group than was the *P. putida* group.

All replicate experiments showed similarity coefficients of  $\geq 94\%$  (r = 0.88), and all different strains were distinguished at similarity values of  $\leq 91\%$  (r = 0.82). Hence, similarity coefficients of < 92 to 94% (r = 0.84 to 0.88) can reliably define different hybridization groups. Using the regression equation from Fig. 2, a cutoff value of 77% was calculated to correspond to a 70% DNA similarity value to define "species" (22). This cutoff resolved the *P. fluorescens, P. chlororaphis, P. aeruginosa*, and *P. putida* species but did not resolve *P. marginalis* from *P. fluorescens* or *P. aureofaciens* from *P. chlororaphis*, but these latter pairs of species are known to be very similar by other methods. *P. aureofaciens* ATCC 13985<sup>T</sup> and *P. chlororaphis* ATCC 9447 show 85% DNA similarity (13), and the 16S rDNA similarity between *P. aureofaciens* and *P. chlororaphis* and be-



FIG. 5. Similarity dendrogram (UPGMA) of hybridization profiles of 338 genome fragments spotted on the microarray. Clusters F (98.7%), C (94.1%), A (91.8%), P (100%), and Y (100%) comprise genome fragments from the reference strains *P. fluorescens* ATCC 13525<sup>T</sup>, *P. chlororaphis* ATCC 9447, *P. aeruginosa* ATCC 15692, and *P. putida* ATCC 12633<sup>T</sup>. Clusters V to Z comprise genome fragment from different reference strains, except cluster Y.



FIG. 6. (a) Evenness value ( $\theta_E$ ) scatter diagram, with average and SD of log hybridization signal ratio. (b)  $\theta_E$  values by genome fragment, ID 1 to 92, 93 to 182, 183 to 278, and 279 to 338 originated from *P. fluorescens* ATCC 13525<sup>T</sup>, *P. chlororaphis* ATCC 9447, *P. putida* ATCC 12633<sup>T</sup>, and *P. aeruginosa* ATCC 15692, respectively. The solid line and dotted lines (horizontal) indicate average and SD, respectively.

tween *P. fluorescens* and *P. marginalis* is 99.5 and 99.9%, respectively (results from different strains) (11). *P. marginalis* is also reported to have very similar characteristics to *P. fluorescens* and was previously classified as *P. fluorescens* (10, 14, 18, 21). Hence, it appears that this method has promise for providing reliable guideline values for species or genomovar resolution.

Either the array hybridization profiles (signal ratios) or raw images (microarray scans) can be archived in a World Wide Web server to establish the central database so that researchers can compare their results with the database and consequently identify their strains, analogous to retrieval of RDP data.

The more reference strains whose genome fragments are spotted on the array and more genome fragments spotted from the reference strains, the greater the resolution and the consistency of this approach are likely to be. We used 338 genome fragments from four reference strains in our test model. Considering that the average genome size of fluorescent *Pseudo*-



FIG. 7. Proposed relationship between  $\theta_E$  value and taxonomic distance in taxonomic continuum. Taxonomic continuum is multidimensional, and hence, genetic similarity peaks could also be a multidimensional structure, but diagram is drawn as shown (two-dimensional) for convenience. Broken lines indicate the degree of conservation of genome fragments with different  $\theta_E$  values.

*monas* strains is approximately 5 Mb and that the size of the genome fragments used is 1 to 2 kb, the array used in this study samples approximately 1 to 3% of a genome. However, supposing that each spot (genome fragment) tests individual genetic characteristics quantitatively, the array performed 338 individual tests for determining the similarity coefficients for one test strain. Sokal and Sneath (19) suggest that more than 60 characters give significant reliability for similarity coefficients and enough information for numerical taxonomy. In fact, all of our similarity coefficients were statistically significant (P < 0.0001).

Cluster analysis was also performed on the hybridization patterns of all 338 spotted fragments across all strains tested (Fig. 5). Four main clusters were found at a cophenetic similarity of 70%. Main clusters F, C, A, and P mainly comprised the fragments from the four reference strains, *P. fluorescens* (98.7%), *P. chlororaphis* (94.1%), *P. aeruginosa* (91.8%), and *P. putida* (100%). Minor clusters V, W, X, and Z comprised the genome fragments from different reference strains. In a gene expression data analysis, such clusters indicate that these genes tend to turn on and off simultaneously, but the grouping in this study indicates only that the hybridization patterns of the cluster members are similar to a certain degree. If the genome fragments from the different reference strains form such a cluster, it suggests but does not confirm conserved sequences.

To conveniently find conserved and unique (variable) sequences in our fragment collection, we calculated an evenness index (*E*) (7, 15) from hybridization signal ratio profiles of each spotted genome fragment across the test strains (Fig. 6). If a fragment is extremely conserved in all test strains (e.g., rRNA genes), the angle ( $\theta_E$ ) would show its minimum value (0°). Genomic fragments showing a small angle (high evenness) tend to show high hybridization signal ratio with low standard deviation, indicating that they showed as high a hybridization signal as many genomes tested and hence can be considered conserved sequences. In contrast, genomic fragments with a large angle (low evenness) tend to show a low average signal ratio with high standard deviation, indicating that they showed appreciable hybridization signal only to the closely related strains and hence are considered variable sequences. The average angle ( $\theta_E$ ) for all data was 35.0° ± 12.5°. Fiftyone (15.1%) fragments with  $\theta_E$  values lower than 1 standard deviation (SD) below the mean (<22.5°) (Fig. 6b) showed appreciable hybridization signal (R' > 1) for the genomic DNAs from closely related species, e.g., species pair *P. fluorescens* and *P. marginalis* and pair *P. chlororaphis* and *P. aureofaciens*. The majority of these originated from two reference strains, *P. fluorescens* ATCC 13525<sup>T</sup> and *P. chlororaphis* ATCC 9447, including only five fragments from the clusters V, W, X, and Z in Fig. 5. Fragments showing appreciable hybridization (R' > 1) for all strains tested (e.g.,  $\theta_E << 10^\circ$ ) were not found on our array.

Sixty-eight (20.1%) fragments with  $\theta_E$  values of 1 SD above the mean (>47.5°) showed appreciable hybridization only when hybridized to the reference strains. The rest of the fragments (64.8%) showed an intermediate level of conservation. While four main clusters (F, P, C, and A) (Fig. 5) contain all genome fragments with  $\theta_E$  values of 22.5° to 47.5° (species level), the groups of highly variable sequences ( $\theta_E$  value > 47.5°) (strain level) are also located in the main clusters (Fig. 5). It is noteworthy that the variable and conserved sequences cannot be reliably identified by cluster analysis (Fig. 5), but are easily revealed by  $\theta_E$  values.

Using the calculated  $\theta_E$  values, we can also construct a relationship between  $\theta_E$  value and taxonomic distance (Fig. 7), where valley-shaped regions could be considered to be caused by selection pressure, resulting in subsequent speciation events. The genome fragments with low  $\theta_E$  values have almost identical sequences and are distributed over a wide taxonomic range, while the fragments with high  $\theta_E$  values are distributed over a narrow taxonomic range. When our empirical results  $(\theta_F \text{ values})$  were applied to this diagram, the degree of conservation within strain level, species level, closely related species level, and genus level correspond roughly to  $\theta_E$  values of  ${>}50^\circ\!\!,~50^\circ\!\!$  to  $20^\circ\!\!,~20^\circ\!\!$  to  $10^\circ\!\!,$  and  ${<}10^\circ\!\!,$  respectively. Additionally, a taxonomic distance  $[D_{1/\tan(\theta)}]$  can be calculated  $\{D_{1/\tan(\theta)} = 1/[\tan(\theta_E)]\}$ . The range of  $\theta_E$  values for species level (>20°) in this study resulted in a  $D_{1/\tan(\theta)}$  of 2.74, indicating a radius of taxonomic range for a species. This alternative to calculating taxonomic distance by using genome-wide analyses may be useful for delineating species, although the values would be expected to vary with microbial groups.

**Conclusions.** This study provides a proof of concept for identification of bacteria using DNA-DNA hybridization with DNA microarrays, as tested on four reference strains of fluorescent *Pseudomonas* spp. With current technology of microarray fabrication, 100,000 genomic fragments can be spotted on a chip. Hence, it is feasible to test 1,000 reference strains with 100 genome fragments from each reference strain. Such an array should be enough to cover the full taxonomic range of either gram-negative or gram-positive bacteria. This approach also appears to be useful for determining the genetic distance among bacteria as well as for identifying bacterial species.

Major improvements and advantages over the traditional DNA-DNA reassociation approach are that our method does not require cross-hybridization to find genetic relationship between test strains, does not use an isotope, and can utilize an open database of hybridization profiles when standard genome chips for bacteria are available.

#### ACKNOWLEDGMENTS

We thank Alison Murray for helpful discussion, the MSU-Arabidopsis facility, and Syed Hashsham for microarray facilities.

This research was supported by NSF grant DEB-0075564 and the Center for Microbial Ecology under NSF grant DEB-9120006.

#### REFERENCES

- Cho, J.-C., and J. M. Tiedje. 2000. Biogeography and degree of endemicity of fluorescent *Pseudomonas* in soil. Appl. Environ. Microbiol. 66:5448–5456.
- DeParasis, J., and D. A. Roth. 1990. Nucleic acid probes for identification of phytobacteria: identification of genus-specific 16S rRNA sequences. Phytopathology 80:618–621.
- Devereux, R., S. H. He, C. L. Doyle, S. Orkland, D. A. Stahl, J. LeGall, and W. B. Whitman. 1990. Diversity and origin of *Desulfovibrio* species: phylogenetic definition of a family. J. Bacteriol. 172:3609–3619.
- Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk, Jr. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. 42:166–170.
- Keswani, J., S. Orkand, U. Premachandran, L. Mandelco, M. J. Franklin, W. B. Whitman. 1996. Phylogeny and taxonomy of mesophilic *Methanococcus* spp. and comparison of rRNA, DNA hybridization, and phenotypic method. Int. J. Syst. Bacteriol. 46:727–735.
- 6. Lander, E. S. 1999. Array of hope. Nat. Genet. 21:3-4.
- 7. Legendre, P., and L. Legendre. 1998. Numerical ecology. Elsevier Science, Amsterdam, The Netherlands.
- Lessie, T. G., W. Hendrickson, B. D. Manning, and R. Devereux. 1996. Genomic complexity and plasticity of *Burkholderia cepacia*. FEMS Microbiol. Lett. 144:117–128.
- Martinez-Murcia, A. J., S. Benlloch, and M. D. Collins. 1992. Phylogenetic interrelationships of members of the genera *Aeromonas* and *Plesiomonas* as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridizations. Int. J. Syst. Bacteriol. 42:412–421.
- Misaghi, I., and R. G. Grogan. 1969. Nutritional and biochemical comparisons of plant-pathogenic and saprophytic fluorescent pseudomonads. Phytopathology 59:1436–1450.
- 11. Moore, E. R. B., M. Mau, A. Arnscheidt, E. C. Böttger, R. A. Hutson, M. D. Collins, Y. van de Peer, R. de Wachter, and K. N. Timmis. 1996. The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationship. Syst. Appl. Microbiol. 19:478–492.
- Palleroni, N., R. Kunisawa, R. Contopoulou, and M. Doudoroff. 1973. Nucleic acid homologies in the genus *Pseudomonas*. Int. J. Syst. Bacteriol. 23:333–339.
- Palleroni, N. J., R. W. Ballard, E. Ralston, and M. Doudoroff. 1972. Deoxyribonucleic acid homologies among some *Pseudomonas* species. J. Bacteriol. 110:1–11.
- Pecknold, P. C., and R. G. Grogan. 1973. Deoxyribonucleic acid homology groups among phytopathogenic *Pseudomonas* species. Int. J. Syst. Bacteriol. 23:111–121.
- Pielou, E. C. 1966. The measurement of diversity in different types of biological collections. J. Theor. Biol. 13:131–144.
- Rademaker, J. L. W., B. Hoste, F. J. Louws, K. Kersters, J. Swings, L. Vauterin, P. Vauterin, and F. J. deBruijn. 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: Xanthomonas as a model system. Int. J. Syst. Evol. Microbiol. 50:665–677.
- Rademaker, J. L. W., F. J. Louws, and F. J. de Bruijn. 1998. Characterization of the diversity of ecologically important microbes by rep-PCR fingerprinting, p. 1–26. *In* A. D. L. Akkermans, J. D. van Elsas, and F. J. de Bruijn (ed.), Molecular microbial ecology manual, suppl. 3. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Sands, D. C., M. N. Schroth, and D. C. Hildebrand. 1970. Taxonomy of phytopathogenic pseudomonads. J. Bacteriol. 101:9–23.
- Sokal, R. R., and P. H. Sneath. 1963. Principles of numerical taxonomy. W. H. Freeman and Company, San Francisco, Calif.
- Stackebrandt, E., and B. M. Goebel. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44:846–849.
- Stanier, R. Y., N. J. Palleroni, and M. Doudoroff. 1966. The aerobic pseudomonads: a taxonomic study. J. Gen. Microbiol. 43:159–271.
- Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, and H. G. Truper. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37:463–464.
- Weisburg, W. G., S. M. Barns, D. A. Pelletier, and D. J. Lane. 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173:697– 703
- 24. Woese, C. R. 1987. Bacterial evolution. Microbiol. Rev. 51:221-271.