

## Intraspecific Variation in Small-Subunit rRNA Sequences in GenBank: Why Single Sequences May Not Adequately Represent Prokaryotic Taxa

REBECCA A. CLAYTON,\* GRANGER SUTTON, PAUL S. HINKLE, JR.,  
CAROL BULT, AND CHRIS FIELDS†

*The Institute for Genomic Research, Gaithersburg, Maryland 20878*

**Small-subunit rRNA (SSU rRNA) sequencing is a powerful tool to detect, identify, and classify prokaryotic organisms, and there is currently an explosion of SSU rRNA sequencing in the microbiology community. We report unexpectedly high levels of intraspecific variation (within and between strains) of prokaryote SSU rRNA sequences deposited in GenBank. A total of 82% of the prokaryote species with two published SSU rRNA sequences had more variable positions than a 0.1% random sequencing error would predict, and 48% of these sequence pairs had more variable positions than predicted by a 1.0% random sequencing error. Other sources of sequence variability must account for some of this intraspecific variation. Given these results, phylogenetic studies and biodiversity estimates obtained by using prokaryotic SSU rRNA sequences cannot proceed under the assumption that rRNA sequences of single operons from single isolates adequately represent their taxa. Sequencing SSU rRNA molecules from multiple operons and multiple isolates is highly recommended to obtain meaningful phylogenetic hypotheses, as is careful attention to accurate strain identification.**

The technological advances that have made sequencing rRNA feasible are revolutionizing prokaryotic systematics (22). Because small-subunit rRNA (SSU rRNA) is present in all cells and because it is of sufficient size and variability to provide information, SSU rRNA sequences (16S rRNA sequences in prokaryotes) are currently the most widely used informational macromolecules used in tracing evolution. rRNA sequences obtained from uncultured environmental samples have made the discovery of novel groups of microorganisms possible (5, 13–16). Because so much SSU rRNA sequence information has already been analyzed, it is now possible for microbiologists to sequence the SSU rRNA of an unknown isolate and, by comparing it with the sequences aligned by the Ribosomal Database Project (17), identify the most similar species. Amann et al. (2) call rRNA sequencing the “gold standard” of bacterial identification and classification. The International Journal of Systematic Bacteriology provides a representative view of how SSU rRNA sequences are being used in bacterial systematics. Volume 42 (1992) contained 26 articles (out of 102 articles; 25%) in which SSU rRNA sequence data were used to substantiate a systematic conclusion; volume 43 (1993) contained 40 SSU rRNA articles (out of 130 articles; 31%); and volume 44 (1994) contained 66 SSU rRNA articles (out of 119 articles; 55%). SSU rRNA sequence analyses now dominate bacterial systematic studies at the family, genus, species, and subspecies levels.

Interpretation of these analyses depends critically on the informative utility of SSU rRNA sequences at shallow divergence and on the assumption that the rRNA sequences obtained from reference strains represent functional rRNA molecules typical of their taxa. The level of intraspecific variability in SSU rRNA is generally assumed to be low (9, 11, 15, 27).

Where differences in published sequences for a given strain

or species are found in the literature, some authors have assumed that the differences represent errors in laboratory procedure (23) and have attempted to identify the correct sequence that should be used in analyses. However, the assumption that one of two alternative sequences must be wrong is not justified in the absence of data which demonstrate that the polymorphism rate is negligible. In this paper we describe unexpectedly high levels of SSU rRNA sequence variation within strains and between conspecific strains that we observed in prokaryote sequences deposited in the GenBank nucleotide sequence database. While there are many potential sources for this intraspecific variation, given our results, systematic studies and biodiversity estimates in which prokaryotic SSU rRNA sequences are used cannot proceed under the assumption that rRNA sequences of single operons from single isolates adequately represent their taxa. Sequencing SSU rRNA molecules from multiple operons and multiple isolates is highly recommended to obtain meaningful phylogenetic hypotheses.

**Sequence selection.** We searched GenBank (Release GenBank87, 15 February 1995) for prokaryote SSU rRNA sequences and identified 336 species for which multiple full-length or nearly full-length sequences were available. Of these, 250 were same-species pairs and 86 were groups of three or more (Tables 1 and 2).

**Alignment and analysis.** For each species, the sequences were aligned using a multiple-sequence alignment program (23b). The alignments were then hand edited to further minimize the number of multistate positions. Where there was a choice between introducing gaps and increasing misalignment of bases, both options were tried, and the variable positions were counted for each option. The alignment chosen for our subsequent analysis was the one with the fewest variable positions. Each gap was counted as a single difference, regardless of its length. Ambiguous bases were not included in polymorphism counts. Every attempt was made to minimize the number of variable positions within each alignment in order to obtain the most conservative possible estimate of within-species variability. Because our goal was to obtain minimal estimates of sequence dissimilarity, no attempt was made to force

\* Corresponding author. Mailing address: The Institute for Genomic Research, 932 Clopper Road, Gaithersburg, MD 20878. Phone: (301) 216-9589. Fax: (301) 869-9423. Electronic mail address: rclayton@tigr.org.

† Present address: The National Center for Genome Resources, Santa Fe, NM 87505.

TABLE 1. Summary of sequence pair data

Source of sequences	No. of strains with:				Total no. of pairs
	0–0.1% variable positions	0.11–1.0% variable positions	1.1–2.0% variable positions	>2% variable positions	
Same strain	26 (26) <sup>a</sup>	55 (56)	9 (10)	8 (8)	98
Different strains	17 (21)	38 (47)	10 (12)	16 (20)	81
No strain identification available	5 (7)	43 (60)	17 (24)	6 (8)	71
All sequence pairs	48 (19)	136 (54)	36 (14)	30 (12)	250

<sup>a</sup> The numbers in parentheses are percentages.

alignments by using predicted SSU rRNA secondary structure. Because of space limitations, the sequences used in this study are not presented in this paper; they are available via Internet by the using file transfer protocol from the address ftp.tigr.org (password, 'anonymous'). The sequence alignments on which our conclusions were based are also available in GDE format (23a) from this Internet address.

**Observed variation in GenBank.** Tables 1 and 2 summarize the data which we obtained for variable positions of conspecific sequences. Only 19% of the sequence pairs (and 28% of the species for which three or more sequences were available [Table 3]) exhibited levels of sequence variation that fell within the 0.1% range of error (0 to 2 variable positions in the SSU rRNA sequences) that has been estimated, on average, for sequences in GenBank (7). Clark and Whittam (7) showed that error rates in the 0.1% range have little effect on phylogenies.

**Estimated error rates.** It is possible that SSU rRNA sequences from prokaryotes have a higher error rate than the sequences in GenBank as a whole. Most of the studies in which these SSU rRNA sequences have been used have not directly addressed the question of sequencing accuracy, and for a significant subset of the published sequences the authors did not mention the sequencing methods used (e.g., directly sequenced rRNA, reverse transcriptase, ribosomal DNA [rDNA] sequencing, etc.). Single-pass sequencing of expressed sequence tags (ESTs) has an error rate of approximately 1.25% (1) for most of the sequence length. This type of study tolerates a high error rate. If a random error rate of 1.0% (10 times the error rate estimated for GenBank as a whole [7] and a value which approaches the single-pass sequencing error rate) is assumed for these prokaryotic SSU rRNAs, 26% of the sequence pairs still exhibit a level of variability greater than the random error rate (Table 1). Although in this survey we could not identify

the actual causes of within-species SSU rRNA sequence variability, the results of our comparisons strongly suggest that much of the observed sequence variability is not due to sequencing error. A number of other sources of variation must be considered.

**Other laboratory error.** It is possible that some sequence heterogeneity reflects misidentification of strains. While most researchers check their strain identities and include type strains in taxonomic studies, in many of the publications examined in this survey the authors did not mention what independent verification techniques were used and some did not give the sources of the strains sequenced. In cases where sequence discrepancies are observed, it would be valuable to have such information (either in the publication or in the GenBank accession).

**Biological sources of variation.** Cases in which phenotypically divergent bacteria have identical or nearly identical SSU rRNA sequences have been identified (12), and conversely, significant intrastrain variation in SSU rRNA sequences has been described in a few species (3, 18). However, low levels of intraspecific variability and detectable interspecific variability in SSU rRNAs are widely assumed to be the norm (2).

**Within-strain variation versus between-strain variation.** For 179 of the 250 sequence pairs examined, strain information was available for both sequences. These sequences were grouped as same-strain pairs and different-strain pairs. Contrary to our expectation of within-strain consistency, roughly the same level of variation was observed within strains as between strains (Table 1). This suggests that the observed intraspecific sequence variation represents something other than interstrain variability. If heterogeneity of species definitions alone were responsible for sequence variation, we would expect to see significantly less variation within strains than between strains.

**Interoperon differences.** One potential source of intraspecific variation in SSU rRNA is interoperon differences. Most bacterial species have multiple rRNA operons (20). Extrapolating from studies of eukaryotic organisms (15), workers have widely assumed that the SSU rRNA sequences are nearly identical from operon to operon. The results of the few studies that address this assumption range from a finding that there is no interoperon difference to a finding that 5% of the positions are

TABLE 2. Data for species with multiple sequences available

No. of sequences	No. of species	No. of species with:			
		0–0% variable positions	0.11–1.0% variable positions	1.1–2.0% variable positions	>2% variable positions
3	46	6 (13) <sup>a</sup>	21 (46)	9 (20)	10 (22)
4	21	1 (5)	13 (62)	1 (5)	6 (29)
5	6	1 (17)	2 (33)	1 (17)	2 (33)
6	2	1	0	0	0
7	3	0	0	1	2
9	4	0	1	0	0
10	2	0	0	0	2
13	1	0	0	0	1
17	1	0	0	1	0
18	1	0	0	0	1
19	1	0	1	0	0

<sup>a</sup> The numbers in parentheses are percentages.

TABLE 3. Data for multiple sequences and percentages of sets in the sequencing error rate range postulated for GenBank sequences<sup>a</sup>

No. of sequences	No. of species	Sequence error range (no. of variable positions)	No. of species in sequence error range
3	46	0–4	14 (30) <sup>b</sup>
4	21	0–6	5 (24)

<sup>a</sup> See reference 7.

<sup>b</sup> The numbers in parentheses are percentages.

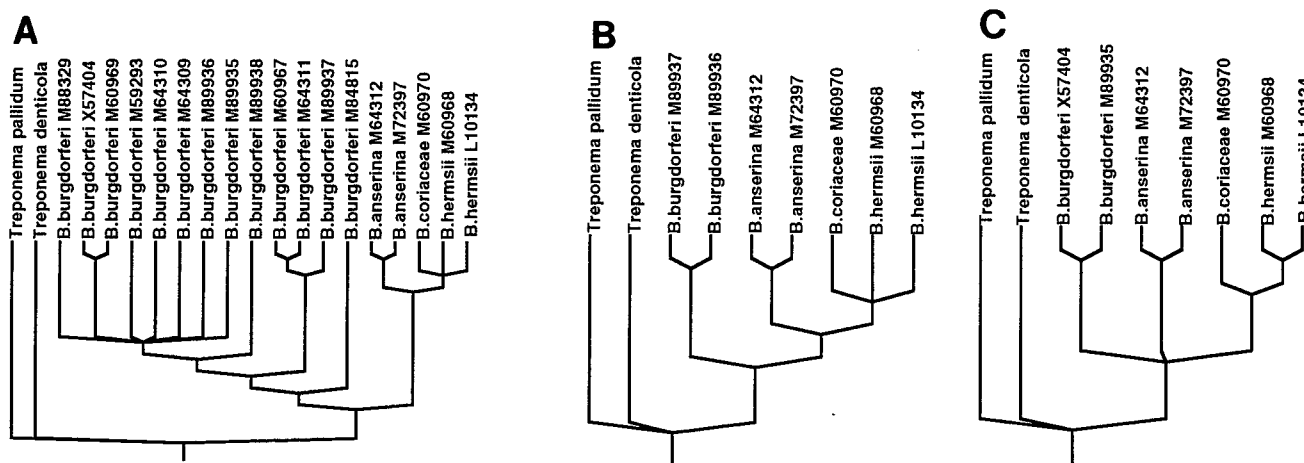


FIG. 1. Effects of sequence choice on phylogeny in the genus *Borrelia*. In the 13 *B. burgdorferi* sequences representing 11 strains, there were 52 variable positions. In the remaining three species represented in the GenBank database, there were 53 variable positions (there was as much variation within *B. burgdorferi* as there was among *B. hermsii*, *B. coriacea*, and *B. anserina*). These parsimony dendrograms demonstrate that the sequence used affects the branching patterns in phylogenetic analyses. Parsimony trees were calculated by using PAUP (24) and the heuristic search option. Strict consensus trees are shown; *Treponema pallidum* and *Treponema denticola* (whose sequences were selected and downloaded from the Ribosomal Database Project [17] on the basis of similarity to *Borrelia* sequences in that database) are included as outgroups. (A) Analysis that included 13 *B. burgdorferi* SSU rRNA sequences. This dendrogram is 160 equally most parsimonious trees having a length of 566. (B and C) Two randomly selected *B. burgdorferi* SSU rRNA sequences were chosen to represent *B. burgdorferi* in the phylogenetic analysis. In panel B the consensus sequence branching pattern matches the branching pattern in panel A; however, in panel C the branching pattern differs from the branching pattern in panel A in that *B. anserina* is as closely related to *B. burgdorferi* as it is to *B. coriacea* and *B. hermsii*, and *B. coriacea* is more closely related to *B. hermsii* than to any other taxon.

variable. The three SSU rRNA sequences of *Rhodobactersphaeroides* (8) are identical to one another. Two of six *Streptomyces coelicolor* rRNA operons have been sequenced (6, 25), and these operons differ at 8 of 1,529 positions (0.5%). The archaean *Haloarcula marismortui* has two nonadjacent rRNA operons (19) which differ from one another at 74 of 1,472 positions (5%) (20). It is quite possible that some of the sequence heterogeneity observed within strains reflects interoperon variability.

**Effects of sequence variation on phylogeny.** To assess the significance of within-species sequence variability, multispecies alignments were constructed for 15 genera (data not shown). Phylogenetic trees and distance dendrograms were calculated from the alignments. Trees were calculated from several sets of alignments of previously published SSU rRNA sequences to assess the implications of within-species variability for phylogenetic analyses. Parsimony trees were calculated using PAUP (24), heuristic search option. Distance trees were calculated using the neighbor-joining/UPGMA method version 3.5 in PHYLIP (10), Kimura two-parameter distance setting, neighbor-joining option.

In some of these genera, the intraspecific variability of the SSU rRNA sequence had only a minimal effect on the phylogenetic analyses. However, when the level of within-species sequence variability approaches the level of between-species variability, phylogenetic analyses may be skewed by the sequences selected. Phylogenetic analyses of the genera *Mycobacterium*, *Borrelia*, and *Cytophaga* illustrate the possible effects of SSU rRNA intraspecific sequence variability on phylogenetic analyses.

The SSU rRNA sequences of *Mycobacterium* species were relatively consistent. In the 19 *Mycobacterium avium* SSU rRNA sequences examined (including the sequences of multiple strains and subspecies), there were only 12 variable positions over a 1,308-base alignment (0.9% of the positions were variable). In contrast, the 43 *Mycobacterium* species examined in this analysis exhibited variability at 441 of 1,308 positions

(33.7% of the positions were variable). The same phylogenetic trees resulted whether all 19 of the *M. avium* sequences were included in the analysis or whether a subset was selected, demonstrating that sequence sampling had only a minimal effect on phylogenetic placement within this consistent group (data not shown; alignments and trees available from ftp.tigr.org by anonymous file transfer protocol).

The genus *Borrelia* exhibited more within-species heterogeneity. In 13 *Borrelia burgdorferi* (sensu Marconi and Garon [18]) sequences representing 11 strains, there were 52 variable positions, while the remaining three species in GenBank had 53 variable positions (there was as much variation within *B. burgdorferi* as there was among *Borrelia hermsii*, *Borrelia coriacea*, and *Borrelia anserina*). If different subsets of *B. burgdorferi* sequences were used in phylogenetic analyses, different dendrograms were produced (Fig. 1). The SSU rRNA variability in *B. burgdorferi*, coupled with phenotypic differences, has led to a revision of the species (4).

The genus *Cytophaga* is a heterogeneous assemblage of species which, along with members of the genera *Flavobacterium* and *Flexibacter*, require extensive taxonomic reevaluation (21). Our survey of SSU rRNA sequences identified a wide range of intraspecific variability in the genus *Cytophaga*. *Cytophaga fermentans* had 3 variable positions in two sequences attributed to the type strain, while *Cytophaga marinoflava* had 139 variable positions in two sequences, both reported to represent the type strain. The sequence choice for *C. fermentans* made no difference in the outcome of a phylogenetic or distance analysis; however, the choice of *C. marinoflava* sequence did affect the branching patterns produced in these analyses (Fig. 2).

**Effects of sequence variation on biodiversity estimates.** SSU rRNA is being used to identify undescribed members of viable but (currently) nonculturable bacterial communities in extreme environments, such as hot springs (5, 26) and the nutrient-poor Sargasso Sea (13). It is feasible to identify the higher phylogenetic affinities of these novel unknown organisms and to group them into "clusters" on the basis of SSU rRNA

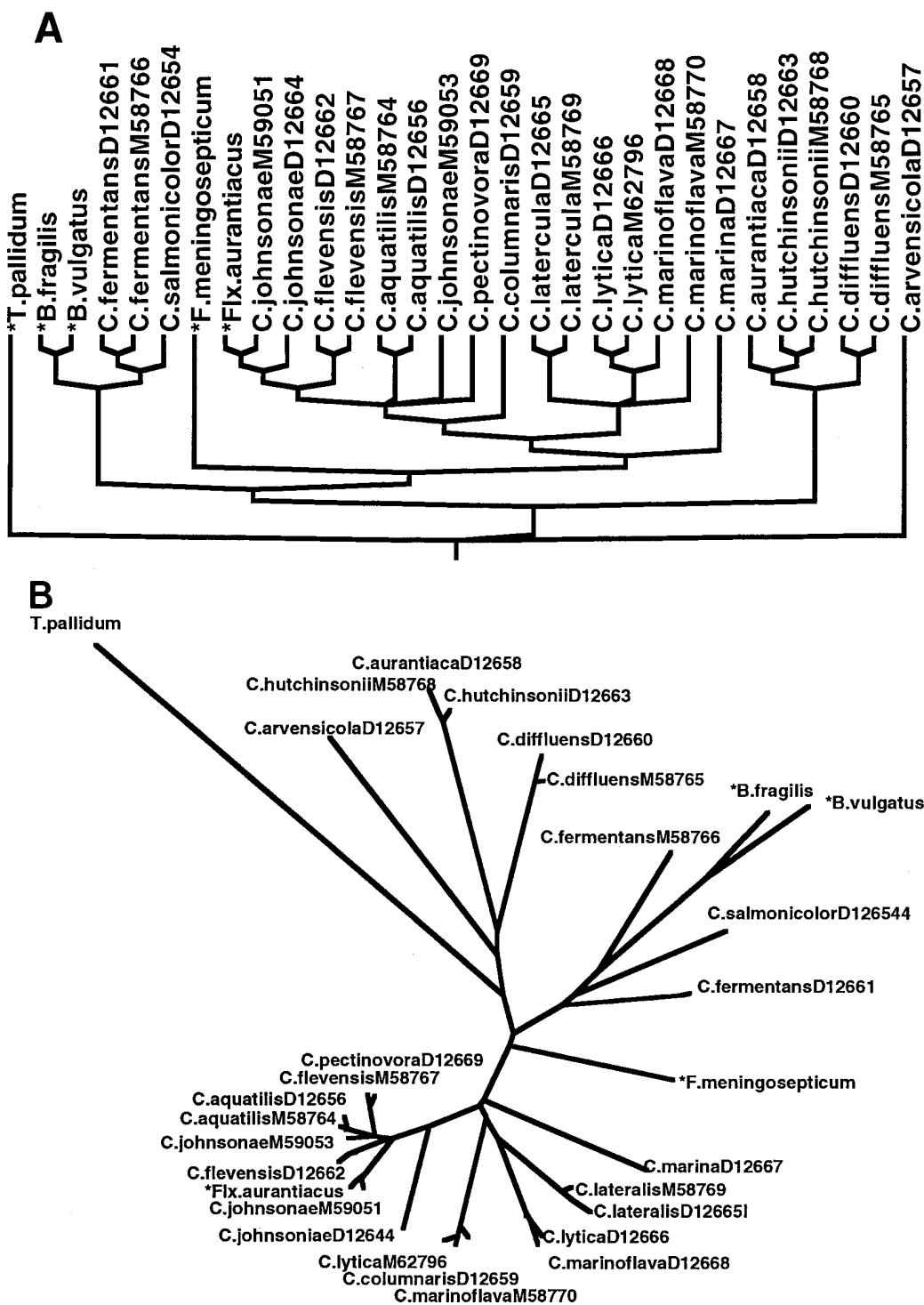


FIG. 2. Effects of sequence choice on phylogeny in the genus *Cytophaga*. While *C. fermentans* had 3 variable positions in two sequences representing the type strain, *C. marinoflava* had 139 variable positions in two sequences representing the type strain. The sequence used for *C. fermentans* made no difference in the outcome of a phylogenetic or distance analysis; however, the *C. marinoflava* sequence used did affect the branching patterns produced in these analyses. The results of cladistic analysis and distance analysis (neighbor-joining method) of *Cytophaga* 16S rRNA sequences are described below. (A) Phylogenetic analysis of *Cytophaga* (strict consensus tree). The sequences of the outgroup taxa used (obtained from the Ribosomal Database Project [17] and selected on the basis of similarity to *Cytophaga* sequences) indicated by asterisks included *T. pallidum*, *Bacteroides fragilis*, *Bacteroides vulgatus*, *Flavobacterium meningosepticum*, and *Flexibacter aurantiacus* sequences. The dendrogram is a consensus dendrogram based on nine equally most parsimonious trees having a length of 2,288. The position of the type strain of *C. marinoflava* depended on which sequence was used to represent it. *C. marinoflava* (GenBank accession number D12668) is the sister taxon of *Cytophaga lytica*, while *C. marinoflava* (GenBank accession number M58770) is the sister taxon of *C. lytica* and *Cytophaga latercula*. The position of *Cytophaga johnsonae* was also dependent on which sequence was used to represent it. GenBank accession numbers M59051 and D12664 were most closely related to the *Flexobacter aurantiacus* sequence, while *C. johnsonae* (GenBank accession number M59053) grouped with these taxa at a lower level of relatedness. (B) Unrooted distance dendrogram obtained by using the neighbor-joining method and the same set of sequences used for panel A. The positions of *C. johnsonae*, *C. marinoflava*, and *C. lytica* varied depending on which sequences were used to represent them.

sequence similarity. However, estimates of species richness for these unusual environments are problematic. If the level of interoperon difference can vary from 0 to 5% and the level of intrastain difference can vary from 0 to 16%, there is no way to determine whether a 1% difference between two sequences represents an interoperon, intrastain, or interspecific level of divergence. Surveys of intraspecific variability would be an invaluable aid in such environmental assays.

**Recommendations.** In this survey of previously published sequences we found that some taxa exhibit substantial levels of SSU rRNA sequence variability; other taxa exhibit very little variability. The variability may represent interoperon variation within a strain, strain-to-strain variation within a species, inadequate taxon delimitation, sequencing error, or other laboratory error. Whatever its underlying cause, undetected SSU rRNA sequence variability can render any phylogenetic, ecological, or clinical conclusions which depend on these sequences ambiguous.

If SSU rRNA sequences are to be used for phylogenetic studies and biodiversity estimates, assessments of interoperon, intrastain, and intraspecific variation are necessary. Multiple PCR products (for rDNA sequencing) or multiple clones (for the reverse transcriptase method) should be sequenced to assess the intrastain variation, and when possible, multiple strains should be examined before primers are designed or phylogenies are postulated. Nucleic acid sequencing technology has advanced considerably since rRNA sequence data were first proposed as a species identifier, and it is now feasible in terms of time and money to isolate and sequence multiple SSU rRNAs. An assessment of variability should also improve the reliability of tentative assignments of taxonomic rank to nonculturable unknown organisms isolated from the environment.

Assessing within-taxon variability of the characters examined has long been standard procedure in animal and plant systematics and ecology, whether the features studied are morphological, behavioral, or biochemical. It should come as no surprise that nucleic acid sequences require similar scrutiny, especially in prokaryotes, a paraphyletic group in which systematic designations are so often provisional.

This work was supported in part by U.S. Department of Energy Genome Program grant 93ER61566 and by National Science Foundation grant DEB-9400861 to C. Fields.

We thank Lisa M. Fitzgerald for helpful discussions and comments.

#### REFERENCES

- Adams, M. D., A. R. Kerlavage, C. Fields, and J. C. Venter. 1993. 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.* **4**:256-267.
- Amann, R., W. Ludwig, and K.-H. Schleifer. 1994. Identification of uncultured bacteria: a challenging task for molecular taxonomists. *ASM News* **60**:360-365.
- Aquino de Muro, M., and F. G. Priest. 1993. Phylogenetic analysis of *Bacillus sphaericus* and development of an oligonucleotide probe specific for mosquito-pathogenic strains. *FEMS Microbiol. Lett.* **112**:205-210.
- Baranton, G., D. Postic, I. Saint Girons, P. Boerlin, J.-C. Piffaretti, M. Assous, and P. A. D. Grimont. 1992. Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS4161 associated with Lyme borreliosis. *Int. J. Syst. Bacteriol.* **42**:378-383.
- Barns, S. M., R. E. Fundyga, M. W. Jeffries, and N. R. Pace. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl. Acad. Sci. USA* **91**:1609-1613.
- Baylis, H. A., and M. J. Bibb. 1987. The nucleotide sequence of a 16S rRNA gene from *Streptomyces coelicolor* A3(2). *Nucleic Acids Res.* **15**:7176.
- Clark, A. G., and T. S. Whittam. 1992. Sequencing errors and molecular evolutionary analysis. *Mol. Biol. Evol.* **9**:744-752.
- Dryden, S. C., and S. Kaplan. 1990. Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter sphaeroides*. *Nucleic Acids Res.* **18**:7267-7277.
- Fell, J. W., A. Statzell-Tallman, M. J. Lutz, and C. P. Kurtzman. 1992. Partial rRNA sequences in marine yeasts: a model for identification of marine eukaryotes. *Mol. Mar. Biol. Biotechnol.* **1**:175-186.
- Felsenstein, J. 1993. PHYLIP, version 3.5. University of Washington, Seattle.
- Field, K. G., S. M. Landfear, and S. J. Giovannoni. 1991. 18S rRNA sequences of *Leishmania enriettii* promastigote and amastigote. *Int. J. Parasitol.* **21**:483-485.
- Fox, G. E., J. D. Wisotzkey, and P. Jurtschuk. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* **42**:166-170.
- Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature (London)* **345**:60-63.
- Haygood, M. G., and D. L. Distel. 1993. Bioluminescent symbionts of flashlight fishes and deep-sea anglerfishes form unique lineages related to the genus *Vibrio*. *Nature (London)* **363**:154-156.
- Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* **66**:411-453.
- Kurtzman, C. P. 1992. rRNA sequence comparison for assessing phylogenetic relationships among yeasts. *Int. J. Syst. Bacteriol.* **42**:1-6.
- Larsen, N., G. J. Olsen, B. L. Maidak, M. J. McCaughey, R. Overbeek, T. J. Macke, T. L. Marsh, and C. R. Woese. 1993. The Ribosomal Database Project. *Nucleic Acids Res.* **21**(Suppl.):3021-3023.
- Marconi, R. T., and C. F. Garon. 1992. Phylogenetic analysis of the genus *Borrelia*: a comparison of North American and European isolates of *Borrelia burgdorferi*. *J. Bacteriol.* **174**:241-244.
- Mevarech, M., S. Hirsch-Twizer, S. Goldman, E. Yakobson, H. Eisenberg, and P. P. Dennis. 1989. Isolation and characterization of the rRNA gene clusters of *Halobacterium marismortui*. *J. Bacteriol.* **171**:3479-3485.
- Mylvaganam, S., and P. P. Dennis. 1992. Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaeobacterium *Haloarcula marismortui*. *Genetics* **130**:399-410.
- Nakagawa, Y., and K. Yamasato. 1993. Phylogenetic diversity of the genus *Cytophaga* revealed by 16S rRNA sequencing and menaquinone analysis. *J. Gen. Microbiol.* **139**:1155-1161.
- Olsen, G. L., C. R. Woese, and R. Overbeek. 1994. The winds of (evolutionary) change: breathing new life into microbiology. *J. Bacteriol.* **176**:1-6.
- Ruimy, R., V. Breittmayer, P. Elbaze, B. Lafay, O. Boussemer, M. Gauthier, and R. Christen. 1994. Phylogenetic analysis and assessment of the genera *Vibrio*, *Photobacterium*, *Aeromonas*, and *Plesiomonas* deduced from small-subunit rRNA sequences. *Int. J. Syst. Bacteriol.* **44**:416-426.
- 23a. Smith, S. Unpublished data.
- 23b. Sutton, G. Unpublished data.
- Swofford, D. L. 1991. PAUP: phylogenetic analysis using parsimony, version 3.1. Illinois Natural History Survey, Champaign.
- van Wezel, G. P., E. Vjgenboom, and L. Bosch. 1991. A comparative study of the ribosomal RNA operons of *Streptomyces coelicolor* A3(2) and sequence analysis of rRNA. *Nucleic Acids Res.* **19**:4399-4403.
- Ward, D. M., R. Weller, and M. M. Bateson. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature (London)* **345**:63-65.
- Woese, C. R. 1994. There must be a prokaryote somewhere: microbiology's search for itself. *Microbiol. Rev.* **58**:1-9.