

Relationship of 16S rRNA sequence similarity to DNA hybridization in prokaryotes

Jyoti Keswanit and William B. Whitman

Author for correspondence: William B. Whitman. Tel: +1 706 542 4219. Fax: +1 706 542 2674.
e-mail: whitman@arches.uga.edu

Department of
Microbiology, University
of Georgia, Athens,
GA 30602-2605, USA

The relationship between 16S rRNA sequence similarity (S) and the extent of DNA hybridization (D) was well described by the equation $\ln(-\ln D) = 0.53 [\ln(-\ln S)] + 2.201$ when D was determined by either the S1 nuclease or membrane filter methods. When the presence of nonultrametric rRNA sequences and differences between genera or families were controlled, this relationship accounted for 78% of the variability of D given S , and it was possible to estimate the distribution of D from S with a known precision. Thus, $D < 0.70$ was expected to occur 50, 95 and 99% of the time when S was 0.998, 0.992 and 0.986, respectively. The relationship between D and S varied between prokaryotic taxa even within the same subphylum, and more precise estimates of D could be made when the relationship for a particular taxon was known. The relationship between D and S was not significantly different between the prokaryotic domains, and S appeared to be a quasi-molecular clock of approximately constant rate when averaging effects and stochastic factors were taken into account. The relationship between $\log D$ and $\log S$ was nonlinear, and D provided a very poor measure of relatedness for distantly related organisms. For instance, within the range $1.0 > S > 0.95$, D decreased from 1.0 to 0.15; and within the range $0.95 > S > 0.90$, D decreased from 0.15 to 0.06. Lastly, at least some of the rRNA sequences from about one-third of the taxa examined had nonultrametric properties where S was much lower than expected from the value of D . For these taxa, S was a poor indicator of relatedness for closely related strains. Thus, the ultrametric properties of rRNA sequences should be tested before making taxonomic or phylogenetic conclusions based upon S .

Keywords: DNA hybridization, 16S rRNA, cophenetic correlation

INTRODUCTION

Because of the importance of DNA hybridization and 16S rRNA sequence similarity in prokaryotic systematics, the relationship between them is of great interest. Devereux *et al.* (1990) proposed that, if the extent of DNA hybridization was considered equivalent to the sequence similarity, then $\log S = K \log D$, where S was the sequence similarity of the 16S rRNA,

D was the extent of DNA hybridization and K was a constant. Consistent with this assumption, a very significant correlation was found between $\log S$ and $\log D$ (Devereux *et al.*, 1990). However, this study was limited by the small number of values compared and because most of the values for S were estimated from S_{ab} values. While subsequent investigations have also found correlations between S and D (Stackebrandt & Goebel, 1994; Hauben *et al.*, 1997, 1999), the relationship between $\log S$ and $\log D$ has not been further tested. Therefore, the goal of the current study was to examine the relationship between S and D more fully.

Although highly correlated to phenotypic similarity (cf. Colwell, 1970; Colwell *et al.*, 1974; Stackebrandt & Goebel, 1994), the precise chemical meaning of the

.....
† **Present address:** Department of Biochemistry, West Virginia University, Morgantown, WV 26506, USA.

Abbreviations: D , extent of DNA hybridization; K , ratio of $\log S / \log D$; r , correlation coefficient; S , 16S rRNA sequence similarity; UPGMA, unweighted pair group method with arithmetic averages.

Table 1 Sources of 16S rRNA sequences and DNA hybridization data

Taxon	Genus	rRNA sequence	DNA hybridization
<i>Aeromonas</i>	<i>Aeromonas</i>	Martinez-Murcia <i>et al.</i> (1992)	Carnahan <i>et al.</i> (1991) Hickman-Brenner <i>et al.</i> (1988)
<i>Bacillus</i>	<i>Bacillus</i>	Ash <i>et al.</i> (1991) Rössler <i>et al.</i> (1991)	Fritze <i>et al.</i> (1990) Kaneko <i>et al.</i> (1978) Seki <i>et al.</i> (1978)
<i>Bdellovibrio</i>	<i>Bdellovibrio</i>	Donze <i>et al.</i> (1991)	Seidler <i>et al.</i> (1972)
<i>Enterococcus</i>	<i>Enterococcus</i>	Martinez-Murcia & Collins (1991)	Collins <i>et al.</i> (1986) Farrow <i>et al.</i> (1983)
<i>Fusobacterium</i>	<i>Fusobacterium</i>	Lawson <i>et al.</i> (1991)	Dzink <i>et al.</i> (1990) Gharbia & Shah (1989)
<i>Legionella</i>	<i>Legionella</i>	Fry <i>et al.</i> (1991)	Brenner <i>et al.</i> (1985)
<i>Leuconostoc</i>	<i>Lactobacillus</i> <i>Leuconostoc</i>	Martinez-Murcia & Collins (1990) Yang & Woese (1989)	Farrow <i>et al.</i> (1989) Schillinger <i>et al.</i> (1989) Shaw & Harding (1989)
<i>Listeria</i>	<i>Listeria</i>	Collins <i>et al.</i> (1991)	Rocourt <i>et al.</i> (1982)
<i>Methanobacteriaceae</i>	<i>Methanobacterium</i> <i>Methanothermobacter</i>	Lechner <i>et al.</i> (1985) Nölling <i>et al.</i> (1993) Østergaard <i>et al.</i> (1987)	Kotelnikova <i>et al.</i> (1993) Patel <i>et al.</i> (1990) Touzel <i>et al.</i> (1992) Zellner <i>et al.</i> (1989)
<i>Methanococcales</i>	' <i>Methanocaldococcus</i> ' <i>Methanococcus</i> ' <i>Methanotorrus</i> '	Burggraf <i>et al.</i> (1990) Jarsch & Bock (1985) Keswani <i>et al.</i> (1996) Woese <i>et al.</i> (1991)	Burggraf <i>et al.</i> (1990) Keswani <i>et al.</i> (1996)
<i>Methanomicrobiaceae</i>	<i>Methanocorpusculum</i> <i>Methanoculleus</i> <i>Methanogenium</i> <i>Methanofollis</i>	Rouviere <i>et al.</i> (1992) Zellner <i>et al.</i> (1999)	Xun <i>et al.</i> (1989)
<i>Methanosarcinaceae</i>	<i>Methanococcoides</i> <i>Methanolobus</i> <i>Methanosarcina</i>	Rouviere <i>et al.</i> (1992)	Boone <i>et al.</i> (1993) Maestrojuan <i>et al.</i> (1992) Sowers <i>et al.</i> (1984)
<i>Methanosarcinaceae</i> (halophilic)	<i>Methanohalobium</i> <i>Methanohalophilus</i> ' <i>Methanosalsus</i> '	Rouviere <i>et al.</i> (1992) Mathrani <i>et al.</i> (1988)	Boone <i>et al.</i> (1993) Wilharm <i>et al.</i> (1991)
<i>Pasteurellaceae</i>	<i>Actinobacillus</i> <i>Haemophilus</i>	Dewhirst <i>et al.</i> (1992)	Borr <i>et al.</i> (1991) Coykendall <i>et al.</i> (1983) Potts & Berry (1983) Potts <i>et al.</i> (1986)
<i>Propionibacterium</i>	<i>Propionibacterium</i>	Charfreitag & Stackebrandt (1989)	Johnson & Cummins (1972)
<i>Pseudonocardiaceae</i>	<i>Amycolata</i> <i>Faenia</i> <i>Kibdelosporangium</i> <i>Pseudonocardia</i> <i>Saccharomonospora</i> <i>Saccharopolyspora</i> <i>Saccharothrix</i>	Bowen <i>et al.</i> (1989)	Bowen <i>et al.</i> (1989)
<i>Serratia</i>	<i>Serratia</i>	Dauga <i>et al.</i> (1990)	Grimont <i>et al.</i> (1982, 1988)
<i>Spirochaeta</i>	<i>Borrelia</i> <i>Serpula</i>	Marconi & Garon (1992) Stanton <i>et al.</i> (1991)	Johnson <i>et al.</i> (1987) Miao <i>et al.</i> (1978)

Table 1 (cont.)

Taxon	Genus	rRNA sequence	DNA hybridization
	<i>Treponema</i>		Postic <i>et al.</i> (1990) Stanton <i>et al.</i> (1991)
<i>Streptococcus</i>	<i>Streptococcus</i>	Bentley <i>et al.</i> (1991) Whiley <i>et al.</i> (1990)	Kilpper-Balz <i>et al.</i> (1985) Welborn <i>et al.</i> (1983) Whiley & Hardie (1989)
<i>Thermus</i>	<i>Thermus</i>	Bateson <i>et al.</i> (1990)	Hensel <i>et al.</i> (1986)

extent of DNA hybridization is ambiguous. It is the change in the melting temperature of the hybrids formed in a DNA reassociation experiment and not the extent of hybridization that is linearly related to sequence similarity, and a change of 1 °C in the melting temperature of the hybrids is equal to a 1.7% change in the sequence similarity (Caccone *et al.*, 1988). The nature of the DNA which does not form hybrids is also not measured. Additional ambiguity results from the fact that reciprocal measurements of the extent of DNA hybridization differ by 4–8% (Sneath, 1989). Thus, when the hybridization experiment is performed by labelling DNA from one organism and comparing the extent of hybridization between homologous and heterologous DNA, the observed value depends to some extent just upon the choice of which DNA is labelled. Lastly, a number of the commonly employed methods of DNA hybridization may give somewhat different results. For instance, the S1 nuclease method is reported to yield 15–20% lower hybridization values than the hydroxyapatite and membrane filter methods (Grimont *et al.*, 1980). In spite of these difficulties in determination and interpretation of the DNA hybridization, it is one of the three criteria used in defining prokaryotic species (Wayne *et al.*, 1987).

METHODS

Sources of 16S rRNA sequences. Published 16S rRNA sequences found in the literature before 1994 for which DNA hybridization values could also be identified for the same strains were compiled (Table 1). In addition, sequences determined in part in our laboratory but published later were included. Partial sequences of 500 bp or less were excluded. Sequences were accessed through the GenBank database.

Determination of S. Sequences belonging to the same genus or closely related genera were aligned by the program PILEUP contained in Genetics Computer Group (GCG) Sequence Analysis package v. 7.01 (Devereux *et al.*, 1984) on a VAX computer. Any position containing undetermined nucleotides in the aligned sequences were excluded from the subsequent analysis. This alignment was used in the DISTANCE program of GCG for computing the similarity matrices (uncorrected for back mutations). This procedure was used for all the sequences despite the availability of similarity matrices in some of the references to avoid

inconsistencies due to the use of different alignment programs.

Determination of cophenetic correlation coefficients. The matrices of 16S rRNA sequence similarity values calculated above were used to calculate the cophenetic matrices with an unweighted pair group method with arithmetic averages (UPGMA) program (Sneath & Sokal, 1973). A cophenetic matrix consisted of the estimated similarity values derived from the calculation of the UPGMA tree. The cophenetic correlation coefficient for each taxon was the correlation coefficient (r) calculated from the linear regression between the corresponding values of the similarity matrix and the cophenetic matrix.

Sources of DNA hybridization values. DNA hybridization values were taken from published reports (Table 1) and classified according to four general methods. The 'S1 nuclease' method refers to those studies based upon the method of Crosa *et al.* (1973). For the 'membrane filter' method, studies using the general technique as described by Johnson (1985) but employing a variety of types of membranes were combined. The 'hydroxyapatite' method refers to those studies based upon the method of Brenner *et al.* (1969). The 'renaturation' method refers to those studies based upon the method of De Ley *et al.* (1970). When applicable, values obtained under optimal conditions, usually at 25 °C below the melting temperature, were used. Data from methods based upon dot-blot hybridizations were not included in this study.

Data analysis. Statistical analysis were performed with the statistical package in Excel 4.0. The analysis of covariance and Cook's distances (Neter *et al.*, 1985) were performed in the statistical analysis package SAS.

RESULTS

Evaluation of the relationship of logD versus logS

Using data obtained from 387 pairs of organisms belonging to 20 different taxa, a highly significant relationship was found between the logarithmic transformations of 16S rRNA sequence similarity (S) and the extent of DNA hybridization (D). The correlation coefficient (r) between $\log S$ and $\log D$ was 0.571, which was significant at $P < 0.0001$ (Fig. 1). The DNA hybridization values were obtained from studies using four different general methods. Because different methods have been reported to yield different

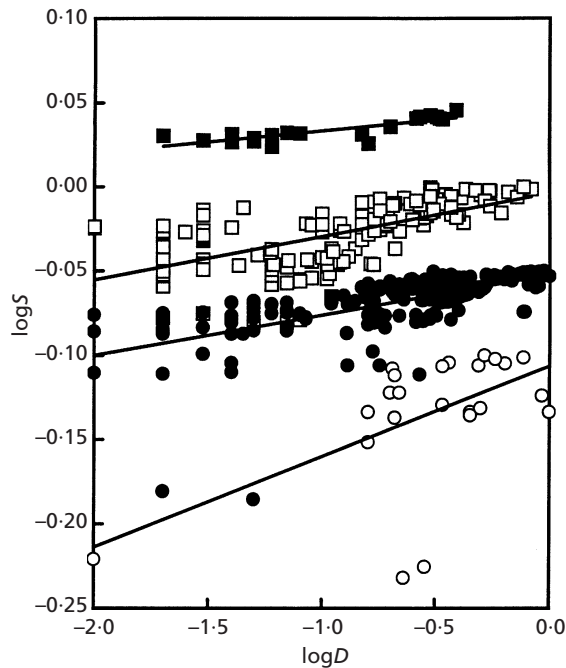


Fig. 1. Relationship between the sequence similarity for the 16S rRNA (S) and the extent of DNA hybridization (D). The correlation between $\log S$ and $\log D$ where D was determined by the S1 nuclease (●), membrane filter (□), hydroxyapatite (■) or thermal renaturation (○) methods. The lines were determined by linear regression of the values for each method. For clarity, the values for the hydroxyapatite, S1 nuclease and thermal renaturation methods were offset by 0.05, -0.05 and -0.10 units, respectively.

hybridization values (Grimont *et al.*, 1980), the relationship of $\log D$ and $\log S$ was also tested for each method separately. For D values obtained by the S1 nuclease, membrane filter, hydroxyapatite and renaturation methods, the correlation coefficients of $\log D$ with $\log S$ were 0.623, 0.636, 0.834 and 0.550 for sample sizes of 188, 153, 23 and 23, respectively. These correlation coefficients were all significant at $P < 0.0001$.

Although the relationship between $\log S$ and $\log D$ was highly significant, the scatter of values in Fig. 1 was also high, and the amount of the variance accounted for (r^2) was low. To determine the sources of variability, the ratio $\log S/\log D$ (K) was examined. K was expected to be a constant. Pairs of organisms for which $S = 1.0$ or $D = 1.0$ were removed from the analysis because the logarithmic transformations of these values were zero, resulting in uninformative or nonsensical K values. The total number of such pairs was only five. In addition, the five remaining values from the genus *Bdellovibrio* were also removed because they were based upon 16S rRNA sequences of less than 1000 positions. Reducing the data set to 377 values had little consequence on the correlation coefficient (data not shown), and $K \pm 1$ SD was 0.030 ± 0.025 .

Ultrametric properties of S

If the ultrametric properties of the 16S rRNA from a specific taxon were poor, S might no longer reflect the phylogeny of the taxon (Sneath, 1993). In this case, K would no longer measure the biological relationships between D and S but would be artefactual. If this reasoning was correct, the inclusion of these taxa would be expected to increase the variability of K . The cophenetic correlation, which is a correlation between the S value calculated during tree-building and the observed S (see above), tests the ultrametric properties of the sequences (Sneath & Sokal, 1973). For the taxa *Aeromonas*, *Fusobacterium*, *Pasteurellaceae*, *Serratia*, *Methanothermobacter* and halophilic *Methanosarcinaceae*, the cophenetic correlation was below 0.85 and lower than found for the other taxa (Table 2 and data not shown). To test whether taxa with low cophenetic correlations were a source of variability of K , these taxa, excepting the halophilic *Methanosarcinaceae*, were removed from the data set. For the halophilic *Methanosarcinales*, visual inspection of the UPGMA matrix of the observed and estimated S indicated that '*Methanosalsus zhilinae*' was the source of the low cophenetic correlation, and removal of the 16S rRNA sequence of this species increased the cophenetic correlation to 0.90. Therefore, only the two K values that involved '*Methanosalsus zhilinae*' were removed from the data set. This analysis indicated that many of the extreme K values were derived from taxa with low cophenetic correlations. For instance, the SD of K decreased when the nonultrametric taxa were removed, and $K \pm$ SD was 0.029 ± 0.018 , $n = 262$ (where n is the number of values tested). Thus, the non-ultrametric properties of the 16S rRNA gene for some taxa were a major source of variability of K .

Inspection of the K values suggested that taxa with low cophenetic correlations contributed greatest to the variability when D was also high. There are a number of biological reasons why S might prove a poor indicator of evolutionary relationships for closely related taxa (Sneath, 1993; Keswani *et al.*, 1996). To examine this effect quantitatively, the set of values where D was determined by the S1 nuclease method was divided into two groups with either $D \leq 0.40$ or $D > 0.40$. $D = 0.40$ was chosen to divide the groups in order to obtain a sufficient number of values with a high D . The values of $K \pm$ SD were 0.028 ± 0.017 ($n = 142$) and 0.040 ± 0.054 ($n = 45$) when D was ≤ 0.40 and > 0.40 , respectively. When taxa with low cophenetic correlations were removed, the $K \pm$ SD were 0.028 ± 0.019 ($n = 81$) and 0.012 ± 0.009 ($n = 18$), respectively. Thus, when D was ≤ 0.40 , the presence of taxa with a low cophenetic correlation had little effect on the mean and SD of K . Removal of taxa with a low cophenetic correlation strongly affected the mean and SD of K only when D was also > 0.40 . However, if the cophenetic correlation was inversely related with the mean S or SD of K , removal of taxa with low cophenetic correlations might also decrease the SD of K or bias the mean value of K . However, this did not appear to be

Table 2 Variation of the relationship between *S* and *D* for different taxa

Data were obtained from analysis of data set 3, which was determined by the S1 nuclease or membrane methods and include all values for taxa with a cophenetic correlation ≥ 0.90 . For other taxa, they include only those where $D \leq 0.70$. NA, Not applicable.

Representative genera	Slope \pm SE*	Intercept \pm SE*	<i>n</i>	Cophenetic correlation
All	0.53 \pm 0.02	2.20 \pm 0.08	327	NA
<i>Aeromonas</i>	0.01 \pm 0.23	0.19 \pm 0.87	9	0.65
<i>Bacillus</i>	0.68 \pm 0.20	2.93 \pm 0.65	40	0.89
<i>Enterococcus</i>	0.18 \pm 0.21	0.82 \pm 0.73	21	0.95
<i>Fusobacterium</i>	NA	NA	1	0.72
<i>Leuconostoc</i>	0.28 \pm 0.19	1.53 \pm 0.64	43	0.99
<i>Listeria</i>	0.68 \pm 0.22	3.17 \pm 0.81	11	0.99
<i>Methanobacteriaceae</i>	1.08 \pm 0.21	4.89 \pm 0.80	15	1.00
<i>Methanococcales</i>	0.55 \pm 0.19	2.16 \pm 0.65	24	0.94
<i>Methanomicrobiaceae</i>	0.30 \pm 0.19	1.55 \pm 0.64	17	0.99
<i>Methanosarcinaceae</i> (halophilic)	0.35 \pm 0.33	1.59 \pm 0.99	6	0.80
<i>Methanosarcinaceae</i>	0.73 \pm 0.21	3.19 \pm 0.69	15	0.99
<i>Pasteurellaceae</i>	0.74 \pm 0.22	2.72 \pm 0.72	38	0.80
<i>Propionibacterium</i>	1.04 \pm 0.35	3.85 \pm 1.11	6	0.87
<i>Pseudonocardiaceae</i>	-0.68 \pm 1.37	-0.70 \pm 3.19	15	0.93
<i>Serratia</i>	0.49 \pm 0.26	1.75 \pm 0.89	29	0.65
<i>Spirochaeta</i>	0.55 \pm 0.20	2.13 \pm 0.68	9	1.00
<i>Streptococcus</i>	0.93 \pm 0.19	3.52 \pm 0.61	28	0.91

*From the analysis of covariance with genera. Slopes and intercepts are for the linear regression equation $\ln(-\ln D) = m[\ln(-\ln S)] + b$, where *m* is the slope and *b* is the intercept.

the case. The correlations between the mean *S* or the SD of *K* of taxa with their cophenetic correlations were poor, 0.226 ($n = 20$) and 0.260, respectively. Therefore, the variability of *K* associated with taxa with low cophenetic correlations was only associated with those members with a high *D*. Visual inspection of *K* from taxa with a low cophenetic correlation further indicated that most of the extreme values were associated with taxa where *D* was > 0.70 . Other methods to identify specific sequences contributing to the low cophenetic correlations were unsuccessful. For instance, exclusion of taxa represented by sequences with high Cook's distance or where the difference between the calculated and actual similarity values were > 2 SDs above the mean did not increase the cophenetic correlations.

Nonlinearity of the relationship between $\log D$ and $\log S$

Upon the removal of taxa with a low cophenetic correlation, the presence of a systematic variation in *K* with $\log S$ (or *S*) was detected. Using the nearly complete data set described above ($n = 377$), the correlation coefficient of *K* with $\log S$ was 0.421. Although highly significant, this value was less than the correlation coefficient between $\log D$ and $\log S$ and appeared to support a simple linear relationship. However, comparison with three other sets of data suggested that the relationship between $\log D$ and $\log S$ could be better described by more complex functions.

Set 1 consisted of data where *D* was determined by either the S1 nuclease or membrane-binding methods ($n = 338$; see below). Set 2 consisted of the same data with the values from taxa with low cophenetic correlations removed ($n = 246$). Set 3 consisted of the same data as set 1 but with only those values from taxa with low cophenetic correlations where *D* was also > 0.70 removed ($n = 327$). For these three sets, the correlation coefficients for *K* versus $\log S$ were 0.408, 0.769 and 0.711, respectively (Fig. 2 and data not shown). These correlation coefficients were all highly significant with $P < 0.0001$. Very similar results were also obtained when *S* was replaced by evolutionary distance to correct for back mutations (data not shown). In contrast, the correlation coefficients of *K* with $\log D$ (or *D*) were less than 0.15 within each of the three data sets, indicating that no significant relationship existed between these parameters (data not shown). Presumably, the systematic variation of *K* with *S* existed because, in contrast to the initial assumptions, *D* was a poor surrogate for sequence information.

In the absence of a theoretical relationship between *D* and *S*, it was still possible to look for an empirical relationship that might be useful to predict *D* from *S*. To further explore this possibility, alternative functions were examined for data set 3. Of the functions tested, only a complementary log log transformation [$\ln(-\ln D)$ versus $\ln(-\ln S)$] produced a higher correlation coefficient, 0.789 with data set 3, than that found with the *K* versus $\log S$ plot (Fig. 2). Although the correlation coefficients of the two plots

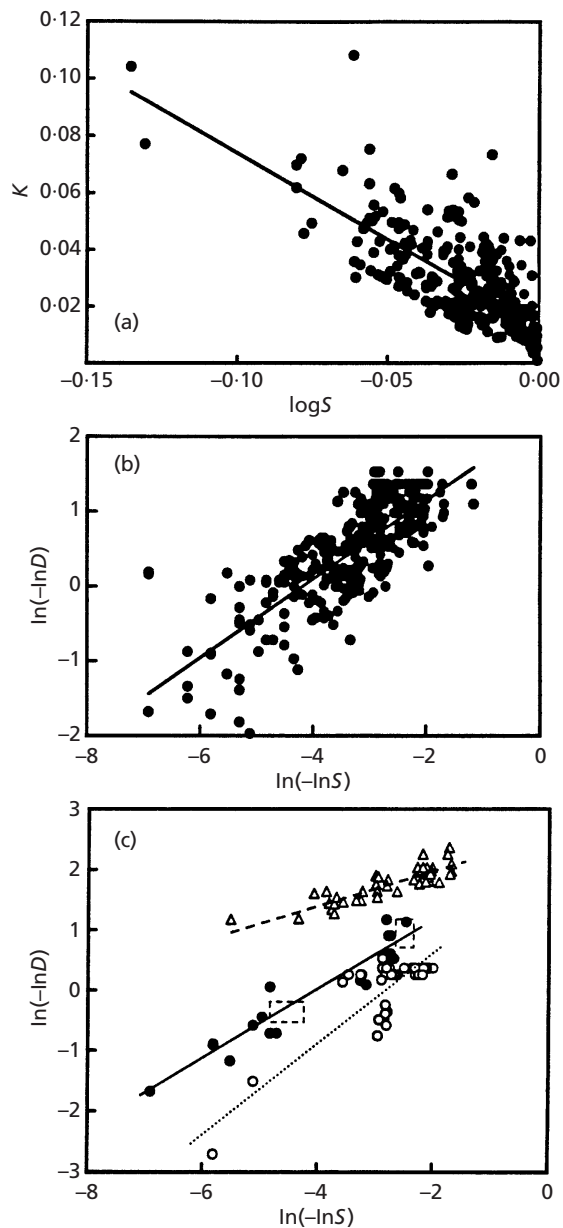


Fig. 2. Nonlinearity of the relationship between $\log S$ and $\log D$. (a) Correlation between $\log S/\log D$ (K) and $\log S$. The line was determined by linear regression of the 327 values in data set 3. (b) Complementary log log plot of the same data. (c) The complementary log log plots for representative taxa. For clarity, the $\ln(-\ln D)$ values for *Bacillus* (\circ) and *Leuconostoc* (\triangle) were offset by -1 and $+1$, respectively. The values for *Methanococcales* (\bullet) were unchanged. The expected SE estimated from the error in reciprocal determinations of D and from the use of S to estimate the evolutionary distance are indicated at $0.99 S$ [$\ln(-\ln S) = -4.6$] and $0.93 S$ [$\ln(-\ln S) = -2.6$] by the rectangles with broken outlines. One SE is shown.

were comparable, the residuals for the K versus $\log S$ plot were significantly skewed ($P < 0.001$). In contrast, the residuals of the complementary log log plot were not skewed and appeared normally distributed (data not shown). For these reasons, further evaluation of

the relationship between D and S was performed with the complementary log log transformation.

Effect of method of measuring D

To determine if the method of measurement of D affected the observed relationship between D and S , values obtained by each of the four methods (with the values removed where the taxon had a low cophenetic correlation and D was also > 0.70) were examined. For values obtained by the S1 nuclease, membrane filter, hydroxyapatite and renaturation methods, the correlation coefficients of $\ln(-\ln D)$ with $\ln(-\ln S)$ were 0.770, 0.734, 0.890 and 0.629 for sample sizes of 178, 149, 23 and 16, respectively. These correlation coefficients were all significant at $P < 0.0001$ except for the renaturation method, which was significant at $P < 0.01$. The slopes for the four methods were 0.514, 0.454, 0.844 and 0.316, respectively, but only the slope from the hydroxyapatite method was significantly different ($P < 0.05$). However, when the residuals were examined by one-way analysis of variance, the values from both the renaturation and hydroxyapatite methods differed significantly ($P < 0.001$) from values of the S1 nuclease and membrane methods. In an analysis of covariance, these latter methods also appeared significantly different ($P < 0.05$). However, the contribution of the method to the variance was very small, and the correlation coefficient only increased from 0.789 to 0.793 when method was included as a variable in data set 3. These analyses indicated that combining the values from the S1 nuclease and membrane methods to form data set 3 was justified. However, it was not possible to conclude that the renaturation and hydroxyapatite methods necessarily affected the observed relationship between D and S . Because of their small sample sizes, the values obtained by these methods represented only a few taxa. Because the relationship between D and S varied between taxa (see below), it was not possible to assign the observed differences solely to differences in methodology.

Taxa-specific effects on the relationship between D and S

In an analysis of covariance, the correlation coefficient for data set 3 increased to 0.884 when taxon was included as a variable, and the effect of taxon was highly significant ($P < 0.0001$). Similar results were also obtained when the values in data set 3 obtained by either the S1 nuclease or membrane methods were analysed separately. These results suggested that the relationship between S and D was not constant across prokaryotic groups. Moreover, even for taxa within the same subphylum or other higher-order group, the relationship varied significantly (Table 2). For instance, within the low G+C group of the Gram-positive bacteria, the slopes of the linear regression (\pm SE) for the taxa represented by *Bacillus*, *Enterococcus*, *Leuconostoc*, *Listeria* and *Streptococcus*

were 0.68 ± 0.20 , 0.18 ± 0.21 , 0.28 ± 0.19 , 0.68 ± 0.22 and 0.93 ± 0.19 , respectively, and these values were significantly different ($P < 0.05$). Similar variation was observed within the representatives of the γ -subclass of the *Proteobacteria* (*Aeromonas*, *Pasteurellaceae* and *Serratia*) and the *Methanosarcinaceae* (Table 2).

In contrast, analysis of covariance indicated no significant difference between members of the bacterial and archaeal domains. For representatives of the bacterial and archaeal groups, the slopes \pm SE were 0.50 ± 0.05 ($n = 250$) and 0.56 ± 0.04 ($n = 77$), respectively, which were not significantly different at $P > 0.1$. Even though the archaea were represented mostly by mesophilic methanogens, these results suggested that the rates of relative change of S and D were about the same for the representatives of both domains.

Prediction of D given S

The r^2 for the analysis of covariance between genera was 0.78, which suggested that this analysis had accounted for most of the variance. The magnitude of the expected error was estimated for comparison. The inherent statistical error for estimating evolutionary distance from S was taken from Kimura (1983): $\sigma = \sqrt{([1 - S]/S.N)}$, where σ is the standard error and N is the number of positions determined (assumed to be 1200). This standard error does not include error expected from inaccuracies in the actual sequences or due to differences between multiple copies of the rRNA gene in the same genome and would be expected to underestimate the total error. Similarly, the SD of reciprocal measurements of D is about 5% (W. B. Whitman, unpublished). This factor also probably underestimates the total error because it neglects the error from repeated measures of D . Nevertheless, the expected ranges of values produced by these effects at $S = 0.99$ and 0.93 in complementary log log plots were consistent with observed values (Fig. 2c) and the conclusion that these plots predicted most of the variability in the relationship between D and S when differences between taxa were also taken into consideration.

The data set was also examined for additional biases. For most of the taxa, D values were reported prior to measuring S . If D was high, it seemed likely that many investigators would not sequence the 16S rRNA, and S would be under reported for high values of D . At high S , this bias might lower the observed distribution of D . Tests were performed for this potential bias. In the first test, the distribution of residuals was compared at $S \geq 0.985$ and $S < 0.985$. At high S , about two-thirds of the residuals were positive, as might be expected if values with a high D were under represented. In the second test, log log complementary plots were examined for high values of S . For the data with $0.999 \geq S > 0.970$ ($n = 105$), the plot was not significantly different ($P > 0.05$) from the plot of the entire data set 3 or of those values where $0.99 > S \geq 0.90$ ($n = 250$). In conclusion, while the data set appeared

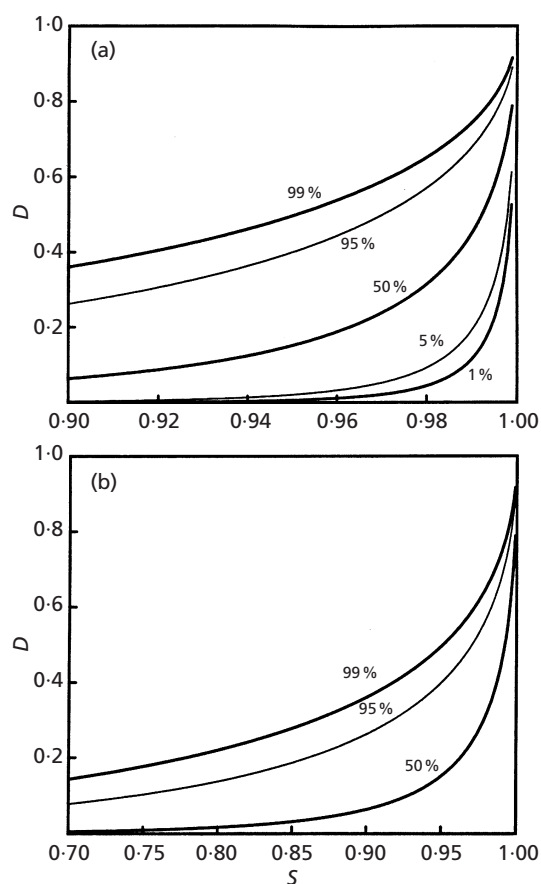


Fig. 3. Probability of D when S is known. The numbers next to the lines indicate the likelihood that D is equal to or below the indicated value. The distribution of D was calculated from the equation $\ln(-\ln D) = 0.5569 [\ln(-\ln S)] + 2.201$ and the SD of the residuals in D , which was 0.4226. (a) Probability of D for S from 0.90 to 1.00. (b) Probability of D for S from 0.70 to 1.00.

biased at high S , this bias did not appear strong enough to affect the usefulness of the observed relationship.

Because the residuals of the complementary log log plots were normally distributed, it was possible to estimate the distribution of D given S (Fig. 3). Important characteristics of this distribution were the precipitous decline in D when S was near 1.0 and a much slower decline in D at low values of S . Thus, it would also be expected that, given an S of 0.998, D would be < 0.70 about 50% of the time. Given an S of 0.992 or 0.986, D would be < 0.70 about 95 or 99% of the time, respectively. Moreover, at $S = 0.963$, D would be expected to be ≤ 0.20 about 50% of the time. At $S = 0.929$, D would be expected to be ≤ 0.10 . To test whether or not the observed distribution was dependent on the type of curve chosen to represent the relationship between D and S , the distribution was also estimated from the plot of K vs $\log S$. Because the residuals of this plot were not normally distributed (see above), the expectations of 50, 95 and 99% were taken

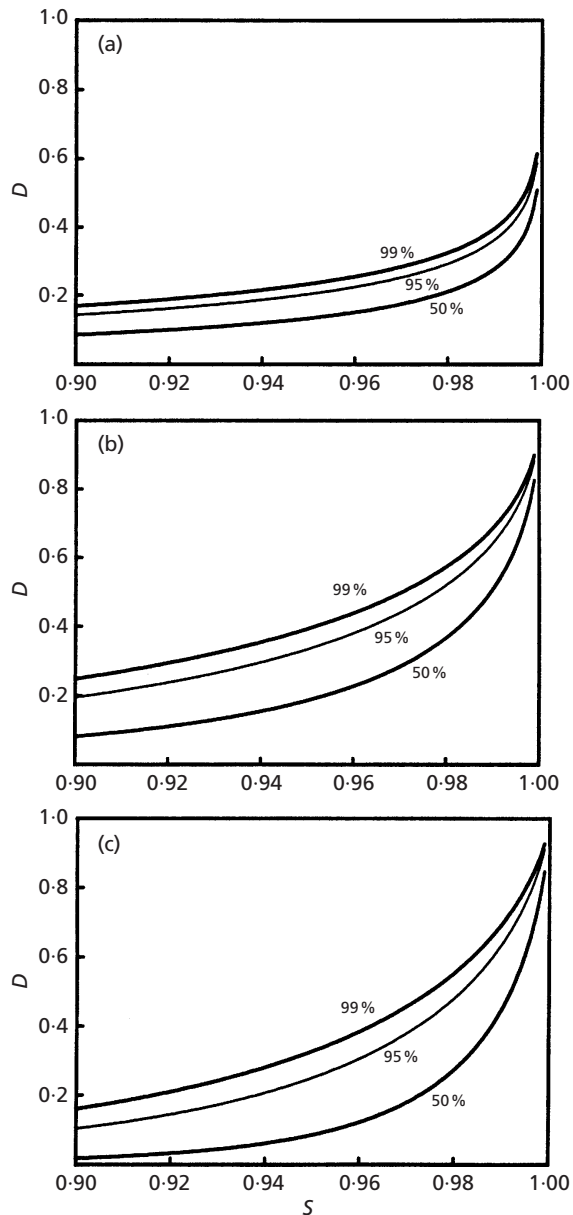


Fig. 4. Probability of D when S is known for the specific taxa *Leuconostoc* (a), *Methanococcales* (b) and *Bacillus* (c). The numbers next to the lines indicate the likelihood that D is equal to or below the indicated value. The distributions of D were calculated from the equation $\ln(-\ln D) = m[\ln(-\ln S)] + b$ and the sd of the residuals in D , where m = the slope and b = the y intercept were as reported in Table 2. The sd of the residuals were, 0.1402, 0.2516 and 0.3383 for *Leuconostoc*, *Methanococcales* and *Bacillus*, respectively.

as the median, the 95% and the 99% highest residuals, respectively (data not shown). From this analysis, given an S of 0.996, 0.986 or 0.982, D would be expected to be < 0.70 about 50, 95 or 99% of the time, respectively. These estimates are in good agreement with those obtained from the complementary log log plots.

Because the relationship of D to S varied significantly between genera, a more precise estimate of D might be made for individual taxa if the complementary log log plots were known for each taxon. To test this possibility, the distribution of D was estimated for three taxa, *Bacillus*, *Leuconostoc* and *Methanococcales*. These taxa were chosen because they each were represented by data from a wide range of S . For *Bacillus* and *Leuconostoc*, the slopes were significantly greater and smaller, respectively, than the slopes for the entire data set used in Fig. 3. The slope for *Methanococcales* was similar to that of the entire data set. As judged from the breadth of the distribution, use of the taxon-specific regression coefficients produced more precise estimates of D (Fig. 4). For instance, for *Bacillus* and *Methanococcales*, given an S of 0.997, 0.993 or 0.991, D would be expected to be < 0.70 about 50, 95 or 99% of the time, respectively. For *Leuconostoc*, given an S of 0.999, D would be expected to be < 0.50 , < 0.58 or < 0.61 about 50, 95 or 99% of the time.

To evaluate the effect of sample size on the taxon-specific complementary log log plots, slopes were calculated for random selections of data. For *Leuconostoc* and *Methanococcales*, which were characterized by an even distribution of measurements over a range of D , the relative sd of the observed slopes for random selections of 10 and 20 measurements were 16 and 8%, respectively (data not shown). In contrast, for *Bacillus*, where most of the data were obtained at low values of D , the relative sd for 10 and 20 randomly selected measurements were 81 and 4%, respectively. Very similar results were also obtained when the intercepts were examined. Therefore, both sample size and the distribution of measurements played important roles in determining the complementary log log plots with a high precision.

DISCUSSION

Although this rather statistical analysis of the relationship between D and S might appear 'dry', it allows for some interesting conclusions to be formed. First, there exists a very significant relationship between D and S . However, this relationship is obscured by the presence of taxa with rRNA sequences with nonultrametric properties, by variation in the relationship between taxa, by differences in the methods for determining D , and by deviation from the ideal model for this relationship. When these factors are controlled for, it is possible to account for 78% of the variability of D given S . Moreover, the experimental error associated with D and the inherent statistical error in using S to estimate evolutionary distance are of sufficient magnitude to account for much of the remaining variability. These results imply that it is possible to predict D from S with a known precision. Given the relative ease in determining S by automated sequencing, the ability to estimate D will be of great utility for systematic studies. For organisms that have never been isolated but have been detected in natural

samples by rRNA sequence alone, the ability to estimate D will provide a clearer understanding of their genetic and phenotypic diversity.

About one-third of the taxa included in this study possess rRNA with nonultrametric properties that confound the estimation of D from S . These taxa all share high values of D and much lower values of S than expected. Sneath (1993) proposed that the nonultrametric properties of the rRNA of *Aeromonas* spp. were due to intergeneric homologous recombination. Homologous recombination appears to be widespread among certain prokaryotic taxa (Dykhuizen & Green, 1991; Smith *et al.*, 1993), so this explanation may be more general. Alternatively, Keswani *et al.* (1996) proposed that the allelic variability expected in the large populations typical of prokaryotes might also contribute to the nonultrametric properties of rRNA. Lastly, sequences which are wrong might also prove nonultrametric.

For taxa where rRNA possess nonultrametric properties, alternative phylogenetic methods may be more appropriate. For the sequences examined in this study, the nonultrametric properties arose from comparisons of very closely related strains where $D > 0.70$. At these high levels of relatedness, classical taxonomic methods may be the methods of choice for determining relatedness. Even when nonultrametric rRNA sequences are excluded, S is expected to be very high, about 0.998, when D is ≥ 0.70 . This result implies that many taxa defined on the basis of $S > 0.97$ may be underspeciated (Stackebrandt & Goebel, 1994). Moreover, at these high levels of S , sequencing errors and gene heterogeneity may have a major effect (Clayton *et al.*, 1995). Together, these factors imply that rRNA sequencing is not a robust method for determining evolutionary relationships at the species level.

These considerations lead to some practical suggestions for systematic studies. First, the ultrametric properties of rRNA should be routinely tested. Second, in taxa where the rRNA sequences do not possess nonultrametric properties and the taxon-specific complementary log log plot is not known, S of ≤ 0.986 provides evidence for different genospecies with a high level of confidence. If the taxon-specific complementary log log plot is known, a higher boundary for S will probably pertain. Third, $S > 0.986$ does not provide evidence for a single genospecies, and S cannot be used to confidently place organisms in the same genospecies under any circumstances. Fourth, in taxa where the rRNA sequences possess nonultrametric properties, S should not be used to make systematic decisions regarding the presence or absence of genospecies.

The relationship between D and S is complex, and D is a poor surrogate for sequence similarity. For closely related organisms, D changes much more rapidly than for distantly related organisms. For instance, within the range $1.0 > S > 0.95$, D decreases from 1.0 to 0.15. Within the range $0.95 > S > 0.90$, D decreases only

from 0.15 to 0.06. Because D is such a poor indicator of relationships among distantly related organisms, ranges of D assigned for intergeneric relationships should be interpreted cautiously (Johnson, 1984). In addition, S is expected to be strongly related to the DNA sequence similarity of genomes, at least in so far as genomic similarity can be represented by a single gene. This implies that the relationship between D and the genomic sequence similarity is also likely to be complex and that D is unlikely to provide reliable, quantitative information about the genomic sequence similarity. Given that the genomic DNA sequence similarity is the reference standard to determine phylogeny and taxonomy in prokaryotes (Wayne *et al.*, 1987), the usefulness of D in taxonomy would appear to be limited.

The relationship between S and D also varied significantly between taxa. While variability in the rates of evolution among rRNA genes from diverse groups of organisms has been noted previously (Woese, 1987), there is no fundamental reason to assume that the rate of change of D is a constant. Thus, differences between taxa cannot, a priori, be attributed to changes in the rates of evolution of either D or S . In complementary log log plots, taxa with a high rate of change of S relative to D would be detected by a low slope (Table 2). Examples of such taxa include *Aeromonas*, *Enterococcus*, *Leuconostoc* and *Methanomicrobiaceae*. Taxa where S changes slowly relative to D include *Methanobacteriaceae*, *Propionibacterium*, *Streptococcus*, *Pasteurellaceae* and *Methanosarcinaceae*. However, attribution of these differences to changes in D , S or both D and S must await further study. Whatever the cause, these observations argue that S and D should be interpreted flexibly when making taxonomic decisions.

In contrast, the relative rates of evolution of S and D were the same in the domains *Bacteria* and *Archaea*. These results are consistent with changes observed in the rates at the genus or family level provided that the variability within each domain is high. Thus, S is a 'quasi-molecular evolutionary clock' that is approximately constant when averaging effects and stochastic factors are taken into account (Kimura *et al.*, 1989). Similarly, Ochman & Wilson (1987) argued that the mean substitution rate for 16S rRNA was nearly the same for prokaryotes, plants and mammals based on calibrations of branching events in the 16S rRNA phylogenetic tree with specific ecological events. Importantly, the similarity of the relative rates of change of D and S suggests that similar evolutionary processes are occurring in the modern populations of both domains and provides strong support for the construction of a universal phylogenetic tree within the prokaryotes based upon this gene.

ACKNOWLEDGEMENTS

The authors are grateful to Wendy Smith and Stephen Rathbun of the Statistics Department of the University of Georgia for assistance with the statistical analyses and

suggesting the complementary log log plots, and to Brian Tindall for helpful discussions. This work was supported in part by NSF grant MCB-0084164.

REFERENCES

- Ash, C., Farrow, J. A. E., Wallbanks, S. & Collins, M. D. (1991). Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. *Lett Appl Microbiol* **13**, 202–206.
- Bateson, M. M., Thibault, K. J. & Ward, D. M. (1990). Comparative analysis of 16S ribosomal RNA sequences of *Thermus* species. *Syst Appl Microbiol* **13**, 8–13.
- Bentley, R. W., Leigh, J. A. & Collins, M. D. (1991). Intrageneric structure of *Streptococcus* based on comparative analysis of small-subunit rRNA sequences. *Int J Syst Bacteriol* **41**, 487–494.
- Boone, D. R., Mathrani, I. M., Liu, Y., Menaia, J. A. G. F., Mah, R. A. & Boone, J. E. (1993). Isolation and characterization of *Methanohalophilus portucalensis* sp. nov., and DNA re-association study of the genus *Methanohalophilus*. *Int J Syst Bacteriol* **43**, 430–437.
- Borr, J. D., Ryan, D. A. J. & MacInnes, J. I. (1991). Analysis of *Actinobacillus pleuropneumoniae* and related organisms by DNA-DNA hybridization and restriction endonuclease fingerprinting. *Int J Syst Bacteriol* **41**, 121–129.
- Bowen, T., Stackebrandt, E., Dorsch, M. & Embley, T. M. (1989). The phylogeny of *Amycolata autotrophica*, *Kibdelosporangium aridum* and *Saccharothrix australiensis*. *J Gen Microbiol* **135**, 2529–2536.
- Brenner, D. J., Fanning, G. R., Rake, A. V. & Johnson, K. E. (1969). Batch procedure for thermal elution of DNA from hydroxyapatite. *Anal Biochem* **28**, 447–459.
- Brenner, D. J., Steigerwalt, A. G., & Gorman, G. W. & 13 other authors. (1985). Ten new species of *Legionella*. *Int J Syst Bacteriol* **35**, 50–59.
- Burggraf, S., Fricke, H., Neuner, A., Kristjansson, J., Rouvier, P., Mandelco, L., Woese, C. R. & Stetter, K. O. (1990). *Methanococcus igneus* sp. nov., a novel hyperthermophilic methanogen from a shallow submarine hydrothermal system. *Syst Appl Microbiol* **13**, 263–269.
- Caccone, A., Desalle, R. & Powell, J. R. (1988). Calibration of the change in thermal stability of DNA duplexes and degree of base-pair mismatch. *J Mol Evol* **27**, 212–216.
- Carnahan, A. M., Chakraborty, T., Fanning, G. R., Verma, D., Ali, A., Janda, J. M. & Joseph, S. W. (1991). *Aeromonas trota* sp. nov., an ampicillin-susceptible species isolated from clinical specimens. *J Clin Microbiol* **29**, 1206–1210.
- Charfreitag, O. & Stackebrandt, E. (1989). Inter- and intrageneric relationships of the genus *Propionibacterium* as determined by 16S rRNA sequences. *J Gen Microbiol* **135**, 2065–2070.
- Clayton, R. A., Sutton, G., Hinkle, P. S., Jr, Bult, C. & Fields, C. (1995). Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. *Int J Syst Bacteriol* **45**, 595–599.
- Collins, M. D., Farrow, J. A. E. & Jones, D. (1986). *Enterococcus mundtii* sp. nov. *Int J Syst Bacteriol* **36**, 8–12.
- Collins, M. D., Wallbanks, S., Lane, D. J., Shah, J., Nietupski, R., Smida, J., Dorsch, M. & Stackebrandt, E. (1991). Phylogenetic analysis of the genus *Listeria* based on reverse transcriptase sequencing of 16S rRNA. *Int J Syst Bacteriol* **41**, 240–246.
- Colwell, R. R. (1970). Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species. *J Bacteriol* **104**, 410–433.
- Colwell, R. R., Johnson, R., Wan, L., Lovelace, T. E. & Brenner, D. J. (1974). Numerical taxonomy and deoxyribonucleic acid reassociation in the taxonomy of some Gram-negative fermentative bacteria. *Int J Syst Bacteriol* **24**, 422–433.
- Coykendall, A. L., Setterfield, J. & Slots, J. (1983). Deoxyribonucleic acid relatedness among *Actinobacillus actinomycetemcomitans*, *Haemophilus aphrophilus*, and other *Actinobacillus* species. *Int J Syst Bacteriol* **33**, 422–424.
- Crosa, J. H., Brenner, D. J. & Falkow, S. (1973). Use of a single-strand specific nuclease for analysis of bacterial and plasmid deoxyribonucleic acid homo- and heteroduplexes. *J Bacteriol* **115**, 904–911.
- Dauga, C., Grimont, F. & Grimont, P. A. D. (1990). Nucleotide sequences of 16S rRNA from ten *Serratia* species. *Res Microbiol* **141**, 1139–1149.
- De Ley, J., Cattoir, H. & Reynaerts, A. (1970). The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem* **12**, 133–142.
- Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.
- Devereux, R., He, S.-H., Doyle, C. L., Orkand, S., Stahl, D. A., LeGall, J. & Whitman, W. B. (1990). Diversity and origin of *Desulfovibrio* species: phylogenetic definition of a family. *J Bacteriol* **172**, 3609–3619.
- Dewhirst, F. E., Paster, B. J., Olsen, I. & Fraser, G. J. (1992). Phylogeny of 54 representative strains of species in the family *Pasteurellaceae* as determined by comparison of 16S rRNA sequences. *J Bacteriol* **174**, 2002–2013.
- Donze, D., Mayo, J. A. & Diedrich, D. L. (1991). Relationships among the bdellovibrios revealed by partial sequences of 16S ribosomal RNA. *Curr Microbiol* **23**, 115–119.
- Dykhuisen, D. E. & Green, L. (1991). Recombination in *Escherichia coli* and the definition of biological species. *J Bacteriol* **173**, 7257–7268.
- Dzink, J. L., Sheenan, M. T. & Socransky, S. S. (1990). Proposal of three subspecies of *Fusobacterium nucleatum* Knorr 1922: *Fusobacterium nucleatum* subsp. *nucleatum* subsp. nov., comb. nov.; *Fusobacterium nucleatum* subsp. *polymorphum* subsp. nov., nom. rev., comb. nov.; and *Fusobacterium nucleatum* subsp. *vincentii* subsp. nov., nom. rev., comb. nov. *Int J Syst Bacteriol* **40**, 74–78.
- Farrow, J. A. E., Jones, D., Phillips, B. A. & Collins, M. D. (1983). Taxonomic studies on some group *D* streptococci. *J Gen Microbiol* **129**, 1423–1432.
- Farrow, J. A. E., Facklam, R. R. & Collins, M. D. (1989). Nucleic acid homologies of some vancomycin-resistant leuconostoc and description of *Leuconostoc citreum* sp. nov., and *Leuconostoc pseudomesenteroides* sp. nov. *Int J Syst Bacteriol* **39**, 279–283.
- Fritze, D., Flossdorf, J. & Claus, D. (1990). Taxonomy of alkaliphilic *Bacillus* strains. *Int J Syst Bacteriol* **40**, 92–97.
- Fry, N. K., Warwick, S., Saunders, N. A. & Embley, T. M. (1991). The use of 16S ribosomal RNA analysis to investigate the phylogeny of the family *Legionellaceae*. *J Gen Microbiol* **137**, 1215–1222.
- Gharbia, S. E. & Shah, H. N. (1989). Glutamate dehydrogenase and 2-oxoglutarate reductase electrophoretic patterns and deoxyribonucleic acid-deoxyribonucleic acid hybridization

- among human oral isolates of *Fusobacterium nucleatum*. *Int J Syst Bacteriol* **39**, 467–470.
- Grimont, P. A. D., Popoff, M. Y., Grimont, F., Coynault, C. & Lemelin, M. (1980). Reproducibility and correlation study of three deoxyribonucleic acid hybridization procedures. *Curr Microbiol* **4**, 325–330.
- Grimont, P. A. D., Irino, K. & Grimont, F. (1982). The *Serratia liquefaciens*-*S. proteamaculeans*-*S. grimesii* complex: DNA relatedness. *Curr Microbiol* **7**, 63–68.
- Grimont, P. A. D., Jackson, T. A., Ageron, E. & Noonan, M. J. (1988). *Serratia entomophila* sp. nov., associated with amber disease in the New Zealand grass grub *Costelytra zealandica*. *Int J Syst Bacteriol* **38**, 1–6.
- Hauben, L., Vauterin, L., Swings, J. & Moore, E. R. B. (1997). Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. *Int J Syst Bacteriol* **47**, 328–335.
- Hauben, L., Vauterin, L., Moore, E. R. B., Hoste, B. & Swings, J. (1999). Genomic diversity of the genus *Stenotrophomonas*. *Int J Syst Bacteriol* **49**, 1749–1760.
- Hensel, R., Demharter, W., Kandler, O., Kroppenstedt, R. M. & Stackebrandt, E. (1986). Chemotaxonomic and molecular-genetic studies of the genus *Thermus*: evidence for a phylogenetic relationship of *Thermus aquaticus* and *Thermus ruber* to the genus *Deinococcus*. *Int J Syst Bacteriol* **36**, 444–453.
- Hickman-Brenner, F. W., Fanning, G. R., Arduino, M. J., Brenner, D. J. & Farmer, J. J. III. (1988). *Aeromonas schubertii*, a new mannitol-negative species found in human clinical specimens. *J Clin Microbiol* **26**, 1561–1564.
- Jarsch, M. & Bock, A. (1985). Sequence of the 16S ribosomal RNA gene from *Methanococcus vannielii*: evolutionary implications. *Syst Appl Microbiol* **6**, 54–59.
- Johnson, J. L. (1984). Nucleic acids in bacterial classification. In *Bergey's Manual of Systematic Bacteriology*, vol. 1, pp. 8–11. Edited by N. R. Kreig & J. G. Holt. Baltimore: Williams & Wilkins.
- Johnson, J. L. (1985). DNA reassociation and RNA hybridization of bacterial nucleic acids. *Methods Microbiol* **18**, 33–74.
- Johnson, J. L. & Cummins, C. S. (1972). Cell wall composition and deoxyribonucleic acid similarities among the anaerobic coryneforms, classical propionibacteria, and strains of *Arachnia propionica*. *J Bacteriol* **109**, 1047–1066.
- Johnson, R. C., Burgdorfer, W., Lane, R. S., Barbour, A. G., Hayes, S. F. & Hyde, F. W. (1987). *Borrelia coriaceae* sp. nov.: putative agent of epizootic bovine abortion. *Int J Syst Bacteriol* **37**, 72–74.
- Kaneko, T., Nozaki, R. & Aizawa, K. (1978). Deoxyribonucleic acid relatedness between *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis*. *Microbiol Immunol* **22**, 639–641.
- Keswani, J., Orkand, S., Premachandran, U., Mandelco, L., Franklin, M. J. & Whitman, W. B. (1996). Phylogeny and taxonomy of mesophilic *Methanococcus* spp. and comparison of rRNA, DNA hybridization, and phenotypic methods. *Int J Syst Bacteriol* **46**, 727–735.
- Kilpper-Balz, R., Wenzig, P. & Schleifer, K. H. (1985). Molecular relationships and classification of some viridans streptococci as *Streptococcus oralis* and emended description of *Streptococcus oralis* (Bridge and Sneath 1982). *Int J Syst Bacteriol* **35**, 482–488.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Kimura, M., Arndt, E., Hatakeyama, T., Hatakeyama, T. & Kimura, J. (1989). Ribosomal proteins in halobacteria. *Can J Microbiol* **35**, 195–199.
- Kotelnikova, S. V., Obratsova, A. Y., Blotevogel, K.-H. & Popov, I. N. (1993). Taxonomic analysis of thermophilic strains of the genus *Methanobacterium*: reclassification of *Methanobacterium thermoalcaliphilum* as a synonym of *Methanobacterium thermoautotrophicum*. *Int J Syst Bacteriol* **43**, 591–596.
- Lawson, P. A., Gharbia, S. E., Shah, H. N., Clark, D. R. & Collins, M. D. (1991). Intrageneric relationships of members of the genus *Fusobacterium* as determined by reverse transcriptase sequencing of small-subunit rRNA. *Int J Syst Bacteriol* **41**, 347–354.
- Lechner, K., Wich, G. & Bock, A. (1985). The nucleotide sequence of the 16S rRNA gene and flanking regions from *Methanobacterium formicicum*: the phylogenetic relationship between methanogenic and halophilic archaeobacteria. *Syst Appl Microbiol* **6**, 157–163.
- Maestrojuan, G. M., Boone, J. E., Mah, R. A., Menaia, J. A. G. F., Sachs, M. S. & Boone, D. R. (1992). Taxonomy and halotolerance of mesophilic *Methanosarcina* strains, assignment of strains to species, and synonymy of *Methanosarcina mazei* and *Methanosarcina frisia*. *Int J Syst Bacteriol* **42**, 561–567.
- Marconi, R. T. & Garon, C. F. (1992). Phylogenetic analysis of the genus *Borrelia*: a comparison of North American and European isolates of *Borrelia burgdorferi*. *J Bacteriol* **174**, 241–244.
- Martinez-Murcia, A. J. & Collins, M. D. (1990). A phylogenetic analysis of the genus *Leuconostoc* based on reverse transcriptase sequencing of 16S rRNA. *FEMS Microbiol Lett* **70**, 73–84.
- Martinez-Murcia, A. J. & Collins, M. D. (1991). *Enterococcus sulfureus*, a new yellow-pigmented *Enterococcus* species. *FEMS Microbiol Lett* **80**, 69–74.
- Martinez-Murcia, A. J., Benlloch, S. & Collins, M. D. (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.
- Mathrani, I. M., Boone, D. R., Mah, R. A., Fox, G. E. & Lau, P. P. (1988). *Methanohalophilus zhilinae* sp. nov., an alkaliphilic, halophilic, methylotrophic methanogen. *Int J Syst Bacteriol* **38**, 139–142.
- Miao, R. M., Fieldsteel, A. H. & Harris, D. L. (1978). Genetics of *Treponema*: characterization of *Treponema hyodysenteriae* and its relationship to *Treponema pallidum*. *Infect Immun* **22**, 736–739.
- Neter, J., Wasserman, W. & Kutner, M. H. (1985). *Applied Linear Statistical Models*. Homewood, IL: Irwin.
- Nölling, J., Hahn, D., Ludwig, W. & De Vos, W. M. (1993). Phylogenetic analysis of thermophilic *Methanobacterium* sp.: evidence for a formate-utilizing ancestor. *Syst Appl Microbiol* **16**, 208–215.
- Ochman, H. & Wilson, A. C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* **26**, 74–86.
- Østergaard, L., Larsen, N., Leffers, H., Kjems, J. & Garrett, R. (1987). A ribosomal RNA operon and its flanking region from the archaeobacterium *Methanobacterium thermoautotrophicum*, Marburg strain: transcriptional signals, RNA structure and evolutionary implications. *Syst Appl Microbiol* **9**, 199–209.
- Patel, G. B., Sprott, G. D. & Fein, J. E. (1990). Isolation and characterization of *Methanobacterium espanolae* sp. nov., a mesophilic, moderately acidophilic methanogen. *Int J Syst Bacteriol* **40**, 12–18.

- Postic, D., Edlinger, C., Richaud, C., Grimont, F., Dufresne, Y., Perolat, P., Baranton, G. & Grimont, P. A. D. (1990). Two genomic species in *Borrelia burgdorferi*. *Res Microbiol* **141**, 465–475.
- Potts, T. V. & Berry, E. M. (1983). Deoxyribonucleic acid-deoxyribonucleic acid hybridization analysis of *Actinobacillus actinomycetemcomitans* and *Haemophilus aphrophilus*. *Int J Syst Bacteriol* **33**, 765–771.
- Potts, T. V., Mitra, T., O'Keefe, T., Zambon, J. J. & Genco, R. J. (1986). Relationships among isolates of oral haemophili as determined by DNA-DNA hybridization. *Arch Microbiol* **145**, 136–141.
- Rocourt, J., Grimont, F., Grimont, P. A. D. & Seeliger, H. P. R. (1982). DNA relatedness among serovars of *Listeria monocytogenes sensu lato*. *Curr Microbiol* **7**, 383–388.
- Rössler, D., Ludwig, W., Schleifer, K. H., Lin, C., McGill, T. J., Wisotzkey, J. D., Jurtshuk, P., Jr & Fox, G. E. (1991). Phylogenetic diversity in the genus *Bacillus* as seen by 16S rRNA sequencing studies. *Syst Appl Microbiol* **14**, 266–269.
- Rouviere, P., Mandelco, L., Winker, S. & Woese, C. R. (1992). A detailed phylogeny for the *Methanomicrobiales*. *Syst Appl Microbiol* **15**, 363–371.
- Schillinger, U., Holzappel, W. & Kandler, O. (1989). Nucleic acid hybridization studies on *Leuconostoc* and heterofermentative lactobacilli and description of *Leuconostoc amelibiosum* sp. nov. *Syst Appl Microbiol* **12**, 48–55.
- Seidler, R. J., Mandel, M. & Baptist, J. N. (1972). Molecular heterogeneity of the bdellovibrios: evidence of two new species. *J Bacteriol* **109**, 209–217.
- Seki, T., Chung, C.-K., Mikami, H. & Oshima, Y. (1978). Deoxyribonucleic acid homology and taxonomy of the genus *Bacillus*. *Int J Syst Bacteriol* **28**, 182–189.
- Shaw, B. G. & Harding, C. D. (1989). *Leuconostoc gelidum* sp. nov., and *Leuconostoc carnosum* sp. nov., from chilled-stored meats. *Int J Syst Bacteriol* **39**, 217–223.
- Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci USA* **90**, 4384–4388.
- Sneath, P. H. A. (1989). Analysis and interpretation of sequence data for bacterial systematics: the view of a numerical taxonomist. *Syst Appl Microbiol* **12**, 15–31.
- Sneath, P. H. A. (1993). Evidence from *Aeromonas* for genetic cross-over in ribosomal sequences. *Int J Syst Bacteriol* **43**, 626–629.
- Sneath, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy: the Principles and Practice of Numerical Classification*. San Francisco: W. H. Freeman.
- Sowers, K. R., Johnson, J. L. & Ferry, J. G. (1984). Phylogenetic relationships among the methylotrophic methane-producing bacteria and emendation of the family *Methanosarcinaceae*. *Int J Syst Bacteriol* **34**, 444–450.
- Stackebrandt, E. & Goebel, B. M. (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.
- Stanton, T. B., Jensen, N. S., Casey, T. A., Tordoff, L. A., Dewhirst, F. E. & Paster, B. J. (1991). Reclassification of *Treponema hyodysenteriae* and *Treponema innocens* in a new genus, *Serpula* gen. nov., as *Serpula hyodysenteriae* comb. nov., and *Serpula innocens* comb. nov. *Int J Syst Bacteriol* **41**, 50–58.
- Touzel, J. P., Conway de Macario, E., Nölling, J., De Vos, W. M., Zhilina, T. & Lysenko, A. M. (1992). DNA relatedness among some thermophilic members of the genus *Methanobacterium*: emendation of the species *Methanobacterium thermoautotrophicum* and rejection of *Methanobacterium thermoformicum* as a synonym of *Methanobacterium thermoautotrophicum*. *Int J Syst Bacteriol* **42**, 408–411.
- Wayne, L. G., Brenner, D. J., Colwell, R. R. & 9, other authors. (1987). International Committee on Systematic Bacteriology. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.
- Welborn, P. P., Hadley, W. K., Newbrun, E. & Yajko, D. M. (1983). Characterization of strains of viridans streptococci by deoxyribonucleic acid hybridization and physiological tests. *Int J Syst Bacteriol* **33**, 293–299.
- Whiley, R. A. & Hardie, J. M. (1989). DNA-DNA hybridization studies and phenotypic characteristics of strains within the 'Streptococcus milleri group'. *J Gen Microbiol* **135**, 2623–2633.
- Whiley, R. A., Fraser, H. Y., Douglas, C. W. I., Hardie, J. M., Williams, A. M. & Collins, M. D. (1990). *Streptococcus parasanguis* sp. nov., an atypical viridans *Streptococcus* from human clinical specimens. *FEMS Microbiol Lett* **68**, 115–122.
- Wilhelm, T., Zhilina, T. N. & Hummel, P. (1991). DNA-DNA hybridization of methylotrophic halophilic methanogenic bacteria and transfer of *Methanococcus halophilus*^{sp} to the genus *Methanohalophilus* as *Methanohalophilus halophilus* comb. nov. *Int J Syst Bacteriol* **41**, 558–562.
- Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.
- Woese, C. R., Achenbach, L., Rouviere, P. & Mandelco, L. (1991). Archaeal phylogeny: re-examination of the phylogenetic position of *Archaeoglobus fulgidus* in light of certain composition-induced artifacts. *Syst Appl Microbiol* **14**, 364–371.
- Xun, L., Boone, D. R. & Mah, R. A. (1989). Deoxyribonucleic acid hybridization study of *Methanogenium* and *Methanocorpusculum* species, emendation of the genus *Methanocorpusculum*, and transfer of *Methanogenium aggregans* to the genus *Methanocorpusculum* as *Methanocorpusculum aggregans* comb. nov. *Int J Syst Bacteriol* **39**, 109–111.
- Yang, D. & Woese, C. R. (1989). Phylogenetic structure of the "leuconostocs": an interesting case of a rapidly evolving organism. *Syst Appl Microbiol* **12**, 145–149.
- Zellner, G., Bleicher, K., Braun, E., Kneifel, H., Tindall, B. J., Conway de Macario, E. & Winter, J. (1989). Characterization of a new mesophilic, secondary alcohol-utilizing methanogen, *Methanobacterium palustre* spec. nov., from a peat bog. *Arch Microbiol* **151**, 1–9.
- Zellner, G., Boone, D. R., Keswani, J., Whitman, W. B., Woese, C. R., Hagelstein, A., Tindall, B. J. & Stackebrandt, E. (1999). Reclassification of *Methanogenium tationis* and *Methanogenium liminatans* as *Methanofollis tationis* gen. nov., comb. nov., and *Methanofollis liminatans* comb. nov., and description of a new strain of *Methanofollis liminatans*. *Int J Syst Bacteriol* **49**, 247–255.