INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY, Oct. 1994, p. 846–849 Vol. 44, No. 4
0020-7713/94/$04.00+0

Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology

E. STACKEBRANDT[1]* AND B. M. GOEBEL[2]

DSM-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, 38124 Braunschweig, Germany,[1] and Department for Microbiology, The University of Queensland, Brisbane 4072, Queensland, Australia[2]

Because a natural entity "species" cannot be recognized as a group of strains that is genetically well separated from its phylogenetic neighbors, a pragmatic approach was taken to define a species by a polyphasic approach (L. G. Wayne, D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper, Int. J. Syst. Bacteriol. 37:463–464, 1987), in which a DNA reassociation value of about 70% plays a dominant role. With the establishment of rapid sequence analysis of 16S rRNA and the recognition of its potential to determine the phylogenetic position of any prokaryotic organism, the role of 16S rRNA similarities in the present species definition in bacteriology needs to be clarified. Comparative studies clearly reveal the limitations of the sequence analysis of this conserved gene and gene product in the determination of relationships at the strain level for which DNA-DNA reassociation experiments still constitute the superior method. Since today the primary structure of 16S rRNA is easier to determine than hybridization between DNA strands, the strength of the sequence analysis is to recognize the level at which DNA pairing studies need to be performed, which certainly applies to similarities of 97% and higher.

Cohn (8, 9) was among the first to address the question of whether bacteria, like animals and plants, can be arranged in distinct taxa. However, the question of whether the form genera and form species were "natural" genera and species was left open for future investigators. Cohn regarded the form genera as natural entities but considered species largely artificial. Since the taxonomic ranks used to describe bacterial form taxa were adopted from those used for the more highly evolved organisms, one can assume that the existence of natural relationships among their members was implied—even when Cohn stressed the artificial character of the classification system.

Almost 100 years later, at the dawn of molecular microbial taxonomy, S. T. Cowan (10) recognized three meanings for the term species: a category, a taxonomic group, and a concept. The category indicates that the species is a taxonomic rank below the genus rank in a hierarchical system. The usefulness of working with the concept of a species is not denied by Cowan, but he reminded the user that the species does not exist and does not represent a natural entity. The zoological definition of a species as "groups of interbreeding or potentially interbreeding natural populations that are reproductively isolated from other such groups" (17) cannot presently be applied to prokaryotes. Even after extensive research, the extent of sexuality is virtually unknown, and reproductive isolation of bacterial strains can probably be excluded.

The inability to define a species or to find a unified species concept reflects the variety of reproductive systems and the dynamic state of biological material. If the term species is used to express memberships of organisms in a taxonomic rank, microbiologists must agree to some guidelines in order to provide stability, reproducibility, and coherency in taxonomy. The past has shown the difficulties in supplying a definition that the majority of taxonomists would regard as a consensus. Realizing this, Cowan (10) was prompted to suggest that "there are as many ideas on species as there are biologists, and many a biologist has changed his idea during the course of his working life." The attempts of taxonomists to work in parallel with different definitions, such as "genospecies," "nomenspecies," and "taxospecies," can be considered a reflection of the inability to integrate subjectively chosen criteria. The circumscription of the species as a taxonomic group defined in terms of the characters of the constituent members (10) took into account the desire of taxonomists to see their individual species concepts verified.

With the arrival of numerical phenetic analysis and molecular techniques in bacterial taxonomy, the inter- and intrarelatedness of species could be determined objectively. Especially, nucleic acid pairing studies on entire genomes or selected genes formed the basis for comparison between species. More recently, gene sequence analysis has been applied. These methods have confirmed, at the level of genetic information, the previously expressed and expected assumption that a natural "species" entity cannot be recognized as a group of strains that is genetically isolated from its phylogenetic neighbors. DNA hybridization is acknowledged as the superior method for the elucidation of relationships between closely related taxa, such as strains and species. Similarly, the potential of sequence analysis of genes coding for rRNA (superseding rRNA cistron similarity studies) and of certain proteins for inferring the evolution of taxa through billions of years is well documented (24). The advent of 16S rRNA sequence analysis and, with some restrictions, nucleic acid reassociation has allowed determination of whether taxa are phylogenetically homogeneous and if not, where the misclassified strains belong within the hierarchical system. Although information obtained from the application of these methods has shown certain species to be phylogenetically isolated, this may merely represent the current limitations of our sequence

* Corresponding author. Mailing address: DSM-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, 38124 Braunschweig, Germany. Phone: 49 531 2616 352. Fax: 49 531 2616 418. Electronic mail address: Stackebrandt@venus.gbf-braunschweig.d400.de.

databases, which are directly coupled to our inability to detect and investigate the full range of bacterial diversity.

The emergence of molecular approaches has led to the formation of several streams within bacterial systematics, e.g., elucidation of phylogenetic relationships, descriptive studies combining phylogenetically and phenotypically based information, environmental studies aiming at revealing uncultured microbial taxa, and the development of diagnostic methodologies. In an attempt to establish a common basis for these separate streams, the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics (23) published a list of taxonomic recommendations with the hope of stimulating discussion on this topic. The committee acknowledged the importance of applying chemotaxonomic approaches (both structural phenetic and phylogenetic [evolutionary]) when inferring or proposing hierarchical levels. There was general agreement that the species was the "only taxonomic unit that can be defined in phylogenetic terms" since, in practice, DNA reassociation approaches the sequence standard and represents the best applicable procedure at the present time. The phylogenetic definition of a species would generally include strains with ". . . . .approximately 70% or greater DNA-DNA relatedness and with 5°C or less ΔT_m," and "Phenotypic characteristics should agree with this definition . . .".

The rationale for using DNA reassociation as the gold standard for species delineation originates from the results of numerous studies, in which a high degree of correlation was found between DNA similarity and chemotaxonomic, genomic, serological, and numerical phenetic similarity. These studies were based on the original finding that single-stranded DNA from two different strains will reassociate to a measurable extent and form a DNA hybrid if the strands contain less than 15% base mispairing (22). Measurements of greater than 70% DNA similarity correlate with differences of less than a few percent in the thermal stability of duplexes formed during reassociation (12). The Ad Hoc Committee's recommendation regarded a thermal stability difference of 5°C or less to be necessary to indicate relationships at the species level. This was based on the results of many hundreds of experiments using several hybridization formats (18). Thermal stabilities have been shown to decrease from 1 to 2.2% for each percent mispairing (6, 22). Although those estimates were made by using experimentally introduced mispairings in only short DNA fragments, one can nevertheless argue that organisms which have 70% or greater DNA similarity will also have at least 96% DNA sequence identity. Considering the total number of bases in the genome of *Escherichia coli* to be 5×10^6, differences of 2%, or 10^5 nucleotides, can account for the significant differences in the phenotype observed among strains of this species. Nevertheless, despite genomic rearrangement caused by horizontal gene transfer and the presence of mobile elements, detectable by bacterial endonuclease restriction patterns, restriction fragment length polymorphism, ribotyping, and randomly amplified polymorphic DNA, the primary structures of the majority of genes are most likely not involved. The species as defined now in microbiology is a rather stable construct because it is defined by a polyphasic approach. Changing the physical map will not markedly influence the extent to which DNA hybridizes, and even if the genetic changes affect one of the characters used in the phenotypic characterization of the species, the DNA similarity values will most likely not change to a measurable extent. This species definition takes into account the possibility that there exists a high degree of genetic rearrangement, gene amplification, mutation, and exchange of genetic material over a nonpredictable range of taxa. What the threshold value of 70% does not take into account is the possibility that the tempo and mode of changes differ in different prokaryotic strains. Sequence analyses of conserved macromolecules have demonstrated that prokaryotes do not evolve isochronically. Whether this effect, however, can be measured by DNA reassociation and whether it influences reassociation values of highly similar DNA sequences are unknown.

The recommendation of the Ad Hoc Committee to continue to apply DNA similarity studies in the circumscription of bacterial species was well received because it was based on a large amount of previous experience and experimental results. The potential of this approach is well proven and has been acknowledged for more than 15 years (12, 21). Numerous hybridization formats have been developed to measure DNA relatedness by hybridization, and excellent congruency exists among these formats (at least among highly related species and strains). In 1987, reports for about 60% of all new species described within existing genera or recognized as synonyms in the *International Journal of Systematic Bacteriology* included DNA reassociation studies, 10% of the new species (mainly spiroplasmas) were described on the basis of serological tests, and 30% of the new species were described with neither of these approaches. In 1993, the corresponding values were 75, 8, and 3%, respectively. However, the remaining 14% of the species descriptions for 1993 contained information inferring the phylogenetic positions exclusively on the basis of 16S rRNA-rRNA analysis.

Through the development of extremely rapid 16S rRNA sequencing protocols and the ensuing large database, it has become obvious that the role of sequence analysis in the circumscription of species must be openly discussed and defined. Either sequence analysis can replace the DNA hybridization component in species delineation, or it can act as an additional parameter. In either instance, it will be necessary to identify whether a threshold 16S rRNA homology value exists which will allow unambiguous recognition of taxa at the species level. The reasons for discussion covering these questions are obvious: taxonomists performing sequence analysis will very soon outnumber those measuring DNA reassociation because of the potential of such analysis to determine the full range of taxonomic ranks; sequences have a rational basis and their accuracy can be tested; and there seems to be no limit to the availability of sequences when we consider the combined application of PCR technology, cloning, and sequencing. However, as more sequence information becomes available, it is evident that the resolution power of 16S rRNA sequences is limited when closely related organisms are being inspected (1, 2, 11).

The primary structure of the 16S rRNA is highly conserved, and species having 70% or greater DNA similarity usually have more than 97% sequence identity. These 3% or 45-nucleotide differences are not evenly scattered along the primary structure of the molecule but are concentrated mainly in certain hypervariable regions. One could argue that these differences could be used as a measure of phylogenetic distance between strains and that for the determination of this level of relatedness, sequence analysis could be restricted to these regions for this close level of relationships. Several arguments speak against this conclusion. First, the amount of difference not concentrated in the hypervariable regions is substantial (20), and by omission one would lose information which already is low in quantity from a statistical point of view. Second, the positions of the hypervariable regions are taxon specific and need to be determined for novel organisms by sequence analysis of the complete molecule. Sequence analysis of 1,540 nucleotides has become rapid and inexpensive, and only complete sequences

TABLE 1. Various degrees of homology in stretches of 200 nucleotides along the primary structure of pairs of 16S rRNAs from organisms with different degrees of relatedness

Position	16S rRNA homology (%)[a] between:		
	Streptomyces ambofaciens and *Streptomyces violaceoruber*	*Mycobacterium phlei* and *Mycobacterium tuberculosis*	*Aeromicrobium erythreum* and *Rhodococcus fascians*
Overall	98.8	96.4	90.9
0–200	**96.3**	94.1	**80.7**
201–400	98.4	97.8	94.6
401–600	100.0	93.1	94.6
601–800	99.0	97.9	85.7
801–1000	100.0	100.0	94.0
1001–1200	98.9	**92.8**	90.0
1201–1400	99.5	100.0	94.0

[a] The lowest values are in boldface.

allow reliable phylogenetic comparison with the available database of complete or almost complete sequences. Very short sequences have a negative influence on the stability of phylogenetic trees, and if because of budget constraints the number of nucleotides analyzed needs to be restricted to only a few hundred, then the regions need to be selected with care in order to obtain the same degree of similarity as that obtained from analyses of full sequences (Table 1). Third, multiple changes which are highly likely to accumulate in the hypervariable regions will certainly obscure the genealogy of the molecule by simulating false identities and masking the actual number of evolutionary events. Even with full sequences, a standardization of similarities or dissimilarities for the delineation of species over the whole range of main lines of descent of the domains *Archaea* and *Bacteria* appears not to be feasible. Recent evolutionary progress, expressed in the diverging primary-structure DNA and changes in phenotype, occurs at different rates in different lines of descent but may not be shown directly at the level of the conservative rRNA genes. Calibration of the 16S rRNA clock of closely related symbionts of aphids revealed approximately 2 to 4% fixed substitutions per 100 million years (14, 15), which is in the same range (1% fixed substitutions per 60 million years) calculated for different bacterial symbionts (16). It appears plausible that within this epoch significant progress can be made in the speciation of prokaryotic species. It is interesting that shortly after the establishment of 16S rRNA cataloging, ranges of S_{AB} values were indicated for each taxonomic level within the group of methanogenic bacteria (5). The species rank was indicated by a range of S_{AB} values of 0.55 to 0.65, which corresponds to 16S rRNA homologies of approximately 90 to 94% (24). The inability of this attempt to reflect the true diversity of methanogenic species was recognized, and it was not continued.

The correlation plot of the two phylogenetic parameters DNA similarity and 16S rRNA homology is not linear (Fig. 1). It is obvious that each method is strong in those areas of relationships in which the other method fails to reliably depict relationships. Sequence analysis of 16S rRNA is superior from the level of domains (starting at about 55% homology) to moderately related species, i.e., below 97.5% similarity. Above this value, DNA reassociation values can either be low or as high as 100%. Several groups of organisms have been identified which share almost identical 16S rRNA sequences but in which DNA hybridization is significantly lower than 70%, thus indicating that they represent individual species. The terms rRNA species complex and rRNA superspecies have been proposed for those organisms that have virtually identical 16S rRNA sequences and which are not likely to be the same species by DNA-DNA hybridization criteria as well (11). This may be true for absolutely identical 16S rRNA sequences, but the number of examples is small. The fact that DNA reassociation values may be as low as 25% at rRNA similarity values

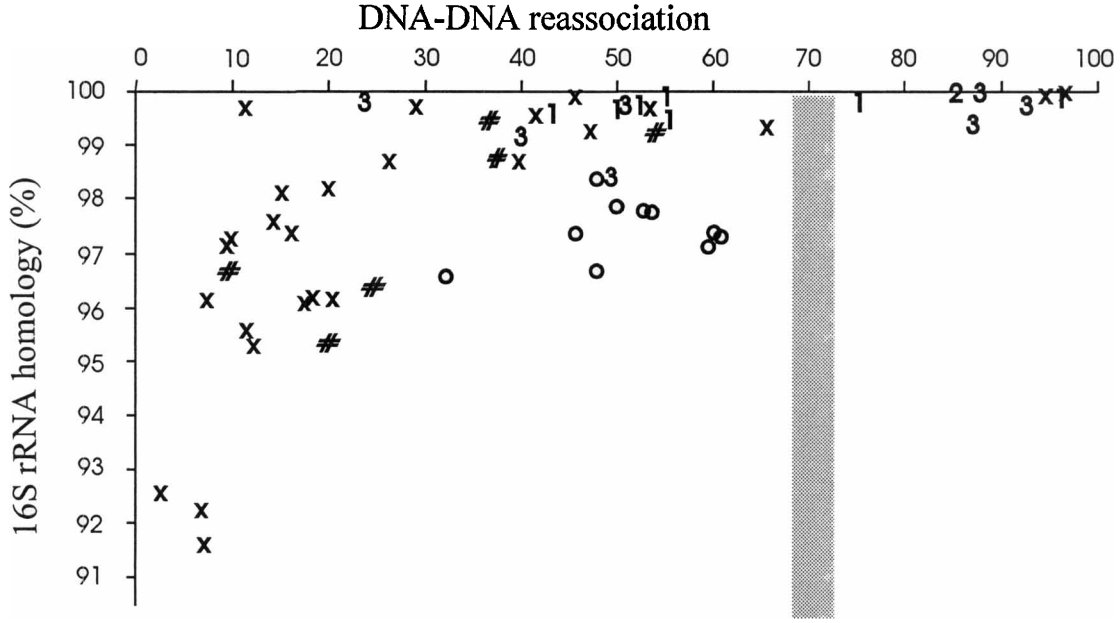

FIG. 1. Comparison of 16S rRNA homology and DNA-DNA reassociation values. Symbols and references for the data are as follows: x, membrane filter method (1); #, renaturation rate method (3), 1, renaturation rate method (4), 2, renaturation rate method (13); 3, renaturation rate method (19); ○, S1 nuclease method (21). The bar indicates the DNA threshold value for species delineation.

of 99.8% indicates that the resolution power of DNA hybridization is significantly higher than that of sequence analysis (1) and that DNA hybridization remains the optimal method for measuring the degree of relatedness between highly related organisms. However, it is interesting that at sequence homology values below about 97.5%, it is unlikely that two organisms have more than 60 to 70% DNA similarity and hence that they are related at the species level.

The rapid and reliable identification of strains remains the most important task in taxonomy. Isolates that can be identified positively become part of the classification system. Organisms not classifiable because of a low character match to any described species and its unique phylogenetic position need to be reinvestigated in order to obtain taxon-specific characters. Only then is this isolate assigned its place in the system. The range of methods to perform the task of identification and classification has been extended significantly through the introduction of molecular techniques. Through the application of PCR technologies, automated sequencing, pulsed-field electrophoresis, and oligonucleotide probing, the speed at which molecular analyses can be performed is breathtaking. Compared with these methods, the several DNA hybridization formats with their noncumulative character and dependence on several physicochemical parameters appear outdated. Nevertheless, evidence is strong that sequence analyses of 16S rRNA is not the appropriate method to replace DNA reassociation for the delineation of species and measurement of intraspecies relationships. Eventually a gene or gene cluster will be found which, except for its highly conservative character, shares all the benefits of a phylogenetic marker with the large 16S rRNA species and whose application to phylogenetic analyses covers the same range that DNA reassociation does today. 16S rRNA analysis is a most valuable addition to the polyphasic approach to bacterial classification, and for the species level it is extremely helpful in deciding whether the laborious DNA reassociation needs to be performed. According to the available compilation of data, organisms that have less than 97.0% sequence homology will not reassociate to more than 60%, no matter which hybridization method is applied. This rRNA threshold value should be lower in order to keep flexibility in the recommendation for a phylogenetic definition of a species by parameters of DNA reassociation. In the end it is the presence or absence of phenotypic coherency among strains that should be the deciding factor about whether to describe species at all and whether to delineate species at the 60% or at the 80% DNA-DNA similarity level.

REFERENCES

1. **Amann, R. I., C. Lin, R. Key, L. Montgomery, and D. A. Stahl.** 1992. Diversity among *Fibrobacter* strains: towards a phylogenetic classification. Syst. Appl. Microbiol. **15**:23–31.
2. **Ash, C., J. A. E. Farrow, M. Dorsch, E. Stackebrandt, and M. D. Collins.** 1991. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int. J. Syst. Bacteriol. **41**:343–346.
3. **Baess, I.** 1982. Deoxyribonucleic acid relatedness among species of rapidly growing mycobacteria. Acta Pathol. Microbiol. Immunol. Scand. Sect. B **90**:371–375.
4. **Baess, I.** 1983. Deoxyribonucleic acid relationships between different serovars of *Mycobacterium avium*, *Mycobacterium intracellulare* and *Mycobacterium scrofulaceum*. Acta Pathol. Microbiol. Immunol. Scand. Sect. B **91**:201–203.
5. **Balch, W. E., G. E. Fox, L. J. Magrum, C. R. Woese, and R. S. Wolfe.** 1979. Methanogens: reevaluation of a unique biological group. Microbiol. Rev. **43**:260–296.
6. **Bautz, E. K. F., and F. A. Bautz.** 1964. The influence of noncomplementary bases on the stability of ordered polynucleotides. Proc. Natl. Acad. Sci. USA **52**:1476–1481.
7. **Brenner, D. J., A. G. Steigerwalt, A. Ibrahim, and E. Stackebrandt.** Unpublished data.
8. **Cohn, F.** 1872. Untersuchungen über Bacterien. Beitr. Biol. Pfl. Heft 2 **1**:127–224.
9. **Cohn, F.** 1876. Untersuchungen über Bacterien IV. Beitr. Biol. Pfl. Heft **2**:249–277.
10. **Cowan, S. T.** 1968. A dictionary of microbial taxonomic usage. Oliver and Boyd, Edinburgh.
11. **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.
12. **Johnson, J. L.** 1985. DNA reassociation and RNA hybridization of bacterial nucleic acids. Methods Microbiol. **28**:33–74.
13. **Ludwig, W., G. Mittenhuber, and C. G. Friedrich.** 1993. Transfer of *Thiosphaera pantotropha* to *Paracoccus denitrificans*. Int. J. Syst. Bacteriol. **43**:363–367.
14. **Moran, N., and P. Baumann.** 1994. Phylogenetics of cytoplasmically inherited microorganisms of arthropods. Trends Ecol. Evol. **9**:15–20.
15. **Moran, N. A., M. A. Munson, P. Baumann, and H. Ishikawa.** 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. London Ser. B **253**:167–171.
16. **Ochman, H., and A. C. Wilson.** 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26**:74–86.
17. **Ravin, A. W.** 1963. Experimental approaches to the study of bacterial phylogeny. Am. Nat. **97**:307–325.
18. **Schleifer, K. H., and E. Stackebrandt.** 1983. Molecular systematics of prokaryotes. Annu. Rev. Microbiol. **37**:143–187.
19. **Spaeth, G.** 1989. Überprüfung der taxonomischen Stellung von *Bacillus anthracis* mit Hilfe der DNA Hybridisierungs Technik. Ph.D. thesis. University Giessen, Giessen, Germany.
20. **Stackebrandt, E., W. Liesack, and D. Witt.** 1992. Ribosomal RNA and rRNA sequence analyses. Gene **115**:255–260.
21. **Steigerwalt, A. G., G. R. Fanning, M. A. Fife-Ashbury, and D. J. Brenner.** 1976. DNA relatedness among species of *Enterobacter* and *Serratia*. Can. J. Microbiol. **22**:121–137.
22. **Ullmann, J. S., and B. J. McCarthy.** 1973. The relationship between mismatched base pairs and the thermal stability of DNA duplexes. Biochim. Biophys. Acta **294**:416–424.
23. **Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper.** 1987. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. Int. J. Syst. Bacteriol. **37**:463–464.
24. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51**:221–271.