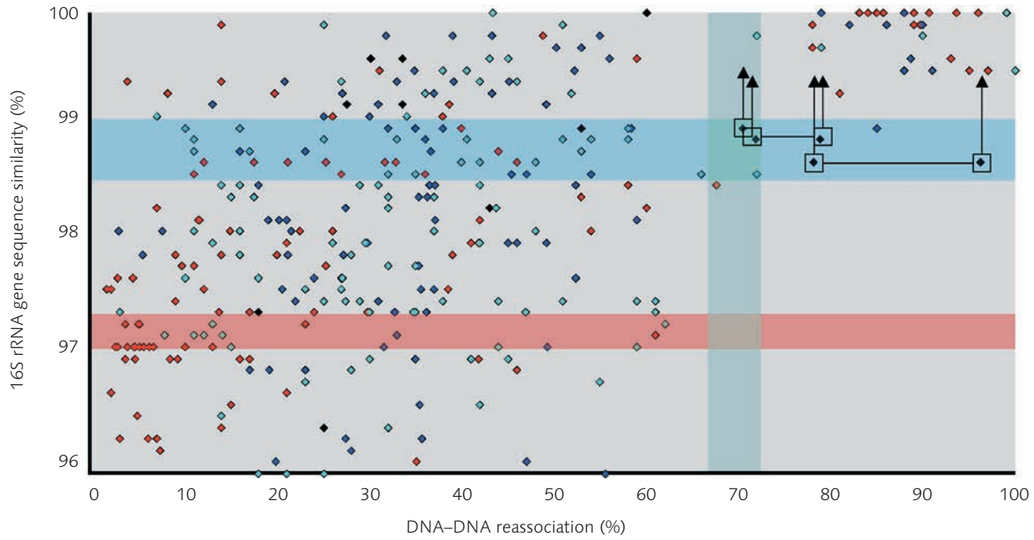
Taxonomic parameters revisited: tarnished gold standards

Molecular properties are included in the definition of a 'species'. Exciting new findings announced here by **Erko Stackebrandt** and **Jonas Ebers** show that a 16S rRNA gene sequence similarity range above 98.7–99 % should be mandatory for testing the genomic uniqueness of a novel isolate. This overturns the old value of 97 % and will greatly facilitate the work of taxonomists.

With the inclusion of defined genomic properties in 'minimal standards' of taxon descriptions, molecular data are now fully acknowledged in systematic studies of prokaryotes. Depending on the rank of a taxon, these approaches are either mandatory or optional. At the taxonomic level of 'species', molecular properties serve two requirements: first, to verify the morphological, biochemical and chemotaxonomic coherence of strains of a 'species' by their similarities (preferably identity) at the genomic level and, second, to delineate this taxon from phylogenetically neighbouring species of the genus (Wayne *et al.*, 1987; Rosselló-Mora & Amann, 2001). As the taxon 'species' represents populations that themselves are the result of different mechanisms and tempi of evolution (Stackebrandt *et al.*, 2002; Gevers *et al.*, 2005), the degree of deviation from nearly absolute phenotypic and genomic identity (as expected to occur in clones) requires from taxonomists a balanced judgement of evolutionary processes that they may possibly not be aware of. In order to facilitate and harmonize taxonomic decisions in a field in which the Biological Species Concept does not apply, an arbitrary and artificial definition has evolved over a century of bacterial taxonomy (Staley & Krieg, 1984; Stackebrandt, 2000); today, the description of the construct 'species' is more stringently controlled by recommendations than that of any other taxon. While in the pre-Approved Lists era, taxonomists were allowed to follow their own subjective judgements, the past 25 years have witnessed a more objective and internationally controlled verification process of 'species' descriptions.

The predictability of the uniqueness of a '*species novum*' has been largely strengthened by the universal applicability of molecular data. Methods applied, to name a few, embrace approximate characterization of the chromosome by determination of the base composition (mol% G+C content) and degree of reassociation of single-stranded DNA (DNA–DNA hybridization) as well as comparison of one-dimensional restriction and PCR patterns (Pukall, 2005); other methods focus on genes and operons, encoding rRNA and proteins, including typing and sequencing. Each of the methods applied has its strength in elucidating a defined range of the 4-billion-year evolution of prokaryotes. Though several molecular methods have their merits in taxonomy, two approaches, the 'gold standards', play a dominant role: DNA–DNA hybridization for 'species' delimitation, and 16S rRNA gene sequence similarities for unravelling more distant relationships among strains. DNA–DNA hybridization can be expressed as percentage reassociation similarity or ΔT_m of reassociated DNA strands (Wayne *et al.*, 1987), but only the first parameter is in general use. This judgement appears objective when browsing through the past 15 or so volumes of the *International Journal of Systematic and Evolutionary Microbiology* (*IJSEM*) (formerly *International Journal of Systematic Bacteriology*), the official publication of the International Union of Microbiological Societies (IUMS). Almost every species description contains a phylogenetic analysis of the type strain based on 16S rRNA gene sequence similarity comparison and many novel species are delineated from their phylogenetic neighbours by DNA–DNA reassociation values below 70 %.

◀ Gold standards. *Digital Vision*



▶ Fig. 1. Comparison of 16S rRNA gene sequence similarities and DNA–DNA reassociation values. Data have been compiled from publications containing species descriptions from *IJSEM* **55** (2005). The different colours refer to broad categories of reassociation methods: red, microtitre plate technique, e.g. Ezaki *et al.* (1989); dark blue, spectrophotometric technique, e.g. De Ley *et al.* (1970); light-blue, membrane filter method, e.g. Tourova & Antonov (1987); black, other methods, e.g. dot hybridization (Amakata *et al.* 2005), or not defined. Horizontal rules between squares indicate data obtained by two different reassociation methods. Arrows point to the position of *in silico*-recalculated binary 16S rRNA gene sequence similarity values of sequences deposited by Amakata *et al.* (2005). The horizontal blue bar indicates the threshold range above which it is now recommended to perform DNA–DNA reassociation experiments; the horizontal red bar indicates the threshold values published previously (Stackebrandt & Goebel, 1994). *E. Stackebrandt & J. Ebers*

Despite the importance of the DNA–DNA reassociation approach, most microbial taxonomists are not in a position to perform these studies and need collaboration with specialized laboratories. Experience is needed in isolation and purification of DNA and, although one can choose from a variety of different hybridization methods (Rosselló-Mora, 2005), none of these is straightforward to apply without thorough training. But these are not the only reasons for the aversion to this technique: the method of reassociation of denatured DNA strands of two different strains unfolds the homologous genome stretches that are involved in the reassociation process. In these times of complete genome sequences and the teaching of sequence techniques to undergraduates, this failure to examine the mechanisms behind the process makes DNA–DNA reassociation seem like a method salvaged from the past. Also, a significant number of physico-chemical parameters, genome size, the presence of large plasmids, DNA purity and other factors, influence the hybridization results; reciprocal values may differ by up to 15 %. Unlike sequences, which must be deposited in public databases for inspection of quality, no reviewer of a new species description is in a position to judge DNA reassociation values. Last, but not least, as the data are not cumulative, studies on a large number of closely related species (i.e. in the genera *Streptomyces*, *Pseudomonas*, *Aeromonas* or certain groups of *Bacillus*) may become a search for the end of the rainbow. However, it must be conceded that the requirement to provide evidence for overall genomic and phenetic similarity among members of a species on the one hand, while proving dissimilarity of character traits between members of different species on the other, works well, and has set the stage for stability in prokaryotic taxonomy, keeping in mind the artificial definition with which 'species' are described.

At the beginning of the 1990s, with the release of the avalanche of 16S rRNA gene sequences, it became obvious that sequence similarities and DNA reassociation values obtained for the same strain pairs do not show a linear relationship (Rosselló-Mora & Amann, 2001; Fox *et al.*, 1992; Stackebrandt & Goebel, 1994). It could be demonstrated on the basis of a limited dataset that, below a threshold value of 98.5 % gene sequence similarity, the corresponding DNA reassociation values were always lower than 70 %. In order to reduce the workload involved in DNA–DNA reassociation experiments, it was suggested that reassociation experiments need only be performed for strains that shared 16S rRNA gene sequence similarities higher than about 97.0 % (Stackebrandt & Goebel, 1994). This value was lower than that determined from the literature, but was suggested from a taxonomically conservative point of view. Having been cited more than 1,350 times since 1994, this demarcation value indeed turned out to be a guide for researchers and reviewers. In order to update the correlation between these two taxonomic parameters, we screened all articles published in volume 55 of *IJSEM* and are now able to revise the results published in 1994 (Fig. 1). Rather than 97.0 %, we now recommend a 16S rRNA gene sequence similarity threshold range of 98.7–99 % as the point at which DNA–DNA reassociation experiments should be mandatory for testing the genomic uniqueness of a novel isolate(s). The graph compiles 380 data points obtained by all major hybridization techniques performed on representatives of most phyla of prokaryotes (only values above 96 % similarity are shown in Fig. 1). Only two studies on three strain pairs revealed that, at 16S rRNA gene sequence similarity lower than 99 %, the corresponding DNA reassociation values were higher than 70 %.

Our criticism is also directed towards a somewhat careless handling of 16S rRNA gene sequences. Many sequences deposited in public databases appear to be direct downloads from computer printouts, lacking rigorous inspection of quality and secondary-structure feasibility. As sequence errors decrease rather than increase similarity values, the relatedness between organisms with erroneous sequences is lowered. Indeed, there are some sequences of highly related strains which show deviations from highly conserved secondary structure features. We have critically analysed the quality and similarity values of 16S rRNA gene sequences for one data set that show higher than 70 % DNA reassociation values at 16S rRNA gene sequence similarities below 99 % (marked in black in Fig. 1). Though we had access to neither the strains nor the original data, *in silico* corrections of nucleotide idiosyncrasies meant that the similarity values increased by up to 0.8 %, pushing them over the 99 % line: these values are indicated by arrows in Fig. 1. It appears to be critically important to check the quality of sequences according to secondary features prior to deposition in public databases.

This recommended increase of about 2.0 % in 16S rRNA gene sequence similarity will significantly facilitate the work of taxonomists without sacrificing the quality and precision of a 'species' description. As indicated above, DNA–DNA hybridization constitutes the bottleneck of taxonomic studies among closely related 'species', and taxonomists should acknowledge the updated correlation curve and welcome the expected reduction in workload. As the artificial cut-off value of around 70 % reassociation may not have phylogenetic significance, rare examples may exist and will arise in the future in which reassociation values around and above 70 % emerge at corresponding 16S rRNA gene sequence similarities around 99.0 %. In these cases, taxonomists are reminded of the article of Wayne *et al.* (1987), summarizing an overall concern of these authors '*that any phylogenetically based taxonomic schemes that result must also show phenotypic consistency*'.

Erko Stackebrandt and **Jonas Ebers**
DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen, Mascheroder Weg 1b, 38124 Braunschweig Germany
(E. Stackebrandt: t +49 53 1261 6352; e erko@dsmz.de)

References

Amakata, D., Matsuo, Y., Shimono, K., Park, J.K., Yun, C.S., Matsuda, H., Yokota, A. & Kawamukai, M. (2005). *Mitsuaria chitosanitabida* gen. nov., sp. nov., an aerobic, chitosanase-producing member of the '*Betaproteobacteria*'. *Int J Syst Evol Microbiol* **55**, 1927–1932.

De Ley, J., Cattoir, H. & Reynaerts, A. (1970). The quantitative measurement of DNA hybridization from renaturation rates. *Eur J Biochem* **12**, 133–142.

Ezaki, T., Hashimoto, Y. & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid–deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *Int J Syst Bacteriol* **39**, 224–229.

Fox, G.E., Wisotzkey, J.D. & Jurtshuk, P., 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.

Gevers, D., Cohan, F.M., Lawrence, J.G. & 8 other authors (2005). Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**, 733–739.

Pukall, R. (2005). DNA fingerprinting techniques applied to the identification, taxonomy and community analysis of prokaryotes. In *Molecular Identification, Systematics, and Population Structure of Prokaryotes*, pp. 52–82. Edited by E. Stackebrandt. Heidelberg: Springer.

Rosselló-Mora, R. (2005). DNA–DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In *Molecular Identification, Systematics, and Population Structure of Prokaryotes*, pp. 23–50. Edited by E. Stackebrandt. Heidelberg: Springer.

Rosselló-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol Rev* **25**, 39–67.

Stackebrandt, E. (2000). Defining taxonomic ranks. In *The Prokaryotes. An Evolving Electronic Resource for the Microbiological Community*, release 3.1. Edited by M. Dworkin and others. New York: Springer. http://link.springer-ny.com/link/service/books/10125/

Stackebrandt, E. & Goebel, B.M. (1994). 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.

Stackebrandt, E., Frederiksen, W., Garrity, G.M. & 10 other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043–1052.

Staley, J.T. & Krieg, N.R. (1984). Bacterial classification I. Classification of procaryotic organisms: an overview. In *Bergey's Manual of Systematic Bacteriology*, vol. 1, pp. 1–4. Edited by N.R. Krieg & J.G. Holt. Baltimore: Williams & Wilkins.

Tourova, T.P. & Antonov, A.S. (1987). Identification of microorganisms by rapid DNA–DNA hybridization. *Methods Microbiol* **19**, 333–355.

Wayne, L., 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.