E. STACKEBRANDT: Richness of Prokaryotic Diversity, *Food Technol. Biotechnol.* **41** (1) 17–22 (2003)

17

The Richness of Prokaryotic Diversity: There Must Be a Species Somewhere

Erko Stackebrandt

DSMZ-German Collection of Microorganisms and Cell Cultures, Mascheroder Weg 1b,
D-38124 Braunschweig, Germany

Summary

The rapidly increasing number of potentially novel species, combined with the methodologically laborious polyphasic approach used in bacterial systematics, makes identification and, consequently, the description of novel taxa a highly demanding discipline. The number of new descriptions of about 200 species a year can apparently not be expanded significantly but new isolation procedures and renewed interest in working with bacteria rather than with DNA clones provide microbiologists with a broad range of different pheno- and genotypes. It seems obvious that some of the worldwide acknowledged techniques will not be continued but the question which methods will be used in the future is still open. At present, DNA-DNA reassociation is the final arbiter for the taxon »species«, but the limitations of these techniques, which are subject to significant errors, are well known. On the other side of the range of molecular techniques stands a whole genome sequencing that demonstrates the evolutionary events which lead to differences in DNA, such as point mutations, insertions and deletions, DNA acquisition and loss, recombination, gene loss and formation of pseudogenes. Each of these evolutionary events has an effect on DNA-DNA reassociation similarity, which, however, cannot be tested by the reassociation approaches used. Further, DNA pairing does not allow the establishment of a cumulative database because the hybridization parameters change from laboratory to laboratory and reference organisms need to be included in each experiment.

This communication reports on the development of ideas that may change the paradigm of the present concept of artificial, arbitrary and pragmatic species into a species definition, driven by insights into population genetics.

Key words: microbial diversity, species, population genetics

Introduction

Imagine the number 5 by 10 to the power 30 ($5 \cdot 10^{30}$)! This is the estimated number of prokaryotic cells on the planet earth against which the estimated number of stars in the Milky Way (10^{12}) appears insignificant. Prokaryotic cells are described to occur roughly to 25 % each in the water column, soil, sediments and deep subsurface, respectively (*1*). The numbers of cells range from 10^5 per mL in seawater to 10^8 in certain soils to 10^{10} in manure. Another staggering number is that of different genomes in 1 g of Norwegian soil, estimated by cot curves of DNA-DNA reassociation to be around 4000 (*2*). In contrast to the mere number of cells, the number of different genomes indeed measures different genomic entities. However, several factors influence

Corresponding author; E-mail: erko@dsmz.de

DNA renaturation, *e.g.*, genome architecture, differences in gene sequence, loss versus possession of genes, gene duplication and the like (*3*), and it is therefore not known whether the 4000 different genomes reflect 4000 different species. This is indeed unlikely, as the present species definition allows for a certain degree of genomic heterogeneity. This is exemplified by the assumption that organisms that share ≥70 % DNA similarities share at least 96 % DNA sequence identity. Considering the totality of bases in the genome of *E. coli* to be approximately $4 \cdot 10^6$, then 4 % differences equates to $1.6 \cdot 10^5$ nucleotide differences (not taking into account the possibility that genome rearrangement is a source of decrease in DNA similarity). This genomic divergence between strains of a single species can easily account for the significant differences in phenotype observed among strains of *e.g. E. coli*. This is seen in *Helicobacter pylori* strains and in *E. coli* K12 and *E. coli* 0157, each pair differing in about 20 % of their DNA, and in pseudomonad species having a genome sizes which range from 3.7–4.7 Mbp in *Pseudomonas stutzeri* or 5.2–7.1 Mbp in *Pseudomonas aeruginosa*.

Even if we allow the number of 4000 genomes in a single gram of randomly selected Norwegian soil to represent only 500 different prokaryotic species, then this sample would contain about one tenth of all validly described species. And if we consider the validly described species to originate not only from North European soil, but from other soil types, marine and freshwater samples, from the Southern hemisphere, from deserts, polar regions and other extreme biotopes as well, one can assume that not only the Norwegian soil sample but any sample worldwide, irrespective of origin, will contain a significant number of novel species. This assumption, though not proven for this very sample, has been verified in many habitats subjected to culture and molecular studies.

Determination of Diversity: Uncultured and Cultured Organisms

Prokaryotic diversity of bacterial 16S rDNA clones, determined for a microbial mat sample from the moated region of Lake Fryxell, McMurdo Dry Valleys, Antarctica, should serve as an example (*4,5*). The molecular diversity was assessed on the basis of more than 350 partial 16S rDNA clone sequences of libraries generated by *Bacteria*- and *Archaea* specific PCR primers. Among the clone sequences, several phyla were represented, including *Proteobacteria*, *Verrucomicrobiales*, the two subphyla of Gram-positive bacteria (class *Actinobacteria*, and *Clostridium/Bacillus* subphylum), and the *Cytophaga-Flavobacterium-Bacteroides*. As judged from the similarity values of clone sequences found among them and with those of cultured organisms the clones formed about 133 potential species, as their sequence similarity was higher than 98 %. Of these, 70 groups that represent potential new species emerged since they showed less than 98 % similarity. Rarely only clone sequences were found to be highly related to Lake Fryxell isolates and to strains of described species. The sample appears to be rich in anaerobic fermenters, as exemplified by the presence of members of the *Clostridium* line of descent of Gram-posi-

tive organisms and members of *Bacteroides*. The community contained putative novel species related to anaerobic saccharolytic organisms (*e.g. C. estertheticum, C. fallax, C. lentocellum, Eubacterium contortum* and *E. saburreum*) forming CO_2, H_2 and C_1-C_4 acids and alcohols, which should be metabolized by other anaerobic members of the community (*e.g.* methanogens, acetogens [*Sporomusa, Acetonema, Acetobacterium*]), as well as by aerobes, facultative anaerobes and phototrophic organisms. Polysaccharides may be derived from primary producers described to be abundant in Lake Fryxell and by decomposers, *e.g. C. lentocellum, C. vincentii* and *C. termitidis*-like clostridia. Other putative novel species can be proteolytic (*C. limosum, C. subterminale*) or specialists, thriving only on a restricted number of compounds (*e.g. Pelobacter* sp.). The findings of some sequences related to those of sulfur compounds metabolizing organisms (*Desulfomonile* sp., *Desulfosporosinus* and the aerobic *Bosea thiooxidans*) indicated the presence of a sulfur cycle.

While the analysis of 16S rRNA gene clone libraries determines the phylogenetic diversity, hence potentially available species in a given environment, the challenge is to match this diversity with that of cultured organisms. This has been attempted by the isolation of more than 800 strains but their identification revealed little similarity with the uncultured diversity. The routine involved in isolation procedures, either in academic laboratories, targeting defined physiological or morphological types, or in industrial laboratories, targeting the »rare« taxa with potentially new metabolic properties, is laborious. The bottleneck in any isolation strategy is not to obtain a high number of sufficiently diverse isolates but to apply sufficiently discriminating techniques that would rapidly separate taxonomically novel from known strains. Besides difficulties connected with the reliable identification of large strain collections, there are two other main problems: the same strain is likely to express different phenotypes on different growth media, therefore the diversity may be overestimated; on the other hand, as the phenotypic appearance of phylogenetically diverse organisms may be similar, the diversity can be underestimated and the reduction of apparently similar types to a workable number of isolates can lead to a man-made reduction of diversity of a given sample.

The challenge is, therefore, to apply rapid methods for identifying strains belonging to the same taxon, even if they may differ slightly in phenotypic and morphological properties. Several molecular tools have been described for the determination of inter- and intra-species relation, based on rapid DNA typing methods (*6,7*), such as those targeting chromosomes (AFLP, RAPD, ERIC, BOX, REP, PFGE), gene clusters (ribotyping of *rrn* operons), individual genes (ARDRA and T-RFLP of 16S rDNA), and intergenic 16S-23S rRNA gene spacer regions. As these methods require a series of molecular techniques, they cannot be considered rapid enough for the investigation of hundreds, if not thousands, of isolates.

The characterization and/or identification of isolates has been improved by applying physical methods to prokaryotic cells, such as Fourier transformed infrared spectroscopy (FT-IR) (*8*), pyrolysis mass spectrometry, and matrix-assisted laser desorption/ionization with

time-of-flight mass (Maldi/Tof) (*9*). Of these, FT-IR is most widely used. It determines the IR spectra of dry bacterial layers in the range of 4000–500 cm^{-1} (2.5–20 µm). These are characterized by many peaks that may overlap each other. Individual peaks can grossly be affiliated to certain compounds, such as H$_2$O (~3500), fatty acids (~2850), amid I (~1700 = carbonyl groups), amid II (~1550 = amino groups) and polysaccharides (~1100); two mixed areas from 1500–1200 cm^{-1} (esters, aromatic compounds, fatty acids, phosphodiesters) and the so called fingerprint area from 900–700 cm^{-1}.

Scientists can choose certain spectral windows and put different weights to selected windows, *e.g.* 3000–2798, 1200–900 and 901–700, with the first range scored three-fold. This is done in order to restrict the analysis to those components which are most discriminative, thus of taxonomic power. Comparative analyses of different sets of spectral windows will show whether or not the composition of clusters will remain stable. The disadvantages of working with FT-IR patterns are: (*i*) the restriction of isolates able to grow on the same growth medium, and (*ii*) small size of the commercial FT-IR database, which is restricted mainly to clinical strains and strains used in food processing. The database does not include the majority of type strains originating from other environments. Thus, FT-IR patterns generated from hundreds of isolates from *e.g.* the Antarctic environment cannot be used for classification, as patterns of reference strains are missing. However, the dendrogram of FT-IR pattern relationships clusters highly related isolates and can thus be used for sorting. Comparative analysis of strain clusters by 16S rRNA gene sequencing, ribotyping, fatty acid analysis and DNA-DNA reassociation revealed that members of tight FT-IR clusters are indeed phylogenetically, genomically and chemo-taxonomically highly related.

FT-IR analysis is inexpensive and rapid but it does not provide any qualitative or quantitative data on cell constituents. Once the isolates are sorted into clusters, individual members need to be included in an identification process. One of the most widely used methods in identification is the analysis of fatty acid methyl esters (FAMEs) determined by *e.g.* the microbial identification system (MIDI, Newark, USA). The patterns, however, are rarely species-specific but confirm affiliation of strain at the genus level. Phylogenetically novel isolates with no previous entries into the fatty acid database cannot be affiliated by this method.

From Strains to Species

Sequence analysis of 16S rRNA genes of cultured organisms is an acknowledged method and the extensive database of sequences makes phylogenetic affiliation likely. However, the question whether or not an isolate can be considered a member of a described species cannot be answered unambiguously. Due to the extremely slow evolution of rRNA genes organisms may belong to different species as judged by DNA-DNA hybridization even with higher than 98 % 16S rRNA gene sequence similarities (*10*). Another factor that influences the quality of 16S rRNA gene sequences is the occur-

rence of microheterogeneity within multiple *rrn* operons (*11*). Length heterogeneity and the presence of different nucleotides at the same position in different operons will lead to undeterminable sequences and positions, respectively, thus to low quality sequences. In order to solve these problems, cloning and renewed sequencing are needed to obtain the accurate sequence. Fig. 1 is an example of two isolates from an Antarctic mat sample containing microheterogeneities in their *rrn* operons. Each of the isolates contains at least four different operons, which are more closely related among themselves than to a sequence of the other strain. The taxonomic affiliation of the two isolates to new clostridial species, originally derived from the low-quality sequence, is not challenged.
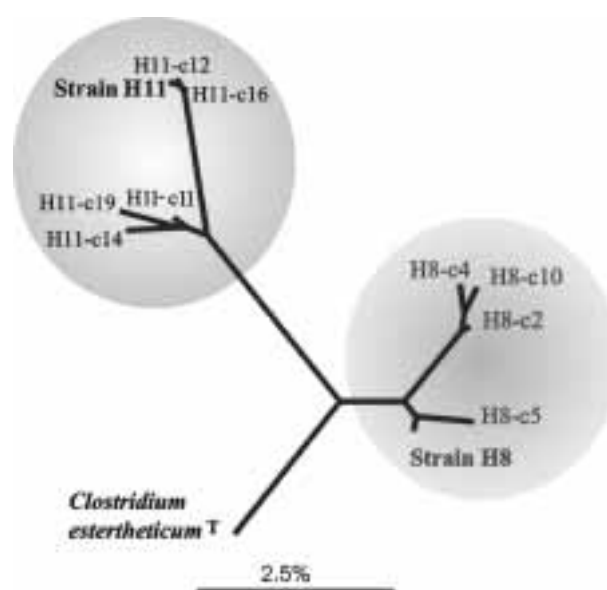

Fig. 1. Diversity of 16S rDNA clones of two strains of Antarctic isolates related to *Clostridium estertheticum*

Sequences obtained from studies on environmental DNA do not show the microheterogeneity effects, as these sequences already derive from cloned operons. The presence of bush-like clusters of certain groups in phylogenetic trees obtained with environmental samples may therefore not be due to the diversity of organisms but to genomic diversity, caused by sequence heterogeneity of individual operons of a few species only (*12*).

Molecular approaches are growth-medium independent, such as stated above, AFLP, ARDRA, T-RFLP and ribotyping. The latter method is widely performed either manually or in its more expensive but highly reproducible variation, by a robot. The patterns, compared to a curated database, are highly specific and appropriate for determination of intraspecific relationships. While in some species the patterns are highly similar and even apt to determine misclassified strains and synonyms (Fig. 2), strains of other species show highly dissimilar riboprints, even when their affiliation to the same species has been verified by DNA-DNA hybridization (*13*).
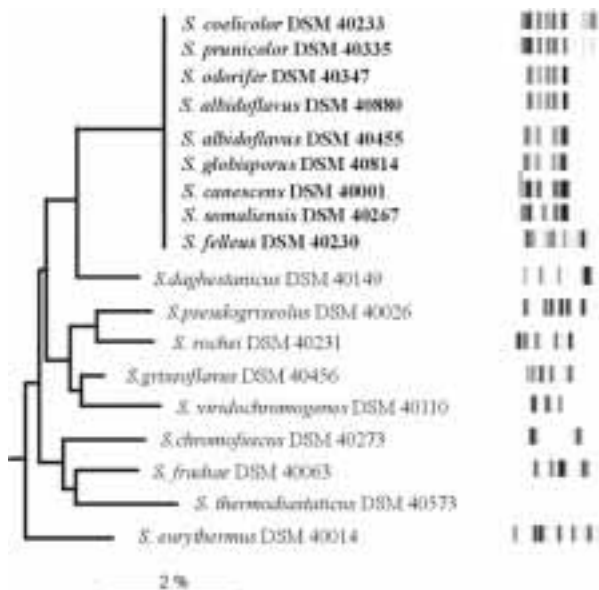
Fig. 2. Characterization of *Streptomyces* strains by riboprint analysis. Strains are arranged according to 16S rRNA gene sequence analyses (Schumann and Stackebrandt, unpublished).

The occurrence of different riboprints in strains of the same species supports the earlier notion derived from multi-locus enzyme typing and from DNA patterns that the genomic diversity within the taxon »species« may be high (*14*). Today the boundaries of the taxon »species« are defined by a high degree of phenotypic and genomic similarity. The phenotypic component includes metabolic and physiological properties, which may indeed differ to a certain degree among the strains of a species, as well as chemotaxonomic properties, that are generally conserved within a species. The genomic similarity is measured by DNA-DNA hybridization and, as outlined by Wayne *et al.* (*15*), a threshold value of around 70 % is recommended, while the strains that share higher than 70 % DNA relatedness, also share high phenotypic similarities. Values from 30 to 70 % reflect a moderate degree of relationship, while values below 30 % level constitute the background level of some methods. One has, however, to consider that the recommendation for species delineation derives mainly from the experience made with enterobacteria. The taxonomic treatment of rapidly evolving eukaryote-associated organisms and the more slowly evolving archaeal and bacterial representatives to the category »species« based on DNA reassociation similarity >70 % grossly underestimates the different mechanisms as well as the mode and tempo at which organisms develop. It should be remembered that in the absence of a better understanding of microbial evolution this delimitation value is an artificial value used to structure the bacterial world at the level of species in a coherent way. Although DNA : DNA similarity values (0–100 %) have been equated to the 15 % range of DNA matching required to see any DNA hybridization and the 70 % DNA : DNA species level equated to ~96 % DNA identity, we see *Helicobacter pylori* strains with 20 % of their DNA completely different and *E. coli* K12 and *E. coli* 0157 with 20 % difference in genome size, pseudomonad species have genome

sizes which range from 3.7 – 4.7 Mbp in *Pseudomonas stutzeri* or 5.2 – 7.1 Mbp in *Pseudomonas aeruginosa*.

Comparison of 16S rDNA sequence similarity and DNA : DNA similarity has shown that the relation between these parameters is not linear but curvilinear (*10*). As outlined (*16*), each approach is strong in those areas of relationships in which the other methods fail to depict relationships reliably. Sequence analysis has proven to be a reliable marker for the phylogeny of organisms between the levels of domains (around 55–60 % sequence similarity) to moderately related species (around 97 % sequence similarity). Above 97 % sequence similarity, DNA values can be as low as 55 or as high as 100 %. Some organisms that belong to different species because the DNA reassociation values are below 70 % threshold value are known to share 99.8 or even 100 % rDNA similarity. For discriminating highly related organisms, the resolution power of DNA-DNA reassociation is, therefore, significantly higher than that of 16S rDNA sequences. Below 16S rDNA similarity value of 97 % the corresponding DNA-DNA reassociation value has never been shown to be higher than 60 % (*10*), from which follows that below a sequence similarity of 97 % strains do not belong to the same species.

Changing the Species Definition to a Species Concept

The present species definition is probably not as artificial as one would have thought some years ago. This has not only been shown by molecular typing methods, but also by whole genome sequences of different strains of *E. coli* which may differ in more than 25 % of their genome size but are still recognized as strains of the same species on the basis of genome architecture and gene sequences. Apparently, taxonomists have learned in the past 120 years to select phenotypic properties which are not subjected to lateral gene transfer and which are expressed by gene coding key metabolic processes.

Despite the importance of DNA-DNA reassociation in bacterial systematics very few laboratories are in a position to actually execute this method. Moreover, this method is often the slowest and most problematic step in the species description; reproducibility of most variations of the method is not convincing and the hybridization similarity values are highly susceptible to physical and chemical modification. Moreover, the data are not cumulative, requiring the inclusion of reference strains. This assessment has recently been made public in the report of the Ad hoc Committee for the Re-Evaluation of the Species Definition in Bacteriology (*17*). The committee came to the conclusion that the parameter DNA similarity and ΔTm, currently applied in the delineation of genomospecies, cannot be improved.

So, is there an alternative to DNA-DNA reassociation that would avoid all the weak points of the »black box« hybridization techniques? The answer is positive and the confidence is based on the results from population genetics and epidemiological studies. It has been postulated that strains within a taxon should be subject to a higher frequency of lateral gene transfer among orthologous genes than between strains of different taxa

(orthology describes genes in different species that derive from a common ancestor). Lack of recombination between genes from strains of neighboring taxa indicates that both taxa have reached a level of isolation that would merit species status (*3*). This approach has opened a new dimension in the elucidation of genomic relatedness at the inter- and intraspecific level by sequence analyses of housekeeping genes (*19*). This technique has been mainly used to trace the spreading of medical strains in epidemiological studies but it also offers the opportunity to incorporate the insights available from population genetics and phylogenetic approaches into bacterial systematics (Fig. 3). Close relationship between *E. coli* and strains of *Shigella* species has indeed been recognized earlier by DNA-DNA binding studies but sequence analysis reveals a significantly higher level of confidence with respect to genomic affiliation.
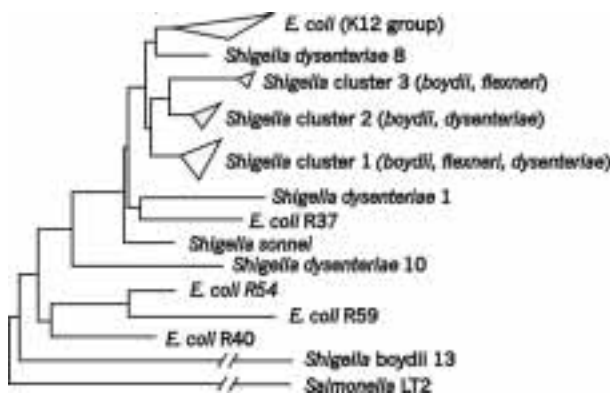

Fig. 3. Combined phylogenetic tree (neighbor-joining) including selected regions of housekeeping genes thrB-thrC, trpC-thrB, purM-purN and mdh-argR. Analysis leads to the recognition of *Shigella* strains as strains of *Escherichia coli* (except for *Shigella bodyii* 13). Modified with permission of Peter R. Reeves (*21*).

It is perhaps surprising that the genomic diversity of strains indeed reveals a »lumpy« character and discrete units of diversity can be observed (*19*), *e.g.* with strains *Escherichia coli*, *Shigella* spp., *Mycobacterium tuberculosis*, *Campylobacter* spp. and *Neisseria meningitis*. These discrete units may or may not correlate with centers determined by DNA-DNA reassociation experiments. Thus, one problem of bacterial systematics remains: while certain strains form a tight center of genomic similarity (the species), there are others that show a larger degree of dissimilarity and the question remains whether or not to affiliate these strains into the species. Other properties will still have to be determined for a strain in order to come to a decision about its taxonomic affiliation.

Where to Go from Here

It can be assumed that the following years will see the application of a strategy that maintains the pragmatic delineation of species in the genomic era whilst integrating new techniques and new knowledge. Taxonomists are encouraged to propose new species based

upon other genomic methods or techniques provided that they can demonstrate that, within the taxa studied, there is a sufficient degree of congruence between the technique used and DNA-DNA reassociation. As little background data are presently available, the new strategy will therefore require an extensive testing period in which genes of well-studied species are subjected to sequence analysis. Only when the strain composition of clusters emerging from the analysis of several genes matches not only each other but also the results of DNA similarity studies, can these genes be considered reliable markers for the classification process.

One important issue relates to the kind of genes to be sequenced. It is obvious that in contrast to the DNA-DNA reassociation studies, in which one molecule, the DNA, serves as a universal marker for determination of relatedness, the restriction to a few selected genes may require to select combinations of different genes for members of different higher taxa. Significant pioneering work is therefore called for in order to select the appropriate genes for the non-clinical species for which very little information on gene sequences (except for fully sequenced genomes) is available. Some information on genes is indeed available but scattered in the literature. Sequences, such as ribosomal *rrn* genes, genes coding ribosomal proteins, the heat shock proteins Hsp60 (GroEL) and Hsp70/DnaK, alanyl-t-RNA, succinyl-CoA synthetase, pyrophosphatase, Lon protease, biotin synthase, DNA gyrase B, UDP-glucose epimerase, PAC-transformylase, and RecA protein have been used to unravel the phylogeny of higher taxa. Genes used for multilocus sequence analysis are, to name a few, the anaerobic DMSO reductase chain A, glutamine synthetase, phosphomannomutase, aspartokinase, thymidylate kinase and anthranylate synthase component I (see also http://campylobacter.mlst.net and http://neisseria.mlst.net). From this follows that additional genes need to be identified for the existing and new species, also including those higher taxa, not presently included in the genome sequencing projects. As the monocistronic genes coding for proteins that are used for multilocus sequence analysis are subject to a higher mutation rate, new primers with wobble nucleotides need to be developed. In order to increase the statistical significance of the emerging clusters, different gene sets need to be developed for optimal recognition of genomic coherence and the rate of transformation. It is obvious that in contrast to abundantly available clinical strains, many environmental species have been described on the basis of a single strain only; as long as more strains are not available the determination of clonal evolution is excluded. Obviously, no recombination events can be determined with low number of strains available; hence the species definition of Dykhuizen and Green (*20*) cannot be applied. However, the availability of sequences from a set of housekeeping genes in a cumulative database provides the basis for subsequent genomic analysis of novel strains.

It is quite clear from what has been said, that the transformation from a species definition to a species concept in bacteriology under the influence of population genetics will be a slow and stony path: genes must be identified, primers developed, phylogenetic and sta-

tistic analyses introduced. The transformation will not be completed tomorrow and changes of the current classification system cannot be anticipated. The Ad hoc Committee reinforces the earlier statement of Wayne *et al.* (*15*) that »new recommendations should be compatible with the current classification« and, for the time being, phenotypic properties will continue to play an important role in the classification process.

References

1. W. B. Whitman, D. C. Coleman, W. J. Wiebe, *Proc. Natl. Acad. Sci. USA, 94* (1998) 6578–6583.

2. V. Torsvik, J. Goksoyr, F. L. Daae, *Appl. Environ. Microbiol. 56* (1990) 782–787.

3. R. Lan, P. R. Reeves, *Trends Microbiol. 9* (2001) 419–424.

4. B. J. Tindall, E. Brambilla, M. Steffen, R. Neumann, R. Pukall, R. M. Kroppenstedt, E. Stackebrandt, *Environ. Microbiol. 2* (2002) 310–318.

5. E. Brambilla, H. Hippe, A. Hagelstein, B. J, Tindall, E. Stackebrandt, *Extremophiles, 5* (2001) 23–33.

6. M. Vaneechoutte, *Mol. Biotech. 6* (1996) 115–143.

7. A. Van Belkum, M. Struelens, A. deVisser, H. Verbrugh, M. Tibayrenc, *Clin. Rev. Microbiol. 14* (2001) 547–560.

8. D. Helm, H. Labischinski, G. Schallehn, D. Naumann, *J. Gen. Microbiol. 137* (1991) 69–79.

9. M. Claydon, S. N. Davey, V. D. Edwards-Jones, V. D. B. Gordon, *Nature Biotech. 14* (1996) 1584–1586.

10. E. Stackebrandt, B. M. Goebel, *Int. J. Syst. Bacteriol. 44* (1994) 846–849.

11. F. A. Rainey, N. Ward-Rainey, P.H. Janssen, H. Hippe, E. Stackebrandt, *Microbiology, 142* (1996) 2087–2095.

12. R. Pukall, O. Päuker, D. Buntefuß, G. Ulrichs, P. Lebaron, L. Bernhard, T. Guindulain, J. Vives-Rego, E. Stackebrandt, *FEMS Microbiol. Ecol. 28* (1998) 335–344.

13. P. L. Manachini, D. Mora, G. Nicastro, C. Parine, E. Stackebrandt, R. Pukall, M. G. Fortina, *Int. J. Syst. Evol. Microbiol. 50* (2000) 1331–1337.

14. C. A. Istock, J. A. Bell, N. Ferguson, N. L. Istock, *J. Ind. Microbiol. 17* (1996) 137–150.

15. L. G. Wayne, D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, L. Krichevsky, L. H. Moore, W. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, H. G. Trüper, *Int. J. Syst. Bacteriol. 37* (1987) 463–464.

16. E. Stackebrandt, *The Prokaryotes*, http://link.springer-ny.com:6335/contents

17. E. Stackebrandt, W. Fredericksen, G. M. Garrity, P. A. D. Grimont, P. Kämpfer, M. C. J. Maiden, X. Nesme, R. Rosselló-Mora, J. Swings, H. G. Trüper, L. Vauterin, A. C. Ward, W.B. Whitman, *Int. J. Syst. Evol. Microbiol. 52* (2002) 1043–1047.

18. M. C. J Maiden, J. A. Bygraves, E. Feil, G. Morelli, J. E. Russel, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, B. G. Spratt, *Proc. Natl. Acad. Sci. USA, 95* (1998) 3140–3145.

19. J. Maynard Smith, N. H. Smith, M. O'Rourke, B. G. Spratt, *Proc. Natl. Acad. Sci. USA, 90* (1993) 4384–4388.

20. D. E. Dykhuizen, L. Green, *J. Bacteriol. 173* (1991) 7257–7268.

21. G. M. Pupo, R. Lan, P.R. Reeves, *Proc. Natl. Acad. Sci. USA, 97* (2000) 10567–10572.

Bogatstvo raznolikosti prokariota

Sažetak

Zbog naglog povećanja broja potencijalnih novih vrsta, povezanih s metodološki složenim polifaznim pristupom pri bakterijskoj sistematizaciji, identifikacija i prema tome opis novih taksona postaje vrlo zahtjevan posao. Broj novih opisa, približno 200 taksona na godinu, vjerojatno se ne može bitno proširiti. Međutim, novim postupcima izolacije te pojačanim interesom za rad s bakterijama, umjesto s DNA klonovima, dobiven je široki raspon feno- i genotipova. Očigledno je da se neke od svjetski priznatih tehnika više neće koristiti, ali je pitanje koji će se postupci koristiti u budućnosti. Zasada, reasocijacija DNA-DNA konačno prosuđuje o »vrsti« taksona, ali su dobro poznata ograničenja ovog postupka podložnog znatnim pogreškama. S druge strane, niz molekularnih tehnika koristi sekvenciju cijeloga genoma, što upućuje na to da su evolucijski procesi uzrokovali razlike u DNA, kao što su točkaste mutacije, insercije i delecije, dobivanje i gubitak DNA, rekombinacija, gubitak gena i stvaranje pseudogena. Svaki od tih evolucijskih procesa utječe na sličnost reasocijacije DNA-DNA koji se, međutim, ne mogu uobičajeno testirati pri reasocijaciji. Nadalje, sparivanjem DNA ne uspostavlja se kumulativna baza podataka jer se parametri hibridizacije mijenjaju od laboratorija do laboratorija, a u svaki je pokus potrebno uključiti referentne organizme. U radu su prikazane ideje koje mijenjaju paradigmu postojećeg umjetnog, arbitrarnog i pragmatičkog koncepta vrste pri definiciji vrste, a kojemu je ključ uvid u populacijsku genetiku.