

Review

# The species concept for prokaryotes

Ramon Rosselló-Mora \*, Rudolf Amann

*Department of Molecular Ecology, Max-Planck-Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany*

Received 4 November 1999; received in revised form 23 August 2000; accepted 24 August 2000

## Abstract

The species concept is a recurrent controversial issue that preoccupies philosophers as well as biologists of all disciplines. Prokaryotic species concept has its own history and results from a series of empirical improvements parallel to the development of the techniques of analysis. Among the microbial taxonomists, there is general agreement that the species concept currently in use is useful, pragmatic and universally applicable within the prokaryotic world. However, this empirically designed concept is not encompassed by any of the, at least, 22 concepts described for eukaryotes. The species could be described as ‘a monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity in many independent characteristics, and is diagnosable by a discriminative phenotypic property’. We suggest to refer it as a phylo-phenetic species concept. Here, we discuss the validity of the concept in use which we believe is more pragmatic in comparison with those concepts described for eukaryotes. © 2001 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

*Keywords:* Species concept; Phenetic analysis; Phylogenetic reconstruction; DNA–DNA similarity

## Contents

1. Introduction . . . . .	40
2. The development of prokaryote taxonomy . . . . .	41
3. Methods and parameters used in prokaryotic species circumscription . . . . .	43
3.1. Retrieving genomic information . . . . .	43
3.2. Phenotypic methods . . . . .	48
4. Species concepts for eukaryotes . . . . .	51
5. The prokaryotic species concept . . . . .	52
5.1. The concept . . . . .	52
5.2. The limits . . . . .	53
5.3. The 16S rRNA sequence data . . . . .	53
5.4. Intraspecific subdivisions . . . . .	54
5.5. Prokaryotic sex . . . . .	56
5.6. Pure cultures and the culture collections . . . . .	56
5.7. The species concept to be used for prokaryotes . . . . .	57
5.8. The phylo-phenetic species concept . . . . .	59
5.9. Guidelines to the recognition of a prokaryotic species . . . . .	60
5.10. Future species circumscriptions . . . . .	60
Acknowledgements . . . . .	61
References . . . . .	61
Glossary . . . . .	65

\* Corresponding author. Present address: Laboratori de Microbiologia, Departament de Biologia, Facultat de Ciències, Universitat de les Illes Balears, Ctra. Valldemossa Km 7, 5, E-07071 Palma de Mallorca, Spain. Tel.: +34 971173140; Fax: +34 971173184; E-mail: rossello-mora@uib.es

## 1. Introduction

*'The true taxonomist is a man with a mission; he often leads a cloistered life, protected from the vexations and frustrations of the everyday world, and he may well wear blinkers as opaque as any worn by a horse.... Living a life of seclusion, safe in his small laboratory, and surrounded by his books, his microscope (and perhaps his computer tape), he affects an unconcern for the mundane application of his work' [1].*

With these words, S.T. Cowan, one of the most avant-gardist and successful taxonomists, described a not very optimistic view of his own collective. However, most of the taxonomists would agree that today's specialist is a much more dynamic scientist adapted to the rush of technical development, actively incorporating new methodologies, open for discussions, and aware that this field of science is slowly being taken more seriously. Microbiology itself is developing and diversifying enormously. One of the 'missions' of the microbial taxonomists is to create a common language for all microbiologists.

It is surprising, particularly for non-microbiologists, that prokaryotes, most of which are invisible to the human eye, constitute an essential component of the Earth's biota. They catalyze unique and indispensable transformations in the biogeochemical cycles of the biosphere, produce important components of the Earth's atmosphere, and represent a large proportion of life's genetic diversity. In contrast to textbook knowledge, the estimated  $4\text{--}6 \times 10^{30}$  prokaryotic cells existing on the Earth (Table 1) might even constitute  $> 50\%$  of the protoplasmic biomass (excluding most of the plant biomass that is made up of extracellular material such as cell walls and structural polymers [2]). It is also interesting to note that the majority of the prokaryotes may be located in oceanic and terrestrial subsurface environments. Less attention than those of the Earth's surface has been paid to these habitats. Thus, research in microbiology might have dealt with a very small portion of the total Earth's prokaryotic community.

Several attempts to estimate the number of species living on Earth have been made, and it seems that the number of recorded species is directly related to the effort and interest of the scientists [3]. However, in spite of the enormous quantitative contribution of prokaryotes to the biosphere,

their diversity and importance has always been underestimated by non-microbiologists (Table 2). One of the reasons is that to date less than 5000 prokaryotic species have been described. This relative low number is caused by the problems encountered for the isolation of microorganisms in pure cultures and their characterization. Problems such as hitherto unculturability, lack of proper research funding and in some cases the underestimation of the isolation efforts are responsible for such numbers. However, the isolation of an organism in pure culture is to date an indispensable requisite for the recognition of prokaryotic species [4]. On the other hand, it is of general knowledge among microbiologists that there is a large potential of prokaryote diversity made up of hitherto uncultured microorganisms [5–8]. Molecular techniques, most notably those based on 16S rRNA, which are directed towards analyzing community composition of environmental samples indicate that the hitherto classified prokaryotic species account for a very small portion of the real prokaryote diversity [9]. Thus, if we only take into account the recognized prokaryotic species for diversity calculations, their total number would never be regarded as a significant proportion of the total Earth's biodiversity. This underestimation of the real prokaryote diversity has obvious negative effects on the distribution of research funding [10,11].

The species concept for prokaryotes has been developed in parallel to the design of laboratory techniques that permitted the retrieval of useful information. The original species concepts based on morphological traits were demonstrated to be wrongly tailored. Improved concepts have developed through the use of new information units (e.g. chemotaxonomic markers, DNA properties, rRNA sequences...). At the present, many, if not all, prokaryote taxonomists agree that the current species circumscription, although not perfect, is acceptable and pragmatic, and covers the primary goals of taxonomy such as a rapid and reliable identification of strains [12]. However, among non-taxonomists, the prokaryotic species concept is criticized. For some, the concept is too conservative, leading to an underestimation of the real prokaryote diversity [2,13]. They consider the conservative nature of the concept to be a significant disadvantage in as much as it is not comparable to the concepts designed for higher eukaryotes [14,15]. The current discussion about the adequacy of the

Table 1  
Number and biomass of prokaryotes in the world (data obtained from [2])

Environment	No. of prokaryotic cells ( $\times 10^{28}$ )	g of C ( $\times 10^{15}$ )
Aquatic habitats	12	2.2
Oceanic subsurface	355	303
Soil	26	26
Terrestrial subsurface	25–250	22–215
Total	415–640	353–546
Plants <sup>a</sup>	–	562

<sup>a</sup>Plant carbon is the sum of protoplasmic biomass, structural polymers and cell wall material.

Table 2  
Estimates on the contribution of major groups of organisms to the total biological diversity [16]

Organism	Percent of contribution
Vertebrates	0.4
Insects	64.4
Other arthropods	7.7
Other invertebrates	6.7
Plants	2.4
Algae	1.6
Protozoa	1.6
Fungi	8.0
Bacteria	3.2
Viruses	4.0

species concept is not restricted to microbiologists, but is also being discussed among eukaryote taxonomists [16]. It is indeed tempting to introduce a more universal concept that covers all major groups of organisms making the species units comparable. In this respect, it should be helpful to analyze in more detail the prokaryote species concept, its history and potential future. This is the aim of this study.

## 2. The development of prokaryote taxonomy

Prokaryote classification is the youngest and most dynamic among the different classifications of living organisms. Prokaryotes were not even known to exist until a few centuries ago, due to their small size and the fact that they cannot normally be seen with the naked eye. In addition, the development of a reliable classification based on morphological traits as these for higher eukaryotes has been difficult because of the relative simplicity of the prokaryotes<sup>1</sup>. The lack of a useful fossil record, together with the difficulties in identifying diagnostic characteristics from these small organisms have contributed to the instability of the prokaryote classification system. The development of new techniques directed to the understanding of useful phenotypic and genomic traits of microorganisms was the rate-determining step towards a reliable taxonomic scheme for prokaryotes.

Most microscopists of the 17th and 18th centuries described meticulously the 'infusion animalcules' which they observed, but no classification was attempted. Initially, prokaryotes were treated as only a single species which could develop a great variety of shapes (pleomorphism). The earliest attempts to create microbial classifications were solely based on morphological observations. At the end of the 18th century, Otto Müller was the first to attempt

a systematic arrangement of microorganisms [18]. He created two form genera, *Monas* and *Vibrio*, which encompassed bacteria and differentiated the punctiform and the elongated types. In the early 19th century, Christian Ehrenberg extended Müller's nomenclature and added the helical bacteria. Some of these species designations are still in use (e.g. *Spirochaeta plicatilis* and *Spirillum volutans*). Subsequent workers devised simpler classifications, although still based upon microscopic morphology. They assumed constancy of form at a time when theories of spontaneous generation and pleomorphism were still widely held. In the 1870s, Ferdinand Cohn still supported the idea that bacterial forms were constant irrespective of environmental conditions, but he already recognized the existence of a wide diversity of bacteria. He recognized the similarity between the cyanobacteria (schizophyceae, 'fission algae') and bacteria (schizomycetes, 'fission fungi') and combined them as schizophytae ('fission plants') [19,20]. He arranged bacteria into six form genera, but appreciated that physiologies, end products and pathogenesis of similar-shaped organisms might differ. Robert Koch in 1876 proved the truth of the germ theory of disease previously postulated by Louis Pasteur with his studies on *Bacillus anthracis* [19,20], and later concluded that the different morphologies of pathogenic bacteria must be regarded to belong to a single distinct and constant species [18].

One of the most important steps in the development of microbiology was the ability to isolate organisms in pure cultures. In 1872, Cohn's co-worker Joseph Schroeter cultivated pure colonies of chromogenic bacteria, and in 1878 Joseph Lister obtained a pure culture of a milk-souring organism by dilution [18]. In 1881, Koch published the technique of cultivation on solidified gelatin media, which was subsequently replaced by agar, and this was the start of what he called 'the golden age of the medical microbiology' [19]. By cultivating microorganisms in pure culture, researchers were able to retrieve direct information on the organisms. Many tests for distinguishing bacteria were developed, and these formed the basis for their classification [18]. They permitted the phenotypic description of these organisms.

The amount of bacteria that were described at the end of the 19th century and the first two decades of 20th century is impressive. K.B. Lehman and R. Neumann published in 1896 their 'Atlas und Grundriss der Bakterien' in where several new genera were described (*Corynebacterium*, *Mycobacterium*, *Actinomyces*). And W. Migula in 1897 compiled in his 'Das System der Bakterien' all bacteria that had been previously described [19]. Physiological characters took then a predominant role in bacterial classifications. Researchers like S.N. Winogradsky and M.W. Beijerinck published a good number of new genera which names described the ecology, physiology and biochemistry of the organisms bearing them [19]. S. Orla-Jensen, in his publication 'Die Hauptlinien des natürlichen Bakteriensystems' ('the mainlines of the natural bacterial system',

<sup>1</sup> Indeed, morphology itself does not provide enough information for the development of an operative and predictive classification system. However, once genealogy is known, bacterial morphologies are consistent with their phylogenetic reconstructions [17].

1909), attempted to construct a classification system based on genealogical relationships, in where he presented the lithoautotrophic bacteria as the most primitive [20]. Later on, several classifications based on morphological characters were published and intended again to reconstruct the 'natural system': Pringsheim (1923), Buchanan (1925, the beginnings of Bergey's Manual), Kluver and van Niel (1936), latter emended by Stanier and van Niel (1941), and Prévot (1940) [21]. It is, however, in 1923 when the 'Bergey's Manual of Determinative Bacteriology' was published, when a modern identification key for bacteria was first provided. At the time there was no common agreement on prokaryotic classification [22], this manual and the subsequent editions became the reference works on bacterial classification [18]. These publications provided a framework for the unification of criteria among microbiologists and avoided nomenclatural problems that occurred relatively frequently in the early years of bacterial classification. At this time, the lack of scientific exchange among microbiologists, together with a strong tendency towards special purpose classification (artificial schemes in which one or a few properties of the organism are given undue prominence; [23]) were responsible for a confusing nomenclature for quite a large number of distinct bacteria. A good example is *Pseudomonas stutzeri*, which during the first half of this century appeared in the literature under at least seven different names, often simultaneously [24].

As the number of different methods for characterizing bacteria increased, bacterial taxonomists suffered more and more from the lack of an objective overview of these approaches to classification. In the late 1950s, numerical taxonomy was developed in parallel to the onset of the computer age as a part of multivariate analyses. Its aim was to devise a consistent set of methods for classification of organisms. Much of the impetus for the development of numerical taxonomy in bacteriology came from the problem of handling the large tables of data on physiological, biochemical and other properties of numerous strains. There was thus a need for an objective method of taxonomic analyses, aimed at sorting individual strains of bacteria into homogeneous groups, conventionally species, and the arrangement of species into genera and higher groupings [25]. The period of numerical taxonomy coincided with the rise of chemotaxonomy and the application of modern biochemical analytical techniques, principally chromatographic and electrophoretic separation methods, to the study of distributions of specific chemical constituents such as amino acids, proteins, sugars and lipids in bacteria [18].

During the early 1960s, the increasing knowledge of the properties of DNA and the development of molecular biological techniques supported the idea that bacteria might best be classified by comparing their genomes. Initially, overall base compositions of DNAs (mol% G+C values) were used. Bacteria whose mol% G+C values differed markedly were obviously not of the same species. How-

ever, single values obtained by the analysis of DNA base compositions allowed only very superficial comparisons, and a much more precise method was needed. Thus, DNA-DNA hybridization techniques were developed [26]. A great practical advantage of this method was that it often produced sharply defined clusters of strains than those solely circumscribed by phenotypic traits [27]. Organisms tended to be either closely related or not. DNA-DNA hybridization consequently became a standard technique for the circumscription of bacterial species. However, as these experiments were used increasingly in bacterial classification, some microbiologists became worried that the data might merely be sets of figures with little practical value. If these data are obtained as an end in themselves, this certainly would be true [27]. Their practicality depends on subsequently determining phenotypic characters that can be used to describe a DNA similarity group, and also on determining which phenotypic characters can be used to identify new isolates easily, rapidly and reliably [28]. Indeed, the Committee on Reconciliation of Approaches to Bacterial Systematics [4] recommended that a bacterial species classification must provide diagnostic phenotypic properties.

In the late 1970s, a remarkable breakthrough in the attempts to determine relationships between distantly related bacteria was achieved by cataloging ribosomal ribonucleic acids (rRNA) [29] and DNA-RNA hybridization [30], and in the mid 1980s by the full sequence analysis of rRNA. The rRNA sequences were shown to be a very useful molecular marker for phylogenetic analyses [31]. Among the three rRNA molecules, 16S rRNA has been

Table 3  
Differences in the description of new species between the years 1989 and 1999

Classifications with	1989 ( <i>n</i> = 44) <sup>a</sup>	1999 ( <i>n</i> = 156) <sup>a</sup>
1 strain	25 (11)	61 (95)
2 strains	7 (3)	14 (22)
3–5 strains	16 (16)	9 (14)
≥6 strains	52 (23)	16 (25)
16S rRNA data	2 (1)	100 (156)
DNA–DNA similarities	75 (33)	56 (87)
Chemotaxonomy <sup>b</sup>	30 (13)	51 (79)
SDS–PAGE <sup>c</sup>	5 (2)	21 (33)
Commercial tests <sup>d</sup>	27 (12)	26 (41)
Relation to health <sup>e</sup>	30 (13)	29 (45)

In 10 years of difference, there is a significant increase of new classifications based on one or two strains. 16S rRNA data are provided in all current classifications. Values are expressed as percentages.

<sup>a</sup>Between brackets are indicated the absolute numbers of new classified species.

<sup>b</sup>One or more chemotaxonomical markers (cell wall, polyamines, fatty acids quinones...) are provided.

<sup>c</sup>Whole cell protein electrophoresis is used to discriminate the different isolates.

<sup>d</sup>Metabolic characters have been analyzed by the use of some commercially available tests (e.g. API or Biolog).

<sup>e</sup>The strains in these studies have been isolated from samples related to human, animal or plant health.

the most widely studied. Thus most information is available for this molecule [32]. The era of the 16S rRNA sequencing brought new information such as for the definitive recognition of *Archaeobacteria* or *Archaea* as an independent cellular lineage [33], and important rearrangements of the prokaryotic classification scheme [34]. rRNA sequencing has become routine in most of microbiology laboratories, and although not regarded to be essential for new classifications, this information is provided in the description of most of the newly classified bacterial species. It is also becoming increasingly popular to propose new bacterial species using data generated from 16S rRNA sequencing studies (Table 3) [12]. Unfortunately, the resolving power of the 16S rRNA is insufficient to guarantee correct delineation of bacterial species [35,36]. Furthermore, its validity as a marker for phylogenetic inferences is being questioned [37–39].

Nevertheless, at the present, the vast majority of bacterial taxonomists accept that 16S rRNA sequence analysis provides a stable and quite satisfactory framework for prokaryotic classification. However, it is also widely accepted that an adequate classification of prokaryotes, in particular of the lower taxonomic ranks such as species, will be achieved when a ‘polyphasic approach’ is undertaken [40], thus combining as many different techniques as possible. This would include a fine-tuning of the circumscription that takes into account the differing properties of the different bacterial groups [41].

### 3. Methods and parameters used in prokaryotic species circumscription

Today, prokaryote taxonomists agree that a reliable classification can only be achieved by the exploration of the internal diversity of taxa by a wide range of techniques in what is generally known as the ‘polyphasic approach’ [40]. This approach implies that two sources of information must be investigated as extensively as possible: genomic information and phenotype. Genomic information is gained from all data that can be retrieved from nucleic acids, either directly through sequencing or indirectly through parameters like DNA–DNA similarity or G+C mol%. Phenotype refers to the way in which the genotype is expressed, the visible or otherwise measurable physical and biochemical characteristics of an organism, a result of the interaction of genotype and environment. There is a tendency among microbiologists to use the term genotypic as a synonym for genomic information. However, genotype is the genetic information, the genetic constitution of an organism, which acts together with environmental factors to determine phenotype. For example, the information harbored (genotype) in the nucleotide sequence (genomic information) coding of an enzyme (phenotype). Genotype can therefore not include any parameter that can be retrieved from the genome. Thus, when referring

to large amounts of information in the genome, or derived from it, it is better to use the term genomic instead of genotypic [42].

For taxonomic purposes, both sources of information, genomic and phenotype, need to be investigated and the lack of either can result in the rejection of the proposed classification. Such a combined description is essential for the delineation of new species in prokaryotes [4]. Today, the accepted species classification can only be achieved by the recognition of genomic distances and limits between the closest classified taxons (DNA–DNA similarity), and of those phenotypic traits that are exclusive and serve as diagnostic of the taxon (phenotypic property).

Prokaryote systematics has undergone spectacular changes in recent years by taking full advantage of developments in chemistry, molecular biology and computer science to improve the understanding of the relationships between microorganisms and the underlying genetic mechanisms on which they are based [23]. A relatively large set of techniques are being used routinely for prokaryote classification. However, it is of primary importance to understand at which level these methods carry information. The kind of information that each technique retrieves is directly related to its resolving power, and the correct use of this information is essential to guarantee the adequate classification of a taxon. An extensive review on the application of different techniques to prokaryote taxonomy and their resolving power has been published by Vandamme et al. [40]. In this chapter, we will give an overview of techniques that are commonly used for species circumscription.

#### 3.1. Retrieving genomic information

The methods of genomic information retrieval are mostly directed toward DNA or RNA molecules. Undoubtedly, these methods presently dominate modern taxonomic studies not only as a consequence of technological progress, but because of the present view that classification should reflect genotypic relationships as encoded in their DNA [40]. Unlike other cell constituents used for chemotaxonomy (see Section 3.1.1), only the amounts and not the composition of RNA and chromosomal DNA are affected by growth conditions, and are thus independent of environmental changes. Furthermore, nucleic acids are universally distributed and these are excellent tools to be used as standards for wide-ranging comparisons.

The most complete genomic source of information is of course the entire bacterial genome. As large-scale sequencing of complete genomes is not feasible at present, several alternative approaches have been taken. They include estimating the mean overall base composition of DNA, comparing genomic similarities by DNA–DNA pairing studies, generating unique sets of DNA fragments by digestion with restriction endonucleases (low-frequency restriction fragment analysis (LFRFA), pulsed field gel elec-

trophoresis (PFGE), RFLP...), sequence comparisons of selected genes, DNA–rRNA hybridization and sequencing of rRNA.

### 3.1.1. DNA base ratio (mol% G+C; G+C content; G+C%)

The primary structure of DNA results from the linear succession of the four nucleotide bases adenine (A), thymine (T), guanine (G) and cytosine (C), and this succession determines the genetic information of an organism's genome. Because of the double-stranded nature of DNA, where both strands are complementary with base pairing G-C and A-T, the ratios G/C and A/T usually remain constant at 1. However, the relative ratio  $[G+C]/[A+T]$  varies from genome to genome. The base ratio of a DNA molecule is generally described as the relative abundance of the pair G+C, and is commonly called G+C content. The DNA base ratio is calculated in percentage of G+C:  $[G+C]/[A+T+C+G] \times 100$ . This was the first nucleic acid technology applied to prokaryote systematics [43], and initially proved to be a useful and routine way of distinguishing between phenotypically similar and genomically different strains [23]. It is usually one of the genomic characteristics recommended for the descriptions of species and genera.

Among the prokaryotes, G+C contents vary between 20 and 80 mol% [44]. The greater the difference between two organisms, the less closely related they are. Theoretically, DNA molecules with differences of greater than 20–30 mol% can have virtually no sequences in common [18]. Empirically, it has been shown that organisms that differ by more than 10 mol% do not belong to the same genus and that 5 mol% is the common range found within a species. While firm guidelines have yet to be set for the range of variation, values higher than 15 mol% can be taken as a strong indication for heterogeneity within a genus [41]. However, it should be noted that although differences in mol% are taxonomically useful for separating groups, similarities in base compositions do not necessarily indicate close relationships because the determinations do not take the linear sequences of bases in the DNA molecules into account (the criterion can only be used negatively).

### 3.1.2. DNA–DNA similarity (DNA–DNA pairing; DNA–DNA homology; DNA–DNA relatedness)

The determination of whole genome DNA–DNA similarity is today still the standard technique for species delineation. The rationale for using this parameter to set the borders of the species circumscription originates from the results of numerous studies, in which a high degree of correlation was found between genomic DNA similarity and phenotypic similarity (i.e. chemotaxonomic, serological... [12]).

A characteristic property of DNA and RNA is its ability for reassociation or hybridization. The complementary strands of DNA, once denatured, can, under appropriate experimental conditions, reassociate to reform native duplex structures. The specific pairings are between the base pairs A-T and G-C, and the overall pairing of the nucleic acid fragments is dependent upon similar linear arrangements of these bases along the DNA. Under standardized conditions, DNAs from different organisms reassociate depending on the similarity of their nucleotide sequences, thereby allowing quantification of the degree of relatedness, usually expressed as % similarity or homology. It is important to note that the term homology has been replaced by the term similarity due to the inaccuracy of its use [12]. There is no linear correlation between actual sequence identities and hybridization values; the latter gives only relative similarity values between genomes. There are several methodologies to measure DNA–DNA relatedness, but all of them rely on the same principle (Fig. 1). DNAs of two different organisms are mixed and denatured to give a solution of a mixture of single-stranded DNA molecules. Under controlled experimental conditions, DNA reassociation occurs and results in hybrid molecules: the higher the genetic similarity of the two organisms, the more nucleotide base sequences they have in common, and the more hybrid formation (hybridization) will occur. The comparison between the results obtained with the mixture of DNAs and pure reference DNA (homoduplex DNA) yields a degree of similarity.

There are two main parameters that are used to measure the degree of relatedness: the relative binding ratio (RBR) and the difference in thermal denaturation midpoint ( $\Delta T_m$ ). Although both parameters result from the determi-

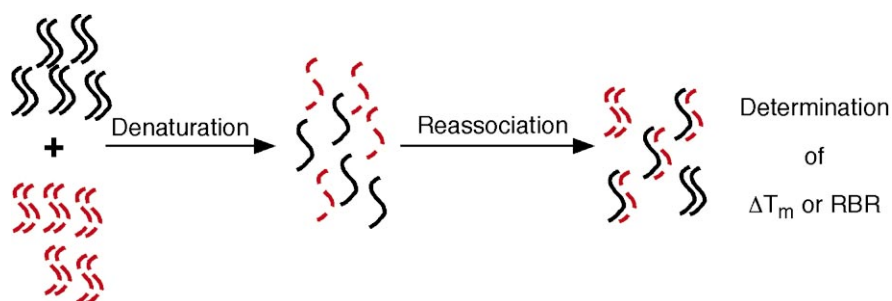


Fig. 1. DNA–DNA reassociation assay.

nation of different features, they are correlated and can be independently used for the species circumscription [45]. In a mixture of two different DNAs, after the hybridization procedure, the amount of double-stranded units is dependent on the degree of identity between both DNAs. In this case, double-stranded DNA occurs to a lesser extent than in the homoduplex mixture. RBR reflects the relative amount of heterologous in comparison to the homoduplex DNA, which is considered to represent 100% reassociation.

$\Delta T_m$  is a reflection of the thermal stability of the DNA duplexes. Double-stranded DNA denaturation is mainly dependent on three factors: the G+C mol%, the ionic strength of the solution in which the DNA is dissolved, and the temperature. G+C mol% is a constant parameter characteristic of each DNA, and the ionic strength of the hybridization solution is normally kept constant. Thus, the single variable parameter is the temperature. In a denaturation kinetic curve of a double-stranded DNA, the temperature at which 50% of the DNA strands are already denatured is called the melting temperature or the thermal denaturation midpoint ( $T_m$ ). The heteroduplex DNAs account for a lower number of paired bases than those from the homoduplex, i.e. less hydrogen bonds are formed. Thus, the duplexes are less stable and in denaturation kinetics the  $T_m$  is reached at a lower temperature (Fig. 2). In this case, the parameter used for measuring the DNA–DNA relatedness,  $\Delta T_m$ , is the difference between the homoduplex DNA  $T_m$  and the heteroduplex DNA  $T_m$ .

There is no direct transformation of RBR into  $\Delta T_m$  or viceversa. However, results of both analyses are well correlated (Fig. 3). The advantage of  $\Delta T_m$  over RBR is that the former parameter is independent of the method used

for hybridization and the data do not have to be transformed. In contrast, RBR results are subjected to differences that are related to the hybridization technique used, and this has to be taken into account when comparing results obtained with different hybridization protocols [46].

Based on numerous studies with well defined prokaryotic species, it is currently recommended that values of 70% or higher RBR, and 5°C or lower of  $\Delta T_m$ , are reasonable borders for the species circumscription [4]. However, it is important to realize that DNA similarity values do not reflect the actual degree of sequence similarity at the level of the primary structure. Indeed, it has been estimated that prokaryotic DNA heteroduplexes will not be formed, even under non-stringent conditions, unless the DNA strands show at least 80% sequence complementarity. Thus, depending on the sequence similarity of the reassociating single strands, a difference of about 20% of sequence identity is spread between 0% and 100% DNA–DNA similarity [23]. Additionally, DNA–DNA similarity studies are time-consuming and they are hampered by the fact that they rely on pair-wise comparisons, which means that experiments are normally carried out with a relatively small set of organisms [47]. Therefore, unless the reference strains chosen are representative of the constituent species, incorrect conclusions can easily be drawn [48].

Despite these problems, the advantages of DNA–DNA similarity analyses outweigh their limitations, and it is an attractive measure as it can be applied to all cultivable prokaryotes irrespective of their growth requirements. These analyses provide a unified measure for the delineation of bacterial species, and among other properties, can also be used to detect and identify unknown isolates [41].

### 3.1.3. rRNA analysis

Over the last 25 years, techniques involving the analysis of rRNA or of the genes coding rRNA (rDNA) have revolutionized prokaryotic taxonomy. The conclusions drawn from these studies are based on the assumption that the rRNA genes are highly conserved because of the fundamental role of the ribosome in protein synthesis. rRNAs are molecules with universal, constant and highly constrained functions that were established at an early stage in evolution and that are not affected by changes in the organism's environment. Therefore, and because they are large molecules containing considerable genetic information, they have been chosen as the molecular basis for phylogenetic reconstructions at least in the prokaryotic world [21]. Two additional assumptions are basic for the validity of this approach, namely that lateral gene transfer has not occurred between rRNA genes, and that the amount of evolution or dissimilarity between rRNA sequences of a given pair of organisms is representative of the variation shown by the corresponding genomes [41]. If this holds true, the variations in the rRNA primary structures among the prokaryotes will reflect evolutionary distances among organisms.

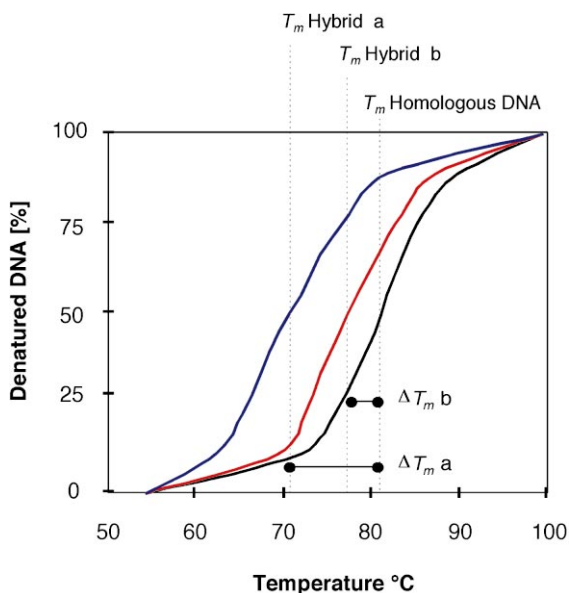


Fig. 2. Thermal denaturation curves of a homoduplex DNA and two heteroduplex DNAs.

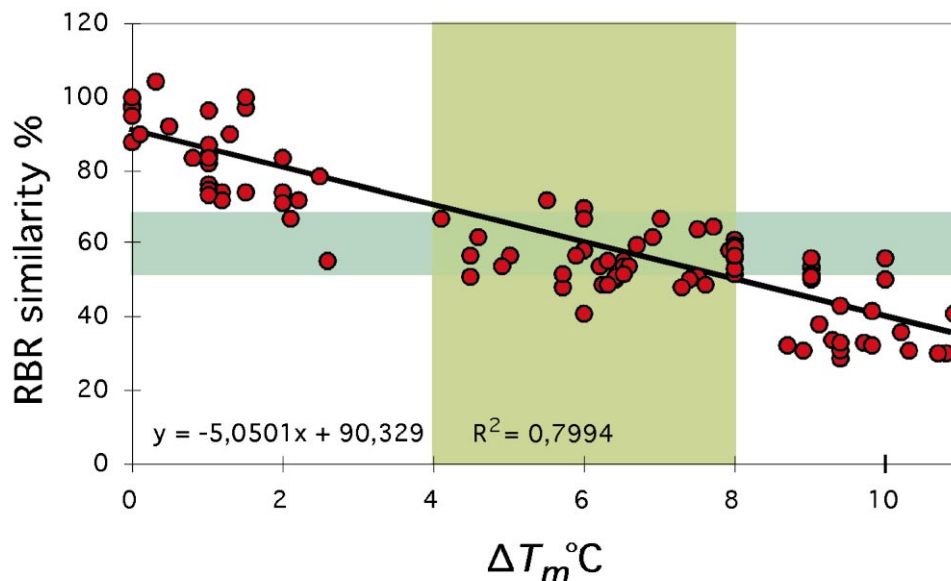


Fig. 3. Correlation between RBR and  $\Delta T_m$  values. Generally, values of  $\Delta T_m$  below 4–5°C are correlated with RBR values above 50%; those above 8–9°C with RBR values below 50%. Commonly accepted values for species boundaries are indicated in green. This plot is obtained by combining 111 datasets available in vol. 49 of IJSB (1999).

The three rRNAs are classified by their sedimentation rates during ultracentrifugation as 23S, 16S and 5S. They have chain lengths of about 3300, 1650 and 120 nucleotides, respectively. Until recently, direct and complete sequencing of the larger rRNA molecules was not feasible as a routine approach. Instead, sequence data were analyzed indirectly by DNA–rRNA hybridization or partial sequences were obtained by oligonucleotide cataloging [29,30]. Initially, for prokaryotes, complete sequencing of 5S rRNAs was used for phylogenetic inferences. However, this technique waned in favor of 16S rRNA sequencing for various reasons, including the greater information content of the larger molecule. Nowadays, almost complete sequencing of the latter is routine. The 23S rRNA molecule is a larger information unit than the 16S, and in many cases has higher resolving power for phylogenetic reconstructions [49,50]. However, due to its length, its sequencing has not been as popular as 16S and the number of 23S rRNA sequences in the databases is much smaller.

Meanwhile, the 16S rRNA approach is one of the most widely used standard techniques in microbial taxonomy. Consequently, a comprehensive sequence dataset (around 18 000 entries in 1999) is available in widely accessible databases. The phylogenetic reconstruction with these data provides a basis for an ongoing evaluation and restructuring of the current bacterial systematics accompanied by emendation, reclassification and hence renaming of bacterial taxa. It is also widely accepted to apply the rRNA technology as an integrated part of a ‘polyphasic approach’ for new descriptions of bacterial species or higher taxa [49,51]. The congruence of 16S rRNA-based reconstructions of phylogenetic trees with those based on alternative molecules, such as 23S rRNA, ATPase sub-

units, elongation factors and RNA polymerases, has been tested and resulted in very similar tree topologies [49,50].

An important feature of the 16S rRNA molecule in its use as a universal standard parameter for phylogenetic inferences is the relative ease of sequence alignment [52]. Alignment is the first critical step of sequence-based phylogenetic analyses. Given that positions with a common ancestry have to be compared for reliable phylogenetic conclusions, homologous positions have to be arranged in common columns in correct alignment [49]. All rRNA molecules share a common secondary and higher-order structure. Many of these structural elements are identical or similar with respect to their position within the molecule as well as number and position of paired bases, or internal and terminal loops, while the primary structures differ [49]. The conservation of helical elements is maintained independently of primary structure conservation by compensating base changes at positions involved in base pairing. Therefore, checking a primary structure for its potential for higher-structure formation usually helps to improve the primary structure alignment. However, there are always variable regions which cannot be unambiguously aligned and it is the subjective decision of the researcher how to arrange the data [49]. In the case of rRNAs, an additional fact facilitates the arrangement of data. There are comprehensive databases of aligned sequences accessible to the public. These alignments have been established and are maintained by specialists. The databases contain secondary structure information, and can be used as a guide to inserting new sequence data [49].

What is the best method for inferring phylogenetic rela-



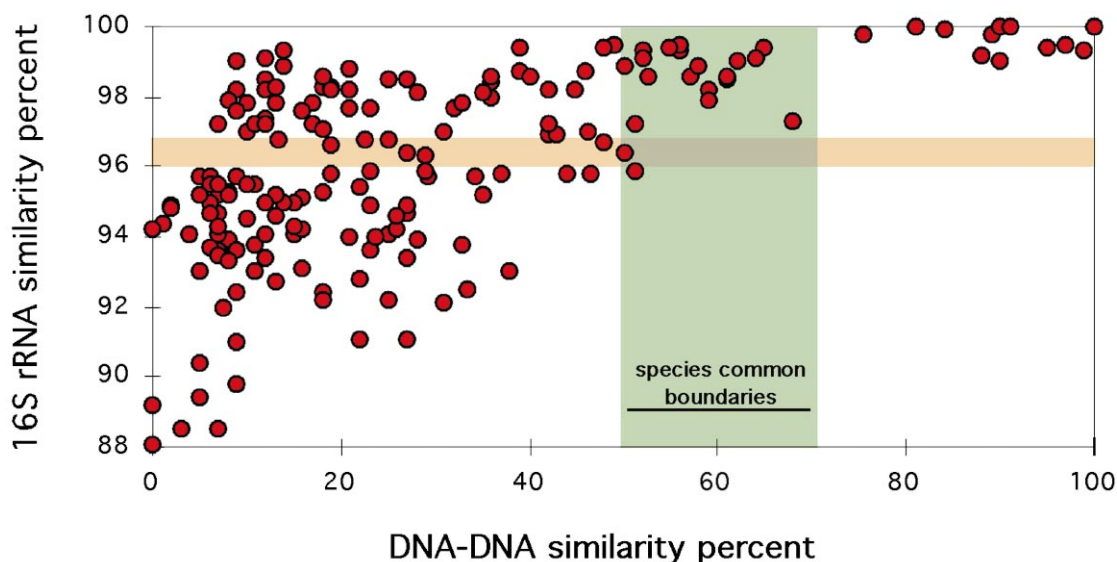


Fig. 4. Comparison of DNA–DNA and 16S rRNA similarities. The dataset is based on 180 values from 27 independent articles of the IJSB vol. 49 (1999). These data combine intrageneric values obtained for members of *Proteobacteria*, *Cytophaga-Flavobacterium-Bacteroides* and Gram positives of high GC phyla.

tionships from sequence data? Computer simulations and experiments have revealed that all methods fail when the assumptions upon which they are based are badly violated [52]. Answers to such questions, pitfalls of the methods, and recommendations for novel researchers have been addressed [49–51]. There are three major approaches for tree reconstruction: distance matrix, maximum parsimony and maximum likelihood methods. These approaches are based on models of evolution<sup>2</sup> and, in general, operate by selecting trees which maximize the congruency of tree topology and the measured data under the criteria of the model [49]. Given that the treeing methods are based on different models and treat the data in different ways, a perfect match of the tree topologies cannot be expected. It has to be taken into consideration that the models only partially reflect reality. Thus for a final reconstruction of a tree, the use of different methods together with calculations based on several data subsets is strongly recommended. In many cases, the separation of clusters or subtrees is stable whereas their relative branching orders differ with the applicable alternative treeing methods. In such cases, a fairly acceptable compromise is to use a consensus tree which shows detailed branching patterns where stable topologies emerged, and multifurcations where inconsistencies or uncertainties could not be resolved [49].

16S rRNA sequencing and comparison analyses have demonstrated high resolving power for measuring the degree of relatedness between organisms above the species level. However, as more sequence information becomes available, it is evident that the resolving power of 16S

rRNA sequences is limited when closely related organisms are being inspected (Fig. 4) [12,53]. Thus, as discussed in a later chapter, there can be no bacterial species definition based solely on sequence similarity of rRNAs or their genes. Absolute values for delineating species cannot be set because of the low resolving power of 16S rRNA at this level [41]. However, the rRNA sequencing approach has additional advantages e.g. for the exploration of uncultured prokaryote species diversity [52], and design of rRNA-targeted probes for a cultivation-independent monitoring of microbial communities [54].

#### 3.1.4. DNA-based typing methods (DNA fingerprinting)

DNA-based typing methods generally allow the detection of intraspecific diversity, e.g. the subdivision within species into a number of distinct types. They are an additional support to the phenotypic analyses trying to reveal diversity of close relative organisms [40]. One can differentiate between two basic techniques: (i) the first-generation typing methods, based on whole genome restriction fragment analysis, (ii) and the polymerase chain reaction (PCR) methods, based on the amplification of genome fragments. All these techniques rely on the electrophoretic separation and subsequent visualization of DNA fragments.

With first-generation DNA-based methods, fragments of the whole genome are generated by using restriction enzymes. Common restriction enzymes recognize specific combinations of 4–6 bases. Due to the size of the bacterial genome (between 0.6 and 9.5 megabases; [55]), the digestion with common enzymes results in a complex mixture of DNA fragments of different sizes that, in most cases, is difficult to analyze. However, the number of DNA fragments can be reduced by selecting restriction enzymes which only rarely cut DNA, recognizing a specific combi-

<sup>2</sup> Although it is true that the distance matrix is generated after phenetic analyses, the tree reconstructions are based on algorithms (Jukes-Cantor, Kimura, De Soete...) which are based on models of evolution.

nation of 6–8 bases. The technique is referred to as LFRFA. The fragments, however, are too large to be separated by conventional agarose gel electrophoresis, and therefore PFGE has to be used [40,56,57]. The result is an electrophoretic pattern for each strain. The comparison of patterns by numerical analysis leads to the establishment of similarity groups within the species. This technique, combined with Southern blot hybridization, yields data on the genome size and organization, that in the future could be of relevance for a comprehensive description of any organism [58]. The number and distribution of rRNA operons within the genome may be a simple, but taxonomically useful feature of the genome that can be revealed by this technique [59].

Alternatively, the complex DNA patterns generated by regular restriction digestion with common enzymes can be transferred to a membrane and then hybridized with a labeled probe, which reveals the hybridized fragments. A typical example of one of these developments is the ribotyping method, which uses rRNA as a probe [60].

The introduction of the PCR methodology into the microbiology laboratory has opened a vast array of applications [40]. Among others, a battery of different typing methods was developed. Different methods in which short arbitrary sequences were used as primers in the PCR assay were described: oligonucleotides of about 20 bases are used in arbitrary primed PCR ([61]); oligonucleotides of about 10 bases are used in randomly amplified polymorphic DNA analysis (RAPD; [62]). These are only few of the many examples of the application of PCR to the typing of bacteria. However, they are only applicable to the understanding of intraspecific diversity. These techniques are not suitable for the circumscription of prokaryotic species as well as for higher taxonomic units, and therefore their use in prokaryotic taxonomy is rather limited.

### 3.2. Phenotypic methods

The phenotype is the observable expression of the genotype. Before molecular techniques were available to prokaryote taxonomists, the classification was exclusively based on morphology, physiology and growth conditions of the organisms. These investigations were directly linked to the use of pure cultures, and the laboratory capabilities to cultivate the organisms and analyze their properties. Thus, the classification schemes were biased towards aerobic heterotrophic microorganisms for which an extensive retrieval of information was easy. One of the disadvantages of analyzing the phenotype is that the whole information potential of a prokaryotic genome is never expressed. Gene expression is directly related to the environmental conditions (e.g. growth conditions in the laboratory). Prokaryotic phenotype cannot be based on the simple observation of the organism. The prokaryotes lack complex morphological features. Most do not show life cycles with different morphological stages, and lack an

ontogenetic development. Thus, the analysis of the prokaryotic phenotype mostly relies on the development of experimental techniques that test directly or indirectly different phenotypic properties, e.g. enzyme activities, substrate utilization profiles and growth conditions.

Phenotypic data, in contrast to gene sequences, can mostly be compared phenetically, that means through the comparison of a large set of independent covarying characters (see [42,63]). These comparisons produce results that reflect the degree of similarity of the units under analysis. It is difficult to perform cladistic analyses (those that produce phylogenetic reconstructions, see above) based on phenotypic data. This is because one of the basic problems is that unless we know the genes that are responsible for the phenotypic traits that we observe, we cannot distinguish between autapomorphic characters (homologous characters thought to have originated in the most recent common ancestor of the taxon, and exclusive for it) and synapomorphic characters (homologous characters common to two or more taxa and believed to have originated in their most recent common ancestor). Prokaryote taxonomy has traditionally disregarded this problem. However, due to the development of molecular techniques, and the establishment of a reliable phylogenetic scheme, taxonomists are beginning to consider the autapomorphic or synapomorphic nature of the investigated phenotypic traits. Soon, complete genome sequencing will shed light on the homology of phenotypic traits.

Several problems must be considered when planning a phenotypic study of bacteria. Two of them, concerning analyses based upon different character sets might show poor agreement (congruence), and that, as mentioned above, the phenotype represents a very small part of each organism's genome [18]. These problems are addressed by basing the classification on a large number of characters from a wide phenotypic range. Other problems are concerned with analysis of data. Many methods exist for the calculation of similarities and for the arrangement of strains according to these similarities, and the application of different combinations of methods to a set of data can lead to a wide variety of interpretations. Fortunately, bacteriologists tend to restrict themselves to relatively few methods (i.e. those available in the computer programs) [18].

Phenotypic analysis is the most tedious task in the classification of microorganisms. It requires much time and skill, and the techniques should be standardized to avoid subjective observations. An important aspect when analyzing the phenotype of a prokaryotic species is that the strains should be chosen (when this is possible) to represent the known diversity and environmental niches of the group being studied. For this purpose, it is most important to use both, recent isolates and culture collection strains (type and reference strains). However, in classifying prokaryotes, it is desirable to use an orderly approach based on common sense, and to use tests that are pertinent [64].

### 3.2.1. Classical phenotypic analyses

The classical or traditional phenotypic tests are used in identification schemes in the majority of microbiology laboratories. They constitute the basis for the formal description of taxa, from species and subspecies up to genus and family. While genomic data alone are sufficient to allocate taxa in a phylogenetic tree and very helpful in drawing the major borderlines in classification systems, the consistency of phenotypic and genomic characters is required to generate useful classification systems [40]. The classical phenotypic characteristics of bacteria comprise morphological, physiological and biochemical features. Individually, many of these characteristics have been shown to be insufficient as parameters for genetic relatedness, yet as a whole, they provide descriptive information enabling us to recognize taxa. The morphology of a bacterium includes both cellular (shape, endospore, flagella, inclusion bodies, Gram staining...) and colonial characters (color, dimensions, form...). The physiological and biochemical features include data on growth at different temperatures, pH values, salt concentrations, or atmospheric conditions (e.g. aerobic/anaerobic), growth in the presence of various substances such as antimicrobial agents, and data on the presence or activity of various enzymes, metabolization of compounds, etc. Reproducibility of results within and between laboratories is a major problem which can be addressed with highly standardized procedures [40].

### 3.2.2. Numerical taxonomy applied to phenotypic analyses (phenetic evaluation)

Numerical taxonomy is also known as computer-assisted classification [23]. Sneath and Sokal [63] define numerical taxonomy as 'the grouping by numerical methods of taxonomic units into taxa on the basis of their character states'. The methods involved require the conversion of information on taxonomic entities into numerical quantities. Numerical taxonomy is often incorrectly used as a synonym for phenetic analyses of phenotypic data, and such use should be avoided. Phenetic principles, based on the concept of Adansonian taxonomy, state that maximum information content should be achieved: i.e. all possible characters should be studied for the strains, they should be equally weighted, and taxa should be defined on the basis of overall similarity according to the results of the analyses.

A phenetic evaluation of phenotypic data involves five essential steps [65]:

1. *Selection of strains.* There are no rules controlling the range of diversity among the strains to be examined. It is recommended that the set of strains should be as large as possible, and should include cultures of historical, pathological or environmental importance. It is important to include reference strains (type strains) whose identity has been established for comparative purposes. As mentioned above, it is recommendable

that the strains represent fresh isolates, so that little modifications have occurred due to laboratory adaptations.

2. *Test selection.* Routine tests should represent a broad spectrum of the biological activities of the organism and include morphological, colonial, biochemical, nutritional and physiological characters. Tests that are not highly reproducible, including some routinely employed in conventional bacterial taxonomy, should be avoided. An optimum number of tests for numerical taxonomy is considered to be between 100 and 200. Standardization of treatment, inoculation and incubation of the strains is also required.
3. *Data coding.* Generally data are given in a binary numerical format. Positive responses (plus) are coded as 1 and negative responses (minus) are coded as 0. Weighting of characters is usually avoided.
4. *Computer analysis.* Coded data are computerized by one of the several available programs. Among the different similarity coefficients available, bacteriologists generally employ either the simple matching coefficient ( $S_{SM}$ ), or the Jaccard coefficient ( $S_J$ ). The data analyses generate similarity matrices containing information about relationships among the strains. Subsequently, cluster analyses are performed to finally generate the dendrograms.
5. *Presentation and interpretation of results.* This last step is mainly dependent on the dataset used. Results can be presented in sorted similarity matrices, as well as dendrograms. These branched diagrams are partially informative because they are generated with the highest similarity values linking a pair of organisms, but give an easy visual overview. One problem with dendrograms is that inexperienced researchers are not aware that they do not necessarily reflect phylogenetic relationships between strains (i.e. relationships based on ancestry of the organisms), and sometimes can lead to misinterpretations.

Numerical taxonomy has supported the development of stable prokaryotic classifications, especially the determination of homogeneous groups that can be equated with species. Furthermore, the databases generated are essentially information storage and identification systems.

### 3.2.3. Chemotaxonomy

Phenotypic methods comprise all methods that are not directed toward DNA or RNA, and thus also include chemotaxonomic techniques. The term 'chemotaxonomy' refers to the application of analytical methods for collecting information on various chemical constituents of the cell in order to classify bacteria. The introduction of chemotaxonomy is generally considered one of the essential milestones in the development of modern bacterial classification, it is often treated as a separate unit in taxonomic reviews [40]. However, as the parameters measured are a

direct reflection of the expression of the genetic information of an organism, they should be regarded as phenotype. As for other phenotypic and genomic information retrieval techniques, some of the chemotaxonomic methods have been widely applied on vast numbers of bacteria whereas others were so specific that their application was restricted to particular taxa [40].

Chemotaxonomy is concerned with the discontinuous distribution of specific chemicals, notably amino acids, lipids, proteins and sugars, and in this sense can be considered to provide good characters for classification and identification [23]. It is, however, important that the observed variation in chemical composition is the result of genetic differences and not due to variation in cultivation conditions. Therefore, it is usually necessary to grow cultures under carefully standardized growth regimes before comparative chemotaxonomic work can be undertaken. Rigorously standardized cultivation conditions are particularly important in studies involving quantitative analyses of chemical data [23]. Several techniques are increasingly being used routinely in prokaryotic taxonomy, e.g.:

*Cell wall composition.* This character is generally used for the classification of Gram positive organisms. The peptidoglycan type and teichoic acids are analyzed [66,67].

*Lipids.* The composition and relative ratio of fatty acids (hydroxylated, non-hydroxylated, branched...), polar lipids, lipopolysaccharides, isoprenoid quinones (ubiquinones, menaquinones...), are generally analyzed by chromatography and are used successfully for discriminating among taxa of various ranks [67,68].

*Polyamines.* These are polycationic compounds with an important but unclear role in the prokaryotic cell. Their composition and relative ratio can be discriminative for taxa above the rank of genus [69].

### 3.2.4. Phenotype typing methods

As described in Section 3.1.4 (DNA typing methods), these are techniques that are useful for establishing relationships within a prokaryote species, but generally lack resolution above this taxon level [70–72]. There are several methods that have been successfully used for discriminating strains as well as for the understanding of intraspecific variability: (i) serotyping, based on the presence of variability in the antigenic constituents of the cells (capsules, cell envelopes, flagella, fimbria... [70,73]), (ii) electrophoretic protein profiles, based on the extraction of proteins and separation on polyacrylamide gels (whole cell protein profiles, Gram negative outer membrane protein profiles and multilocus enzyme electrophoresis... [70,71,74,75]), (iii) lipopolysaccharide electrophoretic profiles, where variations on the O-side chains are reflected in different ladder-like electrophoretic patterns [70,76,77], (iv) pyrolysis mass spectrometry, Fourier transform infrared spectroscopy

and UV resonance Raman spectroscopy [78]. These are sophisticated analytical techniques which examine the total chemical composition of bacterial cells. Due to the complexity of the analytical apparatus, these techniques have so far only been used on particular groups of bacteria [23].

### 3.2.5. Identification keys and diagnostic tables

One of the goals of phenotypic characterization is the construction of a framework for an accurate identification of organisms. This framework can consist of dichotomous identification keys where the identity of an isolate is tested in an orderly, step-like series of questions. However, diagnostic tables are more common in microbiology. Diagnostic tables contain more information than the dichotomous keys and are much more useful as a determinative aid [79]. These tables are based upon the sharing of several (usually unweighted) characters, which are characteristic and identify the taxon (the species' 'phenotypic property'; [4]). In diagnostic tables, variable characters within the studied taxon are also recorded, and this is a good index for intra-specific diversity. The success in the identification of new isolate to already established species is dependent on how accurate the description of the species is, and the accuracy is dependent on the size of the dataset analyzed. It is postulated that for a rather accurate description of a species, a minimum of 10 but better 25 strains should be studied [80]. However, in the majority of cases, new species and genera are described on the basis of only a few strains or even only one. Poor descriptions based on a small set of strains can lead to improper phenotypic circumscription of taxa, thus hindering the identification of new isolates.

### 3.2.6. Microbial identification systems

A significant contribution of industry specially to the clinical microbiology was the development of miniaturized identification systems based on classical methods. Several systems are commercially available (e.g. API, Analytab Products, Plainview, NY, USA; Biolog, Biolog, Inc., Hayward, CA, USA; Vitek, Vitek Systems, Inc. Hazelwood, MO, USA), mostly based on modifications of classical methods [81]. First-generation systems were addressed towards the identification of members of the family *Enterobacteriaceae*, and consisted of miniaturized tubes containing individual substrates, multicompartiment tubes or plates with multiple substrates, and paper strips or disks impregnated with dehydrated substrates. These methods were improved by the incorporation of highly sophisticated, computer-generated identification databases tailored for each system [81]. The major problem of such methods is that the identification results are dependent on the quality of the database. The computer will always give an identification result, then the researcher should interpret it. Incomplete or sparse databases will tend to give wrong identification results. In this regard, most of

the computerized identification systems are mainly addressed to the identification of organisms with high medical importance. All of them consist of a relative short number of key tests useful to identify a particular group of microorganisms. Therefore, these systems have been less successful in the identification of environmental isolates due to a lack of knowledge of the phenotypic diversity of microorganisms in natural environments. Thus, they should be applied very cautiously to samples that are not of clinical origin [82]. Miniaturized systems are commonly used for phenotype exploration in classification attempts. However, some taxonomists regret their use in classification because of the reduced set of tests [83]. Non-commercial, miniaturized systems with large numbers of physiological tests (over 200 different tests) have been developed for classification purposes [84]. These systems have been used successfully for the examination of the physiological diversity of environmental isolates [71,84].

#### 4. Species concepts for eukaryotes

To judge the significance and usefulness of the species concept adopted for prokaryotes, it is necessary to analyze the current state of the concept for eukaryotes. Eukaryote classification, in particular of plants and animals, has a much longer history than that of prokaryotes, and thus has been discussed and debated more broadly. Indeed, all hierarchic taxonomic ranks in the prokaryotic classification scheme have been adopted as an analogy of those designed for eukaryotes. As discussed later, taxonomic ranks above species (genus and higher classes) can be considered abstract entities [85], thus easily comparable among any taxonomic classifications. However, species are regarded as practical entities for whom the requirements for their circumscription depend on the concept adopted [85]. A direct comparison among the different species units designed for any living entity is only possible if a universally applicable concept is adopted. We will show that this is not the case for prokaryotes and eukaryotes, nor is it the case among the different eukaryotes. Some microbiologists criticize how the species unit in current use has been circumscribed for prokaryotes [2,13–15] qualifying it as too conservative, and some of them support this view by referring to the debate about primate taxonomy [14,15]. Unfortunately, these authors are not aware of the controversy occurring among the eukaryote taxonomists, where the unit that successfully circumscribes the different primate species might be the least universally applicable circumscription.

Some of the terms and early concepts in taxonomy come from the Aristotelian system of logic, where definition, genus, differentia and species were terms that inspired the naturalists of the 17th and 18th centuries to attempt to classify living organisms. However, modern classification really first began with the work of Linnaeus

in the 18th century. He was the founder of the modern binomial system of nomenclature. Linnaeus' notion of species was characterized by three different attributes: (a) distinct and monotypic, (b) immutable and created as such, and (c) breeding true [16]. He already implied that a species should be defined in terms of sexuality. During the 18th and 19th centuries, naturalists and museums started to collect large amounts of specimens from all over the world to be classified. Thus, the tradition was reinforced that species, and indeed higher taxa, must be based on morphological characters recognizable in preserved specimens, although little was known about their habits and habitats. This is the earliest species concept, a morphological species concept or morphospecies, which is not a true concept but the description of a technique which can be stated as '*a community, or a number of related communities, whose distinctive morphological characters are, in the opinion of a competent systematist, sufficiently definite to entitle it, or them, to a specific name*' [16]. However, during these two centuries, two distinct tendencies were followed: the naturalists emphasizing breeding criteria and systematists emphasizing morphological differences.

With the beginning of the 20th century, and particularly with the publication of Ernst Mayr 'Systematics and the Origin of Species' [86], the biological species concept (BSC) was established unifying genetics, systematics and evolutionary biology [16]. This is both the most widely known and most controversial species concept to date. Most of the non-taxonomists recur to a 'true breeding concept' when trying to define a species. The BSC, which considered '*species as groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups*' [86], has been refined over the past 50 years, and has been successfully applied to many of the animal lineages for which the concept was originally conceived. However, over time, taxonomists of all different disciplines have shown this concept to be unsuccessful in accommodating the smallest recognizable units that they created. We are currently experiencing a period of controversy among taxonomists and philosophers about the most adequate species concept.

At the present, at least 22 different concepts have been developed to accommodate species [87]. They can be grouped in at least three categories with different theoretical commitment [88]. Philosophers are currently disagreeing about which is the most universally applicable concept. Some of them would prefer a pragmatic concept, a small amount of theory like the phenetic species concept (PhSC '*a similarity concept based on statistically covarying characteristics which are not necessarily universal among the members of the taxa*', [88,89]). Others regard the highly theoretical evolutionary species concept (ESC, '*an entity composed of organisms which maintains its identity from other such entities through time and over space, and which has its own independent evolutionary fate and historical ten-*

dencies') as the primary concept to be universally applied [87]. Not only philosophers, but taxonomists of several disciplines are also fighting to find an appropriate circumscription of their unit. The BSC has been shown to be successful for animals, particularly insects [90,91], most of the invertebrates [92,93] and vertebrates [94,95]. It is, however, difficult to apply the BSC to animals who reproduce parthenogenetically [96,97]. On the other hand, for non-animal taxonomy, the situation is quite different. In most cases, species are described by morphological discontinuities simply because the BSC is too difficult to apply. Such is the situation with algae [98], lichens [99], fungi [100] and plants [101]. Thus, although it is the best known species concept among non-taxonomists, there is a common agreement that the BSC should be abandoned towards an ESC or at least a phylogenetic species concept (PSC, 'the smallest diagnosable monophyletic unit with a parenteral pattern of ancestry and descent'). These last two concepts regard species as monophyletic group products of natural selection and descent [87], and seem to be generally applicable among the different eukaryotic lineages [102,103].

## 5. The prokaryotic species concept

As outlined in Section 1, the field of prokaryote taxonomy experienced most of its growth during the 20th century. The classification system as well as the Linnean nomenclature were adopted as an analogy to established systems for eukaryotes, in particular the botanical code [104]. The adopted system has been satisfactory for all levels of bacterial classification but the species. Supraspecific classes, being regarded as abstract entities [85], can be compared to the classification systems established for eukaryotes. However, the concept of a prokaryote species is different because no universal concept exists [87,88], and species are (regarded as individuals [105]) practical entities [85] for whom circumscriptions may vary depending on the species concept adopted (biological, phenetic, evolutionary...; [87]). Unfortunately, through the history of prokaryote taxonomy, much more attention has been paid to the nomenclature of taxa (e.g. [1,104,106–111]) than to the practical circumscription of the species concept applied to prokaryotes.

Today's prokaryotic species concept results from empirical improvements of what has been thought to be a unit. The circumscription of the species has been optimized through the development of microbiological methods that reveal both genomic and phenotypic properties of prokaryotes, which cannot be retrieved through simple observation. Recently, the current definition of the prokaryotic species has been heavily criticized by some non-taxonomists as too conservative and ill-defined [2,13,15]. On the other hand, many other microbiologists find the present concept acceptable [112]. We will argue in the

following that based on the current state of available techniques and information on microorganisms, the current concept has shortcomings, but is the most practicable one for the moment. It also fulfills several important requirements for a concept, e.g. the resulting classification scheme is stable, operational and predictive.

### 5.1. The concept

Early definitions of bacterial species were often based on monothetic groups<sup>3</sup> described by subjectively selected sets of phenotypic properties [41]. This concept had severe limitations as, for example, strains which varied in key characters could not be identified as a member of an already classified taxon. Additionally, these original classifications were produced simultaneously by different microbiologists that applied different criteria to the classification of the same group of organisms. The number of species in a genus was influenced by the aims of the taxonomist, the extent to which the taxon had been studied, the criteria adopted to define the species and the ease by which the strains could be brought into pure culture. Some classifications were defined unevenly, for example when members of environmentally and medically important genera had been underclassified and those in industrially significant taxa overclassified [41]. Moreover, this practice often lead to nomenclatural confusions, where a single species could be simultaneously classified under several different names [24].

Until the discovery of DNA as an information-containing molecule, prokaryote classification was based solely on phenotypic characteristics. The development of numerical taxonomy [63], in which the individuals are treated as operational taxonomic units that are polythetic (they can be defined only in terms of statistically covarying characteristics), resulted in a more objective circumscription of prokaryotic units. The discovery of genetic information gave a new dimension to the species concept for microorganisms. Parameters like G+C content and overall DNA–DNA similarity have additionally been used for a more objective circumscription and as discussed below, these parameters enable at least a first rough insight into phylogenetic relationships. Thus, the species concept for prokaryotes evolved into a mostly phenetic or polythetic. This means that species are defined by a combination of independent, covarying characters, each of which may occur also outside the given class thus not being exclusive of the class [85].

There is no official definition of a species in microbiology. However, from a microbiologist's point of view 'a microbial species is a concept represented by a group of strains, that contains freshly isolated strains, stock strains

<sup>3</sup> Monothetic groups are based on a unique set of features considered to be both sufficient and necessary for the definition of the group.

*maintained in vitro for varying periods of time, and their variants (strains not identical with their parents in all characteristics), which have in common a set or pattern of correlating stable properties that separates the group from other groups of strains'* [113]. This definition only applies to prokaryotes which have been isolated in pure culture (essential for the classification of new prokaryotic species), and excludes uncultured organisms which constitute the largest proportion of living prokaryotes. However, a prokaryote species is generally considered to be '*a group of strains that show a high degree of overall similarity and differ considerably from related strain groups with respect to many independent characteristics*', or '*a collection of strains showing a high degree of overall similarity, compared to other, related groups of strains*' [114].

There are, in the literature, at least three different species definitions that, to date, tend to disappear due to the unification of criteria: (i) *taxospecies*, defined as a group of organisms (strains, isolates) with mutually high phenotypic similarity that form an independent phenotypic cluster, (ii) *genomic species* as a group showing high DNA–DNA hybridization values, and (iii) *nomenspecies* as a group that bears a binomial name [114]. The simultaneous occurrence of these three conceptual units has been definitively avoided by the agreement that a species classification can only be achieved by the integration of both phenotypic and genomic parameters. Indeed, the Committee on Reconciliation of Approaches to Bacterial Systematics [4] recommended '*that a distinct genospecies that cannot be differentiated from another genospecies on the basis of any known phenotypic property not be named until it can be differentiated by some phenotypic property*'.

It is now accepted among microbial taxonomists that a prokaryotic species should be classified after the analysis and comparison of as many parameters as possible, combining phenotypic and genomic markers in what is known to be 'polyphasic taxonomy' [40]. The evaluation of several distinct independent phenotypic and genomic characters has promoted a more unified approach to the objective delineation of a bacterial species concept. This delineation responds to a relative relaxed and pragmatic concept that can universally accommodate prokaryotic species. However, the extent to which a species is delimited has to be empirically set and takes into account the differing behavioral properties of the members of a taxa [41].

### 5.2. The limits

Unfortunately, there are no absolute boundaries for the circumscription of prokaryotic species, and this is a problem for non-taxonomists attempting to identify new isolates belonging to hitherto unclassified species. As said above, a species can only be classified through the analysis of a large set of phenotypic and genomic characters in a polyphasic approach. It is necessary to show that the studied group of strains form an independent and diag-

nosable unit within the established classification scheme. The species circumscription approach is a tedious task often underestimated in its importance.

The most accepted parameter for a numerical and quasi-absolute boundary for the species circumscription is overall DNA similarity. Values expressed as percentage of similarity or in degrees of  $\Delta T_m$  (see Section 3) are considered to be, to some extent, crude measures of genomic distances among microorganisms. These values are an indirect reflection of the genomic sequence similarity at the level of the primary structure [23], so that DNA reassociation approaches represent the best applicable procedure for the inference of genotypic relationships among closely related prokaryotes [4]. Based on numerous studies in which a high degree of correlation was found between DNA similarity, and chemotaxonomic, genomic, serological and numerical similarity, DNA reassociation has been used as standard for species delineation [12]. Empirically, it has been observed that most of the well defined prokaryotic species harbor strains with genomic similarities above 70% when optimal stringence hybridization conditions are applied [115]. This observation led the Committee on Reconciliation of Approaches to Bacterial Systematics to recommend that the boundaries for species circumscription are described in terms of DNA–DNA binding. The Committee wrote that a '*species generally would include strains with approximately 70% or greater DNA–DNA relatedness and with 5°C or less  $\Delta T_m$* ' [4]. It is important to note that the values recommended by the Committee are not absolute numbers in defining the genomic boundary of a single prokaryotic species. In some cases, these values seem to be too narrow to harbor all the strains of a single species [72,116–118], in which case a more relaxed delimitation for the unit is recommended [40,41,119]. Additionally, the second most used genomic parameter in prokaryotic taxonomy, the G+C content, also gives some numerical boundaries for the unit. Empirically, it has been observed that a single species does not usually contain strains whose G+C mol% differs more than 5%, but of course lower percentages do not guarantee that these genomes would share high DNA–DNA similarity percentages. Both parameters are necessary for an adequate species classification.

### 5.3. The 16S rRNA sequence data

The application of molecular techniques to bacterial systematics introduced a new parameter that has an enormous influence on prokaryotic classification, 16S rRNA sequence analysis. It enabled remarkable breakthroughs in the determination of relationships between distantly related bacteria [23]. Molecular sequencing is dominated by the possibility of drawing genealogical trees that represent lines of descent [21,120]. Markers with different functional pressures report on different periods of evolutionary time. A fast evolving gene can only report on recent develop-

ments whereas a conserved molecule fails in this respect but does well on ancient events [49]. In these trees, organisms represent terminal points in a genealogical tree, and it is important to note that due to the absence of a useful fossil record, no time-scaled patterns of ancestry can be drawn.

Phylogenetic reconstructions have opened the door for a more objective classification system among prokaryotes, specially for taxa above species. It allowed for example the recognition of the phyletic nature of the classified taxa. A group of taxa can be monophyletic if they are derived from a single common ancestor; polyphyletic if they are derived from more than one common ancestor; and paraphyletic if the taxa are derived from a common ancestor but the group does not include all descendent taxa of the same common ancestor (Fig. 5). Of course, one of the goals of any taxonomy is to create a classification scheme that reflects the genealogy of the organisms, thus circumscribing all taxa in monophyletic groups. Phylogenetic reconstruction based on rRNA sequence analyses allowed the recognition of badly circumscribed taxa and their further reclassification. Two of the many examples of polyphyletic groups are: the family *Pseudomonadaceae* [121] that harbored taxa that have been reclassified into different genera or families [122], and the species *Zoogloea ramigera* that harbored strains belonging to quite different genera [123]. Examples of paraphyletic taxa are also common in the literature. Examples are genera like *Caulobacter* or *Brevundimonas* that harbored species with an incorrect genus designation [124]. Many more examples of polyphyletic and paraphyletic groups can be found in the reconstructed phylogenetic trees at web pages of Bergey's Manual of Systematic Bacteriology [125], or the Microbiology Department of the Munich Technical University [126].

The analysis of 16S rRNA sequences produces numerical values of 16S rRNA similarities that can be used as circumscription limits for taxa. This has been especially useful for classes above the species level where, for example, a genus could be defined by species with 95% sequence similarity [49]. It has been observed that organisms with genomic similarities above 70% usually share more than 97% 16S rRNA sequence similarity [12,127]. This value could be used as an absolute boundary for the species circumscription. However, 16S rRNA lacks resolving

power at the species level. Thus, there are examples (Fig. 6) of different species with identical [128] or nearly identical 16S rRNA sequences [35,36], a micro heterogeneity of the 16S rRNA genes within a single species [129,130] or, in exceptional cases, single organisms with two or more 16S rRNA genes with relatively high sequence divergence [131,132]. Due to the highly conserved nature of the 16S rRNA, there is no linear correlation between DNA–DNA similarity percent and 16S rRNA similarity for closely related organisms [12,53] and the bacterial species definition can never be solely based on sequence similarity of rRNAs. However, comparative analysis of 16S rRNA is a very good method for a first phylogenetic affiliation of both potentially novel and poorly classified organisms [41]. Due to the practical advantage of the 16S rRNA approach for identification purposes [51], it is recommended to include the ribosomal sequence in new descriptions of prokaryotic species. In the near future, it may become a necessary parameter, together with DNA–DNA similarity and G+C content, for any classification.

#### 5.4. Intraspecific subdivisions

Subspecies is the lowest taxonomic rank that has official standing in nomenclature [104]. A species may be divided into two or more subspecies based on minor but consistent phenotypic variations within the species or on genetically determined clusters of strains within the species [22]. However, many bacterial species are endowed with a relative internal heterogeneity that does not show consistency for a subspecific subdivision. Strains of a single species can sometimes be grouped in terms of some independent special characteristics. These groups or infrasubspecific subdivisions are not arranged in any order of rank, and may overlap one another [104]. For example, members of any species can be grouped in terms of biochemical or physiological properties (biovar or biotype), of pathogenic reactions (pathovar or pathotype), of reactions to bacteriophages (phagovar or phagotype), antigenic characteristics (serovar or serotype), and so on.

DNA reassociation experiments, however, produce intraspecific subdivisions that are often seen as potential species, but their independent classification is hampered by the lack of diagnostic phenotypic property [4]. There is, in the literature, a terminological confusion about the

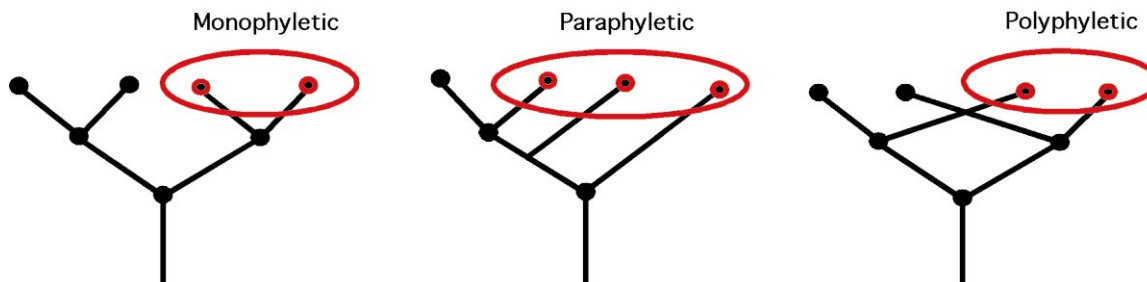


Fig. 5. Description of the terms mono-, para- and polyphyletic. The ovals indicate the circumscribed taxa.



phenotypically similar but genotypically distinct groups, which have been referred to as genomic species [114], genospecies [4], DNA groups [130], genomospecies [116,117], genomic groups [133] or genomovars [118,119]. There is an open discussion about the adequacy of these terms used to designate these subdivisions [40]. Actually, each of the different intraspecific units defined by DNA similarity values cannot be considered a single species per se. The polythetic nature of the currently accepted species concept does not allow the recognition of a species based on a single characteristic. This means that although the DNA reassociation values give a numerical boundary for the species circumscription, this unit cannot be recognized unless there is an overall agreement on the distinct characters

analyzed in a polyphasic study. Thus, the suffix ‘-species’ is inadequate for this term as far as DNA similarity groups per se lack a specific standing. Similarly, the term ‘group’ is informal and has no nomenclatural standing, and among the suffixes, ‘-var’ is recommended [104]. The term genomovar has been suggested to accommodate different DNA similarity groups within a nomenclature [118,119]. This term has been positively accepted by some taxonomists [41,72] because it indicates that a genomic species is an integral part of a nomenclature. It is suggested that genomovars encompassed in species should be numbered, and not named [118]. Ultimately, genomovars could be given a formal name when a determinative phenotype is described.

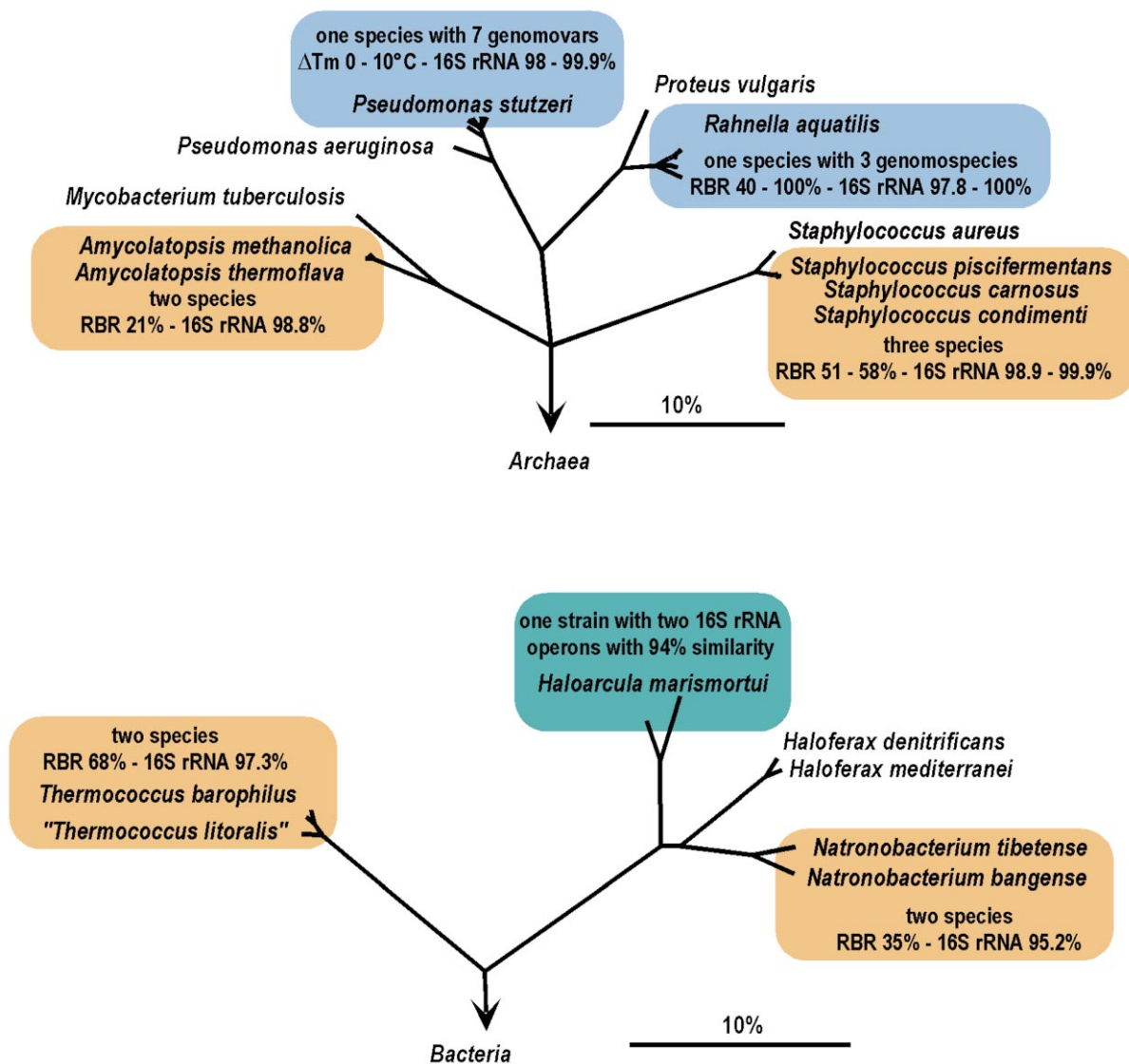


Fig. 6. Correlation among DNA–DNA similarity data and 16S rRNA-based phylogenetic reconstructions. Cases like *Natronobacterium* species and *Thermococcus* species can be regarded as optimal correlation of both genomic data. Exceptions to this rule can easily be found: highly similar 16S rRNA gene sequence and very low genomic similarity (*Amycolatopsis methanolica* and *Amycolatopsis thermoflava*). Nearly identical 16S rRNA but phenotypically and genomically different species (*Staphylococcus piscifermentans*, *Staphylococcus carnosus* and *Staphylococcus condimentii*). Genomically heterogeneous but phenotypically indistinguishable like *P. stutzeri* or *Rahnella aquatilis*. And a single strain harboring different rRNA operons with 6% of sequence differences as reported for *Haloarcula marismortui*.

It has been empirically observed that the circumscription of the prokaryotic species should be more relaxed in absolute values of genomic similarity [40,119]. The level of 70% binding and  $5^{\circ}\text{C } \Delta T_m$  is very strict, and often phenotypic consistency would not exist if these recommendations were strictly applied. More relaxed boundaries for the species delineation would be a group of strains sharing 50–70% DNA reassociation and 5–7°C difference in thermal stability between the homoduplex and heteroduplex [119], which may be a more realistic standard [40]. Allowing internal genomic heterogeneity (represented by genomovars) and more relaxed genomic boundaries, the circumscription of the prokaryotic species will have an even more conservative nature, but it may be a more pragmatic definition for facilitating diagnosis.

### 5.5. Prokaryotic sex

Attempts to circumscribe the prokaryotic species in terms of genetic exchange, in what could be analogous to the 'BSC' ([86,134,135]), have been made [136,137]. However, this prokaryotic species concept has no direct theoretical analogy to the BSC developed for eukaryotes. It has been adapted to the particular way in which prokaryotes share gene pools. Prokaryotes do not have the same reproductive mechanisms as eukaryotes, i.e. meiosis, fertilization. They are haploid and reproduce asexually by binary fission in which the genetic material is passed vertically from mother to daughter cell. Thus the evolutionary pattern would only be linked to genome organization rearrangements and mutation rate. The latter is considered to be one of the major sources of genetic diversity [2]. However, prokaryotes have several common horizontal gene transfer systems, i.e. conjugation, transformation and transduction [138]. Originally, it was assumed that these horizontal gene transfer systems were restricted to closely related organisms, and that the homologous recombination would follow the same patterns observed for eukaryotes [137]. However, genetic exchange in prokaryotes is less frequent but more promiscuous than that in eukaryotes [37]. In eukaryotes with sexual reproduction, populations that are separated by only 2% sequence divergence are frequently unable to exchange genes. Prokaryotic genomes, in contrast, may undergo homologous recombination with related species that are up to 25% (and possibly more) divergent in the sequences of homologous genes; they can also accept and express new genes on plasmids from extremely divergent sources [139]. Prokaryote taxonomy has consequently, to date, not benefited greatly from studies involving genetic exchange of chromosomal material, and the BSC is far from being realized [140]. Indeed, there is no sense in applying the BSC for prokaryotes because it fails in one of the basic statements, the interbreeding discrimination.

Independent of the best species concept for microbiology, taxonomists are aware that horizontal gene transfer

may have important consequences for bacterial systematics [140]. It may play a role in one of the most questioned aspects within prokaryotic systematics, namely current phylogenetic reconstructions. Genealogical relationships are basically inferred by the analysis of a single gene encoding 16S rRNA. Although 16S rRNA trees are consistent with trees generated from other molecules [49], it remains unclear whether microbiologists are dealing with a true organismal tree or merely a gene tree [141]. The skepticism about the value of 16S rRNA as a biological chronometer originates from the release of the first prokaryotic whole genome sequences. It seems that horizontal gene transfer might have played a more important role in evolution than originally assumed, and some incongruities are found with the current phylogenetic scheme [39,142]. A better understanding of the extent of gene exchange among prokaryotes will have to await the analysis of further whole genome sequences.

Fortunately, the prokaryotic species concept as it is currently conceived needs not be troubled by the problem of gene exchange. It is true that the acquisition of genomic material through horizontal gene transfer, and/or the presence of extrachromosomal elements can lead to misclassifications or misidentifications of prokaryotes because of their influence on the phenotype [140]. However, such characters, particularly those coded on extrachromosomal elements, should be excluded from taxonomic studies once they are known to exist. In spite of these problems, which may be minimized by computer-assisted extensive phenotypic analyses [140], the species concept for prokaryotes is appropriate because it is based on whole genome similarities [12]. As explained in Section 3, DNA reassociation values are an indirect expression of the real genome sequence identity. One can estimate that for example 96% sequence identity between two genomes compares to 70% DNA similarity [12], or  $4^{\circ}\text{C}$  of  $\Delta T_m$  [143]. Genome sequence variation within the range of values recommended for the description of a single species can account for significant differences in phenotype. However, despite genomic rearrangement caused by horizontal transfer and the presence of extrachromosomal elements, the primary structure of the majority of genes is most likely not involved [12]. 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. The current microbiological species definition consequently is relatively insensitive to genetic rearrangement, gene amplification, mutation and exchange of genetic material over a non-predictable range of taxa [12].

### 5.6. Pure cultures and the culture collections

The prerequisite of the current prokaryotic species concept is the isolation of the microorganisms in pure cul-

tures. Biochemical tests, genome analyses or chemical component analysis cannot be performed unless the organism under study is analyzed separately [83]. The isolation and maintenance of organisms in pure cultures requires time and skill and remains one of the prime challenges for microbiologists. However, the classification of a prokaryotic species cannot rely on a single strain or isolate, but rather on as many strains as possible in order to evidence the real intraspecific diversity. It has been calculated that 25 strains are necessary for an accurate description of a species, and that the lowest tolerable limit is 10 strains [80]. This number is, unfortunately, seldom achieved in taxonomic publications. In the two most renowned journals for prokaryotic taxonomy, i.e. the 'International Journal of Systematic Bacteriology' (IJSB) and 'Systematic and Applied Microbiology', most of the newly classified species are based on a single isolate (see Table 3 in Section 2), thus providing a relatively poor description that can lead to wrong classifications, or the inability to assign new isolates to this species.

After the classification of a new species, it is important to identify one of the isolates as the type strain, that is the reference strain for other scientists to be used for comparisons. This is named 'type strain and consist of living cultures of an organism which are descended from a strain designated as the nomenclatural type. The strain should have been maintained in pure culture and should agree closely in its characters with those in the original description' [104]. It is also recommended to select reference strains for each infraspecific subdivision particularly for genomovars which might be reclassified as different species later on.

Another important aspect of the pure culture technique is that isolated strains need to be preserved and made available to the scientific community for information and comparison. Most microbiologists tend to establish their own culture collections. However, reference collections of bacteria (culture collections) have been set up for the maintenance of large numbers of strains of microorganisms, as well as to make available reference strains, particularly type strains, that are needed for comparative work [80,144]. It is therefore important to deposit the type strain of a species (or the reference strains of a genomovar; [118]) in one of the public all accessible reference culture collections [144].

It is currently impossible to retrieve sufficient information for taxonomic studies on uncultured organisms from environmental samples. However, the use of the rRNA approach in exploring uncultured prokaryotes in natural samples has given valuable insights into prokaryotic diversity [52,145]. Unfortunately, as stated above, the rRNA approach lacks resolution at the species level. Also it is very difficult to infer physiological characteristics of an organism based only on its rRNA sequence. Phylogenetically related prokaryotes can have diverse physiologies and physiologically similar organisms can occur in differ-

ent phylogenetic lineages [146]. A novel rRNA sequence isolated from nature therefore merely indicates that there is a currently unknown microorganism in the environment. Knowing the phylogenetic affiliation of new microorganisms might, however, help in their cultivation, since growth conditions can be based on those for the closest cultivable relative [147]. Additionally, the rRNA approach allows the design of rRNA-targeted probes for unique sequence motifs of the unknown microorganism. These probes can subsequently be used for the in situ identification of the organism from which the sequence was retrieved [54,148]. This approach permits, among many other applications, the monitoring of the isolation of new organisms [149], as well as the determination of the morphology, abundance, distribution in certain habitats, and even indications on growth rates and physiological activities of uncultured organisms [150,151]. These new methods allow the retrieval of some information on uncultured organisms and permit the recognition of their uniqueness within the hitherto established classification scheme. In this regard, the International Committee on Systematic Bacteriology implemented the category of *Candidatus* to record the properties of putative taxa of prokaryotes. This category is used for 'describing prokaryotic entities for which more than a mere sequence is available but for which characteristics required for description according to the International Code of Nomenclature of Bacteria are lacking' [152]. Such descriptions should include not only phylogenetic information, but also information on morphological and ecophysiological features as far as they can be retrieved in situ, together with the natural environment of the organism. It is important to note that *Candidatus* is not a rank but a provisional status, and efforts to isolate and characterize the members of the putative taxa should be made to enable their definite classification.

### 5.7. The species concept to be used for prokaryotes

Recently, the current 'polyphasic' species concept has been heavily criticized [2,13–15]. Besides the claim that the current concept is too conservative, it was argued that it lacks congruency with the concept delineated for higher organisms. Indeed, it is very difficult to compare species concepts for prokaryotes and eukaryotes, and as May stated [3], the basic notions about what constitutes a species would be necessarily different for vertebrates than for bacteria. Comparisons among the different species units of living organisms can only be made after a universal species concept has been devised that applies to all organisms. This is, however, not easy. Eukaryote taxonomists are currently in disagreement about the adequate concept to be used among the > 20 concepts currently being discussed [16,87–89,102,153–156]. There is, however, a general tendency that Mayr's BSC should be abandoned because of its lack of practicability. Among the species concepts, there seem to be two candidates which are uni-

versally applicable and thus could serve for the classification of all living organisms [87,88]: the PhSC (polythetic) and the ESC.

1. *The PhSC.* The phenetic or polythetic [85] species concept (PhSC) is a similarity concept based on statistically covarying characteristics which are not necessarily universal among the members of the taxa [88]. This is the concept that has empirically been adopted to circumscribe the prokaryotic species, which to date appears to be rather stable as well as operational. This concept has no theoretical commitment as it is considered theory-neutral or theory-free [88]. A theoretical foundation has traditionally been seen as a valuable characteristic of a species concept [87,88]. However, the significance of theory in a species concept is a controversial issue between philosophers and scientists. It has been argued that the more theoretically significant a concept is, the more difficult it is to apply [88]. Scientists, in general, see no real need to include the criteria of applicability in their theory-based concepts. On the other hand, pragmatism in their concepts is seen as a valuable virtue. As analyzed by Hull [88], the PhSC covers most of the primary requirements for being a concept, i.e. universality, monism and applicability, which are considered to be the most valuable characteristics for scientists. It is, to date, the most similar concept to the one applied to prokaryotic species, it has also been recommended for higher organisms [89], and although it is considered to be theory-free, this is one of the most persuasive concepts as yet conceived [88].
2. *The ESC.* The evolutionary species concept (ESC) has been considered the most theoretically committed of the species concepts [87,88]. It is regarded as the only one that can serve as a primary concept because it can accommodate all types of species known [87]. Evolutionary species is a lineage concept which is explicitly temporal, treating these units as lineages extended in time (space–time worms; [88]). This concept, however, has no pragmatic significance for the prokaryotes when we analyze the current state of knowledge about this group of organisms. Among the prokaryotes, we cannot recognize an evolutionary fate nor historical tendencies because of the lack of a useful fossil record. Morphological features, in general, have little information content in prokaryotes, and therefore, the rare prokaryotic fossils found so far do not provide sufficient information on genomic and phenotypic characteristics of the ancestors of present prokaryotic species. The same is true for a prospective on prokaryote evolution. Currently, the knowledge of evolutionary tempo and mode of evolutionary changes in prokaryotes is rather incomplete. Different groups have been demonstrated to evolve non-isochronically [12]. The predictions become even more difficult if we take into account the possibilities of horizontal gene transfers between

distant groups. Thus to date, the adoption of an ESC for prokaryotes is not yet possible.

There is a third type of concept, that although not considered to be the most universally applicable [87,88], has been recommended to be used for animals [94]:

3. *The PSC.* There are two different versions of the polygenetic species concept (PSC): the monophyletic (or autapomorphic) species concept and the diagnostic species concept [88]. Both of them are defined as phylogenetic (or genealogical) concepts with a minimal time-dimension. In spite of the lack of useful fossil record for prokaryotes, modern molecular techniques permitted the establishment of genealogical trees among the prokaryotes through 16S rRNA gene sequence analysis [49]. As discussed above, this phylogenetic reconstruction is based on a single gene analysis, and its ability to represent the organismal phylogeny is questioned today. However, based on the congruency of the reconstructed trees with similar slowly evolving molecules [49], and the coherency of the phenetically designed taxa with the phylogenetic reconstructions [157], it is most likely that at least the local branches represent stable genealogical relationships.

The monophyletic species concept considers that ‘a species is the least inclusive monophyletic group definable by at least one autapomorphy<sup>4</sup>’ [88]. Disregarding the low resolution of 16S rRNA analyses, we can recognize each of the species as a monophyletic group. The problem is, however, the recognition of which characters are truly autapomorphies. The only possibility of recognizing an autapomorphy among members of a taxon is the availability of a gene sequence that is exclusive to all members of a species. The sequences should show to be homologous, unique for the taxon, and excluded from horizontal gene transfer. This is a nearly impossible task and thus this concept is not operational for the prokaryotes.

The diagnostic species concept considers that ‘a species is the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent’ [88]. We can recognize a pattern of ancestry among the prokaryote species in our phylogenetic tree, as well as the diagnosable units are circumscribed after the polyphasic approach. This concept would indeed serve as a primary concept for prokaryotes. However, there is some danger in attempting to recognize the smallest diagnosable clusters of prokaryotes. On one hand, one could argue that each genomically coherent unit (genomovar within a given species) would be such a diagnosable cluster. This could lead, following a ‘splitter’ tendency [23], to a subdivision of microorganisms into smaller, more nu-

<sup>4</sup> In phylogenetics, an autapomorphy denotes a homologous character common to all members of a single taxa and thought to be exclusive of the group.

merous species than systematists are used to recognizing [88]. On the other hand, microbiologists tend to recognize units by the independent use of different approaches (i.e. serology, physiology, pathogenesis, phage-specificity, whole genome similarity...), and it would be possible to argue that each of the single approaches would lead to a diagnosis. Thus, the strict application of this concept to prokaryotes could lead to the simplification of the unit circumscription, which contrasts empirical observations that a prokaryotic species cannot be regarded as a smallest diagnosable unit [40,119].

### 5.8. The phylo-phenetic species concept

As argued, it is difficult to find a concept that accommodates what microbial taxonomists understand to be a species. This unit has been modeled since the origins of microbiology, and through the years of development and modernization of this science. Most of the prokaryote taxonomists would agree that the current circumscription of the species harbors most of the requirements of this taxonomic unit. It is universally applicable among prokaryotes. It is operational, when a group is taxonomically well characterized, and it can then be recognized by the independent use of currently available identification tools. Furthermore, its application has been shown to give, in most of the cases, a rather stable, objective and predictable classification system.

The current species concept is circumscribed by the use of three different approaches. The first approach is the demarcation of genomic boundaries of the unit after whole genome hybridization, and additionally the analysis of the G+C content, genome size, etc. This corresponds to Mallet's genotypic cluster definition of a species [155]. There are no absolute numerical borders that are universally applicable among prokaryotes. These have to be delimited in agreement with the other two characterization approaches. However, one can assume that genomes sharing less than 70–50% of DNA–DNA similarity (or more than 5–7°C of  $\Delta T_m$ ) belong to strains of different species [12]. Although DNA–DNA similarity results cannot be regarded as a result of cladistic analysis, they reflect a very tight genealogical relationship among strains that share high similarity values. Indeed, as discussed in Section 3.1.2 (DNA–DNA similarity) and Section 5.5 (Prokaryotic sex), it is expected to find values above 90% identity among genomic sequences of strains sharing higher values than 50% DNA–DNA similarity. Such observations guarantee a monophyletic nature of related strains sharing high DNA–DNA similarity values.

The second approach, which is of the same importance as the demarcation of the genomic borders, is the description of the phenotype of the taxon. This description should be exhaustive [1], because it serves in finding a determinative and discriminative 'phenotypic property' for classification [4], and also in describing the internal

diversity of a group. This last point, commonly forgotten by many researchers, is of extreme importance. When studying the phenotype of a taxon, we cannot recognize whether we are dealing with the analysis of homologous characters, or with homoplasies (false homologies, evolutionary convergences). To minimize the importance of homoplastic characters in classification, it is important to determine as many characters as possible, and to analyze the characters by treating each of them equally [63]. Phenetics does not only apply to the analysis of biochemical or physiological properties, but also to the analysis of chemotaxonomical markers, such as fatty acid profiles. Additionally, both typing methods, the DNA-based (e.g. PFGE, RAPD) and phenotypic methods (e.g. whole cell protein gel electrophoresis, multilocus enzyme electrophoresis), can be phenetically analyzed. Indeed, when the number of characters used in a phenetic analysis is sufficiently high, the clustering produced tends to reproduce the genomic grouping [71,158].

The third approach is the recognition of the monophyletic nature of members of a taxon, and their position within a reconstructed genealogy. The inclusion of individuals in a genomic cluster after DNA–DNA hybridization is a rough estimation of their monophyletic origin [4]. The analysis of the 16S rRNA sequence, although not sufficient for establishing numerical borders of the prokaryotic species, can indicate the ancestry pattern of the taxon studied, as well as confirm the monophyletic nature of the members of a group. The comparative analysis of 16S rRNA sequences is overdue to become a common additional information in the classification of prokaryotes, and is in fact already part of the vast majority of new species descriptions.

The currently practiced species concept in microbiology corresponds partially to several concepts designed for eukaryotes. It is phenetic in that the basis for the understanding of the taxon's coherency and internal diversity is based upon the numerical analysis of independently covarying characters, which are not necessarily universally present in the taxon. It is phylogenetic in that the members of these units have to show a common pattern of ancestry, i.e. they must be monophyletic. Finally, it is based on a genotypic (genomic) cluster definition [155], on the way that genome comparisons, although indirect, give objective numerical frontiers to the unit circumscription, and as discussed above guarantees the close genealogical relationship of the strains included in a cluster. We suggest that it is referred to as the phylo-phenetic species concept, indicating the combination of a phenetic evaluation of the unit with requirements for monophyly of its components. A phylo-phenetic species is '*a monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity with respect to many independent characteristics, and is diagnosable by a discriminative phenotypic property*'.

This definition includes the following requisites: (1) a

species should be a monophyletic group of organisms, with a high degree of genomic similarity. (2) The absolute genomic boundaries for the circumscription of each independent species should be particularly defined after the analysis of their phenotype. (3) The internal homogeneity or heterogeneity of the group can only be understood after the phenetic analysis of as many characters as possible. (4) A prokaryotic species should not be classified unless it can be recognized by several independent identification approaches, and given a phenotypic set of determinative properties.

#### 5.9. Guidelines to the recognition of a prokaryotic species

Microbiologists have achieved a rather stable classification system using the current circumscription of prokaryotic species. However, it is important to make correct use of the definition in order to avoid confusions and misclassifications. We would like to formulate recommendations for classifying a new prokaryotic species.

1. Try to isolate or collect an adequate number of strains of the taxon to be studied, and use all of them for comparisons. Avoid, although sometimes impossible, the description of a species based on a single strain. This could hamper the identification of new isolates.
2. Try to recognize the closest related taxa through 16S rRNA analysis and phenotypic characteristics. Include, at least, the type strains of these related taxa in the taxonomic analyses.
3. Do not use values of 70% DNA similarity (or  $5^{\circ}\text{C } \Delta T_m$ ) as absolute limits for circumscribing the species. The current concept allows more relaxed DNA–DNA similarity frontiers, and an internal genomic heterogeneity is permitted. A single species can consist of several genomic groups (genomovars) which do not necessarily have to be classified as different species. This will be possible when a phenotypic property that identifies each of them is found.

4. Make an effort to characterize the phenotype of the organisms. Although commercially available tests are useful (API, Biolog), the information retrieved might be insufficient. The phenotype is not only described by metabolism, there are for example chemotaxonomic markers (fatty acids, polyamines, quinones...) that produce important information on organisms. The more exhaustively the phenotype is described, the better the circumscription.
5. Do not be sparse in time, nor in effort, when taxonomically analyzing your strains. The classification of a species is not an easy task, and should not be underestimated.
6. Follow the nomenclature rules [104,111]. This is the best discussed and established side of prokaryotic taxonomy. Avoid using words that are hard to pronounce if you do not want to annoy your colleagues.

Of course, these are recommendations and may sometimes be difficult to follow. It is still necessary to obtain pure cultures of microorganisms to recognize them as a new species. We cannot retrieve enough information from uncultured organisms to achieve a correct and stable classification.

#### 5.10. Future species circumscriptions

*'The adequacy of characterization of a bacterium is a reflexion of time; it should be as full as modern techniques make possible. Unfortunately, one now regarded as adequate is likely, in 10 years time, to be hopelessly inadequate!'* Cowan, 1965 [1]. This statement also describes the evolutionary fate of the taxonomy of prokaryotes and also applies to our current activities. This is mainly because the development and improvement of the taxonomic classification for prokaryotes is linked to the development of modern molecular techniques. During the last years, the number of new isolates as well as the amount of informa-

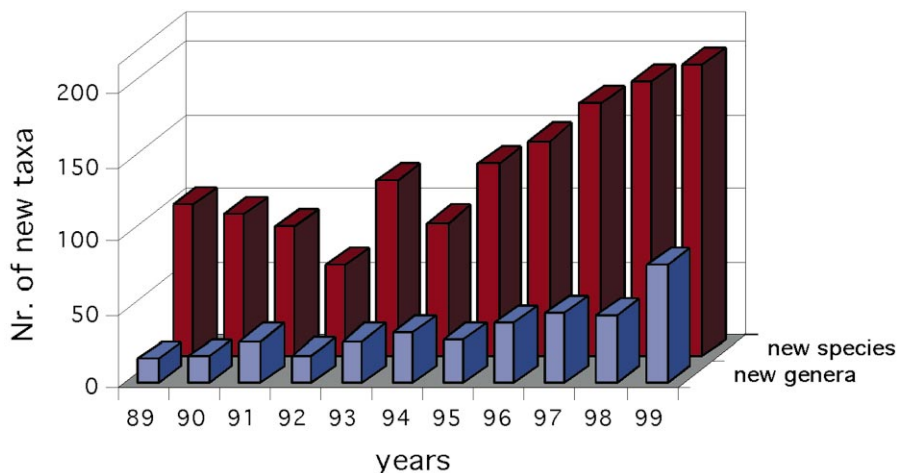


Fig. 7. Descriptions of new taxa through the years 1989 and 1999. New combinations of already existing taxa have not been included.

tion useful for systematics has increased significantly (Fig. 7). Also, the availability of information has been improved by the generation of diverse databases (16S rRNA, fatty acids, metabolic markers...) as well as software packages to handle them [40,159]. Among the different sources of information that are currently emerging, there is one for which it is difficult to predict how significant it will be for prokaryotic taxonomy: the complete genome sequence. Just as rRNA-based reconstructions of prokaryotic phylogeny have dramatically changed the classification system, genomics is likely to give insights into the naturalness or artificialness of the current classification system for prokaryotes. Dozens of prokaryote genomes have currently been sequenced or will be sequenced in the near future (www.tigr.org). Microbiology will soon be able to access an enormous flood of sequences. Preliminary results show a conserved genome organization and sequence identity within a single species [160]. Strain-specific genes, which are assumed to have been acquired by lateral gene transfer, are mostly clustered in single hypervariable regions in what can be considered hot spots of recombination [161].

So far, little attention has been paid to the genome organization as an additional parameter for species circumscription. A simplified, and less expensive and time-consuming source of information on genome organization as compared to whole genome sequencing is the physical mapping of prokaryotic genomes [55]. From experiments based mainly on PFGE of large genome fragments and Southern hybridization, one can retrieve information on genome size, plasmids, number of chromosomes and their topology, number and distribution of house-keeping genes such as *rrn* operons, genome rearrangements... In a nearly forgotten publication, Krawiec [162] proposed the use of the organization of chromosomal loci as an identifying characteristic of bacterial species. He argued that the chromosomal organization like the presence/absence of genes is an important selective feature and fundamental character of an evolving population, that the structural organization of the genome demarcates functional units of the genome, and that the organization of a genome is directly linked to the niche that this organism has adapted to. He proposed that the order of loci establishes the identity of a species as well as preserves the identity by creating a barrier to the exchange of genes. There are currently too few prokaryotes whose genome has been analyzed to evaluate those assumptions. However, these analyses give valuable information for a polyphasic approach such as genome size, and number and distribution of *rrn* operons. Indeed, both parameters appear to be fairly conservative within a single species, as described in *P. stutzeri* [59] and *Helicobacter pylori* [160]. On the other hand, some incongruences have been observed in members of the species *Vibrio cholerae* [163].

Taxonomists should have a close look at prokaryote genomes determined for medical or biotechnological reasons. This might prepare us for a time when whole genome

sequencing might be a part of the species circumscription. This would also allow a circumscription of uncultured bacteria based on large chromosomal fragments or even full genomes directly retrieved from the environment [164]. It is, indeed, too early to evaluate the impact of genomics on prokaryotic taxonomy and particularly on the species concept. In 10 years time, we shall look back to the bacterial classification and the polythetic species concept of the late 90s with Cowan's prophecy in mind.

### Acknowledgements

The present work is dedicated to the memory of Professor Jan Ursing who taught R.R.-M. how to be a taxonomist. We are grateful to P. Kämpfer and W. Ludwig for critically reviewing this manuscript and helpful comments. To Nicole Dubillier for her efforts in reviewing and improving the English of this manuscript. To Bernd Stickfort for his helpful discussions and provisions of literature. This work has been financially supported by the Max-Planck-Gesellschaft and the Europäische Akademie.

### References

- [1] Cowan, S.T. (1965) Principles and practice of bacterial taxonomy – a forward look. *J. Gen. Microbiol.* 39, 148–159.
- [2] Whitman, W.B., Coleman, D.C. and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. USA* 95, 6578–6583.
- [3] May, R.M. (1988) How many species are there on Earth? *Science* 241, 1441–1449.
- [4] Wayne, L.G. et al. (1987) Report of the Ad Hoc Committee on reconciliation of approaches to Bacterial Systematics. *Int. J. Syst. Bacteriol.* 37, 463–464.
- [5] Pace, N.R. (1996) New perspective on the natural microbial world: molecular microbial ecology. *ASM News* 62, 463–470.
- [6] Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740.
- [7] Torsvik, V., Sørheim, R. and Goksøyr, J. (1996) Total bacterial diversity in soil and sediment communities – a review. *J. Industr. Microbiol.* 17, 170–178.
- [8] Ward, D.M., Weller, R. and Bateson, M.M. (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345, 63–65.
- [9] Bull, A.T., Goodfellow, M. and Slater, J.H. (1992) Biodiversity as a source of innovation in biotechnology. *Annu. Rev. Microbiol.* 46, 219–252.
- [10] Garrity, G.M., Boone, D.R., Brenner, D.J., Krieg, N.R. and Staley, J.T. (1999) The need for a new division of environmental prokaryotic biology in the NSF. *ASM News* 65, 459.
- [11] Triplett, E.W. (1999) Proposal and justification for a new division of environmental prokaryotic biology in the NSF. *ASM News* 65, 63–64.
- [12] Stackebrandt, E. and 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.
- [13] Dykhuizen, D.E. (1998) Santa Rosalia revisited: why are there so many species of bacteria? *Antonie van Leeuwenhoek* 73, 25–33.
- [14] Staley, J.T. (1997) Biodiversity: are microbial species threatened? *Curr. Opin. Biotechnol.* 8, 340–345.

- [15] Ward, D.M. (1998) A natural species concept for prokaryotes. *Curr. Opin. Microbiol.* 1, 271–277.
- [16] Claridge, M.F., Dawah, H.A. and Wilson, M.R. (1997) Practical approaches to species concepts for living organisms. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 1–15. Chapman and Hall, London.
- [17] Siefert, J.L. and Fox, G.E. (1998) Phylogenetic mapping of bacterial morphology. *Microbiology* 144, 2803–2808.
- [18] Logan, N.A. (1994) Blackwell Scientific Publications, London.
- [19] Schlegel, H.G. (1999) Vol. 28. Deutsche Akademie der Naturforscher Leopoldina, Halle.
- [20] Schlegel, H.G. and Köhler, W. (1999) Bacteriology paved the way to cell biology: a historical account. In: *Biology of the Prokaryotes* (Lengeler, J.W., Drews, G. and Schlegel, H.G., Eds.) Thieme, Stuttgart.
- [21] Woese, C.R. (1992) Prokaryote systematics: the evolution of a science. In: *The Prokaryotes*. Second Edition, Vol. 1 (Balows, A., Trüper, H.G., Dworkin, M., Harder, W. and Schleifer, K.-H., Eds.), pp. 3–18. Springer-Verlag, Berlin.
- [22] Staley, J.T. and Krieg, N.R. (1989) Classification of procaryotic organisms: an overview. In: *Bergey's Manual of Systematic Bacteriology*, Vol. 4 (Williams, S.T., Sharpe, M.E. and Holt, J.G., Eds.), pp. 2299–2302. Williams and Wilkins, Baltimore, MD.
- [23] Goodfellow, M. and O'Donnell, A.G. (1993) Roots of bacterial systematics. In: *Handbook of New Bacterial Systematics* (Goodfellow, M. and O'Donnell, A.G., Eds.), pp. 3–54. Academic Press Ltd., London.
- [24] Van Niel, C.B. and Allen, M.B. (1952) A note on *Pseudomonas stutzeri*. *J. Bacteriol.* 64, 413–422.
- [25] Sneath, P.H.A. (1989) Numerical taxonomy. In: *Bergey's Manual of Systematic Bacteriology*, Vol. 4 (Williams, S.T., Sharpe, M.E. and Holt, J.G., Eds.), pp. 2303–2305. Williams and Wilkins, Baltimore, MD.
- [26] Brenner, D.J., Fanning, G.R., Rake, A.V. and Johnson, K.E. (1969) Batch procedure for thermal elution of DNA from hydroxyapatite. *Anal. Biochem.* 28, 447–459.
- [27] Krieg, N.R. (1988) Bacterial classification: an overview. *Can. J. Microbiol.* 34, 536–540.
- [28] Johnson, J.L. (1973) Use of nucleic acid homologies in the taxonomy of anaerobic bacteria. *Int. J. Syst. Bacteriol.* 23, 308–315.
- [29] Stackebrandt, E., Ludwig, W. and Fox, G.E. (1985) 16S ribosomal RNA oligonucleotide cataloging. *Methods Microbiol.* 18, 75–107.
- [30] De Ley, J. and De Smedt, J. (1975) Improvements of the membrane filter method for DNA–rRNA hybridization. *Antonie van Leeuwenhoek* 41, 287–307.
- [31] Ludwig, W. and Schleifer, K.-H. (1994) Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbiol. Rev.* 15, 155–173.
- [32] Maidak, B.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.J. and Woese, C.R. (1997) The RDP (Ribosomal Database Project). *Nucleic Acids Res.* 25, 109–110.
- [33] Wheelis, M.L., Kandler, O. and Woese, C.R. (1992) On the nature of global classification. *Proc. Natl. Acad. Sci. USA* 89, 2930–2934.
- [34] Olsen, G.J., Woese, C.R. and Overbeek, R. (1994) The winds of (evolutionary) change: breathing new life into microbiology. *J. Bacteriol.* 176, 1–6.
- [35] Fox, G.E., Wisotzky, J.D. and Jurtschuk Jr., P. (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* 42, 166–170.
- [36] Martínez-Murcia, A.J., Benlloch, S. and 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.
- [37] Cohan, F.M. (1994) Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol. Evol.* 9, 175–180.
- [38] Gribaldo, S., Lumia, V., Creti, R., Conway de Macario, E., Sanangelantoni, A. and Cammarano, P. (1999) Discontinuous occurrence of the *hsp70* (*dnaK*) gene among *Archaea* and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J. Bacteriol.* 181, 434–443.
- [39] Gupta, R.S. (1998) What are archaeobacteria: life's third domain or monoderm prokaryotes related to Gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. *Mol. Microbiol.* 29, 695–707.
- [40] Vandamme, P., Pot, B., Gillis, M., De Vos, P. and Swings, J. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60, 407–438.
- [41] Goodfellow, M., Manfio, G.P. and Chun, J. (1997) Towards a practical species concept for cultivable bacteria. In: *Species: the Units of Biodiversity* (Claridge, M.F. and Dawah, H.A., Eds.), pp. 25–59. Chapman and Hall, London.
- [42] 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.
- [43] Lee, K.Y., Wahl, R. and Barbu, E. (1956) Contenu en bases purique et pyrimidiques des acides deoxyribonucleiques des bacteries. *Ann. Inst. Pasteur* 91, 212–224.
- [44] Tamaoka, J. (1994) Determination of DNA base composition. In: *Chemical Methods in Prokaryotic Systematics* (Goodfellow, M. and O'Donnell, A.G., Eds.), pp. 463–469. John Wiley and Sons, Chichester.
- [45] Johnson, J.L. (1989) Nucleic acids in bacterial classification. In: *Bergey's Manual of Systematic Bacteriology*, Vol. 4 (Williams, S.T., Sharpe, M.E. and Holt, J.G., Eds.), pp. 2306–2309. Williams and Wilkins, Baltimore, MD.
- [46] Grimont, P.A.D., Popoff, M.Y., Grimont, F., Coynault, C. and Lemelin, M. (1980) Reproducibility and correlation study of three deoxyribonucleic acid hybridization procedures. *Curr. Microbiol.* 4, 325–330.
- [47] Hartford, T. and Sneath, P.H.A. (1988) Distortion of taxonomic structure from DNA relationships due to different choice of reference strains. *Syst. Appl. Microbiol.* 10, 241–250.
- [48] Hartford, T. and Sneath, P.H.A. (1990) Experimental error in DNA–DNA pairing: a survey of the literature. *J. Appl. Bacteriol.* 68, 527–542.
- [49] Ludwig, W., Strunk, O., Klugbauer, S., Klugbauer, N., Weizenegger, M., Neumaier, J., Bachleitner, M. and Schleifer, K.-H. (1998) Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* 19, 554–568.
- [50] Ludwig, W. and Schleifer, K.-H. (1999) Phylogeny of Bacteria beyond the 16S rRNA standard. *ASM News* 65, 752–757.
- [51] Ludwig, W. (1999) The role of rRNA as a phylogenetic marker in the context of genomics. *USFCC Newsl.* 29, 2–6.
- [52] Embley, T.M. and Stackebrandt, E. (1997) Species in practice: exploring uncultured prokaryote diversity in natural samples. In: *Species: The Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 61–81. Chapman and Hall, London.
- [53] Grimont, P. (1988) Use of DNA reassociation in bacterial classification. *Can. J. Microbiol.* 34, 541–546.
- [54] Amann, R.I., Ludwig, W. and Schleifer, K.-H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169.
- [55] Fonstein, M. and Haselkorn, M. (1995) Physical mapping of bacterial genomes. *J. Bacteriol.* 177, 3361–3369.
- [56] Maslow, J.N., Mulligan, M.E. and Arbeit, R.D. (1993) Molecular epidemiology: Application of contemporary techniques to the typing of microorganisms. *Clin. Infect. Dis.* 17, 153–164.
- [57] Tenover, F.C., Arbeit, R.D., Goering, R.V., Mickelsen, P.A., Murray, B.E., Persing, D.H. and Swaminathan, B. (1995) Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J. Clin. Microbiol.* 33, 2233–2239.
- [58] Palleroni, N.J. (1993) Structure of bacterial genome. In: *Handbook*



- of New Bacterial Systematics (Goodfellow, M. and O'Donnell, A.G., Eds.), pp. 57–113. Academic Press Ltd., London.
- [59] Ginard, M., Lalucat, J., Tümmler, B. and Römmling, U. (1997) Genome organization of *Pseudomonas stutzeri* and resulting taxonomic and evolutionary considerations. *Int. J. Syst. Bacteriol.* 47, 132–143.
- [60] Grimont, F. and Grimont, P.A.D. (1991) DNA fingerprinting. In: *Nucleic Acid Techniques in Bacterial Systematics* (Stackebrandt, E. and Goodfellow, M., Eds.) John Wiley and Sons Ltd., West, Sussex.
- [61] Welsh, J. and McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18, 7213–7218.
- [62] Williams, J.G.K., Kubelic, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18, 6531–6535.
- [63] Sneath, P.H.A. and Sokal, R.R. (1973) W.H. Freeman and Company, San Francisco, CA.
- [64] Smibert, R.M. and Krieg, N.R. (1994) Phenotypic characterization. In: *Methods for General and Molecular Bacteriology* (Gerhardt, P., Murray, R.G.E., Wood, W.A. and Krieg, N.R., Eds.), pp. 607–654. American Society for Microbiology, Washington, DC.
- [65] Colwell, R.R. and Austin, B. (1981) Numerical taxonomy. In: *Manual of Methods for General Bacteriology* (Gerhardt, T.P., Murray, R.G.E., Costilow, R.N., Nester, E.W., Wood, W.A., Krieg, N.R. and Phillips, G.B., Eds.), pp. 444–449. American Society for Microbiology, Washington, DC.
- [66] Schleifer, K.-H. and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* 36, 143–187.
- [67] Suzuki, K., Goodfellow, M. and O'Donnell, A.G. (1993) Cell envelopes and classification. In: *Handbook of New Bacterial Systematics* (Goodfellow, M. and O'Donnell, A.G., Eds.), pp. 195–250. Academic Press Ltd., London.
- [68] Kämpfer, P. (1998) Some chemotaxonomic and physiological properties of the genus *Sphaerotilus*. *Syst. Appl. Microbiol.* 21, 245–250.
- [69] Busse, J. and Auling, G. (1988) Polyamine pattern as a chemotaxonomic marker within the *Proteobacteria*. *Syst. Appl. Microbiol.* 11, 1–8.
- [70] Rossello, R., Garcia-Valdes, E., Macario, A.J.L., Lalucat, J. and Conway de Macario, E. (1992) Antigenic diversity of *Pseudomonas stutzeri*. *Syst. Appl. Microbiol.* 15, 617–623.
- [71] Rossello-Mora, R.A., Lalucat, J., Dott, W. and Kämpfer, P. (1994) Biochemical and chemotaxonomic characterization of *Pseudomonas stutzeri* genomovars. *J. Appl. Bacteriol.* 76, 226–233.
- [72] Vandamme, P. et al. (1998) *Pelitestga europaea* gen. nov., sp. nov., a bacterium associated with respiratory disease in pigeons taxonomic structure and phylogenetic allocation. *Int. J. Syst. Bacteriol.* 48, 431–440.
- [73] Henriksen, S.D. (1978) Serotyping of bacteria. *Methods Microbiol.* 12, 1–13.
- [74] Selander, R.K., Caugent, D.A., Ochman, H., Musser, J.M., Gilmour, M.N. and Whittam, T.S. (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol.* 51, 873–884.
- [75] Vauterin, L., Swings, J. and Kersters, K. (1991) Grouping of *Xanthomonas campestris* pathovars by SDS-PAGE of proteins. *J. Gen. Microbiol.* 137, 1677–1687.
- [76] DeWeger, L., Jann, A.B., Jann, K. and Lugtenberg, B. (1987) Lipopolysaccharides of *Pseudomonas* spp. that stimulate plant growth composition and use for strain identification. *J. Bacteriol.* 169, 1441–1446.
- [77] Siverio, F., Cambra, M., Gorris, M.T., Corzo, J. and Lopez, M.M. (1993) Lipopolysaccharides as determinants of serological variability in *Pseudomonas corrugata*. *Appl. Environ. Microbiol.* 59, 1805–1812.
- [78] Magee, J. (1993) Whole-organism fingerprinting. In: *Handbook of New Bacterial Systematics* (Goodfellow, M. and O'Donnell, A.G., Eds.) Academic Press Ltd., London.
- [79] Trüper, H.G. and Schleifer, K.-H. (1992) Prokaryote characterization and identification. In: *The Prokaryotes*. Second Edition, Vol. 1 (Baldwin, A., Trüper, H.G., Dworkin, M., Harder, W. and Schleifer, K.-H., Eds.), pp. 126–148. Springer-Verlag, Berlin.
- [80] Sneath, P.H.A. (1977) The maintenance of large numbers of strains of microorganisms and the implications for culture collections. *FEMS Microbiol. Lett.* 1, 333–334.
- [81] D'Amato, E.E., Taylor, R.H., Blannon, J.C. and Reasoner, D.J. (1991) Substrate profile systems for the identification of bacteria and yeasts by rapid and automated approaches. In: *Manual of Clinical Microbiology* (Balows, A., Hausler, W.J.J., Herrmann, K.L., Isenberg, H.D. and Shadomy, H.J., Eds.), pp. 128–136. American Society for Microbiology, Washington, DC.
- [82] Klingler, J.M., Stowe, R.P., Obenhuber, D.C., Groves, T.O., Mishra, S.K. and Pierson, D.L. (1992) Evaluation of the Biolog automated microbial identification system. *Appl. Environ. Microbiol.* 58, 2089–2092.
- [83] Palleroni, N.J. (1997) Prokaryotic diversity and the importance of culturing. *Antonie van Leeuwenhoek* 72, 3–19.
- [84] Kämpfer, P., Kroppenstedt, R.M. and Dott, W. (1991) A numerical classification of the genera *Streptomyces* and *Streptoverticillium* using miniaturized physiological tests. *J. Gen. Microbiol.* 137, 1831–1891.
- [85] Van Regenmortel, M.H.V. (1997) Viral species. In: *Species: the Units of Biodiversity* (Claridge, M.F. and Dawah, H.A., Eds.), pp. 17–24. Chapman and Hall, London.
- [86] Mayr, E. (1942) Columbia University Press, New York.
- [87] Mayden, R.L. (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 381–324. Chapman and Hall, London.
- [88] Hull, D.L. (1997) The ideal species concept-and why we can't get it. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 357–380. Chapman and Hall, London.
- [89] Sokal, R.R. and Crovello, T.J. (1970) The biological species concept: a critical evaluation. *Am. Nat.* 104, 127–153.
- [90] Claridge, M.F., Dawah, H.A. and Wilson, M.R. (1997) Species in insect herbivores and parasitoids-sibling species, host races and biotypes. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 247–272. Chapman and Hall, London.
- [91] Lane, R. (1997) The species concept in blood-sucking vectors of human diseases. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 273–290. Chapman and Hall, London.
- [92] Knowlton, N. and Weigt, L.A. (1997) Species of marine invertebrates: a comparison of the biological and phylogenetic species concept. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 199–220. Chapman and Hall, London.
- [93] Minelli, A. and Foddai, D. (1997) The species in terrestrial non-insect invertebrates (earthworms, arachnids, myriapods, woodlice and snails). In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 309–325. Chapman and Hall, London.
- [94] Cracraft, J. (1997) Species concepts in systematics and conservation biology – an ornithological viewpoint. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 325–340. Chapman and Hall, London.
- [95] Corbet, G.B. (1997) The species in mammals. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 341–356. Chapman and Hall, London.
- [96] Hunt, D.J. (1997) Nematode species: concepts and identification strategies exemplified by the Longidoridae, Steinernematidae and Heteroharbitidae. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 221–246. Chapman and Hall, London.
- [97] Footitt, R.G. (1997) Recognition of parthenogenetic insect species. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 291–308. Chapman and Hall, London.

- [98] John, D.M. and Maggs, C.A. (1997) Species problems in eukaryotic algae: a modern perspective. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 83–108. Chapman and Hall, London.
- [99] Purvis, O.W. (1997) The species concept in lichens. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 109–134. Chapman and Hall, London.
- [100] Brasier, C.M. (1997) Fungal species in practice. Identifying species units in fungi. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 135–170. Chapman and Hall, London.
- [101] Gornall, R.J. (1997) Practical aspects of the species concept in plants. In: *Species: the Units of Biodiversity* (Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds.), pp. 171–190. Chapman and Hall, London.
- [102] Ereshefsky, M. (1992) The MIT Press, Cambridge, MA.
- [103] Templeton, A.R. (1989) The meaning of species and speciation: a genetic perspective. In: *Speciation and its Consequences* (Otte, D. and Endler, J.A., Eds.), pp. 3–27. Sinauer Ass., Inc., Sunderland, MA.
- [104] Sneath, P.H.A. (1992) American Society for Microbiology, Washington, DC.
- [105] Hull, D.L. (1976) Are species really individuals? *Syst. Zool.* 25, 174–191.
- [106] Buchanan, R.E., John-Brooks, R.S. and Breed, R.S. (1948) International bacteriological code of nomenclature. *J. Bacteriol.* 55, 287–306.
- [107] Buchanan, R.E. (1955) *Taxonomy*. *Ann. Rev. Microbiol.* 9, 1–20.
- [108] Heise, H. and Starr, M.P. (1946) Nomenifers: are they christened or classified? *Syst. Zool.* 17, 458–467.
- [109] Sneath, P.H.A. (1989) Bacterial nomenclature. In: *Bergey's Manual of Systematic Bacteriology*, Vol. 4 (Williams, S.T., Sharpe, M.E. and Holt, J.G., Eds.), pp. 2317–2321. Williams and Wilkins, Baltimore, MD.
- [110] Trüper, H.G. (1996) Help! Latin! How to avoid the most common mistakes while giving Latin names to newly discovered prokaryotes. *Microbiol. SEM* 12, 473–475.
- [111] Trüper, H.G. (1999) How to name a prokaryote? Etymological considerations, proposals and practical advice in prokaryote nomenclature. *FEMS Microbiol. Rev.* 23, 231–249.
- [112] Stahl, D.A. (1996) Molecular approaches for the measurement of density, diversity and phylogeny. In: *Manual of Environmental Microbiology* (Hurst, C.J., Knudsen, G.R., McInerney, M.J., Stetzenbach, L.D. and Walter, M.V., Eds.), pp. 102–114. ASM Press, Washington, DC.
- [113] Gordon, R.E. (1978) A species definition. *Int. J. Syst. Bacteriol.* 28, 605–607.
- [114] Colwell, R.R., Clayton, R.A., Ortiz-Conde, B.A., Jacobs, D. and Russek-Cohen, E. (1995) The microbial species concept and biodiversity. In: *Microbial Diversity and Ecosystem Function* (Allsopp, D., Colwell, R.R. and Hawksworth, D.L., Eds.), pp. 3–15. CAB International, Oxon.
- [115] Schleifer, K.-H. and Stackebrandt, E. (1983) Molecular systematics of prokaryotes. *Ann. Rev. Microbiol.* 37, 143–187.
- [116] Brenner, D.J., Grimont, P.A.D., Steigerwalt, A.G., Fanning, G.R., Ageron, E. and Riddle, C.F. (1993) Classification of citrobacteria by DNA hybridization designation of *Citrobacter farmeri* sp. nov., *Citrobacter youngae* sp. nov. *Citrobacter braakii* sp. nov., *Citrobacter werkmanii* sp. nov., *Citrobacter sedlakii* sp. nov., and three unnamed *Citrobacter* genomospecies. *Int. J. Syst. Bacteriol.* 43, 645–658.
- [117] Brenner, D.J., Muller, H.E., Steigerwalt, A.G., Whitney, A.M., O'Hara, C.M. and Kämpfer, P. (1998) Two new *Rahnella* genomospecies that cannot be phenotypically differentiated from *Rahnella aquatilis*. *Int. J. Syst. Bacteriol.* 48, 141–149.
- [118] Rossello, R., Garcia-Valdes, E., Lalucat, J. and Ursing, J. (1991) Genotypic and phenotypic diversity of *Pseudomonas stutzeri*. *Syst. Appl. Microbiol.* 14, 150–157.
- [119] Ursing, J.B., Rossello-Mora, R.A., Garcia-Valdes, E. and Lalucat, J. (1995) Taxonomic note: a pragmatic approach to the nomenclature of phenotypically similar genomic groups. *Int. J. Syst. Bacteriol.* 45, 604.
- [120] Woese, C.R. (1987) Bacterial evolution. *Microbiol. Rev.* 51, 221–271.
- [121] Palleroni, N.J. (1984) *Pseudomonadaceae*. In: *Bergey's Manual of Systematic Bacteriology*, Vol. 1 (Krieg, N.R. and Holt, J.G., Eds.), pp. 141–199. The Williams and Wilkins, Baltimore, MD.
- [122] Palleroni, N.J. (1991) Present situation of the taxonomy of aerobic Pseudomonads. In: *Pseudomonas*, Molecular Biology and Biotechnology (Galli, E., Silver, S. and Witholt, B., Eds.), pp. 105–115. ASM, Washington, DC.
- [123] Rossello-Mora, R.A., Ludwig, W. and Schleifer, K.-H. (1993) *Zoogloea ramigera*: a phylogenetically diverse species. *FEMS Microbiol. Lett.* 114, 129–134.
- [124] Abraham, W.R. et al. (1999) Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundimonas* and *Caulobacter*. *Int. J. Syst. Bacteriol.* 49, 1053–1073.
- [125] Bergey's (1998) <http://www.cme.msu.edu/bergeys/>.
- [126] Ludwig, W. (1999) <http://www.mikrobiologie.tu-muenchen.de/pub/ARB>, Department of microbiology, TU-München, Munich.
- [127] Amann, R.I., Lin, C., Key, R., Montgomery, L. and Stahl, D.A. (1992) Diversity among fibrobacter isolates: towards a phylogenetic classification. *Syst. Appl. Microbiol.* 15, 23–31.
- [128] Probst, A., Hertel, C., Richter, L., Wassill, L., Ludwig, W. and Hammes, W.P. (1998) *Staphylococcus condimenti* sp. nov., from soy sauce mash, and *Staphylococcus carnosus* (Schleifer and Fischer 1982) subsp. *utilis* subsp. nov.. *Int. J. Syst. Bacteriol.* 48, 651–658.
- [129] Bennisar, A., Rossello-Mora, R., Lalucat, J. and Moore, E.R.B. (1996) 16S rRNA gene sequence analysis relative to genomovars of *Pseudomonas stutzeri* and proposal of *Pseudomonas balearica* sp. nov.. *Int. J. Syst. Bacteriol.* 46, 200–205.
- [130] Ibrahim, A., Gerner-Smidt, P. and Liesack, W. (1997) Phylogenetic relationship of the twenty-one DNA groups of the genus *Acinetobacter* as revealed by 16S ribosomal DNA sequence analysis. *Int. J. Syst. Bacteriol.* 47, 437.
- [131] Mylvaganam, S. and Dennis, P.P. (1992) Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaeobacterium *Haloarcula marismortui*. *Genetics* 130, 399–410.
- [132] Nübel, U., Engelen, B., Felske, A., Snaidr, J., Wieshuber, A., Amann, R.I., Ludwig, W. and Backhaus, H. (1996) Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *J. Bacteriol.* 178, 5636–5643.
- [133] Ziemke, F., Höfle, M., Lalucat, J. and Rossello-Mora, R. (1998) Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov.. *Int. J. Syst. Bacteriol.* 48, 179–186.
- [134] Mayr, E. (1970) The Belknap Press of Harvard University Press, Cambridge, MA.
- [135] Mayr, E. (1987) The ontological status of species: scientific progress and philosophical terminology. *Biol. Phil.* 2, 145–166.
- [136] Ravin, A.W. (1960) The origin of bacterial species: Genetic recombination and factors limiting it between bacterial populations. *Bact. Rev.* 24, 201–220.
- [137] Ravin, A.W. (1963) Experimental approaches to the study of bacterial phylogeny. *Am. Nat.* 97, 307–318.
- [138] Lanka, E. and Pansegrau, W. (1999) Genetic exchange between microorganisms. In: *Biology of the Prokaryotes* (Lengeler, J.W., Drews, G. and Schlegel, H.G., Eds.), pp. 386–415. Blackwell Science, Stuttgart.
- [139] Cohan, F.M. (1996) The role of genetic exchange in bacterial evolution. *ASM News* 62, 631–636.

- [140] Jones, D. (1989) Genetic methods. In: *Bergey's Manual of Systematic Bacteriology*, Vol. 4 (Williams, S.T., Sharpe, M.E. and Holt, J.G., Eds.), pp. 2310–2312. Williams and Wilkins, Baltimore, MD.
- [141] Woese, C. (1998) The universal ancestor. *Proc. Natl. Acad. Sci. USA* 95, 6854–6859.
- [142] Pennisi, E. (1998) Genome data shake tree of life. *Science* 280, 672–674.
- [143] Werman, S.D., Springer, M.S. and Britten, R.J. (1996) Nucleic acids I: DNA–DNA hybridization. In: *Molecular Systematics* (Hillis, D.M., Moritz, C. and Mable, B.K., Eds.), pp. 169–203. Sinauer Associates, Inc., Sunderland, MA.
- [144] Gibbons, E., Sneath, P.H.A. and Lapage, S.P. (1989) Reference collections of bacteria – the need and requirements for type strains. In: *Bergey's Manual of Systematic Bacteriology*, Vol. 4 (Williams, S.T., Sharpe, M.E. and Holt, J.G., Eds.), pp. 2325–2327. Williams and Wilkins, Baltimore, MD.
- [145] Pace, N.R., Stahl, D.A., Lane, D.J. and Olsen, G.J. (1985) Analyzing natural microbial populations by rRNA sequences. *ASM News* 51, 4–12.
- [146] Murray, R.G.E. and Schleifer, K.-H. (1994) Taxonomic note: a proposal for recording the properties of putative taxa of prokaryotes. *Int. J. Syst. Bacteriol.* 44, 174–176.
- [147] Kane, M.D., Poulsen, L.K. and Stahl, D.A. (1993) Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S rRNA sequences. *Appl. Environ. Microbiol.* 59, 682–686.
- [148] Amann, R., Ludwig, W. and Schleifer, K.-H. (1994) Identification of uncultured bacteria: a challenging task for molecular taxonomists. *ASM News* 60, 360–365.
- [149] Huber, R., Burgraff, S., Mayer, T., Barns, S.M., Rossmagel, P. and Stetter, K.O. (1995) Isolation of a hyperthermophilic archaeum predicted by in situ RNA analysis. *Nature* 376, 57–58.
- [150] Poulsen, L.K., Ballard, G. and Stahl, D.A. (1993) Use of rRNA fluorescence in situ hybridization for measuring activity of single cells in young established biofilms. *Appl. Environ. Microbiol.* 59, 1354–1360.
- [151] Rossello-Mora, R., Thamdrup, B., Schäfer, H., Weller, R. and Amann, R. (1999) The response of the microbial community of marine sediments to organic carbon input under anaerobic conditions. *Syst. Appl. Microbiol.* 22, 237–248.
- [152] Murray, R.G.E. and Stackebrandt, E. (1995) Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described prokaryotes. *Int. J. Syst. Bacteriol.* 45, 186–187.
- [153] Coyne, J.A., Allen Orr, H. and Futuyama, D.J. (1988) Do we need a new species concept? *Syst. Zool.* 37, 190–200.
- [154] Donoghue, M.J. (1985) A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist* 88, 172–181.
- [155] Mallet, M. (1995) A species definition for the modern synthesis. *Trends Ecol. Evol.* 10, 294–299.
- [156] Nelson, G. (1989) Species and taxa: systematics and evolution. In: *Speciation and its Consequences* (Otte, D. and Endler, J.A., Eds.), pp. 60–81. Sinauer Associates, Sunderland, MA.
- [157] Stackebrandt, E. (1992) Unifying phylogeny and phenotypic diversity. In: *The Prokaryotes*, Second Edition, Vol. 1 (Balows, A., Trüper, H.G., Dworkin, M., Harder, W. and Schleifer, K.-H., Eds.), pp. 19–47. Springer-Verlag, Berlin.
- [158] Sikorski, J., Rossello-Mora, R. and Lorenz, M.G. (1999) Analysis of genotypic diversity and relationships among *Pseudomonas stutzeri* strains by PCR-based genomic fingerprinting and multilocus enzyme electrophoresis. *Syst. Appl. Microbiol.* 22, 393–402.
- [159] Perez Canhos, V., Manfio, G.P. and Blaine, L.D. (1993) Software tools and databases for bacterial systematics and their dissemination via global networks. *Antonie van Leeuwenhoek* 64, 205–229.
- [160] Alm, R.A. et al. (1999) Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397, 176–180.
- [161] Holloway, B.W. and Morgan, A.F. (1986) Genome organization in *Pseudomonas*. *Ann. Rev. Microbiol.* 40, 79–105.
- [162] Krawiec, S. (1985) Concept of bacterial species. *Int. J. Syst. Bacteriol.* 35, 217–220.
- [163] Nandi, S., Khetawat, G., Sengupta, S., Majumder, R., Kar, S., Bhadra, R.K., Roychudhury, S. and Das, J. (1997) Rearrangements in the genomes of *Vibrio cholerae* strains belonging to different serovars and biovars. *Int. J. Syst. Bacteriol.* 47, 858–862.
- [164] Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H. and DeLong, E.F. (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* 178, 591–599.
- [165] Cowan, S.T. (1968) Oliver and Boyd Ltd., Edinburgh.
- [166] Lawrence, E. (1995) Addison Wesley Longman Limited, Essex.

## Glossary

(definitions extracted from: [165] and [166])

**Adansonian, Adansonism:** terms applied when equal weight is given to each character of an organism used in the construction of a classification. As a principle of classification, equal weighting is generally attributed to Adanson, a French naturalist who lived in 1727–1806.

**Analogous:** characters that are similar in function but not in structure, and developmental and evolutionary origin, e.g. the wings of insects and birds.

**Ancestor:** in taxonomy a taxon from which others are thought to have descended. Not a higher rank in a hierarchical system.

**Archaea:** or Archaeobacteria, are a heterogeneous group of prokaryotes which differ markedly from other prokaryotes (bacteria) in their 16S ribosomal RNA sequences and in other important characteristics of cellular composition.

**ARDRA:** amplified rDNA restriction analysis.

**Autapomorphy:** in phylogenetics an autapomorphy denotes a homologous character common to all members of a single taxa and thought to be exclusive of the group.

**Bacteria:** a diverse group of prokaryotes which differ from the Archaea in their 16S ribosomal RNA sequences and in other important characteristics of cellular composition.

**Bacteriological code:** the short title of the International Code of Nomenclature of Bacteria, approved at the Ninth International Congress of Microbiology, Moscow, 1966. The code includes principles, rules, recommendations and provisions (means by which the rules can be altered and the procedures for the conservation or rejection of names).

**Basionym, Basonym:** the name-bearing or epithet-bearing synonym which occurs in a new combination, e.g. *Bacillus coli* Migula is the basonym of *Escherichia coli* (Migula) Castellani and Chalmers.

**Binary system:** a system of naming in which binomial nomenclature is used for species.

**Binomial system:** a system of biological nomenclature in which a species is named by two words. The system is usually attributed to Linnaeus (though it has been used earlier by others), a Swedish naturalist who developed classifications of plants and animals, to the different species which he gave two names, one generic and the other specific; together formed the binomial.

**Biospecies:** biological species concept in which a species consists of interbreeding forms that are reproductively isolated from all other forms.

**Biotaxonomy:** classification of living things.

**Characterization:** the determination and listing of the characters of an organism to form a description.

**Circumscription:** statement defining the limits of a taxon, and showing by implication how it differs from similar taxa.

**Cladistics#**refers to a branch (Greek: Klados; branch) of an evolutionary lineage. Indicates the degree of relatedness, as shown by the pathways or phyletic lines by which taxa are linked. Method of classification of living organisms that makes use of lines of descent only to deduce evolutionary relationships, and which groups organisms strictly on the relative recency of common ancestry. Cladistic methods of classification only permit taxa in which all the members share a common ancestor.

**Classification:** is often confused with identification. Classification may be defined as the orderly arrangement of individuals into units composed of 'likes', each unit being homogeneous but different from every other unit. Classification is used both for the act and the result of the act.

**Classify:** the act of arranging objects into groups of similar objects; these groups can then be taken as units to which other objects may be referred. Sometimes loosely (and wrongly) used to mean identify.

**Dendrogram:** any branching tree-like diagram illustrating the relationship between organisms or objects.

**Description:** a list of characters of an organism or group by which subsequent workers will be able to recognize similar organisms or groups.

**Genome:** the genetic complement of a living organism or a single cell, more specially the total haploid genetic complement of a diploid organism or the total number of genes carried by a prokaryotic microorganism or a virus.

**Genotype:** genetic constitution of an organism, which acting together with environmental factors determines phenotype.

**Holotype:** the specimen (strain, isolate) designated by the original author of a name.

**Homology:** having a common origin. Homologous features are those that can be traced back to a feature in a common ancestor.

**Homonym:** the same name given to two or more different organisms; the adjective *senior* may be attached to the first published name, and *junior* to the same name (attached to a different organism) published later.

**Homoplasy, homoplasty:** resemblance in form or structure between different characters or organisms due to evolution along similar lines rather than a common descent; convergent evolution.

**Identification:** the act and result of a comparison in which an unknown is shown to be similar to (identical with) a known. Because organisms are never identical, some prefer the terms determine or determination.

**LFRFA:** low-frequency restriction fragment analysis.

**Monism:** philosophical theory that expresses that the plurality of the world can be explained in a single principle. In taxonomy, a monistic view is to think that a single level of organization exists across all organisms that deserves to be recognized as the species level.

**Monophyletic:** derived from a common ancestor. Taxa derived from and including a single founder species.

**Monothetic:** a classification that determines group membership according to the states of just one or few characters, but it may use different characters at different stages of the process.

**Nomenclature:** the scheme by which names are attached to organisms.

**Nomimifer:** nomenclatural type to which a name is permanently attached.

**Ontogenesis, ontogenetic:** the history of development and growth of an individual.

**Orthologous:** genes in different species that are homologous because they are derived from a common ancestral gene.

**OTU (operational taxonomic unit):** a convenient unit for the purpose in hand; it may be an individual, a population, a species or even a genus.

**Overall similarity:** defined by Sneath as the 'proportion of agreements between two organisms over the characters being studied'.

**Paralogous:** two genes that are similar because they derive from a gene duplication, e.g.  $\alpha$ - and  $\beta$ -globin.

**Paraphyletic:** groups which have evolved from and include a single ancestral species (known or hypothetical) but which do not contain all the descendants of that ancestor.

**PCR:** polymerase chain reaction, a molecular biological technique that allows the enzymatic amplification of a defined region of DNA.

**PFGE:** pulsed field gel electrophoresis.

**Phenetic:** Adj. Applied to a classification based on overall similarity, as determined by equal weighting of all known characters. Since all observable characters are used, a phenetic classification will make use of molecular genetic data if these are available. In contrast to phyletic and phylogenetic, the term phenetic does not have any evolutionary implications, other than in the

sense of showing the end product of evolution. Phenetic classifications may include phenotypic and genotypic characters.

*Phenogram*: dendrogram intended to show the phenetic relations of the organisms. A tree-like diagram showing the conclusions of numerical taxonomy.

*Phenome*: all the phenotypic characteristics of an organism determined by its genome.

*Phenon*: taxonomic group in which the degree of similarity is established by numerical methods. Group of organisms placed together by numerical taxonomy.

*Phenotype*: the visible or otherwise measurable physical and biochemical characteristics of an organism, a result of the interaction of genotype and environment.

*Phylogenetic tree*: a diagram setting out the genealogy of a species or other taxon.

*Phylogenetics*: approach to classification that attempts to reconstruct evolutionary genealogies and the historical course of speciation.

*Phylogeny*: the evolutionary history and line of descent of a species or higher taxonomic group. The expression of evolutionary relationships between organisms. Used as synonym of genealogy.

*Plesiomorphic*: the original pre-existing member of a pair of homologous characters. In phylogenetic reconstructions based on molecular data plesiomorphic sites are those false identities, identical residues at a particular alignment position which are treated a priori as evolutionarily identical independently if they have resulted from multiple base changes during the course of evolution.

*Polyphasic taxonomy*: term used to signify successive or simultaneous taxonomic studies of a group of organisms using an array of techniques designed to yield both molecular and phenotypic data.

*Polyphyletic*: a taxonomic group having origin in several different lines of descent.

*Polythetic*: a classification based on many characters, not all of which are necessarily shown by every member of the group.

*Prokaryote*: the etymology of the word indicates the absence of a true nucleus, separated from the cytoplasm by a nuclear membrane. Prokaryotes lack additional membrane-containing structures characteristic of eukaryotes (e.g. chloroplasts and mitochondria) and unlike the latter have ribosomes with a sedimentation constant of 70S. Introns are rare in prokaryotes and DNA is usually present in a single molecule. The prokaryotes comprise two domains, *Bacteria* and *Archaea*.

*RAPD*: randomly amplified polymorphic DNA.

*RFLP*: restriction fragment length polymorphism.

*Synapomorphy*: in cladistic phylogenetics denotes a homologous character common to two or more taxa and thought to have originated in their most recent common ancestor.

*Systematics*: a plural adjective used as a noun to embrace the many ways used for the study of organisms with the ultimate object of characterizing and arranging them in an orderly manner. Systematics include not only taxonomy but also disciplines like ecology, biochemistry, genetics...

*Taxon*: refers to any taxonomic group, but to be distinguished from *category* which indicates the rank of a group.

*Taxonomy*: the theoretical study of classification including its bases, principles, procedures and rules.

*Wild-type*: a term used to describe the microorganism as it exists in the nature.