Available online at www.sciencedirect.com
ScienceDirect
Systematic and Applied Microbiology 32 (2009) 533–542
SYSTEMATIC AND APPLIED MICROBIOLOGY
www.elsevier.de/syapm

MINIREVIEW

Classification of *Bacteria* and *Archaea*: Past, present and future ☆

Karl Heinz Schleifer*

Department of Microbiology, Technical University Munich, Am Hochanger 4, D-85350 Freising, Germany

Received 15 July 2008

Abstract

The late 19th century was the beginning of bacterial taxonomy and bacteria were classified on the basis of phenotypic markers. The distinction of prokaryotes and eukaryotes was introduced in the 1960s. Numerical taxonomy improved phenotypic identification but provided little information on the phylogenetic relationships of prokaryotes. Later on, chemotaxonomic and genotypic methods were widely used for a more satisfactory classification. *Archaea* were first classified as a separate group of prokaryotes in 1977. The current classification of *Bacteria* and *Archaea* is based on an operational-based model, the so-called polyphasic approach, comprised of phenotypic, chemotaxonomic and genotypic data, as well as phylogenetic information. The provisional status *Candidatus* has been established for describing uncultured prokaryotic cells for which their phylogenetic relationship has been determined and their authenticity revealed by *in situ* probing.

The ultimate goal is to achieve a theory-based classification system based on a phylogenetic/evolutionary concept. However, there are currently two contradictory opinions about the future classification of *Bacteria* and *Archaea*. A group of mostly molecular biologists posits that the yet-unclear effect of gene flow, in particular lateral gene transfer, makes the line of descent difficult, if not impossible, to describe. However, even in the face of genomic fluidity it seems that the typical geno- and phenotypic characteristics of a taxon are still maintained, and are sufficient for reliable classification and identification of *Bacteria* and *Archaea*. There are many well-defined genotypic clusters that are congruent with known species delineated by polyphasic approaches. Comparative sequence analysis of certain core genes, including rRNA genes, may be useful for the characterization of higher taxa, whereas various character genes may be suitable as phylogenetic markers for the delineation of lower taxa. Nevertheless, there may still be a few organisms which escape a reliable classification.

Keywords: *Bacteria*; *Archaea*; History; Numerical taxonomy; Chemotaxonomy; Genotypic characterization; Concepts of classification; Polyphasic approach; Candidatus; Genome structure

Introduction

A reliable classification system is a prerequisite for scientists and professionals dealing with microorganisms in order to keep track of their tremendous variety. The ultimate objective of biological classification is the

☆ This article is based on the Lwoff Award Lecture 2009 presented at the 3rd European Congress of Microbiology in Gothenburg, Sweden, June 28–July 2, 2009 and a lecture presented at the symposium "Recent Advances in Microbial Taxonomy" by the "Deutsche Akademie der Naturforscher Leopoldina" in Zürich, Switzerland, March 30–31, 2009.

*Tel.: +49 8161 71 5441; fax: +49 8161 71 5475.
E-mail address: schleife@mikro.biologie.tu-muenchen.de

0723-2020/$ - see front matter © 2009 Elsevier GmbH. All rights reserved.
doi:10.1016/j.syapm.2009.09.002

characterization and orderly arrangement of organisms into groups. Classification is often confused with identification but, as a matter of fact, classification is a prerequisite for identification.

Currently, there is no official classification of *Bacteria* and *Archaea* available. Many bacteriologists think that Bergey's Manual represents the official classification but this is a misunderstanding. The editors of Bergey's Manual try to provide a classification that is as accurate and up-to-date as possible but it is not official, in contrast to bacterial nomenclature where each taxon has one valid name. The closest to an official classification system is the one that is widely accepted by the community [5].

History of classification

The history of the classification of bacteria clearly demonstrates that changes were caused by the availability of new techniques (Table 1). The late 19th century was the beginning of bacterial taxonomy and Ferdinand Cohn in 1872 [11] was the first to classify six genera of bacteria (as members of the plants) mainly based on their morphology. However, at that time, the majority of scientists were interested in the description of pathogenic bacteria. Actually, many of the pathogenic bacteria known today were described between 1880 and 1900. At that time, besides morphology, growth requirements and pathogenic potential were the most important taxonomic markers [24].

At the beginning of the 20th century more and more physiological and biochemical data were used, in addition to morphology, as important markers for the classification and identification of microorganisms. Numerous biochemical and physiological properties of bacterial cultures were determined for their characterization and identification. Later, enzymes were studied and metabolic pathways were elucidated. The first edition of Bergey's Manual of Determinative Bacteriology [3] classified the *Bacteria* in 1923 on the basis of these phenotypic properties as "typically unicellular plants", the so-called *Schizomycetes*. Even in the 7th edition of Bergey's Manual [4], published in 1957, *Bacteria* were still classified as members of plants (*Protophyta*, primitive plants). Based on the partial sequences of 16S ribosomal RNA (rRNA) genes, *Archaea* (originally named *Archaebacteria*) were first classified as a separate kingdom in 1977 [45].

Table 1. History of the classification of *Bacteria* and *Archaea*.

Time span	Classification mainly based on	References
Late 19th century	Morphology, Growth Requirements, Pathogenic potential	[11,24]
1900–1960	Morphology, Physiology, Biochemistry	[3,6]
1960–1980	Chemotaxonomy, Numerical Taxonomy, DNA–DNA Hybridization	[7,32,34,37]
1980–today	Genotypic Analyses, Multilocus Sequence Analyses, Average Nucleotide Identity, Whole Genome Analysis	[16,18,22,31, 39,44,46,47]

The French protistologist Edouard Chatton, the mentor and long time colleague of A. Lwoff, mentioned for the first time in 1925 [9] the two categories prokaryotes and eukaryotes but only to distinguish prokaryotic from eukaryotic protists. However, his proposal did not become generally known. Later on, A. Lwoff propagated this distinction and finally convinced R. Stanier, together with C.B. van Niel in 1962, to describe a detailed and well-accepted division of prokaryotic (bacteria) and eukaryotic (animals, plants) organisms [42].

In the 8th edition of Bergey's Manual, which was published in 1974 [7], *Bacteria* were no longer considered as plants and were recognized as members of the kingdom *Procaryotae*. However, all former ideas about phylogeny and relationships were discarded and bacteria were arranged in groups based mainly on the Gram-stain, morphology and oxygen requirement. A typically bad example for the classification of phenotypically similar but genetically quite different bacteria is the treatment of the family *Micrococcaceae* in this 8th edition. The two genera of this family, *Micrococcus* and *Staphylococcus*, are definitely not related (see below).

Numerical taxonomy based on phenetic analyses

Numerical taxonomy improved phenotypic identification by increasing the number of tests used and by calculating the coefficients of phenetic similarities between strains and species [37]. For numerical studies, the results are tabulated in a table of t organisms versus n characters and the term OTU (operational taxonomic unit) is used for an individual strain. The characters are equally weighted and should come from the various different categories of properties (morphology, physiology, biochemistry, etc.). The number of common characteristics is considered as a quantitative measure of taxonomic relatedness, although this does not mean that the organisms are also phylogenetically related.

Chemotaxonomy

The chemical composition of cell constituents is a useful property for improving the classification and

identification of prokaryotes. Chemotaxonomic methods are widely used, in particular, for those groups of prokaryotes where morphological and physiological characters have largely failed or have not been sufficient to provide a satisfactory classification [34].

The **DNA base composition**, guanine–cytosine (GC) content of DNA, is one of the required characteristics on the minimum list of data needed for the description of a new genus. However, it is only an exclusionary determinant in the classification of bacteria, in that two strains differing by more than 10 mol% should not be considered as members of the same genus, whereas, on the other hand, a similar DNA base composition does not necessarily imply that the two strains are closely related. In practice, it has proved to be a valuable character for distinguishing between non-related bacteria, such as staphylococci (30–35 mol% GC) and micrococci (70–75 mol% GC).

The occurrence of alkyl glycerol ether **lipids** instead of fatty acid ester lipids is a very characteristic property of *Archaea* and can be used for distinguishing them from *Bacteria* and *Eukarya*. *Bacteria* contain a wide variety of fatty acids (unbranched or branched fatty acids, hydroxy fatty acids, cyclopropane fatty acids, saturated and unsaturated ones). Most fatty acids of bacteria are in the range of C_{12} to C_{20}. Fatty acid patterns can be determined rather easily and quickly, and automatic identification is even possible. However, the bacteria have to be cultivated under carefully controlled conditions since fatty acid patterns may alter in response to exogenous and endogenous parameters, such as growth temperature, pH composition of the medium, or age of the culture.

Isoprenoid quinones play an important role in electron transport. Different bacteria not only synthesize different quinone types (ubiquinone, menaquinone, demethylmenaquinones) but, in particular, the length and the degree of saturation of polyprenyl side chains are of considerable value in classification [12]. The cyanobacteria contain neither ubiquinones nor menaquinones but phylloquinones and plastoquinones, which are normally associated with green plants. Most strictly aerobic, Gram-negative bacteria produce only ubiquinones, whereas facultatively anaerobic, Gram-negative bacteria additionally contain menaquinones and/or demethylmenaquinones. Aerobic and facultatively anaerobic, Gram-positive bacteria produce only menaquinones. Strictly anaerobic bacteria lack isoprenoid quinones or contain only menaquinones.

Bacterial **cytochromes** are involved in a wide variety of redox processes, such as aerobic and anaerobic respiration and photosynthetic electron transfer. Most cytochromes are associated with the cytoplasmic membrane. Cytochrome c is often absent in both Gram-negative and Gram-positive bacteria, and enterobacteria can be easily separated from pseudomonads since the former do not contain cytochrome c and are therefore oxidase negative. The cytochrome pattern is also helpful for distinguishing staphylococci (which usually lack cytochromes c and d) from micrococci.

The ultrastructure and chemical composition of the **cell walls** of Gram-positive and Gram-negative bacteria are quite different. In profile, the cell wall of Gram-positive bacteria reveals a single thick and more or less homogeneous layer, whereas Gram-negative bacteria have a thinner, distinctly layered cell wall with an outer membrane resembling the typical trilaminar cytoplasmic membrane. The polymers found in the cell walls of these two groups of bacteria are chemically quite different. The walls of **Gram-negative** cells are mainly composed of lipopolysaccharide, phospholipid, protein, lipoprotein, and relatively little peptidoglycan (usually less than 10% of the total cell wall). The **Gram-positive** cells contain peptidoglycan (usually more than 30% of the total cell wall), polysaccharides or teichoic acid (or both), or teichuronic acid, as major components. Thus, in contrast to the Gram-negative bacteria, the Gram-positive bacteria contain hardly any lipids in their cell walls. There is, however, one exception: acid-fast bacteria. They are resistant to decolorization with acidic ethanol after staining with fuchsin (Ziehl-Nielsen staining). These acid-fast bacteria (*Mycobacterium*, *Nocardia*, and *Corynebacterium sensu stricto*) are Gram-positive bacteria which contain large amounts of lipids in their cell walls; in particular, mycolic acids (high-molecular-weight, 3-hydroxy acids with a long alkyl branch in position 2).

Peptidoglycan (murein) is the only cell wall polymer found in both Gram-positive and Gram-negative bacteria. It is a heteropolymer consisting of glycan strands that are cross-linked through short peptides. The glycan strand is made up of alternating ß-1,4-linked residues of N-acetylglucosamine and N-acetylmuramic acid, a derivative of glucosamine and the unique constituent of peptidoglycan. The peptide moiety is linked to N-acetylmuramic acid and contains both L- and D-amino acids. The peptidoglycan of Gram-negative bacteria is remarkably uniform [32]. Gram-positive bacteria contain a multilayered peptidoglycan and reveal, in contrast to Gram-negative organisms, a great variation in the chemical composition of their peptidoglycans. Minor variations can be found in the glycan strand. In mycobacteria and nocardia, the N-acetyl group of muramic acid is oxidized to N-glycolyl. The greatest variation occurs within the peptide moiety of the peptidoglycan. Both the peptide subunit (stem peptide) and the interpeptide bridges, which cross-link the peptide subunits, can vary in their composition and primary structure. Based on the mode of cross-linkage, two main groups of cross-linkage, A and B, can be distinguished [32]. A given peptidoglycan structure is a fairly stable character and fulfils the most important

prerequisites of a useful taxonomic marker. No single-step mutations are known so far that lead to an altered peptidoglycan structure. Phenotypic variations are also rather limited and can be easily controlled [33].

Chemotaxonomic data are very useful for reliable classification and identification of *Bacteria* and *Archaea* but they are not sufficient for a comprehensive reconstruction of their phylogeny. Therefore, it is not surprising that Stanier et al. [41] wrote in their textbook "…for bacteria, the general course of evolution will never be known, and there is simply not enough objective evidence to base their classification on phylogenetic grounds".

However, in the late 1970s and the beginning of the 1980s a breakthrough was achieved by Carl Woese and co-workers when they were able to derive a tree of life consisting of three distinctly different branches (*Bacteria*, *Archaea* and *Eukarya*) by comparing first partial and later on complete small subunit rRNA gene sequences [45,46]. This study revolutionized bacterial taxonomy and for the first time bacteriologists were able to classify prokaryotes on the basis of their phylogenetic relatedness. In subsequent years, further genotypic studies, including comparative large subunit rRNA gene as well as protein-coding gene sequence analyses, allowed an even better insight into the relationships of prokaryotes. More recently, multilocus sequence analysis (MLSA) has been applied for distinguishing between closely related bacterial species.

Concepts of classification

In biology there are two quite different concepts for the classification of the organisms available. The **theory-based model** that is, for example, comparable to the biological species concept proposed by Ernst Mayr [25]. It is mainly applied to animals and some plants. It is based on an explicit and predictive theory of the mechanisms of speciation. Species are defined as groups of interbreeding natural populations that are reproductively isolated from other such groups. Such a model is currently not readily identifiable for microorganisms because there is a lack of understanding of the selection and diversity processes that may drive the creation of microbial species. A theory-based model for the classification of microorganisms would also have to take into account the effect of chemical, physical and functional boundaries on speciation. It should be based on cohesive evolutionary forces. A distinct taxonomic cluster should have evolved separately from other lineages.

Very recently, theory-based concepts have been proposed for the classification of bacteria, for instance, the **ecotype-based approach** [20] and the **metapopulation concept** [1]. However, both the approaches have severe drawbacks. From a conceptual standpoint, ecotype boundaries can only be established by elucidating the ecological niche of a strain. Moreover, some species encompass, on the one hand, multiple ecotypes and, on the other hand, there are also single ecotypes that can comprise multiple clusters of genotypes. Gene swapping is so frequent among prokaryotes that recombination, and not ecological adaptation, is the main cause of diversity [29]. The question is whether ecotypes are really true biological groupings.

The metapopulation concept is not suited for a pragmatic classification system that can also serve as a tool in routine identification. It reveals a severe weakness, namely that there is currently no known methodology that correlates most strongly with this concept.

The **operational-based model**, on the other hand, is not based on a unifying theory of speciation but depends on data-driven analyses. A set of criteria and cut-off levels for delineating groups that share functional and phylogenetic similarities is empirically defined. For example, groups of strains that show more than approximately 70% DNA–DNA similarity values and/or less than a 5% difference in the melting temperature of their DNA–DNA hybrids are considered to belong to the same species.

There are two quite different approaches for the operational-based classification of bacteria. On the one hand, the **phenotypic** approach, which is based on similarities of morphological, physiological and chemotaxonomical properties. This is especially useful for routine identification but provides only limited information on their phylogenetic relationship. On the other hand, the **genotypic** approach, which is based on genetic relatedness, deduced mainly from DNA–DNA hybridization (DDH) studies and comparative sequence analyses of homologous macromolecules (e.g. rRNA). This approach provides insight into the (phylo)genetic relatedness of microorganisms. However, in contrast to higher organisms, it is not possible to infer the natural relationships of microorganisms from fossil records and from comparative studies on their ontogeny and morphology, respectively.

Genotypic classification

16S, as well as 23S, **rRNA gene sequence** comparisons are currently considered as the gold standard for deducing the phylogenetic relationships of prokaryotes (Table 2). These genes fulfil all the properties of useful molecular markers (ubiquitous, functionally constant, conserved, homologous). They are stable markers and less subjected to lateral gene transfer. There is also a

Table 2. Advantages and problems of comparative 16S rRNA gene sequence analysis.

Advantages	Problems
rRNAs fulfil all properties of useful molecular markers (ubiquitous, functionally constant, conserved).	The resolution at the species level is often not sufficient, since the gene is too conserved.
rRNAs are stable markers. They are less subjected to lateral gene transfer.	
Good congruence for branching pattern of phylogenetic trees derived from conserved, mostly informational, genes involved in translation and transcription, respectively	Multiple 16S rRNA genes exist (showing in most cases a sequence divergence range of 1–2%, but sometimes even higher).
Genome-based studies are in good agreement with the rRNA data.	
Facilitates identification of uncultivated prokaryotes.	At the phylum level, it is often difficult to organize relative branching orders.

Table 3. Advantages and drawbacks of DNA–DNA hybridization studies.

Advantages	Drawbacks
Current reference standard for species delineation.	Describe only a rough measurement of average genetic relationship.
Strains with DDH values of more than 70% under standardized conditions and/or a difference in melting temperature of less than 5% belong to the same species.	Only closely related species or subspecies can be distinguished (above 90% genome similarity).
Strains with DDH similarity values of less than 70% correlate well with 16S rRNA. Strains with sequence identity of 98.7%, or lower, can be generally considered as members of different species.	Methods are rather tedious and time-consuming and incremental databases cannot be developed. Not applicable for uncultivated bacteria.

good congruence for the branching pattern of phylogenetic trees derived from conserved, mostly informational, genes involved in translation (e.g. EF-Tu) and transcription (rpoB, rpoC), respectively. Moreover, genome-based studies are consistent with the rRNA data [10]. A curated All-Species Living Tree, reconstructed from a 16S rRNA tree comprising all sequenced type strains of validly published species of *Bacteria* and *Archaea* up to the end of 2007, has been published [47] and will be updated soon. 16S rRNA gene sequencing has also been widely used in the identification of currently non-cultivable prokaryotes occurring in the environment.

However, species definition using rRNA sequences is often not possible because the molecule is too conserved to distinguish between closely related species (Table 2). Prokaryotes with 98.7%, or lower, 16S rDNA sequence identity can be considered as members of different species, because such differences in rRNA correlate well with **DDH similarity** values of less than 70% [38]. However, the opposite is not necessarily true because in a very few cases distinct species have been described that share 16S rDNA sequence identity of more than 98.7% [15]. Relationships inferred from 16S rRNA genes may also be distorted by the presence of multiple 16S rRNA genes in most of the prokaryotes. They show, in many cases, a sequence divergence range of 1–2%, and, in a few cases, even higher. At the phylum level, it is also difficult to organize relative branching orders based on rRNA sequence comparisons because the power of resolution is not sufficient. For species delineation, DDH was not proposed as a ''gold standard'' but as a reference standard [39]. It has its weaknesses but it cannot be replaced until a better method becomes available (Table 3).

Current classification

It has to be emphasized that there is still no officially recognized system for the classification of prokaryotes. The currently applied classification systems rely – for practical reasons – on methods and do not depend on theoretical concepts. The most widely used system is the so-called **polyphasic approach** [43]. This approach includes phenotypic, chemotaxonomic and genotypic data, as well as phylogenetic information. 16S rRNA gene sequencing is applied for determining the phylogenetic position of the organisms. Based on these results, organisms are selected for DDH studies and species are defined using the 70% DDH cut-off criterion [44]. Each taxon should be described and differentiated from related taxa by its phenotypic, chemotaxonomic and

Table 4. Strengths and weaknesses of multilocus sequence analysis.

Strengths	Weaknesses
Multiple genes provide more informative nucleotide sites and buffers against the distorting effects of recombination of one of the loci.	Selection of a proper set of genes is not clear. Different sets of genes are required for the classification of different taxa and lineages. One set cannot fit all.
House-keeping genes are essential and evolve relatively slowly but faster than rRNA genes.	Designing of primers facilitating amplification of genes in all strains is often difficult or impossible.
Resolves at lower taxonomic levels than rRNA.	Even concatenated sequences of 7–12 genes correspond only to a minor fraction of the genome.
Can be combined with rRNA data.	Depth of the clustering that defines a taxon is not clear.
Correlates well with classical species definitions (e.g. *Burkholderia*, *Streptococcus*).	Clustering may not occur or may include different groups of strains.
Method is amenable to automation and large curated databases can be developed.	Cannot be applied to uncultivated bacteria.

genotypic characteristics. Distinguishing phenotypic differences are required for the description of a new species. If such differences are not found, groups of similar bacteria that appear to be genetically distinct have to be described by other terms (e.g. genospecies/genomovars). It should be kept in mind that the end-users need a pragmatic classification system that can serve as a tool in routine identification. A classification that is of little use to microbiologists, no matter how sophisticated a scheme is, will soon be ignored or significantly modified [40]. The polyphasic approach has proven its value in bacterial taxonomy, but apparently it is still not satisfactory to many end-users [16]. It cannot cope with the huge microbial diversity that remains to be revealed. Therefore, some end-users avoid the time-consuming and tedious descriptive classification and prefer just a single-step phylogenetic taxonomy on the basis of 16S rRNA gene sequencing.

More recent genotypic classification approaches

Attempts are underway to test whether the data from genome comparison can be used for taxonomic purposes. Goris et al. [18] have compared completely sequenced genomes and their corresponding hybridization values. Pairwise comparison of complete genome sequences showed that the **average nucleotide identity** (ANI) of all conserved genes between any two genomes correlated well with 16S rRNA sequence identity and DNA–DNA similarity values. It has also been shown that 70% DNA–DNA similarity corresponds to 95% ANI [21]. Moreover, all pairs of genomes showing 95%, or higher, ANI also showed at least 98.5% 16S rRNA gene identity. ANI was found to be the genome-derived parameter that most correlated with DDH and may in future be able to substitute the tedious DDH method [22,31].

Recently, **multilocus sequence analysis** (MLSA) has been proposed as a replacement for DDH in the classification of prokaryotes [16]. MLSA is a method for the genotypic characterization of a diverse group of prokaryotes by comparing sequences of multiple house-keeping genes. Multiple genes provide more informative nucleotide sites and buffers against the distorting effects of recombination of one of the loci. The best approach is to concatenate the sequences of at least 12 genes from a set of strains and to use the concatenated sequences to reconstruct a phylogenetic tree which can identify deeply branching clusters and help to delineate genotypic clustering within a genus or species [35]. However, the selection of a proper set of genes is not clear and has not been systematically explored. The same holds true for the size of sequenced fragments of each gene. A different set of genes is often necessary for different groups of organisms and it is often difficult or even impossible to design primers facilitating amplification of genes in all strains (see also Table 4).

Provisional classification of uncultured prokaryotes

The provisional status ***Candidatus*** has been established for certain putative taxa that could not be described in sufficient detail as a novel taxon [26,27]. The designation *Candidatus* is not a rank but a status that is currently not formally recognized in the *International Code of Nomenclature of Bacteria*. The category *Candidatus* should be used for describing uncultured prokaryotic cells for which their phylogenetic relatedness has been determined and their authenticity revealed by *in situ* probing (e.g. fluorescence *in situ* hybridization, [2]) or similar techniques. In addition to the genomic information, all phenotypic information, including structural, metabolic, physiological and reproductive features, should be included in the description (Table 5).

Table 5. Characteristics recommended for the description of the category *Candidatus*.

Characteristics	Examples
Phylogenetic information	Comparative sequence analysis, e.g. 16S rRNA; *gyrA*, *recA*; metagenomic data
Morphology and Gram reaction	Coccus, rod, filament, etc., Gram staining
Specific identification	Nucleic acid probe, FISH, genomic data
Habitat and/or source	Free-living, symbiotic, syntrophic
Physiology, Metabolism	Aerobic, anaerobic, electron acceptor, electron donator, unusual metabolism
Growth temperature	Meso-, thermo-, psychrophilic

Identification by *in situ* hybridization or other similar techniques for cell identification should be performed in their natural environment. Stackebrandt et al. [39] encouraged bacteriologists to use the "*Candidatus*" concept for well-characterized but as yet uncultured organisms. The names included in the category *Candidatus* should be written as follows: *Candidatus* in italics and the subsequent name(s) in roman type with an initial capital letter for the genus name. The entire name should be in quotation marks (e.g. "*Candidatus* Magnospira bakii"). The Judicial Commission of the International Committee on Systematics of Prokaryotes decided that the concept *Candidatus* should be mentioned in the main body of the Bacteriological Code, despite the fact that such names have no standing in nomenclature [14]. Currently, there are more than 200 bacteria and *Archaea*, respectively, described as *Candidatus*. They are, in particular, endosymbionts or parasites of eukaryotes, many of them belonging to the phyla *Mollicutes*, *Chlamydiales* and *Rickettsiales*, respectively. However, there are also other not yet cultured prokaryotes, such as special enrichments or organisms living in co-cultures (e.g. syntrophic organisms) or occurring in unusual habitats, revealing unusual metabolic properties (Anammox, Fe-oxidizing bacteria) that have been described as *Candidatus*. The information for a description of a *Candidatus* is summarized in Table 5.

Classification in the age of genomics

It is now apparent that prokaryotes reveal a mosaic genome structure. The bacterial genome consists of three pools of genes. The first gene pool consists of a conserved core of essential genes common to all genomes of a phylogenetically coherent group of bacteria. They make up the so-called core genome. The **core genome** of bacteria preferentially contains informational or house-keeping genes that are rather stable and less prone to lateral gene transfer. Studies on almost 300 bacterial genomes have shown that these shared genes are present in at least 99% of these genomes. They belong to about 250 gene families and constitute about 8% of an average bacterial genome [23]. The second gene pool, consisting of the so-called **character or lifestyle genes**, is essential for colonization, survival and adaptation to a particular environment. Specific metabolic properties are often controlled or coded by these genes. Character genes have been found to be the main component (64%) of the 300 compared bacterial genomes. However, they belong to only 7900 gene families, whereas the genes of the third pool, the so-called **accessory genes** represent about 28% of an average bacterial genome but belong to almost 140,000 gene families [23]. These accessory genes are non-essential, less conserved and often strain specific. They show high turnover rates and some of them have descended from bacteriophages and plasmids, respectively. The sum of all genes found within the various strains of a distinct taxonomic cluster is known as the **pan-genome**. Recent findings indicate that the pan-genome is probably of infinitive size (i.e. the domain *Bacteria* as a whole has an open pan-genome).

Future prospectives

The ultimate goal is to achieve a theory-based classification system based on a phylogenetic/evolutionary concept that also provides the basis for a pragmatic, reliable identification system. Links between phylogeny and phenotypic features have to be found and a distinct taxon should be characterized by genomic and phenetic coherence. Moreover, data from metagenomic studies may be helpful for improving the classification of uncultured bacteria.

However, there are currently two contradictory opinions about the future classification of bacteria: a pessimistic and a more optimistic perspective, respectively (Table 6). There is a group of mostly molecular biologists who doubt that evolutionary processes always lead to clearly resolved clusters of organisms which derive from a common ancestor. The reason for the rather pessimistic perspective is that the yet-unclear effect of gene flow, in particular lateral gene transfer (LTG), makes the line of descent difficult, if not impossible, to describe [13] and may erase the boundaries between species or any other taxonomic units [17]. Papke [29] has come to the conclusion that "the only recourse is to adopt species concepts that

Table 6. Future prospectives for the classification of *Bacteria* and *Archaea*.

Pessimistic view	Optimistic view
Frequent gene swapping between different taxa makes lines of descent difficult, if not impossible, to describe.	Gene flow is more pronounced among closely related organisms and comprises mainly genes that are not suitable candidates for phylogenetic analysis.
Non-arbitrary classification of bacteria is not possible because of lack of evidence for discrete clusters having a common ancestor.	16S rRNA-based phylogeny of higher taxa is in good agreement with analyses retrieved from genomic approaches.
Gene flow, in particular LGT, may create patterns of similarity that mimic patterns produced by vertical descent.	Well-resolved genotypic clusters are congruent with species assigned by a polyphasic approach.
	Comparative genomic analyses may disclose characteristic genes for classification of bacteria (e.g. core genes for higher taxa and character genes for delineation of species).
There is no real species concept and it will always be a compromise solution. In some cases, there may be monophyletic clusters, in others not.	A theory-based classification system should be based on a phylogenetic/evolutionary concept, it has to be pragmatic and supported by phenotypic features.

require pragmatic and subjective approaches which do not necessitate species to be monophyletic groups''. Recently, it has also been reported that changes in gene flow caused by ecological factors may even lead to an incipient merging of two bacterial species, *Campylobacter jejuni* and *C. coli*, reversing the process of speciation [36]. After 10,000 years of shared adaptation to a particular habitat the genome of *C. coli* is becoming ever more like that of *C. jejuni* and the two may eventually become indistinguishable.

Despite this pessimistic view of some scientists on a reliable classification of prokaryotes, it is quite clear that gene swapping is more pronounced between closely related organisms, and mainly comprises genes that are not suitable candidates for phylogenetic analysis [28,30]. There are also other indications that support the feasibility of a stable classification of bacteria. There are many discrete well-defined genotypic clusters that are congruent with known species delineated by polyphasic approaches [19,22]. Even in the face of substantial genomic fluidity it seems that the typical geno- and phenotypic characteristics of a taxon are still maintained and, as shown in many cases, are sufficient for a reliable classification and identification of bacteria [30].

The rather conserved and ubiquitous genes of the core genome are potential phylogenetic markers for the genotypic classification of less related prokaryotes. Thus, comparative sequence analysis of certain core genes, including rRNA genes, may be useful for retrieving the phylogenetic relationships of higher taxa. For the classification of lower taxa, such as the delineation of species, various character genes may be suitable as phylogenetic markers. In particular, genes that code for key phenotypic differences may play an important role in the improved classification of bacteria. Moreover, core and/or character genes coding for functional proteins detected in the proteome could provide some information on the characteristic phenotypic features [8] and they can be used for defining a taxon. However, it has to be stressed that different sets of genes and phenotypic traits are required for the classification of different taxa and lineages, and there may also be some bacteria which escape a reliable classification because they do not form coherent clusters.

References

[1] M. Achtman, M. Wagner, Microbial diversity and the genetic nature of microbial species, Nat. Rev. Microbiol. 6 (2008) 431–440.
[2] R.I. Amann, W. Ludwig, K.H. Schleifer, Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation, Microbiol. Rev. 59 (1995) 143–169.
[3] D.H. Bergey, F.C. Harrison, R.S. Breed, B.W. Hammer, F.M. Hantoon, (Eds.), Bergey's Manual of Determinative Bacteriology, 1st ed., The Williams and Wilkins Co., Baltimore, USA, 1923.
[4] R.S. Breed, E.G.D. Murray, N.R. Smith (Eds.), Bergey's Manual of Determinative Bacteriology, 7th ed, The Williams & Wilkins Co, Baltimore, USA, 1957.
[5] D.J. Brenner, J.T. Staley, N.R. Krieg, Classification of procaryotic organisms and the concept of bacterial speciation, in: 2nd ed, in: D.R. Boone, R.W. Castenholz, G.M. Garrity (Eds.), Bergey's Manual of Systematic Bacteriology, Vol. 1, Springer, New York, USA, 2001, pp. 27–31.
[6] R.E. Buchanan, Taxonomy, Ann. Rev. Microbiol. 9 (1955) 1–20.
[7] R.E. Buchanan, N.E. Gibbons, (Eds.), Bergey's Manual of Determinative Bacteriology, 8th ed., The Williams and Wilkins Co., Baltimore, USA, 1974.
[8] S.J. Callister, L.A. McCue, J.E. Turse, M.E. Monroe, K.J. Auberry, R.D. Smith, J.N. Adkins, M.S. Lipton,

Comparative bacterial proteomics: analysis of the core genome concept, PLoS ONE 3 (2008) e1542.
[9] E. Chatton, Pansporella perplexa, Ann. Sci. Nat. Zool. 8 (1925) 5–85.
[10] F.D. Ciccarelli, T. Doerks, C. von Mering, C.J. Creevey, B. Snel, P. Bork, Toward automatic reconstruction of a highly resolved tree of life, Science 311 (2006) 1283–1287.
[11] F. Cohn, Grundzüge einer neuen natürlichen Anordnung der kryptogamischen Pflanzen, Jahresb. Schles. Ges. Vaterl. Kultur 49 (1872) 83–89.
[12] M.D. Collins, D. Jones, Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implications, Microbiol. Rev. 45 (1981) 316–354.
[13] W.F. Doolittle, E. Bapteste, Pattern pluralism and the tree of life hypothesis, Proc. Natl. Acad. Sci. USA 104 (2007) 2043–2049.
[14] P. De Vos, H.G. Trüper, B.J. Tindall, Judicial Commission of the International Committee on Systematics of Prokaryotes. Xth International (IUMS) Congress of Bacteriology and Applied Microbiology. Minutes of the Meetings, Int. J. Syst. Evol. Microbiol. (2005) 525–532.
[15] G.E. Fox, J.D. Wisotzkey, P. Jurtshuk Jr, How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity, Int. J. Syst. Bacteriol. 42 (1992) 166–170.
[16] D. Gevers, P. Dawyndt, P. Vandamme, A. Willems, M. Vancanneyt, J. Swings, P. de Vos, Stepping stones towards a new prokaryotic taxonomy, Phil. Trans. R. Soc. London B 361 (2006) 1911–1916.
[17] J.P. Gogarten, W.F. Doolittle, J.G. Lawrence, Prokaryotic evolution in the light of gene transfer, Mol. Biol. Evol. 19 (2002) 2226–2238.
[18] J. Goris, K.T. Konstantinidis, J.A. Klappenbach, T. Coenye, P. Vandamme, J.M. Tiedje, DNA–DNA hybridization values and their relationship to whole-genome sequence similarities, Int. J. Syst. Evol. Microbiol. 57 (2007) 81–91.
[19] W.P. Hanage, C. Fraser, B.G. Spratt, Sequences, sequence clusters and bacterial species, Philos. Trans. R. Soc. B 361 (2006) 1917–1927.
[20] A. Koeppel, E.B. Perry, J. Sikorski, D. Krizanc, A. Warner, D.M. Ward, A.P. Rooney, E. Brambilla, N. Connor, R.M. Ratcliff, E. Nevo, F.M. Cohan, Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics, Proc. Natl. Acad. Sci. USA 105 (2008) 2504–2509.
[21] K.T. Konstantinidis, J.M. Tiedje, Genomic insights into the species definition for prokaryotes, Proc. Natl. Acad. Sci. USA 102 (2005) 2567–2572.
[22] K.T. Konstantinidis, J.M. Tiedje, Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead, Curr. Opin. Microbiol. 10 (2007) 504–509.
[23] P. Lapierre, J.P. Gogarten, Estimating the size of the bacterial pan-genome, Trends Genet. 25 (2009) 107–110.
[24] K.B. Lehmann, R.O. Neumann, Atlas und Grundriss der Bakteriologie und Lehrbuch der speciellen bakteriologischen Diagnostik, Lehmann's Medicin. Handatlanten Band X, Verlag J.F. Lehmann, München, 1896.
[25] E. Mayr, Systematics and the origin of species, Columbia University Press, New York, 1942.
[26] R.G.E. Murray, K.H. Schleifer, Taxonomic notes: a proposal for recording the properties of putative taxa of prokaryotes, Int. J. Syst. Bacteriol. (1994) 174–176.
[27] R.G.E. Murray, E. Stackebrandt, Taxonomic note: implementation of the provisional status Candidatus for incompletely described prokaryotes, Int. J. Syst. Bacteriol. 45 (1995) 195–196.
[28] H. Ochman, E. Lerat, V. Daubin, Examining bacterial species under the specter of gene transfer and exchange, Proc. Natl. Acad. Sci. USA 102 (2005) 6595–6599.
[29] T.R. Papke, A critique of prokaryotic species concept, Methods Mol. Biol. 532 (2009) 379–395.
[30] M.A. Riley, M. Lizotte-Waniewski, Population genomics and bacterial species concept, Methods Mol. Biol. 532 (2009) 367–377.
[31] R. Rossello-Mora, Updating prokaryotic taxonomy, J. Bacteriol. 187 (2005) 6255–6257.
[32] K.H. Schleifer, O. Kandler, Peptidoglycan types of bacterial cell walls and their taxonomic implications, Bacteriol. Rev. 36 (1972) 407–477.
[33] K.H. Schleifer, W.P. Hammes, O. Kandler, Effect of endogenous and exogenous factors on the primary structures of bacterial peptidoglycan, Adv. Microb. Physiol. 13 (1976) 245–292.
[34] K.H. Schleifer, E. Stackebrandt, Molecular systematics of prokaryotes, Ann. Rev. Microbiol. (1983) 143–187.
[35] V. Soria-Carrasco, M. Valens-Vadell, A. Pena, J. Anton, R. Amann, J. Castresana, R. Rossello-Mora, Phylogenetic position of *Salinibacter ruber* based on concatenated protein alignments, Syst. Appl. Microbiol. 30 (2007) 171–179.
[36] S.K. Sheppard, N.D. McCarthy, D. Falush, M.C.J. Maiden, Convergence of *Campylobacter* species: implications for bacterial evolution, Science (2008) 237–239.
[37] P.H.A. Sneath, R.R. Sokal, Numerical taxonomy, the principles and practice of numerical classification, W.H. Freeman, San Francisco, 1973.
[38] E. Stackebrandt, J. Ebers, Taxonomic parameters revisited: tarnished gold standards, Microbiol. Today 33 (2006) 152–155.
[39] E. Stackebrandt, W. Fredriksen, G.M. Garrity, P.A.D. Grimont, P. Kämpfer, M.C.J. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H.G. Trüper, L. Vauterin A.C. Ward, W.B. Whitman, Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology, Int. J. Syst. Evol. Microbiol. (2002) 1045–1047.
[40] J.T. Staley, N.R. Krieg, Classification of prokaryotic organisms: an overview, in: N.R. Krieg, J.G. Holt (Eds.), Bergey's Manual of Systematic Bacteriology, Vol. 1, Williams and Wilkins, Baltimore, USA, 1984, pp. 1–4.
[41] R.Y. Stanier, M. Doudoroff, E.A. Adelberg, The Microbial World, 3rd ed, Prentice-Hall, 1970.
[42] R.Y. Stanier, C.B. Van Niel, The concept of a bacterium, Arch. Mikrobiol. 42 (1962) 17–35.

[43] P. Vandamme, B. Pot, M. Gillis, P. de Vos, K. Kersters, J. Swings, Polyphasic taxonomy, a consensus approach to bacterial systematics, Microbiol. Rev. 60 (1996) 407–438.

[44] L.G. Wayne, D.J. Brenner, R.R. Colwell, P.A.D. Grimont, O. Kandler, M.L. Krichevsky, L.H. Moore, W.E.C. Moore, R.G.E. Murray, E. Stackebrandt, M.P. Starr, H.G. Trüper, Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics, Int. J. Syst. Bacteriol. 37 (1987) 463–464.

[45] C. Woese, G. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms, Proc. Natl. Acad. Sci. USA 74 (1977) 5088–5090.

[46] C.R. Woese, Bacterial evolution, Microbiol. Rev. 51 (1987) 221–271.

[47] P. Yarza, M. Richter, J. Peplies, J. Euzeby, R. Amann, K.-H. Schleifer, W. Ludwig, F.O. Glöckner, R. Rossello-Mora, The all-species living tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains, Syst. Appl. Microbiol. 31 (2008) 241–250.