SYSTEMATICS OF ARCHAEO AND BACTERIA

A. Oren

Department of Plant and Environmental Sciences, The Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel.

Keywords: Archaea, Bacteria, 16S rRNA, Taxonomy, Systematics, Nomenclature, Bacteriological Code, Bergey’s Manual, Polyphasic taxonomy.

Contents

1. Introduction
2. Systematics, Taxonomy and Nomenclature of Procaryotes - A Few Definitions
3. Procaryote Systematics - A Historical Overview
4. The Formal Framework of Description and Nomenclature of Procaryotes
5. Approaches to the Classification of Procaryotes
5.1. Experimental Tools of Procaryote Taxonomy
5.2. The Species Concept in the Procaryote World
5.3. The Number of Described Procaryote Species
6. Bacteria and Archaea, the two Domains of the Procaryotic World
   6.1. The Recognition of the Archaea as a Separate Domain
   6.2. The Differences between Archaea and Bacteria
6.3. Systematics of Archaea and the Properties of the Main Groups of Archaea
6.4. Systematics of Bacteria and the Properties of the Main Groups of Bacteria
7. How Many Procaryote Species are there in Nature?
8. Conclusions
Appendices
Glossary
Bibliography
Biographical Sketch

Summary

This article presents the current views relating to the taxonomy of procaryotes. Following a historic survey of the methods used in the past in bacterial classification and taxonomy, the present-day concepts of classification are discussed, concepts based both on 16S ribosomal RNA sequence comparisons and on polyphasic approaches, taking into account a wealth of phenotypic and genotypic data. All models now support the division of the procaryotes into two domains, the Archaea (formerly named Archaebacteria or Archaeobacteria) and the Bacteria (formerly the Eubacteria).

The properties of the Archaea and the Bacteria are discussed, including the division of both domains into different kingdoms. Use of the 16S rRNA sequence database in environmental studies has made it clear that the presently known species (about 7500 described) form only a small percentage of the total species richness of procaryotes in Nature.
1. Introduction

This contribution will discuss the systematics of the procaryotes. It is generally accepted that procaryotes have inhabited Earth for at least 3.5 billion years, and that they have since developed into an extremely varied group of organisms, displaying a tremendous metabolic diversity (see Metabolic Diversity in Procaryotes and Eucaryotes), greatly exceeding the metabolic diversity shown by the eucaryotes (see Biological Science Fundamentals).

When molecular approaches were introduced to procaryotic systematics in the late 1970, it became clear that the procaryotes do not form a single, phylogenetically coherent group, but consist of two, fundamentally different groups, the Archaea (formerly called Archaebacteria) and the Bacteria (formerly named Eubacteria). Most microbiologists now divide the organisms inhabiting Earth into three domains: the Archaea, the Bacteria, and the Eucarya (eucaryotes). Accordingly, this chapter will deal with the systematics of two out of these three domains. After a short historical overview on the taxonomic ideas brought forward by microbiologists in the past, the methods and approaches currently used in this branch of science will be discussed. It will be shown that procaryote systematics is a highly dynamic science, in which the concepts rapidly change in accordance with the development of new techniques and approaches. Furthermore, procaryotic systematics still lacks a firm theoretical basis: there is not even a firm definition of the basic taxonomic unit - the species - used in classification. In spite of these obvious problems, a classification scheme has emerged that is now generally accepted. The chapter will present an overview of the present-day classification of the approximately 7,500 species of procaryotes currently recognized. Finally it will also be made clear that the number of species known today probably amounts to only a few percent at most of the total number of species of procaryotes that inhabit our planet.

2. Systematics, Taxonomy and Nomenclature of Procaryotes - A Few Definitions

Systematics is the scientific study of organisms with the ultimate objective of characterizing and arranging them in an orderly manner. The term has also sometimes been defined as "the study of organismal diversity and interrelationships" (see Biological Science Fundamentals).

Taxonomy is often used as a synonym for systematics. Other definitions can be found in the literature, such as "the art of biological classification", or "the theoretical study of classification, including its bases, principles, and rules" (see Classification and Diversity of Life Forms: Historical Review of Systematic Biology and Nomenclature). Taxonomy is often defined as the theory and practice of classifying organisms into groups (taxa) on the basis of similarities and relationships, systematics being a broader concept that includes the evolutionary and phylogenetic concepts. Cowan stated in 1968 that taxonomy consists of (1) Classification, i.e., the orderly arrangement of units into groups. Classification includes the theory and process of ordering the characterized organisms into one or more systems, (2) Nomenclature, i.e., the labeling of units defined by classification, and (3) Identification of unknowns with the units defined by classification and labeled by nomenclature. Identification is thus the practical use of the
arts of classification and nomenclature to determine the identity of an isolate as a member of an established taxon or as a member of a previously unidentified species. Classification is motivated by the desire of taxonomists to provide the user with a system that in their opinion optimally reflects the natural relatedness between the taxa. The basis of any classification system of living organisms, procaryotic as well as eucaryotic, is the species and the genus, in accordance with the binomial system pioneered by Linnaeus in the 18th century.

Nomenclature is the assignment of names to the taxonomic groups according to international rule, by naming of the appropriate taxon within a classification system. Nomenclature is subject to changes that occur whenever novel insights alter the taxonomic weight of characters and thereby the rank of taxa.

3. Procaryote Systematics - A Historical Overview

From the early days of microbiology as a science, microbiologists have realized the difficulties in establishing a satisfactory classification system for bacteria. Procaryotes are morphologically very simple, so morphological characters, so useful in the systematics of eucaryotes, are of little help. Moreover, a useful fossil record is altogether lacking, as those fossils that do exist are phylogenetically uninformative.

One of the first attempts toward a classification of the bacteria was made by Ferdinand Cohn in the 1870s. Cohn grouped the bacteria according to their overall morphological appearance, and he discerned six genera based on morphological criteria (e.g., cocci, short rods, spirals). He also clearly pointed out that morphological properties are insufficient, as similarly shaped bacteria may have different physiological characters, and may differ in important properties such as metabolites produced, pathogenesis, etc. He also perceived the close relationship between the "common bacteria" and the cyanobacteria (then called blue-green algae), and grouped them together as the Schizophyta.

The recognition in the 1930s of the procaryotic cell structure of bacteria being fundamentally different from the more complex eucaryotic cell provided a structural basis for the development of a bacterial taxonomy. Since the early 20th century, taxonomic systems were proposed based on the presumed evolutionary relationships of bacteria. However, the shortcomings of the deductive phylogenetic approach to bacterial classification were fully recognized. These shortcomings were mainly seen to be the lack of suitable parameters that allow differentiation between "relationship" and mere "resemblance". Physiological characters were suspect because of the wide range of adaptation that is manifested in bacteria.

For that reason, early procaryote taxonomists suggested that rather than searching for a natural system, taxonomists should limit themselves to the development of determinative keys to provide easy identification of species and genera. Stanier wrote, in 1970, that "For (most) major biological groups [including the bacteria], the general course of evolution will probably never be known, and there is simply not enough objective evidence to base their classification on phylogenetic grounds. For these and other reasons, most modern taxonomists have explicitly abandoned the phylogenetic
approach”.

The recognition that phylogenetic information can be obtained by comparison of nucleotide or amino acid analyses in DNA, RNA, and protein, together with the development of the tools to elucidate these sequences, has fundamentally changed our views on the systematics of the procaryotes. A phylogenetic approach to the classification of procaryotes became possible in the 1970s with the comparison of nucleotide sequences of small-subunit (16S) ribosomal RNA, as pioneered by Carl Woese. This approach also led to the recognition that the procaryotes do not form a phylogenetically coherent group, but rather consist of two major groupings, the Archaea (Archaebacteria) and the Bacteria (Eubacteria). The evidence that led to this recognition and the impact of the concept are discussed in greater depth in section 6.1.

With all the progress made in recent years in procaryote systematics with the reconstruction of phylogenetic trees, it should be stated that classification is still a dynamic process. Phylogenetic reconstructions based on inferred homologies cannot be considered definitive unless witnessed by the evolutionary history of taxa, i.e., by fossil data. It may be doubted whether such data will ever become available.

4. The Formal Framework of Description and Nomenclature of Procaryotes

The rules of nomenclature appear in the "International Code of Nomenclature of Bacteria" (see the Bibliography). These rules are framed in legalistic language. Since the publication of the Approved Lists of Bacterial Names (1980), valid publication of new names and new nomenclatural combinations can only be obtained by publication in the International Journal of Systematic and Evolutionary Microbiology (until 1999, the International Journal of Systematic Bacteriology), either in the form of an original article, or in the "Validation Lists" regularly appearing in that journal. The code and the Lists are under the aegis of the International Committee on Systematics of Prokaryotes (formerly named the International Committee on Systematic Bacteriology), which is a constituent part of the International Union of Microbiological Societies. The Committee is assisted by a number of Taxonomic Subcommittees on different groups of bacteria, and by the Judicial Commission, which considers amendments to the Code and any exceptions that may be needed to specific Rules. An updated edition of the revised Code was published in 1992, and a new version will soon be published ("International Code of Nomenclature of Prokaryotes").

Description of an isolate as a member of a new taxon (genus, species) requires extensive documentation of its properties, morphological, and physiological, as well as 16S rRNA sequence determination, base composition of the DNA, DNA-DNA similarity with its closest relatives, and as many other characteristics as possible, such as records on enrichment and isolation, motility, colony characterization, optimal growth conditions, and growth requirements. A type strain should be established, which has to be deposited in at least two culture collections located in different countries. In case a procaryote cannot be described in sufficient detail to warrant establishment of a novel taxon, e.g., because the organism cannot yet be cultured, the possibility exists to describe the organism with the status “Candidatus”, i.e., awaiting positioning as a putative taxa. As of September 4, 2008, 87 such ‘Candidati’ have been described in the International
Journal of Systematic Bacteriology / International Journal of Systematic and Evolutionary Microbiology.

A formal system of bacterial classification, based both on morphological and on physiological characteristics, was established in the second decade of the 20th century. This system provided the basis for "Bergey's Manual of Determinative Bacteriology", the first edition of which appeared in 1923. Eight editions of this manual have been published, the last one in 1974, each edition being updated with the newly discovered bacterial species and being revised according to the changing concepts on bacterial taxonomy. All editions provided identification keys to enable the identification of bacterial isolates. "Bergey's Manual of Determinative Bacteriology" was succeeded by "Bergey's Manual of Systematic Bacteriology" (first edition 1984-1990, the first of the five planned volumes of the 2nd edition appeared in 2001, the second in 2005), in which the classification is based on phylogenetic relationships, and attempts to provide identification keys have been abandoned.

The highest taxonomic rank used in Bergey's Manual is called a domain. All procaryotic organisms are placed within two domains: *Archaea* and *Bacteria*. Phylum, class, order, family, genus, species, and subspecies are successively smaller, non-overlapping subsets of the domain (see Appendix 1 and Appendix 2). Many bacteriologists have the impression that there exists an "official classification" of procaryotes, and that the classification in Bergey's Manual represents this classification. However, there is no "official classification" of bacteria - this is in contrast with bacterial nomenclature, where each taxon has one valid name, according to internationally agreed-upon rules, and judicial decisions are rendered in instances of controversy about the validity of a name. The closest approximation to an "official classification" of procaryotes would be one that is widely accepted by the community of microbiologists.

A special situation exists in the case of the phylum *Cyanobacteria* (domain *Bacteria*). While being procaryotic, the cyanobacteria are also claimed by botanists to be part of the plant world, as they can perform oxygenic photosynthesis. In botanical classification systems they are named Cyanophyta or blue-green algae. Under the rules of the International Code of Botanical Nomenclature, types are not represented by live reference cultures but by preserved herbarium specimens, photographs, etc. The nomenclature of the cyanobacteria is still highly confusing, and many species still appear under different names in the literature. The 2001 edition of Bergey's Manual refrains from dividing the phylum into classes, orders and families, but prefers a provisional division of the group into five subsections, each subsection consisting of several "form-genera" (see Appendix 2).

5. Approaches to the Classification of Procaryotes

The description of species, genera and higher taxa is based on a combination of genotypic and phenotypic characteristics. This section will describe some of the experimental approaches in use in procaryote systematics.

While total genome sequencing is not yet a routine procedure in bacterial taxonomy, the number of procaryotes whose genome has been completely sequenced is rapidly
increasing. It may be expected that in the future, whole genome sequences may become an integral part of the characterization of new species. Until then, the genotypic characterization of strains is necessarily more limited in scope.

5.1. Experimental Tools of Procaryote Taxonomy

Certain properties of the total genome have obtained an important place in the characterization of procaryotes. The DNA base ratio (expressed as the mol% G+C) is an important parameter in any species description. DNA-DNA hybridization is the most rapid and inexpensive of the phylogenetic methods for measuring an average nucleotide similarity of the entire genome, but gives no indication of which genes contribute to or detract from the similarity (see *Evolutionary and Molecular Taxonomy*). This method is extensively used to elucidate the phylogenetic relationship between closely related taxa (e.g., to determine whether two isolates belong to the same species); strains (clones made up of the descendants of a single isolation in pure culture) that share at least 70% DNA similarity, corresponding with at least 96% DNA sequence identity, are generally considered to belong to the same species (see section 5.2). At the level of genera or higher taxa, DNA-DNA hybridization has limited resolving power. Another powerful genotypic characteristic for use in taxonomy is the 16S tRNA nucleotide sequence (see section 6.1).

Individual phenotypic tests are insufficient as parameters for genetic relatedness. Yet, as a whole, they provide descriptive information enabling the recognition of taxa. Phenotypic parameters of value in bacterial systematics and taxonomy include cell shape, cell size, motility, mode of flagellation, sporulation, presence of cellular inclusions, color, colonial morphology, ultrastructural characteristics, Gram-staining behavior, the chemical composition of cellular components such as the cell membrane, the polar ester lipids of *Bacteria*, hopanoids, the ether lipids in *Archaea*, aliphatic hydrocarbons, isoprenoid quinones, polyamines, bacteriochlorophylls in photosynthetic *Bacteria*, cytochromes, cell wall ultrastructure and composition, lipopolysaccharides, cell wall polysaccharides and/or teichoic acids, mycolic acids in those organisms that possess them, the structure of the peptidoglycan (where applicable), the nature of cell envelope polymers in *Archaea*, and the presence of exopolysaccharides. Also valuable are physiological properties, such as basic energy metabolism, nutritional and metabolic characters, special nutritional requirements, diagnostic enzymes, ecological parameters (requirements for pH, redox potential, salinity, etc.), and susceptibility to bacteriophages and production of - and susceptibility to bacteriocins and other antimicrobial agents.

Procaryote taxonomy is now generally based on a polyphasic approach. The term polyphasic taxonomy was introduced in 1970 to refer to taxonomy that assembles and accumulates many levels of information, both based on phylogenetic and phenotypic data, from molecular to ecological. Nowadays, polyphasic taxonomy refers to a consensus type of taxonomy and aims to utilize all the available data in delineating consensus groups (see *Schools of Taxonomy*). All phylogenetic and phenotypic parameters mentioned above are useful in polyphasic taxonomy, in addition to additional sources of information such as DNA restriction patterns (Restriction Fragment Length Polymorphism and related methods), genome size, PCR-based DNA fingerprinting, use of DNA probes, electrophoretic patterns of total cell proteins or cell
envelope proteins, electrophoretic enzyme patterns), serological tests, etc. The above list is by no means exhaustive (see *Evolutionary and Molecular Taxonomy*).

A special type of polyphasic taxonomy is numerical taxonomy (also called phenetic taxonomy or Adansonian taxonomy), an approach developed in the late 1950s as part of multivariate analysis. Here, a large number (at least 50-60, preferentially 100 or more) of properties belonging to different categories (morphology, physiology, biochemistry, serology, etc.) are determined. These characters should be independent and should come about equally from the various different categories. Numerical taxonomy principally allots the same weight to each character, enabling the construction of similarity matrices after calculating coefficients of similarity between strains and classification of isolates in different groups ("phenons") (see *Schools of Taxonomy*).

### 5.2. The Species Concept in the Procaryote World

The basis of the taxonomic hierarchy is essentially the species. However, the concept of a species still lacks a theoretical consensus (see *Evolution and the Species Concept*). Species description is still subjective, and all existing definitions are pragmatic. To give a few examples found in the literature: "A species consists of an assemblage of individuals (or, in microorganisms, of clonal populations) that share a high degree of phenotypic similarity, coupled with an appreciable dissimilarity from other assemblages of the same general kind", or "A microbial species is a concept represented by a group of strains, that contain freshly isolated strains, stock strains maintained *in vitro* for varying periods of time, and their variants, which have in common a set or pattern of correlating stable properties that separates the group from other groups of strains", or "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 "The smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descendents", or even "A group of organisms defined more or less subjectively by the criteria chosen by the taxonomist to show to best advantage as far as possible and putting into practice his individual concept of what a species is". The species definition is so extremely subjective because one cannot consistently determine and define such concepts as "a close resemblance", "essential features", or how many "distinguishing features" are sufficient to create a species.

Probably the most widely accepted pragmatic definition of a procaryotic species defines a species as a group of strains, including the type strain, sharing at least 70% total genome DNA-DNA hybridization and having less than 5°C $\Delta T_m$ (= the difference in the melting temperature between the homologous and the heterologous hybrids formed under standard conditions). Thus, total genome DNA-DNA hybridization values are the key parameter in this species delineation, and DNA reassociation values are an indirect expression of the real genome sequence identity. As a result, the procaryotic species concept is not troubled by the problem of horizontal gene exchange mediated by processes such as conjugation, phage transduction, or DNA transformation, as the species concept is based on whole genome similarities. Values from 30-70% DNA relatedness reflect a moderate degree of relationship. It should be realized that the delineation value of 70% is artificial, but has proven satisfactory in most cases.
There are no clear-cut recommendations for the delineation of procaryote genera or higher taxonomic levels, except that it is generally accepted that genera should reflect phylogenetic relationships. As a result, the ranks of species and genera are determined by subjective decision-making by bacteriologists on the basis of a phylogenetic framework. One can define the genus as "a collection of species with many characters in common", but there are no clear guidelines that indicate the extent of this sharing of characters, and it is purely a matter of personal judgment as to what constitutes a genus. The definition of a genus has not been changed by the availability of molecular data. As with the species, the genus is a subjective concept without any well-defined foundation.

5.3. The Number of Described Procaryote Species

As of 4 September 2008, the number of validly described names of different species of procaryotes, *Archaea* and *Bacteria* combined, was 7,534, divided over 1,837 genera (for updates see http://www.bacterio.cict.fr; last accessed: October 7, 2008). Thus, the number of recognized species of procaryotes is very small, as compared to the eukaryotes, when realizing that for example the number of orchid species described exceeds 20,000, and that more than a million different species of insects have been recognized, including more than 300,000 species of beetles (order *Coleoptera*).

This small number of recognized procaryote species is in part due to the fact that most of the species existing in nature are still waiting to be isolated and described. This insight has been obtained from the time that 16S rRNA sequencing techniques were applied to DNA isolated from natural communities of procaryotes (see section 7). However, the number of known species is also low as a result of the way species are defined in the bacterial world. When comparing the species concept in procaryotes with that in higher organisms using criteria such as DNA G+C content, DNA-DNA relatedness and 16S and 18S rRNA sequence similarity respectively, it becomes apparent that the bacterial species concept is much broader than that of higher eucaryotes. Humans and chimpanzees are 98.4% related on the basis of DNA-DNA hybridization. Indeed, even lemurs (78% DNA relatedness with humans) would be included in the same species as humans if the definition of a bacterial species was used!

6. Bacteria and *Archaea*, the two Domains of the Procaryotic World

Since the late 1970s –and the early 1980s, the determination of phylogenetic relationships among procaryotes has become achievable thanks to the development of appropriate techniques based on DNA and RNA sequencing. The theoretical basis for the new systematics was laid down by Zuckerkandl and Pauling, who in 1965 introduced the concept that because today's organisms are the products of historical events, these historical events can be deduced from sequence comparisons of homologous and phylogenetically informative molecules, i.e., DNA, RNA, and proteins. Useful phylogenetic markers are informative molecules that are ubiquitously distributed, functionally equivalent and homologous housekeeping molecules. Through the comparison of sequences, using treeing algorithms of proven resolving power and statistical significance, trees are constructed that provide information on phylogenetic relationships (see *Evolutionary and Molecular Taxonomy*).
Bibliography


Biographical Sketch