

Universal species concept: pipe dream or a step toward unifying biology?

James T. Staley

Received: 3 August 2009 / Accepted: 11 September 2009 / Published online: 25 September 2009
© Society for Industrial Microbiology 2009

Abstract The Universal Tree of Life, which is based on phylogenetic analysis of the RNA sequence from the small ribosomal subunit, was a breakthrough in understanding the relatedness among all living organisms. The result has had a major impact on taxonomy by separating life into three domains: Eukarya, Bacteria, and Archaea. Indeed, microbiologists have used the 16S ribosomal RNA (rRNA) of the small ribosomal subunit to construct the hierarchical classification of Bacteria and Archaea from the level of domain to genus. However, the 16S rRNA of the Bacteria and Archaea and the corresponding 18S rRNA of the Eukarya are too highly conserved to be useful phylogenetically at the species level. For this reason, I propose that biologists adopt a phylogenomic species concept that utilizes both phylogenetic analyses of less highly conserved genes and proteins as well as genomic analyses for the circumscription of species. If biologists adopt a phylogenomic concept for species, the classification of all living organisms from domain to species could be completed. Furthermore, this universal species concept could help provide a more equitable circumscription among all species, as well as aid in the unification of biologists and biology.

Keywords Taxonomy · Phylogeny · Speciation · Phylogenomic · Species concepts

As a science, microbiology developed more slowly than botany and zoology due to the late discovery of

microorganisms and the time required to develop the special techniques and instruments required for their investigation. This unavoidable situation resulted in the early separation of microbiology from biology. This separation persists in some forms today as microbiologists typically have their own separate journals, societies, and organizational units in universities and governmental agencies. This began to change in the twentieth century when it was recognized that all organisms shared common metabolisms and physiologies, a view that has been referred to as the “comparative biochemistry” of life. However, it was not until the latter half of the century following the discovery of DNA and the genetic code that a more complete recognition of the unity of biology could be realized. To illustrate this, consider the role that molecular sequences have played in unifying biology. Even a cursory visual comparison of the sequence of the 5S rRNA molecule of *Escherichia coli* shows its striking similarity to that of *Homo sapiens* (Fig. 1).

One of the major achievements of molecular sequencing was the construction of the phylogenetic Universal Tree of Life, which relied on comparing representatives of all organisms using phylogenetic analyses of the same macromolecule, the 16S rRNA of the Bacteria and Archaea and the 18S rRNA of the Eukarya [1]. The Tree of Life has been one of the major unifying advances in biology. The highly conserved 16S rRNA became the most important macromolecule for bacterial taxonomy because of its fidelity in inferring evolution at taxonomic levels at and above the genus. Bergey’s Manual Trust adopted the phylogenetic approach for taxonomy of the Archaea (Archaeobacteria) in the first edition of Bergey’s Manual of Systematic Bacteriology [2, 3]. Furthermore, the entire second edition uses the phylogenetic approach for the Bacteria as well as the Archaea (www.bergeys.org). Therefore, for the first time, a complete hierarchical taxonomy based on the phylogeny of

J. T. Staley (✉)
Department of Microbiology, University of Washington,
Campus Box 357242, Seattle, WA 98195, USA
e-mail: jtstaley@u.washington.edu

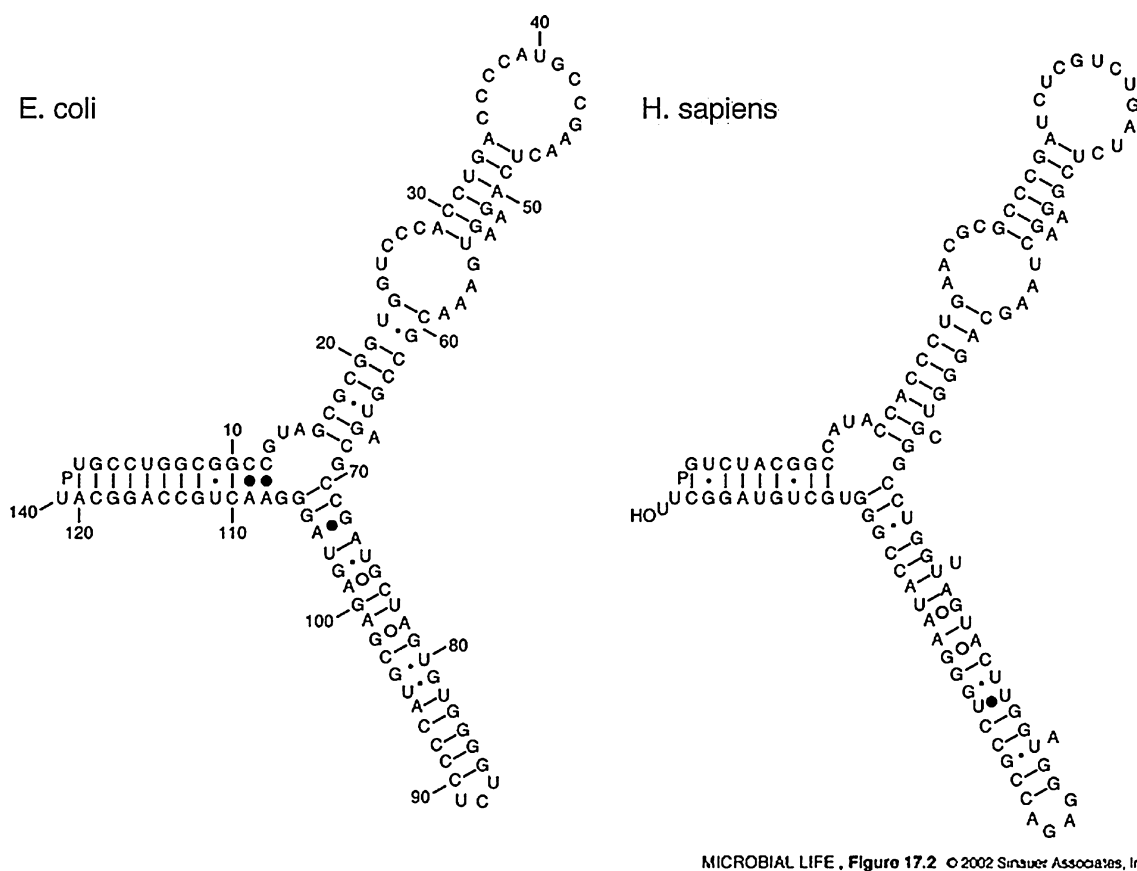


Fig. 1 The structure of 5S ribosomal RNA (rRNA) from *Escherichia coli* and *Homo sapiens* illustrates their overall similarity, although these organisms belong to two separate domains of life. From *Microbial Life*, 2nd edition. Courtesy of Sinauer Publications, Massachusetts

16S rRNA gene sequences is in place for the Bacteria as well as the Archaea from the domain down to the genus. Unfortunately, however, the 16S rRNA of Bacteria and Archaea and the 18S rRNA of eukaryotes are too highly conserved to recover relationships at lower taxonomic levels, especially the species [4]. This paper proposes that biologists adopt a universal species concept, in particular, the phylogenomic concept that uses less highly conserved genes and proteins tailored to the group of organisms being investigated and coupling this with genomic approaches for identifying species.

Speciation and species concepts

Speciation is an evolutionary process in which a population containing distinct individuals diverges to form two separate and novel clades that eventually become separate species. Speciation is of interest to population biologists and taxonomists, as this process is at the juncture of both biological subdisciplines. Botanists and zoologists regard two principal factors as drivers of speciation. The primary factor is

allopatry, which occurs when a population becomes separated into two different geographic areas. This may be caused by a geological or environmental change or by long-distance dispersal of some members of a population across pre-existing barriers. In both examples, the new habitat allows the species to diverge from the ancestral (and sister) lineage through mutation, selection, and/or genetic drift so that it eventually forms a novel species. Species may also diverge to form new species in the same habitat as the ancestral species by a process termed sympatric speciation. For example, hybridization in plants may result in the formation of polyploids that become reproductively isolated from both parents.

Many different species concepts have been proposed, but most do not have the potential of becoming universally accepted by biologists. One example is arguably the most highly recognized concept, the biological species concept, proposed by Ernst Mayr [5]. This concept is based on reproductive isolation within a species, which typically occurs through allopatry. The population diverges into two subpopulations that are sufficiently disparate that they can no longer interbreed to produce fertile progeny. Although

this applies well to many animal and plant species, it is not at all applicable to most microorganisms. For example, although Bacteria and Archaea are known to exchange DNA, sexual reproduction is unnecessary, as they reproduce by asexual means.

By far the primary species concept used by botanists and zoologists is the morphological species concept, which relies on the structural attributes of organisms. It is also used by paleontologists to describe and name fossils. Protistologists and those who study diatoms use this concept to identify species by visible features they determine microscopically. However, bacteriologists cannot use this concept because most of these microorganisms are indistinguishable morphologically. Some population geneticists have supported a species concept that emphasizes gene coalescence at a single node [6]. One variation of this concept allows for interbreeding within a zone, referred to as a hybrid zone, which lies between two species. As with the phylogenomic species concept (PSC) proposed here, this may also be broadly applicable across biology.

This paper proposes the PSC, which was originally termed the genomic–phylogenetic species concept [7]. The PSC is based on phylogenetic theory in which the evolution of an organism is inferred from sequence analyses of its genes and proteins [8] as well as genomic data that already provides important information about bacterial species [9, 10]. Synteny, hybridization, and expression technology can be used in genomic analyses, not only to evaluate divergence at the species level but to help resolve problems such as those caused by horizontal gene transfer (HGT) (Box 1).

The PSC proposal is based on the realization that speciation is a universal process that occurs in all organisms and can be assessed using phylogenetic and genomic analyses. Phylogenetic analyses need not be restricted to tree topology but can also be based on individual character analyses of gene (DNA base) and protein (amino acid) sequence differences [11, 12]. Likewise, genomic analyses entail approaches in addition to sequence analyses such as the use of expression arrays to confirm phenotype. The decision about the extent of divergence within a clade that justifies two separate species should be based on whether there are

diagnosable differences that allow one to distinguish between them.

The phylogenomic species concept

Actually, the PSC is being applied to microorganisms already. Spratt's laboratory has pioneered the use of multiple-locus sequence analysis (MLSA) to assess the relatedness of bacterial pathogens [13–15]. The sequences of several genes, typically from five to eight, are concatenated and subjected to phylogenetic analysis. In this procedure, numerous strains of a genus are used for the analysis, which results in a tree for each known species (Fig. 2). In most instances, the strains fall into one or another of the named species of the genus. Occasionally, strains fall between two species. These may represent a separate, as yet undescribed, species or are perhaps the result of interspecies HGT. Genomic studies should lead to a better understanding of whether or not these represent examples of HGT, and if so, what the nature is of their origin.

Some major advantages of the PSC are:

- (i) It relies on phylogenetic theory, which is based on ascertaining the evolution of an organism through gene and protein sequence analyses.
- (ii) It is pragmatic, as it relies on the methodology of phylogenomic analyses.
- (iii) Unlike DNA hybridization, data are readily portable, i.e., sequences can be transmitted anywhere electronically.
- (iv) Unlike DNA hybridization and phenotypic traits that vary depending on laboratory conditions, data are archival in that sequences can be stored for future comparison.
- (v) Its universal applicability to all organisms.

The bacterial species definition and phenotype

Bacteriologists do not have a species concept. Instead, they have adopted a highly specific definition: those strains that

Box 1 Do microbial species evolve?

Some biologists, especially microbiologists, believe there is no traceable evolution of organisms, only of their genes. Those who endorse this view believe that extensive horizontal gene transfer (HGT), in particular, in the Bacteria and Archaea, has erased the evolutionary record of their lineages. Although there is considerable gene exchange among microorganisms, this does not belie the fundamental place and role of the organism in its own evolution. The organism coordinates the use of its large repertoire of genes and proteins to enable it to obtain energy, reproduce, carry out its ecological activities, and survive in a dynamic environment. Occasionally, DNA from the organism's environment is taken up through HGT and used for its benefit. Although the evolution of organisms is more complex to study than that of their genes, organisms do have an evolutionary history, and it is the complexity of their evolutionary pathway that makes the fabric of life so rich. The role of biologists is to understand how this occurs and why. After all, it is not genes that inhabit the world, but the organisms that contain them. The view of those who regard a bacterium as a "a bunch of independently evolving genes" is analogous to the earlier view of some physiologists who treated bacteria as nothing more than a "bag of enzymes".

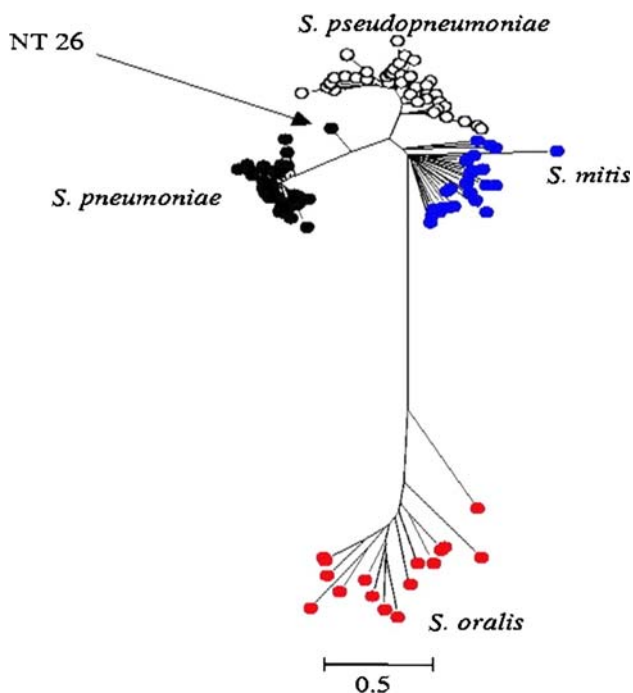


Fig. 2 Phylogeny of the *Streptococcus* genus using multiple-locus sequence analysis (MLSA). Each dot represents a strain that has been analyzed by MLSA. Note that each species is separated into its own clade based on phylogenetic analyses. Note that one strain, NT 26, lies between two species and is therefore not identifiable by this analysis. Courtesy of Brian Spratt and Phil. Trans. R. Soc.

exhibit >70% whole-cell DNA–DNA hybridization and a distinctive phenotype are considered to be members of the same species. If they show <70% DNA hybridization, they are considered to be separate species [16]. This definition has served bacteriologists well since it was adopted, because it has allowed for greater uniformity in delineating species of Bacteria and Archaea. Also, the analysis assesses the degree of molecular (DNA–DNA) cohesion within strains of a species. However, it is neither conceptual nor based on the evolutionary history of the organism. It is noteworthy that phenotypic properties are also considered to be important in the PSC, as it is recognized that all species have a distinct phenotype of expressed properties that aid in circumscribing a species' boundary. The traditional use of phenotypic data should be continued where it provides valuable information. However, “genomospecies” have been identified for some bacteria in which there is no differentiating phenotype, even though DNA hybridization supports their being separate species [17]. So, at finer levels of resolution, it may be difficult or even impossible to identify some species by phenotype, even though phylogenetic analysis provides convincing support. However, character analyses of gene and protein sequences as well as genomic approaches should help resolve this issue.

Why is a universal concept for species desirable?

Most biologists would agree that it is unrealistic to expect that the adoption of a species concept could lead to the complete equivalency of species of plants, animals and microorganisms. In contrast, the current situation is far from that ideal. As an example, consider the inequity among the species of Bacteria when compared with the Eukarya. Based on the Tree of Life, there are literally tens of “kingdoms” or “phyla” of Bacteria versus only one kingdom of animals. However, it is curious to note that there are fewer than 8,000 species of Bacteria compared with more than a million species of animals. Mayr used this information to conclude that Bacteria are much less diverse than the Eukarya [18]. To wit, he stated: “Archaeobacteria...even where combined with the eubacteria, as prokaryotes, this group does not reach anywhere near the size and diversity of the eukaryotes.” In the same article he says: “Approximately 10,000 eubacteria have been named. The number of species of eukaryotes exceeds 30 million: in other words, it is greater by several orders of magnitude. The numbers of species of birds alone is ~10,000, and there are many millions of species of insects.” These statements rely on the assumption that species of Bacteria and Archaea are essentially equivalent to those of animals, which is incorrect for the following reasons.

Several technical reasons account in part for the marked disparity between the large number of bacterial phyla and the paucity of their species. First, it is not easy to identify and describe new bacterial or archaeal species. For most organisms, several weeks or months of laboratory work are required. Not only is this labor intensive, but it requires appropriate materials and equipment as well as a laboratory and personnel to carry out the studies. In fact, only a few specialized laboratories conduct DNA hybridization analyses now. Second, the naming of bacteria requires that they are isolated in pure culture and described. It has been estimated that fewer than 1% of all Bacteria and Archaea have been isolated in pure culture. Therefore, many thousands of species yet remain to be isolated, described, and named.

However, aside from these purely technical reasons, there is an intrinsic underlying problem that remains at the heart of why there are so few species: the species definition for bacteria and archaea is very broad in comparison with that of animals and plants. This point can be illustrated by considering the DNA–DNA hybridization definition with a cutoff of >70% that is used to circumscribe a bacterial species. If this were applied to primates, all primates – from humans to lemurs – would comprise a single species [19]! In point of fact, the current bacterial species definition is so broad that it has been very difficult to detect allopatric species. This issue of biogeography has haunted

microbiologists for years, ever since the work of Baas-Becking who stated that “Everything is everywhere, the environment selects” [20]. This statement is interpreted to mean that all bacteria have a cosmopolitan distribution, i.e., a species will be found anywhere on Earth where environmental conditions are conducive to its growth. Of course, if this were true, then there would be many fewer bacterial species than would otherwise actually exist. Do endemic species, i.e., those that are found at only one location on Earth exist?

Some protistologists, who use the morphospecies concept, have argued in accordance with Baas-Becking’s view, that smaller organisms such as protists and bacteria show no biogeography; that is, they are not endemic [21, 22]. This interpretation suggests that as the size of an organism decreases, there is no allopatry because of its rapid and comprehensive global dispersal. Therefore, they conclude that there are fewer species of small organisms. In contrast, if microbial endemism is common, then there are many more unidentified species. The morphospecies concept used for protists is so broad that it cannot detect allopatric species. Therefore, detection and recognition of endemic microbial species will require the adoption of a new species concept, such as the PSC.

Solid evidence for endemism has been obtained using MLSA studies of the hot-springs thermoacidophile *Sulfolobus islandicus* [23]. This study, which investigated 78 strains, identified separate clades of this archaeon in Icelandic, North American, and Russian hot springs. The taxonomic question is: Should these clades be called emerging species, subspecies, or actual species? If data analysis shows that all strains from a particular location had diagnosable sequence differences that permitted those strains to be distinguished from clades at other locations, these could be named as separate species by the PSC.

Another reason that bacterial biogeography has been difficult to study is the vast diversity of microbial types that reside at each location. Therefore, it is virtually impossible to conclude that a particular species is endemic to one locale because it may reside at another location also, albeit at undetectably low concentrations. This problem is not unique to microbiology, as other organisms show similar patterns. However, genomic studies in progress support the view that microbial endemism occurs and is likely to be very common.

Issues of uniformity—symbiotic associations

A major justification for a universal species concept is that it will lead to greater uniformity and equity in what constitutes a species across all biology. To illustrate this issue, consider the situation with respect to the gamma-

proteobacterial genus, *Buchnera*, which lives in an obligate mutualistic symbiosis with aphids. Although there are about 4,500 species of aphids, there is only a single bacterial species, *B. aphidicola*, named for all of the aphid families, genera and species. This is incredible. As this, as claimed, is truly a coevolutionary process [24], it must be a cospeciation process, too. Therefore, this additional diversity within *B. aphidicola* should be recognized taxonomically. Although in theory there could be as many as one species of *Buchnera* for each species of aphid, this would be warranted only if justified by appropriate phylogenetic and genomic analyses, keeping in mind that other events such as extinction, host switching, and failure to speciate when hosts separate may also be occurring. Correspondingly, additional families and genera should be named if phylogenetic and genomic analyses provide supportive evidence.

Some commensal bacteria may also undergo a coevolutionary, cospeciation process. For example, the oral bacterial genus, *Simonsiella*, which is a member of the Betaproteobacteria family, may coevolve and cospeciate in the mammalian oral cavity where it resides on the epithelial cells of its host. If this can be shown to be an example of cospeciation based on further research, there could be many additional species as well as a substantial number of genera and families of the current *Simonsiella* clades that correspond to the 5,000 known mammalian species [7, 25]. If only a small fraction of animal and plant commensal species coevolve and cospeciate with their hosts, this could account for many thousands of new bacterial species.

Rationale for a universal species concept

The species is the fundamental unit of organismal biology. The species is the basic unit for understanding biodiversity and extinction as well as population and community ecology. Therefore, a universal species concept is clearly important. Considering the importance of having a universal species concept raises the question of why one still has not been adopted. I believe this has not happened for two principal reasons. First, it relates in part to the tardiness of microbiology to the debate. Until recently, it was unknown whether Bacteria, Archaea, and eukaryotic microorganisms actually speciate by the same mechanisms as plants and animals, so there could be no universal species concept. However, emerging data on allopatry support the view that endemism occurs.

Another important aspect of the debate centers on the “spare” or simple biology of microorganisms. Morphology and sexuality are at their most fundamental level in the Bacteria and Archaea, so it is not possible to include microorganisms in certain concepts that are based on

organisms with complicated life styles, such as plants and animals. As J. B. Lamarck (Philosophie Zoologique, 1809) stated “The most important discoveries of the laws, methods and progress of Nature have nearly always sprung from the examination of the small objects which she contains.” Based on this premise and the conclusions of this article, it is clear that it is timely now, with the development of phylogenetic and genomic approaches, for biologists to address and resolve the species concept issue.

Recommendations for moving forward

A decision on the species concept cannot be made without a concerted effort on behalf of the community of biologists. I believe there is no more important issue in biology today. As with great projects in science, resources will need to be allocated to address this issue. Fortunately, its resolution does not require enormous resources or funding. What is needed is leadership within the biological community to establish workshops of population biologists and taxonomists interested in speciation, the species concept, and a universal taxonomy of life. This means that international meetings need to be organized that are highly focused on the issue with the aim of identifying a concept that can be universally applied across biology. Although this article supports the PSC, perhaps another species concept will emerge as being more suitable. The major goal is to adopt a concept that can be applied universally and is acceptable scientifically among all biological disciplines. The adoption of a universal species concept would provide a major step toward further unification of biology and biologists.

Acknowledgments The author thanks Joel Cracraft and Micah Krichevsky for their helpful comments on the paper. I also acknowledge the U.S. Federation of Culture Collections for their support, in part, for the development of this paper.

References

1. Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 87:4576–4579
2. König H, Stetter KO (1989) Section 25 Archaeobacteria. In: Staley JT, Bryant MP, Pfennig N, Holt JG (eds) *Bergey's manual of systematic bacteriology*, vol 3. Williams and Wilkins, Baltimore, MD, pp 2171–2173
3. Staley JT (1989) On using the manual. In: Staley JT, Bryant MP, Pfennig N, Holt JG (eds) *Bergey's manual of systematic bacteriology*, vol 3. Williams and Wilkins, Baltimore, MD, pp xix–xxi
4. Fox GE, Wisotzkey JW, Jurtschuk 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
5. Mayr E (1942) *Systematics and the origin of species*. Columbia University Press, New York
6. Templeton AR (1996) Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. *Genetics* 144:1263–1270
7. Staley JT (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. *Phil Trans R Soc B* 361:1899–1909
8. Cracraft J (1989) Speciation and ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: Otte D, Endler JA (eds) *Speciation and its consequences*. Sinauer Associates, Inc, Sunderland, MA, pp 28–59
9. Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 102:2567–2572
10. Konstantinidis KT, Ramette A, Tiedje JM (2006) The bacterial species definition in the genomic era. *Phil Trans R Soc B* 361:1929–1940
11. Goldstein PZ, DeSalle R (2000) Phylogenetic species, nested hierarchies and character fixation. *Cladistics* 16:364–384
12. Sneath P (1989) Analysis and interpretation of sequence data for bacterial systematics: the view of a numerical taxonomist. *Syst Appl Microbiol* 12:15–31
13. Bishop CJ, Aanensen DM, Jordan GE, Kilian M, Hanage WP, Spratt BG (2009) Assigning strains to bacterial species via the internet. *BMC Biol* 7:3
14. Hanage WP, Fraser C, Spratt BG (2005) Fuzzy species among recombinogenic bacteria. *BMC Biol* 3:6–13
15. Hanage WP, Fraser C, Spratt BG (2006) Sequences, sequence clusters and bacterial species. *Phil Trans R Soc B* 361:1917–1928
16. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky ML, Moore LH, Moore WEC, Murray RGE, Stackebrandt E et al (1987) Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int J Syst Bacteriol* 37:463–464
17. Ursing JB, Roselló-Mora RA, García-Valdés E, Lalucat J (1995) Taxonomic note: a pragmatic approach to the nomenclature of phenotypically similar genomic groups. *Int J Syst Bacteriol* 45:604
18. Mayr E (1998) Two empires of life or three? *Proc Natl Acad Sci USA* 95:9720–9723
19. Staley JT (1997) Biodiversity: are microbial species threatened? *Curr Opin Biotechnol* 8:340–345
20. Baas-Becking LGM (1934) *Geobiologie of inleiding tot de milieukunde*. Van Stockkum & Zoon, The Hague, The Netherlands, p 263
21. Finlay BJ, Esteban GF, Fenchel T (1996) Global diversity and body size. *Nature* 383:132–133
22. Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. *Science* 296:1061–1063
23. Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic barriers isolate endemic populations of hyperthermophilic *Archaea*. *Science* 301:976–978
24. Moran NA, Munson MA, Baumann P, Ishikawa H (1993) A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Phil Trans R Soc B* 253:161–171
25. Hedlund BP, Staley JT (2002) Phylogeny of the genus *Simonsiella* and other members of the Neisseriaceae. *Int J Syst Evol Microbiol* 52:1377–1382