PERSPECTIVES

OPINION

# Re-evaluating prokaryotic species

*Dirk Gevers, Frederick M. Cohan, Jeffrey G. Lawrence, Brian G. Spratt, Tom Coenye, Edward J. Feil, Erko Stackebrandt, Yves Van de Peer, Peter Vandamme, Fabiano L. Thompson and Jean Swings*

Abstract | There is no widely accepted concept of species for prokaryotes, and assignment of isolates to species is based on measures of phenotypic or genome similarity. The current methods for defining prokaryotic species are inadequate and incapable of keeping pace with the levels of diversity that are being uncovered in nature. Prokaryotic taxonomy is being influenced by advances in microbial population genetics, ecology and genomics, and by the ease with which sequence data can be obtained. Here, we review the classical approaches to prokaryotic species definition and discuss the current and future impact of multilocus nucleotide-sequence-based approaches to prokaryotic systematics. We also consider the potential, and difficulties, of assigning species status to biologically or ecologically meaningful sequence clusters.

Attempts to bring order through categorization to the bewildering variety of organisms with which we share the planet have been an ongoing human endeavour. For higher organisms, the species has a special status as, unlike broader taxonomic categories, evolutionary and ecological processes underpin the concept of species. By contrast, species demarcation in prokaryotes is not defined by a theory-based concept and tends to be more arbitrary, anthropocentric or rooted in practical necessity. In the case of prokaryotic pathogens, species are historically defined on the basis of the disease they cause, regardless of other ecological or evolutionary considerations (for example, *Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Bacillus anthracis*, or *Mycobacterium tuberculosis*). Although the status of a species is difficult to justify from a broad conceptual viewpoint, these familiar labels do have a vital role in clinical settings, the food-processing industry, agriculture, bioremediation, public health, environmental science and biosafety. The names of species of human interest yield an implicit understanding of what that organism is likely to do, or is capable of doing, based on past experience. It is less clear, however, whether our species demarcations provide this information for the vast majority of prokaryotes that are never going to cause gum disease, will never be weaponized and will never prove useful in treating raw sewage or making yoghurt. Whereas practicalities dictate that many, if not most, existing species labels remain intact, we are forced to concede that the status of species confers little evolutionary and ecological meaning in prokaryotes in general. Given the recent advances in sequencing technology and commensurate advances in our understanding of evolutionary processes in prokaryotes, it is timely to reconsider the extent to which it is possible to provide a closer marriage between an operational species 'definition' and a theory-based species 'concept' for those taxa that are currently poorly characterized or yet to be discovered.

## Defining prokaryotic species

***Defining prokaryotic species by DNA–DNA hybridization.*** Prokaryotic species are currently characterized using a polyphasic approach that incorporates genotypic and phenotypic (including chemotaxonomic) properties[1,2] (BOX 1). Since the 1970s, the cornerstone of genotypic characterization has been the measurement of overall genetic similarity among isolates, assessed by the degree to which their genomes hybridize under standard conditions (DNA–DNA hybridization or DDH). Here, both similarity in gene content and nucleotide similarity of shared genes contribute to a measure of the overall relatedness of their genomes. Prior to the availability of DNA sequence information, DDH provided a standardized means for identifying and classifying prokaryotes that lack well defined morphological or phenotypic characteristics.

Despite being introduced several decades ago, DDH data are consistent with recent results from complete genome sequences and multilocus data. Nevertheless, DDH has serious limitations: it is a time-consuming procedure, carried out properly by few laboratories only, and it is ill-suited for rapid identification of prokaryotes. It is also unavailable for the classification of prokaryotes that are currently non-culturable, now recognized to be most prokaryotes in the biosphere[3]. More importantly, classification by DDH requires pairwise comparisons of two prokaryotic genomes, and individual strains cannot be analysed and compared with a database using a codified set of criteria to assign it to a known taxon or to propose a new taxon. We have no ecological or evolutionary theory for how species demarcations should relate to the extent of gene sharing between organisms. The recommendation to delineate species using a 70% DNA–DNA binding criterion does not correspond to a theory-based concept of what properties a species should have, but was calibrated empirically to yield many of the phenotype-based species already recognized at the time of its inception[4].

Box 1 | **Polyphasic taxonomy of prokaryotes**

In polyphasic taxonomy, microbiologists aim to come to a consensus classification by integrating different kinds of data and information into a classification of the biological entities that contains a minimum of contradictions[2]. This includes phenotypic data (for example, the results of biochemical tests, fatty-acid composition), genotypic data (for example, DNA fingerprint data) and phylogenetic information (for example, rRNA gene sequences).

At present, a prokaryotic species is defined as "a category that circumscribes a (preferably) genomically coherent group of individual isolates/strains sharing a high degree of similarity in (many) independent features, comparatively tested under highly standardised conditions"[1]. Practically, a prokaryotic species is considered to be a group of strains (including the type strain) that are characterized by a certain degree of phenotypic consistency, showing 70% of DNA–DNA binding and over 97% of 16S ribosomal RNA (rRNA) gene-sequence identity[2].

Since the early 1970s, prokaryotic species have been delineated based on data generated in DNA–DNA-hybridization experiments. In this approach, the overall genetic similarity among isolates is assessed by the degree to which their genomes hybridize under standardized conditions. Isolates that show more than 70% DNA–DNA-binding values and less than 5% difference in their melting temperature (ΔTm) are considered to belong to the same species (although this threshold value might be adjusted by expert practitioners that are familiar with the vagaries of individual taxonomic groups), whereas isolates that share less than 50% DNA–DNA-binding values definitely belong to different species. To facilitate identification, phenotypic consistency within the species and differences between species are required. Comparative sequence analysis of the 16S rRNA gene is extensively used to determine the phylogenetic position of novel isolates. Strains that show <97% 16S rRNA sequence similarity to all known taxa are considered to belong to a new species, as there are no examples in which strains with this extent of divergence in 16S rRNA sequence meet the criteria of >70% DNA–DNA binding.

In practice, the taxonomic characterization of a collection of isolates starts with a screening that allows the more-closely related isolates to be clustered and to be distinguished from unrelated isolates. Often-used screening methods include whole-cell fatty-acid analysis (a chemotaxonomic method in which the lipids that are present in bacterial cells are analysed and used to delineate clusters)[31] and a wide range of DNA-based typing methods, including amplified fragment length polymorphism fingerprinting[32] and several PCR-based methods using random primers[33]. These methods have the advantage of being relatively quick and easy to carry out and can be automated, making them particularly useful for the standardized screening of a large number of isolates. Subsequently, 16S rRNA gene-sequence analysis is carried out on representatives of the different clusters, and these sequences are compared with those of known species. In this way, the phylogenetic position is determined. Based on the results of this analysis, organisms are selected that need to be included in DNA–DNA-hybridization experiments.

***Defining prokaryotic species by a rRNA gene-sequence approach.*** The advent of rapid and cost-effective DNA sequence analysis has circumvented the need for physicochemical measures of genomic similarity. As pioneered by Woese[5], ribosomal RNA (rRNA) gene-sequence similarity allows the classification of prokaryotes using a universally distributed trait. Importantly, classification of organisms in this manner employs sequence databases. By comparing the rRNA gene sequence of an isolate to those of all known prokaryotic species, microbiologists can rapidly identify prokaryotic strains. In addition, microbiologists who explore the vast diversity of prokaryotic life using molecular approaches can place their findings in a universal context even when the isolates cannot be cultured.

Although the advantages of the direct genotypic approaches are clear, classification by rRNA gene sequence alone — an increasingly common practice — is unsatisfying for several reasons. The rRNA gene sequences often lack resolution when compared with DDH. Whereas isolates that have less than 97% rRNA gene-sequence similarity usually share <70% DDH and belong to different species, isolates that have ≥97% identity might or might not meet the 70% DDH criterion for inclusion in the same species[6,7] (FIG. 1). Consequently, near identity of rRNA gene sequences does not eliminate the need to apply other methods to further explore whether isolates are sufficiently similar to be assigned to the same species. Most importantly, despite the perceived reliability of 16S rRNA gene sequence as a phylogenetic marker, any single measure of sequence similarity is subject both to simple stochastic variation and to the influence of recombination or horizontal gene transfer[8,9].

***Defining prokaryotic species by a multilocus sequencing approach.*** Sequence-based approaches to defining species require loci that evolve more rapidly than rRNA genes, and multiple genes provide a buffer against the distorting effects of recombination at a single locus. This is the premise that underlies the development of multilocus sequence typing (MLST (see Glossary)), a method for the genotypic characterization of prokaryotes at the infraspecific level using the allelic mismatches of a small number (usually 7) of housekeeping genes[10]. Groups of isolates with identical allelic profiles define strains or clones and form the basis for a genotypic classification system. Therefore, whereas the rRNA gene sequence might assign an isolate to a genus, MLST might assist in grouping isolates into the major genetic lineages within a species[11].

MLST refers to a specific tool designed for molecular epidemiology and for defining strains within named species. To be applied to the problem of SPECIES DEFINITION, more diverse groups of strains, incorporating whole genera, would need to be characterized. This would invalidate the use of simple clustering procedures based on the number of allelic mismatches, as most taxa would differ at all sequenced loci, and therefore phylogenetic procedures based on the nucleotide sequences of the alleles would be employed instead (BOX 2). To emphasize the distinction from MLST, we propose that the more general term multilocus sequence analysis (MLSA) be used for such purposes.

Rapid and robust classification with MLSA could use a universal set of genes, which would allow for the hierarchical classification of all prokaryotes[12,13]. However, this approach might be impractical, as genes that are informative within a genus or family might not be useful or even present in more distantly related taxa. Also, genes that are conserved enough to be amplified by a set of primers common to all species might not evolve quickly enough to distinguish closely related taxa. Because species-level taxonomy is not primarily concerned with the deeper phylogenetic relationships but instead with demarcating species within a given genus or family, our priority is to look for sets of genes that can be used with all strains from a particular group (a genus or family). Nevertheless, some genes might be informative in more than one group-specific set, and these more widely distributed genes could provide tools for broader comparisons.

MLSA should use genes that are ubiquitous (at least in the taxon under study) and in single copy, and should avoid those genes

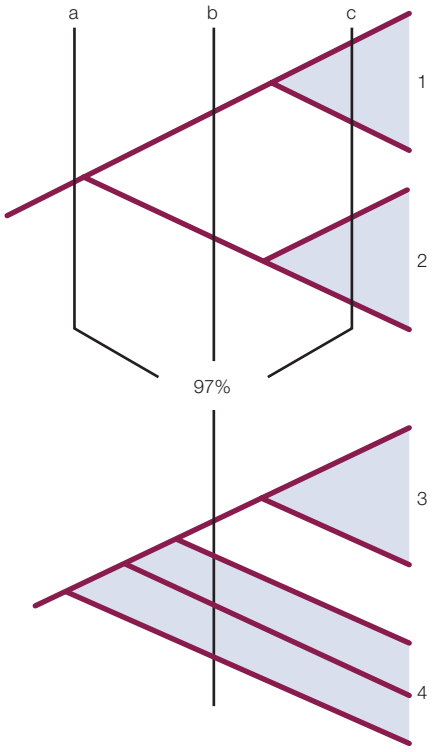



Figure 1 | **Failure of threshold methods in delineating prokaryotic species.** Arbitrary thresholds, such as 97% sequence identity, will fail when more than one cohesive group (1 & 2) is encompassed (position a) or when a single cohesive group is split (position c). Any threshold will fail when one group (taxon 4) is paraphyletic and contains another group (taxon 3) within its diversity.

in which recombination might confer a selective advantage (for example, virulence and antigen-encoding genes) or closely linked genes.

The simplest approach is to concatenate the sequences of the multiple genes and to use the concatenated sequences to construct a tree, which can identify deeply branching clusters that can help the division of the genus into species. Prokaryotic identification in this scenario is a two-step process: rRNA sequencing assigns an unknown strain to a group (genus or family), which defines which genes and primers are to be used for MLSA to assign the strain to a species.

This MLSA approach has been used in several recent studies[14–18]. For example, MLSA showed the clear separation of *Burkholderia pseudomallei* isolates from those of *Burkholderia thailandensis*, which is consistent with their status as separate species (BOX 2), and showed that *Burkholderia mallei* is a clone within *B. pseudomallei*, raised to species status owing to its ability to cause a distinctive disease in equines. Isolates assigned to the *Bacillus cereus* group (*Bacillus cereus*, *B. anthracis*, *Bacillus thuringiensis* and *Bacillus mycoides*) have been similarly characterized[17], producing results that are consistent with other approaches. Strains of the insect pathogen *B. thuringiensis* correspond to several distinct lineages within the phylogeny of *B. cereus*, a free-living soil bacterium. *B. anthracis*, however, is more homogenous and is a specialized clone of *B. cereus*. Were it not for the capacity of this clone to cause anthrax, it is unlikely that it would have been recognized as a separate species[17].

## What are prokaryotic species?

MLSA is clearly capable of yielding sequence clusters at a wide range of taxonomic levels, from the infraspecific clonal complexes of MLST, over clusters corresponding to named species, to clusters at higher levels. An important problem is deciding at what depth of clustering to define species and how to incorporate ecology into species definitions. The demarcation of familiar species could in principle be used to calibrate the approximate range of diversity that is expected to encompass newly characterized species. However, as mentioned previously, many named species demarcations are not rooted in evolutionary or ecological theory, so that many species are extremely diverse in their metabolic capabilities[19], in the gene content of their genomes[20] and in their ecology[21]. Indeed, various molecular techniques, including PCR-based and restriction-based methods[21] as well as MLST[22,23], have subclassified named species into multiple clusters that might frequently be ecologically distinct. Conversely, other named species are extremely narrowly defined, particularly among those pathogens in which ecological specialization is all too obvious, resulting in the onset of distinctive diseases, such as anthrax (*B. anthracis*), gonorrhoea (*N. gonorrhoeae*), glanders[18], tuberculosis[24] and the plague (*Yersinia pestis*).

Defining species limits by using levels of sequence similarity typically found within existing named species is clearly inappropriate. A more attractive approach is to seek ecological, genomic or phenotypic differences among the major clusters resolved by MLSA that would justify their separation into species. Ideally, we would like a theory-based approach to the definiton of species based on MLSA data. Cohan has proposed that bacterial species could be split into smaller, more meaningful units by incorporating the concept of the ecotype, and that an 'ecotype model' could provide a rational basis for demarcating bacterial taxa[4,25]. Ecotypes are defined as populations that are genetically cohesive and ecologically distinct. Cohesion results from periodic selection events that recurrently purge each ecotype of its genetic diversity (BOX 3). Moreover, ecotypes are expected to be irreversibly separate from one another. Ecotypes therefore hold all the quintessential properties of species as understood in systematics outside of microbiology[26].

From the standpoint of taxonomy, there are many circumstances under which ecotypes are expected to be discerned as distinct sequence clusters, although there are other plausible scenarios in which they are not (BOX 3). When putative ecotypes are shown to have both a history of coexistence (by being in long-divergent sequence clusters) and a prognosis for future coexistence (by using different ecological resources in nature), we could rationally give these groups taxonomic status, either as species or as infraspecific taxa. We note that such an approach raises two fundamental difficulties. First, many ecological adaptations will be difficult to identify, especially when we are interested in differences that allow coexistence of populations in nature. Nevertheless, genomic approaches might facilitate investigations of ecological specialization, for example, by comparing genome content and genome-wide gene expression[27]. Second, splitting of existing species into multiple closely related species would yield a cumbersome taxonomy and would be unpalatable to many taxonomists. One possible solution brought up for discussion is to grant legacy status to existing

### Glossary

MLST
Multilocus sequence typing, a method for the genotypic characterization of prokaryotes at the infraspecific level, using the allelic mismatches of a small number (usually 7) of housekeeping genes. Designed as a tool in molecular epidemiology and used for recognizing distinct strains within named species.

MLSA
Multilocus sequence analysis, a method for the genotypic characterization of a more diverse group of prokaryotes (including entire genera) using the sequences of multiple protein-coding genes.

SPECIES CONCEPT
A framework to understand how and why an observer can sort organisms into species; that is, what kind of unit do we think the term species embraces, and what characteristics are shared between all members of a species.

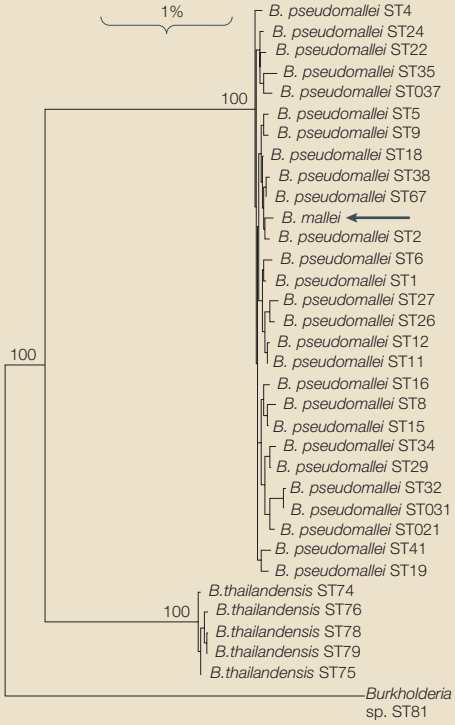
SPECIES DEFINITION
A more practical outline of how to assign isolates to a named species or identify new species.

Box 2 | **Taxonomic relationships of *Burkholderia* spp.: insights from MLSA**

***Burkholderia mallei*** **is the causative agent of equine glanders, an acute infection that is characterized by either pneumonia and necrosis of the tracheobronchial tree if the organism is inhaled, or pustular skin lesions, multiple abscesses and sepsis if the skin is the portal of entry.** ***Burkholderia pseudomallei*** **causes melioidosis, an endemic disease in northern Australia and South-East Asia. Melioidosis is characterized by a broad spectrum of clinical manifestations, ranging from asymptomatic colonization to fulminant sepsis. In contrast to *B. mallei*, which is an obligate parasite of horses, mules and donkeys with no other known natural reservoir, *B. pseudomallei* is a saprophytic organism that is broadly distributed in water and soil in its endemic regions. *B. mallei* and *B. pseudomallei* are phylogenetically closely related, and these two species were shown to represent a single genomic species by DNA–DNA hybridization (DDH) criteria[34]. The observed differences in 'behaviour' between *B. mallei* and *B. pseudomallei* might relate to differences in gene content, as revealed by study of the genome sequences[35,36].**

**Recently, several avirulent, environmental *B. pseudomallei*-like organisms were formally classified as *Burkholderia thailandensis*[37]. The 16S rRNA gene sequences of these three *Burkholderia* species are more than 99% identical. Whereas it was previously shown that, based on DDH, *B. mallei* and *B. pseudomallei* belong to the same taxon (as they show levels of DNA–DNA binding higher than 76%)[38], *B. thailandensis* is clearly different from *B. pseudomallei*, as DDH values as low as 47% were found between representatives of both taxa[39].**

**A thorough assessment of the genetic relationships between *B. mallei*, *B. pseudomallei* and *B. thailandensis* was urgently needed. Godoy *et al.*[18] recently reported on the use of multilocus sequence analysis (MLSA) to elucidate these relationships. In this study, they determined the sequence of internal fragments (~550 bp) of seven housekeeping genes (*ace*, *gltB*, *gmhD*, *lepA*, *lipA*, *narK* and *ndh*) for a large collection of *B. mallei*, *B. pseudomallei* and *B. thailandensis* isolates. These sequences were then concatenated into a single sequence of 3,399 bp, and this sequence was used to construct a phylogenetic tree (see figure). The analysis showed that all *B. pseudomallei* isolates were tightly clustered and well separated from the *B. thailandensis* isolates. In addition, all *B. mallei* isolates had identical sequences and clustered with the *B. pseudomallei* isolates. The position of the *B. mallei* clone within *B. pseudomallei* is indicated with an arrow.**




**These data unambiguously show that *B. pseudomallei* and *B. thailandensis* are different species (although they have virtually identical 16S rRNA gene sequences), and that *B. mallei* is a clone (or a specific 'ecotype'[4]) of *B. pseudomallei*. This means that, on taxonomic and on population-genetics grounds, *B. pseudomallei* and *B. mallei* should not be given separate species names, although the important differences in their biochemical activities and in the clinical symptoms and epidemiology justify this two-species concept.**

**This example clearly illustrates the value of the MLSA approach for resolving taxonomic ambiguities.**

species, while naming confirmed ecotypes within these species by an infraspecific label such as ecovar[4]. Defining species as ecotypes might remain reserved for those groups whose ecological distinctness is obvious and important (for example, pathogens), as well as species yet to be discovered.

## Beyond MLSA

MLSA is attractive in that it avoids the vagaries of a single-gene approach and provides an objective method to cluster strains within a genus. Moreover, it is a high-throughput methodology that is amenable to automation and, by using digital data, allows taxonomic assignment through the internet, a solution to the burden of routine species identification. But this approach is not without its flaws, in that species assigned as clusters of multilocus genotypes remain empirically defined and subject to sampling bias. For example, members of a species cannot simply be delineated as strains with some defined level of similarity to a type strain. The composition of the clusters, and their relationships, will change as more strains are genotyped and large numbers of isolates will need to be characterized to produce robust groupings of strains. However, a recent study provides some encouragement. Using a set of strains of two closely related species (*N. meningitidis* and *Neisseria lactamica*) that both colonize the human nasopharynx and are known to exchange genes, and strains of *N. gonorrhoeae*, a tree based on the concatenated sequences of the seven MLSA loci provided a good separation of the two named nasopharyngeal taxa, whereas, owing to recombination, single loci completely failed to do so[28]. If nothing had been known about the taxonomic status of these strains, the MLSA clustering patterns, combined with a consideration of the ecological context, would lead to three clear groups — a cluster associated with nasopharyngeal carriage and meningitis (*N. meningitidis*), one with carriage but not disease (*N. lactamica*) and one with genital colonization and genital disease (*N. gonorrhoeae*).

Another problem is that a single set of criteria or rules cannot be applied fairly to groups in which significantly different biological processes have led to their diversification and are responsible for their cohesion. For example, MLSA clusters of highly recombinogenic strains of *Neisseria* spp.[29] probably correspond to groups with a different ecological breadth than comparable clusters of clonal *M. tuberculosis*[24]. This distinction is not surprising for microbiologists familiar with these organisms; therefore, effective use of an MLSA database should incorporate important aspects of each organism's biology and a certain pragmatism, not strict definitions and rules. It should be stressed that discrete sequence clusters might exist in some genera but not in others, and that new approaches to taxonomy should not be judged on their ability to unfailingly identify species or to produce the same groupings identified by traditional taxonomic approaches. Bacterial species are not necessarily natural entities that arise as a consequence of evolutionary processes, and disagreement and frequent taxonomic revision within a group might in some cases

reflect an underlying reality that genetic variation is not neatly partitioned into consistent entities that we can define as species, whatever method we apply.

Although the MLSA approach outlined above is practical in grouping strains together, it uses core genes and ignores genes that lead to potentially significant differences among strains. Important phenotypic or ecological differences might exist among strains that are assigned to related clusters by MLSA, but showing ecological differentiation that justifies their separation into distinct species or ecotypes might be difficult. Our current ignorance of the genetic basis of ecological differentiation makes it unlikely that even the availability of complete genome sequences will provide assistance in predicting whether clusters are probably ecologically distinct.

In traditional taxonomy, the discrimination between prokaryotic species implies that organisms within a group share phenotypic characteristics that are fundamentally different to strains from other groups and which can be used to distinguish these species. But the identification of significant phenotypic differences is neither easy nor objective and, if species are split to recognize ecological differences among clusters, simple tests to identify ecotypes might not be readily achieved.

Despite these difficulties, the combined analysis of complete genome sequences, MLSA data and ecological data will greatly help the improved delineation of biologically meaningful taxonomic groups in prokaryotes[30]. Species definition and delineation requires a taxonomic framework and suitable software tools, which can be achieved if MLSA data are augmented, in central curated databases, with additional information — relating to ecological adaptation, population structure, rates of recombination, amount of horizontal gene transfer or source of isolation — that would arbitrate species assignment. In the absence of a universal SPECIES CONCEPT, MLSA presents a highly valuable 'baseline' for creating and maintaining operational protocols for species assignment.

## Future prospects

The rapid advances in molecular and genome biology have forced microbiologists to revisit one of the oldest branches of their discipline, systematics. A sequence-based approach, in which comparisons between strains are carried out by a computer and not a person at a laboratory bench, is clearly the future of microbial systematics and taxonomy. Beyond the lack of ecological context and necessity to retain 'legacy' nomenclature reflecting important, previously described groups,

### Box 3 | Models for a theory-based taxonomy of prokaryotes

In the stable ecotype model, ecotypes are created and extinguished at a low rate. Each ecotype (indicated by colour — panel a) undergoes a series of periodic selection events (indicated by asterisks) during its long history of divergence from other such ecotypes. In each periodic selection event, an adaptive mutant out-competes to extinction the competing lineages within the ecotype; owing to rare recombination, periodic selection purges sequence diversity at nearly all loci[4]. During its lifetime, an ecotype is recurrently purged of its diversity by these selection events, whereas divergence among ecotypes is not[25,27,40]. Most such ecotypes (of sufficient age) should be distinguishable from other ecotypes as separate sequence clusters, provided that effective population sizes of the bacteria are large ($>10^9$)[4,41,42].

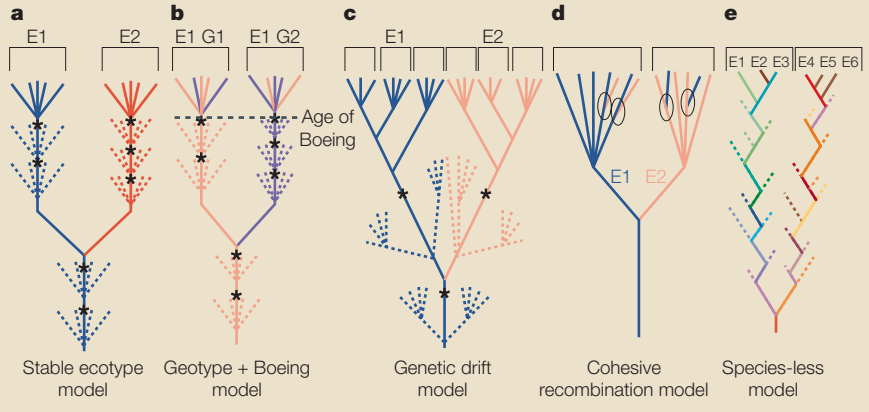
There are alternative scenarios in which ecotypes might each contain multiple sequence clusters. In the geotype-plus-Boeing model, geographically isolated populations of the same ecotype (or 'geotypes', indicated by colour — panel b) first diverge historically into separate sequence clusters and then, in recent decades, jet planes carry all the endemic clusters of a single ecotype into each region of the world[25,27,43]. In this transitional era when air travel (and to some extent even sea travel) is still new, we might see multiple sequence clusters within one ecotype at one place[25].

Genetic drift can also produce multiple clusters within each ecotype (panel c), when effective population sizes are not enormous (for example, in pathogens, owing to bottlenecks of transmission)[44].

We next consider evolutionary models that yield multiple ecotypes per sequence cluster. This might occur in the cohesive recombination model, in which multiple, closely related ecotypes genetically recombine at a high rate (recombination indicated by ovals — panel d), retarding sequence divergence at loci that are not responsible for adaptive divergence (for example, loci used in multilocus sequence analysis (MLSA)). The outcome can be a reduction in the distinctness of ecotypes for genes used in MLSA[41,45]. In the species-less model, there are no cohesive groups[27]. There is frequent invention and extinction of ecologically distinct populations and little periodic selection (panel e). The diversity within a population is not constrained into the indefinite future by periodic selection; instead, diversity is constrained only by the short amount of time from the population's founding from a single mutant (or recombinant) to the time the population goes extinct. Here, a single sequence cluster might contain multiple, ecologically distinct (and young) populations.

In summary, sequence clusters can be guides to discovering ecologically distinct populations, especially under the stable ecotype model, in which ecotypes and sequence clusters correspond nearly one-to-one, but other relationships between ecotypes and sequence clusters are possible. Nevertheless, a general sequence-based approach can be developed to identify ecologically significant groups while accommodating these different models of evolution.

One possibility is that systematists might base their classification on the smallest groups that have a history of coexistence as separate lineages and a prognosis of future coexistence as separate lineages[25]. A history of coexistence can be inferred from the groups being in long-divergent multilocus sequence clusters; past coexistence of extremely young groups might be inferred from clustering according to quickly evolving VNTR (variable number of tandem repeats) markers[46]. A prognosis for future coexistence can be based on an assessment that the groups are ecologically distinct in the resources they use in nature[25]. Identifying taxa as the smallest groups known to be ecologically distinct and to form distinct sequence clusters has successfully delimited the functionally distinct players within a bacterial community; moreover, this is the *de facto* means by which microbial ecologists discover groups of ecological significance[47].




Ecotypes are labeled as E1, E2, E3 and so on. In panel b, geotypes within ecotype 1 are labeled E1 G1 and E1 G2.

several issues remain to be resolved. First, it might not be possible to delineate groups within a continuous spectrum of genotypic variation; that is, clustering might not occur, or might include different groups of strains depending on the sample used. Second, it might be difficult or impossible to recognize a new group when only one or few genotypes have been isolated. Third, important phenotypes might not be conferred by stable chromosomal loci, resulting in marked differences among strains that are closely related. Indeed, important pathogenicity determinants in *Salmonella enterica* and *B. anthracis* are plasmid-encoded. Fourth, with new taxa, it might be unclear where to draw the distinction between clusters that should be considered to define species and those that are subgroups within a single species.Last, it is not straightforward to apply MLSA to uncultured material unless the population that is being sampled is dominated by a single species or strain, although approaches are being developed to analyse linked sets of genes from a single uncultured organism[27].

Despite these limitations, an MLSA framework moves prokaryotic taxonomy towards a rapid, flexible and objective standard with sufficient flexibility to accommodate the vast differences in biology presented by prokaryotes, the most abundant organisms on the planet. Incorporation of ecological data into the MLSA framework will allow for meaningful taxonomic assignments, whereby taxa are delineated by virtue of important ecological traits as well as degrees of genetic relatedness.

We propose that large-scale studies of well known genera, preferably with differing levels of recombination, should be carried out to establish the patterns of clustering obtained with MLSA data and to provide a framework for further studies that can explore the biological, ecological and genetic differentiation between the observed clusters. It would be interesting to include some genera in which species resolution has been highly contentious and subject to frequent revision, to establish whether there are situations in which there is a genetic continuum where no procedures can provide robust species clusters. These studies should provide criteria (pragmatic instead of rules and cut-offs) to determine how we should use MLSA data to define species and subdivisions within species for poorly studied or newly discovered taxa. The next few years will undoubtedly be both challenging and fascinating for taxonomists, and rewarding for those working to understand species delineation and species concepts in prokaryotes.

*Dirk Gevers is at the Laboratory of Microbiology and the Bioinformatics and Evolutionary Genomics Research Group, Ghent University/VIB, Ghent, Belgium.*

*Frederick M. Cohan is at the Department of Biology, Wesleyan University, Middletown, Connecticut, USA.*

*Jeffrey G. Lawrence is at the Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.*

*Brian G. Spratt is at the Department of Infectious Disease Epidemiology, Imperial College London, London, UK.*

*Tom Coenye and Peter Vandamme are at the Laboratory of Microbiology, Ghent University, Ghent, Belgium.*

*Edward J. Feil is at the Department of Biology and Biochemistry, University of Bath, Bath, UK.*

*Erko Stackebrandt is at the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany.*

*Yves Van de Peer is at the Bioinformatics and Evolutionary Genomics Research Group, Ghent University/VIB, Ghent, Belgium.*

*Fabiano L. Thompson is at the Centro Pluridisciplinar de Pesquisas Químicas, Biológicas e Agrícolas, Universidade Estadual de Campinas, Campinas, Brazil.*

*Jean Swings is at the Laboratory of Microbiology and the BCCM™/LMG Bacteria Collection, Ghent University, Ghent, Belgium.*

*Correspondence to D.G.*
*e-mail: dirk.gevers@UGent.be*

doi:10.1038/nrmicro1236
Published online 10 August 2005

1. Stackebrandt, E. *et al.* Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* **52**, 1043–1047 (2002).
2. Vandamme, P. *et al.* Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* **60**, 407–438 (1996).
3. Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).
4. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487 (2002).
5. Woese, C. R. The Use of Ribosomal RNA in Reconstructing Evolutionary Relationships Among Bacteria. in *Evolution at the Molecular Level* (eds Selander, R. K., Clark, A. G. & Whittam, T. S.) Ch. 1, 1–24 (Sinauer Associates Inc., Sunderland, 1991).
6. Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P. Jr. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int. J. Syst. Bacteriol.* **42**, 166–170 (1992).
7. Stackebrandt, E. & Goebel, B. M. A place for DNA–DNA reassociation and 16S ribosomal-RNA sequence-analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* **44**, 846–849 (1994).
8. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19**, 2226–2238 (2002).
9. Boucher, Y., Douady, C. J., Sharma, A. K., Kamekura, M. & Doolittle, W. F. Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J. Bacteriol.* **186**, 3980–3990 (2004).
10. Maiden, M. C. J. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl Acad. Sci. USA* **95**, 3140–3145 (1998).
11. Cooper, J. E. & Feil, E. J. Multilocus sequence typing — what is resolved? *Trends Microbiol.* **12**, 373–377 (2004).
12. Zeigler, D. R. Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int. J. Syst. Evol. Microbiol.* **53**, 1893–1900 (2003).
13. Santos, S. R. & Ochman, H. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ. Microbiol.* **6**, 754–759 (2004).
14. Thompson, F. L. *et al.* Phylogeny and molecular identification of vibrios based on multilocus sequence analysis. *Appl. Environ. Microbiol.* (in the press).
15. Devulder, G. , Perouse de Montclos, M. & Flandrois, J. P. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int. J. Syst. Evol. Microbiol.* **55**, 293–302 (2005).
16. Kotetishvili, M. *et al.* Multilocus sequence typing for studying genetic relationships among *Yersinia* species. *J. Clin. Microbiol.* **43**, 2674–2684 (2005).
17. Priest, F. G. , Barker, M., Baillie, L. W. J., Holmes, E. C. & Maiden, M. C. J. Population structure and evolution of the *Bacillus cereus* group. *J. Bacteriol.* **186**, 7959–7970 (2004).
18. Godoy, D. *et al.* Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J. Clin. Microbiol.* **41**, 2068–2079 (2003).
19. Feldgarden, M., Byrd, N. & Cohan, F. M. Gradual evolution in bacteria: evidence from *Bacillus* systematics. *Microbiology* **149**, 3565–3573 (2003).
20. Welch, R. A. *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 17020–17024 (2002).
21. Schloter, M. , Lebuhn, M., Heulin, T. & Hartmann, A. Ecology and evolution of bacterial microdiversity. *FEMS Microbiol. Rev.* **24**, 647–660 (2000).
22. Jolley, K. A. *et al.* Carried meningococci in the Czech Republic: a diverse recombining population. *J. Clin. Microbiol.* **38**, 4492–4498 (2000).
23. Enright, M. C., Spratt, B. G., Kalia, A., Cross, J. H. & Bessen, D. E. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect. Immun.* **69**, 2416–2427 (2001).
24. Sreevatsan, S. *et al.* Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl Acad. Sci. USA* **94**, 9869–9874 (1997).
25. Godreuil, S. , Cohan, F. M., Shah, H. & Tibayrenc, M. Which species concept for pathogenic bacteria?: an e-debate. *Infect. Genet. Evol.* 20 Jan 2005 (doi/10.1016/j.meegid.2004.03.004).
26. de Queiroz, K. in *Endless Forms: Species and Speciation* (eds Howard, D. J. & Berlocher, S. H.) Ch. 5, 57–75 (Oxford University Press, Oxford, 1998).
27. Ward, D. M. & Cohan, F. M. Microbial Diversity in Hot Spring Cyanobacterial Mats: Pattern and Prediction. in *Geothermal Biology and Geochemistry in Yellowstone National Park* (eds Inskeep, W. P. & McDermott, T.) in the press (Thermal Biology Institute, Bozeman).
28. Hanage, W. P. , Fraser, C. & Spratt, B. G. Fuzzy species among recombinogenic bacteria. *BMC Biol.* **3**, 6 (2005).
29. Feil, E. J. *et al.* Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl Acad. Sci. USA* **98**, 182–187 (2001).
30. Field, D., Feil, E. J. & Wilson, G. Databases and software for the comparative genomic study of collections of genomes. *Microbiology* (in the press).
31. Welch, D. F. Applications of cellular fatty acid analysis. *Clin. Microbiol. Rev.* **4**, 422–438 (1991).
32. Vos, P. *et al.* AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414 (1995).
33. van Belkum, A. DNA fingerprinting of medically important microorganisms by use of PCR. *Clin. Microbiol. Rev.* **7**, 174–184 (1994).
34. Coenye, T. & Vandamme, P. Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ. Microbiol.* **5**, 719–729 (2003).
35. Holden, M. T. G. *et al.* Genornic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc. Natl Acad. Sci. USA* **101**, 14240–14245 (2004).
36. Nierman, W. C. *et al.* Structural flexibility in the *Burkholderia mallei* genome. *Proc. Natl Acad. Sci. USA* **101**, 14246–14251 (2004).
37. Brett, P. J., Deshazer, D., & Woods, D. E. *Burkholderia thailandensis* sp. nov., a *Burkholderia pseudomallei*-like species. *Int. J. Syst. Bacteriol.* **48**, 317–320 (1998).
38. Rogul, M., Brendle, J. J., Haapala, D. K. & Alexander, A. D. Nucleic acid similarities among *Pseudomonas pseudomallei*, *Pseudomonas multivorans*, and *Actinobacillus mallei*. *J. Bacteriol.* **101**, 827–835 (1970).

39. Yabuuchi, E. *et al. Burkholderia uboniae* sp. nov., L-arabinose-assimilating but different from *Burkholderia thailandensis* and *Burkholderia vietnamiensis*. *Microbiol. Immunol.* **44**, 307–317 (2000).
40. Cohan, F. M. Concepts of Bacterial Biodiversity for the Age of Genomics. in *Microbial Genomes* (eds Fraser, C. M., Read, T. D. & Nelson, K. E.) Ch. 11, 175–194 (Humana Press, Totowa, 2004).
41. Lawrence, J. G. Gene transfer in bacteria: speciation without species. *Theor. Popul. Biol.* **61**, 449–460 (2002).
42. Palys, T., Nakamura, L. K. & Cohan, F. M. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int. J. Syst. Bacteriol.* **47**, 1145–1156 (1997).
43. Papke, R. T., Ramsing, N. B., Bateson, M. M. & Ward, D. M. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**, 650–659 (2003).
44. Fraser, C., Hanage, W. P. & Spratt, B. G. Neutral microepidemic evolution of bacterial pathogens? *Proc. Natl Acad. Sci. USA* **102**, 1968–1973 (2005).
45. Cohan, F. M. Does recombination constrain neutral divergence among bacterial taxa. *Evolution* **49**, 164–175 (1995).
46. Keim, P. *et al.* Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infect. Genet. Evol.* **4**, 205–213 (2004).
47. Dunlap, P. V. and Ast, J. C. Genomic and phylogenetic characterization of luminous bacteria symbiotic with the deep-sea fish *Chlorophthalmus albatrossis* (Aulopiformes: Chlorophthalmidae). *Appl. Environ. Microbiol.* **71**, 930–939 (2005).



## Acknowledgements

This paper is the outcome of a workshop entitled 'The prokaryotic species: genome plasticity, microevolution and taxonomy', organized by F. Thompson and J. Swings in Ghent, Belgium, 25 October 2004, and has the intention to stir up the interdisciplinary debate on the prokaryotic species. We thank all participants as well as D. Mazel, G. Manfio and T. Iida for their contributions. T.C., P.V. and J.S. are indebted to the Fund for Scientific Research — Flanders (Belgium) for a position as postdoctoral fellow and research funding, respectively. F.M.C. acknowledges funding from the National Science Foundation.


## Competing interests statement

The authors declare no competing financial interests.

## Online links

**FURTHER INFORMATION**

**Get into the debate on the prokaryotic species at:**
http://bioinformatics.psb.ugent.be/publications/prokspec
**Access to this interactive links box is free online.**