

Available online at www.sciencedirect.com

Infection, Genetics and Evolution xxx (2005) xxx–xxx

Infection, Genetics
and Evolutionwww.elsevier.com/locate/meegid

Discussion

Which species concept for pathogenic bacteria? An E-Debate

Sylvain Godreuil^a, Frederick Cohan^b, Haroun Shah^c, Michel Tibayrenc^{a,*}^a*Génétique et Evolution des Maladies Infectieuses (GEMI), UMR IRD/CNRS 2724, IRD, BP 64501, 34374 Montpellier Cedex 5, France*^b*Department of Biology, Wesleyan University, Middletown, CT 06459-0170, USA*^c*Molecular Identification Services, NCTC, Health Protection Agency, Specialist and Reference Microbiology Division (SRMD), 61 Colindale Avenue, London NW9 5HT, UK*

Accepted 12 March 2004

First question: There is no unified evolutionary concept of species in bacteria. However, don't you think that the present state of art is OK for routine medical practice? If you disagree, please show why the present state of art could be misleading even for strict medical/veterinary purposes.

Response from Haroun Shah

It has long been recognised that the study of infectious diseases could not proceed without clear circumscription of the microbe. Thus, microbiology has been developed largely using tests that will delineate a disease-associated entity from one that is not, viz. biochemical tests, serology, phage typing, etc. The arrival of genomics (based largely upon 16S rDNA sequence comparisons has meant that no taxon (family, genus, species, etc.) has been left untouched and the new edition of Bergeys Manual (in press) has already adopted the system. Genomics is therefore changing this long familiar system and, some aspects of microbiology which is focused at the clinical edge, is finding these changes difficult to reconcile. For example, the Gram stain (see e.g., Shah et al., 1997) has held a pivotal role in diagnostic microbiology but has little value in a modern phylogenetically based classification. The systematist and the diagnostic microbiologist find themselves at loggerheads with each other in many areas and, for pragmatic reasons, this debate is likely to continue for the foreseeable future. However, these differences are perhaps not as great as they might seem, since it is at the sub-specific levels (subspecies, serovars, biotypes, etc.) that controversies exist at the practical level and while it may take time to resolve, modern genomic approaches are sure to lead to more accurate and

reproducible schemes. For example, the separation of *Neisseria meningitidis* into serotypes has a direct bearing on treatment (vaccine, antibiotics, etc.) but is at variance with groupings based upon MLST types, with many of the latter crossing several sero-type boundaries. However, as more work is done and epidemiology based upon genomic data begin to accumulate, genes that are more discriminative for groups that represent the pathogenic potential of various taxa will become more apparent and lead to a better understanding and application of these methods across all areas of microbiology.

Response from Sylvain Godreuil

Because of their peculiar and variable model of reproduction, bacteria species cannot be defined on the basis of the classical mixiologic concept of species (Mayr, 1940). The definition criteria remain arbitrary and from a long time and until now still appear as a matter of debate. Indeed, we can find in the literature several proposals of species concept in bacteria: as examples, Lan and Reeves (2001) suggest that the species definition in bacterial should be based on analysis of sequence variation in housekeeping genes, other authors propose that bacteria are organised into discrete phenotypic and genetic clusters, which are separated by large phenotypic and genetic gaps, and these clusters are recognised as species (Sneath, 1985). From a medical point of view, the main interest to describe bacteria species is to give the means of identifying the agent responsible for a pathogenic process. In this framework, two fundamental criteria of species definition, as proposed by Goodfellow et al. (Sneath, 1985) appear essential and inseparable: the phenotypic and genotypic characters considered together. Indeed, for a clinical purpose the basis of species classification must allow the identification of

* Corresponding author. Tel.: +33 4 67 41 62 07; fax: +33 4 67 41 62 99.
E-mail address: michel.tibayrenc@mpl.ird.fr (M. Tibayrenc).

pathogens but should make it possible to have a predictive approach. About the last point, the identification of incriminated bacterial species should help, first at the identification of the epidemiological properties in order to take the adapted measures to stop the propagation of the pathogens and second, at the decision for a better treatment for protection and thus, cure. In addition, with antibiotic selection and the increase of immunosuppressed populations, we observe in hospital emergence of new bacterial pathogens which are normally non-pathogenic, or even environmental microorganisms. This shows that, and particularly for the nosocomial problem, it is fundamental to consider concurrently the phenotypic and genotypic characters, the ecological and evolutionary aspects for the species concept of bacteria. This is fundamental in order to understand why these non-pathogenic strains become pathogenic.

Response from Fred Cohan

Like organisms from all walks of life, bacteria fall into clusters of similar organisms on the basis of phenotype, DNA sequence, and ecology. It is the aim of our systematics to discover, describe, and classify these distinct clusters of organisms. From the point of view of public health and medical and veterinary practice, our urgent aim is to understand the full ecological diversity within a taxon of pathogens. I will argue that microbial systematics, and particularly, commonly held concepts of bacterial species, are not up to this task.

Mainstream bacterial systematics has principally viewed bacterial species as clusters of organisms, as defined by phenotypic or molecular criteria (Rosselló-Mora and Amann, 2001; Vandamme et al., 1996), but this is not the only possible approach. Species may alternatively be defined as groups having a set of dynamic properties (Cohan, 2002; de Queiroz, 1998). In this view, a species is a group of organisms whose divergence is constrained by some force of cohesion (e.g., genetic exchange in the case of the highly sexual animals); in contrast, divergence between different species is not constrained, and so species are irreversibly separate in their evolution. Finally, species are deemed to be ecologically distinct groups.

In both microbial and macrobial systematics, species are in practice demarcated simply as clusters. The difference is that while microbial systematists have not embraced a theory-based concept of species, macrobial systematists have attempted to fit their cluster-based demarcations in accordance with a theory of species. In macrobial systematics, cluster- and theory-based approaches have generally yielded the same species demarcations. This is because, having a theory of species, practicing macrobial systematists may continually re-set their visions of how large a cluster should be to fit within a species.

Because mainstream bacterial systematics does not aspire to base its species on theory, there is no opportunity for recalibrating the size of a bacterial species cluster according to

theory. As a result, the group of organisms typically jammed into a named species is extremely diverse in its ecological adaptations and genomic content (Feldgarden et al., 2003). For example, the various sequence-based clusters within the human pathogen *N. meningitidis* differ in their carriage and disease-causing properties (Jolley et al., 2000; Maiden et al., 1998). By pooling a huge assembly of ecological diversity under one species name, we tend to delude ourselves into acting as if everything with the same name has the same properties.

If instead we adopt a theory-based concept of bacterial species, I believe we would be more highly motivated to find the full ecological diversity within existing named species. One such concept is based on the “ecotype”, which is defined as a group of organisms that are ecologically similar to another, such that an adaptive mutant from one ecotype can outcompete to extinction all other individuals from the same ecotype; an adaptive mutant does not, however, outcompete to extinction members of other ecotypes, owing to the ecological differences among ecotypes (Cohan, 1994, 2002). I have shown ecotypes to have the quintessential dynamic properties of species, since the whole-genome diversity within an ecotype is recurrently purged by natural selection favoring each adaptive mutant, yielding cohesion within an ecotype. In contrast, natural selection does not purge the diversity among ecotypes, making inter-ecotype divergence irreversible. Moreover, I have shown that ecotypes so defined can be discovered by universal molecular approaches, whereby each longstanding, ecologically distinct population can be discerned as a sequence cluster (as based on multilocus sequence typing or various PCR- and restriction-based proxies for sequencing) (Cohan, 2002). While there are some (resolvable) issues regarding the size of sequence cluster that corresponds to an ecotype, sequence clustering has successfully identified ecologically distinct groups, in some cases with different pathogenic properties, within named species (Maiden et al., 1998; Ramsing et al., 2000; Rocop et al., 2002; Schloter et al., 2000).

I have previously argued that infraspecific groups that are each ecologically distinct and appear as separate sequence clusters are likely to have the universal, dynamic properties of species, and should be given a trinomial name (e.g., the sequence cluster and putative ecotype that yields meningitis epidemics in subsaharan Africa might be named *N. meningitidis* ecotypus *africana*) (Cohan, 2002; Finlay, 2004). Such naming practices would encourage the routine practice of identifying pathogens to go a step further than the named species; identifying an unknown pathogen to ecotype may yield more specific information about the strain's properties of carriage, transmission, disease, and treatment.

Of course, the present state of bacterial systematics is not blind to infraspecific diversity when variants have unmistakably different impacts on human health. For example, *N. meningitidis* and *N. gonorrhoeae* would have been considered conspecific on the basis of their metabolic and genomic

similarities, but were divided into separate species on the basis of differences in the diseases they cause (Brenner et al., 2001). However, I argue that cases such as this are just the extremes on a continuum: from very closely related ecotypes yielding clearly different clinical features, which we name as separate species, to equally related ecotypes that are more subtly different in their clinical features, which we currently choose not to name. It is important to note that many (perhaps most) named species are diverse in their ecological properties and appear to have multiple ecotypes (Cohan, 2002; Schloter et al., 2000). Perhaps if we were to use a theory-based concept of species to discover all the ecologically distinct and irreversibly separate ecotypes within a named species, the routine practices of public health would be more open for seeing the ecological richness (and subtle pathogenic diversity) among strains given the same Latin binomial.

Second question will be based on two comments about Fred's relevant concept of *ecotype*: (i) in highly recombining species such as *N. gonorrhoeae* or *Helicobacter pylori*, or, in eucaryotic microorganisms, *Plasmodium falciparum*, ecotypes are bound to be very unstable due to abundant horizontal gene transfer, (ii) in the case of the predominantly asexual parasite *Trypanosoma cruzi* (the agent of Chagas disease), clonal genotypes (I mean laboratory-cloned lineages with verification under the microscope) have unstable phenotypic (= ecological) properties: growth in culture, pathogenicity, susceptibility to drugs, etc.: after sets of passages, a given lineage may get quite different phenotypic properties, probably due to gene regulation. Do you think therefore that a complicated Latin nomenclature is adequate for entities that could be made unstable by either genetic exchange or phenotypic flexibility or both?

Response from Fred Cohan

It is true that some bacteria recombine at a much higher rate than others. Nevertheless, even in the most frequently recombining bacteria (e.g., *N. meningitidis* and *H. pylori*), rates of genetic exchange are quite low, within an order of magnitude or two of mutation rates (Feil et al., 1999; McVean et al., 2002). I will argue that such infrequent recombination is unlikely to destabilize bacterial ecotypes, at least in any sense that would prevent us from recognizing ecotypes as taxa.

Let us consider how ecotypes might be destabilized by recombination. Most profoundly, consider first whether the very existence of distinct ecotypes is possible in the face of recurrent, albeit infrequent, recombination. The ecological distinctness of an ecotype is due to its "niche-determining" genes, which are alleles or gene loci that are adaptive in the genetic background of their own ecotype, but are maladaptive in the genetic backgrounds of other ecotypes (Cohan, 1994). I have shown that even in the most frequently recombining of bacteria, natural selection should have no trouble eliminating the occasional recipient of a niche-

determining gene from another ecotype, and thus the existence of bacterial ecotypes is not threatened by recombination (Cohan, 1994).

Next, consider the possibility that recombination might prevent us from detecting ecotypes as sequence clusters based on neutral molecular markers (e.g., the housekeeping genes of MLST, or various random-primer- or restriction-based method) (Maiden et al., 1998; Schloter et al., 2000). A possible disadvantage of neutral molecular markers is that there is (by definition) no fitness penalty for transfer of alleles from one ecotype to another. Thus, given a high enough recombination rate, the niche-neutral variation within and among ecotypes may become totally blended, such that different ecotypes do not ever diverge into separate sequence clusters.

We have previously provided a theoretical rationale for why recombination in bacteria is not frequent enough (with the possible exception of *H. pylori*) to prevent ecotypes from diverging into sequence clusters (Cohan, 1995; Palys et al., 1997). Moreover, the recent MLST literature has demonstrated that while the phylogeny of strains within *N. meningitidis* and some other species may vary from gene to gene, owing to recombination (Feil et al., 2001), the memberships of clonal complexes are stable from multilocus enzyme electrophoresis to multilocus sequence typing, and the clonal complexes in MLST are stable from one set of genes to another (Maiden et al., 1998). In addition, these clonal complexes are known at least in some cases to be ecologically distinct (Maiden et al., 1998). That ecologically distinct groups can be detected as sequence clusters for niche-neutral genes, even in the frequently recombining taxon *N. meningitidis*, suggests that recombination will not prevent us from detecting ecotypes through universal, molecular approaches.

Having said this, I believe that proponents of molecular typing approaches could do much more to investigate the robustness of their methods. In MLST, for example, we would like to know more definitively how many genes are required to identify all of a named species' ecologically distinct groups (at least those of a certain age), and what is the appropriate criterion for including strains within a putative ecotype (e.g., should it be complete sequence identity to a central strain at six of seven loci, as in the Burst algorithm (Feil et al., 1999), or should it be the more inclusive criterion of eBurst (Feil et al., 2004)). Now that we are obtaining full genome sequences of multiple strains within named species (Baba et al., 2002), we can use an approach based on Rokas et al. (2003) to find out how many genes, it takes to obtain the full assemblage of ecologically distinct groups within a named species. Also, our criteria for demarcating putative ecotypes can be adjusted to yield groups whose phylogenies are consistent with predictions for a single ecotype (Cohan, 2002).

Consider next what the existence of two clearly distinct, infraspecific sequence clusters tells us about the stability of these clusters. Our laboratory and others (Dykhuisen, 1998;

Palys et al., 1997) have argued that two long-divergent clusters must represent ecologically distinct populations, each subject to its own private periodic selection events (provided that the clusters have always been sympatric). Otherwise, if the clusters were members of a single ecotype, adaptive mutations throughout the long history of the two clusters would have recurrently purged diversity from both clusters within the ecotype, leaving only a single lineage. The existence of two sequence clusters within a single ecotype (in one place) is extremely unstable and implies that the clusters must represent stably coexisting ecotypes.

Does the persistent coexistence of multiple, closely related ecotypes mean that the ecological niches of the ecotypes have never changed? No, the continuing coexistence means only that in each periodic selection event, the ecotypes have always been ecologically distinct from one another. The ecotypes have almost certainly evolved as their niches became warmer or cooler, or more or less crowded with competitors, potential hosts, and natural and human-made toxins and enemies. And much of this adaptation has no doubt been facilitated by recombination, both homologous (Maynard Smith et al., 1991) and heterologous (Lawrence, 1999).

Recombination appears to destabilize ecotypes only in the sense that it helps an ecotype to gain new adaptations and shed old ones. In spite of recombination, two ecotypes can coexist and maintain the integrity of their separate adaptations, and provided that they are old enough, ecotypes can be identified as sequence clusters.

Regarding the second issue of environmental variation, a strain's ability to take on different phenotypes in different environments certainly makes identification of ecotypes more difficult. Suppose, for example, that we identify multiple sequence clusters based on MLST or a restriction-based method. We would then like to confirm that the sequence clusters correspond, one-to-one, with ecotypes. Environmentally induced variation could yield a one-to-many relationship between sequence clusters and groups with ecologically distinct properties. The investigator would then have to go to extra lengths to determine whether the ecological diversity within clusters is due to environmental induction within a single ecotype, or whether there are many newly arisen ecotypes that are still too young to have diverged into separate sequence clusters for neutral markers.

The question of environmental induction of phenotypes leads us to the issue of reliability of phenotypic versus molecular approaches to bacterial systematics. A taxonomy based only on phenotypes has the potential to be fooled by environmental induction of phenotypic differences; a taxonomy based on DNA sequences will not. Phenotypic approaches are more generally at a disadvantage in the discovery of ecologically distinct groups. It is difficult to discover unknown ecotypes through their phenotypes when we do not know, a priori, the ecotypes' unique ecological

and phenotypic characteristics. In contrast, molecular approaches provide us with working hypotheses about the existence of new ecotypes, even before we are knowledgeable about their unique ecology, etiology, and phenotypes.

Likewise, molecular approaches can be fooled because sequence clusters do not necessarily correspond one-to-one with ecotypes. Newly originated ecotypes are too young to be discerned as separate sequence clusters (Palys et al., 1997), and geographical isolation can cause a single ecotype to diverge into separate clusters ("geotypes") in every geographical region that is beyond the range of frequent migration (Papke et al., 2003).

A full accounting of ecological diversity within a named species will usually require discovery of putative ecotypes through molecular approaches, and then confirmation of ecological distinctness through phenotypic approaches. After an ecotype has been discovered and characterized, phenotypic diagnostics can be developed to help the clinician rapidly identify a strain to its ecotype (Keys et al., 2004). Beyond the diagnostic benefits, phenotypic analysis is also a valuable foray into the natural history and etiology of pathogenic ecotypes, and as Haroun Shah points out, our attempts to find more discriminative diagnostics can help us to understand a taxon's pathogenic potential.

For most bacterial pathogens, I believe that neither recombination nor environmental variation should discourage us from using universal molecular approaches to identify the full diversity of ecotypes within a named species. And we should not be discouraged from naming the diversity that we discover.

Response from Sylvain Godreuil

The definition of an ecotype given in medical textbooks, is a subgroup of a species that has adapted to its local environment and as a result is distinctive from other members of the species, but, in the case of a sexual organism, can still successfully interbreed with other members of the species. From this definition, it appears that the ecotypes of a same species can present different phenotypic or genotypic properties and can exchange genes between them. Thus, in the case of highly recombining species, to determine the status of a new genetic or phenotypic variant as new ecotype or new species can be based on the potentiality to exchange genes with the closer organisms. The difficulty appears within asexual organisms for which the criteria of interbreeding cannot be used. Indeed, Lan and Reeves (2001) claim that the distinction between two bacterial species is often arbitrary based on the phenotypic properties and evidence the need to use phylogenetic analyses to perform species identification. For these authors, the difference between two bacterial species would be often an adaptation of its environment and would only correspond to a new ecotype. For example, the use of mobility, lactose fermentation and lysine decarboxylase allow us to distinguish *Shigella* from *Escherichia coli*.

Meanwhile, a phylogenetic approach of housekeeping genes suggests that they belong to the same species. Actually, *Shigella* is a set of pathogenic clones of *E. coli* that has developed virulence factors and occupied the new ecological niche.

It is my view that an ecotype can be named by a Latin nomenclature when the characters differentiating the ecotypes appear stable in space and time. From a medical point of view, the organisms are generally identified and named according to the disease and phenotypic characters. Then, it is crucial that the Latin nomenclature be used only for stable ecotypes. The *Shigella* ecotype seems very stable, indeed this ecotype has a predictive value for disease symptoms (fever, diarrhoea, etc.) and for physiopathological mechanisms (enteroinvasive) and allows to determine a medical management (antibiotics).

Another case is the acquisition of genes by lateral transfer. This process is a major factor in the adaptation of clones and bacterial evolution. This transfer of genes by plasmids or phages can lead to new properties such as virulence or resistance to antibiotics. This can produce an adaptation to new ecological niches or/and expansion of a new clone. Bacterial adaptation to antibiotics is the most spectacular example to illustrate the evolutive changes and adaptation of bacteria. The evolution of bacteria resistance has been considerably accelerated by the selective pressure exercised by overprescription of drugs in clinical settings and their heavy use as growth promoters for farm animals. This resistance of antibiotics may emerge from the mutation in a chromosomal gene or by acquisition of exogenous genetic material bearing resistance determinants. With a phenotypic approach, we can consider that resistant and sensitive strains are two separate species, in fact, the resistance clone is an adaptive ecotype (antibiotic pressure) of the wild clone. The bacterial resistance group corresponds to the definition of new ecotype but these properties do not appear stable without antibiotic pressure. This suggests that the Latin nomenclature is not appropriate with the resistant phenotype of new ecotypes. Then, for the medical purpose, the solution is to keep the Latin nomenclature of species and to specify the resistance type. For example, the methicillin-resistant *Staphylococcus aureus* (SARM) is very significant for medical personnel: (i) the SARM is synonymous of the serious infection, (ii) the SARM has got an important epidemic power (the patient must be isolated in order to prevent an epidemic), (iii) by definition, beta-lactams and other antibiotic families are ineffective with SARM organisms.

To my mind, the Latin nomenclature must be reserved to the ecotypes with phenotypic and genotypic properties showing a stability in space and time. To identify the phenotypic and/or genotypic variants of new emergent unstable ecotypes, the Latin nomenclature of species would be used associated with a code indicating the phenotypic properties of resistance (as SARM) or virulence (as *E. coli* enteroinvasive EIEC).

Response from Haroun Shah

Designating a general Latin trinomial to reflect the diverse range of microbes that colonise various habitats is challenging but is perhaps not feasible with the technology and resources currently available. Full genomes of multiple strains of some species (e.g., seven strains of *S. aureus*) have now been completed and as predicted variable and hypervariable regions of the DNA exist that may be exploited to provide markers of diversity. Genetic variation is the basis of dynamic adaptation in bacterial populations and is the ultimate evolutionary basis of most species and community level diversity. However, this is too incoherent across the microbial kingdom to index these in a consistent manner. Thus, the genomes of *Mycobacterium tuberculosis*, irrespective of their isolation site are so conserved that IS elements or variable number tandem repeat sequences are employed to discern their diversity while others such as *N. meningitidis* possess immensely fluid genomes.

Throughout the history of microbiology, various systems have been introduced largely for epidemiological purposes and tracing the movement of strains in the environment. Notable examples include phage typing, serology and later techniques such as multilocus enzyme electrophoresis (MLEE). Early examples using MLEE for tracing contaminated grain for agricultural products and transmission of parasitic agents demonstrated the importance of sub-typing methods. Currently these have been superseded by DNA-based methods such as multilocus sequencing typing (MLST), amplified fragment length polymorphism, variable number tandem repeat sequences, etc. in which web-based databases can be interrogated electronically. A nucleotide base is the smallest and most accurate unit in biology and because the genetic code is universal, these methods are likely to expand and used widely in foreseeable future. They are accurate, additive, portable, have inbuilt quality controlled parameters, can be curated, and are accessible to microbiologists throughout the world. Standard protocols are available and data may be submitted from any laboratory and appropriate nomenclature adopted for each group of microorganisms.

A system based on ecotype would be cumbersome and difficult to standardise. Previous work in the author's laboratory have shown that strains isolated from clinical samples had very restricted amino acid profiles for *Fusobacterium nucleatum*, a non-fermentative species that relies on nitrogenous substrates as sources of energy. A few amino acids such as glutamate or aspartate or dibasic acids such lysine or arginine were readily taken up and metabolised. However, as these strains were subcultured over several months, the amino acid up-take profiles dramatically broadened to include a larger number of other amino acids (Gharbia and Shah, 1989). Such experiments serve to highlight the disparity between in vivo and in vitro culture and emphasise the difficulties of retaining the physiological status of a strain isolated under different environmental conditions.

The major international culture collections, such as, NCTC, ATCC, DSMZ, etc., that act as repositories of microbial species are presently struggling financially to maintain a limited number of type strains. Very few collections have attained recognised accredited status for their strains (NCTC accredited for ca. 6000 strains) and, as such, are not in a position to act as a reliable source of designated strains. Strains isolated from specific studies are usually the property of academic or research establishments that have, more often than not, little experience in maintaining a culture collection. Therefore, establishment and maintenance of a microbial collection of strains from designated sources would be difficult to envisage due to the large number of resources required, including technical expertise, quality control, and the massive financial support that these criteria demand.

Third question: Do you think that powerful new biotechnologies such as microarrays will bring significant progress in our view of bacteria classification?

Response from Fred Cohan

Sequence-based approaches allow systematists to identify ecologically distinct populations of bacteria, even before they know what is ecologically distinct about them (Cohan, 2004). This is because sequence clusters are predicted, under certain circumstances, to correspond to ecotypes. A major challenge is to confirm that putative ecotypes, originally discovered as sequence clusters, do indeed make their livings in different ways (e.g., by infecting different hosts or tissues).

Finding the physiological basis for ecological distinctness is especially difficult when we do not know what specific adaptations we are looking for, as will often be the case when putative ecotypes are known only as distinct sequence clusters. However, recent advances in genomic technology allow us to focus on possible adaptive differences (Feldgarden et al., 2003). First, genomic approaches can help us identify genes that are not shared across populations. Given that many bacterial adaptations have been acquired through horizontal transfer (Gogarten et al., 2002), genes that are unshared among close relatives can point us to the genetic basis of ecological distinctness. For example, Zhang et al. (2003) found consistent differences in genome content among sequence clusters within *Listeria*, and these differences were found to affect regulation of virulence genes.

Microarray assays can help us identify differences among putative ecotypes in their genome-wide patterns of gene expression. For example, Townsend et al. (2003) used this approach to find differences in gene expression among yeast isolates from nature, but to my knowledge this approach has not yet been applied to characterizing ecological divergence in bacteria.

Differences among sequence clusters in their gene content or gene expression will not by themselves prove

that the clusters represent distinct ecotypes. However, such results can suggest experiments that would more definitively demonstrate ecological divergence.

Sequence cluster analysis and subsequent genomic analysis of gene content and gene expression together provide a universal molecular approach to bacterial systematics. We can imagine a time when systematic analysis can be fully automated, from identification of putative ecotypes as sequence clusters to characterizing the differences that allow them to coexist as ecologically distinct players in the biological community.

Response from Sylvain Godreuil

Classification refers to how we group and categorize extinct and living species of organisms. Initially, it was based on shared physical characteristics. At this time, species classification called also molecular systematics are more and more based on genomic DNA analysis and especially for microorganisms such as bacteria, which are often morphologically indistinguishable from each other. Thanks to the continuous technical progress in molecular biology, bacterial classification is constantly evolving.

Knowledge has been strongly detailed with the beginning of DNA studies. The first step of bacterial classification based on genome study was the determination of DNA base ratio (percentage G–C base pairs) and DNA–DNA hybridization. These methodologies were useful in clarifying relationships at the species level, and were considered at the beginning as standards for the classification of a lot of Bacteria (Schleifer and Stackebrandt, 1983) (Stackebrandt and Goebel, 1994) For example, the G + C content allowed to divide the *Campylobacter* genus in different species: *Campylobacter fetus* had a mean G + C content of 35.7% whereas the thermophilic species *C. jejuni*, *C. coli* and the NARTC group (*C. laridis*) had a G + C content between 31.5 and 32.6%. Furthermore, DNA–DNA hybridizations showed also clear differences between most *Campylobacter* taxa (Owen, 1983).

This last technique was used as an indirect parameter of the sequence similarity between two entire genomes. Two bacteria were considered as belonging to the same species if the value of DNA–DNA hybridization between the two genomes was higher than 70% (Wayne et al., 1987). This method allowed a better understanding of taxonomy of some bacterial genera such as *Xanthomonas* or *Campylobacter*.

DNA sequencing, first designed in 1975, has been an extraordinary progress to study distantly related organisms as well as within-species comparison of multiple strains. This allowed considerable progress in knowledge of bacterial classification. At present, the whole sequencing of more than 200 bacterial genomes has been completed. The whole sequencing would be the most informative technique to improve Bacteria classification, but it is hardly feasible, because too heavy, time-consuming and too

expensive. It is the reason for which the challenge, today, is to discover a technique (low cost, fast) giving the best picture of the whole bacterial genome and allowing to make a classification as finest as possible.

rRNA sequence analysis has been accepted as the best target for studying phylogenetic relationships and bacterial classification. This method has a lot of advantages: (i) an important number of rDNA sequences are accessible via international databases, (ii) the presence of universal bacteria sequences in the rRNA molecule allows to perform a global classification of all bacteria.

Another molecular tool appeared in the late 1990s, multilocus sequence typing (MLST). It has been proposed as a nucleotide sequence-based approach that could be applied to many bacterial species (Maiden et al., 1998). This technique presents various advantages: microorganism culturing is spared, sequencing data are unambiguous, the technique is easy to standardize, and electronically portable. Nevertheless, the limit for bacterial classification is the specificity of this technique. However, it appears as a powerful new DNA-typing tool for the evaluation of intraspecies genetic diversity (Robles et al., 2004).

After the sequencing of many bacteria genome, the development of DNA microarray technology has proved to be a valuable tool for genomic screening and can be efficiently used for bacterial classification. Indeed, this technique offers the possibility of examining large amounts of sequence with a single hybridization step and thus can give a good view of the whole bacterial genome. Edwards-Ingram et al. (2004) demonstrated that this technology provided new insights into the molecular taxonomy of the *Saccharomyces sensu stricto* complex. This technique remains today expensive, but in the future the array technology could be routinely used for a large range of applications (molecular epidemiology, diagnosis, etc) and especially for a powerful and accurate bacterial classification.

All the examples detailed above show that each time new biotechnologies brought or will bring significant progress in our view of bacterial classification. But one has to keep in mind that each of them presents limits of application. Thus, it is indispensable to select the more efficient tool according to the organism under study and to the scientific question under study.

Response from Haroun Shah

Biological sciences are witnessing a staggering array of powerful new technologies for genomics and proteomics. The ability to undertake high throughput analyses at the nanoscale level is the hallmark of these new technologies. For biosciences such as microbiology and, in particular areas such as microbial systematics and evolution, the limitations of earlier methods have been in general a failure to include a sufficiently broad spectrum of strains in most experiments. Because of its inherent diversity, most taxonomists and judicial systematic committees would prefer that studies

aimed at defining new species include the type, a large number of reference and wild-type strains but the cumbersome nature of previous techniques did not facilitate this. For example, it is generally accepted that DNA–DNA reassociation is the gold standard for defining a species and while this method is robust, scientifically sound and yields unequivocal data, because the technique is so tedious it is now being phased out. However, the limitation of this method has been its reliance on the pre-selection of a few reference ‘target’ strains to which other strains are compared. The resulting homology groups may therefore contain experimental bias but more importantly it gives no indication of the functional relatedness of the strains being compared. Traditional microbiological tests, which rely on the ability of the test organism to ferment a particular carbohydrate are today widely regarded as outmoded. However, these tests provide a measure of functional relatedness of a taxon in that ‘glucose positive’ for example implies the presence of several enzymes related to the metabolism of this substrate. Conversely, a negative result may indicate the absence of an entire metabolic pathway. These tests have provided such an excellent framework for microbial classification that, up to the present time, it is still considered necessary to include relevant tests in the description of taxonomic units.

Modern microbial systematics should strive to encompass both comparative DNA sequence relatedness and functional analysis. Consequently, microarrays, as do many modern technologies provide an ideal platform for measuring both these parameters. DNA arrays, at the current time, are limited by the information provided through full genome analysis but it nevertheless provides an objective means of comparing a large number of test strains. DNA microarrays have already had a considerable impact on disease diagnosis and expression profiling. The latter via transcriptomics is expanding rapidly, however, protein microarrays, which potentially could provide some of the most interesting data, particularly in relation to microbial classification and evolution, presents a significant technical challenge due largely to the complexity of the molecules being analysed. Problems that may arise due to post-translational modification have been largely overcome, however, devising a platform that can analyse the vast range of analyte concentrations present in the cell is a major task in making these systems as accessible as DNA microarrays. For example, the detection of low abundance proteins in the complex biological milieu of the cell, which also contains high abundance proteins, is currently very tedious and may be affected by levels of signal to background ratios.

Apart from DNA binding proteins, current methods generally rely on some form of serological detection and therefore require the availability of high affinity antibodies. However, rapid progress is being made and it is evident that proteomics will herald a new era of microbial systematics and evolution due mainly to the progress being made in protein sequence identification via mass spectrometry. A

simple preamble to such analysis could be via selective protein capture from a cell lysate using protein affinity chips. For example, ProteinChip arrays using the Ciphergen's SELDI-TOF mass spectrometer enables a search for taxon-specific biomarkers to be readily identified. Initial identification together with MS/MS or LC/MS/MS directly provides a novel method of obtaining sequence information on particular biomarkers.

Microbial systematics has had a long period (ca. 100 years) of using morphological and physiological criteria to describe taxonomic units and is still used widely today. This was followed by a period of chemical analyses (e.g., peptidoglycan chemotypes, fatty acids, polar lipids, respiratory quinines, etc.) and limited applications of peptide analysis (e.g., SDS-PAGE). This was superseded by nucleic acid-based methods with the greatest impact today being the wide acceptance of 16S rDNA sequence analysis.

The latter provided a molecular chronometer for microbiological studies and for the first time placed phylogeny on a firm footing. This approach has gained such widespread acceptance that 16S rDNA sequence analysis is often being used as the sole criterion for defining a species.

The presence of very novel powerful technologies such as DNA microarrays, pyrosequencing, denaturing high-pressure liquid chromatography, nanoarrays, mass arrays SNP discovery chips, electrical probe addressing systems, multidimensional LC–MS–MS, etc. together with advances being made in microfluidics and robotic liquid handling systems, opens up great opportunities to undertake large scale microbial studies. Microbial systematics and evolution should not be based solely on genomics and particularly not restricted to a single gene (e.g., 16S rDNA) to represent the overall profile of a taxonomic unit. It is this author's view, it should also encompass several genes (far more than MLST) and also include the functional aspects of the cell. With the advances being made in proteomics and transcriptomics, the time is now near to including these data into a new era of 'proteosystematics' and introduce the functional aspects of microbial systematics and evolution that has been long overdue.

Fourth question: If you were the man who has the power to decide right now: what would you propose as bacterial species concept that would be equally satisfactory for theoreticians/evolutionists and medical doctors and could be printed in textbooks both for scientific and medical students? You are free to develop your thought, however try and summarize your bacterial species concept in a few sentences.

Response from Haroun Shah.

Systematic bacteriology has now been taking place over a century and throughout this period, a pivotal role has been accredited to various editions of Bergey's manual. Even up to the present time, every student or experienced micro-

biologist still utilises these volumes as a reference source. It continues to be a pillar of this science taking into consideration the continual changes in microbiology as new technologies and analytical tools enable deeper insights into the complexity of the bacterial cell. Consequently, the concept of a bacterial species should encompass recent information gained from cellular structure and genomics but should not exclude those that have stood the test of time such as the gram stain, morphological criteria, e.g., the cell shape or spore-formation, growth characteristics, e.g., aerobic versus anaerobic and especially the functional nature of the cell, such as its capacity to metabolise carbohydrates or its non-fermentative physiology. Similarly considerable work has been undertaken on the chemical structure of the cell, e.g., the nature of the cross-linkage of the peptidoglycan, respiratory quinones, long-chain cellular fatty acid composition and the structure of complex macromolecules such as lipopolysaccharide or cell surface polymers.

However, it is the genome and the proteome that modern concepts should embrace. It is this author's view that the current type strain for a species should give way to a strain that has been used to obtain the full sequence of the genome. In cases such as *S. aureus*, where multiple strains have been sequenced, the type strain or the one that most closely resembles the type strain should be selected. The reliance placed on a single gene such as 16S rRNA should be reduced and substituted by multiple gene analysis, e.g., *gyrB* and *rpoA/B*, *recA*, *fusA*, etc.

New advanced technologies particularly in mass spectrometry (MALDI-TOF-MS, SELDI-TOF-MS, etc.) are making high throughput analysis of low abundance molecules or unique biomarkers available. Therefore, a species should not be proposed or reclassified until a large number of strains have been studied except in rare instances where there are no additional isolates. Key proteins and biomarkers that are now a hallmark of the functional nature of the cell should be included.

Finally, attempts should be made to provide a measure of the diversity (diversity index, e.g., use of MLST, VNTRs, etc.) of a species particularly for pathogens and emerging pathogens as they leave the environment and enter the mammalian ecosystem. The capacity of each for horizontal gene transfer and recombination characteristics would provide an index of the stability of the genome. For pathogens, a measure of its pathogenic potential from available genes (e.g., pathogenicity island characteristics), is essential. Therefore, in summary, a holistic approach should be taken to describe a species and not as commonly seen, the use of a single criterion such as 16S rRNA. The description should include as many characters as possible based upon selected historical criteria but in line with modern concepts include key features of the genome and proteome.

Response from Sylvain Godreuil

During the last years, the development of new technologies in particularly the complete genome sequence

of bacteria allowed to revisit the bacterial species concept. Thanks to gene sequencing and the collected informations a bacterial species concept could be elaborated: “the species genome concept” (Lan and Reeves, 2001). In this concept, the genome of bacterial species has been divided into two components: the core set of genes found in the majority of the members of a species and auxiliary genes found in some but not in all the members of a species. The core genes determine characteristic properties of all members of a species. They include principally the housekeeping genes which allow, for a great part of them, species classification. The maintenance of species specificity of core genes can be explained by the existence of recombination barrier between species. The auxiliary genes correspond for example to acquisition of genes by horizontal transfers. These genetic elements can allow bacterial species to have adapted responses against a selection pressure (e.g., antibiotic pressure) or to invade a new ecological niche. These genes bring new phenotypic properties within a bacterial species such as antibiotic resistance, virulence properties or new immunological characteristics.

From these bases “the species genome concept” allows to make the difference between an adaptive clone with new phenotypic properties corresponding only to an intraspecies variation and not to the emergence of a new species.

To my opinion, “the species genome concept” could be a relevant definition of bacterial concept that can satisfy the theoreticians/evolutionists and medical doctors. The fact that the bacterial species can be defined on the basis of multilocus sequence typing (MLST) of the housekeeping genes is an attractive argument for medical practice for two reasons: (i) a large number of medical microbiology laboratories owns the tools to realise the MLST technique, (ii) an important number of bacterial housekeeping gene sequences is available on electronic network banks. However, to my mind, the notion of adaptive clone of bacterial species should be cautiously considered. As I have already described above, a phylogenetic approach of housekeeping genes has demonstrated that *E. coli* and *Shigella* belong to the same species, and *Shigella* would be an adaptive virulent clone of *E. coli*. In this case, the auxiliary genes coding for the virulent properties are stable in space and time and determine a real clinical and biological entity. Thus, in some cases, auxiliary genes can be used as markers for species definition, but only if they present a genetic stability and a medical relevance. This demonstrates that the “species concept” whatever it be cannot be a fixed concept, it must be able to adapt in function of the progress of knowledge about the microorganisms under study.

Response from Fred Cohan

Our concepts of bacterial species and smaller taxa should help us discover and demarcate the full diversity of ecologically distinct populations within any group of bacteria. I have argued that this goal could be reached by giving taxonomic status to ecotypes (defined as populations

that are ecologically distinct, cohesive, and irreversibly separate from other ecotypes). While it is difficult to demarcate taxa by these properties directly, I will show that a combination of ecological and sequence data should allow us to infer demarcations among ecotypes, under a variety of models of bacterial evolution.

Consider first what we might call the Stable Ecotype Model, in which ecotypes are created and extinguished at a very low rate, and during its long lifetime any given ecotype is recurrently purged of its diversity by periodic selection. Assuming also a high rate of migration across the geographic range of an ecotype, most ecotypes (of sufficient age) should be easily distinguishable from other ecotypes as a sequence cluster. This is because the sequence diversity within each ecotype is limited by periodic selection, and there is ample time for neutral sequence divergence to accumulate among ecotypes (Palys et al., 1997).

Next consider the Geotype Model (Papke et al., 2003), in which ecotypes are long-lived as above, but there is only rare migration among the geographic regions of the ecotype. In this model, ecologically identical populations in different regions can diverge into different sequence clusters. When we observe that each geographic region has its own endemic sequence cluster, it is possible that the various endemic clusters represent ecologically distinct ecotypes, but alternatively they could represent *geotypes* (i.e., ecologically identical but geographically isolated populations of the same ecotype) (Papke et al., 2003). As in the systematics of any organism, endemic sequence clusters are often difficult to interpret.

One could argue that geographic isolation should not prevent us from using sequence clustering to discover ecotypes that are found *in one region* (Palys et al., 1997). However, in a model I call the Geotype-Plus-Boeing Model, geographically isolated populations of the same ecotype diverge 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. In this transitional era when air travel (and to some extent even sea travel) is still new, we may see multiple sequence clusters within one ecotype at one place. Therefore, it would be foolish to conclude from sequence data alone that two sympatric sequence clusters are separate ecotypes. We would also need evidence of ecological distinctness, particularly for bacteria that are no strangers to jet travel.

Consider finally the Species-Less Model (Cohan, 2004; Lawrence, 2002), in which there is frequent invention of ecologically distinct populations, owing to horizontal genetic transfer, and there is little, if any, periodic selection. If we add to Lawrence (2002) model the assumption that populations are extinguished at about the same high rate that they are created, this gives us an alternative to periodic selection for explaining the modest sequence diversity typically found within bacterial populations. That is, each new ecotype is founded by a single mutant or recombinant clone, and there is only limited time for accumulation of

sequence diversity within the population before it goes extinct. Under this model, a single sequence cluster might contain multiple, ecologically distinct (and very young) populations.

In summary, sequence clusters can be guides to discovering ecologically distinct populations, especially under the Stable Ecotype Model. However, the Geotype-Plus-Boeing model yields a one-to-many relationship between ecotypes and sequence clusters, and the Species-Less Model yields a many-to-one relationship.

Considering the vagaries of the rates of niche-invasion, periodic selection, and migration, how can we readily demarcate groups with the dynamics of ecotypes? I offer an operational method that I believe is practical, robust with respect to variations in the rates of these dynamic processes, and consistent with the aim of discovering ecotypes. It focuses on ecotypes having both a history of coexistence as separate lineages, and a prognosis for further coexistence—in other words, the stability that Sylvain Godreuil has demanded in a taxon. Our prognosis for further coexistence can be based on an assessment that populations are ecologically distinct in the resources they utilize in nature; a history of coexistence can be inferred from the populations being in long-divergent sequence clusters. Thus, I propose the following operational criteria for demarcating ecotypes: an ecotype may be demarcated as the smallest group that (1) is ecologically distinct in nature from other such groups, and (2) forms a separate sequence cluster from other such groups (but we reserve the right to waive the latter criterion; see below).

Why is ecological distinctness alone not sufficient to demarcate ecotypes? First, given the potential for horizontal genetic transfer, any two closely related isolates or populations are likely to differ somewhat in their ecology, at least in the context of some laboratory test of metabolic capacity (Joyce et al., 2002; Lawrence and Hendrickson, 2003). In addition, Haroun Shah has made the point that ecological capacities, as measured in the laboratory, can change spontaneously with mutation and natural selection. Clearly, what we want to know goes beyond assessment of physiological differences that have no bearing on ecological niche in nature. Rather, we need to ascertain that populations are ecologically distinct in a way that allows them to partition resources *in nature*, and to thereby coexist. Sequence data provides a means for inferring that ecological differences observed in the laboratory are important in nature. When two ecologically distinct populations fall into distinct sequence clusters, we may infer that the populations are longstanding in their coexistence, possibly owing to their ecological differences (alternatively to previous geographic separation).

There is a second problem in identifying populations as different ecotypes when they are not yet separate sequence clusters. The populations might not be irreversibly separate, a case most likely when the populations owe their ecological distinctness entirely to the gain or loss of a plasmid. In a

process, I call “ecological conversion”, members of a plasmid-free population (e.g., a plasmid-free *Rhizobium* population adapted to soil) can acquire a plasmid and thereby become converted to another population (e.g., a symbiosis-plasmid-bearing *Rhizobium* population adapted to a mutualism with its legume host), and the reverse conversion can occur with the loss of the plasmid. If these reciprocal ecological conversions recur repeatedly, then the populations are not irreversibly separate lineages. However, if we demarcate ecotypes only when they form separate sequence clusters, we can be assured that the populations are not recurrently undergoing ecological conversion, but instead have a history of divergence as separate lineages. I believe that the synergy of demanding both ecology- and sequence-based distinctness should obviate my co-authors’ concerns that the ecotypes we name must be stable.

Why should we not demarcate ecotypes solely by sequence clustering? To the extent that the Stable Ecotype Model is correct, different sequence clusters are indeed likely to represent different ecotypes. However, to the extent that the Geotype-Plus-Boeing Model applies (especially likely with human, veterinary, and agricultural pathogens and commensals), different clusters could represent formerly geographically isolated populations of the same ecotype that have recently been flown or shipped to the same locations. Therefore, sequence clusters must be verified to be ecologically distinct before they can be declared ecotypes.

Finally, we must take into account the Species-Less Model, in which ecologically distinct populations are frequently too new to be distinguishable as sequence clusters. I believe that we will not normally want to grant ecotype status to a new population that has not yet demonstrated its ability to coexist with others (by forming a separate sequence cluster), but there are clearly some cases where we would want to waive the sequence-cluster requirement. Some newly arisen pathogens, for example, are difficult to distinguish from closely related populations by sequences of protein-coding genes (e.g., *Bacillus anthracis* versus *B. cereus*) (Keim and Smith, 2002), but the ecological distinctness we observe (regarding virulence) is clearly relevant to the ways that the bacteria make a living in nature, and the ecological distinctness of these groups is not readily reversible (e.g., with the gain or loss of a plasmid) (Welkos et al., 1993). It is therefore reasonable to give a prognosis for the continued coexistence of such populations as separate lineages, and they could be put into separate taxa on the basis of their ecological distinctness alone. An alternative approach is to further investigate DNA sequence clustering of very closely related populations with a more rapidly evolving sequence, such as IS or VNTR sequences (Henderson et al., 1995; Keim et al., 2004).

Note that we do not need to make a special exception for pathogens by invoking the public health danger of

misdiagnosing them (the rationale for *nomen periculosum*) (Stackebrandt et al., 2002). Generally, whenever we are extremely confident that two populations are ecologically distinct *in nature*, and that their ecological distinctness is not readily reversible, we may regard them as separate ecotypes, regardless of their danger or the similarity of their DNA sequences.

Some important, pragmatic issues remain. First, how do we determine that putative ecotypes are indeed ecologically distinct? As I have mentioned, critical evidence will surely come from post-genomic approaches, such as comparing genome content, or comparing genome-wide gene expression, either at the transcription level or the “proteosystematic” level envisioned and developed by Haroun Shah. But much of the evidence would draw on the existing skills of polyphasic taxonomists, who are trained in testing the capabilities of growth of organisms with different resources and under different physical conditions. I would therefore recommend a new charge for polyphasic taxonomy, to move from finding diagnostic phenotypic characters (Vandamme et al., 1996) to using the broadest diversity of techniques to assess ecological differences.

I recognize that taxonomists are not currently trained to study ecology. This may change, but for the moment I recommend that taxonomists provide highly detailed information about the microhabitat and etiology of isolates used to describe possible ecotypes (e.g., including not just the state or province, but also details of the soil type and its chemical content; including not just that a pathogen causes an upper respiratory infection, but also the cell types infected and transmissibility, severity, and duration of infection). Then, in a second stage of analysis, interested ecologists could attempt to find an inherited basis for microhabitat or etiological differences among putative ecotypes.

Another issue regards the criterion for being in different sequence clusters. Sequence-based phylogenies often contain a hierarchy of clusters, subclusters, sub-subclusters, and so on, but which level of cluster corresponds to the ecotype? This issue is not yet fully resolved, but some clonal complexes yielded by multilocus sequence typing appear to correspond to ecotypes (Cohan, 2002; Feil et al., 2000, 2004). Also, Jason Libsch and I are developing a method for objectively demarcating sequence clusters based on the phylogenetic complexity (i.e., number of significant nodes in the phylogeny) expected within one ecotype (Cohan, 2002), and Dan Krizanc and I are developing a method for estimating the number of ecotypes in a sample of environmental DNA, without any assumptions about the sequence divergence within or between ecotypes.

Finally, while I have advocated taxonomic status for ecotypes, I do not favor overhauling the current Linnaean binomial system for bacteria. It is true that bacterial systematists have lumped ecologically and genomically distinct groups within a given named species (Feldgarden et al., 2003; Gordon and Cowling, in press; Schlöter et al.,

2000; Staley, 2003), but they have been very careful to ascertain that each species is monophyletic, and therefore worth maintaining as a taxon (Stackebrandt et al., 2002). A typical named species appears to contain multiple ecotypes, and so the ecotypes we demarcate could be given a trinomial name with a special epithet such as “ecovar” (or possibly “ecotypus”) following the genus–species binomial (Cohan, 2002). By giving a name to each ecotype, microbiologists will be encouraged to describe the full ecological diversity within the bacterial world. This will be a profound challenge to the infrastructure of systematics, as noted by Haroun Shah. However, this fuller understanding of bacterial diversity that I am promoting will allow public health microbiologists to more precisely determine the ecological properties of new bacterial isolates.

MT I thank you very much for participating in this e-debate.

Acknowledgements

Frederick Cohan would like to thank Jean Swings and Fabiano Thompson for organizing the recent conference, “The prokaryotic species: genome plasticity, microevolution, and taxonomy”, as well as the conference participants for formative discussions on bacterial species. He also thanks his son Willie for shooting his portrait. This work was supported by grants from the National Science Foundation and Wesleyan University.

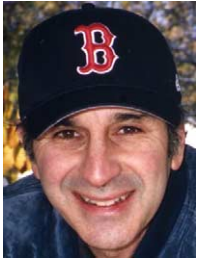
References

- Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K., Hiramatsu, K., 2002. Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* 359, 1819–1827.
- Brenner, D., Staley, J.T., Krieg, N.R., 2001. Classification of prokaryotic organisms and the concept of bacterial speciation. In: Boone, D.R., Castenholz, R.W. (Eds.), *Bergey’s Manual of Systematic Bacteriology*. Springer-Verlag, New York, pp. 27–31.
- Cohan, F.M., 1994. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am. Nat.* 143, 965–986.
- Cohan, F.M., 1995. Does recombination constrain neutral divergence among bacterial taxa? *Evolution* 49, 164–175.
- Cohan, F.M., 2002. What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487.
- Cohan, F.M., 2004. Concepts of bacterial biodiversity for the age of genomics. In: Fraser, C.M., Read, T.D., Nelson, K.E. (Eds.), *Microbial Genomes*. Humana, Totowa, NJ.
- de Queiroz, K., 1998. The general lineage concept of species, species criteria, and the process of speciation. In: Howard, D.J., Berlocher, S.H. (Eds.), *Endless Forms: Species and Speciation*. Oxford University Press, Oxford.
- Dykhuizen, D.E., 1998. Santa Rosalia revisited: why are there so many species of bacteria? *Anton. Leeuw.* 73, 25–33.
- Edwards-Ingram, L.C., Gent, M.E., Hoyle, D.C., Hayes, A., Stateva, L.I., Oliver, S.G., 2004. Comparative genomic hybridization provides new

- insights into the molecular taxonomy of the *Saccharomyces sensu stricto* complex. *Genome Res.* 146, 1043–1051.
- Feil, E.J., Holmes, E.C., Bessen, D.E., Chan, M.S., Day, N.P., Enright, M.C., Goldstein, R., Hood, D.W., Kalia, A., Moore, C.E., Zhou, J., Spratt, B.G., 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. U.S.A.* 98, 182–187.
- Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., Spratt, B.G., 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* 186, 1518–1530.
- Feil, E.J., Maiden, M.C., Achtman, M., Spratt, B.G., 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol. Biol. Evol.* 16, 1496–1502.
- Feil, E.J., Maynard Smith, J., Enright, M.C., Spratt, B.G., 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 154, 1439–1450.
- Feldgarden, M., Byrd, N., Cohan, F.M., 2003. Gradual evolution in bacteria: evidence from *Bacillus* systematics. *Microbiology* 149, 3565–3573.
- Finlay, B.J., 2004. Protist taxonomy: an ecological perspective. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.*
- Gharbia, S.E., Shah, H.N., 1989. The uptake of amino acids from a chemically defined medium by *Fusobacterium* species. *Curr. Microbiol.* 18, 189–193.
- Gogarten, J.P., Doolittle, W.F., Lawrence, J.G., 2002. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238.
- Gordon, D.M., Cowling, A., in press. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology*.
- Henderson, I., Yu, D., Turnbull, P.C., 1995. Differentiation of *Bacillus anthracis* and other 'Bacillus cereus group' bacteria using IS231-derived sequences. *FEMS Microbiol. Lett.* 128, 113–118.
- Jolley, K.A., Kalmusova, J., Feil, E.J., Gupta, S., Musilek, M., Kriz, P., Maiden, M.C., 2000. Carried meningococci in the Czech Republic: a diverse recombining population. *J. Clin. Microbiol.* 38, 4492–4498.
- Joyce, E.A., Chan, K., Salama, N.R., Falkow, S., 2002. Redefining bacterial populations: a post-genomic reformation. *Nat. Rev. Genet.* 3, 462–473.
- Keim, P., Smith, K.L., 2002. *Bacillus anthracis* evolution and epidemiology. *Curr. Top. Microbiol. Immunol.* 271, 21–32.
- Keim, P., Van Ert, M.N., Pearson, T., Vogler, A.J., Huynh, L.Y., Wagner, D.M., 2004. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infect. Genet. Evol.* 4, 205–213.
- Keys, C.J., Dare, D.J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M., Shah, H.N., 2004. Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infect. Genet. Evol.* 4, 221–242.
- Lan, R., Reeves, P.R., 2001. When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends Microbiol.* 9, 419–424.
- Lawrence, J.G., 1999. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microbiol.* 2, 519–523.
- Lawrence, J.G., 2002. Gene transfer in bacteria: speciation without species? *Theor. Popul. Biol.* 61, 449–460.
- Lawrence, J.G., Hendrickson, H., 2003. Lateral gene transfer: when will adolescence end? *Mol. Microbiol.* 50, 739–749.
- Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145.
- Maynard Smith, J.M., Dowson, C.G., Spratt, B.G., 1991. Localized sex in bacteria. *Nature* 349, 29–31.
- Mayr, E., 1940. Speciation phenomena in bird. *Am. Nat.* 74, 249–278.
- McVean, G., Awadalla, P., Fearnhead, P., 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160, 1231–1241.
- Owen, R.J., 1983. Nucleic acids in the classification of *Campylobacter*. *Eur. J. Clin. Microbiol.* 24, 367–377.
- Palys, T., Nakamura, L.K., Cohan, F.M., 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int. J. Syst. Bacteriol.* 47, 1145–1156.
- Papke, R.T., Ramsing, N.B., Bateson, M.M., Ward, D.M., 2003. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* 5, 650–659.
- Ramsing, N.B., Ferris, M.J., Ward, D.M., 2000. Highly ordered vertical structure of *Synechococcus* populations within the 1 mm-thick photic zone of a hot spring cyanobacterial mat. *Appl. Environ. Microbiol.* 66, 1038–1049.
- Robles, J.C., Koreen, L., Park, S., Perlin, D.S., 2004. Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of *Candida albicans*. *J. Clin. Microbiol.* 426, 2480–2488.
- Rocap, G., Distel, D.L., Waterbury, J.B., Chisholm, S.W., 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* 68, 1180–1191.
- Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.
- Rosselló-Mora, R., Amann, R., 2001. The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67.
- Schleifer, K.H., Stackebrandt, E., 1983. Molecular systematics of prokaryotes. *Annu. Rev. Microbiol.* 37, 1433–1487.
- Schlöter, M., Leubhn, M., Heulin, T., Hartmann, A., 2000. Ecology and evolution of bacterial microdiversity. *FEMS Microbiol. Rev.* 24, 647–660.
- Shah, H.N., Gharbia, S.E., Collins, M.D., 1997. The Gram-stain; a declining synapomorphy in an emerging evolutionary tree. *Rev. Med. Microbiol.* 8, 1–8.
- Sneath, P.H.A., 1985. Future numerical taxonomy. In: Goodfellow, M., Jones, D., Priest, F.G. (Eds.), *Computer-Assisted Bacterial Systematics*. Academic Press, Orlando, pp. 415–431.
- Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A., Kampfer, P., Maiden, M.C., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B., 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047.
- Stackebrandt, E., Goebel, B.M., 1994. Taxonomic note: a place for DNA–DNA reassociation and 16s rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44, 846–849.
- Staley, J.T., 2003. Speciation and bacterial phylospecies. In: Bull, A.T. (Ed.), *Microbial Diversity and Bioprospecting*. American Society for Microbiology Press, Washington, DC, pp. 40–48.
- Townsend, J.P., Cavalieri, D., Hartl, D.L., 2003. Population genetic variation in genome-wide gene expression. *Mol. Biol. Evol.* 20, 955–963.
- Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., Swings, J., 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60, 407–438.
- Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, P., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., Trüper, H.G., 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37, 463–464.
- Welkos, S.L., Vietri, N.J., Gibbs, P.H., 1993. Non-toxicogenic derivatives of the Ames strain of *Bacillus anthracis* are fully virulent for mice: role of plasmid pX02 and chromosome in strain-dependent virulence. *Microb. Pathog.* 14, 381–388.
- Zhang, C., Zhang, M., Ju, J., Nietfeldt, J., Wise, J., Terry, P.M., Olson, M., Kachman, S.D., Wiedmann, M., Samadpour, M., Benson, A.K., 2003. Genome diversification in phylogenetic lineages I and II of *Listeria monocytogenes*: identification of segments unique to lineage II populations. *J. Bacteriol.* 185, 5573–5584.



Sylvain Godreuil is a French medical doctor. He works in the bacteriology service at the Montpellier Hospital. This lab has an activity of clinical diagnosis and research on molecular epidemiology of nosocomial infections and on antibiotic resistance. He is preparing a PhD work at Michel Tibayrenc's lab, on molecular epidemiology of *M. tuberculosis* in Burkina Faso (West Africa), under the supervision of Anne-Laure Baniuls. He has worked on the antibiotic resistance (quinolones) in *Listeria monocytogenes* and beta-lactamase.



Frederick Cohan studies the evolutionary genetics of speciation and adaptation. Under the mentorship of Richard Lewontin and Timothy Prout, he began his career by using *Drosophila* to study the forces of cohesion within animal species. As he grew weary of changing flies, he seized an opportunity to reinvent himself as an evolutionary bacteriologist, with the guidance of Conrad Istock, John Spizizen, and Richard Michod. While he first saw bacteria as a convenient model system for studying very general questions about

evolution that one might rather study in elephants (if one could), he has grown to see bacteria as very cool creatures in their own right. He is intrigued by what is the same and different about species and speciation

across all walks of life, and has investigated how the unique combination of enormous population size and rare but promiscuous genetic exchange in bacteria affects bacterial speciation and diversity. He is a professor of biology at Wesleyan University. He has been a proud fan of the Boston Red Sox for nearly 30 years, but is now prouder than usual.



Haroun N. Shah did his PhD thesis in microbial biochemistry of human pathogens (1980) at University of London. He was appointed as clinical scientist, head, Molecular Identification Services Unit, Health Protection Agency, Colindale (until March 2003, Public Health Laboratory Service). He was a chairs in Several universities, both in the UK and abroad. He was a lecturer, senior lecturer and reader in University of London in 1980–1997. He was fellow of the Royal College of Pathologists in 1996 and Herman fellow at

University of Western Australia in 1998. He is currently a PhD supervisor for five students in the fields of molecular biology and proteomics of human bacterial pathogens. Since 1994 he has been a chair in International Committee on Systematic Bacteria (ICSB), Subcommittee on Gram-negative Anaerobic Rods. He has been a senior editor in *Anaerobe* (from 1995 till date). His researches were on biochemical and molecular basis of microbial pathogenicity, microbial systematics and evolution, genomic and proteomic applications in diagnostic microbiology. He has credit of publishing over 120 peer reviewed journals, 20 chapters in books and one completed book.