Bacterial species and evolution: theoretical and practical perspectives

CA Istock[1], JA Bell[2], N Ferguson[1] and NL Istock[1]

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ; [2]Department of Botany and Plant Pathology, Michigan State University, East Lansing, MI, USA

A discussion of the species problem in modern evolutionary biology serves as the point of departure for an exploration of how the basic science aspects of this problem relate to efforts to map bacterial diversity for practical pursuits—for prospecting among the bacteria for useful genes and gene-products. Out of a confusing array of species concepts, the Cohesion Species Concept seems the most appropriate and useful for analyzing bacterial diversity. Techniques of allozyme analysis and DNA fingerprinting can be used to put this concept into practice to map bacterial genetic diversity, though the concept requires minor modification to encompass cases of complete asexuality. Examples from studies of phenetically defined *Bacillus* species provide very partial maps of genetic population structure. A major conclusion is that such maps frequently reveal deep genetic subdivision within the phenetically defined species; divisions that in some cases are clearly distinct genetic species. Knowledge of such subdivisions is bound to make prospecting within bacterial diversity more effective. Under the general concept of genetic cohesion a hypothetical framework for thinking about the full range of species conditions that might exist among bacteria is developed and the consequences of each such model for species delineation, and species identification are discussed. Modes of bacterial evolution, and a theory of bacterial speciation with and without genetic recombination, are examined. The essay concludes with thoughts about prospects for very extensive mapping of bacterial diversity in the service of future efforts to find useful products. In this context, evolutionary biology becomes the handmaiden of important industrial activities. A few examples of past success in commercializing bacterial gene-products from species of *Bacillus* and a few other bacteria are reviewed.

Keywords: bacteria; species concepts; genetics; diversity; evolution

Introduction

'Whenever we wish to apply our knowledge of bacteriology, we will probably need to know some population genetics.'

John Maynard Smith, 1995 [31]

Beyond sheer intellectual curiosity, why should industrial microbiologists, and the companies that employ them, care about bacterial species and their evolution? In this essay we assert that biological prospecting among microbes is like prospecting for geological resources. To achieve a high measure of success one needs good maps, and an understanding of what they promise. As with underground minerals, metals, coal, gas, and petroleum, good maps of bacterial genetic resources are hard to come by. Indeed, they hardly exist for most of bacterial diversity, particularly at the fine scale of species and subspecific levels needed to explore microbial genetic resources in the future. One way the concept of such a map can be realized is through the hierarchical ordering of substantial numbers of wild isolates from many different habitats in combination with strains from stock collections using computations that create dendograms or phylogenetic trees from genotypic or phenotypic data. Such representations, somewhat analogous to the stratigraphic maps of geologists, can provide orderly

images of the natural structure of bacterial variation at many taxonomic levels. As new isolates are obtained they can be placed within such cluster diagrams. Below, we illustrate the use of dendograms to explore bacterial variation at species and subspecific levels.

In addition to enhancing our chances of discovering novel, useful compounds and enzymes made by bacteria, such maps are urgently needed for vaccine development, and by those studying the epidemiology of newly emerging diseases, the increase and spread of antibiotic resistance among pathogens, and the use of microorganisms for bioremediation and pest control. Maps of bacterial diversity based on the analysis of evolutionary lineages and natural species and subspecific differences can be constructed and potentially turned to practical pursuits.

Currently, our grasp of bacterial diversity consists of: (1) collections of many strains with little information about their relation to natural populations and habitats of genetically related bacteria; and (2) phenotypic species descriptions, sometimes augmented by ribosomal RNA phylogenies or dendrograms based on monophasic or polyphasic analyses [41]. However, it is becoming increasingly clear that collections of strains having a uniform, or nearly uniform, phenotype can contain subgroups of strains that are only distantly related genetically (examples are given below). As we shall see, the utility of bacterial species designations would be greatly enhanced by much greater and more systematic knowledge of genetic relationships. For example, consider a hypothetical bacterial species that contains two genetic subgroups, A and B. If a mem-

ber of A produces an enzyme of commercial interest and a member of B produces an antibiotic, a rational search for a better enzyme would be focused on allies of group A and a search for a better antibiotic would be focused on allies of group B. Such focused searches are only possible if the genetic subgroups have been mapped and if a speedy and reliable method of recognizing members of different subgroups has been developed. Thus, to the extent that genetic subdivisions within classically described bacterial species have not been examined, our current understanding of bacterial diversity is of limited use—supporting nothing better than a prospecting strategy of blind or haphazard searching.

It is not our intent to discuss the relative merits of the many different kinds of genotypic and phenotypic descriptions of bacterial diversity obtained by bacterial taxonomists—this topic has been recently and thoroughly reviewed by Sneath [49], by O'Donnell et al [41], and by Vandamme et al [55] in their reviews of the application of polyphasic taxonomy to bacteria. The latter authors point out that a knowledge of both population genetic structure and evolutionary descent has become essential for a complete systematics of microorganisms, and it is these issues we wish to explore. It is also not our intent to advance yet another definition of species. Like John Maynard Smith [31], we do not believe that a universal species concept is appropriate or can be achieved for bacteria.

Our discussion is directed to both basic and applied microbiologists. After opening with a brief history of the 'species problem' we proceed to an outline of current species concepts, then to a set of nine, non-mutually exclusive, genotypic/phenotypic models of bacterial species. From there we pass to a consideration of how these models might play out in bacterial evolution and speciation and affect our efforts to describe and exploit bacterial diversity. Then we illustrate how dendograms provide orderings of genomic diversity. Next we explore practical perspectives on the use of bacterial genetic resources. A final section looks briefly toward the future and the role industry might play in mapping bacterial diversity. Throughout our discussion, we employ only a few illustrative examples, examples drawn mainly from the genus Bacillus, partly because of our experience with these particular bacteria and partly because of the commercial importance of many species in this genus. Additional examples are drawn from the genera Rhizobium, Agrobacterium, Escherichia, and Neisseria and a few other bacteria.

We hope readers will find this intellectual journey into the realm of bacterial species and evolution as challenging—and as much fun—as we do, and that they will gain a better appreciation of its relevance to practical pursuits.

The species problem

Since the formulation of the so-called 'modern (or evolutionary) synthesis' over a half century ago [21,33], evolutionary biology has engaged in a continuous struggle to make the concept of species clear, rigorous, and universal [20]. The struggle has not brought a satisfying result; the 'species problem' remains unsolved. As Mayr [34] pointed out, Darwin himself was unclear on the nature of species. Views closely allied to the biological species con-

cept (discussed below) predominate in his notebooks (1837–1852), but the view he expressed in the Origin of Species is largely a typological one—the intellectual ancestor of the modern phenetic species concept (discussed below). The idea that species are discrete genetic entities was put forward powerfully by Dobzhansky [11] and Mayr [35], two of the architects of the neo-Darwinian, or 'modern', synthesis. Since the modern synthesis, conceptions and discussions of the nature of species in biology have tended to become more confusing with time, although many biologists today hold tenaciously to the 'biological species concept' that was central to the synthesis [42]. This concept applied only to sexually reproducing organisms, asserting that species were populations of actually or potentially interbreeding populations of organisms [36]. Furthermore, almost all of the thinking in the modern synthesis derived from studies of sexual, multicellular eukaryotes. The microbial world was conceptually invisible to architects of the synthesis, largely putting the first three billion years of evolution beyond its reach.

Prior to the modern synthesis the phenotypically delineated or phenetic species of taxonomists prevailed, as it still does in practice throughout most of taxonomy, though many biologists recognize that a phenetic approach can lead to species delineations having little concordance with natural genetic discontinuities and hence to a very shaky picture of biological diversity. A purely phenetic approach may overestimate or underestimate the number of species, fail to diagnose natural species boundaries, or simply misclassify many specimens, eg strains of microorganisms. The issue is thus an important one if useful, reliable maps of bacterial diversity are to be constructed [41].

The failure to achieve a universal species concept is partly due to competing theoretical notions, principally those of geneticists versus systematists. Geneticists focus on the search for clear delineation of species boundaries using genetic data [11], while systematists seek a phylogenetic (evolutionary tree) reconstruction of evolutionary history [9], and often use phenetic (character-based) [49] methods to distinguish species. Conflicting ideas about what species might be, or should be, have been shaped by the distinctive knowledge and traditions of these disciplines (Table 1). More importantly, attempts by both camps to define species have often foundered on the confusing arrays of natural variation observed among many types of organisms, something especially true for bacteria—possibly the genetically most variable organisms on earth [22,46,47]. At times, pluralistic approaches have been suggested [39]. While these efforts did little more than assemble collections of species concepts, they wisely sought to preserve something of the pax specia that the modern synthesis brought earlier. They have bought time toward the day when greater knowledge will enlighten all those interested in the nature of species.

It has been suggested from time to time that species simply do not exist as a distinct and fundamental taxon, that the problem of species delineation is insoluble, and that if species did exist they would not be different from higher taxa (genus, family, etc) [40]. We do not accept the contention that species are never real, never exist, or are in no way different from higher taxa, though we will entertain

Table 1 Current concepts of evolutionary biology pertaining to species

Category/concept	Definition
(1) *Purely genetic*	
(a) Biological Species Concept	a population or group of populations of actually or potentially interbreeding organisms distinct from all other such populations of organisms, also called the isolation species concept [11,33,35]; also may be used as a non-dimensional, spatially local, concept [36].
(b) Recognition Species Concept	a population or group of populations that share a common mate recognition system leading to effective reproduction, essentially a variant of the preceding concept with a specific emphasis on mating behavior [43].
(2) *Purely phenetic*	
(a) Phenetic Species Concept	a population or group of populations statistically distinguishable solely by one or more phenotypic characteristics, includes the 'morphological species concept', and is the modern rendition of the species concept of traditional taxonomy [50,51]; also may be used as a non-dimensional, spatially local, concept.
(b) Polytypic Species Concept-1	identical to the Phenetic Species Concept, except that multivariate data, usually extensive, are required [50].
(c) Phenetic Bacterial Species	**'a collection of similar strains that differ sufficiently from other groups of strains to warrant recognition as a basic taxonomic unit', quoted from [6].**
(3) *Combined genetic and phenetic concepts*	
(a) Phylogenetic Species Concept	a cluster of individual organisms in one or more populations sharing a most recent common ancestor (monophyletic) and diagnosably distinct from other such clusters based on one or more shared and derived characters; originated in cladist systematics; can include phenotypic and genetic characters [9].
(b) Evolutionary Species Concept	a population or group of populations in a descending lineage that thereby share a common evolutionary fate including both evolutionary changes with time and duration of existence; can include fossil and extant forms, and sexual and asexual forms [39]. It is not clear how this abstract concept would be applied rigorously.
(c) Cohesion Species Concept	various mechanisms—genetic, developmental, and ecological (demographic)—cause a population or group of populations to retain an array of similar genotypes and phenotypes distinct from other such arrays [54].
(d) Polytypic Species Concept-2	similar to the phenetic species concept except that the multivariate data used to differentiate species include both phenotypic and genetic characters; different from the phylogenetic species concept because a phylogenetic reconstruction is not required; analyses proceed using the same methods developed for the pure phenetic species concept [eg 50].
(4) *Non-existence of species*	the view that species are not phenomenologically real; a named species is merely a useful, arbitrary taxon devoid of evolutionary meaning [40].
(5) *Non-dimensional species*	populations of different species in a single locality are readily distinguished one from another even though there are confusing intergradations over larger geographical areas (ie rassenkreisen or ring species, clines, or hybrid zones); the kinds of distinguishing characters locally are likely to be phenetic, including behavioral traits, but could be genetic as well; sometimes referred to as the local naturalist's idea of species [36].

the possibility below that species in a pragmatic, delineating sense, may *sometimes* not exist. Conversely, we will also present examples where species of bacteria are readily distinguished genetically, just as many similar examples can be found among the eukaryotes.

Where species can be unambiguously delineated as genetically circumscribed entities, they are primary in evolution. We refer to these as 'absolute species'. Such a species is a group of populations, a genetically coherent metapopulation, within which genetic variation originates by mutation and recombination, is subject to random drift, migration with gene flow among populations, natural selection, adaptive structuring, and extinction. In the abstract, at least, each absolute species is on a genetically independent evolutionary journey through time (anagenesis). Gathering species under higher taxonomic categories according to their shared properties, however logical phylogenetically, is a different process altogether. Classification into higher taxa depends on the ancestral similarities among species left by evolution rather than the dissimilarities erected by

evolutionary divergence during and after speciation. Given present knowledge, the fundamental laws of genetics and evolution appear to operate only at population and species levels. Attempts to argue otherwise [15,18,57] are not persuasive.

It is inescapable that any rigorous resolution of the species problem rests on our ability to discern the gaps in biological variation to the same extent and as clearly as evolution has created them. However, it is necessary to recall that evolution is more or less continuous through time and unresolvable genetic and phenotypic situations are bound to exist at many times for many groups of evolving lineages— especially when species formation is at some intermediate state between its inception and its conclusion in the absolute sense, if indeed it ultimately reaches that stage. Further, vagueness may arise because the populations and species of completely asexual forms are simply collections of clones, and the clarity with which gaps in variation appear will depend on the extent to which natural selection creates them in opposition to endless diversification by mutation.

Theoretical perspectives: a bestiary of species concepts

In Table 1 we have arranged the most prominent concepts pertaining to species and their definitions into four categories: (1) purely genetic; (2) purely phenetic; (3) combined genetic and phenetic; (4) the non-existence of species; and (5) Mayr's [36] non-dimensional, or local naturalist's, view of species.

Of all the concepts, the one widely applied to bacteria, the Phenetic Bacterial Species (bold in Table 1) is clearly the weakest. It is vague and lacks rigor. It cannot provide a sound basis for explorations of the natural variation of bacterial populations and species. Ancillary use of an arbitrary standard such as 70% similarity from DNA–DNA hybridization measurements (plus a difference of <5°C in the melting temperature of the DNA hybrid compared to the DNA melting temperature of the reference strain) as suggested in the Report of the *ad hoc* Committee on Reconciliation of Approaches to Bacterial Systematics [58] does not improve the situation [31], because such a cutoff does not correspond to any aspect of the biology or evolutionary history of bacteria. Equally arbitrary is the choice of any other percent DNA–DNA similarity, or of some degree of DNA difference in G+C/G+C+A+T ratios [6]. In practice very few natural isolates are used when such techniques are applied, hence no comprehensive mapping of bacterial diversity is achieved.

The Biological Species Concept and Recognition Species Concept have been difficult to apply even to many eukaryotes for several reasons: (1) it may be impossible to test potential species boundaries by experimental mating; (2) experimental matings that produce fertile offspring in captivity may be interspecific matings that do not succeed in nature where the species in fact remain distinct; and (3) the organisms may be strictly asexual in which case these definitions cannot apply. For the same reasons the two concepts are not applicable to bacteria, even when a sexual, genetic exchange mechanism such as transformation, transduction or conjugation exists. Figures 1 and 2, and other information, indicate that *B. subtilis* and *B. licheniformis* are phenotypically and genetically distinct entities in nature even when living close together (ie sympatric, indeed sharing the same soil particles [12]), but they will exchange genes spontaneously in laboratory soil culture [13]. Additionally, there appear to be genetically distinct species within natural populations of these named entities [12,46].

Figure 1 Mapping of genetic diversity based on allozyme variation at ten loci within samples of natural isolates of the phenetic species *B. subtilis* and *B. licheniformis* from a single site in the Sonoran Desert of Arizona, near Tucson. The phenetic identifications were based on metabolic characters (phenotypes). The dendrogram was produced using the unweighted pair-group method with averages (UPGMA). Three standard laboratory strains, M28, 168wt, and W23 are included. Deep genetic divisions within each phenetic species can be seen. The *B. licheniformis* sample contains two distinct species (groups I and II) based on linkage disequilibrium analyses [12], with group II more dissimilar from group I than the latter is from *B. subtilis*. Group A under *B. subtilis* has been named recently as a distinct species, *B. mojavensis* [46]. Figure adapted from Duncan *et al* [12]. Two of the laboratory strains, M28 and 168wt, have divergent genomes relative to the wild isolates of their respective phenetic species and subgroups from the desert populations.
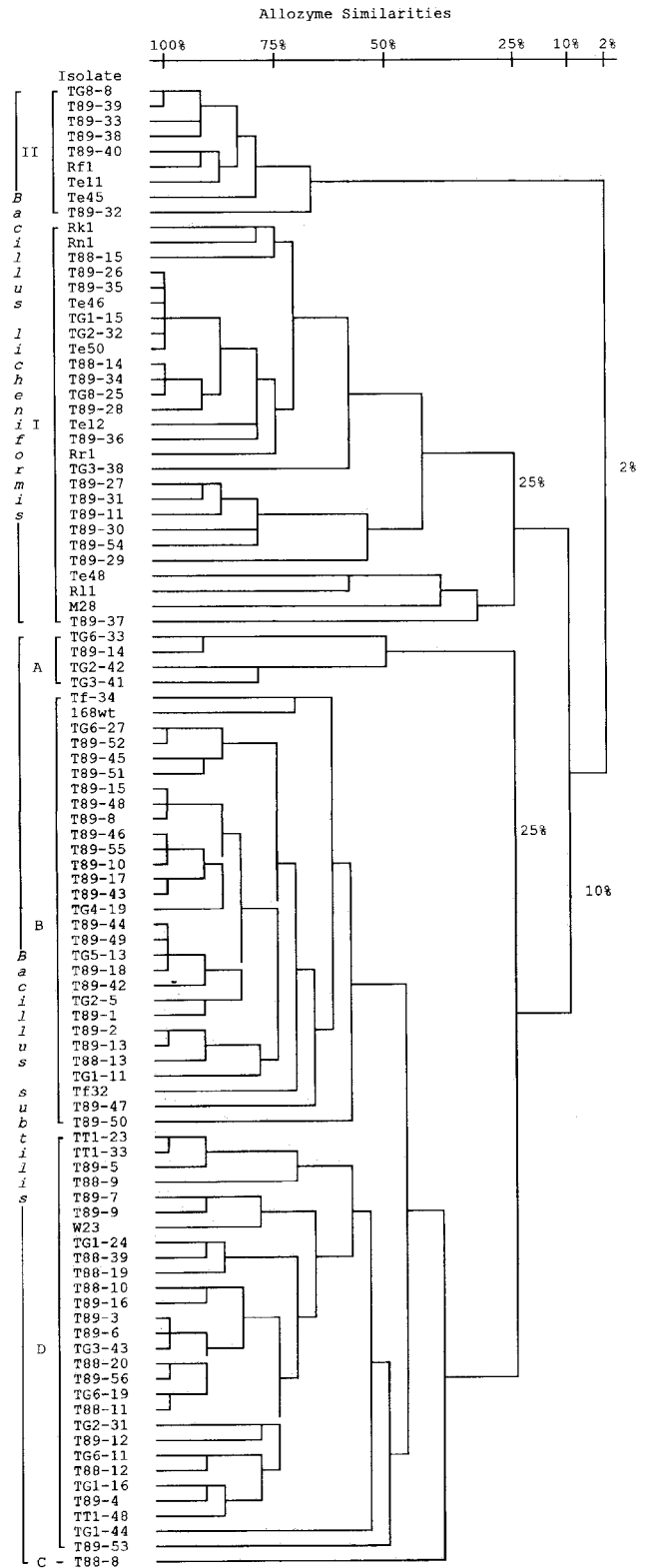
At present, bacterial species delineation can usually be done best using allozyme variation (Figure 2), or one of the DNA fingerprinting methods [10,25,29,59,60,62] (Figure 3), applied to fairly large samples of natural isolates. (Allozymes are alternative molecular forms resulting

Electromorphs of *Bacillus subtilis* and *B. licheniformis*

6PGD ■ ■ ■ ▨ ▢ ▢
AK ▨ ▢ ▢ ■ ■ ▢
AlaDH ■ ■ ■ ■ ■ ■ ■ ▢ ▢ ▢ ▢ ▢ ▢ ▢
EstA ■ ■ ■ ■ ■ ■ ▢ ▨ ▨ ▢ ▢ ▢ ▢ N■
EstB ▢ ▨ ▢ ▨ ▢ ■ ■ ▨ ▢ ▨ ▢ ▨ ▨ ▨ ■ N■
G6PDH ■ ■ ■ ■ ▢ ▢ ▢ ▢
MDH ■ ▢ ▢ ■ ▢
CAT1 ▢ ▢ ▨ ■ ■ ▢ ▢ ■ ■ ■ ■
●CAT2 ■ ■ ■ ■ ■ ■ ■ ■ ■
LeuDH ▢ ■ ▢ ▨ ■
PGI ■ ■ ▢ ▢ ▢ ▨ ▨ ▢
●LDH ■ ■ ■ ■ ■ ■ ■ ■ ■
●PGM ■ ■ ■ ■ ■ ■

B. subtilis	■
B. licheniformis	▢
Both spp.	▨

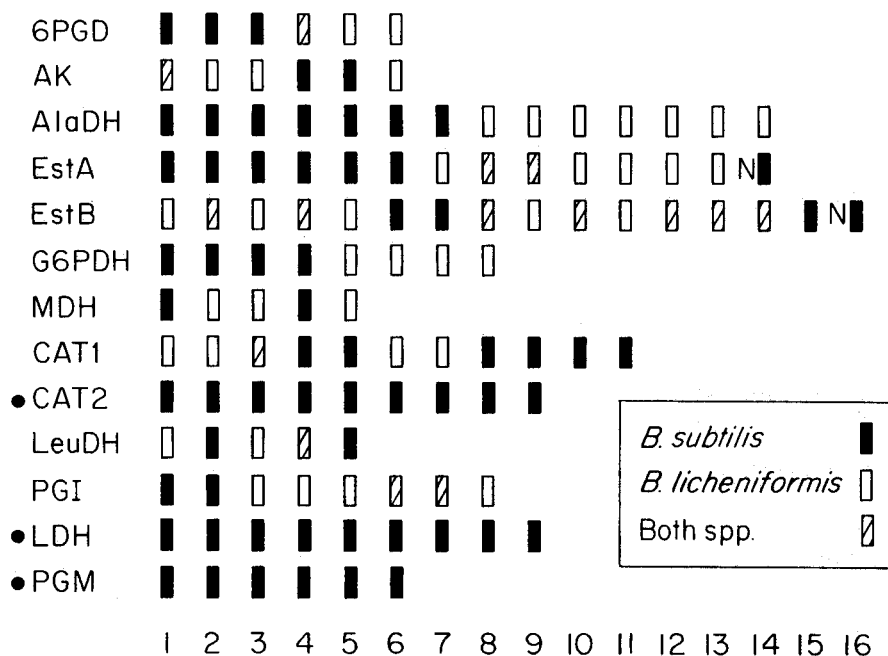1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure 2 Schematic diagram of the allozyme variants (electromorphs) found in the samples of the phenetic species *B. subtilis* and *B. licheniformis* of Figure 1. Abbreviations for the various loci appear down the left-hand side. The variants are numbered from left to right in order of the distance the enzyme molecules migrate down a starch gel, except for those marked N for null alleles (no active enzyme). The three loci marked with a solid circle were assayed only for *B. subtilis*. The gene pools represented by these samples show that the two phenetic species are genetically quite distinct in nature. Figure from Duncan *et al* [12].

from allelic variation of enzyme-encoding genes within the same locus, as distinguished from isozymes that are enzymes of identical function coded for by different loci.) To be valid, this assertion requires that species are made 'cohesive' either by the exchange of genetic material with reasonably high frequency, or by natural selection in a circumscribed ecological niche that tightly constrains mutational diversification of genotypes in purely clonal (asexual) forms of bacteria (but see models 4 and 7 of the next section for potential problems). Thus it appears to us that Templeton's Cohesion Species Concept [54] (Table 1) has the most promise as a theoretical basis for understanding bacterial species and it brings welcome simplification, clarification and generality. However, on occasion we may need to relax its staunchly genetic foundation to cover the wide range of population genetic and phenetic patterns bacteria might exhibit, as we explain below.

One of the most attractive features of the Cohesion Species Concept is that it allows for a range of kinds of species, from strictly asexual species through 'good' biological species to syngameons composed of interbreeding but phenetically recognizable species [19]. In addition, the Cohesion Species Concept explicitly recognizes the population genetic forces that define species—gene flow or the lack thereof, genetic drift, and natural selection—and thus recognizes them as evolutionary entities, something the phenetic species concept cannot readily do [31].

One might ask, why not go beyond analyses using allozymes and DNA fingerprinting to the sequencing of complete genomes, as was done recently with one strain of

Haemophilus influenzae [17]? The answer is that mapping of bacterial diversity requires inclusion of genomic analyses of many isolates. Even if it soon requires only a few months, or even weeks, for complete sequencing of one entire bacterial genome, sufficient sampling of population variation will never be obtained in this way, not to mention the enormous data storage and computer analysis problems that would arise. Once allozyme or DNA fingerprinting methods have produced good general mappings of within-species diversity, selective whole-genome sequencing of isolates from major subdivisions, such as those found in Figures 1 and 4, should provide evolutionary and practical information of great value; ie these maps can serve as useful guides for better directed whole-genome sequencing.

It is tempting to think that the Polytypic Species Concept-2 that combines genetic and phenetic data, the so-called polyphasic analysis of bacterial taxonomy, should be superior to genetic analysis alone. Sometimes this may be true, but one needs to be cautious. As phenetic species, *B. subtilis* and *B. licheniformis* in the Arizona populations were 83% similar based on about 50 metabolic tests, while they were only 2–10% similar based on allozyme (genetic) data (Figure 1), and there were clearly two distinct genetic species within the phenetic species typing as *B. licheniformis* [12], and possibly one or more genetic species within *B. subtilis* [46] (Figure 3). Similarly, the 'good' phenetic species *B. megaterium* (Figure 4) and *B. mycoides* [3] harbor genetic subdivisions just as deep as those between *B. subtilis* and *B. licheniformis*. Based on these preliminary results, we are inclined to think that the
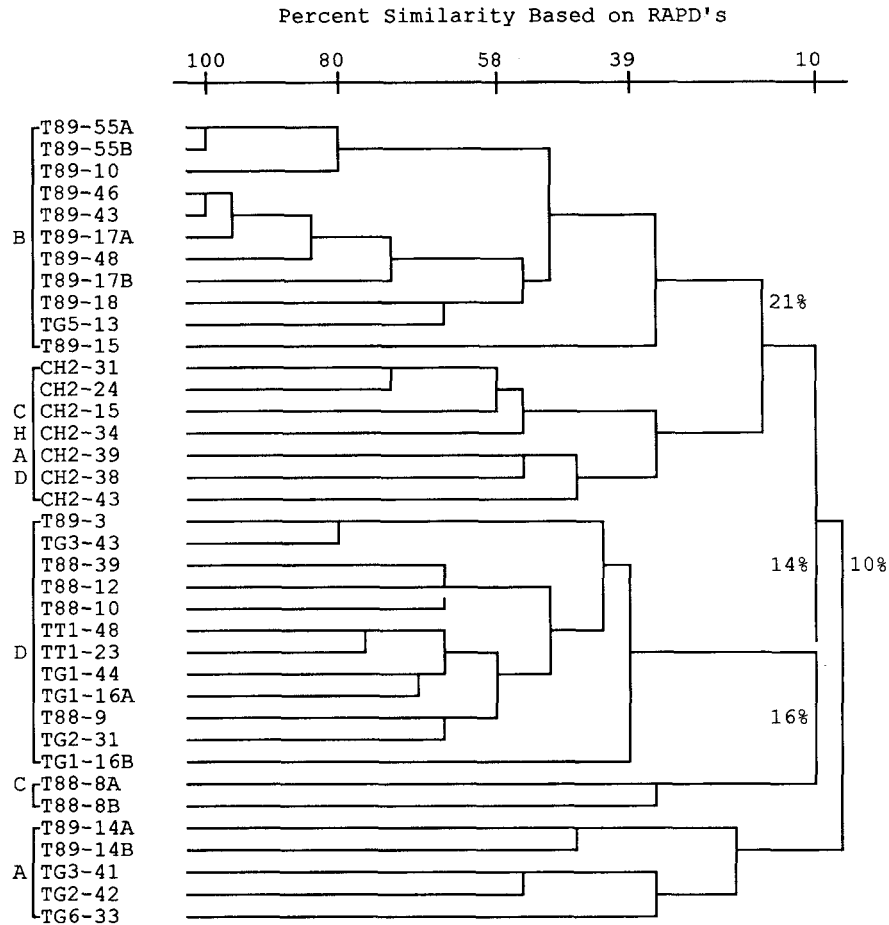
Percent Similarity Based on RAPD's

Figure 3 This mapping of genetic diversity illustrates how samples from different locations can be combined and a map progressively extended. It also provides a comparison of results from allozyme analysis and DNA fingerprinting. Twenty-five of the Arizona *B. subtilis* isolates in Figure 1 were analyzed using the randomly amplified polymorphic DNA (RAPD) method of DNA fingerprinting with two random 10 base-pair primers and PCR. The RAPD agarose gels were analyzed using image-analysis software that catalogs the DNA fragments in each fingerprint. The four isolates with A and B extensions to their isolate labels are from independent fingerprinting reactions, as an indication of repeatability. An exact replica of the four-part structure (A–D) in Figure 1 is obtained using RAPDs. Then, seven isolates typing metabolically as *B. subtilis* from a site in Chad, Africa, were similarly 'fingerprinted' and combined with the Arizona data to produce the full dendrogram. The Chad isolates group together suggesting local genetic differentiation, and the subdivision they add to the map is allied with the Arizona B group at an average of 21% genomic similarity.
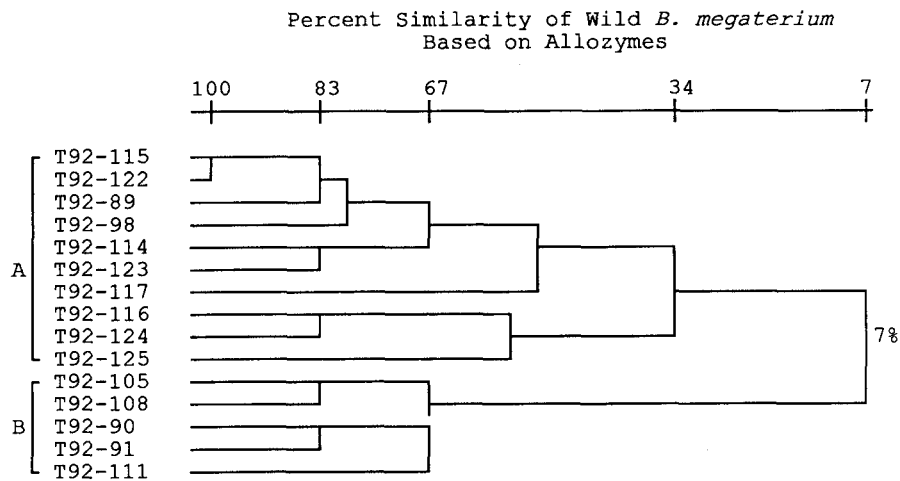
Percent Similarity of Wild *B. megaterium*
Based on Allozymes

Figure 4 Mapping of genetic diversity based on allozyme variation at six loci (Table 3) within a sample of natural isolates of the phenetic species *B. megaterium* from a single site in the Sonoran Desert of Arizona, near Tucson. The dendrogram was produced using the unweighted pair-group method with averages (UPGMA). A deep division appears within the phenetically defined species identified by metabolic characters and large cell-size.

phenetic similarity of allied species will typically be much greater than their underlying genetic similarity, often to the point of masking multiple genetic species. Hence, it is not correct to assume that phenetic and genetic species delineations will generally be concordant, though sometimes they may be [44].

The Phylogenetic Species Concept is difficult to apply to bacteria. Genetically distinct asexual clones would each be species. Conversely, when extensive lateral transfer of genes has occurred, phylogenetic methods lose resolving power. (This is not true of dendrograms produced using the unweighted pair-group method with averages or UPGMA, which some people consider to be a phylogenetic method, but others do not.) The Evolutionary Species Concept is vague and not obviously operational, though logically it appears to be subsumed under the Cohesion Concept.

Therefore, to reiterate, we suggest that application of the Cohesion Species Concept when it involves primary reliance on genetic analysis using allozyme variation or various forms of DNA fingerprinting with RAPDs, Rep-PCR, and RFLPs [10,25,29,59,60,62] and the like, provide the best initial approach currently available for the differentiation of natural genetic clusters in nature. When we thereby affix the status of species to repeatedly diagnosable genomic clusters we do so with at least some confidence that we have discerned something of the natural order of bacterial diversity, though with strictly asexual forms there may often be lingering arbitrariness. With samples from sympatric (ie with no spatial or geographical separation) bacterial populations that actually engage in extensive genetic exchange (sexuality), there are standard statistical methods for pairwise and multi-locus analyses of linkage disequilibrium that can help in testing putative species boundaries [12,22,31,32].

A hypothetical framework for population structure and species of bacteria

We can frame the bacterial species problem in a general way through a series of imaginary, alternative models of the species or non-species structure of a monophyletic bacterial genus. Figure 5 provides a schematic representation of this hypothetical framework and possible transitions between species states during bacterial evolution. This framework is meant to portray a broad spectrum of situations that we might find if we were able to determine the genotype and phenotype of every extant cell within our hypothetical genus. The nine models defined below involve shifting influences from population structure, dispersal (gene flow), recombination, spatial scale, and selection among genomes based on their phenotypic attributes. While the clonality in local populations of sexual bacteria may be higher or lower due to absolute population densities, the models depict states arising when average densities are high enough to permit genetic recombination. Obviously, it is impossible to determine the genotype and phenotype of every cell for even one genus. The point is that extensive sampling will be necessary to determine natural species structures. It is important here to note that in bacterial population genetics, the term 'sexual' is used to mean 'recombining' and the term 'clonal' is used to indicate organisms that are both

'non-recombining' and 'genomically identical by descent.' In the names for some of the models the term 'panmictic' is used to signify that sexuality, and hence extensive genetic recombination, is possible, and may in the extreme lead to panmixis, ie the random association of alleles at different loci.

Here, 'phenotype' refers to metabolic or morphological characteristics, and the terms 'genotype' or 'genetic' refer to markers such as allozymes and various forms of DNA fingerprinting that can assess genomic variation. Prospects for classification and identification of species are noted with each of the nine models. Throughout, it is assumed that a monophyletic *genus* of bacteria is the entity within which species do or do not form. Different types of species and non-species structures might evolve within the same genus. Whether or not a species is geographically 'cosmopolitan' (ie genetically or phenetically very similar everywhere it occurs on earth) adds an additional dimension to these models, one of particular interest in the biology of bacteria. The ensuing models fall at various points within or, in the last two models, beyond the continuum of possibilities under the Cohesion Species Concept [54]. For simplicity, each model is presented as if the entire genus conforms uniformly to one type of species or non-species formation. In reality, more complex situations where a genus conforms to two or more of the models are likely to occur.

(1) Cosmopolitan panmictic biological species
Species arise as separate, genotypic and phenotypic divergences between global metapopulations with extensive *within*-species recombination. Local populations of sufficient density tend toward panmixis (random or free recombination). Species are geographically cosmopolitan. Genetic exchange *between* species is not sufficient to erase well-marked species metapopulations. Classification and identification of species are easily done genetically.

(2) Non-cosmopolitan panmictic biological species
Species arise, but undergo local genetic and phenotypic differentiation. Extensive *within*-species recombination occurs. Local populations of sufficient density tend toward panmixis. Selection causes phenotypic divergence between local populations, and selection or drift or both cause genetic differentiation between localities. A diffuse system of dispersal and recombination unites a vast array of local subpopulations into a geographically varying metapopulation. Most genotypes within species can still recombine with each other. Genetic exchange *between* species is not sufficient to erase the species metapopulations. However, the species are difficult to classify and identify due to their protean nature. Nonetheless, strong species boundaries actually exist. It requires a large amount of analytical work, with large samples of isolates, to delineate species boundaries accurately, but there are good species because the members of each species belong to a strictly monophyletic group.

(3) Phenetically cosmopolitan panmictic biological species
This model is identical to the preceding one except that strong ecological (natural) selection stabilizes the same cir-

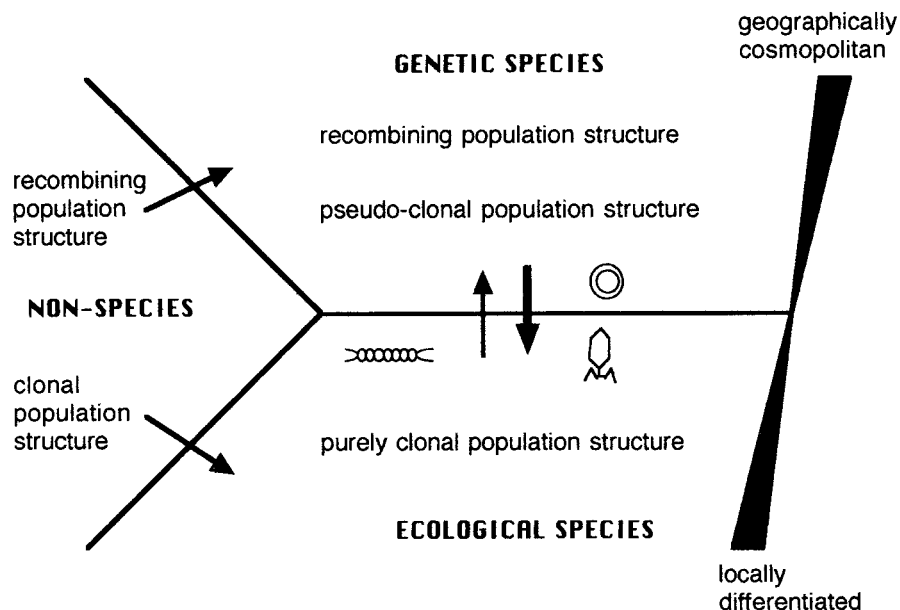# SPECIES STATES AND TRANSITIONS

Figure 5 Schematic rendering of the elements used to define the nine models of the 'Hypothetical Framework for Population Structure and Species of Bacteria' (see text). On the far left is the non-species condition of models 8 and 9, possibly reflecting a primordial stage in early bacterial evolution before either genetic or ecological species emerged, or a subsequent condition in a population formed by widespread genetic recombination but preceding its 'crystallization' into a cluster of species. Strictly ecological species are completely asexual, and thus purely clonal (models 6–7). Genetic species (models 1–3) may have recombining populations if recombination is frequent, but if gene transfer is rare they will appear clonal, ie as pseudo-clonal species (models 4 and 5). The three figures near the center line represent conjugative plasmids, transducing bacteriophages, and transforming DNA (ie direct DNA exchange between cells), which along with conjugative transposons are the known mechanisms for bacterial sexuality. Genetic species may evolve to become ecological species by losing their sexual mechanism (larger downward arrow), but a reverse transition from an ecological species to a genetic species can occur through the origin of one or more of the sexual mechanisms. On the right (perpendicular to the plane of the figure) it is suggested that the geographical structure of many of the species states may range from cosmopolitan to highly differentiated from one place on earth to another.

cumscribed range of phenotypic variation (narrow niche) within each species everywhere, even though considerable local, non-adaptive, genetic differentiation develops via mutation. Phenotypic differences between species are well marked. The species are more easily classified and identified phenetically than by genetic analysis, though extensive genetic analysis confirms the monophyly of each species.

(4) Cosmopolitan clonal biological species
Species form as restricted sets of clones repeatedly found worldwide, but retain some capacity for recombination, much more *within* than *between* species, but never with enough recombination to make local populations even quasi-panmictic. The successful clonal sets of each species are the consequence of strong ecological (natural) selection, and are readily recognized and classified phenotypically and genetically. These are the pseudo-clonal species of Figure 5.

(5) Weakly cosmopolitan clonal biological species
Species form as sets of divergent, highly variable, clonal, local populations. Each species experiences enough recombination globally *within* the species to preserve sufficient genetic and phenotypic similarity (weakly cosmopolitan) to make it vaguely distinguishable from other species. Genetic exchange *between* species does not erase the species meta-

populations. Genetic and phenotypic variation between local populations of the same and different species is great enough to make classification and identification difficult, but sufficiently extensive data for many isolates can probably map the species metapopulations. Since the species themselves are real, sufficiently extensive polyphasic analysis will likely yield correct species delineations. These, too, are pseudo-clonal species.

(6) Cosmopolitan clonal ecological species
Species form as non-overlapping, small, *global* sets of genetically and phenotypically distinguishable clones. Most clones of a given species set are found world-wide, a few occur more locally, but the metapopulations of species sets are recognizable because mutational spread of genetic variation is opposed by the strong force of normalizing selection that allows propagation of only those genotypes of each species that have high fitness in a narrowly circumscribed ecological niche. There is no, or exceedingly rare, recombination *within* or *between* species. The clonal sets of each species are maintained by persistent differential extinction of less fit clones which would otherwise make the clonal variation across species more or less continuous and confusing both genetically and phenotypically. Classification and identification are not difficult. Lacking recombination, the species are *ecologically* enforced clonal sets.

(7) Phenetically cosmopolitan ecological species

Species form that have circumscribed phenotypic variation, restricted *local* sets of genetic clones, and no, or rare, recombination *within* or *between* species. The clonal sets at many localities are genetically different, almost unique. However, phenotypic resemblance is sufficiently consistent within presumptive species to make phenetic metapopulations recognizable, presumably due to selection among genetic systems of common ancestry that have diverged greatly through mutation and drift but continue to be forced toward phenotypic conformity due to ecologically conservative natural selection. Phenetic classification and identification is possible. Here the Cohesion Species Concept is modified with cohesion resting on the preservation of phenetic similarity that is statistically so strong as to rule out convergence of unrelated lineages, even though the evidence from genetic similarity on which Templeton's [54] original concept relies is too weak to define species.

(8) Phenetically continuous clonal non-species

Species do not form among the organisms of this bacterial genus. There are restricted sets of genetic clones, *without recombination*, at all localities, but the clonal sets tend to be genetically different, almost unique, at each location. Strains isolated from different localities are so widely and continuously variable phenotypically that they cannot be classified, except possibly for small areas (the 'non-dimensional species' of Mayr [36]). Mutation, drift, and selection have such a diversifying effect on phenotypes over larger spatial scales as to preclude even the description of ecological species, or the recognition of metapopulations. It is possible that only part of a genus may conform to this model, while species of types 1–7 arise as well.

(9) Highly variable partially recombining non-species

Species do not form among the organisms of this genus. There is seemingly endless genetic and phenotypic variation of genomes everywhere. Variable amounts of *recombination occur* via complex networks of DNA sequence homology creating confusing avenues and barriers to recombination in space and time, and thus indistinct, seemingly unbounded metapopulations. Local populations of sufficiently high density are quasi-panmictic. As the sample of isolates from different locations becomes large, objective clustering and classification become impossible. For this model, one might argue that the species (highly polytypic) is the same as the genus (monotypic). However, the issue is not semantic if the amount of phenotypic and genetic variation encountered in this genus is similar to that for all the species combined within other reasonably speciose genera adhering to models 1–7. Furthermore, the fact that many genomes may or may not recombine with each other, and the absence of ecologically enforced phenotypic species subdivisions, puts this system in an altogether different, amorphous realm. The subgeneric taxonomic problem is fundamentally intractable, and the genus as an evolutionary unit is meaningless because it lacks cohesion and will continue to do so unless at some point it 'crystallizes' or 'collapses' into separate, smaller genetic or phenetic entities, possibly through considerable extinction of considerable

portions of its field of variation. (Taxonomists fervently hope that models 8 and 9 are purely hypothetical.)

Many evolutionary transitions between the various models can be imagined. In particular, intermediates and transitions between model nine and the other models with recombination might involve partial 'crystallization' of species with some genetic exchange among species still occurring; such a group of species has been called a 'genospecies' [45] because the species, to whatever extent they are discernible, remain embedded in a larger and somewhat confusing metapopulation.

Accessory, self-transmissable, genetic elements such as plasmids, bacteriophage, and transposons will also have varying genetic consequences within different species structures. When they routinely pass between host species these elements may have separate metapopulations of their own or they can expand or contract those of their hosts. In some cases, plasmid-borne traits are used to define species (*Agrobacterium* spp, *Bacillus thuringiensis*, *Bacillus anthracis*, *Rhizobium* spp). Indeed, accessory elements may control the evolution of species diversification in some cases.

Table 2 summarizes major aspects of the nine models. The Cohesion Species Concept serves well as the theoretical basis for models 1, 2, 4, 5, 6, and to a lesser extent for model 3. Model 7 requires the reinterpretation of cohesion discussed under that model, becoming a phenetically based modification of the concept, but one consistent with Templeton's [54] idea of 'demographic', ie phenotypic, 'exchangeability'. Also in this model, the concordance of genetic and phenetic species that Templeton [54] assumes to be universal is not so, as we found earlier in the empirical examples from *Bacillus* studies. Models 8 and 9 postulate situations where genetic and phenetic cohesion are absent.

Evolution and speciation in bacteria

The nine models of the preceding section suggest a number of contrasting modes of evolution for bacteria, sometimes involving accessory genetic elements:

(1) evolution of strictly asexual (clonal) bacteria with intense normalizing selection among mutationally vary-

Table 2 Defining characteristics of nine hypothetical models for species states of bacteria discussed in the 'Hypothetical Framework' section

Characteristic	Model number								
	1	2	3	4	5	6	7	8	9
Biological (genetic) species	+	+	+	+	+	−	−	−	−
Local populations potentially panmictic	+	+	+	−	−	−	−	−	+
Metapopulations exist	+	+	+	+	+	+	+	−	+
Metapopulations discernible	+	+	+	+	?	+	+	−	−
Species phenetically cosmopolitan	+	−	+	+	w	+	+	−	−
Species genetically cosmopolitan	+	−	−	+	w	+	−	−	−
Ecological species	−	−	−	+	−	+	+	−	−

+, Present; −, absent; w, weak; ?, uncertain.

ing clones when natural selection opposes mutational spread and genetic drift and thus enforces genotypic and phenotypic conformity (ecological species);

(2) similar evolution of potentially sexual bacteria when density is too low for the frequent outcrossing required to overcome clonal selection;

(3) classical evolution by selection within large fields of mutating and recombining variation within well-circumscribed biological species, though the degree of 'clonal structure' in local populations may be variable;

(4) genospecies evolution by selection within fields of recombining variation open to some extent between biological species [45], with species more or less distinct depending on the amount of gene flow between them; and

(5) evolution due to shifting local patterns of mutation, recombination, and selection within metapopulations without formation of species.

The evolution of a single bacterial species or lineage might be shaped by several of these modes operating simultaneously in different places, or at different times. The various modes suggest great potential variety and flexibility in bacterial evolution and adaptation.

Species formation (species splitting or cladogenesis), when it does occur, may not always or even typically conform to the geographical, or allopatric, process championed by the architects of the evolutionary synthesis. They thought the universal form of speciation was a process where complete geographical separation (allopatry) of two or more populations (or metapopulations) persisting for sufficient time allowed mutation, recombination, genetic drift, and natural selection to create complete genetic incompatibility between populations (reproductive isolation). Clearly, the concept cannot apply to asexual forms.

Both sexual and asexual species of bacteria may originate by quite different processes under complete sympatry. A clue to how such a process of speciation might occur comes from the observation of deep genetic subdivision within the population structures discussed earlier. In simplest form this model of bacterial speciation involves an initial event in which a clone appears that has high fitness (ie possesses a high rate of binary fission) relative to its ancestral population. The initiating event may involve mutation, recombination, the creation of a new coalition between host and accessory element, or a combination of these. At the outset the 'upstart' clone proliferates rapidly, creating a distinctive, spatially localized, subpopulation within the ancestral population.

If asexual, the upstart will form a new species if the genetic and ecological reasons for its high fitness are enduring, and if natural selection constrains its phenotypic or genetic variation, or both, subsequently arising by mutation. Thereby the new species remains recognizable, and thus, an ecological species is born.

Suppose, however, the upstart is sexual and engages in recombination through both incrossing among its descendants and outcrossing with its ancestral population. Whether it becomes a new species or not will depend on the relative strengths of the incrossing and outcrossing rates as mutation and recombination create variation within the upstart population. If outcrossing is frequent and genetic and ecological divergence from the ancestor are slow to develop, the upstart will simply fade back into the ancestral field of variation, losing its initial fitness advantage. The upstart population could retain its fitness advantage if the incrossing rate remains high while mutation, limited outcrossing, and possibly coevolution with one or more accessory element(s), along with natural selection foster rapid divergence from the ancestral population. As this happens, the outcrossing rate falls and the upstart population will 'escape' genetically and probably ecologically from its ancestral population. In the case of bacteria where sexuality depends on transformation, divergence between the upstart and ancestral populations in DNA sequences may be decisive because the rate of recombination is dependent on DNA sequence similarity [8]. If sexuality occurs by transformation it follows that when it bursts forth the upstart, initially a single clone, has an 'incrossing' rate that is maximal and higher than its outcrossing rate. To 'escape' as a new species it requires 'a run of good luck' through favorable mutation, and some fitness-enhancing, recombination events (both incrossing and outcrossing) that carry it to genetic independence [12].

Speciation of asexual and sexual bacteria may also occur allopatrically in geographically isolated populations, but if the ancestral species is cosmopolitan speciation in sympatry is the only possibility. At present we do not know the extent to which non-pathogenic bacterial species are cosmopolitan. Among pathogenic and commensal bacteria we know that both cosmopolitan and locally differentiated forms occur, eg in the case of *E. coli*.

Examples of mapping of bacterial diversity

The dendrogram in Figure 1 resolves patterns of genetic relatedness among 60 wild isolates of *B. subtilis* and 34 isolates of *B. licheniformis* from populations at a tiny field site in Arizona [12]. Genomic variation among the isolates of *B. subtilis* was determined using ten enzyme-encoding loci, commonly referred to as allozyme variation in the population genetics literature [48]. *B. licheniformis* was typed for the same ten loci. Allozyme variation for *B. subtilis* and *B. licheniformis* is illustrated schematically in Figure 2. Extensive statistical and molecular-biological analysis suggests that this population genetic structure of *B. subtilis* has resulted from the interplay of past clonal differentiation and considerable genetic exchange within and possibly among its four major subdivisions (Figures 1 and 3, groups A–D) [12]. The group labeled A has recently been named as a distinct species, *B. mojavensis*, based on restriction fragment comparisons [46]. An example of how existing strains of commercial value could be mapped into larger descriptions of natural diversity is illustrated by the inclusion of standard laboratory strains M28, 168wt, and W23 in Figure 1.

Figure 2 makes another point about genetic variation among the *B. subtilis* and *B. licheniformis* isolates. The two phenetically defined species share only 15% of the allozyme variants observed for ten loci [12]. Closer scrutiny of the allozyme data and analysis of degrees of linkage disequilibrium (ie degrees of recombination involving

alleles at different loci) showed that there are definitely two distinct genetic species within the isolates that originally typed phenetically as *B. licheniformis* (groups I and II) [12]. Similarly, allozyme data reveal a previously unknown, deep, genetic subdivision within the phenetic 'species' *B. megaterium* from a single Arizona desert population (Figure 4 and Table 3; Istock *et al*, unpublished). There is only 7% similarity between the two subdivisions because five of six loci are strongly divergent. Also remarkable is the similarity of allelic frequencies at the glucose-6-phosphate dehydrogenase locus, suggesting that for this locus genetic exchange between subdivisions may be frequent and unopposed by natural selection. Such limited genetic bridges among allied bacterial species may or may not be common in nature. However, interspecific genetic exchange will not prevent good resolution of species boundaries so long as a majority of loci or DNA sequence comparisons define sharp and concordant discontinuities after extensive sampling.

Another example, shown in Figure 4, indicates that rapid DNA fingerprinting techniques, of which many are appearing [10,25,59,60], can produce concordant and extendible maps of bacterial diversity. A subset of 25 of the 60 *B. subtilis* isolates (Figure 1) were reanalyzed using the Randomly Amplified Polymorphic DNA (RAPD) technique [62] followed by computer-assisted image analysis of the resulting DNA fragments in each genomic 'fingerprint'. With only two 10-base random primers in separate polymerase chain reactions, we obtained exactly the same four

Table 3 Allozyme frequencies for a sample of 15 *Bacillus megaterium* isolates from Tumamoc Hill, Tucson, Arizona. The numbering of the alleles within loci is not consecutive because the alleles here are part of larger numbering system involving other species

Enzyme locus	Allele	Allele frequency	
		Group A	Group B
Alanine dehydrogenase	6	0.20	
	10		1.00
	16	0.80	
Esterase A	8	0.20	
	10	0.20	
	11	0.50	
	12	0.10	
	16		1.00
Glucose-6-phosphate dehydrogenase	8	0.20	0.20
	9	0.30	0.20
	10	0.50	0.60
Malate dehydrogenase	1		0.20
	7	1.00	
	8		0.80
Catalase 1	4		0.60
	5	0.10	
	10	0.20	
	11	0.50	
	13	0.20	0.20
	14		0.20
Leucine dehydrogenase	1	0.80	
	5	0.20	
	7		1.00

major subdivisions that resolved with allozymes. Furthermore, when an African sample from Chad was added to the analysis it resolved as a distinct subgroup most closely allied with group B from Arizona (Figures 1 and 3; Istock *et al*, unpublished data).

Evidence is accumulating that gene transfer not only occurs in bacteria [12,13,27,28,30], but that it plays a significant role in their evolution [26]. We will briefly mention a few examples chosen for their illustrative value. Although early studies using allozyme variation among commensal isolates of *Escherichia coli* indicated that this resident of the human gut was strictly clonal [7], subsequent comparison of DNA sequences from a number of genes provided evidence of genetic recombination in the form of 'mosaic' genes containing nucleotide segments traceable to different parental isolates [14,37,38]. Allozyme analysis also indicates that strains of *E. coli* pathogenic to chickens are much less clonal [12,61]. When five populations of symbiotic *Rhizobium etli* bv *phasioli* were examined, two possessed clonal and three recombining population structures [52]. Transfer of nodulation genes from introduced *Rhizobium loti* to non-symbiotic soil rhizobia allowed recipients to form nodules on the host *Lotus corniculatus* [53]. *Neisseria meningitidis*, an important pathogen of humans, possibly has recombining populations in two serogroups, and a clonal structure in a third serogroup that in addition has mosaic genes for several loci including those coding for penicillin-binding proteins [30]. Mosaic sequences have also been found in *Neisseria gonorrhoeae* [32]. We know that rare genetic exchange can occur between evolutionarily distant lineages of bacteria [24]. For example, Bostin and Reeves showed that part of the O-antigen gene in *E. coli* may be derived from a very distantly related bacterium having a low percent G+C in its DNA [5]. Recombining populations of *Neisseria meningitidis* also serve as an example of the 'epidemic' population structure described by Maynard Smith [32]. In such a structure, individual clones can arise and, if more fit than their competitors, come to comprise a large proportion of the population. This state of affairs will last until they are in turn displaced by more fit clones or until they lose their identity by undergoing recombination. This population structure is possible even in frequently-recombining bacteria because recombination and reproduction by fission are not obligatorily tied together.

Several lessons come from the foregoing examples. First, even bacteria that have clonal population structures have been shown to undergo recombination, although possibly infrequently. Second, populations of the same species at different places, or in different ecological settings, may exhibit either clonal or recombining population structures. Third, gene transfer, as in the *Rhizobium loti* case, can play a dramatic role in opening a new niche to a recipient or by allowing the recipient to survive changes in its present environment [53].

Practical perspectives

In this section we explore the relation between the mapping of species and subspecific population genetic structure and strategic prospecting for useful products derived from bac-

terial diversity. To assure us all that prospecting within bacterial diversity is worth the candle, we present some examples of past successes.

Strains assigned to several phenetic species in the genus *Bacillus* are currently important for the commercial production of enzymes [16], peptide antibiotics [63], and other products [56]. The species involved include *B. subtilis*, *B. licheniformis*, *B. megaterium*, *B. amyloliquifaciens*, *B. brevis*, *B. cereus*, *B. circulans*, *B. polymyxa*, and *B. coagulans*. At least 11 antifungal, antibacterial, and antitumor agents found in *B. subtilis* are commercially produced, and another 13 such agents from six other *Bacillus* species (*licheniformis*, *brevis*, *circulans*, *cereus*, *polymyxa*, and *mesentericus*) have been brought to production [63]. In addition, *B. thuringiensis* and *B. sphaericus* provide living insecticides. The record of success with *B. subtilis* is not surprising since it has been the most intensively studied species in the genus. This fact, combined with more limited rewards from the other species, suggests that many more genetic variants and species will prove useful for pharmaceutical and industrial applications in the future.

Antibiotics that have different chemical structures and modes of action are usually produced by different organisms. *Bacillus* species make peptide antibiotics that attack cell membranes, for example gramicidins are linear peptides made by *B. brevis*, bacitracins are cyclic peptides made by *B. subtilis* and *B. licheniformis*, and polymixins are cyclic peptides substituted with fatty acids made by *B. polymyxa*. In contrast, aminoglycosides interfere with protein synthesis and are mostly made by bacteria in the genera *Streptomyces* and *Micromonospora*, eg streptomycin by *S. griseus* and other species, neomycin by *S. fradiae*, spectinomycin by *S. spectabilis* and *S. flavopersicus*, and gentamycin by several *Micromonospora* species [4].

For *B. subtilis* and *B. licheniformis*, or any other *Bacillus* species, it is unknown to which subdivisions (Figures 1 and 3) the commercially valuable strains belong and thus the best place to look in nature for other genotypes with more potent versions of the currently useful products. Alternatively, once we determine where the currently valuable strains fall in the larger field of natural variation, other subgroups of *B. subtilis* and *B. licheniformis* may be good places to prospect for unique products. In short, it will be possible to base prospecting strategies on 'maps' of bacterial diversity, similar to Figure 1, that faithfully resolve the lineages evolution has produced. The foregoing examples argue strongly that maps placing these and many other valuable bacteria within a much larger rendering of natural bacterial diversity based on genetic relatedness at species and subspecific levels would provide a scientific foundation for future prospecting.

Genetic engineering will undoubtedly accomplish much by fashioning useful new bacterial genomes [1,16], but full realization of the promise of genetic engineering will ultimately depend on exploitation of the large repository of natural variation, the vast pool of building blocks, among organisms living in the wild. Only by having a much better genetic, ecological and geographical mapping of bacterial diversity can we hope to make prospecting for natural products an efficient endeavor. The current culture collections of bacteria hold only a tiny fraction of the earth's store of

bacterial variety. It is possible that present-day evolutionary change within the whole world of bacteria is generating new genes, physiological pathways, species, and potential natural products faster than we can hope to discover them. An intriguing example comes from a study by Barns *et al* [2]. They retrieved archaebacterial, ribosomal RNA sequences from a single hot spring using the polymerase chain reaction and found many distinct sequences belonging to organisms not closely related to known Archaea, many of which may come from new species.

Where among the many varied habitats on earth should one search for new useful genes and natural products from bacteria? There is no ready or simple answer to this question, although some might recommend extreme environments such as hot springs, thermal vents, extraordinarily arid deserts, Antarctic lakes, or highly saline lakes. Extreme environments are certain to provide isolates with unusual genetic and physiological adaptations, but less extreme environments are likely to harbor greater species and genetic diversity. The kinds of mapping we have been discussing also provide help on this front if sampling from many kinds of habitats is included when obtaining the genetic or phenetic data used to construct these maps. For example, we have recently found (unpublished) that particular subgroups of *B. subtilis* are better represented in some places than others: the Chad group in Chad (Figure 4), the C group in a Tunisian desert, the A group at a site in the Atacama Desert of Chile, and the B and D groups in the Sonoran Desert of Arizona.

Prospectus

For more than three centuries, countless naturalists have set about describing the earth's divergent life forms, with primary emphasis on the eukaryotes. More than a million species later their work is far from done, and extensive description of prokaryotic life only beginning. Two and a half centuries ago Linnaeus imposed order upon the emerging picture of biological diversity with the binomial naming of species and hierarchical taxonomic classification. In the 19th century it became apparent to Lamarck that evolution had occurred, and to Darwin and Wallace that evolutionary diversification had a single cause—evolution by natural selection—and that life seemed to come packaged as discrete species formed during evolution. The 20th century brought the modern synthesis and its aftermath of confusion over the nature of species. Now comes a rising concern that our own species is engaged in activities that have caused and will increasingly cause the extinction of other species, described and undescribed alike, and a concomitant awareness that useful biological resources lie encrypted within the genes of many of these species, especially the microbes. Suddenly, evolutionary biology is not just an academic enterprise pursued by curious minds; it is the sound scientific foundation for a new gold rush.

Metaphorically speaking, the earth's evolving bacterial lineages are like a vast array of genetic factories turning out new gene products under the aegis of mutation, recombination, and natural selection. The 21st century mapping of the earth's bacterial diversity must search out definable entities at strain (clone), population, and species levels—

entities with recognizable genetic or phenetic boundaries within which to search for novel and repeatable sources of natural products. But, like sourdough prospectors of gold-rush days we do not have good maps, and without such maps many searches will be expensive and fruitless.

Recent studies of bacterial population structure and bacterial species suggest both that the terrain of bacterial diversity to be mapped is rugged and potentially complex, with many deep genetic divisions within and between the species recognized by traditional bacterial taxonomy, and that many new species remain to be discovered. From the welter of current conceptions of species we found that only a modified Cohesion Concept of species is consistent with the variety of ways in which the genetic, phenetic, ecological, and geographical properties of bacterial populations may be diagnosed. Implementation of this approach to the delineation of bacterial species and subspecific genomic groups must rely first of all on the use of modern methods of genetic analysis. When genetic data cannot resolve species boundaries it is still possible that phenetic methods may do so, but with less certainty that natural evolutionary lineages have been uncovered, as opposed to ones that have converged phenotypically from genetically long-separated ancestral lineages. Combined genetic and phenetic analyses may also uncover instances where the terrain cannot be charted at the species level because species do not exist, then only strain by strain mapping is possible.

Comprehensive mapping of bacterial diversity for even a few genera is an enormous undertaking. Large numbers of isolates need to be typed genetically and phenotypically. The effort can start with genera that have traditional phenetic species that are readily cultured and perhaps within which commercial value has already been found. Eventually there will be a need for techniques that do not require laboratory colonization; such techniques do not yet exist in a form that can match the resolving power of allozymes and DNA fingerprinting. Companies with large strain collections are best positioned to undertake such a program of bacterial diversity mapping. However, beyond existing strain collections it will be necessary to include large samples of wild isolates of precisely known geographical and ecological (habitat) origin. Only in this way can strain collections be related to the structure of bacterial diversity in nature. Similarly, in the case of plant, animal and human pathogens, mapping of genomic diversity can reveal relationships within the genetic structure of these bacteria across many host populations.

Acknowledgements

References

1 Arbige MV, BA Balthuis, J Schultz and D Crabb. 1993. Fermentation of *Bacillus*. In: *Bacillus subtilis* and Other Gram-Positive Bacteria (Sonenshein AL, JA Hoch and R Losick, eds), pp 872–895, Am Soc Microbiology, Washington, DC.

2 Barns SM, RE Fundyga, MW Jeffries and NR Pace. 1994. Remarkable archael diversity detected in a Yellowstone National Park hot-spring environment. Proc Natl Acad Sci USA 91: 1609–1613.

3 Bell JA and SB Friedman. 1995. Genetic structure and diversity within local populations of *Bacillus mycoides*. Evolution 48: 1678–1714.

4 Berdy J, A Aszalos, M Bostain and KL McNitt. 1980. CRC Handbook of Antibiotic Compounds vol 1; vol 4, part 1. CRC Press, Boca Raton, Florida.

5 Bostin DA and PR Reeves. 1995. Sequence and analysis of the O-antigen gene (rfb) cluster of *Escherichia coli* 111. Gene 164: 17–23.

6 Brock TD, MT Madigan, JM Martinko and J Parker. 1994. Biology of Microorganisms. Prentice Hall, Englewood Cliffs.

7 Caugant DA, BR Levin and RK Selander. 1981. Genetic diversity and temporal variation in the *E. coli* population of a human host. Genetics 98: 467–490.

8 Cohan FM, MS Roberts and EC King. 1991. The potential for genetic exchange by transformation within a natural population of *Bacillus subtilis*. Evolution 45: 1383–1421.

9 Cracraft J. 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: Speciation and its Consequences (Otte D and J Endler, eds), pp 28–59, Sinauer Associates, Sunderland.

10 de Bruijn FJ. 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergenic consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium melilloti* and other soil bacteria. Appl Environ Microbiol 58: 2180–2187.

11 Dobzhansky T. 1937. Genetics and the Origin of Species. Columbia University Press, New York.

12 Duncan KE, N Ferguson, K Kimura, X Zhou and CA Istock. 1994. Fine-scale genetic and phenotypic structure in natural populations of *Bacillus subtilis* and *Bacillus licheniformis*: implications for bacterial evolution and speciation. Evolution 48: 2002–2025.

13 Duncan KE, CA Istock, JB Graham and N Ferguson. 1989. Genetic exchange between *Bacillus subtilis* and *Bacillus licheniformis*: variable hybrid stability and the nature of bacterial species. Evolution 43: 1585–1609.

14 Dykhuizen D and L Green. 1991. Recombination in *Escherichia coli* and the nature of bacterial species. J Bacteriol 173: 7257–17268.

15 Eldredge N. 1985. Unfinished Synthesis: Biological Hierarchies and Modern Evolutionary Thought. Oxford University Press, Oxford.

16 Ferrari E, AS Jarnigan and BF Schmidt. 1993. In: *Bacillus subtilis* and Other Gram-Positive Bacteria (Sonenshein AL, JA Hoch and R Losick, eds), pp 917–937, Am Soc Microbiology, Washington, DC.

17 Fleischmann RD, MD Adams, O White, RA Clayton, EF Kirkness, AR Kerlavage, CJ Bult, J Tomb, BA Dougherty, JM Merrick, K McKenney, G Sutton, W Fitshugh, C Fields, JD Gocayne, J Scott, R Shirley, L Liu, A Glodek, JM Kelley, JF Weidmann, CA Phillips, T Spriggs, E Hedblom, MD Cotton, TR Utterback, MC Hanna, DT Nguyen, DM Saudek, RC Brandon, LD Fine, JL Fritchman, JC Fuhrmann, NMS Geoghagen, CL Gnehm, LA McDonald, KV Small, CM Fraser, HO Smith and JC Venter. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269: 496–512.

18 Gould SJ. 1980. Is a new and general theory of evolution emerging? Paleobiology 6: 119–130.

19 Grant V. 1981. Plant Speciation, 2nd edn. Columbia University Press, NY.

20 Howard DJ. 1988. The species problem. Evolution 42: 1111–1112.

21 Huxley J. 1942. Evolution: The Modern Synthesis. Allen and Unwin, London.

22 Istock CA, KE Duncan, N Ferguson and X Zhou. 1992. Sexuality in a natural population of bacteria—*Bacillus subtilis* challenges the clonal paradigm. Mol Ecol 1: 95–103.

23 Reference deleted in proof.

24 Istock CA. 1990. Genetic exchange and genetic stability in bacterial populations. In: Ecological Risks of Biotechnology (Ginzburg LR, ed), pp 123–147, Butterworths, London.

25 Judd AK, M Schneider, M Sadowsky and FJ de Bruijn. 1993. Use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. Appl Environ Microbiol 59: 1702–1708.

26 Lan R and PR Reeves. 1996. Gene transfer is a major factor in bacterial evolution. J Molec Biol Evol 13: 47–55.

27 Lorenz NA, BW Ardema and W Wackernagel. 1988. Highly efficient genetic transformation of *Bacillus subtilis* attached to sand grains. J Gen Microbiol 134: 107–112.

28 Lorenz MG and W Wackernagel. 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Rev 58: 563–602.

29 Lynch M and BG Milligan. 1994. Analysis of population genetic structure with RAPD markers. Mol Ecol 3: 91–99.

30 Maiden MCJ. 1993. Population genetics of a transformable bacterium: the influence of horizontal gene exchange on the biology of *Neisseria meningitidis*. FEMS Microbiol Lett 112: 243–250.

31 Maynard Smith J. 1995. Do bacteria have population genetics? In: Population Genetics of Bacteria (Baumberg S, JPW Young, EMH Wellington and JR Saunders, eds), pp 1–12, Cambridge University Press, Cambridge.

32 Maynard Smith J, NH Smith, M O'Rourke and BG Spratt. 1993. How clonal are bacteria? Proc Natl Acad Sci USA 90: 4384–4388.

33 Mayr E. 1988. Toward a New Philosophy of Biology. Harvard University Press, Cambridge.

34 Mayr E. 1991. One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Thought. Harvard University Press, Cambridge, MA, pp 26–34.

35 Mayr E. 1942. Systematics and the Origin of Species. Columbia University Press, New York.

36 Mayr E. 1963. Animal Species and Evolution. Harvard University Press, Cambridge.

37 Milkman R and A Stoltzfus. 1988. Molecular evolution of the *Escherichia coli* chromosome. II. Clonal segments. Genetics 120: 359–366.

38 Milkman R and MM Bridges. 1990. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics 126: 505–517.

39 Mischler BD and MJ Donoghue. 1982. Species concepts: a case for pluralism. System Zool 31: 491–503.

40 Nelson G. 1989. Species and taxa: systematics and evolution. In: Speciation and its Consequences (Otte D and J Endler, eds), pp 60–81, Sinauer Associates, Sunderland.

41 O'Donnell AG, M Goodfellow and DL Hawksworth. 1994. Theoretical and practical aspects of the quantification of biodiversity among microorganisms. Phil Trans R Soc Lond B 345: 65–73.

42 Otte D and J Endler. 1989. Speciation and its Consequences. Sinauer Associates, Sunderland.

43 Paterson HEH. 1985. The recognition species concept. In: Species and Speciation (Vrba ES, ed), pp 21–29, Transvaal Museum Monograph No. 4, Pretoria.

44 Pennington TH. 1994. Molecular systematics and traditional medical microbiologists—problems and solutions. J Med Microbiol 41: 371–373.

45 Ravin AW. 1960. The origin of bacterial species: genetic recombination and factors limiting it between bacterial populations. Bacteriol Rev 24: 201–220.

46 Roberts MS, LK Nakamura and FM Cohan. 1994. *Bacillus mojavensis* sp nov, distinguishable from *B. subtilis* by sexual isolation, divergence in DNA sequence and differences in fatty acid composition. Int J Syst Bacteriol 44: 256–264.

47 Selander RK, DA Caugant and TS Whittam. 1987. Genetic structure and variation in natural populations of *Escherichia coli*. In: Cellular and Molecular Biology, Vol 2: *Escherichia coli* and *Salmonella typhimurium* (Neidhardt FC, JL Ingraham, KB Low, B Megasanik, M Schaechter and HE Umbarger, eds), pp 1625–1648, American Society of Microbiology, Washington, DC.

48 Selander RK, DA Caugant, H Ochman, JM Musser, MN Glimour and TS Whittam. 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl Environ Microbiol 51: 873–884.

49 Sneath PHA. 1989. Analysis and interpretation of sequence data for bacterial systematics: the view of a numerical taxonomist. Syst Appl Microbiol 12: 15–31.

50 Sneath PHA and RR Sokal. 1973. Numerical Taxonomy. WH Freeman, San Francisco.

51 Sokal RR and TJ Crovello. 1970. The biological species concept: a critical evaluation. Am Nat 104: 104–127.

52 Souza V, L Eguiarte, G Avila, R Capello, C Gallardo, J Montoya and D Pinero. 1994. Genetic structure of *Rhizobium etli* biovar *phasioli* associated with wild and cultivated beans (*Phaseolus vulgaris* and *Phaseolus coccineus*) in Morelos, Mexico. Appl Environ Microbiol 60: 1260–1268.

53 Sullivan JT, HN Patrick, WL Lowther, DB Scott and CW Ronson. 1995. Nodulating strains of *Rhizobium loti* arise through chromosomal gene transfer in the environment. Proc Natl Acad Sci USA 92: 8985–8989.

54 Templeton AR. 1989. The meaning of species and speciation: a genetic perspective. In: Speciation and its Consequences (Otte D and J Endler, eds), pp 3–27, Sinauer Associates, Sunderland.

55 Vandamme P, B Pot, M Gillis, P De Vos, K Kersters and J Swings. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Rev 60: 407–438.

56 Vary P. 1992. Development of genetic engineering in *Bacillus megaterium*. In: Biology of the Bacilli: Applications to Industry (Doi RH and M McGloughlin, eds), pp 251–310, Butterworth-Heinemann, Boston.

57 Vrba ES. 1980. Macroevolutionary trends: new perspectives on the roles of adaptation and incidental effect. Science 221: 387–389.

58 Wayne LG, DJ Brenner, RR Colwell, PAD Grimont, O Kandler, MI Krichevsky, LH Moore, WEC Moore, RGE Murray, E Stackebrandt, MP Starr and HG Truper. 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. Intl J Syst Bact 37: 463–464.

59 Welsh J and M McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucl Acid Res 18: 7213–7218.

60 Welsh J, C Pretzman, D Postic, I Saint Girons, G Baranton and M McClelland. 1993. Genomic fingerprinting by arbitrarily primed polymerase chain reaction resolves *Borrelia burgdorferi* into three distinct phyletic groups. Int J Syst Bacteriol 42: 370–375.

61 Whittam TS and RA Wilson. 1988. Genetic relationships among pathogenic strains of avian *Escherichia coli*. Infect Immun 56: 2458–2466.

62 Williams JGK, AR Kubelik, KJ Livak, JA Rifelski and SV Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acid Res 18: 6531–6535.

63 Zuber P, MM Nakano and MA Marahiel. 1993. Peptide antibiotics. In: *Bacillus subtilis* and Other Gram-Positive Bacteria (Sonenshein AL, JA Hoch and R Losick, eds), pp 897–916, Am Soc Microbiology, Washington, DC.