Current Biology *17*, R373–R386, May 15, 2007 DOI 10.1016/j.cub.2007.03.032

A Systematics for Discovering the Fundamental Units of Bacterial Diversity

Review

Frederick M. Cohan and Elizabeth B. Perry

Bacterial systematists face unique challenges when trying to identify ecologically meaningful units of biological diversity. Whereas plant and animal systematists are guided by a theory-based concept of species, microbiologists have yet to agree upon a set of ecological and evolutionary properties that will serve to define a bacterial species. Advances in molecular techniques have given us a glimpse of the tremendous diversity present within the microbial world, but significant work remains to be done in order to understand the ecological and evolutionary dynamics that can account for the origin, maintenance, and distribution of that diversity. We have developed a conceptual framework that uses ecological and evolutionary theory to identify the DNA sequence clusters most likely corresponding to the fundamental units of bacterial diversity. Taking into account diverse models of bacterial evolution, we argue that bacterial systematics should seek to identify ecologically distinct groups with evidence of a history of coexistence, as based on interpretation of sequence clusters. This would establish a theory-based species unit that holds the dynamic properties broadly attributed to species outside of microbiology.

Introduction

> "*The sure and definite determination (of species of bacteria) requires so much time, so much acumen of eye and judgment, so much of perseverance and patience that there is hardly anything else so difficult*."
>
> — Otto F. Müller

For decades, the *International Journal of Systematic Bacteriology* featured on its cover this testimonial to the challenge of studying bacterial diversity. Indeed, compared to zoologists and botanists, bacterial systematists face unique difficulties when beholding a small phylogenetic group to identify its ecologically distinct populations and the different roles that the populations play within a community. Bacterial systematists are handicapped to some extent by the paucity of morphological differences that could help us demarcate closely related species. More profoundly, bacteriologists cannot predict with confidence what will be the traits determining the ecological differences between closely related species; this is because prokaryotes often adapt to new niches by acquiring genes from distant relatives through horizontal genetic transfer [1]. Consequently, the ecological differences among closely related bacteria are often invisible to systematists.

We can imagine how evolutionary biology might have fared if Charles Darwin had arrived on the Galapagos Islands with the handicaps of a bacterial systematist. Would he have noticed 13 distinct finch species, each with a bill morphology adapted for consuming a different set of seeds or insects? Or would these birds simply have appeared as a flock of related organisms — all much of a muchness of finchdom? The challenge of not *seeing* bacterial species — with the help of morphology — has been surmounted by systematists in several ways, but we shall see that bacterial systematics still suffers deeply for not readily sensing the ecological differences among close relatives.

Closely related bacterial species were first distinguished by careful analysis of phenotype (typically metabolism). In recent decades, systematists have adopted molecular approaches that have allowed standardized species demarcation and have ensured that each taxon is a monophyletic group — a true evolutionary group, including all and only the descendants of a given ancestor [2]. Ironically, just as these molecular techniques have promised a more efficient and confident systematics, they have also revealed a daunting task ahead. Surveys of gene sequence diversity from environmental DNA have indicated that fewer than 1% of bacterial species are cultivable at present [3]; so traditional, laboratory-based studies of pure-culture isolates have been blind to over 99% of bacterial diversity. Recently, cultivation-independent molecular methods have indirectly estimated that bacterial species may number in the millions or even billions [4,5]. This massive expansion of the scope of systematics, following on the heels of molecular promise, may strike us as a Sisyphusian curse. However, we will show how molecular and genomic approaches, when combined with advances in ecological and evolutionary theory, can bring us critical steps forward in the venture to fully characterize our planet's biological diversity.

The key to molecular discovery of biodiversity is to find organisms that fall into highly distinct sequence clusters for a given gene or set of genes. Because such clusters have each had a long history of separate evolution, they are likely to have evolved unique adaptations shared by the entire cluster. Systematists have applied this sequence-based approach to discover prokaryote diversity at all levels — from the urkingdoms, such as the Archaea, to species, and to even lower levels of diversity [6,7]. By identifying and then characterizing ever smaller phylogenetic groups, each with its unique history and adaptations, systematists have reached a fuller understanding of the ways that bacteria can make a living.

Department of Biology, Wesleyan University, Middletown, Connecticut 06459-0170, USA.
Email: fcohan@welseyan.edu

But how far must we delve up the tree of life, identifying smaller and smaller clades, before we have fully characterized ecological diversity within the bacterial world? A comprehensive study of any type of biological diversity would be complicated beyond feasibility if nearly every individual organism were ecologically unique. Fortunately, organisms from all walks of life — including bacteria, fungi, plants, and animals — and from every known community, appear to fall into discrete clusters of ecologically interchangeable individuals [8,9]. We will argue that systematics should seek to recognize and characterize all these irreducible, ecologically distinct groups within a clade or within a community, as these are the fundamental units playing unique roles in community assembly, ecosystem function and biotic interactions [10,11]. We will explain that this is not the present aim of bacterial systematics, and that the recognized 'species' of bacterial systematics frequently contain a diversity of populations that are distinct in their biochemistry, physiology, genome content and ecology; classifying an unknown organism to its species thus tells us only vaguely about the organism's way of life. Here we propose a systematics that demarcates and names the fundamental, ecologically distinct groups within the bacteria. This approach aims to satisfy Hutchinson's central mission of systematics — to ensure that the taxonomic name of an organism can inform us precisely about the organism's ecological and physiological properties [12].

Theory-Free Approaches to Bacterial Systematics

Species demarcation in bacteria has been historically handicapped by lacking a theory-based conceptual framework. Systematists have yet to agree upon a set of ecological and evolutionary properties that can be expected for the set of organisms within a bacterial species. Instead, bacterial species have been demarcated *empirically* as clusters of similar organisms. Bacterial species were first demarcated as phenotypic clusters based largely on metabolic capabilities [13]. In the 1970s, bacterial systematics began incorporating molecular methods to help distinguish closely related species. Beginning with whole-genome comparisons via DNA–DNA hybridization, systematists established molecular criteria that would correspond to the species groups that had already been defined by their metabolic characteristics. Through whole-genome hybridization, members of different named species were usually found to share less than 70% of their genome content, while members of the same species share greater than 70% of their genome content [14]. In 1987, this molecular cutoff became part of the canon of species demarcation [15].

Systematists have recently sought to replace the DNA hybridization standard of divergence with criteria based on sequence divergence of homologous genes [2]. One principal advantage to a sequence-based criterion is that any newly discovered organism can be compared, *in silico*, to every existing sequence in a growing data base for 16S ribosomal (r)RNA sequences [16]. Stackebrandt and Goebbel [17] found that, if two strains are at least 2.5% divergent in 16S rRNA, they are sure to fall into different species on the basis of DNA–DNA hybridization (although the converse is not necessarily true). More recently, a 16S divergence level of ~1% has been deemed sufficient to consider strains as sufficiently divergent to be in different species [18]. Similarly, Konstantinidis and Tiedje [19] have found that a genome-wide average nucleotide identity of at least 94% in homologous protein-coding genes is typical for members of the species recognized by bacterial systematics. Efforts are now underway to make possible a universal classification based on protein-coding sequences [20–22].

Another empirical approach is to find molecular criteria that yield clear clusters of closely related organisms, without the constraint that the clusters need to coincide with existing species. For example, Hanage *et al.* [23] have shown that phylogeny based on a concatenation of several genes can yield sequence clusters that are robust with respect to recombination, even within bacterial groups with relatively high recombination rates. Also, studies of genomic content may have the potential to help identify clusters of diversity.

We must note that these molecular and genomic approaches, however promising, are not designed to infuse a *theory* of species into systematics. They are merely adding new empirical criteria for dividing organisms into clusters with little attempt to correlate clusters with the fundamental, ecologically distinct populations within a natural community.

What is wrong with demarcating bacterial species without a theory-based concept of species? To illustrate the importance of theory, we turn to the systematics of animals and plants, for which a robust theory of species was developed long ago [24]. While species concepts for macrobes are by no means without dispute, all modern species concepts embrace the following dynamic attributes for species [25]: a species is a cohesive group (there are forces that limit genetic diversity within a species); a species is monophyletic (invented only once); different species are irreversibly separate; and different species are ecologically distinct (allowing them to coexist into the future).

In the case of animals and plants, there is one quintessential property that determines membership within a given species: successful interbreeding in nature. Interbreeding acts as a force of genetic and phenotypic cohesion among the members of an animal or plant species, and loss of the ability to interbreed allows two species to diverge phenotypically and to follow irreversibly separate evolutionary paths [26]. Different macrobial species, so defined, are typically distinct in morphology, behavior, physiology and ecology; organisms of the same species are expected to be functionally interchangeable. Thus, zoologists and botanists benefit from a systematics that satisfies Hutchinson's [12] fundamental mission of systematics — classifying an organism to an animal or plant species yields detailed and specific information about the organisms' physiology, biochemistry, and ecology.

The Enormous Diversity within Named Bacterial Species

Regrettably, lacking a concept of species, bacterial systematics has failed to identify taxa that might satisfy this mission. That is, a typical named bacterial

species contains huge diversity at all levels of analysis; so species identification does not provide specific ecological information about any of its members. Even though bacterial species were originally demarcated as phenotypic (usually metabolic) clusters, named bacterial species hold an enormous amount of phenotypic diversity [27–29]. Species also show a significant amount of genomic diversity. DNA–DNA hybridization studies have demonstrated that members of a named species frequently share only 80–90% of their genes [30]. These results have been more recently corroborated by physical mapping of genomes [31,32], and by genome sequence comparisons [33–36]. Even for genes shared across an entire named species, there exists a great deal of sequence variation within the species. At the 16S rRNA locus, sequence diversity within a recognized bacterial species is frequently at 1%. This is equal to the level of divergence typically found between orders of mammals at the homologous nuclear gene 18S rRNA [37]. When we correct for the faster rate of molecular evolution in bacteria, we can estimate the time of divergence among conspecific bacteria to be about five times greater than that for eukaryotic species [38].

Recent ecological studies show that a named bacterial species is typically an assemblage of closely related but ecologically distinct populations [39–41]. For example, a species of free-living heterotrophs may contain numerous sequence clusters that differ in the carbon sources they can utilize, as seen in the aquatic *Shewanella putrefaciens* [42]. Clusters within a species of free-living soil heterotrophs may differ in the solar radiation (and covarying parameters) to which they are adapted, as seen in *Bacillus simplex* [43] and in *B. licheniformis* [44]. Very closely related phototrophs may partition resources by light quantity and spectral quality, as seen in *Synechococcus* [10,44], and in the mineral nutrient resources they can utilize or store, as seen in *Prochlorococcus marinus* [45]. Within a pathogen species, populations can differ in their host ranges [41], their target tissues [46], or in the mode of transmission [47].

Even though fine-scale ecological differences are frequently overlooked by modern systematics, we should note that these differences within named species are extremely interesting from an evolutionary point of view. These differences tell us about the kinds of ecological adaptations that can accrue over very short time periods, and that can foster coexistence among the most closely related of populations. We expect that finding the basis of coexistence among closest relatives will be a rewarding challenge for the next generation of bacterial community ecologists and systematists.

There is one realm of bacteriology where systematics has been scrupulous about naming every ecologically distinct sequence cluster — medical microbiology. When one bacterial sequence cluster can kill us, but a close relative cannot, it becomes worth our while to put this distinction into our taxonomy [48]. This is seen in the naming of *Bacillus anthracis* (causing anthrax) as distinct from *B. cereus* [49], and the naming of *Yersinia pestis* (causing plague) as distinct from *Y. pseudotuberculosis*.

We would prefer that *all* bacteria be as aptly demarcated into ecologically distinct groups as are the most virulent agents of human death. We believe an understanding of bacterial population dynamics can help us identify these groups possessing the fundamental attributes of species.

The Peculiarities of Bacterial Population Dynamics

In contrast to most animals and plants, prokaryotes reproduce clonally, and the genetic exchange among prokaryotes proceeds by various parasexual processes that are not tied to reproduction. When genetic exchange occurs in bacteria, a short segment from a 'donor' individual replaces the homologous segment in a 'recipient' individual. Recombination in bacteria occurs at much lower frequencies than in most animals and plants, with the recombination rate ranging from less than the mutation rate (per gene segment) to about ten times the mutation rate in the most frequently recombining organisms [50–52]. Only one bacterial species, *Helicobacter pylori*, is known to undergo recombination many orders of magnitude more frequently than mutation [53].

The rarity of recombination is expected to simplify the formation of new species of bacteria. In order for adaptive divergence to occur between highly sexual animal or plant populations, gene flow must be restricted by some geographical or ecological barrier [26]. In contrast, recombination in bacterial populations is not frequent enough to hinder adaptive divergence [54]. Because of the extremely low rates of recombination in prokaryotes, the influx of genes from other populations is unlikely to disrupt the integrity of each population's specific adaptations. Evolution of sexual isolation is thus not a necessary step toward permanent divergence between ecologically distinct bacterial populations [54], and sympatric speciation may be a common occurrence.

The typical rarity of recombination in prokaryotes is also expected to lead to a recurrent purging of diversity within an ecological population. When natural selection increases the frequency of an adaptive mutation at a particular gene locus within a population, there is little opportunity for separation of the adaptive mutation from the genome in which it originated. Thus, the entire genome of the adaptive mutant increases in frequency and diversity is purged at nearly all loci. This kind of diversity-purging event, called 'periodic selection' [55,56], can be an important agent limiting the sequence diversity within a typical prokaryotic population (excluding the frequently recombining *H. pylori* [53,57]). In contrast, in the highly sexual animals and plants, each bout of natural selection has only very limited ability to purge diversity — generally affecting only the chromosomal region near the locus under selection.

What then are the forces limiting genetic diversity within an animal or plant species, and to what extent do they apply to bacteria? Genetic drift is the primary force limiting sequence diversity within a highly sexual population of animals or plants [58]. Random elimination of lineages limits sequence diversity. While genetic drift applies to populations of all sizes, the effects of genetic drift are quite weak in populations much

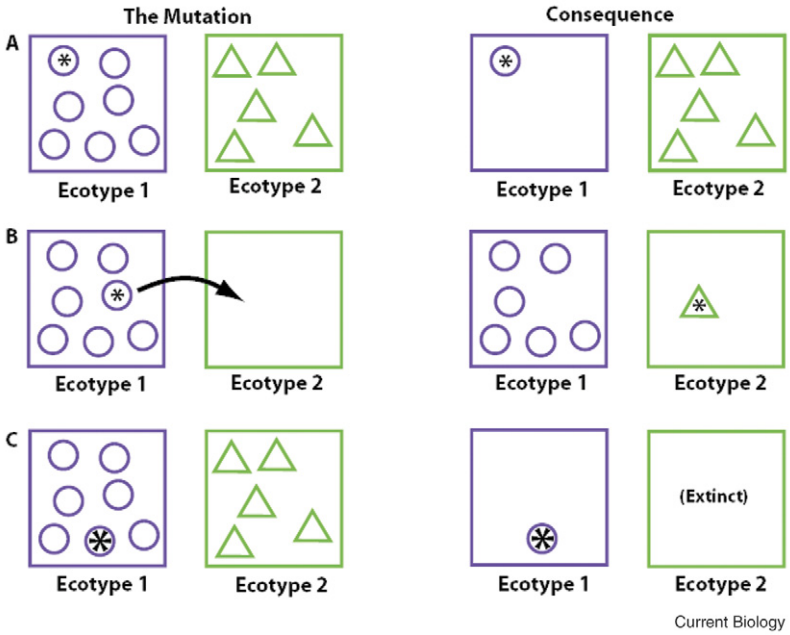

Figure 1. Three classes of mutation and recombination events that determine ecotype diversity in bacteria.

Circles represent different genotypes, and asterisks represent adaptive mutations. (A) Periodic selection mutations. These improve the fitness of an individual such that the mutant and its descendants out-compete all other cells within the ecological niche (ecotype); these mutations do not affect the diversity within other ecotypes because ecological differences prevent direct competition. Periodic selection leads to the distinctness of ecotypes by purging the divergence within but not between ecotypes. (B) Ecotype-formation mutations. Here a mutation or recombination event allows the cell to occupy a new ecological niche, founding a new ecotype. The ecotype-formation mutant, as well as its descendants, can now escape periodic selection events from its former ecotype. (C) Speciation-quashing mutation. Even if two ecotypes have sustained a history of separate periodic selection events, an extraordinarily adaptive mutant genotype may out-compete to extinction another ecotype. Competitive extinction of another ecotype (Ecotype 2) is possible only if all of Ecotype 2's resources are also utilized by Ecotype 1. Speciation-quashing mutations are expected in the Nano-niche model [64]. (Used with permission from Landes Publishers.)

larger than 10^9 [44]. In large bacterial populations, if drift were the only factor limiting the sequence diversity, we would expect to see saturation in variation at every neutral nucleotide site [54]. When we do not see evidence of such saturation in large populations, we can conclude that a factor other than drift (such as periodic selection) is operating to limit sequence diversity.

In some bacteria, however, population sizes may be small enough for genetic drift to be an important factor. The power of genetic drift may be especially important when considering obligate pathogens and commensals, which must be transmitted from host to host. In cases where very few bacterial cells colonize an individual host, and very few cells successfully leave an individual host, the effective population size of the bacteria is reduced to nearly that of the host species.

The Ecotype Model of Bacterial Species

Taking into account these features of bacterial population biology, we consider first a model of bacterial species based on the 'ecotype'. An ecotype is defined here as a group of bacteria that are ecologically similar to one another, so similar that genetic diversity within the ecotype is limited by a cohesive force, either periodic selection or genetic drift, or both. The present concept of ecotype is more general than in our previous work, where periodic selection was viewed as the only cohesive force [38,54].

In this model, diversity within an ecotype is ephemeral, persisting only until the next periodic selection event, when diversity is next crushed to near zero at all loci, or until purged by genetic drift (Figure 1A). What, then, is the source of *permanent* divergence among closely related bacteria? Divergence can become permanent when a mutation (or recombination event) places the organism into a new ecological niche and founds a new ecotype. Because the new ecotype is ecologically distinct from the parental ecotype, periodic selection events in the parental ecotype cannot extinguish the founding organism and its descendants (Figure 1B). The new ecotype thus escapes the periodic selection of the parental ecotype, and the two new ecotypes are free to diverge indefinitely.

Each ecotype so defined bears all the dynamic properties attributed to species [25]. Each ecotype is a cohesive group whose diversity is limited by periodic selection and/or genetic drift. Different ecotypes are irreversibly separate because they are out of range of one another's periodic selection events, and because recombination is too rare to prevent their adaptive divergence [59,60]; they are by definition ecologically distinct, which allows them to coexist into the future. Finally, such ecotypes are monophyletic groups because they are founded by a single individual (except in the Recurrent Niche Invasion model, below).

A variant of the ecotype model is particularly promising for its utility in systematics. This is the 'Stable

Ecotype' model, where ecotypes are created and extinguished at a very low rate, and during its long lifetime an ecotype is recurrently purged of its diversity by periodic selection events [2,9,61] (Figure 2A). Most such ecotypes (of sufficient age) should be distinguishable from other ecotypes as separate sequence clusters under most rates of recombination encountered in bacteria [7,60,62]. Ideally, we may identify ecotypes as DNA sequence clusters, provided that the Stable Ecotype model applies. However, the Stable Ecotype model does not take into account several factors that may be important in some lineages, for example, geographic isolation, adaptation through plasmid transfer and loss, genetic drift and very rapid speciation. We will examine each of these complicating factors in turn, but we will first consider whether the basic Stable Ecotype model appears correct.

Guttman and Dykhuizen provided the first evidence for the diversity-purging effects of selection in natural populations of bacteria [63]. They found a chromosomal region (near *gapA)* in *Escherichia coli* that is nearly homogeneous, even while the rest of the chromosome shows substantial heterogeneity and forms discrete sequence clusters. They originally interpreted this pattern as evidence that an adaptive mutation in the *gapA* region swept through the "population" of *E. coli*, but that recombination rescued existing variation in all areas outside of this chromosomal region. Such an interpretation could lead to the generalization that in bacteria, as in the highly sexual animals, each adaptive mutation purges diversity only in the chromosomal region near the adaptive mutation.

We disagree with this interpretation, because we have previously shown recombination to be rare enough to support periodic selection in nearly all bacteria, provided that the rate of recombination is within an order of magnitude of mutation (per gene) [64]; this certainly includes *E. coli*, as well as all bacteria studied for recombination rates, with the exception of *Helicobacter pylori* [53]. Also, this interpretation begs the question of the origin of the sequence clusters within *E. coli*. If all of *E. coli* is one population through which one adaptive mutation can advance, why should there be multiple, apparently coexisting sequence clusters?

We have previously proposed an alternative model for the small chromosomal regions of homogeneity that one encounters in genomic comparisons, as well as in the original *gapA* survey. In our 'Adapt Globally Act Locally' model [65,66], we interpret the various sequence clusters of *E. coli* as distinct ecotypes that undergo their own periodic selection events (Figure 3). Certain mutations arise, however, that are adaptive in many different ecotypes, and they can be passed to them through recombination. The adaptive mutation can then trigger a periodic selection event within each ecotype into which it is transferred. A series of periodic selection events in different ecotypes results in homogenization of the small chromosome region that is transferred between ecotypes, but the sequence clusters corresponding to the different ecotypes are otherwise left intact.

As many closely related genomes are compared, we believe that the Adapt Globally model will be the most compelling explanation for small chromosomal regions of homogeneity. For example, we have found a clade of hot spring *Synechococcus* to be ecologically diverse, with different ecotypes (revealed as sequence clusters) living at different temperatures and in different photic zones [10,44]. However, this clade is homogeneous within and between ecotypes in the nitrogen-fixing region of their chromosomes [67]. Clearly, the homogeneity in this region is not due to one adaptive mutant outcompeting cells in the entire, ecologically heterogeneous *Synechococcus* clade. Rather, an adaptive mutant was most likely able to extinguish cells within its ecological niche, followed by the spread of the adaptive mutation to other ecotypes through recombination, with subsequent periodic selection within each recipient ecotype.

We next address other arguments that periodic selection may not be a significant force in bacterial evolution. Roumagnac *et al.* [68] noted that existing sequence variation within *Salmonella enterica* serovar Typhi appears neutral with respect to fitness — nonsynonymous and synonymous substitution rates are similar — and claimed that this result was inconsistent with a history of periodic selection. However, periodic selection is expected simply to reduce the sequence variation to near zero; so the subsequent accumulation of sequence variation after periodic selection does not constitute a test of periodic selection.

Also, Roumagnac *et al.* [68] have noted that a particular adaptive mutant clone of *S. enterica* serovar Typhi (H58), conferring resistance to the antibiotic nalidixic acid, has rapidly increased in frequency but does not appear to be headed toward 100% frequency; they have argued that this failure to reach 100% is evidence against periodic selection. However, when an adaptive mutation does not reach 100% within a clade, it is most likely because the clade contains multiple ecotypes [69,70]. In the case of H58, the inability to reach 100% within the *S. enterica* serovar Typhi clade, either in local geographic regions or globally, provides evidence of at least two ecotypes within serovar Typhi, perhaps one adapted to the niche of humans treated with nalidixic acid, and one or more ecotypes where nalidixic acid is not part of the environment.

We are aware of only one taxon where periodic selection is known not to effect a genome-wide purging of diversity. This is the extremely frequently recombining *H. pylori* [53], an obligate pathogen of humans that shows nearly the same pattern of geographical diversity as its human host [57]; all other bacterial groups studied appear clonal enough to support periodic selection [64]. What is not clear is the extent to which genetic drift might also limit the diversity within a given ecotype. This is an issue that will probably not be resolved by analysis of sequence diversity, because periodic selection and drift can both, separately or together, limit sequence diversity within an ecotype. In order to assess the relative contribution of these forces, we must instead consider the population size and recombination rates of the organism in its natural environment. Genetic drift can be ruled out as a major factor in populations larger than 10^9, and periodic selection is unlikely to be a major cohesive force in populations with extremely high recombination rates.

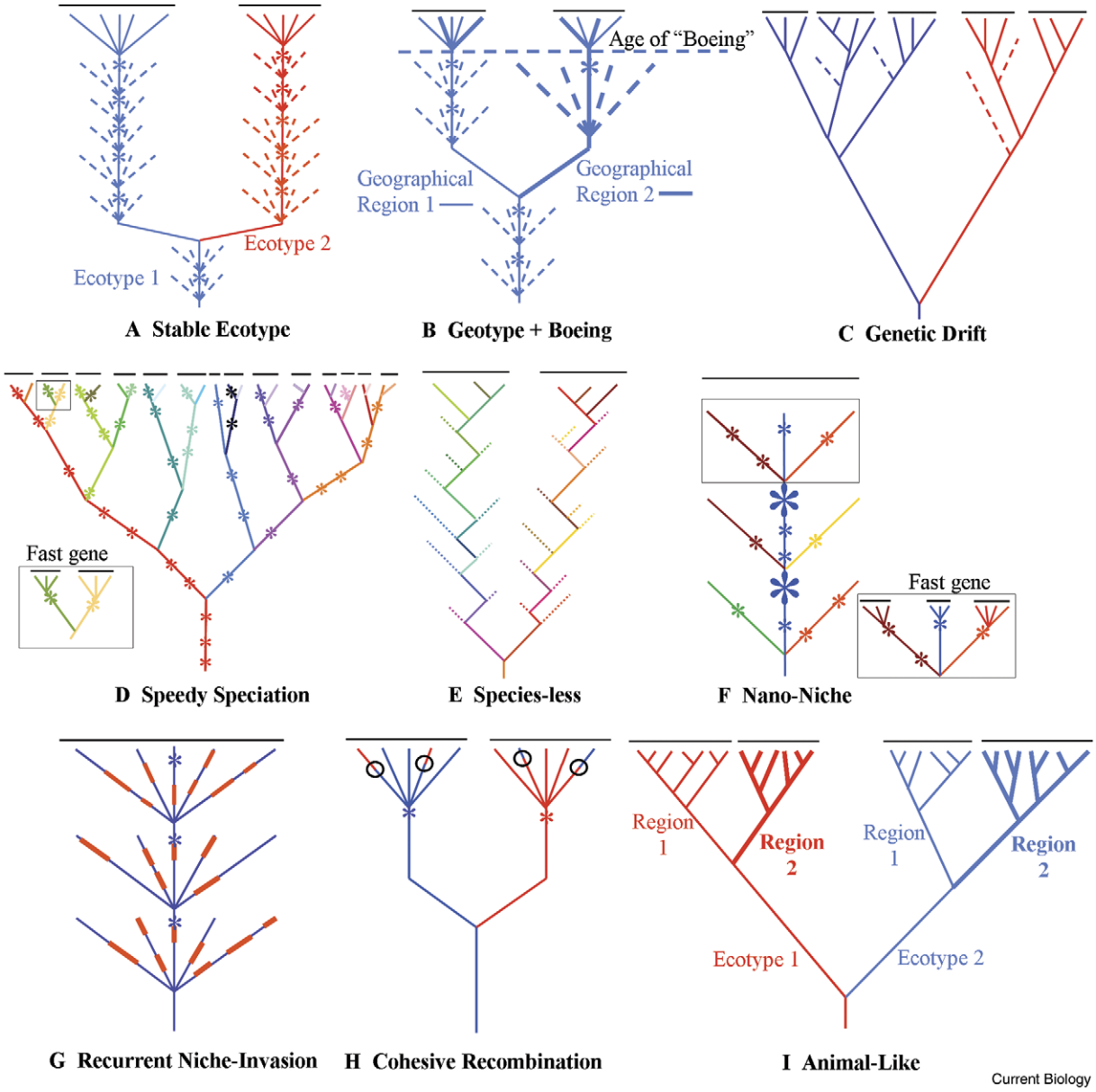

Figure 2. A diversity of evolutionary models for the relationship between ecologically distinct populations and DNA sequence clusters. Ecotypes are represented by different colors; periodic selection events are indicated by asterisks; extinct lineages are represented by dashed lines; clades that may be perceived as sequence clusters are marked by a horizontal black line at the top of the phylogeny. (A) The Stable Ecotype model. This model is marked by a much higher rate of periodic selection than ecotype formation, such that each ecotype endures many periodic selection events during its lifetime. The Stable Ecotype model generally yields a one:one correspondence between ecotypes and sequence clusters. (B) The Geotype-plus-Boeing model. A history of geographic isolation, followed by recent human-aided transport (during the 'age of Boeing'), can lead to multiple sympatric sequence clusters within a single ecotype. Lineages from different regions are represented by different line thicknesses. (C) Genetic Drift model. Genetic drift can also produce multiple clusters within each ecotype, provided that effective population sizes are relatively small. (D) The Speedy Speciation model. Here new ecotypes are formed at a high rate, such that many young ecotypes have not diverged sufficiently enough to be distinguished as separate sequence clusters. For example, the olive and tan ecotypes within the rectangle form a single sequence cluster when analyzed using a gene with a slow evolutionary clock. However, when analyzed by a more rapidly evolving molecular marker (the 'fast gene' indicated), a one:one correspondence between ecotypes and sequence clusters may be seen. (E) The Species-less model. Here, invention and extinction of ecotypes occur very frequently. Diversity within an ecotype may be constrained primarily by the short amount of time between the ecotype's founding by a single mutant (or recombinant) to the time the population goes extinct. (F) The Nano-niche model. Here an ecotype contains a set of subtly ecologically distinct subpopulations that may use the same set of resources but in different proportions. While each of the subpopulations may undergo its own periodic selection events for some time (small asterisks), eventually a particularly adaptive mutation (large asterisk) extinguishes all but one of the subpopulations and purges diversity throughout the whole ecotype. While a slowly evolving gene might not give enough resolution to distinguish the subpopulations, a faster gene might (inset box). (G) Recurrent Niche Invasion model. Here a lineage may move, frequently and recurrently, from one population to another, usually by acquisition and loss of niche-determining plasmids. (H) Cohesive Recombination model. With recurrent recombination between ecotypes, distinct clusters may never be seen in genes that are not niche-determining.

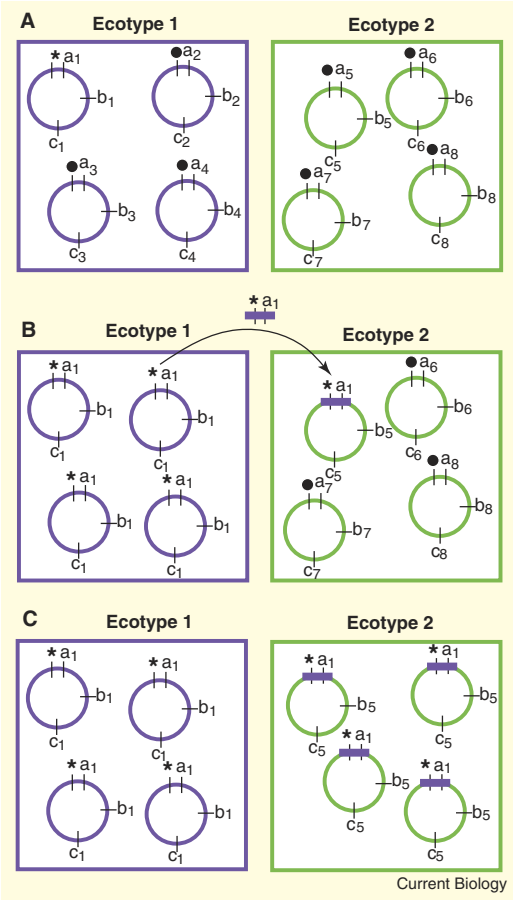

Figure 3. The Adapt Globally Act Locally model.

Adaptive mutants compete only with cells within their ecotype, but the mutation may be adaptive in many ecotypes and can be transferred through recombination. (A) An adaptive mutation (asterisk) occurs in Ecotype 1. Periodic selection purges diversity (black circles) within the ecotype. (B) The adaptive allele is transferred to a cell in Ecotype 2 via recombination. (C) Because the mutation is adaptive in this ecological niche as well, a periodic selection event occurs within Ecotype 2 [65]. (Used with permission from the Society of Systematic Biologists.)

Systematics and the Diversity of Models of Bacterial Evolution

We next consider alternatives to the Stable Ecotype model, where the correspondence between ecotypes and sequence clusters is not expected to be 1:1. In some alternative models of bacterial evolution, one ecotype may contain multiple sequence clusters. Consider first a model in which ecotypes are long-lived, as in the Stable Ecotype model, but there is only rare migration among the geographic regions of the ecotype. Ecologically identical populations in different regions can thus diverge into different sequence clusters. Papke *et al.* [71] have coined the term 'geotype' to refer to closely related populations from different geographic regions that have diverged as a result of their geographic isolation and not ecological differences.

How likely is the formation of geotypes in the bacterial world? For nearly a century, the theory of bacterial biogeography was dominated by the idea that "everything is everywhere, and the environment selects" [72]. That is, what determines the presence of a species at a location is not its ability to get there, but only its ability to thrive once it arrives [73,74]. However, recent studies suggest that at least some bacteria are severely restricted in their migration. Some extremophiles show only rare dispersal across uninhabitable, mesic habitats to other favorable, extreme locales [71,75,76], and rare migration is not limited to extremophiles [77]. Some environments are especially prohibitive of dispersal [38], for example, the highly viscous substrate of deep-rock bacteria and the perennially frozen lakes of Antarctica. Many pathogens and commensals can be only as mobile as their hosts.

A history of geographic separation among ecologically interchangeable populations can lead to difficulties for sequence-based taxonomy. In a model we call the 'Geotype-plus-Boeing' model [2,9,61], geographically isolated populations of the same ecotype could diverge into separate sequence clusters in the time before rapid human transport; then, in recent decades (or centuries), human shipping and travel could carry all the endemic geotypes within a single ecotype into each region of the world (Figure 2B). In this transitional era when air travel (and even transoceanic sea travel) is still relatively new, we may see multiple sequence clusters (the pre-'Boeing' geotypes) within one ecotype at one place, as seen in the plague bacterium *Yersinia pestis* [78] and in the tuberculosis agent *Mycobacterium tuberculosis* [79]. Therefore, we cannot conclude from sequence clustering alone that two sympatric clusters are separate ecotypes.

Genetic drift can also yield subclusters of closely related organisms of the same ecotype, especially in pathogens and endosymbionts, where effective population size can be severely limited. Thus, under either the Geotype-plus-Boeing or the Drift models (Figure 2C), the relationship between ecotypes and sequence clusters can be one:many.

We next consider models yielding a many:1 correspondence between ecotypes and sequence clusters, beginning with the 'Speedy Speciation' model, where new ecotypes form at a rapid rate, and new ecotypes are expected to coexist and diverge from one another into the indefinite future (Figure 2D). While the numerous new ecotypes are expected to form their own sequence clusters *eventually*, at any one moment a large number of newly formed ecotypes may be invisible to detection as sequence clusters for a given molecular marker. In these cases of rapid speciation, it may be important to use more rapidly evolving molecular markers to detect the most recently formed ecotypes, such as variable number tandem repeated (VNTR) sequences [80]. Slowly evolving genes will not highlight newly evolved ecotypes if the rate of nucleotide substitution is much less than the rate of ecotype formation and periodic selection [81]. A Speedy Speciation

(I) The Animal-like model. Recombination is frequent enough that periodic selection does not limit sequence diversity within ecotypes. The effective population size is low enough that genetic drift can effectively limit diversity within ecotypes. Also, migration is infrequent enough that geographic regions have their own geotypes within each ecotype.

regime that generates a many:one relationship between ecotypes and sequence clusters, as seen through protein-coding gene diversity, might actually generate a one:one relationship when VNTR or insertion sequences are used to reveal more microdiverse clustering [70].

Under what circumstances might the Speedy Speciation model apply? Speciation may be most rapid when many empty ecological niches are available, following an evolutionary innovation [82], although it is not clear what evolutionary innovations might precipitate rapid speciation in bacteria. Rapid speciation in bacteria may also depend on access to horizontal genetic transfer (HGT), whereby bacteria can acquire new genes and operons, sometimes from extremely divergent bacteria [1]. HGT-facilitated speciation may be promoted by a bacterium's intrinsic ability to absorb a genetic shock — such as a metabolic imbalance caused by acquiring a novel biochemical pathway — without substantial ill effect [83] and to the bacterium's ability to evolve changes that ameliorate adaptive changes initially carrying some deleterious side-effect [84,85]. Also rapid speciation may be promoted by a bacterium's ability to independently tweak the expression of newly acquired genes, so as to maintain metabolic balance [86], as may be the case in *Pseudomonas* [87].

Geographic isolation may be a necessary step facilitating bacterial speciation, depending on the extent to which evolution of a new adaptation diminishes the population's previously existing adaptations. When there are intrinsic trade-offs between new and preexisting adaptations, a set of adaptive mutations can immediately produce a new population that can coexist with the parental population. However, when evolution of new adaptations does not diminish old adaptations, evolution of ecologically distinct, coexisting populations can occur only in geographic isolation, where each population loses the subset of adaptations not needed in its respective environment [88]. Thus, rapid evolution of new species in bacteria could depend on low dispersal and geographic isolation. Finally, a rapid increase in the number of ecotypes may also be due to a lower rate of extinction, but little is known about extinction in the bacterial world. Thus, for a number of reasons, some bacteria may engage in rapid speciation, perhaps too quickly to ever discern through clustering of DNA sequences.

We next consider a variant of the Speedy Speciation model, the 'Species-Less' model [2,9,60,61,89], where there is a high rate of both formation and extinction of ecotypes, most likely when environments change rapidly. In the Species-Less model, the diversity within an ecotype need not be constrained into the indefinite future by any force of cohesion; instead, diversity may be constrained principally by the short amount of time from the ecotype's founding from a single mutant (or recombinant) to the time the ecotype goes extinct (Figure 2E). Here a single sequence cluster might contain multiple very young ecotypes.

We next consider two models in which there is a rapid and frequent invasion of ecological niches, but the ecologically distinct populations are not expected to diverge indefinitely. Both models yield a many:one correspondence between ecologically distinct populations and sequence clusters. The 'Nano-niche' model postulates a great diversity of ephemeral habitats, for example, small particles in the marine water column in the case of *Vibrio splendidus* (M. Polz, personal communication) (Figure 2F). In this model, subgroups within one ecotype each become adapted in nuanced ways to the subtleties of their own habitats; they may even have their own separate periodic selection events. Nevertheless, it may be possible for one especially adaptive mutant to outcompete to extinction all the ecological diversity among the ecotype's subgroups (a 'speciation-quashing' event; Figure 1C) [64]. In this case, the various ecologically distinct subgroups within an ecotype are not irreversibly separate, and do not have a chance to diverge into separate sequence clusters.

We note that recombination may have the potential to foster an extended coexistence of these ecologically distinct subgroups. If the adaptive mutation arising in one subgroup can be transferred to other subgroups and confer adaptation there, as in the Adapt Globally Act Locally model [66], the first subgroup might fail to outcompete the other subgroups to extinction [65].

In the 'Recurrent Niche Invasion' model [61] (Figure 2G), members of each ecotype frequently and recurrently lose the adaptations of their present ecotype and acquire the adaptations of another. Recurrent niche invasion is most likely when the populations owe their ecological distinctness entirely to the facile gain or loss of a plasmid. For example, in *Bacillus thuringiensis*, some clades can host a number of alternative 'crystal toxin' plasmids, each adapted to killing a different order of insect. A lineage may recurrently move from one ecological niche to another by losing one crystal toxin plasmid and then acquiring another. If these reciprocal ecological conversions recur repeatedly, then the populations are not irreversibly separate lineages, and they may never appear as separate sequence clusters.

A complex relationship between ecotypes and sequence clusters may occur if recurrent recombination prevents irreversibly separate ecotypes from diverging into separate sequence clusters for most genes — the 'Cohesive Recombination' model (Figure 2H) [54,90]. However, recombination between ecotypes is expected to decrease over time, owing to a positive feedback between sexual isolation and sequence divergence among ecotypes [60,62].

Finally, we propose the 'Animal-like' model (Figure 2I), inspired by the population dynamics of *Helicobacter pylori* [53,57]. In this model, recombination is extremely frequent, such that periodic selection is not a cohesive force. Effective population sizes are low enough to allow genetic drift to maintain sequence diversity at a low level in the absence of periodic selection. Also, dispersal is rare enough such that there is a trace of migration history in the sequence record. In other words, the recombination, population size, and migration parameters are as expected for many animal species.

One consequence of the high recombination rates in the Animal-like model is that speciation would require recombination between nascent ecotypes to be

reduced by some form of sexual isolation, as is the case for animal speciation. Possible mechanisms of sexual isolation are: adoption of different restriction enzymes in different ecotypes [91]; non-sharing of vectors of recombination (phage or plasmids) across ecotypes [92]; DNA sequence divergence, which hinders recombination [93–95]; and most importantly, ecological differences preventing access to other populations' DNA [96].

Incorporating Ecology and Evolution into a Systematics of Ecotypes

Animal and plant systematists have the advantage that, for a given clade, the traits determining niche differences can be anticipated. Finch species are specialized by differences in the shapes and sizes of their bills; herbaceous species on the American prairie can specialize to different growth seasons through differences in their root architecture, and so on. Thus, animal and plant systematists can demarcate species by identifying clades that are homogeneous for the traits most likely to be niche specifying [97].

In the case of bacteria, horizontal genetic transfer makes the niche-defining differences between species very difficult, if not impossible, to predict. Indeed, close relatives can infect entirely different hosts [98,99] or can have profoundly different roles in mineral cycling [45]. Because ecologists will not know ahead of time the traits distinguishing the most closely related ecotypes, we suggest that identification of ecotypes should begin with a sequence-based approach to formulate hypotheses about putative ecotypes, followed by an ecological approach to confirm these hypotheses.

We begin by assuming the Stable Ecotype model, where ecotypes are in principle discoverable as sequence clusters, and later take into account that other models may apply instead. Even under the Stable Ecotype model, however, ecological interpretation of sequence-based phylogenies is not straightforward. Any sequence-based phylogeny is likely to contain a hierarchy of subclusters within clusters, and it is not clear which level of cluster corresponds to ecotypes [59]. We have therefore developed and tested a conceptual framework for identifying ecotypes from sequence clusters under the assumptions of the Stable Ecotype model [10,38,44,100].

This 'ecotype simulation' approach begins by characterizing the sequence diversity within a clade as the number of sequence clusters (or bins) present for different sequence-identity criteria, following Martin [101] and Acinas *et al.* [102] (Figure 4). The number of sequence clusters at a particular sequence identity level represents the number of lineages at some point in the past that have survived to the present [101]. The simulation estimates the rates of periodic selection and drift, the net rate of ecotype formation (taking into account ecotype extinction), and the number of ecotypes so as to yield the clade sequence diversity pattern of Figure 4 with maximum likelihood (Figure 5).

We have employed the ecotype simulation approach to examine three clades whose ecological diversity and habitats have been intensively studied [44], including isolates of *Bacillus* primarily from 'Evolution Canyon' III of the Negev Desert [103], sequences from uncultured members of the *Synechococcus* A-A′ clade from Yellowstone hot springs [9,10,104], and world-wide clinical and environmental isolates of *Legionella pneumophila* [100].

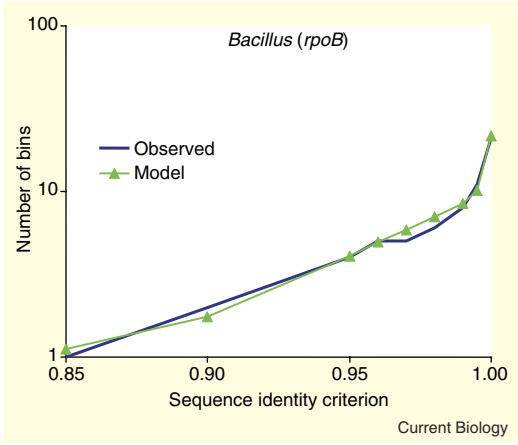

Figure 4. Clade sequence diversity, for the *B. subtilis–B. licheniformis* clade, for gene *rpoB*.

A set of 88 sequences from the clade was binned into clusters with different levels of minimum pairwise identity. The log-linear portion of the curve to the left of ~0.99 identity indicates a constant net rate of ecotype formation; the flair of diversity to the right of ~0.99 indicates the facile sequence diversity within ecotypes [102]. The ecotype simulation analysis yielded estimates of the parameters for periodic selection rate, net ecotype formation rate, rate of genetic drift, and the number of ecotypes, and the estimated quartet of parameter values generated the 'model' values shown [44].

Ecotype simulation estimated more ecotypes than the number of existing species and subspecies, for each clade analyzed, corroborating previous evidence that the demarcations of bacterial systematics frequently lump many ecologically distinct populations into a single species [9,40,41,43,105]. We extended the ecotype simulation analysis to identify all the individual ecotypes of a clade resolvable with DNA sequence data [38,44]. The rationale was first to quantify the sequence diversity within a given subclade, and then to determine the number of ecotypes that yields the subclade's sequence diversity pattern with maximum likelihood. We then demarcated putative ecotypes as the largest clades that were each consistent with a single ecotype.

In each clade we have analyzed, ecotype simulation demarcated some putative ecotypes that appear ecologically distinct in nature [10,38,44,100]. Most of these ecological distinctions were inferred from differences in the microhabitats where the organisms (or sequences) were most frequently isolated. In *Synechococcus*, microhabitat distribution suggested that very closely related putative ecotypes are specialized to different temperatures and photic zones [10,44]. In *Bacillus*, very closely related putative ecotypes were found to be adapted to different conditions of solar insolation (or co-varying factors) on different faces of the canyon [38,44] (Figure 6). In the case of *Legionella*, putative ecotypes differed in the species of amoebae they can infect [100].

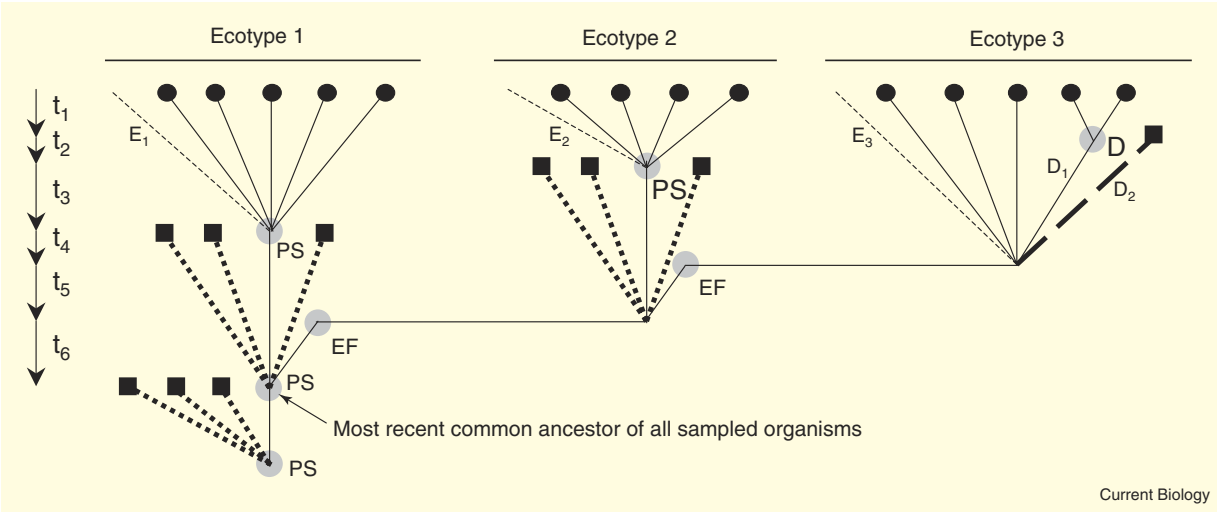

Figure 5. The ecotype simulation algorithm.

A computer algorithm simulates the evolutionary history of the ν organisms sampled from nature, under different quartets of values for the net rate of ecotype formation (EF), the rates of periodic selection (PS) and drift (D), and the number of ecotypes in the sample. In the coalescence approach taken, the algorithm considers only the lineages that are directly ancestral to the ν sampled organisms (circles). These focal lineages are represented by solid lines; the many contemporary lineages not sampled from each ecotype are indicated by light dashed lines (E_1, E_2, and E_3); the lineages extinguished by past periodic selection and drift are represented by bold, short-dashed lines and long-dashed lines, respectively, with each extinction represented by a square. The program begins with a 'backward' simulation that stochastically produces a phylogenetic representation of the history of the community, establishing nodes of coalescence of lineages (indicated by grey circles) and time between nodes (t_1, t_2, etc.); this phylogeny is then taken as a scaffold for the forward simulation. The purpose of the forward simulation is to produce nucleotide substitutions throughout the history of the clade, according to the phylogenetic scaffold. To begin a simulation, a set of ν contemporary organisms (representing the ν organisms sampled from nature) are distributed randomly among the n ecotypes. (In the case of this figure, $\nu = 14$ and $n = 3$.) Working backwards from the ν organisms in the present, the processes of ecotype formation, periodic selection, and drift occur stochastically in time according to their respective rates (Ω, σ, and d). The backwards phase of the simulation ends when all of the branches have coalesced into a single node; this represents the most recent common ancestor of all the sampled organisms. Then the forward simulation begins when a sequence is assigned to this most recent common ancestor. Nucleotide substitutions then occur stochastically, going forward in time, between each pair of nodes in the phylogeny derived from the backward simulation, according to the time between the events determining the nodes [44].

Ecotype simulation promises to be an effective way to identify the fundamental units of bacterial ecology and evolution. This approach has the advantage that it analyzes a particular clade's population dynamics to obtain the appropriate demarcation threshold for that clade. It is unlikely that the *universal* molecular thresholds routinely used in systematics [7,15,18], or indeed any universal molecular threshold, could demarcate these fundamental units [44].

Nevertheless, sequence-based hypotheses about putative ecotypes must take into account the great diversity of possible models of bacterial evolution — that sequence clusters, even if identified by ecotype simulation or a similar algorithm, do not always correspond to ecotypes as in the Stable Ecotype model. To this end, we suggest that ecotypes should be demarcated as the smallest groups that: (1) show a history of coexistence as separate, ecologically distinct lineages, as inferred from ecotype simulation (or an equivalent sequence-based approach); and (2) show a prognosis for future coexistence, as inferred from the ecological distinctness of the groups in nature [38,44].

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, different clusters could represent formerly isolated populations of the same ecotype that have recently been flown or shipped to the same locations. Alternatively, in cases where drift is likely to be an important force, an ecotype could contain multiple sequence clusters caused by genetic drift. Therefore, sequence clusters must be verified to be ecologically distinct before they can be declared ecotypes.

Why is ecological distinctness alone insufficient 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 physiology [106,107]. Clearly what we want to know goes beyond laboratory 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 thereby coexist into the future. Sequence data provide 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 long-standing in their coexistence, possibly owing to their ecological differences (alternatively to previous geographic separation) [61].

There is a second problem in identifying populations as different ecotypes when they are not yet separate

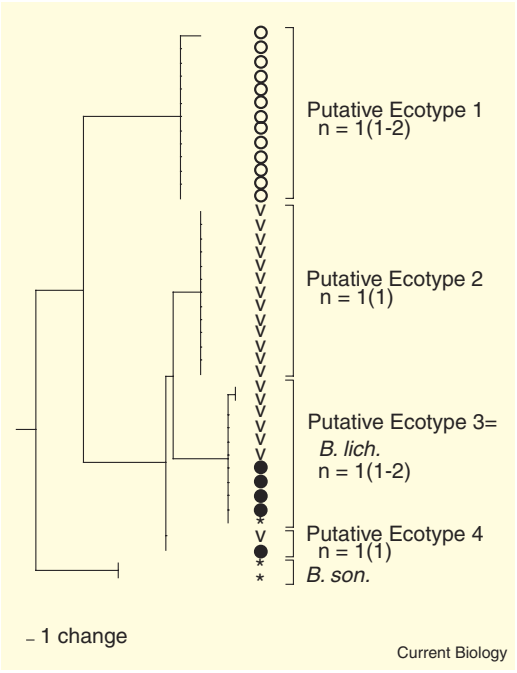

Figure 6. Phylogeny and ecotype demarcation of the *B. licheniformis* subclade of *Bacillus*.

The four putative ecotypes of this subclade differ significantly in their associations with the three major microhabitats of 'Evolution Canyon' in the Negev Desert. Microhabitat sources were the south-facing slope (○), the north-facing slope (●), and the canyon bottom (V); asterisks indicate strains isolated elsewhere. Ecotypes were also demarcated based on the more rapidly evolving *gyrA* gene, and what appeared to be one ecotype (Ecotype 2) for *rpoB* was further split into three more ecotypes (as illustrated schematically in Figure 2D) [44].

sequence clusters. The populations might not be irreversibly separate, a case most likely when populations owe their ecological distinctness entirely to the gain or loss of a plasmid (the Recurrent Niche Invasion model), or when ecological differences between populations are not sufficient to indefinitely evade one another's periodic selection events (Nano-niche model). However, if we demarcate ecotypes only when ecologically distinct groups form separate sequence clusters, we can be assured that the populations have had a history of divergence as separate lineages [38,61].

Finally, we must take into account models where ecologically distinct populations are frequently too new to be distinguishable as sequence clusters. We 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 might 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 (for example, *Bacillus anthracis* versus *B. cereus*) [80], but the ecological distinctness we observe (regarding virulence) is clearly relevant to the ways that the bacteria make a living in nature. When the ecological distinctness of such groups is not readily reversible (for example, with the gain or loss of a plasmid), it is reasonable to give a prognosis for the continued coexistence of these populations as separate lineages and to declare them ecotypes [61].

In summary, to accommodate models yielding a one:many correspondence between ecotypes and sequence clusters, systematists would need to confirm that putative ecotypes identified as sequence clusters are ecologically distinct from one another. To accommodate models with a many:1 correspondence, we will need to confirm that each putative ecotype is ecologically homogeneous within itself.

How will systematists determine whether putative ecotypes are ecologically distinct? We anticipate that in future applications of ecotype simulation, microbiologists will confirm the ecological distinctness of putative ecotypes through microhabitat distribution studies, as well as comparisons of genome content and analyses of genome-wide gene expression and comprehensive metabolic phenotype [29]. Much of the evidence will draw on the existing skills of taxonomists, who are trained in testing the capabilities of the growth of organisms with different resources and under different growth conditions. We recommend a new charge for taxonomists, to move from finding diagnostic phenotypic characters [108] to using the broadest diversity of techniques to assess ecological differences [61]. The challenge of an ecotype-based systematics will thus be no less demanding than the "acumen of eye and judgment" described by O. F. Müller two centuries ago, but as we discuss, the rewards should be worth the investment.

An Ecotype-Based Systematics

We have proposed a systematics for identifying ecotypes, the fundamental units of bacterial ecology and evolution. We recommend that these ecotypes be recognized also as the fundamental units of bacterial systematics [44], by issuing a name for every ecotype with a history of coexistence and a prognosis for future coexistence with other ecotypes. In the case when a legacy, named species is found to have multiple ecotypes, we recommend that each confirmed ecotype be given a trinomial name with an ecovar epithet, for example, *Legionella pneumophila* ecovar *pneumophila* for the ecotype containing the originally described Philadelphia strain [2,100]. We recommend that newly discovered clades be demarcated such that each confirmed ecotype is named as a separate species (with a binomial name). We believe that the fullness of ecological diversity within the bacterial world will be taken most seriously when each ecotype is given its own name [61].

By identifying taxa at the level of ecotypes, systematics will allow microbiologists to focus on groups most likely to differ in adaptations of physiology, genome content, and gene expression. For example, an ecotype-based systematics will help genomicists optimize their choices of organisms to be fully sequenced. Because comparative genomics can yield details of ecological divergence between organisms [45,67], it is important to choose organisms from different ecotypes for genome sequencing — then differences in gene content will have a greater chance of determining the niche. Ecotype-based systematics will help avoid

choosing organisms that have a high chance of being ecologically interchangeable. Likewise, in comparative physiology, we should choose model organisms from different ecotypes. For example, in the case of *Legionella pneumophila*, three strains have been chosen as models (Lens, Philadelphia and Paris), but Paris and Philadelphia appear to be from the same ecotype and indeed are very similar in their physiology [100,109].

In preparation for future epidemics, whether natural or the result of biowarfare, we should try to discover all the long-standing ecotype diversity within each named pathogenic species. We could then anticipate and prepare for future epidemics by characterizing the disease-causing properties of each ecotype.

Biotechnologists may also take advantage of an ecotype-based systematics. After discovering a strain with a valuable enzyme, one could then search for homologs in each ecotype closely related to the strain. This may allow discovery of similar enzymes with different substrates or with optima at different conditions.

An ecotype-based systematics would simplify the burden of industrial testing of bacterial strains for their safety and efficacy in agricultural applications. For example, for any named species that is heterogeneous for characteristics of safety concern — for example, secreted metabolites and persistence in the environment — the European Union requires that any new strain developed for release be tested for these characteristics of concern; however, individual strains from a species known to be homogeneous for these features need not be tested [110]. Thus, demarcating taxa as ecologically homogeneous units would obviate or at least lessen the burden of these tests.

An ecotype-based systematics will allow quantification of the ecological diversity within a community. Recent studies have demarcated and counted bacterial taxa in a community by binning DNA sequences for a given gene (often 16S rRNA) into sequence clusters and then attributing each sequence cluster to an operational taxonomic unit [111]. The operational taxonomic units typically used in these studies have no theoretical justification, but ecotype simulation offers a quantification of ecological diversity as the number of ecotypes present in a sample of sequences from a habitat.

Finally, a systematics of ecotypes will allow us to identify and characterize the ecologically distinct groups of bacteria, a critical step forward in our venture to understand the myriad ecological interactions within natural microbial communities.

Acknowledgments

We are grateful to Mark Achtman, Brendan Bohannan, William Hanage, Alex Koeppel, Danny Krizanc, Jeffrey Lawrence, Eviatar Nevo, Martin Polz, Johannes Sikorski, Alan Templeton, David Ward, and Stephen Wessels for many valuable discussions that have profoundly shaped our view of bacterial diversity. This research was funded by the FIBR program of NSF and by the Exobiology program of NASA.

References

1. Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002). Prokaryotic evolution in light of gene transfer. Mol. Biol. Evol. *19*, 2226–2238.
2. Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., *et al*. (2005). Opinion: Re-evaluating prokaryotic species. Nat. Rev. Microbiol. *3*, 733–739.
3. Giovannoni, S.J., and Stingl, U. (2005). Molecular diversity and ecology of microbial plankton. Nature *437*, 343–348.
4. Dykhuizen, D.E. (1998). Santa Rosalia revisited: why are there so many species of bacteria? Antonie Van Leeuwenhoek *73*, 25–33.
5. Gans, J., Wolinsky, M., and Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science *309*, 1387–1390.
6. Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science *276*, 734–740.
7. Palys, T., Nakamura, L.K., and 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.
8. Claridge, M.F., Dawah, H.A., and Wilson, M.R. (1997). Species: The Units of Biodiversity (London: Chapman & Hall).
9. Ward, D.M., and Cohan, F.M. (2005). Microbial diversity in hot spring cyanobacterial mats: pattern and prediction. In Geothermal Biology and Geochemistry in Yellowstone National Park, W.P. Inskeep and T. McDermott, eds. (Bozeman: Thermal Biology Institute), pp. 185–202.
10. Ward, D.M., Bateson, M.M., Ferris, M.J., Kühl, M., Wieland, A., Koeppel, A., and Cohan, F.M. (2006). Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. Phil. Trans. Roy. Soc. Ser. B *361*, 1997–2008.
11. Staley, J.T. (2006). The bacterial species dilemma and the genomic-phylogenetic species concept. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *361*, 1899–1909.
12. Hutchinson, G.E. (1968). When are species necessary? In Population Biology and Evolution, R.C. Lewontin, ed. (Syracuse: Syracuse University Press), pp. 177–186.
13. Jones, D., Sackin, M.J., and Sneath, P.H. (1972). A numerical taxonomic study of streptococci of serological group D. J. Gen. Microbiol. *72*, 439–450.
14. Johnson, J. (1973). Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria. Int. J. Syst. Bacteriol. *23*, 308–315.
15. Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., *et al*. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. *37*, 463–464.
16. Harris, J.K., Kelley, S.T., and Pace, N.R. (2004). New perspective on uncultured bacterial phylogenetic division OP11. Appl. Environ. Microbiol. *70*, 845–849.
17. Stackebrandt, E., and 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.
18. Stackebrandt, E., and Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. Microbiol. Today *33*, 152–155.
19. Konstantinidis, K.T., and Tiedje, J.M. (2005). Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. USA *102*, 2567–2572.
20. Konstantinidis, K.T., Ramette, A., and Tiedje, J.M. (2006). Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. Appl. Environ. Microbiol. *72*, 7286–7293.
21. Santos, S.R., and Ochman, H. (2004). Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. Environ. Microbiol. *6*, 754–759.
22. Zeigler, D.R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. Int. J. Syst. Evol. Microbiol. *53*, 1893–1900.
23. Hanage, W.P., Fraser, C., and Spratt, B.G. (2006). Sequences, sequence clusters and bacterial species. Phil. Trans. Roy. Soc. Ser. B *361*, 1917–1927.
24. Mayr, E. (1944). Systematics and the Origin of Species from the Viewpoint of a Zoologist (New York: Columbia Univ. Press).
25. de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. Proc. Natl. Acad. Sci. USA *102* (Suppl. 1), 6600–6607.
26. Coyne, J.A., and Orr, H.A. (2004). Speciation (Sunderland: Sinauer Associates).
27. De Clerck, E., Rodriguez-Diaz, M., Forsyth, G., Lebbe, L., Logan, N.A., and DeVos, P. (2004). Polyphasic characterization of *Bacillus coagulans* strains, illustrating heterogeneity within this species, and emended description of the species. Syst. Appl. Microbiol. *27*, 50–60.

28. Logan, N.A., and Berkeley, R.C. (1984). Identification of *Bacillus* strains using the API system. J. Gen. Microbiol. *130*(Pt 7), 1871–1882.

29. Bochner, B.R., Gadzinski, P., and Panomitros, E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res. *11*, 1246–1255.

30. Feldgarden, M., Byrd, N., and Cohan, F.M. (2003). Gradual evolution in bacteria: evidence from *Bacillus* systematics. Microbiology *149*, 3565–3573.

31. Bergthorsson, U., and Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol. Biol. Evol. *15*, 6–16.

32. Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J., Sarma-Rupavtarm, R., Distel, D.L., and Polz, M.F. (2005). Genotypic diversity within a natural coastal bacterioplankton population. Science *307*, 1311–1313.

33. Boucher, Y., Douady, C.J., Sharma, A.K., Kamekura, M., and Doolittle, W.F. (2004). Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. J. Bacteriol *186*, 3980–3990.

34. Lindsay, J.A., and Holden, M.T. (2004). *Staphylococcus aureus*: superbug, super genome? Trends Microbiol. *12*, 378–385.

35. Nelson, K.E., Fouts, D.E., Mongodin, E.F., Ravel, J., DeBoy, R.T., Kolonay, J.F., Rasko, D.A., Angiuoli, S.V., Gill, S.R., Paulsen, I.T., *et al*. (2004). Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species. Nucleic Acids Res. *32*, 2386–2395.

36. Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. *57*, 81–91.

37. Staley, J.T. (2003). Speciation and bacterial phylospecies. In Microbial Diversity and Bioprospecting, A.T. Bull, ed. (Washington, D.C.: American Society for Microbiology Press), pp. 40–48.

38. Cohan, F.M. (2006). Toward a conceptual and operational union of bacterial systematics, ecology, and evolution. Proc. Roy. Soc. Lond. Series B *361*, 1985–1996.

39. Lopez-Lopez, A., Bartual, S.G., Stal, L., Onyshchenko, O., and Rodriguez-Valera, F. (2005). Genetic analysis of housekeeping genes reveals a deep-sea ecotype of *Alteromonas macleodii* in the Mediterranean Sea. Environ. Microbiol. *7*, 649–659.

40. Schloter, M., Lebuhn, M., Heulin, T., and Hartmann, A. (2000). Ecology and evolution of bacterial microdiversity. FEMS Microbiol. Rev. *24*, 647–660.

41. Smith, N.H., Kremer, K., Inwald, J., Dale, J., Driscoll, J.R., Gordon, S.V., van Soolingen, D., Hewinson, R.G., and Smith, J.M. (2006). Ecotypes of the *Mycobacterium tuberculosis* complex. J. Theor. Biol. *239*, 220–225.

42. Höfle, M.G., Ziemke, F., and Brettar, I. (2000). Niche differentiation of *Shewanella putrefaciens* populations from the Baltic as revealed by molecular and metabolic fingerprinting. In Microbial Biosystems: New Frontiers, Proc. Eighth Int. Symp. Micr. Ecol., C.R. Bell, M. Brylinsky, and P. Johnson-Green, eds. (Halifax: Atlantic Canada Society for Microbial Ecology), pp. 135–142.

43. Sikorski, J., and Nevo, E. (2005). Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II Israel. Proc. Natl. Acad. Sci. USA *102*, 15924–15929.

44. Cohan, F.M., Perry, E., Koeppel, A., Krizanc, D., Ward, D.M., Bateson, M., Rooney, A., Sikorski, J., Nevo, E., and Ratcliff, R.M. (2007). Identifying the fundamental units of bacterial diversity. (in preparation).

45. Coleman, M.L., Sullivan, M.B., Martiny, A.C., Steglich, C., Barry, K., Delong, E.F., and Chisholm, S.W. (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. Science *311*, 1768–1770.

46. Enright, M.C., Spratt, B.G., Kalia, A., Cross, J.H., and Bessen, D.E. (2001). Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between *emm* type and clone. Infect. Immun. *69*, 2416–2427.

47. Paraskevopoulos, C., Bordenstein, S.R., Wernegreen, J.J., Werren, J.H., and Bourtzis, K. (2006). Toward a *Wolbachia* multilocus sequence typing system: discrimination of *Wolbachia* strains present in *Drosophila* species. Curr. Microbiol. *53*, 388–395.

48. 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., *et al*. (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.

49. Helgason, E., Okstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., and Kolsto, A.B. (2000). *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*–one species on the basis of genetic evidence. Appl. Environ. Microbiol. *66*, 2627–2630.

50. Cohan, F.M. (2002). Population structure and clonality of bacteria. In Encyclopedia of Evolution, *Volume 1*, M. Pagel, ed. (New York: Oxford University Press), pp. 161–163.

51. Feil, E.J., Maiden, M.C., Achtman, M., and 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.

52. Maynard Smith, J., Smith, N., O'Rourke, M., and Spratt, B.G. (1993). How clonal are bacteria? Proc. Natl. Acad. Sci. USA *90*, 4384–4388.

53. Falush, D., Kraft, C., Taylor, N.S., Correa, P., Fox, J.G., Achtman, M., and Suerbaum, S. (2001). Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. Proc. Natl. Acad. Sci. USA *98*, 15056–15061.

54. Cohan, F.M. (1994). The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. Am. Nat. *143*, 965–986.

55. Koch, A.L. (1974). The pertinence of the periodic selection phenomenon to prokaryote evolution. Genetics *77*, 127–142.

56. Levin, B.R. (1981). Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. Genetics *99*, 1–23.

57. Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S.W., *et al*. (2007). An East African origin for the intimate association between humans and *Helicobacter pylori*. Nature *445*, 915–918.

58. Kimura, M. (1983). The Neutral Theory of Molecular Evolution (Cambridge: Cambridge Univ. Press).

59. Cohan, F.M. (2002). What are bacterial species? Annu. Rev. Microbiol. *56*, 457–487.

60. Lawrence, J.G. (2002). Gene transfer in bacteria: speciation without species? Theor. Popul. Biol. *61*, 449–460.

61. Godreuil, S., Cohan, F., Shah, H., and Tibayrenc, M. (2005). Which species concept for pathogenic bacteria? An E-Debate. Infect. Genet. Evol. *5*, 375–387.

62. Cohan, F.M. (1995). Does recombination constrain neutral divergence among bacterial taxa? Evolution *49*, 164–175.

63. Guttman, D.S., and Dykhuizen, D.E. (1994). Detecting selective sweeps in naturally occurring *Escherichia coli*. Genetics *138*, 993–1003.

64. Cohan, F.M. (2005). Periodic selection and ecological diversity in bacteria. In Selective Sweep, D. Nurminsky, ed. (Georgetown, Texas: Landes Bioscience), pp. 78–93.

65. Cohan, F.M. (2001). Bacterial species and speciation. Syst. Biol. *50*, 513–524.

66. Majewski, J., and Cohan, F.M. (1999). Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. Genetics *152*, 1459–1474.

67. Bhaya, D., Grossman, A.R., Steunou, A.S., Khuri, N., Cohan, F.M., Hamamura, N., Melendrez, M., Bateson, M.M., Ward, D.M., and Heidelberg, J.F. (2007). Genomic, metagenomic and functional analyses of cyanobacteria from hot-spring microbial mats reveal an unexpected diversity in nutrient utilization strategies. ISME J. (in review).

68. Roumagnac, P., Weill, F.X., Dolecek, C., Baker, S., Brisse, S., Chinh, N.T., Le, T.A., Acosta, C.J., Farrar, J., Dougan, G., *et al*. (2006). Evolutionary history of *Salmonella typhi*. Science *314*, 1301–1304.

69. Rainey, P.B., and Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. Nature *394*, 69–72.

70. Rozen, D.E., and Lenski, R.E. (2000). Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. Am. Nat. *155*, 24–35.

71. Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. (2003). Geographical isolation in hot spring cyanobacteria. Environ. Microbiol. *5*, 650–659.

72. Beijerinck, M.W. (1913). Jaarboek van de Koninklijke Akademie v. Wetenschappen (Amsterdam: Müller).

73. Finlay, B.J. (2002). Global dispersal of free-living microbial eukaryote species. Science *296*, 1061–1063.

74. Roberts, M.S., and Cohan, F.M. (1995). Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. Evolution *49*, 1081–1094.

75. Staley, J.T., and Gosink, J.J. (1999). Poles apart: biodiversity and biogeography of sea ice bacteria. Annu. Rev. Microbiol. *53*, 189–215.

76. Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science *301*, 976–978.

77. Papke, P.T., and Ward, D.M. (2004). The importance of physical isolation in microbial evolution. FEMS Microbiol. Ecol. *48*, 293–303.

78. Achtman, M., Morelli, G., Zhu, P., Wirth, T., Diehl, I., Kusecek, B., Vogler, A.J., Wagner, D.M., Allender, C.J., Easterday, W.R., *et al*. (2004). Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc. Natl. Acad. Sci. USA *101*, 17837–17842.

79. Gutacker, M.M., Mathema, B., Soini, H., Shashkina, E., Kreiswirth, B.N., Graviss, E.A., and Musser, J.M. (2006). Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J. Infect. Dis. *193*, 121–128.

80. Keim, P., and Smith, K.L. (2002). *Bacillus anthracis* evolution and epidemiology. Curr. Top. Microbiol. Immunol. *271*, 21–32.

81. Palys, T., Berger, E., Mitrica, I., Nakamura, L.K., and Cohan, F.M. (2000). Protein-coding genes as molecular markers for ecologically distinct populations: the case of two Bacillus species. Int. J. Syst. Evol. Microbiol. *50*(Pt 3), 1021–1028.

82. Maynard Smith, J., and Szathmáry, E. (1995). The Major Transitions in Evolution (Oxford: Oxford Univ. Press).

83. Wagner, A. (2005). Robustness and Evolvability in Living Systems (Princeton: Princeton Univ. Press).

84. Cohan, F.M., King, E.C., and Zawadzki, P. (1994). Amelioration of the deleterious pleiotropic effects of an adaptive mutation in *Bacillus subtilis*. Evolution *48*, 81–95.

85. Levin, B.R., Perrot, V., and Walker, N. (2000). Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics *154*, 985–997.

86. Vermeij, G.J. (1973). Adaptation, versatility, and evolution. Syst. Zool. *22*, 466–477.

87. Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., *et al*. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature *406*, 959–964.

88. MacLean, R.C., and Bell, G. (2002). Experimental adaptive radiation in *Pseudomonas*. Am. Nat. *160*, 569–581.

89. Cohan, F.M. (2004). Concepts of bacterial biodiversity for the age of genomics. In Microbial Genomes, C.M. Fraser, T.D. Read, and K.E. Nelson, eds. (Totowa, New Jersey: Humana), pp. 175–194.

90. Cohan, F.M. (1996). The role of genetic exchange in bacterial evolution. ASM News *62*, 631–636.

91. Cohan, F.M., Roberts, M.S., and King, E.C. (1991). The potential for genetic exchange by transformation within a natural population of *Bacillus subtilis*. Evolution *45*, 1393–1421.

92. Turner, K.M., and Feil, E.J. (2007). The secret life of the multilocus sequence type. Int. J. Antimicrob. Agents *29*, 129–135.

93. Majewski, J., and Cohan, F.M. (1999). DNA sequence similarity requirements for interspecific recombination in Bacillus. Genetics *153*, 1525–1533.

94. Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F.M., and Dowson, C.G. (2000). Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. J. Bacteriol. *182*, 1016–1023.

95. Vulić, M., Dionisio, F., Taddei, F., and Radman, M. (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc. Natl. Acad. Sci. USA *94*, 9763–9767.

96. Cohan, F.M. (2002). Sexual isolation and speciation in bacteria. Genetica *116*, 359–370.

97. Templeton, A.R., Maskas, S.D., and Cruzan, M.B. (2000). Gene trees: a powerful tool for exploring the evolutionary biology of species and speciation. Plant. Species Biol. *15*, 211–222.

98. Dobrindt, U. (2005). (Patho-)Genomics of *Escherichia coli*. Int. J. Med. Microbiol. *295*, 357–371.

99. Ron, E.Z. (2006). Host specificity of septicemic *Escherichia coli*: human and avian pathogens. Curr. Opin. Microbiol. *9*, 28–32.

100. Cohan, F.M., Koeppel, A., and Krizanc, D. (2006). Sequence-based discovery of ecological diversity within *Legionella*. In Legionella: State of the Art 30 Years after Its Recognition, N.P. Cianciotto, Y. Abu Kwaik, P.H. Edelstein, B.S. Fields, D.F. Geary, T.G. Harrison, C. Joseph, R.M. Ratcliff, J.E. Stout, and M.S. Swanson, eds. (Washington, D.C.: ASM Press), pp. 367–376.

101. Martin, A.P. (2002). Phylogenetic approaches for describing and comparing the diversity of microbial communities. Appl. Environ. Microbiol. *68*, 3673–3682.

102. Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. Nature *430*, 551–554.

103. Grishkan, I., Beharav, A., Kirzhner, V., and Nevo, E. (2007). Adaptive spatiotemporal distribution of soil microfungi in "Evolution Canyon" III, Nahal Shaharut, extreme southern Negev Desert, Israel. Biol. J. Linn. Soc. *90*, 263–277.

104. Ferris, M.J., Kühl, M., Wieland, A., and Ward, D.M. (2003). Cyanobacterial ecotypes in different optical microenvironments of a 68 degrees C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. Appl. Environ Microbiol. *69*, 2893–2898.

105. Gordon, D.M., and Cowling, A. (2003). The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. Microbiology *149*, 3575–3586.

106. Joyce, E.A., Chan, K., Salama, N.R., and Falkow, S. (2002). Redefining bacterial populations: a post-genomic reformation. Nat. Rev. Genet. *3*, 462–473.

107. Lawrence, J.G., and Hendrickson, H. (2003). Lateral gene transfer: when will adolescence end? Mol. Microbiol. *50*, 739–749.

108. Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., and Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol. Rev. *60*, 407–438.

109. Brüggemann, H., Hagman, A., Coppée, J., Dillies, M., Heuner, K., Steinert, M., Gouyette, C., and Buchrieser, C. (2006). Bringing the genome of *L. pneumophila* to life: Gene expression patterns of the replicative and transmissive phase *in vitro*. In Legionella, N.P. Cianciotto, Y. Abu Kwaik, P.H. Edelstein, B.S. Fields, D.F. Geary, T.G. Harrison, C. Joseph, R.M. Ratcliff, J.E. Stout, and M.S. Swanson, eds. (Washington, D.C.: ASM Press).

110. European Commission (2005). Guideline developed within the Standing Committee on the Food Chain and Animal Health on the taxonomic level of micro-organisms to be included in Annex I to Directive 91/414/EEC.

111. Bohannan, B.J., and Hughes, J. (2003). New approaches to analyzing microbial biodiversity data. Curr. Opin. Microbiol. *6*, 282–287.