Annu. Rev. Microbiol. 2002. 56:457–87
doi: 10.1146/annurev.micro.56.012302.160634
Copyright © 2002 by Annual Reviews. All rights reserved
First published online as a Review in Advance on May 10, 2002

WHAT ARE BACTERIAL SPECIES?

Frederick M. Cohan
Department of Biology, Wesleyan University, Middletown, Connecticut 06459-0170; e-mail: fcohan@wesleyan.edu

Key Words speciation, ecotype, species concept, genetic exchange, systematics, MLST

■ **Abstract** Bacterial systematics has not yet reached a consensus for defining the fundamental unit of biological diversity, the species. The past half-century of bacterial systematics has been characterized by improvements in methods for demarcating species as phenotypic and genetic clusters, but species demarcation has not been guided by a theory-based concept of species. Eukaryote systematists have developed a universal concept of species: A species is a group of organisms whose divergence is capped by a force of cohesion; divergence between different species is irreversible; and different species are ecologically distinct. In the case of bacteria, these universal properties are held not by the named species of systematics but by ecotypes. These are populations of organisms occupying the same ecological niche, whose divergence is purged recurrently by natural selection. These ecotypes can be discovered by several universal sequence-based approaches. These molecular methods suggest that a typical named species contains many ecotypes, each with the universal attributes of species. A named bacterial species is thus more like a genus than a species.

CONTENTS

INTRODUCTION

These are formative times for systematics. Owing to recent technological developments, the study of biological diversity has matured into a powerful science. Systematists have invented robust methods for deriving evolutionary trees (35) as well as powerful computer algorithms for implementing these methods (91). Also, the technology for DNA sequencing is readily accessible, so that systematic research today is typically based on sequence variation and sequence-derived phylogenies. Together, the new methods of systematics and a wealth of sequence data are allowing systematists to reconstruct the history of life and the origins of biodiversity with confidence.

For microbiologists, the new systematics has been particularly fruitful. Microbial systematists have built a universal tree of life, such that any newly discovered organism can be placed near its closest relatives (18, 104). Sequence surveys have fostered discovery of new bacterial taxa at all levels. Sequence data have frequently turned up organisms with no known close relatives; these organisms represent new divisions within the bacterial world (18). At the other extreme, sequence surveys have fostered the discovery of new species. For example, sequence data revealed the Lyme disease spirochete *Borrelia burgdorferi* (sensu lato) to consist of several species, each with its own etiology (7, 8). Sequence-based approaches have allowed systematists to characterize the species diversity even among uncultured bacteria, and species names have now been given to many uncultured organisms, pending further characterization (62).

Despite these remarkable successes, however, bacterial systematics has not yet reached a consensus for defining the fundamental unit of biological diversity, the species. Bacterial species exist—on this much bacteriologists can agree. Bacteriologists widely recognize that bacterial diversity is organized into discrete phenotypic and genetic clusters, which are separated by large phenotypic and genetic gaps, and these clusters are recognized as species (85). Beyond agreeing on the existence of species, however, bacteriologists differ on operational procedures for identifying species most appropriately (32). Moreover, we fail to agree on whether a bacterial species should be conceived simply as a cluster of phenotypically and genetically similar organisms, or whether we should also expect a species to have special genetic, ecological, evolutionary, or phylogenetic properties.

I argue that there are bacterial taxa called “ecotypes” (15), which share the quintessential set of dynamic properties held by all eukaryotic species. I demonstrate that, alas, the species generally recognized in bacterial systematics do not have these universal properties: Each named “species” appears to contain many ecotypes, each with the dynamic properties of a species. I present several universal sequence-based approaches for discovering ecotypes and recommend a means of incorporating ecotypes into bacterial taxonomy.

BACTERIAL SPECIES AS PHENOTYPIC AND GENETIC CLUSTERS

Bacterial systematics began in much the same way as the systematics of animals and plants. Before there was a widely accepted theory of species, systematists of all organisms identified species simply as phenotypic clusters (87). While macrobiologists generally surveyed morphological characters and microbiologists generally investigated metabolic characters, systematists of all major groups were successful in carving biological diversity into the phenotypic clusters they identified as species (4).

Macrobial and microbial systematics split profoundly with Mayr's publication of the Biological Species Concept in 1944 (57), in which evolutionary theory was incorporated into systematics. The Biological Species Concept and several later concepts of species changed zoologists' and botanists' views of what a species should represent. A species was no longer merely a cluster of similar organisms; a species was now viewed as a fundamental unit of ecology and evolution, with certain dynamic properties. In the case of Mayr's Biological Species Concept, a species was viewed as a group of organisms whose divergence is opposed by recombination between them. However, as we shall see, the mainstream of bacterial systematics has not incorporated theory-based concepts of species.

Instead, the past half-century of bacterial systematics has been characterized by improvements in methods for demarcating species as clusters. Technological advances in assaying phenotypes have increased our ability to discern closely related species. For example, gas chromatographic techniques for characterizing a strain's fatty acid content (81) frequently provide important diagnostic phenotypic characters (67). Automated equipment for assaying metabolism, such as the microtiter plate reader, allows many more strains to be assessed for many more metabolic traits (10, 53).

Bacterial systematists have also developed improved statistical methods for demarcating phenotypic clusters. Numerical taxonomy, developed by Sneath & Sokal (86), was designed as an objective, mathematical approach for demarcating clusters. In this method, large numbers of strains are assayed for many phenotypic traits, including degradation or metabolism of certain chemicals, the ability to produce and survive various antibiotics, the ability to grow on various carbon or nitrogen sources, staining reactions, and morphology (32). The multidimensional space of phenotypic diversity can then be collapsed onto two or three axes, for example, by principal component analysis. Practitioners of numerical taxonomy have clearly illustrated the discrete nature of phenotypic clusters within many bacterial genera (9). Note that phenotype-based numerical taxonomy is ultimately much more than a method for delimiting species; it is also a venture into the natural history of a bacterial group. By studying the phenotypic variation within a taxon, we learn about the ecological diversity that gives meaning to the taxonomic enterprise.

Over the past three decades, bacterial systematists have added molecular techniques to their arsenal for demarcating clusters. In the early 1970s bacteriologists adopted a genomic approach for discovering clusters. With whole-genome DNA-DNA hybridization, systematists could assay the genomic similarity of two strains, as measured by the fraction of their genomes that are homologous. Johnson (44) determined that strains from the same species, as defined by phenotypic clustering, nearly always shared 70% or more of their genomes and that strains from different species nearly always shared less than 70%. A 70% level of homology over the genome was thus adopted as a gold standard for determining whether two strains should be considered different species (100).

Systematists have more recently utilized DNA sequence divergence data, particularly divergence in the 16S rRNA genes, for demarcating species. Stackebrandt & Goebel (88) found that strains that are more than 3% divergent in 16S rRNA are nearly always members of different species, as determined by DNA-DNA hybridization, whereas strains that are less than 3% divergent may or may not be generally members of different species. A cutoff of 3% divergence was therefore recommended as a conservative criterion for demarcating species.

Molecular approaches have made bacterial systematics much more accessible. The existence of multiple species within a group can now be inferred by general molecular techniques, even when we are unaware of the phenotypic and ecological differences between the clusters. In the case of uncultivated bacteria, molecular methods are all that is available for identifying the diversity of species.

Nevertheless, I argue that molecular approaches have been under-utilized in bacterial systematics. This is because the molecular cutoffs for demarcating species have been calibrated to yield the species groupings already determined by phenotypic clustering. The 70% cutoff for DNA-DNA hybridization was calibrated to yield the phenotypic clusters previously recognized as separate species, and the 3% cutoff for 16S rRNA divergence was calibrated to yield the species previously determined by DNA-DNA hybridization and phenotypic clustering. Because bacterial systematics is lacking a theory-based concept of species, all we can do is calibrate each new molecular technique to yield the clusters previously determined by phenotypic criteria. I demonstrate that, if we adopt an ecological and evolutionary theory of species, molecular approaches will give us much more than the phenotypic clusters of yore; we will be able to identify nearly every ecologically distinct population of bacteria.

THE UNIVERSAL DYNAMIC PROPERTIES OF SPECIES

Evolutionary biologists and systematists of animals and plants have widely believed that species are more than just clusters of closely related and similar organisms. Species are believed to have some quintessential, dynamic properties as well. Most fundamentally, genetic diversity within a species is thought to be constrained by one or more forces of cohesion.

The principal insight of Mayr's (57) Biological Species Concept is that there is a reason why organisms form the tight clusters discovered by systematists. Every species has a force of cohesion that hinders genetic divergence among its members, and (at least for highly sexual animal and plant species) this cohesive force is genetic exchange. So long as organisms can successfully interbreed, argued Mayr, they will remain phenotypically and genetically similar, but when they lose the ability to interbreed, they become free to diverge without bound. Mayr thus defined species as a reproductive community, a group of organisms with the potential to interbreed and produce viable and fertile offspring.

The Biological Species Concept was widely accepted by zoologists and botanists for two compelling reasons, one based in theory and the other empirical. First, population genetic theory predicts that recurrent interbreeding between populations tends to homogenize the populations at all gene loci (105). Thus, Mayr (58) argued that even low levels of genetic exchange between populations should limit their genetic divergence [although probably not in the way Mayr envisioned; see (30)]. Second, Mayr pointed out a striking empirical correspondence between the phenotypic clusters previously recognized as species and groups of organisms with the potential to interbreed.

Systematists have pointed out that the Biological Species Concept does not felicitously accommodate hybridization between species of plants (89) and species of bacteria (75, 76). That is, there are many pairs of species (as defined by phenotypic clustering) that occasionally exchange genes yet retain their integrity as distinct phenotypic clusters. In these cases, genetic exchange is clearly not acting effectively as a force of cohesion.

The problem of hybridization has been accommodated by the Cohesion Species Concept by Meglitsch (61) and Templeton (92). According to the Cohesion Species Concept, a species is a group of organisms whose divergence is capped by one or more forces of cohesion. In the case of sexual species, the predominant cohesive force is understood to be genetic exchange. But the cohesion species concept takes into account that genetic exchange between two groups is not always sufficient to restrict their divergence. In cases where two clusters of organisms retain their separate adaptations, in spite of occasional genetic exchange between them, the clusters are considered separate species, according to the Cohesion Species Concept. With regard to highly sexual species, the Cohesion Species Concept may be understood as a paraphrasing of the Biological Species Concept: A species is a group of organisms whose divergence is constrained by genetic exchange. We see that the Cohesion Species Concept is especially useful in accommodating bacterial groups that form separate phenotypic clusters despite recurrent recombination between them.

Another limitation of the Biological Species Concept is that it does not accommodate asexual species. Again, the Cohesion Species Concept proves to be general enough to take into account species that fall outside the box built by the Biological Species Concept. Templeton (92) argues that asexual species are subject to their own powerful force of cohesion. This force is natural selection, which can purge all genetic diversity from an asexual population. Consider the fate of an adaptive

mutation, which grants its bearer superior competitive ability compared to all other members of the population. For example, Ferea et al. (29) found that in asexual experimental populations of *Saccharomyces cerevisiae*, evolutionary adaptation to an oxygenated environment involved many mutations, some suppressing anaerobic metabolism and others augmenting oxidative phosphorylation. Each of these mutations was brought to fixation (i.e., to a frequency of 100%) by natural selection. In the absence of recombination, the entire genome of the successful mutant is brought to fixation, and so all the genetic diversity within the population (at all loci) is purged to zero. An asexual species may then be understood as a group of organisms whose divergence is constrained and recurrently reset to zero by intermittent bouts of natural selection. [This diversity-purging process is called "periodic selection" (5).]

What, then, constitutes speciation in the asexual world? A new species is formed when an asexual lineage evolves into a new ecological niche (i.e., uses a different set of resources or microhabitats), such that the new species cannot be extinguished by adaptive mutants from its former population (Figure 1). For example, within

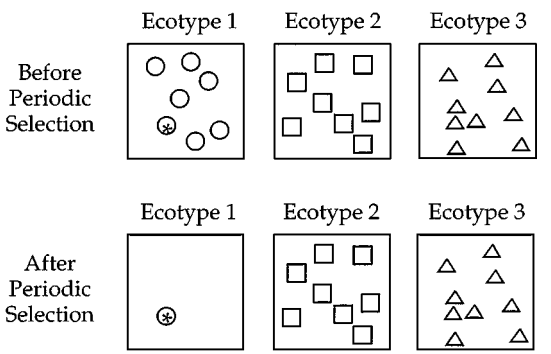

Figure 1 The transience of diversity within an asexual (or rarely sexual) species and the permanence of divergence among species. Each individual symbol represents an individual organism, and the distance between symbols represents the genetic divergence between organisms. The *asterisk* represents an adaptive mutation in one individual of Species 1. Because of the absence (or rarity) of recombination, the adaptive mutant and its clonal descendants replace the rest of the genetic diversity within the species. Genetic diversity within an asexual (or rarely sexual) species is thus transient, awaiting its demise with the next periodic selection event. Because asexual (or rarely sexual) species differ in the resources they use, the adaptive mutant from one species (e.g., Species 1) does not out-compete the organisms of other species, and the genetic diversity within these other species remains untouched. Once populations of organisms are divergent enough to escape one another's periodic selection events, these populations are free to diverge permanently and have reached the status of distinct species. Used with permission from the American Society of Microbiology (14a).

an experimental population of *Escherichia coli* inoculated with a single clone, a mutant population arose that fed on a waste metabolite from the original (95). While the original population and the cross-feeding population continued to coexist, each population endured its own periodic selection events. That is, adaptive mutants appearing in each population purged the diversity within their own populations. Owing to the differences in ecology between populations, however, periodic selection from one population failed to affect the diversity within the other. When two asexual populations reach the point that they can survive each other's periodic selection events, a force of cohesion no longer caps their divergence, and so they may be considered separate species.

In summary, any species, whether highly sexual (such as animals and plants) or completely asexual (such as experimental microbial populations engineered to lack sex), is subject to forces of cohesion. Within highly sexual populations, genetic exchange is a powerful force of cohesion, although genetic exchange between groups that recombine only rarely (as in hybridizing species) does not constrain divergence. Within asexual species, periodic selection is a powerful force of cohesion that recurrently resets the genetic diversity within the species to zero.

While cohesion is the fundamental attribute claimed for all species (13, 15, 61, 92), several corollary attributes follow from cohesion. Because different species are not bound together by any cohesive force, they are free to diverge without constraint from one another. Different species are thus understood to be irreversibly separate, with distinct evolutionary fates [the Evolutionary Species Concept by Simpson (83) and Wiley (103)].

Another corollary is that species occupy different ecological niches either by utilizing different kinds of resources or by utilizing the same kinds of resources at different times or within different microhabitats [the Ecological Species Concept by van Valen (98)]. This is most clearly true for asexual species, as asexual organisms must diverge ecologically before they can escape one another's periodic selection events. However, it is also true for any pair of species, if they are to coexist without constraint from one another. Early twentieth-century ecologists demonstrated that two species cannot coexist unless they differ at least somewhat in the resources they consume [the competitive exclusion principle (31)]. Thus, in the highly sexual world of animals and plants, speciation requires both reproductive divergence and ecological divergence (25, 31, 54, 59).

Although the various species concepts I have discussed emphasize different attributes of species, all modern species concepts endow species with one quintessential property: that species are evolutionary lineages that are irreversibly separate, each with its own evolutionary tendencies and historical fate (19).

Systematists have thus delineated general attributes of species that apply well for organisms at the extremes of sexuality, including species that exchange genes every generation as well as species that never exchange genes. As we shall see, bacteria fall into neither extreme, but nevertheless they fall into species that fit all the universal attributes of species I have delineated.

THE PECULIAR SEXUAL HABITS OF BACTERIA

Genetic exchange in bacteria differs profoundly from that in the highly sexual eukaryotes, the animals and plants, for which species concepts were invented. First, recombination in bacteria is extremely rare in nature. Several laboratories have taken a retrospective approach to determining the historical rate of recombination in nature (27, 28, 79). Based on surveys of diversity in allozymes, restriction recognition sites, and DNA sequences, recombination rates have been estimated from the degree of association between genes or parts of genes [(37, 39, 40); for limitations of this approach, see (16, 56)]. Survey-based approaches have shown that in most cases a given gene segment is involved in recombination at about the same rate or less, as mutation (27, 28, 56, 79, 82, 102). Less commonly, as in the cases of *Helicobacter pylori* and *Neisseria gonorrhoeae*, recombination occurs at least one order of magnitude more frequently than mutation, although survey methods do not allow us to determine by how much (66, 90).

Also in contrast to the case for animals and plants, recombination in bacteria is promiscuous. Whereas animal groups typically lose the ability to exchange genes entirely by the time their mitochondrial DNA sequences are 3% divergent (6), bacteria can undergo homologous recombination with organisms at least as divergent as 25% in DNA sequence (21, 52, 99).

There are, nevertheless, some important constraints on bacterial genetic exchange. Ecological differentiation between populations may prevent them from inhabiting the same microhabitat at the same time, as may be the case, for example, for *Streptococcus pyogenes* populations adapted for throat versus skin infection at different times of year (26). Recombination that depends on vectors, such as transduction or conjugation, is limited by the host ranges of the respective phage and plasmid vectors. Also, restriction-endonuclease activity can greatly reduce the rate of recombination by transduction (24, 60), although not by transformation (17, 94). Finally, homologous recombination is limited by the resistance to integration of divergent DNA sequences because mismatch repair tends to reverse integration of a mismatched donor-recipient heteroduplex (77, 99) and because integration requires a 20–30-bp stretch of nearly perfectly matched DNA (51, 52, 74).

In addition, recombination events in bacteria are localized to a small fraction of the genome. Segments transduced or transformed in the laboratory are frequently less than several kilobases in length (60, 107). Surveys of sequences in nature show that recombination in nature is likewise highly localized within the chromosome (27, 28, 50, 55).

Finally, recombination in bacteria is not limited to the transfer of homologous segments. Bacteria can acquire new gene loci from other organisms, which in some cases are extremely distantly related. This may occur as a side effect of homologous recombination, whereby a heterologous gene from a donor is integrated along with flanking homologous DNA (34, 52). Alternatively, heterologous genes may be integrated along with a transposable element brought into the recipient on a plasmid or phage (97). Genomic analyses have recently shown that a sizeable

fraction of bacterial species' genomes (frequently 5%–15%) has typically been acquired from other species (65).

WHY THE BIOLOGICAL SPECIES CONCEPT IS INAPPROPRIATE FOR BACTERIOLOGY

Systematists of animals and plants are indebted to Mayr's (57) Biological Species Concept for the infusion of evolutionary theory into systematics. Even though animal and plant species are usually discovered as phenotypic clusters (87), systematists have understood that there is a force of cohesion that holds each cluster together and that organisms in different clusters are no longer bound by a force of cohesion (92).

Why has bacterial systematics not been similarly transformed by evolutionary theory? It is not for lack of wrestling with the concept. For example, in the early 1960s, Ravin (75, 76) attempted to apply the Biological Species Concept to bacteria. Recognizing that bacteria are sexual and that bacteria can exchange genes even with distant relatives, Ravin (76) defined "genospecies" as groups of bacteria that could exchange genes and "taxospecies" as the phenotypic clusters of mainstream bacterial systematics. In contrast to the case for animals and plants, the genospecies and taxospecies so defined did not correspond well: Many clusters retained their phenotypic distinctness despite their inclusion within a genospecies (76). The lack of correspondence between genospecies and taxospecies suggested that the ability to exchange genes had little effect on the evolution of phenotypic divergence in bacteria. Perhaps discouraged, bacteriologists did not attempt to apply the Biological Species Concept (or any other theory-based concept of species) for another three decades.

Dykhuizen & Green (23) proposed in 1991 to classify bacteria into species according to the Biological Species Concept, delimiting bacterial species as groups of strains that recombine with one another but not with strains from other such groups. They suggested a phylogenetic approach using sequence data, which would identify groups that have and have not been exchanging genes.

One minor problem is that this proposal does not recognize the intrinsic promiscuity of genetic exchange: Bacteria do exchange genes both within and between the clusters we recognize as named species (48, 55). Nevertheless, in practice it is often the case that phylogenetic evidence for recombination is much more common within named bacterial species than between them.

My principal objection to applying the Biological Species Concept to bacteria is that there is no biological motivation for doing so. The Biological Species Concept is appropriate for the highly sexual animals and plants because divergence between two closely related animal or plant populations cannot be permanent until the rate of recombination between the populations is severely reduced compared to the rate of recombination within populations. Owing to the high rate of recombination within animal and plant populations, interpopulation recombination would rapidly

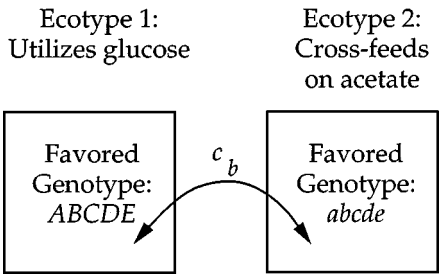

Figure 2 Ecological divergence between ecotypes is stable with respect to recurrent recombination. The genetic basis of ecological divergence among ecotypes is assumed to be due to differences in alleles at several gene loci (as in the figure), or divergence may be due to acquisition of different gene loci. Recombination occurs between ecotypes at rate c_b. The fitness penalty for recombination at any of the genes responsible for ecological divergence is s, such that the fitness of each nonrecombinant genotype (e.g., *ABCDE*) is 1, and the fitness of a single-locus recombinant (e.g., *ABCDe*) is $1 - s$. A mathematical model shows the equilibrium frequency of maladaptive foreign alleles in each ecotype to be c_b/s, which, given the low rate of recombination in bacteria, is a negligible frequency (14). Used with permission from American Society for Microbiology (14a).

eliminate interpopulation divergence if it were to proceed at the same rate as recombination within populations.

In contrast, because recombination in bacteria is so rare, recurrent recombination between bacterial species cannot hinder their divergence (14). Even if recombination between species were to occur at the same rate as recombination within them, natural selection against interspecies recombinants could easily limit the frequency of recombinant genotypes to negligible levels (Figure 2). While the evolution of sexual isolation is an important milestone in the origin of animal and plant species, it is irrelevant to the evolution of permanent divergence in the bacterial world. The Biological Species Concept is thus a red herring for bacterial systematics.

Nevertheless, bacteriologists need not envy the macrobial world for its tidy application of the Biological Species Concept. It turns out that there is an appropriate species concept for bacteria. Moreover, bacteria and eukaryotes both fit comfortably within a universal concept of species.

BACTERIA FORM SPECIES LIKE EVERYONE ELSE

Let us begin by defining a bacterial "ecotype" with respect to the fate of an adaptive mutant (14, 15): An ecotype is a set of strains using the same or similar ecological resources, such that an adaptive mutant from within the ecotype out-competes to extinction all other strains of the same ecotype; an adaptive mutant does not,

however, drive to extinction strains from other ecotypes (Figure 1). For example, an adaptive mutant from an ecotype of *Streptococcus pyogenes* that is genetically adapted to infecting our throats would out-compete to extinction other members of its own ecotype but would not out-compete closely related ecotypes genetically adapted to infecting our skin.

If they were entirely asexual, bacterial ecotypes defined in this way would have the universal properties of species. I earlier discussed how asexual populations defined by the domains of periodic selection have all the attributes of species. Each such asexual ecotype would be subject to a force of cohesion (the diversity-purging effect of its own periodic selection events), different ecotypes would be irreversibly separate (free to diverge from one another indefinitely), and the ecotypes would be ecologically distinct.

Let us now add the reality of rare but promiscuous genetic exchange to these ecotypes. Would they still retain the universal qualities of species? We might imagine that periodic selection would not be an effective force of cohesion within a rarely recombining population. Perhaps even rare genetic exchange would rapidly place the adaptive mutation into many genetic backgrounds within the ecotype, such that selection would fail to purge sequence diversity at all loci. However, this is not the case. Under rates of recombination typical of bacteria, selection will purge each locus, on average, of 99.9% of its sequence diversity (13). Thus, natural selection does act as a potential force of cohesion within rarely recombining ecotypes.

In contrast, there is no effective force of cohesion binding different ecotypes. Ecotypes are defined to be free to diverge without the constraint of one another's periodic selection events; moreover, as we have seen, the rare recombination occurring in bacteria is unable to prevent adaptive divergence between ecotypes.

In summary, the bacterial ecotypes defined here share the fundamental properties of species. They are each subject to an intense force of cohesion. Once different ecotypes have diverged to the point of escaping one another's periodic selection events, there is no force that can prevent their divergence; and bacterial ecotypes are ecologically distinct. Bacterial ecotypes are therefore evolutionary lineages that are irreversibly separate, each with its own evolutionary tendencies and historical fate (19, 83, 103). A species in the bacterial world may be understood as an evolutionary lineage bound together by ecotype-specific periodic selection.

BACTERIAL SPECIATION AS AN EVERYDAY PROCESS

How frequently do bacterial populations split irreversibly into lineages with separate evolutionary fates? By applying the principles of population genetics and ecology, we can predict an enormous potential for speciation in the bacterial world, much greater than that in the highly sexual world of animals and plants.

First, speciation in highly sexual eukaryotes requires both reproductive (59) and ecological (25, 31) divergence, but speciation in bacteria requires only ecological divergence (13).

Second, speciation in highly sexual eukaryotes requires allopatry (i.e., that the incipient species inhabit different geographical regions) (58), or at least microallopatry (i.e., that they inhabit different microhabitats) (11, 101). This is because highly sexual populations cannot diverge as long as they are exchanging genes at a high rate; allopatry (or microallopatry) provides the only mechanism for reducing genetic exchange between populations in early stages of speciation. In contrast, as I have shown, genetic exchange is too rare to hinder divergence between bacterial populations, and so the need for allopatry in bacterial speciation is greatly reduced (but not necessarily eliminated, as we shall see).

Third, the extremely large population sizes of bacteria make rare mutation and recombination events much more accessible to a bacterial population than is the case for macroorganisms.

Fourth, whereas each animal and plant species is genetically closed to all other species (except for hybridization with closely related species), a bacterial species is open to gene transfer from many other species, even those that are distantly related (65, 106). So, while animal and plant species must evolve all their adaptations on their own, bacteria can take up existing adaptations from a great diversity of other species. Homologous recombination can substitute an adaptive allele from another species into an existing gene in the recipient (55); recombination can also introduce entirely novel genes and operons from other species (3, 33, 46, 47, 65). By granting an entirely new metabolic function, heterologous gene transfer has the potential to endow a strain with a new resource base, such that the strain and its descendants are instantaneously a new species—beyond the reach of periodic selection within the strain's former population. Since 5%–15% of the genes in a typical bacterial genome have been acquired from other species (65), it is possible that many speciation events in the past have been driven by the acquisition of new genes.

The transfer of adaptations across species is facilitated by the peculiar characteristics of bacterial genetic exchange. Incorporation of highly divergent DNA is fostered not only by the promiscuous nature of bacterial genetic exchange, but also by the localized nature of bacterial recombination, whereby only a small fraction of the donor's genome is integrated. This allows for the transfer of a generally useful adaptation (i.e., useful in the genetic backgrounds and the ecological niches of both the donor and recipient), without the co-transfer of narrowly adapted donor segments that would be deleterious for the recipient (107). This is in contrast to the case for most eukaryotes, where the processes of meiosis and fertilization yield hybrids that are a 1:1 mix of both parents' genomes.

Finally, genetic exchange between ecotypes (48) may enhance speciation by preventing a nascent ecotype from being extinguished by an adaptive mutant from the parental ecotype (15) (Figure 3). This can occur if ecological divergence between incipient ecotypes involves several mutational steps. In the early stages of such divergence, nearly every periodic selection event may be limited to purging the diversity within its own ecotype. Occasionally, however, an extraordinarily fit adaptive mutant from the parental ecotype might out-compete all strains from

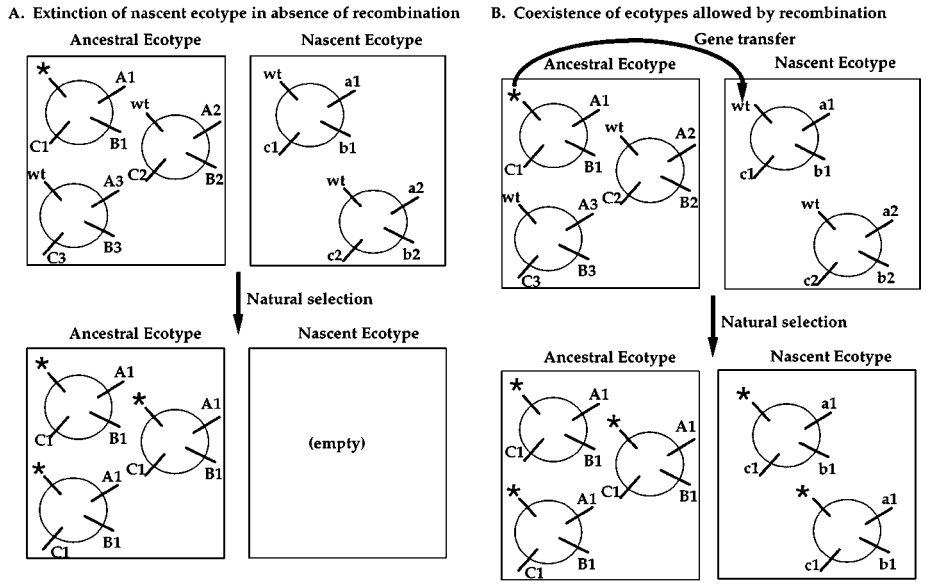

Figure 3 Facilitation of speciation by recombination among ecotypes. It is assumed that each newly divergent ecotype in the figure has already undergone several private periodic selection events. However, in the figure we suppose that an extraordinarily competitive adaptive mutant (with *asterisk*) has appeared in the ancestral ecotype, such that this mutant would out-compete the membership of the nascent ecotype as well as its own ecotype membership. (*A*) When there is no recombination between the newly divergent ecotypes, the adaptive mutant could extinguish the membership of the other ecotype, and the speciation process would be terminated. (*B*) When the adaptive mutation can be transferred from one ecotype to the other, periodic selection is less likely to cause extinction of one ecotype by another. The transfer of the adaptive mutation would cause a private periodic selection event within the nascent ecotype. Because the two ecotypes would then share the adaptive mutation, one ecotype would not be able to extinguish the other. Used with permission from the Society for Systematic Biology (15).

the nascent ecotype (as well as all the other strains from its own ecotype). In this case, the speciation process would be quashed by a periodic selection event before the two incipient ecotypes had diverged sufficiently to be completely free of one another (Figure 3*A*). However, recombination between two incipient species could potentially prevent this (Figure 3*B*). Through genetic exchange, the adaptive mutation could be transferred from the parental ecotype to a recipient in the other ecotype, and the new ecotype would lose its disadvantage.

Although recurrent recombination is not sufficient to prevent ecological divergence between ecotypes (13), recombination should be sufficient to allow an adaptive mutation to pass between ecotypes and enable the recipient ecotype to

become fixed for the adaptation by natural selection. Preventing divergence between ecotypes requires recurrent recombination at a high rate, but initiating a natural selection event in a recipient population requires only a single recombinational transfer into the recipient ecotype. Given the enormous population sizes of bacterial populations, such a transfer event is not unlikely.

In summary, population genetic principles suggest that the rare but promiscuous nature of bacterial genetic exchange, as well as the large population sizes of bacteria, should foster a much higher rate of speciation in bacteria than is possible in plants and animals. Nevertheless, important questions remain unresolved.

It is not clear, for example, whether bacterial speciation can proceed without allopatry. As I have discussed, allopatry is unnecessary for evolution of sexual isolation between incipient ecotypes because sexual isolation is not a necessary step in the origin of bacterial ecotypes. However, allopatry may be necessary to give a nascent ecotype a chance to gradually build up its ecological distinctness from the parental ecotype before being exposed to periodic selection from the parental ecotype.

In addition, we do not know the typical source of adaptations that enable invasion of new niches: Is it mutational change in existing genes, or acquisition of new gene loci from other species? Finally, it is not clear how frequently speciation in bacteria actually occurs in nature.

Fortunately, these issues can now be addressed by model experimental systems developed for studying the origins of ecological diversity in the bacterial world. In these model systems, a clone and its descendants are cultured in liquid in the laboratory and are allowed to evolve on their own. In one system, using *E. coli*, bacteria are cultured in a chemostat (95); in another (also using *E. coli*), the bacteria are maintained in serial batch culture (43, 80). In yet another system (using *Pseudomonas fluorescens*), the culture medium is neither replenished nor stirred (72). In all these systems, no extrinsic source of DNA is provided, so novel genes cannot be introduced by horizontal transfer. Moreover, all vectors of recombination have been eliminated from the *E. coli* systems.

From research in all these systems, it appears inevitable that a bacterial clone can evolve into multiple ecotypes by mutation alone. Treves et al. (95) found replicable evolution of a new ecotype, which utilized acetate secreted by the original clone. In experiments in a nonstirred environment, as performed by Rainey & Travisano (72), ecotypes have replicably arisen that are specifically adapted to different parts of the structured environment (i.e., the surface, the bottom, and the water column). In other experiments, molecular markers have demonstrated the existence of a diversity of ecotypes (43). Here, each periodic selection event has purged the diversity in only a subset of the population, indicating that multiple ecotypes are present. In some cases, the putative ecotypes have coexisted over years of evolution (80, 95).

The rate at which new ecotypes can be formed is striking. In the case of the physically structured environment, new ecotypes originated with high replicability in the course of several days. In the unstructured environments of the chemostat and serial batch culture, ecotypes originated within several weeks.

These experiments also demonstrate that allopatry is not required for ecotypes to gradually build up their ecological distinctness. Because some incipient ecotypes have coexisted for years (80, 95), it appears that nascent ecotypes have evolved to escape periodic selection from the parental ecotype without the benefit of allopatry.

The work by Treves et al. is especially notable in demonstrating the vast potential for speciation even in the simplest of environments: The chemostat environment does not have daily or seasonal fluctuations; the stirring eliminated the possibility of adaptation to different microhabitats; no extrinsic DNA was present; and only one carbon source was introduced into the system. The bacterial metabolism itself created a diversity of resources (by secreting acetate), and this was all that was needed to foster speciation. Thus, the evolution of new ecotypes would appear to be an ineluctable process in the bacterial world.

SEQUENCE-BASED APPROACHES TO IDENTIFYING BACTERIAL SPECIES

Discovery of Ecotypes as Sequence Clusters

The theory of evolutionary genetics provides a compelling rationale for using sequence data to characterize bacterial diversity (69). Given enough time, each bacterial ecotype is expected to be identifiable as a sequence cluster, where the average sequence divergence between ecotypes is much greater than the average sequence divergence within them, for any gene shared by the ecotypes. In addition, each ecotype is expected to be identifiable as a monophyletic group in a phylogeny based on DNA sequence data (Figure 4).

The rationale can be outlined from a phylogenetic perspective. Suppose a new ecotype is derived clonally from one mutant cell that is adapted to a new ecological niche. The nascent ecotype constitutes a monophyletic group descending from this original recombinant [i.e., the ecotype consists of all and only the descendants of the original mutant (Figure 4*A*)]. However, this ecotype is not yet a sequence cluster; one would not conclude from the sequence-based phylogeny that two populations exist within this group. After periodic selection, however, the diversity within the new ecotype is purged (Figure 4*B*). Likewise, periodic selection events within the ancestral ecotype will purge diversity within that ecotype as well (Figures 4*C*,*D*). Note that owing to the diversity-purging effect of periodic selection within each ecotype, the ecotypes eventually appear as separate sequence clusters and each is a monophyletic group (Figure 4*E*). Although this result is seen most clearly in the case of no recombination, Palys et al. (69) showed that under the extremely low rates of recombination occurring in bacteria different ecotypes are nevertheless expected to fall eventually into different sequence clusters for any gene shared across ecotypes.

Conversely, ecotypes are not expected to split into two or more sequence clusters (22, 69) because multiple clusters within an ecotype would be unstable with respect to periodic selection (Figure 5). Each adaptive mutant within the ecotype would

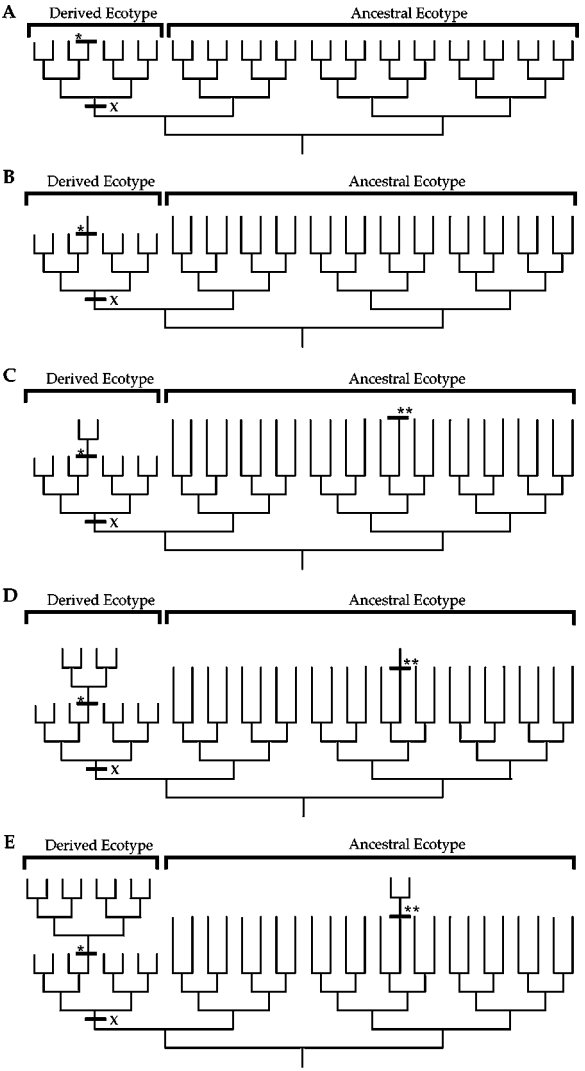

Figure 4 A phylogenetic perspective on periodic selection. As demonstrated here, two ecotypes will become distinct sequence clusters. (*A*) The derived ecotype consists of the descendants of a mutant (X) capable of utilizing a new ecological niche. The adaptive mutant in the derived ecotype (*) is capable of out-competing all other members of the derived ecotype. (*B*) The adaptive mutant (*) has driven all the other lineages within the derived ecotype to extinction. (*C*) With time, the derived ecotype becomes more genetically diverse. One cell in the ancestral ecotype (**) has developed a mutation that allows it to out-compete other members of its ecotype. (*D*) The adaptive mutant (**) has out-competed other members of the ancestral ecotype. (*E*) The ancestral ecotype is becoming more genetically diverse. At this point, each ecotype is a distinct sequence cluster as well as a monophyletic group. Used with permission from the Society for Systematic Biology (15).

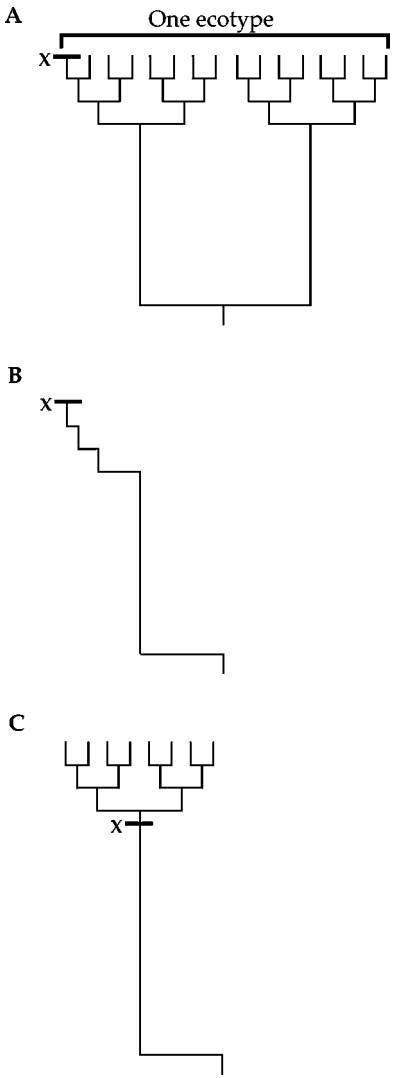

Figure 5 Sympatric members of a single ecotype cannot be split among multiple sequence clusters. (*A*) The ecotype initially contains two distinct sequence clusters. Then an adaptive mutation occurs in lineage X. (*B*) Because the adaptive mutant can purge diversity from the entire ecotype, only one cluster survives periodic selection. (*C*) After periodic selection, variation within the ecotype is re-established, but the population now forms a single sequence cluster. Used with permission from the American Society for Microbiology (69).

drive to extinction cells from all the clusters within the ecotype, and the cluster bearing the adaptive mutant would be all that survives this purge of diversity. If two highly divergent clusters have coexisted long enough to survive periodic selection, then the clusters must belong to different ecotypes.

There is one exception to this conclusion: Geographically isolated populations could diverge into separate sequence clusters, even if they are members of the same ecotype. In this case, an adaptive mutant from one geographical region would not be able to compete with subpopulations from other regions, so sequence divergence between the geographically isolated subpopulations could proceed indefinitely. Divergence among geographically isolated members of the same ecotype would be especially likely for bacteria with low mobility (perhaps pathogens of nonmobile hosts) but would not be possible for highly mobile organisms like *Bacillus*, where intercontinental migration of spores occurs extremely frequently (79). In any case, we can be sure that two highly divergent sequence clusters from the same geographic region (i.e., within migration range) must represent different ecotypes.

In summary, sequence clusters are expected to correspond, more or less, to ecotypes. Indeed, surveys of sequence diversity show a good correspondence between sequence clusters and groups known to be ecologically distinct (69, 96).

The correspondence between ecologically distinct populations and sequence clusters has proven useful for bacterial systematics in several ways (68). First, previously characterized sequence differences between taxa can be used diagnostically to identify unknown isolates.

Second, the correspondence between ecotypes and sequence clusters has enabled us to discover ecological diversity among uncultured bacteria. Increasingly often, uncultured taxa are being described on the basis of forming sequence clusters for 16S rRNA (20, 41, 62). For example, David Ward and coworkers have found that sequence clusters of uncultured *Synechococcus* strains from Yellowstone hot springs correspond to populations inhabiting distinct microenvironments defined by temperature, photic zone, and stage of ecological succession [(73); D. Ward, personal communication].

Finally, the correspondence between ecotypes and sequence clusters is useful for discovering cryptic ecological diversity within a named species. In several cases, a survey of sequence diversity within a named species has revealed multiple sequence clusters that were later found to be ecologically distinct (69). For example, a survey of sequence diversity of 16S rRNA in the genus *Frankia* uncovered previously unknown taxa with unique host specificities (64).

Discovery of Ecotypes as Star Clades

While sequence clusters provide a useful criterion for discovering ecotypes, a serious problem remains. A sequence-based phylogeny from almost any named bacterial species reveals a hierarchy of clusters, subclusters, and sub-subclusters. This raises the possibility that a typical named bacterial species may contain many cryptic and uncharacterized ecotypes, each corresponding to some small subcluster. The challenge is to determine which level of subcluster, if any, corresponds to

ecotypes. Fortunately, the peculiar population dynamics of bacteria allows us to identify the clusters that correspond to ecotypes.

Jason Libsch and I have developed a model for identifying the clusters corresponding to ecotypes [(15); J. Libsch & F.M. Cohan, unpublished results]. Our "star clade" approach assumes that the sequence diversity within an ecotype is constrained largely by periodic selection and much less by genetic drift (random fluctuation in gene frequencies within a population, most notably within populations of small size). This assumption is correct if the population size of a bacterial ecotype is 10^{10} or greater. [If sequence diversity in populations of this size were limited only by genetic drift, sequence diversity would be far greater than the 0.5%–1.0% generally seen within sequence clusters (14)].

Consider next the consequences of periodic selection on the phylogeny of an ecotype. Nearly all strains randomly sampled from an ecotype should trace their ancestries directly back to the adaptive mutant that caused and survived the last periodic selection event. Thus, the phylogeny of an ecotype should be consistent with a star clade, with only one ancestral node, such that all members of the ecotype are equally closely related to one another (Figure 6). In contrast, a population whose sequence diversity is limited by genetic drift will have a phylogeny with many nodes.

In an asexual ecotype, a sequence-based phylogeny would yield a perfect star clade, with only minor exceptions due to homoplasy (i.e., convergent nucleotide

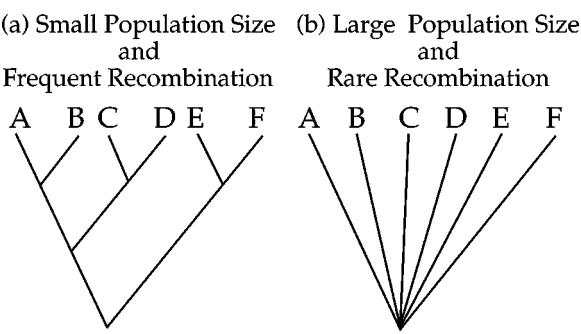

Figure 6 The phylogenetic signatures of populations whose diversity is controlled by periodic selection versus genetic drift. (*a*) In a population of small size, genetic drift causes coalescence of many pairs of lineages. Moreover, if recombination is frequent, there is no opportunity for genome-wide purging of diversity. Consequently, the phylogeny has many nodes. (*b*) In a bacterial population, characterized by large population size and rare recombination, the population's phylogeny is expected to resemble a star. Following periodic selection, each strain traces its ancestry directly back to the adaptive mutant that precipitated the periodic selection event. In addition, population sizes are too large for genetic drift to create coalescences between pairs of strains with appreciable frequency.

substitutions in different lineages and nucleotide substitutions reversing to a former state). However, in an ecotype subject to high rates of recombination, particularly with other ecotypes, the sequence-based phylogeny can deviate significantly from a perfect star clade. For example, suppose that a large segment from a divergent ecotype is recombined into one recipient within the ecotype and that this recipient later donates this foreign segment to another member of the ecotype (Figure 7). In a sequence-based phylogeny, these two strains would appear as closest relatives, a deviation from a perfect star clade.

We have used a computer simulation to determine how closely an ecotype's sequence-based phylogeny should resemble a perfect star clade (15). In general, within groups recombining only rarely (e.g., *Staphylococcus aureus*) (E. Feil, personal communication), the phylogeny of an ecotype is expected to closely

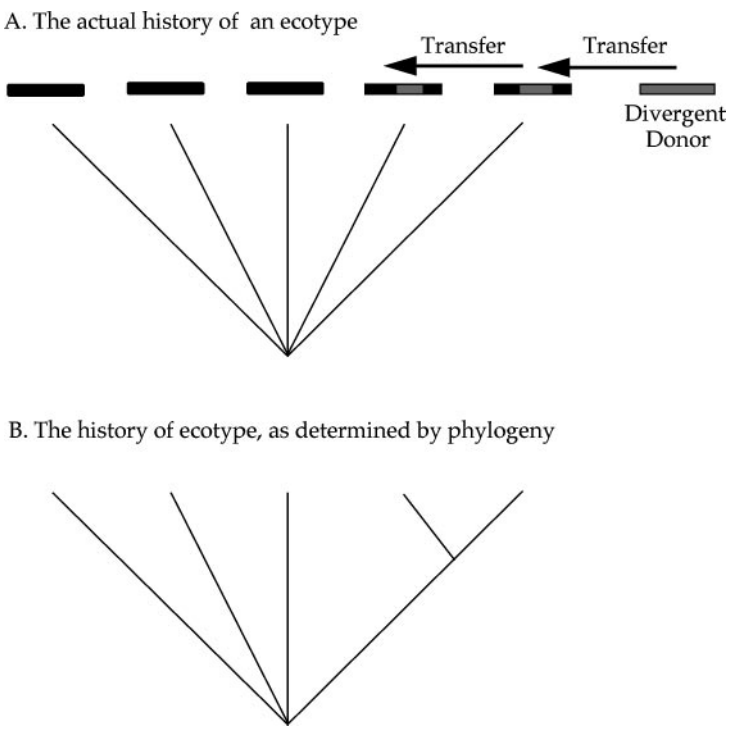

Figure 7 Recombination causes a bacterial ecotype's phylogeny to deviate from a star. Here a member of the ecotype has received a divergent donor's sequence in a gene on which the phylogeny is based, and the divergent sequence is in turn transferred to another member of the ecotype. Each of these strains then appears to be the other's closest relative, and a node is added to the phylogeny. The star clade computer simulation determines how many nodes an ecotype's phylogeny is expected to have, taking into account the recombination and mutation parameters of the taxon.

resemble a perfect star clade; within a taxon with more frequent recombination [e.g., *Neisseria meningitidis* (27)], the phylogeny of an ecotype is expected to deviate to a greater extent from the star form. Our approach is to determine, for a given taxon, how closely an ecotype's sequence-based phylogeny should resemble a star clade and then to identify the largest groups of strains that are each consistent with what is expected for an ecotype. Here the phylogenies are based on a concatenation of several gene loci, usually seven.

The number of nodes within a tree quantifies the degree of resemblance of an ecotype's phylogeny to a star clade: A perfect star has one node, and each additional coalescence of two or more lineages yields an additional node (Figure 8). In the case of *S. aureus*, where individual alleles are subject to mutation three times more frequently than recombination per gene (E. Feil, personal communication),

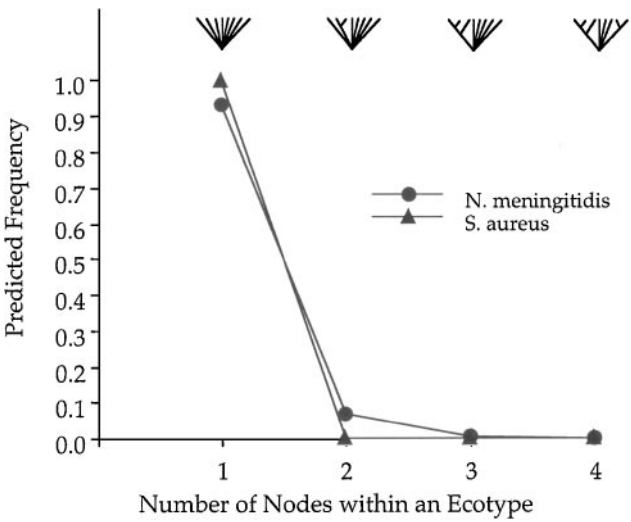

Figure 8 The number of nodes predicted to occur within an ecotype's phylogeny. The Star computer program simulated sequence evolution within an ecotype, at the seven loci used by MLST. The program took into account the recombination and mutation parameters for *N. meningitidis* and *S. aureus*, as estimated from sequence data. Sequence evolution was simulated over many replicate runs. From each replicate run, a 95% bootstrap-supported phylogeny of the ecotype was determined using PAUP*'s heuristic parsimony algorithm (91), based on a concatenation of the seven loci. The figure indicates for each taxon the fraction of times a particular number of nodes was found in the ecotype's phylogeny, over all replicate runs. Owing to the low frequency of recombination in *S. aureus*, an ecotype in this taxon usually has just one significant node; therefore, any collection of strains from *S. aureus* that has more than one node likely contains more than one ecotype. In the more frequently recombining *N. meningitidis*, an ecotype can have up to two significant nodes.

our simulations have shown that an ecotype is only rarely, by chance, expected to have more than one significant node in its phylogeny (F.M. Cohan, unpublished results) (Figure 8). Ecotypes may be identified, then, as the largest groups of strains whose phylogeny contains one significant node. On the other hand, the greater recombination rates within *N. meningitidis* [where recombination occurs four times more frequently than mutation per gene (28)] allow for greater deviation from a perfect star clade (F.M. Cohan, unpublished results) (Figure 8). An ecotype within *N. meningitidis* is expected to contain at most one or two significant nodes. Accordingly, we may tentatively identify ecotypes within *N. meningitidis* as the largest clusters whose phylogenies contain at most two significant nodes. As we shall see, there appear to be many ecotypes within each of these species and perhaps within most named species of bacteria.

Although the star clade approach produces a theory-based criterion for testing whether a set of strains are members of the same ecotype (i.e., the maximum number of nodes expected within an ecotype's phylogeny), this approach does not help us choose the groups of strains to be tested for membership within an ecotype. As we shall see, the MLST approach developed by B. Spratt and coworkers (50) produces accurate hypotheses for demarcating strains into ecotypes.

Discovery of Ecotypes Through Multilocus Sequence Typing

In MLST, strains of a named species are surveyed for partial sequences (usually ~450 bp) of seven gene loci that produce "housekeeping proteins" (proteins that are not involved in niche-specific adaptations and are presumably interchangeable between ecologically distinct groups). The evolutionary distance between strains is quantified in MLST as the number of loci that are different. Two strains are scored as different for a locus whether they differ by one nucleotide substitution or by scores of nucleotides (possibly due to a recombination event). Strains are then classified into "clonal complexes": All strains that are identical with a particular strain at five or more loci (in some cases, six or more loci) are deemed members of a clonal complex (Figure 9). The MLST website (http://www.mlst.net/new/index.htm) provides the "Burst" computer algorithm developed by E. Feil for assigning strains into clonal complexes according to criteria set by the user.

The clonal complexes defined by MLST correspond remarkably well to ecologically distinct clusters. The various hypervirulent lineages within *N. meningitidis* (1, 12, 50, 109) and within *Streptococcus pneumoniae* (36, 108) have been distinguished by MLST as separate clonal complexes. For example, one clonal complex of serogroup A in *N. meningitidis* causes pandemic and epidemic meningitis, particularly in sub-Saharan Africa; the other clonal complex within serogroup A is not associated with disease. One clonal complex of serogroup C causes localized outbreaks in primary schools, university dormitories, and prisons, where conditions are crowded; and one clonal complex of serogroup B causes disease more sporadically (1, 12, 50, 109). It is especially impressive that recombination within *N. meningitidis*, which occurs at a higher rate than in most bacteria, has not prevented

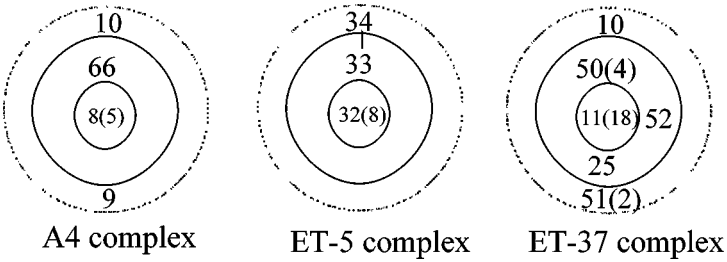

Figure 9 Demarcation of clonal complexes by MLST. The figure indicates three of the ten clonal complexes identified within *N. meningitidis* (27). In the case of *N. meningitidis*, the Burst computer algorithm by E. Feil has identified groups of strains that are identical to a "central" strain at five or more loci; in other species, a criterion of identity at six or more loci has been used (28). The numbers indicate multilocus sequence types, as listed at the MLST website. The numbers in parentheses indicate the number of strains with a given sequence type. Within the inside circle is the "central" sequence type that is identical to the rest of the clonal complex at five or more loci. In the next circle are sequence types identical to the central strain at exactly six loci. In the outside circle are sequence types identical to the central strain at five loci. The straight line indicates identity between peripheral strains at six loci. The clonal complexes identified in this way have been shown to be ecologically distinct, and the star clade approach shows almost every clonal complex to have the phylogenetic properties of an ecotype.

MLST from correctly identifying ecologically distinct groups (50). It will be interesting to see how well MLST holds up in analysis of the most frequently recombining bacteria, including *H. pylori* (90) and *N. gonorrhoeae* (66).

Why does MLST work so well? I have previously hypothesized that the clonal complexes identified by MLST are actually ecotypes, as defined here (16). Because each periodic selection event is recurrently purging the diversity within an ecotype, ecotypes are expected to accumulate little sequence diversity. It was the intuitive insight of E. Feil and B. Spratt that ecotypes would have only enough time between periodic selection events for a given strain to accumulate divergence at one or two loci out of seven, whether by mutation or recombination—this yields the 6/7 and 5/7 criteria used in MLST. Because MLST's 6/7 and 5/7 criteria are intuitively based, we should test analytically whether MLST's clonal complexes are indeed ecotypes.

Fortunately, the star clade approach can test whether the clonal complexes identified by MLST have phylogenies consistent with ecotypes, taking into account the recombination and mutation parameters estimated for the particular taxon. To this end, I have tested the strains of each of the ten clonal complexes within *N. meningitidis* for inclusion within a single ecotype. As shown in Figure 8, using recombination rates and other parameters estimated for *N. meningitidis*, an ecotype

is expected to have at most two significant nodes in its phylogeny. As it turns out, the phylogenies of all the MLST clonal complexes within *N. meningitidis* contain one or two nodes, with the exception of the ET-37 complex, which has three. Because an *N. meningitidis* ecotype is so unlikely to contain three nodes, these results suggest that the ET-37 complex contains two ecotypes.

We may conversely address whether more than one MLST clonal complex may be subsumed within a single ecotype. In all but one case, the pool of strains from different *N. meningitidis* clonal complexes contained too many nodes to fit within a single ecotype.

The same pattern has emerged from analysis of ecotypical diversity within *S. aureus*: With few exceptions the clonal complexes demarcated by MLST are each consistent with the phylogeny of one ecotype, and each pair of complexes represents more than one ecotype.

It will be interesting to determine how MLST's criterion for demarcating clonal complexes should change with the rate of recombination. Intuitively, one might expect that a taxon with extremely frequent recombination (such as *H. pylori*) might undergo changes in more than two out of seven loci between periodic selection events. It will be interesting to simulate evolution within an ecotype (using the simulation developed for the star clade algorithm) to determine how the optimal criterion for demarcating MLST complexes changes with recombination rate.

In summary, the star clade approach demonstrates that, in most cases, the clonal complexes yielded by MLST have phylogenies consistent with ecotypes, at least within *S. aureus* and *N. meningitidis*. The clonal complexes produced by MLST thus produce reliable hypotheses about the membership of ecotypes, and these hypotheses can be tested by the star clade approach. These hypotheses can of course also be tested directly by an ecological investigation of the putative ecotypes.

What is striking is that each named species studied by MLST has so many clonal complexes, 10 in the case of *N. meningitidis* (27), 26 in *S. aureus* (E. Feil, personal communication), and 28 in *S. pneumoniae*. Because each of these clonal complexes appears to be a distinct ecotype, with the universal properties of a species, a named bacterial species may actually be more like a genus than a species.

RECOMMENDATIONS FOR BACTERIAL SYSTEMATICS

A principal aim of systematics is to discover, describe, and classify the diversity of living organisms. Systematists have concluded that the basic unit of biological diversity is the species, with these quintessential properties: Species are groups of organisms whose divergence is hindered by one or more forces of cohesion; they are ecologically distinct from one another; and they are irreversibly separate. In the case of bacteria, these universal properties of species are held not by the named species of bacterial systematics but by ecotypes. These are populations of organisms occupying the same ecological niche, whose divergence is purged recurrently by natural selection. Named species appear to contain many such ecotypes.

It should not come as a surprise that named species contain this magnitude of ecological diversity. For decades, systematists have known that there is considerable variation in metabolic traits within named species (49); also, DNA-DNA hybridization experiments have demonstrated a great diversity in genomic content within named species (63). More recently the sequencing of multiple genomes within species has shown considerable variation in the genes contained (3, 70, 71, 78, 93).

We have before us an urgent but accessible goal: to characterize the ecotypical diversity within our most familiar and important named species. A comprehensive study of ecological diversity would demand no less. From a clinical point of view, identifying a pathogen with its ecotype, and thus its distinct virulence properties, will be invaluable; indeed, this was a primary motivation in developing MLST (50). Finally, from an evolutionary genetic point of view, statistical techniques for estimating dynamic properties [such as recombination (37) and migration (84) rates] from sequence survey data require that the strains sampled come from a single ecotype.

A practical first step toward identifying ecotypes is to use a universal, molecular approach to form hypotheses about ecotypical diversity within named species. The clonal complexes yielded by MLST give a good first approximation for ecotype demarcation. The star clade method can then test whether the clonal complexes obtained are consistent with ecotypes. Apparently, in most cases they will be.

Next, one can test whether the ecotypes suggested by sequence-based clustering are indeed ecologically distinct, and the ecological differences can be characterized. Subtractive hybridization is a promising method for discovering the sets of genes not shared by two ecotypes and may suggest the nature of their ecological differences (2, 45). Also, ecological differences between ecotypes can be characterized by differences in the expression levels of all the genes they share by using microarray technology (29).

Sequence-based approaches are particularly important for discovering the ecological diversity among uncultured bacteria, which we now know to constitute the great majority of bacterial diversity (18). The most discerning sequence-based methods must be utilized, or we will likely underestimate the ecological diversity among uncultured bacteria. For example, limiting sequence surveys of uncultured bacteria to 16S rRNA data would likely miss many closely related ecotypes because multilocus sequence typing has revealed multiple ecotypes within named species of cultivated bacteria, and these ecotypes are typically identical or nearly identical in their 16S rRNA sequences.

We are left, then, with the practical, taxonomic matter of classifying the ecotypical diversity within named species. I recommend that when putative ecotypes demarcated by MLST and/or the star clade approach are confirmed to be ecologically distinct, we should recognize their existence with a Latin trinomial, giving the genus, species, and ecotype name. For example, the virulent serogroup A clonal complex and ecotype of *N. meningitidis* might be named *N. meningitidis* ecotypus *africana* for its role in epidemics and pandemics in sub-Saharan Africa. The

clusters we identify in this fashion are the fundamental units of ecology and evolution. They deserve our attention and they deserve a name.

ACKNOWLEDGMENTS

I am grateful to E. Feil for sharing unpublished data about *S. aureus* and for many enlightening discussions about MLST. I am grateful to M. Dehn for help in clarifying the manuscript. This review was supported by NSF grant DEB-9815576 and research grants from Wesleyan University.

The *Annual Review of Microbiology* is online at http://micro.annualreviews.org

LITERATURE CITED

1. Achtman M, van der Ende A, Zhu P, Koroleva IS, Kusucek B, et al. 2001. Molecular epidemiology of serogroup A meningitis in Moscow, 1969 to 1997. *Emerg. Infect. Dis.* 7:420–27
2. Akopyants NS, Fradkov A, Diatchenko L, Hill JE, Siebert PD, et al. 1998. PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* 95:13108–13
3. Alm RA, Ling LL, Moir DT, King BL, Brown ED, et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176–80
4. American Society for Microbiology. 1957. *Bergey's Manual of Determinative Bacteriology*. Baltimore: Williams & Wilkins. 7th ed.
5. Atwood KC, Schneider LK, Ryan FJ. 1951. Periodic selection in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 37:146–55
6. Avise JC. 2000. *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard Univ. Press
7. Balmelli T, Piffaretti J. 1996. Analysis of the genetic polymorphism of *Borrelia burgdorferi* sensu lato by multilocus enzyme electrophoresis. *Int. J. Syst. Bacteriol.* 46:167–72
8. Baranton G, Postic D, Saint Girons I, Boerlin P, Piffaretti JC, et al. 1992. Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS461 associated with Lyme borreliosis. *Int. J. Syst. Bacteriol.* 42:378–83
9. Barrett SJ, Sneath PH. 1994. A numerical phenotypic taxonomic study of the genus *Neisseria*. *Microbiology* 140:2867–91
10. Bochner BR. 1989. Sleuthing out bacterial identities. *Nature* 339:157–58
11. Bush GL. 1994. Sympatric speciation in animals: new wine in old bottles. *Trends Ecol. Evol.* 9:285–88
12. Chan MS, Maiden MC, Spratt BG. 2001. Database-driven multi locus sequence typing (MLST) of bacterial pathogens. *Bioinformatics* 17:1077–83
13. Cohan FM. 1994. Genetic exchange and evolutionary divergence in prokaryotes. *Trends in Ecol. Evol.* 9:175–80
14. Cohan FM. 1994. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am. Nat.* 143:965–86

14a. Cohan FM. 1996. The role of genetic exchange in bacterial evolution. *ASM News* 62:631–36

15. Cohan FM. 2001. Bacterial species and speciation. *Syst. Biol.* 50:513–24
16. Cohan FM. 2002. Clonal structure: an overview. In *Encyclopedia of Evolution*, ed. M Pagel, pp. 159–61. Oxford: Oxford Univ. Press

17. Cohan FM, Roberts MS, King EC. 1991. The potential for genetic exchange by transformation within a natural population of *Bacillus subtilis*. *Evolution* 45:1393–421
18. DeLong E, Pace N. 2001. Environmental diversity of Bacteria and Archaea. *Syst. Biol.* 50:470–78
19. de Queiroz K. 1998. The general lineage concept of species, species criteria, and the process of speciation. In *Endless Forms: Species and Speciation*, ed. DJ Howard, SH Berlocher, pp 57–75. Oxford: Oxford Univ. Press
20. Dojka MA, Harris JK, Pace NR. 2000. Expanding the known diversity and environmental distribution of an uncultured phylogenetic division of Bacteria. *Appl. Environ. Microbiol.* 66:1617–21
21. Duncan KE, Istock CA, Graham JB, Ferguson N. 1989. Genetic exchange between *Bacillus subtilis* and *Bacillus licheniformis*: variable hybrid stability and the nature of species. *Evolution* 43:1585–609
22. Dykhuizen DE. 1998. Santa Rosalia revisited: Why are there so many species of bacteria? *Antonie Van Leeuwenhoek* 73:25–33
23. Dykhuizen DE, Green L. 1991. Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* 173:7257–68
24. Edwards RA, Helm RA, Maloy SR. 1999. Increasing DNA transfer efficiency by temporary inactivation of host restriction. *Biotechniques* 26:892–98
25. Eldredge N. 1985. *Unfinished Synthesis: Biological Hierarchies and Modern Evolutionary Thought*. New York: Oxford Univ. Press
26. Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect. Immun.* 69:2416–27
27. Feil EJ, Maiden MC, Achtman M, Spratt BG. 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol. Biol. Evol.* 16:1496–502
28. Feil EJ, Smith JM, Enright MC, Spratt BG. 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 154:1439–50
29. Ferea TL, Botstein D, Brown PO, Rosenzweig RF. 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. USA* 96:9721–26
30. Futuyma DJ. 1987. On the role of species in anagenesis. *Am. Nat.* 130:465–73
31. Gause GF. 1934. *The Struggle for Existence*. Baltimore: Williams & Wilkins
32. Goodfellow M, Manfio GP, Chun J. 1997. Towards a practical species concept for cultivable bacteria. In *Species: The Units of Biodiversity*, ed. MF Claridge, HA Dawah, MR Wilson, pp 25–29. London: Chapman & Hall
33. Groisman EA, Ochman H. 1997. How *Salmonella* became a pathogen. *Trends Microbiol.* 5:343–49
34. Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR. 1989. New method for generating deletions and gene replacements in *Escherichia coli*. *J. Bacteriol.* 171:4617–22
35. Hennig W. 1966. *Phylogenetic Systematics*. Urbana: Univ. Illinois Press
36. Henriques Normark B, Kalin M, Ortqvist A, Akerlund T, Liljequist BO, et al. 2001. Dynamics of penicillin-susceptible clones in invasive pneumococcal disease. *J. Infect. Dis.* 184:861–69
37. Hey J, Wakeley J. 1997. A coalescent estimator of the population recombination rate. *Genetics* 145:833–46
38. Deleted in proof
39. Hudson RR. 1987. Estimating the recombination parameter of a finite population model without selection. *Genet. Res.* 50:245–50
40. Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination

events in the history of a sample of DNA sequences. *Genetics* 111:147–64
41. Hugenholtz PB, Goebel M, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180: 4765–74
42. Deleted in text
43. Imhof M, Schlotterer C. 2001. Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc. Natl. Acad. Sci. USA* 98:1113–17
44. Johnson JL. 1973. Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria. *Int. J. Syst. Bacteriol.* 23:308–15
45. Lai YC, Yang SL, Peng HL, Chang HY. 2000. Identification of genes present specifically in a virulent strain of *Klebsiella pneumoniae*. *Infect. Immun.* 68:7149–51
46. Lan R, Reeves PR. 1996. Gene transfer is a major factor in bacterial evolution. *Mol. Biol. Evol.* 13:47–55
47. Lawrence JG, Ochman H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* 95: 9413–17
48. Linz B, Schenker M, Zhu P, Achtman M. 2000. Frequent interspecific genetic exchange between commensal Neisseriae and *Neisseria meningitidis*. *Mol. Microbiol.* 36:1049–58
49. Logan NA, Berkeley RC. 1984. Identification of *Bacillus* strains using the API system. *J. Gen. Microbiol.* 130:1871–82
50. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* 95:3140–45
51. Majewski J, Cohan FM. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* 148:13–18
52. Majewski J, Cohan FM. 1999. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* 152:1459–74
53. Mauchline WS, Keevil CW. 1991. Development of the BIOLOG substrate utilization system for identification of *Legionella* spp. *Appl. Environ. Microbiol.* 57:3345–49
54. May RM. 1973. On relationships among various types of population models. *Am. Nat.* 107:46–57
55. Maynard Smith J, Dowson CG, Spratt BG. 1991. Localized sex in bacteria. *Nature* 349:29–31
56. Maynard Smith J, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* 90:4384–88
57. Mayr E. 1944. *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York: Columbia Univ. Press
58. Mayr E. 1963. *Animal Species and Evolution.* Cambridge: Belknap Press/Harvard Univ. Press
59. Mayr E. 1982. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*, Chapter 6. Cambridge: Harvard Univ. Press
60. McKane M, Milkman R. 1995. Transduction, restriction and recombination patterns in *Escherichia coli*. *Genetics.* 139: 35–43
61. Meglitsch PA. 1954. On the nature of species. *Syst. Zool.* 3:491–503
62. Murray RGE, Stackebrandt E. 1995. Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described procaryotes. *Int. J. Syst. Bacteriol.* 45:186–87
63. Nakamura LK, Roberts MS, Cohan FM. 1999. Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: a proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov. and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. *Int. J. Syst. Bacteriol.* 49:1211–15
64. Normand P, Orso S, Cournover B, Jeannin P, Chapelon C, et al. 1996. Molecular phylogeny of the genus *Frankia* and

related genera and emendation of the family Frankiaceae. *Int. J. Syst. Bacteriol.* 46:1–9
65. Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304
66. O'Rourke M, Spratt BG. 1994. Further evidence for the non-clonal population structure of *Neisseria gonorrhoeae*: extensive genetic diversity within isolates of the same electrophoretic type. *Microbiology* 140:1285–90
67. Palmisano MM, Nakamura LK, Duncan KE, Istock CA, Cohan FM. 2001. *Bacillus sonorensis* sp. nov., a close relative of *Bacillus licheniformis*, isolated from soil in the Sonoran Desert, Arizona. *Int. J. Syst. Evol. Microbiol.* 51:16717–19
68. Palys T, Berger E, Mitrica I, Nakamura LK, Cohan FM. 2000. Protein-coding genes as molecular markers for ecologically distinct populations: the case of two *Bacillus* species. *Int. J. Syst. Evol. Microbiol.* 50:1021–28
69. Palys T, Cohan FM, Nakamura LK. 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int. J. Syst. Bacteriol.* 47:1145–56
70. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, et al. 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 404:502–6
71. Perna NT, Plunkett G 3rd, Burland V, Mau B, Glasner JD, et al. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–33
72. Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment. *Nature* 394:69–72
73. Ramsing NB, Ferris MJ, Ward DM. 2000. Highly ordered vertical structure of *Synechococcus* populations within the one-millimeter-thick photic zone of a hot spring cyanobacterial mat. *Appl. Environ. Microbiol.* 66:1038–49
74. Rao BJ, Chiu SK, Bazemore LR, Reddy G, Radding CM. 1995. How specific is the first recognition step of homologous recombination? *Trends Biochem. Sci.* 20: 102–13
75. Ravin AW. 1960. The origin of bacterial species: genetic recombination and factors limiting it between bacterial populations. *Bacteriol. Rev.* 24:201–20
76. Ravin AW. 1963. Experimental approaches to the study of bacterial phylogeny. *Am. Nat.* 97:307–18
77. Rayssiguier C, Thaler DS, Radman M. 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch repair mutants. *Nature* 342:396–401
78. Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, et al. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* 28:1397–406
79. Roberts MS, Cohan FM. 1995. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* 49:1081–94
80. Rozen DE, Lenski RE. 2000. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *Am. Nat.* 155:24–35
81. Sasser M. 1990. Identification of bacteria by gas chromatography of cellular fatty acids. *Tech. Note 101. Microbial ID, Inc.*, Newark, Del.
82. Selander RK, Musser JM. 1990. Population genetics of bacterial pathogenesis. In *Molecular Basis of Bacterial Pathogenesis*, ed. BH Iglewski, VL Clark, pp. 11–36. San Diego: Academic
83. Simpson GG. 1961. *Principles of Animal Taxonomy*. New York: Columbia Univ. Press
84. Slatkin M, Maddison WP. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123:603–13
85. Sneath PHA. 1985. Future of numerical taxonomy. In *Computer-Assisted*

Bacterial Systematics, ed. M Goodfellow, D Jones, FG Priest, pp. 415–31. Orlando: Academic
86. Sneath PHA, Sokal RR. 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. San Francisco: Freeman
87. Sokal RR, Crovello TJ. 1970. The biological species concept: a critical evaluation. *Am. Nat.* 104:127–53
88. Stackebrandt E, Goebel BM. 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.* 44:846–49
89. Stebbins GL. 1957. *Variation and Evolution in Plants*. Irvington: Columbia Univ. Press
90. Suerbaum S, Maynard Smith J, Bapumia K, Morelli G, Smith NH, et al. 1998. Free recombination within *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* 95: 12619–24
91. Swofford DL. 1998. *Phylogenetic Analysis Using Parsimony* (**And Other Methods*). Version 4. Sunderland, MA: Sinauer
92. Templeton AR. 1989. The meaning of species and speciation: a genetic perspective. In *Speciation and Its Consequences*, ed. D Otte, JA Endler, pp. 3–27. Sunderland, MA: Sinauer
93. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, et al. 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 287:1809–15
94. Trautner TA, Pawlek B, Bron S, Anagnostopoulos C. 1974. Restriction and modification in *B. subtilis*. *Mol. Gen. Genet.* 131:181–91
95. Treves DS, Manning S, Adams J. 1998. Repeated evolution of an acetate cross-feeding polymorphism in long-term populations of *Escherichia coli*. *Mol. Biol. Evol.* 15:789–97
96. Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60:407–38
97. Van Spanning RJ, Reijnders WN, Stouthamer AH. 1995. Integration of heterologous DNA into the genome of *Paracoccus denitrificans* is mediated by a family of IS1248-related elements and a second type of integrative recombination event. *J. Bacteriol.* 177:4772–78
98. Van Valen L. 1976. Ecological species, multispecies, and oaks. *Taxon* 25:233–39
99. Vulic M, Dionisio F, Taddei F, Radman M. 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc. Natl. Acad. Sci. USA* 94:9763–67
100. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, et al. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37:463–64
101. White MJD. 1978. *Modes of Speciation*. San Francisco: Freeman
102. Whittam TS, Ake SE. 1993. Genetic polymorphisms and recombination in natural populations of *Escherichia coli*. In *Molecular Paleopopulation Biology*, ed. N Takahata, AG Clark, pp. 223–45. Tokyo: Jpn. Sci. Soc. Press
103. Wiley EO. 1978. The evolutionary species concept reconsidered. *Syst. Zool.* 27:17–26
104. Woese CR. 2000. Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA* 97:8392–96
105. Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159
106. Young JPW. 1989. The population genetics of bacteria. In *Genetics of Bacterial Diversity*, ed. DA Hopwood, KF Chater, pp. 417–38. London: Academic
107. Zawadzki P, Cohan FM. 1995. The size and continuity of DNA segments integrated in *Bacillus* transformation. *Genetics* 141:1231–43
108. Zhou J, Enright MC, Spratt BG. 2000.

Identification of the major Spanish clones of penicillin-resistant pneumococci via the Internet using multilocus sequence typing. *J. Clin. Microbiol.* 38:977–86

109. Zhu P, van der Ende A, Falush D, Brieske N, Morelli G, et al. 2001. Fit genotypes and escape variants of subgroup III *Neisseria meningitidis* during three pandemics of epidemic meningitis. *Proc. Natl. Acad. Sci. USA* 98:5234–39