


January 2004

# Concepts of bacterial biodiversity for the age of genomics

Frederick M. Cohan

Wesleyan University, fcohan@wesleyan.edu

Follow this and additional works at: <http://wescholar.wesleyan.edu/div3facpubs>

 Part of the [Biodiversity Commons](#), [Ecology and Evolutionary Biology Commons](#), and the [Environmental Microbiology and Microbial Ecology Commons](#)

---

## Recommended Citation

Cohan, Frederick M., "Concepts of bacterial biodiversity for the age of genomics" (2004). *Division III Faculty Publications*. Paper 370. <http://wescholar.wesleyan.edu/div3facpubs/370>

This Contribution to Book is brought to you for free and open access by the Natural Sciences and Mathematics at WesScholar. It has been accepted for inclusion in Division III Faculty Publications by an authorized administrator of WesScholar. For more information, please contact [dschnaidt@wesleyan.edu](mailto:dschnaidt@wesleyan.edu), [ljohnson@wesleyan.edu](mailto:ljohnson@wesleyan.edu).

## Concepts of Bacterial Biodiversity for the Age of Genomics

---

Frederick M. Cohan

### INTRODUCTION

A full accounting of ecological diversity in the bacterial world will no doubt require genome-based and sequence-based approaches. Because only a small fraction of bacterial species is currently cultivable, the best hope of identifying the full complement of bacterial biodiversity is from sequence surveys of genes that can be amplified from natural bacterial habitats (1,2). In addition, genome- and sequence-based approaches are uncovering ecological diversity even within the most familiar (and cultivated) species: Ecologically distinct populations within named species are being discovered as sequence clusters even when there is ignorance of their ecology (3–5); populations are also being discovered as clusters based on the content of their genomes (6,7). Beyond discovery of ecologically distinct populations, genomic approaches promise to elucidate the ecological functioning of each community member (8,9) and how the various populations partition the environment and manage to coexist. The role that horizontal transfer has played in allowing bacteria to occupy new ecological niches (10) can be discovered. Further, the donors of horizontal transfer events (11) can be identified, and the genetic and ecological barriers to gene transfer can be discovered (12).

For each of these goals now made accessible by the genomic revolution, it is critical to identify ecologically distinct populations of strains and to recognize which strains are ecologically interchangeable and thus members of the same population. For example, consider future investigations into the role of horizontal transfer in fostering invasion of new ecological niches. A difference in the genes present in two ecologically distinct populations could represent a horizontal transfer event responsible for the populations' ecological divergence. On the other hand, horizontal transfer events that distinguish two ecologically interchangeable strains of the same population would be ecologically meaningless craters on the chromosome brought about by the meteors of horizontal transfer (13). Here, I describe recent approaches for discovering and classifying the ecological diversity of bacteria, with the aim of providing a sound ecological basis for studies of genome content and expression.

We may first turn to bacterial systematics for guidance in classifying strains into ecologically distinct groups. In the systematics of bacteria, organisms fall into clusters of

phenotypically, genetically, and ecologically similar organisms, with large gaps between clusters (14–16). In bacteria, as in other organisms, these clusters are the recognized, named species. Bacterial species were originally discerned as phenotypic clusters, usually based on presence vs absence of metabolic capabilities (see Chapter 9) (4,16). Later, as molecular techniques became available, they have been incorporated into species demarcation by calibrating each to provide the familiar clusters yielded by phenotypic analyses (4). For example, whole genome deoxyribonucleic acid (DNA)–DNA hybridization has become a principal criterion for distinguishing species; this is based on the observation that groups whose genomes anneal for 70% or more of the chromosome correspond to the phenotypic clusters of yore (17,18). More recently, 16S ribosomal RNA (rRNA) sequence similarity has demarcated species; this is based on the observation that 16S rRNA sequence divergence greater than 2.5% usually corresponds to different species, although some distinct species show less than this divergence (19).

Do these named bacterial species, discernible by phenotypic and whole genome and sequence similarity, correspond to ecologically distinct groups, and do they correspond to anything resembling species for other kinds of organisms (e.g., animals and plants)? These are the critical questions as we attempt to use sequence and genomic data to glean information about biodiversity and the ecological functioning of a microbial community.

In the world of frequently sexual higher eukaryotes, species are understood to have special dynamic properties that guarantee a high degree of homogeneity within a species, as well as heterogeneity between species (20). Most fundamentally, species have the property that genetic diversity within a species is constrained by a force of cohesion (4,21,22). In the case of the highly sexual eukaryotes, the ability to exchange genes is understood as the primary force impeding divergence (23,24). A second universal property is that different species are irreversibly separate; once populations reach a critical threshold of divergence, they become free to diverge without bound (25,26). In the case of the highly sexual eukaryotes, this threshold is most likely the point of reproductive isolation (23). Finally, while members of a single species are ecologically interchangeable, different species can coexist by partitioning resources as well as the conditions at which they thrive (27). As discussed here, these dynamic properties of species apply beyond the plants and animals to groups with peculiar sexual characteristics, including the bacteria.

There is a growing consensus among microbial evolutionary geneticists that the named species of bacterial systematics do not exhibit the special dynamic properties of species. First, bacterial systematics has demonstrated a great deal of metabolic and presumably ecological diversity within a typical species (13,28,29). Second, for decades DNA–DNA hybridization studies in systematics have revealed an enormous level of genomic diversity within named species (30), which has recently been corroborated by genomic sequencing of two or more strains from each of several species (31–36). Third, multilocus sequence typing (MLST) has revealed the existence of multiple sequence clusters within named species (3,37,38), and these are likely to correspond to ecologically distinct populations (5,37,39). Finally, natural history studies of ecological and sequence variation in the environment have revealed the existence of multiple ecologically distinct populations that are similar enough in sequence to be subsumed within a single named species (40–43).

It is thus clear that an inventory of ecological diversity in the bacterial world must go beyond counting named species. There is less consensus, however, as to whether each ecologically distinct group within a named species has the dynamic properties ascribed to species. Here, I present several contrasting, contemporary views on the nature of biological diversity: the biological species concept as applied to bacteria (39,44,45), the ecotype concept (4,46), and the species-less concept of bacterial diversity (12). Because these concepts begin with somewhat different assumptions about the nature of genetic exchange in bacteria, I review the properties of bacterial genetic exchange and the consequences of different rates of genetic exchange on bacterial population dynamics.

## THE CHARACTER OF GENETIC EXCHANGE IN BACTERIA

### *The Rarity of Genetic Exchange*

Bacteria can reproduce clonally for an indefinite number of generations, with pure clonality punctuated occasionally by recombination events, when a short segment from a donor replaces the homologous segment in a recipient. The rates of bacterial recombination in nature have been estimated by “retrospective” approaches utilizing surveys of sequence or allozyme variation in natural populations. The rationale is that low recombination rates can be inferred when alleles at different loci show high degrees of association among individuals (i.e., linkage disequilibrium) or when different DNA segments yield congruent phylogenies (47).

These retrospective approaches have shown that recombination in most bacterial species occurs in a gene segment at about the rate of mutation or somewhat higher (48–50). For example, *Staphylococcus aureus* is among the most clonal of bacterial taxa, and a gene segment undergoes recombination at a rate three times lower than mutation (51, 52); *Neisseria meningitidis* is one of the most frequently recombining bacterial taxa, and here a gene segment undergoes recombination about 3.6 times more frequently than by mutation (37).

Because a given recombination event affects many more nucleotides than does a point mutation, the rate at which recombination affects a given nucleotide can be up to 80 times greater than the rate of mutation (37). This result has fueled the notion that recombination is not really rare in bacteria, and that models of bacterial evolution that depend on rare recombination are not valid (12,45). However, recent sequence-based estimates of recombination yield essentially the same recombination rates per gene segment obtained by earlier allozyme-based approaches (i.e., with recombination occurring at a rate less than 10 times that of mutation). As discussed here, this rarity of recombination allows natural selection to have a profound effect on genetic diversity within a bacterial population.

### *Promiscuity of Genetic Exchange*

While bacteria recombine only rarely, they are not fussy about their choice of partners in genetic exchange. Bacteria can undergo homologous recombination with organisms that differ by as much as 25% in their DNA sequence (53–56).

There are, nevertheless, some important constraints on genetic exchange between divergent bacteria, including the requirement that recipient and donor share vectors of

recombination (for phage- and plasmid-mediated recombination) and microhabitats (57,58). Also, homologous recombination is limited by molecular constraints on integration of divergent donor sequences. Recombination requires the ends of the donor segment to match the recipient's homolog nearly perfectly, and this is unlikely when the donor and recipient are highly divergent (54,59,60). In addition, mismatch repair systems tend to reverse integration when they detect nucleotide mismatches between recipient and donor (55,61); mismatch repair has been shown to play a major part in sexual isolation in the Enterobacteriaceae (55,61), although not in Gram-positive bacteria (53,59).

Recombination in bacteria is not limited to the transfer of homologous segments. In heterologous recombination, bacteria can "capture" new gene loci and gene operons from other organisms, sometimes from organisms that are extremely distantly related (62). Genomic analyses have recently shown that a sizable fraction (frequently 5–10%) of genomes of bacterial species has typically been acquired from other species (63).

While horizontal transfer has clearly made a substantial mark on bacterial genomes, it should not necessarily be concluded that horizontal transfer occurs at a high rate, especially on a per capita (i.e., per individual cell) basis. Lawrence (45) estimated that successful horizontal transfer events have occurred in *Escherichia coli* at a rate of 6 to 7 per million years. Assuming even a modest population size of  $10^{12}$ , this leads to a per capita, per generation rate of successful horizontal transfer events of  $7 \times 10^{-22}$  (assuming 1 division per hour) (64). Even if horizontal transfers succeed at a rate as low as 1 in 1 million, the per capita rate of all horizontal transfer events (whether successful or unsuccessful) would be  $7 \times 10^{-16}$ .

## EVOLUTIONARY CONSEQUENCES OF RARE, BUT PROMISCUOUS, GENETIC EXCHANGE

### *Introduction of Adaptive DNA from Other Taxa*

Genetic exchange in bacteria is clearly frequent enough to effect transfer of adaptive alleles from one species to another. For example, antibiotic resistance alleles have spread across species of *Neisseria* and *Streptococcus*, replacing antibiotic-sensitive alleles through homologous recombination (65). Also, heterologous recombination has been frequent enough to have introduced hundreds of gene loci into a typical bacterial genome (63). Interspecific transfer of adaptive DNA, whether an allele or a novel operon, is the least challenging feat for recombination in that extremely low rates of recombination can accomplish it. As calculated in the preceding section, the history of horizontal transfer in *E. coli* could have been accomplished by a per capita recombination rate of  $7 \times 10^{-16}$ , even if only 1 in 1 million horizontal transfer events were adaptive. Introduction of adaptive DNA requires so little recombination simply because only a single recombination event is required to introduce an adaptive gene into a lineage; once the gene is present, natural selection can increase the abundance of the newly adaptive genotype. Moreover, if an acquired gene allows invasion of a new niche, the recombinant genotype is especially likely to be successful (12,66).

### *Effects of Recombination on Neutral Sequence Diversity*

Maiden and coworkers developed MLST, a sequence-based system for classifying strains into clusters called *clonal complexes* (3). MLST is based on sequencing genes

for seven housekeeping proteins whose diversity is assumed to be neutral in fitness; also, the proteins are assumed not to be involved in niche-specific adaptations and should be interchangeable between ecologically distinct populations. Therefore, MLST data can be used to view empirically the effect of recombination on neutral sequence diversity.

MLST has shown recombination to impact the sequence-based phylogeny of strains. For example, the sequence-based phylogeny for strains of *N. meningitidis* varies depending on the gene (67): There is no one organismal phylogeny, but each strain is truly a composite of genes from throughout the named species and from outside as well (12).

As I discussed, MLST also shows recombination has greater impact on neutral sequence diversity than mutation, owing to recombination occurring at about the rate of mutation per gene, with each recombination event involving hundreds of nucleotides or more.

There is, however, one realm in which recombination does not impact neutral sequence diversity. This is the ability to use MLST to classify strains into clonal complexes (3). Here, the evolutionary distance between strains is quantified as the number of loci that are different, with two strains scored as different for a locus whether they differ by one nucleotide substitution or by scores of nucleotides (possibly because of a recombination event). Strains are then classified into clonal complexes: All strains that are identical to a particular central strain at five or more (in some cases, six or more) of the seven loci are deemed members of a clonal complex.

In development of the MLST approach, Maiden and coworkers found that the clonal complexes obtained with 7 loci were the same as obtained with up to 11 loci, and the subset of 7 loci chosen did not affect the classification of strains into complexes (3). While a minority of loci may be recombinant in any given strain, the complexes appear to be a robust signal classifying the diversity of strains. Even *N. meningitidis*, the most frequently recombining of the species studied, yields robust clonal complexes. As will be discussed the MLST clonal complexes empirically correspond to ecologically distinct groups, and there is a theoretically compelling reason for this correspondence.

### *Effect of Recombination on Adaptive Divergence Between Ecologically Distinct Populations*

In the highly sexual world of animals and plants, divergence between two closely related populations requires sexual isolation between the populations (23). If recombination between animal or plant populations were to proceed at the same rate as within populations, recombination would rapidly eliminate any adaptive divergence between populations.

In contrast, because recombination in bacteria is so rare, recurrent recombination between bacterial species cannot hinder their divergence (46). Suppose, for example, that two populations are ecologically specialized on different substrates, and that alleles at several loci are responsible for the adaptive divergence. This model implies a cost of recombination: If the multilocus genotype *ABCDE* confers one population's adaptations, and genotype *abcde* confers the other population's adaptation, then the fitness of a recombinant genotype at these critical niche-determining loci (e.g., *Abcde*) would be reduced to  $1 - s$  (where  $s$  is the intensity of selection against recombinants).

I have previously shown that the equilibrium frequency of a maladaptive foreign allele is  $c_b/s$ , where  $c_b$  is the rate of recombination between populations (46). Given that the

rate of recombination within bacterial populations is already quite rare ( $c_w \approx 10^{-6}$ , per gene segment), the frequency of maladaptive foreign alleles is expected to be negligible, even if the rate of recombination between populations were as high as recombination within populations. (A similar argument applies if the basis of ecological divergence is the presence vs absence of horizontally transferred gene loci.) Thus, 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 adaptive divergence in the bacterial world (4,46).

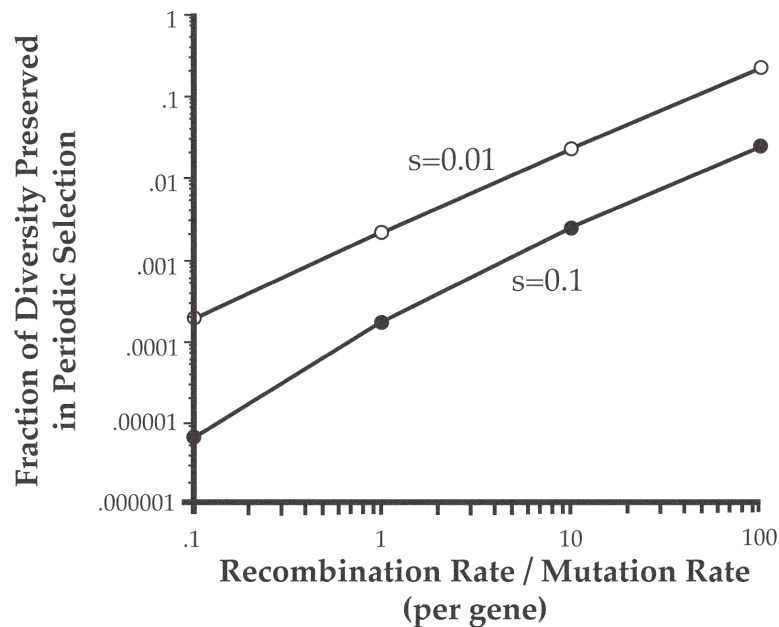
### *Effect of Recombination on Diversity Within a Population*

It has long been understood that natural selection will purge all genetic diversity from an entirely asexual population in a process known as *periodic selection* (68). In the absence of recombination, any adaptive mutant and its clonal descendants eventually replace the other cells of the population, thus extinguishing genetic diversity at all loci.

Frequent recombination can clearly quash the diversity-purging effect of periodic selection. If the adaptive mutation recombines into another genetic background, then the entire genome of the recipient is saved from extinction; alternatively, if a segment from a strain lacking the adaptive mutation should recombine into a strain with the adaptive mutation, then that segment (only) will be saved from extinction. This quashing of periodic selection is the most difficult challenge for recombination: Extremely frequent recombination, with recombination nearly obligately tied to sex, is required to prevent the purging of diversity (69,70) (Fig. 1). When the intensity of periodic selection is strong (i.e., fitness advantage for the adaptive mutation is 10%), each bout of periodic selection purges nearly all diversity within a population. Over recombination rates typically observed in nature (from 0.3 to 3.6 times the mutation rate), periodic selection purges all but 0.001–0.2% of the sequence diversity (Fig. 1). Over more modest selection intensity (i.e., fitness advantage of 1%), periodic selection purges all but 0.02–2% of sequence diversity over naturally occurring recombination rates. Thus, even if recombination rates in bacteria were orders of magnitude greater than estimated, recombination would be ineffective in quashing the diversity-purging effects of periodic selection.

Lawrence (45) argued that periodic selection does not have a significant role in reducing genetic diversity in bacterial populations. His argument is based in part on a sequence survey by Guttman and Dykhuizen that demonstrated a periodic selection event in *E. coli* (71). Guttman and Dykhuizen demonstrated that a segment of at least 30 kb near *gapA* was anomalously homogeneous throughout *E. coli*, whereas the rest of the chromosome showed much greater diversity. Lawrence claimed that this observation proves periodic selection has an insignificant effect on bacterial diversity genomewide, much like a selective sweep within an animal population, owing to rampant recombination.

Majewski and I previously pointed out that Guttman and Dykhuizen's (71) original interpretation of the periodic selection is flawed because it assumes that one adaptive mutant (or recombinant) would out-compete all other variants in *E. coli* (72). This is clearly impossible owing to what both bacterial systematists (16) and evolutionary geneticists (12,73,74) understand about the tremendous ecological diversity in *E. coli* or in most other named species: Because a named species contains strains adapted to distinct niches, one adaptive mutant could not out-compete the rest of the species' diversity. Also, as noted, theory shows that periodic selection would purge nearly all of a bacterial



**Fig. 1.** The relationship between recombination rate and the diversity-purging effect of periodic selection over different intensities of selection  $s$  favoring the adaptive mutation. The ratios of recombination rate to mutation rate seen in the figure reflect the range of ratios typically observed in nature. The ordinate is based on the mean fraction of a cell's genome at the end of periodic selection that is not descended from the genome of the original mutant. These results are based on a Monte Carlo simulation. (Used with permission from ref. 70, Landes Bioscience.)

population's diversity, even if recombination was substantially more frequent than typically observed (69,70) (Fig. 1). The correct interpretation of Guttman and Dykhuizen's (71) data requires a more nuanced view of ecological diversity in bacteria, as I describe next.

## MODELS OF ECOLOGICAL DIVERSITY IN BACTERIA

Bacterial systematics has largely been based on demarcation of species as phenotypic and genetic clusters (16), but the last decade has seen the introduction of several concepts of bacterial diversity based on the dynamics of bacterial evolution.

### *Species Concepts Based on Genetic Exchange and Sexual Isolation*

The biological species concept of Mayr (23,24) may be credited for infusing evolutionary theory into systematics. The biological species concept changed zoologists' and botanists' views of what a species should represent: A species is not merely a cluster of similar organisms, but rather a fundamental unit of ecology and evolution, with certain dynamic properties. In the case of Mayr's biological species concept, a species is viewed as a group of organisms whose divergence is opposed by recombination between them (21–23).



Dykhuizen and Green (39) proposed classifying bacteria into species according to the biological species concept, demarcating bacterial species as groups of strains that recombine within the group, but not with strains from other such groups. They suggested a phylogenetic approach to demarcation; the phylogeny of members of a single species would be expected to vary from gene to gene, while the phylogeny of members of different species would be congruent across genes.

This approach has been criticized for not taking into consideration the intrinsic promiscuity of genetic exchange in bacteria (4,12,46): Bacteria do exchange genes both within and between the clusters recognized as named species (12,65,75–77). Generally, phylogenetic evidence for recombination is more common within named bacterial species than between them, but only if the donor species are not included in the phylogeny (12). When donor species are included, horizontally transferred alleles become evident in a phylogeny of multiple species.

Two recent concepts of bacterial diversity explain the evolutionary origin of sexual isolation in bacteria. In Lawrence's (45) "speciation without species" model begins with two populations that are ecologically distinct owing to modest differences in their sets of acquired genes. Any interpopulation genetic exchange that includes a gene involved in the populations' adaptive divergence is strongly selected against. Thus, regions of the chromosome near these genes are protected from the homogenizing effects of genetic exchange, and neutral sequence divergence will be allowed to accumulate in these regions. This in turn reduces the efficiency of successful recombination in these regions (44,54). As each population evolves further into its respective niche, more genetic changes accumulate throughout the chromosome, and each such change protects its neighboring region of the chromosome. Eventually, the entire chromosome is protected from recombination, and the consequent sequence divergence impedes genetic exchange anywhere on the chromosome.

I believe that Lawrence's model is correct in that any gene responsible for adaptive divergence will lead to lower successful genetic exchange in the flanking region and will thereby accelerate neutral sequence divergence. However, it is not clear that this local prevention of genetic exchange is necessary for eventual sequence divergence between ecologically distinct populations. This is because neutral sequence divergence between populations is intrinsically a self-accelerating process in which any random neutral sequence divergence (in any part of the genome, whether "protected" from recombination or not) tends to increase sexual isolation in that region, and this increased sexual isolation results in lower recombination, which further increases sequence divergence, and so on. I have previously shown that the positive feedback between sequence divergence and sexual isolation in *Bacillus* results eventually in unbounded neutral sequence divergence between ecologically distinct populations (78). Nevertheless, I agree with Lawrence (45) that the accumulation of niche-specific genes throughout the chromosome will speed up the process.

The "molecular keys to speciation" concept notes that the adaptive changes allowing a niche invasion can be promoted by modulation of the mismatch repair system (44,55). Because mismatch repair is the primary agent of sexual isolation in some bacteria (e.g., the Enterobacteriaceae) (55), a debilitated mismatch repair system more readily allows uptake of potentially adaptive genes from other species. However, once a population becomes adapted to a new environmental challenge, mismatch repair deficien-

cies can be disadvantageous because they yield high mutation rates (79). The population may then regain a functional mismatch repair system, often by horizontal transfer (80). In the time that an incipient species is lacking a fully functional mismatch repair system, it rapidly acquires sequence divergence from its parental population, and this acquired sequence divergence contributes to sexual isolation when mismatch repair is reinstated (44).

The biological species concept motivates these dynamic models for the evolution of sexual isolation, as well as Dykhuizen and Green's (39) plan to classify strains by their recombination history: If groups of bacteria that recombine at a high rate can be identified, then species can be identified; if the accumulation of sexual isolation between populations can be understood, then the origins of species can be understood. However, as I have discussed here, recurrent recombination between bacterial species cannot hinder their divergence (12,46). At the moment that two populations acquire ecologically distinct traits, there is already too little recombination between them to threaten the integrity of their separate adaptations. Moreover, recombination cannot prevent the accumulation of further niche-specific adaptations in each. The evolution of sexual isolation is irrelevant to the evolution of permanent divergence in the bacterial world, so the biological species concept is a red herring for bacterial systematics (4).

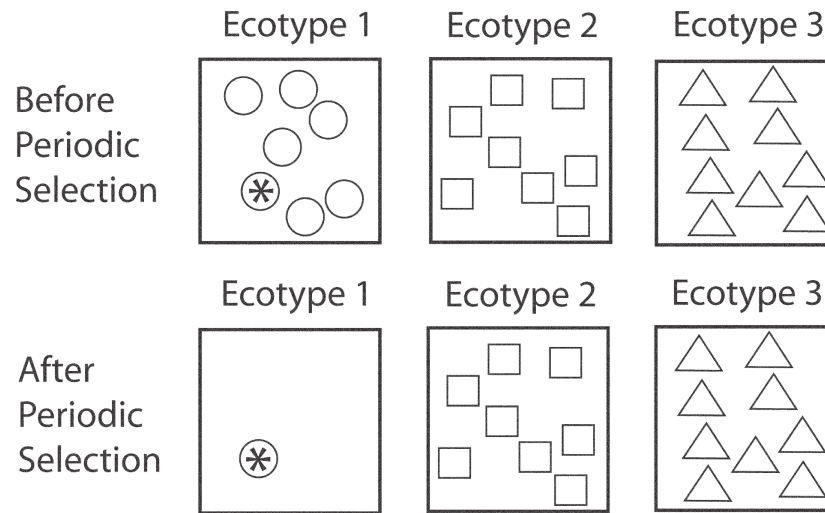
While accurately predicting the course of increase in sexual isolation, the models of Lawrence (45) and Vulic et al. (44) are not necessary for understanding the origin of bacterial species. As I discuss next, the quintessential step in the origin of bacterial species is the ecological divergence of populations.

### *Species Concept Based on Periodic Selection*

I have previously defined a bacterial "ecotype" with respect to the fate of an adaptive mutant (or recombinant): An *ecotype* is a set of strains using the same or similar ecological niche, such that an adaptive mutant from within the ecotype out-competes to extinction all other strains from the ecotype; an adaptive mutant cannot, however, drive to extinction strains from other ecotypes (4,46,81) (Fig. 2). Thus, an ecotype is the set of strains whose diversity is purged through periodic selection favoring each adaptive mutant. Periodic selection is expected to be a powerful force of cohesion within a bacterial ecotype in that it recurrently resets the genetic diversity to near zero.

When two populations become ecologically distinct, they may each undergo their own private periodic selection events. At this point, natural selection favoring an adaptive mutant purges the diversity only within the mutant's own population. This is a milestone toward forming new species; such populations are now irreversibly separate because periodic selection cannot prevent further divergence and, as I have explained, neither can recombination (46).

Bacterial ecotypes, as defined by the domains of periodic selection, are expected to share the fundamental properties of species: Ecotypes are each subject to an intense force of cohesion, periodic selection, which recurrently purges diversity within an ecotype; divergence between ecotypes is irreversible; ecotypes are expected to form distinct phenotype- and sequence-based clusters (as discussed in the next section on predictions); and bacterial ecotypes are ecologically distinct (4,81). According to the ecotype concept, a true species in the bacterial world may be understood as an evolutionary lineage bound together by ecotype-specific periodic selection (4).



**Fig. 2.** Periodic selection purges the diversity within, but not between, ecotypes. Each individual symbol represents an individual organism, and the distance between symbols represents the sequence divergence between organisms. The asterisk represents an adaptive mutation in one individual of ecotype 1. Once populations are divergent enough to escape one another's periodic selection events (as separate ecotypes), those populations are free to diverge permanently. (Used with permission from ref. 81a.)

### *Bacterial Diversity Without Species*

Finally, in one view, bacterial taxa with the dynamic attributes of species may not exist (12). This model assumes that recombination is too frequent to allow purging of diversity by periodic selection, such that there is no force of cohesion within a bacterial population. Without cohesion, there are no species (4,12,22). Gogarten et al. (12) suggest that, owing to frequent recombination, sequence data will not be a reliable marker for ecologically distinct populations, and sequence data will not provide a true indication of the phylogenetic relationships among such populations.

## **PREDICTIONS OF THE ECOTYPE AND SPECIES-LESS CONCEPTS OF DIVERSITY**

### *Predictions on Correspondence Between Ecotypes and Sequence Clusters*

The species-less concept makes no specific predictions about the correspondence between sequence diversity and ecologically distinct populations, only that recombination will make sequence diversity an unreliable indicator of an organism's history and ecological characteristics (12). In contrast, the ecotype concept provides a rationale for sequence-based discovery of bacterial diversity (4,5). Because periodic selection purges diversity within, but not between, ecotypes, each bacterial ecotype is expected eventually to be identifiable as a sequence cluster, distinct from all closely related ecotypes. Palys et al. (5) showed that, under the recombination rates typical of bacteria, the average sequence divergence between ecotypes should greatly exceed that within

ecotypes. Nevertheless, recombination may result in misdiagnosis of an occasional strain to a donor's ecotype, especially if strain diagnosis is based on the sequence of a single gene.

The MLST approach of Maiden and coworkers (3) appears to yield a more robust approach to ecotype discovery and classification than the method of Palys et al. (5), and as will be discussed here, it also is based on the diversity-purging effects of periodic selection. These clonal complexes generated by MLST are robust with respect to recombination in that they are only rarely affected by the choice of loci (3).

I previously hypothesized that the clonal complexes of MLST are ecotypes, and that the diversity-purging effect of periodic selection causes a correspondence between clonal complexes and ecotypes (4,49,50). Because periodic selection is recurrently purging the diversity within an ecotype, ecotypes are expected to accumulate only a limited level of sequence diversity between periodic selection events. It may be speculated (and later tested) that ecotypes typically have only enough time between periodic selection events to accumulate divergence at 1 or 2 loci, at most, of 7, whether by mutation or recombination. This would justify MLST's 5/7 and 6/7 criteria for including strains within a clonal complex. In general, it would be expected that frequently recombining bacteria would diverge at more loci between periodic selection events, compared to rarely recombining bacteria, for which nearly all divergence accumulates simply through mutation. In any case, the hypothesis that MLST clusters correspond to ecotypes appears reasonable and should be rigorously tested.

In contrast, the species-less concept of diversity does not acknowledge the diversity-purging effect of periodic selection and so denies any mechanism for ecologically distinct populations to be visualized as sequence clusters. No mechanism is provided even for the existence of multiple discrete sequence clusters within named species.

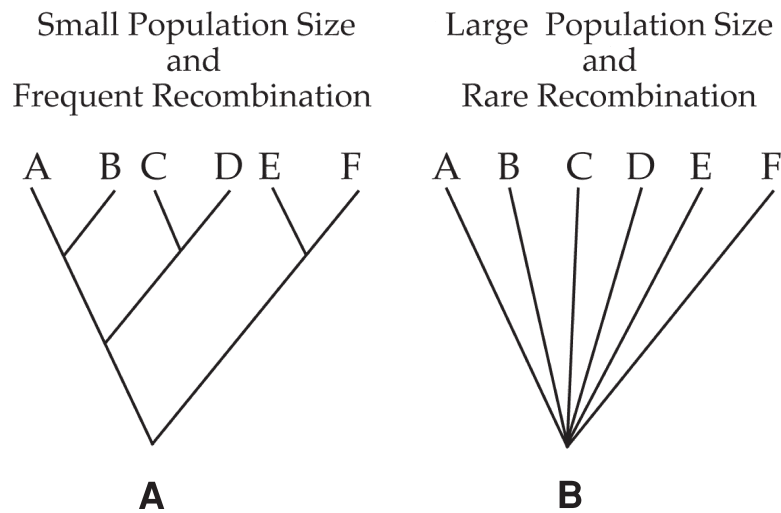
Consider next how one can rigorously test whether MLST clonal complexes correspond to ecotypes, each dominated by periodic selection, or whether these groups are ecologically meaningless, as expected under the species-less concept.

### *Predictions Regarding Ecological Distinctness*

The ecotype concept predicts that sequence clusters should be ecologically distinct (with the caveat of genotypes, discussed in a later section). One test of this prediction is that each sequence cluster might be associated with a different microhabitat. Indeed, several studies have shown that very closely related sequence clusters form a series of discrete populations occupying different ecological niches (40,41,43). For example, Ramsing et al. (41) found that very closely related sequence clusters of *Synechococcus* in Yellowstone's hot springs are distributed at different depths, with different light conditions.

Another direct test of ecological distinctness, at least for pathogens, is that the various putative ecotypes should have different disease-causing properties. MLST was originally designed for the purpose of diagnosing unknown strains of pathogenic species into ecologically distinct populations; indeed, several of the MLST clusters have been shown to correspond to populations with different virulence and transmission characteristics (3).

Also, genomic approaches can test whether putative ecotypes are ecologically distinct. Using either subtractive hybridization or microarray technology, strains can be



**Fig. 3.** The phylogenetic signatures of populations with diversity that is controlled by periodic selection vs genetic drift. (Used with permission from ref. 4.)

assayed for the sets of genes they contain. The ecological distinctness of putative ecotypes will be supported when members of the same putative ecotype tend to share nearly all their genes, and members of different putative ecotypes share many fewer genes (6). For example, Salama et al. (7) found that strains of *Helicobacter pylori* resolved clearly into clusters on the basis of the genes they contain. Unfortunately, these genome-content clusters cannot be compared to putative ecotypes demarcated by MLST clonal complexes because *H. pylori* is the one taxon known to recombine at too high a rate to yield robust clonal complexes (82). In addition, messenger ribonucleic acid (mRNA) microarray approaches can address whether different putative ecotypes have different patterns of gene expression for the genes they share (13,83). These genomic approaches are particularly promising in that they should yield information about the nature of the ecological differences between ecotypes (6).

### **Phylogenetic Predictions**

Sequence diversity within an ecotype is assumed to be limited largely by periodic selection and not by genetic drift, so nearly all strains randomly sampled from an ecotype should trace their ancestries directly back to the adaptive mutant that caused and survived the last selective sweep (4,70). Thus, the phylogeny of an ecotype should be consistent with a star clade with only one ancestral node, such that all members of an ecotype are equally closely related to one another (Fig. 3). In contrast, a population with sequence diversity that is limited by genetic drift would have a phylogeny with many nodes.

In a strictly asexual ecotype, a sequence-based phylogeny would yield a perfect star clade, with only minor exceptions due to homoplasy (convergent nucleotide substitutions in different lineages) (4). However, in an ecotype subject to modest rates of recombination, particularly with other ecotypes, the sequence-based phylogeny can deviate significantly from a perfect star. Libsch and I have developed a computer simulation (Star) to determine how closely a phylogeny based on multilocus sequencing should

resemble a star clade (4,84). Taking into account a taxon's mutation and recombination parameters, the Star simulation determines the likelihood that the phylogeny of strains from a single ecotype would have only one significant node (i.e., a perfect star) vs two, three, four, or more significant nodes.

It was found that, for *S. aureus*, which is among the least frequently recombining bacteria (51,52), an ecotype should almost never have more than one node (4,84). In *N. meningitidis*, which is among the most frequently recombining bacteria (37), the phylogeny of an ecotype is expected to have either one or two significant nodes, but almost never three or more.

Let us consider how well the clonal complexes of MLST fit Star's phylogenetic predictions. I have tested whether the phylogenies of each of the 10 clonal complexes found within *N. meningitidis* were consistent with the Star simulation's expectations for a single ecotype (4). The phylogenies of all but 1 clonal complex were found to contain 1 or 2 nodes, as expected given this taxon's recombination parameters. Similarly, all but 1 of the 26 clonal complexes within *S. aureus* contained only 1 significant node, consistent with the expectation for a single ecotype in this taxon. The pooled strains of most pairs of clonal complexes contained 2 significant nodes, indicating that pairs of complexes represented 2 ecotypes. However, 3 exceptional clonal complexes, when pooled together, contained only 1 significant node among them, suggesting that they are members of the same ecotype. I previously argued that, in *S. aureus*, a more stringent criterion for inclusion within a clonal complex (e.g., 6/7 identical loci) might yield a more perfect match between clonal complexes and the phylogenetic expectation for an ecotype (4).

In summary, the MLST clonal complexes appear to pass two tests of their correspondence with ecotypes: At least some are known to be ecologically distinct, and the phylogenies of the clonal complexes are generally consistent with the expectations for a single ecotype.

### ***Prediction of Private Periodic Selection Within Each Ecotype***

The ecotype concept predicts that each putative ecotype should show evidence of having undergone its own private periodic selection events. To explore tests of this prediction, let us begin with Guttman and Dykhuizen's (71) sequence-based evidence for periodic selection within *E. coli*: Most genes fell into four major sequence clusters, but in the chromosomal region near *gapA*, all strains were anomalously homogeneous in sequence. As I have discussed, Guttman and Dykhuizen's (71) interpretation that an adaptive mutant (or recombinant) out-competed all other variants within *E. coli* is unlikely. The strains of *E. coli* are ecologically diverse, so a single adaptive mutant within *E. coli* could not out-compete the entire species; moreover, even if all of *E. coli* strains were a single ecotype, periodic selection would be expected to purge nearly all diversity over the entire chromosome (Fig. 1).

Majewski and I (72) proposed the adapt globally, act locally model to explain anomalous homogeneity around a small chromosomal region, as seen in *gapA* of *E. coli*. We proposed that there may be multiple ecotypes within *E. coli* (perhaps corresponding to the four major sequence clusters, or smaller subclusters), and that the adaptive mutation around *gapA* was generally useful for all of the ecotypes of *E. coli*. We proposed that the adaptive mutation first caused a purging of diversity within its original ecotype and was

then passed by recombination into other ecotypes, precipitating a “local” purging of diversity within each recipient ecotype.

The adapt globally, act locally model is so named because the adaptive *mutation* (i.e., the allele) improves fitness in all ecotypes, but the adaptive *mutant* (i.e., the cell) out-competes only members of its own ecotype (72). Thus, for genes closely linked to the adaptive mutation (which are transferred between ecotypes along with the adaptive mutation), there would be a nearly total purging of diversity both within and between ecotypes, but for genes not linked to the adaptive mutation, selection would purge only the diversity within ecotypes. Whenever a small chromosomal segment is homogenized across strains that otherwise form distinct clusters, the model predicts that a generally useful adaptive mutation has passed from ecotype to ecotype, causing local periodic selection in each.

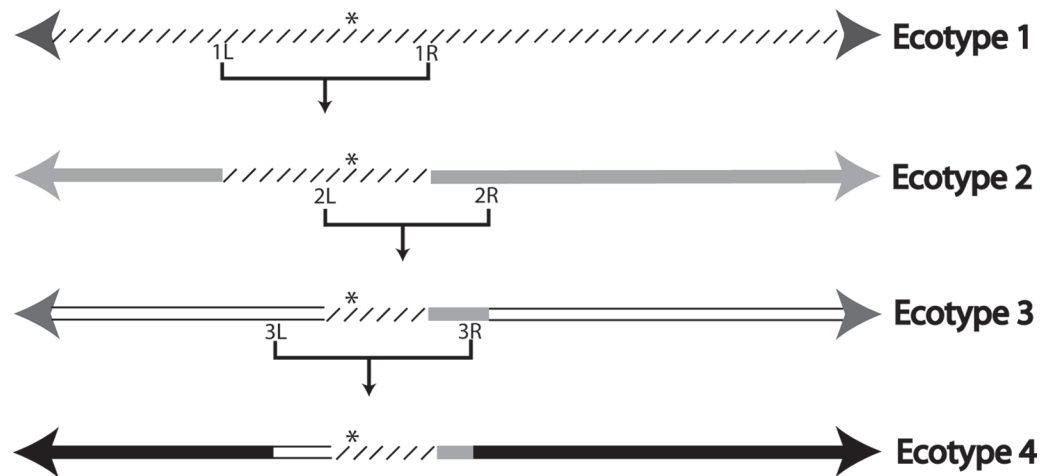
I have recently proposed a genomic approach for verifying the adapt globally, act locally model (70). While Guttman and Dykhuizen’s (71) discovery of a periodic selection event was based on a serendipitous choice of loci to survey, today comparisons of whole genomes should readily yield Guttman–Dykhuizen islands of anomalous homogeneity flanked by regions that fall into discrete sequence clusters. The adapt globally, act locally model predicts that the adaptive allele driving the periodic selections in all of the ecotypes is passed to each ecotype in a separate recombination event. Therefore, the region that is homogenized should be somewhat different for each pair of ecotypes, reflecting the junctions of the recombination events that transferred the adaptive mutation across ecotypes (70) (Fig. 4). It thus may be predicted that, if MLST clonal complexes correspond to ecotypes, the junctions of the homogeneous region will be unique for each pair of sequence clusters. Such a result would confirm that the clonal complexes are indeed separate ecotypes, having undergone their own periodic selection events, albeit caused by the same allele.

The species-less concept of diversity makes no specific prediction for the junctions of homogeneity in periodic selection (12). It is possible that, because recombination is presumed to be so frequent, many of the surviving lineages would obtain the adaptive mutation in a separate recombination event, so every pair of strains sampled within a named species may have unique junctions of homogeneity.

### *Distinguishing Ecotypes From Geotypes*

Finally, in any sequence-based method of discovering and classifying ecological diversity, care must be taken to ensure that distinct clusters are not the result of geographic separation (5,70,85). Geographically isolated populations that are ecologically identical could diverge into separate sequence clusters (termed *geotypes* by Papke et al. in ref. 85). It is sometimes difficult to rule out the geotype hypothesis even when bacterial sequence clusters are sampled from the same geographic region. This is because geotypes from different regions may have migrated only recently into the same region (e.g., aided by human transport); the various geotypes, now living in the same place, may not yet have endured a periodic selection event that would purge diversity throughout all geotypes within the ecotype (70).

One possible approach to ruling out the geotype hypothesis is to find that one adaptive mutation has precipitated separate periodic selection events within each putative ecotype, as I have just described (70). This would indicate that a single periodic selec-



**Fig. 4.** In the adapt globally, act locally model, the region that is homogenized is expected to differ between each pair of ecotypes. The adaptive mutation (indicated by an asterisk) originally occurs in ecotype 1. After the selective sweep, a small region of chromosome around the adaptive mutation (between 1L and 1R) enters ecotype 2, causing a selective sweep in that ecotype. Then, a segment around the adaptive mutation enters ecotype 3 and causes a selective sweep there, and so on. The entire ensemble of ecotypes becomes homogenized for the sequence near the adaptive mutation, but the boundaries of the homogeneous region differ for each pair of ecotypes. (Used with permission from ref. 70.)

tion event cannot extinguish the diversity among putative ecotypes, and that the putative ecotypes are indeed ecologically distinct.

#### *Can Recombination Destroy the Molecular Signal of Ecologically Distinct Populations?*

The ecotype concept predicts that sequence clusters are separate ecotypes, which are expected to be ecologically distinct, to resemble a star phylogeny, and to have undergone their own private periodic selection events. The first two predictions have been tested successfully to some extent and appear to corroborate the ecotype concept. The last prediction will have to await application of the genomic approach that I have outlined.

The species-less concept of diversity may be correct in its assertion that attempts to discover the true phylogeny of closely related organisms and ecologically distinct populations are futile (12). Gogarten et al.'s (12) quotation from Feil et al. (67) states their case well: "Over the long term, the impact of relatively frequent recombination is to obliterate the phylogenetic signal in gene trees such that the relationships between many lineages of many bacterial species should be depicted as a network rather than a tree." However, the species-less concept appears incorrect in extrapolating that recombination can also obliterate the sequence-based signal of ecologically distinct populations. What Gogarten et al. (12) do not mention is that the clonal complexes derived from MLST are robust with respect to the choice of genes and thus to the effects of recombination, even in frequently recombining bacteria such as *N. meningitidis* (3). That is, ecotypes and the sequence clusters they generate are stable with respect to recombination,



even when the phylogenetic relationships among ecotypes are not. This is why sequence diversity can help us identify ecologically distinct populations.

### RECOMMENDATIONS FOR THE SCIENCE OF GENOMICS

With thanks to genomics, bacteriologists are now poised to make enormous progress in discovering and characterizing bacterial diversity: Ecological diversity can be discovered in terms of the sets of genes that are contained within the genome and in terms of the expression levels of genes that are shared; the genes responsible for invasions of new niches can even be identified, as can the donor sources of each gene (6). However, before embarking on this great adventure, we should make a point of focusing on those genomic differences that determine differences in ecological niche. Efforts should be made, therefore, to try to distinguish strains that are ecologically distinct from those that are not.

To this end, genomic approaches should be used to test for the existence of ecotypes, as defined here, and to hone sequence-based approaches to discovering ecotypes. I suggest that we first assign strains to putative ecotypes based on multilocus sequence clustering (using MLST) and then test whether these putative ecotypes fit the ecological, phylogenetic, and genomic expectations of a single ecotype. Tests up to this point have shown good correspondence between MLST clonal complexes and ecotypes, and further tests will be important. There are several important contributions that genomics can make toward validating the ecotype concept: use of microarray approaches to show that genomic patterns of gene expression differ between putative ecotypes, use of subtractive hybridization and microarrays to identify gene acquisitions that might suggest the nature of ecological differences between putative ecotypes, and genomic comparisons to show that putative ecotypes have undergone their own periodic selection events.

How should genomics proceed to characterize diversity once we are confident that sequence-based approaches can successfully classify organisms into ecotypes, and that there are many ecotypes contained within a typical named species? The ecotype concept should allow organization of the apparent chaos of genomic diversity within a named species by shifting the focus from the diversity of genes among strains of a named species (73,74) to the genomic differences among ecotypes. It should be recognized that variation in gene content and gene expression within a putative ecotype is likely to represent merely the random changes occurring within a population between periodic selection events. However, any genomic differences that correspond to different ecotypes could represent the changes responsible for invasion of new ecological niches and sustained coexistence among populations. These are the genomic differences that matter, and these are the differences that should demand attention.

### ACKNOWLEDGMENTS

I am grateful to Michael Dehn for many suggestions that improved the clarity of the manuscript. This research was supported by National Science Foundation grant DEB-9815576 and by research funds from Wesleyan University.

### REFERENCES

1. DeLong, E, Pace N. Environmental diversity of bacteria and archaea. *Syst Biol* 2001; 50:470–478.
2. Ward DM, Weller R, Bateson MM. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 1990; 345:63–65.

3. Maiden MC, Bygraves JA, Feil E, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998; 95:3140–3145.
4. Cohan FM. What are bacterial species? *Annu Rev Microbiol* 2002; 56:457–487.
5. Palys T, Nakamura LK, Cohan FM. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* 1997; 47:1145–1156.
6. Joyce EA, Chan K, Salama NR, Falkow S. Redefining bacterial populations: a post-genomic reformation. *Nat Rev Genet* 2002; 3:462–473.
7. Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci USA* 2000; 97:14668–14673.
8. Hihara Y, Kamei A, Kanehisa M, Kaplan A, Ikeuchi M. DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. *Plant Cell* 2001; 13:793–806.
9. Harrington CA, Rosenow C, Retief J. Monitoring gene expression using DNA microarrays. *Curr Opin Microbiol* 2000; 3:285–291.
10. Nesbo CL, Nelson KE, Doolittle WF. Suppressive subtractive hybridization detects extensive genomic diversity in *Thermotoga maritima*. *J Bacteriol* 2002; 184:4475–4488.
11. Sandberg R, Winberg G, Branden CI, Kaske A, Ernberg I, Coster J. Capturing whole-genome characteristics in short sequences using a naive Bayesian classifier. *Genome Res* 2001; 11:1404–1409.
12. Gogarten JP, Doolittle WF, Lawrence JG. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 2002; 19:2226–2238.
13. Feldgarden M, Byrd N, Cohan FM. Gradual evolution in bacteria: evidence from *Bacillus* systematics. *Int J Syst Bacteriol* 2003; 149:3565–3573.
14. Sneath PH. Future of numerical taxonomy. In: Goodfellow M, Jones D, Priest F (eds). *Computer-Assisted Bacterial Systematics*. Orlando, FL: Academic, 1985, pp. 415–431.
15. Rossello-Mora R, Amann R. The species concept for prokaryotes. *FEMS Microbiol Rev* 2001; 25:39–67.
16. Goodfellow M, Manfio GP, Chun J. Towards a practical species concept for cultivable bacteria. In: Claridge MF, Dawah HA, Wilson MR (eds). *Species: The Units of Biodiversity*. London: Chapman and Hall, 1997.
17. Johnson J. Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria. *Int J Syst Bacteriol* 1973; 23:308–315.
18. Wayne LG, Brenner DJ, Colwell RR, et al. Report of the Ad Hoc Committee on reconciliation of Approaches to Bacterial Systematics. *Int J Syst Bacteriol* 1987; 37:463–464.
19. Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA:DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* 1994; 44:846–849.
20. de Queiroz K. The general lineage concept of species, species criteria, and the process of speciation. In: Howard DJ, Berlocher SH (eds). *Endless Forms: Species and Speciation*. Oxford, UK: Oxford University Press, 1998, pp. 57–75.
21. Meglitsch P. On the nature of species. *Syst Zool* 1954; 3:491–503.
22. Templeton A. The meaning of species and speciation: a genetic perspective. In: Otte D, Endler J (eds). *Speciation and Its Consequences*. Sunderland, MA: Sinauer, 1989, pp. 3–27.
23. Mayr E. *Animal Species and Evolution*. Cambridge, MA: Belknap Press of Harvard University Press, 1963.
24. Mayr E. *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York: Columbia University Press, 1944.
25. Simpson G. *Principles of Animal Taxonomy*. New York: Columbia University Press, 1961.

26. Wiley E. The evolutionary species concept reconsidered. *Syst Zool* 1978; 27:17–26.
27. Eldredge N. *Unfinished Synthesis: Biological Hierarchies and Modern Evolutionary Thought*. New York: Oxford University Press, 1985.
28. Yohalem DS, Lorbeer JW. Intraspecific metabolic diversity among strains of *Burkholderia cepacia* isolated from decayed onions, soils, and the clinical environment. *Antonie Van Leeuwenhoek* 1994; 65:111–131.
29. Logan NA, Berkeley RC. Identification of *Bacillus* strains using the API system. *J Gen Microbiol* 1984; 130(Pt 7):1871–1882.
30. Lan R, Reeves PR. Gene transfer is a major factor in bacterial evolution. *Mol Biol Evol* 1996; 13:47–55.
31. Edwards RA, Olsen GJ, Maloy SR. Comparative genomics of closely related Salmonellae. *Trends Microbiol* 2002; 10:94–99.
32. Parkhill J, Achtman M, James KD, et al. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 2000; 404:502–506.
33. Perna NT, Plunkett G 3rd, Burland V, et al. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 2001; 409:529–533.
34. Read TD, Brunham RC, Shen C, et al. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res* 2000; 28:1397–1406.
35. Tettelin H, Saunders NJ, Heidelberg J, et al. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 2000; 287:1809–1815.
36. Alm RA, Ling LS, Moir DT, et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 1999; 397:176–180.
37. Feil EJ, Maiden MC, Achtman M, Spratt BG. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol Biol Evol* 1999; 16: 1496–1502.
38. Feil EJ, Smith JM, Enright MC, Spratt BG. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 2000; 154:1439–1450.
39. Dykhuizen DE, Green L. Recombination in *Escherichia coli* and the definition of biological species. *J Bacteriol* 1991; 173:7257–7268.
40. Roca G, Distel DL, Waterbury JB, Chisholm SW. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* 2002; 68:1180–1191.
41. Ramsing NB, Ferris MJ, Ward DM. Highly ordered vertical structure of *Synechococcus* populations within the one-millimeter-thick photic zone of a hot spring cyanobacterial mat. *Appl Environ Microbiol* 2000; 66:1038–1049.
42. Ward DM. A natural species concept for prokaryotes. *Curr Opin Microbiol* 1998; 1:271–277.
43. Beja O, Koonin E, Aravind L, et al. Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl Environ Microbiol* 2002; 68:335–345.
44. Vulic M, Lenski RE, Radman M. Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proc Natl Acad Sci USA* 1999; 96:7348–7351.
45. Lawrence JG. Gene transfer in bacteria: speciation without species? *Theor Popul Biol* 2002; 61: 449–460.
46. Cohan FM. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am Naturalist* 1994; 143:965–986.
47. Posada D. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol* 2002; 19:708–717.
48. Smith JM, Smith N, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci USA* 1993; 15:4384–4388.

49. Cohan FM. Clonal structure: an overview. In: Pagel M (ed). Encyclopedia of Evolution. Vol. 1. New York: Oxford University Press, 2002, pp. 159–161.
50. Cohan FM. Population structure and clonality of bacteria. In: Pagel M (ed). Encyclopedia of Evolution. Vol. 1. New York: Oxford University Press, 2002, pp. 161–163.
51. Feil EJ, Cooper JE, Grundman H, et al. How clonal is *Staphylococcus aureus*? J Bacteriol 2003; 185:3307–3316.
52. Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc Natl Acad Sci USA 2002; 99:7687–7692.
53. Majewski J, Zawadzki P, Pickerill P, Cohan FM, Dowson CG. Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. J Bacteriol 2000; 182: 1016–1023.
54. Majewski J, Cohan F. M. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. Genetics 1999; 153:1525–1533.
55. Vulic M, Dionisio F, Taddei F, Radman M. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc Natl Acad Sci USA 1997; 94:9763–9767.
56. Duncan KE, Ferguson N, Kimura K, Zhou X, Istock CA. Fine-scale genetic and phenotypic structures in natural populations of *Bacillus subtilis* and *Bacillus licheniformis*: important implications for bacterial evolution and speciation. Evolution 1994; 48:2002–2025.
57. Majewski J. Sexual isolation in bacteria. FEMS Microbiol Lett 2001; 199:161–169.
58. Cohan FM. Sexual isolation and speciation in bacteria. Genetica 2002; 116:359–370.
59. Majewski J, Cohan FM. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. Genetics 1998; 148:13–18.
60. Rao BJ, Chiu SK, Bazemore LR, Reddy G, Radding CM. How specific is the first recognition step of homologous recombination? Trends Biochem Sci 1995; 20:109–113.
61. Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature 1989; 342:396–401.
62. Doolittle WF. Lateral genomics. Trends Cell Biol 1999; 9:M5–M8.
63. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature 2000; 405:299–304.
64. Licht TR, Christensen BB, Krogfelt KA, Molin S. Plasmid transfer in the animal intestine and other dynamic bacterial populations: the role of community structure and environment. Microbiology 1999; 145(Pt 9):2615–2622.
65. Maynard Smith JM, Dowson CG, Spratt BG. Localized sex in bacteria. Nature 1991; 349: 29–31.
66. Arthur W. Mechanisms of Morphological Evolution: A Combined Genetic, Developmental and Ecological Approach. New York: Wiley, 1984, pp. 182–186.
67. Feil EJ, Holmes EC, Bessen DE, et al. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc Natl Acad Sci USA 2001; 98:182–187.
68. Atwood KC, Schneider LK, Ryan FJ. Periodic selection in *Escherichia coli*. Proc Natl Acad Sci USA 1951; 37:146–155.
69. Cohan FM. Genetic exchange and evolutionary divergence in prokaryotes. Trends Ecol Evol 1994; 9:175–180.
70. Cohan FM. Periodic selection and ecological diversity in bacteria. In: Nurminsky D (ed). Selective Sweep. Georgetown, TX: Landes Bioscience, in press.
71. Guttman DS, Dykhuizen DE. Detecting selective sweeps in naturally occurring *Escherichia coli*. Genetics 1994; 138:993–1003.

72. Majewski J, Cohan FM. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* 1999; 152:1459–1474.
73. Lan R, Reeves PR. Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol* 2000; 8:396–401.
74. Boucher Y, Nesbo CL, Doolittle WF. Microbial genomes: dealing with diversity. *Curr Opin Microbiol* 2001; 4:285–289.
75. Ravin A. The origin of bacterial species: genetic recombination and factors limiting it between bacterial populations. *Bacteriol. Rev* 1960; 24:201–220.
76. Ravin A. Experimental approaches to the study of bacterial phylogeny. *Am Nat* 1963; 97:307–318.
77. Linz B, Schenker M, Zhu P, Achtman M. Frequent interspecific genetic exchange between commensal *Neisseriae* and *Neisseria meningitidis*. *Mol Microbiol* 2000; 36:1049–1058.
78. Cohan FM. Does recombination constrain neutral divergence among bacterial taxa? *Evolution* 1995; 49:164–175.
79. Giraud A, Matic I, Tenaillon O, et al. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* 2001; 291:2606–2608.
80. Denamur E, Lecointre G, Darlu P, et al. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* 2000; 103:711–721.
81. Cohan FM. Bacterial species and speciation. *Syst Biol* 2001; 50:513–524.
- 81a. Cohan FM. The role of genetic exchange in bacterial evolution. *ASM News* 1996; 62:631–636.
82. Suerbaum S, Smith JM, Bapumia K, et al. Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci USA* 1998; 95:12619–12624.
83. Gibson G. Microarrays in ecology and evolution: a preview. *Mol Ecol* 2002; 11:17–24.
84. Cohan FM, Libsch J. Sequence-based evidence for numerous ecologically distinct and irreversibly separate taxa within bacterial species.
85. Papke T, Ramsing NB, Bateson MM, Ward DM. Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* 2003; 5:650–659.