Theoretical Population Biology **61**, 449–460 (2002)
doi:10.1006/tpbi.2002.1587

TPB

Gene Transfer in Bacteria: Speciation without Species?

Jeffrey G. Lawrence
Pittsburgh Bacteriophage Institute and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received January 15, 2002

Although Bacteria and Archaea reproduce by binary fission, exchange of genes among lineages has shaped the diversity of their populations and the diversification of their lineages. Gene exchange can occur by two distinct routes, each differentially impacting the recipient genome. First, homologous recombination mediates the exchange of DNA between closely related individuals (those whose sequences are sufficient similarly to allow efficient integration). As a result, homologous recombination mediates the dispersal of advantageous alleles that may rise to high frequency among genetically related individuals via periodic selection events. Second, lateral gene transfer can introduce novel DNA into a genome from completely unrelated lineages via illegitimate recombination. Gene exchange by this route serves to distribute genes throughout distantly related clades and therefore may confer complex abilities—not otherwise found among closely related lineages—onto the recipient organisms. These two mechanisms of gene exchange play complementary roles in the diversification of microbial populations into independent, ecologically distinct lineages. Although the delineation of microbial ''species'' then becomes difficult—if not impossible—to achieve, a cogent process of speciation can be predicted.

Key Words: **horizontal gene transfer; speciation; homologous recombination; bacteria**

INTRODUCTION

The Linnean paradigm of biological classification—that is, the hierarchical organization of organisms into increasingly narrower groups based on their shared characteristics—has provided a useful framework for interpreting Darwinian evolution. Here, two sister groups in a classification hierarchy (frequently represented by a phylogenetic tree) share characters found in their most recent common ancestor and are distinguished by characters derived from genetic material also found in this ancestor. Intuitively, characteristics shared by all members at any taxonomic level would represent features shared with, or derived from, their most recent common ancestor.

In biological classification, the most narrowly defined group is the species, and the formation of new lineages—that is, speciation—entails the diversification of one species into two. While many concepts have been applied to delineate their boundaries, species are often typified by a free exchange of genetic information among its members that is curtailed between heterospecific individuals. Outside of the species (that is, at higher taxonomic levels), genetic information is transmitted solely by vertical inheritance, since lineages have become genetically isolated following the act of speciation.

While this strict interpretation of the Linnean paradigm applies well to many eukaryotic lineages, whose members fall into well-defined hierarchical groupings, it fails to represent accurately the evolution of microörganisms, particularly the Bacteria and the Archaea, since information is transmitted horizontally at all levels of taxonomic inclusiveness. Gene exchange in prokaryotic lineages involves two distinct mechanisms, and here I explore the idea that while prokaryotic species may be impossible to delineate rigorously in the classical sense, speciation events can still occur by well-defined processes, with critical roles being played by gene exchange at multiple taxonomic levels.

0040-5809/02 $35.00

GENE EXCHANGE BY HOMOLOGOUS RECOMBINATION: A BRIEF HISTORY OF MICROBIAL ''SPECIES''

Unlike most familiar eukaryotes (the diploid, obligately sexual kinds for which traditional species concepts were devised), Bacteria and Archaea reproduce by binary fission, with a single mother cell replicating its DNA and apportioning it equally between two daughter cells. The exchange of DNA between individuals, so tightly coupled with organismal reproduction in these eukaryotes via meiosis and syngamy, is not required for microbial reproduction. This observation clearly predicts that exchange of genes between primarily asexual microbial cells would occur at a rate far lower than that observed for diploid, obligately sexual eukaryotes. As a result, microbial populations should be dominated by clones of nearly identical individuals, derived from common ancestors, which rise to high frequency by virtue of their superior fitness.

This model of primarily clonal inheritance was seemingly upheld when widespread analysis of genetic variation by multilocus enzyme electrophoresis (MLEE) showed extensive linkage disequilibrium between loci in *Escherichia coli* (Ochman *et al*., 1983; Whittam *et al*., 1983, 1984; Caugant *et al*., 1984; Ochman and Selander, 1984). Since exchange of genetic information among strains by any means would serve to disrupt disequilibrium, these data strongly reinforced the clonal model of microbial population biology, and the effects of gene transfer were not considered to be important. Moreover, far from displaying ''bushy'' phylogenies predicted by clonal inheritance in the absence of selection, strains of *E. coli* could be sorted into major groups by their electrotypes (allele patterns determined by MLEE) and serotypes (e.g., Ochman *et al*., 1983; Caugant *et al*., 1984; Ochman and Selander, 1984; Achtman *et al*., 1986). This clustering was attributed to periodic selection events, which would allow strains with beneficial mutations to rise to high frequency (Levin, 1981). Mechanisms for gene exchange among microörganisms (e.g., conjugation, transduction and transformation) were known to exist, were used extensively in the laboratory for genetic manipulations, and could even act at moderate rates and not disrupt the observed patterns of linkage disequilibrium. Yet at the time, Occam's razor favored the exclusion of such ''intraspecific'' recombination in models of *E. coli* population structure (Hartl and Dykhuizen, 1984), and the issue of the impact of gene exchange was set aside until sufficient data were amassed to address this issue.

The collection of nucleotide sequence data beginning in the 1980s brought the role of gene exchange into the forefront of microbial population genetics. Nucleotide sequences for regions of the *trp* operon (Milkman and Crawford, 1983), the *phoA* gene (DuBose *et al*., 1988) and the *gnd* gene (Dykhuizen and Green, 1991) were determined for the same subset of strains of *E. coli*. Although reproduction by binary fission predicted that the phylogenetic relationships among the strains inferred from each of these sets of gene sequences should be congruent, they were not. Two conclusions could be drawn from these pivotal data sets. First, gene exchange via homologous recombination had clearly played a role in shuffling genetic information among strains of *E. coli*, sometimes involving mere fragments of genes (DuBose *et al*., 1988). The mechanisms for transmission of DNA between cells—bacteriophage-mediated transduction, transformation by naked DNA, and transfer by direct conjugative contact—moved fragments of DNA smaller than the entire chromosome, making recombination a physically restricted, chromosomally local event. Further fine-scale analyses uncovered a critical role for restriction endonucleases in reducing the size of the DNA fragments that were eventually incorporated into the recipient chromosome (McKane and Milkman, 1995).

Second, periodic selection events—now termed selective sweeps—could be plausibly documented, since the average nucleotide divergence among strains differ greatly between loci; while *trp* genes showed only 0–2% nucleotide divergence among this set of strains, *phoA* alleles were 2–4% divergent and *gnd* alleles varied up to 16% among the same sets of strains. The difference in absolute levels of nucleotide divergence can be interpreted to reflect differences in the time since the last periodic selection event at each locus. Mechanistically, then, the spread of alleles via homologous recombination appeared necessarily restricted to small regions of the chromosome bearing selectively advantageous information. The influx of nucleotide sequence data allowed the identification of selective sweeps at many other loci (e.g., the *gapA* locus in *E. coli*; Guttman and Dykhuizen, 1994b).

As more genes were investigated, the genetic divergence at the *gnd* locus was recognized as being unusually high. This unexpectedly high level of genetic diversity was thought to result from the proximity of the *gnd* gene to the *rfb* locus, which encodes enzymes responsible for synthesis of the O-antigen, the outermost portion of the lipopolysaccharide (LPS). Since the O-antigen is a target for diversifying and/or frequency-dependent selection (Reeves, 1993; Lan and Reeves, 1996), purifying

periodic selection events would be counterselected at the *rfb* locus, and hence at any closely linked genes (like *gnd*). These data showed that although homologous recombination could act to distribute variant alleles among strains of *E. coli*, it did not affect all loci uniformly (Milkman, 1997).

Homologous recombination plays a critical role in mitigating the scope of periodic selection events which increase the representation of advantageous alleles. Rather than allowing an entire chromosome bearing a single advantageous mutation to sweep a population of *E. coli*, periodic selection could carry to high frequency only the region local to the advantageous allele. In this way, linkage disequilibrium between distantly situated loci on the *E. coli* chromosome could be left undisturbed, making relatively high-frequency recombination completely consistent with the observations of linkage disequilibrium. The accumulation of DNA sequence information provided initial estimates of rates of homologous recombination in *E. coli*—regardless of whether recombination was involved in a selective sweep or not—to be on the order of the mutation rate (Guttman and Dykhuizen, 1994a,b); that is, the probability of a strain inheriting a variant allele by recombination was comparable to the probability of gaining a variant allele by mutational processes (1:1). More recent analyses by Multi-Locus Sequence Typing have raised this estimate to be between 20:1 and 50:1 for *E. coli*, and much higher for some bacterial lineages (Feil *et al.*, 2000, 2001), where natural transformation allows for high-frequency, lineage-specific gene exchange [e.g., *Haemophilus* (Smith *et al.*, 1995) or *Neisseria* (Elkins *et al.*, 1991)]. The finding that 94% of natural isolates of *Salmonella enterica* are lysogens bearing generalized transducing phages (Schicklmaier *et al.*, 1998) provides a plausible mechanism for frequent gene exchange among enteric bacteria.

The observations that different genes from the same strains had different evolutionary histories led Dykhuizen and Green (1991) to postulate that homologous recombination was a unifying force for bacterial species, allowing for the distribution of advantageous mutations among all its members. Their model predicted that phylogenies based on different gene sequences would be incongruent within a bacterial species as a result of homologous recombination (as seen for the *trp*, *phoA* and *gnd* loci at that time). However, relationships among these genes would be congruent between species, since these individuals would lie beyond the scope of homologous recombination, which is precluded by the intervention of the mismatch-correction systems once sequences have become dissimilar (Zawadzki *et al.*, 1995; Vulic *et al.*, 1997, 1999; Majewski and Cohan, 1998, 1999). The "cohesion" provided by homologous recombination fits well with microbiological data (e.g., strains of sister lineages *E. coli* and *S. enterica* are rarely, if ever, confused), and established Mayr's Biological Species Concept (Mayr, 1942, 1963) as a plausible framework for understanding the evolution of bacterial populations.

In this model, however, the mismatch-correction barrier, which prevents facile gene exchange between more distantly related individuals via homologous recombination, can merely serve as an upper boundary when delineating a microbial "species." Although high-frequency homologous recombination between dissimilar genes is precluded by this barrier, the converse is not necessarily true. That is, homologous recombination does not necessarily act freely between genes that are, for the most part, highly similar; this is certainly the case for genes near the *gnd* locus as discussed above, where such recombination is counterselected. The possibility of multiple bacterial "species" or "ecotypes" inhabiting the genetic space encompassed by "freely recombining" individuals has been explored (Cohan, 2001) and will be discussed further below. It is also clear from an empirical standpoint that organisms falling within the purview of a bacterial "species" by the Dykhuizen and Green criteria may bear distinctly different characteristics; for example, strains of *E. coli* include both the benign laboratory strain K12 and the notoriously pathogenic strain O157:H7. The genome sequences of these organisms (Blattner *et al.*, 1997; Perna *et al.*, 2001) show dramatic differences reflecting additional recombinatorial processes (Kudva *et al.*, 2002) involving very distantly related individuals that are discussed below.

GENE TRANSFER BETWEEN SPECIES

As predicted by Dykhuizen and Green (1991), phylogenies constructed from sequences of genes shared among closely related bacterial taxa were indeed congruent (Lawrence *et al.*, 1991), and such molecular methods gradually replaced cruder measures [e.g., numbers of shared biotypic characters (Ewing, 1984), or similarity metrics based on chromosomal DNA:DNA hybridization (Brenner and Falkow, 1971)] as the method of choice for inferring relationships among microbial taxa. The sequences of universally distributed rRNA loci even provided a framework for inferring

relationships among all Bacteria, Archaea, and Eukaryotes (Fox *et al.*, 1980; Woese, 1987, 1991; Olsen and Woese, 1993). Yet these methods focused on genes that were shared among the taxa being analyzed; genes unique to a lineage, or subset of lineages, were not useful from a phylogenetic standpoint and were essentially ignored.

As more sequence data were collected, instances where the Dykhuizen and Green predictions were not upheld began to be recognized. Genes unique to a taxon were sometimes limited in their distributions due to the deletion of their cognate sequences from related taxa [like the *phoA* gene from *Salmonella* (DuBose and Hartl, 1990), or the *cadA* gene from *Shigella* (Maurelli *et al.*, 1998)]. But more often, these genes appeared to have been introduced into a genome from an unrelated source and incorporated by illegitimate or site-specific recombination, rather than by homologous recombination with a closely related sequence. While individual genes may have a different likelihoods of being mobilized, no gene appears immune to this sort of "lateral" gene transfer; genes encoding central metabolic functions (Doolittle *et al.*, 1990; Olendzenski *et al.*, 2000), complete biosynthetic pathways (Kranz and Goldman, 1998; Boucher *et al.*, 2001), portions of the transcription and translation machinery (Ibba *et al.*, 1997; Wolf *et al.*, 1999; Woese *et al.*, 2000), even ribosomal proteins (Brochier *et al.*, 2000) and ribosomal RNA (Mylvaganam and Dennis, 1992; Yap *et al.*, 1999) have been shown to be transferred across large phylogenetic distances.

If the donor taxa are not included in a phylogenetic analysis, then the predictions of Dykhuizen and Green would be upheld (Fig. 1), and phylogenies of different genes from different taxa would remain congruent. Yet when distantly related taxa were included in phylogenetic analyses, the inferred relationships were rarely congruent with those inferred from the rRNA molecule (Woese *et al.*, 2000; Friedich, 2002). While individual genes often provided compelling cases for lateral gene transfer between distantly related organisms, even between prokaryotes and eukaryotes (Buchanan-Wollaston *et al.*, 1987; Heinemann and Sprague, 1989), a thorough assessment of the impact of laterally transferred sequences required more comprehensive data sets.

Ultimately, the determination of complete genome sequences revealed that lateral (or horizontal) gene transfer between very distantly related organisms was far more prevalent than had been suspected. By detecting unusually high levels of similarity between genes found in otherwise unrelated taxa, phylogenetic methods showed that large fractions of prokaryotic genomes arose from gene transfer; for example, up to 24% of the *Thermotoga maritima* genome was predicted to have been obtained from Archaeal lineages (Nelson *et al.*, 1999), perhaps facilitating its invasion of high-temperature habitats (although other interpretations of these data have been offered, Logsdon and Fuguy, 1999), and both phylogenetic analyses and examination of atypical sequences predict that more than 20% of the *E. coli* genome has been recently introduced by gene transfer (Lawrence and Ochman, 1998, 2002). In detecting incongruities in the relationships among organisms reflected by different genes, phylogenetic methods are powerful tools for detecting lateral gene transfer. However, they are limited by the depth and breadth of the sequence database, by potentially rapid rates of sequence evolution that may confound phylogenetic analyses, by the ambiguous identification of orthologues among diverse gene families and by incomplete taxon sampling that may lead to spurious conclusions of potential gene transfers (exemplified quite dramatically by the apparently premature claim that hundreds of genes in the human genome were acquired by lateral transfer; International Human Genome Sequencing Consortium, 2001; Salzberg *et al.*, 2001; Stanhope *et al.*, 2001).

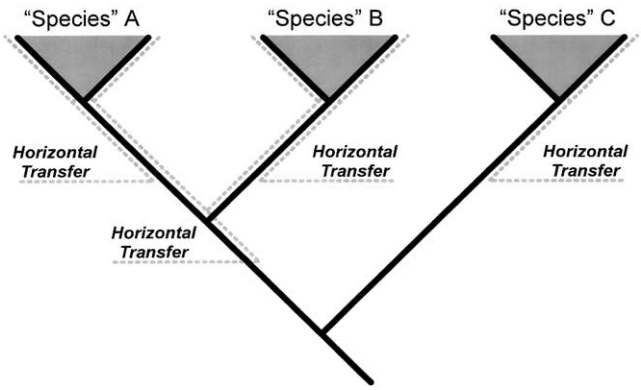

FIG. 1. Relationships among bacterial taxa, as predicted by Dykhuizen and Green (1991). Due to homologous recombination, phylogenies derived from different genes are not congruent within bacterial "species," as denoted by the gray triangles encompassing the genetic variation within the clade. However, the lack of homologous recombination between taxa—precluded by the mismatch correction barrier—results in congruent phylogenies for different genes. Horizontal transfer (denoted by dashed lines) may introduce genes into these lineages, but so long as the donor taxa are not included in the phylogenetic analysis, the prediction of the Dykhuizen and Green model are upheld. Therefore, this model strictly applies only to "local" phylogenies, that is, those of groups of closely related taxa.

Alternatively, genes introduced by horizontal transfer can be detected as those which do not resemble long-term residents of bacterial genomes. These atypical genes can be recognized as those whose nucleotide composition (Lawrence and Ochman, 1997, 1998), dinucleotide patterns (Karlin and Burge, 1995; Karlin, 1998) codon usage biases (Médigue *et al.*, 1991; Whittam and Ake, 1992; Lawrence and Ochman, 1997, 1998) or other sequence features (Hayes and Borodovsky, 1998) fail to reflect patterns resulting from the directional mutation pressures (Sueoka, 1962, 1988, 1992) characteristic of their resident genome. Directional mutation pressures cause genes native to, or long-term residents of, a microbial genome to resemble one another in these features, and significantly atypical genes are strong candidates for those introduced recently by horizontal processes.

A strength of detecting horizontally transferred genes by these parametric methods is its independence from both the sequence database and the vagaries of phylogenetic comparisons. A caveat, however, is that the atypical features of newly acquired genes, so useful in their initial identification, ameliorate over time as a result of the directional mutation pressures characteristic of the new host (Lawrence and Ochman, 1997), eventually causing foreign genes to resemble long-term residents of bacterial genomes. Both phylogenetic and parametric approaches to the detection of horizontally acquired genes provide complementary means for assessing the extent and impact of lateral gene transfer among bacterial lineages (Lawrence and Ochman, 2002), and show that introduction of foreign genes is an ongoing process, not merely comprising ancient events.

Parametric methods for the detection of recently acquired genes, while not without their faults, offer an opportunity to quantitate the rate of horizontal transfer (Lawrence and Ochman, 1997, 1998). Atypical genes ameliorate over time to resemble native genes, whose nucleotide compositions follow predictable patterns (Muto and Osawa, 1987). During the process of amelioration, atypical genes do not conform to these patterns, and their deviations from the Muto and Osawa relationships can be quantitated to estimate their time of introduction into a bacterial genome (Lawrence and Ochman, 1997, 1998). These amelioration analyses indicated that stably maintained DNA was introduced at a rate of 16 kb/Myr into the *E. coli* genome (Lawrence and Ochman, 1998), and surveys of numerous bacterial genomes (Ochman *et al.*, 2000) demonstrated that introduction of significant numbers of genes by lateral transfer is not limited to this lineage alone.

This frequent transfer of DNA between lineages would seem to deal a devastating blow to the application of Mayr's Biological Species Concept for bacterial lineages as interpreted by Dykhuizen and Green (1991). Since DNA is readily exchanged among all lineages, one cannot define a species as a group of organisms sharing a common gene pool. From the recipient organism's perspective, the introduction of DNA from exogenous sources more closely approximates a mutational event than a recombination event, allowing horizontal transfer to be ignored in the phylogenetic reconstruction of closely related groups of taxa. Much like an advantageous point mutation, the new information introduced by lateral transfer may be distributed among closely related lineages by homologous recombination at the sequences flanking the acquired DNA. Yet from a global perspective, the hierarchical organization of organisms encapsulated by the Linnean paradigm is infeasible, since it cannot be applied uniformly to all of the genes in bacterial chromosomes. As discussed next, without a firm delineation of a bacterial "species," how can the process of speciation be described?

LATERAL TRANSFER AS A SPECIATION CATALYST

The recognition that DNA was being introduced into bacterial genomes at appreciable rates also led to a reconsideration of the forces that drive lineage diversification (that is, speciation) (Lan and Reeves, 1996; Lawrence, 1997, 1999; Lawrence and Roth, 1998; Woese, 1998; 2000; Doolittle, 1999a, 1999b, 2000; Levin and Bergstrom, 2000). Unlike point mutational processes, which can gradually alter existing genes to allow their encoded products to perform modified functions, lateral gene transfer could introduce in a single step all of the genes required for the deployment of complex metabolic processes, thereby allowing immediate, efficient exploitation of competitive environments (Lawrence and Roth, 1996; Lawrence, 1997). This potential for saltational change in organismal phenotype is due, in many instances, to the underlying organization of bacterial genes into clusters or operons of cotranscribed genes.

Often, all of the genes required for complex biological processes (like the synthesis of an amino acid or cofactor, or the transport and degradation of a molecule for energy or for its carbon or nitrogen, or even the

synthesis and deployment of cytochromes; Kranz and Goldman, 1998) are found in gene clusters or operons. This gene arrangement facilitates the distribution of these genes among bacterial lineages since all mechanisms for gene transfer are limited by the size of the fragments they mobilize. Moreover, transcription of acquired operons can be supplied by a host promoter at the site of insertion, thereby circumventing the necessity of adapting to the peculiarities of the host's transcriptional apparatus. In this way, gene clustering allows microörganisms to acquire complex phenotypes in a single step (e.g., the acquisition of virulence plasmids by *Yersinia pestis*, allowing it to invade a significantly different niche than its parental lineage, *Y. pseudotuberculosis*; Achtman *et al.*, 1999; Dykhuizen, 2000).

The transfer of operons may play an important role in prokaryotic diversification that lacks a eukaryotic counterpart. Complex eukaryotes are able to invade and adapt to novel ecological niches by the modification of existing characters (e.g., altering beak shape among Darwin's finches allows for the use of a new food sources), entailing the differential expression of genes involved in the development of the salient morphological features. Among prokaryotes, however, the invasion of new ecological niches often requires the action of new biochemical activities (e.g., the breaking of a newly encountered glycosidic bond to allow the use of a novel sugar as a food source). The acquisition of genes by lateral transfer allows recipient organisms to exploit novel biochemical functions already refined by selection in donor organisms, thereby facilitating the efficient and effective exploitation of novel ecological niches (Lawrence, 1999, 2001; Lawrence and Roth, 1999). Although acquisition of novel biochemical pathways by lateral transfer has been observed in eukaryotes (e.g., the acquisition of genes by ruminant fungi from colocalized bacteria; Garcia-Vallve *et al.*, 2000), the role for lateral transfer in higher organisms in mediating the evolution of ecological distinctiveness is less clear. For example, the genetic alterations that allow for different beak shapes among Darwin's finches manifest their ecological importance only in the context of the developmental pathways of the host, and would likely not confer a similar ecological advantage if the altered genes from finches were transferred to a different kind of organism.

The ability for gene acquisition to change the character of a bacterial lineage makes lateral transfer an ideal candidate for the evolutionary force driving lineage diversification (that is, speciation). But homologous recombination would act to distribute newly acquired genes among closely related organisms, thereby coalescing nascent "species" as they attempted to diverge. How can these processes be reconciled to provide a uniform view of bacterial species and speciation?

GENE EXCHANGE AND THE FUZZY SPECIES BOUNDARY

Consider a population of bacteria, freely recombining at all of their loci, wherein a horizontal transfer event introduces genes which allow the newly created organism to exploit effectively and efficiently a niche dissimilar to the niche of its maternal parent. Also, consider a second event in this incipient lineage, either (a) the acquisition of additional genes by horizontal transfer which are beneficial in this new niche, (b) the loss of genes that are actively problematic in the new niche (e.g., Maurelli, 1994), or (c) the appearance of a point mutation beneficial in this new environment that is not beneficial in the old environment. Alternatively, modifications may arise in the "parental" lineage which improve its performance in the ancestral niche, or allow it to exploit a new niche different from the first incipient lineage. At this point, two independent loci confer fitness differences between the two organisms, which are now adapted to new environments. Homologous recombination of genes found in these twice-derived lineages would still proceed unfettered, except when it transferred DNA corresponding to one of the two loci which conferred ecological distinctiveness (Fig. 2). If recombination reassorted alleles at one of these two loci, the hybrid progeny bacterium would be less fit than either of its parents. Such recombination is not unlikely; recall that even in moderately recombining taxa (like *E. coli*), the recombination rate is 20–50 times greater than the substitution rate (Feil *et al.*, 2000, 2001) and, as predicted, phytogenies of *E. coli* strains inferred from sequences of different loci are always different, or show evidence for recombination [e.g., the *trp* (Milkman and Crawford, 1983), *phoA* (DuBose *et al.*, 1988), *gnd* (Dykhuizen and Green, 1991), *sppA*, *gapA*, *pabB*, and *zwf* (Guttman and Dykhuizen, 1994a), *mdh* (Boyd *et al.*, 1994) and *putP* (Nelson and Selander, 1992) loci], testifying to the strong impact gene exchange by homologous recombination has had among closely related strains.

This situation is analogous to the "privatization" of periodic selection events to ecologically distinct lineages as proposed by Cohan (1994a, b, 1995, 1996). In our case, rates of homologous recombination would be

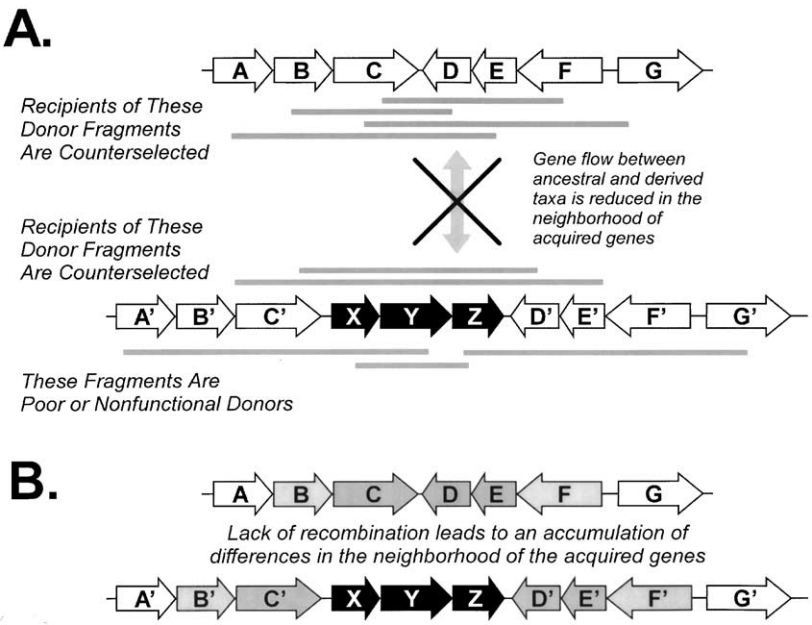

FIG. 2. The effect of horizontal transfer on the rate of homologous recombination. (A) A three gene cluster (genes X, Y and Z) is introduced into a population by lateral transfer; a secondary event then occurs that distinguished this derived lineage from the parental lineage. As a result, the rate of homologous recombination between the two lineages in genes flanking the XYZ genes decreases for two reasons. First, recombinants resulting from donor fragments which carry the join-point of the integration site between groups are counterselected, either because they introduce genes X, Y and Z into the ancestral background, or because they remove them from the derived background. In addition, donor fragments for which one or more dsDNA end lies within genes X, Y or Z will serve as poor donors for non-XYZ bearing strains. (B) As a result of the decreased rates of recombination between XYZ-bearing and non-XYZ-bearing strains, the regions adjacent to these acquired genes will diverge at a greater rate since they will not participate in local periodic selection events (denoted by gray regions).

reduced in the neighborhood of horizontally acquired genes since such exchange would produce less-fit progeny (Fig. 2). As a result, periodic selection events will not purge variation in the vicinity of horizontally transferred DNA if these acquired genes conferred useful functions, thereby leading to a "forbidden zone," so to speak, where rates of homologous recombination are reduced in the neighborhoods of horizontally acquired loci. An analogous result is observed at the *gnd* locus, where homologous exchange occurs at a reduced rate in the neighborhood of the *rfb* locus (reflected in higher overall genetic diversity indicating the absence of purifying selective sweeps).

On the whole, these two populations—distinguished by their horizontally acquired genes—would be freely exchanging DNA at most, but not all of their loci. As a result, they would conform to the Biological Species Concept, but only for parts of their chromosomes. Over time, lineage-specific changes would accumulate at more locations across the chromosome, thereby curtailing recombination further (Fig. 3). Genetic drift would allow the accumulation of neutral mutations in the associated forbidden zones (as they have for the *gnd* locus), leading to higher levels of diversity which reinforce and expand the neighborhood of reduced recombination (Figs. 2 and 3). Eventually, sufficient loci would distinguish the two lineages as being ecologically distinct and post-mating reproductive isolation (the reduced fitness of progeny recombinants) would curtail gene flow between the lineages. The accumulation of neutral mutations would cause the DNA to diverge to the point where the mismatch-correction system would impose pre-mating reproductive isolation (Fig. 3).

If we estimate the "forbidden zone" to extend $\sim$ 5 kb on either side of an acquired gene (the distance from the *gnd* locus to the *rfb* operon, a very conservative estimate), then each horizontally acquired gene "protects" about 10 kb of the surrounding chromosome from homologous exchange between the two nascent lineages. For a genome the size of *E. coli* (4500 kb), several hundred lineage-specific loci—either genes obtained by horizontal transfer or advantageous alleles arising by point mutation—would be required to effect

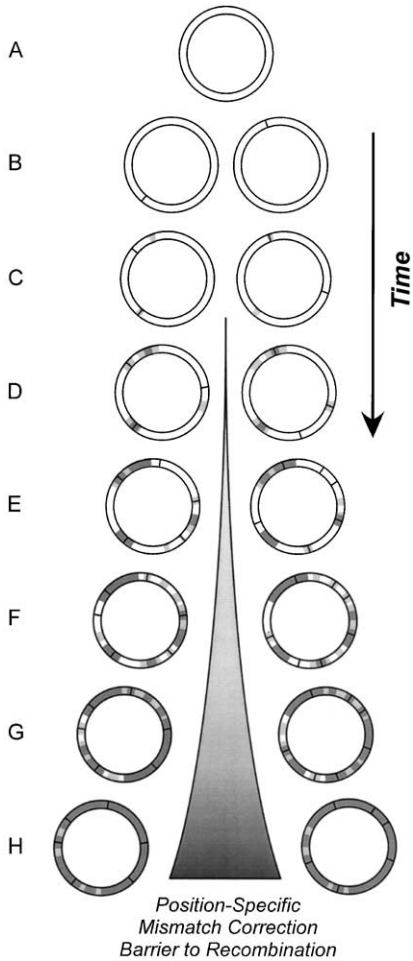

FIG. 3. Accumulation of neutral mutations in the neighborhoods of horizontally acquired genes leads to mosaic chromosomes, whereby some genes between lineages freely recombine whereas those proximal to acquired genes do not. In the two diverging lineages, acquired genes are denoted as lines interrupting the solid white chromosome. As a result, homologous recombination between the two lineages is reduced in the vicinity of the acquired genes (gray areas; see text). As the number of acquired genes conferring lineage-specific functions increases, the accumulation of differences reaches a threshold where the mismatch correction system will impart pre-mating genetic isolation, preventing facile exchange of genes by homologous recombination between these groups, completing lineage diversification (speciation).

complete post-mating reproductive isolation (Fig. 3) at surrounding loci. If the forbidden zone is larger, fewer events would be required to isolate the lineages completely.

Amelioration analyses suggested that *E. coli* acquired $\sim$ 16 kb of stably maintained (that is, retained for a sufficiently long period of time to infer that the genes experienced selection for function), novel DNA per million years (Lawrence and Ochman, 1998), representing about 16 genes 1 kb in length (the average size in *E. coli*), or about $\sim$ 6–7 events (approximating an average operon length as $\sim$ 2–3 genes). Therefore, in considering only horizontal transfer events, the lineage splitting process likely occurs over a period of tens of millions of years; any "burst" in the rate of successful horizontal transfer events during the initial period of lineage diversification would shorten this period. It is difficult to assess the rate at which lineage-specific point mutations would accumulate, since no feature known to discriminate between *E. coli* and its sister species *S. enterica* can be attributed to mutational processes.

Overall we can see that gene exchange by homologous recombination imparts, at best, a "fuzzy" species boundary, where ecological distinctiveness counterselects recombination at some loci, but not at others. Only after these lineages have been well established can neutral mutations accumulate to the point where mismatch correction systems prevent gene exchange at all loci, thereby privatizing all periodic selection events to each lineage. So, while microbial "species" cannot be delineated in the same fashion as eukaryotic species, speciation (that is, lineage separation) can occur by a predictable, well-defined mechanism. Moreover, the mismatch-correction barrier can only be used as an upper boundary for any potential microbial "species."

Before reproductive isolation between lineages has been achieved at all loci—that is, where individuals appear to recombine at some loci and not at others—organisms will necessarily fall into non-hierarchical groups, since more than two ecologically distinct sets of strains may coevolve within populations that exchange genes via homologous recombination (Fig. 4). That is, the collection genes "private" to a particular group (exchanged among group members, but not with individuals outside of the group) depends upon the outside group used in the comparison. For example, groups "A" and "B" may exchange some genes (denoted by the gray area of overlap between circles "A" and "B" in Fig. 4), but fail to exchange others (non-gray regions within circles "A" and "B"). However, these groups of "privatized" genes may each be shared differentially with members of group "C" (hatched areas in Fig. 4). Making matters worse, a horizontal gene transfer event that alters the dynamic niche of group "A" strains (discussed below) may then allow exchange of genes that were previously private to group A with members of group "B" or "C", if those genes were no longer important for ecological distinctiveness of group "A" strains in its new niche. Since the

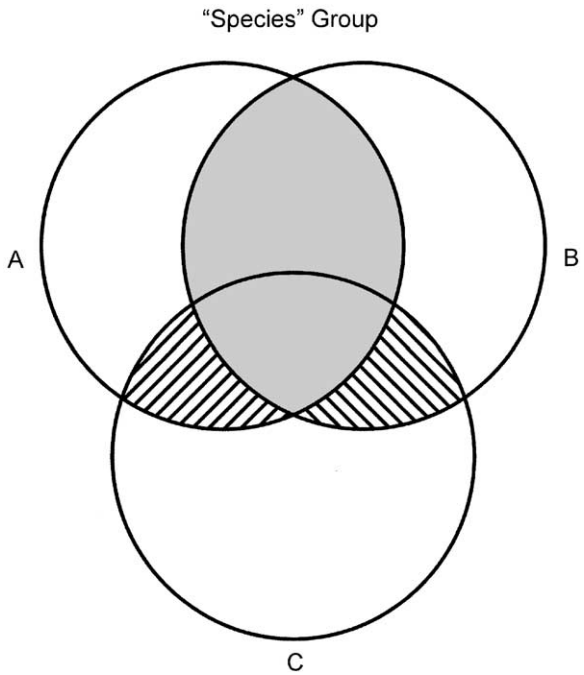

FIG. 4. Clusters of strains which share portions of their genomes cannot be classified in a hierarchical fashion (see text). Circles denote groups of strains; gray and hatched areas of overlap between the circles denoted subsets of genes which experience free recombination between the groups, while regions outside these area are "privatized" between any two groups.

distinctions between groups are fluid, it is difficult to place a boundary around a group of similar strains—even if they appear to be ecologically distinct—to provide this group a label (like a species name) that then connotes permanence while they are still exchanging genes via homologous recombination. The influx of new DNA by lateral gene transfer would serve to constantly shift the boundaries of such groups within recombining populations, making their rigorous delineation impossible.

This "fuzzy" species boundary may have unexpected effects upon comparisons of genes in even well-separated lineages. If the time required to obtain complete genetic isolation between two lineages is large, genes between the two lineages, but located in the neighborhoods of loci first conferring ecological distinctiveness between the two population, will have been in genetic isolation for a longer period of time than the last genes to become reproductively isolated. As a result, comparisons of relative rates of evolution of different genes (e.g., by measuring K_a and K_s) will be confounded by variation in their time of divergence, which cannot be considered constant between genes.

LATERAL TRANSFER AND THE DYNAMIC NICHE BOUNDARY

The model detailed above considers horizontal transfer as a primary force allowing the exploitation of novel environments. This viewpoint is reinforced by the analysis of the genomes of closely related organisms (e.g., *E. coli* and *S. enterica*), wherein all characteristics that distinguish between these organisms (at least in laboratory environments) can be attributed to genes acquired by one lineage or lost from the other. In no case can a physiological difference be attributed to an ancestral gene adopting different roles in the two taxa. This viewpoint is attractive in that genes acquired by lateral transfer can confer novel functions—including complex physiological capabilities encoded by operons—that allow exploitation of new ecological niches. This role for niche expansion is very much different than that typically envisioned for point mutations, which are traditionally viewed as increasing the fitness of an organism within a defined niche. This view has been propagated, for example, by examining the evolution of bacteria within strictly defined environments, wherein mutations do accumulate to increase the fitness of the resident organisms (Papadopoulos *et al.*, 1999). While it is clear that mutations can serve to increase the fitness of an organism within a particular niche, it is not clear that such "niche refinement" is a common practice in natural environments. Since horizontal transfer events may introduce completely novel physiological capabilities, these events would—almost by definition—change the ecological niche of the recipient organism, making "niche refinement" a difficult task that must operate on a moving target.

Therefore, the role of horizontal gene transfer in introducing novel information into bacterial chromosomes forces us to reconsider the role of the bacterium in defining its own ecological niche. Rather than being a passive substrate, whereby beneficial mutations are those which confer higher organismal fitness within a "chosen" niche, one may consider a bacterium to be the active arbiter of its niche; here the niche is defined as the environment best exploited by the complement of genes it currently possesses. Hence, the niche does not dictate the genotype of the organism (by arbitrating which mutations are beneficial and which are not), but rather is the environment best exploited by the genotype (which redefines the niche upon any change). Here, beneficial mutations allow for the successful exploitation of an environment, while detrimental mutations shift an organism into a niche in which it cannot compete

successfully. The bacterium's niche would be redefined continually by virtue of the constant influx of DNA, rather than occurring only at the time of lineage diversification. Even in the cases of rigorously controlled laboratory environments, bacteria adapt via point mutation to create their own niches within this seemingly constant ecological space (Treves *et al.*, 1998), supporting the hypothesis that a bacterium will fit into whatever environmental space is best exploited by its current genetic makeup.

ACKNOWLEDGMENTS

This work was supported by a grant from the David and Lucile Packard Foundation. I thank E.A. Presley and H. Hendrickson for helpful discussions and four anonymous reviewers for helpful comments on the manuscript.

REFERENCES

Achtman, M., Heuzenroeder, M., Kusecek, B., Ochman, H., Caugant, D., Selander, R. K., Vaisanen-Rhen, V., Korhonen, T. K., Stuart, S., Orskov, F., and Orskov, I. 1986. Clonal analysis of *Escherichia coli* O2: K1 isolated from diseased humans and animals, *Infect. Immun.* **51**, 268–276.

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., and Carniel, E. 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*, *Proc. Natl. Acad. Sci. USA* **96**, 14,043–14,048.

Blattner, F. R., Plunkett, G. R., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. 1997. The complete genome sequence of *Escherichia coli* K-12, *Science* **277**, 1453–1474.

Boucher, Y., Huber, H., L' Haridon, S., Stetter, K. O., and Doolittle, W. F. 2001. Bacterial origin for the isoprenoid biosynthesis enzyme hmg-Coa reductase of the archaeal orders thermoplasmatales and archaeoglobales, *Mol. Biol. Evol.* **18**, 1378–1388.

Boyd, E. F., Nelson, K., Wang, F. S., Whittam, T. S., and Selander, R. K. 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*, *Proc. Natl. Acad. Sci. USA* **91**, 1280–1284.

Brenner, D. J., and Falkow, S. 1971. Molecular relationships among members of the enterobacteriaceae, *Adv. Genet.* **16**, 81–118.

Brochier, C., Philippe, H., and Moreira, D. 2000. The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome, *Trends Genet.* **16**, 529–533.

Buchanan-Wollaston, V., Passiatore, J. E., and Canon, F. 1987. The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants, *Nature* **328**, 170–175.

Caugant, D. A., Levin, B. R., and Selander, R. K. 1984. Distribution of multilocus genotypes of *Escherichia coli* within and between host families, *J. Hyg. Camb.* **92**, 377–384.

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

Cohan, F. M. 1994b. Genetic exchange and evolutionary divergence in prokaryotes, *Trends Ecol. Evol.* **9**, 175–180.

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

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

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

Doolittle, R. F., Feng, D. F., Anderson, K. L., and Alberro, M. R. 1990. A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote, *J. Mol. Evol.* **31**, 383–388.

Doolittle, W. F. 1999a. Lateral genomics, *Trends Cell. Biol.* **9**, M5–M8.

Doolittle, W. F. 1999b. Phylogenetic classification and the universal tree, *Science* **284**, 2124–2129.

Doolittle, W. F. 2000. Uprooting the tree of life, *Sci. Am.* **282**, 90–95.

DuBose, R. F., Dykhuizen, D. E., and Hartl, D. L. 1988. Genetic exchange among natural isolates of bacteria: recombination within *the phoA* gene of *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **85**, 7036–7040.

DuBose, R. F., and Hartl, D. L. 1990. The molecular evolution of alkaline phosphatase: correlating variation among enteric bacteria to experimental manipulations of the protein, *Mol. Biol. Evol.* **7**, 547–577.

Dykhuizen, D. E. 2000. *Yersinia pestis*: an instant species?, *Trends Microbiol.* **8**, 296–298.

Dykhuizen, D. E., and Green, L. 1991. Recombination in *Escherichia coli* and the definition of biological species, *J. Bacteriol.* **173**, 7257–7268.

Elkins, C., Thomas, C. E., Seifert, H. S., and Sparling, P. F. 1991. Species-specific uptake of DNA by gonococci is mediated by a 10-base-pair sequence, *J. Bacteriol.* **173**, 3911–3913.

Ewing, W. H. 1984. Taxonomy and nomenclature, *in* "Identification of Enterobacteriaceae" (P.R. Edwards and W.H. Ewing, Eds.), pp. 1–25, Elsevier, New York, Amsterdam, Oxford.

Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M. S., Day, N. P., Enright, M. C., Goldstein, R., Hood, D. W., Kalia, A., Moore, C. E., Zhou, J., and Spratt, B. G. 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences, *Proc. Natl. Acad. Sci. USA* **98**, 182–187.

Feil, E. J., Smith, J. M., Enright, M. C., and Spratt, B. G. 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data, *Genetics* **154**, 1439–1450.

Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N., and Woese, C. R. 1980. The phylogeny of prokaryotes, *Science* **209**, 457–463.

Friedich, M. W. 2002. Phylogenetic analysis reveals multiple lateral transfers of adenosine-5′-phosphosulfate reductase genes among sulfate-reducing microorganisms, *J. Bacteriol.* **184**, 278–289.

Garcia-Vallve, S., Romeu, A., and Palau, J. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi, *Mol. Biol. Evol.* **17**, 352–361.

Guttman, D. S., and Dykhuizen, D. E. 1994a. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation, *Science* **266**, 1380–1383.

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

Hartl, D. L., and Dykhuizen, D. E. 1984. The population genetics of *Escherichia coli*, *Ann. Rev. Genet.* **18**, 31–68.

Hayes, W. S., and Borodovsky, M. 1998. How to interpret an anonymous bacterial genome: machine learning approach to gene identification, *Genome Res.* **8**, 1154–1171.

Heinemann, J. A., and Sprague, G. F. J. 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast, *Nature* **340**, 205–209.

Ibba, M., Morgan, S., Curnow, A. W., Pridmore, D. R., Vothknecht, U. C., Gardner, W., Lin, W., Woese, C. R., and Soll, D. 1997. A euryarchaeal lysyl-tRNA synthetase: resemblance to class I synthetases, *Science* **278**, 1119–1122.

International Human Genome Sequencing Consortium 2001. Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.

Karlin, S. 1998. Global dinucleotide signatures and analysis of genomic heterogeneity, *Curr. Opin. Microbiol.* **1**, 598–610.

Karlin, S., and Burge, C. 1995. Dinucleotide relative abundance extremes: a genomic signature, *Trends Genet.* **11**, 283–290.

Kranz, R. G., and Goldman, B. S. 1998. Evolution and horizontal transfer of an entire biosynthetic pathway for cytochrome *c* biogenesis: *Helicobacter*, *Deinococcus*, Archae and more. *Mol. Microbiol.* **27**, 871–874.

Kudva, I. T., Evans, P. S., Perna, N. T., Barrett, T. J., Ausubel, F. M., Blattner, F. R., and Calderwood, S. B. 2002. Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms, *J. Bacteriol.* **184**, 1873–1879.

Lan, R., and Reeves, P. 1996. Gene transfer is a major force in bacterial evolution, *Mol. Biol. Evol.* **13**, 47–55.

Lawrence, J. G. 1997. Selfish operons and speciation by gene transfer, *Trends Microbiol.* **5**, 355–359.

Lawrence, J. G. 1999. Gene transfer, speciation, and the evolution of bacterial genomes, *Curr. Opin. Microbiol.* **2**, 519–523.

Lawrence, J. G. 2001. Catalyzing bacterial speciation: correlating lateral transfer with genetic headroom, *Syst. Biol.* **50**, 479–496.

Lawrence, J. G., and Ochman, H. 1997. Amelioration of bacterial genomes: rates of change and exchange, *J. Mol. Evol.* **44**, 383–397.

Lawrence, J. G., and Ochman, H. 1998. Molecular archaeology of the *Escherichia coli* genome, *Proc. Natl. Acad. Sci. USA* **95**, 9413–9417.

Lawrence, J. G., and Ochman, H. 2002. Reconciling the many faces of gene transfer, *Trends Microbiol.* **10**, 1–4.

Lawrence, J. G., Ochman, H., and Hartl, D. L. 1991. Molecular and evolutionary relationships among enteric bacteria, *J. Gen. Microbiol.* **137**, 1911–1921.

Lawrence, J. G., and Roth, J. R. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters, *Genetics* **143**, 1843–1860.

Lawrence, J. G., and Roth, J. R. 1998. Roles of horizontal transfer in bacterial evolution, *in* "Horizontal Transfer" (M. Syvanen and C. I. Kado, Eds.), pp. 208–225, Chapman & Hall, London.

Lawrence, J. G., and Roth, J. R. 1999. Genomic flux: genome evolution by gene loss and acquisition, *in* "Organization of the Prokaryotic Genome" (R. L. Charlebois, Ed.), pp. 263–289, ASM Press, Washington, DC.

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

Levin, B. R., and Bergstrom, C. T. 2000. Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes, *Proc. Natl. Acad. Sci. USA* **97**, 6981–6985.

Logsdon, J. M., and Fuguy, D. M. 1999. *Thermotoga* heats up lateral gene transfer, *Curr. Biol.* **9**, R747–R751.

Majewski, J., and Cohan, F. M. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*, *Genetics* **148**, 13–18.

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

Maurelli, A. T. 1994. Virulence protein export systems in *Salmonella* and *Shigella*: a new family or lost relatives, *Trends Cell. Biol.* **4**, 240–242.

Maurelli, A. T., Fernández, R. E., Bloch, C. A., Rode, C. K., and Fasano, A. 1998. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence *of Shigella* spp. and enteroinvasive *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **95**, 3943–3948.

Mayr, E. 1942. "Systematics and the Origin of Species," Columbia Univ. Press, New York.

Mayr, E. 1963. "Animal Species and Evolution," Harvard Univ. Press, Cambridge.

McKane, M., and Milkman, R. 1995. Transduction, restriction and recombination patterns in *Escherichia coli*, *Genetics* **139**, 35–43.

Médigue, C., Rouxel, T., Vigier, P., Hénaut, A., and Danchin, A. 1991. Evidence of horizontal gene transfer in *Escherichia coli* speciation, *J. Mol. Biol.* **222**, 851–856.

Milkman, R. 1997. Recombination and population structure in *Escherichia coli*, *Genetics* **146**, 745–750.

Milkman, R., and Crawford, I. P. 1983. Clustered third-base substitutions among wild strains of *Escherichia coli*, *Science* **221**, 378–379.

Muto, A., and Osawa, S. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution, *Proc. Natl. Acad. Sci. USA* **84**, 166–169.

Mylvaganam, S., and Dennis, P. P. 1992. Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaebacterium *Haloarcula marismortui*, *Genetics* **130**, 399–410.

Nelson, K., and Selander, R. K. 1992. Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*, *J. Bacteriol.* **174**, 6886–6895.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., and Fraser, C. M. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*, *Nature* **399**, 323–329.

Ochman, H., Lawrence, J. G., and Groisman, E. 2000. Lateral gene transfer and the nature of bacterial innovation, *Nature* **405**, 299–304.

Ochman, H., and Selander, R. K. 1984. Evidence for clonal population structure in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **81**, 198–201.

Ochman, H., Whittam, T. S., Caugant, D. A., and Selander, R. K. 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*, *J. Gen. Microbiol.* **129**, 2715–2726.

Olendzenski, L., Liu, L., Zhaxybayeva, O., Murphey, R., Shin, D. G., and Gogarten, J. P. 2000. Horizontal transfer of archaeal genes

into the deinococcaceae: detection by molecular and computer-based approaches [In Process Citation], *J. Mol. Evol.* **51**, 587–599.

Olsen, G. J., and Woese, C. R. 1993. Ribosomal RNA: a key to phylogeny, *FASEB J.* **7**, 113–123.

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R. E., and Blot, M. 1999. Genomic evolution during a 10,000-generation experiment with bacteria, *Proc. Natl. Acad. Sci. USA* **96**, 3807–3812.

Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A., and Blattner, F. R. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7, *Nature* **409**, 529–533.

Reeves, P. 1993. Evolution *of Salmonella* O antigen variation by interspecific gene transfer on a large scale, *Trends Genet.* **9**, 17–22.

Salzberg, S. L., White, O., Peterson, J., and Eisen, J. A. 2001. Microbial genes in the human genome: lateral transfer or gene loss?, *Science* **292**, 1903–1906.

Schicklmaier, P., Moser, E., Wieland, T., Rabsch, W., and Schmieger, H. 1998. A comparative study on the frequency of prophages among natural isolates of *Salmonella* and *Escherichia coli* with emphasis on generalized transducers, *Antonie Van Leeuwenhoek* **73**, 49–54.

Smith, H. O., Tomb, J. F., Dougherty, B. A., Fleischmann, R. D., and Venter, J. C. 1995. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome, *Science* **269**, 538.

Stanhope, M. J., Lupas, A., Italia, M. J., Koretke, K. K., Volker, C., and Brown, J. R. 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates, *Nature* **411**, 940–944.

Sueoka, N. 1962. On the genetic basis of variation and heterogeneity in base composition, *Proc. Natl. Acad. Sci. USA* **48**, 582–592.

Sueoka, N. 1988. Directional mutation pressure and neutral molecular evolution, *Proc. Natl. Acad. Sci. USA* **85**, 2653–2657.

Sueoka, N. 1992. Directional mutation pressure, selective constraints, and genetic equilibria, *J. Mol. Evol.* **34**, 95–114.

Treves, D. S., Manning, S., and Adams, J. 1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*, *Mol. Biol. Evol.* **15**, 789–797.

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

Vulic, M., Lenski, R. E., and Radman, M. 1999. Mutation, recombination, and incipient speciation of bacteria in the laboratory, *Proc. Natl. Acad. Sci. USA* **96**, 7348–7351.

Whittam, T. S., and Ake, S. 1992. Genetic polymorphisms and recombination in natural populations of *Escherichia coli*, *in* "Mechanisms of Molecular Evolution" (N. Takahata and A. G. Clark, Eds.), pp. 223–246, Japan Scientific Society Press, Tokyo.

Whittam, T. S., Ochman, H., and Selander, R. K. 1983. Multilocus genetic structure in natural populations of *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **80**, 1751–1755.

Whittam, T. S., Ochman, H., and Selander, R. K. 1984. Geographical components of linkage disequilibrium in natural populations of *Escherichia coli*, *Mol. Biol. Evol.* **1**, 67–83.

Woese, C. 1998. The universal ancestor, *Proc. Natl. Acad. Sci. USA* **95**, 6854–6859.

Woese, C. R. 1987. Bacterial evolution, *Microbiol. Rev.* **51**, 221–271.

Woese, C. R. 1991. The use of ribosomal RNA in reconstructing evolutionary relationships among bacteria, *in* "Evolution at the Molecular Level" (R. K. Selander, A. G. Clark, and T. S. Whittam, Eds.), pp. 1–24, Sinauer Associates Inc., Sunderland, MA.

Woese, C. R. 2000. Interpreting the universal phylogenetic tree, *Proc. Natl. Acad. Sci. USA* **97**, 8392–8396.

Woese, C. R., Olsen, G. J., Ibba, M., and Soll, D. 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process, *Microbiol. Mol. Biol. Rev.* **64**, 202–236.

Wolf, Y. I., Aravind, L., Grishin, N. V., and Koonin, E. V. 1999. Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events, *Genome Res.* **9**, 689–710.

Yap, W. H., Zhang, Z., and Wang, Y. 1999. Distinct types of rRNA operons exist in the genome of the Actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon, *J. Bacteriol.* **181**, 5201–5209.

Zawadzki, P., Roberts, M. S., and Cohan, F. M. 1995. The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust, *Genetics* **140**, 917–932.