Biol Philos (2010) 25:569–588
DOI 10.1007/s10539-010-9215-5

The myth of bacterial species and speciation

Jeffrey G. Lawrence · Adam C. Retchless

Published online: 5 May 2010

Abstract The Tree of Life hypothesis frames the evolutionary process as a series of events whereby lineages diverge from one another, thus creating the diversity of life as descendent lineages modify properties from their ancestors. This hypothesis is under scrutiny due to the strong evidence for lateral gene transfer between distantly related bacterial taxa, thereby providing extant taxa with more than one parent. As a result, one argues, the Tree of Life becomes confounded as the original branching structure is gradually superseded by reticulation, ultimately losing its ability to serve as a model for bacterial evolution. Here we address a more fundamental issue: is there a Tree of Life that results from bacterial evolution without considering such lateral gene transfers? Unlike eukaryotic speciation events, lineage separation in bacteria is a gradual process that occurs over tens of millions of years, whereby genetic isolation is established on a gene-by-gene basis. As a result, groups of closely related bacteria, while showing robust genetic isolation as extant lineages, were not created by an unambiguous series of lineage-splitting events. Rather, a temporal fragmentation of the speciation process results in cognate genes showing different genetic relationships. We argue that lineage divergence in bacteria does not produce a tree-like framework, and inferences drawn from such a framework have the potential to be incorrect and misleading. Therefore, the Tree of Life is an inappropriate paradigm for bacterial evolution regardless of the extent of gene transfer between distantly related taxa.

Keywords Species · Speciation · Recombination · Gene transfer · Tree of Life

J. G. Lawrence (✉) · A. C. Retchless
Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA
e-mail: jlawrenc@pitt.edu

Species: a central concept in biological inference

Species are the fundamental units of biological organization. Species names are intended to convey information about the characteristics of a group of individuals, allowing properties and characteristics of previously encountered organisms to inform us about the behavior and capabilities of other members of that group. In general, the delineation of species boundaries for plants and animals has little impact outside of scientific communities. Recognition of species may guide conservation efforts and influence constraints on the transport of plants and animals across large geographical distances, but otherwise remain a casual interest to the average person. In contrast, bacterial species carry gravity in agriculture, directed biosynthesis, biotechnology, geochemistry, mineralogy and mining, food safety, disease diagnosis, epidemiology and public health, and bioterrorism. In the latter health-sciences applications, potential life-and-death decisions are guided by the species identification of the microorganisms involved. Yet it is precisely in these arenas that species groups are defined with the least amount of rigor.

There are, of course, reasons why species play such a central role in understanding organismal diversity. Members of a species not only share genotypic and phenotypic similarity, but genetic connectivity as well. For the vast majority of well-defined species, genetic modification arising within a population may be shared with other members of that species. It is this tight coupling of the evolutionary potentials of conspecific individuals that sets the species apart from higher levels of biological classification. Biological organization at higher taxonomic levels (genus, family, class, order, phylum) are arguably passive, historical and somewhat arbitrary. There is no active process leading to phenotypic cohesion at these levels of organization; the similarity of the organisms within those groups reflects properties of common ancestors whose descendents have long since parted ways. In contrast, members of a species share similarity that is maintained by active processes, and speciation events entail the active apportionment of shared biological diversity into evolutionarily independent lineages.

Thus, it has been argued that successive speciation events created the Tree of Life, the widely held notion that organismal diversity can be organized as "groups within groups" representing the process of successive lineage separations; bacterial (and archaeal) lineages would have contributed the majority of its branches. Despite their primacy in populating the tree, and the gravitas of their biological character, the delineation of bacterial species and the events which lead to their formation remain muddled (Gevers et al. 2005). Methods traditionally applied to the description of eukaryotic species fail with bacteria. They have few morphological characters and highly variable physiological properties, the ecological roles of bacterial strains are difficult to ascertain, and sexual exchange is both infrequent and difficult to assess. In addition, the stability of the Tree of Life has also been called into question due to the role of inter-lineage gene transfer in confounding an organism's parentage (Bapteste et al. 2005, 2009; Doolittle 2000; Doolittle and Bapteste 2007; Koonin and Wolf 2009). But here we address an arguably more fundamental issue: do the nature of bacterial species—if they exist at all—and the mechanisms of bacterial lineage separation (speciation) allow for a Tree of Life in

the first place? Here we explore bacterial species and speciation and argue that the Tree of Life, as most widely popularized by Darwin (1859), is not an appropriate model for bacterial evolution. That is, the slow, diffuse process of lineage separation in bacteria not only casts doubt on well-defined speciation events, but leads one to conclude that bacterial species exist only as theoretical constructs not to be seen in nature.

Characteristics of species reflect rules of gene exchange

To begin, we examine the nature of conventionally described bacterial species. Much has been written on this topic (Gevers et al. 2005) and we seek here not to review the arguments but to place ours within an established framework. As formulated by Mayr, species membership can be dictated by rules of gene exchange (Mayr 1942, 1954, 1963). Put simply, gene exchange occurs within a species, providing the force maintaining genotypic and phenotypic similarity within this group; by definition, gene exchange is prohibited across species boundaries. Such gene exchange in eukaryotes is often very frequent (tied to reproduction) and involves the formation of diploid offspring following syngamy of two haploid genomes from distinct parents. Thus within-species gene exchange occurs for all genes in a species' genome during every syngamy event, and (barring rare cases of introgression) between-species genetic isolation for one gene is associated with genetic isolation for all other genes. Following genetic isolation, similarities between related species reflect either convergence or properties of their common ancestors. While there are many other species concepts available for consideration, for example those based on evolutionary trajectory (Wiley 1978) or shared ecological roles (Van Valen 1976), we focus on Mayr's model because of its focus on the genetic relationships between organisms; this parallels the use of genetic relationships in evaluating the extent of microbial species boundaries (e.g., Hanage et al. 2005). While many exceptions to the Mayrian species concept can be enumerated (e.g., ring species in salamanders, diffuse species in oaks), their status as exceptions highlights the utility in the Mayrian concept in describing "typical" species for multicellular eukaryotes, which have served as a conceptual model for the delineation of bacterial species.

In contrast to archetypal eukaryotes, bacteria grow clonally whereby asexual cell division produces two daughter cells from a single parent. Therefore, it may seem odd to employ the Mayrian Biological Species Concept, which was formulated for eukaryotic species wherein obligate genetic exchange occurs during reproduction. Yet bacteria are not strictly asexual (Lederberg 1947; Lederberg and Tatum 1946), and may exchange DNA with very distantly related partners (Ochman et al. 2000). Bacteria may exchange genes by several mechanisms, including conjugation (direct DNA transfer via cytoplasmic bridges), transduction (movement by DNA encapsulation within virus particles) and transformation (direct uptake of naked DNA). Following the introduction of DNA into the cytoplasm, allele conversion requires that the donor and recipient have sufficient genotypic similarity to allow homologous recombination to occur. As strains become more dissimilar, allelic

exchange between them becomes increasingly less frequent as mismatch-repair systems preclude efficient recombination (Vulic et al. 1997, 1999; Zawadzki et al. 1995). Thus, as in eukaryotes, beneficial alleles may spread through a population via recombination and gene exchange can lead to cohesion within groups of bacteria.

However, the rules of prokaryotic gene exchange differ from those of typical eukaryotes. First, gene exchange is not tied to reproduction and is therefore far less frequent. Second, recombination in bacteria involves the unidirectional transfer of genes from donor to recipient, rather than syngamy of two haploid genomes to produce a diploid offspring. Third, gene transfer often involves third parties as mediators (plasmids or bacteriophages), thus eliminating the need for partners to be in the same place at the same time; bacteria can receive genes from donors that lived on the other side of the planet decades or centuries earlier. These frequent, long-distance gene transfers preclude the development of allopatric models of bacterial speciation. Lastly, and most importantly, it is always size-limited, whereby only a small portion of a bacterial genome is mobilized in any single transfer event. This critical difference potentially uncouples the genetic isolation of genes from that of other genes physically distant to them in the genome. The question we pose here is, what patterns result from these processes and how do they influence our interpretation of bacterial species and speciation events?

Deconstructing a bacterial species concept

As discussed above, any species concept in bacteria constructs a framework whereby genotypic and phenotypic similarity among a group of organisms—genetic cohesion sensu Templeton (1989); see also Boyd (1999)—reflects an active process. Here, organisms share an evolutionary trajectory as novel beneficial alleles arising within the population are shared; they are not merely a group of bacterial strains that exhibit similarity imparted by a common $(great)_n$-grandparent. Above the level of the species, evolutionary trajectories are far more distinct, as genetic information is shared far less frequently and in a more diffuse fashion via horizontal gene transfer (Ochman et al. 2000), although transfers may more frequently involve members of the same large taxonomic groups (Gogarten et al. 2002); as a result, similarities among individuals of the same family or division reflect primarily their shared ancestry, not an active process of cohesion.

Overall, the degree of genetic diversity in bacterial populations represents the balance of mutation, selection and recombination. The accumulation of mutations increases the nucleotide diversity of a set of strains; but, diversity may be reduced in two ways. First, a selective sweep of a population by a single strain may occur if a beneficial mutation arises within that strain (Fig. 1), initiating a periodic selection event (Levin 1981). Here, all genes within this candidate genome go to fixation within the population by hitchhiking on the one gene bearing a beneficial mutation (Cohan 2001). The extent of such a selective sweep is limited by the ability of the strain carrying the beneficial allele to out-compete ecologically similar individuals.

Alternatively, individual alleles may themselves experience selective sweeps by being transferred among strains via homologous recombination (Guttman 1997;

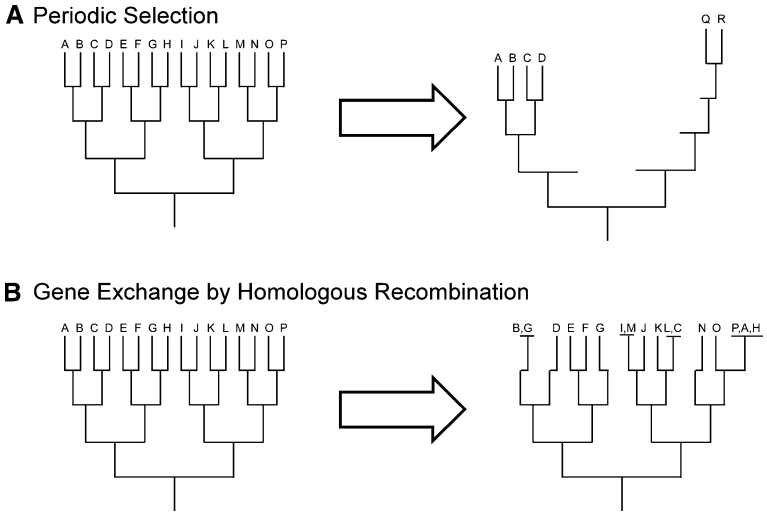

Fig. 1 Models for maintaining genotypic cohesion. In the periodic selection model the arrival of a beneficial mutation in lineage 'P' leads to the extinction of lineages 'E'–'O', thus reducing variability among strains in this population. Strains 'A'–'D' represent a different ecotype and are not out-competed in the periodic selection event. In the recombination model, variability is reduced at individual loci as allelic exchanges occur. Recombination events among strains B&G, I&M, L&C and P, A & H are denoted here. Figure after Lawrence and Retchless (2009)

Guttman and Dykhuizen 1994a, b). There is wide variability in the rate of allelic exchange within bacterial species, but in many strains—and perhaps in the majority—the likelihood that a variant allele is introduced by recombination exceeds the likelihood that a variant allele will arise by mutation (Feil et al. 2001; Feil and Spratt 2001). Thus allelic exchange between strains decreases overall diversity via position-specific selective sweeps (Fig. 1). (Selective sweeps do not occur at higher taxonomic levels since their members exploit different ecologies and do not experience homologous recombination.) If the rate of diversification due to mutation outweighs the rate of homogenization via recombination, then lineage diversification will occur. If the purging of variability by recombination keeps pace with the introduction of variant alleles by mutation, then genotypic cohesion will result, at least at those loci not encoding niche-specific genes, even within groups of strains that are ecologically too diverse to allow periodic selection of a single genotype to purge variability at all loci. Recent studies have examined the influence of these parameters on lineage formation within simulated neutrally evolving populations (Falush et al. 2006; Fraser et al. 2007).

Current exploration of bacterial species concepts typically encompass strains whose diversity is limited by one of these two mechanisms. The ecotype model focuses on groups of very closely related strains, wherein periodic selection events allow particular genotypes to sweep populations (Cohan 2001). Here, periodic selection events are limited to strains which are ecologically identical to the strain wherein the beneficial mutation arose. Strains belonging to different ecotypes are, by definition, not affected by the selective sweep. In this way, ecotype boundaries

have been equated to species boundaries (Gevers et al. 2005). Yet practical difficulties lie in defining ecotype boundaries: how different must strains be to classify them as ecologically distinct? More importantly, how can different ecotypes represent distinct bacterial species if recombination is actively producing similarity among them? That is, recombination acts as an active cohesion mechanism above the level of the ecotype. Because such transfers mediate selective sweeps and not merely neutral exchanges, the evolutionary trajectories of ecotypes are not independent. This is not to say that ecotypes are not meaningful levels of organization within bacterial population. Rather, we must focus our attention at domains of gene exchange above the level of the ecotype if we are to map the evolutionary history of microbes onto a potential Tree of Life.

The high frequency of recombination relative to mutation in some bacteria led Dykhuizen and Green (1991) to extend Mayr's Biological Species Concept; here, bacterial species are delimited by the extent of allelic exchange among strains, even if exchange occurs on a gene-by-gene basis. Boundaries are established by examining phylogenetic trees of orthologous genes found in different strains. If trees for different genes are not congruent (that is, the taxa show different relationships), then the strains are proposed to belong to the same species, wherein allelic exchange has led to different sets of relationships for each gene. In contrast, if phylogenies are congruent then strains belong to different species. Dykhuizen and Green offered conflicting topologies for three genes found in *Escherichia coli* as evidence for recombination within this long-recognized species (Dykhuizen and Green 1991). Evidence for the lack of gene exchange between groups was provided by Wertz et al. (2003); here, several strains of well-separated groups of enteric bacteria were characterized at several genes. The relationships of the strains within the named species depended on which gene one examined, consistent with the Dykhuizen and Green observation. But strains from different species remained in separate clades and the relationships between these clades were, for the large part, congruent. Therefore, allelic exchange had taken place only between strains within each named species, not between strains belonging to different species. This was consistent with the Dykhuizen and Green species concept.

Despite the attractive simplicity and clarity of this model, it lacks rigor. In the Mayrian model, species are positively defined, whereby organisms that shared a common gene pool belong to the same species; organisms which fail this criterion belong to different species. In contrast, the Dykhuizen and Green model effectively establishes when organisms belong to different species; but, the criteria for belonging to the same species are less clearly defined. Between species, congruent phylogenies of shared genes imply that gene exchange has not occurred; therefore, organisms would not share the same gene pool and must belong to different species. Yet incongruent phylogenies for shared genes merely show that the organisms' ancestors had a history of gene exchange; they do not speak to the potential for gene exchange among extant strains. It may be true that the strains could still exchange genes, and therefore belong to the same species; but it also may be true that they do not. That is, a history of gene exchange does not necessitate ongoing gene exchange.

Moreover, that potential for gene exchange need not be equal among all genes in the genome. For example, three strains of *Escherichia coli* may share less than 40%

of their combined gene pool, where each strain bears hundreds of genes not found in other strains (Welch et al. 2002). The broad distribution of some of these type-specific genes—such as the common occurrence of O157-specific pathogenicity islands among enterohemorrhagic *E. coli* that are absent from non-pathogenic *E. coli* (Kudva et al. 2002a, b; Ogura et al. 2009)—suggests that these genes play ecology-specific functions; their removal via recombination would be disadvantageous, raising the possibility that gene exchange at this locus would be counterselected. Therefore, how can this group of organisms be fully exchanging genes at all loci—which is an implication of the Dykhuizen and Green species concept—if such recombination events are counterselected at some loci? This potential disparity reflects a fundamental difference in the nature of recombination between Bacteria and eukaryotes. Although gene exchange in eukaryotes is an all-or-none event involving entire haploid genomes, genetic exchange in bacteria occurs only within selected regions within the genome at any one time.

If recombination were not permissible among all strains at all loci, then groups of strains that are recombining at some loci would be partially isolated (at some genes) from other such groups within the same species. At one conceptual extreme, any strains being partially isolated in this way could be termed a species because they would be performing a somewhat different ecological role owing to their differing physiological repertoires; here, the species concept has been reduced to the ecotype, wherein all strains are ecologically identical, gene exchange is permissible at all loci, and periodic selection limits genetic diversification. This is elegant and theoretically well-founded but not practical, as it would fragment named species into hundreds or thousands of groups, thereby defeating the utility of species identification as a tool to predict the general properties of organisms one may encounter. At the other extreme, any strain still recombining at any locus could be included in a species; this is taxonomically practical (allowing a Dykhuizen and Green method of species evaluation) but not theoretically sound, as numerous partially isolated, ecologically distinct groups would be classified under a single name. Strains that recombine at only a few genes would be classified under a single species label, despite the near-total genetic isolation that would result from the acquisition of numerous lineage-specific genes. In the most extreme cases, two populations could be included in the same species due to their shared ability to recombine with a third population, even if they recombine at non-overlapping sets of loci and therefore have no way to acquire DNA from each other. This would defeat the purpose of the "species" name in allowing it to report faithfully on the physiological capabilities of the bacterial strains belonging to that group.

If this scenario holds true, then species delineation must follow an intermediate path, including only those strains which recombine at large numbers of genes, excluding those which recombine to lesser degrees. Such a 'fuzzy' species boundary—as described for *Neisseria* (Hanage et al. 2005)—introduces the specter of arbitrary delineation of species boundaries. Division into nested groups (e.g., subspecies or tribes) yields the same problem: there are no clear boundaries that warrant demarcation. Rather than being unusual cases that reflect ecological or biogeographical oddities (e.g., eukaryotic ring species), such "fuzzy" boundaries are recognized in every taxon where sufficient numbers of strains are examined. In

addition, this difficulty in defining species boundaries leads to an even larger problem when examining the course of lineage separation. If the process of ceasing recombination at all loci is rapid, then the step-wise isolation of lineages (that is, successive lineage splitting events) is not problematic. Descendent lineages will be isolated before the next lineage splitting event is initiated. But if it is gradual, then lineages experiencing a slow separation may begin to split again before they themselves have been completely isolated. This would lead to different evolutionary histories for different genes in the chromosomes of the descendent lineages. It is these two problems that we explore below.

Bacterial speciation: species in pieces

Recombination interference provides a robust model for the stepwise separation of lineages (Lawrence 2002). Independent of selection, the accumulation of nucleotide polymorphisms will reduce the efficacy of recombination. In addition, the acquisition of lineage-specific characters—either point mutational differences that lead to functional distinctiveness, or the gain or loss of genes that confer a selectable difference—prevents recombination in several ways (Fig. 2). First, recombination from donors lacking these characters would eliminate them; if the genes are providing an important function, such recombinants would be counter-selected. Second, movement of these genes into lineages that lack them would introduce features that the recipient would not find useful so that the transfer would be, at best, neutral; they could even perform potentially problematic functions, leading to counterselection of this recombinant. For example, the loss of the *Shigella ompT* gene is advantageous, and reintroduction leads to a disadvantageous decrease in virulence (Nakata et al. 1993). Because recombinants at niche-specific loci would be

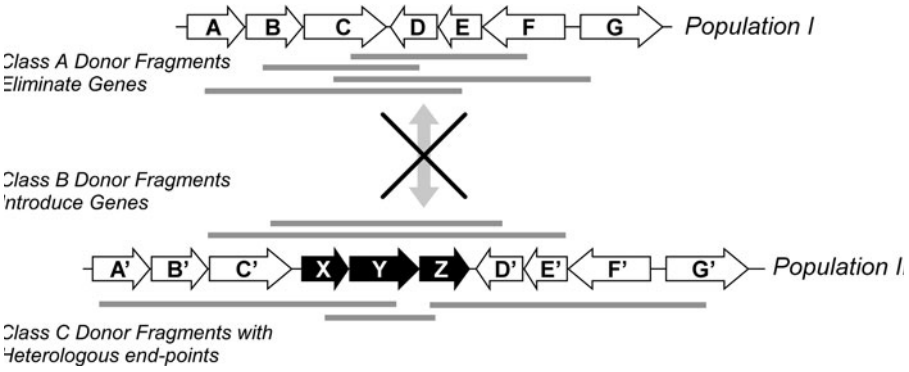

Fig. 2 Acquisition of lineage-specific genes interferes with recombination between ecologically dissimilar organisms. The gain of genes X, Y and Z in one population can interfere in several ways with recombination between populations. Movement of fragment class A would result in loss of genes X, Y & Z from population II; movement of fragment class B would result in potentially problematic gain of genes X, Y & Z in population I, should their expression there be counterproductive (see text). Introduction of fragment class C does not result in recombination due to a lack of homologous sequences at the dsDNA end. Figure after Lawrence (2002)

counterselected, lowered rates of recombination would also be observed at shared genes that lie adjacent to lineage-specific genes (Fig. 2). The lack of recombination among strains would prevent selective sweeps from crossing this "species" boundary at genes proximal to niche-specific loci. Because gene transfer is limited in size, this deficit would be observed only in the region immediately proximal to the niche-specific locus. As a result, shared genes accumulate differences between lineages, but only at those loci that are linked to lineage-specific genes. As more and more lineage-specific genes are acquired at different locations across the chromosome, the genomes of the two newly separating lineages become genetically isolated (Lawrence 2002).

An example of such a position-specific barrier to recombination is seen at the *Escherichia coli rfb* locus (Milkman et al. 2003). This large operon encodes proteins which direct the synthesis of the O-antigen polysaccharide. The O-antigen is the outer-most structure of the cell wall in gram-negative bacteria and is the most abundant molecule on the cell surface. It is recognized by potential predators, including predatory amoebae within the intestinal lumen as well as white blood cells when bacteria invade host tissues (Wildschutte et al. 2004). The *rfb* operons in different strains of *E. coli* and *Salmonella enterica* are hypervariable (Stevenson et al. 1994; Xiang et al. 1993), whereby numerous alleles are found in the population. Because predators in different habitats have different preferences based on the nature of their prey's O-antigen (Wildschutte and Lawrence 2007), this variability enables strains to escape predation by amoebae that are found in different environments. As a result, diversifying selection prevents any one *rfb* allele from sweeping the population, as each allele provides a selective advantage in only a specific set of environments.

When we examine nucleotide diversity within a set of *E. coli* strains, we see that the average divergence of genes increases in the vicinity the *rfb* locus (Fig. 3); this increase reflects the lack of selective sweeps which would purge variant alleles, consistent with the counterselection of strains that experienced recombination events in the vicinity of the *rfb* locus (Butela and Lawrence 2009). Thus the *rfb*

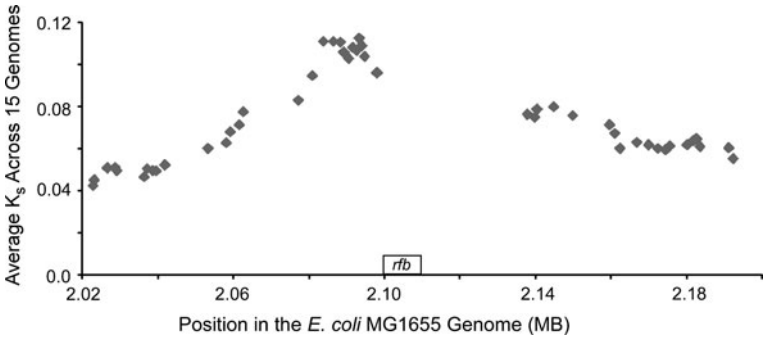

Fig. 3 Average divergence at synonymous sites for genes flanking the *E. coli rfb* operon. K_s values were averaged for all pairwise comparisons of genes shared among 16 completely sequenced *E. coli* genomes; values are plotted by the position of the gene in the *E. coli* K12 genome. Figure after Butela and Lawrence (2009)

locus represents a position-specific barrier to recombination. We propose that, over time, groups of strains harboring different O-antigens could diverge into separate species; if so, then this locus would represent a "pioneer gene" in the exploration of a newly distinct habitat, initiating the process of imparting genetic isolation to all regions of the genome.

This model of stepwise speciation can be tested using genome sequence data from two sibling "species." To begin, we need to establish that the genes shared between the two species show different times of divergence. Regardless of one's model for lineage diversification, the size-limited nature of bacterial gene exchange means that different genes must experience their last recombination events at different times. This was done by using the degree of codon usage bias as a time-independent metric for selection on synonymous sites, thereby generating an expected value of divergence at synonymous sites (Retchless and Lawrence 2007). As a model system, the genomes of the enteric bacteria *E. coli* and *S. enterica* were analyzed. By comparing observed divergence values to expected levels of divergence, regions of the chromosomes were identified which had been separated (not experiencing recombination) for greater or lesser periods of time (Retchless and Lawrence 2007). The correlation of relative divergence time with genomic position, whereby physically proximate genes showed common divergence times (Fig. 4), confirmed that the data reflected the time to the last recombination event and not merely stochastic variation in divergence values. A lack of similar structure in non-recombining bacteria argues against a mechanistic source for position-specific variation in mutation rates (Retchless and Lawrence 2007).

From this analysis, the local region of recombination interference measured about 10–20 genes (Fig. 4), or about the size of transducing fragments carried by lambdoid bacteriophages. Moreover, genetic isolation took place very slowly. Remarkably, when benchmarked to an average time of lineage separation of 140 MYr ago (Ochman and Wilson 1988), the time frame of lineage separation spanned more than 100 MYr. The earliest diverging genes were 1.4-times as divergent as expected, corresponding to a time of separation of $\sim$200 MYr ago, while the latest diverging genes were less than one-fifth as divergent as expected, corresponding to a separation time of only 20 MYr. Between these extremes, the separation of 95% of the genes shared between the *E. coli* and *Salmonella* lineages took place during a 70 MYr period. This time period of lineage separation (70 MYr

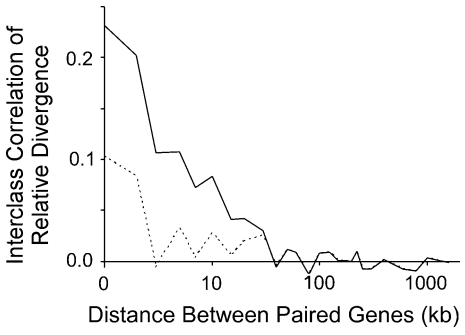

Fig. 4 Intraclass correlations of relative divergence for gene pairs as a function of distance (in kb); solid line, all gene pairs; dotted line, gene pairs not within runs of consecutive genes transcribed in the same direction. Figure after Retchless and Lawrence (2007)

for the majority of genes, 180 MYr for the entire chromosome) is far longer than the near instantaneous isolation of species that accompanies a traditional allopatric speciation event envisioned for eukaryotic lineages. For bacteria that divide only once per day, the course of 100 MYr represents 3×10^{10} generations. Indeed, the entire mammalian radiation occurred within the time frame of this single bacterial 'speciation' event.

To test whether selection for the maintenance of different sets of genes was responsible for the onset of genetic isolation in the nascent *E. coli* and *Salmonella* lineages, the relative divergence of orthologous genes was then assessed relative to the position of lineage-specific or strain-specific genes (Retchless and Lawrence 2007). In addition to the common backbone of genes found in strains of both species, two other sets of genes were identified. Some genes were lineage-specific, and were found in all strains of one lineage but none of the other. We posit that a subset of these lineage-specific loci were acquired in the early stages of lineage separation. In other cases, a more disjoint distribution—e.g., the gene is found in only one strain of one lineage—suggests that the gene was acquired long after lineage separation. By the model proposed above, genetic isolation would be initiated by the acquisition of lineage-specific genes or mutations. Therefore, lineage-specific genes—those found in multiple strains of one lineage but absent from the other—would be preferentially located in 'older' regions of the genome, adjacent to genes which diverged earlier than average. In contrast, recently acquired genes should have inserted at random with respect to the age structure of the shared gene backbone. This is what was observed, where lineage-specific genes were preferentially inserted adjacent to genes which had diverged earlier (Retchless and Lawrence 2007); thus, these data support the hypothesis that genetic isolation of early-diverging genes was initiated by the acquisition of ecology-defining traits.

Asynchrony in lineage separation

We have argued that different genes in bacterial genome diverge at different times, and that genetic isolation is driven by the acquisition of lineage-specific (presumably niche-specific) genes. This asynchrony in the genetic isolation of different genes within a genome has the potential to confound organismal phylogeny when considering successive lineage diversification (speciation) events. If speciation events occur within daughter lineages only after they have established complete genetic isolation, then no problems will arise (Fig. 5). This will occur if internodes in phylogenetic trees represent periods of time longer than the interval required to establish genetic isolation ($\sim$100 MYr in the case discussed above). When such well-separated species are examined (e.g., Wertz et al. 2003), congruent phylogenies will result and the Dykhuizen and Green model will be supported.

Yet if speciation events occur more rapidly than the interval required to establish genetic isolation, then nascent lineages will begin to split into descendent lineages—that is, develop genetic isolation between themselves—before they are completely separated from their evolutionary siblings. Consider a lineage splitting twice to form three descendent taxa (X, Y and Z in Fig. 5). Now, consider the

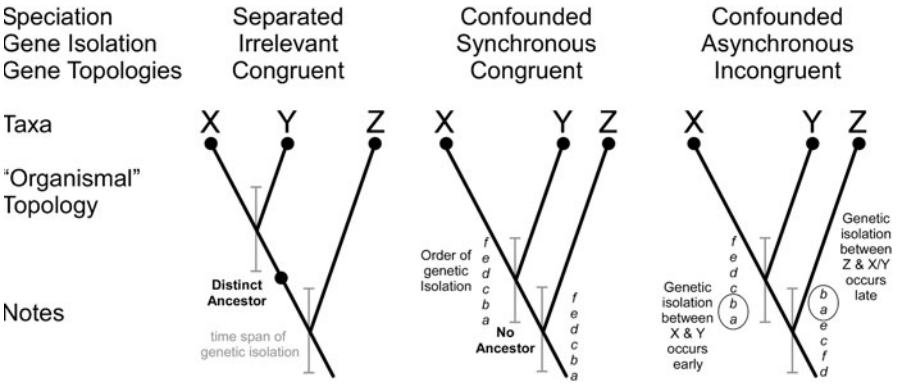

Fig. 5 Ambiguity in lineage separation. In the fragmented speciation model, different genes in the chromosome may acquire genetic isolation at different times. With sufficient time between consecutive lineage separations, a distinct ancestral state may be established at all loci, and phylogenies would be congruent (*left*). If consecutive lineage separations overlap chronologically, congruence can be maintained only if loci become isolated in the same order during both separations (*center*). If the order of genetic isolation is not identical during both separation processes (*right*), then incongruence will result

population ancestral to lineages X and Y beginning to diversify to form these taxa, but before it has separated completely from the lineage forming taxon Z. Congruent relationships among genes will result *only* if genetic isolation occurs in the same temporal order for genes when forming lineages X and Y as when lineage Z is formed. Aside from some hyper-variable loci that are regularly under diversifying selection, we know that genetic isolation is driven by the acquisition of position-specific adaptive changes (e.g., acquired genes as shown in Fig. 2). Therefore, this is unlikely to be the case. Rather, it is highly likely that genes will diverge asynchronously, leading to phylogenetic incongruency as the result of the protracted period of lineage isolation. In Fig. 5, this is indicated by the genetic isolation of genes *a* and *b* occurring between lineages X and Y before one of those lineages has become isolated from lineage Z.

We tested this prediction by examining the relationships between three clades of enteric bacteria (Fig. 6; Table 1): *Escherichia* spp. (Esc), *Salmonella enterica* (Sen) and *Citrobacter* spp. (Cit). Using other enteric bacteria as outgroups (see Fig. 6), there are three possible phylogenies: *Escherichia* will group with *Salmonella*, *Escherichia* will group with *Citrobacter*, or *Salmonella* will group with *Citrobacter*. These three strains show very similar distances from one another (note the short internodes in Fig. 6), suggesting that the successive speciation events will have occurred relatively rapidly when compared to the time frame of complete lineage separation. If each speciation event were independent, we would expect the majority of genes to support a single topology (as in Wertz et al.), with perhaps the phylogenies of a few genes supporting alternate topologies (though less robustly) due to convergence, ambiguity in the data or gene transfer between separated species. Instead, as expected from our model, we found significant number of genes that robustly supported each of the three possible rooted phylogenies (Table 1,

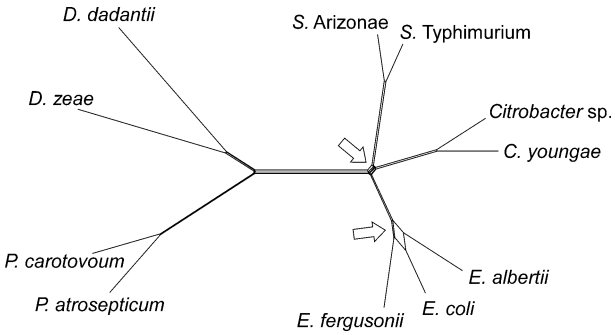

Fig. 6 Relationships among selected enteric bacteria. Dendrogram was constructed by SplitsTree (Huson and Bryant 2006) using 1185 aligned protein-coding genes shared among all taxa; genes were chosen as being reciprocal best matches with sufficient conservation of flanking gene context to ensure orthology. Positions of ambiguous branches discussed in the text are noted with *arrows*. Strain identities are indicated in Table 1

Table 1 Support of organismal topology

Bootstrap support (%)	Number of genes supporting topology[a]		
	(Cit, Sen), Esc	(Cit, Esc), Sen	(Sen, Esc), Cit
50	548	308	267
70	327	150	126
90	126	35	41
	(Efe, Eco), Eal	(Efe, Eal), Eco	(Eco, Eal), Efe
50	615	76	529
70	500	37	458
90	300	8	379

[a] Lineages were defined by one or two genomes: Cit, *Citrobacter* sp. 30_2 and *C. youngae* ATCC29220; Sen, *Salmonella enterica* Typhimurium LT2 and *S. enterica* Arizonae 62:z4; Esc, Eco and Efe; Eco, *Escherichia coli* K12; Efe, *Escherichia fergusonii* ATCC 35469; Eal, *Escherichia albertii* TW07627. Outgroups included *Dickeya dadantii* Ech703, *D. zeae* Ech1591, *Pectobacterium carotovorum* PC1 and *P. atrosepticum* SCRI1043 (Fig. 6); these taxa were sufficiently distinct from other lineages that branching order was unambiguous

upper lines). Because similar numbers of genes supported each of the alternative topologies when the stringency of bootstrap support was increased, we conclude that this phylogenetic ambiguity is not merely the product of noisy data sets.

Further analyses show that this pattern is robust in the choice of outgroup, and is not the result of recombination occurring between extant lineages (rather than at the time of lineage separation). In addition, it is unlikely that this pattern has resulted from incomplete lineage sorting of ancestral polymorphisms (e.g., see Pamilo and Nei 1988). For ancestral polymorphism to explain our data, the majority of loci in the chromosome would need to be sheltered from selective sweeps for far longer than has been observed in extant populations of *E. coli* and *Salmonella* (A. C. Retchless and

Table 2 Number of genes supporting each topology with 70% bootstrap support

Topology for enteric clade	Topology for Escherichia clade[a]			
	(Efe, Eco), Eal	(Efe, Eal), Eco	(Eco, Eal), Efe	Total
(Cfr, Sen), Esc	130	9	115	254
(Cfr, Esc), Sen	54	5	61	120
(Sen, Esc), Cfr	47	6	38	91
Total	231	20	214	465

[a] Lineages as defined in Table 1

J. G. Lawrence, unpublished results). Moreover, genes and sites supporting each of the topologies are clustered on the chromosome, pointing to recombination as the source of the phylogenetic disparities, not random convergence (A. C. Retchless and J. G. Lawrence, unpublished results). Taken together, these data support the hypothesis that the history of recombination has yielded genes with different evolutionary histories being found in the same genome.

In addition, we examined the divergence of the distinct but closely related groups in the genus *Escherichia*: *E. coli, E. fergusonii and E. albertii* (Walk et al. 2007, 2009). The split between these three groups occurred after their separation from the *Salmonella* lineage (Fig. 6). Similar results were obtained, whereby multiple genes supported each of the three alternative phylogenies (Table 1, lower lines). Fewer genes showed *E. albertii* and *E. fergusonii* as sister taxa (Table 1), reflecting rare recombination between these lineages; we attribute this deficit possibly to likely lower population sizes of these clades relative to *E. coli,* as evidenced by their surplus of synonymous substitutions (Walk et al. 2009), which would cause *E. coli* to be the dominant recombination partner. Considered together, this set of five strains has nine possible rooted phylogenies, assuming monophyly of *Escherichia* in this analysis; we observe all nine topologies with a good degree of confidence at each of the two branches of interest (70% bootstrap, Table 2), with eight of these topologies supported by genes with very strong bootstrap values (90% bootstrap support; data not shown). These data support the conclusion that the different genes within a single genome acquire different evolutionary histories during the process of lineage formation itself. A single 'organismal' topology will fail to represent this diversity; if the 'majority rule' tree were to be constructed, one supported by 130 of 465 orthologous genes (Table 2), this would represent the relationships of a minority (28%) of the genes in the genome. Lateral gene transfer between distantly related genomes will further confound organismal relationships, but these data strongly suggest that an unambiguous 'Tree of Life' never existed in the first place.

The inevitable failure of the tree of life

This data-bound ambiguity in organismal relationships is not equivalent to the perceived ambiguity that results from incomplete lineage sorting (Joly et al. 2009; Pamilo and Nei 1988). In the latter case, ancestral polymorphisms have not gone to

fixation following a lineage splitting event when a second split occurs. As a result, the phylogenies of genes may not be congruent. Here, lineage splitting events are distinct events and ancestral populations were genetically isolated. It is merely ambiguity in the extant data—a result of the differential fixation of genes in the resulting lineages—that makes the speciation events difficult to reconstruct. In the case of fragmented bacterial speciation, where recombination is subject to locus-level influences, there never was an established ancestral population. The ancestral population represented by nodes within the Tree of Life simply did not exist. Indeed, the same mechanisms which can lead to rapid genetic isolation in eukaryotes—e.g., strong assortative mating—act to drive fragmented speciation in bacteria (Fig. 2).

This obligate reticulation in organismal relationships effectively invalidates hierarchical classification. The central supposition that biological species can be classified as 'groups within groups' rests on the idea that lineage splitting events are independent of one another, and they are not; descendent taxa are not separated from one another before they begin splitting again. As a result, descendent populations cannot be traced back to an ancestral population represented by a node in a phylogenetic tree. These "coalescent" species implied by phylogenetic inference methodologies simply did not exist. In the absence of a phylogenetic network, such ambiguous nodes could be collapsed into a single trifurcation, connecting three descendents to a single node.

Such trifurcations could occur in eukaryotic lineages as well, whereby a single species simultaneously splits into more than two descendent lineages. These events do not invalidate hierarchical classification; they merely assign a single common ancestor to more than one descendent lineage. The situation in bacteria differs in two important ways. First, the obligatorily reticulate nature of the lineage splitting event has invalidated the concept of the ancestral taxon. In the case of *Escherichia*, *Salmonella* and *Citrobacter*, we are not suggesting that the three lineages separated simultaneously from a single ancestral population. Rather, there were indeed two successive lineage splitting events, but the putative ancestor formed by the first event did not exist as an evolutionarily distinct group. Second, the long time frame for lineage separation combined with the immense diversity of bacterial life effectively means that *all* branches on the bacterial tree of life are confounded in this way. That is, we are not simply collapsing three descendent taxa as a single trichotomy. Due the very long time frame of lineage separation in Bacteria, essentially all branches of the Tree of Life will be similarly confounded, at least in those branches containing bacteria which exchange genes by homologous recombination; this process has been detected in all free-living bacteria and archaea, being absent only from obligate intracellular parasites. Because rates of recombination vary between taxa (Feil et al. 2001; Maynard Smith et al. 1993), the degree of lineage entanglement will also vary, being most pronounced in groups experiencing higher rates of recombination. While the resulting phylogenetic failures are evident now only among the most heavily populated regions of the Bacterial tree, this simply reflects sampling bias; as more taxa are added to the tree, more problematic taxa will arise. If we were to collapse one problematic branch of the bacterial Tree of Life, we would inevitably do so for all its branches. If a well-resolved tree can

only be constructed for branches that are spaced far apart, pruning inconvenient internal lineages that confound well-ordered branching patterns, then such a tree is not useful. More importantly, the topology of the resulting tree would depend on which branches were pruned.

Lastly, the reticulate nature of lineage splitting casts a shadow over the use of recombination as a cohesion mechanism for bacterial species. Because lineage separation takes such a long time, existing bacterial groups necessarily comprise numerous emerging lineages, each in partial genetic isolation with its siblings. That is, ecologically distinct sets of strains within bacterial "species"—such as enterohemorrhagic vs. commensal *E. coli*, or those preferentially colonizing different groups of mammals (Gordon and Cowling 2003)—represent nascent species which both recombine at loci shared between all subgroups and maintain sets of niche-specific genes that are not shared with strains outside the subgroup. Bacterial groups defined by the presence of a subset of freely recombining genes are also fragmented by sets of lineage-specific genes. Therefore, according to a Mayrian or Dykhuizen and Green species concept, such strains may belong to the same species at some genes but different species at others. This situation is a direct result of the biology of bacteria whereby gene exchange at different loci occur via different events at different times. Thus the biology of bacteria precludes the use of traditional concepts of species and speciation, and necessarily prevents a Tree of Life from being applied to bacteria, at least in its conventional formulation.

Implications of the failed model

One could argue that the failure of the Tree of Life model we outlined above is a technicality, that evolution is still described—at least in broad terms—by a bifurcating structure whereby descendent taxa show similarity due to their relationship with a common ancestor. This was, after all, the main point proffered by Darwin, that the Tree of Life explains organismal diversity in the context of the evolutionary process. Yet phylogenetic trees serve as frameworks for interpreting evolutionary change, and it is in this arena where a confounded Tree of Life can be problematic. In their original, unambiguous forms, phylogenetic trees summarizing organismal history posit the existence of well-resolved ancestral taxa. On this framework, one can examine the distribution of variable characters. If most members of a clade share a character that is missing from one member, then a parsimonious explanation is that the character has been lost from that taxon (Fig. 7). In our example, the model supposes that taxon A lost a gene that is present in taxa B and C. The implication is that the ancestor of taxa A and B carried that gene and that it was lost as taxon A became a distinct lineage, indicating that this gene no longer served an important purpose in the nascent taxon A lineage.

Yet this model requires that the set of 'organismal' relationships dictate the ancestry of all genes in the genome, and we now know that this is far from the case. When viewed through the lens of temporally fragmented speciation, other possibilities arise. For example, the gene in question may have been acquired by lineages B and C during the time of lineage separation, when they were still in

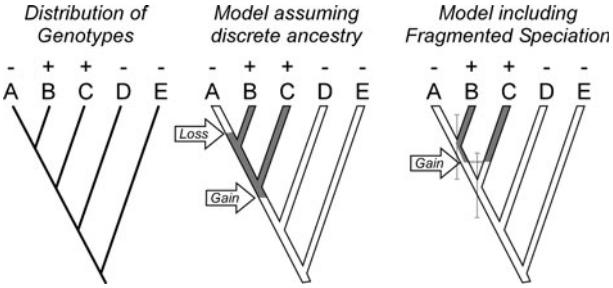

Fig. 7 Ambiguity in lineage separation leads to problems in interpreting character states. The assumption of distinct lineage-forming events implies the fixation of a variant at some point in history with subsequent regeneration of variation (*center*). A complex lineage-forming process facilitates the retention of ancestral diversity in the emergent taxa (*right*)

genetic communication at this location in the chromosome even though lineages A and B were genetically isolated (see Fig. 5). Therefore, a single acquisition event may explain the data (Fig. 7). According to this scenario, taxon A did not lose the gene in question because its genome never contained it. Here, parsimony fails because it relies on an established framework from which the likelihood of alternative events can be interpreted. The 'fuzziness' of species boundaries has reduced our confidence in this framework, thus making our deciphering of the evolutionary process based on interpretation of extant characters all the more difficult.

Speciation and the myth of bacterial species

By our argument, the gradual, diffuse process of bacterial lineage separation has two effects. First, it invalidates the use of well-resolved, bifurcating trees to summarize the relationships among bacterial taxa. Second, the boundaries of bacterial species become difficult to define, as the propensity for high-frequency gene exchange is no longer an attribute that can be used to delineate species groups. As a result, species themselves are almost mythological, as extant species are either narrowly defined as groups that have specific properties of interest but are able to exchange genes with related species (e.g., the pathogen *Bacillus anthracis* is a biovar of *B. cereus*) or so broadly defined that they include numerous, ecologically distinct protospecies (e.g., *Escherichia coli*). Yet the delineation of bacterial species shapes our policy and response in the arenas of agriculture, geochemistry, food safety, disease diagnosis, epidemiology and public health, and bioterrorism. How can we reconcile these two positions, whereby species boundaries impact public policy and health care decisions, yet lack a firm conceptual basis for their identification?

Perhaps a solution can be found in closer inspection of extant species definitions. Unlike species concepts, whereby boundaries between groups reflect the limits of action of cohesion mechanisms, species definitions are practical descriptions of those organisms belonging to groups of our choosing. Species definitions are inherently arbitrary, but they are useful; more importantly, their utility derives from

their focus on characters which define a group according to properties we find important. For example, *Bacillus anthracis* may be distinguished from strains of *Bacillus cereus* only by the presence of virulence plasmids; the choice is arbitrary in that no other group of *B. cereus* strains is elevated to species status by virtue of their plasmid content. Yet other groups of *B. cereus* do not cause anthrax; while arbitrary, such a species definition is quite useful. Similarly, strains of *Shigella dysenteriae* could be considered to be simply strains of *E. coli*, albeit ones that cause particular illness. If such groups of strains exploit a particular ecological niche, and recombination events between such groups are counter-selected if they affect niche-specific genes, then species which were initially described via convenient and useful definitions may, in fact, correspond to protospecies experiencing the initial events in the long road to complete lineage separation. While species definitions lack the theoretical rigor of species concepts, they are practical, being founded in the recognition of similarity among members within species and meaningful distinctions between members of different groups. Bacterial species defined this way may be mythological, but they are eminently useful.

Acknowledgments This work was support by grant GM078092 from the National Institutes of Health. The paper was first presented at the workshop, Perspectives on the Tree of Life, sponsored by the Leverhulme Trust and held in Halifax, Nova Scotia, July, 2009.

References

Bapteste E, Susko E, Leigh J, MacLeod D, Charlebois RL, Doolittle WF (2005) Do orthologous gene phylogenies really support tree-thinking? BMC Evol Biol 5:33

Bapteste E, O'Malley MA, Beiko RG, Ereshefsky M, Gogarten JP, Franklin-Hall L, Lapointe FJ, Dupré J, Dagan T, Boucher Y, Martin W (2009) Prokaryotic evolution and the tree of life are two different things. Biol Direct 4:34

Boyd R (1999) Homeostasis, species, and higher taxa. In: Wilson R (ed) Species: new interdisciplinary essays. MIT Press, Cambridge, pp 141–185

Butela K, Lawrence JG (2009) Selection for antigenic diversity in *Salmonella*. In: Falush D, Robinson A (eds) Bacterial population genetics in infectious disease. Wiley-Blackwell, London (in press)

Cohan FM (2001) Bacterial species and speciation. Syst Biol 50:513–524

Darwin C (1859) On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life. John Murray, London

Doolittle WF (2000) Uprooting the tree of life. Sci Am 282:90–95

Doolittle WF, Bapteste E (2007) Pattern pluralism and the tree of life hypothesis. Proc Natl Acad Sci USA 104:2043–2049

Dykhuizen DE, Green L (1991) Recombination in *Escherichia coli* and the definition of biological species. J Bacteriol 173:7257–7268

Falush D, Torpdahl M, Didelot X, Conrad DF, Wilson DJ, Achtman M (2006) Mismatch induced speciation in *Salmonella*: model and data. Philos Trans R Soc Lond B Biol Sci 361:2045–2053

Feil EJ, Spratt BG (2001) Recombination and the population structures of bacterial pathogens. Annu Rev Microbiol 55:561–590

Feil EJ, Holmes EC, Bessen DE, Chan MS, Day NP, Enright MC, Goldstein R, Hood DW, Kalia A, Moore CE, Zhou J, Spratt BG (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

Fraser C, Hanage WP, Spratt BG (2007) Recombination and the nature of bacterial speciation. Science 315:476–480

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J (2005) Re-evaluating prokaryotic species. Nat Rev Microbiol 3:733–739

Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. Mol Biol Evol 19:2226–2238

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

Guttman DS (1997) Recombination and clonality in natural populations of *Escherichia coli*. Trends Ecol Evol 12:16–22

Guttman DS, Dykhuizen DE (1994a) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science 266:1380–1383

Guttman DS, Dykhuizen DE (1994b) Detecting selective sweeps in naturally occurring *Escherichia coli*. Genetics 138:993–1003

Hanage WP, Fraser C, Spratt BG (2005) Fuzzy species among recombinogenic bacteria. BMC Biol 3:6

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Joly S, McLenachan PA, Lockhart PJ (2009) A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am Nat 174:E54–E70

Koonin EV, Wolf YI (2009) The fundamental units, processes and patterns of evolution, and the tree of life conundrum. Biol Direct 4:33

Kudva IT, Evans PS, Perna NT, Barrett TJ, Ausubel FM, Blattner FR, Calderwood SB (2002a) Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. J Bacteriol 184:1873–1879

Kudva IT, Evans PS, Perna NT, Barrett TJ, DeCastro GJ, Ausubel FM, Blattner FR, Calderwood SB (2002b) Polymorphic amplified typing sequences provide a novel approach to *Escherichia coli* O157:H7 strain typing. J Clin Microbiol 40:1152–1159

Lawrence JG (2002) Gene transfer in bacteria: speciation without species? Theor Popul Biol 61:449–460

Lawrence JG, Retchless AC (2009) The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. Methods Mol Biol 532:29–53

Lederberg J (1947) Gene recombination and linked segregations in *Escherichia coli*. Genetics 32: 505–525

Lederberg J, Tatum EL (1946) Gene recombination in *Escherichia coli*. Nature 158:558

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

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

Mayr E (1942) Systematics and the origin of species. Columbia University Press, New York

Mayr E (1954) Change of genetic environment and evolution. In: Huxley JS, Hardy AC, Ford EB (eds) Evolution as a process. Allen and Unwin, London, pp 156–180

Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge

Nakata N, Tobe T, Fukuda I, Suzuki T, Komatsu K, Yoshikawa M, Sasakawa C (1993) The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: the relationship between the *ompT* and *kcpA* loci. Mol Microbiol 9:459–468

Ochman H, Wilson AC (1988) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol 26:74–86

Ochman H, Lawrence JG, Groisman E (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304

Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T (2009) Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc Natl Acad Sci USA 106:17939–17944

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. Mol Biol Evol 5:568–583

Retchless AC, Lawrence JG (2007) Temporal fragmentation of speciation in bacteria. Science 317:1093–1096

Stevenson G, Neal B, Liu D, Hobbs M, Packer NH, Batley M, Redmond JW, Lindquist L, Reeves P (1994) Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. J Bacteriol 176:4144–4156

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

Van Valen L (1976) Ecological species, multispecies, and oaks. Taxon 25:223–239

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

Vulic M, Lenski RE, Radman M (1999) Mutation, recombination, and incipient speciation of bacteria in the laboratory. Proc Natl Acad Sci USA 96:7348–7351

Walk ST, Alm EW, Calhoun LM, Mladonicky JM, Whittam TS (2007) Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. Environ Microbiol 9:2274–2288

Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, Whittam TS (2009) Cryptic lineages of the genus *Escherichia*. Appl Environ Microbiol 75:6534–6544

Welch RA, Burland V, Plunkett G 3rd, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J, Stroud D, Mayhew GF, Rose DJ, Zhou S, Schwartz DC, Perna NT, Mobley HL, Donnenberg MS, Blattner FR (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci USA 99:17020–17024

Wertz JE, Goldstone C, Gordon DM, Riley MA (2003) A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. J Evol Biol 16:1236–1248

Wildschutte H, Lawrence JG (2007) Differential *Salmonella* survival against communities of intestinal amoebae. Microbiology 153:1781–1789

Wildschutte H, Wolfe DM, Tamewitz A, Lawrence JG (2004) Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. Proc Natl Acad Sci USA 101:10644–10649

Wiley EO (1978) The evolutionary species concept reconsidered. Syst Zool 27:17–26

Xiang SH, Haase AM, Reeves PR (1993) Variation of the *rfb* gene clusters in *Salmonella enterica*. J Bacteriol 175:4877–4884

Zawadzki P, Roberts MS, Cohan FM (1995) The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. Genetics 140:917–932