Antonie van Leeuwenhoek **81:** 143–153, 2002.

Does experimental evolution reflect patterns in natural populations? *E. coli* strains from long-term studies compared with wild isolates

V. Souza[1,*], M. Travisano[2], P.E. Turner[3] & L.E. Eguiarte[1]

[1]*Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70–275, México D.F. 04510, México;* [2]*Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA;* [3]*Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA (*Author for correspondence; E-mail: souza@servidor.unam.mx)*

Key words: Escherichia coli, experimental evolution, genetic structure, population genetics, recombination, wild isolates

Abstract

Our results show that experimental evolution mimics evolution in nature. In particular, only 1000 generations of periodic recombination with immigrant genotypes is enough for linkage disequilibrium values in experimental populations to change from a maximum linkage value to a value similar to the one observed in wild strains of *E. coli*. Our analysis suggests an analogy between the recombination experiment and the evolutionary history of *E. coli*; the *E. coli* genome is a patchwork of genes laterally inserted in a common backbone, and the experimental *E. coli* chromosome is a patchwork where some sites are highly prone to recombination and others are very clonal. In addition, we propose a population model for wild *E. coli* where gene flow (recombination and migration) are an important source of genetic variation, and where certain hosts act as selective sieves; i.e., the host digestive system allows only certain strains to adhere and prosper as resident strains generating a particular microbiota in each host. Therefore we suggest that the strains from a wide range of wild hosts from different regions of the world may present an ecotypic structure where adaptation to the host may play an important role in the population structure.

Introduction

Surveys of wild populations have revealed that bacteria comprise the most diverse group among recognized taxa, and are found in virtually all habitats suitable for life. The combination of high diversity and broad distribution suggests that these features are a simple function of bacterial ability to adapt to a wide variety of environmental conditions. Whereas survey work can demonstrate the extreme diversity of bacteria and their ability to persist under many environmental conditions, experimental evolution provides a more powerful means to explore the underlying principles affecting the origin and maintenance of diversity. Experimental evolution uses microbes as living simulations, by allowing direct observations of evolutionary change over hundreds (or even thousands) of generations (Lenski 1995; Travisano & Rainey 2000). The experimenter chooses the environmental conditions and initial genotypes, ensuring that evolutionary outcomes are attributable to the particular selective conditions in which the experiment was performed. However, little effort has been made to combine both approaches to gain a general understanding of the factors affecting bacterial diversity. Here we compare survey data with those from laboratory selection experiments to examine one critical issue in the maintenance of bacterial diversity: the importance of recombination in bacterial evolution. We focus on recombination because of its potential to greatly increase or decrease extant diversity, depending upon the rate of recombination and its selective benefit. Our primary goal in this paper is to demonstrate the value of the comparison and to suggest the importance of further work. The limitations of this and other current studies will be made evident.

Recombination *in situ* and *in vitro*

The evolutionary importance of recombination in bacteria has been intensely debated. Bacteria repro-

duce clonally, but they may experience recombination through mechanisms of gene transfer, which are independent of cell division. The extent of recombination in natural populations of bacteria is unclear; bacterial populations may exist as very large inter-mixing gene pools that are analogous to panmictic species, or as individual lineages that are essentially isolated from one another (Whittam & Ake 1993). Initial population genetic studies on pathogenic and non-pathogenic *Escherichia coli* associated with humans and domestic or zoo animals suggested clonal structure (Ochman & Selander 1984), and high genetic diversity within species (heterozygosity (H) = 0.41 − 0.61, Whittam & Ake 1993; Pupo & Richardson 1995). However, the importance of recombination in *E. coli* has been detected at the molecular level through its signature in clonal frames (Milkman & McKane Bridges 1990; Guttman & Dykhuizen 1994). Extensive signals of recombination were found also in genome comparisons among *E. coli* (Perna et al. 2001), and between *E. coli* and *Salmonella* (Koshi et al. 2001).

The most extensive experimental evolution study has been that of Lenski and collaborators, in which 12 replicate populations of the bacterium *E. coli* B/6 were propagated for more than 25 000 generations (Lenski et al. 1991, Lenski & Travisano 1994). A single ancestral genotype was used to found 12 (hereafter referred to as long-term) populations maintained through serial culture, in minimal medium with glucose as the growth-limiting carbon and energy source. The forces shaping evolution in this experiment were limited to adaptation and chance, because the replicate populations were of large size, experienced the same selective conditions, and were initially isogenic. Recombination was not a factor because the founding cell was plasmid-free, and featured no detectable recombination.

Experimental evolution studies have often excluded recombination due to the added complexity inherent in this process. In the absence of recombination, evolution is strictly clonal and diversity is generated only through *de novo* mutations. Although *E. coli* has been the dominant model organism in bacterial molecular and population genetics (Selander et al. 1987), experimental studies of the evolutionary effects of recombination have not been performed. It is critical to examine recombination within the context of experimental evolution to provide insight into the importance of recombination in wild populations of *E. coli*.

To that end, we studied the effects of recombination on genetic and evolutionary change in the long-term populations that had previously experienced 7000 generations of strictly clonal evolution. In particular, we tested whether recombination would increase genetic diversity in the long-term populations, and whether the added variability would increase the rate of adaptation when mutation alone had forced the populations onto a selective plateau (Souza et al. 1997). To do so, one clone from each long-term population was propagated for 1000 additional generations in the presence of recombination (imposed every 5 days, or ∼33 generations). Controls were propagated in an identical environment but in the absence of recombination. Recombination was imposed using immigrant genotypes consisting of a mixture of four HFR (high frequency recombination) strains of *E. coli* K12 that differed from the recipients based on plate, MLEE, and metabolic (Biolog) markers (see below). Although in natural populations genetic exchange is likely to occur as or more frequently by phage mediated transduction than HFR recombination, our goal in this initial study was not to precisely replicate the mechanisms of recombination that occur in nature. Rather, we wished to investigate the potential effects of recombination on adaptation in the most easily controlled manner.

In a separate study we examined the prevalence and importance of recombination in wild populations of *E. coli* by gathering over 3000 isolates from different organisms across several continents; a representative sample of the collection was analyzed in terms of diversity, niche specialization, presence of plasmids, sugar utilization, and toxin and antibiotic resistance (Souza et al. 1999). We combined samples from the aforementioned laboratory study and field studies to examine the effects of recombination on bacterial population structure. Specifically, 120 experimental evolution laboratory strains and 215 wild-mammal associated strains of *E. coli* were used to estimate several population-genetic parameters: genetic diversity (H_e) and linkage disequilibrium (D', V_o/V_e). In addition, we analyzed the phylogenetic effect of recombination in a distance analysis where wild and experimental isolates were pooled. Neither the genetic analysis of the experimental isolates, nor the linkage disequilibrium analyses of wild and experimental populations have been published elsewhere.

Our results show that recombination in experimental laboratory populations can generate striking evolutionary patterns that show some similarities to

Table 1. Genetic markers of the parental strains in a long-term evolution experiment with recombination, where *Escherichia coli* B6 was the recipient and four Hfr *E. coli* K12 strains were donors.

	Plate markers[1]										Electrophoretic markers[2]						Biological markers[3]					
Strain	Ara	Lac	Tet[4]	Str	Arg	Leu	Ilv	Val	T1	T6	IDH	6PGD	ADH	MPI	PEP	FUC	AAC	FOR	SAC	SUC	GLU	SR
MAP[5]	1.4	10	–	64	40	76	76	76	4	11	26	41	27	36	6	60	91	80	16	17	82	51
B/6	–/+	+	s	r	+	+	+	+	s	r	2	1	1	1	2	0	3	0	2	1	3	3
288	+	+	84	s	+	+	–	–	s	s	1	2	2	2	1	3	3	2	3	0	0	3
291	+	–	89	s	–	+	+	+	r	r	1	2	2	2	1	3	3	2	0	0	0	3
296	–	–	2	s	–	–	+	+	r	s	1	2	2	2	1	3	1	2	0	0	0	1
298	+	–	62	s	–	–	+	+	r	r	1	2	2	2	1	3	2	1	2	0	0	3

[1]Plate Markers: Arabinose (Ara); Lactose (Lac); Resistance to tetracycline (Tet); Streptomycin resistance (Str); auxotrophy to Arginine (arg); Leucine (Leu); Isoleucine (Il); and valine (Val); Resistance to phages T1 (T1), and T6 (T6).
[2]Electrophoretic markers: isocitrate dehydrogenase (idh); 6-phosphogluconate dehydrogenase (6pgd); alcohol dehydrogenase (adh); mannose phosphate isomerase (mpi); and peptidase (pep).
[3]Biolog markers: 3 indicates activity within 24 h, 2 within 48 h, 1 within 72 h, and 0 for no activity by 72 h. The carbon sources were: L-fucose (fuc); acetic acid (aac); formic acid (for); D-saccaric acid (sac); succinic acid (suc); L-glutamic acid (glu); and D-serine (ser).
[4]Hfrs are resistant because each contains a TN10 transposon inserted in a different region of the chromosome several minutes after the origin of conjugation. The site of the tetracycline gene in the HFR is expressed in minutes, B/6 is sensitive to tetracycline.
[5]Origins of conjugation of the different Hfr's are situated in different regions of the chromosome before the Tn10 transposon. Hfr 288 has its origin at 67′, 291 at 2′, 296 at 12′, and 298 at 84′. The direction of conjugation is from the point of origin to Tn10; e.g., if the origin of Hfr 288 is at 67′ and Tn10 is at 84′, the direction of conjugation is upstream, whereas Hfr 291 has its origin at 2′ and Tn10 at 89′, making the direction of conjugation downstream.

those observed in nature. In particular, only 1000 generations of periodic recombination with immigrant genotypes was sufficient for linkage disequilibrium values in experimental populations to change from a maximum linkage value to that observed in wild *E. coli*. Moreover, we observed that recombination in both the field and laboratory studies generated a patchwork of genes laterally inserted in a common backbone. We propose a population model for wild *E. coli* where gene flow (recombination and migration) is an important source of genetic variation, and where certain hosts act as selective sieves; i.e., the host digestive system allows only certain strains to adhere and prosper as resident strains. More generally, this preliminary comparison of field and experimental studies indicates that additional studies are necessary to distinguish the causes of the patterns we observed.

Materials and methods

Strains and genetic markers in the recombination experiment

Bacterial strains and experimental procedures in the recombination study are described elsewhere (Souza et al. 1997). Briefly, we used a total of 18 markers (allozyme, plate, and MLEE) to distinguish donors (four *E. coli* K12 mutants that are unable to grow in minimal media) and recipients (*E. coli* B/6), and to gauge the production of recombinant genotypes at 100 generation intervals (i.e., every three bouts of imposed recombination); location of recombination events in the genome was based on the published genome map for the *E. coli* chromosome (Ridley & Labedan 1996). The present study uses Biolog metabolic markers and multi-locus electrophoresis of metabolic enzymes (MLEE) as a means to further examine the extent of recombination in the previous study. Biolog MicroPlates (Biolog Inc., Hayward, CA) test the ability to utilize 95 different carbon sources, yielding a characteristic metabolic fingerprint of the test strain. Strains were grown in a minimal glucose medium (DM25, medium used in the long term experiment) for 24 h, and 150 μl of the stationary culture was deposited in each well of a Biolog EG plate using a multi-channel pipetter. Plates were scored every 24 h for 3 days to assess indicator (color) changes. The wells that turned purple on day one were scored as '3', day two as '2', day three as '1', and no activity after 28 h as '0'. Each donor and recipient was tested with replication ($n = 4$), and similarly ($n = 2$) for each recombinant and control genotype. Of the 95 metabolic markers analyzed, only nine (two of which are identical to plate markers) were considered for the final analysis. These markers differed among donors and recipients and

were consistent across replicates, and their location in the chromosome is known (Table 1).

Strains and genetic markers in wild isolates

Vertebrate hosts and geographic origins of the wild *E. coli* isolates have been previously described (Souza et al. 1999). We examined 215 strains isolated from fresh feces of healthy wild animals: 94 mammalian species (representing 39 families and 14 orders from Australia, the Americas and reference strains from Africa), and 10 families of birds from Mexico. The analyzed loci and MLEE procedures are described by Souza et al. (1999), with the exception of one additional locus (G-6PDH, glucose 6-phosphate dehydrogenase, EC 1.1.1.49) (Herbert & Beaton 1993).

Phylogenetic analysis

We visualized the effect of recombination on genetic relatedness using a tree approach. To do so, we compared five divergent recombinant genotypes (generated through parental strains K12 and B/6) with two long-term genotypes isolated at 10 000 generations. To compare these experimental strains with natural isolates we randomly selected 100 electrotypes from the collected wild isolates and included these genotypes in the phylogenetic analysis. MLEE electrophoresis was performed for both the experimental strains and the wild strains using the same loci. A maximum parsimony tree was obtained using PAUP* (Macintosh beta test version 4.0b2; Swofford 2001), using a heuristic search. We chose an arbitrary tree from a large set of equally parsimonious trees in order to illustrate the effect of recombination in phylogenetic reconstruction.

Genetic diversity and linkage disequilibrium

Average genetic diversity was estimated as $H_e = \sum h_j/m$, where m equals the number of loci scored and $h_j = [n/(n-1)]\ (1 - \sum {p_{ij}}^2)$, where p_{ij} is the frequency of allele i at locus j, and n is the number of strains (Selander et al. 1986).

One way to measure linkage disequilibrium among pairs of loci is to obtain the coefficient D' between a pair of loci using the formula

$$D_{ik,jl} = g_{ik,jl} - p_{ij}\,p_{lk}$$

where $g_{ik,jl}$ is the frequency of two haploid genotypes, and $p_{ij}p_{lk}$ is the product of the frequencies of the corresponding alleles. This coefficient is standardized by dividing by the maximum disequilibrium obtainable considering only allele frequencies above 0.1 (D_{max}); the absolute value of this ratio is the index D'. Below we report the average D' for all pairs of loci using the optimum allelic frequency of 0.1 (Souza et al. 1992). In simulations with different sample sizes (data not shown), we observed that D′ is very sensitive to sample sizes less than 100 individuals, and a false signal of clonality was obtained in samples of 50 individuals or less.

Another useful index to describe multilocus linkage disequilibrium in haploid populations is based on the ratio of the variance (V_o/V_e) in mismatches observed for a particular sample (V_o) to the expected variance (V_e) under the assumption of linkage equilibrium $D' = 0$ (Brown et al. 1980; Souza et al. 1992; Haubold et al. 1999):

$$V_e = \sum h_j(1 - h_j)$$

where h_j is the genetic diversity per locus. The mean number of mismatches, for a given set of loci, provides a measure of the allelic diversity of the sample. The ratio of the variances provides a measure of linkage disequilibrium, where the denominator scales for the effect of single locus genetic diversities (Souza et al. 1992). V_o/V_e ranges from 1.0 (panmixis) to very large. To assess the probability of rejecting the null hypothesis (panmixis), we obtained a distribution of V_o/V_e by performing 10 000 iterations using a Monte Carlo simulation (LDV program, Souza et al. 1992). If V_o/V_e is significantly greater than 1.0 ($P < 0.01$), we conclude that the population is not panmictic. Based on the total sample of wild isolates, for each sample size we generated a random sample where the expected D' (using an allele frequency of 0.1), and the multilocus expected disequilibrium (exp V_o/V_e) was calculated with the LDD and the LDV programs (Souza et al. 1992, 1994). In simulations with different sample sizes (data not shown), we observed that V_o/V_e is not very sensitive to sample sizes above 30 individuals, but is very sensitive to the number of polymorphic loci. This is not surprising because this measurement of linkage disequilibrium relies on the variance between mismatches among loci. We determined that when fewer than five polymorphic loci exist, this could give a biased signal. In our study the experimental populations show on average three polymorphic loci, while the wild populations average twelve polymorphic loci.

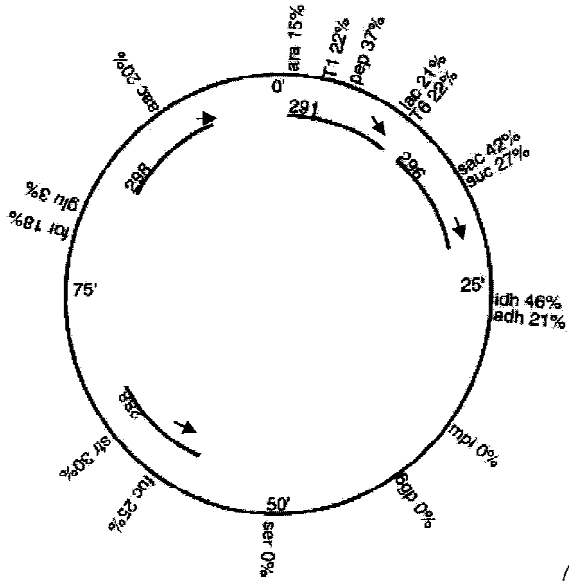

Figure 1. Genetic map of *Escherichia coli* according to Ridley & Laberdan (1996), indicating the location of 17 studied genetic markers (see Methods), including the percentage of recombinants between K12 and B6 parental strains that were detected for each locus. The arrows show the origin and direction of conjugation of the four K12 HFR *E. coli* strains that were used in the experiment.

Results

Genetic analysis of the experimental recombination populations

We studied eighteen polymorphic genetic markers in ten strains from each experimental population in both treatments (recombination and control) after 1000 generations. As expected, none of the 120 strains from the 12 control populations showed changes in their B/6 markers, indicating the experiment was contamination-free. In contrast, 111 of 120 strains from the twelve recombination populations showed substitution of at least one K12 marker for a B/6 marker. We consider this to be clear evidence of successful recombination within the treatment populations ($P < 0.0001$, paired t-test with df = 12 – 1). Based on the *E. coli* genetic map (Ridley & Laberdan 1996), our mapping of genetic substitutions for each recombinant genotype indicated that the recipient chromosome was not homogeneous for recombination events. Rather, we observed a zone between positions 35′ and 51′ that did not feature any recombination events. This coincides with the presence of several genes for termination of replication and ribosomal genes. In contrast, a 'hot spot' for recombination was observed between 4′ and 28′, whereas a 'cold spot' was seen between 75′ and 2′ (Figure 1). These results suggest that differences between K12 and B/6 cause recombinational exchange to be strongly selected against in several regions of the *E. coli* chromosome.

Table 2. Genetic diversity and linkage disequilibrium estimates in *Escherichia coli* isolated from different geographic regions, compared with strains from the worldwide ECOR collection and those from an experimental evolution study involving recombination

Sample	N	$H_e \pm$(SD)	V_o/V_e[1]	Exp[2] V_o/V_e	D'[3] $\pm$(SD)	Exp D'
Mexico	143	0.72 (0.07)	1.54	1.49	0.23 (0.03)	0.39
Australia	40	0.62 (0.06)	2.72	2.11	0.66 (0.03)	0.70
Experimental study (100 gen)	120	0.01 (0.01)	1.50	1.56	1	0.42
Experimental study (1000 gen)	120	0.36 (0.07)	1.52	1.56	0.36 (0.02)	0.42
ECOR	19	0.58 (0.06)	3.40	2.64	0.95 (0.02)	0.85

[1]Significance estimated using a Monte Carlo procedure. All values obtained exceed $P < 0.001$.
[2]Estimated based on a random sample of the same size.
[3]Estimated using allelic frequencies greater than 0.1 (see text).

Genetic diversity increased with generation time in the recombination lines from an average of zero mismatches at generation 0 to an average of 1.96 at generation 1000, and from $H_e = 0$ at generation 0 to $H_e = 0.36$ at generation 1000; in contrast, genetic diversity did not change in the controls (Table 2). Similarly, we observed that linkage disequilibrium (D′) decreased over time in the recombinant populations (Figure 2). By generation 600, recombinant genotypes were numerous enough to be detected in each population, coinciding with a drastic change in linkage disequilibrium. Prior to 600 generations it is likely that rare recombinant alleles (frequency less than 0.1) were present, but were not detected and therefore unaccounted for in the statistics. At 1000 generations diversity was high and the signature of mixis was evident; the initial $D' = 1$ declined to an average of $D' = 0.36$ (Table 2, Figure 2). In contrast, V_o/V_e was extremely large at time 0 because the observed variance is maximal with only 2 initial lineages (K12 and B/6). But it very quickly reaches equilibrium close to panmixia (V_o/V_e [100] = 1.50, V_o/V_e [1000]= 1.52). This apparent equilibrium can be due to a lack of statistical resolution, or to the fact that recombination is repeatedly shuffling the same handful of markers. Even though the observed linkage disequilibrium indicates

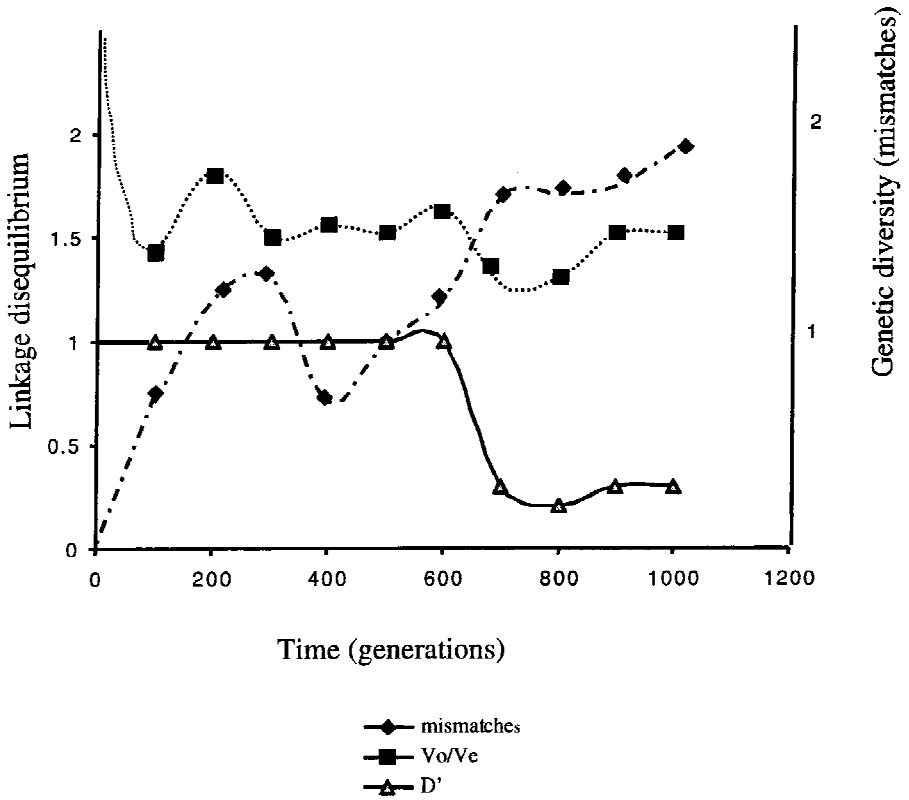

Figure 2. Mean trajectories for linkage disequilibrium D' (triangles), multilocus linkage disequilibrium V_o/V_e (squares), and genetic diversity measured as the number of mismatches (diamonds) analyzed every 100 generations for the 12 *E. coli* populations in the recombination experiment.

Table 3. Genetic diversity and linkage disequilibrium estimates in *Escherichia coli* isolated from different host orders.[1]

Order	N	$H_e \pm$(SD)	V_o/V_e[2]	Exp[3] V_o/V_e	D'[4] $\pm$ (SD)	Exp D'
Carnivora	36	0.69 (0.07)	2.30	2.30	0.76 (0.03)	0.75
Rodentia	52	0.69 (0.07)	2.15	2.16	0.68 (0.03)	0.71
Marsupialia	33	0.67 (0.07)	2.75	2.36	0.78 (0.03)	0.76
Primates[5]	29	0.66 (0.07)	2.86	2.39	0.88 (0.02)	0.76
Humans	17	0.53 (0.07)	3.27	2.71	0.96 (0.01)	0.90
Chiroptera	14	0.67 (0.07)	3.47	2.87	0.93 (0.02)	0.91
Artiodactyla	19	0.65 (0.07)	3.56	2.64	0.95 (0.02)	0.85

[1]Most ungulates (Perisodactyla, Sirenida, Probosidea, Cetacea) as well as Insectivora and Birds were excluded due to small sample sizes.
[2]Significance estimated using a Monte Carlo procedure. All values obtained exceed $P < 0.001$.
[3]Estimated based on a random sample of the same size.
[4]Estimated using allelic frequencies greater than 0.1 (see text).
[5]Primates include 17 strains from humans.

clonality (statistically greater than 0), it is closer to panmixia than the random sample obtained from the wild isolates (Table 2).

Genetic diversity and linkage disequilibrium in wild isolates

Wild *E. coli* isolates were diverse (H_e = 0.58–0.72; Table 2) and, as previously observed (Souza et al. 1999), diversity was structured by geographic location (Table 2). However, genetic diversity was very similar among *E. coli* isolates from different host orders (Table 3).

In the linkage disequilibrium analysis by geographic location, all multilocus linkage disequilibrium estimates V_o/V_e were significantly different from 1.0, indicating departures from a panmictic population (Table 2). However, the Mexican isolates showed far lower linkage disequilibrium (V_o/V_e = 1.54, D' = 0.23) than the human related ECOR collection (V_o/V_e = 3.40, D'= 0.95), while the collection from Australia was intermediate (V_o/V_e = 2.72, D'= 0.66) (Table 2). This may have been due to differences in host organisms across samples, or differences in sample size (Table 3). In general, the ECOR collection shows higher linkage disequilibrium. Over the entire sample

(Table 2), both measures of linkage disequilibrium are relatively low (V_o/V_e = 1.51 and D' = 0.39) suggesting some importance of genetic exchange. In our collection, the expected linkage value was relatively close to that expected (due to the sufficient sample size) in the different regions of the world, while the ECOR collection had a much larger value than expected due to its small sample size (Table 2).

We also analyzed linkage disequilibrium within different groups of hosts (Table 3). We observed that strains from rodents and carnivores had the least disequilibrium (V_o/V_e = 2.15 and 2.30, D' = 0.68 and 0.76, respectively), while strains from artiodactyls and bats had higher values (V_o/V_e = 3.56 and 3.47, D' = 0.95 and 0.93, respectively). The samples for primates and for marsupials were intermediate (V_o/V_e = 2.86 and 2.75, D' = 0.88 and 0.78, respectively). None of the host orders had lower linkage disequilibrium than the geographical Mexican sample (Table 2). We found that the V_o/V_e distributions of the groups with low disequilibrium (Mexico, carnivores and rodents) did not overlap with the distributions of V_o/V_e from the more clonal groups (ECOR, bats and Artiodactyla; data not shown). This may suggest that some groups of *E. coli* are more clonal than others, and that genetic exchange in less clonal groups is not frequent enough to disrupt the linkage disequilibrium.

Effect of recombination on phylogeny

The genetic structure of the natural isolates of *E. coli* showed high genetic diversity, although this diversity was strongly influenced by both geography and host type (Souza et al. 1999). When we included the experimental evolution strains in this parsimony analysis we observed that two recombinant strains were in the same group as K12, while another two recombinant strains were associated with the B/6 group along with the 10 000-generation long-term strains. Interestingly, one recombinant strain was placed in a group independent from the parental lines, demonstrating that recombination can generate divergent genotypes (Figure 3). Furthermore, this analysis shows that recombination can change the genetic relationship of strains and alter their position in the phylogenetic tree.

Discussion

The goal of this study was to investigate the effects of recombination on genetic diversity and population structure, by comparing and contrasting the results from experimental evolution and survey studies. Recombination was observed to have profound effects both in the laboratory and in nature. After 1000 generations of laboratory selection, substantial allozyme variation was observed in populations experiencing recombination while no allozyme variation was observed in control non-recombining populations. The evolution and persistence of recombinant genotypes occurred despite 7000 generations of prior selection to the same environmental conditions. In nature, extensive recombination was evident, although linkage disequilibrium estimates indicated significant geographic and host taxa subdivision.

Genetic diversity and niche adaptation

Recombination allows for both the spread of an adaptive mutation through a population without concomitant loss of diversity at other linked loci, and also the genesis of novel genotypes by mixis. Many studies of the importance of recombination in bacterial diversity have focused solely on the first potential effect (see Milkman 1997), placing much less emphasis on the second. In our laboratory study, we exclusively examined the effect of recombination on the formation of novel genotypes, with no possibility for the spread of an adaptive mutation by recombination among existing genotypes within populations. Despite the inability for recombination to prevent selective sweeps within populations, genetic diversity, as measured by MLEE, increased in all recombination populations (Table 2). Although the average genetic diversity within the recombining laboratory populations (H_e = 0.36) was substantially less than among wild isolates (H_e = 0.732), this was not unexpected, given that there were only 30 rounds of mixis during 1000 generations with relatively little diversity among the donor and recipients. In nature, wild *E. coli* populations have undoubtedly experienced more mixis, between many more genotypes and over millions of generations.

An important component for the maintenance of MLEE diversity within the evolving populations was ecological niche diversification. Niche specialization was observed in the long-term experiment in the absence of recombination (Elena & Lenski 1997), with crossfeeding as the most likely explanation. Rozen & Lenski (2000) observed that coexistence of multiple genotypes can occur due to the evolution of both a glucose specialist genotype that excretes metabol-

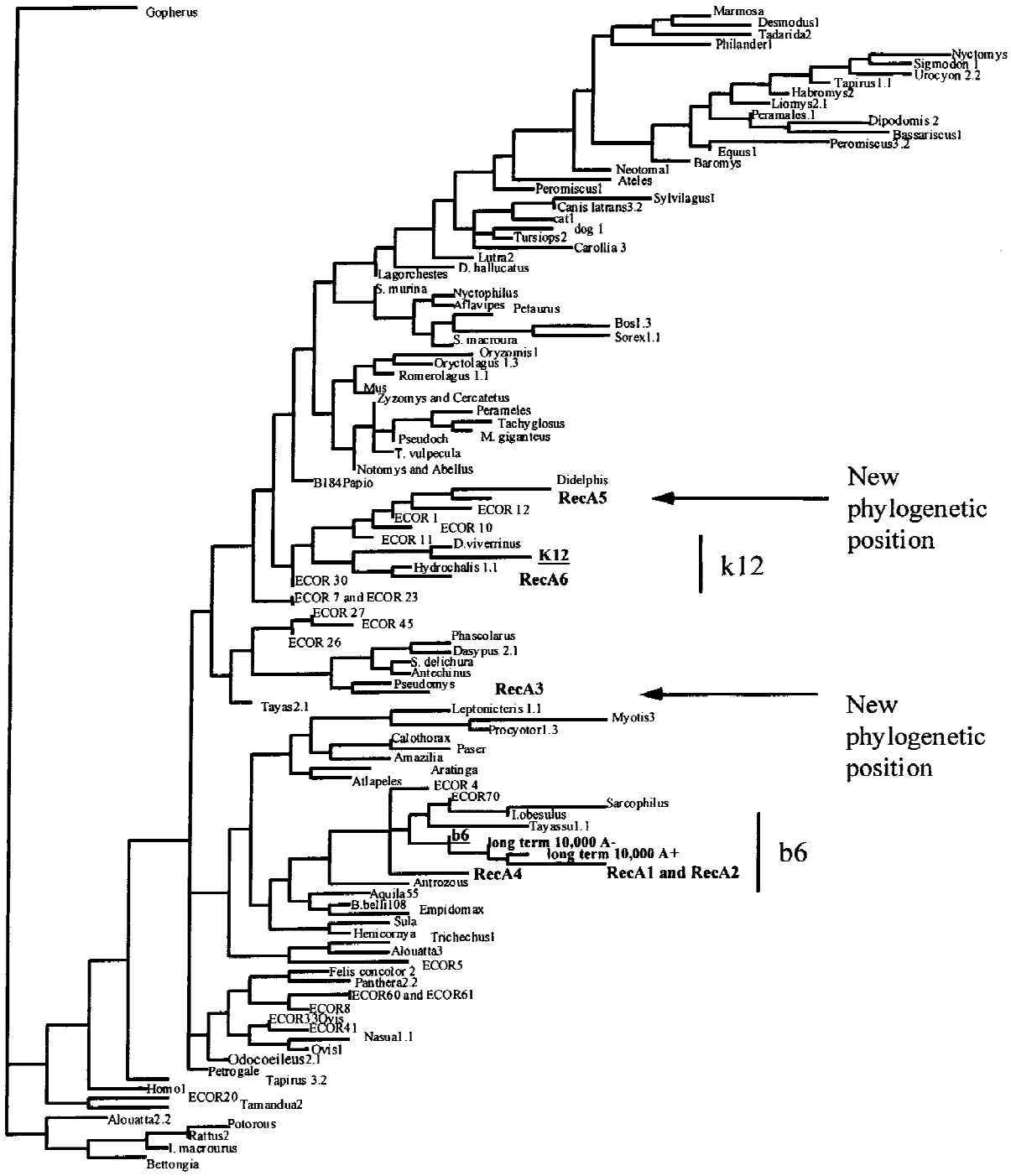

Figure 3. A Maximum Parsimony tree depicting the apparent relationships of the *E. coli* strains used and obtained in the recombination experiment in relation to the wild *E. coli* strains. The strains included in the analysis are the five divergent recombinant genotypes (generated through parental strains K12 and B6), and parental strains, K12, B6, and the two long-term original genotypes (isolated at 10 000 generations of the initial Lenski & Travisano (1994) experiment). To compare these experimental strains with natural isolates, we randomly selected 112 electrotypes from the collected wild isolates. MLEE electrophoresis was performed for 12 loci on both the experimental strains and the wild strains using the same loci as described in Methods. The analysis was done with PAUP*(Macintosh beta test version 4.0b2; Swofford 2001), using a heuristic search. One arbitrary most parsimonious tree (the first one) of several was selected, the length of the branches reflects the estimated number of changes.

ites of glucose metabolism and a generalist genotype that can grow on glucose and the excreted metabolites. However, such ecological specialization was observed to have a large effect in only one of the 12 evolving populations. In contrast, two or more distinct genotypes were commonly observed to co-occur in the recombining populations (Turner 1995; Turner et al. 1996), suggesting that recombination provided the additional genetic variation required for ecological specialization to readily evolve.

Niche specialization is also observed in the wild. In our previous study (Souza et al. 1999), geographic effects, host diet and host taxonomic group accounted for as much as 14% of the diversity. The latter results suggest that the genetic diversity of *E. coli* strains isolated from humans is about two thirds of that found in strains isolated from mammals in general.

Linkage disequilibrium

In the present study, we assessed whether recombination is sufficient to explain the genetic structure that we previously observed in *E. coli* from wild mammals (Souza et al. 1999). To do so, we compared the effects of strictly controlled recombination in experimental populations to recombination that occurs in wild populations. It is commonly accepted that human-associated *E. coli* is considered a good example of a clonal bacterium (see review by Whittam 1996). In the recombination experiment, clonal structure was deliberately broken apart, and as expected linkage disequilibrium diminished over time from D' = 1 at generation 0 to D' =0.36. These values are lower than the expected value based on a random sample of the same size from the wild isolates, where we found that the population is not as clonal as previously observed. Therefore, based on a sample of the same size of alleles of wild isolates, we observe that experimental evolution with a high efficiency of recombination attains lower linkage disequilibrium in 1000 generations than does nature in several millions of generations.

The ability for recombination to affect inferred relatedness was evident after the combined analysis of the experimental and natural isolates. Even after 10 000 generations of selection, no MLEE diversification was evident among the control long-term lines and they grouped together on the phylogenetic tree, while the recombination lines were distributed throughout the tree after only 30 mixis events. Although this suggests that caution is required in the interpretation of inferred phylogenetic relationships, the wide distribution of the recombination lines also demonstrates the power of recombination to rapidly generate genetic diversity.

Our survey results suggest that recombination and niche specialization have also been important in the diversification of *E. coli* in the wild. Even though the multilocus analyses (V_o/V_e) showed significant linkage disequilibrium at all levels, the linkage disequilibrium values are lower than expected, especially in Mexico. For instance, the distribution of V_o/V_e from Mexico does not overlap with the distribution of the ECOR collection, with Mexico closer to panmixia and the ECOR group near the maximum possible linkage disequilibrium value. We suggest that the wide array of hosts sampled in Mexico provide a more complete representation of the different niches of *E. coli*, because each group of hosts may involve a different part of the total genetic pool of *E. coli.* In contrast to the sample from Mexico, the ECOR collection is mostly from humans and relatively few captive mammals. We suggest that the genetic structure observed previously in *E. coli* could be the result of strong selective pressure in the human gut, where only certain clones are selected from the wide genetic pool of *E. coli.*

Following the idea that each mammalian group is sampling a different section of the gene pool of *E. coli*, we analyzed each host order independently. We found that bats and artiodactyls feature highly clonal populations of *E. coli*, that marsupials and primates feature intermediate values, and that carnivores and rodents represent more recombinant populations. In general, our results agree with the ecotypic structure proposed by Maynard Smith (1991). If there is no recombination as a source of new variation, newly established clones that are locally adapted may not easily colonize other environments, because they lack the needed adaptations to initially compete well (Haubold & Rainey 1996). Our interpretation is that strains from some host groups are more clonal than others. This could result from differences in gut biochemistry and physiology of the host, as well as from the feeding habits and ecology of these animals (Prosser 1973; Souza et al. 1999).

Limitations of the current work and future directions

The most severe limitation of the analysis presented in this paper is that the two studies are comparable at only the most general level: the effects of recombination on evolution in *E. coli*. No effort was

made to control for the recombination mechanism, strength of selection, or frequency of recombination (or other factors) across the studies, and such factors can drastically alter patterns of recombination. Nevertheless, we believe that the approach we have taken is both innovative and necessary. While experimental evolution studies are exceedingly powerful, they can only demonstrate *potential* outcomes, and may not reflect the processes occurring in nature. Comparative studies can delineate patterns in nature, but are often less effective in discerning underlying processes. We argue that there are substantial benefits for tightly connecting experimental evolution studies in the laboratory with comparative field studies. The value of such combined studies is that patterns in nature can be associated with specific ecological and evolutionary processes. To date, few have undertaken such a combined approach.

Acknowledgements

We wish to thank the numerous colleagues who assisted us in obtaining samples: D. Gordon collected all the Australian samples. B. Hernández, J. Goluvov, M. Mandrujano, R. Medellin, S. Aguilar, O. Gaona, A. Mendoza, G. Pérez, L. Medrano, A. Navarro, J. Castillo, A. Zavala, B. Morales, R. Frías, A. Velázques, H. Arita, C. Alvarez, J. Charles, H. Charles, A. Miranda, D. Valenzuela, L. Sandner, R. Cerritos J. Ortega and several others collected the Mexican samples. We also thank Antonio Cruz, Laura Espinoza and Aldo Valera for valuable technical assistance. We are especially grateful to Richard Lenski who not only made this work possible by teaching us experimental evolution but also by showing us how to be critical with our work. This work was supported by DGAPA IN 218698 and CONACyT 27557 to VS, and CONACyT 27983-N to LEE.

References

Brown AHD, Feldman MW & Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics, 96: 523–536.

Caugant DA, Levin BR & Selander RK (1981) Genetic diversity and temporal variation in the *E.coli* population of a human host. Genetics 98: 467–490.

Elena S & Lenski RE (1997) Long-term experimental evolution in *Escherichia coli*. VII. mechanisms maintaining genetic variability within populations. Evolution 51: 1058–1067.

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

Haubold B & Rainey PB (1996) Genetic and ecotypic structure of a fluorescent *Pseudomonas* population. Mol. Ecol. 5: 747–761.

Hedrick PW & Thomson G (1985) A two-locus neutrality test: applications to human, *E. coli* and lodgepole pine. Genetics 112: 135–156

Herbert PDN & Beaton MJ (1993) Methodologies for Allozyme Analysis Using Cellulose Acetate Electrophoresis. Helena Laboratories, USA.

Koshi LB, Morton RA & Golding GB (2001) Codon bias and base composition are poor indicators of horizontally transfered genes. Mol. Biol. Evol. 18: 404–412.

Lenski RE & Travisano M (1994). Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc. Natl. Acad. Sci. USA 91: 6808–6814.

Lenski RE, Rose MR, Simpson SC & Tadler SC (1991). Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138: 1315–1341.

Maynard Smith J (1991) The population genetics of bacteria. Proc. R. Soc. London 245: 37–41.

Maynard Smith J & Smith NH (1998) Detecting recombination from gene trees. Mol. Biol. Evol. 15: 590–599.

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

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

Milkman R & McKane-Bridges M (1990) Molecular Evolution of the *Escherichia coli* chromosome. III. Clonal Frames. Genetics 126: 505–517.

Milkman R & Stolzfus A (1988) Molecular evolution of *Escherichia coli* chromosome. II. Clonal segments. Genetics 120: 359–366.

Ochman H & Selander RK (1984) Evidence for clonal population structure in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 81: 198–201.

Ochman H & Wilson AC (1987) Evolutionary History of Enteric Bacteria. In: Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M & Umbarger MHE (Eds) *Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology (pp 1625–1648). American Society of Microbiology, Washington, DC.

O'Rourke M & Stevens E (1993) Genetic structure of *Neisseria gonorrhoeae* populations: a non clonal pathogen. J. Gen. Microbiol. 139: 2603–2611.

Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE & Blot M (1999). Genomic evolution during a 10,000-generation experiment with bacteria. Proc. Natl. Acad. Sci. USA 96: 3807–3812

Perna NT, Plunkett III G, Burland V, Mau B, Glasner JD, Rose JD, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Posfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dinalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch R & Blattner FR (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157: H7. Nature 409: 529–533.

Prosser CL (1973) Comparative Animal Physiology. 3rd Edn. Saunders, Philadelphia, PA.

Pupo GM & Richardson BJ (1995) Biochemical genetics of a natural population of *Escherichia coli*: seasonal changes in alleles and haplotypes. Microbiology 141: 1037–1044.

Ridley M & B Labedan (1996) *Escherichia coli* gene products: physiological functions and common ancestries. In: Neidhardt

FC, Curtis R, Ingraham J, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Magasanik, B, Riley M, Schaechter M & Umbarger, MHE (Eds) *Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology, 2nd Edn (pp 2108–2140). American Society of Microbiology, Washington, DC.

Rozen DE & Lenski RE (2000). Long-term experimental evolution in *Escherichia coli.* VIII. Dynamics of a balanced polymorphism. Am. Nat. 155: 24–35.

Selander RK, Caugant DA Ochman H, Muser JM & Whittam TS (1986) Methods of multilocus electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. 51: 873–884.

Selander RK, Caugant DA & Whittam TS (1987) Genetic structure and variation in natural populations of *Escherichia coli*. In: Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M & Umbarger MHE (Eds) *Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology (pp 1625–1648). American Society of Microbiology, Washington, DC.

Souza V, Nguyen TT, Hudson RR, Piñero D & Lenski RE (1992) Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for localized sex? Proc. Natl. Acad. Sci. USA 89: 8389–8393.

Souza V, Eguiarte L, Avila G & Gallardo C (1994) Genetic structure of *Rhizobium etli* biovar *phaseoli* associated with wild and cultivated been plants (*Phaseolus vulgaris* and *Phaseolus coccineus*) in Morelos, Mexico. Appl. Environ. Microbiol. 60, 1260–1268.

Souza V, Turner PE, & Lenski RE (1997). Long-term experimental evolution in *Escherichia coli*. V. effects of recombination with immigrant genotypes on the rate of bacterial evolution. J. Evol. Biol. 10: 743–769.

Souza V, Rocha M, Valera A & Eguiarte LE (1999) Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. Appl. Environ. Microbiol. 65: 3373–3385.

Swofford D (2001) Version 4.0 beta version, Laboratory of Molecular Systematics. Smithsonian Institution, Beta Documentation http://www.lms.si.edu/PAUP.

Travisano M & Rainey P (2000) studies of adaptative radiation using model microbial systems. Am. Nat. 156: 35–44.

Turner PE (1995) Associations between bacteria and conjugative plasmids: model systems for testing evolutionary theory. Dissertation. Michigan State University, East Lansing, MI, USA.

Turner PE, Souza V & Lenski RE (1996) Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. Ecology 77: 2119–2129.

Whittam TS (1989). Clonal dynamics of *Escherichia coli* in its natural habitat. Antonie van Leeuwenhoek. 55: 23–32.

Whittam TS (1996) Genetic variation and evolutionary processes in natural population of *Escherichia coli*. In: Neidhardt FC, Curtis R, Ingraham J, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Magasanik, B, Riley M, Schaechter M & Umbarger, MHE (Eds) *Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology, 2nd edn. (pp 2708–2720). American Society of Microbiology, Washington, DC.

Whittam TS & Ake SE (1993) Genetic Polymorphisms and recombination in natural populations of *Escherichia coli.* In: Takahata N & Clark AG (Eds) Mechanisms of Molecular Evolution (pp 223–245). Japan Scientific Society Press, Sinauer Associates, Inc., Tokyo.

Whittam TS, Wolfe ML, Wachsmuth IK, Ørskov F, Ørskov I & Wilson RA (1993) Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect. Immun. 61: 1619–1629.