JOURNAL OF BACTERIOLOGY, Oct. 1995, p. 5784–5789 Vol. 177, No. 20
0021-9193/95/$04.00+0
Copyright © 1995, American Society for Microbiology

Heterogeneity of Genome Sizes among Natural Isolates of *Escherichia coli*

ULFAR BERGTHORSSON* AND HOWARD OCHMAN

Department of Biology, University of Rochester, Rochester, New York 14627

Received 24 April 1995/Accepted 10 August 1995

Comparisons of the genetic maps of *Escherichia coli* K-12 and *Salmonella typhimurium* LT2 suggest that the size and organization of bacterial chromosomes are highly conserved. Employing pulsed-field gel electrophoresis, we have estimated the extent of variation in genome size among 14 natural isolates of *E. coli*. The *Bln*I and *Not*I restriction fragment patterns were highly variable among isolates, and genome sizes ranged from 4,660 to 5,300 kb, which is several hundred kilobases larger than the variation detected between enteric species. Genome size differences increase with the evolutionary genetic distance between lineages of *E. coli*, and there are differences in genome size among the major subgroups of *E. coli*. In general, the genomes of natural isolates are larger than those of laboratory strains, largely because of the fact that laboratory strains were derived from the subgroup of *E. coli* with the smallest genomes.

The most comprehensive information concerning the evolution of bacterial chromosomes has been based on alignments of the genetic maps of *Escherichia coli* K-12 and *Salmonella typhimurium* LT2 (5, 24, 42, 44). Although these enteric species diverged an estimated 120 to 160 million years ago (36), the overall similarity in their linkage maps and the relative paucity of repetitive DNA in bacterial genomes have led to the view that the size and organization of bacterial chromosomes are highly conserved (41–43). Recent comparisons of the physical maps of *E. coli* K-12, *Shigella flexneri* 2a, and representatives of several serovars of *Salmonella enterica* also support the notion that the sizes of bacterial chromosomes are evolutionarily stable (26, 27, 37). Laboratory strains of *E. coli* and *Shigella flexneri* are estimated to have chromosomes ranging from 4.5 to 4.7 Mb in length, while those of *Salmonella* spp. range from 4.6 to 4.9 Mb.

The conservation in chromosome size and organization among enteric bacteria is surprising in light of both the estimated divergence times among these species and the high frequency of large-scale rearrangements observed in laboratory populations (2, 3, 20, 49). Microorganisms readily acquire exogenous DNA, and changes in chromosome organization can occur through deletions, duplications, inversions, and translocations. However, little is known about the contribution of these processes to chromosomal evolution and genetic variation within natural populations of *E. coli* (18).

In contrast to the apparent similarity among the chromosomes of *E. coli* K-12, *Shigella flexneri*, and serovars of *Salmonella enterica*, Brenner et al. (8) detected substantial variations in genome size among clinical isolates of *E. coli*. Employing DNA renaturation procedures, this study revealed genome size differences of more than 1 Mb between certain isolates. This variation could not be wholly attributed to extrachromosomal sequences since many strains had genomes smaller than that of an *E. coli* K-12 strain known to lack plasmids.

Pulsed-field gel electrophoresis (PFGE) currently offers the most expedient means to both analyze genome sizes and construct low-resolution physical maps of bacterial chromosomes. To date, studies of chromosomal variation within *E. coli* by PFGE have included laboratory and clinical isolates but have not examined the variation within the species at large (4, 7, 10, 15, 19, 20, 38–40, 46, 50). Harsono et al. (17) reported wide variations in genome sizes (ranging from 3.5 to 5.5 Mb) for *E. coli* O157:H7 isolates; however, these estimates are questionable since the sizes of individual strains often varied with the restriction enzyme.

To examine the diversity in genome size and the rate of chromosomal evolution in natural populations of *E. coli*, we employed PFGE in an analysis of natural isolates of *E. coli* of known genetic and genealogical relationships. For each strain, genome sizes, as estimated from digestions with different restriction enzymes, were very similar, and the total size variation within the species was estimated to be less than 1 Mb. Furthermore, the genomes of natural isolates were larger than those of laboratory strains.

MATERIALS AND METHODS

Bacterial strains. Fourteen strains of *E. coli* were selected from the ECOR reference collection, which includes isolates from a wide variety of hosts and geographic regions (35). The phylogenetic relationships of these strains have been established by multilocus enzyme electrophoresis (MLEE) of 38 polymorphic loci (21), and the nucleotide sequences of several genes in many of these strains have been determined (12, 14, 16, 33). The phylogenetic relationships of the 14 strains used in this study are shown in Fig. 1. Derivatives of *E. coli* K-12, EMG2, MG1655, and W3110, which had previously been characterized by PFGE (40) were used as controls.

Preparation of DNA. Cells were grown overnight in 1.0 ml of Luria-Bertani (LB) broth and harvested by centrifugation. The cellular pellet was washed twice in 5 ml of TEN (10 mM Tris [pH 7.5], 100 mM EDTA [pH 8], 250 mM NaCl) and resuspended in 0.5 ml of TEN. Then cells were mixed with 0.75 ml of 1.5% InCert agarose (FMC, Rockland, Maine) in TEN and dispensed into 100-μl plug molds (Pharmacia, Piscataway, N.J.). Agarose plugs were incubated for at least 8 h in lysis solution (0.1% lysozyme, 0.002% RNase, 0.5% Sarkosyl, 10 mM Tris [pH 7.5], 100 mM EDTA [pH 8], 250 mM NaCl), with subsequent overnight incubation at 45°C in 0.1% proteinase K–1% Sarkosyl–250 mM EDTA. To inactivate excess proteases, agarose plugs were incubated in 0.01 mM phenylmethylsulfonyl fluoride for 1 h, washed, and stored in 10 mM Tris–100 mM EDTA at 4°C.

Restriction endonuclease digestion. Agarose plugs were washed five times, each for 30 min., in 50 volumes of distilled H_2O, and equilibrated in the appropriate restriction enzyme buffer. Fifteen units of enzyme (*Bln*I or *Not*I) was added to initiate digestion, and after incubation overnight at 37°C, EDTA was added to each sample to a final concentration of 0.1 M. Samples were stored at 4°C until subjected to PFGE.

PFGE. Electrophoresis was performed with a CHEF-DR II apparatus (Bio-Rad Laboratories, Richmond, Calif.). Approximately 10 μl of an agarose plug was inserted into a 0.9% agarose gel, and electrophoresis was conducted in 0.5×

* Corresponding author. Phone: (716) 275-8389. Fax: (716) 275-2070. Electronic mail address: ulfar@ho-lab.biology.rochester.edu.

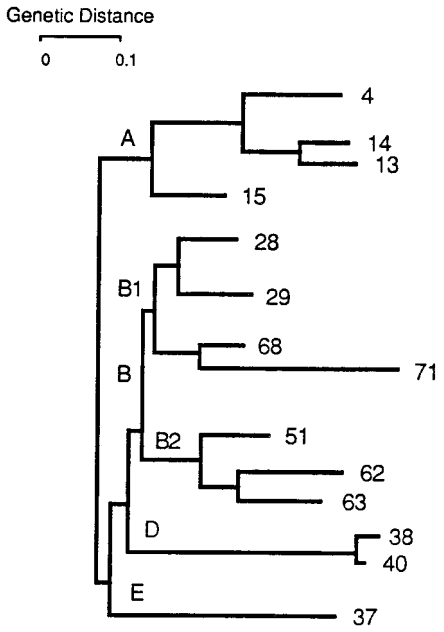

FIG. 1. Phylogenetic relationships of the ECOR strains analyzed (based on MLEE [21]). Major phylogenetic subgroups within *E. coli* are labeled with the letters A to E, and numbers refer to strain designations.

Tris-borate-EDTA at 14°C. For most gels, electrophoresis proceeded for 40 h at 180 V, and pulse times varied according to the intended range of resolution. To resolve fragments in the range from 30 to 600 kb, pulse times were ramped from 10 to 40 s over the 40-h period; for fragments ranging from 50 to 900 kb, pulse times were ramped from 30 to 70 s. For larger fragments, pulse times of 2 min were applied for 24 h with subsequent 4-min pulse times for 30 h at 150 V, resolving fragments in the range of 0.7 to 2 Mb. Gels were stained in 0.01% ethidium bromide and photographed under UV light. Lambda ladder, Low-Range PFG marker (New England Biolabs, Beverly, Mass.), and yeast chromosomes were used for molecular size markers.

Genome size measurements. Genome sizes were estimated by adding up the sizes of the restriction fragments produced in these digests. Since ethidium bromide binds stoichiometrically to DNA, comigrating fragments were resolved by the intensity of staining. Densitometric tracings of photographs of each gel were used to determine the number of comigrating fragments contained in bands of unusual intensity.

Growth rate estimates. Estimates of growth rates in LB broth were made by sampling cultures every 15 min and recording optical densities at 550 nm on a Spectronic 21 spectrophotometer (34). Regression lines of increases in density during log phase were used to calculate doubling times. The growth rates of these strains in M9 minimal medium have already been estimated (31).

Plasmid analysis. Plasmid DNAs were isolated from each strain by the method of Anderson and McKay (1). Agarose plugs containing undigested genomic DNA were also subjected to PFGE in an attempt to detect nonlinearized plasmid molecules.

Statistical analysis. An analysis of variance of the differences in genome size among isolates of the four major subgroups of *E. coli* that are represented by more than one strain in our sample (subgroups A, B1, B2, and D) was performed by applying a general formula that takes unequal sample sizes into account (47).

To test for an association between differences in genome size and evolutionary genetic distance between strains, we calculated regression and correlation coefficients of genome size differences to genetic distances estimated from the tree of Herzer et al. (21). Regression coefficients were also calculated after log transformations of the data. (The rationale behind log transformations of genome size differences is that the variance in genome size is expected to increase over time after divergence from a common ancestor; when the variance is correlated with the mean, log transformation stabilizes the variance [47].) These strains are not evolutionarily independent and therefore cannot be treated as such in correlations between genome size difference and other measures, such as genetic distance or growth rates (13). To avoid the problems associated with the nonindependence of data caused by a shared phylogenetic history, we based this analysis on $(n - 1)$ independent contrasts, as advised by Felsenstein (13). Each point in Fig. 2 is based on the size differences and genetic distances calculated for 13 phylogenetically independent pairs of strains. According to the relationships shown in Fig. 1, one point is based on comparisons between ECOR 13 and 14, a second point is based on comparisons between ECOR 4 and the inferred ancestor of ECOR 13 and 14, a third is based on comparisons between ECOR 15 and the last common ancestor of ECOR 4, 13, and 14, and so on.

RESULTS

The genome sizes of these 14 strains of *E. coli* ranged from 4.66 to 5.30 Mb, as estimated by PFGE. Strains were originally selected to represent the five major subgroups in the ECOR collection (21, 35, 45), as defined by MLEE, and the estimates of genome sizes were calculated from digests with two rare-cutting restriction enzymes, *Bln*I and *Not*I (Table 1). Although other enzymes have been employed to construct low-resolution restriction maps of laboratory strains of *E. coli* and *Salmonella* spp., we have found that *Sfi*I (39, 46) and *Xba*I (40) often yield partial digests for natural isolates of *E. coli*, probably because of the overlap of their recognition sequences with methylation sites in the *E. coli* genome (40). Furthermore, the homing endonuclease I-*Ceu*I, which cleaves in the *rrn* genes, normally produces a chromosomal fragment of over 2 Mb (25–28), and it is difficult to accurately determine the lengths of fragments of this size.

We were able to resolve restriction fragments that ranged from 5 to 1,800 kb. The fragment sizes listed in Table 1 are average values based on 4 to 14 pulsed-field gels, and the coefficient of variation for size estimates of fragments greater than 100 kb was normally below 5%, with a mean of 2.2%. The number of restriction sites per genome for *Not*I in these natural isolates ranges from 11 to 27, with a mean of 24, and the number ranges from 11 to 21 for *Bln*I, with a mean of 14. Laboratory strains derived from *E. coli* K-12 normally produce 22 fragments upon digestion with *Not*I and 17 fragments upon digestion with *Bln*I (40). Like *E. coli* K-12 derivatives, natural isolates also contain more *Not*I sites than *Bln*I sites, except for ECOR 62 and ECOR 37, which contains only 11 *Not*I sites.

For most strains there is good correspondence between the genome sizes estimated from the two restriction enzymes, and the size estimates for 8 of these 14 strains differ by less than 100 kb for the two enzymes. ECOR 29 has the largest disparity between the estimates based on *Bln*I and *Not*I (360 kb). In total, the correlation coefficient between estimates based on *Not*I and *Bln*I is 0.8 ($P < 0.001$; $n = 14$). Since variations in genome size estimates between enzymes could result either from plasmids cleaved by only one of the enzymes or from unresolved comigrating fragments, we applied two procedures to resolve these confounding variables. Partial digests and comigrating fragments were readily detected by examining staining intensity by densitometry, and in each case, genome sizes were adjusted accordingly. About half of the strains in our sample harbor plasmids that are larger than 100 kb. To test the effects of nonlinearized large plasmids, undigested genomic DNA from each strain was subjected to PFGE. In these experiments, we observed no bands that might correspond to large plasmids; therefore, the differences in genome size estimates from using different restriction enzymes probably result from plasmids that have not entered the gel and contain restriction sites for only one of the enzymes.

In rare cases, fragments of less than 20 kb may not have been detected in our gels; however, these would not considerably alter genome size estimates. For *E. coli* K-12 derivatives, Perkins et al. (40) estimated that there were one to four *Not*I or *Bln*I fragments of less than 20 kb in each strain and that the sum of these fragments did not exceed 30 kb for a given strain.

The restriction fragment patterns of the natural isolates that we tested are highly variable, even between closely related strains. For example, ECOR 38 and ECOR 40 differ at only 1 of 38 polymorphic enzyme loci, but about half of the restriction

TABLE 1. Sizes of *Bln*I and *Not*I restriction fragments in natural isolates of *E. coli*

Strain	Fragment size(s) (kb)		
	*Bln*I	*Not*I	Mean
ECOR 4	26, 60, 71, 77, 95, 110, 158, 236, 291, 511, 517, 580, 634, 1,310	28, 32, 40, 83, 100, 132, 163, 165, 212, 245, 273, 280, 290, 298, 309, 344, 394, 1,365	4,714
Total	4,676	4,752	
ECOR 13	32, 50, 46, 89, 89, 115, 116, 150, 159, 403, 450, 502, 628, 710, 1,053	19, 32, 83, 92, 92, 92, 104, 104, 134, 172, 189, 210, 219, 227, 240, 242, 253, 331, 640, 1,250	4,658
Total	4,592	4,725	
ECOR 14	13, 22, 32, 35, 41, 49, 83, 83, 134, 214, 219, 468, 474, 708, 1,135, 1,400	13, 26, 33, 35, 72, 80, 84, 90, 90, 100, 133, 138, 155, 162, 162, 190, 210, 220, 240, 255, 283, 300, 350, 400, 1,070	5,000
Total	5,110	4,891	
ECOR 15	29, 29, 59, 80, 129, 446, 556, 634, 690, 1,030, 1,230	14, 15, 23, 30, 33, 39, 63, 67, 71, 106, 113, 133, 137, 163, 179, 183, 192, 205, 248, 253, 267, 285, 292, 308, 350, 1,200	4,940
Total	4,912	4,969	
ECOR 28	31, 34, 49, 49, 102, 103, 217, 283, 356, 424, 514, 696, 725, 1,400	31, 31, 31, 55, 75, 90, 99, 102, 105, 105, 120, 140, 150, 180, 200, 216, 235, 250, 258, 280, 290, 305, 354, 590, 690	4,982
Total	4,983	4,981	
ECOR 29	5, 18, 34, 44, 96, 155, 155, 223, 270, 306, 346, 1,020, 1,090, 1,360	26, 32, 71, 73, 81, 85, 102, 133, 134, 158, 194, 198, 254, 257, 260, 266, 275, 315, 334, 405, 1,110	4,942
Total	5,121	4,763	
ECOR 37	13, 29, 79, 97, 97, 131, 131, 161, 237, 303, 330, 339, 425, 449, 476, 485, 500, 595	30, 49, 69, 124, 196, 287, 428, 453, 472, 940, 1,760	4,842
Total	4,877	4,808	
ECOR 38	23, 29, 50, 59, 61, 87, 97, 120, 146, 165, 200, 216, 216, 282, 300, 354, 378, 388, 388, 412, 540, 769	17, 23, 32, 44, 63, 68, 86, 117, 134, 175, 179, 196, 204, 207, 230, 290, 298, 318, 404, 467, 496, 1,200	5,264
Total	5,280	5,248	
ECOR 40	18, 32, 59, 87, 110, 120, 144, 171, 205, 225, 260, 296, 318, 354, 378, 413, 437, 548, 582, 615	10, 19, 21, 30, 48, 68, 77, 96, 149, 180, 183, 202, 206, 229, 247, 300, 328, 334, 395, 411, 525, 1,175	5,302
Total	5,372	5,233	
ECOR 51	10, 29, 29, 68, 108, 108, 134, 149, 201, 376, 412, 454, 493, 510, 589, 1,670	17, 35, 35, 56, 56, 56, 77, 90, 90, 98, 109, 110, 147, 147, 152, 161, 171, 173, 206, 217, 227, 244, 244, 244, 275, 381, 648, 700	5,253
Total	5,340	5,166	
ECOR 62	29, 32, 37, 48, 68, 71, 77, 84, 92, 94, 94, 112, 192, 212, 226, 252, 257, 370, 428, 431, 495, 610, 750	24, 31, 38, 40, 59, 79, 93, 136, 150, 175, 180, 194, 194, 228, 255, 268, 276, 310, 332, 401, 763, 910	5,098
Total	5,061	5,136	
ECOR 63	32, 32, 38, 110, 112, 130, 140, 245, 280, 310, 480, 555, 669, 1,800	19, 25, 36, 36, 61, 96, 99, 110, 112, 137, 160, 165, 169, 192, 218, 232, 247, 269, 288, 340, 346, 680, 940	4,952
Total	4,933	4,977	
ECOR 68	22, 30, 48, 70, 90, 101, 135, 190, 300, 303, 330, 360, 365, 395, 510, 785, 1,100	33, 33, 38, 69, 76, 80, 87, 104, 125, 135, 158, 208, 251, 273, 276, 291, 325, 340, 400, 417, 595, 800	5,122
Total	5,131	5,114	
ECOR 71	31, 31, 55, 61, 86, 125, 136, 207, 309, 357, 493, 583, 900, 1,590	36, 70, 76, 81, 90, 93, 102, 122, 136, 153, 175, 186, 186, 245, 253, 260, 266, 274, 281, 310, 388, 1,075	4,916
Total	4,964	4,858	

fragments generated by *Not*I and *Bln*I differ in size. The average genome sizes for subgroups A, B1, B2, and D are 4,800, 5,000, 5,100, and 5,300 kb, respectively. In a one-way analysis of variance, the differences in genome size among these major subgroups of the ECOR collection are significant ($F = 5.95$; $df = 12$; $P = 0.016$), whereas the nonparametric Kruskal and Wallis test on these data resulted in only borderline significance ($H = 7.18$; $P = 0.066$).

The chromosomal sizes of laboratory isolates of *E. coli*, *Shigella flexneri*, and serovars of *Salmonella enterica* are fairly similar, suggesting strong selective constraints on genome size over long evolutionary periods. Since natural strains display an even wider range of variation, we examined the rates of change in genome size by plotting the differences in size among pairs of strains against their evolutionary genetic distances, as determined by MLEE (Fig. 2). The linear regression of the logarithm of genome size differences against the logarithm of evolutionary genetic distances is significant ($R^2 = 0.40$; $P = 0.021$); without the log transformations, however, the regression is only marginally significant ($R^2 = 0.27$; $P = 0.064$). Together with the differences among subgroups of *E. coli*, these results suggest a phylogenetic component in genome size variation, i.e., closely related strains tend to have genomes that are more similar in size and genome sizes change gradually on an evolutionary timescale. If there are constraints on genome size, we expect the variation in genome size to plateau and not to go beyond a certain value as the genetic distance increases; unfortunately, the data are too limited to test this prediction. Since it is generally believed that bacterial genomes are streamlined to promote replication, we compared genome

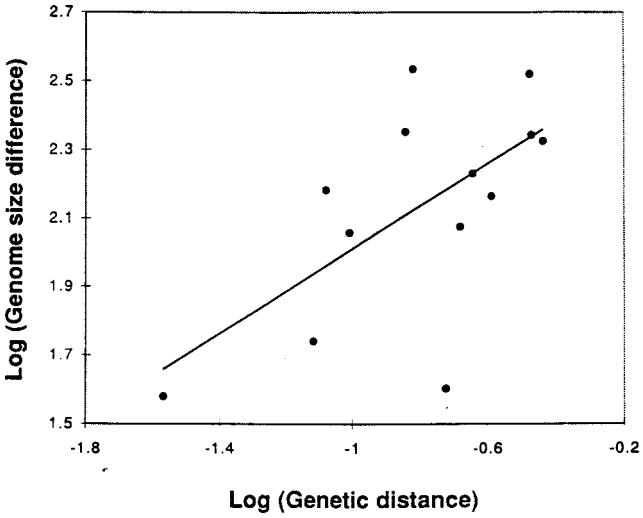

FIG. 2. Relationships between genome size differences among ECOR lineages and their evolutionary genetic distances. Each point is based on the size differences and genetic distances calculated for 13 phylogenetically independent pairs of strains. The regression line is significant ($R^2 = 0.40$; $P = 0.021$); without the log transformations, it is only marginally significant ($R^2 = 0.27$; $P = 0.064$).

sizes with growth rates in LB and minimal media; however, there was no association between growth rates and genome sizes in these media (data not shown).

DISCUSSION

Previous studies examining the extent of variations in genome size within and among enteric species have yielded contradictory information about the rates of chromosome evolution in bacteria. Comparisons of the physical and genetic maps of *E. coli* K-12 and *Salmonella typhimurium* LT2 suggest that strong selection maintains the size and organization of bacterial genomes (27, 42), whereas studies examining closely related strains of *E. coli* have detected large variations in chromosome size (8, 17). Despite uncertainties associated with some of these previous estimates, we have discovered that the genome sizes of natural isolates of *E. coli* may differ by as much as 650 kb, more than three times the difference observed between *E. coli* K-12 and *Salmonella typhimurium* LT2 (27, 46).

Studies employing PFGE to construct physical maps of *E. coli* K-12 derivatives have reported chromosome lengths ranging from approximately 4.5 to 4.7 Mb, often accompanied by inversions, duplications, and deletions (10, 19, 40, 46). Because of the rate at which spontaneous chromosome rearrangements are known to occur (2, 3, 20, 49), it is perhaps not surprising that certain laboratory isolates vary with respect to their chromosome sizes and organizations. The differences in laboratory strains are not representative of genetic variation in the species at large, and many of the observed changes in chromosome size and organization have probably resulted from mutagenic treatment in the laboratory environment (40).

Two previous analyses of clinical isolates of *E. coli* revealed a greater degree of genome size variation than that detected among the natural strains that we selected to encompass the range of genetic diversity in the species as a whole. On the basis of DNA renaturation kinetics, Brenner et al. (8) noted large differences in the genome sizes of two serotypes of *E. coli*, with estimates ranging from 2.29×10^9 to 2.97×10^9 Da (3.8 to 4.8 Mb). Since the rates of DNA reassociation depend on the GC content of the genome as well as the amount of unique sequences, it is somewhat difficult to establish the degree to which these estimates reflect actual variations in genome size.

Examining *Xba*I and *Sfi*I digests of *E. coli* by PFGE, Harsono et al. (17) detected considerable size variations among strains serotyped to O157:H7. However, these estimates varied widely with the enzyme used to fractionate the genome. In *Xba*I digests, genome sizes ranged from 3.6 to 5.6 Mb; with *Sfi*I, the same strains had genomes ranging from 3.3 to 4.0 Mb. Furthermore, certain pairs of strains that were indistinguishable for their *Sfi*I restriction fragments differed by as much as 1.2 Mb when they were analyzed with *Xba*I.

The set of strains analyzed in this study contains members of each major subgroup of *E. coli*, as defined by MLEE (21, 45), and the diversity observed in this sample is likely to represent the entire range of genome size variation in the species at large. To examine the rate of divergence in genome size, we plotted the size differences among strains against their evolutionary genetic distances (Fig. 2). This bivariate plot considers samples at each node in the phylogenetic tree, which successively eliminates the correlation among phylogenetically related lineages. Comparisons of genome size differences and the evolutionary genetic distances among strains show that closely related lineages tend to have genomes of similar sizes; therefore, overall genome size is a slowly evolving trait. This reinforces the view that most events leading to the acquisition and deletion of sizable portions of the genome, which are common in laboratory populations, are not stable over the evolutionary history of *E. coli*. In one case, however, two closely related lineages, ECOR 13 and 14, differed by more than 300 kb, providing evidence that some changes in genome size occur over a relatively short timescale.

In the event that the association between these variables plateaus beyond a certain genetic distance, in that more distantly related strains do not have more disparate genome sizes, the effect is apt to be the consequence of stabilizing selection. Although the size of our sample is too small to detect the effects of stabilizing selection on genome size in *E. coli*, the sizes of four serovars of *Salmonella enterica* are well within the range detected in natural populations of *E. coli* (despite the tendency for genome sizes to diverge over time), suggesting that such constraints exist (27). On the other hand, it can also be argued that selective constraints on genome size are probably not as strong as previously assumed since natural strains of *E. coli* can have genomes as large as 5.3 Mb.

The estimated sizes of genomes of natural isolates of *E. coli* were generally larger than those of laboratory strains. It is likely that two factors, selection for replication rates and the phylogenetic ancestry of strains, contribute to the relatively small chromosome of *E. coli* K-12. According to the results presented here, the ancestry of the *E. coli* K-12 strain largely accounts for this difference in genome size between laboratory strains and natural isolates. *E. coli* K-12 is most closely related to strains in subgroup A; in our sample, both ECOR 4 and 13 from subgroup A have a genome size of about 4.7 Mb, which is similar to the length calculated for the *E. coli* K-12 chromosome (40, 46). The group A ECOR strains have an average genome size of 4.8 Mb, which is smaller than the average sizes of group B (5.05 Mb) and group D (5.3 Mb) strains. We have not fully resolved the source of these size variations, but the preliminary physical maps of these strains indicate that the increases in genome size for groups B and D are not due to a single event that occurred in the lineage ancestral to these strains.

The following two broad classes of events may be responsible for the observed diversity in genome size within *E. coli*: duplications or deletions of the existing chromosomal DNA

and the acquisition of exogenous sequences. While the frequency of duplications in enteric bacteria ranges from 10^{-2} to 10^{-5} per cell (2, 3, 20), which is considerably higher than the spontaneous mutation rate of 10^{-9} to 10^{-10} mutations per bp per generation (11), most of these events are unstable and probably do not contribute to the differentiation among closely related strains or species of enteric bacteria (22, 48). In contrast, it is thought that deletions of chromosomal regions occur much less frequently than do duplications; however, the elimination of large chromosomal regions from pathogenic strains of *E. coli* occurs at high rates and provides a means to modulate virulence (23, 38).

A disproportionately large amount of the chromosome variation among laboratory strains has been mapped to the region surrounding the replication terminus (40). For the most part, this variation is caused by small deletions and insertions and probably results from high levels of recombination in this region (29). In comparisons of the *E. coli* K-12 and *Salmonella typhimurium* LT2 genetic maps, however, regions unique to one of these species seem to be distributed at random around the chromosome. Although variations among natural isolates of *E. coli* are probably not confined to a single variable region, such a situation is thought to exist in *Bacillus cereus*. Its chromosome sizes vary from 2.4 to 5.3 Mb, and a highly variable portion is missing in strains with smaller chromosomes (9).

What are the consequences of these differences in genome size among strains of *E. coli*? Since bacterial chromosomes are composed principally of coding regions, strains with smaller genomes could lack scores of genes; therefore, one might hypothesize that strains with smaller genomes are adapted for rapid growth under complex nutrient-rich conditions, whereas strains with larger genomes are at a selective advantage under simple or nutrient-poor conditions. Among the strains examined in this study, there were no significant associations between growth rate and genome size in either minimal or LB medium. Mikkola and Kurland (31, 32) demonstrated that factors such as the translational efficiency of ribosomes have a profound effect on the growth rates of natural isolates of *E. coli*; this would probably obscure any variations in growth rates due to differences in genome size.

The origin(s) and nature(s) of the sequences contributing to the genome size variations in natural isolates of *E. coli* are still unclear. Some of the variation is probably due to mobile genetic elements; however, transposons, phages, and plasmids rarely total several hundred kilobases. Pathogenic *E. coli* strains are known to harbor large arrays of the chromosomal genes required for virulence, termed pathogenicity islands, and natural populations are polymorphic with respect to the presence of these genes. For example, uropathogenic *E. coli* 536 contains two such islands, which are 70 and 190 kb in length (6). Three strains in our sample, ECOR 40, 62, and 71, were originally isolated from urinary tract infections (35); however, their genomes are not significantly larger than those of closely related nonpathogenic strains.

Horizontal transfer has certainly contributed to the evolution of bacterial genomes. While the chromosomes of *E. coli* K-12 and *Salmonella typhimurium* LT2 are approximately the same size, alignments of their genetic maps have revealed several regions (15% of their chromosomes) that are unique to one species (42). Moreover, examinations of GC contents and codon usage of sequenced genes from *E. coli* suggest that at least 6% (51) and as much as 16% (30) of the *E. coli* genome has been acquired through horizontal transfer. On the basis of these estimates, the contribution of horizontal transfer to the divergence in genome size of distantly related ECOR strains is likely to be 50 to 100 kb.

ACKNOWLEDGMENTS

We thank the Weinstock laboratory at the University of Texas Medical School for providing strains, Allen Orr for discussions, and Pilar Francino for reviewing the manuscript.

This work was supported by grants from the NIH and NSF.

REFERENCES

1. **Anderson, D. G., and L. L. McKay.** 1983. Simple and rapid method for isolating large plasmid DNA from lactic streptococci. Appl. Environ. Microbiol. **46:**549–552.
2. **Anderson, P., and J. Roth.** 1981. Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. Proc. Natl. Acad. Sci. USA **78:**3113–3117.
3. **Anderson, R. P., and J. R. Roth.** 1977. Tandem genetic duplications in phage and bacteria. Annu. Rev. Microbiol. **31:**473–505.
4. **Arbeit, R. D., M. Arthur, R. Dunn, C. Kim, R. K. Selander, and R. Goldstein.** 1990. Resolution of recent evolutionary divergence among *Escherichia coli* from related lineages: the application of pulsed field electrophoresis to molecular epidemiology. J. Infect. Dis. **161:**230–235.
5. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
6. **Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschäpe, and J. Hacker.** 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606–614.
7. **Böhm, H., and H. Karch.** 1992. DNA fingerprinting of *Escherichia coli* O157:H7. J. Clin. Microbiol. **30:**2169–2172.
8. **Brenner, D. J., G. R. Fanning, F. J. Skerman, and S. Falkow.** 1972. Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms. J. Bacteriol. **109:**953–965.
9. **Carlson, C. R., and A.-B. Kolstø.** 1994. A small (2.4 Mb) *Bacillus cereus* chromosome corresponds to a conserved region of a larger (5.3 Mb) *Bacillus cereus* chromosome. Mol. Microbiol. **13:**161–169.
10. **Daniels, D. L.** 1990. The complete *Avr*II restriction map of the *Escherichia coli* genome and comparisons of several laboratory strains. Nucleic Acids Res. **18:**2649–2651.
11. **Drake, J. W.** 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:**7160–7164.
12. **DuBose, R. F., D. E. Dykhuizen, and D. L. Hartl.** 1988. Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85:**7036–7040.
13. **Felsenstein, J.** 1985. Phylogenies and the comparative method. Am. Nat. **125:**1–15.
14. **Guttman, D. S., and D. E. Dykhuizen.** 1994. Detecting selective sweeps in naturally occurring *Escherichia coli*. Genetics **138:**993–1003.
15. **Hacker, J., L. Bender, M. Ott, J. Wingender, B. Lund, R. Marre, and W. Goebel.** 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. Microb. Pathog. **8:**213–225.
16. **Hall, B. G., and P. M. Sharp.** 1992. Molecular population genetics of *Escherichia coli*: DNA sequence diversity at the *celC*, *crr*, and *gutB* loci of natural isolates. Mol. Biol. Evol. **9:**654–665.
17. **Harsono, K. D., C. W. Kaspar, and J. B. Luchansky.** 1993. Comparison and genomic sizing of *Escherichia coli* O157:H7 isolates by pulsed-field gel electrophoresis. Appl. Environ. Microbiol. **59:**3141–3144.
18. **Hartl, D. L., M. Medhora, L. Green, and D. E. Dykhuizen.** 1986. The evolution of DNA sequences in *Escherichia coli*. Philos. Trans. R. Soc. Lond. B **312:**191–204.
19. **Heath, J. D., J. D. Perkins, B. Sharma, and G. M. Weinstock.** 1992. *Not*I genomic cleavage map of *Escherichia coli* K-12 strain MG1655. J. Bacteriol. **174:**558–567.
20. **Heath, J. D., and G. M. Weinstock.** 1991. Tandem duplications of the *lac* region of the *Escherichia coli* chromosome. Biochimie (Paris) **73:**343–352.
21. **Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam.** 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:**6175–6181.
22. **Horiuchi, T., S. Horiuchi, and A. Novick.** 1963. The genetic basis of hypersynthesis of β-galactosidase. Genetics **48:**157–169.
23. **Karch, H., T. Meyer, H. Rüssmann, and J. Heesemann.** 1992. Frequent loss of Shiga-like toxin genes in clinical isolates of *Escherichia coli* upon subcultivation. Infect. Immun. **60:**3464–3467.
24. **Krawiec, S., and M. Riley.** 1990. Organization of the bacterial chromosome. Microbiol. Rev. **54:**502–539.
25. **Liu, S.-L., A. Hessel, H.-Y. M. Cheng, and K. E. Sanderson.** 1994. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella paratyphi* B. J. Bacteriol. **176:**1014–1024.
26. **Liu, S.-L., A. Hessel, and K. E. Sanderson.** 1993. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella enteritidis* shows an inversion relative to *Salmonella typhimurium* LT2. Mol. Microbiol. **10:**655–664.
27. **Liu, S.-L., A. Hessel, and K. E. Sanderson.** 1993. Genomic mapping with I-*Ceu*I, an intron-encoded endonuclease specific for genes for ribosomal

RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. Proc. Natl. Acad. Sci. USA **90:**6874–6878.
28. **Liu, S.-L., and K. E. Sanderson.** 1995. Rearrangements in the genome of the bacterium *Salmonella typhi*. Proc. Natl. Acad. Sci. USA **92:**1018–1022.
29. **Louarn, J., F. Cornet, V. François, J. Patte, and J.-M. Louarn.** 1994. Hyperrecombination in the terminus region of the *Escherichia coli* chromosome: possible relation to nucleoid organization. J. Bacteriol. **176:**7524–7531.
30. **Médigue, C., T. Rouxel, P. Vigier, A. Hénaut, and A. Danchin.** 1991. Evidence for horizontal gene transfer in *Escherichia coli* speciation. J. Mol. Biol. **222:**851–856.
31. **Mikkola, R., and C. G. Kurland.** 1991. Is there a unique ribosome phenotype for naturally occurring *Escherichia coli*? Biochimie (Paris) **73:**1061–1066.
32. **Mikkola, R., and C. G. Kurland.** 1992. Selection of laboratory wild-type phenotype from natural isolates of *Escherichia coli* in chemostats. Mol. Biol. Evol. **9:**394–402.
33. **Milkman, R., and M. M. Bridges.** 1990. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics **126:**505–517.
34. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 31–32. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.
36. **Ochman, H., and A. Wilson.** 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26:**74–86.
37. **Okada, N., C. Sasakawa, T. Tobe, K. A. Talukder, K. Komatsu, and M. Yoshikawa.** 1991. Construction of a physical map of the chromosome of *Shigella flexneri* 2a and the direct assignment of nine virulence-associated loci identified by Tn*5* insertions. Mol. Microbiol. **5:**2171–2180.
38. **Ott, M.** 1993. Dynamics of the bacterial genome: deletions and integrations as mechanisms of bacterial virulence modulation. Zentralbl. Bakteriol. **278:** 457–468.
39. **Perkins, J. D., J. D. Heath, B. R. Sharma, and G. M. Weinstock.** 1992. *Sfi*I genomic cleavage map of *Escherichia coli* K-12 strain MG1655. Nucleic Acids Res. **20:**1129–1137.
40. **Perkins, J. D., J. D. Heath, B. R. Sharma, and G. M. Weinstock.** 1993. *Xba*I and *Bln*I genomic cleavage maps of *Escherichia coli* K-12 strain MG1655 and comparative analysis of other strains. J. Mol. Biol. **232:**419–445.
41. **Riley, M., and A. Anilionis.** 1978. Evolution of the bacterial genome. Annu. Rev. Microbiol. **32:**519–560.
42. **Riley, M., and S. Krawiec.** 1987. Genome organization, p. 967–981. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
43. **Sanderson, K. E.** 1971. Genetic homology in the *Enterobacteriaceae*. Adv. Genet. **16:**35–51.
44. **Sanderson, K. E., and J. R. Roth.** 1988. Linkage map of *Salmonella typhimurium*, edition VII. Microbiol. Rev. **52:**485–532.
45. **Selander, R. K., D. A. Caugant, and T. S. Whittam.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
46. **Smith, C. L., J. G. Econome, A. Schutt, S. Klco, and C. R. Cantor.** 1987. A physical map of the *Escherichia coli* K12 genome. Science **236:**1448–1453.
47. **Sokal, R. R., and F. J. Rohlf.** 1981. Biometry. Freeman, San Francisco.
48. **Sonti, R. V., and J. R. Roth.** 1989. Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. Genetics **123:**19–28.
49. **Starlinger, P.** 1977. DNA rearrangements in procaryotes. Annu. Rev. Genet. **11:**103–126.
50. **Tschäpe, H., R. Prager, L. Bender, M. Ott, G. Blum, and J. Hacker.** 1993. Dissection of pathogenetic determinants and their genomic positions for the evaluation of epidemic strains and infection routes. Zentralbl. Bakteriol. **278:**425–435.
51. **Whittam, T. S., and S. E. Ake.** 1993. Genetic polymorphisms and recombination in natural populations of *Escherichia coli*, p. 223–245. *In* N. Takahata and A. G. Clark (ed.), Mechanisms of molecular evolution. Japan Scientific Societies Press, Tokyo.