Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7036–7040, September 1988
Population Biology

Genetic exchange among natural isolates of bacteria: Recombination within the *phoA* gene of *Escherichia coli*

(evolution/DNA sequences/clonality)

ROBERT F. DUBOSE*, DANIEL E. DYKHUIZEN†, AND DANIEL L. HARTL*‡

*Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110; and †Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY 11794

Communicated by Robert K. Selander, June 6, 1988 (received for review February 22, 1988)

ABSTRACT **An 1871-nucleotide region including the *phoA* gene (the structural gene encoding alkaline phosphatase, EC 3.1.3.1) was cloned and sequenced from eight naturally occurring strains of *Escherichia coli*. Alignment with the sequence from *E. coli* K-12 made apparent that there were 87 polymorphic nucleotide sites, of which 42 were informative for phylogenetic analysis. Maximum parsimony analysis revealed six equally parsimonious trees with a consistency index of 0.80. Of the 42 informative sites, 22 were inconsistent with each of the maximum parsimony trees. The spatial distribution of the inconsistent sites was highly nonrandom in a manner implying that intragenic recombination has played a major role in determining the evolutionary history of the nine alleles. The implication is that different segments of the *phoA* gene have different phylogenetic histories.**

The nucleotide sequence data reported in this paper resolve an apparent paradox relating to the effects of recombination in natural populations of bacteria. Although recombination is a fundamental process in most organisms, concepts of recombination as applied to natural populations of bacteria are rather vague. The concept of recombination that is prevalent in evolutionary genetics is based on meiosis in eukaryotic organisms, in which entire genomes come into contact and large segments of genetic information are reciprocally exchanged. However, prokaryotes provide several alternative pathways of recombination—conjugation, transformation, transduction, sexduction, etc.—that differ not only in mechanism but also in their implications at the population level (1). Linkage disequilibrium occurring between widely separated genes supports the view that lineages in *Escherichia coli* are largely clonal. Nevertheless, the data reported here demonstrate that recombination has played a major role in shaping the nucleotide sequences of alleles of *phoA*, the structural gene encoding alkaline phosphatase (EC 3.1.3.1), in this organism. In *E. coli*, the presence of largely clonal lineages at the chromosomal level is consistent with the occurrence of recombination, because the mechanism of recombination appears to involve primarily short stretches of nucleotide sequence.

Early evidence of largely clonal patterns of genetic variation in *E. coli* came from the finding of a highly nonrandom distribution of serotypic antigen markers among strains (2). Subsequently, Ochman and Selander (3) demonstrated evidence of clonality among 142 pathogenic isolates of *E. coli* by using protein electrophoresis. However, an objection can be raised that clonality might be expected in pathogenic isolates, because pathogenicity involves many coadapted genes, and selection might act to retain linkage disequilibrium between the coadapted gene complexes (4). More recently, Whittam *et al.* (5) and Selander *et al.* (6) have studied the electrophoretic mobilities of polymorphic enzymes among natural isolates of *E. coli*, the majority of them nonpathogenic. In addition to finding large amounts of linkage disequilibrium among enzyme-coding genes, they also found that the amount of linkage disequilibrium was independent of the map distance between the genes. This suggests that recombination of large pieces of the bacterial chromosome, as occurs in conjugation, is rare in natural populations of *E. coli* (7, 8).

However, linkage disequilibrium between genetic markers does not necessarily exclude the occurrence of recombination involving short stretches of DNA, as might be expected if the mode of transfer of DNA between strains were through bacteriophage-mediated transduction. To be specific, consider two polymorphic genetic markers separated by 500 kilobase pairs of DNA (about 10 min in the chromosome of *E. coli* K-12). If the rate of reciprocal recombination resulting in the exchange of flanking markers occurs at the rate r per million nucleotide pairs per generation, then linkage disequilibrium between the markers will be dissipated at the rate of $r/2$ per generation. However, if r means the rate of recombination exchanging stretches of, for example, 500 nucleotide pairs, then linkage disequilibrium will be affected by an exchange event only when one of the genetic markers is actually included in the stretch of exchanged material. If the average length of the exchanged segments is such that it cannot include both markers, this principle has the unexpected implication that the rate of decay of linkage disequilibrium depends only on the average length of the exchanged segments and not on the genetic distance between the markers.

To study population processes at this level, comparison of DNA sequences is necessary. With this in mind, we have determined the nucleotide sequences of the alleles of *phoA* that occur in eight natural isolates of *E. coli*.§ When the sequences are compared with the sequence from the standard laboratory strain *E. coli* K-12 for shared nucleotides at sites that are polymorphic, it is apparent that the shared sites are clustered in such a way as to strongly suggest that they originated from intragenic recombination. The implication is that different segments of the *phoA* gene have different phylogenetic histories. This inference is consistent with genetic exchange processes that have been experimentally identified in *E. coli* (9, 10).

MATERIALS AND METHODS

The bacterial strains used are a subset of the ECOR reference collection of strains studied by Milkman and Crawford (11). Our allele designations are 1–8, corresponding to RM39A, RM191F, RM201C, RM217T, RM45E, RM70B, RM224H, and RM202I, respectively. This subset was chosen to be

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

§The sequences reported in this paper are being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04079).

representative of the genetic variation that occurs at the *trp* locus (11). In addition, we have found substantial variation in the nucleotide sequences of the *gnd* genes in these same strains (12), and extensive protein variation detected by electrophoresis also occurs (6). The *phoA* sequence in *E. coli* K-12 has been published previously; we use the same numbering scheme as refs. 13 and 14. The *E. coli* host used in preparation of M13mp18 phage containing the *phoA* genes was strain BD56 [Δ(*phoA–proC*), *phoR*, *tsx*::Tn*5*, F$^+$::Tn*3*], constructed by conjugative transfer of the F factor (F$^+$::Tn*3*) from WB351 (15) into Mph44 (16). Recombinant phage were identified by their ability to form blue plaques on B plates containing 5-bromo-4-chloro-3-indolyl phosphate (XP, Sigma) Wild-type *phoA* DNA used as probe in Southern blots to determine suitably sized fragments was isolated from plasmid pJC2431 (17), which contains a 2.7-kilobase (kb) *Hin*dIII–*Xho* I DNA fragment bearing *phoA* from *E. coli* K-12. All *phoA* genes were cloned by using a common *Hin*dIII site 282 nucleotides upstream from the initiation codon of *phoA*. In strains 1, 2, 3, 5, and 7 the cloned *phoA*-containing fragment was 2.7-kb *Hin*dIII–*Xho* I restriction fragment, and in strains 4, 6, and 8 it was a 4.1-kb *Hin*dIII–*Pst* I fragment. Cloning was carried out with restriction fragments isolated from agarose gels and ligated, directly after purification, into double-stranded M13 vectors digested with either *Hin*dIII–*Sal* I or *Hin*dIII–*Pst* I.

Single-stranded DNA was prepared essentially as in ref. 18, and dideoxy sequencing was performed with the Sequenase kit (United States Biochemical, Cleveland). The universal M13 primer was used initially, and subsequent reactions utilized a series of 20-base-pair synthetic oligonucleotides. Five sequencing reactions were sufficient to cover the region of interest. Sequences were read with a sonic digitizer (Science Accessories) and entered into an IBM PC/XT by using Microgenie DNA software (Beckman).

RESULTS

Sequence Polymorphisms. We determined the sequence of 1871 nucleotides extending from the *Hin*dIII site 282 nucleotides upstream from the *phoA* initiation codon to 173 nucleotides downstream from the *phoA* ochre termination codon. The upstream region includes an additional open reading frame, and the downstream region includes two (13, 14). The sequence data are summarized in Fig. 1. Among 87 polymorphic nucleotide sites, 42 were informative for phylogenetic analysis (that is, two different nucleotides were present in at least two strains each). Ten amino acid polymorphisms were observed within the *phoA* alleles: one within the signal sequence, the remainder within the mature peptide. The average divergence between nucleotide sequences among pairs of strains for the entire region was 1.7%.

Phylogenetic Analysis. Analysis of the sequences with the phylogenetic analysis program PAUP (version 2.4.1, David Swofford, Illinois Natural History Survey, Champaign, IL) revealed six equally parsimonious unrooted trees, each with a length of 109 mutational steps and a consistency index of 0.80. The consistency index is defined as 1 − (parallel steps/total steps). This index may be artificially high, since each of the changes unique to one allele is consistent with any tree that is created. If only the informative sites are considered, the consistency index drops to 0.66.

Three of the six trees are variations on the branching order of alleles K-12, 2, and 3, which have no informative sites to separate them. Each of these three has two alternative configurations, which differ in placing allele 7 with either the 4, 5 cluster or the K-12–2–3, 8 cluster. Tree P in Fig. 2 shows the consensus. For all trees, 22 of the informative sites show at least one reverse or parallel substitution. That so many of the sites showed apparent parallelisms or reversals was surprising in light of the low overall divergence among the *phoA* genes. For example, the most divergent alleles, 6 and 8, differ by only 2.5%. If *E. coli* lineages are exclusively clonal, then these 22 sites must represent actual instances of parallel mutation. However, if some sort of recombination is occasionally exchanging nucleotide sequences among *phoA* genes, then each segment of the gene might have a different phylogenetic history, and a large number of nucleotides that are inconsistent with the consensus phylogeny might be

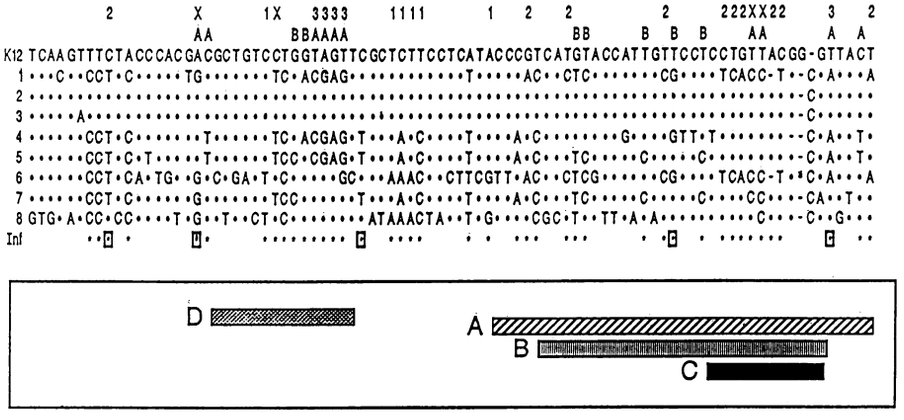

FIG. 1. Distribution of polymorphisms within the sequenced region of *phoA*. Note that only the 87 polymorphic nucleotide sites (see below) are given; the invariant sites have been deleted for a compact presentation. Phylogenetically informative sites are each designated with an asterisk on the "Inf" line. The symbols A, B, 1, 2, and 3 above the nucleotide sites represent the partitions shown in Fig. 3 with which each site is consistent. Informative sites which could not be unambiguously assigned to one of the five partitions are blank; sites not consistent with *any* of the partitions 1–3 are denoted with an X above the sequence. The *P* values associated with each partition (see ref. 19 and Fig. 3) are highly significant. Note especially the clustering of sites consistent with partition 2 at the 3′ end of the region. The shaded bars labeled A–D in the box beneath the sequences indicate the regions of the gene proposed to have undergone the recombinational events described in the text. Informative sites congruent with neither the recombination "base" phylogeny shown in Fig. 2 nor the four recombination events described in the text are indicated with an open box below the sequence, on the Inf line. The actual positions of the polymorphic nucleotides within the sequence are provided below. Numbering starts with 1 at the *Hin*dIII site, as in ref. 19. Amino acid replacement sites are underlined (sites 1360 and 1361 are adjacent within the same codon and represent only one amino acid replacement). 23/53/166/291/388/414/420/438/513/534/573/579/627/651/660/681/763/831/840/861/864/894/897/942/951/966/999/1059/1061/1068/1074/1077/1081/1098/1110/1113/1114/1116/1119/1154/1176/1179/1185/1188/1200/1212/1233/1239/1257/1260/1296/1320/1341/1360/1361/1386/1425/1428/1432/1435/1464/1474/1476/1479/1485/1488/1497/1509/1524/1551/1560/1590/1597/1620/1635/1683/1712/1727/1731/1755/1759/1769/1782/1817/1826/1850/1853.

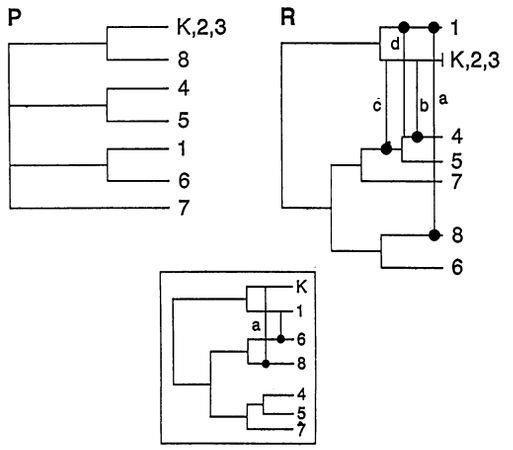

FIG. 2. Consensus and recombinational phylogeny. Tree P shows the consensus phylogeny. Tree R denotes the phylogeny used as the starting point for inferring recombinational events. Hypothesized recombinational events are depicted as lines connecting branches of phylogeny R, which are labeled a–d, corresponding to A–D in Fig. 1. The direction of the exchange is indicated by the dot on the recipient branch. Although one of the events is depicted as bidirectional, it cannot be resolved from two separate events. This is demonstrated in the boxed phylogeny at the bottom, which resolves the reciprocal exchange between strains 1 and 8 (a) into two separate nonreciprocal events involving strains K-12 and 8 and 1 and 6. For each tree, the unresolvable K-12–2–3 cluster is indicated by the stem leading to K.

expected. In such a case, the inconsistent nucleotides would be expected to be clustered in predictable ways (19).

Evidence for Intragenic Recombination. To be conservative, alleles 2 and 3 were not included in the statistical analysis of recombination because the polymorphisms that distinguish the 2, 3 cluster from the K-12 allele, and allele 2 from allele 3, are phylogenetically uninformative. Including such alleles artifically inflates the scores for Sawyer's test (described below).

As one test of intragenic recombination, we used a nonparametric method of S. Sawyer (personal communication), which requires no phylogenetic inference. In Sawyer's method, test statistics calculated from pairs of actual sequences are compared with values obtained from random permutations of the same sequences to determine an empirical significance level (*P* value). Any pair of homologous nucleotide sequences includes sites that differ between the sequences, which are interspersed with stretches of nucleotides that are identical in the two sequences. Each stretch of contiguous nucleotides that are identical in a pair of alleles defines an "inner gap." Within each inner gap are nucleotide sites that, while identical in a particular pair of alleles, are polymorphic in the entire sample. The number of such sites defines the "inner run" for the particular inner gap. Intragenic recombination will have the effect of making both inner gaps and inner runs larger than expected. The test statistics are the sums of squares of inner gaps and inner runs in all pairwise comparisons of the nucleotide sequences, compared with the sums of squares obtained by random permutations of the polymorphic sites among the sequences (keeping their positions fixed). Note that while the inner-gap sum of squares scores will be affected by mutational hot spots or cold spots across the sequence, the inner-run sum of squares is not (Sawyer, personal communication). The *P* value is the proportion of inner-gap or inner-run scores in the simulation that exceed the observed values among 10,000 random permutations. Applied to the *phoA* data, Sawyer's test for intragenic recombination yielded a *P* value of 0.002 for the inner-gap sum of squares and $P < 0.0001$ for the inner-run sum of squares. The length of the maximum run (41 nucleotides) was also highly significant ($P = 0.002$).

Intragenic recombination in *phoA* is also supported by the method of Stephens (19), which associates with each polymorphic site a particular "partition" (grouping of strains) that the site supports. In this method, once each site has been assigned to a particular partition, those partitions that are not mutually exclusive are grouped together and the distribution of sites supporting each group is examined. With no recombination, the distribution of sites supporting a given partition or group of partitions is expected to be random along the sequences. Intragenic recombination results in clustering of sites that support particular partitions. Stephens' test statistics for each possible partition are the number of nucleotides between the most widely spaced sites that support the partition (d_o) and the maximum number of nucleotides between consecutive sites that support the partition (g_o). Recombination results in d_o values that are too small and g_o values that are too large.

Applied to the *phoA* data, the significance level P_d of d_o (calculated as in ref. 19) over all polymorphisms equals 0.42, and that of g_o (P_g) equals 0.17. Thus, Stephens' test indicates no overall clustering of polymorphic sites, which justifies using the length of the entire sequence when determining *P* values in the subsequent analysis. For the groups of partitions that correspond to A, B, 1, 2, and 3 in Fig. 3, the P_d values are 0.02, 0.002, 0.0004, 0.10, and 0.07, respectively, and the P_g values are 0.02, 0.02, 0.15, 0.001, and 0.00001, respectively. That is, for each group, the d_o value is signif-

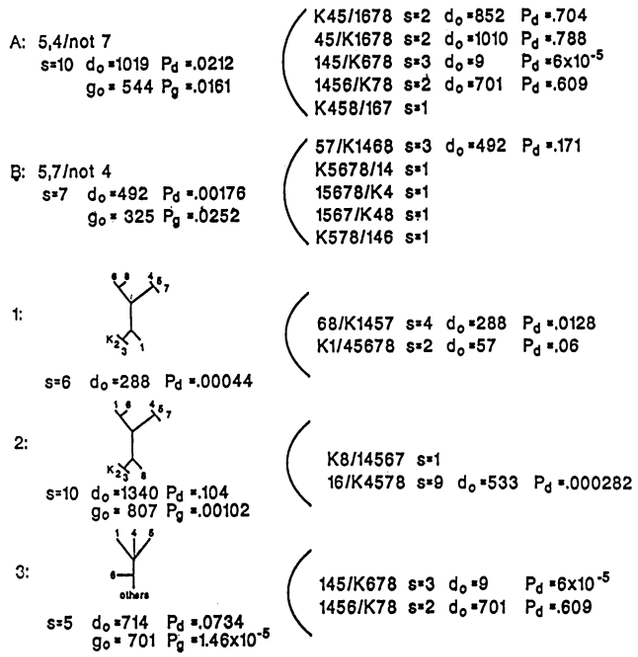

FIG. 3. The composite partitions A, B, and 1, 2, 3, along with supportive partitions observed in the data. For each composite partition, the results of the Stephens test relative to that partition are shown on the left. Composite partitions A and B distinguish between two alternative placements for allele 5 (grouped either with allele 4 but not 7 or with allele 7 but not 4). In composite partitions 1 and 2, alleles 4, 5, and 7 are always grouped, and thus are represented by a trifurcation. Composite partition 3 separates the alleles 1, 4, 5, and 6 from the rest, designated by "others." Each composite partition is resolved at the right into individual two-way partitions (partitions in which some group of the strains shares one nucleotide, and the remainder share a different nucleotide) observed in the data, along with the test statistics for each. *s* is the number of polymorphisms associated with each partition; other statistics are defined in the text. Alleles 2 and 3 are not included in the two-way partitions, since no informative sites exist that distinguish them from allele K-12.

icantly small, the g_o value is significantly large, or both. The runs of sites that support these partition-groups are denoted in Fig. 1 by letters or numbers above the sequences. The clustering of polymorphic sites supporting each is evident, the most striking example being the contrast of 1 versus 2.

DISCUSSION

Two independent statistical tests indicate the occurrence of clustered polymorphism within the *phoA* gene among natural *E. coli* isolates, which we have interpreted as being due to intragenic recombination. We can extend the analysis by postulating a small number of genetic exchange events that together account for the distribution of all but five of the phylogenetically informative nucleotide sites. By "genetic exchange" we mean the *result* of apparent replacement of one DNA sequence by another, without specifying the *mechanism* of its occurrence. The model is summarized in the boxes in Fig. 1. Each proposed recombination terminates at the first major run of informative polymorphic sites that is inconsistent with it. The phylogeny depicted in tree R in Fig. 2 is the inferred ancestral phylogeny that accounts for the observed data, assuming a minimum number of recombinational events. The consensus tree obtained by maximum parsimony analysis is represented in tree P of the same figure. The multiple branching at the base of the tree cannot be resolved in greater detail because of the occurrence of sites that support distinct and mutually exclusive bifurcating trees. The phylogeny R chosen as the basis for inferring recombination is also a representation of one of the partition groups observed in the sequences (i.e., the one labeled 1 in Fig. 1). Thus, phylogeny R represents a phylogeny for the strains that is supported by a number of nucleotides within the sequenced region that are significantly clustered according to Stephens' test. By superimposing the intragenic recombinational events described below onto tree R, it is possible to generate the observed distributions of polymorphic sites.

In the model in Fig. 1, two replacements (box A) between strains K-12, 8, 1, and 6 in the downstream half of *phoA* switch places in the phylogeny, making sequence 8 appear like K-12 and 1 like 6. Whether this event, which includes a minimum of 597 nucleotides, represents one reciprocal exchange between 1 and 8 or two superimposed exchanges (involving, respectively, K-12 and 8 and 1 and 6) cannot be determined, but we will treat it as one event. The next largest event (box B) spans sites 1360 to 1769, and it explains 13 of the 14 informative sites in the region as arising from a K-12-like sequence replacing a strain 4-like sequence. Box C indicates a third postulated event of minimally 210 nucleotides (including seven informative sites), which are explained by replacement of a sequence ancestral to that in strains 4 and 5 by a K-12-like sequence. The last major run of informative sites occurs from nucleotides 861 to 1098 (box D), which includes nine such sites. These sites can be explained by a strain 4-like sequence replacing a strain 1-like sequence. With these four recombinational events (illustrated in Fig. 2 as vertical lines connecting branches of tree R) the distribution of 37 of the 42 informative sites in Fig. 1 can be explained.

It would be surprising if all of the postulated recombinational events occurred between members of this small set of nine strains. However, two of the alleles (6 and 4) each possess runs of unique polymorphic sites that are significantly clustered according to Stephens' test (for allele 4, $s = 5$, $d_o = 257$, and $P = 0.001$; for allele 6, $s = 13$, $d_o = 853$, $P < 0.001$). While our focus has been on exchange events between sequences represented in this set of strains, the runs of unique polymorphic sites could be explained by recombination with sequences not represented in the sample.

The evidence for intragenic recombination in *phoA* is corroborated by the DNA sequences of other genes in the same strains. Among the *gnd* (6-phosphogluconate dehydrogenase) genes, the nucleotide sequences are very highly polymorphic (12, 20). However, Sawyer's test for recombination gives highly significant inner-gap and inner-run sums of squares, both indicative of intragenic recombination (Sawyer, personal communication). A region of the *trp* (tryptophan) operon has also been sequenced in these strains (11). Although much less polymorphic than either *phoA* or *gnd*, the *trp* region has variation that is also consistent with intragenic recombination in that the polymorphic nucleotide sites are spatially clustered.

Of the known mechanisms promoting genetic exchange in bacteria, two in particular (transduction and plasmid-mediated recombination) can be invoked as underlying the events proposed here. The size of the exchanged segments is too small to be consistent with conjugational transfer, but it is within the range that might be expected of transduction or plasmid-mediated events. Bacteriophage P1, for example, can transfer approximately 1 min (approximately 40 kb) of the *E. coli* chromosome (10). Other phages transfer molecules within the range 35–100 kb. Plasmid-mediated recombination involves still shorter segments (21). Conjugation-like events are less plausible explanations, as they typically involve regions of the chromosome an order of magnitude larger, and previous studies have shown that events of this scale are too rare to be consistent with finding four such events within one gene (7).

While there are several mechanisms for exchanging 20- to 100-kb segments, the events in Fig. 1 involve only a few hundred nucleotides. However, since the sequenced region was only 2 kb, events involving larger segments would not have been detected. Moreover, the proposed exchange event denoted A could involve a rather long segment. Since no sites upstream from the 3′ endpoint of event A are inconsistent with the proposed exchange, the 3′ endpoint of event A cannot be determined.

The size of DNA segments transferred by transducing phage or contained in plasmids imposes a physical limit on the size of the DNA segments involved in recombination. The presence of restriction/modification systems also poses a potentially significant limitation. Since the strains we studied represent the three major groupings of *E. coli* identified by Selander *et al.* (6), each of the strains might possess different restriction systems. Restriction endonucleases with four-base recognition sequences would be expected to generate fragments of 50–500 base pairs, which is consistent with the sizes of the events we postulate. Indeed, restriction/modification systems may serve to provide small segments of DNA with recombinogenic ends used as raw material for DNA recombination and repair.

Intragenic recombination observed in *phoA* confirms our earlier conjecture that different segments of the bacterial chromosome might have different phylogenetic histories (1). If the results with *phoA*, *gnd*, and *trp* are representative of the *E. coli* genome, then the chromosome must be a mosaic of phylogenetic histories, even within genes. An important implication is that any two homologous genes may be a mosaic of segments differing in their times of divergence from a common ancestral sequence. For example, consider two adjacent segments X and Y in a pair of recently diverged genes. If the X segment in one gene is replaced by a corresponding segment from a divergent homologous sequence X' Y' with which it shares a more remote common ancestor, the result is a pair of genes X' Y and X Y that contain closely related segments (Y versus Y) and more distantly related segments (X' versus X). When compared with a homologous sequence in a suitable outgroup, the X and Y segments will show different apparent rates of nucleotide substitution, as if each segment of the gene were governed by its own molecular clock, and the overall phylogeny of the

gene would be a sort of average phylogeny, of dubious biological significance. On the other hand, rates of molecular evolution and phylogenetic reconstruction are unaffected by genetic exchange events provided they are based on sequences derived from long-standing species between which genetic exchange does not regularly occur (22). The contrast is evident in intraspecific comparisons of *phoA*, *gnd*, and *trp* in *E. coli* versus interspecific comparisons with homologous sequences from *Salmonella typhimurium*. Within *E. coli*, *trp* has a relatively low level of nucleotide polymorphism, *phoA* intermediate, and *gnd* relatively large. However, in the two cases where homologous sequences exist (*trp* and *gnd*), these genes have approximately the same level of nucleotide substitution when compared with their counterparts in *Salmonella* (11, 12, 23, 24).

The finding of intragenic recombination of *E. coli* also encourages consideration of the meaning of DNA relatedness values among bacterial species. Brenner and collaborators (25, 26) have proposed that the species *E. coli* can be defined as isolates that have DNA relatedness values of 85% or greater, defined on the basis of DNA hybridization. Extant bacterial species do not show a continuous series of DNA relatedness values with *E. coli* that are arbitrarily close to 85%. Rather, various species show discrete gaps in relatedness values when compared with *E. coli* and with each other. The genetic similarity at the DNA level between conspecific bacterial isolates reflects in part their shared common ancestry, but it may also demarcate those groups of strains that can exchange genes among themselves by means of one or more of the genetic mechanisms that have been documented in bacteria (9, 10). If this is the case, the classical definition of eukaryotic species as groups of actual or potentially interbreeding natural populations that are reproductively isolated from other such groups may also extend to those groups of prokaryotes that are actually or potentially capable of exchanging small segments of chromosomal DNA.

We thank Louis Green for technical assistance, Stanley Sawyer for discussions on his method of recombinational analysis, and R. K. Selander, Howard Ochman, members of the Hartl laboratory, and the Population Biology Discussion Group at Stony Brook for comments and helpful discussion on earlier versions of the manuscript. We also thank Thomas S. Whittam and an anonymous reviewer for suggestions on improving the manuscript. This work was supported by National Institutes of Health Grant GM 30201 to D.E.D. and D.L.H. and National Institutes of Health Training Grant GM 08036 (R.F.D.).

1. Hartl, D. L. & Dykhuizen, D. E. (1984) *Annu. Rev. Genet.* **18,** 31–68.
2. Ørskov, I., Ørskov, F., Jann, B. & Jann, K. (1977) *Bacteriol. Rev.* **41,** 667–710.
3. Ochman, H. & Selander, R. K. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 198–201.
4. Smith, H. (1977) *Bacteriol. Rev.* **41,** 475–500.
5. Whittam, T. S., Caugant, D. A. & Selander, R. K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1751–1754.
6. Selander, R. K., Caugant, D. A. & Whittam, T. S. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, eds. Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. & Umbarger, H. E. (Am. Soc. Microbiol., Washington, DC), pp. 1625–1648.
7. Levin, B. R. (1981) *Genetics* **99,** 1–23.
8. Selander, R. K. & Levin, B. R. (1980) *Science* **210,** 545–547.
9. Smith, G. R. (1987) *Annu. Rev. Genet.* **21,** 179–201.
10. Scaife, J., Leach, D. & Galizzi, A., eds. (1985) *Genetics of Bacteria* (Academic, London).
11. Milkman, R. & Crawford, I. P. (1983) *Science* **221,** 378–380.
12. Sawyer, S., Dykhuizen, D. E. & Hartl, D. L. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6225–6228.
13. Chang, C. N., Kuang, W.-J. & Chen, E. Y. (1986) *Gene* **44,** 121–125.
14. Shuttleworth, H., Taylor, J. & Minton, N. (1986) *Nucleic Acids Res.* **14,** 8689.
15. Barnes, W. M. & Tuley, E. (1983) *J. Mol. Biol.* **165,** 443–459.
16. Michaelis, S., Inouye, H., Oliver, D. & Beckwith, J. (1983) *J. Bacteriol.* **154,** 366–374.
17. Lazzaroni, J. C., Atlan, D. & Portalier, R. C. (1985) *J. Bacteriol.* **164,** 1376–1380.
18. Barnes, W. M., Bevan, M. & Son, P. H. (1983) *Methods Enzymol.* **101,** 98–122.
19. Stephens, J. C. (1985) *Mol. Biol. Evol.* **2,** 539–556.
20. Dykhuizen, D. E. & Green, L. (1986) *Genetics* **113,** s71 (abstr.).
21. Shapiro, J., ed. (1983) *Mobile Genetic Elements* (Academic, Orlando, FL).
22. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
23. Barcak, G. J. & Wolf, R. E., Jr. (1988) *J. Bacteriol.* **170,** 365–371.
24. Barcak, G. J. & Wolf, R. E., Jr. (1988) *J. Bacteriol.* **170,** 372–379.
25. Brenner, D. J., Fanning, G. R., Johnson, K. E., Citarella, R. V. & Falkow, S. (1969) *J. Bacteriol.* **98,** 637–650.
26. Brenner, D. J., Fanning, G. R., Skerman, F. J. & Falkow, S. (1972) *J. Bacteriol.* **109,** 953–965.