Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4384–4388, May 1993
Population Biology

How clonal are bacteria?

(recombination/linkage disequilibrium/population structure/genetic transformation/parasite evolution)

JOHN MAYNARD SMITH, NOEL H. SMITH, MARIA O'ROURKE, AND BRIAN G. SPRATT*

School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom

Contributed by John Maynard Smith, January 25, 1993

ABSTRACT **Data from multilocus enzyme electrophoresis of bacterial populations were analyzed using a statistical test designed to detect associations between genes at different loci. Some species (e.g., *Salmonella*) were found to be clonal at all levels of analysis. At the other extreme, *Neisseria gonorrhoeae* is panmictic, with random association between loci. Two intermediate types of population structure were also found. *Neisseria meningitidis* displays what we have called an "epidemic" structure. There is significant association between loci, but this arises only because of the recent, explosive, increase in particular electrophoretic types; when this effect is eliminated the population is found to be effectively panmictic. In contrast, linkage disequilibrium in a population of *Rhizobium meliloti* exists because the sample consisted of two genetically isolated divisions, often fixed for different alleles: within each division association between loci was almost random. The method of analysis is appropriate whenever there is doubt about the extent of genetic recombination between members of a population. To illustrate this we analyzed data on protozoan parasites and again found panmictic, epidemic, and clonal population structures.**

Observations of the association of particular serotypes with disease initially suggested that bacterial populations are clonal (1). Recently, the methods, theory, and statistics of population biology have been applied to bacteria, notably through the application of multilocus enzyme electrophoresis (MLEE) (2–14). MLEE, which indexes the allelic variation in multiple chromosomal genes, has been highly successful in generating large data sets for the statistical analysis of bacterial populations and has shown that many species of bacteria exhibit strong linkage disequilibrium (the nonrandom association of alleles), with the frequent recovery of only a few of all the possible multilocus genotypes. Superficially these observations suggest that the population structure of most species of bacteria is clonal—that the rate of recombination of large chromosomal segments is not high enough to randomize genomes or break up clonal associations (15, 16).

The data supporting the clonality of *Escherichia coli* populations is extensive and persuasive; coefficients of linkage disequilibrium are near their theoretical maxima even when samples are taken from a restricted location (3, 17) and are independent of the map distance between the genes (18); strains of identical allozymic profile can be recovered from geographically and temporally unassociated hosts. Furthermore, in *E. coli* the dendrograms generated by MLEE are generally concordant with phylogenetic trees based on extensive analysis of nucleotide sequence (19–21). Although recombination is invoked to explain a few discrepancies between dendrograms derived from MLEE and nucleotide sequences, it is apparent that, at the population and the chromosomal level, *E. coli* is basically clonal.

That *E. coli* populations are clonal has been elevated, with notable exceptions (16, 17, 22), to the status of a paradigm extending to all bacterial populations. The data supporting clonality in other species of bacteria tend to rely heavily upon the demonstration of high coefficients of linkage disequilibrium by MLEE studies and the frequent recovery of one or a few multilocus genotypes. However, linkage disequilibrium can arise in bacterial populations in which recombination is frequent, in several ways:

(*i*) The samples analyzed may consist of a mixture of several populations, within each of which recombination is common, although it is rare between populations. The populations may be geographically (17) or ecologically isolated, or there may be biological barriers to gene exchange (23, 24).

(*ii*) An "epidemic" population structure can lead to temporary disequilibrium, as discussed below.

(*iii*) Epistatic fitness interactions between loci can maintain disequilibrium.

(*iv*) Disequilibrium can also arise by genetic drift.

A clonal structure of bacterial populations has also been deduced from the global distribution of a single multilocus genotype. Before the frequent recovery of such genotypes is taken as evidence of clonality, one must show that they cannot be explained simply by the random association, by recombination, of the most common alleles at each locus.

Therefore, there is a critical need for the development and application of statistical tests of population genetic data that will unequivocally index the extent of clonality within bacterial populations and can be used to compare the population structures of species from diverse ecological niches. We here report the application of a statistical test that can be used to determine the population structure of bacterial species. We have applied the test to several bacterial MLEE data sets and we have identified four types of bacterial populations: a fully sexual population such as *Neisseria gonorrhoeae*; a population such as *Neisseria meningitidis*, which is sexual but, because of its epidemic epidemiology, is superficially clonal; *Rhizobium*-like populations, which are sexual at the fine scale but do not recombine between populations; *Salmonella*-like populations that are clonal at all levels. Finally, we analyze some published data on protozoan parasites, and we ask whether they support the claim that such populations are clonal (25).

This analysis shows that bacterial populations are not invariably clonal but that they can occupy a spectrum of population structures ranging from the highly sexual *N. gonorrhoeae* to the almost strictly clonal *Salmonella*.

MATERIALS AND METHODS

A Measure of Association. Suppose that m loci have been analyzed in n individuals. Let p_{ij} be the frequency of the ith allele at the jth locus. Then $h_j = 1 - \Sigma p_{ij}^2$ is the probability that

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.

Abbreviations: MLEE, multilocus enzyme electrophoresis; ET, electrophoretic type.

*To whom reprint requests should be addressed.

two individuals are different at the jth locus. Let K be the "distance" between two individuals—that is, the number of loci at which they differ. There are in all $n(n-1)/2$ such pairs. The mean difference between two individuals is

$$\overline{K} = \sum h_j. \quad [1]$$

If the alleles present at different loci in an individual are independent—that is, if there is no linkage disequilibrium—the expected value of the variance of K is

$$V_E = \sum h_j(1 - h_j). \quad [2]$$

Knowing the observed values of p_{ij} it is easy to calculate V_E. By comparing V_O, the observed variance of K, with V_E one can obtain a measure of the degree of association between loci. A convenient "index of association" is

$$I_A = V_O/V_E - 1. \quad [3]$$

This index was first used by Brown *et al.* (26) to measure population structure in *Hordeum spontaneum*, a wild self-fertilizing grass, and was later used by Whittam *et al.* (5) to demonstrate nonrandom association of alleles in *E. coli*. It is a generalized measure of linkage disequilibrium and has an expected value of zero if there is no association between loci. The error variance of I_A, assuming no association, can be calculated, knowing that (26)

$$\mathrm{var}(V_E) = 1/n\left\{\sum h_j - 7\sum h_j^2 + 12\sum h_j^3 - 6\sum h_j^4 + 2\left[\sum h_j - \sum h_j^2\right]^2\right\}. \quad [4]$$

Expected Values of the Index of Association. If the "individuals" analyzed are gametes from a single large random-mating population, the expected value of I_A is zero. For bacteria reproducing by binary fission, the value will depend on the relative rates of two processes: the divergence of lineages through the fixation, by drift or selection, of new mutations and the transfer of genes between lineages by recombination.

In the absence of recombination, the value of I_A will depend on a number of factors, including the form of the phylogenetic tree, the number of loci analyzed, and the likelihood that electrophoretically indistinguishable mutations occur independently in different lineages. It is therefore best to distinguish only between values of I_A that differ significantly from zero and those that do not. However, the following result holds for an "infinite alleles" model (all mutations unique), a constant probability of change per unit time, and a "uniform tree" (all lineages have the same constant probability of splitting per unit time):

$$E(V_O) = \overline{K} + 1/3\overline{K}^2. \quad [5]$$

Note that $E(V_O)$ increases as $\overline{K}^2$, and hence as m^2, where m is the number of loci examined. In contrast, V_E increases only as m. Thus, if there is no recombination, and if the phylogeny is not a "star" phylogeny, diverging from a single common ancestor, the value of I_A increases with m. For this reason it was stated above that we should take note only of values of I_A that differ significantly from zero: absolute values of I_A are not a reliable guide to the degree of clonality.

Non-Zero Values of I_A. Values of I_A that differ significantly from zero indicate that recombination has been rare or absent. This is likely to be for either of two reasons:

(*i*) Lineages are spatially isolated (e.g., they live in different continents, in different hosts, or in different parts of the same host).

(*ii*) There is no mechanism for genetic exchange (e.g., between reproductively isolated sexual species or between bacterial strains that are too distant genetically to permit homologous recombination) or mechanisms detected in the laboratory are not significant in nature.

These two causes of non-zero values of I_A cannot be distinguished merely from the absolute value of I_A: additional information about the origin of the strains and their biology would be needed. It is possible, however, to gain further insight by calculating I_A for the complete data set and for subsets of the data. Two analyses can be made:

(*i*) *Analysis of subgroups.* The organisms can be subdivided by some type of phylogenetic analysis, either using the electrophoretic data themselves or in some other way. Does the evidence for association disappear when I_A is calculated separately for each subgroup? If the population structure is clonal at all levels (Fig. 1*A*), then $I_A > 0$ for subgroups as well as for the whole data set (see data on *Salmonella*, Table 1).

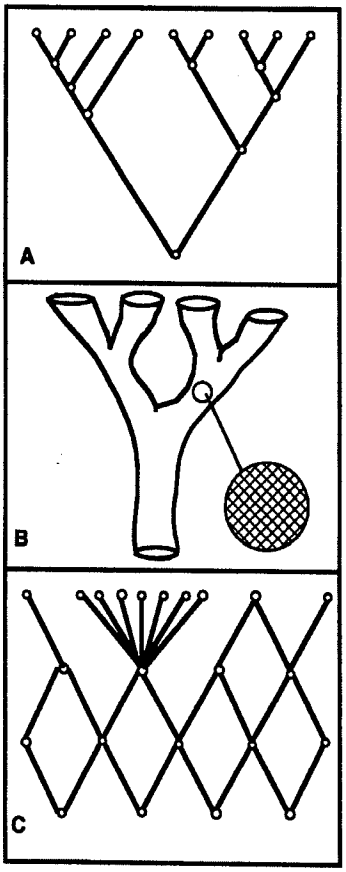

FIG. 1. Representations of population structures. (*A* and *B*) Populations consisting of isolates that are separated into two major branches. The population structure in *A* is clonal at all levels, such that the dendrogram is an evolutionary tree: no recombination occurs, either between isolates in the same, or different, branches of the tree. In *B*, recombination does not occur between isolates from the two major branches, but frequent recombination occurs between isolates within each major branch. The structure within these branches is thus net-like rather than tree-like, as represented by the expanded section of one of the major branches. (*C*) An epidemic structure in which there is frequent recombination within all members of the population, such that the structure is a net rather than a tree. However, occasionally a highly successful individual arises and increases rapidly in frequency to produce an epidemic clone.

Table 1. Measures of association between loci in bacteria

Bacteria	No. of isolates	No. of ETs	No. of loci	All isolates D	All isolates I_A*	ETs only D	ETs only I_A*	Ref.
N. gonorrhoeae	227	89	9	0.309	0.04 ± 0.09	0.414	−0.16 ± 0.17	†
N. meningitidis	688	331	15	0.507	1.96 ± 0.05	0.547	0.21 ± 0.08	9
Haemophilus influenzae	2209	56‡	17	0.226	5.40 ± 0.03	0.534	1.25 ± 0.18	27
"Division 1"	2117	42	17	0.186	4.42 ± 0.03	0.452	0.90 ± 0.21	
"Division 2"	92	14	17	0.361	2.30 ± 0.14	0.454	1.20 ± 0.37	
Salmonella	1495	106	24	0.355	3.11 ± 0.04	0.366	1.38 ± 0.14	28
S. panama	99	13	24	0.038	1.34 ± 0.16	0.138	0.28 ± 0.39	
S. paratyphi B	118	14	24	0.074	2.68 ± 0.13	0.168	0.36 ± 0.37	
S. typhimurium	340	17	24	0.035	1.03 ± 0.10	0.117	1.22 ± 0.33	
S. paratyphi C	100	9	24	0.054	4.66 ± 0.15	0.143	4.07 ± 0.46	
S. choleraesuis	174	11	24	0.036	2.57 ± 0.12	0.178	1.36 ± 0.42	
Rhizobium meliloti	232	50	14	0.238	6.34 ± 0.09	0.495	4.74 ± 0.19	14
Division A	208	34	14	0.101	0.47 ± 0.10	0.233	−0.24 ± 0.24	
Division B	23	15	14	0.219	0.24 ± 0.29	0.249	0.01 ± 0.35	
Legionella	170	62	22	0.359	4.63 ± 0.11	0.420	3.8 ± 0.18	6
L. pneumophila	143	50	22	0.252	1.49 ± 0.12	0.312	0.69 ± 0.20	
L. "species 1"	24	9	22	0.107	2.45 ± 0.29	0.182	1.93 ± 0.47	
Bordetella bronchiseptica	304	21	15	0.097	1.29 ± 0.08	0.248	0.99 ± 0.30	10
Pseudomonas syringae	23	10	26	0.447	18.35 ± 0.29	0.686	12.11 ± 0.44	11
P. syringae tomato	17	4	26	0.076	3.39 ± 0.33	0.180	1.68 ± 0.67	
P. syringae syringae	6	6	26	0.480	1.42 ± 0.57	0.480	1.42 ± 0.57	

ET, electrophoretic type; D, mean genetic distance per locus between strains.
*Standard errors are for the null hypothesis of independence between loci.
†M.O'R., unpublished data.
‡Fifty-six clusters identified by Musser *et al.* (27).

Alternatively, recombination may be occurring between similar lineages but not more distant ones (Fig. 1*B*): if so, I_A for the subgroups will approach zero (see data on *Rhizobium*, Table 1).

(*ii*) *Analysis of all isolates, of ETs, and of clusters.* In some cases, $I_A > 0$ because, although the population is effectively sexual in the long term, one or a few ETs have recently become abundant and widespread (Fig. 1*C*). In such cases, if each ET is treated as a single individual, the evidence of association will largely disappear (see data on *N. meningitidis*, Table 1, and *T. brucei*, Table 2). These are examples of what we shall call an epidemic population structure. If a single epidemic lasts long enough to accumulate some genetic variability, the epidemic nature of the structure can still be detected by taking clusters of ETs, instead of single ETs, as units. Thus, in *N. meningitidis*, if single ETs are taken as units, $I_A = 0.21 \pm 0.08$, a value that is just significant. However, Caugant *et al.* (9) identified 37 clusters of ETs in *N. meningitidis*: if each of these is treated as a unit, evidence for clonal structure disappears ($I_A = -0.14 \pm 0.17$).

This method of identifying an epidemic structure should be used with caution for two reasons. First, the method of sampling could lead to a significant positive value of I_A, even in a population with a high frequency of recombination. To give an extreme example, if an additional 100 isolates of *N. gonorrhoeae* had been taken from a single patient, I_A for the full data set could have been positive. A second difficulty is that with complete panmixis the expected value of I_A, calculated on ETs only, is not zero but negative. The effect is small unless the expected number of individuals with a particular ET, assuming independence, is large. We have checked by simulation that this bias is not large enough to alter our conclusions for the allele frequencies and numbers of loci we have observed.

RESULTS

Analysis of Bacterial Electrophoretic Data. Table 1 summarizes our analysis of a number of studies of electrophoretic variability in bacteria. The sources of the data, and some comments, follow:

(*i*) *N. gonorrhoeae*. Two hundred twenty-seven clinical isolates, collected worldwide between 1963 and 1989, were analyzed (M.O'R., unpublished data). The data suggest that the gonococcus is effectively sexual. This is confirmed by the following additional observations. Of the 36 linkage disequilibria calculated in pairs of alleles, none was statistically significant. The genetic variability within geographic regions is as great as within the whole population.

The recovery of the same ET in isolates that are separated in time and space is sometimes taken as evidence of a clonal population structure. The most prevalent ET in the *N. gonorrhoeae* data set, which exhibited the most common allele at each locus, and included isolates from several continents, occurred 35 times. However, the widespread occurrence of this ET cannot be taken as evidence of clonality since 32.2 isolates of this ET would be expected if loci are randomly assorting. Table 3 shows that the frequencies of the other more commonly recovered ETs also fit closely with the assumption of random assortment in a freely

Table 2. Measures of association between loci in protozoan parasites

Protozoa	No. of isolates	No. of ETs	No. of loci	All isolates D	All isolates I_A	ETs only D	ETs only I_A	Ref.
Plasmodium falciparum	29	29	13	0.500	0.19 ± 0.26	0.500	0.19 ± 0.26	29
Trypanosoma brucei	321	74	6	0.532	0.45 ± 0.07	0.622	0.09 ± 0.15	30
Trypanosoma cruzi	524	19	4	0.541	2.63 ± 0.06	0.781	0.86 ± 0.32	31

D, mean genetic distance per locus between strains.

Table 3. Observed and expected frequency of recovery of *N. gonorrhoeae* ETs, based on the random association of alleles

ET	Observed	Expected
1	35	32.23
3	12	10.59
33	9	10.95
40	11	8.99
All others	160	164.24
Total	227	227

χ^2 (4 df) = 1.34 (not significant). Unpublished data of M.O'R.

recombining population. It is relevant that *N. gonorrhoeae* is competent for transformation, and there is direct evidence for recombination in the evolution of drug resistance (32) and the generation of antigenic variation (33). It must also be the case that human behavior ensures frequent opportunities for gonococci of different genotypes to meet.

(*ii*) *N. meningitidis*. The data analyzed are for 688 clinical isolates, collected worldwide, mainly in the period 1970–1984, with a few earlier isolates from the United States. They were kindly provided by D. Caugant and are, in fact, the data analyzed by Caugant *et al.* (9) with a few additions. They provide a beautiful example of an epidemic population structure—a high value of I_A for the full data set, which reduces to 0.21 ± 0.08 (Table 1) when each ET is treated as a unit and which disappears when clusters of ETs are taken as units, as explained above. The most frequently recovered ET occurred 156 times. This ET (ET5) was first detected in the 1970s and has since caused epidemics in many countries. We predict that as a result of frequent recombination, this particular association of electromorphs will reduce in frequency and eventually fail to mark this epidemic. We note that variants of ET5 that differ at one or more enzyme-encoding loci are becoming increasingly common (D. Caugant, personal communication). *N. meningitidis* is also transformable and recombination has been detected in genes encoding cell surface structures (34–36) and genes involved in antibiotic resistance (37, 38).

(*iii*) *Haemophilus influenzae*. Musser *et al.* (27) analyzed 2209 isolates, collected worldwide, mainly in the period 1965–1987. The data include six serotypes, but most (1975 isolates) are of serotype b, the commonest cause of invasive disease. These authors found 280 distinct ETs, which they group into 56 clusters for phylogenetic analysis: our analysis is based on these clusters. *Haemophilus*, like *Neisseria*, is competent for transformation, and there is evidence that recombination occurs in nature (39). Surprisingly, however, evidence for clonal structure remains, even when each cluster is treated as a single individual and when each of the two main divisions identified by Musser *et al.* (27) is analyzed separately. There are, however, good reasons, independent of our estimates of I_A, to think that *Haemophilus* is clonal: particular ETs tend to be associated with particular serotypes and to have restricted geographical distributions (27).

(*iv*) *R. meliloti* is defined by the ability to form root nodules in association with one of three plant genera, *Medicago*, *Melilotis*, or *Trigonella*. *Medicago* includes the widely grown forage plant, alfalfa: *R. meliloti* has been inoculated into soils where alfalfa is to be grown. Eardly *et al.* (14) divided 232 isolates into two divisions, A and B, on the basis of their electrophoretic data: the reduced genetic distance within these divisions justifies the split. Division A was found worldwide and included the nine commercial inoculant strains examined. Of the 23 division B isolates, 21 were from annual *Medicago* species of the eastern Mediterranean. The I_A values confirm the genetic isolation between the two divisions and suggest that recombination within the divisions is common.

(*v*) *Salmonella*. The electrophoretic data for 1495 isolates of *Salmonella* are presented in Selander *et al.* (28). The traditional division into "species" is based not on the electrophoretic data but on serotype and the clinical symptoms caused. Recently, the most frequently encountered strains have been redesignated as serovars of a single species, *Salmonella enterica* (40), and it has been shown by MLEE that the genomic diversity in the salmonellae is roughly equal to that in *E. coli* (41). The striking reduction in genetic distance within serovars shows that they correspond to real genetic divisions and, in contrast to *R. meliloti*, the I_A values indicate a clonal structure within, as well as between, these genetic divisions (Table 1).

Protozoan Electrophoretic Data. It has been argued that many parasitic protozoa have a clonal population structure (25) and in Table 2 we present some data on protozoans. In the analysis, we have treated each diploid genotype at a locus as an "allele." This throws away some information but it is justified for our analysis because, in a panmictic population, the genotype at one locus is independent of that at others.

(*i*) *P. falciparum*. The data on *P. falciparum* are for 29 isolates from an African village (29): they are consistent with frequent sexual recombination, as expected since *Plasmodium* has an obligate sexual stage. A more thorough analysis by Day *et al.* (42) also concluded that *P. falciparum* populations are effectively panmictic, although these authors also find evidence of what we have called an epidemic population structure.

(*ii*) Trypanosomes. In contrast, Trypanosomes belong to a group, the Protomonads, for which cytological evidence of meiosis and syngamy is absent. However, Jenni *et al.* (43) infected tsetse flies simultaneously with two strains of *T. brucei* and recovered recombinants, so it seems that sex does take place in the secondary host, even if it is hard to observe cytologically. Mihok *et al.* (30) analyzed 321 strains of *T. brucei* from a valley in Kenya, isolated from tsetse flies, and from various primary hosts (humans, cattle, antelopes). The data, although less extensive than those for *N. meningitidis*, also suggest an epidemic population structure: sex is frequent, but occasionally a single ET becomes abundant. Although 74 ETs were found, 45 of the 144 isolates from humans belonged to a single ET, and 107 belong to one of four ETs. In contrast, the data on *T. cruzi* (31) do suggest a clonal population structure that is not merely the result of the explosive spread of a single type: however, it is not clear whether it is caused by spatial isolation or biological barriers to mating.

DISCUSSION

Our analysis shows that bacterial population structure ranges from effectively panmictic to one that is clonal at all levels. Two intermediate types of structure can also be recognized. *N. meningitidis* and, to give a protozoan example, *T. brucei*, show an epidemic structure. If all isolates are treated as individuals, there is a significant association between loci, but this arises largely because one or a few ETs have recently become widespread. On a larger time scale, recognized by treating ETs, or clusters of ETs, as units, the population structure is sexual. An opposite pattern is found in *R. meliloti*, which consists of two species, one panmictic and the other, perhaps, showing some association.

Other authors have used pairwise linkage disequilibrium values for electromorphs to show that some bacterial populations may be sexual. Souza *et al.* (17) analyzed data on *Rhizobium leguminosarum* biovar *phaseoli*, over several spatial scales, from individual host plants to the Western Hemisphere. They found that linkage disequilibrium values were high on large spatial scales and small, although still significant, on the scale of a single cultivated plot. They contrast the low degree of association in local populations with the high values

in *E. coli* and suggest that recombination may be common in *Rhizobium*, in agreement with our findings.

Istock *et al.* (22) studied *Bacillus subtilis*, another bacterium that, like *Neisseria* and *Haemophilus*, is competent for transformation. A strength of their study is that they analyzed samples from a single microsite, 200 cm^3 of desert soil. A single sample of ≈1 g of soil was taken in three successive years. In a total of 60 isolates, they found 50 unique ETs and five pairs of clone mates (clonal identity was checked by restriction fragment length polymorphism analysis). The majority of two-locus comparisons did not depart significantly from linkage equilibrium and they reasonably conclude that recombination must be common and suggest, in agreement with our findings, that bacterial population structures are likely to range from strictly clonal to panmictic. Interestingly, those bacterial species where recombination appears to be common (*N. gonorrhoeae*, *N. meningitidis*, and *B. subtilis*) are naturally transformable. However, transformability *per se* does not necessarily imply extensive recombination, since *H. influenzae* is also transformable but shows little evidence of frequent recombination in nature.

Phylogenetic trees are commonly derived from electrophoretic data sets. However, for those data sets where I_A is not significantly greater than zero, there is no justification for constructing trees: one might as well construct a tree for the members of a panmictic sexual population.

Several questions remain. Are any populations strictly clonal? It is relevant that in *E. coli* (reviewed in refs. 16 and 19) and in *Salmonella* (44) there is evidence of localized horizontal gene transfer in nature. However, the frequency of such transfer is too low to destroy linkage disequilibrium and must be detected by sequence analysis. Milkman and Bridges (19) estimate that in *E. coli* point mutations are about five times as common as localized recombination events.

Second, how frequent must recombination be to give an apparently panmictic population ($I_A = 0$)? This question is hard to answer analytically, but our unpublished simulations of populations with binary fission, mutation, and localized recombination (simulated by choosing two individuals at random, and changing a random locus in one of them to resemble that in the other) suggest that the "effective" recombination of a gene (that is, one that alters the ET of the recipient to resemble that of the donor, if they are different) must be 10–20 times as frequent as electrophoretically recognizable mutations in that gene for I_A to approach zero. If ≈70% of single nucleotide changes alter an amino acid, and 30% of amino acid changes are electrophoretically recognizable (45), then one-fifth of all point mutations will cause an electrophoretic change. If we suppose that all recombination events are effective, in the sense defined above, this suggests that to produce apparent panmixis, localized recombination affecting a gene must be 2–4 times as frequent as point mutations in that gene. This underestimates the frequency of recombination because not all recombination events will be effective.

In pathogenic bacteria, at least some capacity for localized recombination may be essential for the maintenance of useful variation (e.g., in cell surface genes) within the population against the purifying effects of repeated waves of periodic selection. However, the reason(s) for the apparently large variations in the extent of recombination in bacteria in nature, among pathogens and nonpathogens, remains obscure.

We are grateful to D. A. Caugant for supplying a complete electrophoretic data set from *N. meningitidis*. We are indebted to Gill Carpenter for technical assistance. This work was supported by The Wellcome Trust. B.G.S. is a Wellcome Trust Principal Research Fellow.

1. Ørskov, F. & Ørskov, I. (1983) *J. Infect. Dis.* **148,** 346–357.
2. Selander, R. K. & Levin, B. R. (1980) *Science* **210,** 545–547.
3. Caugant, D. A., Levin, B. R. & Selander, R. K. (1981) *Genetics* **98,** 467–490.
4. Ochman, H., Whittam, T. S., Caugant, D. A. & Selander, R. K. (1983) *J. Gen. Microbiol.* **129,** 2715–2726.
5. Whittam, T. S., Ochman, H. & Selander, R. K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1751–1755.
6. Selander, R. K., McKinney, R. M., Whittam, T. S., Bidd, W. F., Brenner, D. J., Nolte, F. S. & Pattison, P. E. (1985) *J. Bacteriol.* **163,** 1021–1037.
7. Young, J. P. W. (1985) *J. Gen. Microbiol.* **131,** 2399–2408.
8. Musser, J. M., Barenkamp, S. J., Granoff, D. M. & Selander, R. K. (1986) *Infect. Immun.* **52,** 183–191.
9. Caugant, D. A., Mocca, L. F., Frasch, C. E., Frøholm, L. O., Zollinger, W. D. & Selander, R. K. (1987) *J. Bacteriol.* **169,** 2781–2792.
10. Musser, J. M., Bemis, D. A., Ishikawa, H. & Selander, R. K. (1987) *J. Bacteriol.* **169,** 2793–2803.
11. Denny, T. P., Gilmour, M. N. & Selander, R. K. (1988) *J. Gen. Microbiol.* **134,** 1949–1960.
12. Musser, J. M., Kroll, J. S., Moxon, E. R. & Selander, R. K. (1988) *Infect. Immun.* **56,** 1837–1845.
13. Piñero, D., Martínez, E. & Selander, R. K. (1988) *Appl. Environ. Microbiol.* **54,** 2825–2832.
14. Eardly, B. D., Materon, L. A., Smith, N. H., Johnson, D. A., Rumbaugh, M. D. & Selander, R. K. (1990) *Appl. Environ. Microbiol.* **56,** 187–194.
15. Selander, R. K. & Musser, J. M. (1990) in *Molecular Basis of Bacterial Pathogenesis*, eds. Iglewski, B. H. & Clark, V. L. (Academic, San Diego), pp. 11–36.
16. Maynard Smith, J., Dowson, C. G. & Spratt, B. G. (1991) *Nature (London)* **349,** 29–31.
17. Souza, V., Nguyen, T. T., Hudson, R. R., Piñero, D. & Lenski, R. E. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8389–8393.
18. Whittam, T. S., Ochman, H. & Selander, R. K. (1983) *Mol. Biol. Evol.* **1,** 67–83.
19. Milkman, R. & Bridges, M. M. (1990) *Genetics* **126,** 505–517.
20. Nelson, K., Whittam, T. S. & Selander, R. K. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6667–6671.
21. Nelson, K. & Selander, R. K. (1992) *J. Bacteriol.* **174,** 6886–6895.
22. Istock, C. A., Duncan, K. E., Ferguson, N. & Zhou, X. (1992) *Mol. Ecol.* **1,** 95–103.
23. Cohan, F. M., Roberts, M. S. & King, E. C. (1991) *Evolution* **45,** 1393–1421.
24. Rayssiguier, C., Thaler, D. S. & Radman, M. (1989) *Nature (London)* **342,** 396–401.
25. Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Breniere, S. F., Darde, M.-L. & Ayala, F. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5129–5133.
26. Brown, A. H. D., Feldman, M. W. & Nevo, E. (1980) *Genetics* **96,** 523–536.
27. Musser, J. M., Kroll, J. S., Granoff, D. M., Moxon, E. R., Brodeur, B. R., Campos, J., *et al.* (1990) *Rev. Infect. Dis.* **12,** 75–111.
28. Selander, R. K., Beltran, P., Smith, N. H., Helmuth, R., Rubin, F. A., Kopecko, D. J., Ferris, K., Tall, B. T., Cravioto, A. & Musser, J. M. (1990) *Infect. Immun.* **58,** 2262–2275.
29. Babiker, H. A., Creasey, A. M., Fenton, B., Bayoumi, R. A. L., Arnot, D. E. & Walliker, D. (1991) *Trans. R. Soc. Trop. Med. Hyg.* **85,** 572–577.
30. Mihok, S., Otieno, L. H. & Darji, N. (1989) *Parasitology* **100,** 219–233.
31. Zhang, Q., Tibayrenc, M. & Ayala, F. J. (1988) *J. Protozool.* **35,** 81–85.
32. Spratt, B. G. (1988) *Nature (London)* **332,** 173–176.
33. Seifert, H. S., Ajioka, R. S., Marchal, C., Sparling, P. F. & So, M. (1988) *Nature (London)* **336,** 391–395.
34. Connell, T. D., Black, W. J., Kawula, T. H., Barritt, D. S., Dempsey, J. A., Kverneland, K., Stephenson, A., Schepart, B. S., Murphy, G. L. & Cannon, J. G. (1988) *Mol. Microbiol.* **2,** 227–236.
35. Feavers, I. M., Heath, A. B., Bygraves, J. A. & Maiden, M. J. C. (1992) *Mol. Microbiol.* **6,** 489–495.
36. Halter, R., Pohler, J. & Meyer, T. F. (1989) *EMBO J.* **8,** 2737–2744.
37. Spratt, B. G., Zhang, Q.-Y., Jones, D. M., Hutchison, A., Brannigan, J. A. & Dowson, C. G. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8988–8992.
38. Spratt, B. G., Bowler, L. D., Zhang, Q.-Y., Zhou, J. & Maynard Smith, J. (1992) *J. Mol. Evol.* **34,** 115–125.
39. Kroll, J. S. & Moxon, E. R. (1990) *J. Bacteriol.* **172,** 1374–1379.
40. Ewing, W. H. (1986) *Edwards and Ewing's Identification of Enterobacteriaceae* (Elsevier, New York), 4th Ed.
41. Selander, R. K. & Smith, N. H. (1990) *Rev. Med. Microbiol.* **1,** 219–228.
42. Day, K. P., Koella, J. C., Nee, S., Gupta, S. & Read, A. F. (1992) *Parasitology* **104,** 35–52.
43. Jenni, L., Marti, S., Schweizer, J., Betschart, B., Le Page, R. W. F., Wells, J. M., Tait, A., Paindavoine, P., Pays, E. & Steinert, A. (1986) *Nature (London)* **322,** 173–175.
44. Smith, N. H. & Selander, R. K. (1990) *J. Bacteriol.* **172,** 2209–2216.
45. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).