Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 198–201, January 1984
Genetics

Evidence for clonal population structure in *Escherichia coli*

(electrophoretic types/serotypes/outer membrane proteins/chromosomal genotypes)

HOWARD OCHMAN AND ROBERT K. SELANDER*

Department of Biology, University of Rochester, Rochester, NY 14627

Contributed by Robert K. Selander, September 12, 1983

ABSTRACT **Genotypes of 142 K1 isolates of four O serogroups of *Escherichia coli* from human hosts in Europe and the United States were characterized by an electrophoretic analysis of allozymic variation in 12 chromosomally encoded enzymes. The genetic structure of natural populations revealed by this analysis is closely similar to that indicated in earlier studies by Achtman and colleagues of the electrophoretic migration pattern for four outer membrane proteins and the chemical structure of the cell-wall lipopolysaccharides. The combined evidence demonstrates that most of the K1 isolates belong to a small number of geographically widespread clones. The distribution of O serogroups among the isolates does not consistently correspond to the clonal structure; O1:K1 isolates represent at least two distantly related, geographically widespread clones, one of which is genetically similar to a clone of the O18:K1 serotype. These findings for K1 isolates add to a growing body of evidence supporting the hypothesis that the genetic structure of natural populations of *E. coli* is basically clonal, with very limited recombination of chromosomal genes. Clonal structure has important implications for the study of the determinants of pathogenicity and disease specificity in *E. coli*.**

Of the many methods employed to distinguish strains of *Escherichia coli* isolated from natural populations in epidemiological and other research, serotyping of the O (lipopolysaccharide), K (capsular), and H (flagellar) antigens has long been the standard (1–5). Research over the past 40 years has revealed extensive variation in these and other serological characters (6) and associations of a relatively small number of serotypes with various intestinal and extraintestinal infections in humans and domesticated animals (7, 8). Serotypes have sometimes been considered uniform genetic entities and, therefore, meaningful, if not fundamental, units of population structure (7, 9), but recently this view has been questioned (10–12). In an analysis of the sodium dodecyl sulfate/polyacrylamide gel electrophoretic migration patterns of four outer membrane proteins (OMP) in a large number of K1 isolates recovered from human hosts in Europe and North America over a 40-year period, Achtman *et al.* (11) detected extensive variation; moreover, among O1:K1 isolates, they identified several markedly different OMP patterns and associated biotypes.

Over the past few years, we have studied genotypic variation in natural populations of *E. coli* and *Shigella* by a multilocus electrophoretic method whereby isolates are characterized by the relative mobilities of allelic variants (electromorphs) at 11 to 20 enzyme loci (13–19). Because the analysis by Achtman *et al.* (11) indicated that most of their K1 isolates could be assigned to a small number of geographically widespread clones, to which O serogroups did not closely correspond, we undertook the present study to determine the extent to which the population structure identified by OMP pattern actually reflects the organization of genetic diversity among isolates. Specifically, we have electrophoresed 12 chromosomally encoded enzymes to estimate genotypic diversity and relatedness among 142 of the K1 isolates of Achtman *et al.* (11), representing four O serogroups and six widespread OMP patterns.

MATERIALS AND METHODS

The isolates analyzed and their variable properties are listed in Table 1. With the exception of number 191, the isolates were recovered from human hosts. Dates of collection range from 1941 to 1981.

Each isolate was grown in Difco nutrient broth and processed to obtain an aqueous extract of proteins for electrophoresis (14). Techniques of horizontal starch-gel electrophoresis and the selective demonstration of specific enzyme activity were similar to those of Selander *et al.* (20). The 12 enzymes assayed were malate dehydrogenase (MDH), 6-phosphogluconate dehydrogenase (6PG), glucose-6-phosphate dehydrogenase (G6P), isocitrate dehydrogenase (IDH), alcohol dehydrogenase (ADH), phosphoglucose isomerase (PGI), mannose phosphate isomerase (MPI), adenylate kinase (AK), aconitase (ACO), indophenol oxidase (IPO), glutamate-oxaloacetate transaminase (GOT), and peptidase 2 (PE2). Each enzyme is encoded by a separate chromosomal gene (19, 21).

Electromorphs of each enyzme were numbered in order of decreasing anodal mobility, and strains lacking activity of a particular enzyme were scored as having a "null" allele at the corresponding enzyme locus. In this manner, a profile of electromorphs for the 12 enzymes was obtained for each isolate. Distinctive profiles, which are designated as electrophoretic types (ETs), were identified by letter (Table 1).

RESULTS

All but 2 of the 12 enzyme loci (IPO and IDH) were polymorphic, being represented by from two to five alleles (Table 1). Fourteen distinctive ETs were distinguished among the 142 K1 isolates examined. In the same set of isolates, there were four O serogroups and eight OMP patterns.

To graphically represent degrees of similarity among the 14 ETs, we plotted their scores for the first two factors derived from a principal components analysis of genotypes (Fig. 1). These two factors together account for 63% of the total variance. For purposes of this analysis, each of the total of 26 electromorph alleles at the 12 enzyme loci was considered to be a two-state variable, present or absent (see ref. 18). Also indicated in Fig. 1 are the OMP patterns associated with the various ETs.

O1:K1 Isolates. Most of the 56 O1:K1 isolates examined were either ET*a* or ET*e*, which have different alleles at 5 of the 12 loci (MDH, AK, PE2, ACO, and MPI) (Table 1). ET*a*

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: OMP, outer membrane protein(s); ET, electrophoretic type; LPS, lipopolysaccharide.

*To whom reprint requests should be addressed.

Table 1. List of K1 isolates and properties, including enzyme electromorph designations

Sequential number*	Number of isolates	Geographic source†	O serogroup	OMP pattern	ET	Polymorphic enzyme‡									
						MDH	6PG	AK	PE2	GOT	PGI	ACO	MPI	G6P	ADH
1–19, 21, 23–26	24	GA, Belg., Fin., Sw., Eng.	01	5	*a*	8	3	3	5	3	8	5	8	3	6
22	1	Eng.	01	3	*b*	8	6	3	5	1	6	2	8	3	4
36	1	Fin.	01	9	*c*	2	3	6	4	3	8	3	6	3	0
47	1	Eng.	01	9	*d*	2	3	6	4	3	8	3	6	3	10
27–34, 37–46, 48–55	26	Den., Fin., Sw., Ho., Eng.	01	9	*e*	2	3	6	4	3	8	3	6	3	6
56–58	3	Sw., Eng.	01	6	*e*	2	3	6	4	3	8	3	6	3	6
63, 64, 67–71, 74–77, 81, 83, 87–91	18	Fin., Sw., Ho., Eng.	018§	9	*f*	2	7	6	4	3	8	3	6	3	6
66	1	Fin.	018	9	*g*	2	7	6	4	3	8	3	6	3	0
72	1	Fin.	018	9	*h*	2	7	6	4	3	8	3	6	2	6
78	1	Eng.	018	9	*i*	2	7	6	4	3	8	3	6	1	0
92–96, 98–101, 103, 105, 107, 108, 114–116	16	NY, KY, AL, TX, WA, CA, Fin., Sw., Eng.	018§	6	*f*	2	7	6	4	3	8	3	6	3	6
102	1	NY	018	6	*g*	2	7	6	4	3	8	3	6	3	0
106, 111	2	Fin., Sw.	018	6	*h*	2	7	6	4	3	8	3	6	2	6
201–210	10	CA, Sw., Eng.	016	12	*j*	2	5	6	4	3	8	3	6	5	6
212	1	Sw.	016	12¶	*k*	2	4	6	4	3	8	3	6	2	6
151, 152, 155–158, 160–162, 164, 166–170, 172, 176–179, 181–186, 188, 189	28	NY, TX, WA, CA, Den., Fin., Sw., Eng.	07	3	*l*	8	4	6	5	3	8	5	8	3	6
153, 173–175	4	WA, Eng.	07	3	*m*	8	4	6	5	3	8	5	8	3	0
190	1	Sw.	07	16	*l*	8	4	6	5	3	8	5	8	3	6
191‖	1	Ger.	07	24	*n*	2	4	6	5	3	8	3	6	3	6
192	1	Sw.	07	15	*l*	8	4	6	5	3	8	5	8	3	6

Additional information on geographic source, biotype, plasmid content, and production of colicin and hemolysin are given in ref. 11.
*Numbers assigned by Achtman *et al.* (11), who also list the original strain designations.
†GA, Georgia; Belg., Belgium; Fin., Finland; Sw., Sweden; Eng., England; Den., Denmark; Ho, Holland; NY, New York; KY, Kentucky; AL, Alabama; TX, Texas; WA, Washington; CA, California; Ger., Germany.
‡Electromorphs are numbered in order of decreasing anodal mobility; IPO and IDH were monomorphic.
§Indicates isolates typed as 018, 018ab, and 018ac.
¶Described by Achtman *et al.* (11) as "similar" to 12.
‖Recovered from "rats" in Hannover, Federal Republic of Germany.

was associated with OMP pattern 5 (designated OMP5) in 24 isolates, and ET*e* was associated with OMP9 in 26 isolates. Two isolates (numbers 36 and 47) that did not have distinctive OMP patterns or biotypes were differentiated, as ET*c* and ET*d*, respectively, from the ET*e*/OMP9 isolates by variation in ADH. Three O1:K1 isolates (numbers 56–58) having OMP6 also were ET*e*, which is not surprising, inasmuch as OMP6 differs from OMP9 only by the absence of a plasmid-encoded protein (11). Isolate 22, with OMP3, was very distinctive genetically (ET*b*), differing from ET*a* and ET*e* at five enzyme loci.

O18:K1 Isolates. Isolates of the O18:K1 serotype were typed by Achtman *et al.* (11) as either OMP6 or OMP9. Among the 40 isolates we studied, ET*f* was associated with OMP9 in 18 and with OMP6 in 16. Four minor variations of the ET*f* profile were distinguished: ET*g*, represented by isolates 66 and 102; ET*h*, isolates 72, 106, and 111; and ET*i*, isolate 78.

O16:K1 Isolates. All 13 O16:K1 isolates studied by Achtman *et al.* (11) were OMP12 and were uniform in biotype. Similarly, 10 of the 11 isolates we examined (numbers 201–210) were indistinguishable, all having ET*j*; the exception was isolate 212, with ET*k*, which differs from ET*j* at two loci (6PG and G6P). Although Achtman *et al.* (11) assigned this isolate to OMP12, they noted that its pattern was atypical.

O7:K1 Isolates. Of the 35 isolates of the O7:K1 serotype examined, 28 were OMP3 and ET*l*; the genotypes of four other OMP3 isolates (numbers 153 and 173–175) differed from ET*l* only in having a null allele at the ADH locus (ET*m*). Isolates 190 and 192 were also ET*l*; these were assigned to OMP16 and OMP15, respectively, but, as shown in figure 3 of Achtman *et al.* (11), these patterns are similar to OMP3. In contrast, isolate 191, with OMP24, which is unlike OMP3, had a very distinctive enzyme profile (ET*n*).

DISCUSSION

The results of our electrophoretic analysis of allelic variation at 12 chromosomal enzyme loci confirm in all essential features the findings of Achtman *et al.* (11), based on sodium dodecyl sulfate/polyacrylamide gel electrophoresis of four

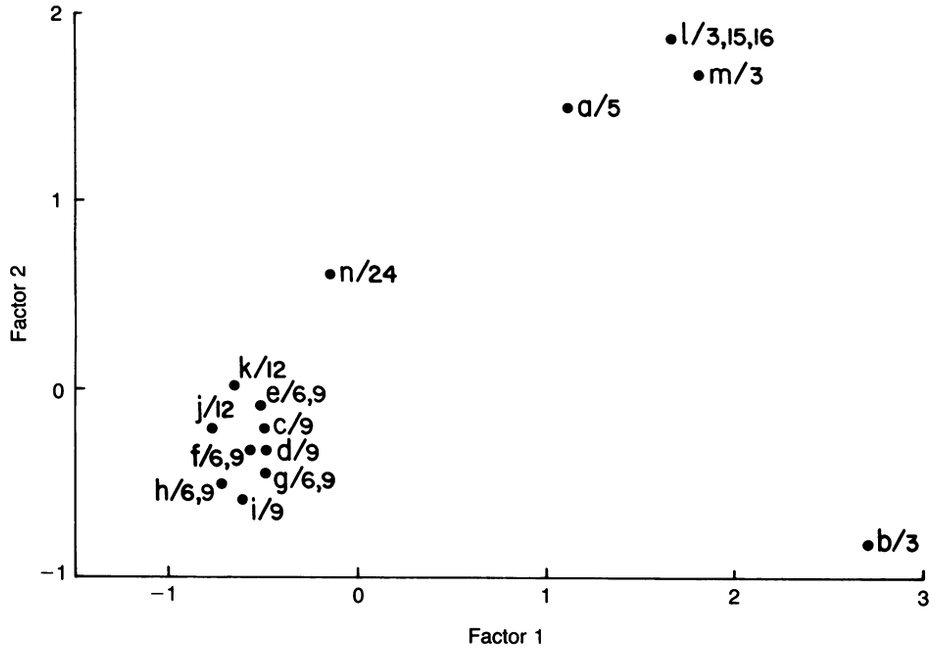

FIG. 1. Factor scores of 14 ETs of *E. coli*, with associated OMP patterns.

outer membrane proteins, with respect to both the apparent clonal genetic structure of K1 isolates from Europe and North America and the lack of a consistent relationship between chromosomal genotype and O serogroup. Because we assayed variation in more proteins than did Achtman *et al.* (11), it is not surprising that we were able to identify a larger number of distinctive genetic types among the 142 isolates studied. However, the principal findings of the two studies are remarkably similar. They clearly demonstrate that among O1:K1 isolates there are two common, geographically widespread clones (ET*a*/OMP5 and ET*e*/OMP6 or 9), which have no special genetic relatedness, and a third, apparently less common, genotype (ET*b*/OMP3) that is very different from either of the widespread clones and, for that matter, from all other genotypes represented in the sample of 142 isolates examined. The two widespread clones apparently are of independent origin and have converged with respect to serotype through either selection or recombination. Further analysis by Kusecek *et al.* (22) has shown that their lipopolysaccharides (LPSs) differ in chemical structure and antigenic properties, although they are indistinguishable by conventional methods of serotyping. These clones also differ in the average amount of capsular polysaccharide synthesized and in fimbriation. Most isolates of the ET*a*/OMP5 clone either lack fimbriae or have only the type I fimbriae, whereas almost all isolates of ET*e*/OMP6 or 9 have both type I and type P fimbriae.

It is noteworthy that O1:K1 isolate number 22, with ET*b*/OMP3, which is genotypically very distinctive (Fig. 1), synthesizes a unique LPS (22).

The two groups of O18:K1 isolates that were distinguished as subclones by Achtman *et al.* (11) by having either OMP6 or OMP9 were uniform in enzyme genotype (ET*m*). However, the results of the two studies are not inconsistent, inasmuch as the only difference between OMP6 and OMP9 is the absence or presence, respectively, of the PCP protein, the structural gene of which is carried on a plasmid (11). Multilocus enzyme electrophoresis clearly demonstrated that O1 and O18 isolates with either OMP6 or OMP9 are very closely related; indeed, their enzyme profiles differed at only 1 of the 12 loci assayed. Analysis of the LPSs of K1 isolates has yielded results consistent with those obtained by multilocus enzyme electrophoresis, in that O1 and O18 isolates with OMP9 have different LPS types, whereas the LPSs of O18 isolates of OMP6 and OMP9 are identical in structure, as are the LPSs of O1 isolates typed as OMP6 and OMP9 (22).

Electrophoretic studies of both OMPs and enzymes have identified O16:K1 isolates as a common, widespread clone (ET*j*/OMP12) and an apparently less common subclone. The results of OMP and enzyme analyses of O7:K1 isolates are also in close agreement. The major finding is that most isolates belong to one widespread clone, characterized by ET*l*/OMP3.

Biotype (based on nine tests), fimbriation, and amount of capsular polysaccharide synthesized tend to be uniform within clones identified by OMP pattern (11, 22). We note that O1:K1 isolates of the genetically very dissimilar ET*a*/OMP5 and ET*e*/OMP9 clones differ in about half of the biotype characters, whereas the genotypically similar ET*e*/OMP9 and ET*f*/OMP6 or 9 genotypes are closely similar in biotype. Among the O7:K1 isolates, number 191 (ET*n*/OMP24), which is highly distinctive genotypically, has a unique LPS and also makes less capsular polysaccharide than do isolates of the widespread O7 clone (ET*l*/OMP3), from which it can also be distinguished by an absence of type P fimbriae. In contrast, isolate number 192 (ET*l*/OMP15), which is genotypically similar to the common clone (ET*l*/OMP3), resembles that clone in LPS structure and fimbriation.

Our findings add to a growing body of evidence supporting the hypothesis that the genetic structure of natural populations of *E. coli* is basically clonal, with very limited recombination of chromosomal genes among cell lineages (6, 14, 18, 23, 24). The clone concept was formulated for *E. coli* to account for the global distribution and temporal stability of certain serotypes associated with pathogenicity (7). However, the present results, together with those of Ochman *et al.* (12), indicate that serotypes are unreliable indices of clonal identity. Indeed, as shown here, K1 isolates of the same O serogroup may be no more closely related than isolates chosen randomly from the *E. coli* population at large. Elsewhere we have shown that isolates of a variety of identical O:K, O:H, and O:K:H serotypes may be genotypically dissimilar (12).

In discussing the population structure of K1 isolates, Achtman *et al.* (11) noted that clonal assignments would be very different if plasmid pattern or H serotype, rather than OMP pattern, were considered. This observation raises the question of what types of characters are to be used in characterizing clones of *E. coli*. Achtman *et al.* (11) rejected plasmid contents and H serotypes because they are subject to "minor evolutionary variation" and do not, in fact, distinguish all the clones known to exist even on the basis of OMP pattern alone. If one is interested in studying the genetic structure of populations, serotyping may be an especially poor method, for reasons already discussed. Another serious weakness of serotyping is that a large portion of isolates cannot be typed, particularly for the K and H serogroups.

We believe that the analysis of the genetic structure of natural populations of bacteria should be founded on an assessment of the chromosomal genotype over a large number of genes. At present the only practical way of determining chromosomal genotypes in the large numbers of isolates required for studies of genetic variation in natural populations is by multilocus electrophoresis of enzymes, OMPs, and other proteins. A particularly attractive feature of enzyme polymorphisms is that the electrophoretic variants (allozymes) may be selectively neutral or nearly so (18, 25–27) and, therefore, minimally subject to evolutionary convergence. As shown elsewhere (17), estimates of genetic distance among strains of *E. coli* and *Shigella* based on our technique are correlated (r = 0.65–0.75) with estimates of polynucleotide sequence divergence obtained by DNA hybridization (28, 29).

Periodic selection and epistatic selection in clonally structured populations should lead to associations among the various genetically determined characters employed to differentiate isolates of *E. coli* (18). Indeed, there is evidence that this is generally so with regard to enzyme electrophoretic type, OMP pattern, LPS structure, production of capsular polysaccharide, and fimbriation (22). However, given the heterogeneity of serotype, biotype, and plasmid content among isolates of the same genotypic clone, it is likely that these characters will be of limited value in determining the genetic structure of natural populations.

The demonstration that O1:K1 isolates of the ET*e*/OMP9 clone are genetically very similar to O18:K1 isolates of the ET*f*/OMP9 clone has implicated the LPS (O antigen) as a determinant of virulence (22). Isolates of clone ET*e*/OMP9 are frequently recovered from urinary tract infections but rarely from cases of neonatal meningitis, whereas the reverse is true for isolates of the ET*f*/OMP9 clone (30–32). Experimental work with newborn rats has shown that isolates of ET*e*/OMP9 are efficient invaders of the blood stream after penetration of the gut mucosa, whereas ET*f*/OMP9 isolates are incapable of producing high levels of bacteremia (32). These findings indicate that knowledge of the genetic structure of natural populations may be essential for an understanding of the causes of pathogenicity and disease specificity in *E. coli*.

We are indebted to M. Achtman for supplying the isolates. This research was supported by grants to the University of Rochester from the National Institutes of Health and the Center for Naval Analysis.

1. Edwards, P. R. & Ewing, W. H. (1972) *Identification of Enterobacteriaceae* (Burgess, Minneapolis, MN).
2. Kauffmann, F. (1944) *Acta Pathol. Microbiol. Scand.* **21,** 20–45.
3. Kauffmann, F. (1975) *Classification of Bacteria* (Munksgaard, Copenhagen).
4. Ørskov, I., Ørskov, F., Jann, B. & Jann, K. (1977) *Bacteriol. Rev.* **41,** 667–710.
5. Jann, K. & Jann, B. (1983) *Prog. Allergy* **33,** 53–79.
6. Ørskov, I. & Ørskov, F. (1983) *Prog. Allergy* **33,** 80–105.
7. Ørskov, F., Ørskov, I., Evans, D. J., Jr., Sack, R. B., Sack, D. A. & Wadstrom, T. (1976) *Med. Microbiol. Immunol.* **162,** 73–80.
8. Bettelheim, K. A. (1978) *J. Hyg.* **80,** 83–111.
9. Bettelheim, K. A., Ismail, N., Shinebaum, R., Shooter, R. A., Moorehouse, E. & Farrell, W. (1976) *J. Hyg.* **76,** 403–406.
10. Achtman, M., Mercer, A., Kusecek, B. & Pohl, A. (1982) in *Bacterial Vaccines*, Seminars in Infectious Disease, eds. Robbins, J. B., Hill, J. C. & Sadoff, J. C. (Thieme-Stratton, New York), Vol. 4, pp. 233–241.
11. Achtman, M., Mercer, A., Kusecek, B., Pohl, A., Heuzenroeder, M., Aaronson, W., Sutton, A. & Silver, R. P. (1983) *Infect. Immun.* **39,** 315–335.
12. Ochman, H., Wilson, R. A., Whittam, T. S. & Selander, R. K. (1984) in *Proceedings of the Fourth International Symposium on Neonatal Diarrhea* (Veterinary Infectious Diseases Organization Publications, Univ. of Saskatchewan, Saskatoon), in press.
13. Selander, R. K. & Levin, B. R. (1980) *Science* **210,** 545–547.
14. Caugant, D. A., Levin, B. R. & Selander, R. K. (1981) *Genetics* **98,** 467–490.
15. Caugant, D. A., Levin, B. R., Lidin-Janson, G., Whittam, T. S., Svanborg Eden, C. & Selander, R. K. (1983) *Prog. Allergy* **33,** 203–227.
16. Ochman, H. & Selander, R. K. (1984) *J. Bacteriol.*, in press.
17. Ochman, H., Whittam, T. S., Caugant, D. A. & Selander, R. K. (1983) *J. Gen. Microbiol.* **129,** 2715–2726.
18. Whittam, T. S., Ochman, H. & Selander, R. K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1751–1755.
19. Whittam, T. S., Ochman, H. & Selander, R. K. (1983) *Mol. Biol. Evol.* **1,** 67–83.
20. Selander, R. K., Smith, M. H., Yang, S. Y., Johnson, W. E. & Gentry, J. B. (1971) *Univ. Tex. Publ. Stud. Genet.* **6,** 49–90.
21. Bachmann, B. J. (1983) *Microbiol. Rev.* **47,** 180–230.
22. Kusecek, B., Wloch, H., Mercer, A., Vaisanen, V., Pluschke, G., Korhonen, T. & Achtman, M. (1984) *Infect. Immun.* **43,** in press.
23. Milkman, R. & Crawford, I. P. (1983) *Science* **221,** 378–380.
24. Silver, R. P., Aaronson, W., Sutton, A. & Schneerson, R. (1980) *Infect. Immun.* **29,** 200–206.
25. Hartl, D. L. & Dykhuizen, D. E. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6344–6348.
26. Dykhuizen, D. E. & Hartl, D. L. (1980) *Genetics* **96,** 801–817.
27. Dykhuizen, D. E. & Hartl, D. L. (1983) *Genetics* **105,** 1–18.
28. Brenner, D. J., Fanning, G. R., Skerman, F. J. & Falkow, S. (1972) *J. Bacteriol.* **109,** 953–965.
29. Brenner, D. J., Fanning, G. R., Miklos, G. V. & Steigerwalt, A. G. (1973) *Int. J. Syst. Bacteriol.* **23,** 1–7.
30. Robbins, J. B., McCracken, G. H., Gotschlich, E. C., Ørskov, F., Ørskov, I. & Hanson, L. A. (1974) *N. Engl. J. Med.* **290,** 1216–1220.
31. Sarff, L. D., McCracken, G. H., Jr., Schiffer, M. S., Glode, M. P., Robbins, J. B., Ørskov, I. & Ørskov, F. (1975) *Lancet* **i,** 1099–1104.
32. Pluschke, G., Mercer, A., Kusecek, B., Pohl, A. & Achtman, M. (1983) *Infect. Immun.* **39,** 599–608.