JOURNAL OF BACTERIOLOGY, Feb. 1984, p. 690–693
0021-9193/84/020690-04$02.00/0

Vol. 157, No. 2

Standard Reference Strains of *Escherichia coli* from Natural Populations

HOWARD OCHMAN AND ROBERT K. SELANDER*

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

Received 29 August 1983/Accepted 3 November 1983

A set of 72 reference strains of *Escherichia coli* isolated from a variety of hosts and geographical locations has been established for use in studies of variation and genetic structure in natural populations. The strains, which have been characterized by multilocus enzyme electrophoresis, are representative of the range of genotypic variation in the species as a whole.

Most information on genotypic and phenotypic variation in *Escherichia coli* has been derived from a small number of laboratory strains (1), but in recent years there has been increasing interest in extending the analysis of variation to natural populations. In response to requests from a number of our colleagues whose research is concerned with the distribution of variable characters among *E. coli* isolates from natural sources, we have assembled a set of 72 strains that are representative of the genotypic diversity, as indexed by multilocus enzyme electrophoresis, in the species as a whole. Because the genetic distance between pairs of strains determined by multilocus electrophoretic profiles is positively correlated (r = 0.65 to 0.75) with percent divergence in chromosomal nucleotide sequence, as estimated by DNA hydridization (12), the enzyme loci assayed are believed to be a random sample of the chromosomal genome.

The selection of these 72 reference strains (Table 1) from our collection of 2,600 *E. coli* isolates from natural populations was based on the following considerations. (i) Previous use in published studies from our or other laboratories. The set includes many strains originally collected by Roger Milkman and studied by him and other investigators over the past decade. Table 1 lists references to studies in which the various strains have been used. (ii) Representation of the three subspecific groups of *E. coli* characterized by different associations of alleles at enzyme loci, as demonstrated electrophoretically (12, 15). We have chosen strains whose positions in factor space, produced by a principal components analysis of allelic diversity at 11 enzyme loci, encompass the range of variation within each group (Fig. 1). (Principal components analysis assesses interrelationships among a number of characters from a correlation or covariance matrix and reduces the total variation to a smaller number of "factors," which are combinations of correlated variables.) (iii) Host species and geographical distribution. Within each of the three subspecific groups, we selected strains that were isolated from different continents and from both human and animal hosts. Additionally, there are within each group some strains that have identical electromorph profiles over the 11 enzyme loci but which were recovered from different hosts; and each group includes a pair of pathogenic and nonpathogenic isolates with the same profile. A few pairs of strains with distinct profiles that were isolated from the same host individual are also included.

Electromorph profiles of the 72 reference strains are shown in Table 1. Electrophoretic and staining techniques are described elsewhere (3, 14). The 11 enzymes assayed were malate dehydrogenase, 6-phosphogluconate dehydrogenase, adenylate kinase, peptidase-2, glutamic oxaloacetic transaminase, isocitrate dehydrogenase, phosphoglucose isomerase, aconitase, mannose phosphate isomerase, glucose-6-phosphate dehydrogenase, and alcohol dehydrogenase.

For information on the availability of the reference strains for research, contact the authors. A set of strains has been deposited in the American Type Culture Collection (12301 Parklawn Dr., Rockville, MD 20852).

---

* Corresponding author.

TABLE 1. Standard reference strains and electromorph mobility profiles

Strain[a]		Source		References	Group[b]	Enzyme[c]										
No.	Previous designation[a]	Host (sex)	Location			MDH	6PG	ADK	PE2	GOT	IDH	PGI	ACO	MPI	G6P	ADH
1	RM74A	Human (F)	Iowa	8, 9, 10, 12, 13, 15, 16	I	2	6	4	5	3	2	4	7	3	2	1
2	STM1	Human (M)	New York	12, 15	I	2	6	4	5	3	2	4	7	3	2	1
3	WIR1(a)	Dog	Massachusetts	12, 15	I	2	6	4	5	3	2	4	7	3	2	1
4	RM39A	Human (F)	Iowa	8–10	I	2	15	4	7	3	2	4	6	3	2	1
5	RM60A	Human (F)	Iowa	8, 9, 12, 13, 15, 16	I	2	4	4	5	3	2	4	7	3	2	1
6	RM66C	Human (M)	Iowa	5, 6, 8, 9, 11–13, 15, 16	I	2	13	4	5	3	2	4	6	3	2	1
7	RM73C	Orangutan	Washington (zoo)	5, 8, 9, 12, 13, 15	I	2	5	4	7	3	2	4	7	3	1	1
8	RM77C (b)	Human (F)	Iowa	4, 7–9, 12, 13, 15, 16	I	2	9	4	5	3	2	4	7	3	2	1
9	FN98	Human (F)	Sweden	2, 12, 15, 16	I	2	9	4	5	3	2	4	7	3	2	1
10	ANI	Human (F)	New York	12, 15	I	2	9	4	5	3	2	4	7	3	2	1
11	C97	Human (F)	Sweden	2, 12, 15, 16	I	2	9	4	5	3	2	4	7	3	2	1
12	FN59	Human (F)	Sweden	2, 12, 15, 16	I	2	6	4	5	3	5	4	7	3	2	4
13	FN10	Human (F)	Sweden	2, 12, 15, 16	I	2	6	4	7	3	2	4	7	3	1	1

TABLE 1—*Continued*

Strain[a]		Source		References	Group[b]	Enzyme[c]										
No.	Previous designation[a]	Host (sex)	Location			MDH	6PG	ADK	PE2	GOT	IDH	PGI	ACO	MPI	G6P	ADH
14	P62	Human (F)	Sweden	2, 12, 15, 16	I	2	6	4	7	3	2	4	7	3	1	1
15	FN3	Human (F)	Sweden	2, 12, 15, 16	I	2	6	4	5	3	5	4	7	8	2	0
16	RM191F (c)	Leopard	Washington (zoo)	8–10	I	2	2	4	7	3	2	4	7	8	2	1
17	RM200Q	Pig	Indonesia	4, 7–9	I	2	18	4	5	3	2	4	7	3	2	1
18	RM210F (d)	Celebese ape	Washington (zoo)	6, 8, 9	I	2	6	4	7	3	2	4	7	3	1	1
19	RM210J (d)	Celebese ape	Washington (zoo)	6, 8–11	I	2	19	4	7	3	2	4	7	3	1	1
20	RM213I (e)	Steer	Bali	6, 8, 9, 11	I	2	4	4	7	3	5	4	7	3	1	1
21	RM213K (e)	Steer	Bali	6, 8, 9, 11	I	2	4	4	7	3	5	4	7	3	1	1
22	RM215C	Steer	Bali	4–9	I	2	6	4	7	3	2	4	6	8	2	1
23	RM183E	Elephant	Washington (zoo)	8, 9, 12, 13, 15	I	2	6	4	5	3	5	4	7	3	1	0
24	FN33	Human (F)	Sweden	2, 12, 15, 16	I	2	6	4	5	3	2	4	5	7	2	1
25	MSI	Dog	New York	12, 15	I	2	5	4	5	3	5	4	7	3	2	1
26	LL	Human infant	Massachusetts	12, 13, 15	II	2	6	2	5	3	5	5	6	7	2	6
27	RM24J	Giraffe	Washington (zoo)	8, 9, 12, 13, 15	II	2	6	2	5	3	5	5	6	7	2	6
28	RM52B	Human (F)	Iowa	8, 9, 12, 15, 16	II	2	6	4	5	3	5	7	6	8	2	6
29	RM3A	Kangaroo rat	Nevada	8, 9, 12, 13, 15	II	2	8	4	5	3	5	7	6	9	2	6
30	RM10A	Bison	Alberta	6, 8, 9, 11	II	2	8	4	5	3	5	7	6	9	2	6
31	RM12 (c)	Leopard	Washington (zoo)	8, 9, 12, 13, 15	II	2	6	4	5	3	5	4	6	5	2	6
32	RM28 (f)	Giraffe	Washington (zoo)	8, 9, 12, 13, 15	II	2	6	4	5	3	5	7	6	9	2	6
33	RM56C	Sheep	California	8, 9, 12, 13, 15	II	2	6	4	5	3	5	7	6	9	2	6
34	WIR2 (a)	Dog	Massachusetts	12, 15	II	2	6	4	5	3	5	7	6	9	2	6
35	RM42B	Human (M)	Iowa	8, 9, 12, 13, 15, 16	II	6	2	4	5	3	5	7	7	7	2	6
36	RM77B (b)	Human (F)	Iowa	8, 9, 12, 13, 15, 16	II	6	2	4	5	3	5	7	7	7	2	6
37	RM44B	Marmoset	Washington (zoo)	8, 9, 12, 13, 15	II	2	16	4	5	2	3	4	6	7	2	6
38	RM75A	Human (F)	Iowa	6, 8, 9, 11–13, 15, 16	II	6	4	5	5	3	5	7	7	7	2	6
39	FN104	Human (F)	Sweden	2, 12, 15, 16	II	6	4	5	5	3	5	7	7	7	2	6
40	P60	Human (F)	Sweden	2, 12, 15, 16	II	6	4	5	5	3	5	7	7	7	2	6
41	T44	Human (M)	Tonga	16	II	6	4	5	5	3	5	7	7	7	2	6
42	DAR1	Human (M)	Massachusetts	12, 15	II	2	6	4	5	3	5	4	6	7	2	6
43	FN36	Human (F)	Sweden	2, 12, 15, 16	II	2	6	4	5	3	5	4	6	5	2	6
44	RM189I	Cougar	Washington (zoo)	8, 9	II	6	4	4	5	3	5	4	2	7	2	6
45	RM201C	Pig	Indonesia	8–10	II	2	6	4	5	3	5	7	6	8	2	6
46	RM202F	Celebese ape	Washington (zoo)	4, 7–9	II	6	2	4	2	3	5	8	6	7	2	6
47	RM211C	Sheep	New Guinea	6, 8, 9, 11	II	7	4	4	5	3	5	4	2	7	2	0
48	C90	Human (F)	Sweden	2, 12, 15, 16	II	2	6	4	5	4	5	6	6	7	2	1
49	FN90	Human (F)	Sweden	2, 12, 15, 16	II	5	14	4	5	3	5	9	5	7	2	6
50	P97	Human (F)	Sweden	2, 12, 15, 16	II	5	14	4	5	3	5	9	5	7	2	6
51	DD	Human infant	Massachusetts	12, 13, 15, 16	III	2	6	5	2	3	5	7	6	3	2	6
52	RM73A	Orangutan	Washington (zoo)	8, 9, 12, 13, 15	III	2	6	5	2	3	5	7	6	3	2	6
53	RM33B	Human (F)	Iowa	8, 9, 12, 13, 15, 16	III	2	6	4	4	3	5	7	6	3	2	6
54	RM64A	Human	Iowa	8, 9, 12, 13, 15, 16	III	2	6	5	2	3	5	7	6	3	2	6
55	FN4	Human (F)	Sweden	2, 12, 15, 16	III	2	6	5	2	3	5	7	6	3	2	6
56	P106	Human (F)	Sweden	2, 12, 15, 16	III	2	6	5	2	3	5	7	6	3	2	6
57	RM71B	Gorilla	Washington (zoo)	8, 9, 12, 13, 15	III	2	15	5	2	3	5	7	6	3	2	6
58	RM185S	Lion	Washington (zoo)	8, 9, 12, 13, 15	III	2	6	4	2	4	5	7	7	8	2	2
59	SIL8	Human (M)	Massachusetts	12, 15	III	2	6	4	4	3	5	7	6	3	2	6
60	C89	Human (F)	Sweden	2, 12, 15, 16	III	2	6	4	4	3	5	7	6	3	2	6
61	FN23	Human (F)	Sweden	2, 12, 15, 16	III	2	14	5	4	3	5	7	6	3	2	0
62	P69	Human (F)	Sweden	2, 12, 15, 16	III	2	14	5	4	3	5	7	6	3	2	0
63	FN21	Human (F)	Sweden	2, 12, 15, 16	III	2	3	5	4	3	5	7	5	3	2	6
64	C70	Human (F)	Sweden	2, 12, 15, 16	III	2	6	5	4	3	5	7	5	7	2	6
65	RM202I	Celebese ape	Washington (zoo)	8–10	III	2	6	5	4	3	5	8	7	9	2	0
66	RM209I	Celebese ape	Washington (zoo)	6, 8, 9, 11	III	2	6	4	1	4	5	7	6	6	2	8
67	RM217T	Goat	Indonesia	8–10	III	2	10	4	5	3	5	7	6	3	2	6
68	RM224H (f)	Giraffe	Washington (zoo)	8–10	III	2	6	4	4	3	5	4	6	7	2	6
69	RM45EM	Celebese ape	Washington (zoo)	8–10	III	2	9	2	5	3	5	7	6	3	2	6
70	RM70B	Gorilla	Washington (zoo)	8–10, 12, 13, 15	III	2	6	4	7	3	5	2	7	3	2	0
71	ABU84	Human (F)	Sweden	2, 12, 15, 16	III	2	5	4	7	3	5	2	6	3	2	0
72	P68	Human (F)	Sweden	2, 12, 15, 16	III	2	2	4	7	3	5	1	6	3	2	7

[a] Strain designations prefixed by RM are from the collection of Roger Milkman. Strains prefixed with C, P, or ABU were isolated from the urine of women with urinary tract infections. All other strains were obtained from feces of healthy hosts. Pairs of strains isolated from the same host individuals are indicated by letters in parentheses.

[b] Group assignments are based on multilocus electrophoretic genotypes (13, 16); see Fig. 1.

[c] Electromorphs are numbered in order of decreasing anodal mobility and are cognate with those reported in references 12, 15, and 16. Abbreviations: MDH, malate dehydrogenase; 6PG, 6-phosphogluconate dehydrogenase; ADK, adenylate kinase; PE2, peptidase-2; GOT, glutamic oxaloacetic transaminase; IDH, isocitrate dehydrogenase; PGI, phosphoglucose isomerase; ACO, aconitase; MPI, mannose phosphate isomerase; G6P, glucose-6-phosphate dehydrogenase; ADH, alcohol dehydrogenase.

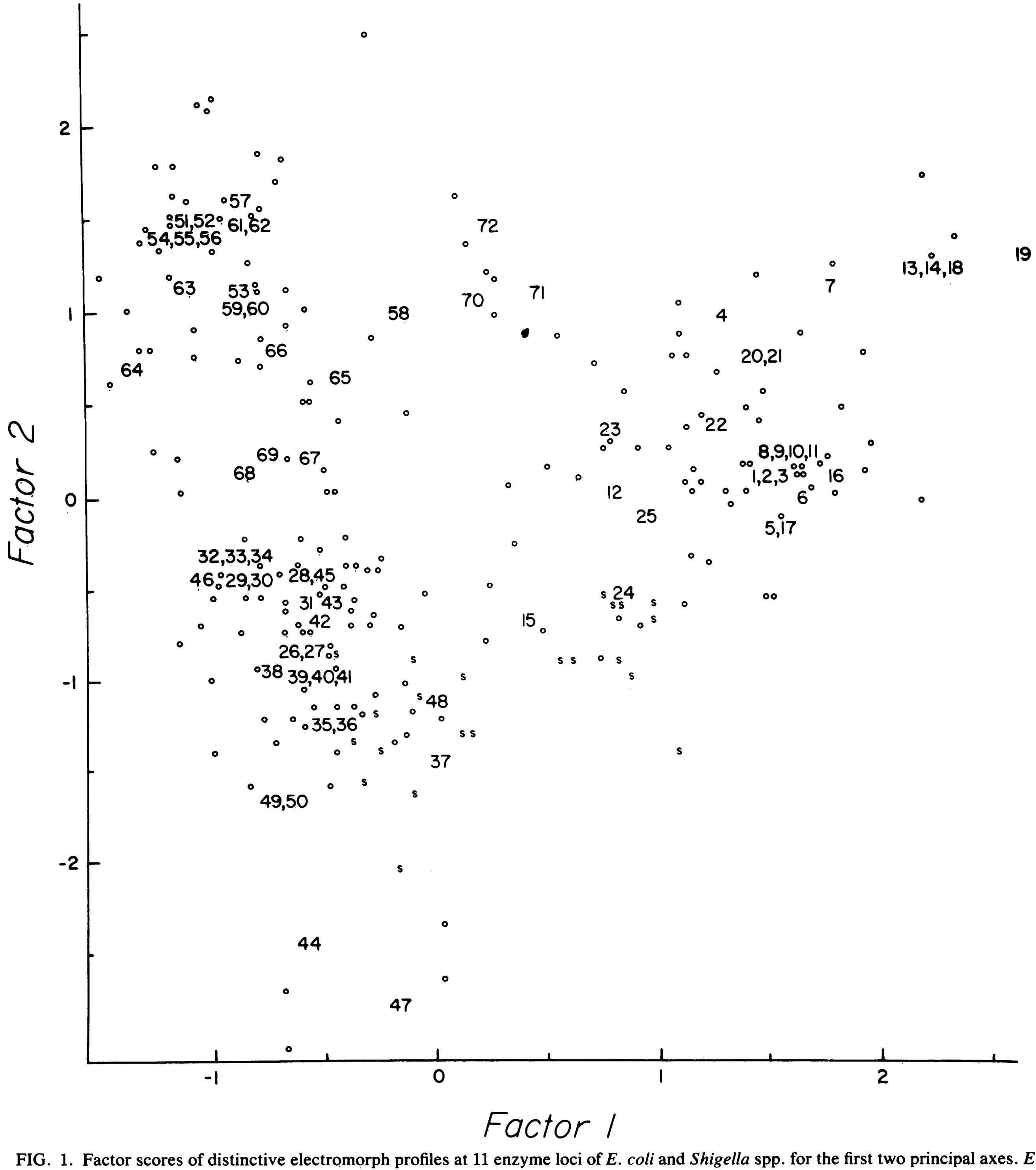

FIG. 1. Factor scores of distinctive electromorph profiles at 11 enzyme loci of *E. coli* and *Shigella* spp. for the first two principal axes. *E. coli* reference strains are indicated by number. Symbols: O, other strains of *E. coli*; s, *Shigella* spp. (For a description of the principal components analysis, see references 12 and 15.)

We thank D. Davison, D. Dykhuizen, B. R. Levin, and R. Milkman for supplying strains.

Research leading to the selection and characterization of the reference strains was supported by grants from the National Institutes of Health and the National Science Foundation.

LITERATURE CITED

1. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli*, edition 7. Microbiol. Rev. **47:**180–230.
2. **Caugant, D. A., B. R. Levin, G. Lidin-Janson, T. S. Whittam, C. Svanborg Eden, and R. K. Selander.** 1983. Genetic diversity and

relationships among strains of *Escherichia coli* in the intestine and those causing urinary tract infections. Prog. Allergy **33**:203–227.

3. **Caugant, D. A., B. R. Levin, and R. K. Selander.** 1981. Genetic diversity and temporal variation in the *E. coli* population of a human host. Genetics **98**:467–490.
4. **Dykhuizen, D. E., and D. L. Hartl.** 1980. Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. Genetics **96**:801–817.
5. **Dykhuizen, D. E., and D. L. Hartl.** 1983. Functional effects of PGI allozymes in *Escherichia coli*. Genetics **105**:1–18.
6. **Harshman, L., and M. Riley.** 1980. Conservation and variation of nucleotide sequences in *Escherichia coli* strains isolated from nature. J. Bacteriol. **144**:560–568.
7. **Hartl, D. L., and D. E. Dykhuizen.** 1981. Potential for selection among nearly neutral allozymes of 6-phosphogluconate dehydrogenase in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **78**:6344–6348.
8. **Milkman, R.** 1973. Electrophoretic variation in *Escherichia coli* from natural sources. Science **182**:1024–1026.
9. **Milkman, R.** 1975. Allozyme variation in *Escherichia coli* of diverse natural origins, p. 273–305. *In* C. L. Markert (ed.), Isozymes, vol. 4: Genetics and evolution. Academic Press, Inc., New York.
10. **Milkman, R., and I. P. Crawford.** 1983. Clustered third-base substitutions among wild strains of *Escherichia coli*. Science **221**:378–380.
11. **Nyman, K., H. Ohtsubo, D. Davison, and E. Ohtsubo.** 1983. Distribution of insertion element IS*1* in natural isolates of *Escherichia coli*. Mol. Gen. Genet. **189**:516–518.
12. **Ochman, H., T. S. Whittam, D. A. Caugant, and R. K. Selander.** 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J. Gen. Microbiol. **129**:2715–2726.
13. **Selander, R. K., and B. R. Levin.** 1980. Genetic diversity and structure in *Escherichia coli* populations. Science **210**:545–547.
14. **Selander, R. K., M. H. Smith, S. Y. Yang, W. E. Johnson, and J. B. Gentry.** 1971. Biochemical polymorphism and systematics in the genus *Peromyscus*. I. Variation in the old-field mouse (*Peromyscus polionotus*). Stud. Genet. **6**:49–90.
15. **Whittam, T. S., H. Ochman, and R. K. Selander.** 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **80**:1751–1755.
16. **Whittam, T. S., H. Ochman, and R. K. Selander.** 1984. Geographic components of linkage disequilibrium in natural populations of *Escherichia coli*. Mol. Biol. Evol. **1**:67–83.