Journal of General Microbiology (1991), **137**, 601–606. *Printed in Great Britain*

Reference collection of strains of the *Salmonella typhimurium* complex from natural populations

PILAR BELTRAN,[1] SHEILA A. PLOCK,[1] NOEL H. SMITH,[1] THOMAS S. WHITTAM,[1] DAVID C. OLD[2] and ROBERT K. SELANDER[1]*

[1]*Institute of Molecular Evolutionary Genetics, Mueller Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[2]*Medical Microbiology Department, University of Dundee Medical School, Ninewells Hospital, Dundee DD1 9SY, UK*

(*Received 17 April 1990; revised 3 December 1990; accepted 6 December 1990*)

A collection of 72 reference strains of the *Salmonella typhimurium* complex of clones recovered from a variety of hosts and environmental sources in diverse geographic locations has been established for use in studies of variation in natural populations. Included are strains of the serovars *S. typhimurium*, *S. saintpaul*, *S. heidelberg*, *S. paratyphi B* (including variety *java*) and *S. muenchen*. The strains, which have been characterized by enzyme electrophoresis for allelic variation in 24 chromosomal structural genes and represent 48 distinctive multilocus genotypes (electrophoretic types or ETs), exemplify the full range of genotypic variation in the *S. typhimurium* complex. Evolutionary genetic relationships among the ETs are indicated in a phylogenetic tree generated by the neighbour-joining method from a matrix of Nei's standard genetic distance.

Introduction

Classification and identification of strains of *Salmonella* within the seven subgroups (variously termed subgenera or subspecies) defined by DNA hybridization and biochemical characteristics (Crosa *et al.*, 1973; Le Minor *et al.*, 1986; Reeves *et al.*, 1989) have been based on the Kauffmann–White scheme of serotyping developed in the 1930s, augmented to some degree by biotyping (Ewing, 1986; Le Minor, 1984). Within subgenus I, which includes the great majority of the more than 2200 serovars that have been distinguished on the basis of antigenic variation in the somatic lipopolysaccharide (O) and the phase 1 (H1) and phase 2 (H2) flagellins, serovars are assigned to groups on the basis of their major O antigenic factors. But because serotyping and biotyping do not provide an adequate basis for estimating overall levels of genetic relatedness among strains, there has been no phylogenetic framework within which to study the evolution of host adaptation, pathogenicity, and other traits in this medically important group of bacteria.

Recently, an evolutionary genetic framework for the genus *Salmonella* has been constructed through application of the technique of multilocus enzyme electrophoresis, which estimates overall genetic relatedness among strains by indirectly assessing allelic variation in a large number of chromosomal structural genes encoding metabolic enzymes (Selander *et al.*, 1986). Limitations to the Kauffmann–White serological scheme of classification have been emphasized by the demonstration that strains of the same serovar may be distantly related in multilocus enzyme genotype and that the O antigen groups of subgenus I may similarly be polyphyletic (Beltran *et al.*, 1988; Selander *et al.*, 1990*a*, *b*). This work has also shown that the population structure of the salmonellae is basically clonal, with each serovar being represented by one or a few predominant clones of widespread if not global distribution (Beltran *et al.*, 1988; Reeves *et al.*, 1989; Selander *et al.*, 1990*a*, *b*). Comparative sequencing of the flagellin genes in strains of several serovars has implicated both intragenic and interstrain recombination as mechanisms generating new serovars (Frankel *et al.*, 1989; Smith & Selander, 1990; Smith *et al.*, 1990).

Most information on genetic and phenotypic variation in *Salmonella* has been derived from studies of *S. typhimurium* strain LT2 and a small number of other laboratory strains (Neidhardt *et al.*, 1987). But in view of increasing interest on the part of microbiologists, molecular biologists and population geneticists in extending analyses of variation to natural populations, we

Abbreviations: ECOR, *Escherichia coli* reference collection; ET, electrophoretic type; SARA, *Salmonella* reference collection A.

Table 1. *Properties of the 72 strains of the Salmonella reference collection A (SARA), representing 48 electrophoretic types (ETs) of the five serovars of the S. typhimurium complex*

SARA no.	RKS no.	ET	Original number*	Source	Locality	Biotype†	Date
Salmonella typhimurium (1,4,[5],12:i:1,2)‡							
1	284	Tm 1	INSP 24	Human	Mexico	1 f	–
2	4939	Tm 1	LT2	–	Lab. strain§	–	–
3	145	Tm 1	NVSL 7095	Horse	Rhode Island	11 di	1987
4	183	Tm 1	NVSL 5820	Rabbit	Indiana	25 e	1986
5	810	Tm 1	IVB 232	–	Mongolia	9 fi	–
6	345	Tm 2	CDC B1213	Human	Ohio	17 a	–
7	821	Tm 3	IVB 665/81	–	Norway	1 a	–
8	811	Tm 5	IVB 5560	–	Finland	1 a	–
9	203	Tm 7	NVSL 2816	Parrot	California	1 a	1987
10	154	Tm 9	NVSL 6814	Opossum	California	1 b	1987
11	829	Tm 10	IVB 276/25	–	Thailand	2 a	–
12	147	Tm 11	NVSL 6993	Horse	Louisiana	32 begi	1987
13	837	Tm 12	IVB 1430	–	France	29 b	–
14	842	Tm 13	IVB 75/67	–	Panama	1 a	–
15	149	Tm 14	NVSL 6968	Dog	Texas	17 a	1987
16	350	Tm 15	CDC B1236	Human	North Carolina	3 d	–
17	1164	Tm 16	IVB 48/81	–	Yugoslavia	3 d	–
18	151	Tm 17	NVSL 6938	Horse	Iowa	3 d	1987
19	93	Tm 21	INSP 85	Human	Mexico	1 f	–
20	839	Tm 22	IVB 1544	–	France	1 a	–
21	4535	Tm 23	USFW 318	Heron	Oregon	–	–
Salmonella saintpaul (1,4,[5],12:e,h:1,2)							
22	1688	Sp 1	CDC B1605	Human	Massachusetts	–	–
23	1689	Sp 2	CDC B1722	Human	Pennsylvania	–	–
24	1690	Sp 3	CDC B2076	Human	Texas	–	–
25	1380	Sp 3	IVB 516	–	France	–	–
26	3748	Sp 3	IP 67/88	Human	France	–	1988
27	3755	Sp 3	IP 78/88	Human	France	–	1988
28	3763	Sp 3	IP 86/88	Human	France	–	1988
29	1686	Sp 4	CDC B1400	Human	Florida	–	–
Salmonella heidelberg (1,4,[5],12:r:1,2)							
30	539	He 1	NVSL 7039	Chicken	Pennsylvania	9 i	1987
31	560	He 1	NVSL 5876	Swine	Maryland	9 i	1987
32	562	He 1	NVSL 5145	Dog	Texas	9 i	1986
33	576	He 1	INSP 94	Human	Mexico	9 gi	–
34	1364	He 1	IVB 7135/1990	–	Israel	9 i	–
35	1389	He 2	IVB 126/82	–	Brazil	9 i	–
36	1391	He 3	IVB 588/24	–	Thailand	1 a	–
37	543	He 4	NVSL 5208	Turkey	Colorado	9 i	1987
38	540	He 5	NVSL 4960	Turkey	Arizona	9 i	1987
39	646	He 7	CDC B2487	Human	North Carolina	9 gi	–
40	1347	He 8	IVB 218/82	–	USA	9 gi	–
S. paratyphi B (including variety *java*) (1,4,[5],12:b:[1,2])							
41	3222	Pb 1	DMS 155/76	Human	France	3 gh	1976
42	3279	Pb 1	DMS 724/74	Human	Scotland	27 bg	1974
43	3305	Pb 1	DMS 220/82	Human	Africa	11 ghi	1982
44	3265	Pb 1	DMS 2434	Human	Middle East	19 gh	<1965
45	3596	Pb 1	IP 7/88	Cow	France	–	1988
46	3294	Pb 1a	DMS 3254/7/81	Human	Europe	11 bgi	1981
47	3249	Pb 2	DMS 3205/83	Sewage	Scotland	9 a	1983
48	3237	Pb 2a	DMS 843/82	Human	Scotland	1 dh	1982
49	3267	Pb 2b	DMS 2442	Sewage	UK	1 b	<1965
50	3202	Pb 3	DMS 106/76	Food	Middle East	2 a	1976
51	3193	Pb 3	DMS 53/76	Human	France	2 bg	1976
52	3614	Pb 3	IP 87/87	Cow	France	–	1987
53	3605	Pb 3	IP 16/88	Human	France	–	1988
54	3597	Pb 3	IP 8/88	Human	France	–	1988
55	3211	Pb 3a	DMS 47/81	Human	France	2 bg	1981
56	3201	Pb 4	DMS 83/76	Human	France	13 i	1976
57	3274	Pb 5	DMS 2471	Water	UK	11 i	<1965
58	3218	Pb 5a	DMS 59/81	Human	France	26 i	1981

Table 1—*continued*

SARA no.	RKS no.	ET	Original number*	Source	Locality	Biotype†	Date
59	3219	Pb 5b	DMS 61/81	Human	France	10 i	1981
60	3192	Pb 5c	DMS 52/76	Food	France	9 bi	1976
61	3277	Pb 6	DMS 203/74	Water	Scotland	1 a	1974
62	3215	Pb 7	DMS 53/81	Human	Africa	1 g	1981
Salmonella muenchen (6,8 : d : 1,2)							
63	4283	Mu 1	IP 6/88	Human	France	–	1988
64	4129	Mu 1	NVSL 519	Cow	Kentucky	–	1986
65	4135	Mu 1	NVSL 2817	Chicken	Florida	–	1987
66	4277	Mu 1	CDC B2026	Human	Massachusetts	–	–
67	4317	Mu 1	INSP 46	Human	Mexico	–	–
68	4292	Mu 1a	IP 15/88	Human	France	–	1988
69	4288	Mu 2	IP 11/88	Human	France	–	1988
70	4300	Mu 3	IP 25/88	Human	France	–	1988
71	4272	Mu 4	CDC B1293	Human	North Carolina	–	–
72	4306	Mu 4a	IP 31/88	Human	France	–	1988

* Abbreviations: CDC, Centers for Disease Control (Atlanta, Georgia, USA); DMS, University of Dundee Medical School (Ruth Barker collection) (Dundee, UK); INSP, Instituto Nacional de Salud Publica (Cuernavaca, Mexico); IP, Institut Pasteur (Paris, France); IVB, Institut fur Veterinarmedizin des Bundesgesundheitsamtes (Berlin, FRG); NVSL, National Veterinary Services Laboratories (Ames, Iowa, USA); USFW, United States Fish and Wildlife collection (Ruth Duncan collection) (Madison, Wisconsin, USA).

† For explanation of biotype designations, see Duguid *et al.* (1975) and Barker *et al.* (1988).

‡ The serotype formula, indicating the O : H1 : H2 antigenic factors, is shown in parentheses for each serovar. See Le Minor (1984) for detailed explanation.

§ Obtained from Dr C. F. Higgins.

have assembled several collections of isolates of *Salmonella* that have been characterized genotypically by multilocus enzyme electrophoresis and are representative of the total range of genomic variation in natural populations. A comparable standard reference collection of isolates of *Escherichia coli* (ECOR) (Ochman & Selander, 1984; Selander *et al.*, 1987) has been widely used for several years in studies of population genetics and molecular evolution (e.g. DuBose *et al.*, 1988; Hall *et al.*, 1989; Hartl & Sawyer, 1988; Stoltzfus *et al.*, 1988).

We here report the availability of the *Salmonella* reference collection A (SARA), which consists of 72 strains (laboratory strain LT2 and 71 strains from natural populations) selected to represent the total span of genomic diversity among clones of the *S. typhimurium* complex of subgroup I (Selander *et al.*, 1990*a*), which includes the serovars *S. typhimurium*, *S. saintpaul*, *S. heidelberg*, *S. paratyphi B* (including variety *java*) and *S. muenchen* (Table 1). As shown in Table 1, the H1 antigens of the five serovars are distinctive, but all serovars share H2 antigenic factors 1 and 2. *S. typhimurium*, *S. saintpaul*, *S. heidelberg* and *S. paratyphi B* (including variety *java*) express O antigenic factors 1, 4, 5 and 12 and are placed in O antigen group B on the basis of possession of antigenic factor 4, but *S. muenchen* has O antigenic factors 6 and 8 and is assigned to O antigen group C_2 (Le Minor, 1984).

Results and Discussion

In an analysis of electrophoretically detectable allelic variation in 24 chromosomal genes encoding metabolic enzymes, we identified 48 electrophoretic types (ETs), marking clones, among a total of 916 isolates of these five serovars examined. The enzymes assayed and the allele profiles of the 48 ETs are indicated in Table 2. Genetic variation and evolutionary relationships among strains of *S. paratyphi B* and *S. java* in relation to biotype characteristics (Barker *et al.*, 1988) have been analysed in detail by Selander *et al.* (1990*a*). Nucleotide sequencing of the H1 flagellin genes of multiple strains of the serovars of the *S. typhimurium* complex and in strains of other serovars has indicated that horizontal gene transfer and recombination is an important evolutionary mechanism generating serotypic variation; consequently, variation in the phase 1 flagellins is not phylogenetically informative (Smith & Selander, 1990; Smith *et al.*, 1990).

SARA includes one isolate of each of these 48 ETs and four or five additional isolates of the predominant clone of each serovar: Tm 1, Sp 3, He 1, Pb 1, Pb 3 and Mu 1 (Table 1). Tm 1 is the ET of laboratory strain LT2, which is represented by SARA strain 2. In selecting multiple isolates of the predominant clones, an effort was made to maximize the representation of biotypic, geographical and host diversity.

Table 2. *Allele profiles for 20 polymorphic enzyme loci in the 48 electrophoretic types of the Salmonella reference collection A (SARA)*

		Allele at indicated enzyme locus‡:																			
ET*	*N*†	IDH	ACO	CAK	AP1	AP2	6PG	PGI	NSP	CAT	HEX	LG1	PLP	MDH	G6P	M1P	GDH	PGM	GLU	GOT	SKD
Tm 1	258	3	3	3	3	10	2	3	7	4	3	3	3	3	3	3	3	3	3	3	0·5
Tm 2	5	2	3	3	3	10	2	3	7	4	3	3	3	3	3	3	3	3	3	3	0·5
Tm 3	2	2	3	3	3	7	2	3	7	4	3	3	3	3	3	3	3	2	3	3	0·5
Tm 5	1	3	3	3	3	7	2	3	7	4	3	3	3	3	2·5	3	3	2	3	3	0·5
Tm 7	2	3	3	3	3	10	2	3	7	4	4	3	3	3	4	3	3	3	3	3	0·5
Tm 9	17	3	3	3	3	5	2	3	7	4	3	3	3	3	3	3	3	3	3	3	0·5
Tm 10	9	3	3	3	3	10	2	3	7	5	3	3	3	3	3	3	3	3	3	3	0·5
Tm 11	5	3	3	3	3	10	2	3	7·5	0	3	3	3	3	3	3	3	3	3	3	0·5
Tm 12	27	3	3	3	3	10	2	3	0	0	4	3	3	3	3	3	3	3	3	3	0·5
Tm 13	3	3	3	3	3	10	2	3	7	4	3	3	3	3	2·5	3	3	3	3	3	0·5
Tm 14	1	3	3	3	3	10	2	3	7	4	3	3	3	2	3	3	3	3	3	3	0·5
Tm 15	1	3	3	3	3	10	2	3	7	4	3	3	3	1	3	3	3	3	3	3	0·5
Tm 16	1	3	3	3	3	10	2	3	7	4	3	3	3	2·5	3	3	3	3	3	3	0·5
Tm 17	1	3	3	3	3	10	2	3	7	4	3	3	3	3	3	3	3	3	2	3	0·5
Tm 21	1	3	3	3	3	10	2	1	7	4	3	3	3	3	3	3	3	3	3	3	0·5
Tm 22	4	3	3	3	3	10	2	4	7	4	3	3	3	3	3	3	3	3	3	3	0·5
Tm 23	2	3	3	3	3	10	2	3	0	4	3·5	3	3	3	3	3	3	2	3	3	0·5
Sp 1	5	3	3	3	3	7	2	3	7	4	3	3	3	3	3	3	3	3	3	3	0·5
Sp 2	1	2	3	3	3	8·2	2	3	7	4	3	3	3	3	3	3	3	3	3	3	0·5
Sp 3	27	3	3	3	3	8·2	2	3	7	4	3	3	3	3	3	3	3	3	3	3	0·5
Sp 4	1	2	3	3	5	10·5	5·5	3	5	3	3	5	3	3	3	3	3	3	3	3	1
He 1	173	3	3	3	3	10	2	3	7	2·5	3	3	3	3	3	3	3	3	3	3	1
He 2	1	3	3	3	3	10	2	4	7	2·5	3	3	3	3	3	3	3	3	3	3	1
He 3	3	2	3	3	3	10	2	4	7	2·5	3	3	3	3	3	3	3	3	3	3	6
He 4	15	2	3	3	3	10	2	3	7	2·5	3	3	3	3	3	3	3	3	3	3	1
He 5	6	3	3	3	3	10	2	3	7	2·5	3	3	3	3	3	3	3	3	3	6	1
He 7	1	3	3	2	3	10	2	3	7	2·5	3	3	3	3	3	3	3	3	3	3	1
He 8	1	3	3	3	3	10	2	3	7	2·5	3	3	3	3	3·5	3	3	3	3	3	1
Pb 1	139	3	5	3	3	14	2	3	3	3	3	3	3	3	3	3	3	3	3	3	0·5
Pb 1a	1	3	5	3	3	14	4	3	3	3	3	3	3	3	3	3	3	3	3	3	0·5
Pb 2	11	2	5	3	3	14	2	3	3	3	3	3	3	3	3	3	3	3	3	3	0·5
Pb 2a	1	2	3	3	3	14	2	3	3	3	3	3	3	3	3	3	3	3	3	3	0·5
Pb 2b	1	2	5	3	3	14	2	3	5	3	3	3	3	3	3	3	3	3	3	3	0·5
Pb 3	46	2	5	3	3	0	2	3	3	2	3	3	3	3	3	3	3	3	3	3	0·5
Pb 3a	2	3	5	3	3	0	2	3	3	2	3	3	3	3	3	3	3	3	3	3	0·5
Pb 4	62	3	3	3	5	14	2	3	3	2	3	3	3	3	3	5	3	3	3	3	0·5
Pb 5	1	3	3	3	3	10	2	3	3	1·5	3	3	3	3	3	3	3	3	3	3	0·5
Pb 5a	1	3	3	3	3	10	2	3	3	1·5	3	3	2	3	3	3	3	3	3	3	0·5
Pb 5b	1	2	3	3	3	10	2	3	3	1·5	3	3	3	3	3	3	3	3	3	3	0·5
Pb 5c	1	2	3	3	3	0	2	3	3	1·5	3	3	3	3	3	3	3	3	3	3	0·5
Pb 6	1	3	3	3	3	10·5	3	3	5	3	3	3	3	3	3	3	3	2	3	3	0·5
Pb 7	1	3	3	3	3	14	3	3	5	2	3	3	3	3	3	3	3	2	3	3	1
Mu 1	46	3	5	3	3	14	2	3	7	3	3	3	3	3	3	3	3	3	3	3	0·5
Mu 1a	1	2	5	3	3	14	2	3	7	3	3	3	3	3	3	3	3	3	3	3	0·5
Mu 2	19	3	5	3	5	14	2	3	5	4	3	3	3	3	3	3	2	2	3	3	1
Mu 3	4	2	3	3	5	14	3	3	3	3·5	3	3	3	3	3	3	3	3	3	3	0·5
Mu 4	2	3	3	3	3	7	3	3	5	3	3	2	3	2	3	3	3	3	3	3	0·5
Mu 4a	1	3	3	3	3	7	3	3	5	3	3	2	3	2	3	4	3	3	3	5	0·5

* Serovar symbols: Tm, *S. typhimurium*; Sp, *S. saintpaul*; He, *S. heidelberg*; Pb, *S. paratyphi B*; Mu, *S. muenchen*.

† Number of isolates examined.

‡ Enzyme locus symbols: IDH, isocitrate dehydrogenase; ACO, aconitase; CAK, carbamylate kinase; AP1, acid phosphatase-1; AP2, acid phosphatase-2; 6PG, 6-phosphogluconate dehydrogenase; PGI, phosphoglucose isomerase; NSP, nucleoside phosphorylase; CAT, catalase; HEX, hexokinakse; LG1, leucylglycyl-glycine peptidase-1; PLP, phenylalanyl-leucine peptidase; MDH, malate dehydrogenase; G6P, glucose-6-phosphate dehydrogenase; M1P, mannitol-1-phosphate dehydrogenase; GDH, glucose dehydrogenase; PGM, phosphoglucomutase; GLU, glutamate dehydrogenase; GOT, glutamic oxaloacetic transminase; and SKD, shikimate dehydrogenase. Four additional enzyme loci (not shown) were each monomorphic: ADK, adenylate kinase; LG2, leucylglycyl-glycine peptidase-2; IPO, indophenol oxidase; and MPI, mannose phosphate isomerase. For each enzyme locus, alleles are numbered in order of decreasing electrophoretic mobility of their corresponding electromorphs; 0 indicates a null allele.

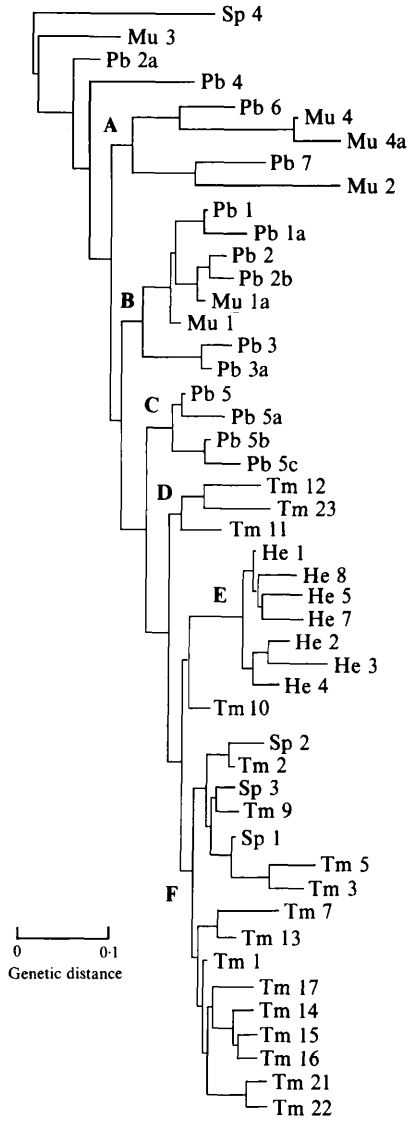

Fig. 1. Phylogenetic tree showing relationships among the 48 ETs represented by the 72 SARA strains. See text for explanation. Abbreviations of serovars: Tm, *S. typhimurium*; Sp, *S. saintpaul*; He, *S. heidelberg*; Pb, *S. paratyphi B* (including *S. java*); Mu, *S. muenchen*.

Evolutionary genetic relationships among the 48 ETs are indicated in the phylogenetic tree in Fig. 1, which was generated by the neighbour-joining algorithm (Saitou & Nei, 1987) from a matrix of Nei's standard genetic distance (Nei, 1987), based on pairwise comparison of ETs at 24 enzyme loci. This algorithm has been shown in simulation studies to recover the correct tree topology when evolutionary rates vary among lineages (Saitou & Nei, 1987). We rooted the tree at the node connecting the longest branch (Sp 4). The distance between two ETs in the tree is the sum of the lengths of the horizontal branches in the path connecting them and is interpreted as the number of electrophoretically detected codon differences per locus.

In terms of total genotypic diversity, the SARA strains are roughly equivalent to strains of group A of the ECOR collection (Selander *et al.*, 1987).

For convenience of reference, six clusters of ETs in Fig. 1 are designated by letter. Thirteen of the 17 ETs of *S. typhimurium* fall in cluster F, but Tm 11, Tm 12 and Tm 23 are in cluster D, and Tm 10 stands apart near cluster E, which consists of the six ETs of *S. heidelberg*. Except for Sp 4, which is a highly divergent genotype, the ETs of *S. saintpaul* do not cluster separately from those of *S. typhimurium*. *S. paratyphi B* (including variety *java*) is markedly heterogeneous, with ETs in clusters A, B and C, and Pb 2a and Pb 4 representing separate lineages; cluster C includes all the monophasic strains of this serovar. Five of the six ETs of *S. muenchen* fall in clusters A and B, with ETs of *S. paratyphi B*, but Mu 3 represents a divergent lineage.

Two additional reference collections of *Salmonella* strains are currently being prepared, one representative of genomic variation among strains of many of the common serovars of subgenus I, and another consisting of isolates of serovars of all the subgenera of *Salmonella*.

Information on the availability of the *Salmonella* reference collections for research may be obtained from R.K.S. A set of SARA has been deposited in the *Salmonella* Genetic Stock Centre (contact Dr K. E. Sanderson, Department of Biological Sciences, University of Calgary, 200 University Drive NW, Calgary, Alberta, Canada T2N 1N4).

Research on bacterial population genetics was supported by grant AI22144 (to R.K.S.) and grant AI24566 (to T.S.W.) from the National Institutes of Health. N.H.S. was supported by a postdoctoral fellowship from the Alfred P. Sloan Foundation.

References

BARKER, R. M., KEARNEY, G. M., NICHOLSON, P., BLAIR, A. L., PORTER, R. C. & CRICHTON, P. B. (1988). Types of *Salmonella paratyphi B* and their phylogenetic significance. *Journal of Medical Microbiology* **26**, 285–293.

BELTRAN, P., MUSSER, J. M., HELMUTH, R., FARMER, J. J., III, FRERICHS, W. M., WACHSMUTH, I. K., FERRIS, K., MCWHORTER, A. C., WELLS, J. G., CRAVIOTO, A. & SELANDER, R. K. (1988). Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S. choleraesuis*, *S. derby*, *S. dublin*, *S. enteritidis*, *S. heidelberg*, *S. infantis*, *S. newport*, and *S. typhimurium*. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7753–7757.

CROSA, J. H., BRENNER, D. J., EWING, W. H. & FALKOW, S. (1973). Molecular relationships among the salmonellae. *Journal of Bacteriology* **115**, 307–315.

DUBOSE, R. F., DYKHUIZEN, D. E. & HARTL, D. L. (1988). Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7036–7040.

DUGUID, J. P., ANDERSON, E. S., ALFREDSSON, G. A., BARKER, R. M. & OLD, D. C. (1975). A new biotyping scheme for *Salmonella typhimurium* and its phylogenetic significance. *Journal of Medical Microbiology* **8**, 149–166.

EWING, W. H. (1986). *Edwards and Ewing's Identification of Enterobacteriaceae*, 4th edn. New York: Elsevier.

FRANKEL, G., NEWTON, S. M. C., SCHOOLNIK, G. K. & STOCKER, B. A. D. (1989). Intragenic recombination in a flagellin gene: characterization of the *H1 :j* gene of *Salmonella typhi*. *EMBO Journal* **8**, 3149–3152.

HALL, B. C., PARKER, L. L., BETTS, P. W., DUBOSE, R. F., SAWYER, S. A. & HARTL, D. L. (1989). IS*103*, a new insertion element in *Escherichia coli*: characterization and distribution in natural populations. *Genetics* **121**, 423–431.

HARTL, D. L. & SAWYER, S. A. (1988). Why do unrelated insertion sequences occur together in the genome of *Escherichia coli*? *Genetics* **118**, 537–544.

LE MINOR, L. (1984). *Salmonella* Lignières 1900, 389. In *Bergey's Manual of Systematic Bacteriology*, vol. 1, pp. 427–458. Edited by N. R. Krieg & J. G. Holt. Baltimore: Williams & Wilkins.

LE MINOR, L., POPOFF, M. Y., LAURENT, B. & HERMANT, D. (1986). Individualisation d'une septième sous-espece de *Salmonella*: *S. choleraesuis* subsp. *indica* subsp. nov. *Annales de l'Institut Pasteur/ Microbiologie* **137B**, 211–217.

NEI, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

NEIDHARDT, F. C., INGRAHAM, J. L., LOW, K. B., MAGASANIK, B., SCHAECHTER, M. & UMBARGER, H. E. (editors) (1987). *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*, vols 1 and 2. Washington, DC: American Society for Microbiology.

OCHMAN, H. & SELANDER, R. K. (1984). Standard reference strains of *Escherichia coli* from natural populations. *Journal of Bacteriology* **157**, 690–693.

REEVES, M. W., EVINS, G. M., HEIBA, A. A., PLIKAYTIS, B. D. & FARMER, J. J., III (1989). Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. *Journal of Clinical Microbiology* **27**, 313–320.

SAITOU, N. & NEI, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.

SELANDER, R. K., CAUGANT, D. A., OCHMAN, H., MUSSER, J. M., GILMOUR, M. N. & WHITTAM, T. S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Biology* **51**, 873–884.

SELANDER, R. K., CAUGANT, D. A. & WHITTAM, T. S. (1987). Genetic structure and variation in natural populations of *Escherichia coli*. In *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*, vol. 2, pp. 1625–1648. Edited by F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter & H. E. Umbarger. Washington, DC: American Society for Microbiology.

SELANDER, R. K., BELTRAN, P., SMITH, N. H., BARKER, R. M., CRICHTON, P. B., OLD, D. C., MUSSER, J. M. & WHITTAM, T. S. (1990*a*). Genetic population structure, clonal phylogeny, and pathogenicity of *Salmonella paratyphi* B. *Infection and Immunity* **58**, 1891–1901.

SELANDER, R. K., BELTRAN, P., SMITH, N. H., HELMUTH, R., RUBIN, F. A., KOPECKO, D. J., FERRIS, K., TALL, B. D., CRAVIOTO, A. & MUSSER, J. M. (1990*b*). Evolutionary genetic relationships of clones of *Salmonella* that cause human typhoid and other enteric fevers. *Infection and Immunity* **58**, 2262–2275.

SMITH, N. H. & SELANDER, R. K. (1990). Sequence invariance of the antigen-coding central region of the phase 1 flagellar filament gene (*fliC*) among strains of *Salmonella typhimurium*. *Journal of Bacteriology* **172**, 603–609.

SMITH, N. H., BELTRAN, P. & SELANDER, R. K. (1990). Recombination of *Salmonella* phase 1 flagellin genes generates new serovars. *Journal of Bacteriology* **172**, 2209–2216.

STOLTZFUS, A., LESLIE, J. F. & MILKMAN, R. (1988). Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between *trp* and *tonB*. *Genetics* **120**, 345–358.