J Mol Evol (1994) 39:129–133

Journal of Molecular Evolution

Genetic Variation in IncI1-ColIb Plasmids

Francisco José Ayala, Dan E. Krane,* Daniel L. Hartl

Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Received: 17 December 1993 / Revised and accepted: 7 February 1994

Abstract. Nucleotide sequences of portions of three plasmid genes (*cib*, *cir*, and *abi*) present in IncI1-ColIb colicin plasmids obtained from strains of *Salmonella typhimurium* isolated in either 1974 (Barker strains) or between 1935 and 1941 (Murray strains) were examined along with sequences of the chromosomal gene for 6-phosphogluconate dehydrogenase (*gnd*). Our principal findings were: (1) The plasmid genes were virtually identical to those in IncI1-ColIb plasmids from *E. coli*, suggesting that *Salmonella* and *E. coli* share overlapping pools of these plasmids. (2) The plasmid genes were much less polymorphic than *gnd* or any other known chromosomal gene from *Salmonella*, further suggesting horizontal transfer with rapid transmission and turnover. (3) No characteristic differences were found in either the plasmid genes or the chromosomal gene between the 1974 isolates and the Murray strains, indicating that these plasmids have been stable for at least several decades. (4) There was an excess of amino-acid replacement polymorphisms, relative to synonymous polymorphisms, in the plasmid genes, which is consistent with the hypothesis of diversifying selection among colicin-producing plasmid families. (5) The *abi* (abortive infection) gene present in each of the plasmids contained two single-nucleotide insertions relative to the published sequence. These result in a putative *abi* protein of 114 amino acids instead of 89.

Key words: *Salmonella typhimurium* — Plasmid — Colicin Ib — Abortive infection gene (*abi*) — Diversifying selection

**Present address:* Department of Biological Sciences, Wright State University, Dayton, OH 45435, USA
Correspondence to: D.L. Hartl

Introduction

The colicin plasmids comprise a diverse group of plasmids containing genes that code for proteins toxic to bacteria as well as genes conferring specific immunity to the host which produces those toxins (Konisky 1982). The distribution and abundance of colicin plasmids have been characterized in both *S. typhimurium* (Barker 1980) and *E. coli* (Riley and Gordon 1992). Approximately 12% of natural isolates of *S. typhimurium* (Barker 1980) and 35% of a reference collection of natural isolates of *E. coli* (Riley and Gordon 1992) produce colicins; among the most frequent colicin classes are ColIa and ColIb. Colicins specifying related bacterial toxins are identified by the presence of cross-immunity. At present, approximately 20 distinct types of colicins have been identified (Pugsley 1985; Riley 1993a).

Colicins that are related physiologically are encoded in genes that are similar in nucleotide sequence. Comparisons of pairs of related colicin operons in three different colicin families (colicins E3 vs E6, E2 vs E9, and Ia vs Ib) yielded nucleotide-sequence identities ranging from 85 to 96% (Riley 1993b). In comparing the rates of synonymous, nonsynonymous, and intergenic nucleotide substitution in these three pairs of closely related colicins, Riley (1993b) observed an excess of nonsynonymous substitutions in the immunity gene and in the immunity-binding domain of the colicin gene. The excess of amino acid replacements, relative to synonymous nucleotide substitutions, was hypothesized to result from selection for diversity in the immunity protein and in the immunity-binding region of the colicin (Riley 1993a,b).

The present analysis was carried out, in part, to determine whether independent plasmid isolates coding for a particular type of colicin also show an excess of

amino-acid replacement polymorphisms, relative to synonymous polymorphisms. The study also addresses the issue of nucleotide diversity among genes present in plasmids. Most previous work on plasmid polymorphisms has been biased toward finding diversity because the studies have focused on isolates of plasmids that differ in one or more phenotypic attributes. For example, regions of several genes suspected to be responsible for the diversity of plasmid phenotypes have been sequenced and compared from a wide range of F-related plasmids (Finlay et al. 1986a–c). The level of DNA sequence identity in these regions ranged from 68 to 100%. These data contrast with studies of several types of transposable insertion sequences (IS) indicating little or no genetic variation within species (Lawrence et al. 1992; Bisercic and Ochman 1993), a result attributed to rapid turnover and plasmid-mediated horizontal transmission of IS elements (Lawrence et al. 1992). Hence, a reduced level of genetic variation in plasmid genes, relative to chromosomal genes, might also be expected.

We have examined nucleotide-sequence variation in three genes in conjugative plasmids of the incompatibility group IncI1 containing the ColIb colicin operon. The plasmids were present in strains of *Salmonella typhimurium* obtained as natural isolates either from Scotland in 1974 (Barker 1980; Barker et al. 1980) or from geographically diverse locations and collected by E.D.G. Murray in 1935–1941 (Hughes and Datta 1983; Jones and Stanley 1992). The latter strains represent those of the preantibiotic era and so provide an opportunity to detect evolutionary changes in plasmid genes over a period of approximately 40 years. The plasmids were chosen solely because they were IncI1-ColIb and so have not been preselected for phenotypic or genetic diversity. The loci analyzed were *cib, cir,* and *abi.* The *cib* and *cir* loci are tightly linked and code for the colicin Ib toxin and immunity proteins, respectively (Mankovich et al. 1986); the *abi* locus codes for the abortive-infection protein that is responsible for abortive infection by the bacteriophages BF23 or T5 (Gupta and McCorquodale 1988). We also examined sequences of the chromosomal gene, *gnd,* coding for 6-phosphogluconate dehydrogenase, from the same bacterial strains.

Materials and Methods

Bacterial Strains and Plasmids. S. typhimurium strains M102, M140, M303, M496, M498, and M501 from the Murray collection were obtained from Dr. John Stanley and the National Collection of Type Cultures, London. M140 was added to the Murray collection in 1935; the other strains were added in 1941. The Murray strains are sometimes referred to as preantibiotic era (PAE) strains (Hughes and Datta 1983; Jones and Stanley 1992). The IncI1-ColIb plasmids in the strains were identified by Jones and Stanley (1992). More recent *S. typhimurium* strains 205/74 (biotype 17g) and 375/74 (biotype 9f) were collected in 1974 from water in Glasgow and from a cattle farm in Wigtownshire, respectively, and characterized by Barker et al. (1980). These strains were provided by Dr. Jordan Konisky.

DNA Amplification. Oligonucleotide primers for the polymerase chain reaction (PCR) and for DNA sequencing were designed from published sequences of *cib* and *cir* (Mankovich et al. 1986), *abi* (Gupta and McCorquodale 1988), and *gnd* (Reeves and Stevenson 1989) using the OLIGO 4.0 program for Macintosh (Rhoads and Rychlik 1989). The following PCR primers were used: for the region containing parts of the *cib* and *cir* loci, 5′-AAGCAGTTAATTCATTTGTTT-3′ and 5′-ACAGGAGCACCTCCGTTCCAT-3′; for the *abi* locus, 5′-ACAGGAGGATGCGAATGAACA-3′ and 5′-GCGTTGCGGCTACTTACTTTA-3′; and for the *gnd* locus, 5′-GAGTATGTAATGTCCAAGCAA-3′ and 5′-CTCAGCTCGCCGTTATTCCAC-3′.

The optimal annealing temperature of each primer pair was estimated using the OLIGO-4.0 program. Following amplification (Saiki et al. 1988), the PCR products were purified by diluting the 30-μl reaction volume to a final concentration of 100 μl with distilled and dionized H_2O, followed by phenol extraction, precipitation in the presence of 2.5 M sodium acetate and one volume of absolute ethanol, and washing with 70% ethanol (Sambrook et al. 1989).

DNA Sequencing. Sequencing of PCR products was performed with an Applied Biosystems model 373A automated DNA sequencing system using the *Taq* DyeDeoxy terminator cycle-sequencing kit using reaction conditions and other methods recommended by the manufacturer. Organic contaminants were removed by means of the phenol-chloroform extraction protocol provided with the sequencing kit. Both strands of all regions were sequenced in their entirety. Sequences were aligned and analyzed by hand.

Nucleotide sequences from the following regions (numbered according to the published sequences) were determined: for the *cib* locus, sites 1399–1519, 1640–1716, 2244–2520, 2568–2598, and 2601–2621 (total 527 bp); for the *cir* locus, sites 3057–3163 and 3201–3260 (total 167 bp); for the *abi* locus, sites 1343–1508 (total 166 bp); and for the *gnd* locus, sites 235–385 (total 151 bp). All strains were sequenced in these regions except strain 205/74, in which the *abi* and *gnd* regions were not analyzed. Nucleotide sequence was also determined for the *abi* locus between sites 1509 and 1617 in strain M140 and sites 1509–1596 in strain M496.

Results

Nucleotide-Sequence Variation

Analysis of nucleotide variation in the IncI1-ColIb plasmids was based on approximately 7 kb of sequence from three plasmid genes in IncI1 plasmids present in two strains isolated in 1974 as well as in IncI1 plasmids present in six Murray strains isolated in 1935–1941. The sequenced regions included 527 nucleotides of *cib* (the colicin Ib structural gene), 167 nucleotides of *cir* (the colicin Ib immunity gene), and 166 nucleotides of *abi* (abortive infection gene). For comparison, a region of approximately 150 nucleotides was determined for the chromosomal gene *gnd,* coding for 6-phosphogluconate dehydrogenase, in the same strains.

The polymorphic nucleotides in each of the three plasmid loci, as well as in the *gnd* locus, are summarized in Fig. 1, in which the nucleotides are classified according to position, type of nucleotide change, and ef-

Locus	Site		Strain 2	3	M 1	M 1	M 3	M 4	M 4	M 5	Type
			0	7	0	4	0	9	9	0	
cib	1404	C	.	.	.	.	.	.	.	A	Rep
	1655	G	.	.	A	.	.	.	.	.	Syn
	2420	T	.	.	.	.	C	.	.	.	Syn
	2428	T	.	.	.	.	.	C	.	.	Rep
	2432	T	.	C	.	.	C	C	.	C	Syn
	2455	A	G	.	.	.	.	.	.	.	Rep
	2484	G	.	.	.	T	.	.	.	.	Rep
	2514	G	.	.	.	A	.	.	.	.	Rep
	2570	A	T	T	T	T	T	T	T	T	Rep
	2597	A	.	.	.	C	.	.	.	.	Rep
cir	3063	A	.	.	.	.	.	G	.	.	Rep
	3084	T	C	.	.	.	.	.	.	.	Rep
	3103	A	.	.	C	.	.	.	C	C	Rep
	3126	A	C	.	.	.	.	.	.	.	Rep
	3139	T	C	.	.	.	.	.	.	.	Syn
abi	>1354	-		T	T	T	T	T	T	T	-
	1366	T		.	.	A	.	.	.	.	Rep
	1442	T		G	G	G	.	.	.	.	Rep
gnd	237	T		.	.	.	.	.	.	C	Syn
	240	T[b]		.	.	.	.	.	.	C	Syn
	252	C		.	.	.	.	.	T	.	Syn
	261	G		.	.	C	.	.	C	.	Syn
	276	A		.	.	.	T	T	.	.	Rep
	282	C		T	.	.	.	T	.	.	Syn
	300	C		T	T	.	T	.	.	.	Syn
	330	T		C	.	.	.	.	.	.	Syn
	353	A		.	.	.	.	.	.	G	Rep
	371	G		.	.	.	.	.	C	.	Rep
	378	T		G	.	.	.	.	.	.	Syn
	381	T		.	G	G	.	.	.	.	Syn

Fig. 1. Site refers to the numbering of the nucleotides in the published sequences of Mankovich et al. (1986) for *cib* and *cir*, Gupta and McCorquodale (1988) for *abi*, and Nasoff et al. (1984) for *gnd*. Nucleotides identical to the published sequences are indicated by *dots*. Type (replacement of synonymous) refers to the effect on the amino acid sequence. The symbol *>1354* in the *abi* gene indicates a single nucleotide insertion following site 1354.

fect on amino acid sequence (replacement difference vs synonymous difference). The proportion of polymorphic nucleotides (K) in each gene is summarized in the second column in Table 1. All three plasmid genes are polymorphic: taken together, the three plasmid genes have $K = 0.021 \pm 0.005$. The third column in Table 1 summarizes the polymorphism data in terms of the nucleotide diversity, π, which is the average number of nucleotide differences per site between two randomly chosen alleles. The nucleotide diversity has the convenient property that it is independent of both sample size and total number of sites (Nei 1987, p 256). The weighted average π for the three plasmid loci (weighted by the reciprocal of the estimated variance) is 0.0071.

There is a difference in character of the polymorphisms, as well as in the level, between the plasmid and chromosomal genes. In particular, the proportion of replacement polymorphisms is significantly greater in the colicin genes *cib* and *cir* than in the chromosomal gene *gnd* (Fig. 1; $\chi^2 = 6.2$, $P < 0.02$). This difference suggests either relaxed selective constraints on the amino acid sequence of the colicin genes or positive selection for amino acid diversity.

Table 1. Genetic variation in IncI1-ColIb plasmid genes and *gnd*

Locus	K[a] ± SE	π[b] ± SE
cib	0.019 ± 0.006	0.005 ± 0.039
cir	0.030 ± 0.013	0.009 ± 0.056
abi	0.018 ± 0.010	0.005 ± 0.049
gnd	0.080 ± 0.022	0.030 ± 0.105

[a] Proportion of polymorphic nucleotides in the sample ± binomial standard error

[b] Average number of nucleotide differences per site between randomly chosen pairs of alleles ± standard error (Nei 1987, p 276)

Plasmids from 1935–1941 Isolates vs 1974 Isolates

There are some pairs of duplicates in the Murray strains (Hughes and Datta 1983), and so a serious consideration a priori was whether the strains added to the collection at the same time and with similar numbers might be duplicates of the same isolate. This consideration would especially include strains M496, M498, and M501, which were all added to the collection in 1941. However, the DNA sequencing results clearly indicate that the strains are different. Not only does the *gnd* gene differ in all three strains, but so does *cib* (*cir* also differs between M496 and M498/M501).

There are no characteristic differences in any of the plasmid genes between the 1974 Scotland isolates and the Murray strains. For example, dendrograms of the strains constructed from the plasmid genes by means of the unweighted pair-group method using arithmetic averages (UPGMA), as implemented in the Clustal V program (Higgins et al. 1992), showed no grouping of the 1974 isolates apart from the Murray isolates (data not shown).

The UPGMA dendrograms also showed no tendency to group the plasmid gene sequences determined for IncI1-ColIb plasmids derived from *E. coli* apart from the plasmid-gene sequences determined for IncI1-ColIb plasmids derived from *S. typhimurium* (data not shown). This observation provides evidence that the IncI1-ColIb plasmid pools in *E. coli* and *S. typhimurium* are not distinct, genetically divergent plasmid subpopulations.

Extended Open Reading Frame in the abi Gene

Unexpectedly, all of the *abi* sequences determined in this study differed in a significant way from the published sequence, and also from a *Salmonella* IncI1-ColIb plasmid (Gupta and McCorquodale 1988). In particular, all sequences contain a single-nucleotide insertion relative to the published sequence, following nucleotide site 1354. The resulting frameshift generates a termination codon 33 amino acid residues from the

published putative initiation codon. Downstream sequence determined for *abi* in the Murray strains M140 and M496 revealed another single-nucleotide insertion—a guanosine following nucleotide site 1570—which extends the published open reading frame by 42 codons. The structure of the *abi* gene proposed by Gupta and McCorquodale (1988) was determined on the basis of analyses of open reading frames and the location of appropriate promoter signals within a 0.8-kb restriction fragment known from deletion-mapping experiments to overlap the *abi* locus in an IncI1-ColIb plasmid. Based on our data indicating an additional nucleotide following sites 1354 and 1570, we propose that the *abi* coding region extends from the methionine at site 1357 to the UAA stop codon at site 1698 (numbered as in Gupta and McCorquodale 1988). This open reading frame codes for a putative *abi* protein about 30% longer than that suggested by Gupta and McCorquodale (1988), yielding 114 amino acids vs 89. It is preceded by a Shine-Dalgarno sequence (GGAGG) five nucleotides upstream from the putative initiating methionine and overlaps, in frame, the proposed reading frame of Gupta and McCorquodale (1988) by 71 codons. Computer searches of both protein and nucleic acid databases performed at the National Center for Biotechnology Information using the BLAST network service failed to reveal any significant homologies with the proposed alternative coding sequence for the *abi* gene.

Discussion

An analysis of genetic variation among IncI1-ColIb plasmids was undertaken to elucidate factors affecting plasmid evolution as well as the evolutionary relationships between plasmids in strains collected in 1935–1941 vs 1974 and those between IncI1-ColIb plasmids in *Salmonella* and *E. coli*. To assess the levels of nucleotide polymorphism in plasmid strains, we determined the DNA sequences of three loci—*cib, cir* and *abi*—present in IncI1-ColIb plasmids in isolates of *S. typhimurium* from 1974 and from the Murray strains.

The level of genetic variation among the plasmid genes averaged $K = 0.021$ for the proportion of polymorphic nucleotides and $\pi = 0.0071$ for the pairwise number of nucleotide differences per site. These values are substantially smaller than those for chromosomal genes in *Salmonella:* estimates of π and K for the 6-phosphogluconate dehydrogenase gene are $\pi = 0.030$, $K = 0.080$ (present study); those for the glyceraldehyde-3-phosphate dehydrogenase gene from 16 isolates are $\pi = 0.037$ and $K = 0.128$ (Nelson et al. 1991); and those for the proline permease from 16 isolates are $\pi = 0.046$ and $K = 0.147$ (Nelson and Selander 1992). A reduced level of nucleotide polymorphism in plasmid genes, relative to chromosomal genes, would be expected from a high rate of plasmid turnover and horizontal transmissibility. This effect is observed dramatically in the remarkable homogeneity of insertion sequences in *E. coli* (Lawrence et al. 1992). On the other hand, the reduction may be smaller than first appears because the chromosomal genes that have been sequenced are among the most polymorphic. In general, isolates of *S. typhimurium* have limited genetic variation compared with *E. coli* (Beltran et al. 1991; Ochman and Selander 1984), as evidenced by the fact that 76% of isolates of *S. typhimurium* have an identical electrophoretic pattern for each of 20 polymorphic enzyme loci (Beltran et al. 1991).

The average level of nucleotide-sequence divergence between homologous chromosomal genes in *Salmonella* and *E. coli* is 15.6% (Bisercic and Ochman 1993); however, the genes that were sequenced from the IncI1-ColIb plasmids from *Salmonella* are not more similar to each other than they are to those from IncI1-ColIb plasmids from *E. coli* (Isaacson and Konisky 1974; Mankovich et al. 1986). This result would be expected if *Salmonella* and *E. coli* shared overlapping pools of IncI1-ColIb plasmids. It should also be noted that some chromosomal copies of the transposable element IS1 found in *Salmonella* are virtually identical to the IS1 found in *E. coli* (Bisercic and Ochman 1993), and it is thought that plasmids play a major role in the horizontal transmission of such elements among strains (Sawyer et al. 1987; Hartl and Sawyer 1988). Furthermore, no differences distinguishing the 1974 isolates from the Murray isolates were found in either the plasmid genes or the chromosomal genes. This result is consistent with the finding that many of the same plasmid incompatibility groups are present among plasmids of 1935–1941 as among those of 1974 (Jones and Stanley 1992; Hughes and Datta 1983). On the other hand, because there can be considerable sequence variation among genes in plasmids of the same incompatibility group (Finlay et al. 1986a–c), some differences in nucleotide sequence might have been expected. The lack of such differences implies that IncI1-ColIb plasmids have been widespread in distribution and stable for at least a number of decades. Although our data support intergeneric exchange of IncI1-ColIb plasmids, it should be noted that this is not necessarily true of all plasmids. For example, Mercer et al. (1984) found no evidence that common plasmids are transferred among *E. coli* lineages in nature, although such plasmids can readily be transferred in the laboratory.

The levels of nucleotide polymorphism in the IncI1-ColIb plasmids reported here are lower than might be expected from results of Riley and Gordon (1992). These authors used restriction enzyme digests to study polymorphisms among a set of either ColE1 plasmids and five ColIa plasmids isolated from a reference collection of *E. coli*. Based on the number of restriction-fragment differences pooled over four restriction endonucleases, the nucleotide diversity among the ColE1

plasmids were estimated as $\pi = 0.204 \pm 0.059$ (to be compared with an estimated $\pi = 0.0071$ for the IncI1-ColIb plasmids). The divergence among the ColIa plasmids was even greater: each had a unique pattern of restriction fragments, and no common set of fragments was shared among all ColIa plasmids. However, the data reported here and those of Riley and Gordon (1992) are not strictly comparable because, in our data, the plasmids were selected a priori for being IncI1-ColIb whereas, in their study, the plasmids were selected solely based on colicin type. Plasmids selected solely by colicin type may very well be more diverse than those selected by more stringent criteria. Alternatively, ColE1 and ColIa plasmids from *E. coli* may have more genetic variation at the DNA sequence level than ColIb plasmids from *S. typhimurium.*

The higher proportion of replacement polymorphisms, relative to synonymous polymorphisms, observed in the colicin and immunity genes as compared with *gnd* suggests either relaxed selective constraints on the amino acid sequences or diversifying selection for variation in the colicin and immunity genes. Because the same pattern was observed in three pairs of related colicin operons, Riley (1993b) suggested positive selection for genetic diversity. The excess of replacement substitutions in our data is consistent with the postulated process of diversification among the allelic operons of a single colicin type present in IncI1-ColIb plasmids.

Acknowledgments. The authors thank Howard Ochman and Mari Bisercic for sharing data prior to publication. This work was supported by grant number GM40322 from the National Institutes of Health.

References

Barker R, Old DC, Sharp JCM (1980) Phage type/biotype groups of *Salmonella typhimurium* in Scotland 1974–1976: variation during spread of epidemic clones. J Hygiene 84:115–126

Barker RM (1980) Colicinogeny in *Salmonella typhimurium.* J Gen Microbiol 120:21–26

Beltran P, Plock SA, Smith NH, Whittam TS, Old DC, Selander RK (1991) Reference collection of strains of the *Salmonella typhimurium* complex from natural populations. J Gen Microbiol 137:601–606

Bisercic M, Ochman H (1993) The ancestry of insertion sequences common to *Escherichia coli* and *Salmonella typhimurium.* J Bacteriology 175:7863–7868

Finlay BB, Frost LS, Paranchych W (1986a) Nucleotide sequences of five IncF plasmid *finP* alleles. J Bacteriol 167:754–757

Finlay BB, Frost LS, Paranchych W (1986b) Nucleotide sequences of the R1-19 plasmid transfer genes *traM, finP, traJ,* and *traY,* and the *traY*Z promoter. J Bacteriol 166:368–374

Finlay BB, Frost LS, Paranchych W (1986c) Origin of transfer of IncF plasmid and nucleotide sequences of the Type II *oriT, traM,* and *traY* alleles from ColB4-K98 and the Type IV *traY* allele from R100-1. J Bacteriol 168:132–139

Gupta SK, McCorquodale DJ (1988) Nucleotide sequence of a DNA fragment that contains the *Abi* gene of the ColIb plasmid. Plasmid 20:194–206

Hartl DL, Sawyer SA (1988) Why do unrelated insertion sequences occur together in the genome of *Escherichia coli?* Genetics 118:537–541

Higgins DG, Bleasy AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Hughes VM, Datta N (1983) Conjugative plasmids in bacteria of the "pre-antibiotic" era. Nature 302:725–726

Isaacson RE, Konisky J (1974) Studies on the regulation of colicin Ib synthesis. I. Isolation of the Col Ib-P9 plasmid. Mol Gen Genet 132:215–221

Jones C, Stanley J (1992) *Salmonella* plasmids of pre-antibiotic era. J Gen Microbiol 138:189–197

Konisky J (1982) Colicins and other bacteriocins with established modes of action. Ann Rev Microbiol 36:125–144

Lawrence JG, Ochman H, Hartl DL (1992) The evolution of insertion sequences within enteric bacteria. Genetics 131:9–20

Mankovich JA, Hsu C, Konisky J (1986) DNA and amino acid sequence analysis of structural and immunity genes of colicins Ia and Ib. J Bacteriol 168:228–236

Mercer AA, Morelli G, Heuzenroeder M, Kamke M, Achtman M (1984) Conservation of plasmids among *Escherichia coli* K1 isolates of diverse origins. Infect Immun 46:649–657

Nasoff MS, Baker HVI, Wolf REJ (1984) DNA sequence of the Escherichia coli gene, gnd, for 6-phosphogluconate dehydrogenase. Gene 27:253–264

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York City

Nelson K, Selander RK (1992) Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli.* J Bacteriol 172:6886–6895

Nelson K, Whittam TS, Selander RK (1991) Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli.* Proc Natl Acad Sci USA 88:6667–6671

Ochman H, Selander RK (1984) Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol 157:690–693

Pugsley AP (1985) *Escherichia coli* K12 strains for use in the identification and characterization of colicins. J Gen Microbiol 131:369–376

Reeves P, Stevenson G (1989) Cloning and nucleotide sequence of the *Salmonella typhimurium* LT2 *gnd* gene and its homology with the corresponding sequence of *Escherichia coli* K12. Mol Gen Genet 217:182–184

Riley MA (1993a) Molecular mechanisms of colicin evolution. Mol Biol Evol 10:1380–1395

Riley MA (1993b) Positive selection for colicin diversity in bacteria. Mol Biol Evol 10:1048–1059

Riley MA, Gordon DM (1992) A survey of col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of col-plasmid lineages. J Gen Microbiol 138:1345–1352

Rhoads RE, Rychlik W (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucleic Acids Res 17:8543–8551

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi RG, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sawyer S, Dykhuizen D, DuBose R, Green L, Mutangadura-Mhlanga T, Wolczyk D, Hartl D (1987) Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli.* Genetics 115:51–63